



université de bretagne
occidentale



THÈSE / UNIVERSITÉ DE BRETAGNE OCCIDENTALE

sous le sceau de l'Université européenne de Bretagne

pour obtenir le titre de

DOCTEUR DE L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE

Mention : Traitement du signal et des images

École Doctorale SICMA

présentée par

Isabelle LEONARD

Préparée dans l'équipe VISION

ISEN Brest

Reconnaissance des objets manufacturés dans des vidéos sous-marines

Thèse soutenue le 28 septembre 2012

devant le jury composé de :

Mohammad ALAM

Professeur, University of South Alabama / *rapporteur*

François GOUDAIL

Professeur, Institut d'optique Graduate School / *rapporteur*

Ayman ALFALOU

Professeur, ISEN Brest / *Directeur de thèse*

Badr-Eddine BENKELFAT

Professeur, Telecom & Management SudParis / *co-directeur de thèse*

Andreas ARNOLD BOS

Ingénieur d'études, Thales Underwater Systems / *co-encadrant de la thèse*

Christian BROSSEAU

Professeur, Université de Bretagne Occidentale / *examineur*

Jihad ZALLAT

Professeur, Université de Strasbourg / *examineur*

Jean Pierre KERNIN

Chef du groupe guerre des mines, DGA / *invité*

Pierre Guthmann

Directeur Technique, Thales Underwater Systems / *invité*



université de bretagne
occidentale



THÈSE / UNIVERSITÉ DE BRETAGNE OCCIDENTALE

sous le sceau de l'Université européenne de Bretagne

pour obtenir le titre de

DOCTEUR DE L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE

Mention : Traitement du signal et des images

École Doctorale SICMA

présentée par

Isabelle LEONARD

Préparée dans l'équipe VISION

ISEN Brest

Reconnaissance des objets manufacturés dans des vidéos sous-marines

Thèse soutenue le 28 septembre 2012

devant le jury composé de :

Mohammad ALAM

Professeur, University of South Alabama / *rapporteur*

François GOUDAIL

Professeur, Institut d'optique Graduate School / *rapporteur*

Ayman ALFALOU

Professeur, ISEN Brest / *Directeur de thèse*

Badr-Eddine BENKELFAT

Professeur, Telecom & Management SudParis / *co-directeur de thèse*

Andreas ARNOLD BOS

Ingénieur d'études, Thales Underwater Systems / *co-encadrant de la thèse*

Christian BROSSEAU

Professeur, Université de Bretagne Occidentale / *examineur*

Jihad ZALLAT

Professeur, Université de Strasbourg / *examineur*

Jean Pierre KERNIN

Chef du groupe guerre des mines, DGA / *invité*

Pierre Guthmann

Directeur Technique, Thales Underwater Systems / *invité*

Remerciements

“Un seul mot, usé, mais qui brille comme une pièce de monnaie : merci!” (Pablo Neruda). C’est ce mot que je voudrais adresser à plusieurs personnes.

Tout d’abord à Ayman AL FALOU pour avoir dirigé cette thèse, m’avoir conseillée et orientée tout au long de ces trois années. Merci aussi à l’équipe Vision du laboratoire de recherche et à tout le personnel de l’ISEN pour leur accueil, leur intérêt et leurs encouragements, plus particulièrement à Marwa pour toutes nos discussions, à Alain et Yann pour leurs remarques bienveillantes lors de nos pauses, à Michael et Laure pour leur compagnie lors du déjeuner ainsi qu’à tous ceux qui nous ont rejoint, à Dominique pour son aide en anglais !

Merci aussi à Andreas ARNOLD qui a encadré cette thèse pour m’avoir donné tant d’éclaircissements, d’idées et de pistes de réflexions. Je remercie également le service Études Générales Sonar de Brest pour leur accueil.

Je remercie aussi tous ceux qui ont pris du temps pour relire la totalité ou des parties de ce manuscrit : Ayman ALFALOU, Andreas ARNOLD BOS, Matthieu DUBREUIL, Alain LOUSSERT, Gilles KERYER et Nicolas MANDELERT.

Je tiens à remercier les membres du jury qui ont accepté d’examiner cette thèse, et particulièrement messieurs M.S. ALAM et F. GOUDAIL pour avoir accepté d’être les rapporteurs de ce manuscrit. Merci à B.-E. BENKELFAT pour avoir été co-directeur de cette thèse et présent dans le jury. Merci à messieurs J. ZALLAT et C. BROUSSEAU pour leur présence dans ce jury.

Cette thèse a été financée par Thales et la région Bretagne. Les images Topvision et les informations qui en ont été extraites sont des données de l’Etat français qui ont été fournies par le GESMA (Groupe d’Etudes Sous-Marines de l’Atlantique) dans le cadre du projet Topvision coordonné par Thales Underwater Systems SAS. (Pour des informations plus détaillées, on consultera le site <http://topvision.gesma.fr>)

Enfin, merci aux thésards, plus particulièrement à Mohammed et Yousri pour toutes nos discussions sur nos recherches, les règles de français et surtout sur les autres sujets (Je l’aime bien mon petit âne!!!). Merci aussi à Medjber, Nihel et Philippe.

Pour terminer, je dédie cette thèse à mes parents, Matthieu et Antoine. Merci pour votre soutien et vos encouragements (maintenant je suis une experte en démineur!).

Résumé

Reconnaissance des objets manufacturés dans des vidéos sous marines

Les mines sous marines sont très utilisées dans les conflits. Pour contrer cette menace, les marines s'équipent de moyens de lutte anti-mine autonomes afin d'éviter l'intervention humaine. Une mission de guerre des mines se découpe en quatre étapes distinctes : la détection des objets, la classification et l'identification puis la neutralisation. Cette thèse propose des solutions algorithmiques pour l'étape d'identification par caméra vidéo.

Le drone d'identification connaît la position approximative de l'objet à identifier. La première mission de ce drone est de re-détecter l'objet avant de le classifier et de l'identifier.

Le milieu sous-marin perturbe les images acquises par la caméra (absorption, diffusion). Pour faciliter la détection et la reconnaissance (classification et identification), nous avons prétraité les images.

Nous avons proposé deux méthodes de détection des objets. Tout d'abord nous modifions le spectre de l'image afin d'obtenir une image dans laquelle il est possible de détecter les contours des objets. Une seconde méthode a été développée à partir de la soustraction du fond, appris en début de séquence vidéo. Les résultats obtenus avec cette seconde méthode ont été comparés à une méthode existante.

Lorsqu'il y a une détection, nous cherchons à reconnaître l'objet. Pour cela, nous utilisons la corrélation. Les images de référence ont été obtenues à partir d'images de synthèse 3D des mines.

Pour les différentes méthodes utilisées, nous avons optimisés les résultats en utilisant les informations de navigation. En effet, selon les déplacements du drone, nous pouvons fixer des contraintes qui vont améliorer la détection et réduire le temps de calcul nécessaire à l'identification.

Mots clés : détection, reconnaissance, corrélation, détection de mines, images sous-marines, utilisation des informations de navigation

Abstract

Recognition of man-made objects in underwater videos

To avoid the underwater mine threat and to limit human interventions, navies use autonomous underwater vehicles. An underwater mine warfare mission is divided into four steps : object detection, classification, identification and neutralization. This PhD thesis proposes algorithmic solutions for the identification step done with a video camera.

Thanks to the detection step, the identification vehicle knows approximately the object position. First, the vehicle has to detect and position this object exactly. Then it will be classified and identified.

The underwater medium affects the images acquired with a video camera through absorption and scattering. The first step of our algorithm is to preprocess the images to help the detection and recognition (classification and identification) steps.

We have proposed two detection methods. The first one consists in modifying image spectrum in order to obtain an image in which we will be able to detect edges of objects. The second method, based on region segmentation, has been developed from background subtraction methods. The background image has been learned at the beginning of the video when there is no object. The results of the latter have been compared to those obtained with a state-of-art method, on data acquired at sea.

Once we have detected an object, we want to recognize it. For that, we use the correlation technique. The reference images have been obtained from 3D computer generated images of mines. This novel approach gives promising results.

For each developed method, we have optimized the results through the use of navigational information. Indeed, depending on vehicle's motion, we can set constraints to improve the detection step and reduce processing time.

Keywords : detection, recognition, correlation, underwater images, mine detection, use of navigational information

Table des matières

Remerciements	v
Résumés	vii
Table des matières	xv
Table des figures	xx
Liste des tableaux	xxi
Glossaire, acronymes, notations	xxiii
Introduction	1
I Contexte et état de l'art	5
1 Contexte	9
1.1 Enjeux de la guerre des mines	9
1.1.1 Les enjeux géopolitiques et militaires	9
1.1.2 Les types de mines	10
1.1.3 Notre position par rapport aux enjeux	10
1.2 Le milieu sous marin	11
1.2.1 La vidéo	11
1.2.2 Le sonar	13
1.2.3 Sonar ou vidéo ?	14
1.2.4 Les capteurs	16
1.3 Les étapes d'une opération de guerre des mines actuelle	18
1.3.1 La détection	18
1.3.2 La classification	18
1.3.3 L'identification	19
1.3.4 Le désarmement	19
1.4 Les moyens de guerre des mines	19
1.4.1 Le scénario actuel	19
1.4.2 Le scénario futur	20
1.5 Solution proposée	21
1.5.1 Les systèmes existants et futurs	22
1.5.2 Les objectifs de cette thèse	23
1.5.3 Les détails de l'étape d'identification	24
1.5.4 L'utilisation des capteurs	25
1.6 Conclusion	25

2	Les prétraitements	27
2.1	État de l'art	27
2.1.1	Les prétraitements basés sur la physique	28
2.1.2	Les prétraitements empiriques	30
2.2	Méthodes utilisées	31
2.2.1	Le bruit électronique	32
2.2.2	Le temps de calcul	34
2.2.3	Filtrage du bruit	35
2.2.4	Utilisation du modèle de Phong	35
2.2.5	La détection des contours	38
2.2.6	Les prétraitements annexes	41
2.3	Résultats et conclusion	43
2.3.1	Les images Topvision	43
2.3.2	Les images VAMA	44
3	Choix de l'approche	45
3.1	Méthode basée sur la couleur	46
3.2	Méthode basée sur la transformée de Hough	46
3.3	Méthodes basées sur les contours	47
3.3.1	Utilisation des détecteurs de contours	47
3.3.2	Utilisation des contours actifs	48
3.4	Méthodes basées sur l'attention visuelle	49
3.5	Méthodes basées sur la séparation premier plan arrière plan	49
3.6	Le suivi d'objet	50
3.7	Le positionnement et la cartographie	50
3.8	Utilisation d'images polarisées	50
3.9	Méthodes du milieu terrestre	51
3.9.1	Les méthodes utilisant une base de données	51
3.9.2	Les méthodes à base d'histogramme	51
3.9.3	Les réseaux de neurones	51
3.9.4	Les classificateurs	52
3.9.5	La corrélation	52
3.10	Conclusion	52
II	La corrélation	53
4	Généralités sur la corrélation	57
4.1	La corrélation	57
4.1.1	Définition	57
4.1.2	Implémentation	58
4.1.3	Approche	59
4.2	Quantifier la reconnaissance	59
4.2.1	La robustesse au bruit	59
4.2.2	L'efficacité optique	60
4.2.3	La finesse du pic de corrélation	60
4.2.4	Les courbes ROC	60
5	La corrélation selon l'architecture VanderLugt	63
5.1	État de l'art	64
5.1.1	Le filtre adapté	64
5.1.2	Le filtre de phase pure	64
5.1.3	Le filtre de phase pure binaire	65
5.1.4	Le filtre inverse	66

5.1.5	Le filtre à compromis optimal	66
5.1.6	Le filtre composite	66
5.1.7	Le filtre segmenté	66
5.1.8	Comparaison des différents filtres	68
5.2	Comparaison de l'architecture VanderLugt avec la ZNCC	77
5.3	Le filtre ASPOF et le critère PCEA	78
5.3.1	Le principe du filtre	78
5.3.2	Adaptation du critère PCE	79
5.3.3	Comparaison du filtre ASPOF et des filtres composites et segmentés	80
5.4	Conclusion	83
6	La corrélation selon l'architecture JTC	85
6.1	État de l'art	86
6.1.1	Le JTC sans ordre zéro	86
6.1.2	Le JTC binaire	86
6.1.3	Le JTC non-linéaire	87
6.1.4	Le JTC fringe-adjusted	87
6.1.5	Comparaison des différentes architectures	87
6.2	Utilisation du JTC pour faire du suivi	91
6.3	Conclusion	97
III	La détection et la reconnaissance des mines sous marines	99
7	Application de la corrélation aux mines	103
7.1	Application de la corrélation aux mines	103
7.1.1	Détection et identification des mines sous marines avec l'architecture VanderLugt	103
7.1.2	Détection et identification des mines sous marines avec l'architecture JTC	105
7.2	Problématique	107
8	Détection	111
8.1	Choix de la représentation des résultats	112
8.2	Méthode basée sur la phase	115
8.2.1	Algorithme utilisé	115
8.2.2	Résultats obtenus	120
8.3	Méthode basée sur la soustraction du fond	120
8.3.1	Algorithme utilisé	121
8.3.2	Résultats obtenus et améliorations apportées	122
8.3.3	Application de l'algorithme aux vidéos Topvision	124
8.4	Évaluation de notre algorithme	124
8.5	Perspectives	126
8.5.1	Optimisations à apporter à nos algorithmes	126
8.5.2	Des méthodes pour formaliser nos algorithmes	127
8.6	Conclusion	129
9	Classification et identification	131
9.1	Premiers tests	132
9.2	Les filtres de corrélation	132
9.2.1	Base d'images de référence	133
9.2.2	Fusion des images	134
9.3	Corrélation avec tous les filtres	138
9.3.1	Algorithme utilisé	138
9.3.2	Résultats obtenus	139

9.3.3	Optimisations apportées à notre algorithme	140
9.3.4	Distinction entre les objets manufacturés dangereux et les autres objets	141
9.3.5	Conclusion sur cette première méthode	141
9.4	Utilisation des données de navigation	141
9.4.1	Algorithme utilisé	141
9.4.2	Résultats obtenus	143
9.5	Conclusion	144
Conclusion		147
Conclusion		147
Perspectives		149
Annexes		161
A Comparaison entre les filtres		161
A.1	Le principe du filtre AMPOF	161
A.2	Le filtre MACE	161
A.2.1	Principe du filtre	161
A.2.2	Résultats obtenus	161
A.3	Robustesse au bruit du filtre ASPOF	163
B Création des fichiers de navigation pour les vidéos Topvision		167
C Détails des résultats de la détection		171
C.1	Résultats sur la mine sphérique	171
C.1.1	Méthode basée sur la phase	171
C.1.2	Méthode basée sur la soustraction du fond	172
C.1.3	Algorithme de Cybernetix	173
C.2	Résultats sur la mine Manta	173
C.2.1	Méthode basée sur la phase	173
C.2.2	Méthode basée sur la soustraction du fond	174
C.2.3	Algorithme de Cybernetix	175
C.3	Résultats sur la mine cylindrique	175
C.3.1	Méthode basée sur la phase	175
C.3.2	Méthode basée sur la soustraction du fond	176
C.3.3	Algorithme de Cybernetix	177
C.4	Résultats sur les autres objets	178
C.4.1	Méthode basée sur la phase	178
C.4.2	Méthode basée sur la soustraction du fond	178
C.4.3	Algorithme de Cybernetix	179
C.5	Résultats sur les vidéos vides	180
C.5.1	Méthode basée sur la phase	180
C.5.2	Méthode basée sur la soustraction du fond	180
C.5.3	Algorithme de Cybernetix	181
D Intersection d'un rayon et d'un triangle		183
E Caractéristiques de PCs et logiciels Matlab utilisés		185
E.1	PC ISEN	185
E.2	PC Thales	185

F Mes publications

187

Bibliographie

263

Table des figures

1.1	Les différentes mines étudiées	11
1.2	Diffusion de la lumière incidente par une particule selon la théorie de Mie	12
1.3	Profondeurs à partir desquelles une longueur d'onde est complètement absorbée par le milieu sous marin	12
1.4	Schéma représentant l'image reçue et les trois composantes lumineuses qui la composent	13
1.5	Géométrie de l'image sonar	14
1.6	Comparaison des valeurs des coefficients d'absorption pour une onde acoustique et pour une onde optique	15
1.7	Les étapes d'une mission de guerre des mines	18
1.8	Chasseur de mine tripartite	19
1.9	PVDS de Thales Underwater Systems	19
1.10	Scénario actuel de guerre des mines	20
1.11	Chasseur de mines de classe Avenger	20
1.12	Tableau résumant les étapes, les moyens et les capteurs du scénario actuel de guerre des mines	21
1.13	Scénario du futur	21
1.14	USV	22
1.15	AUV Muscle de Thales Underwater Systems et Bluefin Robotics	22
1.16	Tableau résumant les étapes, les moyens et les capteurs du scénario futur de guerre des mines	22
1.17	Détails de l'étape d'identification	24
1.18	Schéma représentant la mission d'un drone d'identification - neutralisation	24
2.1	Algorithme de rehaussement des contours	32
2.2	Effet sur l'image de phase du rehaussement de contours	33
2.3	Problèmes identifiés dans nos images	33
2.4	Image initiale	33
2.5	Image dont l'effet de moiré a été supprimé avec la méthode proche de celle de Sidorov et Kokaram	34
2.6	Image initiale (figure 2.4) sous échantillonnée	34
2.7	Effet du filtrage médian sur une image	35
2.8	Définition des angles utilisés	36
2.9	Disque de Secchi vu de dessus	36
2.10	Images avant et après application du modèle de Phong	39
2.11	Images avant et après application du modèle de Phong	40
2.12	Schéma expliquant la création d'une image de phase	40
2.13	Lena et l'image de phase de Lena	40
2.14	Combinaison des phases et des amplitudes des spectres	41
2.15	Image de phase calculée à partir de l'image prétraitée issue d'une vidéo Topvision	41
2.16	Image de phase calculée à partir de l'image prétraitée issue d'une vidéo VAMA	41
2.17	Schéma expliquant l'obtention d'une image de contours via l'approche multi-échelle	42
2.18	Image issue de la vidéo	42

2.19	Image après découpe des bords et des écritures	42
2.20	Image prétraitée	43
2.21	Image de contours de l'image initiale	43
2.22	Image de contours sur laquelle le manchon a été supprimé	43
2.23	Algorithme des prétraitements utilisés sur les images Topvision	44
2.24	Algorithme des prétraitements utilisés sur les images VAMA	44
4.1	Principe de la corrélation	57
4.2	Implémentation optique de la corrélation	58
4.3	Valeurs calculées pour réaliser une courbe ROC	61
4.4	Représentation de l'espace ROC	61
5.1	Schéma synoptique du corrélateur proposé par Vander-Lugt	63
5.2	Autocorrélation de Lena avec un filtre adapté	64
5.3	Autocorrélation de Lena avec un filtre POF	65
5.4	Multiplexage de références	67
5.5	Comparaison d'un filtre limité en taille et d'un filtre non limité en taille	67
5.6	Création d'un filtre segmenté	67
5.7	Création d'un filtre segmenté à partir de 3 références	68
5.8	Résultats obtenus avec les différents filtres composites adaptés testés dans le cas de la corrélation de lettres binaires non bruitées	69
5.9	Comparaison des filtres POF composite et segmentés contenant 10 références	73
5.10	Répartition de la base d'images de référence en deux sous bases d'images pour créer les filtres	78
5.11	Assemblage des deux filtres segmentés	78
5.12	Exemples de cas de gestion des pixels isolés	79
5.13	Plan résultant de la corrélation entre la lettre V et un filtre ASPOF pour la lettre A	80
5.14	Comparaison des filtres sur la lettre A	80
5.15	Test de la robustesse du filtre	81
5.16	Courbes ROC	81
5.17	Visages de la base PHPID	82
5.18	Comparaison des filtres sur les visages	82
5.19	Test de la robustesse du filtre	83
5.20	Courbes ROC	83
6.1	Montage du corrélateur JTC	86
6.2	Architecture du JTC binaire	87
6.3	Utilisation du JTC fringe-adjusted pour faire du suivi de personne	91
6.4	Utilisation du JTC fringe-adjusted pour faire du suivi de personne, mise à jour du fond dans le filtre	92
6.5	Utilisation du JTC fringe-adjusted pour faire du suivi de personne avec le critère enhanced PCE	93
6.6	Utilisation du JTC fringe-adjusted non linéaire pour faire du suivi de personne avec le critère enhanced PCE	93
6.7	Utilisation du JTC fringe-adjusted non linéaire pour faire du suivi de personne avec le nouveau critère	94
6.8	Schéma de l'architecture proposée	95
6.9	Plans de corrélation correspondant aux différentes étapes du calcul de l'écart	95
6.10	Exemple de calcul du critère en cas de reconnaissance	96
6.11	Exemple de calcul du critère en cas de non reconnaissance	96
7.1	Mine Manta	104
7.2	Image de phase calculée à partir de l'image prétraitée issue d'une vidéo Topvision	104
7.3	Résultats de la corrélation entre une vidéo Topvision et des filtres de secteur créés à partir de cette vidéo	105

7.4	Résultats obtenus avec le JTC fringe-adjusted non-linéaire sans ordre zéro avec notre critère et post-traitement des résultats	106
7.5	Test de la robustesse du JTC fringe-adjusted non-linéaire sans ordre zéro avec notre critère et post-traitement des résultats	107
7.6	Diagramme de l'algorithme utilisé	109
8.1	Résultats de la détection sur une vidéo de mine Manta	112
8.2	Définition des différentes probabilités utilisées	112
8.3	Exemple d'une bonne détection	113
8.4	Exemples de fausses alarmes	113
8.5	Exemple d'une non détection fausse	113
8.6	Exemple d'une non détection vraie	114
8.7	Schéma expliquant le calcul de la distance au sol pour les images Topvision	114
8.8	Résultats de la détection sur une vidéo de mine Manta	115
8.9	Image utilisée pour la détection d'une mine Manta	116
8.10	Image utilisée pour la détection d'une mine cylindrique	116
8.11	Valeur médiane du fond	117
8.12	Image illustrant les bandes verticales sur lesquelles sont basées la méthode	117
8.13	Valeurs des moyennes auxquelles a été soustraite la médiane du fond (en bleu). En rouge : seuil	118
8.14	Résultats de l'algorithme de détection basé sur la phase sur les images Topvision	118
8.15	Résultats après optimisation de la détection de la mine Manta	119
8.16	Résultats après optimisation de la détection de la mine cylindrique	119
8.17	Résultat de l'algorithme de détection basé sur la soustraction du fond pour une image de mine sphérique	122
8.18	Résultat de l'algorithme de détection basé sur la soustraction du fond pour une image de mine Manta	122
8.19	Etats d'une chaîne de Markov	128
8.20	Etats et observation d'une chaîne de Markov cachée	128
9.1	Schéma indiquant la partie traitée dans ce chapitre	131
9.2	Images des modèles de synthèse des mines	132
9.3	Images des modèles de synthèse des mines	133
9.4	Images de contours d'une mine cylindrique	135
9.5	Exemple de filtre non satisfaisant	135
9.6	Représentation d'un octaèdre après 2 étapes de récurrence, représentant l'espace	137
9.7	Représentation d'un octaèdre et de ses sommets projetés sur une sphère représentant l'espace	137
9.8	Schéma synoptique de l'algorithme utilisé avec la première méthode	138
9.9	Méthode de la détection de contours à partir de la segmentation basée sur les contours actifs	139
9.10	Image résultat de la détection de contours basée sur une approche multi-échelle. Vidéo Tvision008 représentant une mine sphérique	142
9.11	Cercle calculé à partir de l'image de contours	142
9.12	Schéma synoptique de la méthode de reconnaissance utilisant les informations de navigation	143
9.13	Résultats de la reconnaissance	144
9.14	Image issue d'une vidéo réelle	149
9.15	Étude de la position du polariseur	151
9.16	Étude de la forme du dôme	151
9.17	Image d'intensité	152
9.18	Image en degré de polarisation	153
9.19	Image d'intensité	154
9.20	Image en degré de polarisation	155

9.21	Plans de corrélation obtenus	156
9.22	Performances des cibles programmables	157
A.1	Reconnaissance de la lettre A par un filtre MACE contenant 10 références	162
A.2	Reconnaissance de la lettre A bruitée par un bruit blanc par un filtre MACE contenant 10 références	163
A.3	Reconnaissance de la lettre A bruitée par une texture par un filtre MACE contenant 10 références	164
A.4	Comparaison des filtres sur les lettres A perturbées par un bruit blanc additif	164
A.5	Comparaison des filtres sur les lettres A perturbées par un bruit texturé	165
B.1	Image Topvision	167
B.2	Images références	167
B.3	Les différentes étapes pour extraire les informations souhaitées de l'image	168
B.4	Plans de corrélation pour les chiffres 0, 5 et 8	168
C.1	Résultats obtenus avec la méthode basée sur la phase, développée pour les images Topvision	171
C.2	Résultats bruts de la détection pour la mine sphérique	172
C.3	Résultats optimisés de la détection pour la mine sphérique	173
C.4	Résultats Cybernetix sur les vidéos de sphère VAMA	173
C.5	Résultats obtenus avec l'algorithme basé sur la phase	174
C.6	Résultats bruts de la détection pour la mine Manta	174
C.7	Résultats obtenus avec l'algorithme optimisé	175
C.8	Résultats Cybernetix sur les vidéos de Manta VAMA	175
C.9	Résultats obtenus avec l'algorithme basé sur la phase	176
C.10	Résultats bruts de la détection pour la mine cylindrique	176
C.11	Résultats obtenus avec l'algorithme optimisé	177
C.12	Résultats Cybernetix sur les vidéos de cylindre VAMA	177
C.13	Résultats obtenus avec l'algorithme basé sur la phase	178
C.14	Résultats bruts de la détection pour les autres objets	178
C.15	Résultats obtenus avec l'algorithme optimisé	179
C.16	Résultats Cybernetix sur les vidéos d'objets VAMA	179
C.17	Résultats obtenus avec l'algorithme basé sur la phase	180
C.18	Résultats bruts de la détection pour les vidéos vides	180
C.19	Résultats optimisés de la détection pour les vidéos vides	181
C.20	Résultats Cybernetix sur les vidéos vides VAMA	181

Liste des tableaux

1.1	Tableau résumant les avantages et inconvénients du sonar et de la vidéo	15
5.1	Comparaison des filtres composites	70
5.2	Comparaison des filtres segmentés	74
5.3	Tableau comparant le filtre POF et la ZNCC	77
6.1	Comparaison des différentes architectures JTC	88
6.2	Comparaison du JTC binaire et du JTC fringe-adjusted	90
7.1	Résultats obtenus sur les vidéos de mine Manta et cylindrique	105
7.2	Résultats obtenus avec les différentes architectures JTC sur la détection de la mine Manta	106
7.3	Test de la robustesse des différentes architectures JTC sur la détection de la mine Manta	107
8.1	Tableau des résultats de détection obtenus sur les images Topvision	120
8.2	Tableau des résultats de détection obtenus sur les images Vama	120
8.3	Tableau des résultats obtenus avec la méthode basée sur la soustraction du fond .	122
8.4	Tableau des résultats obtenus avec la méthode basée sur la soustraction du fond, optimisée	123
8.5	Tableau des résultats obtenus sur les images Topvision avec l’algorithme initial .	124
8.6	Tableau des résultats obtenus sur les images Topvision avec l’algorithme optimisé	124
8.7	Tableau comparant les résultats obtenus avec les méthodes développées, sur les vidéos VAMA	125
9.1	Nombre de filtres obtenus avec la méthode d’assemblage basée sur la corrélation .	136
9.2	Tableau résumant le nombre d’images et de filtres obtenus	138
9.3	Résultats obtenus avec la détection de contours basée sur les contours actifs . . .	139
9.4	Résultats obtenus avec la détection de contours basée sur l’approche multi-échelle	140
9.5	Résultats obtenus avec la détection de contours basée sur les contours actifs . . .	140
9.6	Résultats obtenus avec la détection de contours basée sur l’approche multi-échelle	141
9.7	Résultats obtenus sur les vidéos Topvision	143
9.8	Résultats de la reconnaissance sur les vidéos de mines sphériques VAMA	144
9.9	Résultats de la détection et de la reconnaissance sur les vidéos vides VAMA . . .	144
B.1	Table regroupant les 10 meilleurs pics de corrélation pour chaque chiffre	168
B.2	Table classant les pics de corrélation selon leur position	169
B.3	Table regroupant les pics de corrélation d’intensité maximale	169

Glossaire, acronymes, notations

Glossaire

Détection vraie	un objet a été détecté et il s'agit bien d'un objet à cet endroit de l'image
Fausse alarme	un objet a été détecté et il n'y a pas d'objet à cet endroit de l'image
Géocentrique	position du Soleil par rapport au centre de la Terre
Héliocentrique	position de la Terre par rapport au centre du Soleil
Non détection fausse	le système ne détecte pas d'objet
Non détection vraie	le système ne détecte rien et il n'y a rien à détecter
Topocentrique	position du Soleil par rapport à la position de l'observateur sur la Terre

Acronymes

ACE	Automatic color equalization	
ALMDS	Airborne Laser Mine Detection System	
AMPOF	Amplitude Modulated Phase Only Filter	
ASPOF	Asymmetric Segmented Phase-Only Filter	Filtre de phase pure segmenté asymétrique
AUV	Autonomous Underwater Vehicle	Véhicule sous-marin autonome
BJTC	Binary JTC	JTC binaire
BPOF	Binary Phase Only Filter	Filtre de phase pure binaire
CCD	Charge-Coupled Device	Dispositif à transfert de charge
CJTC	Classical JTC	JTC classique
CMT	Chasseur de Mine Tripartite	
EGS	General Sonar Studies (GSS)	Études Générales Sonar
FAF	Fringe-adjusted filter	
FA JTC	Fringe-adjusted JTC	
FN	False Negative	Faux négatif
FP	False Positive	Faux positif
GESMA	Groupe d'études sous marines de l'Atlantique	
GPD	Groupe des plongeurs-démineurs	
GPS	Global Positioning System	
HSI	Hue, Saturation, Intensity	
ISEN	Institut Supérieur d'Électronique et du Numérique	
JTC	Joint Transform Correlator	Corrélateur à transformée de Fourier conjointe
LED	Light-Emitting Diode	

LIDAR	LIght Detection And Ranging	
MACE	Minimum Average Correlation Energy filter	
MILCO	MIne Like COntact	
MILEC	MIne Like ECho	
NL	Nonlinear	Non linéaire
NL JTC	Nonlinear JTC	JTC non linéaire
NOMBO	NOOn Mine Bottom Object	
NZ	Non Zero	Sans ordre zéro
NZ JTC	Non Zero JTC	JTC sans ordre zéro
OT	Optimal Trade-off filter	Filtre à compromis optimal
OTF	optical transfert function	fonction de transfert optique
PAP		Poisson Auto-Propulsé
PCE	Peak to Correlation Energy	
PCEA	Adapted Peak to Correlation Energy	
Pd		Probabilité de détection
Pdc		Probabilité de bonne classification
Pfa		Probabilité de fausse alarme
Pfac		Probabilité de fausse alarme après reconnaissance
PHPID	Pointing Head Pose Image Database	
Pndf		Probabilité de non détection fausse
Pndfc		Probabilité de non détection fausse après reconnaissance
Pndv		Probabilité de non détection vraie
Pndvc		Probabilité de non détection vraie après reconnaissance
POF	Phase Only Filter	Filtre de phase pure
PSF	point spread function	fonction d'étalement du point
Pype		Probabilité de mauvaise classification d'une mine
PVDS	Propelled Variable Depth Sonar	Sonar propulsé à immersion variable
RGB	Red, Green, Blue	
ROC	Receiver Operating Characteristic	
ROV	Remotely Operated Vehicle	
SAS	Synthetic Aperture Sonar	Sonar à ouverture synthétique
SLAM	Simultaneous Localization And Mapping	
SLAMF		Système de Lutte Anti-Mines du Futur
SLM	Spatial Light Modulator	Modulateur spatial de lumière
SNR	Signal to Noise Ratio	Rapport signal à bruit
SSS	Side Scan Sonar	
STIL	Streak Tube Imaging Lidar	
TF	Fourier Transform (aussi notée FT)	Transformée de Fourier
TN	True Negative	Vrai négatif
Topvision		Tests OPérationnels de Vidéos Sous-marines pour l'Identification d'Objets Nuisibles
TP	True Positive	Vrai positif
TUS	Thales Underwater Systems	
USV	Unmanned Surface Vehicle	Véhicule de surface autonome

VAMA		Véhicule anti-mine autonome
VL	Vander Lugt correlator	corrélateur de Vander Lugt
ZNCC	Zero mean Normalized Cross Correlation	Corrélation normalisée

Notations

Lettres majuscules

A	vecteur de Stokes	
A	fonction qui permet de supprimer les pôles du filtre du FA JTC	
B	transformée de Fourier du bruit	
B	constante ou fonction permettant d'obtenir un gain supérieur à 1 pour le FA JTC	
C	plan de corrélation	
DI	index de directivité du récepteur, dans l'équation du sonar	
DT	seuil de détection dans l'équation du sonar actif	
E	couleur de l'eau	
F	spectre joint du plan d'entrée dans l'architecture JTC	
H	filtre de corrélation (dans le domaine de Fourier)	
H	filtre du FA JTC	
I	image ou intensité	[cd]
I_d	image obtenue après l'opération de division	
I_{LP}	image I filtrée par un filtre passe-bas gaussien	
I_s	image obtenue après l'opération de soustraction	
\vec{L}	vecteur correspondant à la direction source lumineuse - surface	
$L \frac{s}{N}$	rapport signal à bruit de l'équation du sonar	[dB]
M	matrice de Mueller	
M	couleur de la mine	
N, M	taille du plan de corrélation	
N	axe normal à une surface	
NL	niveau de bruit, dans l'équation du sonar	
O	transformée de Fourier de l'objet	
\vec{O}	vecteur correspondant à la direction surface-observateur	
Q, U et V	intensité dans les modes de polarisation horizontal, vertical, à +45°, à -45°, et circulaires gauche et droite	
\vec{R}	vecteur représentant le rayon lumineux réfléchi	
S	transformée de Fourier du signal	
S	rapport des gradients utilisé pour le rehaussement de contours	
SL	intensité de la source, dans l'équation du sonar	
TL	pertes de transmission, dans l'équation du sonar	
TS	réflectivité de la cible, dans l'équation du sonar actif	

Lettres minuscules

a	coefficient d'absorption dans la loi de Beer-Lambert	[m ⁻¹]
b	bruit	
c	coefficient d'atténuation dans la loi de Beer-Lambert	[m ⁻¹]
d	coefficient de diffusion dans la loi de Beer-Lambert	[m ⁻¹]
f	fréquence	[Hz]

h	filtre de corrélation, ou fonction d'illumination locale	
i	image cible	
i,j	indices des pixels d'une image	
k	degré de non linéarité du NL JTC	
l	longueur de l'antenne considérée	
o	objet à reconnaître	
r	image référence	
s	signal ou scène à analyser	
x_0, y_0	position du pic de corrélation	
x, y	position dans un plan	
z	distance entre l'observateur et l'objet	[m]

Lettres grecques

α	brillance du matériau	
α	facteur de pondération du NL JTC	
Γ	densité spectrale du bruit	
η	efficacité optique	
λ	longueur d'onde	[m]
μ, ν	indices des pixels dans le domaine de Fourier	
θ	désigne la phase d'un signal dans le domaine de Fourier	
θ_s	angle formé entre la normale à la surface observée N et le soleil	[rad]
ρ	désigne l'amplitude d'un signal dans le domaine de Fourier	
σ	écart-type	
Ω	angle formé entre la réflexion de la source de lumière et l'observateur	[rad]
ω	pulsation	[rad/s]

Les symboles

*	opération de convolution
$*^c$	opération de corrélation
\vec{N}	vecteur
\bar{h} ou h^*	complexe conjugué

Introduction

Pendant la première guerre mondiale, les Alliés ont voulu assiéger Istanbul afin de contrôler le Bosphore. Les navires alliés se sont alors engagés dans le détroit des Dardanelles. Le détroit étant miné, des opérations de déminage ont été nécessaires. Au cours de la bataille, au vu de la riposte turque, l'ordre de battre en retraite a été donné. Peu de temps après le cuirassé Bouvet heurta une mine non détectée. La voie d'eau provoquée par l'explosion fit rapidement couler le bateau. Près de 650 marins ont trouvé la mort dans cette tragédie. L'Irresistible de la Royal Navy connut le même sort.

Ce fait historique montre que les mines sous marines représentent une menace redoutable, malheureusement toujours d'actualité. Elles permettent de protéger les littoraux et d'empêcher le passage de bateaux ennemis. Cependant, les mines non détectées restent une menace en temps de paix. Les dégâts infligés peuvent être considérables. La menace est donc importante à prendre en compte. Les marines s'équipent donc de moyens de lutte anti-mine sous marine. Actuellement, les systèmes proposés sont principalement basés sur la technologie sonar. La décision sur la dangerosité de l'objet est prise par un opérateur. Lorsqu'un drone effectue la détection d'objets, la communication entre ce drone et l'opérateur peut être une liaison acoustique (avec débit de transmission de données faible) ou filaire (le périmètre exploré est alors limité par cette liaison). La destruction des mines nécessite souvent l'intervention de plongeurs-démineurs. Pour limiter les problèmes de communication et l'intervention d'opérateurs humains, des drones autonomes sont développés pour répondre aux problématiques de la détection, de la reconnaissance et de la destruction de ces mines.

Pour améliorer les taux de reconnaissance des drones, une possibilité est d'allier les performances à grande distance des antennes sonars avec les performances à courte distance des caméras vidéo. Cette thèse vise à fournir des solutions algorithmiques aux problèmes de détection, de classification et d'identification des mines par vidéo et traitement optique. Le temps de calcul de l'algorithme doit être proche de la cadence vidéo dans l'optique d'effectuer par la suite un guidage du drone à l'aide des résultats obtenus. Mes travaux s'inscrivent dans la continuité d'un projet de TUS, "*Thales Underwater Systems*", le projet VAMA (Véhicule Anti-Mine Autonome) que nous présenterons. Ce projet utilise une méthode proche de la corrélation pour reconnaître les mines¹. Contrairement aux méthodes utilisées dans VAMA, nous développerons une méthode basée sur la corrélation optique : la corrélation a été étudiée particulièrement dans le domaine optique. Il est intéressant d'utiliser ses développements pour les appliquer à notre problème. De plus, la corrélation est une méthode qui permet d'obtenir une décision fiable car elle est robuste et discriminante. Cette thèse a donc pour objectif d'établir un lien entre d'une part le monde optique et sa corrélation et d'autre part le monde numérique des roboticiens. Pour atteindre cet objectif, cette thèse a profité d'une collaboration entre l'ISEN Brest (Institut Supérieur d'Électronique et du Numérique) et TUS. L'équipe Vision du laboratoire de recherche L@bISEN est spécialisée dans la reconnaissance par corrélation optique. Le service EGS (Études Générales Sonar) de TUS propose des solutions aux problématiques de détection et de reconnaissance des mines. Cette thèse permet d'utiliser une méthode dans un domaine spécifique où elle n'est pas encore utilisée aussi intensivement que dans les applications terrestres.

1. Pour plus de renseignements, merci de contacter TUS

Contributions

Au cours de cette thèse, nous avons apporté des améliorations aux architectures existantes de la corrélation (architecture de Vanderlugt et architecture JTC, “*Joint Transform Correlator*”). Nous avons aussi proposé de nouvelles perspectives pour l’optimisation des résultats de détection et de reconnaissance.

Les deux architectures de corrélation que nous venons de citer ont été étudiées. L’architecture JTC consiste à étudier une image de référence et une image à identifier conjointement. Différentes architectures ont été développées. L’architecture de Vanderlugt consiste à convoluer une image avec un filtre de corrélation. Il existe dans la littérature plusieurs filtres de corrélation.

Tout d’abord lorsque nous avons étudié les architectures de corrélation, nous avons constaté que le JTC ne propose pas d’architecture à la fois robuste et discriminante. Nous avons donc créé notre propre architecture basée sur l’introduction, dans le plan de Fourier, d’un filtre spécial et adapté à la non linéarité. Il a été conjugué à la création d’un nouveau critère afin d’obtenir un ensemble avec lequel la prise de décision automatique est facilitée.

Ensuite, lorsque nous avons adapté la corrélation selon l’architecture de Vanderlugt à la reconnaissance des mines, nous nous sommes confrontés au nombre de filtres à appliquer à une image pour reconnaître l’objet présent. En effet, lors de l’utilisation de drones autonomes, le résultat de l’étape de corrélation va permettre le guidage du drone. Par conséquent, le temps de calcul doit être très court (proche de la cadence vidéo) donc le nombre de filtres de corrélation faible. Nos réflexions ont permis la création d’un nouveau filtre qui permet de diviser la base de filtres par un facteur deux. La prise de décision a été facilitée grâce à l’adaptation d’un critère de décision. Ce filtre a été comparé aux principaux filtres de l’état de l’art et a permis d’obtenir de meilleurs résultats.

Lorsque l’on souhaite identifier un objet dans une image par corrélation, il faut avoir à disposition toutes les vues possibles de cet objet. Nous avons été confrontés au problème des vues à utiliser pour créer les filtres de corrélation. En effet, toutes les vues de mines ne sont pas présentes dans les vidéos. Pour résoudre ce problème, nous avons décidé de tester la création de filtres à partir d’images de synthèse des mines.

Les caractéristiques du milieu sous-marin (absorption, turbidité) perturbent les images optiques acquises. Pour améliorer les résultats de notre algorithme, nous avons introduit une étape de prétraitements qui permet d’augmenter la qualité des images. Nous verrons par la suite que nous avons aussi ajouté une étape de détection des objets entre les prétraitements de l’image et la reconnaissance par la corrélation. Pour cette étape, nous avons créé deux méthodes basées sur des techniques connues (la phase du spectre d’une image et la soustraction du fond). L’étude de la phase du spectre n’a pas été utilisée dans le milieu sous-marin. Quant à la soustraction du fond, nous avons adapté la méthode de façon à faire ressortir les objets manufacturés tout en tenant compte du fond qui varie au cours de la vidéo en rapport avec le déplacement du drone.

Enfin, lors de l’application de notre algorithme aux mines, nous avons été confrontés au problème de la continuité des résultats au cours du temps. Par exemple, une image peut être analysée comme contenant un objet et l’image suivante comme ne contenant pas d’objet. Pour limiter ces problèmes, nous avons décidé d’utiliser les informations de navigation comme un a priori. Cette approche est nouvelle dans le traitement d’images sous marines par corrélation et nous permet d’améliorer significativement nos résultats.

Certaines parties de cette thèse ont fait l’objet de publications. Le filtre de corrélation pour l’architecture de Vanderlugt et la comparaison des principaux filtres existants ont fait l’objet d’un article dans le journal *Applied Optics* et d’un chapitre du livre “*Face recognition : Methods, Applications and Technology*”, en 2012. L’architecture JTC développée a été présentée à la conférence SPIE Defense Security and Sensing en 2011 et dans un article du journal *Optical Engineering* en 2012 (article en révision). Enfin, nous avons présenté une première version de l’algorithme et les résultats obtenus à la conférence SPIE Defense, Security and Sensing à Orlando en 2010. Les résultats de l’étape de détection seront présentés à la conférence ICoURS’12 en octobre 2012. Ces publications sont placées en annexe de cette thèse.

Plan du manuscrit

Ce manuscrit de thèse comporte quatre parties. Nous allons tout d'abord présenter le contexte de la guerre des mines sous-marines et l'état de l'art des méthodes de reconnaissance des mines sous marines, orienté selon l'approche choisie : la vidéo et le traitement optique. Ensuite nous présenterons la méthode retenue, la corrélation, pour traiter notre problème. Enfin nous appliquerons cette méthode à notre application.

Première partie

La première partie nous permet de présenter en détail le contexte de la thèse. En effet, pour bien comprendre les problématiques qui ont été les nôtres pendant ces trois années, il faut expliquer les enjeux de la guerre des mines sous marines ainsi que le milieu sous marin. Une opération de guerre des mines se déroule par étapes : la détection, la classification, l'identification et le désarmement (ce vocabulaire spécifique sera expliqué dans cette première partie). Chaque étape fait intervenir des systèmes et capteurs spécifiques. Cette thèse a pour objectif d'apporter une solution pour la partie identification des mines. Une première exigence consiste à utiliser une caméra vidéo pour cette étape. La qualité des images vidéo ne permet pas d'effectuer une reconnaissance sur les images brutes. Nous développons dans cette partie, les prétraitements mis en œuvre pour améliorer nos images, afin de faciliter l'étape de reconnaissance d'objet. Enfin, un état de l'art des méthodes existantes a été effectué afin de choisir l'algorithme de reconnaissance que nous voulons mettre en place.

Deuxième partie

Cette partie présente la méthode que nous avons choisie d'utiliser : la corrélation. Nous nous sommes basés sur la méthode optique que nous avons simulée numériquement avec Matlab. Cette méthode est composée de deux architectures (le corrélateur de Vanderlugt et le corrélateur de Weaver et Goodman). Pour chacune de ces architectures, nous avons présenté l'état de l'art qui s'y rapporte ainsi qu'une comparaison des différentes versions de ces architectures. Nous avons contribué à ces deux architectures en proposant de nouveaux filtres et de nouveaux critères de décision. Ces apports seront détaillés dans cette partie. Enfin, nous avons appliqué la corrélation à la reconnaissance des mines, ce qui nous a permis de définir plus en détails la chaîne de traitements que nous allons appliquer aux images de mines.

Troisième partie

Les premiers tests effectués sur la reconnaissance des mines nous ont montré la nécessité d'ajouter une étape de détection à l'algorithme auquel nous pensions initialement. En effet, nous avons une contrainte forte sur la vitesse d'exécution : le temps de calcul de l'algorithme doit être proche de la cadence vidéo, afin de permettre le guidage du drone autonome.

Dans un premier temps, nous avons étudié l'étape de détection. Une première méthode a été trouvée et testée. Cette méthode, basée sur la phase du spectre de l'image, permet de détecter les contours d'un objet. Les résultats sont comparés à des critères permettant de limiter les fausses alarmes et les arrêts de détection. Une seconde méthode, basée sur la soustraction du fond et l'utilisation des informations de navigation, a été adaptée à notre milieu très contraignant pour palier les inconvénients de la première méthode sur les images contenant des objets peu contrastés par rapport au fond. Enfin, nous avons comparé les résultats de ces deux algorithmes à un algorithme existant.

Dans un deuxième temps, nous nous sommes intéressés à la classification et à l'identification des mines. Pour cela, nous avons fait le choix d'utiliser des images de synthèse comme base d'images de référence. Ces images ont été traitées de façon à créer des filtres de corrélation que nous avons appliqués à nos images. Des prétraitements supplémentaires ainsi que des contraintes ont été nécessaires pour améliorer notre algorithme initial, en particulier, en utilisant les informations de navigation.

Quatrième partie

La dernière partie de cette thèse est consacrée au bilan des travaux effectués durant ces trois années. Nous nous sommes aussi intéressés aux travaux futurs qui pourraient prolonger cette thèse. Deux pistes se dégagent de nos réflexions.

Une première solution consiste à changer le système d'enregistrement des images en utilisant la polarisation de la lumière. Nous pensons que la polarisation va permettre d'obtenir des images de meilleure qualité (via l'utilisation des informations apportées par l'utilisation de la polarisation) qui nécessiteront moins de prétraitements que les images actuelles. De plus la polarisation permet d'observer les objets à une plus grande distance que les systèmes optiques sans polarisation. Cette augmentation de la distance peut permettre de confirmer avec davantage de certitude notre décision sur l'objet présent. Des tests sont actuellement effectués pour quantifier l'apport de cette méthode.

Notre algorithme a vocation à être embarqué sur un drone autonome. Il sera donc traduit pour fonctionner sur des cibles programmables. Les travaux effectués dans notre laboratoire montrent que la corrélation 2D est réalisable sur ces cibles et que le temps de calcul nécessaire est réduit par rapport au temps de calcul nécessaire avec un programme comme Matlab.

Première partie

Contexte et état de l'art

Introduction

Dans cette partie, nous allons présenter le contexte de cette thèse ainsi qu'un état de l'art de ce qui se fait actuellement en traitement d'images sous marines, aussi bien au niveau du prétraitement des images que de la détection et de la reconnaissance des objets (nous nous intéresserons aussi à des méthodes du milieu terrestre susceptibles de répondre à nos besoins). Par reconnaissance, nous désignons à la fois la classification et l'identification de l'objet, à la manière des opticiens. Ces présentations nous amèneront à des choix initiaux sur la façon dont nous allons traiter notre problème.

Cette thèse traite de l'identification des mines sous marines. Nous allons évoquer la menace que ces mines représentent. Nous exposerons les enjeux que représentent cette menace. Pour traiter efficacement cette menace, il est nécessaire de comprendre le milieu sous marin. Nous allons orienter notre présentation de ce milieu en fonction des deux principaux capteurs employés pour la guerre des mines : le sonar et la vidéo.

L'une des contraintes de cette thèse est l'utilisation d'une caméra vidéo. Cette contrainte entraîne la nécessité de prétraiter les images à utiliser avant d'effectuer des traitements. Nous présenterons ce qui se fait actuellement pour prétraiter les images vidéos sous marines. Nous utiliserons cette partie pour exposer ceux que nous avons choisis et les résultats obtenus.

Enfin nous détaillerons ce qui se fait actuellement en détection et en reconnaissance d'objets sous marins dans des images vidéos en liaison avec Thales. C'est un sujet d'étude récent, dans lequel toutes les méthodes efficaces dans le domaine terrestre n'ont pas encore été appliquées. Nous détaillerons ainsi quelques méthodes du milieu terrestre avant de conclure sur la méthode que nous allons utiliser.

Chapitre 1

Contexte et sujet de la thèse

Sommaire

1.1	Enjeux de la guerre des mines	9
1.1.1	Les enjeux géopolitiques et militaires	9
1.1.2	Les types de mines	10
1.1.3	Notre position par rapport aux enjeux	10
1.2	Le milieu sous marin	11
1.2.1	La vidéo	11
1.2.2	Le sonar	13
1.2.3	Sonar ou vidéo?	14
1.2.4	Les capteurs	16
1.3	Les étapes d'une opération de guerre des mines actuelle	18
1.3.1	La détection	18
1.3.2	La classification	18
1.3.3	L'identification	19
1.3.4	Le désarmement	19
1.4	Les moyens de guerre des mines	19
1.4.1	Le scénario actuel	19
1.4.2	Le scénario futur	20
1.5	Solution proposée	21
1.5.1	Les systèmes existants et futurs	22
1.5.2	Les objectifs de cette thèse	23
1.5.3	Les détails de l'étape d'identification	24
1.5.4	L'utilisation des capteurs	25
1.6	Conclusion	25

Cette thèse se situe dans le domaine de la guerre des mines sous marines. Les conflits armés de ces dernières années (guerres du Golfe de 1990-1991 et de 2003, conflit en Lybie de 2011 et en Iran en 2012 [1–5]) nous montrent que les mines sont toujours utilisées pour retarder voire empêcher l'accès aux ports ou aux littoraux. La menace est importante et prise en compte par les différentes marines nationales. En effet, les navires doivent être équipés ou escortés par des moyens de lutte anti-mine afin de garantir la sécurité des personnes et des navires. Dans ce chapitre, nous allons exposer la menace et des moyens pour y faire face. La place et les objectifs de ce travail de thèse dans ce scénario complexe seront éclaircis.

1.1 Enjeux de la guerre des mines

1.1.1 Les enjeux géopolitiques et militaires

Les mines sous marines existent depuis plusieurs siècles. On trouve les premières mentions de mines sous marines dans un traité chinois du 14^{ème} siècle [1]. Par la suite et plus récemment,

elles ont été utilisées dans les différents conflits militaires comme la guerre d'indépendance des États Unis, la guerre de Sécession, les deux guerres mondiales, les guerres du Golfe. En plus de l'utilisation des mines sur le plan militaire, les mines sous marines peuvent être utilisées à des fins économiques. En effet, les mines représentent une menace certaine et peuvent être mouillées dans le but de bloquer un passage emprunté par des navires de marchandises (par exemple la menace de l'Iran de bloquer le détroit d'Ormuz, [2]).

Les différents conflits ont montré le côté asymétrique que représente cette menace. Les dégâts causés par les mines sont importants. Au début de la première guerre du Golfe, un bateau militaire a été endommagé par l'explosion d'une mine. Les réparations ont coûté environ 50 millions de dollars. La mine avait coûté entre 1500 et 3000 dollars et le navire environ 300 millions de dollars [3]. Pour cela, les mines sont qualifiées d'arme du pauvre [4]. En plus de la menace de dégâts et de réparations coûteuses, la présence d'une mine implique des opérations de déminage et donc un retard temporel et des mesures de sécurité supplémentaires. La tentative de minage du port de Misrata en Lybie en mai 2011 [5] en est l'exemple : des bateaux avec pour ambition de miner le port ont été découverts mais ils avaient eu le temps de mouiller des mines. La zone a été déminée, retardant l'arrivée de l'aide aux populations. Des mesures ont ensuite été prises afin de maintenir la sécurité des convois humanitaires.

1.1.2 Les types de mines

Les mines sous-marines sont simples de conception [3]. Cependant, elles sont très diversifiées, par leur déclenchement, leur forme, leur taille,...

1.1.2.1 Le déclenchement des mines sous marines

Les mines sous-marines peuvent être déclenchées de plusieurs façons. Les mines de contact explosent au contact d'un objet. Les mines à influence se déclenchent lors de perturbations dues au passage d'un bateau. Il existe plusieurs types d'influence : l'influence acoustique (qui détecte les variations des niveaux acoustiques), l'influence magnétique (qui détecte les variations du champ magnétique terrestre), l'influence dépressionnaire (qui détecte la variation de pression hydrostatique). D'autres systèmes de déclenchement existent. [6, 7]

1.1.2.2 Les formes des mines sous marines

Il est difficile de classer les mines par formes car il existe des mines de toutes tailles, de toutes formes. Elles peuvent être simples et bon marché. D'autres sont furtives et/ou équipées de filets de camouflage. Les mines standard sont au nombre de quatre : la mine sphérique, la mine cylindrique, la mine Manta et la mine Rockan. Ces mines posent déjà des problèmes d'identification par la forme car certains points vues de deux mines de formes différentes peuvent être confondus (par exemple une sphère avec l'extrémité circulaire d'une mine cylindrique). De plus, d'autres formes de mines existent. La complexité des formes les rendent plus furtives et donc plus difficiles à détecter. Enfin, la menace peut aussi venir d'objets transformés en mine (par exemple un baril de pétrole).

Dans cette étude, nous nous sommes restreints aux quatre mines standard (*cf.* figure 1.1), présentes sur les lieux de tests et d'enregistrement des images [8]. Nous disposons ainsi de bases d'images importantes pour ces mines. La mine sphérique peut être dérivante, située entre deux eaux et reliée au sol par un orin ou posée au sol. Les mines cylindriques, Manta et Rockan sont posées au sol.

1.1.3 Notre position par rapport aux enjeux

Cette thèse se situe du côté défensif de la menace : il s'agit de détecter, de classer et d'identifier les mines. Pour cela, différents moyens peuvent être utilisés. Dans l'étude de ces systèmes, il faut prendre en compte l'influence du milieu sous marin qui peut perturber les

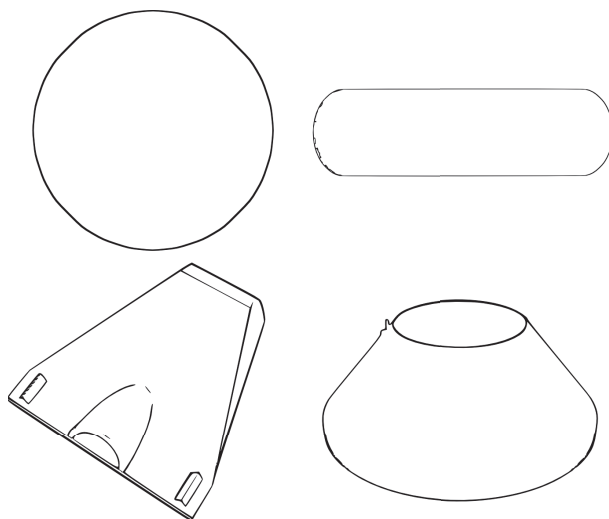


FIGURE 1.1 – Les différentes mines étudiées (images issues de modèles 3D). De gauche à droite et de haut en bas : mine sphérique, mine cylindrique, mine Rockan, mine Manta

signaux utilisés et par conséquent les résultats obtenus. Pour limiter les impacts du milieu sur notre étude, nous allons commencer par l'étudier.

1.2 Le milieu sous marin

Pour détecter et identifier des objets ou des animaux dans le milieu sous marin, le sonar et la vidéo sont utilisés. Le milieu sous marin, par sa nature influe sur les ondes acoustiques et électromagnétiques qui le parcourent. Le phénomène le plus important est l'atténuation des ondes. Cette atténuation est due à l'absorption et à la diffusion. L'absorption dépend de la distance entre la source et le récepteur et de la fréquence de l'onde. La diffusion dépend de la présence de particules dans l'eau.

1.2.1 La vidéo

L'une des possibilités pour étudier la présence d'objets est l'utilisation d'une caméra optique (ou autres systèmes optiques que nous présenterons par la suite). L'effet du milieu marin sur la lumière a fait l'objet de plusieurs études [9–12]. Toutes ces études ont conclu que la lumière reçue par le capteur est atténuée. Cette atténuation, si elle est importante, peut empêcher la détection d'un objet. Elle est due à la diffusion et à l'absorption. Elle varie selon la longueur d'onde utilisée pour éclairer la cible et la profondeur d'observation. Pour modéliser ce phénomène, l'atténuation est considérée comme la somme de la diffusion et de l'absorption [11] :

$$\begin{aligned} \text{atténuation} &= \text{absorption} + \text{diffusion} \\ I(z) &= I(0)e^{-az}e^{-dz} \end{aligned} \tag{1.1}$$

où $I(0)$ représente l'intensité (en cd) à la position 0 et $I(z)$ à la distance z (en m), a représente le coefficient d'absorption (en m^{-1}) et d le coefficient de diffusion (en m^{-1}).

La diffusion est due à la présence d'organismes biologiques transparents et de particules. La solution de Mie aux équations de Maxwell (dont la diffusion de Rayleigh est un cas limite) permet de décrire la diffusion de la lumière par des particules [13]. L'intensité de la diffusion dépend de la taille de la particule, de la direction dans laquelle la lumière est diffusée et de la distance entre la particule et l'observateur. McGlamery [12] montre aussi la dépendance de la diffusion avant avec la distance entre la particule et la caméra. Les variations d'intensité et de direction de la diffusion sont représentées sur les diagrammes polaires, figure 1.2. Plus le diamètre de la particule augmente plus la rétro-diffusion (diffusion vers l'arrière due aux petites particules) est faible par rapport à la diffusion avant (due aux grosses particules).

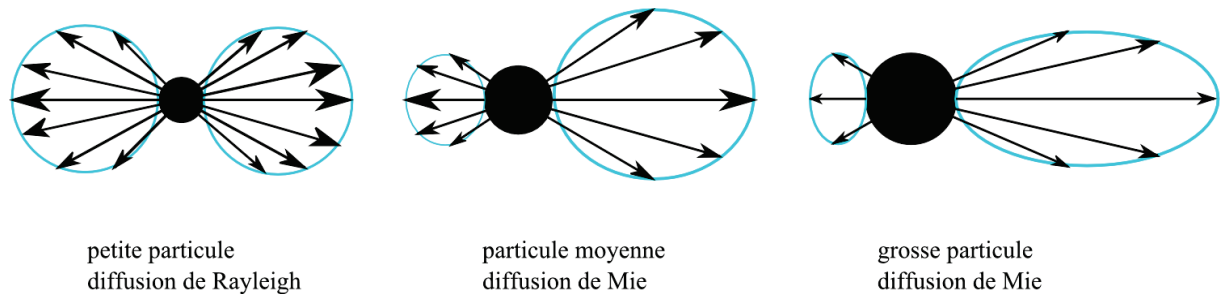


FIGURE 1.2 – Diffusion de la lumière incidente par une particule selon la théorie de Mie. De gauche à droite, intensité de la diffusion dans le cas de la diffusion de Rayleigh, intensité de la diffusion selon la théorie de Mie, dans le cas de particules de tailles plus importantes, d'après [14]. L'onde incidente arrive par la gauche

Ces résultats théoriques sont similaires à ceux présentés par Duntley [9]. La diffusion avant de la lumière implique que l'image d'un objet soit floue tandis que la rétrodiffusion crée un voile sur l'image. La présence de ces particules diffusant la lumière est visible sur les images. Elles sont appelées neige marine [15]. Leur intensité et leur déplacement sur les images sont variables au cours de la vidéo. Ces particules peuvent donc introduire des erreurs dans des algorithmes tels que la détection de contours ou le suivi d'objets.

L'absorption est liée à l'indice complexe de réfraction du milieu et à la présence de particules pigmentées dans ce même milieu [11]. Plusieurs campagnes de mesures ont été effectuées pour mesurer des coefficients d'absorption de l'eau selon la longueur d'onde [16–18]. Sur la figure 1.3, nous avons utilisé les coefficients d'absorption de [16] (mesurés en eau pure) puis calculé l'intensité reçue pour une intensité initiale de 1 selon la distance, en ne tenant compte que de l'absorption. Les distances indiquées sont celles à partir desquelles l'intensité reçue calculée est inférieure à 0.01, soit 1% de l'intensité initiale.

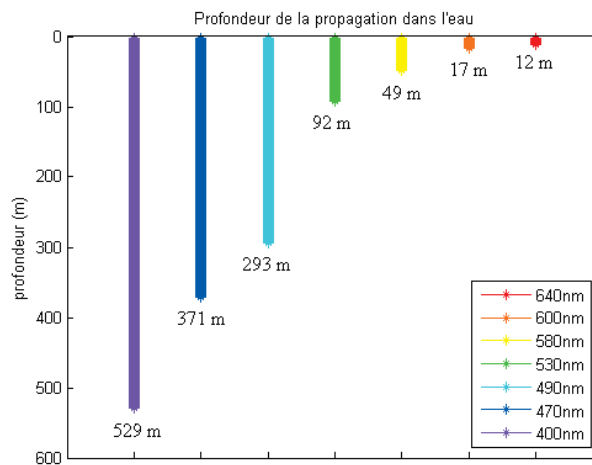


FIGURE 1.3 – Profondeurs à partir desquelles une longueur d'onde est complètement absorbée par le milieu sous marin (d'après [16])

A l'aide de l'atténuation totale et de considérations physiques, Jaffe [11] peut définir l'intensité émise et prédire l'image reçue, qu'il définit comme la superposition linéaire de trois composantes : la composante directe atténuée, la composante qui a subi de la diffusion avant et la rétrodiffusion. L'image reçue peut être écrite comme la somme de trois images :

$$I_{\text{reçue}} = I_{\text{atténuée}} + I_{\text{diffusion avant}} + I_{\text{rétrodiffusée}} \quad (1.2)$$

Cette superposition d'images composant l'image reçue est résumée sur la figure 1.4.

Un autre modèle de formation des images consiste à dire que l'image reçue est le produit de la réflectance de la scène (image émise par la scène) et de l'illumination locale. Ce modèle est

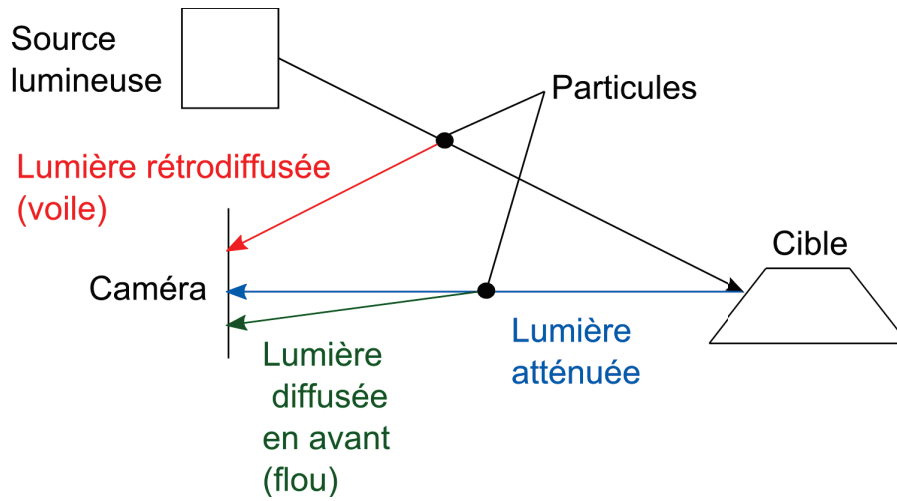


FIGURE 1.4 – Schéma représentant l'image reçue et les trois composantes lumineuses qui la composent

appelé “*illumination-reflectance model*” :

$$I_{\text{reçue}} = I_{\text{émise}} \cdot h \quad (1.3)$$

où h représente l'illumination locale.

1.2.2 Le sonar

La deuxième possibilité pour étudier le milieu sous marin est l'utilisation d'un système basé sur un sonar et sur les ondes acoustiques. Le milieu sous marin agit sur ces ondes acoustiques. L'équation du sonar passif peut être écrite de la façon suivante :

$$L_{\frac{S}{N}} = SL + TL - (NL - DI)$$

avec

$$\begin{aligned} SL &= 10 \log \left(\frac{I_S}{I_0} \right) \\ TL &= 10 \log \left(\frac{I_S}{I_R} \right) \\ NL &= 10 \log \left(\frac{I_N}{I_0} \right) \\ DI &= 10 \log \left(\frac{N_{ND}}{N_D} \right) \end{aligned} \quad (1.4)$$

où $L_{\frac{S}{N}}$ représente le rapport signal à bruit en décibel, SL l'intensité de la source, TL les pertes de transmission, NL le niveau de bruit, DI l'index de directivité du récepteur, I_S l'intensité du signal, I_0 l'intensité référence, I_R l'intensité du signal reçu, N_{ND} le niveau de bruit provenant de toutes les directions et N_D le niveau de bruit généré par le récepteur.

De même, l'équation du sonar actif peut être écrite :

$$L_{\frac{S}{N}} = SL + 2TL + TS - (NL - DI) > DT \quad (1.5)$$

où TS représente la réflectivité de la cible et DT le seuil de détection.

La puissance est émise ou perçue dans une direction donnée par sommation cohérente du signal émis ou perçu sur plusieurs transducteurs. Les phénomènes qui interviennent sont analogues à ceux de la diffraction en vision. Il est possible de montrer que la directivité d'une antenne linéaire dépend de la longueur de cette antenne et de la fréquence du signal [19] :

$$DI = 10 \log \left(\frac{2l}{\lambda} \right) \quad (1.6)$$

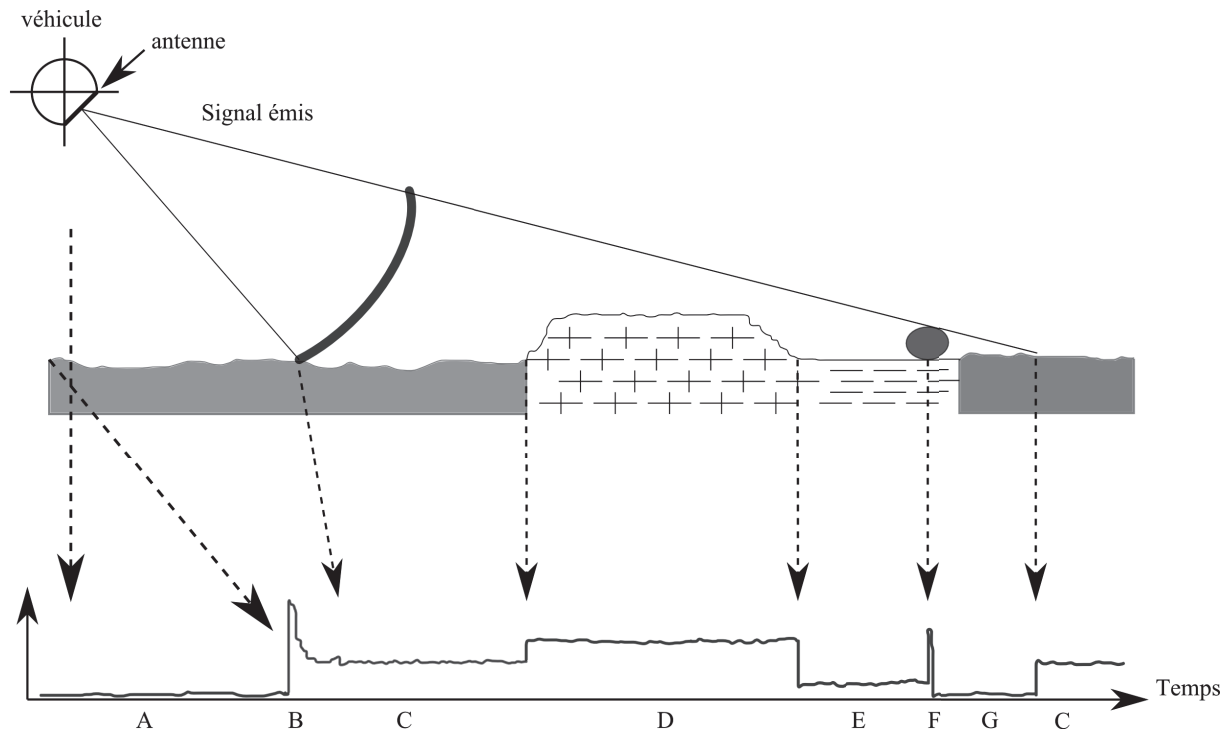


FIGURE 1.5 – Géométrie de l'image sonar, d'après [20]. A : bruit et réverbération dans l'eau, B : écho de fond, C : zone de sable, D : roche, E : vase, F : écho de cible, G : ombre portée par la cible

où l est la longueur de l'antenne, en m et λ est la longueur d'onde du signal, en m. Pour augmenter la directivité d'une antenne, il existe deux possibilités : augmenter la longueur de l'antenne (ce qui peut être compliqué en pratique) ou diminuer la longueur d'onde, ce qui revient à augmenter la fréquence du signal.

Les pertes en transmission regroupent les pertes par divergence et les pertes par atténuation. Il est possible de montrer que ces pertes s'écrivent [19] :

$$TL = -20\log(z) - cz \quad (1.7)$$

où le premier terme ($-20 \log(z)$) correspond aux pertes par divergence, avec z la distance en m et le second terme cz correspond aux pertes par atténuation, avec c le coefficient d'atténuation, en m^{-1} . Plusieurs formules, fonction de la fréquence du sonar, permettent d'obtenir des valeurs de ce coefficient (Francois Garrison, Thorp).

Selon l'application pour laquelle le sonar sera utilisé, il faut trouver un compromis de fonctionnement entre la fréquence et la portée. En effet la directivité dépend de la fréquence et les pertes dépendent de la distance et de la fréquence. Soit la portée est importante et les images seront de mauvaise qualité (la fréquence doit être faible dans ce cas) soit la qualité est privilégiée (une fréquence élevée) mais la portée sera faible.

Le sonar émet l'onde dans un cône. L'image reçue a donc une géométrie particulière, comme le montre la figure 1.5. Lors de l'émission d'une onde acoustique, toute l'onde ne va pas frapper le fond en même temps, cela dépend de l'angle d'inclinaison du sonar (angle formé entre l'horizontale et la direction pointée par le sonar). L'écho reçu va donc s'étaler dans le temps. De plus l'inclinaison du sonar implique une distorsion plus ou moins importante de l'image reçue.

1.2.3 Sonar ou vidéo ?

Nous avons vu la façon dont le milieu sous marin perturbe les ondes émises et par conséquent complique la détection d'objets. Selon les applications, le sonar ou la vidéo sera utilisé. Les principaux avantages et inconvénients des deux systèmes sont résumés dans le tableau 1.1.

	Avantages	Inconvénients
Sonar	portée	à courte distance et résolution équivalente, coût géométrie de l'image distorsion
Vidéo	résolution coût	contraste bruit pas d'exploitation de la couleur

Tableau 1.1 – Tableau résumant les avantages et inconvénients du sonar et de la vidéo

Le sonar a l'avantage de la portée sur la vidéo : un sonar de détection a une portée jusqu'à 400m là où la vidéo a une portée maximale d'une vingtaine de mètres, selon la turbidité. Cela s'explique avec les valeurs des coefficients d'absorption qui sont beaucoup plus faibles en acoustique qu'en optique (*cf.* figure 1.6). Dans le cas d'applications nécessitant une longue portée, la résolution de l'image sonar associée sera faible, selon la formule de la directivité, équation 1.6. La vidéo permet d'obtenir des images résolues à courtes portées pour un coût moindre, comparativement au sonar très hautes fréquences ou au sonar à balayage. Cette option est donc privilégiée lors de la définition d'un drone perdable. Cependant, du fait de l'absorption dépendant de la longueur d'onde, la couleur est difficilement exploitable dans les vidéos sous marines.

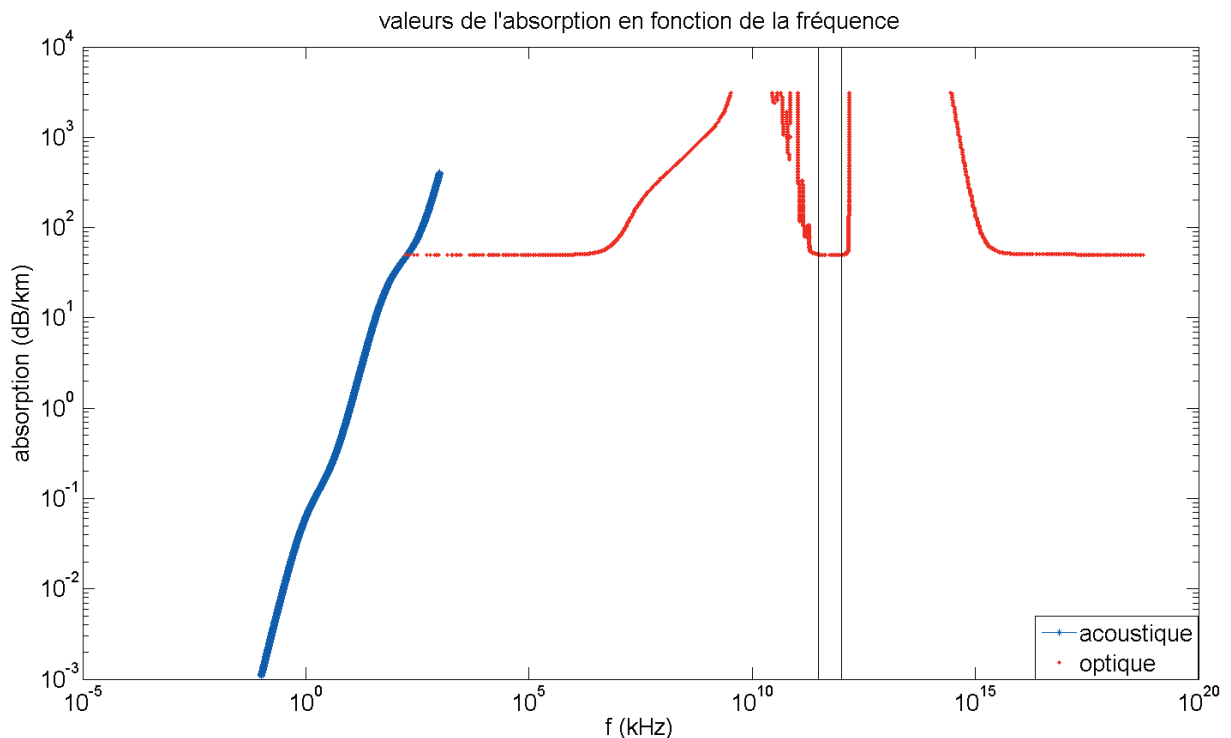


FIGURE 1.6 – Comparaison des valeurs des coefficients d'absorption pour une onde acoustique et pour une onde optique. Les deux traits noirs correspondent aux limites du visible. [21], [22]

Les valeurs des coefficients d'atténuation acoustique de la figure 1.6 ont été calculées à partir de l'article de Ainslie et McColm [21]. Nous nous sommes placés dans les conditions suivantes : une température de 8°C , une salinité de 35 ppt, un pH de 8 et une profondeur de 0 m. Les valeurs de l'absorption pour une onde optique dans l'eau ne proviennent que de mesures et non d'une formule. Les valeurs de la figure 1.6 ont été récupérées à partir d'une figure du livre de Jackson [22].

Dans cette étude où nous souhaitons détecter et identifier des mines, le système choisi est basé sur un compromis entre le sonar et la vidéo. Le sonar détecte une zone potentiellement dangereuse lorsque le bateau est éloigné de cette zone. A courte distance, la vidéo est utilisée de façon à confirmer les résultats obtenus avec les systèmes sonars.

1.2.4 Les capteurs

Plusieurs capteurs sont nécessaires pour que les bateaux et les drones effectuent leur mission. Tout d'abord le positionnement permet de donner une indication de la position des bateaux les uns par rapport aux autres et une indication sur la position des objets détectés. Ensuite la détection et l'identification nécessitent des capteurs spécifiques comme le sonar et/ou la vidéo. Enfin la communication entre les bateaux est importante.

Nous présentons ici, rapidement, quelques capteurs qui peuvent équiper les navires et les drones. La volonté d'utiliser une information de navigation particulière permet de sélectionner un ou plusieurs capteurs. Cependant, pour concevoir un système dans sa globalité, une étude plus approfondie est nécessaire.

1.2.4.1 Le positionnement

Les bateaux de surface, peuvent se repérer et se positionner à partir du système GPS, "Global Positioning System". Sous la surface, ces méthodes fonctionnent mal car les ondes émises (à très hautes fréquences) par les satellites sont très fortement atténuées. Plusieurs capteurs ont été étudiés de façon à remédier à ce problème. Lorsqu'un drone sous-marin effectue une mission, il connaît la position du chasseur de mines et la position de l'objet à aller visiter, aux erreurs de capteurs près. Pour se repérer, le drone doit connaître sa profondeur et son altitude, sa position par rapport au chasseur de mines et son attitude (cap, roulis et inclinaison).

L'altitude est connue à l'aide d'un sondeur acoustique. Il s'agit d'un capteur qui envoie un faisceau acoustique vers le fond. Il détecte le retour de ce faisceau. Connaissant la célérité du son (environ 1500 m.s^{-1} dans l'eau) et le temps écoulé entre l'émission et la réception du signal, la longueur du trajet effectué peut être calculée.

La profondeur est mesurée à partir d'un capteur de pression. La pression absolue est la somme de la pression atmosphérique (1 bar) et de la pression hydrostatique qui augmente linéairement d'un bar tous les 10 m. Connaissant la pression, la distance qui sépare le drone de la surface peut être calculée.

Le drone a également besoin de savoir où il se situe par rapport au bateau, dans le plan (x,y). Plusieurs capteurs et/ou méthodes permettent d'obtenir cette information de position : les centrales inertielles, le positionnement acoustique ainsi que le SLAM, "simultaneous localization and mapping" [23]. La centrale inertielle permet de calculer la position, l'évolution de la vitesse ainsi que de l'attitude (cap, roulis et inclinaison) du véhicule à partir de données issues d'un accéléromètre, de gyroscopes et de gyromètres. Le positionnement acoustique mesure une position par rapport à un réseau de base déployé en surface et/ou en profondeur. L'objectif du SLAM est de cartographier l'environnement du drone tout en retenant les endroits par lesquels il est passé [23].

Les informations sur la position du véhicule peuvent servir à prévoir la position d'un objet connaissant les positions aux instants précédents par rapport à la position du drone. Ces informations peuvent aussi être utilisées pour interpréter et optimiser les résultats obtenus. En effet, connaissant le mouvement du véhicule, les mouvements du ou des objets, détectés dans les images acoustiques et optiques, peuvent être analysés.

1.2.4.2 La détection

Pour la détection, deux types de capteurs sont à notre disposition : les capteurs acoustiques et les capteurs optiques. Le choix du capteur dépend de l'application, de la portée, de la résolution (cf. partie 1.2.3).

Dans la famille des capteurs acoustiques, sont à notre disposition :

- le SSS, "Side Scan Sonar". Le SSS est un sonar latéral qui permet d'imager de grandes parties du fond marin [24]. L'image acquise est perpendiculaire à la trajectoire du véhicule. L'image d'une grande zone est obtenue grâce à la navigation du véhicule.
- le SAS, "Synthetic Aperture Sonar" (sonar à ouverture synthétique). C'est un sonar latéral qui à partir de données reçues par une petite antenne permet de reconstituer, à l'aide d'un

traitement numérique, une image acoustique de bonne résolution. Pour obtenir la même image avec un sonar latéral sans traitement numérique, il faut utiliser une grande antenne, ce qui n'est pas toujours pratique à mettre en œuvre. [24]

- le sonar frontal avec ou sans formation de voies. [25]
- les sondeurs mono-faisceau qui servent aussi à connaître l'altitude [24]
- les sondeurs multi-faisceaux. Ces sondeurs, qui envoient un faisceau acoustique vers le sol permettent de détecter l'altitude du drone par rapport au sol. Ils peuvent aussi détecter un objet. En effet, l'utilisation de plusieurs faisceaux permet d'analyser une bande du fond. Un objet peut alors être détecté à la fois sur la largeur du faisceau mais aussi dans la direction perpendiculaire, en analysant plusieurs images consécutives. [24]
- les caméras acoustiques. Il s'agit de sonars très haute résolution qui produisent des images proches des images obtenues par caméras optiques. [26]

Dans la famille des capteurs optiques, il y a :

- la caméra vidéo. Il s'agit d'un système peu onéreux et très résolu. Cependant l'image obtenue est bruitée par le milieu marin. De plus l'absorption de la lumière limite la portée visuelle.
- le LIDAR, "*light detection and ranging*". Ce système permet de déterminer la distance à un objet en mesurant le temps de trajet de la lumière émise puis renvoyée par un objet. [27]
- le STIL, "*Streak Tube Imaging Lidar*". Ce système est basé sur le LIDAR. Le faisceau laser est élargi pour éclairer une large bande. Le récepteur est une photocathode qui convertit les photons en électrons. Ces électrons sont déviés par le champ électrique créé entre deux plaques vers une ligne particulière du capteur CCD, "*Charge-Coupled Device*". La ligne vers laquelle est déviée le signal dépend du temps. Le système STIL développé par l'entreprise Areté Associates est utilisé par l'armée américaine. Il fait partie intégrante du système ALMDS, "*Airborne Laser Mine Detection System*" utilisé par les hélicoptères MH-60S [28].
- les range-gated cameras. L'utilisateur sélectionne une distance. La caméra émet un faisceau ou un flash laser. Le récepteur est obturé pendant le temps nécessaire au parcours aller-retour de la lumière de la caméra à l'objet positionné à la distance choisie. Cela permet de s'affranchir de la lumière rétrodiffusée par les particules de l'air ambiant ou par les objets situés à une distance inférieure et à une distance supérieure de la distance choisie par l'utilisateur. [29]

Nous avons vu dans la partie 1.2.1 que le milieu sous marin limite la portée de l'onde lumineuse et que la rétrodiffusion et la diffusion avant perturbent l'image reçue par le capteur. Pour réduire l'effet du bruit sur les images, des recherches sur la conception des systèmes d'acquisition des images ont été effectuées. Selon Jaffe [11], des compromis sont à établir entre la distance caméra / source lumineuse et le contraste souhaité tout en restant dépendant des conditions du milieu. Il a développé un modèle qui permet d'étudier les différentes configurations [11]. Il modélise les composantes directes et diffusées ainsi que la composante rétrodiffusée. Ses simulations montrent qu'un système avec une source de lumière séparée de la caméra améliore les performances du système.

Il est aussi possible d'utiliser une caméra polarisée pour enregistrer les images. Entre autres, Y.Y. Schechner a travaillé sur un système de ce type [30–34]. Il s'agit de mettre un polariseur devant la caméra qui enregistre les images. Le système est défini dans [34]. Selon l'orientation du polariseur, l'intensité de l'image enregistrée varie entre une image d'intensité I_{\max} et une image d'intensité I_{\min} . Ces 2 images "extrêmes" sont obtenues pour des orientations orthogonales du polariseur. Le système développé consiste à enregistrer ces deux images "extrêmes" puis à estimer la lumière rétrodiffusée. Ils compensent simplement les canaux de couleurs afin de retrouver une image aux couleurs plus proches des couleurs réelles. Dans [32], ils résolvent mathématiquement ce problème. Le système développé et l'algorithme de traitement des images permettent d'améliorer la qualité des images et d'augmenter la portée visuelle.

Le système qui a enregistré les images qui seront utilisées dans cette thèse est équipé d'un sonar pour la détection à grande distance et d'une caméra vidéo pour la détection à courte distance [8]. Afin d'enregistrer des séquences nocturnes, le système est équipé d'une source lumineuse

placée à proximité de la caméra.

1.2.4.3 La communication

Les drones communiquent avec les bateaux afin de transmettre des informations sur les mines trouvées (transmettre une image pour confirmation par exemple), de guider le drone pour explorer une zone ou un objet précis. Cette communication peut se faire de plusieurs façons. En surface, il est possible d'utiliser les communications aériennes comme les communications hertziennes, satellites, voire wifi. Sous l'eau, nous disposons principalement des communications acoustiques et des communications filaires (souvent par fibres optiques).

En communication acoustique, le débit diminue avec la distance, ce qui peut être limitant lorsque le drone transmet des images au chasseur de mines pour l'identification, par exemple. Une possibilité est de faire suivre le drone sous-marin par un drone de surface. Ces deux drones communiquent entre eux par communication acoustique et le drone de surface communique avec le chasseur de mines par communication aérienne.

L'autre solution de communication est la communication filaire par fibre optique. Le débit peut alors être important. Cependant, une fibre optique présente des inconvénients. Tout d'abord, il y a les inconvénients propres à la fibre optique : la nécessité de convertisseurs optique/électronique ainsi que la présence de répéteurs de gain si la longueur de la fibre est importante. De plus une fibre optique est fragile. En cas de casse, il faut en avoir de rechange sur le chasseur de mines. Si la fibre est perdue en mer, elle n'est pas toujours biodégradable (considération qui peut être prise en compte lors de l'étude du système). Enfin la fibre optique peut s'emmêler, notamment en présence de plusieurs drones.

La communication entre bateaux présente donc des problèmes. Les drones autonomes présentent une alternative : ils ne nécessitent pas d'intervention d'opérateurs humains et par conséquent pas d'utilisation de moyens de communication.

1.3 Les étapes d'une opération de guerre des mines actuelle

Lors d'une mission de guerre des mines, il y a plusieurs étapes à respecter pour vérifier si une zone est minée. Tout d'abord il y a la détection. Ensuite les objets détectés sont classifiés puis identifiés. Les mines sont alors détruites [7].

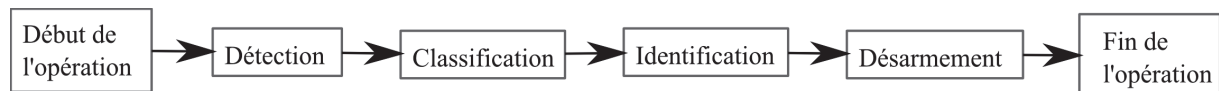


FIGURE 1.7 – Les étapes d'une mission de guerre des mines

1.3.1 La détection

Il s'agit de détecter des objets qui pourraient être des mines. Cette étape s'effectue au sonar à partir du chasseur de mines. Le sonar est caractérisé par une longue portée (environ 400 m), de basses fréquences (entre 10 et 30 kHz). Cela donne des images de mauvaise résolution sur lesquelles des échos qui pourraient être ceux de mines sont recherchés. Les détections sont appelées MILEC, "*mine like echo*".

1.3.2 La classification

Cette étape consiste à distinguer parmi les contacts MILEC ceux qui proviennent d'une mine (MILCO, "*mine like contact*") de ceux qui n'en sont pas (NON MILCO). Pour cela, un sonar courte portée (inférieure à 200 m) et hautes fréquences (entre 200 et 300 kHz) est utilisé. Il permet d'obtenir des images de meilleures résolutions que celles obtenues lors de l'étape de détection. Sur ces images, l'écho et l'ombre des objets détectés sont étudiés. En France, ce sonar

peut être embarqué sur le CMT (chasseur de mines tripartite) (cf. figure 1.8) ou sur un PVDS, “*Propelled Variable Depth Sonar*” (cf. figure 1.9).



FIGURE 1.8 – Chasseur de mine tripartite [35]



FIGURE 1.9 – PVDS de Thales Underwater Systems [36]

1.3.3 L’identification

L’étape suivante est l’identification. Les objets MILCO répertoriés lors de la classification sont classés en deux catégories : mine ou NOMBO, “*non mine bottom object*”. L’identification implique une analyse visuelle, effectuée directement par un groupe de plongeurs démineurs ou à distance par un ROV, “*Remotely Operated Vehicles*” (par exemple le K-ster I fabriqué par ECA Robotics ou le Seafox I fabriqué par Atlas). On peut aussi utiliser un sonar très hautes fréquences (de l’ordre de 1 GHz) qui par conséquent à une portée très courte (10 m), par exemple les sonars Didson [26].

1.3.4 Le désarmement

Une fois que la présence d’un objet manufacturé dangereux est confirmée, deux possibilités existent : la neutralisation et le pétardage. La neutralisation consiste à faire enlever le détonateur par des plongeurs démineurs. Le pétardage signifie l’explosion de la mine par une charge déposée par un ROV ou par les plongeurs ou par l’explosion d’un drone perdable qui se fait exploser à proximité de la mine.

Le scénario du pétardage par le GPD (Groupe de Plongeurs-Démineurs) est fixé de façon à garantir au maximum la sécurité des plongeurs. Un bateau les amène à une distance de sécurité de la mine. Les plongeurs plongent pour aller mettre une charge à proximité de la mine. Ils rejoignent le bateau et déclenchent l’explosif.

1.4 Les moyens de guerre des mines

Une mission de guerre des mines nécessite l’intervention de plusieurs acteurs. Les scénarios actuels nécessitent des bateaux construits selon les normes militaires et l’intervention d’opérateurs humains. Les scénarios pour le futur envisagent des batteries de drones autonomes qui ne nécessitent ni la présence de bateaux aux normes militaires sur la zone étudiée ni l’intervention d’opérateurs.

1.4.1 Le scénario actuel

Les chasseurs de mines effectuent une première recherche d’objets manufacturés dangereux (l’étape de détection, partie 1.3.1). Pour effectuer la classification (partie 1.3.2), on utilise un sonar de portée inférieure à 200 m qui équipe le chasseur de mines ou un ROV. Dans les deux cas, le chasseur de mines doit être sur ou à proximité de la zone à étudier. L’identification et le désarmement sont effectués par des plongeurs démineurs. C’est une étape dangereuse. L’idéal



FIGURE 1.10 – Scénario actuel de guerre des mines [35]. Une mine est présente en bas à droite de l'image. Le bateau représente le CMT qui détecte la mine, le ROV la classe et le plongeur la désarme.

serait de trouver un système efficace qui puisse permettre de ne pas envoyer d'hommes à proximité des mines. Les différentes étapes de ce scénario sont représentées figure 1.10.

Les chasseurs de mines dirigent les opérations de guerre des mines. Ils sont construits selon des normes militaires. Il en résulte des bateaux amagnétiques et avec une coque résistante en cas d'explosion. En France, les chasseurs de mines, *cf.* figure 1.8, aussi appelés CMT car ils sont fabriqués par les marines française, belge et néerlandaise. Leur coque est en composite verre/résine polyester [37]. Aux États Unis, il s'agit de la classe Avenger, *cf.* figure 1.11.



FIGURE 1.11 – Chasseur de mines américain, classe Avenger [14]

Les ROVs, *cf.* figure 1.9, sont des drones mis à l'eau à partir du chasseur de mines. Ils sont reliés et en communication acoustique ou filaire avec le chasseur de mines. Un opérateur les guide et peut ainsi explorer une zone et sélectionner les objets qu'il faut aller étudier de plus près. Certains de ces drones peuvent déposer une charge à côté de la mine (par exemple le PAP 104, Poisson Auto-propulsé, d'ECA Robotics). Cette charge sera déclenchée lorsque le ROV se sera suffisamment éloigné de la mine voire aura été récupéré par le chasseur de mines.

Le tableau 1.12 résume les étapes, les moyens et les capteurs utilisés actuellement en guerre des mines.

1.4.2 Le scénario futur

Dans l'objectif de construire des chasseurs de mines aux normes civiles (bateaux qui ne sont pas amagnétiques, qui n'ont pas de coques particulièrement étudiées pour résister aux explosions), de les éloigner des côtes (pour plus de discrétion) et de limiter l'intervention d'opérateurs

Étapes	Moyens		
	chasseur de mines	ROV	plongeurs
détection	sonar basses fréquences		
classification	sonar hautes fréquences	sonars hautes fréquences	
identification	(par ROV)	sonar très hautes fréquences vidéo optique	visuel

met en oeuvre

FIGURE 1.12 – Tableau résumant les étapes, les moyens et les capteurs du scénario actuel de guerre des mines

humains, des projets sont menés pour étudier la faisabilité d’une flotte de drones autonomes, communiquant entre eux pour réaliser l’ensemble de l’opération de déminage. En France ce projet s’appelle le SLAMF (Système de lutte anti-mines du futur), représenté figure 1.13 [4].

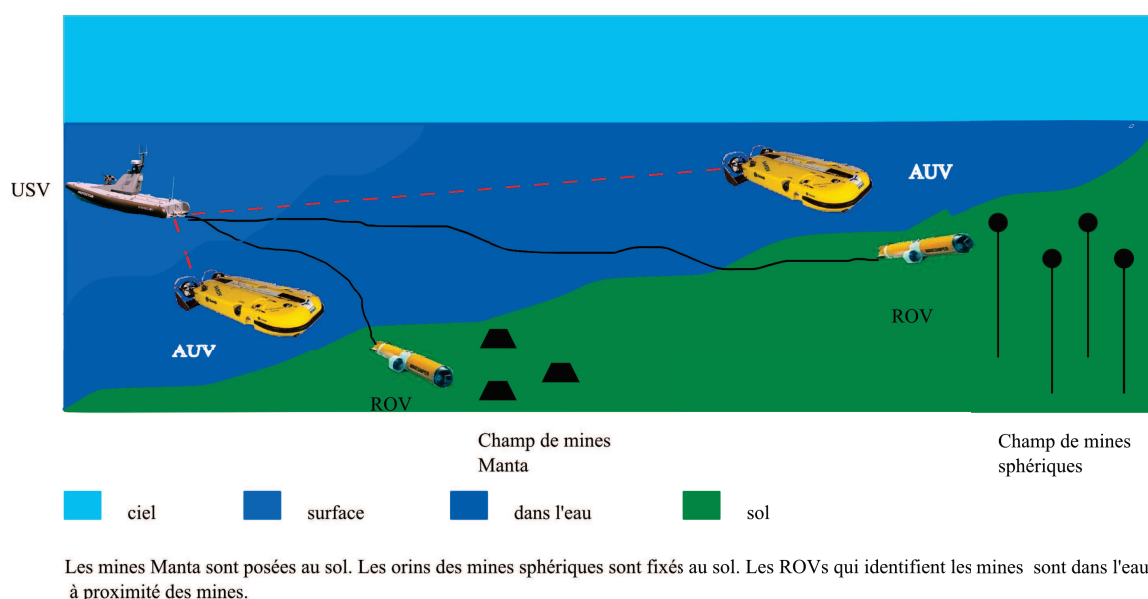


FIGURE 1.13 – Scénario du futur, d’après [4]

L’objectif est d’avoir un USV, “*Unmanned surface vehicle*” (cf. figure 1.14). Il s’agit d’un bateau autonome capable de remplir une mission, de mettre à l’eau des drones et de les récupérer. Cet USV peut mettre à l’eau des sonars remorqués, responsables de la détection, des AUV (“*Autonomous Underwater Systems*”, cf. figure 1.15) pour classifier et identifier et des ROVs pour détruire les mines. Dans le cas où le drone est responsable du pétardage de la mine, il est considéré comme un drone perdable puisqu’il va se faire exploser pour détruire la mine en même temps. L’objectif est de supprimer l’intervention humaine.

Le tableau 1.16 résume les étapes, les moyens et les capteurs prévus dans les scénarios futurs de guerre des mines.

1.5 Solution proposée

Cette thèse n’a pas pour ambition d’étudier les scénarios de guerre des mines dans leur intégralité! Nous avons ciblé l’étape d’identification effectuée par des drones autonomes et plus particulièrement les algorithmes associés. Tout d’abord nous allons faire l’état des lieux des drones autonomes d’identification qui existent et des projets dans lesquels la France a participé. Ensuite nous présenterons plus en détails nos objectifs, l’étape de guerre des mines dans laquelle nous nous situons ainsi que l’utilisation que nous faisons des données fournies par les capteurs

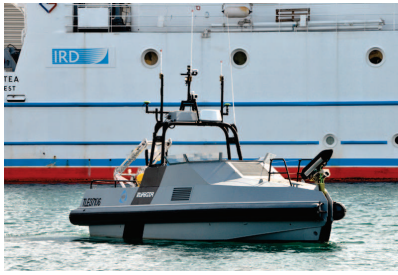


FIGURE 1.14 – USV INSPECTOR MK2 d'ECA Robotics [38]

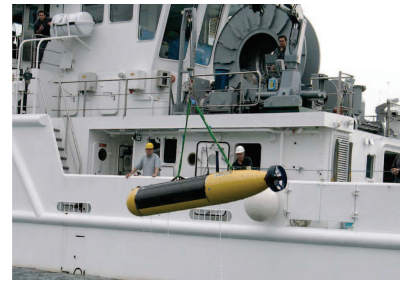


FIGURE 1.15 – AUV Muscle de Thales Underwater Systems et Bluefin Robotics [39]

met en œuvre

Étapes \ Moyens	met en œuvre			
	USV	Sonars remorqués	AUV	ROV
détection		sonar basses fréquences		
classification			sonars hautes fréquences	
identification			vidéo	
neutralisation				charge explosive

FIGURE 1.16 – Tableau résumant les étapes, les moyens et les capteurs du scénario futur de guerre des mines

équipant un drone sous marin.

1.5.1 Les systèmes existants et futurs

Des systèmes de drones autonomes sont à l'étude voire ou en cours d'utilisation pour la détection, l'identification et le désarmement des mines sous marines. Dans cette partie, nous ciblons plus particulièrement les systèmes autonomes, ou pouvant l'être, équipés de sonars et de caméras vidéos (liste non exhaustive). Dans les projets, on trouve le projet international VAMA (véhicule anti-mines autonome). Il y a aussi le projet Topvision de l'État français [40]. Ce dernier projet n'avait pas pour vocation de faire un drone mais d'étudier des algorithmes de traitement d'images sous marines. Plusieurs entreprises proposent des drones autonomes.

1.5.1.1 Les systèmes existants

Parmi les entreprises présentes dans le domaine des drones sous marins, certaines possèdent des ROVs et / ou des AUVs équipés de sonars et de caméras vidéos et qui peuvent être utilisés pour la guerre des mines. Ces drones sont étudiés pour détecter et identifier des objets. Certains de ces drones possèdent une version pour la neutralisation des objets dangereux.

Parmi les ROVs actuellement disponibles, on peut citer :

- le K-Ster d'Eca Robotics [38]
- le Double Eagle de Saab [41]
- le Minesniper Neutron de Kongsberg [42]
- le SeaFox de Atlas Elektronik [43]

Parmi les AUVs actuellement disponibles, on peut citer :

- l'Alister 100 d'Eca Robotics [38]
- le Marlin de Lockheed Martin [44]
- le SeaOtter Mk II de Atlas Elektronik [43]

1.5.1.2 Le projet Topvision

Le projet Topvision (Tests OPérationnels de Vidéos Sous-marines pour l'Identification d'Objets Nuisibles) est un projet organisé par le ministère de la recherche et par le ministère de la défense. Ce projet a été délégué au GESMA (Groupe d'Études sous marines de l'Atlantique) et coordonné par TUS. Ce projet a été réalisé entre 2005 et 2007.

Le but du projet Topvision était de regrouper des laboratoires civils (laboratoires d'écoles d'ingénieurs et d'universités, des instituts et des laboratoires de recherche). Les recherches devaient porter sur l'évaluation d'algorithmes de détection, de localisation, de caractérisation et d'identification automatique d'objets sous marins nuisibles. Certaines de ces recherches ont fait l'objet de publications et seront présentées dans la partie sur l'état de l'art (partie 3).

Les images issues de ce projet et utilisées dans cette thèse sont appelées par la suite images Topvision. Elles contiennent les informations de navigation inscrites en haut et en bas de l'image. Leur taille est de 720×540 pixels. Un exemple est présenté sur la figure B.1, p. 167.

1.5.1.3 Le projet VAMA

Le projet intitulé VAMA est un projet international qui consistait à étudier le fonctionnement d'un drone anti-mines autonome qui contient différents capteurs dont une caméra vidéo pour confirmer la classification et identifier l'objet détecté avec le sonar lors des étapes de détection et de classification (*cf.* parties 1.3.1 et 1.3.2). Cette caméra sert aussi au guidage du drone pour l'attaque et la neutralisation de la mine.

Le drone effectue l'étape d'identification 1.3.3. Pour cela, le drone effectue une étape de détection au sonar. Lorsque la mine est détectée, le drone se guide jusqu'au voisinage de l'objet grâce au suivi d'objets. Ensuite la vidéo effectue la détection et l'identification avant de guider le drone jusqu'à la mine. La détection d'objets dans les images se fait à partir d'une extraction de contours qui sont ensuite comparés à des lignes et des ellipses. Les contours les plus proches de ces formes "manufacturées" sont retenus. Les contours détectés sont comparés aux contours enregistrés dans une base de données. Une fois l'objet identifié, l'étape de suivi consiste, pour la partie traitement d'images, à poursuivre la détection de contours dans les images et à vérifier que l'évolution de l'orientation de la mine est cohérente avec les déplacements du drone. Les détails du scénario sont disponibles dans [8].

La dynamique de la caméra s'adapte à la luminosité ambiante. De plus autour de la caméra une couronne de LEDs, "*Light-Emitting Diodes*" a été ajoutée de façon à éclairer la scène lors de séquences en fin de journée ou de nuit.

Les images issues de ce projet sont appelées par la suite images VAMA et mesurent 640×480 pixels. Un exemple est présenté sur la figure 2.16a, p. 41.

1.5.2 Les objectifs de cette thèse

L'objectif de cette thèse est d'étudier une solution algorithmique pour optimiser l'identification optique d'objets manufacturés présents dans des vidéos sous marines, identification effectuée par un drone autonome. Ce sujet de thèse nous place dans l'étape d'identification d'un scénario futur (*cf.* partie 1.4.2) avec un drone équipé, en plus des capteurs de positionnement et de communication, de sonars et d'une caméra vidéo optique. Nous nous limitons à la recherche des mines sous marines présentées figure 1.1.

Pour valider nos algorithmes, nous disposons d'une base d'images importante constituée des images du projet Topvision et des images du projet VAMA. Cette base de données est constituée de vidéos de mines sphériques à orin ou posées au sol, de mines Manta et de mines cylindriques. Pour tester la robustesse de notre algorithme, nous possédons aussi des vidéos d'objets tels que des bouts et des barils de pétrole. Certaines vidéos ne contiennent pas d'objets.

Le système ayant vocation à être utilisé dans un drone autonome, l'algorithme doit être automatique et nécessiter des temps de calcul proches du temps réel (cadence vidéo) afin d'assurer le guidage du drone de façon cohérente par rapport aux résultats des étapes de détection et d'identification.

1.5.3 Les détails de l'étape d'identification

Pour cette étape, nous nous inspirons des principes développés dans [8]. La détection et la classification (*cf.* figure 1.17, première ligne) auront été effectuées précédemment. Lors de l'étape de classification, une liste de contacts MILCO a été établie (*cf.* partie 1.3.2). Plus un contact est susceptible d'être une mine plus il est en haut de la liste de contacts. Les sous-étapes de l'étapes d'identification sont représentées sur la figure 1.17.

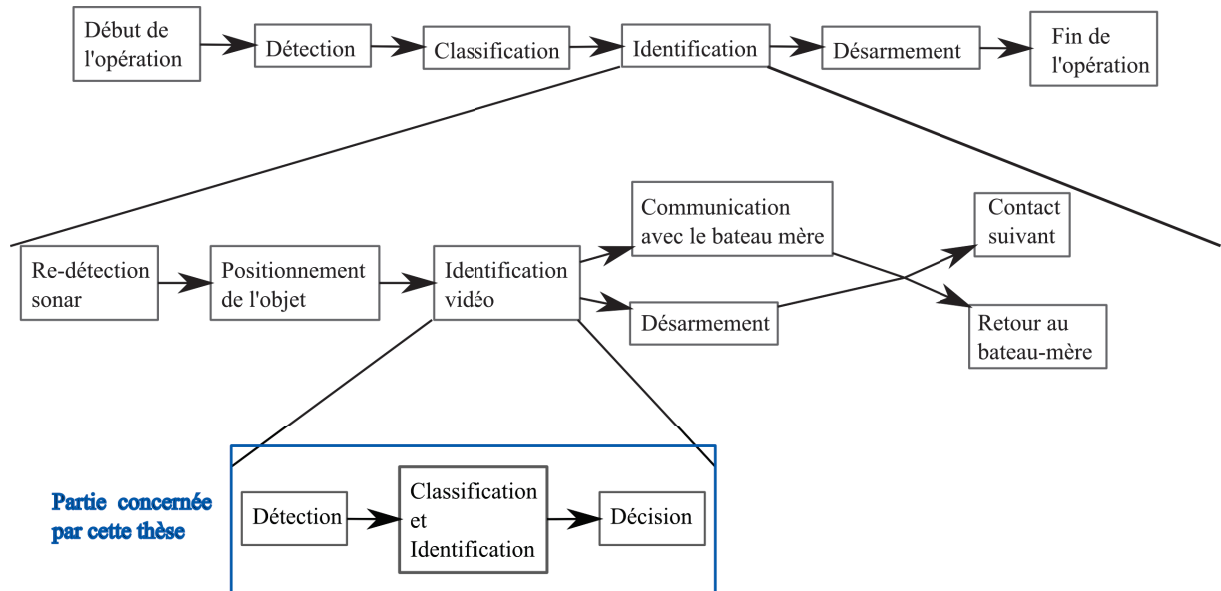


FIGURE 1.17 – Détails de l'étape d'identification

Le drone d'identification reçoit cette liste de contacts avec leur position par rapport au chasseur de mines. La première mission du drone est de retrouver le contact et de le positionner de façon plus précise dans son propre repère (*cf.* ligne 2 de la figure 1.17). Ces étapes sont effectuées au sonar. Toute position est donnée avec une zone d'incertitude due aux précisions des différents capteurs utilisés (*cf.* figure 1.18).

Si le drone est équipé de capteurs vidéos, lorsque la portée visuelle est atteinte (*cf.* figure 1.18), on effectue l'étape d'identification vidéo (*cf.* ligne 3 de la figure 1.17). Cette étape consiste à détecter l'objet recherché dans l'image optique reçue, enregistrée par une caméra vidéo. Lorsque l'objet est détecté, on classe et identifie cet objet. Une décision est prise sur le type de l'objet, sa position et son orientation.

Ensuite le drone communique avec le bateau-mère pour confirmer la décision. Si plusieurs contacts ont été fournis au drone, le drone visite tous les contacts avant de retourner au chasseur de mines. Si le drone a une mission de neutralisation, il dépose sa charge à côté de la mine ou se fait exploser de façon à détruire la menace si elle est avérée.

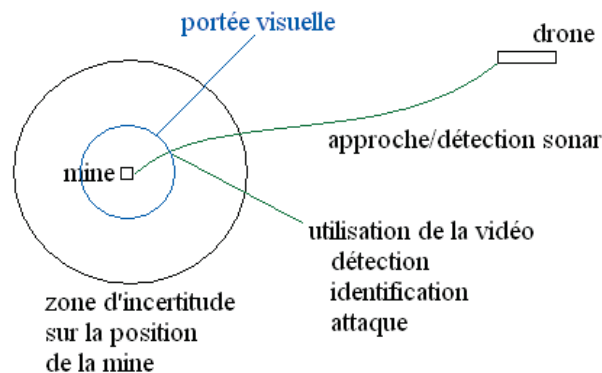


FIGURE 1.18 – Schéma représentant la mission d'un drone d'identification - neutralisation

1.5.4 L'utilisation des capteurs

Le drone d'identification a besoin de capteurs pour remplir sa mission. Tout d'abord il doit se positionner dans l'espace. Pour cela il utilise le sondeur acoustique, le capteur de pression ainsi que la centrale inertielle (voir partie 1.2.4.1 sur les capteurs de positionnement). Ensuite le drone utilise des capteurs de détection. Nous nous plaçons dans le cas où il embarque un sonar et une caméra vidéo optique (voir partie 1.2.4.2 sur les capteurs de détection). Dans l'optique des systèmes du futur, la communication embarquée à bord du drone est de préférence une communication acoustique.

Les données de positionnement mises à notre disposition par ces capteurs seront utilisées dans les travaux de cette thèse comme information à priori pour les traitements et la prise de décision. Sur cette partie de nos travaux, nous sommes en rupture avec les travaux précédents [8, 45–47]. De plus, l'utilisation des données de navigation conditionne les capteurs à utiliser si ce projet est amené à être implémenté sur un drone.

1.6 Conclusion

Dans ce chapitre, nous avons présenté le contexte dans lequel s'intègre cette thèse. Notre premier travail a consisté à étudier les différentes techniques utilisées dans le traitement d'images sous marines : les prétraitements, la détection d'objets, l'identification. Parmi toutes les techniques que nous avons répertoriées nous avons choisi celles qui nous semblaient le mieux adapté à notre sujet. Nous allons dans la suite de ce manuscrit présenter nos recherches, les méthodes utilisées et les résultats obtenus ainsi que nos contributions.

Chapitre 2

Les prétraitements

Sommaire

2.1 État de l'art	27
2.1.1 Les prétraitements basés sur la physique	28
2.1.2 Les prétraitements empiriques	30
2.2 Méthodes utilisées	31
2.2.1 Le bruit électronique	32
2.2.2 Le temps de calcul	34
2.2.3 Filtrage du bruit	35
2.2.4 Utilisation du modèle de Phong	35
2.2.5 La détection des contours	38
2.2.6 Les prétraitements annexes	41
2.3 Résultats et conclusion	43
2.3.1 Les images Topvision	43
2.3.2 Les images VAMA	44

Nous avons vu dans la partie 1.2 que nous souhaitons utiliser un système optique pour la détection de mines sous marines. Les scénarios dans lesquels nos images ont été enregistrées nous placent dans des eaux peu profondes (une trentaine de mètres). Nous avons montré que le milieu sous-marin joue sur la qualité visuelle des images. L'absorption est le phénomène qui influe le plus sur la lumière : les couleurs sont modifiées avec la profondeur et le coefficient d'absorption dépend de la longueur d'onde. De plus la turbidité de l'eau augmente ce coefficient et la portée visuelle en est réduite (la position géographique du lieu de la prise de vue est donc à prendre en compte). Enfin les conditions horaires ont leur importance sur la qualité des images. En effet, la lumière du soleil augmente la visibilité tandis que la nuit une source de lumière artificielle est nécessaire. Le traitement de ces différents phénomènes ont fait l'objet de nombreuses recherches (voir partie 2.1) dont quelques unes sont détaillées ici.

Lorsque l'on cherche à reconnaître un objet, la première idée est d'étudier les contours de ces objets. Dans un premier temps, nous avons souhaité suivre cette idée. Or sur les images sous marines, les contours des objets ne se distinguent pas fortement du reste de l'image. Nous avons orienté nos recherches de prétraitements dans l'optique d'améliorer le contraste et les contours des objets.

2.1 État de l'art

Dans cette partie nous allons nous intéresser plus particulièrement aux prétraitements qui ont été utilisés dans le milieu sous marin et dans des projets semblables au nôtre. Parmi ces méthodes, nous pouvons distinguer les prétraitements basés sur la physique et les prétraitements empiriques. Dans la première catégorie, nous avons répertorié la déconvolution, la correction colorimétrique et la polarisation. Dans la seconde catégorie, nous avons étudié le rehaussement

de contours. Il faut noter que de nombreuses méthodes de prétraitement ont été utilisées sur les images terrestres mais toutes n'ont pas été appliquées au milieu sous marin. Nous allons nous intéresser à quelques unes de ces méthodes et les adapter à notre problème.

2.1.1 Les prétraitements basés sur la physique

2.1.1.1 La déconvolution

Parmi les méthodes de débruitage, nous avons choisi de présenter la déconvolution. Cette méthode a pour but de corriger l'un des termes de bruit : la diffusion avant qui consiste à rendre l'image floue.

Nous avons vu dans la partie 1.2.1 que l'image reçue $I_{\text{reçue}}$ peut s'écrire [11, 12] :

$$I_{\text{reçue}} = I_{\text{atténuée}} + I_{\text{diffusion avant}} + I_{\text{rétrodiffusée}} \quad (2.1)$$

où $I_{\text{atténuée}}$ représente l'image composée de rayons lumineux qui proviennent directement de l'objet, $I_{\text{diffusion avant}}$ représente la partie du rayon lumineux qui a été déviée par la diffusion avant et $I_{\text{rétrodiffusée}}$ représente la partie du rayon lumineux qui a été rétrodiffusée par des particules.

On se place dans l'hypothèse où la scène imagée est plane et parallèle au capteur. De plus on considère un capteur avec une faible ouverture, ce qui nous permet d'être dans l'approximation des petits angles et d'avoir une distance capteur-scène constante. Selon ces hypothèses, l'équation précédente peut aussi s'écrire [12] :

$$I_{\text{reçue}} = I(0)e^{-(a+d)z} + \left(e^{-Gz} - e^{-(a+d)z} \right) I(0) * TF^{-1} \left(e^{-Bzf_{\theta}} \right) + I_{\text{rétrodiffusée}} \quad (2.2)$$

où :

- $I(0)e^{-(a+d)z}$ représente la composante atténuée (*cf.* équation 1.1, partie 1.2.1). I représente l'intensité à la position 0 (en cd), z la distance (en m), a le coefficient d'absorption (en m^{-1}) et d le coefficient de diffusion (en m^{-1}).
- $\left(e^{-Gz} - e^{-(a+d)z} \right) I(0) * TF^{-1} \left(e^{-Bzf_{\theta}} \right)$ représente la diffusion avant. L'expression de la diffusion avant correspond à l'image de la scène à la distance 0 multipliée par la différence de termes d'atténuation (termes dus au volume entre l'objet et l'observateur $e^{-(a+d)z}$ et à l'atténuation due à l'objet e^{-Gz} avec G le coefficient d'atténuation de l'objet, en m^{-1}). L'ensemble est convolué à la fonction de transfert de l'eau $TF^{-1} \left(e^{-Bzf_{\theta}} \right)$ où B est une constante empirique liée à la diffusion et f_{θ} la fréquence angulaire spatiale et TF la transformée de Fourier.
- Pour calculer la composante rétrodiffusée, on considère que l'espace entre l'objet et l'observateur est composé de fines couches. Chaque couche atténue et diffuse la lumière. La lumière provenant de chaque couche est convoluée avec la fonction de transfert de l'eau. La composante rétrodiffusée par le volume correspond à la somme de la lumière rétrodiffusée par toutes les couches.

La déconvolution consiste à estimer la fonction de transfert de l'eau entre l'observateur et l'objet. Cette méthode a largement été utilisée, par exemple dans les articles suivants [12, 45, 48, 49]. On trouve dans cette littérature, des expressions mathématiques pour la fonction de transfert de l'eau [12, 48].

Cette méthodes possède plusieurs inconvénients. Tout d'abord les différents modèles utilisés ne prennent pas en compte l'effet de la caméra sur les images. Un a priori sur la position de l'objet est nécessaire afin de connaître la longueur du trajet lumineux ainsi que sur les coefficients d'atténuation et de diffusion. Cette distance peut être obtenue par des capteurs (laser, stéréovision), par des modèles basés sur les images [50] ou par la "déconvolution aveugle" [45]. La "déconvolution aveugle" considère que la scène étudiée est plane. Ensuite on estime la distance et les paramètres d'atténuation de façon itérative en cherchant à maximiser le contraste après la déconvolution. La déconvolution est une méthode sensible au bruit. Mais le principal inconvénient est que cette méthode ne corrige que le terme de diffusion avant, qui ne représente qu'une faible partie des effets de l'eau sur l'intensité [51]. Cette méthode ne permettra pas de corriger entièrement nos images. Nous n'avons donc pas utilisé cette méthode.

2.1.1.2 La correction colorimétrique

Nous avons vu dans la partie 1.2.1 que les couleurs sont atténuées. Ainsi les images sous marines sont souvent dans les tons bleu-vert. C'est le cas des images Topvision. Nous nous sommes renseignés sur les méthodes traitant la couleur des images.

Plusieurs études ont été menées pour restaurer les couleurs dans les images sous marines. Certaines sont basées sur la loi de Beer-Lambert. D'autres égalisent l'histogramme des canaux colorimétriques, dans les domaine RGB "*Red, Green, Blue*", HSI "*Hue, Saturation, Intensity*".

Helan *et al.* [47] cherchent à corriger les couleurs présentes dans les images à partir du coefficient d'atténuation présent dans la formule de Beer-Lambert (*cf.* équation 2.3). Ils estiment les coefficients d'absorption pour le rouge, le vert et le bleu. Ils corrigent ainsi chaque couleur.

$$I(z) = I(0)e^{-a(\lambda)z} \quad (2.3)$$

où I est l'intensité en 0 et à une distance de z (en m) et a est le coefficient d'absorption (en m^{-1}) qui dépend de la longueur d'onde λ (en m).

La méthode Retinex consiste à calculer la clarté relative le long d'un chemin, en calculant le ratio des réponses visuelles entre les points x et $x + 1$ (deux points successifs du chemin sur lequel est calculée la clarté relative) [52]. A l'aide de cette clarté relative, les nouvelles intensités des canaux R, G ou B sont calculées comme étant la moyenne des clartés sur plusieurs chemins. La méthode ACE, "*automatic color equalization*" corrige aussi les valeurs des canaux R, G et B [52]. La différence d'intensité entre l'intensité au pixel p et au pixel j est calculée, et sert à calculer la valeur de la fonction représentant la clarté relative. Cette valeur est pondérée par une distance représentant la contribution des deux pixels. Ces deux méthodes, comparées dans [52], présentent des résultats proches. La méthode ACE a été testée pour la reconnaissance des poissons à l'aquarium de La Rochelle et semble donner de bons résultats [53].

Torres-Mendez *et al.* [54] utilisent un jeu d'images d'apprentissage découpé en imagerie, dont les couleurs sont correctes. Ils découpent l'image reçue en petites imagerie. A l'aide d'un réseau de Markov, ils essaient d'associer des imagerie apprises avec les imagerie issues de l'image cible. En cas de convergence, ils remplacent la couleur de l'imagerie reçue par la couleur de l'imagerie apprise.

Iqbal *et al.* [55] égalisent le contraste de l'image RGB, en modifiant les histogrammes des canaux rouge et vert afin qu'ils soient proches de l'histogramme du canal bleu. Ensuite ils changent d'espace de couleur pour travailler dans le repère HSI où ils modifient la saturation et l'intensité de l'image.

Ces méthodes semblent donner de bons résultats d'après la littérature. Cependant, dans nos vidéos, la caméra peut se situer loin de la mine et enregistrer des images du volume d'eau comme elle peut se situer très près du sol et de la mine et enregistrer des images d'objets. Pour utiliser ces méthodes, il faudrait avoir à disposition un grand choix d'images d'apprentissage pour créer des imagerie correspondant à toutes les situations et pour estimer les paramètres pour corriger la loi de Beer-Lambert ou les histogrammes. De plus nous aimerions obtenir un algorithme qui fonctionne aussi bien sur les images Topvision (en couleurs) et sur les images VAMA (en niveaux de gris)

2.1.1.3 La polarisation

La polarisation est une méthode utilisée dans le milieu sous marin pour acquérir des images depuis les années 1960. Le milieu sous marin est un milieu qui diffuse et atténue la lumière. Une étude a montré que l'utilisation de la polarisation circulaire permet d'augmenter le contraste et la distance de visibilité de la cible [56]. Ces résultats ont été confirmés par Lewis *et al.* [57], qui ont montré que pour un contraste donné, la distance de visibilité avec une lumière polarisée peut être doublée par rapport à la distance obtenue avec un système qui n'utilise pas la polarisation [57].

Mathématiquement, il est possible de décrire l'état de la polarisation de la lumière par un

vecteur de Stokes A [58] :

$$A = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} \quad (2.4)$$

où I représente l'intensité totale, Q , U et V représentent les intensités dans les modes de polarisation horizontal, vertical, à $+45^\circ$, -45° et circulaires droite et gauche [58]. La transformation de la polarisation par le milieu traversé est représentée par la matrice de Mueller M [58] :

$$M = \begin{pmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{31} & M_{32} & M_{33} & M_{34} \\ M_{41} & M_{42} & M_{43} & M_{44} \end{pmatrix} \quad (2.5)$$

L'état de polarisation de la lumière après passage dans le milieu se calcule de la façon suivante [58] :

$$A' = MA \quad (2.6)$$

Donc si l'action du milieu est connue, représentée par la matrice de Mueller et l'état de la lumière, il est possible de calculer l'intensité et la polarisation de la lumière résultante après traversée du milieu.

Les images que nous avons à disposition sont des images enregistrées avec une caméra non polarisée. Nos images correspondent donc au terme d'intensité du vecteur de Stokes. Dans notre étude nous avons rencontré des problèmes d'atténuation et de visibilité. L'utilisation des images polarisées peut constituer une piste d'amélioration des images et diminuer les prétraitements nécessaires.

2.1.2 Les prétraitements empiriques

2.1.2.1 Le rehaussement de contours

Garcia *et al.* [59] montrent que le rehaussement doit se faire localement et non globalement car la dynamique de l'image varie trop fortement dans l'image. Pour cela, ils comparent plusieurs méthodes. Tout d'abord ils essaient puis abandonnent les méthodes basées sur l'égalisation d'histogramme, qui demandent un temps de calcul très élevé et ne sont pas efficaces sur toute l'image. Ensuite ils travaillent en approximant la luminosité moyenne locale de l'image. Pour cela, ils présentent deux méthodes :

- diviser l'image reçue par sa version filtrée avec un filtre passe-bas, I_{LP} . Cette méthode se réfère au modèle "illumination-reflectance" qui considère l'image reçue comme le produit de la reflectance de la scène par l'illumination locale (*cf.* équation 1.3) :

$$I_{re\grave{c}ue} = I_{\acute{e}mise} \cdot h \quad (2.7)$$

En effet, l'image basses fréquences, I_{LP} peut être comparée à h , l'illumination. Si l'image reçue est divisée par cette image basses fréquences, l'image émise est obtenue.

- soustraire à l'image reçue l'image I_{LP} . Cette deuxième méthode se réfère au modèle additif des trois composantes de l'image (composante atténuée, composante diffuse et composante rétrodiffusée, *cf.* équation 1.2)

$$I_{re\grave{c}ue} = I_{att\acute{e}nu\acute{e}e} + I_{diffusion\ avant} + I_{r\acute{e}trodiffus\acute{e}e} \quad (2.8)$$

L'image basses fréquences, I_{LP} peut aussi être comparée à l'image rétrodiffusée.

Garcia *et al.* constatent empiriquement qu'il vaut mieux diviser que soustraire, tout en privilégiant l'implémentation par filtrage homomorphe.

Arnold-Bos *et al.* [60] montrent dans un premier article que si l'on ajoute une étape de filtrage adaptatif du bruit (filtre qui s'adapte en fonction de l'intensité du bruit local et de l'information locale), l'image obtenue est améliorée. Pour quantifier la bonne visibilité des contours, ils créent un critère semi-empirique basé sur les statistiques du gradient. Dans un deuxième article [51], ils apportent une justification théorique à l'approche de Garcia *et al.* [59] et montrent à l'aide d'une modélisation théorique que selon les zones de l'image, la méthode de la soustraction permet d'obtenir de meilleurs résultats que la méthode utilisant la division. Pour déterminer quelle méthode utiliser, ils utilisent le rapport S du gradient de l'image divisée I_d sur le gradient de l'image issue de la soustraction I_s , rapport dont ils ont démontré qu'il est égal à l'image filtrée par un passe-bas, I_{LP} suivie par une normalisation *ad hoc*. Finalement pour créer une image dont les contours ont été rehaussés, ils proposent une moyenne pondérée par S de l'image divisée et de l'image soustraite.

La méthode est résumée avec la figure 2.1. Nous l'avons appliquée à nos images. Appelons I l'image à débruiter. Sur la figure 2.1, nous avons pris une image Topvision que nous avons transformée en niveaux de gris. Ensuite nous avons coupé les écritures (*cf.* partie 2.2.6.1) et réduit sa taille à 96×128 (*cf.* partie 2.2.2). Pour appliquer la méthode décrite ici, nous avons créé une image I_{LP} qui correspond à l'image I filtrée par un filtre passe-bas gaussien. Cette image I_{LP} contient alors les composantes basses fréquences de l'image I . Dans le modèle "illumination-reflectance", cette image estime l'illumination locale de l'image. Dans le modèle additif de formation des images, l'image estime la composante rétrodiffusée. En soustrayant l'image I_{LP} à l'image I , l'image I_s est obtenue. En divisant l'image I_{LP} à l'image I , l'image I_d est obtenue. Après l'opération de division ou de soustraction, l'histogramme est égalisé en bornant la plage de valeurs et les valeurs de l'image sont mises entre 0 et 1 :

$$I = \begin{cases} \text{mini} & \text{si } I(i, j) < \text{mini} = 1.05 \times \min(I) \\ \text{maxi} & \text{si } I(i, j) > \text{maxi} = 0.95 \times \max(I) \\ I(i, j) & \text{dans les autres cas} \end{cases} \quad (2.9)$$

Ensuite, afin de prendre en compte la valeur la plus importante entre les 2 images, une image débruitée est créée, moyenne pondérée des images I_s et I_d :

$$I_{\text{débruitée}} = \frac{I_{LP} \times I_s + \frac{I_d}{I_{LP}}}{I_{LP} + \frac{1}{I_{LP}}} \quad (2.10)$$

La figure 2.1 montre cette image avec les contours rehaussés. La mine ressort davantage par rapport au fond que dans l'image avant rehaussement de contrastes (*cf.* figure 2.1).

Bazeille *et al.* [46] reprennent le modèle de la division de Garcia *et al.* [59] puis améliorent en suivant Arnold-Bos *et al.* [60], en utilisant des filtres homomorphiques, à ondelettes et anisotropiques. Leur approche traite aussi le moiré présent dans les images.

Nous utiliserons par la suite la méthode de rehaussement des contours sur nos images pour améliorer la détection des contours des objets. En effet, comme le montre la figure 2.2, les contours obtenus par l'image de phase sont plus visibles dans le cas de l'image après rehaussement de contours.

2.2 Méthodes utilisées

Lors de la visualisation des images Topvision et VAMA, nous avons identifié plusieurs problèmes que nous allons essayer de minimiser voire de résoudre (*cf.* figure 2.3). Les principaux problèmes identifiés sur les images Topvision sont la présence des informations de navigation sur les images, la présence de moiré (partie du bruit électronique) plus ou moins accentuée, un faible contraste. Les problèmes identifiés sur les images VAMA sont l'image d'un manchon présent sur

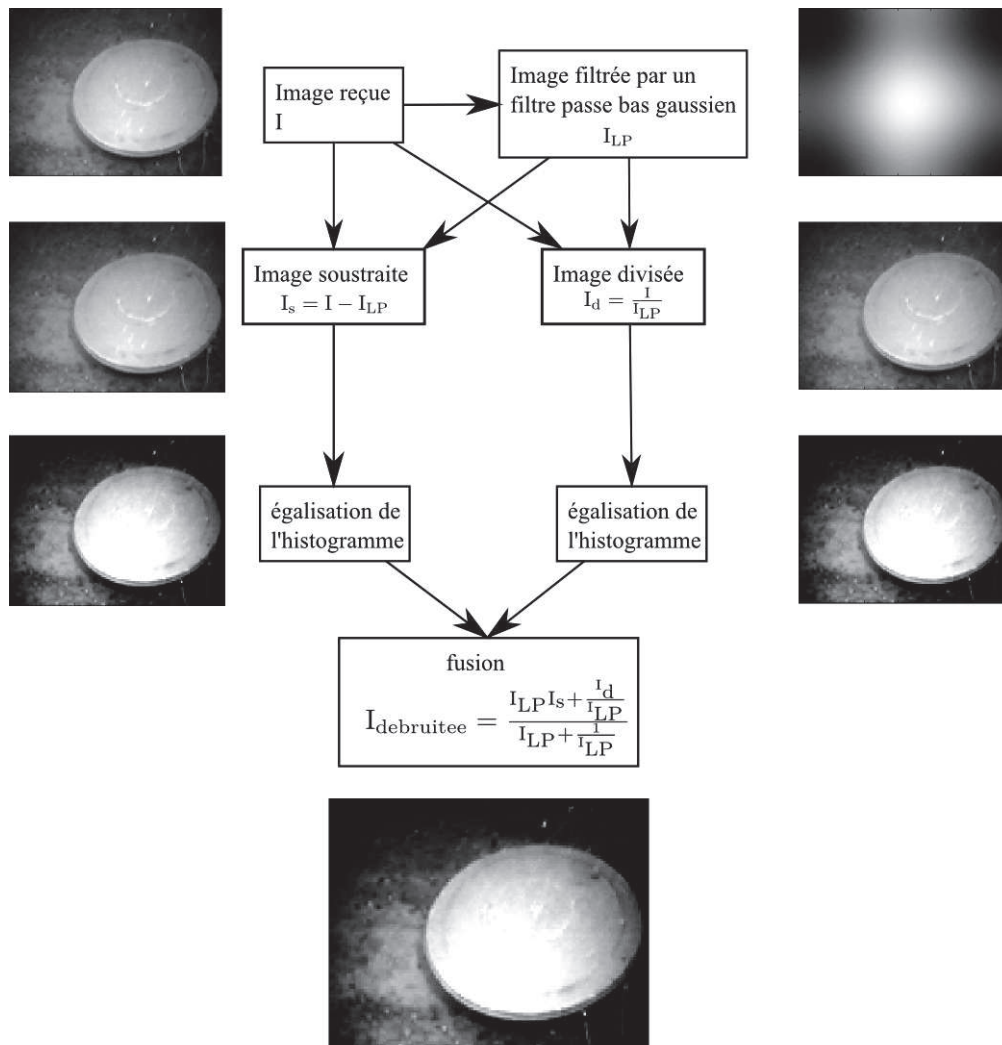


FIGURE 2.1 – Algorithme de rehaussement des contours

les bords de l'image, un faible contraste et un problème de saturation des images dû à la réflexion spéculaire du soleil sur les objets ou à la réflexion sur le sol clair. Excepté le rehaussement des contours (traité partie 2.1.2.1), ces différents problèmes vont être étudiés dans cette partie. Nous commençons par présenter les prétraitements basés sur la physique avant de présenter les autres prétraitements.

2.2.1 Le bruit électronique

La chaîne d'acquisition ajoute du bruit électronique sur les images. Le type de bruit présent sur les images peut provenir des électrons, du capteur ou de la chaîne d'acquisition. Les électrons sont responsables du bruit thermique (bruit généré par l'agitation des électrons qui augmente avec la température du capteur) et du bruit de grenaille (bruit causé par les fluctuations du nombre d'électrons qui arrivent sur les capteurs). Lorsque les éléments du capteur sont défectueux, du bruit poivre et sel est obtenu (auss appelé bruit impulsif). Ce bruit se produit aussi lors de problèmes dans la transmission des données le long de la chaîne d'acquisition. Le bruit de quantification peut aussi être perceptible. L'effet de moiré peut aussi être obtenu. Il se produit entre les motifs de l'image et la trame du capteur.

Les images Topvision présentent un effet de moiré (*cf.* figure 2.4). Cet effet est dû au bruit ajouté par la chaîne d'acquisition lors de l'enregistrement des images. L'image enregistrée est alors plus ou moins hachurée selon l'intensité du bruit.

Une première solution pour limiter cet effet est celle développée par Sidorov et Kokaram [61]. Cette méthode consiste à étudier le spectre de l'image à traiter. En considérant que l'effet de

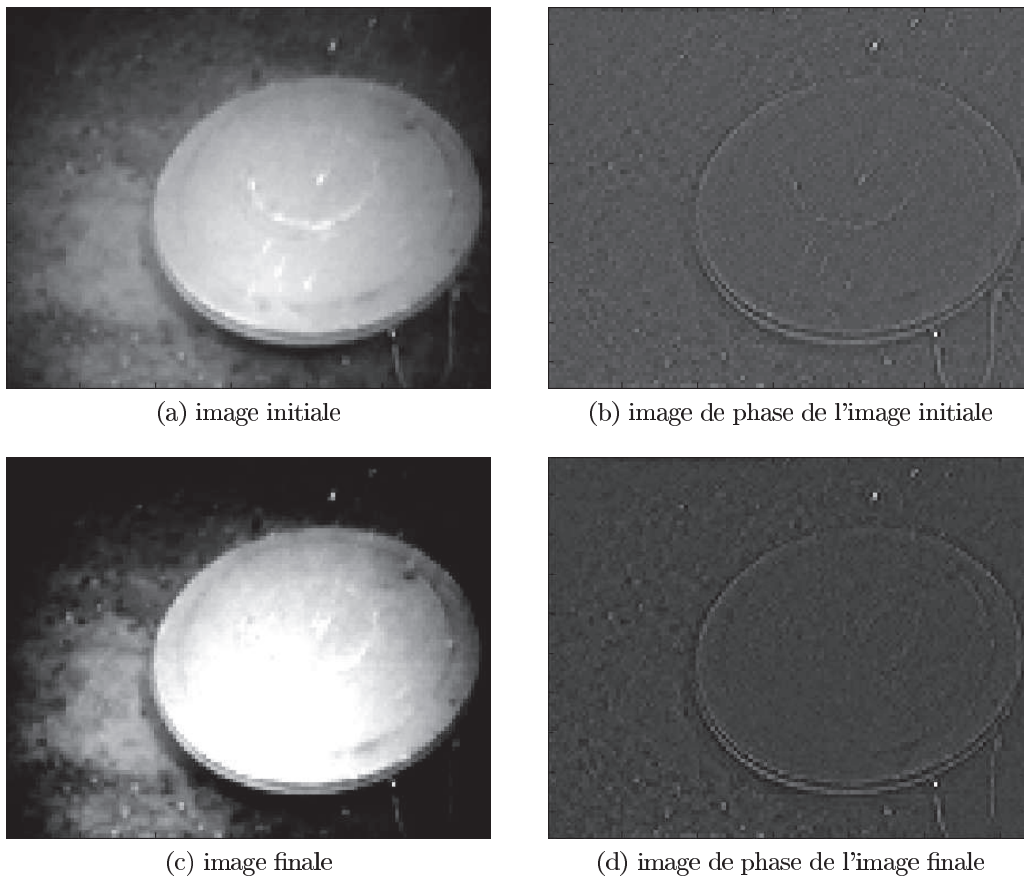


FIGURE 2.2 – Effet sur l'image de phase du rehaussement de contours

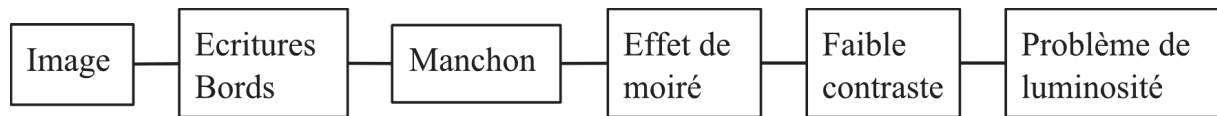


FIGURE 2.3 – Problèmes identifiés dans nos images



FIGURE 2.4 – Image initiale

moiré se matérialise par l'addition d'un motif périodique à l'image initiale, leur idée est de détecter les pics du spectre de l'image dans le domaine de Fourier. Ensuite ils les suppriment en considérant que ces pics correspondent à l'effet de moiré. En effet, plus le motif est périodique (et proche d'une fonction sinus) plus le spectre correspondant se rapprochera d'un pic de Dirac

(et de la transformée de Fourier d'une fonction sinus).

Dans un premier temps nous avons appliqué une méthode proche de la méthode de Sidorov et Kokaram [61]. Nous avons considéré que l'expression mathématique de l'effet de moiré en 1D est la suivante :

$$m(t) = \cos(\omega t) \quad (2.11)$$

De plus nous avons considéré qu'il s'agit d'un bruit additif. L'effet de moiré est donc présent dans le spectre de l'image sous forme de deux pics situés en ω et en $-\omega$. Nous avons ensuite étudié les images et leur spectre. Cela nous a permis de constater que l'information correspondant à l'effet de moiré semble se situer sur les côtés du spectre. Nous avons donc défini deux zones de recherche dans les images et ensuite cherché les pics d'amplitude que nous avons supprimé, de façon analogue à celle utilisée dans [61]. L'effet de moiré n'est plus visible sur l'image modifiée (*cf.* figure 2.5), comparativement à l'image initiale (*cf.* figure 2.4). Nous avons donc adapté cette méthode pour supprimer le problème de l'effet de moiré [62]. De plus, nous envisageons pour la suite de nos travaux de travailler dans le domaine de Fourier. Ce prétraitement dans le domaine de Fourier est donc intéressant car il n'ajoute pas de transformations supplémentaires à notre image.

Une autre méthode plus simple et moins couteuse a été testée. Elle consiste à sous échantillonner l'image. En effet, la méthode utilisée par défaut par Matlab est l'interpolation bicubique. Il s'agit de remplacer un pixel par la moyenne pondérée des pixels dans un voisinage de 4×4 pixels autour du pixel considéré. L'effet de moiré est alors atténué par cette moyenne (*cf.* figure 2.6).

Le temps de calcul est une contrainte importante de ce projet. De plus, l'effet de moiré n'est pas présent et n'a pas la même intensité sur toutes les images. Par conséquent, nous privilégions par la suite la suppression de l'effet de moiré par sous-échantillonnage.



FIGURE 2.5 – Image dont l'effet de moiré a été supprimé avec la méthode proche de celle de Sidorov et Kokaram

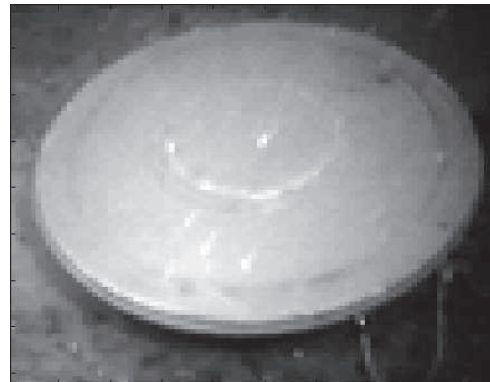


FIGURE 2.6 – Image initiale (figure 2.4) sous échantillonnée

2.2.2 Le temps de calcul

Un des objectifs de notre étude est d'obtenir un programme fonctionnant en temps réel. Un des facteurs importants à prendre en compte est la taille des images. Plus une image sera de grande taille plus les calculs vont demander du temps. La taille des images est importante. Le sous échantillonnage est une méthode qui permet de limiter ce temps de calcul et de répondre au problème de l'effet de moiré quand il est présent. Ainsi, nous avons décidé de redimensionner nos images pour obtenir des images de taille 96×128 pixels.

Prenons l'exemple de la transformée de Fourier sur les images Topvision pour justifier l'utilisation du sous-échantillonnage. Le premier test consiste à lire l'image et faire une transformée de Fourier. Le second test consiste à lire l'image, sous-échantillonner l'image puis appliquer la transformée de Fourier. Les tests ont été effectués avec Matlab sur le PC ISEN (*cf.* annexe E).

- temps de calcul moyen du test 1 sur 100 images : 0,0504 s/im
- temps de calcul moyen du test 2 sur 100 images : 0,0344 s/im

L'écart entre les deux tests est de 0,0160 s/im. L'écart est de 0,0586 s/im si l'on effectue une transformée de Fourier et une transformée de Fourier inverse. Ce résultat est très important pour une seule image, d'autant plus que nous avons besoin de réaliser une transformée de Fourier et une transformée de Fourier inverse. En effet, dans la suite de nos travaux, nous allons comparer l'image traitée à une grande quantité d'images connues. Pour comparer l'image traitée aux images connues, nous aurons besoin d'une transformée de Fourier et de son inverse par image connue testée. Le sous-échantillonnage permettra alors de réduire le temps de calcul de façon significative par rapport au même programme sans l'étape de sous-échantillonnage.

2.2.3 Filtrage du bruit

Nos images présentent un peu de bruit. Pour limiter ce bruit, sans altérer les contours des objets lorsqu'ils sont présents, nous avons appliqué un filtre médian à nos images. Nous pouvons constater sur la figure 2.7 que les particules en suspension présentes dans l'image ne le sont plus. De plus les contours sont effectivement préservés.

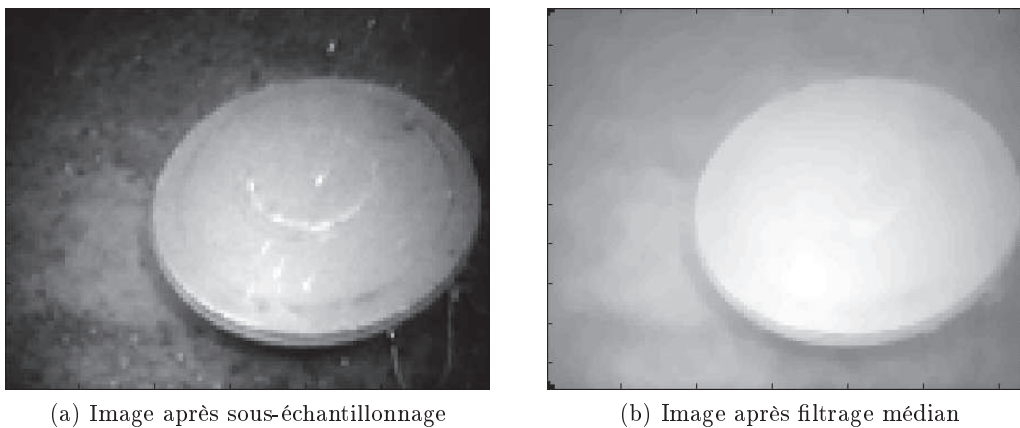


FIGURE 2.7 – Effet du filtrage médian sur une image

2.2.4 Utilisation du modèle de Phong

Lorsque les images sous marines sont enregistrées à faible profondeur, comme c'est le cas pour les images VAMA, le soleil se réfléchit sur le sol et les objets. Certaines zones de l'image ont alors une intensité plus forte que le reste (voire sont saturées), provoquant la détection d'un objet de taille erronée voire une fausse détection. Pour limiter cet effet, nous avons utilisé le modèle de Phong pour retrouver l'image émise.

2.2.4.1 Le modèle de Phong

Selon Phong [63], la lumière reçue par l'observateur est la combinaison de la composante ambiante, de la composante diffuse et de la composante spéculaire de la lumière. L'absorption est aussi à prendre en compte surtout dans notre étude où l'absorption de l'eau influe plus ou moins fortement sur l'image reçue.

La composante ambiante est considérée comme une constante, qui représente la lumière qui arrive sur notre scène à partir d'objets transformés en sources secondaires. La scène vue par l'observateur réfléchit une certaine quantité de lumière, appelée $I_{\text{émission}}$ par l'objet. Cette lumière réfléchie est ensuite soumise à l'absorption du milieu. L'observateur voit la quantité de lumière appelée $I_{\text{reçue}}$. L'absorption, la diffusion et la réflexion spéculaire influent sur l'image émise selon

les lois :

$$\begin{aligned}
 I_{\text{émise par l'objet}} &= I_{\text{ambiante}} + I_{\text{diffuse}} + I_{\text{spéculaire}} \\
 I_{\text{émise par l'objet}} &= I_{\text{ambiante}} + \cos(\theta_s)I_{\text{émise}} + \cos^\alpha(\Omega)I_{\text{émise}} \\
 I_{\text{reçue}} &= e^{-c \cdot z}I_{\text{émise par l'objet}}
 \end{aligned}
 \tag{2.12}$$

où Ω représente l'angle formé entre la réflexion de la source de lumière et l'observateur, θ_s représente l'angle formé entre la normale à la surface observée N et le soleil (*cf.* figure 2.8) et α permet de modéliser la brillance du matériau (plus α est grand, plus la surface est brillante et plus la tache lumineuse due à la composante spéculaire sera petite). L'expression de l'intensité émise par l'objet peut aussi s'écrire sous forme de produit scalaire (*cf.* équation 2.13). Le vecteur \vec{N} représente le vecteur associé à la normale à la surface. Le vecteur \vec{O} représente le vecteur associé à la direction surface-observateur. Le vecteur \vec{L} représente le vecteur source lumineuse-surface. Le vecteur \vec{R} représente le vecteur correspondant au rayon lumineux réfléchi.

$$I_{\text{émise par l'objet}} = I_{\text{ambiante}} + (-\vec{L}\vec{N})I_{\text{émise}} + (\vec{R}\vec{O})^\alpha I_{\text{émise}} \tag{2.13}$$

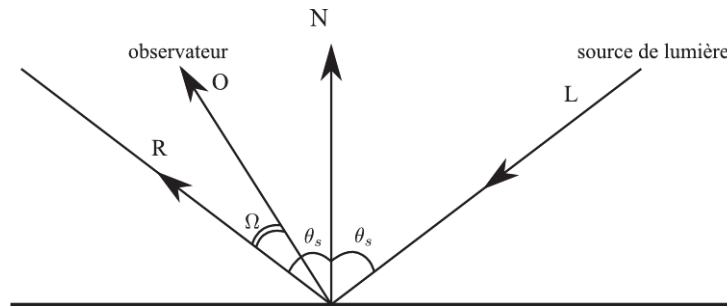


FIGURE 2.8 – Définition des angles utilisés

Ce modèle n'est applicable aux images que si l'on possède des informations sur la position géographique de l'observateur, de l'objet et de la source lumineuse.

2.2.4.2 L'absorption

Sous l'eau, la portée visuelle est limitée par l'absorption qui s'exprime selon la loi de Beer-Lambert [47] :

$$I_{\text{reçue}}(z) = e^{-c \cdot z}I_{\text{émise}} \tag{2.14}$$

où c représente le coefficient d'absorption de l'eau et z la distance, dans notre cas, entre l'observateur et l'objet à identifier.

Pour mesurer le coefficient d'absorption de l'eau, en pratique, le disque de Secchi est utilisé (*cf.* figure 2.9) [64]. Il s'agit d'un disque noir et blanc plongé dans l'eau. On mesure la distance z à laquelle on ne peut plus distinguer le blanc du noir. L'intensité lumineuse est alors égale à la couleur de l'eau E :

$$I(z) = E \tag{2.15}$$



FIGURE 2.9 – Disque de Secchi vu de dessus

L'intensité reçue dépend alors de la couleur de l'eau et de l'intensité reçue de l'objet (un objet manufacturé sur le sol, le sable, un poisson), qui a subi l'absorption selon la loi de Beer-Lambert (*cf.* équation 2.14) :

$$I_{\text{reçue}}(z) = E + e^{-c \cdot z}I_{\text{émise par l'objet}} \tag{2.16}$$

Dans notre cas, nous avons étudié plusieurs vidéos afin de définir une valeur pour le coefficient d'absorption c . Pour cela, nous avons pris comme référence une mine d'une couleur inconnue M . L'équation d'absorption 2.16 est alors modifiée de la façon suivante :

$$I_{\text{reçue}} = E + (M - E)e^{-c.z} \quad (2.17)$$

Lorsque nous sommes contre l'objet, la distance z est nulle, l'intensité reçue est celle provenant de la mine. Lorsque nous sommes très loin et que la valeur de l'exponentielle est devenue négligeable, nous recevons uniquement la couleur de l'eau.

Nous avons cherché la distance observateur-objet à partir de laquelle, nous ne pouvons plus distinguer la mine. Sur les images VAMA, cette distance est de 20 m.

Pour une exponentielle de la forme $e^{-\frac{t}{\tau}}$, on estime que l'équilibre est atteint pour $t = 5\tau$, soit une valeur de l'exponentielle égale à 0.01. Dans notre cas, nous avons atteint cette zone de l'exponentielle pour une distance observateur-objet de 20 m. Nous pouvons alors calculer le coefficient d'absorption c :

$$\begin{aligned} z = 20 &= \frac{5}{c} \\ c &= 0.25 \text{ m}^{-1} \end{aligned} \quad (2.18)$$

Pour les images VAMA, nous disposons pour chaque image de la distance observateur-objet. Nous pouvons ainsi recalculer l'image émise à partir de l'image reçue et atténuer l'effet de l'absorption sur l'image.

2.2.4.3 La réflexion de la lumière

La réflexion de la lumière se décompose en deux parties : la réflexion spéculaire et la réflexion diffuse (aussi appelée diffusion).

La diffusion désigne la réflexion des rayons lumineux incidents dans toutes les directions de l'espace, par une surface irrégulière. L'intensité de la lumière diffusée est la même dans tout l'espace mais dépend du cosinus de l'angle formé entre la normale à la surface observée et la direction d'émission de la source lumineuse (loi de Lambert) (ici appelé θ_s , en cohérence avec la figure 2.8) :

$$I_{\text{reçue}} = \cos(\theta_s)I_{\text{émise}} \quad (2.19)$$

La réflexion spéculaire désigne la réflexion par une surface avec des défauts très petits (de taille inférieure à $\frac{\lambda}{4}$, λ étant la longueur de l'onde lumineuse incidente), par exemple une surface métallique. La lumière est réfléchiée dans une seule direction, selon la loi de réflexion. L'intensité de la réflexion spéculaire vue par l'observateur suit une loi en cosinus mais ce cosinus est élevé à la puissance α , où α représente la brillance du matériau (ne connaissant à priori pas le matériau dans lequel est fabriqué l'objet manufacturé, nous avons fixé α à 1). Cependant, on considère que la direction normale (direction où l'intensité est la plus forte) est la direction de réflexion de la lumière. L'observateur reçoit donc :

$$I_{\text{reçue}} = \cos^\alpha(\Omega)I_{\text{émise}} \quad (2.20)$$

où Ω représente le même angle que sur la figure 2.8.

2.2.4.4 La position du soleil

Pour retrouver l'image émise, nous avons donc besoin de la position du soleil et plus précisément de l'angle zénithal (l'angle du soleil par rapport à la verticale). Il existe un algorithme développé dans l'article de Reda et Andreas [65] qui permet de retrouver la position du soleil connaissant la latitude, la longitude, l'altitude du lieu où se situe l'observation ainsi que la date. Sur les données VAMA, nous avons accès à toutes ces informations, ce qui nous a permis d'utiliser cet algorithme.

L'algorithme demandant beaucoup d'étapes de calculs, le détail peut être trouvé dans l'article [65]. Nous indiquons ici les principales étapes nécessaire pour retrouver l'angle zénithal.

Tout d'abord nous avons besoin de transformer la date de prise de vue en jours juliens. Il s'agit d'un système de datation qui consiste à considérer un instant donné comme une fraction de jours écoulés depuis le 1^{er} janvier -4712 à 12h.

A partir de la date julien, nous calculons la position de la Terre dans le repère héliocentrique.

Ensuite nous changeons de repère pour se placer dans le repère géocentrique. A l'aide de la longitude et de la latitude de la Terre dans le repère héliocentrique, nous pouvons facilement calculer la longitude et la latitude du soleil dans le repère géocentrique. Nous pouvons aussi calculer l'ascension et le déclin du Soleil.

Enfin le passage du repère géocentrique au repère topographique permet de calculer l'angle zénithal.

2.2.4.5 Conclusion

Grâce aux informations fournies par les différents capteurs et les données sur l'enregistrement des images (date et lieu), nous avons pu appliquer le modèle de Phong sur nos images. Considérant que la caméra vise l'objet et connaissant l'angle d'ouverture de la caméra, nous avons pu déterminer la distance parcourue par les différents rayons du cône d'observation de la camera pour atteindre le plan d'observation (plan perpendiculaire aux rayons de la caméra et contenant l'objet). Cela nous permet de calculer des valeurs d'absorption et de réflexion spéculaire pour chaque pixel. Sur la figure 2.10, nous pouvons voir l'effet du modèle de Phong appliqué à une image et l'effet sur la détection d'objet (algorithme de détection présenté dans la suite de ce manuscrit). La mine n'était pas détectée, après application du modèle de Phong elle est détectée. L'image 2.11 montre l'exemple d'une image de sphère où l'application du modèle de Phong permet de supprimer les fausses alarmes.

L'algorithme de détection, pour les images VAMA est basé sur une comparaison entre l'image prétraitée et l'image correspondant à l'arrière plan. Dans le cas des mines sphériques, l'application du modèle de Phong permet de supprimer la détection fausse due au halo. Dans le cas des mines Manta, cet algorithme détecte toujours le halo lumineux. Ce problème sera commenté dans le chapitre consacré à la détection des mines (*cf.* chapitre 8, p. 111).

2.2.5 La détection des contours

Parmi les méthodes de détection et reconnaissance d'objets, plusieurs utilisent les contours (par exemple, la comparaison de motifs, la recherche de caractéristiques, la détection de coins, la segmentation en contours). Notre idée est de s'orienter vers ces méthodes là. Or, la détection de contours avec un filtre de Canny [66] n'est pas efficace sur les images sous marines. En effet, les objets sont pas assez contrastés par rapport au fond. Les contours peuvent aussi être occultés par un poisson, par un nuage de sable.

Nous avons, dans un premier temps, choisi d'utiliser ce que nous appelons les images de phase (*cf.* figure 2.12). Oppenheim [67] a montré que l'on peut retrouver les motifs importants d'une image à partir de la phase du spectre de l'image initiale. Pour obtenir cette image de phase, nous calculons le spectre de l'image initiale par transformée de Fourier. Ensuite nous fixons l'amplitude de ce spectre à 1 avant de revenir dans le domaine spatial via une transformée de Fourier inverse.

Nous avons appliqué les résultats d'Oppenheim [67] sur les images de Lena 2.13a et Barbara 2.14. Notre premier test a consisté à calculer l'image de phase de Lena (image reconstruite à partir de la phase du spectre de Lena). Le résultat est présenté sur la figure 2.13b. Nous pouvons constater que Lena est reconnaissable sur cette image. De plus cette image contient suffisamment de contours pour pouvoir appliquer les méthodes de traitement d'images basées sur les contours. Nous avons ensuite effectué un second test. Dans le domaine spectral, nous avons combiné la phase du spectre de Lena avec l'amplitude du spectre de Barbara. Nous avons aussi combiné la phase du spectre de Barbara avec l'amplitude du spectre de Lena. Les transformées de Fourier inverses de ces deux spectres combinés sont présentés sur la figure 2.14. Dans les deux cas,

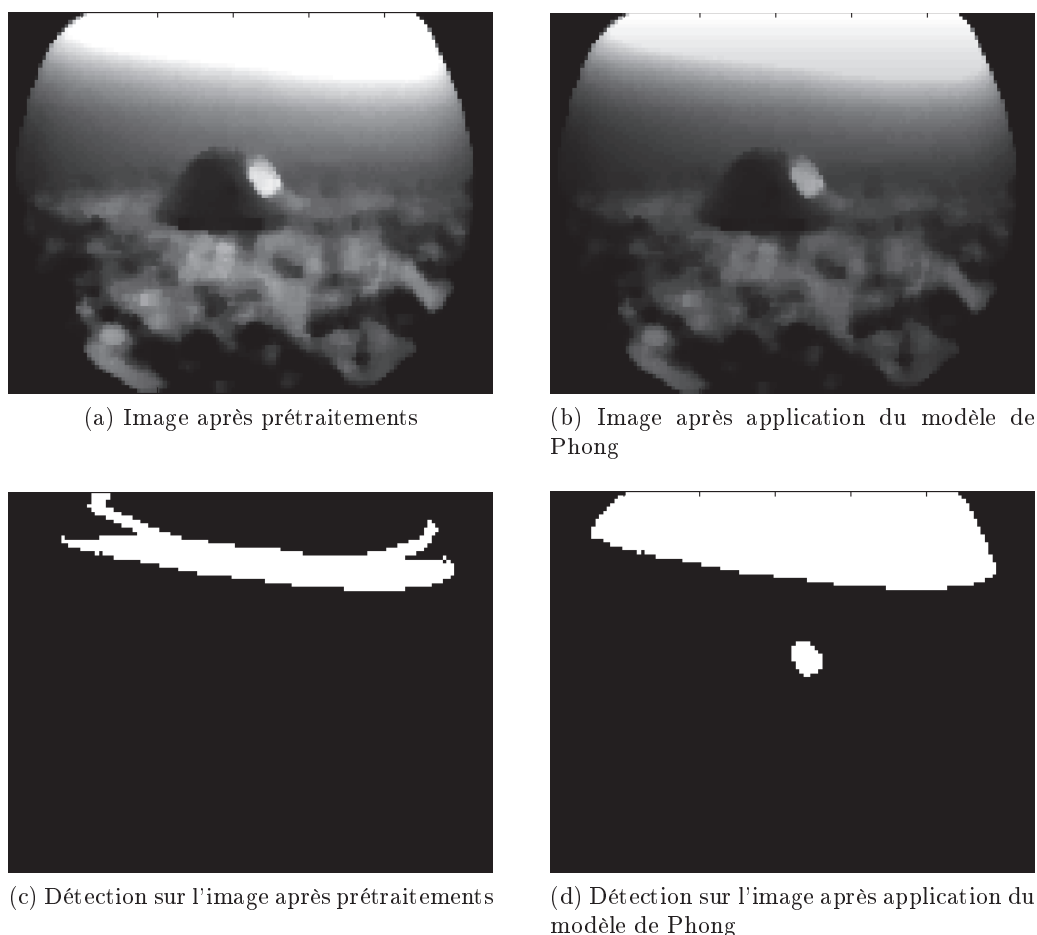


FIGURE 2.10 – Images avant et après application du modèle de Phong

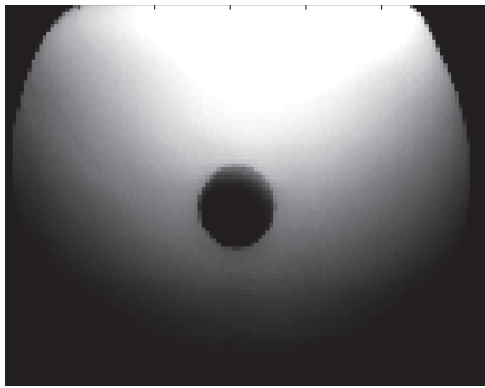
nous pouvons remarquer que la personne qui ressort le plus dans les images combinées sont les personnes dont nous avons utilisé la phase du spectre.

Cette méthode a l'avantage de s'effectuer dans le domaine de Fourier. Elle peut donc s'utiliser efficacement avec une méthode de reconnaissance nécessitant une transformée de Fourier (la corrélation par exemple).

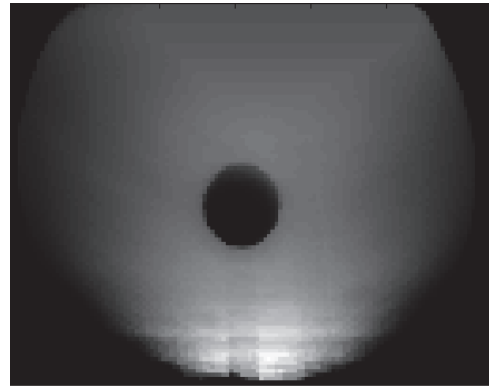
Sur les images Topvision, l'image de phase détecte relativement bien les contours puisque les contours ressortent dans l'image initiale (*cf.* figure 2.15). Cette méthode de détection de contours en vue d'une reconnaissance a fait l'objet d'une publication [62].

Cependant, sur les images VAMA, les objets ne sont pas contrastés et les couleurs sont proches de celles du sol (*cf.* figure 2.16). Nous avons donc cherché une autre solution. Notre choix s'est porté sur les images de contours via l'approche multi-échelle (*cf.* figure 2.17). Pour cela, nous nous sommes inspirés de la méthode de Canny [66]. Nous détectons les contours sur 2 images aux échelles 96×128 et 48×64 puis nous fusionnons les résultats obtenus. Des approches similaires ont été utilisées dans la littérature, basées sur le Laplacien [68] et sur la transformée en ondelette [69].

Nous avons pris l'image issue de la vidéo. D'une part nous effectuons les prétraitements nécessaires décrits dans cette partie (redimensionnement, sous-échantillonnage, rehaussement des contours). Ensuite nous appliquons un filtre de Canny afin de détecter les contours. D'autre part, nous sous-échantillons l'image prétraitée pour obtenir une image de taille 48×64 . Nous appliquons un filtre de Canny sur cette image puis nous sur-échantillons de façon à obtenir une image de taille 96×128 . La multiplication de ces deux images nous permet de ne garder que les contours communs aux deux images.



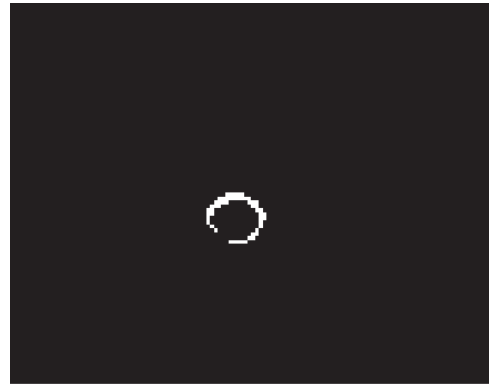
(a) Image après prétraitements



(b) Image après application du modèle de Phong



(c) Détection sur l'image après prétraitements



(d) Détection sur l'image après application du modèle de Phong

FIGURE 2.11 – Images avant et après application du modèle de Phong

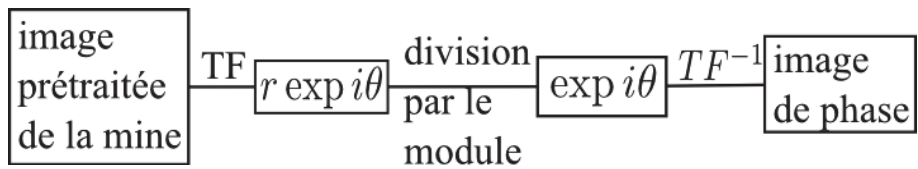


FIGURE 2.12 – Schéma expliquant la création d'une image de phase



(a) Lena



(b) Image de phase de Lena

FIGURE 2.13 – Lena et l'image de phase de Lena

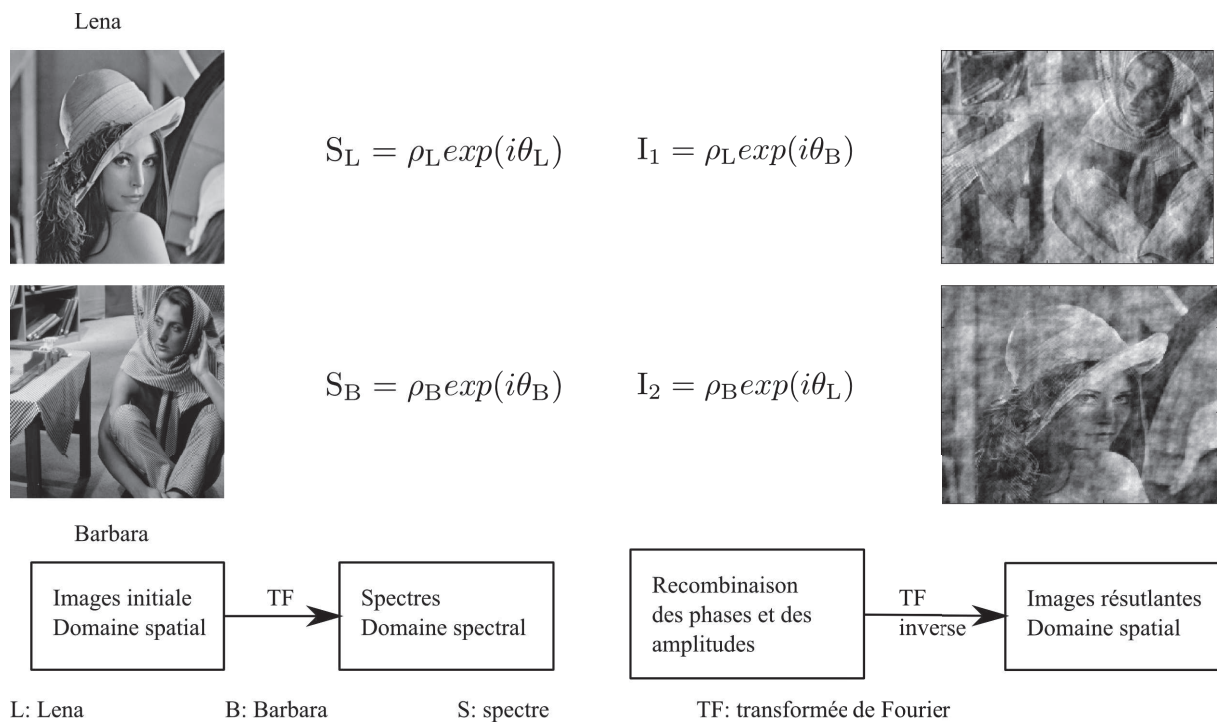


FIGURE 2.14 – Combinaison des phases et des amplitudes des spectres

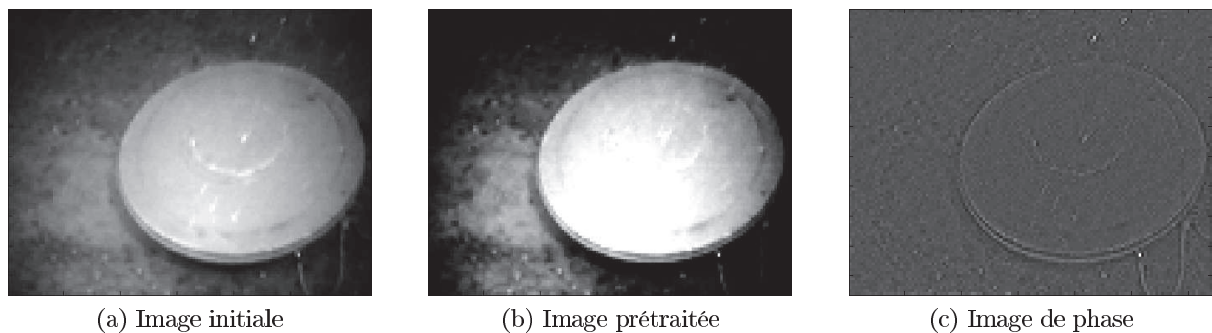


FIGURE 2.15 – Image de phase calculée à partir de l'image prétraitée issue d'une vidéo Topvision

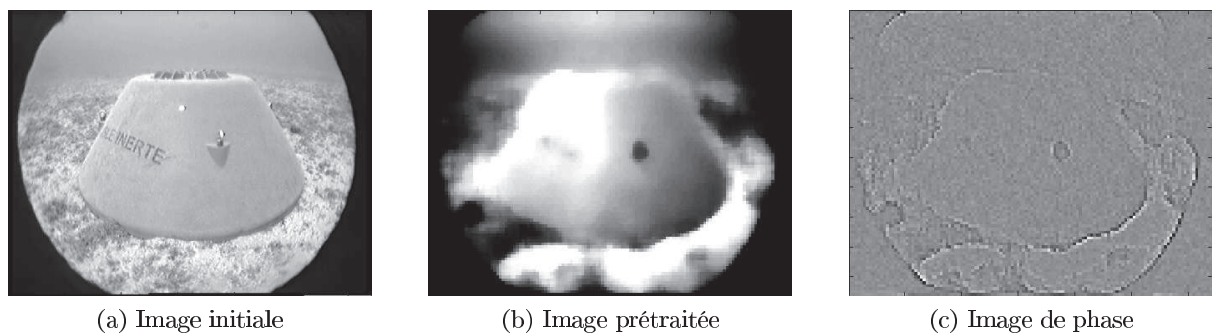


FIGURE 2.16 – Image de phase calculée à partir de l'image prétraitée issue d'une vidéo VAMA

2.2.6 Les prétraitements annexes

2.2.6.1 Les bords des images et les écritures

Les images présentent une ou plusieurs rangées de pixels nuls sur les bords (*cf.* figure 2.18). Lors d'une détection de contours, la séparation entre l'image (où les pixels ont des valeurs variables et proches dans un même voisinage) et ces zones noires vont ressortir et entraîner la

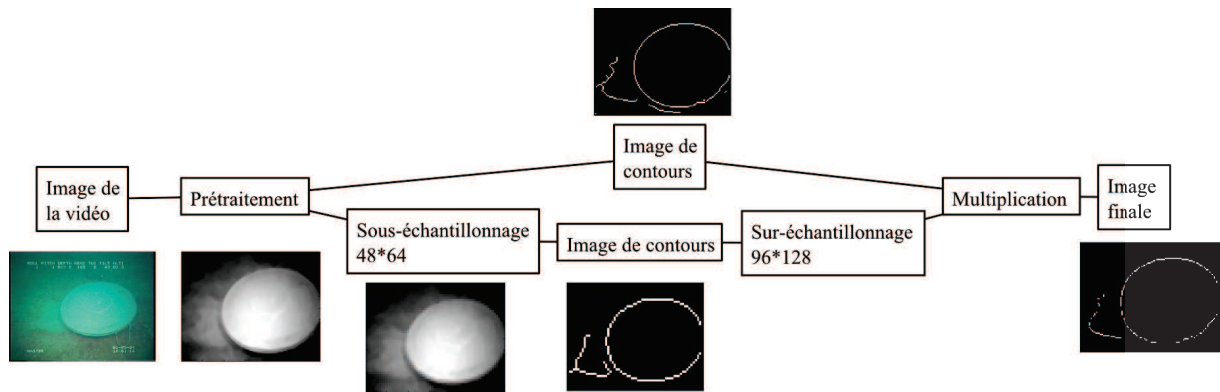


FIGURE 2.17 – Schéma expliquant l'obtention d'une image de contours via l'approche multi-échelle

détection d'objets sur les bords de l'image. Pour supprimer ce risque de fausses alarmes, nous avons décidé de couper ces régions. La taille de la zone découpée est fixe et adaptée selon les images. Cette coupe des bords de l'image peut entraîner la suppression d'une partie de la mine. Cela aura une influence sur le début et la fin de la détection de la mine.

De plus sur les images Topvision, les informations de navigation sont inscrites sur les images. Ces inscriptions, qui créent de forts contours dans les images peuvent perturber notre algorithme. Nous avons donc coupé les images de façon à ne pas être gênés par ces écritures (*cf.* figure 2.19).

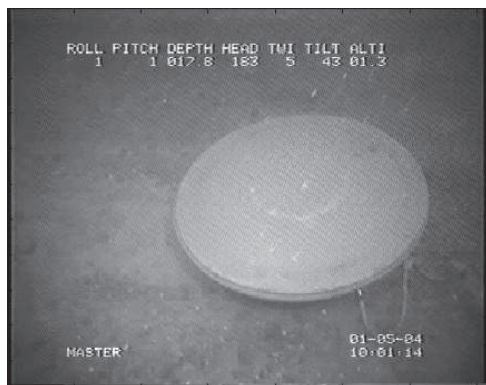


FIGURE 2.18 – Image issue de la vidéo



FIGURE 2.19 – Image après découpe des bords et des écritures

Pour les images Topvision où il faut couper les écritures, la bande est de cent pixels. Les images de taille initiale 720×540 pixels mesurent maintenant 520×340 pixels. Sur les images VAMA où seule la bande noire est à couper, la taille est fixée à dix pixels. Les images de taille initiale 640×480 pixels mesurent alors 620×460 pixels.

2.2.6.2 Suppression du manchon

Sur l'AUV qui a enregistré les données VAMA, une couronne de LEDs est disposée autour de la caméra de façon à éclairer la scène lorsque la luminosité ambiante n'est pas suffisante. La caméra embarquée étant située au centre de cette couronne de LEDs, elle peut être saturée par cette lumière. Pour éviter cette saturation des images, un manchon a été ajouté autour de la caméra. Ce manchon est visible sur les images et introduit des contours qui peuvent perturber la détection et l'identification d'objets (*cf.* figure 2.20).

Nous avons pris une image au début d'une séquence pour le fond de l'image. Nous avons redimensionné et sous échantillonné cette image de façon à obtenir une image de même taille que l'image à prétraiter. Une binarisation de cette image me permet d'obtenir un masque du manchon qui sera multiplié à l'image de contours de l'image cible afin de supprimer ces contours

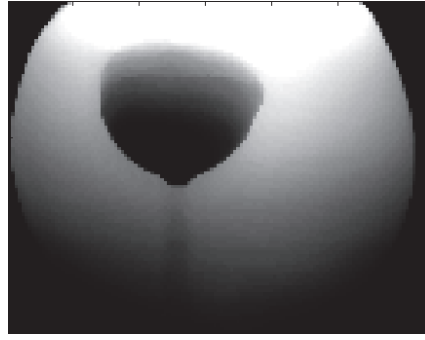


FIGURE 2.20 – Image prétraitée

et de ne pas conclure à la présence d'un objet.

La figure 2.21 montre l'image de contours avant suppression du manchon, obtenue avec la méthode présentée sur la figure 2.17. Le manchon est détecté et peut faire conclure à la présence d'un objet. Sur la figure 2.22, le manchon a été supprimé et n'introduit plus de fausse alarme.



FIGURE 2.21 – Image de contours de l'image initiale



FIGURE 2.22 – Image de contours sur laquelle le manchon a été supprimé

2.3 Résultats et conclusion

Les images Topvision et VAMA ne présentent pas les mêmes caractéristiques et les mêmes problèmes. Nous avons donc dû adapter nos prétraitements aux différentes images.

2.3.1 Les images Topvision

Avec les images Topvision, nous avons utilisé la coupe des écritures pour supprimer le problème des fausses détections à ces endroits de l'image (*cf.* partie 2.2.6.1). Un premier filtrage et une réduction du temps de calcul ont été obtenus grâce au sous-échantillonnage de l'image (*cf.* partie 2.2.2). Nous avons utilisé un filtre médian pour filtrer davantage de bruit présent dans l'image (*cf.* partie 2.2.3). Enfin nous avons rehaussé le contraste de façon à faire ressortir davantage les objets présents dans les images (*cf.* partie 2.1.2.1). Nous n'avons pas rencontré le problème de saturation du soleil sur le sol ou sur les mines. Le modèle de Phong n'a donc pas été utilisé.

Un schéma de l'algorithme de prétraitement est donné figure 2.23 avec des images montrant l'effet de chaque étape. Comparativement à l'image issue de la vidéo, il n'y a plus de contours et de bruits qui peuvent perturber la détection. La mine ressort nettement dans l'image. Les images de phase ont été ajoutées à titre indicatif, pour montrer l'effet des prétraitements sur l'image de phase. Dans ce cas, nous pouvons constater une amélioration des contours. Les images de

contours obtenues par les images de phase seront utilisées pour la détection et l'identification des objets.

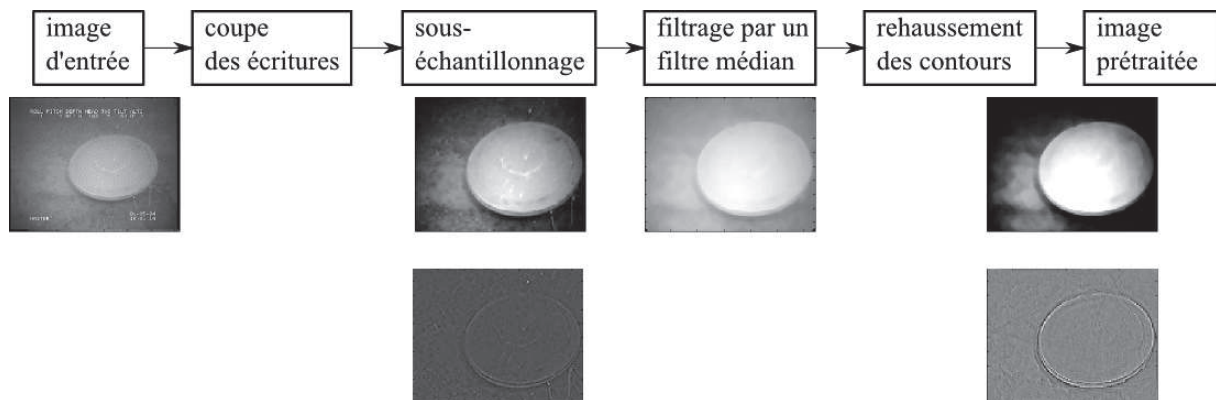


FIGURE 2.23 – Algorithme des prétraitements utilisés sur les images Topvision

2.3.2 Les images VAMA

Sur les images VAMA, nous avons utilisé la coupe des bordures noires (*cf.* partie 2.2.6.1), le sous-échantillonnage (*cf.* partie 2.2.2), le filtre médian (*cf.* partie 2.2.3), le rehaussement de contraste (*cf.* partie 2.1.2.1) ainsi que le prétraitement utilisant le modèle de Phong. L'information de distance fournie par les différents capteurs est nécessaire dans ce cas. La suppression du manchon ainsi que la détection de contours par l'approche multi-échelle ont été utilisées mais dans une étape ultérieure. Un schéma de l'algorithme de prétraitement est donné figure 2.24. Le contraste de l'image a été modifié et les problèmes de saturation limités. Sur les images de phase, nous pouvons constater qu'il y a beaucoup de contours détectés dans l'image initiale. Dans l'image de phase calculée à partir de l'image finale, il y a moins de bruit pour perturber une détection de contours et la mine moins bien détectée qu'initialement est néanmoins toujours bien visible.

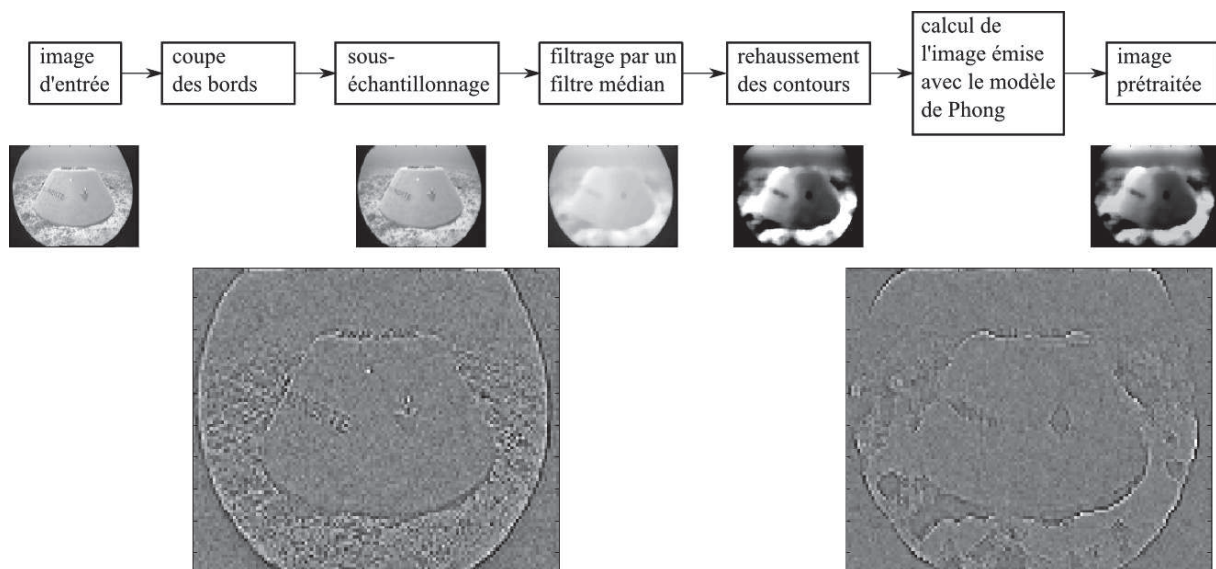


FIGURE 2.24 – Algorithme des prétraitements utilisés sur les images VAMA

L'étude des prétraitements a permis d'utiliser le modèle de Phong dans le milieu sous marin. De plus l'utilisation de l'information de navigation et de l'a priori sur la position de l'objet permet de calculer des valeurs pour l'absorption proches des valeurs réelles.

Chapitre 3

Choix de l'approche

Sommaire

3.1	Méthode basée sur la couleur	46
3.2	Méthode basée sur la transformée de Hough	46
3.3	Méthodes basées sur les contours	47
3.3.1	Utilisation des détecteurs de contours	47
3.3.2	Utilisation des contours actifs	48
3.4	Méthodes basées sur l'attention visuelle	49
3.5	Méthodes basées sur la séparation premier plan arrière plan	49
3.6	Le suivi d'objet	50
3.7	Le positionnement et la cartographie	50
3.8	Utilisation d'images polarisées	50
3.9	Méthodes du milieu terrestre	51
3.9.1	Les méthodes utilisant une base de données	51
3.9.2	Les méthodes à base d'histogramme	51
3.9.3	Les réseaux de neurones	51
3.9.4	Les classificateurs	52
3.9.5	La corrélation	52
3.10	Conclusion	52

Dans ce chapitre, nous allons faire l'état de l'art des méthodes qui existent pour faire de la détection et de la reconnaissance. Nous couvrirons principalement le milieu sous marin, avant de présenter quelques méthodes du milieu terrestre qui nous semblent intéressantes et qui peuvent être adaptées à la résolution de notre problème. Nous avons vu dans le chapitre sur le contexte (*cf.* chapitre 1) que la guerre des mines (l'application développée dans ce travail) nécessite la distinction de trois étapes : la détection, la classification et l'identification pour traiter un objet. Dans un souci de simplification, nous allons nous intéresser à des méthodes qui permettent de fusionner l'étape de la classification avec celle de l'identification. Ces méthodes seront par la suite appelées méthodes de reconnaissance. La fusion de l'étape de classification et de l'étape d'identification est importante du fait que nous utilisons un AUV et que nous avons pour contrainte le temps réel. En effet, l'utilisation de plusieurs méthodes ralentirait le temps d'exécution et nécessiterait davantage de ressources de l'AUV.

De manière générale, les applications de la vision sous marine sont très nombreuses. Nous pouvons recenser la détection, la reconnaissance et le suivi d'objets manufacturés (mines, pipelines,...), le comptage de poissons, le positionnement, etc... Ces applications n'ont pas de techniques spécifiques : une technique utilisée pour le suivi d'objet peut aussi être utilisée pour le comptage des poissons par exemple. Nous avons donc choisi de classer notre état de l'art par méthodes. Nous les commenterons par rapport au projet qui nous intéresse. Enfin nous concluons sur la méthode qui nous semble la plus adaptée à notre problème.

3.1 Méthode basée sur la couleur

En dépit du fait que le milieu sous marin est connu pour son absorption des couleurs, comme on l'a vu dans le chapitre sur le contexte (*cf.* chapitre 1), différentes méthodes utilisant la couleur ont été proposées et validées dans la littérature. Par exemple, Bazeille *et al.* [70] identifient des objets par leur couleur. A partir de la loi de Beer-Lambert, pour chaque couleur, Bazeille *et al.* [70] définissent un ensemble de couleurs qui lui correspondent dans le milieu sous marin (ensemble défini par une équation paramétrique). Ensuite les pixels des images issues de la vidéo sont analysés. S'ils appartiennent à la plage de couleurs correspondant à l'objet recherché (s'ils sont solution de l'équation paramétrique), il est classé comme "objet". Si le nombre de pixels classés "objet" est suffisant, l'image contient un objet. Sinon, l'objet est considéré comme appartenant à la scène. Les tests effectués consistent à détecter une cible rouge dans un bassin. Les résultats sont encourageants mais rien n'assure qu'ils seront aussi élevés sur des images prises en mer avec des cibles de couleur neutre, proche de la couleur du milieu (sol, rocher, algues).

L'utilisation de méthodes basées sur la couleur pour détecter et identifier des objets pose plusieurs problèmes majeurs. Tout d'abord, ces méthodes impliquent une connaissance de la couleur de l'objet. Dans un contexte de guerre des mines, nous ne disposons pas d'informations sur l'objet cherché (l'ennemi ne donne pas les informations sur les mines qu'il place!). De plus, selon les conditions météorologiques, géographiques, etc, le milieu change la couleur perçue. Au cours du temps, les objets peuvent changer de couleur avec le dépôt de sable, d'algues,... Tous ces changements ne peuvent pas être complètement pris en compte dans l'équation paramétrique. Un problème technique posé par cette méthode est la nécessité d'avoir des données en couleurs!

3.2 Méthode basée sur la transformée de Hough

La transformée de Hough est une transformation qui permet de détecter des lignes dans une image. La relation de passage entre les coordonnées cartésiennes et les coordonnées polaires s'écrit :

$$\begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta \end{aligned} \quad (3.1)$$

Il est donc exact d'écrire :

$$\begin{aligned} x \cos \theta + y \sin \theta &= r \\ r \cos \theta \cos \theta + r \sin \theta \sin \theta &= r \end{aligned} \quad (3.2)$$

Il est alors possible de trouver un couple (r, θ) qui permette de décrire l'équation cartésienne

$$\begin{aligned} y &= ax + b \\ y &= -\frac{\cos \theta}{\sin \theta} x + \frac{r}{\sin \theta} \end{aligned} \quad (3.3)$$

Dans l'espace de Hough, une ligne sera représentée par un point de forte intensité, de coordonnées (r, θ) .

A. Balasuriya et T. Ura [71] ont étudié le suivi de câbles à l'aide de la transformée de Hough. Après avoir positionné l'AUV à proximité du câble cible, l'étape d'initialisation consiste à faire osciller l'AUV de façon à détecter le câble. Deux actions sont alors réalisées : la modélisation du câble et le suivi du câble. La modélisation permet de définir une zone probable de présence du câble et de guider l'AUV. Pour le suivi du câble, sur l'image obtenue (tout d'abord prétraitée), on applique une transformée de Hough. En cas de pic dans la zone de présence probable, le câble est considéré comme détecté et le système continue. En cas de non présence du pic, le câble n'est pas détecté. La modélisation sert à guider le véhicule jusqu'à ce que le câble soit détecté dans l'image. Ils ont supposé que le câble ne change pas de direction de façon abrupte.

La transformée de Hough sert dans sa version initiale à détecter des lignes dans les images. Une ligne de l'image est décrite par un point dans le domaine de Hough. Il faut donc traiter le

plan de Hough pour retrouver les lignes de l'image. Malgré de bons résultats obtenus avec cette approche, cette dernière est difficilement applicable à notre cas. En effet, les mines standards (présentées sur la figure 1.1) présentent des formes "simples" mais pas forcément rectilignes comme la mine sphérique. De plus, certaines mines non standards présentent des formes plus aléatoires et proches des formes des objets naturels. La détection de lignes ne nous avancera donc pas dans la résolution de notre problème.

3.3 Méthodes basées sur les contours

3.3.1 Utilisation des détecteurs de contours

Rova *et al.* [72] cherchent à identifier des poissons de forme similaire, qui se distinguent par leur nombre de rayures. Deux types de poissons sont à détecter dans leur application. Pour cela ils utilisent tout d'abord un détecteur de Canny [66] pour détecter les contours des poissons. Seuls cent points sont gardés, sélectionnés aléatoirement. Ces cent points sont reliés avec l'algorithme de Prim [73] qui consiste à relier tous les points entre eux, pondérer toutes ces liaisons et sélectionner l'arbre de poids minimum. Ensuite, ils calculent un histogramme sur les positions des points du contours. Les histogrammes de l'image cible et des images de référence sont comparés de manière à obtenir une image d'alignement estimé. Cette image est filtrée puis classifiée à l'aide de deux classificateurs binaires (un par poisson à reconnaître).

Cette méthode nécessite une bonne détection de contours pour reconnaître efficacement les poissons. En effet, si les contours détectés ne sont pas complets ou ne correspondent pas exactement à la forme du poisson, il est possible de ne pas trouver de correspondance lors de la comparaison des histogrammes et de la classification. Comme le montre la figure 2.17, il est possible sur nos images d'obtenir des contours ne correspondant pas à la mine. Il est aussi possible d'obtenir des contours non fermés ou ne correspondant pas exactement aux contours des objets présents dans les images. De plus cette méthode nécessite un classificateur pour chaque image de référence. Pour recenser tous les points de vues et toutes les échelles pour chaque mine, nous obtiendrons une base très importante, impliquant une fréquence de calcul faible.

Une autre méthode utilisant les contours est proposée et validée dans la référence [74]. G. Foresti [74] cherche à détecter les anodes situées aux jonctions des pipelines sous-marins. Pour se faire, il applique un détecteur de Canny pour extraire les contours de l'image. Parmi les contours obtenus, il cherche ceux qui pourraient appartenir à une ellipse (contour d'un pipeline). Ensuite il fait la correspondance entre les ellipses trouvées et l'état du drone sous marin (selon l'inclinaison du drone, le pipeline n'est pas vu sous le même angle et par conséquent la forme de l'ellipse change).

Cette méthode est efficace si l'on connaît la forme à chercher. Ici la section du pipeline aura toujours une forme elliptique. Nous avons choisi de chercher quatre types de mines différents dont la forme peut changer selon le point de vue. Pour utiliser cette méthode il faudrait trouver une formule mathématique décrivant tous les points de vue de nos mines. Cette méthode a deux problèmes : la forme des mines et le milieu diffusant.

Une troisième méthode proposée et validée par Helan *et al.* [47] privilégie la détection d'objets par les contours. Ils détectent les contours à l'aide du gradient de Di Zenzo [75]. Une étude d'histogrammes basés sur la connexité des contours trouvés permet de filtrer les détections trop petites, ne pouvant pas correspondre à un objet manufacturé. Ensuite ils choisissent les contours rectilignes en appliquant une transformée de Hough sur les contours trouvés. Enfin, une comparaison colorimétrique permet de détecter les contours qui entourent une même zone alors détectée comme étant un objet.

Cette méthode nécessite l'utilisation d'un ensemble de méthodes, ce qui alourdit l'algorithme et augmente considérablement le temps de traitement nécessaire pour traiter une image. La transformée de Hough et l'utilisation de la couleur ont déjà été étudiées et notre position donnée.

Olmos [45] propose et valide une méthode pour distinguer les objets naturels des objets manufacturés. Cette méthode consiste à extraire les contours présents dans l'image avec un détecteur de Canny [66]. Pour classifier les contours détectés, elle définit au préalable les propriétés des contours des objets naturels et des objets manufacturés en fonction de la longueur des contours, de leur nombre et de leur courbures. En mesurant ces grandeurs sur les contours obtenus avec le détecteur de Canny, elle classe les objets détectés en objets naturels ou objets manufacturés.

Cette méthode permet uniquement de classifier des détections, pas d'identifier un objet. De plus, un critère de classification basé sur la longueur des contours détectés demande une bonne détection des contours donc des images contrastées, ce qui n'est pas toujours le cas de nos images. Il est possible que les contours que nous obtenons soient très morcelés, ce qui limite l'intérêt de tels critères.

Le module de Thales [8] permet d'effectuer la détection et la reconnaissance de la mine. L'algorithme de détection a été développé avec l'entreprise Cybernetix. Cet algorithme prétraite les images de façon à pouvoir les segmenter et en extraire les contours. Ces contours sont comparés à des lignes et des ellipses qui permettent de décrire des objets manufacturés avec des contours "lisses" par opposition aux objets naturels qui auront des contours bruités. Le centre du contour détermine la position de la mine. L'identification de la mine se fait par comparaison avec des contours de référence. Si l'identification proposée par le système correspond au déplacement du véhicule, il est possible qu'elle soit correcte. Une mesure de qualité est effectuée pour quantifier l'identification.

Cette méthode présente de bons résultats mais comme toutes les méthodes basées sur la détection de contours, elle est tributaire de la forme de l'objet à détecter et de la qualité de l'image. Ainsi ces deux paramètres (forme et qualité) permettent ou non une bonne détection de contours qui facilite par la suite l'identification.

Globalement, les méthodes basées sur les contours semblent être nombreuses dans le domaine de la vision sous marine. Elles doivent être efficaces pour détecter au mieux les contours des objets présents ou elles nécessitent des images très contrastées. De plus les résultats de la détection de contours doivent être analysés afin de classifier et identifier les objets. Cet enchaînement de deux méthodes (ou plus) permet de différencier la détection de la reconnaissance d'un objet.

3.3.2 Utilisation des contours actifs

La méthode des contours actifs est utilisée par Kocak *et al.* [76] pour détecter et suivre du plancton bioluminescent. Le plancton est segmenté avec les contours actifs. Chaque élément trouvé est labellisé puis suivi dans la vidéo afin de compter les différents planctons des différentes espèces présents dans la vidéo étudiée.

La méthode des contours actifs est une méthode efficace pour trouver des contours mais nécessite un temps de calcul très important. En effet, une zone de départ est fournie à l'algorithme ainsi qu'un nombre d'itérations. A chaque itération, l'algorithme calcule l'énergie à l'intérieur du contour et l'énergie à l'extérieur. La somme de ces deux énergies doit être minimisée. Pour certains points, l'algorithme va trouver un minimum, pour d'autres points l'algorithme va chercher un minimum jusqu'à la dernière itération. Le contour se déforme à chaque itération pour se rapprocher du contour de l'objet. La taille de la zone de départ est importante : si l'objet à trouver est de grande taille et la zone petite, il faudra beaucoup d'itérations pour que le contour actif corresponde au contour de l'objet. Plus le nombre d'itérations est important plus le temps de calcul augmente. Comme nous ne connaissons pas l'objet à reconnaître ni son orientation, il est difficile de définir une taille de zone et un nombre d'itérations qui permettent de détecter efficacement et rapidement les contours des objets.

3.4 Méthodes basées sur l'attention visuelle

Dans la référence [77], C. Barat et R. Phlypo cherchent à détecter les objets contenus dans les images Topvision. Ils extraient des images-caractéristiques de façon à créer une carte de saillance. Leur approche est basée sur l'utilisation de l'attention visuelle [78]. Ils commencent par une étude de l'image cible sur plusieurs échelles. Ensuite, ils extraient des motifs selon la couleur, l'intensité et l'orientation. Les résultats aux différentes échelles sont fusionnés et normalisés pour obtenir une carte de saillance. Finalement, la zone la plus saillante est étudiée par une analyse linéaire discriminante pour décider de la présence ou non d'un objet. En cas de présence d'un objet, ils cherchent à le segmenter en utilisant la méthode des contours actifs [79] qu'ils initialisent avec la région la plus saillante.

Correia *et al.* [80] proposent une autre méthode basée sur l'attention visuelle dans le but de compter les homards. Ils calculent pour chaque image une carte d'intensité, une carte de contours (basée sur un détecteur de Sobel) et une carte de déplacement. Ensuite les cartes sont fusionnées de façon à créer une carte de saillance dans laquelle les homards sont représentés par les zones saillantes.

Cette méthode permet de renforcer la prise de décision en combinant les résultats de plusieurs méthodes. Cependant, comme dans l'approche de C. Barat et R. Phlypo [77], il est nécessaire d'étudier l'image à plusieurs échelles et appliquer sur toutes ces images-échelle les trois méthodes (dans les études présentées ici) utilisées pour détecter des objets. Correia *et al.* [80] considèrent que les zones détectées sont les homards qu'ils cherchent mais il peut être nécessaire, comme dans le cas de C. Barat et R. Phlypo [77] d'analyser ou de reconnaître les zones détectées.

3.5 Méthodes basées sur la séparation premier plan arrière plan

Une autre possibilité pour détecter un objet est de dépister et d'utiliser les changements dans le fond ou l'arrière plan de la scène à étudier.

T.Y.C. Van Valkenburg - Van Haarst et K.A. Scholte [81] se basent pour proposer leur approche sur l'utilisation des variations verticales d'intensité de l'arrière plan de l'image. L'arrière plan est estimé au moyen de polynômes. L'image est sous-échantillonnée horizontalement afin de réduire le temps de calcul et de réduire le bruit (l'affirmation au départ de ce sous-échantillonnage est la faible variation d'intensité horizontalement). Pour chaque pixel de l'image sous-échantillonnée, la variation d'intensité est mesurée. Les intensités sont supposées varier selon une Gaussienne. Pour détecter un objet, il faut que l'intensité entre le pixel mesuré et le pixel modélisé dépasse un seuil fixé (seuil fixé par colonne). Les pixels détectés sont ensuite assemblés au moyen d'opérations morphologiques.

Dans l'approche d'Edginton *et al.* [82], l'image est segmentée afin de séparer le premier plan de l'arrière plan. A partir de cette image segmentée, ils récupèrent des points d'intérêt qui définissent le centre des objets du premier plan. Ces centres seront utilisés pour initialiser les différentes pistes (évolution au cours de la vidéo de la position du ou des objets détectés dans l'image). Ils utilisent comme points d'intérêt les contours orientés. Leur objectif est de faire du suivi d'animaux. Ce suivi s'effectue à l'aide de l'extrapolation des positions de l'objet dans les images précédentes. Si un objet de l'image actuelle correspond à la position estimée d'un objet en cours de suivi, l'objet est assigné à l'objet suivi. Sinon on démarre une autre piste. On arrête de rechercher un objet s'il n'a pas été détecté dans 2 images consécutives. Cependant quand un objet a été perdu dans une image, on interpole sa position dans l'image suivante pour pallier aux défauts du suiveur.

Ces méthodes sont intéressantes pour la détection d'un objet dans une image. Elles ont l'avantage de ne pas nécessiter une détection de contours qui peuvent ne pas être efficaces sur les images peu contrastées. L'inconvénient par rapport à nos images est l'évolution de l'arrière plan. Lorsque l'on image un objet situé dans le volume d'eau, l'arrière plan varie peu lorsque la caméra se rapproche de l'objet. Par contre lorsque l'objet est posé au sol l'arrière plan varie dans la vidéo : au début on image le volume d'eau puis de plus en plus le sol dont l'intensité

change lorsque la caméra s'en approche. Cela peut fausser les résultats si une telle méthode est appliquée. Cependant, vu les bons résultats de ces méthodes, nous allons, par la suite, nous en inspirer pour proposer un système robuste.

3.6 Le suivi d'objet

Certaines applications consistent à suivre des objets pour faire de l'inspection ou pour guider un drone par exemple. Le suivi d'objet ou "*tracking*" nécessite d'un côté un système de détection de l'objet à suivre ainsi qu'un algorithme de "*tracking*". Un inventaire de méthodes récentes (depuis 1996) a été fait par E. Trucco et K. Plakas [83] en 2006. On retrouve la détection de contours et de points d'intérêt, la segmentation en contours et en régions, ainsi que la transformée de Hough. L. Wang, S. Hu et X. Zhang [84] utilisent la soustraction du fond et la soustraction entre images successives pour détecter un objet et son déplacement avant d'appliquer un algorithme de suivi.

3.7 Le positionnement et la cartographie

Le SLAM, "*Simultaneous localization and mapping*" consiste à cartographier une zone donnée. Le drone effectue des aller-retours au dessus de la zone à cartographier. En parallèle de l'opération de cartographie, le drone doit se repérer dans l'espace de façon à créer la carte selon sa position. Pour cela il faut détecter des points d'intérêt, des objets particuliers et établir la correspondance entre deux images. Cette technique est utilisée dans le milieu terrestre, aussi bien que dans le milieu sous marin.

Le drone sous marin utilisé par Eustice et al. [85] est équipé d'une caméra vidéo et des capteurs de navigation. A l'aide des informations de navigation, ils créent un filtre de Kalman étendu [86]. Les informations de navigation servent à calculer l'état futur tandis que les images de la caméra servent à faire la mesure. Pour cela, ils extraient les points d'intérêt de l'image. Ils recherchent ensuite une correspondance entre les points d'intérêt de l'image i et de l'image j . En cas de correspondance, ils calculent les mesures nécessaires à la mise à jour du filtre de Kalman étendu.

Les techniques développées pour faire de la cartographie et du repérage dans l'espace peuvent servir pour faire de la détection d'objet et du suivi d'objets. Cependant ces techniques servent à repérer tout point d'intérêt qui puisse permettre d'apparier deux images. Notre objectif est de détecter les mines sous marines. Nous cherchons donc une méthode qui ne produise pas trop de fausses détections sur des objets naturels.

3.8 Utilisation d'images polarisées

Nous avons vu dans la partie consacrée aux capteurs (*cf.* section 1.2.4.2) qu'il est possible d'utiliser des caméras polarisées pour améliorer la qualité des images ainsi que la portée.

L'étude des images polarisées peut permettre de détecter et identifier des objets. En effet, lorsque la lumière se réfléchit sur un objet, la lumière réfléchie est dépolarisée plus ou moins fortement. Le Brun *et al.* [87] montrent que l'inox dépolarise faiblement la lumière tandis que le sable la dépolarise fortement. De même ils étudient un morceau de métal en partie rouillé et montrent que la rouille change la dépolarisation de la lumière. Wolff [88] applique la segmentation sur des images polarisées terrestres et présente des images sur lesquelles les différents matériaux sont détectés.

L'étude d'images polarisée demande des adaptations matérielles : il faut utiliser une caméra polarisée et éventuellement une source polarisée. De plus les matériaux dépolarisent plus ou moins la lumière incidente réfléchie vers le capteur. L'ajout de peinture ou le dépôt de sable peuvent changer la "signature" du matériau. Cette méthode nécessite des études plus approfondies pour étudier la réponse de différents objets dans diverses situations pour mettre en œuvre un système de détection efficace. Cependant l'utilisation d'images polarisées constitue une perspective

d'amélioration pour cette thèse. Nous avons commencé à effectuer des tests afin d'évaluer dans quelles mesures cela peut améliorer notre détection d'objets.

3.9 Méthodes du milieu terrestre

Un très grand nombre de méthodes de détection et de reconnaissance pour des applications terrestres ont été proposées et validées dans la littérature. Toutes n'ont pas été adaptées au milieu sous marin. Dans cette partie, nous allons présenter quelques méthodes qui nous semblent susceptibles d'être adaptées dans le milieu sous marin.

3.9.1 Les méthodes utilisant une base de données

Cole *et al.* [89] ont proposé une méthode adaptée à la reconnaissance des Lego. Pour cela, ils ont divisé leur base de donnée en classes. Chaque classe représente une pièce et elle-même est représentée par une image moyenne et une image cluster (assemblage des vues caractéristiques de la pièce). Chaque image à identifier est comparée aux images moyennes grâce à une corrélation normalisée et à un seuil. Ensuite l'image est comparée aux images clusters des classes qui pourraient correspondre, selon le résultat obtenu à l'étape précédente.

L'utilisation d'une image représentant une classe est intéressante dans le sens où cela permet de limiter le nombre de comparaisons à effectuer. Ici, pour identifier la pièce de Lego, ils effectuent deux comparaisons avec des bases d'images importantes. Cette méthode ne peut pas être appliquée directement à notre cas (étude des mines sous marines) car le milieu a un effet sur les images des objets que l'on peut obtenir. La méthode est donc à adapter pour être efficace dans le milieu sous marin.

3.9.2 Les méthodes à base d'histogramme

B. Schiele et J.L. Crowley [90] utilisent la correspondance d'histogrammes couleurs pour reconnaître les objets. La première étape consiste à calculer les histogrammes couleurs de l'image. Ensuite, ils comparent l'histogramme ainsi obtenu aux histogrammes références à l'aide d'un test du χ^2 (test statistique permettant de comparer deux séries de données). Leur tests montrent que la méthode est robuste aux changements d'orientation, d'échelle, à l'occlusion partielle et au changement de vue.

Cette méthode est simple, robuste et rapide. Cependant, elle dépend de la couleur et de l'intensité lumineuse. Nous avons déjà évoqué le problème de la couleur. L'intensité lumineuse est une donnée non stable et reproductible dans le milieu sous marin. En effet, selon les conditions météorologiques, l'heure d'enregistrement des images et la position géographique, l'intensité lumineuse ne sera pas identique dans toutes les images.

3.9.3 Les réseaux de neurones

SpikeNet est une méthode en temps réel développée par Thorpe *et al.* [91,92]. La méthode proposée est rapide et consiste à simuler un réseau de neurones avec très peu de paramètres afin de reconnaître les visages. La première version de leur algorithme a quatre étapes : une étape pour connaître l'état auquel le pixel de l'image correspond (par analogie avec la rétine qui possède deux états pour ses cellules : on et off), deux étapes pour reconnaître l'orientation des contours et des caractéristiques spécifiques apprises par le programme. La dernière étape est consacrée à l'analyse des résultats et à la prise de décision. A la suite de cette version, ils ont proposé une version optimisée qui reconnaît toutes sortes d'objets si les caractéristiques spécifiques contiennent des structures orientées.

Cette méthode nécessite une bonne détection de contours afin d'extraire des caractéristiques spécifiques. L'efficacité de cette méthode dépend de la qualité des résultats des détecteurs de contours et donc de la qualité des images. De plus, pour construire notre réseau de neurones, il faut connaître les caractéristiques de tous les points de vue à toutes les échelles ce qui fait

un réseau important et complexe à gérer. En effet, l'identification se base sur la reconnaissance d'une caractéristique spécifique. Il faut donc une caractéristique par point de vue et par échelle qui soit facilement identifiable dans les images de contours. Au vu du nombre de mines et par conséquent de points de vue et d'échelles, le nombre de caractéristiques "uniques" à extraire est important. Il faut être capable par la suite de gérer cette importante quantité d'informations et de les organiser à la façon d'un arbre qui permette de limiter le nombre des possibilités au fur et à mesure de l'avancée dans l'arbre.

3.9.4 Les classificateurs

P. Viola et M. Jones [93] ont utilisé le classificateur AdaBoost pour détecter rapidement les visages dans des images en niveaux de gris. Ils sélectionnent quelques classificateurs qu'ils assemblent en cascade. Ainsi ils obtiennent un très faible taux d'erreur sur la base d'images d'apprentissage.

Cette méthode nécessite aussi une extraction de caractéristiques des images de référence. Comme toutes les mines choisies ne possèdent pas d'axes de symétrie, il faut tenir compte de tous les points de vue différents.

3.9.5 La corrélation

La corrélation est une méthode qui consiste à comparer une image cible à une image de référence [94]. Comme plusieurs méthodes citées ci-dessus, cette méthode nécessite d'avoir une image de référence pour chaque point de vue. Cependant, il existe des techniques pour multiplexer les images de référence et ainsi réduire considérablement le nombre d'opérations à effectuer [94]. De plus la corrélation est une technique très rapide, que ce soit avec un montage optique (la corrélation est alors effectuée à la vitesse de la lumière) ou un montage numérique [95].

3.10 Conclusion

Beaucoup de méthodes citées ici nécessitent réellement l'utilisation et la fusion de deux méthodes : une pour la détection et une pour la reconnaissance. Parmi les méthodes trouvées, plusieurs nécessitent une très bonne qualité d'image pour effectuer une extraction de caractéristiques ou de contours. Le temps de calcul est un point important parmi nos contraintes. Or la succession de méthodes risque de demander un temps de calcul important.

Nous avons repéré deux méthodes intéressantes pour la détection des objets. Les méthodes basées sur les contours semblent être des méthodes privilégiées, bien que la visibilité et le contraste ne soient pas toujours suffisants pour utiliser ces méthodes. La séparation du premier plan et de l'arrière plan peut être intéressante à utiliser. La polarisation est utilisée depuis de nombreuses années mais nécessite un équipement particulier dont nous ne disposons pas.

Parmi les méthodes de reconnaissance que nous avons recensées, plusieurs nécessitent des images de référence pour chaque point de vue des objets. Nous avons choisi dans le chapitre sur le contexte (*cf.* chapitre 1) d'étudier quatre mines dont seule la mine sphérique a le même aspect quelque soit le point de vue. La portée visuelle dans le milieu sous marin peut atteindre plusieurs mètres. Créer des images de référence des quatre mines selon tous les points de vue et pour plusieurs distances observateur - objet revient à créer une base de données très importante. Quelle que soit la méthode employée pour comparer les images cibles à ces images de référence, le temps de calcul sera important.

Notre objectif est de trouver une méthode qui permette d'effectuer conjointement la détection et la reconnaissance. La suite de nos recherches bibliographiques a montré que la corrélation est une méthode correspondant à notre objectif. Il existe parmi les méthodes de corrélation des techniques de multiplexage des images pour limiter la taille de la base de données, de façon similaire à celle utilisée par Cole *et al.* [89].

Deuxième partie

La corrélation

Introduction

Nous avons décidé au vu de l'état de l'art d'utiliser la méthode de la corrélation pour résoudre notre problème. En effet, la corrélation est l'une des techniques les plus fiables car elle est assez robuste (au bruit) et assez discriminante. Elle peut à la fois détecter et identifier un objet cible dans une scène. Avant d'étudier quelques méthodes et architectures qui composent la corrélation, nous allons présenter de manière générale la corrélation et les moyens de prendre une décision.

Nous avons choisi d'utiliser la corrélation d'un point de vue "opticien". En effet, nous privilégions les méthodes qui peuvent s'implémenter optiquement aux méthodes qui ne s'implémentent que numériquement. Une comparaison sera effectuée pour justifier ce choix. Nous avons sélectionné, pour ces études, des images simples et maîtrisées afin de bien comprendre le fonctionnement de la corrélation. Il s'agit d'images binaires de lettres auxquelles nous pouvons ajouter du bruit (blanc ou qui représente un motif, appelé par la suite bruit texturé) ou d'images de visages dans lesquelles de l'information nécessaire à la reconnaissance est présente ainsi que du bruit qui peut perturber la reconnaissance.

Nous présenterons en détails les deux principales méthodes de la corrélation optique. Nous avons effectué des comparaisons des principales architectures développées pour chacune de ces méthodes. Afin de répondre aux besoins créés par notre application, nous avons contribué en proposant deux architectures et les critères de décision associés [96–98]. La première architecture intervient au niveau de la reconnaissance et permet de réduire le nombre d'images de référence d'un facteur 2. L'architecture montre de bons résultats, comparativement aux autres architectures. La seconde architecture a été développée pour faire du suivi d'objet et répond à un besoin de la méthode : allier robustesse et discrimination.

Chapitre 4

Généralités sur la corrélation

Sommaire

4.1	La corrélation	57
4.1.1	Définition	57
4.1.2	Implémentation	58
4.1.3	Approche	59
4.2	Quantifier la reconnaissance	59
4.2.1	La robustesse au bruit	59
4.2.2	L'efficacité optique	60
4.2.3	La finesse du pic de corrélation	60
4.2.4	Les courbes ROC	60

Dans ce chapitre, nous définissons la corrélation de façon générale : nous donnons la définition mathématique et quelques propriétés utiles, nous présentons les deux manières de l'implémenter et l'approche que nous avons choisie pour l'étudier. Ensuite nous détaillons des moyens pour quantifier la reconnaissance. En effet, plusieurs critères sont à notre disposition pour juger la similitude entre l'image à identifier et les images de référence, connues.

4.1 La corrélation

4.1.1 Définition

Une méthode très utilisée dans la reconnaissance d'objets est la corrélation [94, 99]. Il s'agit de comparer l'image à analyser (ou image cible) à une image de référence connue, issue d'une base d'apprentissage (aussi appelée base de référence).

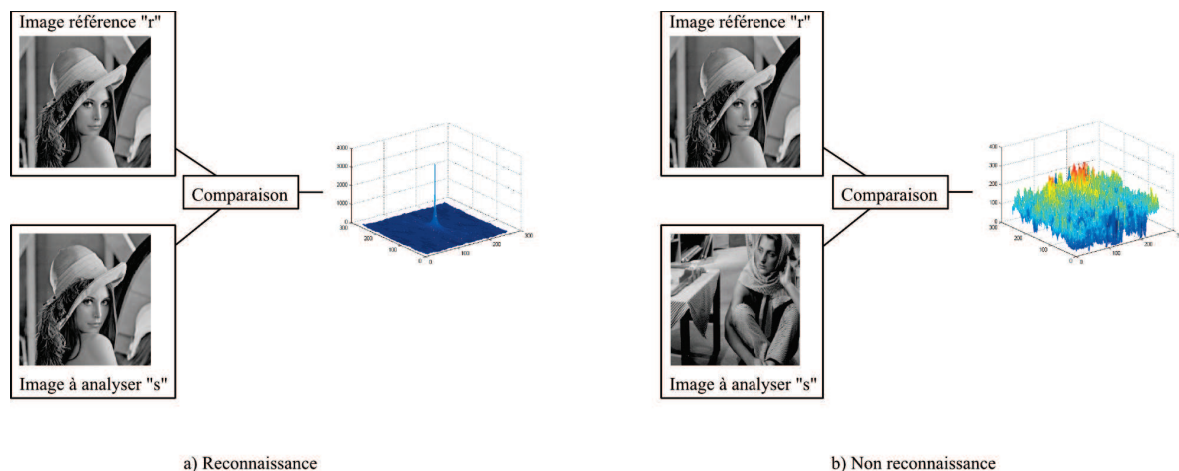


FIGURE 4.1 – Principe de la corrélation. A gauche en cas de reconnaissance, à droite en cas de non reconnaissance

Sur la figure 4.1, nous montrons le principe de la corrélation, basée sur le principe d'une comparaison spéciale. Les deux images sont comparées dans le domaine spectral. Le résultat de cette opération donne un plan de corrélation qui contient un pic de corrélation (*cf.* figure 4.1 a) en cas de reconnaissance et qui n'en contient pas en cas de non reconnaissance (*cf.* figure 4.1 b).

Mathématiquement, la corrélation c entre une image s et une image de référence r (aussi appelée filtre de corrélation) s'écrit :

$$c(x_0, y_0) = s(x_0, y_0) *^c r(x_0, y_0) = \int \int_{-\infty}^{+\infty} s(x, y) r(x - x_0, y - y_0) dx dy \quad (4.1)$$

où $*^c$ représente l'opération de corrélation.

La corrélation est liée à l'opération de convolution par la conjugaison du filtre :

$$c(x_0, y_0) = \bar{r}(-x_0, -y_0) * s(x_0, y_0) \quad (4.2)$$

où \bar{r} représente la conjugaison du filtre et $*$ l'opération de convolution.

Par le théorème de Plancherel, la convolution et par conséquent la corrélation est liée à la TF (transformée de Fourier) :

$$TF[x(t) * y(t)] = TF[x(t)].TF[y(t)] \quad (4.3a)$$

$$TF[x(t).y(t)] = TF[x(t)] * TF[y(t)] \quad (4.3b)$$

où $*$ représente le produit de convolution.

4.1.2 Implémentation

On distingue la corrélation optique (basée sur l'implémentation optique) de la corrélation numérique (basée sur une implémentation numérique, non réalisable en optique). La corrélation optique utilise la lumière et les propriétés des lentilles pour obtenir l'opération de corrélation (*cf.* figure 4.2). En effet, lorsque le plan d'entrée est situé au foyer objet de la lentille, au foyer image se trouve la transformée de Fourier du plan d'entrée. (Les foyers objet et image d'une lentille convergente se situent de part et d'autres de la lentille, à la distance appelée distance focale f). Grâce à cette propriété, le plan d'entrée est placé au foyer objet de la lentille $L1$ sur la figure 4.2. Au foyer image de cette lentille se trouve sa transformée de Fourier. On y place un modulateur spatial de lumière, appelé SLM, "*Spatial light modulator*". C'est un composant dans lequel les filtres sont affichés pour être multipliés optiquement au spectre de l'image. Pour effectuer la transformée de Fourier inverse, on utilise une deuxième lentille, $L2$, placée de façon à ce que le plan de Fourier soit situé au foyer objet. Au foyer image de cette seconde lentille, se trouve le plan de corrélation.

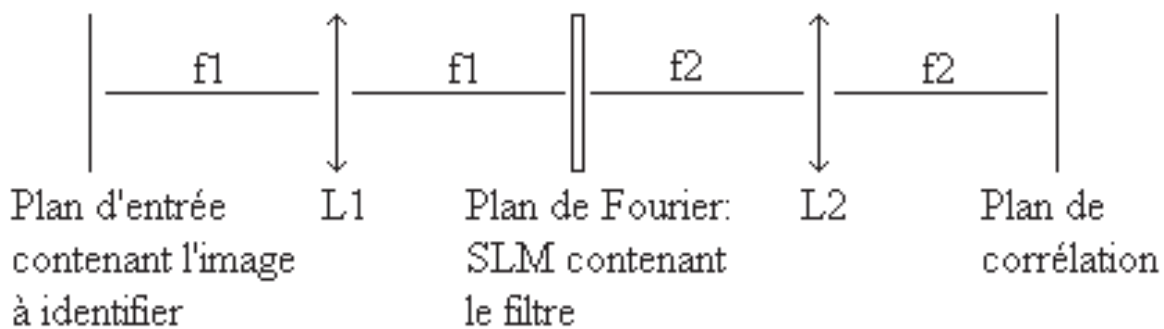


FIGURE 4.2 – Implémentation optique de la corrélation

Une fois que l'image d'entrée et le filtre sont affichés dans le montage, l'opération de corrélation peut s'effectuer à la vitesse de la lumière. La corrélation optique est donc une méthode extrêmement rapide. Cependant lorsque l'on souhaite comparer une image à plusieurs filtres, la

méthode nécessite de prendre en compte le temps de chargement du filtre du modulateur de lumière. De plus le montage optique impose des contraintes physiques (alignement du montage, encombrement spatial).

Cependant, de plus en plus d'applications utilisent des capteurs vidéos pour acquérir les images. Par conséquent les architectures optiques ont été adaptées sous forme numérique. Le traitement numérique des images est moins rapide que la corrélation optique mais est suffisant lors de l'utilisation de caméras vidéos (dans ce cas, le temps de calcul doit être proche de la cadence vidéo).

La corrélation optique regroupe deux architectures : le corrélateur proposé par Vanderlugt [100] et le corrélateur proposé par Weaver et Goodman (le JTC) [101]. Ces deux architectures ont été proposées pour une implémentation optique et ont été adaptées sous forme numérique [95,102]. La corrélation numérique, utilisée par les roboticiens, entre autres, est ici représentée par la corrélation normalisée (ZNCC, “zero mean normalized cross correlation”). Cette architecture consiste à effectuer la corrélation des images cible et référence auxquelles on a soustrait leur moyenne locale ; ce produit est divisé par l'écart type des images. Pour prendre une décision fiable, quelle que soit la méthode utilisée, des critères sont à prendre en compte : la rapidité, la robustesse et la discrimination. Plusieurs optimisations ont été développées pour améliorer la robustesse et / ou la discrimination des méthodes [94,97]. Un système de reconnaissance des formes est robuste lorsqu'il est capable de reconnaître une image cible proche de l'une des images de référence mais pas exactement identique, par exemple une lettre A verticale et une lettre A orientée avec un angle de 5° (l'angle dépend du degré de robustesse choisi). Un système est dit discriminant s'il ne reconnaît pas les images cibles différentes des images de référence mais proches par leur contenu, par exemple la lettre A et la lettre V.

4.1.3 Approche

Nous avons étudié ces différents systèmes ainsi que leurs optimisations afin de choisir une architecture efficace pour résoudre notre problème : l'identification des mines. Une première étape importante est l'évaluation des résultats ; l'utilisation de critères qui vont nous permettre de comparer les différentes optimisations. Pour cela, nous avons étudié des critères de quantification qui permettent de mesurer la robustesse au bruit, la quantité d'énergie, la finesse du pic de corrélation.

De plus, nous avons contribué en créant de nouveaux filtres et de nouveaux critères qui leur correspondent et qui permettent de faciliter la prise de décision sur la ressemblance ou non d'un objet inconnu par rapport à une référence. Cette partie sur la corrélation a fait l'objet de publications [96–98].

4.2 Quantifier la reconnaissance

L'évaluation des résultats nécessite l'utilisation de critères de quantification de la reconnaissance. Ici nous allons présenter quatre critères : le SNR, “Signal to Noise Ratio”, l'efficacité optique, le PCE, “Peak to Correlation Energy” ainsi que les courbes ROC, “Receiver Operating Characteristic”.

4.2.1 La robustesse au bruit

Une méthode de quantification très utilisée est la comparaison du niveau du signal par rapport au niveau du bruit. Le SNR (rapport signal à bruit) [103] permet de comparer les puissances du signal s et du bruit b contenus dans une scène :

$$SNR = \frac{\text{Puissance du signal}}{\text{Puissance du bruit}} = \frac{|S(f)|^2}{|B(f)|^2} \quad (4.4)$$

où $|S(f)|^2$ et $|B(f)|^2$ représentent les densités spectrales d'énergie du signal et du bruit, respectivement.

4.2.2 L'efficacité optique

On peut vouloir comparer l'énergie du plan de corrélation obtenu par rapport à l'énergie de l'image cible. Ce critère permet de quantifier l'énergie du plan de corrélation :

$$\eta = \frac{\text{énergie de la fonction de corrélation}}{\text{énergie de l'image}} = \frac{\sum_{x=1}^{x=N} \sum_{y=1}^{y=M} |C|^2}{\sum_{x=1}^{x=N} \sum_{y=1}^{y=M} |I|^2} \quad (4.5)$$

où C représente le plan de corrélation, I l'image cible et N, M représentent la taille des images.

La corrélation consiste à multiplier le spectre d'une image cible avec un filtre (comme nous l'avons dit précédemment). Dans la littérature plusieurs filtres ont été proposés, créés à partir des différents critères de quantification. Ils ont été validés par différentes applications [94]. Parmi ces filtres, on trouve le filtre de phase pure (filtre POF), présenté dans la suite de cette partie et défini à partir de l'efficacité optique. L'objectif était d'obtenir un filtre dont l'énergie du plan d'entrée et l'énergie du plan de sortie soient les mêmes. Ainsi si le capteur utilisé pour enregistrer la scène est suffisamment résolu, il le sera aussi pour enregistrer le plan de corrélation, ce qui n'est pas nécessairement le cas pour d'autres filtres de corrélation [94, 104].

4.2.3 La finesse du pic de corrélation

Plus une image cible et une image de référence seront proches, plus le pic de corrélation sera haut, fin et énergétique par rapport au plan de corrélation. De plus le plan de corrélation sera faiblement bruité. Pour caractériser cela, on calcule le rapport entre l'énergie contenue dans le pic de corrélation et l'énergie contenue dans le plan de corrélation. Ce ratio s'appelle le PCE [103] :

$$PCE = \frac{\text{énergie contenue dans le pic de corrélation}}{\text{énergie contenue dans le plan de corrélation}} \quad (4.6)$$

$$PCE = \frac{\sum_{x=x_0-t}^{x=x_0+t} \sum_{y=y_0-t}^{y=y_0+t} |C(x, y)|^2}{\sum_{x=1}^{x=N} \sum_{y=1}^{y=M} |C(x, y)|^2}$$

où x_0 et y_0 représentent la position du pic de corrélation. Pour prendre en compte les pics qui ne sont pas exactement centrés sur un pixel, on ajoute au pic de corrélation les valeurs des t premiers voisins. C représente le plan de corrélation.

4.2.4 Les courbes ROC

Les courbes ROC "*Receiver operating characteristic*" consistent à caractériser quantitativement un système selon sa sensibilité et sa spécificité. La sensibilité d'un système TPR , "*true positive rate*" se définit par le rapport entre le nombre de vrais positifs TP , "*true positive*" et la somme des vrais positifs et des faux négatifs FN , "*false negative*". La spécificité d'un système SPC se définit comme le rapport entre le nombre de vrais négatifs TN , "*true negative*" par rapport à la somme des faux positifs FP , "*false positive*" et des vrais négatifs. Ces ratios sont résumés sur la figure 4.3 aussi appelée matrice de confusion et dans l'équation 4.7 [105] :

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN} \quad (4.7)$$

$$SPC = \frac{TN}{N} = \frac{TN}{TN + FP}$$

Les courbes ROC sont souvent représentées dans un graphe représentant l'espace (*cf.* figure 4.4). La probabilité d'une détection aléatoire est représentée (droite en pointillés). Lorsque la courbe ROC (courbe cyan) est située au dessus, le système donne de meilleurs résultats qu'un système aléatoire. Si la courbe est située en dessous, le système est plus mauvais qu'une détection aléatoire... Lorsque la courbe atteint le point (1,1), le système est parfait.

	valeur réelle	
valeur mesurée	vrais positifs TP	faux positifs FP
	faux négatifs FN	vrais négatifs TN

FIGURE 4.3 – Valeurs calculées pour réaliser une courbe ROC

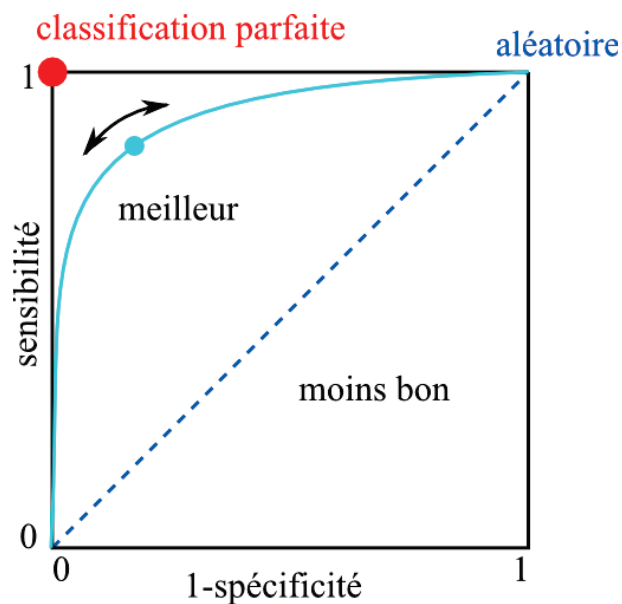


FIGURE 4.4 – Représentation de l'espace ROC

Pour tracer une courbe ROC, il faut avoir un système et un critère de décision. On fixe un seuil sur ce critère et on calcule la sensibilité et la spécificité sur un ensemble d'images à tester (par exemple) (point sur la courbe cyan de la figure 4.4). En faisant varier le seuil sur le critère (variation représentée par la double flèche sur la figure 4.4), on obtient suffisamment de valeurs pour tracer une courbe, la courbe ROC du système.

Il existe d'autres critères pour mesurer la qualité d'une corrélation [103]. Nous avons fait le choix d'utiliser le critère du PCE car il permet de calculer "facilement" le ratio entre l'information utile (le pic de corrélation) et le bruit présent dans le plan de corrélation. L'information utile et le bruit seront plus ou moins importants selon la ressemblance entre l'image cible et l'image de référence et selon le bruit présent dans le plan d'entrée. Lorsque nous utiliserons les courbes ROC, le PCE servira pour le seuillage.

Chapitre 5

La corrélation selon l'architecture VanderLugt

Sommaire

5.1	État de l'art	64
5.1.1	Le filtre adapté	64
5.1.2	Le filtre de phase pure	64
5.1.3	Le filtre de phase pure binaire	65
5.1.4	Le filtre inverse	66
5.1.5	Le filtre à compromis optimal	66
5.1.6	Le filtre composite	66
5.1.7	Le filtre segmenté	66
5.1.8	Comparaison des différents filtres	68
5.2	Comparaison de l'architecture VanderLugt avec la ZNCC	77
5.3	Le filtre ASPOF et le critère PCEA	78
5.3.1	Le principe du filtre	78
5.3.2	Adaptation du critère PCE	79
5.3.3	Comparaison du filtre ASPOF et des filtres composites et segmentés . .	80
5.4	Conclusion	83

La première méthode que nous présentons ici est l'architecture VanderLugt, créé en 1964 [100]. Il s'agit de comparer une image cible à une image de référence en effectuant une corrélation entre les deux images. L'image de référence est transformée en filtre de corrélation. En pratique, le spectre de l'image cible est obtenu via une transformée de Fourier. Il en est de même pour la référence. De façon à obtenir l'opération de corrélation, ce spectre est conjugué. Ces deux spectres sont multipliés afin d'obtenir par Transformée de Fourier inverse le plan de corrélation. La présence d'un pic dans le plan de corrélation témoigne de la ressemblance entre les images cible et référence : plus ce pic est important plus les images se ressemblent. De plus, la position du pic de corrélation donne la position de l'objet dans l'image cible (lorsque l'objet référence est centré dans l'image de référence). Le schéma synoptique de l'architecture proposée par Vander-Lugt est représenté sur la figure 5.1.

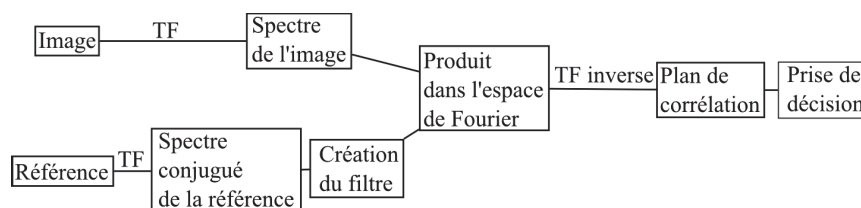


FIGURE 5.1 – Schéma synoptique du corrélateur proposé par Vander-Lugt

5.1 État de l'art

Comme indiqué sur le schéma 5.1, après la conjugaison du spectre de la référence, on crée un filtre de corrélation. Il est possible d'utiliser le spectre de la référence. Mais il est aussi possible de le modifier pour donner des caractéristiques particulières au montage (comme augmenter la robustesse, augmenter la discrimination, réduire le bruit). Ces modifications sont effectuées en ajoutant des informations dans le filtre (le bruit dans le cas du filtre adapté) ou en modifiant l'expression du spectre du filtre (modification de l'amplitude du spectre dans le cas du filtre POF). Les principaux filtres utilisés sont détaillés ci-dessous [94,97].

5.1.1 Le filtre adapté

Dans le cas de l'architecture présentée par Vander-Lugt [100], l'objectif est de maximiser le SNR, dans le cas d'une scène perturbée par un bruit additif. Dans ce cas, l'expression mathématique de la scène devient :

$$s(x, y) = o(x, y) + b(x, y) \quad (5.1)$$

où s représente la scène à analyser (ou image cible), o l'objet à reconnaître et b le bruit. Le filtre adapté h qui permet d'optimiser la reconnaissance de l'objet o s'exprime dans le domaine de Fourier par la relation :

$$H(\mu, \nu) = \frac{O^*(\mu, \nu)}{\Gamma_b(\mu, \nu)} \quad (5.2)$$

où H représente le filtre adapté de corrélation dans le domaine de Fourier, O^* la transformée de Fourier conjuguée de l'objet et Γ_b la densité spectrale du bruit. Dans le cas d'un bruit blanc, la densité spectrale du bruit est supposée égale à "1" et le filtre de corrélation le conjugué de l'objet.

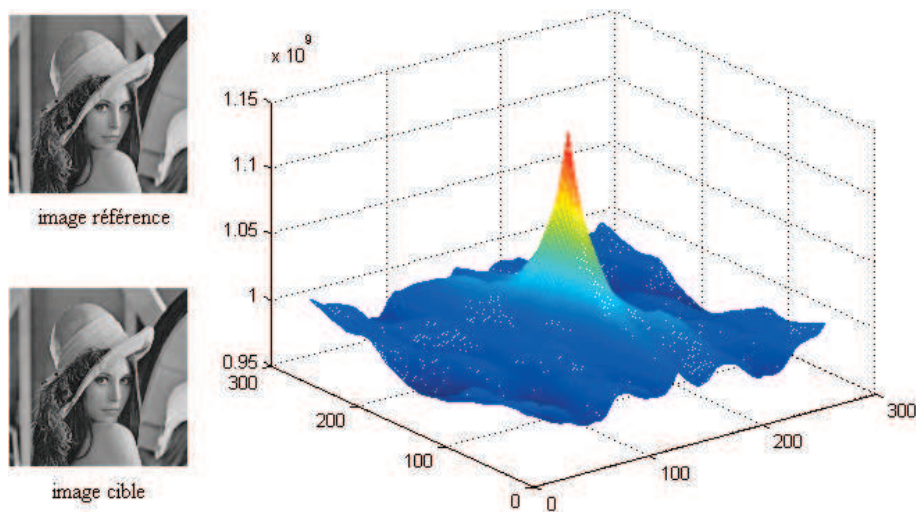


FIGURE 5.2 – Autocorrélation de Lena avec un filtre adapté

L'avantage de ce filtre est d'améliorer le SNR surtout en présence de bruit. Cependant, les pics de corrélation obtenus avec ce genre de filtres sont très larges (*cf.* figure 5.2). La conséquence de cette largeur de pic est la faible discrimination du filtre et par conséquent la faible précision de la position de l'objet (en effet, plus le pic est large plus la position de l'objet est difficile à déterminer précisément). Ce filtre est aussi sensible au changement d'échelle et à la rotation de l'objet.

5.1.2 Le filtre de phase pure

Dans un signal ou une image, l'œil humain est sensible aux contours. Ces contours sont principalement définis par la phase du signal correspondant à l'image. L'importance de la phase a été

démontrée par Oppenheim et Lim [67]. Utilisant ces constats, le filtre POF, “*Phase Only Filter*” (filtre de phase pure), créé par Horner et Gianino [106] prend en compte uniquement la phase du signal. Ce filtre a donc pour objectif d’améliorer l’efficacité optique : en n’utilisant que la phase de l’image de référence, l’énergie du signal n’est pas absorbée. Ce filtre s’écrit :

$$H(\mu, \nu) = \frac{O^*(\mu, \nu)}{|O(\mu, \nu)|} = \frac{\rho_O(\mu, \nu)e^{-i\theta_O(\mu, \nu)}}{|\rho_O(\mu, \nu)e^{-i\theta_O(\mu, \nu)}|} = e^{-i\theta_O(\mu, \nu)} \quad (5.3)$$

où $\rho_O(\mu, \nu)e^{-i\theta_O(\mu, \nu)}$ est une écriture du signal représentant l’objet dans le domaine de Fourier avec $\rho_O(\mu, \nu)$ l’amplitude du signal et $\theta_O(\mu, \nu)$ la phase.

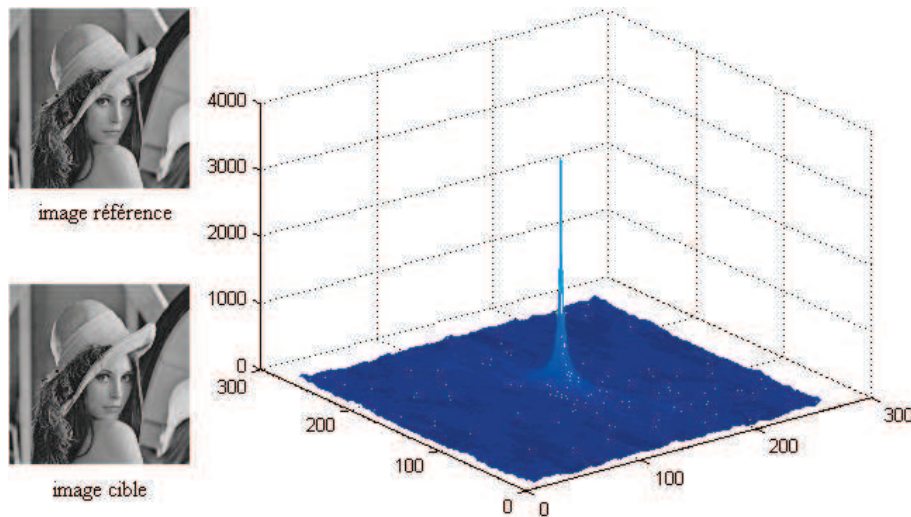


FIGURE 5.3 – Autocorrélation de Lena avec un filtre POF

Ce filtre a l’avantage de donner des pics de corrélation étroits (*cf.* figure 5.3) et d’être très discriminant. Grâce à l’utilisation de la phase, il peut servir de détecteur de contours. L’inconvénient de ce filtre est sa sensibilité au bruit, à l’échelle et à la rotation [107].

Cependant, ce filtre reste très intéressant pour notre étude, car il donne des pics de corrélation fins et d’amplitude élevée lorsqu’il y a reconnaissance, ce qui a pour objectif de faciliter la prise de décision.

5.1.3 Le filtre de phase pure binaire

Pour certaines applications où le filtre ne peut être enregistré sur toute la plage de valeurs, il existe le filtre BPOF, “*binary phase only filter*” (filtre de phase pure binaire) [108]. Pour obtenir ce filtre binaire, on part du filtre de phase pure. Horner *et al.* [108] proposent d’étudier le signe de la partie réelle du filtre POF (il est possible d’utiliser d’autres critères) :

$$H(\mu, \nu) = e^{-i\theta_O(\mu, \nu)} \quad (5.4)$$

$$\begin{cases} H = 1 & \text{si } \text{real}(e^{-i\theta_O(\mu, \nu)}) \geq 0 \\ H = -1 & \text{sinon} \end{cases}$$

Ce filtre présente les mêmes avantages et inconvénients que le filtre POF. Initialement ce filtre a été créé pour les implémentations optiques. En effet, les modulateurs (interfaces utilisées pour introduire les filtres dans le montage optique) utilisés pour afficher le filtre dans le montage optique ne pouvaient enregistrer que deux valeurs (modulateurs rapides à deux niveaux). Dans le cas d’une implémentation numérique, le fait de n’utiliser que deux valeurs permet d’obtenir une taille (en terme de place mémoire) de filtre très faible. Il est donc intéressant pour des applications embarquées.

5.1.4 Le filtre inverse

Le filtre inverse [109] maximise le PCE. Cela a pour effet de donner des pics de corrélation très fins. Le filtre est très discriminant mais très sensible au bruit. L'expression de ce filtre est :

$$H(\mu, \nu) = \frac{O^*(\mu, \nu)}{|O(\mu, \nu)|^2} = \frac{\rho_O v e^{-i\theta_O(\mu, \nu)}}{|\rho_O(\mu, \nu) e^{-i\theta_O(\mu, \nu)}|^2} = \frac{e^{-i\theta_O(\mu, \nu)}}{\rho_O(\mu, \nu)} \quad (5.5)$$

La sensibilité au bruit de ce filtre le rend peu utilisable dans notre étude où les images cibles sont bruitées par le milieu sous marin, contrairement à nos images de référence qui ne sont pas bruitées car fabriquées à partir d'images de synthèse.

5.1.5 Le filtre à compromis optimal

Le filtre adapté maximise le SNR et résiste au bruit. Le filtre inverse maximise le PCE pour être très discriminant. L'objectif du filtre OT, "*optimal trade-off filter*" [110] est de trouver le compromis entre le filtre adapté et le filtre inverse :

$$H(\mu, \nu) = \frac{O^*(\mu, \nu)}{\alpha \Gamma_b(\mu, \nu) + (1 - \alpha) |O(\mu, \nu)|^2} \quad (5.6)$$

$$H(\mu, \nu) = \frac{\rho_O(\mu, \nu) e^{-i\theta_O(\mu, \nu)}}{\alpha \Gamma_b(\mu, \nu) + (1 - \alpha) |\rho_O(\mu, \nu)|^2}$$

Si le paramètre α vaut 0, on retrouve le filtre inverse. S'il vaut 1, on retrouve le filtre adapté. Dans les cas intermédiaires, le filtre est résistant au bruit et discriminant.

5.1.6 Le filtre composite

Les filtres que nous venons de présenter comparent une image cible à une image de référence. Lorsque l'on souhaite une application qui nécessite un nombre important de comparaisons, une solution consiste à multiplexer les références de manière à réduire le nombre d'opérations à effectuer [111, 112]. Cette réduction du nombre d'opérations à effectuer entraîne une réduction du temps de calcul nécessaire pour prendre une décision.

La première méthode de multiplexage consiste à additionner les références. Il est possible d'introduire une pondération :

$$H(\mu, \nu) = a.R_1(\mu, \nu) + b.R_2(\mu, \nu) + c.R_3(\mu, \nu) \quad (5.7)$$

où a , b et c représentent les poids que l'on peut ajouter pour favoriser l'une des références. La méthode est représentée avec un exemple sur la figure 5.4. (Comme la transformée de Fourier est linéaire, la figure représente la méthode dans le domaine spatial pour plus de visibilité)

Ici, le multiplexage a été effectué sous Matlab. Le nombre de bits utilisé pour l'image composite n'est pas limité (64 bits par pixels). Lorsque le nombre d'images à multiplexer est important et que chaque pixel de l'image est enregistré sur 8 bits par exemple, un phénomène de saturation apparaît (*cf.* figure 5.5). Le filtre n'est alors plus discriminant. La limitation de la taille des pixels est nécessaire lorsque l'application est destinée à être implémentée optiquement (en utilisant des modulateurs à plusieurs niveaux de gris comme par exemple les modulateurs SEIKO) ou numériquement (dans le cas d'applications embarquées). Elle est intéressante lorsque l'on souhaite obtenir des filtres de petites tailles (dans les cas où la mémoire de stockage est limitée par exemple).

5.1.7 Le filtre segmenté

Pour remédier au problème de saturation lorsque la taille du filtre est limitée, il est possible de segmenter le filtre [113]. Cela signifie qu'un pixel du filtre est associé à une seule référence et non plus à toutes les références. Le critère de sélection de la référence gagnante pour ce pixel est l'énergie. Cela revient à créer un masque qui dépend de chaque référence et à faire la somme



FIGURE 5.4 – Multiplexage de références

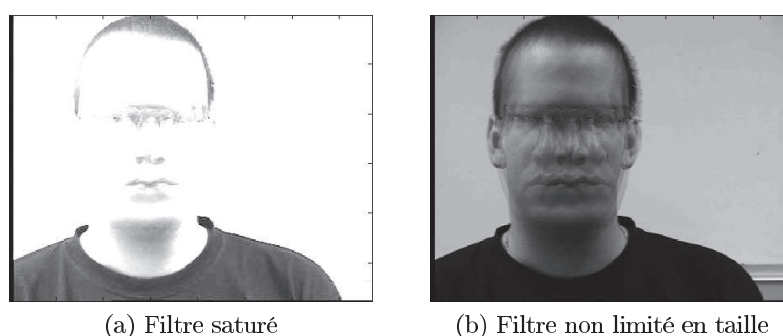


FIGURE 5.5 – Comparaison d'un filtre limité en taille et d'un filtre non limité en taille

des références multipliées avec leur masque. La méthode est schématiquement représentée sur la figure 5.6. Un exemple est présenté sur la figure 5.7. Mathématiquement, on calcule pour chacune des images de référence, l'énergie relative du pixel (i, j) . Le pixel qui a l'énergie relative la plus importante est sélectionné. La même opération est effectuée pour tous les pixels. Il est possible de trouver des pixels isolés (un seul pixel appartenant à la référence i parmi les pixels du voisinage), comme on peut le voir sur la figure 5.7. Plus le nombre de références à multiplexer est élevé plus le nombre de pixels isolés est élevé. Il est alors possible que le filtre ne soit pas très efficace du fait que peu de pixels consécutifs appartiennent à la même référence.

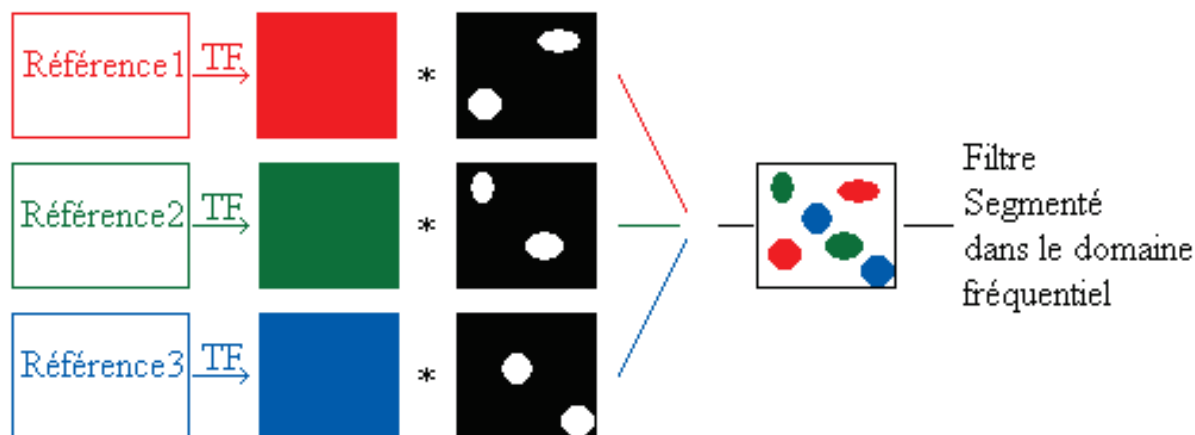


FIGURE 5.6 – Création d'un filtre segmenté

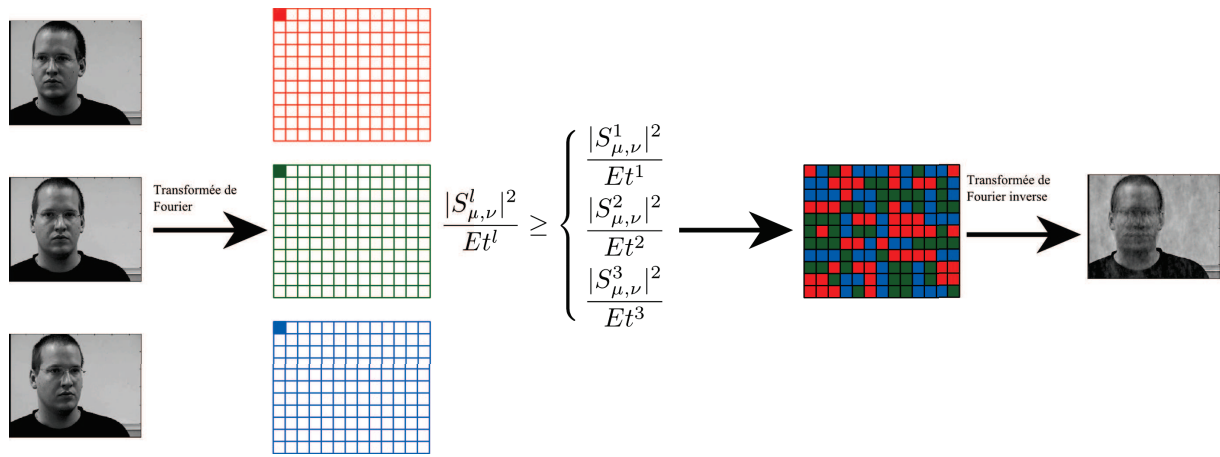


FIGURE 5.7 – Création d'un filtre segmenté à partir de 3 références

Les 3 références utilisées ici n'ont ni les mêmes énergies totales ni les mêmes énergies relatives, ce qui explique ce fond "texturé".

5.1.8 Comparaison des différents filtres

Pour faire notre propre choix et pour tester les avantages et inconvénients des différents filtres (dans le but de proposer une optimisation pour notre application), nous avons testé les différents filtres présentés (adapté, POF, BPOF, inverse, OT). Nous avons créé des filtres contenant une référence et des filtres composites et segmentés contenant trois à trente-sept références. Nous avons choisi, dans un premier temps, comme base d'images des images binaires représentant la lettre A. Ce choix a été effectué car il est facile de calculer les images résultantes d'une rotation de la lettre initiale. Tout d'abord nous avons comparé des images sans bruit afin d'étudier la robustesse des filtres à la rotation de la lettre. Afin de tester la robustesse au bruit, nous avons rajouté du bruit texturé dans le fond et/ou dans la lettre. Nous appelons bruit texturé le bruit qui représente un motif (tronc d'arbre, nuages, cf. 1ère ligne du tableau 5.1). Nous avons aussi testé la robustesse au bruit blanc (cf. 1ère ligne du tableau 5.1). Le bruit blanc est un bruit gaussien dont nous fixons la moyenne et la variance selon le niveau de bruit souhaité. Le facteur d'échelle a été testé avec des lettres deux fois plus petites que les lettres références. Enfin nous avons testé la discrimination des filtres en comparant la lettre A à la lettre V (nous avons effectué ce choix car les deux lettres sont proches, surtout avec les rotations).

Dans le tableau 5.1 sont présentés les résultats obtenus avec les filtres composites. Dans un premier temps, arrêtons nous sur les résultats obtenus avec le filtre adapté dans le cas de la corrélation de lettres binaires non bruitées (cf. figure 5.8). Sur cette figure, l'axe des abscisses représente l'angle de rotation de la lettre testée (entre -90° et $+90^\circ$). L'axe des ordonnées représente la valeur du PCE du plan de corrélation. Pour tester nos filtres, nous avons créés 9 filtres composites adaptés :

- à 1 référence pour reconnaître la lettre orientée à 0° (cas que nous appellerons par la suite autocorrélation)
- à 3 références pour reconnaître les lettres orientées entre -5° et $+5^\circ$
- à 5 références pour reconnaître les lettres orientées entre -15° et $+15^\circ$
- à 11 références pour reconnaître les lettres orientées entre -45° et $+45^\circ$
- à 15 références pour reconnaître les lettres orientées entre -65° et $+65^\circ$
- à 21 références pour reconnaître les lettres orientées entre -85° et $+85^\circ$
- à 25 références pour reconnaître les lettres orientées entre -85° et $+85^\circ$
- à 31 références pour reconnaître les lettres orientées entre -85° et $+85^\circ$
- à 37 références pour reconnaître les lettres orientées entre -90° et $+90^\circ$

Ces différents filtres composites ont été créés afin d'étudier la robustesse des architectures à

la rotation selon le nombre d'images de référence dans le filtre. Nous pouvons constater sur la figure 5.8 que les PCE obtenus avec les filtres jusqu'à 11 références permettent d'obtenir une reconnaissance des images contenues dans les intervalles choisis.

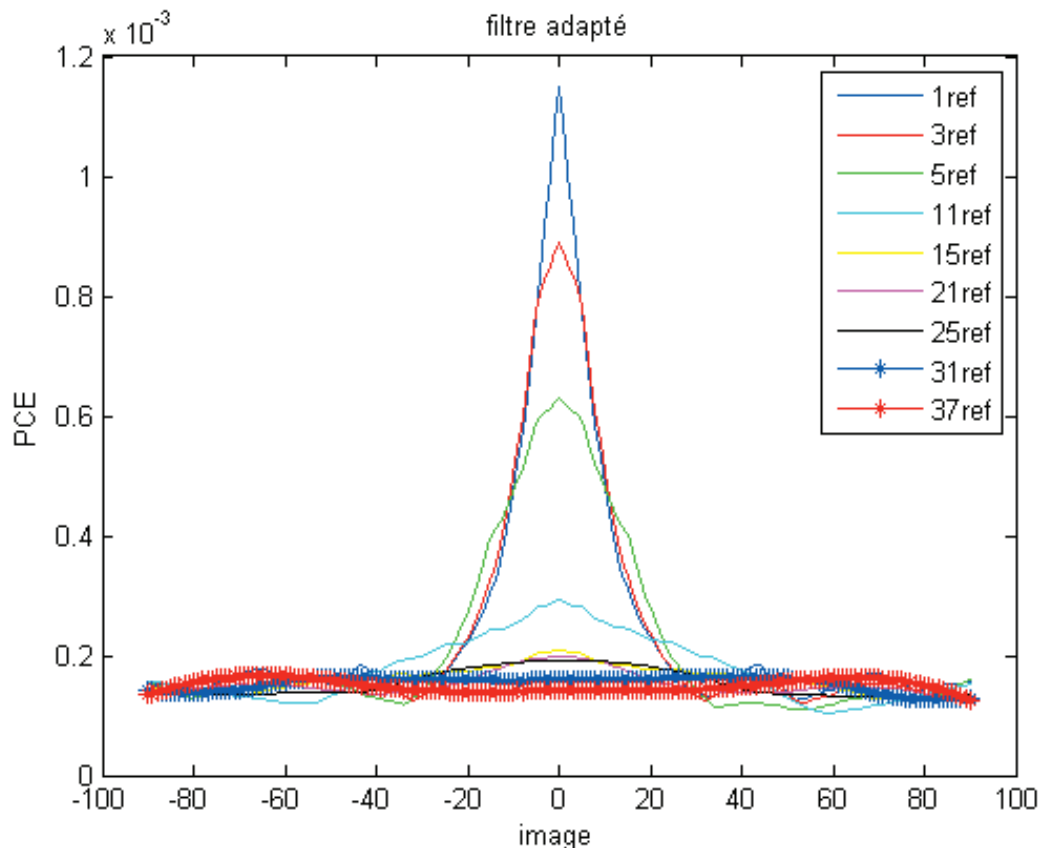


FIGURE 5.8 – Résultats obtenus avec les différents filtres composites adaptés testés dans le cas de la corrélation de lettres binaires non bruitées



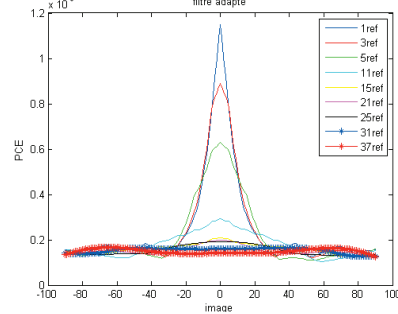
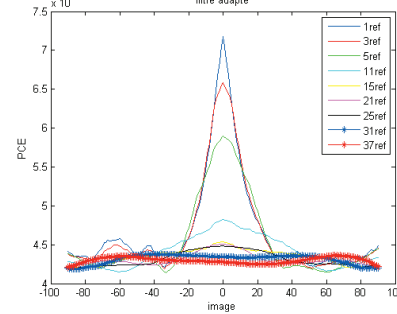
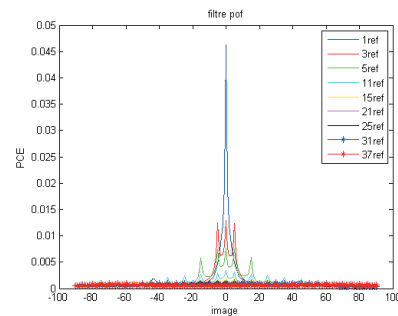
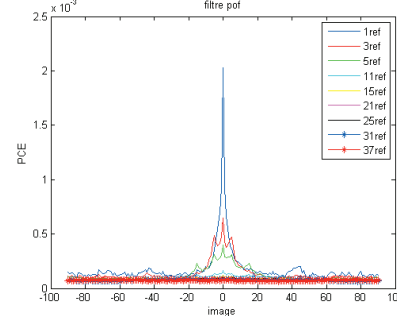
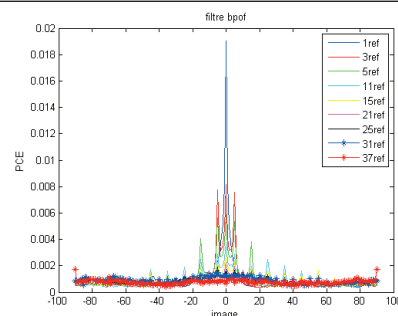
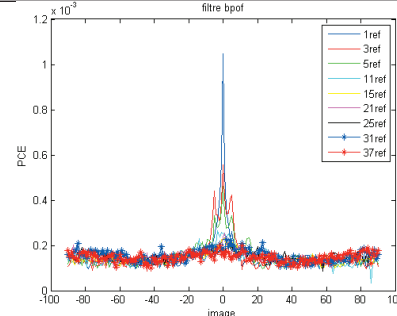
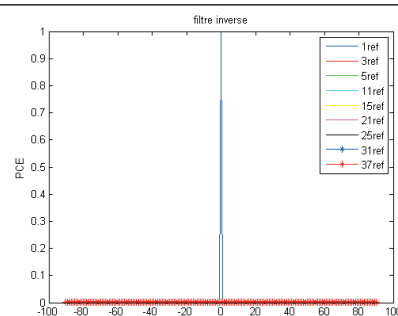
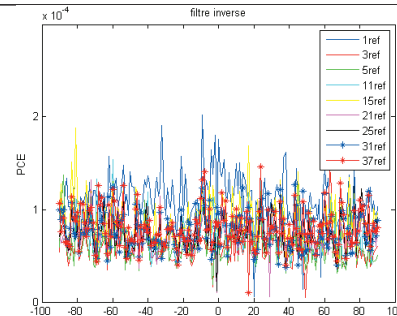
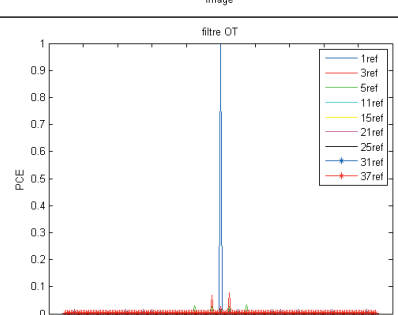
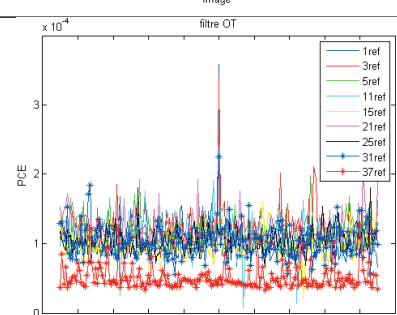
Les filtres adaptés, POF et BPOF composites sont robustes à la rotation et au bruit, aussi bien bruit blanc que bruit texturé. Ces filtres sont sensibles à l'échelle et discriminants. Les filtres inverses et OT sont beaucoup plus sensibles.

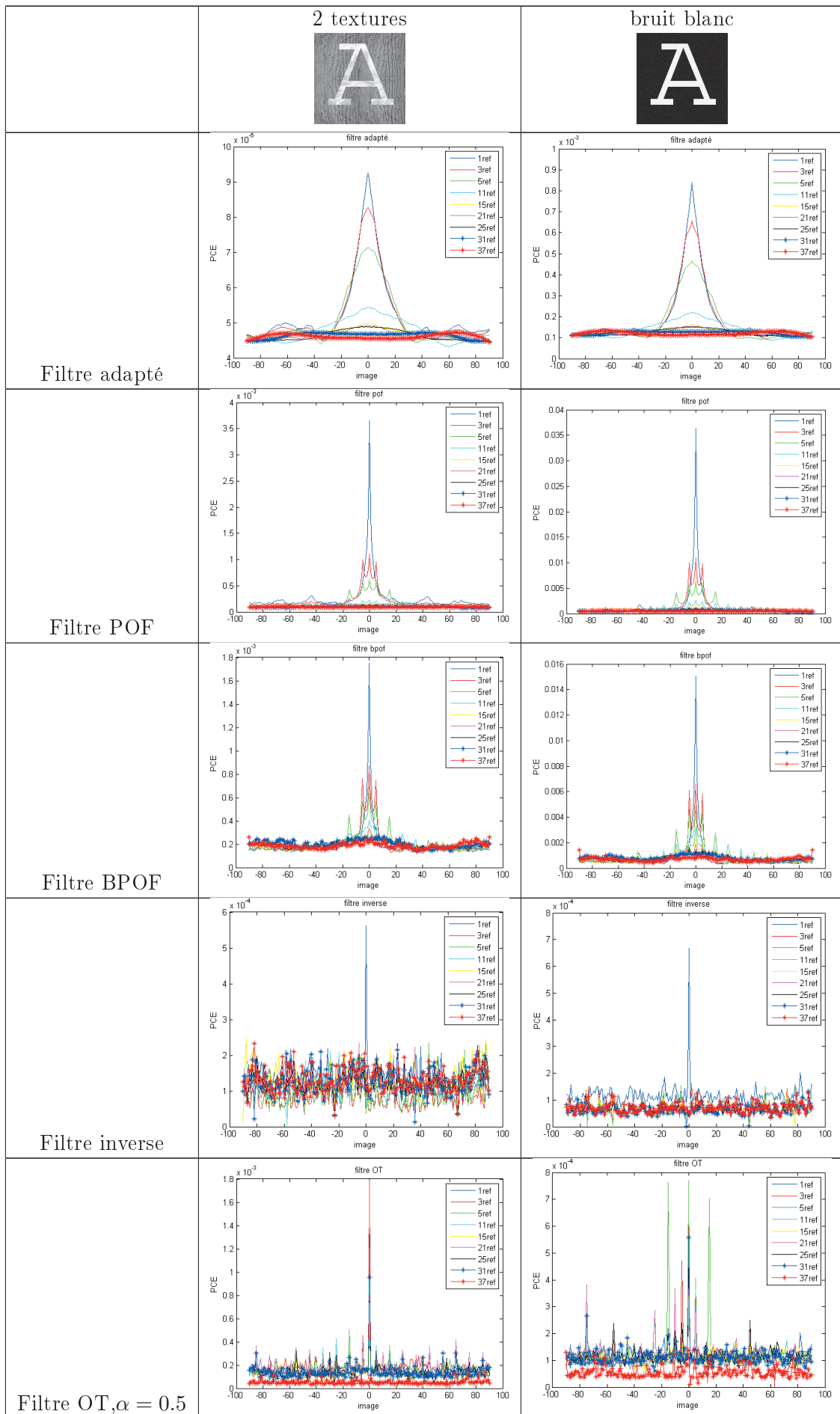
L'étude de la robustesse à la rotation (1ère colonne du tableau 5.1) montre que les filtres adaptés, POF et BPOF reconnaissent les lettres de l'intervalle représenté par les filtres contenant jusqu'à onze références. Nous remarquons aussi que les valeurs du PCE diminuent lorsque le nombre de références comprises dans le filtre augmente. Le filtre inverse ne reconnaît que le résultat de l'autocorrélation. Le filtre OT reconnaît très faiblement les lettres contenues dans les filtres composés de trois et cinq références.

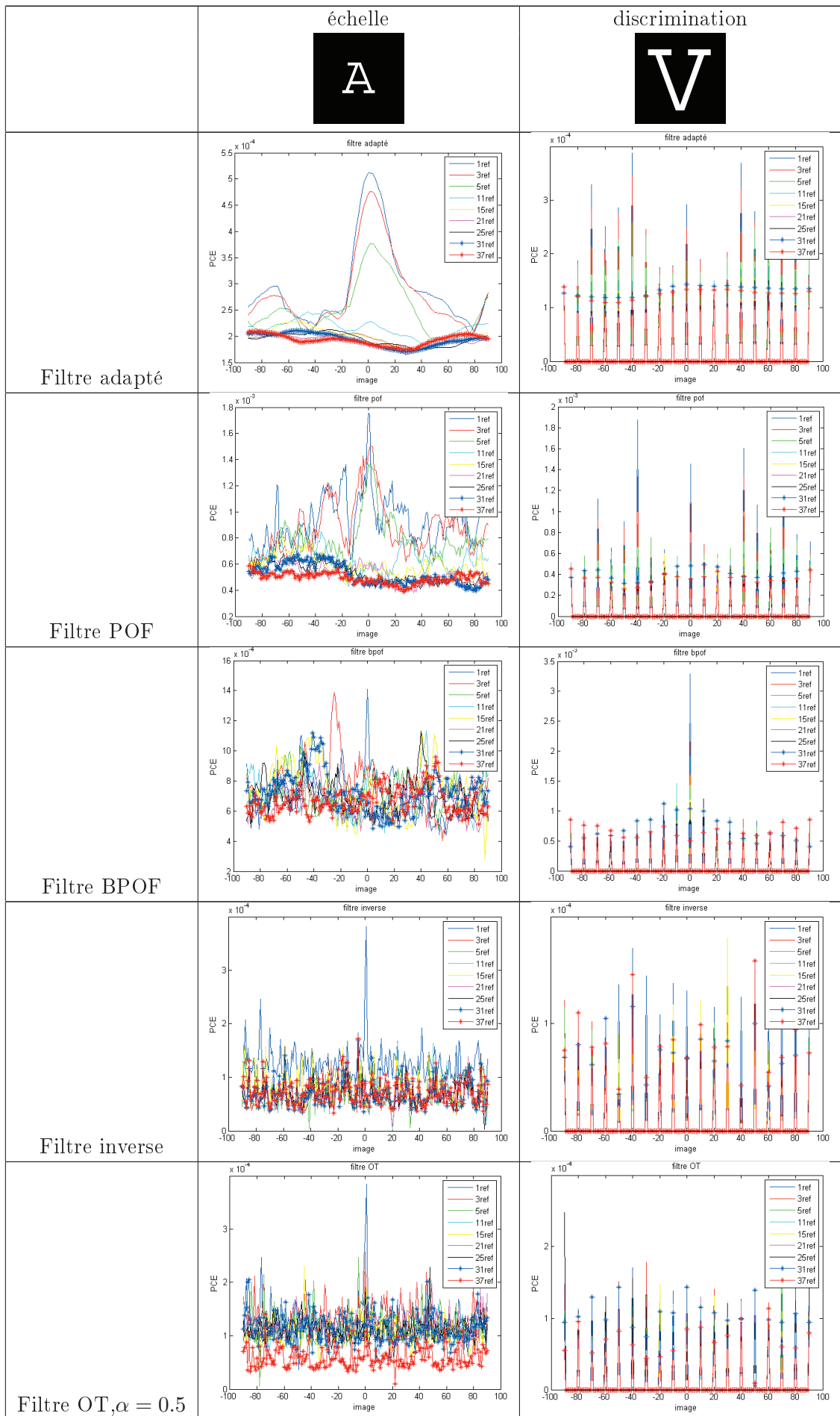
Les tests de la résistance au bruit, que ce soit un bruit blanc ou un bruit texturé dans le fond et dans la lettre, montrent que les valeurs du PCE sont moins élevées que dans le cas de la corrélation entre lettres non bruitées. Les plans de corrélation sont donc plus bruités que dans le cas précédent. Cependant, les filtres adaptés, POF et BPOF reconnaissent les lettres contenues dans l'intervalle décrit par les filtres composés de trois à cinq références. Le filtre inverse ne reconnaît l'autocorrélation que dans le cas où il y a deux bruits texturés dans l'image et dans le cas du bruit blanc. Le filtre OT reconnaît le résultat de l'autocorrélation dans le cas de bruits texturés et quelques références contenues dans les filtres dans le cas de bruit blanc.

Le facteur d'échelle et le changement de lettre n'entraînent pas de reconnaissance par les différents filtres sauf dans le cas de l'autocorrélation entre échelle pour les filtres inverse et OT et dans le cas du filtre BPOF entre la lettre A et la lettre V orientées à 0° .

Tableau 5.1 – Comparaison des filtres composites

	Corrélation 	bruit texturé 
Filtre adapté		
Filtre POF		
Filtre BPOF		
Filtre inverse		
Filtre OT, $\alpha = 0.5$		





Dans le tableau 5.2 sont présentés les résultats obtenus avec les filtres segmentés. Les résultats sont similaires à ceux du filtre composite pour les filtres adaptés, POF et BPOF. Le filtre inverse et le filtre OT reconnaissent davantage de lettres.

Pour ce qui est de la robustesse à la rotation, les filtres segmentés adaptés, POF et BPOF reconnaissent les lettres contenues dans les filtres composés de onze références au maximum. Les filtres inverse et OT reconnaissent les lettres contenues dans les filtres composés au maximum de cinq lettres.

Les mêmes lettres sont reconnues par les filtres segmentés adaptés, POF et BPOF en présence de bruit texturé ou bruit blanc. Les filtres inverse et OT reconnaissent quelques lettres contenues dans les filtres.

Tous les filtres sont sensibles à l'échelle et discriminants. Les lettres ne sont pas reconnues sauf quelques rares exceptions de lettres orientées avec un faible angle par rapport à la lettre initiale orientée à 0° .

Ces deux tableaux 5.1 et 5.2 nous permettent de constater que les filtres les plus robustes et discriminants sont les filtres adaptés, POF et BPOF, aussi bien en filtres composites qu'en filtres segmentés. En observant les plans de corrélation des filtres adaptés, et comme nous l'avons précisé précédemment, les pics de corrélation sont très larges ce qui implique une faible précision sur la position de la mine. Il nous reste donc à choisir le filtre POF ou le filtre BPOF. Les résultats des tableaux montrent que les filtres POF permettent d'obtenir des valeurs de PCE plus élevées que les valeurs obtenues avec un filtre BPOF en cas de reconnaissance. De plus, nous ne sommes pour l'instant pas limité par la taille des filtres. (Cette limitation pourra intervenir si l'algorithme développé est intégré à un drone sous marin). Nous choisissons donc le filtre POF. Cependant, les tableaux permettent difficilement de comparer le filtre composite et le filtre segmenté. Pour faciliter la comparaison entre le filtre composite et le filtre segmenté, sur la figure 5.9, nous avons affiché le filtre POF composite et le filtre POF segmenté, contenant chacun 10 références. Sur cette figure, nous pouvons remarquer que les deux filtres reconnaissent les images orientées entre -20° et $+25^\circ$ (zone B de l'image). Les zones A et C de la courbes correspondent aux images qui ne sont pas reconnues par les filtres. Le filtre POF segmenté permet d'obtenir des valeurs de PCE nettement plus élevées que les valeurs du filtre POF composite. L'écart entre reconnaissance et non reconnaissance est important. Il est ainsi facile de fixer un seuil pour décider automatiquement de la reconnaissance ou non de la lettre.

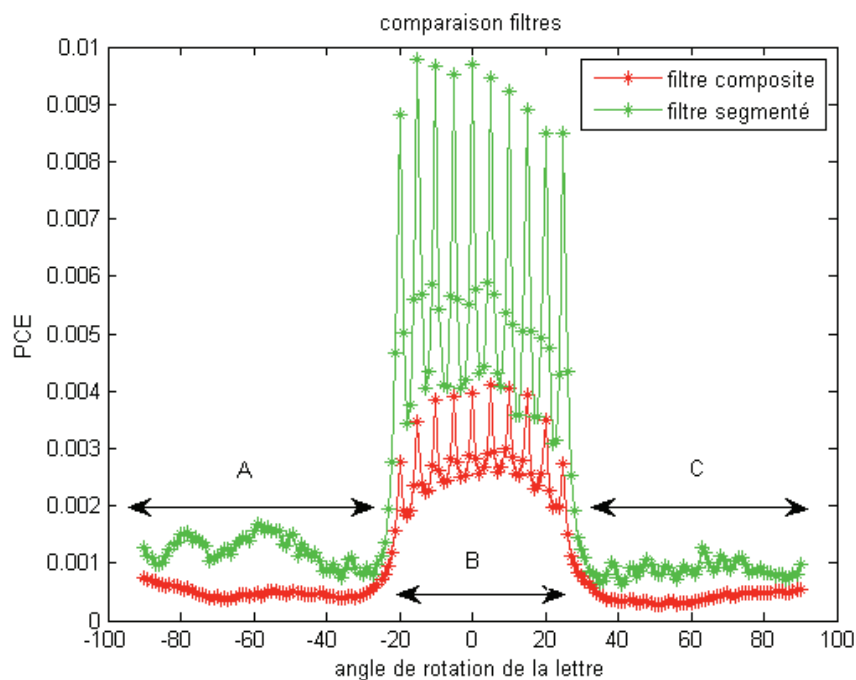


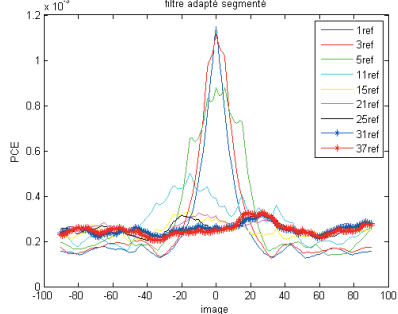
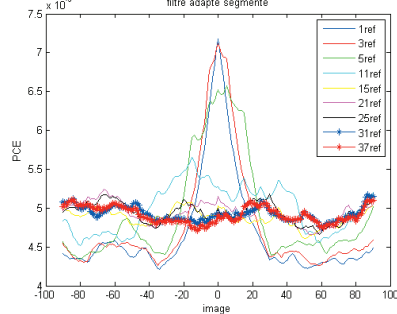
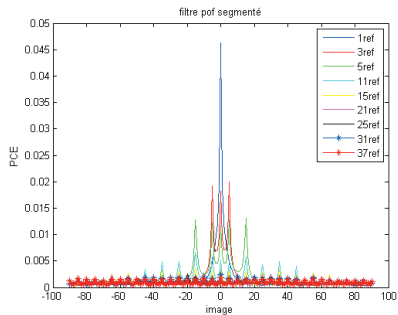
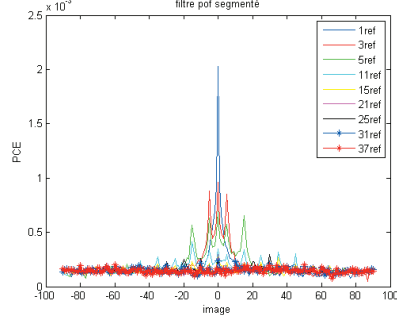
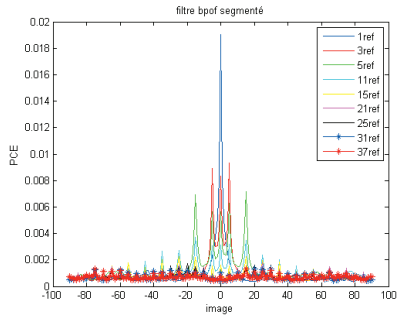
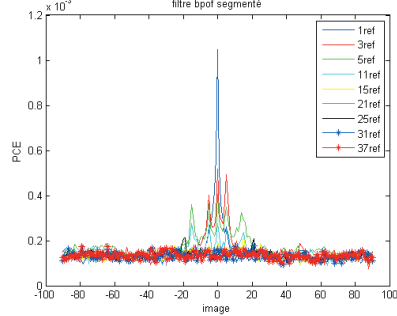
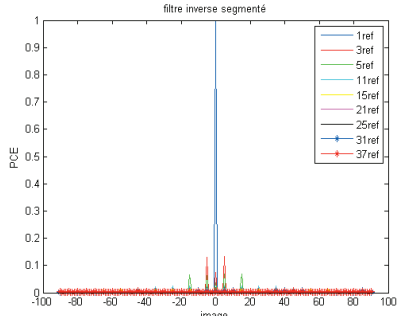
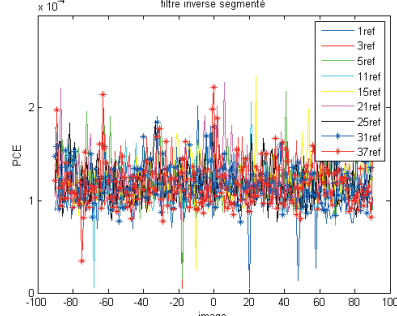
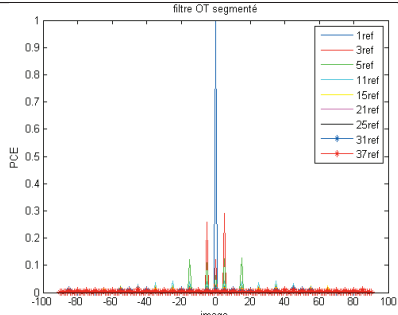
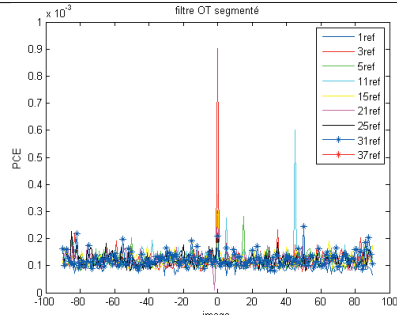
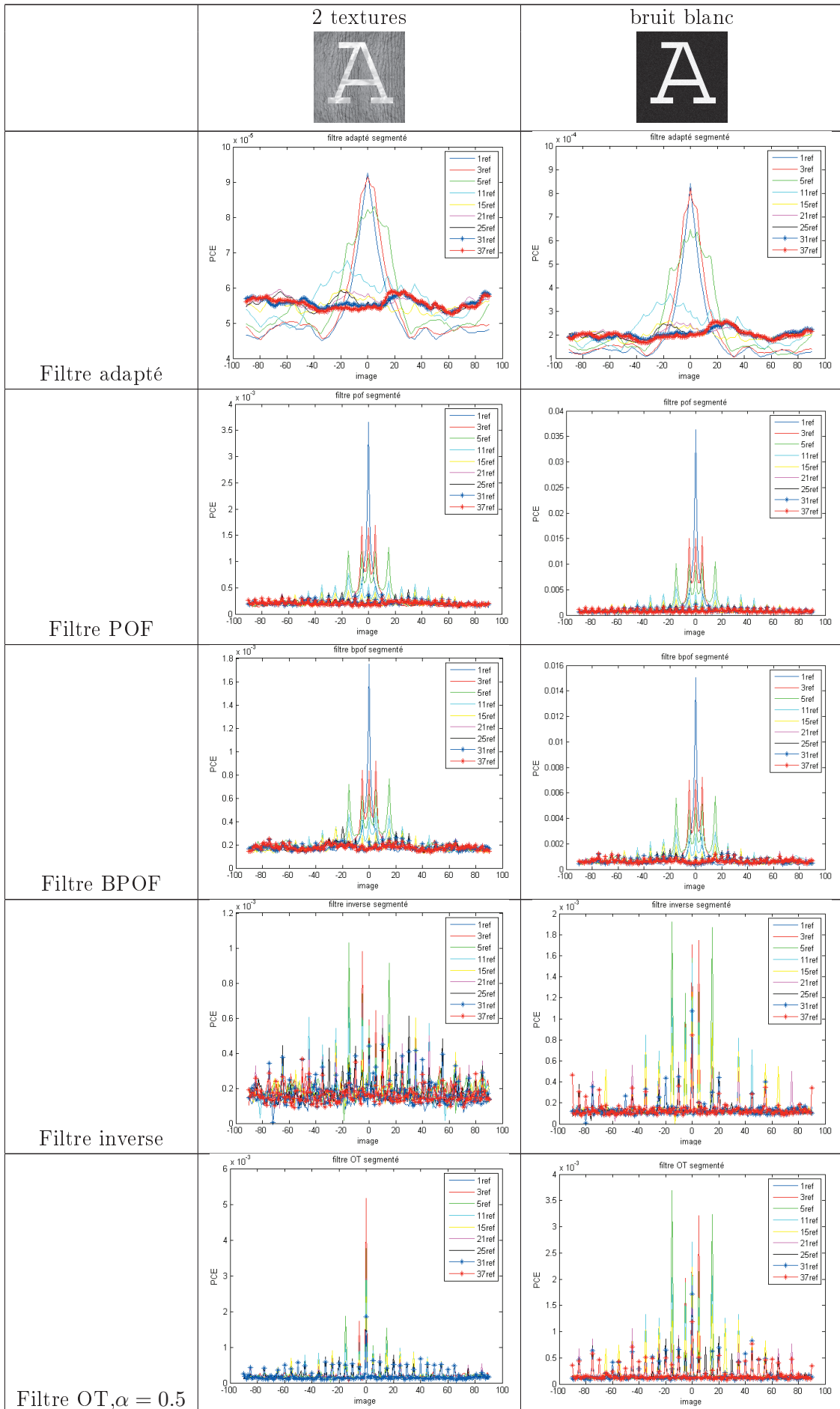
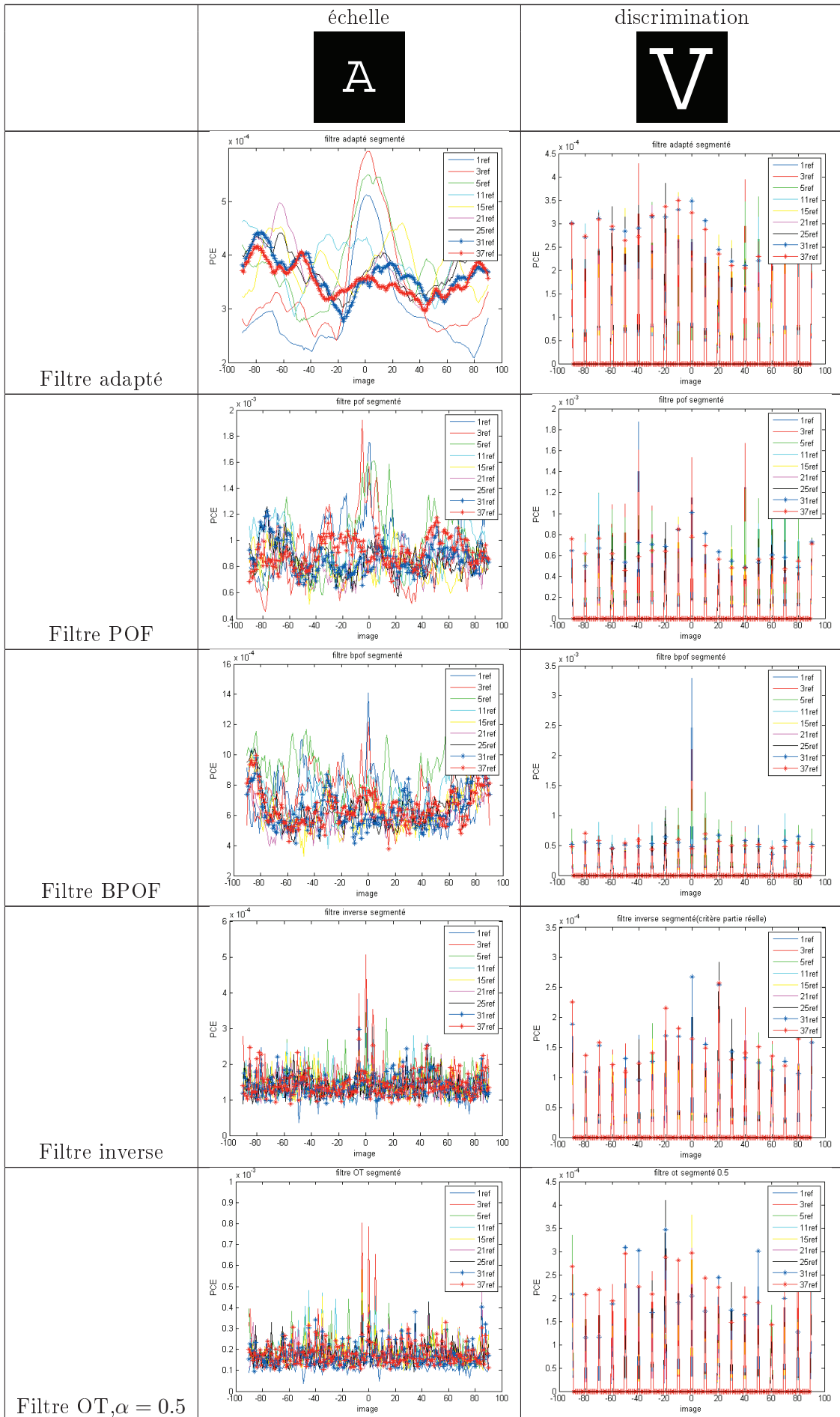


FIGURE 5.9 – Comparaison des filtres POF composite et segmentés contenant 10 références

Tableau 5.2 – Comparaison des filtres segmentés

	Corrélation 	bruit texturé 
Filtre adapté		
Filtre POF		
Filtre BPOF		
Filtre inverse		
Filtre OT, $\alpha = 0.5$		





5.2 Comparaison de l'architecture VanderLugt avec la ZNCC, "zero mean normalized cross correlation"

Nous avons vu que l'architecture VanderLugt permet de reconnaître les lettres avec une robustesse à la rotation et au bruit. Nous avons ensuite voulu comparer cette architecture à la méthode de la corrélation numérique. Pour cela nous avons choisi d'implémenter la ZNCC, "zero mean normalized cross correlation".

L'architecture présentée dans cette partie consiste à soustraire à l'image cible et à l'image de référence leur moyenne locale et à diviser le produit par l'écart-type de ces images [114] :

$$c(x, y) = \frac{1}{N * M} \sum_{x=1}^{x=N} \sum_{y=1}^{y=M} \frac{(r(x, y) - \bar{r})(i(x, y) - \bar{i})}{\sigma_r \sigma_i} \quad (5.8)$$

où r , i , \bar{r} , \bar{i} , σ_r et σ_i représentent respectivement l'image de référence et l'image cible, leur moyenne locale et leur écarts-types.

Nous avons testé cette méthode comparativement au filtre POF de l'architecture VanderLugt (cf. partie 5.1.2). Les résultats obtenus sont présentés dans le tableau 5.3. Dans l'architecture VanderLugt, le filtre doit avoir la même taille que l'image (première ligne du tableau). Dans le cas de la ZNCC, le filtre est un motif de taille inférieure. Nous avons donc découpé le visage de Lena et la mine pour en faire des motifs pour la ZNCC (troisième ligne du tableau). Ces motifs ont été utilisés pour créer des filtres POFs, architecture appelée POF modifié, dans la deuxième ligne du tableau. Les images de test sont présentées dans la première ligne du tableau. Nous avons utilisé le PCE (cf. partie 4.2.3) pour comparer les résultats obtenus. Les filtres POF présentent des pics de corrélation nets et facilement repérables par rapport au reste du plan de corrélation, ce qui n'est pas le cas de la ZNCC, sauf dans le cas de la comparaison de Lena avec elle-même. Ces résultats visuels sont confirmés par les valeurs du PCE calculés sur ces plans de corrélation.




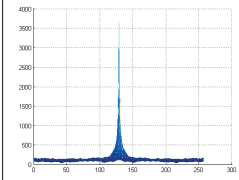
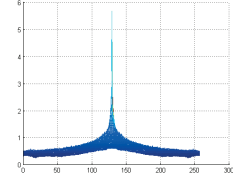
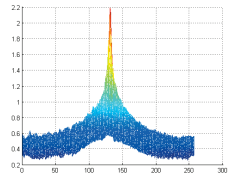
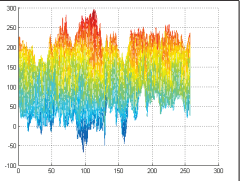
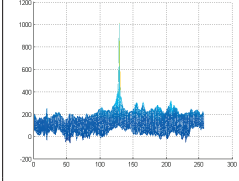
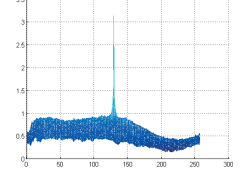
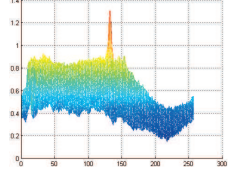
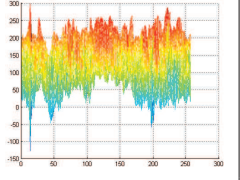
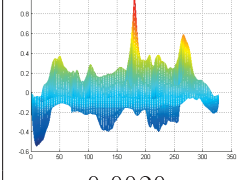
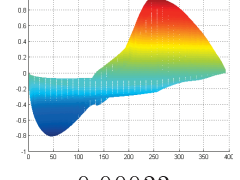
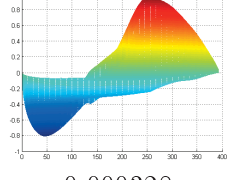
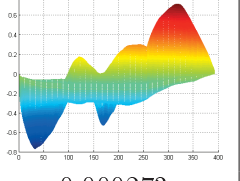
	Lena-Lena 	1082-1082 	1082-1090 	1082-Lena
POF PCE	 0.0327	 0.0041	 0.0019	 0.0006
POF modifié PCE	 0.0032	 0.0012	 0.000679	 0.000515
ZNCC PCE	 0.0020	 0.00033	 0.000328	 0.000273

Tableau 5.3 – Tableau comparant le filtre POF et la ZNCC

5.3 Le filtre ASPOF et le critère PCEA

Comme nous l'avons vu précédemment (*cf.* figure 5.9), le filtre segmenté semble meilleur que le filtre composite. Cependant, lorsque le nombre de références utilisées augmente, on peut remarquer une baisse des performances due à la présence de pixels isolés dans le filtre. La comparaison de l'architecture VanderLugt avec la méthode numérique de la ZNCC a montré que l'architecture VanderLugt présente de meilleurs résultats. Notre choix s'est donc porté sur cette architecture. Cependant, nous allons essayer de l'optimiser de façon à introduire dans les filtres un nombre plus important d'images de référence, tout en gardant un bon pouvoir discriminant. Pour garder un bon pouvoir discriminant avec l'architecture VanderLugt, nous avons essayé de limiter le nombre de références dans les filtres (de l'ordre de 5 références). Lorsque la base de données dans laquelle on cherche l'objet contenu dans l'image est très importante, il reste encore beaucoup de comparaisons à effectuer. Pour réduire le nombre d'opérations d'un facteur 2, nous avons eu l'idée de créer le filtre ASPOF, "*Asymmetric Segmented Phase-Only Filter*".

5.3.1 Le principe du filtre

Le concept du filtre ASPOF est de créer deux filtres POF segmentés comprenant chacun cinq références. Une possibilité est de prendre une référence sur deux pour créer le premier filtre et les références qu'il reste servent à créer le second filtre (*cf.* figure 5.10).

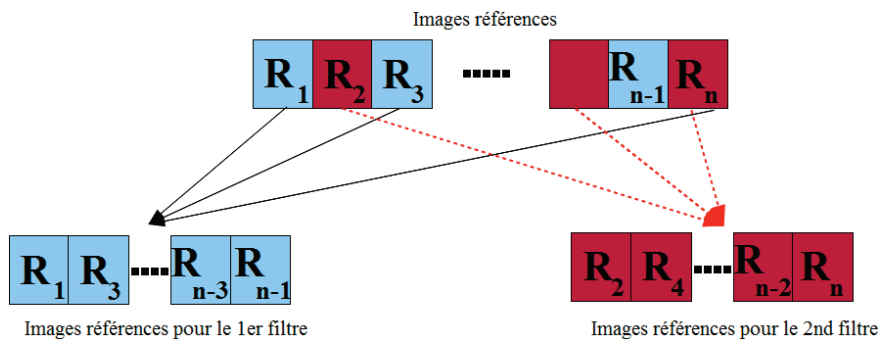


FIGURE 5.10 – Répartition de la base d'images de référence en deux sous bases d'images pour créer les filtres

Ensuite on multiplie le premier filtre segmenté par un filtre symétrique contenant une moitié de 0 et une moitié de 1. Le second filtre segmenté est multiplié par le filtre symétrique inverse au premier (une moitié de 1 et une moitié de 0). Les deux filtres obtenus sont additionnés pour obtenir un seul filtre (*cf.* figure 5.11). Il est possible de pondérer cette addition pour favoriser l'un des filtres segmentés.

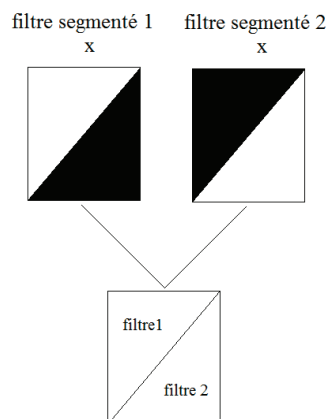


FIGURE 5.11 – Assemblage des deux filtres segmentés

Nous avons aussi cherché à gérer les pixels isolés que nous avons évoqué dans la partie 5.1.7. Pour cela, lors de la création des filtres segmentés POF, nous avons décidé de mettre un seuillage : dans un premier temps, un pixel sera attribué à une référence seulement si son énergie est supérieure de $a\%$ aux énergies des autres références pour ce pixel là, sinon il restera non attribué :

$$H(\mu, \nu) = \begin{cases} S_{\mu, \nu}^l & \text{si } \frac{S_{\mu, \nu}^l}{Et^l} \geq a \frac{S_{\mu, \nu}^k}{Et^k} \\ \text{non attribué, } 0 & \end{cases} \quad (5.9)$$

où H représente le filtre, S le spectre de la référence, Et l'énergie totale de la référence, k et l l'indice de la référence et a le seuil.

Par la suite, nous nous sommes occupés de la gestion des pixels isolés, restés à 0. Nous avons choisi de regarder le voisinage proche (les 8 plus proches voisins). Si une référence est majoritaire alors nous lui attribuons le pixel isolé. Sinon on regarde les 24 plus proches voisins. Quelques exemples de cas sont représentés sur la figure 5.12. Les carrés bleus représentent les pixels attribués au premier filtre, les carrés verts les pixels attribués au second filtre. Les deux premiers schémas représentent les cas où toute une zone est attribuée à l'un des filtres. Dans ce cas, nous attribuons le pixel isolé à ce filtre là. Le troisième schéma représente l'hypothèse d'un pixel isolé entouré par des pixels appartenant aux deux filtres. Lorsque l'un des deux filtres est majoritaire, le pixel isolé lui est attribué. Dans les cas où aucun filtre n'est majoritaire parmi les 8 voisins (quatrième schéma), nous regardons les valeurs des 24 voisins pour attribuer le pixel isolé. Dans les tests effectués, nous n'avons à ce stade plus de pixels isolés. Dans le cas où il reste des pixels isolés, on peut étendre le système jusqu'à avoir attribué tous les pixels isolés.

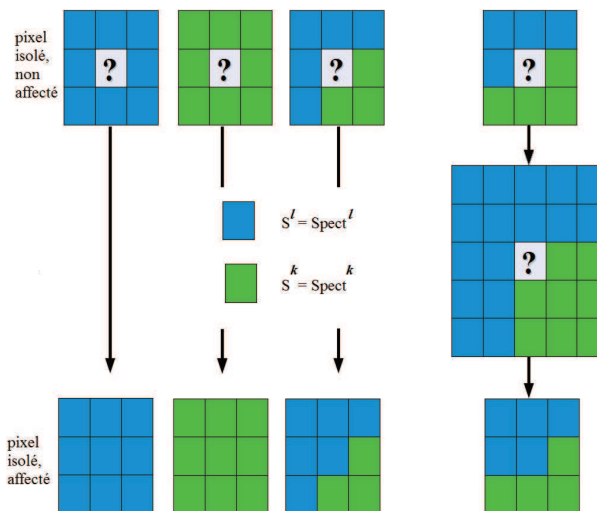


FIGURE 5.12 – Exemples de cas de gestion des pixels isolés

5.3.2 Adaptation du critère PCE

Pour obtenir une comparaison efficace, nous avons aussi dû adapter le PCE. En effet, lorsque les références utilisées dans les deux filtres sont proches, on peut obtenir deux pics de corrélation. Sur la figure 5.13, nous avons représenté le cas de la reconnaissance de la lettre V par un filtre ASPOF créé avec des lettres A. On voit deux pics de corrélation. Cette situation peut aussi se produire dans le cas où les références des deux filtres segmentés sont proches.

Comme seul un demi filtre intervient dans la corrélation, on peut multiplier le pic de corrélation par 2. Il faut dans ce cas multiplier par 2 ce pic lors du calcul de l'énergie du plan (d'où la présence du terme 3 $\left(\sum_{x=x_0-t}^{x=x_0+t} \sum_{y=y_0-t}^{y=y_0+t} |C(x, y)|^2 \right)$ dans la formule suivante). Ce nouveau critère

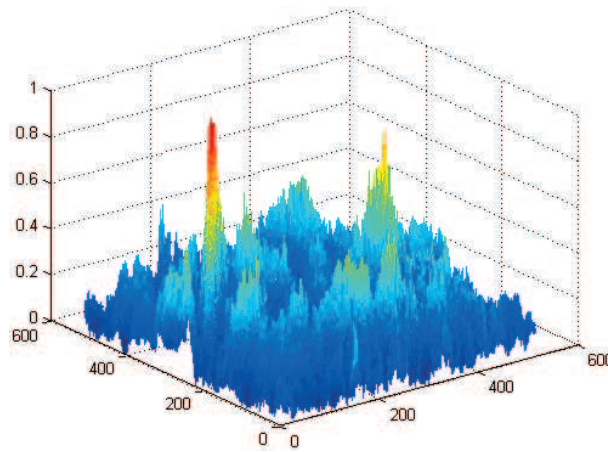


FIGURE 5.13 – Plan résultant de la corrélation entre la lettre V et un filtre ASPOF pour la lettre A

a été appelé PCEA, “*adapted peak to correlation energy*” :

$$PCEA = \frac{\sum_{x=x_0-t}^{x=x_0+t} \sum_{y=y_0-t}^{y=y_0+t} |2C(x, y)|^2}{\sum_{x=N}^{x=1} \sum_{y=M}^{y=1} |C(x, y)|^2 + 3 \left(\sum_{x=x_0-t}^{x=x_0+t} \sum_{y=y_0-t}^{y=y_0+t} |C(x, y)|^2 \right)} \quad (5.10)$$

où N et M désignent la taille du plan de corrélation.

Le pic de corrélation d'intensité maximale est considéré comme le pic de corrélation. Avec ce critère, on favorise ce pic par rapport au plan de corrélation et surtout par rapport au second pic de corrélation lorsqu'il est présent.

5.3.3 Comparaison du filtre ASPOF et des filtres composites et segmentés

Ensuite, nous avons comparé notre filtre ASPOF. La première application est la reconnaissance des lettres A. Notre filtre ASPOF contient l'équivalent de deux filtres de 5 références, soit 10 références. Il est comparé aux filtres POF composites et segmentés contenant 10 références. Nous avons aussi comparé notre filtre aux filtres AMPOF, “*Amplitude Modulated Phase Only Filter*”

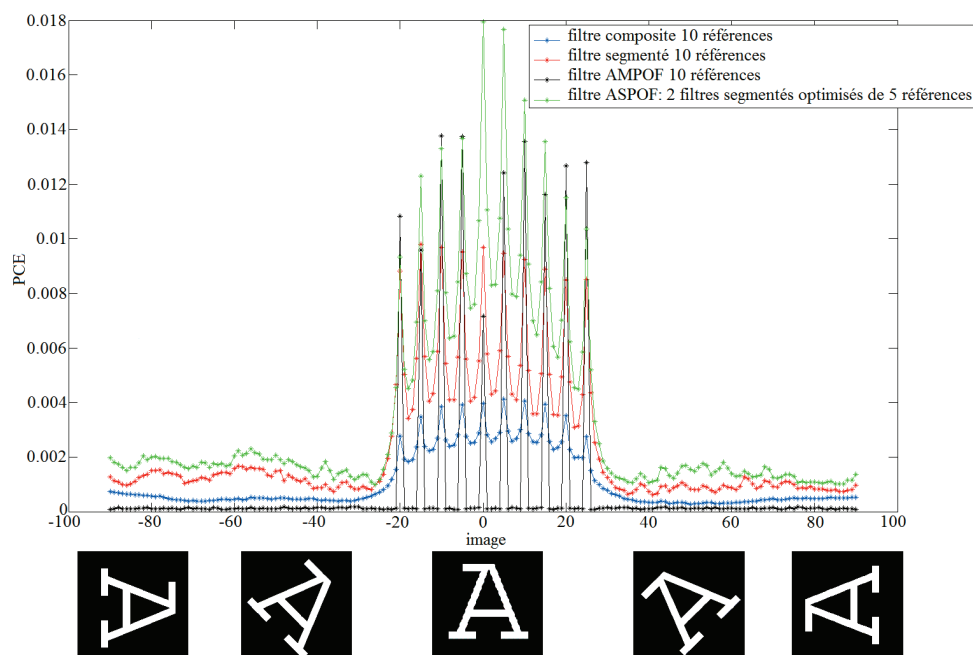


FIGURE 5.14 – Comparaison des filtres sur la lettre A

(présenté dans l'annexe A) et MACE, "*Minimum Average Correlation Energy filter*". Le filtre MACE et les résultats obtenus sont présentés dans l'annexe A. Ce filtre est robuste à la rotation mais ses performances se dégradent lors d'ajout de bruits.

On peut remarquer sur la figure 5.14 que les valeurs du PCEA du ASPOF sont meilleures que les valeurs du PCE pour les filtres composites et segmentés, aussi bien pour les images contenues dans le filtre que pour les images intermédiaires, dont les orientations sont comprises entre -20° et $+25^\circ$. Les valeurs du PCE obtenues avec le filtre AMPOF sont proches voire meilleures que celles obtenues avec le filtre ASPOF mais les images comprises entre -20° et $+25^\circ$ mais non comprises dans le filtre ne sont pas reconnues, ce qui nous pose problème puisque nous cherchons un filtre robuste. Les images à l'extérieur du filtre (avec un angle de rotation compris entre -90° et -20° ou entre $+25^\circ$ et $+90^\circ$) ne sont pas reconnues. L'écart entre reconnaissance et non reconnaissance est important et facilite l'utilisation d'un seuillage pour une application automatique par exemple. Nous avons aussi testé la robustesse au bruit de ces filtres. Le ASPOF permet d'obtenir des valeurs de critère supérieures à celles obtenues avec les autres filtres. Les résultats sont présentés dans l'annexe A. Pour tester la robustesse du filtre, nous avons testé la lettre V qui est proche de la lettre A (*cf.* figure 5.15). Pour cela, nous avons utilisé le même filtre que précédemment, qui reconnaît la lettre A entre -20° et $+25^\circ$ (zone B de la figure 5.15). La lettre A n'est pas reconnue dans les zones A et C où le filtre n'est pas censé la reconnaître. La lettre V n'est pas reconnue par le système.

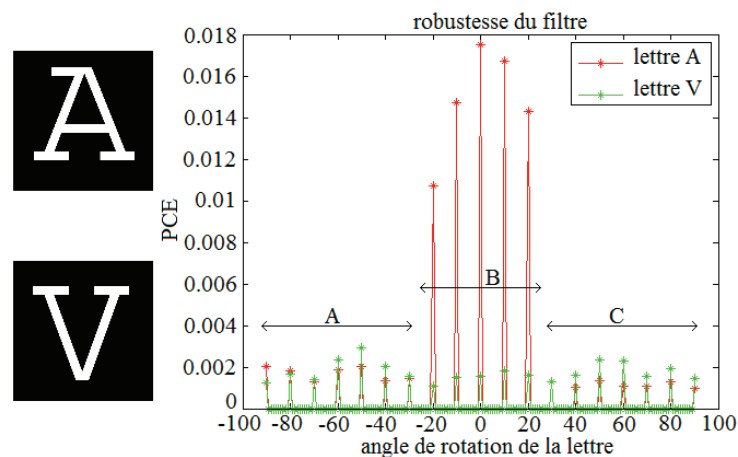


FIGURE 5.15 – Test de la robustesse du filtre

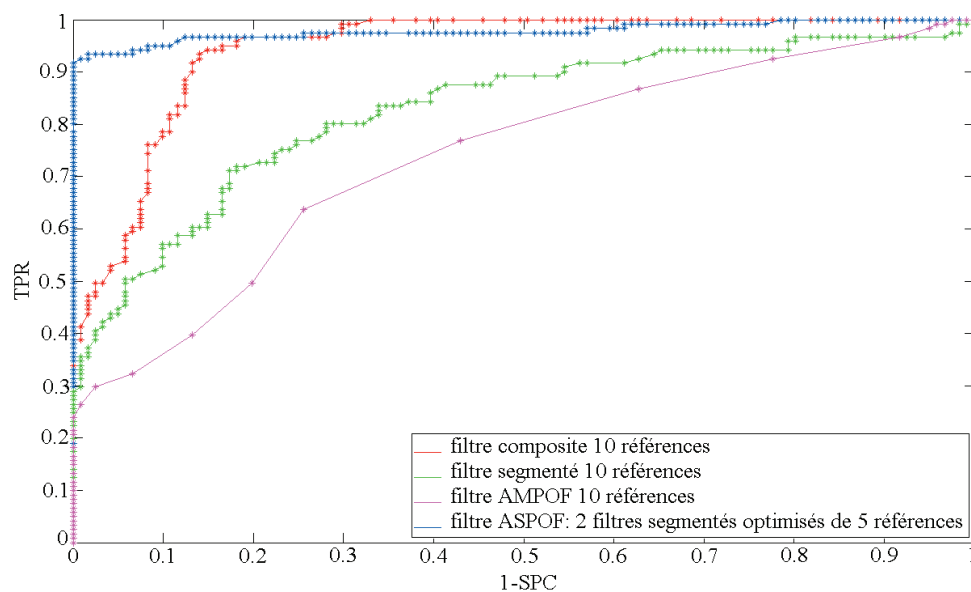


FIGURE 5.16 – Courbes ROC

Nous avons tracé les courbes ROC associée à ces filtres (*cf.* figure 5.16), tracées avec les lettres A et V. Le PCE a servi de seuil pour tracer cette courbe. Nous pouvons constater que le filtre ASPOF est le filtre qui se rapproche le plus du point de coordonnées (1,1), qui représente un système parfait. Le fait que la courbe du filtre AMPOF soit la plus proche de la prédiction aléatoire montre que ce filtre est très discriminant mais peu robuste. Le filtre segmenté présente une courbe inférieure à la courbe du filtre composite à cause des pixels isolés suite à la segmentation qui diminuent les performances du filtre.

Nous avons testé ce filtre sur la reconnaissance des visages. Les visages utilisés sont issus de la base PHPID, "Pointing Head Pose Image Database" [115]. Ceux utilisés sont présentés figure 5.17. La première ligne présente la personne un et la deuxième ligne la personne quatre.



FIGURE 5.17 – Visages de la base PHPID. Sur la première ligne la personne un. Sur la deuxième ligne, la personne quatre. De gauche à droite : orientation à 90° , 45° , 0° , -45° et -90°

Le filtre utilisé ne contient que cinq références : un filtre POF segmenté à trois références et un avec deux références. Il est comparé aux filtres POF composite et segmenté contenant ces cinq références. Comme dans le cas des lettres, les valeurs du PCEA correspondant au filtre ASPOF sont meilleures que les valeurs du PCE des deux autres filtres (*cf.* figure 5.18). Nous avons testé la robustesse de ce filtre en mettant en image cible une autre personne. Cette personne n'est pas reconnue par le système (*cf.* figure 5.19).

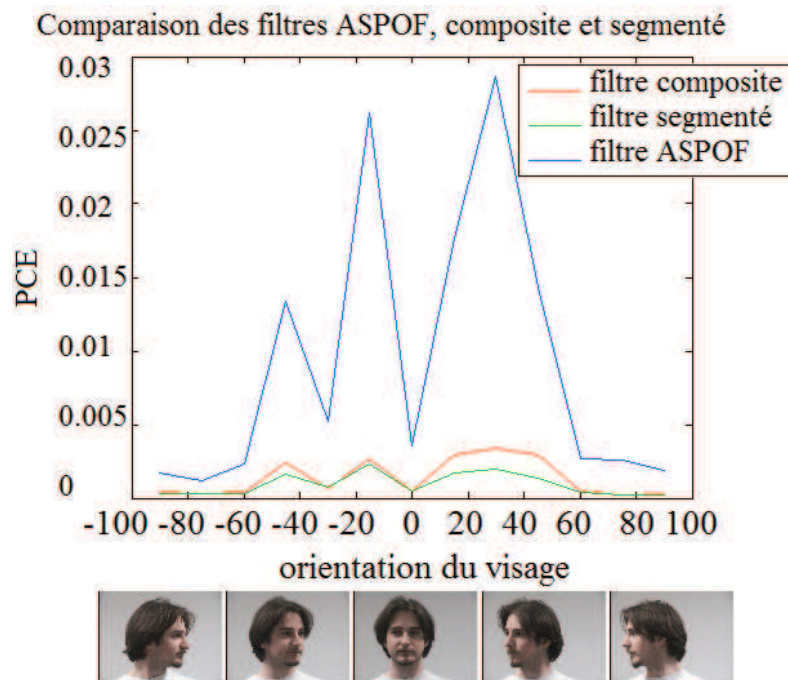


FIGURE 5.18 – Comparaison des filtres sur les visages

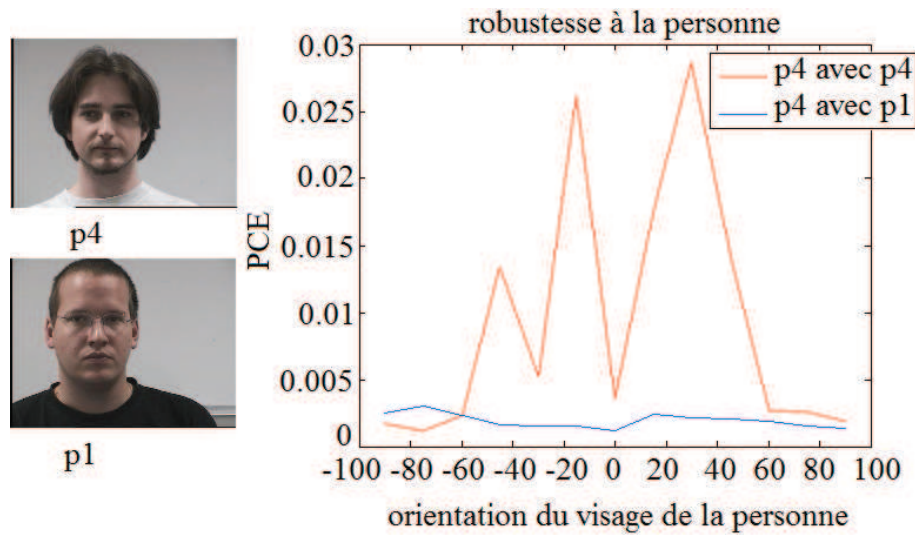


FIGURE 5.19 – Test de la robustesse du filtre

Sur l'application du filtre composite et du filtre ASPOF aux visages, nous avons aussi tracé les courbes ROC associées (*cf.* figure 5.20), en utilisant le PCE comme seuil. Dans les tests effectués, les images correspondent aux deux personnes. Sur cette courbe, le filtre ASPOF donne de meilleurs résultats que le filtre composite.

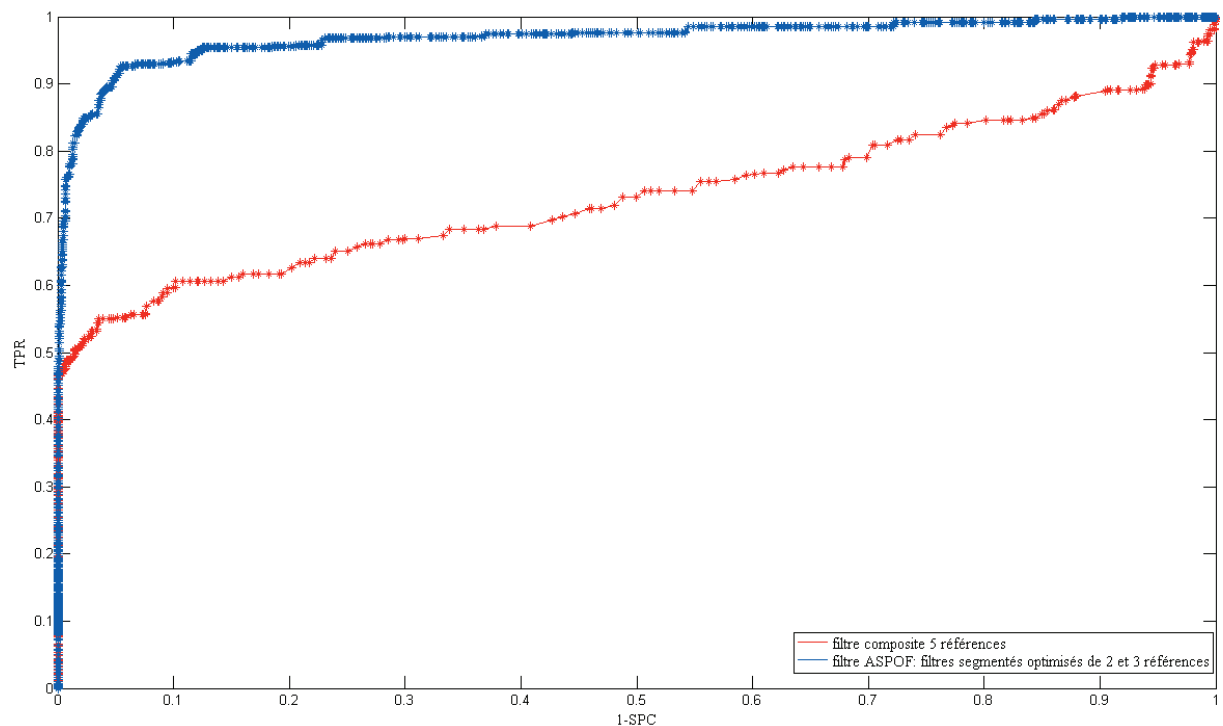


FIGURE 5.20 – Courbes ROC

5.4 Conclusion

Nous avons étudié dans ce chapitre différents filtres qui permettent d'optimiser l'architecture VanderLugt. Pour cela, nous avons créé une base de lettres afin de tester la robustesse à la rotation. Cette base d'images a été complétée avec des images bruitées et des images représentant la même lettre mais plus petite. Ces images ont permis de tester la robustesse aux bruits et à

l'échelle. Enfin, nous avons créé une petite base d'images avec une autre lettre similaire afin de tester la discrimination des filtres. Nous avons aussi étudié l'effet de deux méthodes de multiplexage de références sur les performances des filtres. Les résultats de ces études ont fait l'objet de publications [96,97].

Nous avons ensuite développé notre propre méthode de multiplexage des références. Le critère PCE a été adapté pour quantifier la reconnaissance en prenant en compte les spécificités de notre filtre. Ce nouveau filtre et son critère adapté ont été testés sur deux bases d'images différentes (les lettres binaires et les visages en niveaux de gris), comparativement à plusieurs filtres composites. Notre filtre a montré de bons résultats. Ce filtre a été validé par une publication dans le journal *Applied Optics* [96].

Chapitre 6

La corrélation selon l'architecture JTC

Sommaire

6.1	État de l'art	86
6.1.1	Le JTC sans ordre zéro	86
6.1.2	Le JTC binaire	86
6.1.3	Le JTC non-linéaire	87
6.1.4	Le JTC fringe-adjusted	87
6.1.5	Comparaison des différentes architectures	87
6.2	Utilisation du JTC pour faire du suivi	91
6.3	Conclusion	97

La deuxième architecture optique à laquelle nous nous sommes intéressés est le JTC, aussi appelé CJTC, “*Classical Joint Transform Correlator*”. Elle a été introduite par Weaver et Goodman [101]. Le plan d’entrée f de ce corrélateur est constitué de l’image cible i et de l’image de référence r , décalées horizontalement (ici) du centre du plan d’entrée d’une distance x_0 (*cf.* figure 6.1) :

$$f(x, y) = i(x + x_0, y) + r(x - x_0, y) \tag{6.1}$$

La transformée de Fourier de ce plan d’entrée permet d’obtenir le spectre joint F :

$$F(\mu, \nu) = |I(\mu, \nu)| \exp(\phi_i(\mu, \nu)) \exp(ju x_0) + |R(\mu, \nu)| \exp(\phi_r(\mu, \nu)) \exp(-ju x_0) \tag{6.2}$$

On récupère l’intensité G de ce spectre joint :

$$\begin{aligned} G(\mu, \nu) &= |F(\mu, \nu)|^2 \\ G(\mu, \nu) &= |I(\mu, \nu)|^2 + |R(\mu, \nu)|^2 + 2|I(\mu, \nu)||R(\mu, \nu)| \cos(\phi_i(\mu, \nu) - \phi_r(\mu, \nu) + 2ux_0) \end{aligned} \tag{6.3}$$

Par calcul de la transformée de Fourier inverse, on obtient le plan de corrélation. Ce dernier est composé de trois informations : un pic d’autocorrélation, deux pics d’intercorrélation (en cas de ressemblance des deux images) et du bruit. La méthode et l’allure du plan de corrélation sont représentés sur la figure 6.1.

Cette architecture a l’avantage de ne pas nécessiter la création de filtres de corrélation, contrairement à l’architecture VanderLugt. De plus, du fait de la présence de l’image cible et de la référence dans le même plan, les images doivent être placées de façon à ce que les larges pics ne se chevauchent pas avec le pic central ni ne sortent du plan de corrélation, ce qui implique un plan de corrélation de grandes dimensions. L’espace bande passante n’est ainsi pas totalement utilisé, en entrée et en sortie de l’architecture. Une étude détaillée sur cette limitation a fait l’objet d’une publication [113]. Le pic d’autocorrélation présente un problème dans la prise de décision du fait de son intensité supérieure aux pics d’intercorrélation qui représentent le degré de corrélation entre les deux images. Enfin comme le corrélateur de Vanderlugt, il est sensible au bruit et aux distorsions.

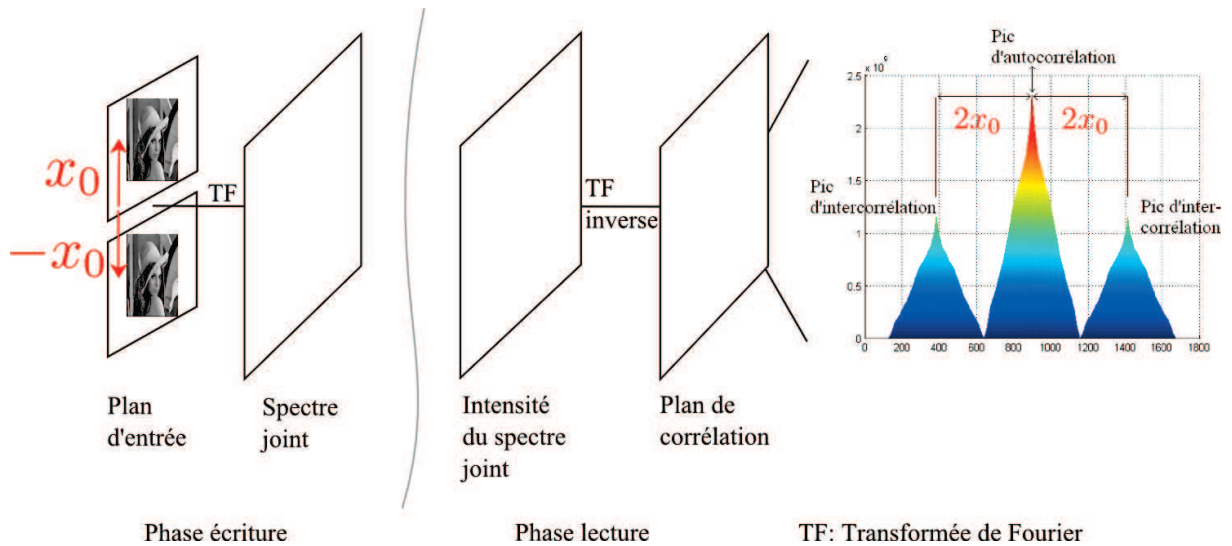


FIGURE 6.1 – Montage du corrélateur JTC

6.1 État de l'art

Pour pallier les inconvénients que nous venons d'évoquer, plusieurs architectures ont été développées [94].

6.1.1 Le JTC sans ordre zéro

Le NZ JTC, “*non-zero joint transform correlator*” est une architecture dans laquelle le pic d'autocorrélation est enlevé mathématiquement. Le pic d'autocorrélation, aussi appelé ordre zéro du fait de sa position centrale dans le plan de corrélation, est dû au terme $|I(\mu, \nu)|^2 + |R(\mu, \nu)|^2$ de l'équation 6.3. I et R représentent respectivement les transformées de Fourier de l'image cible et de l'image de référence. Comme nous connaissons l'image de référence et que nous avons accès à l'image cible, il est possible de calculer leurs densités spectrales de puissance et de les soustraire à l'intensité du spectre joint. Cette architecture a été proposée par [116]. Nous appelons F_{NZ} l'intensité de ce nouveau spectre joint :

$$\begin{aligned} F_{NZ}(\mu, \nu) &= |F(\mu, \nu)|^2 - |I(\mu, \nu)|^2 - |R(\mu, \nu)|^2 \\ F_{NZ}(\mu, \nu) &= 2|I(\mu, \nu)||R(\mu, \nu)|\cos(\phi_i(\mu, \nu) - \phi_r(\mu, \nu) + 2ux_0) \end{aligned} \quad (6.4)$$

Cette architecture résout en partie le problème du pic d'autocorrélation mais les pics d'intercorrélation sont toujours larges.

6.1.2 Le JTC binaire

Le principe du JTC binaire, utilisé par Javidi [117] est de positionner dans le plan de Fourier une caméra CCD qui récupère l'intensité du spectre joint. Cette intensité est binarisée selon un seuil. (Le seuil peut être déterminé par la médiane des pixels de l'histogramme de l'intensité.) L'intensité binarisée est ensuite chargée dans un SLM, qui est binaire. Ensuite la transformée de Fourier inverse est effectuée pour obtenir le plan de corrélation. La méthode est représentée sur la figure 6.2.

Cette architecture permet d'obtenir des pics de corrélation fins. Cependant la recherche du seuil de binarisation peut nécessiter du temps de calcul (selon la taille du plan) et des harmoniques sont introduites par cette binarisation. Sur la figure 6.2, le seuil utilisé a été calculé à partir de l'intensité du spectre joint du NZ JTC. Le pic central est une harmonique due à la binarisation.

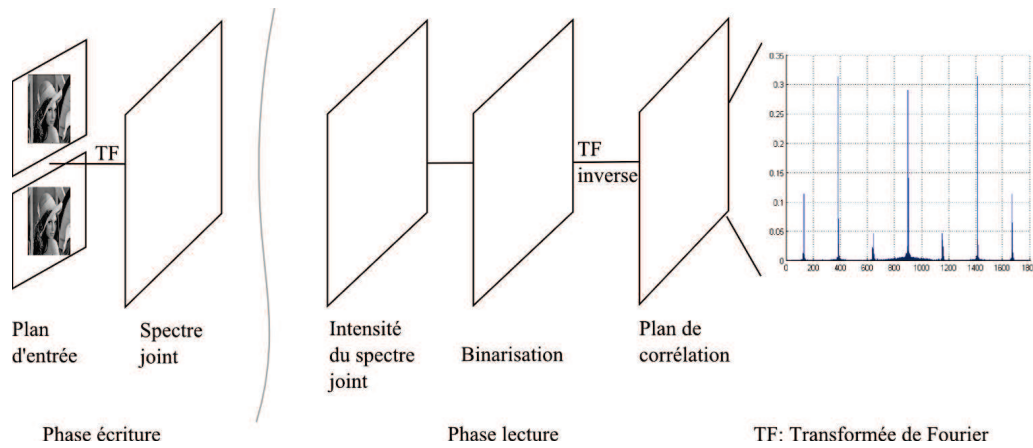


FIGURE 6.2 – Architecture du JTC binaire

6.1.3 Le JTC non-linéaire

Javidi [118] propose un JTC non linéaire, le NL JTC, basé sur un autre seuillage binaire de l'intensité du spectre joint. Pour cela il utilise un composant non linéaire dont la fonction de seuillage, notée F_{NL} peut être modélisée par :

$$F_{NL}(\mu, \nu) = \alpha |F(\mu, \nu)|^k \quad (6.5)$$

où F est le spectre joint de l'équation 6.2, α un facteur de pondération et k est le degré de non linéarité.

Dans le cas où le facteur k vaut 0, on retrouve le JTC binaire et quand il vaut 1 on obtient le JTC classique. Dans le cas particulier où k vaut 0.5, on retrouve un pic semblable à celui obtenu avec le filtre POF du montage Vanderlugt (*cf.* partie 5.1.2). Cette architecture permet selon les besoins d'utiliser davantage les propriétés du JTC classique ou de se rapprocher de celles du JTC binaire.

6.1.4 Le JTC fringe-adjusted

Le FA JTC, "*fringe-adjusted JTC*", proposé par Alam *et al.*, consiste à filtrer l'intensité du spectre joint [119]. Pour cela, un filtre H est multiplié à l'intensité du spectre joint :

$$H(\mu, \nu) = \frac{B(\mu, \nu)}{A(\mu, \nu) + |R(\mu, \nu)|^2} \quad (6.6)$$

où B est une constante ou une fonction qui peut permettre d'obtenir un gain supérieur à l'unité. A est une fonction qui permet de supprimer les problèmes de pôles dans les cas où la densité spectrale de puissance de la référence contient des pixels nuls. Cette fonction peut aussi servir à limiter le bruit. La multiplication de l'intensité du spectre joint par ce filtre permet d'obtenir une architecture très discriminante.

6.1.5 Comparaison des différentes architectures

Nous avons comparé ces différentes architectures sur des images de mines proches (images séparées de quelques images dans une séquence vidéo) et avec Lena. Les architectures binaire (BJTC), non linéaire (NL-JTC) et fringe-adjusted (FA-JTC) ont été testées dans la configuration "*sans ordre zéro*" puisque dans le cas d'une reconnaissance le pic d'autocorrélation peut être assimilé à du bruit. Dans le cas du JTC non linéaire, nous avons choisi la configuration où k vaut 0.5. Les résultats sont présentés dans le tableau 6.1.

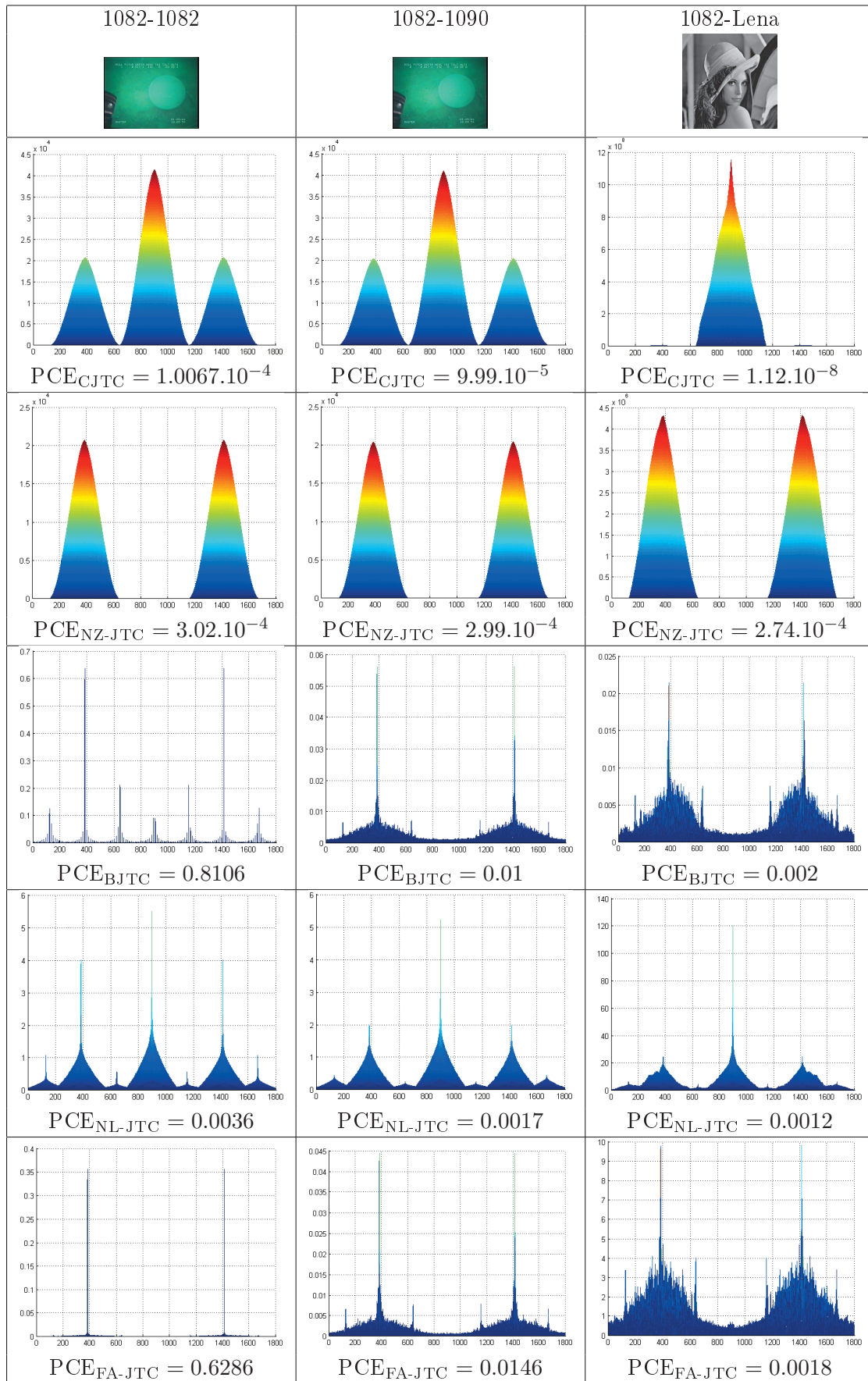


Tableau 6.1 – Comparaison des différentes architectures JTC. 1^{ère} ligne : JTC classique, 2^{ème} ligne : JTC sans l'ordre zéro, 3^{ème} ligne : JTC non-linéaire, 4^{ème} ligne : JTC fringe-adjusted

Les valeurs correspondant aux PCE ont été calculées de la façon suivante :

- on recherche les positions des pics d'intercorrélation
- on calcule l'énergie contenue dans les 8 pixels voisins et dans le pixel du maximum du pic d'intercorrélation, pour les deux pics
- on calcule l'énergie contenue dans le plan de corrélation
- enfin on calcule la valeur du PCE.

Le JTC classique et le JTC sans ordre zéro produisent de larges pics (*cf.* tableau 6.1, lignes 1 et 2). Les valeurs de PCE sont proches et ne permettent pas de fixer un seuil automatiquement, notamment dans le cas du JTC sans ordre zéro où on est en présence de deux pics d'intercorrélation lorsque Lena est comparée à une mine, pics d'intensité très faible par rapport au pic d'autocorrélation dans l'architecture JTC classique. Le JTC binaire (*cf.* tableau 6.1, ligne 3) produit des pics fins mais le plan présente des harmoniques qui font baisser la valeur du PCE. De plus lorsque l'on compare deux images proches (les deux images de mines), la valeur de ce PCE chute fortement (0.81 en cas de comparaison d'images identiques et 0.01 pour deux mines semblables mais non identiques), ce qui fait de cette architecture une architecture peu robuste. La différence des valeurs de PCE obtenues entre la comparaison entre mines et la comparaison d'une mine avec Lena, en fait une architecture discriminante (si deux images sont très proches, la valeur du PCE est très forte, si les images diffèrent un peu, la valeur du PCE est très faible). Le JTC non linéaire est robuste mais peu discriminant : les valeurs du PCE sont proches, que l'on compare deux fois la même mine, deux mines proches ou une mine et Lena. Le JTC fringe-adjusted présente des résultats comparables à ceux obtenus avec le JTC binaire, mais sans la présence d'harmoniques. Cette absence d'harmoniques est un argument important lors du choix entre le JTC binaire et le JTC non linéaire. Cette comparaison nous montre que nous avons à notre disposition trois architectures "*efficaces*" : le JTC non linéaire qui est robuste mais peu discriminant, le JTC binaire qui est discriminant mais peu robuste et qui présente des plans de corrélation avec des harmoniques et le JTC fringe-adjusted qui est discriminant et peu robuste. Parmi les architectures les plus connues présentées ici, il n'y a pas d'architectures à la fois robuste et discriminante.

Le JTC binaire et le JTC fringe-adjusted donnent d'après cette comparaison des résultats proches. Nous les avons donc comparés plus en détails, dans la configuration sans ordre zéro. Pour cela nous avons pris l'image de Lena que nous avons modifiée de façon à se rapprocher des conditions de bruits sous-marins. Nous avons ajouté du flou afin de représenter la diffusion avant qui peut se produire sur les images sous marines. Nous avons modifié la luminosité de l'image en ajoutant une luminosité non uniforme qui est souvent présente dans les images sous marines, surtout lorsqu'une source de lumière artificielle est utilisée. Nous avons combiné le flou et la luminosité. Enfin nous avons retiré le fond de l'image de référence. Cela nous rapproche des conditions dans lesquelles nous voulons identifier les mines. Les résultats de cette étude sont présentés dans le tableau 6.2. Nous pouvons constater que dans les deux cas il y a reconnaissance, ce qui laisse supposer que ces architectures reconnaîtront nos mines. Le JTC fringe-adjusted présente des plans de corrélation dont les pics "*indésirables*" ont une intensité moins importante que les pics présents dans les plans de corrélation obtenus avec le JTC binaire. Nous avons aussi remarqué que le JTC fringe-adjusted présente des plans de corrélation avec un bruit qui semble continu sur tout le plan.

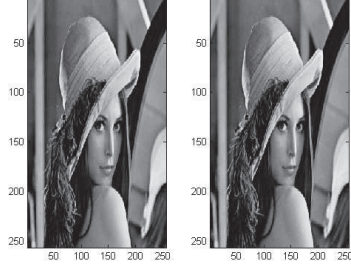
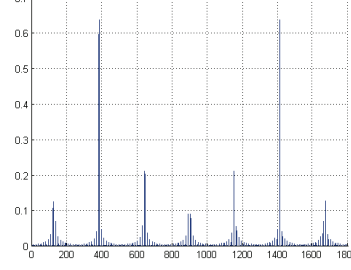
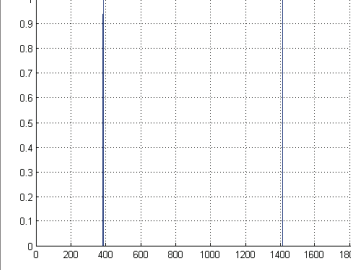
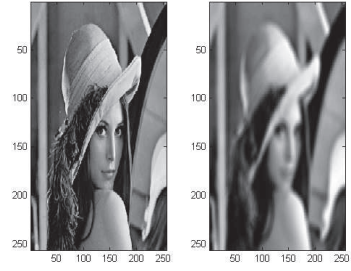
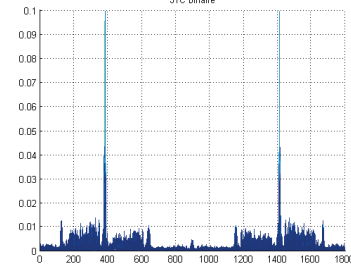
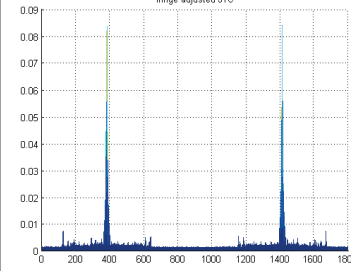
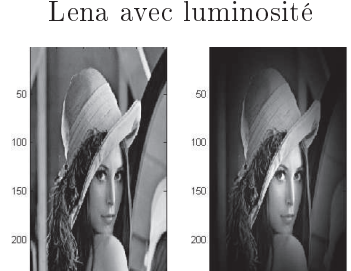
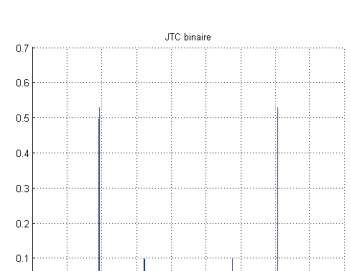
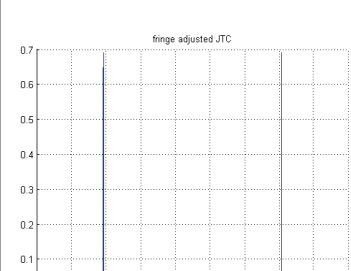
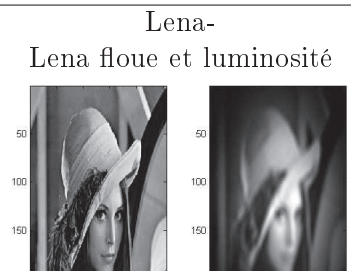
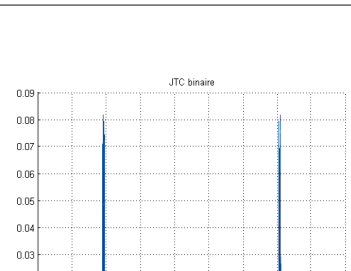
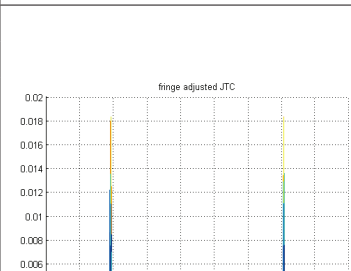
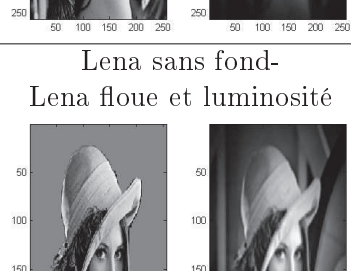
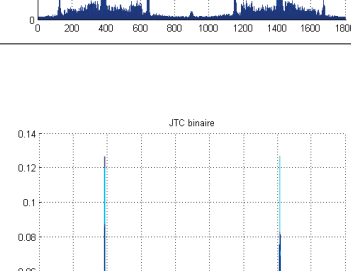
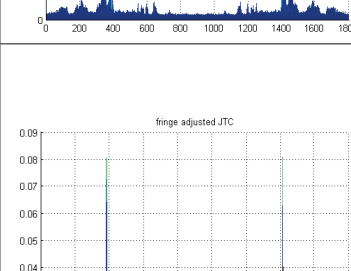
Image de référence et image cible	Plan de corrélation JTC binaire	Plan de corrélation JTC fringe-adjusted
Lena-Lena 		
Lena-Lena floue 		
Lena-Lena avec luminosité 		
Lena-Lena floue et luminosité 		
Lena sans fond- Lena floue et luminosité 		

Tableau 6.2 – Comparaison du JTC binaire et du JTC fringe-adjusted

Malgré la présence de deux pics de corrélation, le JTC présente deux architectures qui sont robustes (le JTC non linéaire) et discriminantes (le JTC fringe-adjusted), comme l'ont montré nos résultats. Pour cette étude, la taille des plans d'entrée, de Fourier et de corrélation a été fixée à partir de l'étude du JTC classique. Pour limiter les risques de chevauchement des pics d'inter-corrélation et d'autocorrélation, et les risques de repliement de spectres, nous avons introduit un espace conséquent entre les images de référence et cible (trois fois la taille de l'image) ainsi qu'un espace autour des images (la largeur et la hauteur de l'image). Les plans d'entrée, de Fourier et de corrélation ont donc de grandes dimensions. Le temps de calcul est par conséquent important, ce qui présente un inconvénient par rapport à l'architecture de Vanderlugt. Si on se concentre sur le JTC non linéaire et le JTC fringe adjusted qui présentent des pics de corrélation fins, la taille de ces plans peut être diminuée. Ainsi le temps de calcul sera réduit. En optimisant la taille des plans, nous pensons qu'il est possible d'obtenir des résultats semblables à ceux obtenus avec l'architecture VanderLugt.

6.2 Utilisation du JTC pour faire du suivi

Dans les résultats que nous venons de commenter, nous n'avons pas à notre disposition une architecture qui est à la fois robuste et discriminante pour identifier les mines sous-marines. En effet, la forme des objets et le milieu changent, ce qui complique l'identification. Nous avons donc voulu tester le JTC pour faire du suivi. Dans ce cas, les deux images successives à comparer sont très proches, nous pensons donc pouvoir obtenir des valeurs élevées en cas de reconnaissance et des valeurs très faibles en cas de non reconnaissance. Pour cela, il faut utiliser une architecture discriminante. Nous avons choisi d'utiliser le JTC fringe-adjusted sans l'ordre zéro. Nous partons du principe qu'une personne est présente dans l'image. Nous utilisons l'image i comme image de référence pour détecter la personne dans l'image $i + 1$. Nous utilisons la base d'images PHPID [115], dans laquelle les personnes tournent la tête par pas de 15° .

Pour le premier test, nous avons utilisé la personne "1". Comme image de fond, nous utilisons le fond de l'image où la personne regarde en face (une orientation de 0°). Nous pouvons remarquer, sur la figure 6.3, qu'il y a bien suivi de la personne entre deux images successives et pas de reconnaissance de l'autre personne (personne numéro quatre). Quand il y a reconnaissance, les valeurs du PCE sont très différentes selon que le visage est orienté à 0° ou non. De plus, l'écart entre reconnaissance et non reconnaissance est très faible. Cela peut s'expliquer par la présence du filtre utilisé dans le FA JTC.

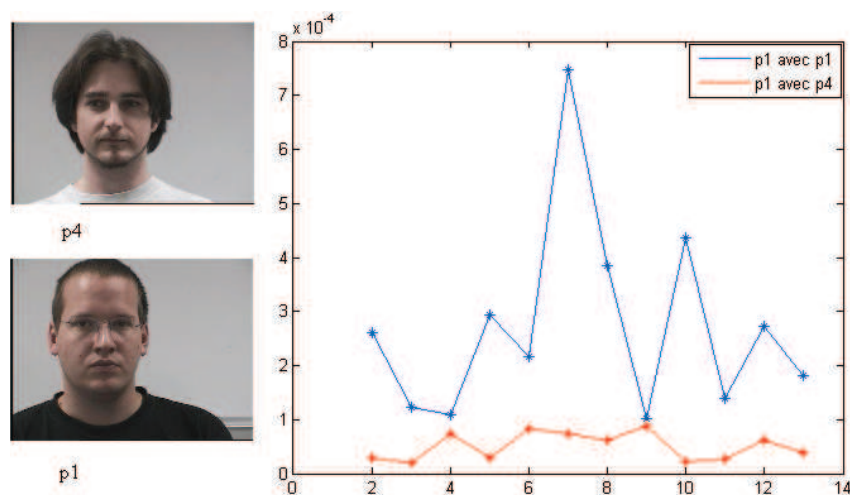


FIGURE 6.3 – Utilisation du JTC fringe-adjusted pour faire du suivi de personne. En bleu : résultats obtenus en comparant la personne 1 avec elle-même. En rouge : résultats obtenus en comparant la personne 1 avec la personne 4.

Nous avons donc refait le test en adaptant le fond à chaque corrélation : nous avons utilisé

chaque image pour lui créer une image fond adapté en supprimant le visage et en le remplaçant par les valeurs de gris situées tout autour du visage. Nous pouvons remarquer (figure 6.4) que les valeurs du PCE en cas de reconnaissance sont meilleures que dans l'étude précédente et que les valeurs en cas de non reconnaissance restent faibles. La valeur de l'écart entre les deux courbes a augmenté. Cependant cette méthode nécessite d'adaptation du fond à chaque image ce qui est lourd en temps de calcul puisqu'il faut calculer le filtre pour chaque image.

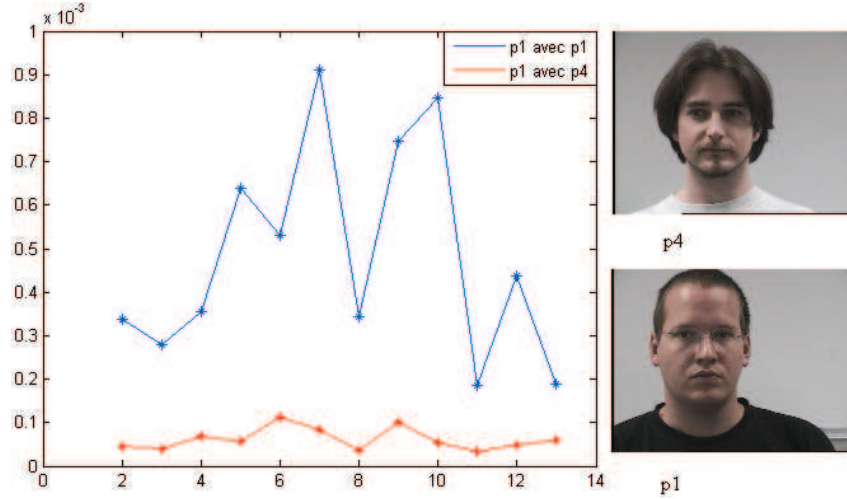


FIGURE 6.4 – Utilisation du JTC fringe-adjusted pour faire du suivi de personne avec mise à jour du fond dans le filtre. En bleu : résultats obtenus en comparant la personne 1 avec elle-même. En rouge : résultats obtenus en comparant la personne 1 avec la personne 4.

Pour obtenir des valeurs proches de 1 en cas de reconnaissance et proches de 0 en cas de non reconnaissance et ainsi faciliter la prise de décision automatique, nous avons décidé de n'utiliser qu'un fond et de changer de critère. Au lieu de calculer l'énergie contenue dans le pic de corrélation par rapport à l'énergie du plan de corrélation, nous nous sommes orientés vers l'écart entre l'énergie contenue dans le pic de corrélation et l'énergie contenue dans les cinq maxima du plan de corrélation (pic de corrélation exclu). Le choix de ces cinq pics est le résultat d'un compromis entre temps de calcul et l'exactitude du critère PCE (en terme de comparaison du pic et du plan de corrélation). L'expression de ce critère que nous avons appelé enhanced PCE est :

$$P = \frac{\left(\sum_{i=i_0-t}^{i=i_0+t} \sum_{j=j_0-t}^{j=j_0+t} |C(i,j)|^2 \right) - \left(\frac{\sum_{k=1}^5 \left(\sum_{i=i_k-t}^{i=i_k+t} \sum_{j=j_k-t}^{j=j_k+t} |C(i,j)|^2 \right)}{5} \right)}{\sum_{i=i_0-t}^{i=i_0+t} \sum_{j=j_0-t}^{j=j_0+t} |C(i,j)|^2} \quad (6.7)$$

Les valeurs obtenues sont bien dans l'intervalle $[0, 1]$. Les écarts entre les valeurs correspondant à la reconnaissance sont resserrés et l'écart entre la reconnaissance et la non reconnaissance s'est agrandi (*cf.* figure 6.5).

Nous avons ensuite décidé d'ajouter une non linéarité au filtre. Après application du filtre fringe-adjusted, l'intensité du spectre est élevée à la puissance k . Cela diminue encore l'écart entre les valeurs de reconnaissance, comme le montre la figure 6.6.

Enfin, nous avons décidé de modifier notre critère. Au lieu de comparer le pic d'intercorrélation aux 5 meilleurs pics, nous comparons le pic d'intercorrélation aux pics dont l'intensité est supérieure à un seuil. Ce seuil correspond aux trois quarts de l'intensité du pic de corrélation auquel on a retiré 3dB, soit 50% de son énergie. Le critère devient :

$$P = \frac{\left(\sum_{|C(i,j)|^2 > \frac{|C(i_0,j_0)|^2}{\sqrt{2}}} |C(i,j)|^2 \right) - \left(\sum_{|C(i,j)|^2 > 0.75 \frac{|C(i_0,j_0)|^2}{\sqrt{2}}} \left(\sum_{i=i_k-t}^{i=i_k+t} \sum_{j=j_k-t}^{j=j_k+t} |C(i,j)|^2 \right) \right)}{\sum_{i=i_0-t}^{i=i_0+t} \sum_{j=j_0-t}^{j=j_0+t} |C(i,j)|^2} \quad (6.8)$$

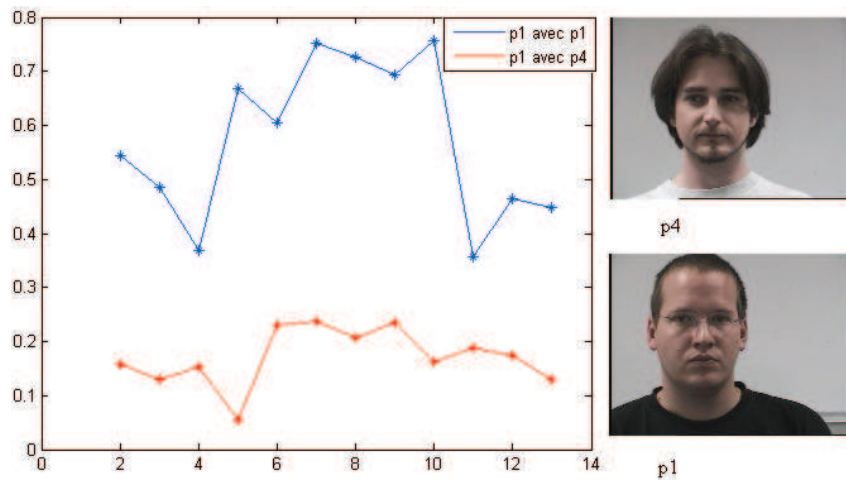


FIGURE 6.5 – Utilisation du JTC fringe-adjusted pour faire du suivi de personne avec le critère enhanced PCE. En bleu : résultats obtenus en comparant la personne 1 avec elle-même. En rouge : résultats obtenus en comparant la personne 1 avec la personne 4.

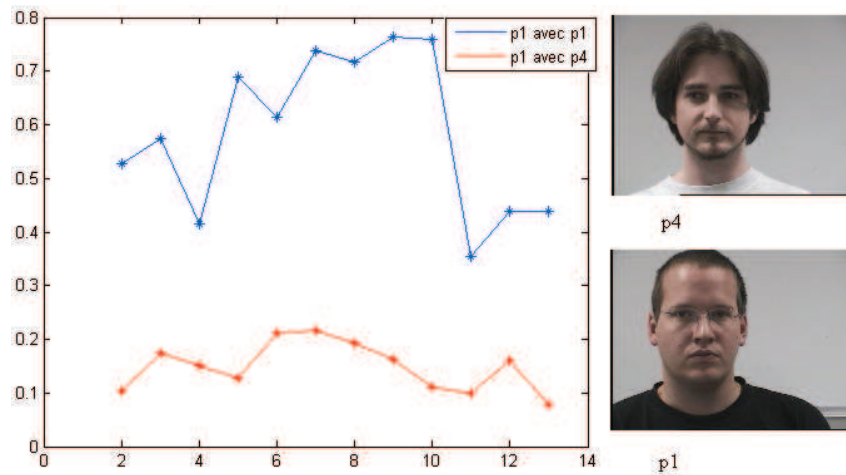


FIGURE 6.6 – Utilisation du JTC fringe-adjusted non linéaire pour faire du suivi de personne avec le critère enhanced PCE. En bleu : résultats obtenus en comparant la personne 1 avec elle-même. En rouge : résultats obtenus en comparant la personne 1 avec la personne 4.

Comme on peut le voir sur la figure 6.7, les valeurs sont proches de 1 en cas de reconnaissance et de 0 en cas de non reconnaissance. Une prise de décision automatique peut être facilement envisagée.

Résumé de nos contributions à l'architecture JTC

Le filtre

Nous venons de montrer que nous pouvons augmenter la fiabilité de la décision en utilisant un nouveau critère de décision. Cependant pour augmenter la robustesse et la discrimination de l'architecture JTC, il faut optimiser l'architecture. Le simple fait d'interpréter le plan de corrélation ne suffit pas. Dans cet objectif nous avons associé notre critère à une optimisation de l'architecture. Nous résumons ici le principe de notre filtre. Notre idée est d'appliquer le filtre fringe-adjusted sur l'intensité du spectre joint sans l'ordre zéro puis d'appliquer le JTC non linéaire au résultat du filtrage. A partir des images de référence r et cible i , le plan d'entrée f est créé. Par transformée de Fourier, on obtient le spectre joint F dont on récupère l'amplitude G (cf. équation 6.9), à laquelle on soustrait les densités spectrales de puissance des images de référence et cible. L'intensité obtenue, F_{NZ} est multipliée par le filtre fringe-adjusted H pour obtenir le

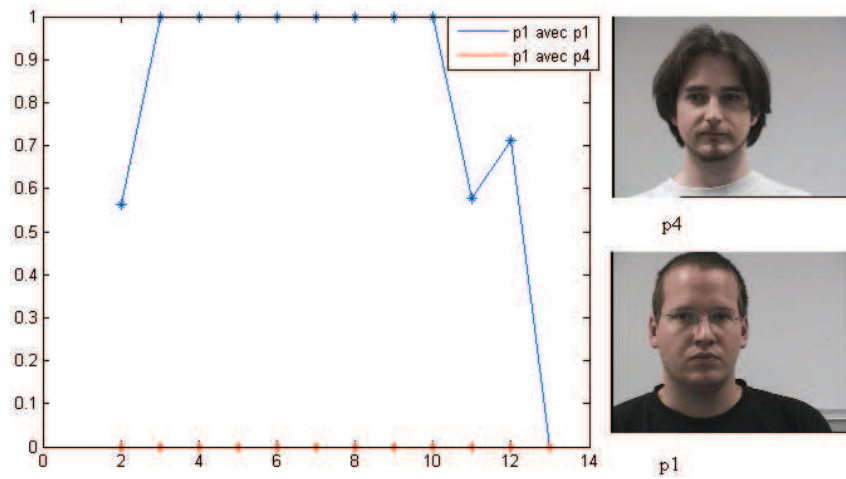


FIGURE 6.7 – Utilisation du JTC fringe-adjusted non linéaire pour faire du suivi de personne avec le nouveau critère. En bleu : résultats obtenus en comparant la personne 1 avec elle-même. En rouge : résultats obtenus en comparant la personne 1 avec la personne 4.

nouveau spectre $G1$. L'application de la non linéarité permet d'obtenir le spectre G_{NL-FAF} . Les opérations mathématiques correspondant à ce nouveau filtre sont résumées sur la figure 6.8 et dans l'équation 6.9 :

$$\begin{aligned}
 f(x, y) &= i(x + x_0, y) + r(x - x_0, y) \\
 F(\mu, \nu) &= |I(\mu, \nu)| \exp(\phi_i(\mu, \nu)) \exp(ju x_0) + |R(\mu, \nu)| \exp(\phi_r(\mu, \nu)) \exp(-ju x_0) \\
 G(\mu, \nu) &= |F(\mu, \nu)|^2 \\
 G(\mu, \nu) &= |I(\mu, \nu)|^2 + |R(\mu, \nu)|^2 + 2|I(\mu, \nu)||R(\mu, \nu)| \cos(\phi_i(\mu, \nu) - \phi_r(\mu, \nu) + 2u x_0) \\
 F_{NZ}(\mu, \nu) &= 2|I(\mu, \nu)||R(\mu, \nu)| \cos(\phi_i(\mu, \nu) - \phi_r(\mu, \nu) + 2u x_0) \\
 G1(\mu, \nu) &= F_{NZ}(\mu, \nu) \cdot * H(\mu, \nu) \\
 G1(\mu, \nu) &= F_{NZ}(\mu, \nu) \cdot * \frac{B(\mu, \nu)}{A(\mu, \nu) + |R(\mu, \nu)|^2} \\
 G_{NL-FAF}(\mu, \nu) &= (G1(\mu, \nu))^k
 \end{aligned} \tag{6.9}$$

Les critères

Nous avons défini un premier critère de décision qui compare le pic de corrélation avec les principaux pics du plan de corrélation. Cela permet de mesurer l'écart entre le pic et le bruit du plan de corrélation.

Tout d'abord on supprime l'ordre zéro du plan de corrélation (quand il est présent) (cf. figure 6.9 a et b). Grâce à la symétrie du plan de corrélation, on peut ne travailler que sur un demi plan (cf. figure 6.9 c). On détecte et on enregistre l'énergie contenue dans le pic de corrélation avant de mettre cette zone à 0. Enfin on calcule la moyenne des énergies des 5 maxima du demi-plan de corrélation restant (cf. figure 6.9 d). Mathématiquement ce critère s'écrit :

$$P = \frac{\left(\sum_{i=i_0-t}^{i=i_0+t} \sum_{j=j_0-t}^{j=j_0+t} |C(i, j)|^2 \right) - \left(\frac{\sum_{k=1}^{k=5} \left(\sum_{i=i_k-t}^{i=i_k+t} \sum_{j=j_k-t}^{j=j_k+t} |C(i, j)|^2 \right)}{5} \right)}{\sum_{i=i_0-t}^{i=i_0+t} \sum_{j=j_0-t}^{j=j_0+t} |C(i, j)|^2} \tag{6.10}$$

Ce critère ne donne pas les valeurs très proches de 0 et de 1 attendues en cas, respectivement, de non reconnaissance et de reconnaissance. Nous l'avons donc modifié. Comme précédemment, on supprime l'ordre zéro s'il est présent et on travaille dans le demi plan de corrélation. On

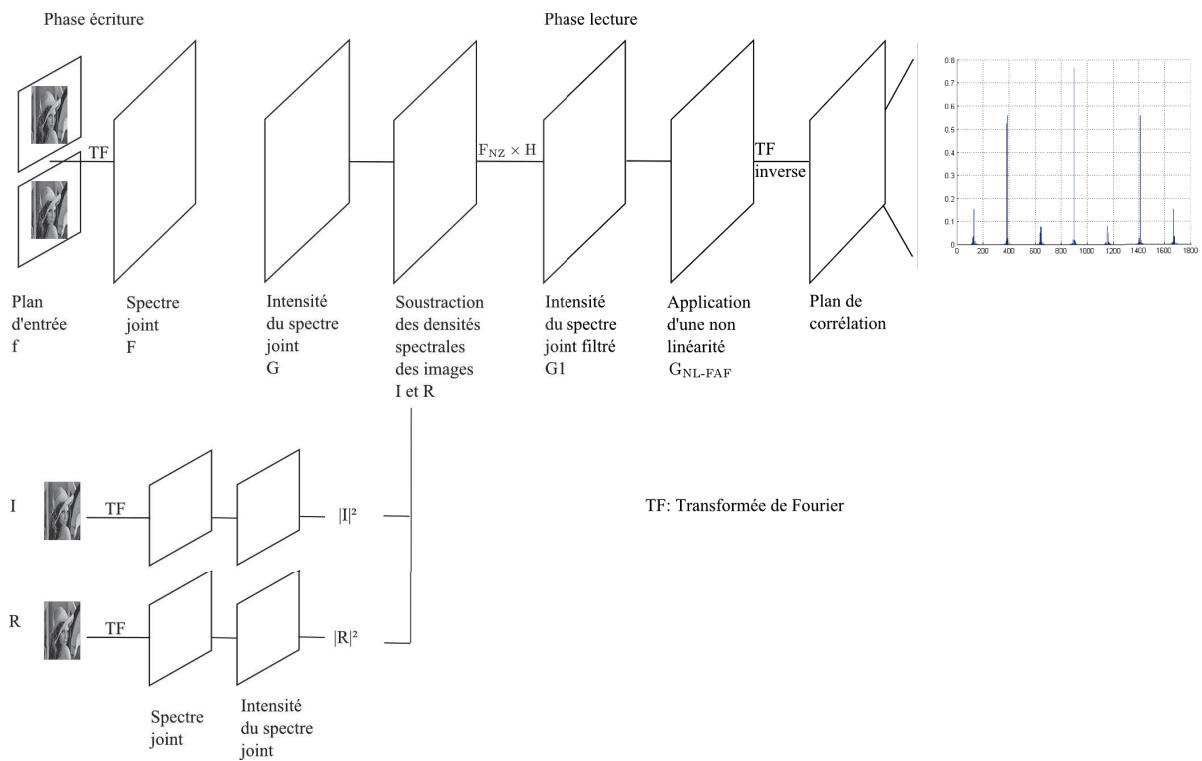


FIGURE 6.8 – Schéma de l'architecture proposée

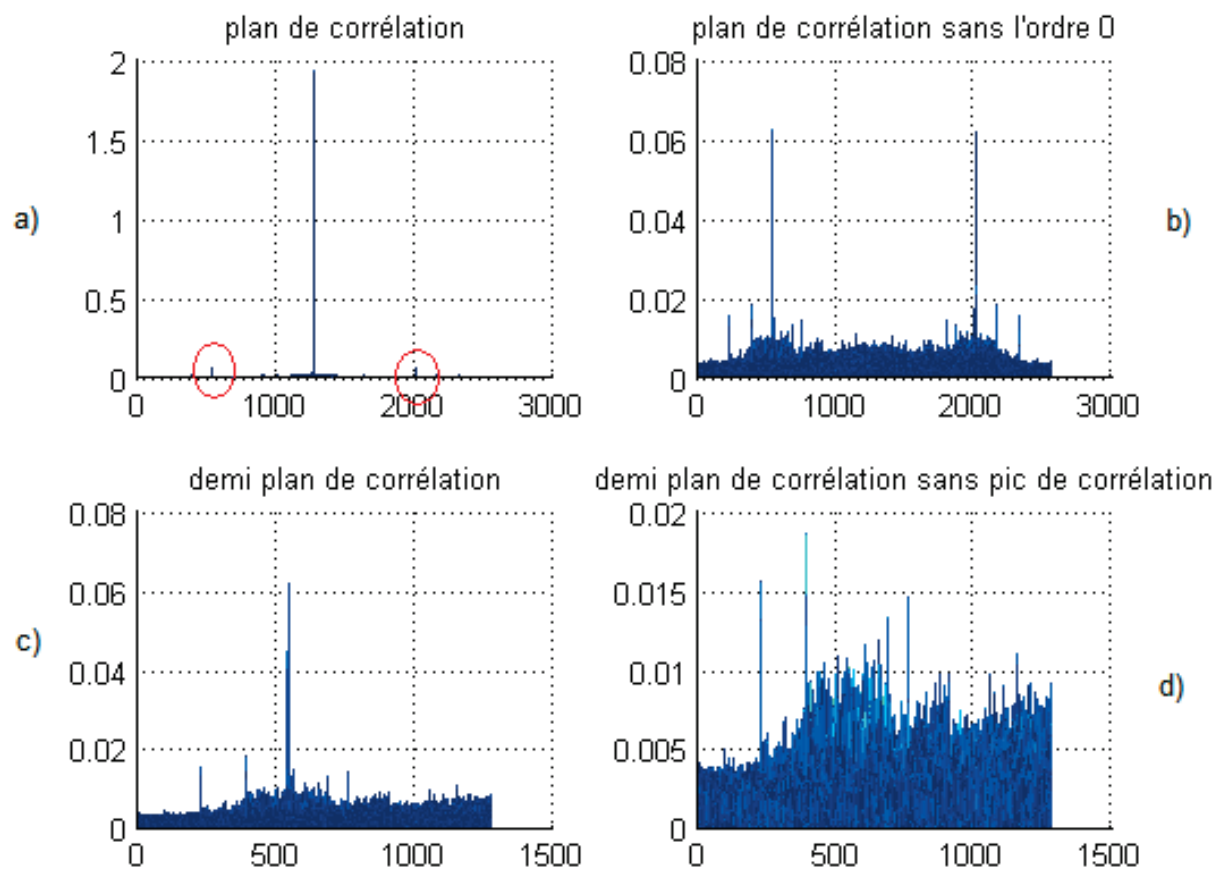


FIGURE 6.9 – Plans de corrélation correspondant aux différentes étapes du calcul de l'écart. a : plan de corrélation. b : plan de corrélation sans l'ordre zéro. c : demi plan de corrélation. d : demi plan de corrélation sans le pic d'intercorrélation

cherche la valeur M du pic de corrélation (ou du maximum du plan de corrélation). On fixe le seuil à M_{-3dB} soit la valeur du pic de corrélation moins 3dB. La valeur attribuée au pic de corrélation est l'énergie des pixels dont la valeur est supérieure au seuil. Cette énergie est appelée $S1$. On compare cette énergie à $S2$ qui correspond à l'énergie de tous les pixels ne faisant pas partie du pic de corrélation dont la valeur est supérieure au seuil multiplié par 0.75. L'écart est ensuite calculé :

$$C = \frac{S1 - S2}{S1} \quad (6.11)$$

Cet écart peut être négatif en cas de non reconnaissance et d'un plan très bruité. Dans ce cas, on fixe la valeur de l'écart à 0. Sur la figure 6.10, il y a reconnaissance. Le pic de corrélation a une valeur très supérieure au reste du plan. On calcule l'énergie de ce pic. Comme aucun pixel n'est supérieur au seuil, le critère vaut 1. Sur la figure 6.11, il n'y a pas reconnaissance. Le pic dont la valeur est maximale sert de référence. On calcule son énergie. On calcule l'énergie des pics dont la valeur dépasse le seuil. La différence des deux sommes est proche de 0 voire négative. Le critère vaut donc 0.

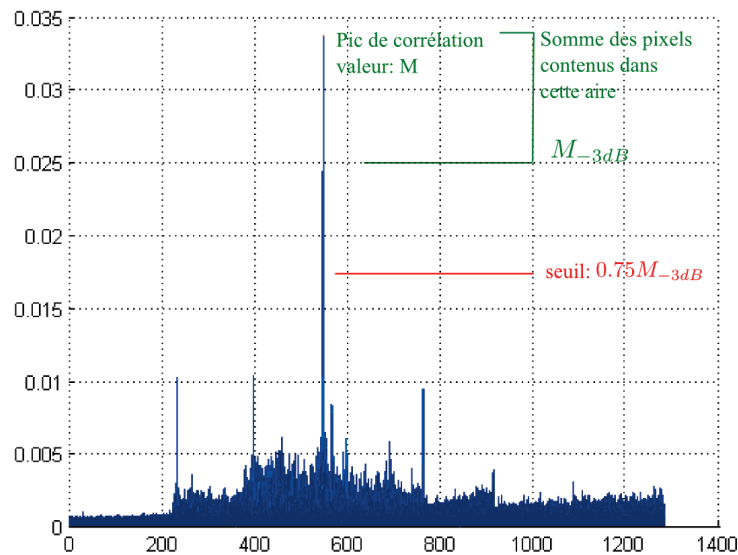


FIGURE 6.10 – Exemple de calcul du critère en cas de reconnaissance

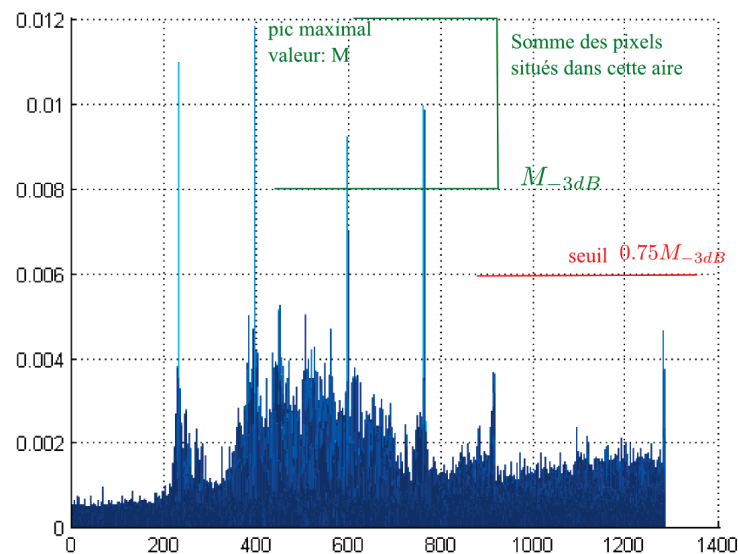


FIGURE 6.11 – Exemple de calcul du critère en cas de non reconnaissance

6.3 Conclusion

Dans ce chapitre, nous avons étudié quelques architectures du JTC. Nous avons étudié la robustesse et la discrimination de ces architectures, avec des images de mines pour la robustesse et une image de mine comparée à Lena pour la discrimination. Les résultats obtenus nous ont amené à comparer plus en détails le JTC binaire et le JTC fringe-adjusted. Pour cela nous avons utilisé l'image de Lena que nous avons modifiée de façon à se rapprocher des conditions réelles de reconnaissance des mines, tout en maîtrisant les modifications apportées. Avec le JTC non linéaire, ces architectures discriminantes peuvent être utilisées pour faire de la reconnaissance robuste. Nous avons donc voulu tester la possibilité de faire du suivi de personne avec le JTC.

La comparaison du JTC binaire au JTC fringe-adjusted nous a permis de sélectionner le JTC fringe adjusted pour cette application. Nous l'avons adapté en ajoutant une non linéarité et en définissant un nouveau critère. Ces avancées nous ont permis d'obtenir une architecture et un outil de décision robustes et discriminants (cf. figure 6.7). Cette architecture a donné lieu à une publication [98]. Ci dessous un tableau montre l'évolution de l'écart entre la valeur de reconnaissance minimale et la valeur maximale de non reconnaissance avec les différents critères que nous avons testé.

	PCE figure 6.3	PCE avec mise à jour du fond figure 6.4	1ère version du critère figure 6.5	1ère version du critère et introduction de la non linéarité figure 6.6	2ème version du critère figure 6.7
écart	0.00001332	0.0000731	0.1202	0.1379	0.562

Troisième partie

La détection et la reconnaissance des mines sous marines

Introduction

Nous avons manifesté précédemment notre intérêt pour la méthode de la corrélation. En effet, les différentes études effectuées dans notre laboratoire, ont montré qu'il s'agit d'une méthode fiable, robuste et discriminante pour la reconnaissance faciale par exemple. Dans cette partie, nous allons l'appliquer à l'étude des mines après avoir apporté une série d'adaptations et d'optimisations.

L'état de l'art montre que les images issues de vidéos sous marines doivent être prétraitées avant d'utiliser un algorithme de détection ou de reconnaissance. Cela a été confirmé par nos premiers tests : la corrélation des images brutes avec des images issues de la vidéo utilisées comme images de référence ne permet pas de reconnaître une mine ! En effet, la corrélation compare une image cible à une image de référence. L'eau ainsi que le bruit électronique changent l'image cible par rapport à l'image de référence. De plus les informations de navigation présentes sur les images Topvision introduisent des contours qui perturbent la reconnaissance. Par conséquent il est nécessaire de prétraiter les images.

Malgré les différents prétraitements réalisés, la corrélation classique appliquée sur ces images ainsi prétraitées pose deux problèmes majeurs : le temps de calcul et les images de référence. Tout d'abord le temps de calcul est anormalement très élevé à cause de la taille des images. Ainsi nous y introduisons une étape de détection pour ne pas réaliser inutilement la corrélation sur des images vides. Cette étape sera suivie d'un post-traitement afin de limiter les fausses alarmes et les non détections au sein d'une zone de présence de la mine. En cas de détection d'un objet, nous l'identifierons. Dans ce cas, nous avons identifié un autre problème à résoudre : les images de référence issues d'images réelles ne permettent pas d'obtenir tous les points de vue voulus. En effet, il n'est pas possible d'aller enregistrer des images de mines à toutes les profondeurs ni de positionner des mines dans toutes les positions (notamment les mines qui ne présentent pas d'axes de symétrie). Nous avons donc orienté notre recherche vers l'utilisation d'images de référence de synthèse pour créer nos filtres de corrélation.

Chapitre 7

Application de la corrélation à l'étude des mines

Sommaire

7.1	Application de la corrélation aux mines	103
7.1.1	Détection et identification des mines sous marines avec l'architecture VanderLugt	103
7.1.2	Détection et identification des mines sous marines avec l'architecture JTC105	
7.2	Problématique	107

La corrélation étant, d'après nos tests, une méthode efficace de reconnaissance, nous l'avons utilisée pour la détection et la reconnaissance des mines. Pour cela, nous avons utilisé les deux architectures optiques : l'architecture de Vanderlugt et l'architecture JTC. Nous avons ainsi pu comparer les résultats obtenus, le temps de calcul et faire des choix pour l'algorithme que nous souhaitons mettre en place pour résoudre notre problème.

Pour créer une vérité terrain qui permette de comparer automatiquement les résultats obtenus à la réalité contenue dans les images, nous avons besoin de définir des zones de présence des mines. Dans un premier temps, nous fixons les bornes de ces zones visuellement : la mine est-elle présente dans l'image ?, est-elle entièrement présente ? est-elle à une distance correcte de l'observateur ? (si une mine est trop loin, elle ne sera pas détectée par les algorithmes). Par la suite les bornes des zones de présence seront fixées selon la distance entre la mine et la caméra afin que les résultats obtenus ne dépendent pas de l'opérateur.

7.1 Application de la corrélation à l'étude des mines sous marines

7.1.1 Détection et identification des mines sous marines avec l'architecture VanderLugt

Pour détecter et identifier les mines dans les vidéos sous marines à notre disposition (vidéos Topvision et vidéos VAMA), nous avons procédé par étapes. Tout d'abord nous nous sommes intéressés à la mine Manta (*cf.* figure 7.1). Nous avons créé cinq filtres composites représentant chacun la mine avec une orientation et une échelle spécifiques. Les images de référence utilisées pour fabriquer les filtres proviennent de la vidéo Topvision003 et ont été choisies pour reconnaître la mine dans une zone particulière, comprise entre les images 650 et 875.

Dans le premier test, nous avons corrélé les filtres composites ainsi créés avec les images issues de la vidéo, sans prétraitements. Les résultats ne sont pas satisfaisants (reconnaissance de la mine tout le temps, qu'il y ait présence ou absence de la mine dans l'image). Pour pallier ce problème, nous avons prétraité les images (*cf.* figure 7.2a) avec la chaîne présentée dans la partie 2.3 (*cf.* figure 7.2b). De plus, pour optimiser le résultat, nous avons transformé les images cibles prétraitées en images de phase (*cf.* partie 2.2.5 et figure 7.2c).

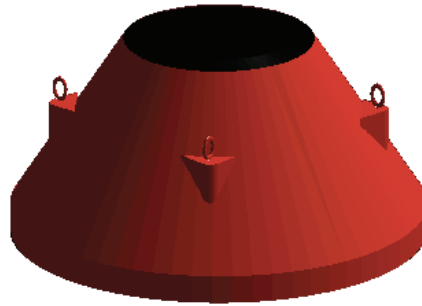


FIGURE 7.1 – Mine Manta

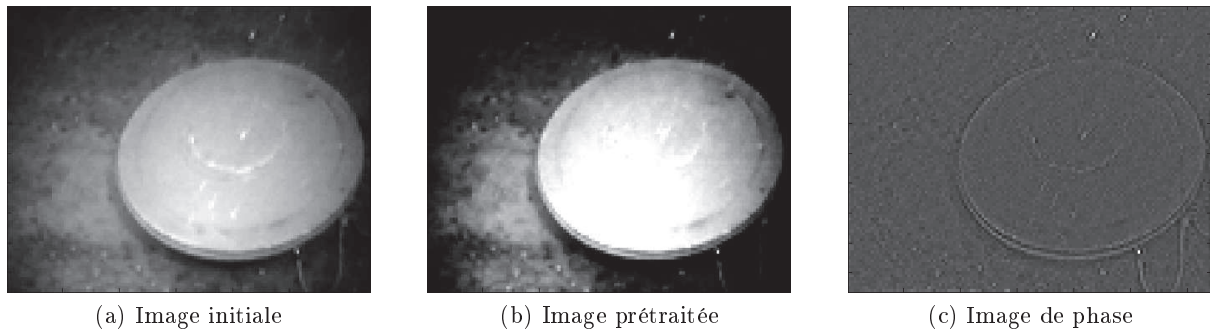


FIGURE 7.2 – Image de phase calculée à partir de l'image prétraitée issue d'une vidéo Topvision

Sur la figure 7.3, le résultat de la corrélation entre les images de la vidéo Topvision003 et les filtres sont présentés. La courbe bleue représente la corrélation de la vidéo Manta avec les filtres Manta. Nous pouvons constater qu'il y a bien détection et identification de la mine dans la zone définie. A la fin de la vidéo, il y a une faible reconnaissance de la mine. En effet, dans cette partie de la vidéo, la mine a une orientation proche de celle de la zone définie mais une taille un peu plus grande. Cela explique cette reconnaissance. La courbe en rouge représente le résultat de la corrélation d'une vidéo représentant une mine cylindrique avec les filtres Manta. Il n'y a pas de reconnaissance de cette mine, hormis quelques fausses alarmes.

Dans le tableau 7.1, nous présentons les résultats chiffrés obtenus. Les filtres Manta ont été créés pour obtenir une reconnaissance entre les images 650-875. La valeur de PCE qui sert de seuil de reconnaissance a été fixée empiriquement à 0.01 après étude des résultats. Comme indiqué dans la ligne 1 du tableau 7.1, le taux de reconnaissance dans la zone couverte par les filtres est donc très élevé. Le taux de reconnaissance de la mine Manta à la fin de la vidéo, lorsque la mine a une forme semblable aux mines présentes dans les filtres, est supérieur à 70%, ce qui montre que nos filtres sont robustes à l'échelle et à la forme (*cf.* ligne 2 du tableau 7.1). Il n'y a pas de taux de fausses alarmes à calculer puisque les filtres ne sont pas faits pour détecter une mine dans ces images. De plus, notre algorithme ne produit pas de fausses alarmes lors de l'étude de cette zone de cette vidéo. Les résultats obtenus avec la même approche de corrélation composite et utilisant des images cibles constituées de mines cylindriques sont présentés dans la ligne 4 du tableau 7.1. L'utilisation d'une mine cylindrique en entrée permet d'étudier la robustesse de la méthode en calculant le taux de fausses alarmes obtenus dans chacune des zones. Il s'élève à 5.38% sur l'ensemble de la vidéo de la mine cylindrique. Comme les deux mines ne sont pas semblables, nous n'avons pas de taux de reconnaissance à calculer. De plus, même dans le cas où une mine a été détectée et l'image classée comme contenant un objet, nous pouvons constater que les valeurs du PCE sont très faibles, à peine supérieures à 0.01, seuil de corrélation (*cf.* figure 7.3).

	Taux de reconnaissance	Taux de fausse alarme
zone 650-875	98.67%	1.33%
zone 1550-1830	72.95%	pas de valeur à calculer
mine cylindrique	pas de valeur à calculer	5.38%

Tableau 7.1 – Résultats obtenus sur les vidéos de mine Manta et cylindrique

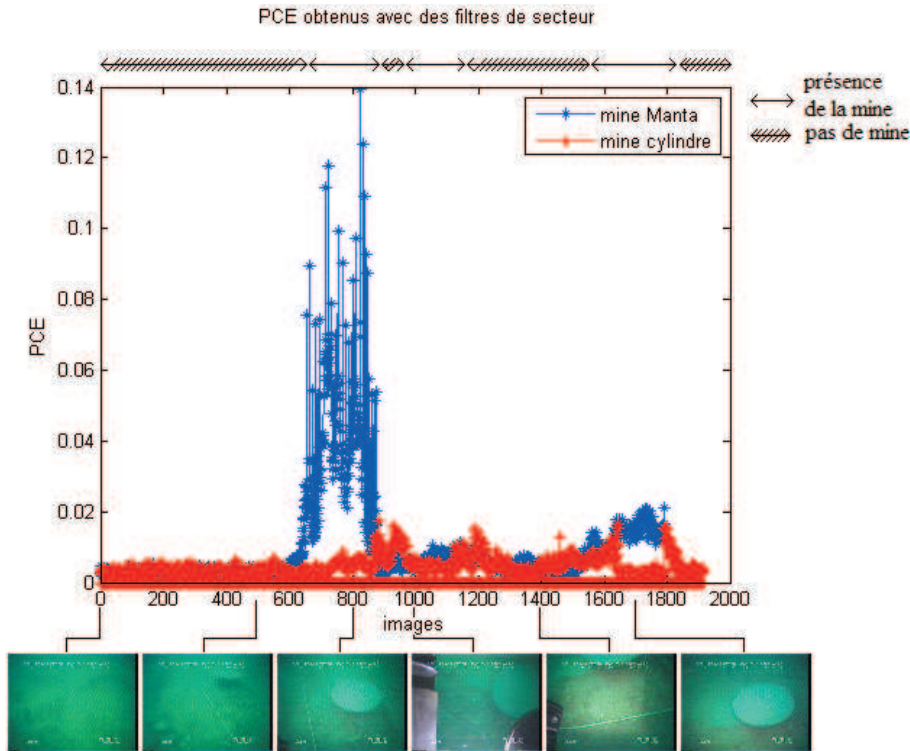


FIGURE 7.3 – Résultats de la corrélation entre une vidéo de mine Manta (Topvision003, courbe bleue) et une vidéo de mine cylindrique (Topvision005, courbe rouge) et des filtres de secteur créés avec des images de référence de mine Manta sélectionnées à partir de la vidéo Topvision003. Les images présentées sont des images issues de la vidéo Topvision003 avant prétraitement

7.1.2 Détection et identification des mines sous marines avec l'architecture JTC

Dans cette étude, nous avons utilisé des filtres créés à partir d'images synthétiques, que nous présenterons dans la partie corrélation, section 9.2. Ces filtres ont pour objectif de détecter la mine dans les images 760 à 850 et 930 à 1130. Nous avons tout d'abord utilisé le JTC fringe-adjusted sans ordre zéro puis le JTC fringe-adjusted non-linéaire sans ordre zéro (*cf.* partie 6.2). Pour la décision, nous avons utilisé le critère que nous avons développé pour le JTC, basé sur la valeur du maximum du plan de corrélation à laquelle on a soustrait 3dB (*cf.* partie 6.2). Sur ce test, nous nous sommes aussi intéressés au post-traitement : comme il est très improbable qu'une mine disparaisse d'une image à l'autre, nous avons considéré que si une mine est détectée dans cinq des dix dernières images, et qu'elle n'est pas détectée dans l'image actuelle, alors elle est présente dans l'image cible étudiée.

La première partie de la vidéo Topvision003, images 1 à 1400, permet de calculer le taux de reconnaissance et le taux de fausses alarmes. En effet, les filtres ont été définis à partir de cette partie de la séquence. Les résultats obtenus sont présentés dans le tableau 7.2. La première ligne présente les résultats obtenus avec l'architecture JTC fringe-adjusted. Le taux de reconnaissance est élevé (93%) ainsi que le taux de fausses alarmes (56%). Cela montre que cette architecture n'est pas assez discriminante au contraire de l'architecture JTC fringe adjusted non linéaire qui est trop discriminante dans notre étude (un taux de reconnaissance de 52% et un taux de fausses

alarmes de 16%). Notre architecture optimisée (troisième ligne du tableau) montre que nous avons réussi à trouver un compromis entre les deux premières architectures avec un taux de reconnaissance de l'ordre de 60% et un taux de fausses alarmes de 10%. Ces derniers résultats sont présentés sur la figure 7.4. Cette figure montre que les fausses alarmes sont situées sur un faible intervalle d'images en début de séquence. Le faible taux de reconnaissance est dû à des arrêts de reconnaissance.

Méthode	Taux de reconnaissance	Taux de fausses alarmes
JTC fringe-adjusted	93.10%	56.44%
JTC fringe-adjusted non-linéaire	52.07%	16.29%
JTC fringe adjusted non-linéaire avec optimisation	65.52%	10.13%

Tableau 7.2 – Résultats obtenus avec les différentes architectures JTC sur la détection de la mine Manta

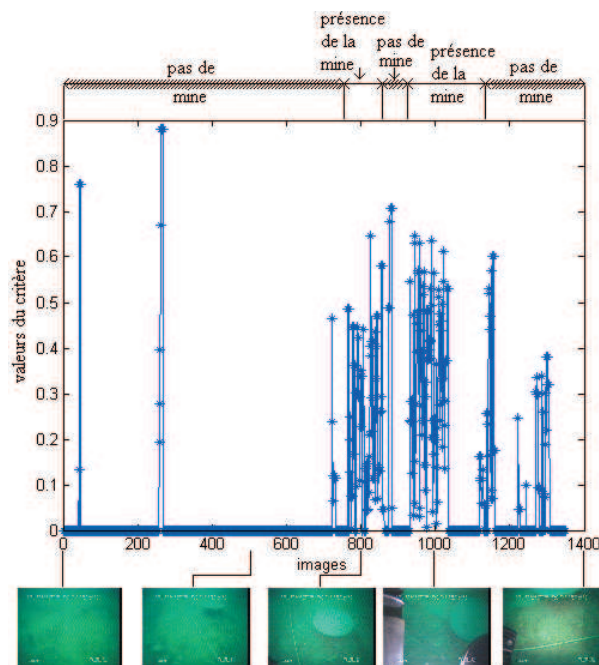


FIGURE 7.4 – Résultats obtenus avec le JTC fringe-adjusted non-linéaire sans ordre zéro avec notre critère et post-traitement des résultats

La deuxième partie de la vidéo Topvision003 (*cf.* figure 7.5) sert à tester la robustesse des architectures. En effet, comme dans la partie précédente, les mines présentent des orientations proches de celles recherchées mais avec une taille plus grande. Les résultats numériques sont présentés dans le tableau 7.3. L'architecture JTC fringe adjusted présente un taux de reconnaissance de 88% et un taux de fausses alarmes de 60%, proche des résultats obtenus sur la première partie de la vidéo. Ce qui confirme que cette architecture est peu discriminante dans la reconnaissance des mines. Le JTC fringe adjusted non linéaire est toujours trop discriminant avec un taux de reconnaissance de 34% et un taux de fausses alarmes de 12%. L'architecture JTC fringe adjusted non linéaire optimisée représente encore un compromis avec un taux de reconnaissance de 41% et un taux de fausses alarmes de 5%. Ces derniers résultats sont présentés sur la figure 7.5. La détection est très "hachée" : il y a peu de grandes zones continues de détection. Cependant cette architecture est relativement robuste comme le montre ce taux de reconnaissance de 41%.

L'association de notre filtre, de notre critère et de notre optimisation permet d'améliorer les résultats en reconnaissance par rapport aux résultats obtenus pour notre filtre et notre critère tout en diminuant fortement les fausses alarmes. Ce filtre permet d'allier robustesse et discrimination.

Méthode	Taux de reconnaissance	Taux de fausses alarmes
JTC fringe-adjusted	88.52%	59.93%
JTC fringe-adjusted non-linéaire	34.07%	12.33%
JTC fringe adjusted non-linéaire avec optimisation	41.11%	5.48%

Tableau 7.3 – Test de la robustesse des différentes architectures JTC sur la détection de la mine Manta

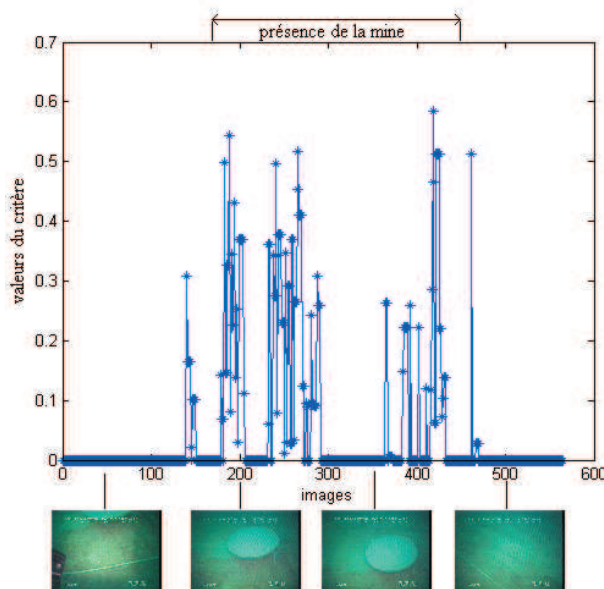


FIGURE 7.5 – Test de la robustesse du JTC fringe-adjusted non-linéaire sans ordre zéro avec notre critère et post-traitement des résultats

Le taux de reconnaissance n'est pas très élevé donc le filtre est un peu trop discriminant mais le taux de fausses alarmes est plus faible que pour les deux autres architectures testées ici, ce qui est un avantage dans la reconnaissance d'objets dangereux.

7.2 Problématique

La comparaison des différentes méthodes de corrélation et les premiers tests effectués sur les images Topvision nous ont permis d'affiner l'algorithme que nous allons utiliser dans la suite de cette thèse. Nous voulons un algorithme qui puisse fonctionner en temps réel, et qui détecte, classe et identifie les objets présents dans l'image. L'algorithme est résumé sur la figure 7.6.

Tout d'abord, nous avons vu dans l'état de l'art sur le sujet que les images nécessitent un premier prétraitement (*cf.* partie 2). De plus la corrélation sur les images non prétraitées ne donne pas de résultats. Cette partie correspond à partie bleue sur la figure 7.6.

La comparaison d'une image avec une large base de données nécessite un temps de calcul important. Pour limiter les corrélations sur des images ne contenant pas d'objets, nous avons décidé d'introduire une étape de détection (partie verte sur la figure 7.6). Cette étape sera suivie d'un post-traitement qui permettra de limiter les non-détections au sein d'une zone de présence de l'objet et de limiter les fausses alarmes. En cas de détection d'un objet, l'image sera traitée de façon à obtenir une image de contours. Ces deux étapes de post-traitements correspondent au cadre vert clair sur la figure 7.6. Cette partie sera développée dans le chapitre 8.

Ensuite les étapes de classification et d'identification sont effectuées en même temps par la corrélation, partie rouge de la figure 7.6. Les résultats en temps de calcul obtenus avec le JTC ne permettent pas d'obtenir un algorithme de calcul en temps réel, du fait de la taille de l'image

calculée. Notre choix s'est porté sur l'architecture Vanderlugt et les filtres de type composites POF fabriqués à partir de références synthétiques (*cf.* cadre gris sur la figure 7.6). Les filtres composites permettent ainsi de réduire le nombre des filtres nécessaires pour prendre une bonne décision. Ce choix est confirmé par une comparaison avec les différents filtres de corrélation classiques (POF, AMPOF,...). Cette comparaison a montré la nécessité de créer un nouveau filtre adapté à notre application. Ce filtre est appelé ASPOF (*cf.* partie 5.3). Le principe du filtre ASPOF a été validé mais il n'a pas pu être utilisé par la suite. Cela sera le sujet d'un travail post-doctoral.

Les images synthétiques représentent un choix intéressant : les images issues des vidéos réelles ne peuvent pas représenter les mines à toutes les distances (effet échelle) ni avec toutes les orientations possibles. De plus ces images réelles sont perturbées par le bruit de la vidéo, présent dans les images prises comme images de référence, bruit qui peut être différent dans la suite de la vidéo. En effet, le bruit dû au milieu sous marin varie au cours du temps et selon le lieu de la prise de vue. Les modèles 3D, utilisés pour obtenir les images synthétiques, présentent l'avantage de ne pas être bruités et de pouvoir fournir des images de références selon les orientations et les distances voulues. Cette partie sera développée dans le chapitre 9.

La décision se fera à partir du critère PCE et d'un post-traitement basé sur le fait que la mine a peu de chance de changer de type d'une image à l'autre (tout en gardant à l'esprit que certaines vues peuvent amener à une confusion du type de mines, par exemple le cylindre vu par l'extrémité ronde, la Manta vue de dessous et la sphère).

Dans la perspective d'effectuer le suivi de l'objet tout au long de la vidéo, nous pensons utiliser l'architecture JTC, déjà utilisée pour ces applications dans la littérature [120].

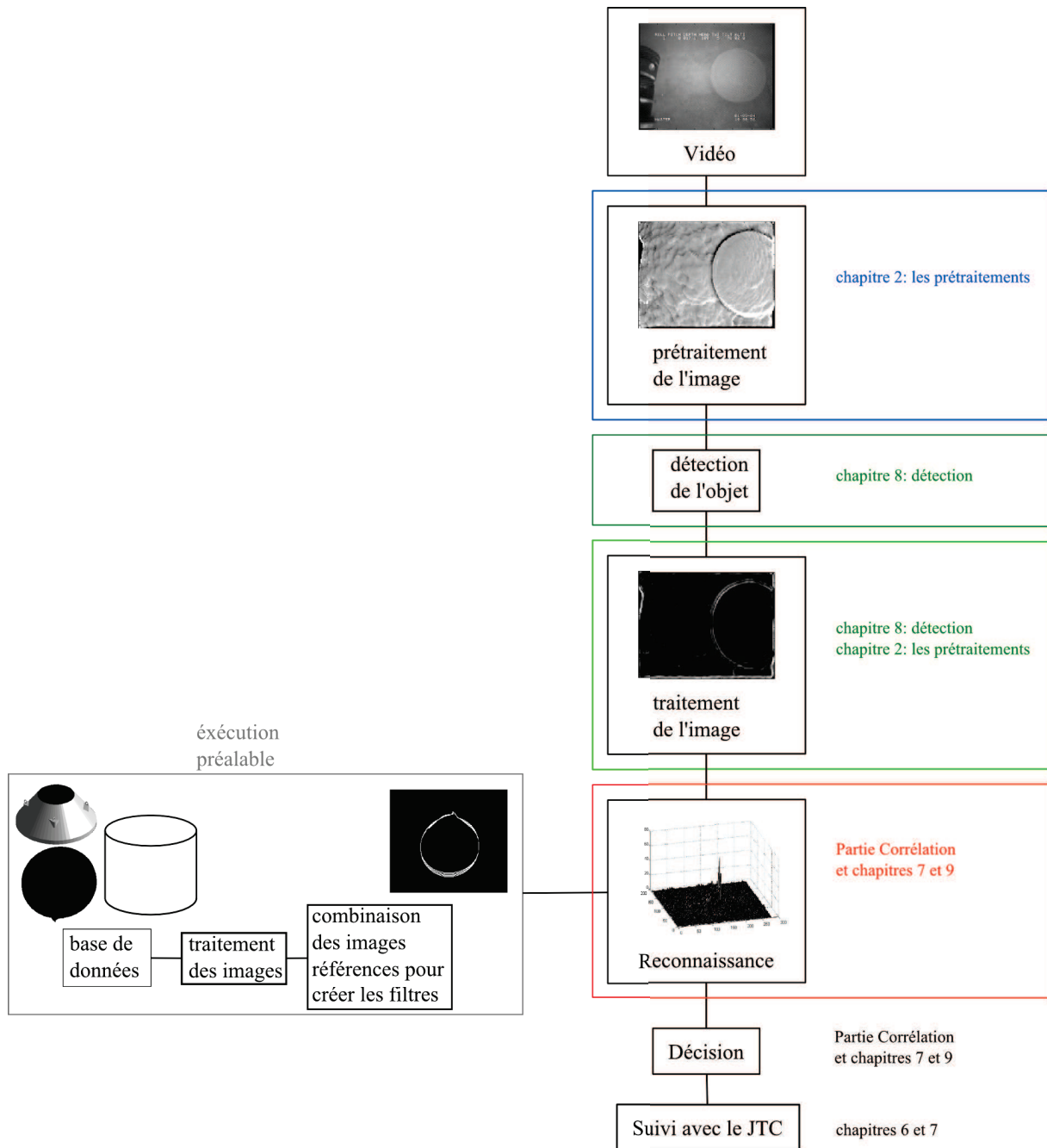


FIGURE 7.6 – Diagramme de l'algorithme utilisé

Chapitre 8

La détection des objets et leur position

Sommaire

8.1	Choix de la représentation des résultats	112
8.2	Méthode basée sur la phase	115
8.2.1	Algorithme utilisé	115
8.2.2	Résultats obtenus	120
8.3	Méthode basée sur la soustraction du fond	120
8.3.1	Algorithme utilisé	121
8.3.2	Résultats obtenus et améliorations apportées	122
8.3.3	Application de l'algorithme aux vidéos Topvision	124
8.4	Évaluation de notre algorithme	124
8.5	Perspectives	126
8.5.1	Optimisations à apporter à nos algorithmes	126
8.5.2	Des méthodes pour formaliser nos algorithmes	127
8.6	Conclusion	129

Les premiers tests effectués (*cf.* partie 7.1) nous ont montré que la corrélation est une solution applicable à la détection, la classification et l'identification des mines sous marines. Ces tests nous ont aussi montré que le temps de calcul nécessaire est trop long pour envisager une application en temps réel (environ 30 s/im). Nous avons donc décidé d'insérer une étape de détection dans notre algorithme. A l'issue de cette étape, si un objet est détecté, l'étape d'identification basée sur la corrélation sera effectuée sinon on traitera l'image suivante. Cette étape de détection a pour but de limiter le nombre de corrélations à effectuer. Il est aussi possible d'utiliser cette étape pour limiter la taille de l'image à identifier. En effet, si on a confiance dans notre détection, on peut réduire la taille de l'image à la taille de la zone dans laquelle on a détecté un objet.

Une des contraintes de notre système est le temps de calcul. L'objectif fixé est d'obtenir une fréquence de 5 Hz pour la détection des objets, soit de traiter 5 images par seconde. Pour cela, nous devons trouver un algorithme de détection efficace et très rapide.

Nous avons tout d'abord testé et validé une méthode basée sur les images de contours, obtenues grâce à la phase de l'image (*cf.* partie 2.2.5). Cette méthode est efficace sur les images contrastées. Mais rappelons que nos images dépendent fortement du milieu sous marin qui absorbe et diffuse la lumière. La turbidité et les conditions météorologiques peuvent changer la qualité des images. En conséquence, certaines images sont peu contrastées. Ainsi, nous avons développé une autre méthode basée sur la soustraction du fond. Ces deux méthodes seront détaillées et validées dans ce chapitre. Ensuite nous avons cherché à comparer la deuxième méthode avec une méthode existante, développée par Thales à base de segmentation et d'extraction de contours [8].

8.1 Choix de la représentation des résultats

Tout d’abord, nous commençons ce chapitre par une exposition des choix que nous avons effectués pour bien représenter et interpréter les résultats visuellement et de façon numérique. Nos réflexions ont évolué au cours de cette thèse pour arriver au choix final fait pour représenter les résultats sur les images VAMA, choix que nous avons appliqué par la suite aux résultats déjà obtenus sur les images Topvision.

Initialement, sur les images Topvision, nous représentions les résultats de la détection en fonction des images traitées (*cf.* figure 8.1). Les zones de présence et d’absence de mines étaient définies manuellement. Une détection sur une image contenant une mine était considérée comme une *bonne détection*, une détection sur une image ne contenant pas de mines était considérée comme une *fausse alarme*. En comptabilisant les résultats sur toute la vidéo, nous obtenions un taux de détection et un taux de fausses alarmes. A titre d’exemple, sur la figure 8.1, nous avons un taux de détection de 73.59% et un taux de fausses alarmes de 3.68%. (La méthode de détection utilisée pour obtenir ces résultats est la méthode basée sur la phase, que nous présenterons dans la suite de ce chapitre.)

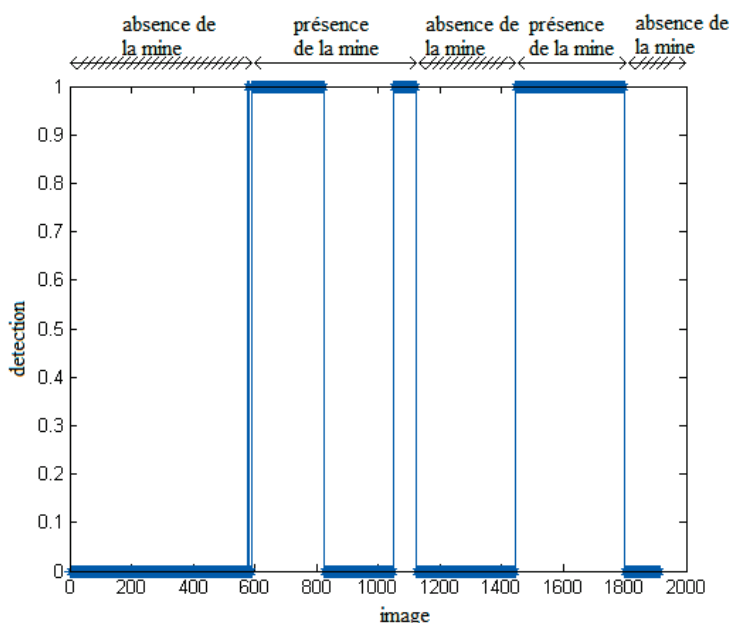


FIGURE 8.1 – Résultats de la détection sur une vidéo de mine Manta

Cette méthode de représentation et d’interprétation avait un inconvénient non négligeable : si une détection avait lieu dans une image contenant une mine mais que l’algorithme ne détectait pas du tout la mine, nous considérons que la mine était détectée. Nous avons résolu ce problème en comparant les détections avec les annotations des vidéos et en définissant quatre types de probabilités (*cf.* figure 8.2).

		Réalité	
		Mine	Rien
Résultats	Mine	Pd	Pfa
	Rien	Pndf	Pndv

FIGURE 8.2 – Définition des différentes probabilités utilisées

- Si le centre de la zone détectée par nos algorithmes est comprise dans la région annotée, nous avons une *bonne détection* (notée P_d , cf. figure 8.3).

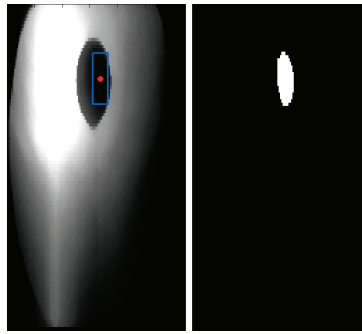


FIGURE 8.3 – Exemple d'une bonne détection

- Si une détection ne correspond pas à la mine dans une image contenant une mine ou que l'image est vide, nous avons une *fausse alarme* (notée P_{fa} , cf. figure 8.4).

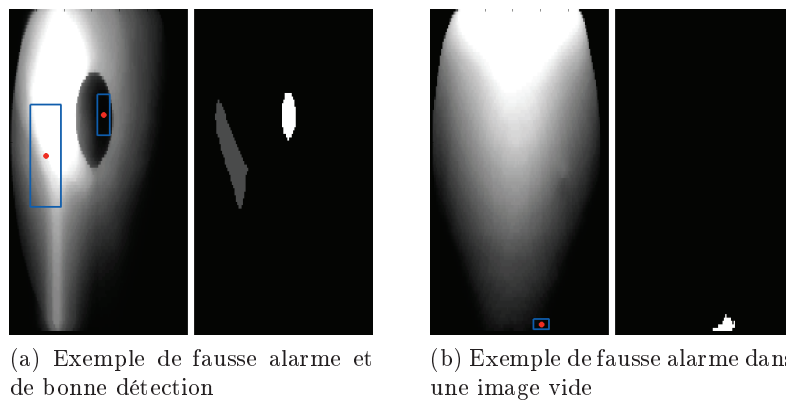


FIGURE 8.4 – Exemples de fausses alarmes

- S'il n'y a aucune détection dans une image contenant une mine, nous avons une *non détection fausse* (notée P_{ndf} , cf. figure 8.5).

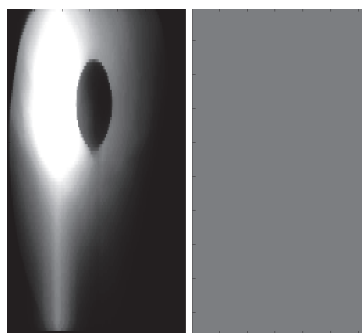


FIGURE 8.5 – Exemple d'une non détection fausse

- S'il n'y a pas de détection dans une image vide, nous obtenons une *non détection vraie* (notée P_{ndv} , cf. figure 8.6).

Un deuxième problème s'est posé : comment définir les bornes des zones de présence de la mine ? Pour mesurer les performances de notre algorithme, il faut savoir quelles images contiennent une mine et quelles images n'en contiennent pas. La méthode la plus intuitive est d'annoter les images manuellement. Le problème avec cette méthode est d'appliquer les mêmes critères à toutes les vidéos : à quel moment considère-t-on qu'une mine est présente dans l'image ?

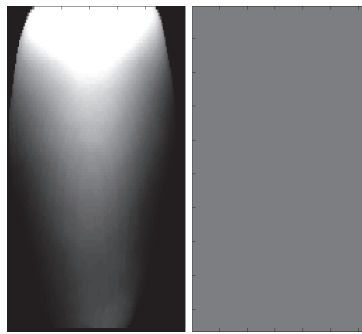


FIGURE 8.6 – Exemple d'une non détection vraie

Est ce lorsqu'il y a quelques centimètres de la mine dans l'image, est ce lorsqu'elle est entièrement dans l'image? De plus, selon la visibilité l'appréciation de ce critère de détection peut varier. Enfin il dépend de chaque opérateur. Pour résoudre ce problème, nous avons choisi d'utiliser un critère objectif : la distance. En effet, nous avons possibilité de connaître la distance entre la caméra et la mine (ou le sol). Après étude de nos images, nous avons décidé que l'algorithme doit détecter un objet lorsque la distance est inférieure ou égale à 6m. Au dessus de cette distance, les objets ont une taille trop petite pour être détectés avec certitude. Ce critère objectif a l'avantage de permettre la comparaison des résultats entre vidéos et l'étude des résultats en fonction de la distance entre la caméra et l'objet.

Après réflexions, nous avons décidé de représenter les résultats selon la distance au sol pour les images Topvision, à la mine pour les images VAMA. Pour les images Topvision, nous n'avons pas accès à la distance à la mine. Pour ne passer trop de temps à la calculer, dans un premier temps, nous avons utilisé la distance au sol, connaissant l'altitude et l'inclinaison de la caméra, *cf.* figure 8.7. La représentation en fonction de la distance à la mine est plus cohérente avec la navigation du véhicule. En effet, le drone et la caméra se rapprochent de l'objet. Il est donc intéressant d'analyser l'évolution des taux de détection et de fausses alarmes par rapport à la distance à la mine.

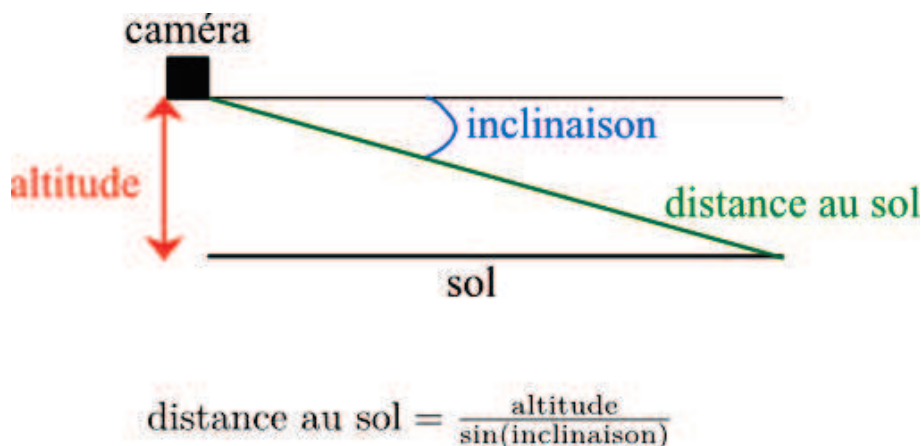


FIGURE 8.7 – Schéma expliquant le calcul de la distance au sol pour les images Topvision

Pour chaque distance, nous avons choisi de représenter le nombre d'images dans lesquelles la mine est correctement détectée (dans le cas de la bonne détection) par rapport au nombre d'images contenant une mine. Comme il n'est pas raisonnable de calculer autant de probabilités que de distances caméra-objet, nous avons décidé de regrouper les images par intervalle de distance de 0.5 m (par exemple les images avec une distance comprise entre 0 m et 0.5 m sont regroupées, de même que les images avec une distance comprise entre 0.5 m et 1 m) pour les images VAMA où la plage de distances couvertes est importante et de 0.1 m pour les images Topvision où la plage de distances couvertes est petite (souvent inférieure à 3 m). Les taux de probabilité sont calculés comme un rapport d'aire. Dans le cas de la probabilité de détection :

- pour chaque intervalle de distance d , nous avons le nombre d’images correctement détectées, $Pd(d)$ et le nombre d’images à cette distance, $nb(d)$
- nous pouvons calculer l’aire occupée par toutes les images contenant une mine :

$$\text{aire}_{\text{tot}} = \sum_d \text{intervalle de distance } d \times nb(d)$$
- nous pouvons calculer l’aire occupée par la probabilité de détection :

$$\text{aire}_{\text{proba}} = \sum_d \text{intervalle de distance } d \times Pd(d)$$
- La valeur numérique de la détection correspond au rapport $\frac{\text{aire}_{\text{proba}}}{\text{aire}_{\text{tot}}}$

Les résultats présentés sur la figure 8.1 ont été représentés avec la nouvelle méthode sur la figure 8.8. Nous obtenons dans ce cas un taux de bonnes détections de 68.21%, un taux de fausses alarmes de 2.74%, un taux de non détection fausses de 24.88% et un taux de non détection vraie de 97.33%. Les astérisques rouges représentent les distances pour lesquelles nous n’avons pas d’images à étudier dans cette vidéo.

La comparaison des résultats obtenus avec les annotations faites sur les images (taille et position de la mine) permettent d’affiner le taux de bonnes détections (68.21% contre 73.59% précédemment). Le taux de fausses alarmes est aussi légèrement différent (2.74% contre 3.68% précédemment). L’évolution de ces résultats confirme la nécessité d’utiliser des critères objectifs indépendants des opérateurs.

Sur la figure 8.8, nous pouvons voir sur les graphes de la probabilité de détection, de la probabilité de non détection fausse et de la probabilité de non détection vraie qu’il y a des intervalles de distance où ces probabilités sont égales à 1. Cela signifie que toutes les images contenant une mine de cet intervalle de distance ont été identifiées comme contenant une mine (dans le cas de la probabilité de détection). Le calcul de la probabilité de non détections vraies (97.33%) montre que notre algorithme donne peu de fausses alarmes dans les images vides. Le taux de non détections fausses sera réduit avec la prise en compte de l’aspect temporel des images. En effet, actuellement, les résultats sont calculés image par image. Or dans une vidéo, les images se suivent et dépendent les unes des autres. Cette dépendance servira à améliorer les résultats.

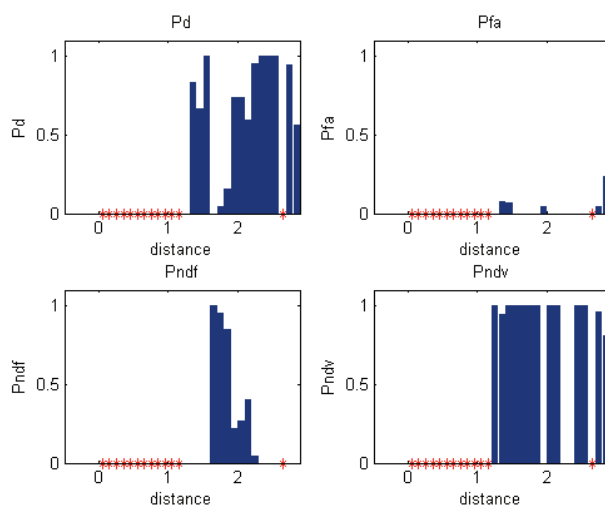


FIGURE 8.8 – Résultats de la détection sur une vidéo de mine Manta

8.2 Méthode basée sur la phase

8.2.1 Algorithme utilisé

Nous avons vu dans la partie 2.2.5 que l’image reconstruite à partir de la phase du spectre de l’image initiale contient essentiellement les contours contenus dans l’image spatiale initiale [67]. Nous utilisons cette propriété pour détecter les contours dans les images Topvision.

Tout d'abord les images sont prétraitées selon la chaîne présentée dans la partie 2.3 (*cf.* figures 8.9a et 8.9b pour la mine Manta et figure 8.10a et 8.10b pour la mine cylindrique). Ensuite, ces images sont transformées en images de phase de façon à obtenir les contours des objets présents dans ces images (*cf.* figures 8.9c et 8.10c). Comme on peut le constater, les contours sont plus ou moins bien visibles. Pour détecter un objet, il suffit de détecter les contours et de les analyser.

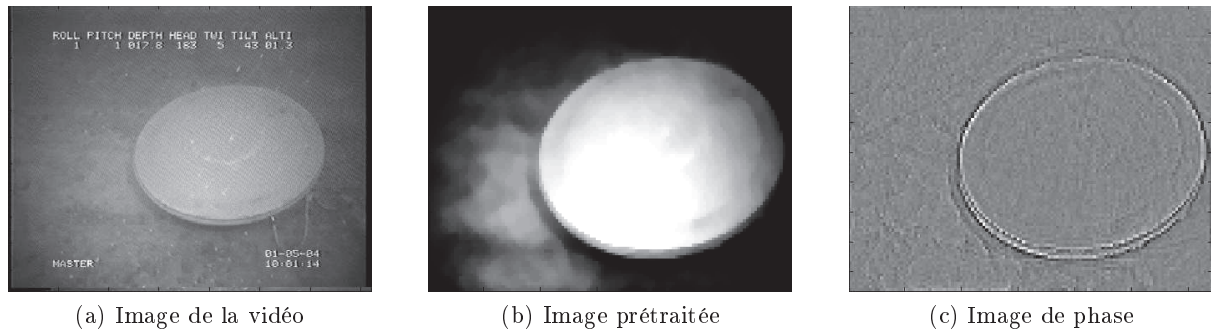


FIGURE 8.9 – Image utilisée pour la détection d'une mine Manta

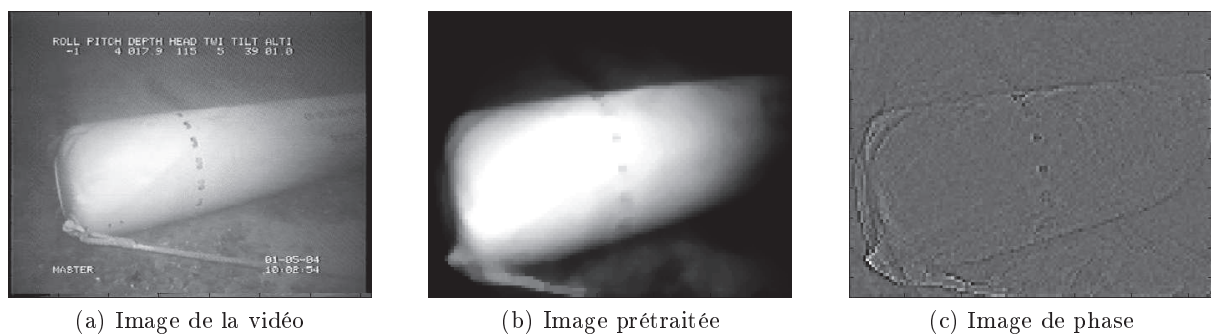


FIGURE 8.10 – Image utilisée pour la détection d'une mine cylindrique

Nous avons étudié des images fixes. Cependant, elles sont issues d'une vidéo. Le bruit présent dans une image est dépendant du présent dans les autres images. Il est donc possible d'étudier le bruit présent dans des images ne contenant pas d'objet pour limiter l'effet du bruit lors de la détection.

Nous avons ainsi traité les cent premières images de la vidéo, images qui ne contiennent pas de mines. Les images de phase du fond obtenues sont moyennées. Ensuite on prend la médiane de chaque ligne afin d'obtenir un vecteur colonne représentatif du fond. Cela nous permet de connaître le bruit de fond que l'on est susceptible de retrouver dans nos images (*cf.* figure 8.11). La différence entre deux pixels appartenant au fond et au contour doit être supérieure au bruit de fond.

Pour chacune des images de la vidéo, pour détecter les sauts de phase, nous avons décidé de découper l'image en six bandes verticales (*cf.* figure 8.12). La médiane du fond (*cf.* figure 8.11) est soustraite à la moyenne de chaque bande. Le résultat de cette soustraction est présenté pour chacune des bandes de la figure 8.11 sur la figure 8.12 avec les courbes bleues. Les sous figures de la figure 8.13 présentent l'intensité des vecteurs colonnes représentatif des bandes. Ensuite on compare le résultat à un seuil fixé à deux fois la médiane de la bande (*cf.* figure 8.13, courbes rouges). Ce seuil dépend donc de la bande de l'image étudiée. Le choix de la médiane de la bande s'est effectué pour éviter d'obtenir beaucoup de sauts de phase dans une bande qui pourrait être bruitée pour une raison quelconque (amas de cailloux, de bouts, problème d'enregistrement de la vidéo). Nous avons fixé un seuil suffisamment élevé de façon à éviter d'obtenir une quantité importante de sauts de phase qui ne correspondraient pas à des contours d'objets. On peut

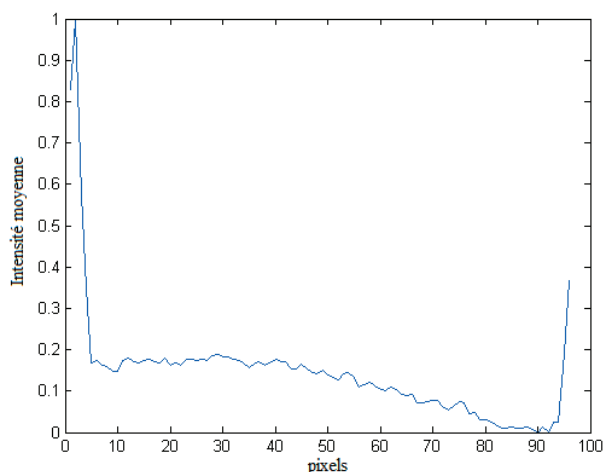


FIGURE 8.11 – Valeur médiane du fond

résumer cette partie par la formule suivante :

$$x - y > 2z \quad (8.1)$$

avec x la moyenne de la bande n , y la médiane du fond et z la médiane de la bande n

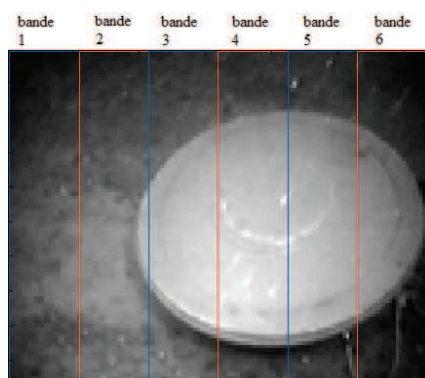


FIGURE 8.12 – Image illustrant les bandes verticales sur lesquelles sont basées la méthode. En bleu les bandes d'indice impair et en rouge les bandes d'indice pair

Nous comparons ensuite les sauts enregistrés. Si au moins deux sauts de phase ont un équivalent dans une autre bande, alors nous avons considéré qu'un objet était probablement dans l'image. Nous cherchons les sauts d'ordonnées la plus faible et la plus forte ainsi que les bandes dans lesquelles ils ont été détectés. La boîte définie par ces deux ordonnées et ces deux largeurs est considérée comme la boîte englobante de l'objet (*cf.* figure 8.14). Cette méthode nous permet d'obtenir une précision de l'ordre du pixel en hauteur et de la largeur de la bande en largeur. Des exemples de résultats sur la mine Manta sont présentés sur la figure 8.14. Ces résultats montrent que la boîte englobante peut avoir la largeur d'une bande (*cf.* figure 8.14b), de plusieurs bandes (*cf.* figure 8.14c) et que l'algorithme fonctionne aussi pour des mines qui ne sont pas entièrement présentes dans l'image (*cf.* figure 8.14a). La différence de largeur entre des boîtes englobantes s'explique facilement : nous comparons la moyenne de la bande à laquelle on a soustrait la médiane du fond au double de la médiane de la bande. Dans le cas de la figure 8.14b, la mine (dont les pixels ont une forte intensité) occupe la majorité de la hauteur de l'image. Les valeurs de la moyenne et de la médiane sont élevées. Le seuillage ne permet pas d'obtenir des sauts de phase dans plusieurs bandes de l'image.

L'algorithme présenté fournit les résultats présentés sur la figure 8.8. Comme nous le disions dans la partie précédente, le taux de bonnes détections est de 68.21%, le taux de fausses alarmes est de 2.74%, le taux de non détection fausses est de 24.88% et le taux de non détection vraie est

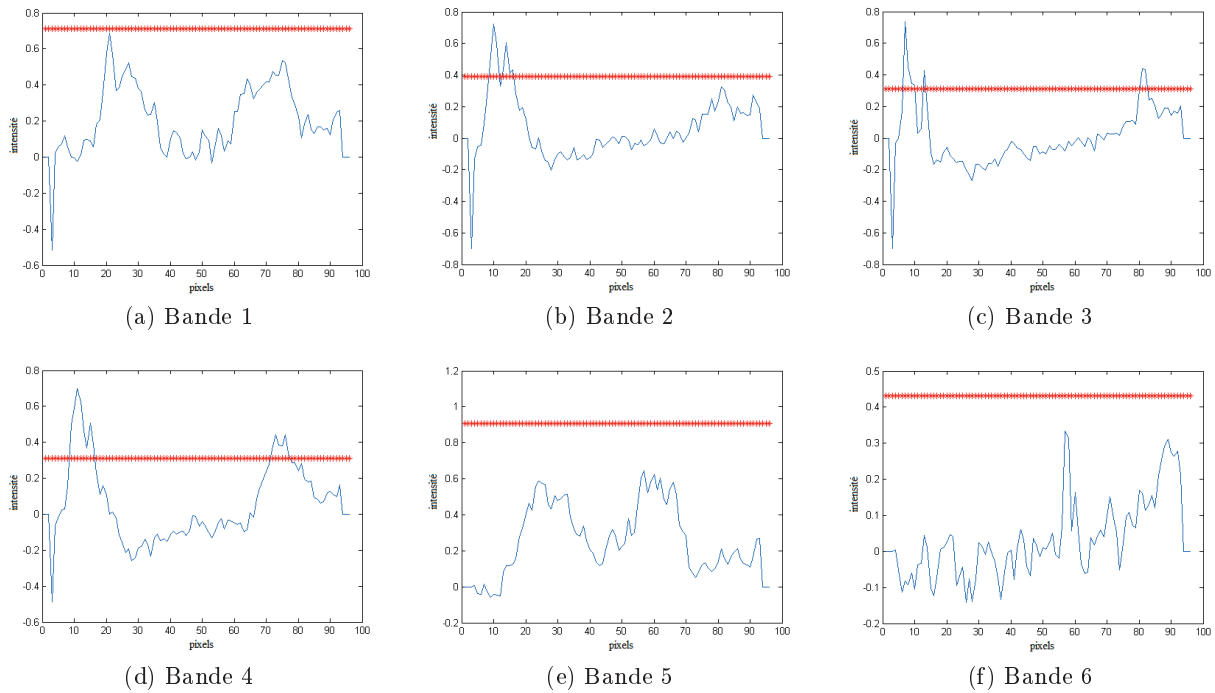


FIGURE 8.13 – Valeurs des moyennes auxquelles a été soustraite la médiane du fond (en bleu). En rouge : seuil

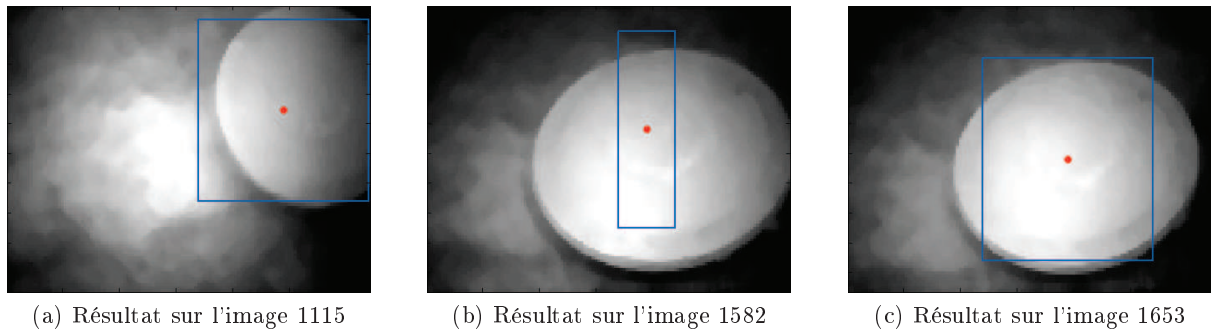


FIGURE 8.14 – Résultats de l'algorithme de détection basé sur la phase sur les images Topvision

de 97.33%. Avec Matlab et le PC ISEN (voir les caractéristiques dans l'annexe E), le temps de calcul nécessaire moyen est de 0.08 s/im soit une fréquence de 12 Hz. En analysant les résultats et en visualisant les zones détectées sur les images, il nous a paru nécessaire d'optimiser ces résultats pour augmenter le taux de bonnes détections et réduire le taux de non détections fausses.

Une première optimisation a été effectuée. Pour limiter les fausses alarmes et les arrêts de détection, nous avons ajouté un lissage simple : lorsqu'aucun saut n'est détecté, on regarde les résultats obtenus sur les dix dernières images. S'il y a eu des détections (au moins cinq sur les dix dernières images), on considère que l'on a une détection dans l'image étudiée (un objet ne doit pas disparaître d'une image à l'autre sauf lorsqu'il est situé sur le bord de l'image). Ensuite on lisse les positions des boîtes englobantes à l'aide de la médiane des trente dernières positions. En effet, les positions détectées peuvent varier beaucoup d'une image à l'autre. Ce lissage permet d'obtenir une certaine continuité dans les détections.

Nous avons aussi identifié des fausses alarmes susceptibles d'être évitées avec un *a priori* : en supposant que le drone ou le plongeur avance, le tilt (l'inclinaison) de la caméra nous renseigne sur la position de la mine dans l'image lors de son apparition. En effet si la caméra est fortement inclinée, la mine apparaît par le haut de l'image. Dans le cas où la caméra est parallèle au sol ou faiblement inclinée, la mine apparaît au centre de l'image. Les détections éloignées de

la zone d'apparition de la mine peuvent ainsi être considérées comme des fausses alarmes que nous pouvons filtrer. Pour mettre en œuvre cette méthode, nous avons eu besoin d'extraire les informations de navigation des images. Pour cela nous avons utilisé l'algorithme présenté dans l'annexe B.

Ces optimisations ont été appliquées. Les nouveaux résultats obtenus sur les images de la vidéo de la mine Manta sont présentés sur la figure 8.15. Le taux de détection est de 73.31%, le taux de fausses alarmes est de 1.76%, le taux de non détections fausses est de 21.66% et le taux de non détections vraies est de 98.77%. Ces optimisations ont donc permis d'augmenter le taux de détection (73.31% contre 68.21%). De plus les taux de fausses alarmes et de non détections fausses ont été réduits (1.76% contre 2.74% et 21.66% contre 24.88%). Sur la figure 8.15, nous pouvons remarquer, comparativement à la figure 8.8, que le taux de détection augmente significativement lorsque la caméra est le plus loin possible de la mine (dans cette vidéo). Les non détections vraies ont elles aussi été améliorées. Cela entraîne une diminution du nombre de fausses alarmes et de non détections fausses. Les optimisations ont permis de rendre notre algorithme plus performant sur cette vidéo de mine Manta.

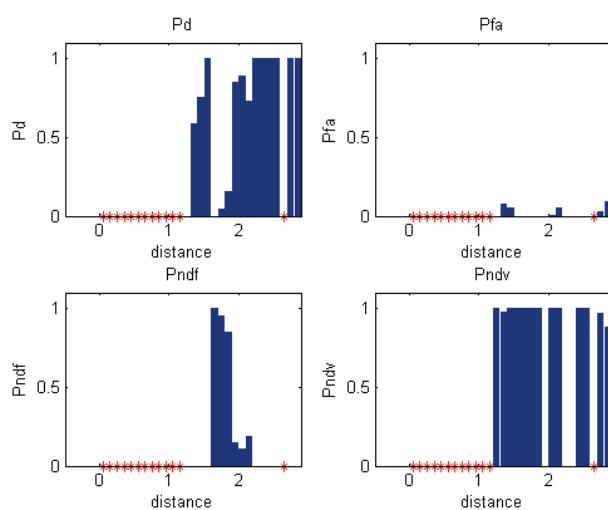


FIGURE 8.15 – Résultats après optimisation de la détection de la mine Manta

Nous avons aussi appliqué cette méthode sur une vidéo de mine cylindrique. Initialement, le taux de détection était de 67.03%, le taux de fausses alarmes de 5.29%, le taux de non détections fausses de 29.29% et le taux de non détections vraies de 96.07%. Après les optimisations, le taux de détection atteint 76.50%. Les résultats sont présentés sur la figure 8.16. Les résultats sur cette vidéo confirment les conclusions que nous avons présentées sur la vidéo de la mine Manta.

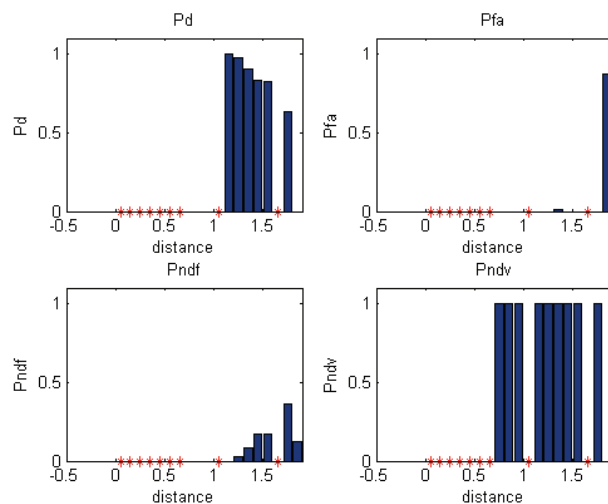


FIGURE 8.16 – Résultats après optimisation de la détection de la mine cylindrique

8.2.2 Résultats obtenus

Dans le tableau 8.1, nous résumons les résultats obtenus. Nous ne possédons qu'une vidéo de mine Manta (avec les annotations utiles) et une vidéo de mine cylindrique. Les résultats des vidéos vides et de mines sphériques sont les résultats de deux vidéos à chaque fois. Cela explique le "faible" nombre d'images étudiées. Les taux de détection sont élevés : 61% pour les mines sphériques, 73% pour les mines Manta et 76% pour les mines cylindriques. Il faut noter le taux très élevé de non détections vraies, proche de 100% pour ces trois types de mines. Ces valeurs montrent l'efficacité de notre algorithme. De plus le faible taux de fausses alarmes (6% pour les mines sphériques, 2% pour les mines Manta et 15% pour les mines cylindriques) montre sa robustesse.

Mine	nombre d'images testées	nombre d'images contenant une mine	Pd	Pfa	Pndf	Pndv
Manta	1915	930	73.31%	1.76%	21.66%	98.77%
Cylindre	1792	1240	76.50%	14.61%	5.54%	98.90%
Sphère	2229	1080	61.27%	6.36%	22.37%	100%
Vidéos vides	3779	0		10.62%		89.38%

Tableau 8.1 – Tableau des résultats de détection obtenus sur les images Topvision

Notre algorithme a l'inconvénient de ne proposer qu'une seule détection par image. En cas de présence de plusieurs objets à détecter, cela peut poser problème. Les taux de détection sont supérieurs à 60%. Le taux de fausses alarmes est un peu élevé pour les mines cylindriques et les vidéos vides. De plus, cet algorithme a une fréquence de 12 Hz sur Matlab et le PC ISEN (pour les caractéristiques, voir annexe E), ce qui remplit les objectifs fixés.

Nous avons ensuite voulu confirmer ces résultats sur les images VAMA. Les résultats obtenus n'étaient pas ceux espérés (*cf.* tableau 8.2). Le taux de détection des mines sphériques est de 31% et un taux de fausses alarmes de 70%. Cela montre que notre algorithme détecte des variations du fond plus que la mine elle-même. Ce qui est confirmé par le taux de non détections vraies qui est très faible, 7%. Ces résultats sont proches de ceux obtenus pour les objets posés au sol qu'ils soient des mines ou non : un taux de bonnes détections faible et un taux de fausses alarmes élevé. Ces valeurs ont plusieurs explications : le faible contraste des objets par rapport à l'arrière plan, le halo lumineux ainsi que le manchon présent sur les images. La fréquence de calcul sur le PC Thales (pour les caractéristiques, voir annexe E), est de 2.42 Hz.

mine	nombre d'images testées	nombre d'images contenant une mine	Pd	Pfa	Pndf	Pndv
Manta	25205	18275	18.87%	71.64%	42.34%	14.74%
Cylindre	49251	37564	32.42%	46.72%	20.96%	33.21%
Sphère	11376	7919	31.45%	69.94%	25.95%	7.10%
Autres objets	13905	10222	17.92%	71.15%	26.90%	14.35%
Vidéos vides	17389			8.97%		74.90%

Tableau 8.2 – Tableau des résultats de détection obtenus sur les images Vama

8.3 Méthode basée sur la soustraction à l'image étudiée d'une image du fond

Nous avons vu dans la partie précédente que la détection de contours ne suffit pas toujours à détecter un objet. Nous avons donc cherché une méthode qui nous permette de segmenter nos images en régions. Après recherches bibliographiques, nous avons orienté notre choix vers les méthodes basées sur la soustraction du fond. Ces méthodes consistent à apprendre le fond d'une

image puis à le soustraire à l'image afin de faire ressortir des régions. Nous avons développé cet algorithme sur les images VAMA avant de l'adapter aux images Topvision.

8.3.1 Algorithme utilisé

8.3.1.1 Apprentissage du fond

Comme précédemment, nous avons choisi les cent premières images comme images pour le fond, après avoir vérifié que ces images ne contenaient pas de mines. Cette méthode ne convient pas à toutes les vidéos. En effet dans certaines vidéos, la mine apparaît rapidement après le début de l'enregistrement. Si nous avons suffisamment d'images ne contenant pas la mine à notre disposition, nous les utilisons pour définir le fond. Dans les cas où il n'y a pas d'images sans mine, nous utilisons le fond d'une autre vidéo après avoir vérifié que les deux fonds sont visuellement semblables. Les images du fond sont traitées de la même façon que les images VAMA (*cf.* partie 2.3).

8.3.1.2 Soustraction du fond et détection d'objets

Les images VAMA sont prétraitées selon la chaîne développée (*cf.* partie 2.3). Dans un premier temps, nous leur avons soustrait l'image du fond et avons cherché l'image de contours associée. Les résultats obtenus étaient proches de ceux obtenus avec la méthode précédente.

Nous avons alors décidé de comparer l'image prétraitée à l'image prétraitée à laquelle on a soustrait le fond. Nous avons remarqué que les parties intéressantes de l'image (objets, mine) sont claires dans une image et sombres dans l'autre (*cf.* figures 8.17 et 8.18). Nous avons donc défini, expérimentalement, deux seuils : le premier pour les zones claires et le second pour les zones sombres. On applique ces deux seuils sur les deux images (*cf.* figures 8.17a et 8.17b ainsi que figures 8.18a et 8.18b). On obtient des images binaires. On multiplie entre elles les images résultats. Les deux images binaires résultantes sont additionnées et contiennent les zones d'intérêt. L'image est labellisée afin d'accéder rapidement à chaque zone d'intérêt. Les équations correspondant à cet algorithme sont détaillées dans l'équation 8.2. Appelons I_p l'image prétraitée et I_s l'image prétraitée à laquelle on a soustrait le fond. L'indice c correspond aux zones claires, l'indice s aux zones sombres et l'indice b aux images binaires résultantes.

$$\begin{aligned}
 I_{1c}(i, j) &= \begin{cases} 1 & \text{si } I_p(i, j) > \text{seuil}_{\text{zones claires}} \\ 0 & \text{sinon} \end{cases} \\
 I_{1s}(i, j) &= \begin{cases} 1 & \text{si } I_p(i, j) < \text{seuil}_{\text{zones sombres}} \\ 0 & \text{sinon} \end{cases} \\
 I_{2c}(i, j) &= \begin{cases} 1 & \text{si } I_s(i, j) > \text{seuil}_{\text{zones claires}} \\ 0 & \text{sinon} \end{cases} \\
 I_{2s}(i, j) &= \begin{cases} 1 & \text{si } I_s(i, j) < \text{seuil}_{\text{zones sombres}} \\ 0 & \text{sinon} \end{cases} \\
 I_{b1} &= I_{1c} \cdot I_{2s} \\
 I_{b2} &= I_{1s} \cdot I_{2c} \\
 I_{\text{labellisée}} &= I_{b1} + I_{b2}
 \end{aligned} \tag{8.2}$$

Nous avons décidé d'inclure une condition sur la taille des zones d'intérêt : elles doivent être supérieures à 20 pixels. Cette méthode a l'avantage de permettre la détection de plusieurs objets dans une image. Les résultats obtenus sur une image de mine sphérique et sur une image de mine Manta sont présentés sur les figures 8.17 et 8.18. Sur l'image de la mine sphérique, la mine est détectée (*cf.* figure 8.17c). Sur l'image de la mine Manta, beaucoup de fausses alarmes sont détectées (*cf.* figure 8.18c). Des améliorations seront apportées à notre algorithme pour limiter ces fausses alarmes.

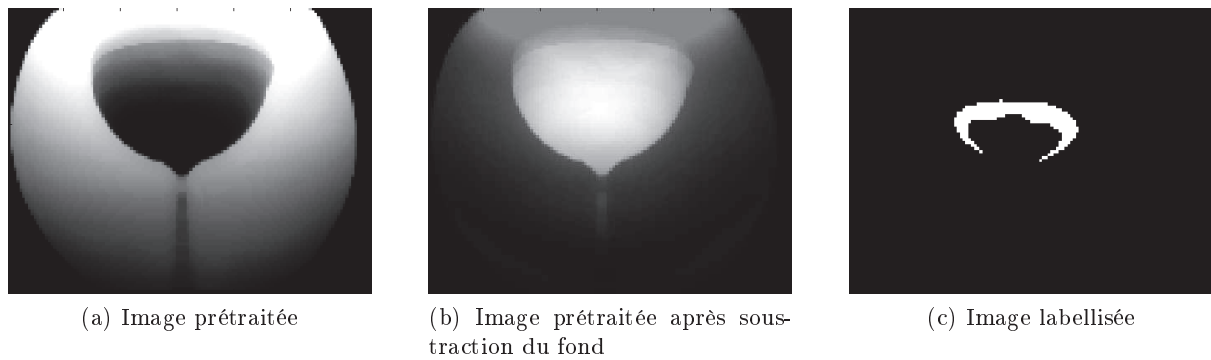


FIGURE 8.17 – Résultat de l’algorithme de détection basé sur la soustraction du fond pour une image de mine sphérique

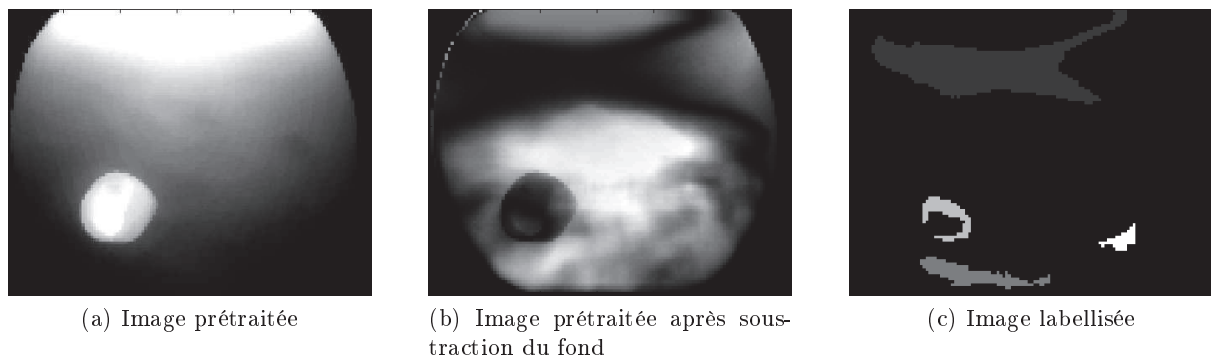


FIGURE 8.18 – Résultat de l’algorithme de détection basé sur la soustraction du fond pour une image de mine Manta

8.3.2 Résultats obtenus et améliorations apportées

La méthode décrite ci dessus nous a donné des résultats intéressants. Le tableau 8.3 donne les résultats numériques pour les différents types de vidéos. Le taux de bonnes détections est meilleur qu’avec la méthode précédente (*cf.* tableau 8.2). Le taux de fausses alarmes est plus faible mais le taux de non détections fausses plus élevé. Le taux de non détections vraies est presque parfait (supérieur à 95 %) pour les mines sphériques et Manta et les vidéos vides. La fréquence de calcul est similaire à l’algorithme précédent (2.38 Hz sur le PC Thales, pour les caractéristiques voir annexe E). Les résultats graphiques sont présentés dans l’annexe C.

Mine	Nombre d’images étudiées	Nombre d’images contenant une mine	Pd	Pfa	Pndf	Pndv
Manta	25205	18275	32.86%	20.83%	26.19%	96.61%
Cylindre	49251	37564	49.41%	62.37%	16.48%	31.61%
Sphère	11376	7919	43.23%	2.10%	56.41%	99.40%
Autres objets	13905	10222	46.82%	71.94%	34.72%	13.96%
Vidéos vides	17389	0		4.13%		95.87%

Tableau 8.3 – Tableau des résultats obtenus avec la méthode basée sur la soustraction du fond

Cependant, il y a comme dans la méthode basée sur la phase des fausses alarmes ainsi que des arrêts de détection. Pour remédier à ces problèmes, nous avons apporté à notre algorithme des améliorations et nous nous sommes imposés différentes contraintes.

Nous nous sommes servis des vidéos de sphères pour améliorer notre programme. Tout d’abord nous avons ajouté l’image traitée filtrée par un filtre passe bas à la moyenne du fond. Cela permet

d'adapter le fond aux changements importants qui peuvent survenir au cours de la vidéo.

Dans cette méthode aussi, nous avons décidé d'utiliser les informations de navigation pour améliorer les résultats. Les informations de navigation consistent en un fichier contenant les mouvements du drone au cours de la vidéo. Nous avons accès, entre autres, à la position (x, y et z) ainsi qu'à l'orientation du drone (inclinaison, cap et roulis). Ces informations peuvent être utilisées pour fixer des contraintes sur les résultats de la détection.

Dans le cas où un objet n'est pas détecté dans l'image, il sera considéré comme présent s'il remplit l'une des conditions suivantes :

- s'il était détecté depuis plus de 10 images
- s'il était détecté au centre de l'image
- si l'AUV (ou le ROV, ou le plongeur) n'a pas bougé

Nous avons vu dans la partie prétraitement que la détection de contours multi-échelle semblait efficace (*cf.* partie 2.2.5). Comme nous utilisons cette méthode pour traiter l'image avant la corrélation, nous pouvons effectuer ce traitement lors de la détection et comparer les résultats de la détection de contours avec l'image labellisée.

Enfin, nous connaissons la position et la taille des objets que nous cherchons ainsi que la distance séparant la caméra de l'objet. Si l'objet détecté est situé dans le volume d'eau, nous pouvons penser qu'il s'agit d'une mine sphérique (seule mine que nous cherchons pouvant être fixée à un orin). Par conséquent, nous pouvons inclure une contrainte basée sur l'estimation de l'objet : si l'objet détecté est beaucoup plus grand que l'objet cherché, il s'agit sans doute d'une fausse alarme. (Nos algorithmes ne détectant pas toute la surface de l'objet nous ne pouvons pas mettre de contraintes sur un objet détecté de taille très inférieure à celle de l'objet cherché). Cette contrainte ne fonctionne pas dans le cas des objets posés au sol puisque toutes les mines que nous cherchons peuvent être posées au sol.

Les résultats sont présentés dans le tableau 8.4. La fréquence de calcul est de 1.55 Hz. Comparativement aux résultats obtenus avant ces optimisations, présentés dans le tableau 8.3, nous pouvons remarquer une nette amélioration des résultats sur la vidéo de mine sphérique. Les résultats sur les vidéos vides sont semblables. Les optimisations permettent de baisser le taux de fausses alarmes et d'augmenter le taux de non détections vraies pour les vidéos d'objets posés au sol. Les optimisations ont été définies à partir des vidéos de mines à orin (mines entre deux eaux, attachées à un poids). La baisse des taux de détection des objets (mines et non mines) posés au sol montre qu'elles ne sont pas adaptées à toutes les détections. En effet, cela a pour conséquence de baisser fortement le taux de bonnes détections de la mine Manta. La modification de l'image d'arrière plan n'est pas suffisante et ne permet plus la détection de la mine. Définir des optimisations par rapport à des interprétations "humaines" des résultats peut entraîner comme ici une spécialisation de l'algorithme pour un objet particulier. Les graphes correspondant aux résultats sont présentés dans l'annexe C.

Mine	Nombre d'images étudiées	Nombre d'images contenant une mine	Pd	Pfa	Pndf	Pndv
Manta	25205	18275	7.98%	12.21%	48.27%	97.61%
Cylindre	49251	37564	37.69%	48%	23.87%	33.17%
Sphère	11376	7919	58.10%	1.13%	39.19%	99.80%
Autres objets	13905	10222	35.35%	44.57%	45.43%	40.52%
Vidéos vides	17389	0		4.18%		95.82%

Tableau 8.4 – Tableau des résultats obtenus avec la méthode basée sur la soustraction du fond, optimisée

8.3.3 Application de l'algorithme aux vidéos Topvision

Nous avons ensuite appliqué cet algorithme aux vidéos Topvision. L'algorithme est moins performant sur ces images contrastées dont l'arrière plan varie peu au cours de la vidéo. L'algorithme basé sur la soustraction du fond, avant optimisations, produit un taux de fausses alarmes très élevé (*cf.* tableau 8.5). La détection est très bonne (supérieure à 80%) pour les trois types de mines. Cependant le taux de fausses alarmes est comparable. Les taux de non détection sont nuls. Cela est dû, entre autres, au fait que cet algorithme autorise plusieurs détections par images. De plus, un objet est détecté dans toutes les images ce qui pose problème si l'on envisage de guider le véhicule à partir des résultats obtenus. Les taux de bonnes détections sont un peu meilleurs que ceux obtenus avec la méthode basée sur la phase (*cf.* tableau 8.1). Le temps de calcul nécessaire est comparable à celui de la méthode précédente (≈ 12 Hz). Les optimisations apportées à l'algorithme, développées à partir des images VAMA n'ont pas les mêmes effets sur les résultats Topvision (*cf.* tableau 8.6). Le taux de bonnes détections chute fortement. Les probabilités de fausses alarmes sont plus faibles et les probabilités de non détections vraies ont augmenté ce qui est positif. La fréquence de calcul a baissé du fait des nouveaux calculs sur les images (ajout de l'image filtrée par un filtre passe-bas à la moyenne du fond, calcul de l'image de contours avec la méthode basée sur l'approche multi-échelle).

Mine	Nombre d'images étudiées	Nombre d'images contenant une mine	Pd	Pfa	Pndf	Pndv	Fréquence de calcul
Manta	1915	930	80.79%	82.75%	0%	0%	11.22Hz
Cylindre	1792	1240	81.7%	78.25%	0%	0%	11.4Hz
Sphère	2229	1080	80.5%	77.81%	0%	0%	13Hz
Vidéos vides	3779	0		100%		0%	11.6Hz

Tableau 8.5 – Tableau des résultats obtenus sur les images Topvision avec l'algorithme initial

Mine	Nombre d'images étudiées	Nombre d'images contenant une mine	Pd	Pfa	Pndf	Pndv	Fréquence de calcul
Manta	1915	930	55.63%	71.49%	18.3%	0%	4Hz
Cylindre	1792	1240	77.84%	16.32%	19.4%	68.36%	4Hz
Sphère	2229	1080	58.77%	24.52%	36.04%	69.95%	4.19Hz
Vidéos vides	3779	0		72.73%		27.27%	4.17Hz

Tableau 8.6 – Tableau des résultats obtenus sur les images Topvision avec l'algorithme optimisé

8.4 Évaluation de notre algorithme

Nous avons deux algorithmes à notre disposition : le premier basé sur la phase et le second basé sur la soustraction du fond. Pour évaluer nos algorithmes, il serait intéressant de les comparer avec un algorithme existant.

Pour cette étape de comparaison, nous avons choisi l'algorithme développé par Cybernetix et Thales [8], appelé par la suite algorithme de référence. Cet algorithme est basé sur une détection de contours. Ceux qui correspondent à des lignes ou des ellipses sont sélectionnés comme formant les contours d'un objet. Cet algorithme est à notre disposition sous forme d'un logiciel exécutable fourni par TUS. De plus nous savons qu'il fonctionne sur les vidéos VAMA. Nous n'avons appliqué cet algorithme que sur les images VAMA.

Les choix effectués dans la partie 8.1 nous ont permis de comparer efficacement les algorithmes. En effet, les fichiers de résultats fournis par les différents algorithmes contiennent la présence ou non d'un objet et sa position. Nous possédons les annotations que nous avons effectuées sur les vidéos. Nous pouvons ainsi calculer les probabilités et les aires que nous avons choisies pour quantifier les algorithmes.

Dnas le tableau 8.7, nous avons résumé tous les résultats obtenus sur les images VAMA :

- les résultats obtenus avec l'algorithme basé sur la phase (en noir), présenté dans la partie 8.2.
- les résultats obtenus avec l'algorithme basé sur la soustraction du fond (en bleu), présenté dans la partie 8.3.1
- les résultats obtenus avec l'algorithme basé sur la soustraction du fond et optimisé (en vert), présenté dans la partie 8.3.2
- et enfin les résultats obtenus avec l'algorithme de référence (en rouge) [8]

Mine	Nombre d'images étudiées	Nombre d'images contenant une mine	Pd	Pfa	Pndf	Pndv
Manta	25205	18275	18.87%	71.64%	42.34%	14.74%
			32.86%	20.83%	26.19%	96.61%
			7.98%	12.21%	48.27%	97.61%
			19.78%	3.54%	70.51%	99.02%
Cylindre	49251	37564	32.42%	46.72%	20.96%	33.21%
			49.41%	62.37%	16.48%	31.61%
			37.69%	48%	23.87%	33.17%
			42.91%	10.68%	40.42%	84.20%
Sphère	11376	7919	31.45%	69.94%	25.95%	7.10%
			43.23%	2.1%	56.41%	99.4%
			58.10%	1.13%	39.19%	99.80%
			36.05%	3.83%	54.04%	99.64%
Autres objets	13905	10222	17.92%	71.15%	26.90%	14.35%
			46.82%	71.94%	34.72%	13.96%
			35.35%	44.57%	45.43%	40.52%
			18.30%	9.63%	73.02%	82.55%
Vidéos vides	17389	0		8.97%		91.03%
				4.13%		95.87%
				4.18%		95.82%
				8.57%		91.43%

Tableau 8.7 – Tableau comparant les résultats obtenus avec les méthodes développées, sur les vidéos VAMA. En noir les résultats obtenus avec l'algorithme basé sur la phase. En bleu, les résultats de l'algorithme basé sur la soustraction du fond. En vert, les résultats de l'algorithme basé sur la soustraction du fond optimisé. En rouge, les résultats de l'algorithme de référence.

Le taux de détection des mines Manta (deuxième ligne du tableau 8.7) est de 19% avec l'algorithme basé sur la phase. Cela s'explique par le fait que les images sont peu contrastées. Par conséquent, la détection de contours basée sur la phase du spectre ne donne pas de bons résultats. L'algorithme de référence obtient un taux de détection comparable. En effet, cet algorithme est basé sur une détection de contours. L'utilisation de la soustraction du fond permet d'améliorer significativement ce taux de détection (32%). Cependant les optimisations que nous avons apportées ne sont pas adaptées aux objets posés au sol (taux de détection de 8%). L'algorithme basé sur la soustraction du fond permet d'améliorer significativement les résultats obtenus avec l'algorithme basé sur la phase. Les deux algorithmes basés sur la soustraction du fond (non optimisé ou optimisé) permettent d'obtenir un taux de non détections fausses inférieur à celui obtenu avec l'algorithme de référence, le taux de non détections vraies est comparable. Par contre nous

obtenons davantage de fausses alarmes.

Les résultats obtenus sur la mine cylindrique (troisième ligne du tableau 8.7) avec l'algorithme basé sur la phase et avec l'algorithme basé sur la soustraction du fond optimisé donnent sensiblement les mêmes résultats. Les taux de détection obtenus avec ces deux algorithmes sont les plus faibles. L'algorithme basé sur la soustraction du fond (non optimisé) donne les meilleurs taux de détection et de non détections fausses. Le meilleur taux de fausses alarmes est obtenu avec l'algorithme de référence. Il en est de même pour le taux de non détections vraies.

L'optimisation de l'algorithme basé sur la soustraction du fond a été développée à partir des vidéos de sphères (quatrième ligne du tableau 8.7). Par conséquent nous obtenons de très bons résultats en détection comparativement aux autres algorithmes. Les autres taux ont aussi été améliorés. Cet algorithme permet d'obtenir de meilleurs résultats que l'algorithme de référence, sur tous les taux calculés.

Les autres objets (cinquième ligne du tableau 8.7) sont détectés pareillement par l'algorithme basé sur la phase et par l'algorithme de référence. La soustraction du fond apporte une amélioration sensible. Le taux de fausses alarmes le plus faible est obtenu avec l'algorithme de référence, de même que le meilleur taux de non détections vraies. Le taux le plus faible de non détections fausses est obtenu avec l'algorithme basé sur la phase.

Les algorithmes basés sur la soustraction du fond permettent d'obtenir les meilleurs résultats sur les vidéos vides (dernière ligne du tableau 8.7). Les résultats obtenus avec l'algorithme basé sur la phase ou l'algorithme de référence sont comparables.

Comme nous le disions dans la partie 8.2, les résultats de l'algorithme basé sur la phase (résultats en noir dans le tableau 8.7) montrent que notre algorithme n'est pas efficace sur des images peu contrastées. Comparativement aux résultats obtenus avec l'algorithme de référence (résultats en rouge dans le tableau 8.7), seul le taux de non détections fausses est meilleur.

L'algorithme basé sur la soustraction du fond (résultats en bleu dans le tableau 8.7) permet d'améliorer les bonnes détections par rapport à l'algorithme développé par référence (résultats en rouge dans le tableau 8.7). Le taux de non détection fausses est inférieur (sauf dans le cas des mines sphériques où il est comparable). Les taux de fausses alarmes et de non détections vraies sont moins bons, sauf dans le cas des vidéos vides. Cela s'explique par les multiples détections autorisées par l'algorithme. L'algorithme optimisé permet d'améliorer toutes les performances de l'algorithme de référence, sur les vidéos de mines sphériques et sur les vidéos vides. Les résultats graphiques sont présentés dans l'annexe C.

8.5 Perspectives

L'analyse des différents résultats obtenus montre le besoin d'optimisations supplémentaires pour améliorer la détection. Tout d'abord nous présentons des réflexions pour améliorer encore les résultats par rapport aux résultats présentés précédemment dans le tableau 8.7. Ensuite nous présenterons des méthodes pour formaliser notre algorithme.

8.5.1 Optimisations à apporter à nos algorithmes

Les résultats obtenus avec les deux algorithmes développés (algorithme basé sur la phase et algorithme basé sur la soustraction du fond, optimisé ou non) sur les deux jeux de vidéos nous montrent que l'algorithme basé sur la phase est plus adapté aux images Topvision tandis que l'algorithme basé sur la soustraction du fond est adapté aux images VAMA. Une première optimisation serait de définir un critère pour sélectionner la méthode la plus adaptée selon les images à traiter. Dans l'optique où notre algorithme sera un jour intégré à un drone, l'idéal serait que ce critère décide de la méthode à employer dès les premières images acquises par la vidéo.

Une piste de réflexion nous mène vers un critère de contraste puisque les objets présents dans les images Topvision ressortent par rapport au sol tandis que les objets présents dans les images VAMA sont plus "fondus" dans le fond. Cependant, les premières images ne contenant pas d'objet, ce critère peut ne pas être efficace.

Les optimisations apportées à l'algorithme basé sur la soustraction du fond doivent être améliorées. En effet, notre algorithme basé sur la soustraction du fond (*cf.* partie 8.3.1) a été développé à partir des vidéos de sphères. Le fond enregistré en début de vidéo correspond au volume d'eau. Ce fond change très peu au cours de la vidéo. Par contre, lors de l'étude des objets posés au sol, le fond enregistré correspond aussi au volume d'eau tandis que lorsque l'objet est présent dans l'image le fond correspond au sol et plus au volume d'eau. Dans les optimisations proposées (*cf.* partie 8.3.2), nous ajoutons l'image filtrée par un filtre passe bas à la moyenne du fond enregistré en début de vidéo. Cette optimisation ne compense pas complètement le changement et produit des fausses alarmes. Il faudrait donc étudier ce problème. Sa résolution entrainerait à notre avis une baisse des fausses alarmes.

8.5.2 Des méthodes pour formaliser nos algorithmes

Les améliorations apportées à nos algorithmes de détection sont simples et basées sur l'interprétation qu'un opérateur peut faire des résultats. Les images sont traitées image par image. Utiliser l'aspect temporel de la vidéo et par conséquent la succession des images, peut permettre d'améliorer les résultats. Relier les mouvements du drone à la position de la mine dans l'image peut aussi permettre d'apporter de nouvelles contraintes efficaces sur les résultats de la détection. Il existe des méthodes plus "mathématiques" qui pourraient prendre en compte les améliorations que l'on a apportées. Nous avons commencé à regarder les méthodes qui existent. Nous avons étudié le principe du track-before-detect [121, 122] et des chaînes de Markov [123].

8.5.2.1 Le track-before-detect

La première méthode que nous avons étudiée est le "*track-before-detect*". Cela consiste à combiner les étapes de tracking et de détection. Le tracking est effectué sur quelques images avant de déclarer une cible détectée [121]. Il faut dans un premier temps prédire le déplacement de la cible à l'aide d'un modèle et dans un deuxième temps détecter et mesurer son emplacement. Pour cela plusieurs algorithmes ont été utilisés : l'estimation bayésienne, la programmation dynamique, le filtrage particulaire [122].

Nous avons décidé d'utiliser la méthode la plus simple dans un premier temps : le filtre de Kalman [124]. Nous nous sommes entraînés sur un problème factice : nous avons créé une trajectoire de points dont certains sont aberrants. Nous avons essayé de suivre les points de cette trajectoire. Puis nous avons ajouté la gestion de plusieurs pistes : si un point est loin de la trajectoire en cours, on initialise une nouvelle piste ; mise à jour des différentes pistes ; lorsqu'une piste n'a pas été utilisée depuis longtemps, on la supprime de pistes étudiées.

Notre idée avant d'essayer d'adapter cet algorithme à nos résultats est de prédire la position de la mine et de faire coïncider cette prédiction avec la position mesurée par l'algorithme de détection. Le résultat n'est pas probant : la probabilité de détection est de 46% alors qu'on avait 48.23% et la probabilité de fausses alarmes est de 40.34% alors qu'on avait 37.65%. Ces résultats s'expliquent : il y a parfois plusieurs détections dans l'image. Le suivi d'une seule piste place l'objet au centre de la boîte englobant toutes les détections. Cette position n'est donc pas cohérente avec les annotations effectuées sur les images. L'ajout du suivi de plusieurs pistes peut améliorer les résultats obtenus. Ayant une solution qui fonctionne correctement nous avons décidé de ne pas passer trop de temps sur cette méthode.

8.5.2.2 Les chaînes de Markov

Pour formaliser les considérations et contraintes que nous avons apportés à nos programmes initiaux, il est possible d'utiliser les chaînes de Markov [125]. En effet, les chaînes de Markov représentent un système à plusieurs états et les conditions pour passer d'un état à l'autre. Le système est décrit par une hypothèse sur l'état initial, x^0 . Une matrice de transition P nous donne la probabilité du système de se trouver dans l'état N au temps n : $x^n = x^0 P^n$, où x^n représente le système au temps n . Représentons un système à 3 états $N1, N2, N3$ (*cf.* figure 8.19).

$$x^0 = [1 \ 0 \ 0]$$

$$P = \begin{pmatrix} N1 \rightarrow N1 & N1 \rightarrow N2 & N1 \rightarrow N3 \\ N2 \rightarrow N1 & N2 \rightarrow N2 & N2 \rightarrow N3 \\ N3 \rightarrow N1 & N3 \rightarrow N2 & N3 \rightarrow N3 \end{pmatrix} \quad (8.3)$$

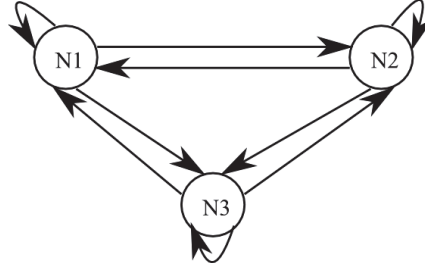


FIGURE 8.19 – Etats d’une chaîne de Markov

Les chaînes de Markov sont utilisées lorsque l’on connaît les états, les probabilités de transition entre états et que les observations correspondent aux états. Dans un système comme notre détection d’objets sous marins, les observations effectuées ne correspondent pas forcément aux états (par exemple les fausses alarmes ou la non détection fausse). Pour ces problèmes il existe les chaînes de Markov cachées [123]. Nous avons comme dans le cas des chaînes de Markov des états, une matrice de transition entre états. Nous avons aussi une matrice de génération des symboles ainsi que la séquence d’observation. Représentons un système à 2 états (*cf.* figure 8.20). Lorsqu’une séquence d’observation est reçue, pour retrouver le système initial, on estime les différents paramètres du système. Des algorithmes comme l’algorithme de Viterbi ou de Baum-Welch permettent itérativement de retrouver les paramètres du système [123].

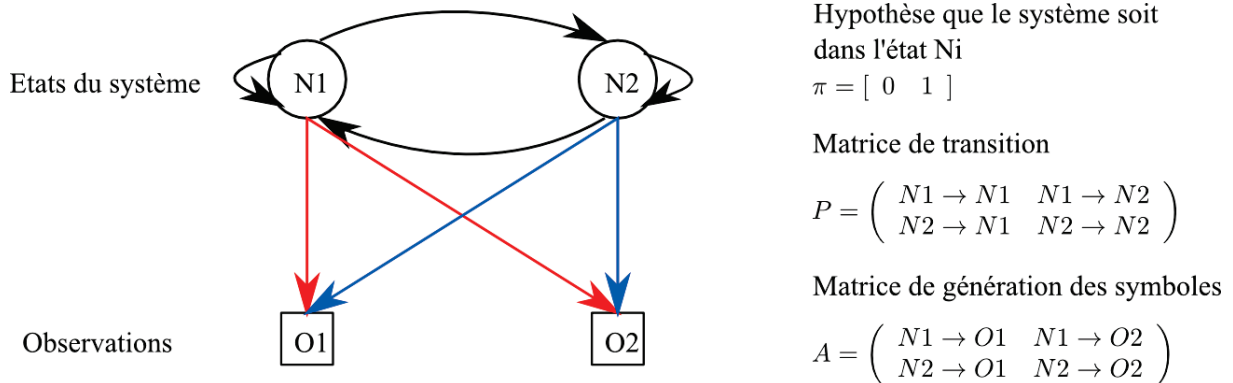


FIGURE 8.20 – Etats et observation d’une chaîne de Markov cachée

Nous nous sommes dans un premier temps exercés sur un problème synthétique : nous avons créé un système à partir duquel nous avons créé une séquence d’observations. Puis nous avons implémenté l’algorithme de Baum-Welch pour retrouver le système initial. Une fois l’algorithme maîtrisé, nous nous sommes intéressés à la mise en œuvre sur une séquence de résultats. Nos conclusions sont qu’il faut utiliser un système à quatre états (détection, fausse alarme, non détection fausse et non détection vraie) et deux observations (la détection et la fausse alarme génèrent la présence d’une mine tandis que la non détection fausse et la non détection vraie génèrent l’absence de mines). Les résultats obtenus ne sont pas tout à fait ceux attendus. En effet, le système converge vers une détection dans toutes les images. Nous en avons donc conclu qu’il ne faudrait utiliser les chaînes de Markov cachées uniquement lorsque l’on est sûr de la présence d’une mine. De plus, nous avons estimé manuellement les paramètres du système. Il s’agit par la suite d’intégrer dans ces paramètres les contraintes que nous avons appliquées grâce aux informations de navigation.

8.5.2.3 Conclusion

Nous avons commencé l'étude de deux méthodes pour créer des versions optimisées robustes de nos algorithmes en utilisant des contraintes "mathématiques". Les études menées n'ont pas permis d'aboutir à une solution fiable. Cependant, nous avons présenté dans cette partie quelques pistes de poursuite de ces travaux. De plus nous sommes confrontés à la traduction "mathématiques" des contraintes que nous avons fixées. Apporter une réponse à cette question permettra d'améliorer l'utilisation des deux algorithmes présentés.

8.6 Conclusion

Nous avons développé deux algorithmes de détection sur les images à notre disposition. Les résultats fournis par ces algorithmes ont été améliorés avec des considérations "*simples*" basées sur les informations de navigation. Cette introduction des informations de navigation pour améliorer la détection des objets au cours de la vidéo est une méthode nouvelle qui a montré son efficacité sur nos résultats, à la fois en terme de taux de détection et en terme de temps de calcul (les améliorations basées uniquement sur les informations de navigation ne demandent pas plus de temps de calcul, contrairement aux améliorations nécessitant des calculs sur les images). De plus la comparaison avec un autre algorithme montre l'avancée apportée par nos méthodes. Des réflexions ont ensuite été menées dans la perspective d'apporter davantage de robustesse à notre programme par la formalisation des améliorations et des contraintes que nous avons ajoutées.

Chapitre 9

Classification et identification des objets

Sommaire

9.1 Premiers tests	132
9.2 Les filtres de corrélation	132
9.2.1 Base d'images de référence	133
9.2.2 Fusion des images	134
9.3 Corrélation avec tous les filtres	138
9.3.1 Algorithme utilisé	138
9.3.2 Résultats obtenus	139
9.3.3 Optimisations apportées à notre algorithme	140
9.3.4 Distinction entre les objets manufacturés dangereux et les autres objets	141
9.3.5 Conclusion sur cette première méthode	141
9.4 Utilisation des données de navigation	141
9.4.1 Algorithme utilisé	141
9.4.2 Résultats obtenus	143
9.5 Conclusion	144

Après l'étape de détection, nous obtenons un ou plusieurs contacts (cibles) à classifier et identifier. Nous appellerons par la suite cette étape "reconnaissance". En effet, nous voulons effectuer la classification (différenciation entre objet manufacturé et objet naturel) et l'identification (différenciation des mines des autres objets) en une seule étape. Pour se faire, nous avons utilisé la méthode de corrélation comme cela est présenté dans la partie II. La figure 9.1, qui est une sous partie de la figure 1.17, montre où nous en sommes dans la chaîne de traitement.

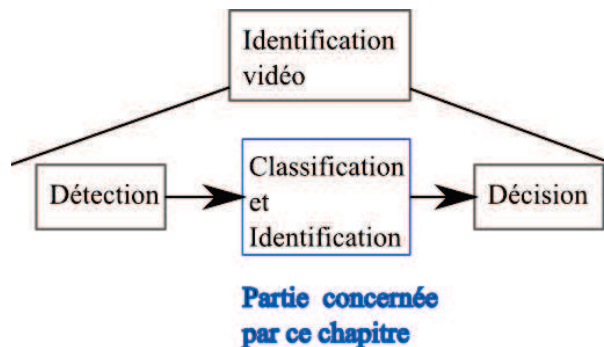


FIGURE 9.1 – Schéma indiquant la partie traitée dans ce chapitre

Dans un premier temps, nous avons travaillé sur les images Topvision. Nous avons créé des filtres de corrélation à partir d'images réelles issues de la vidéo Topvision003. Cette méthode présente des inconvénients que nous avons voulu éviter en utilisant des images de synthèse des

mines (*cf.* chapitre 7 et partie 9.1). Nous présenterons dans ce chapitre les bases d'images de référence de synthèse ainsi créées, la fusion de ces images et l'application aux vidéos Topvision et VAMA via deux méthodes.

9.1 Premiers tests

Comme présenté dans le chapitre 7, nous avons tout d'abord créé des filtres de corrélation à partir d'images issues de la vidéo Topvision003. Ces images ont été prétraitées selon la chaîne adaptée aux images Topvision (*cf.* section 2.3.1). Ensuite nous avons créé cinq filtres POF dans le but de reconnaître la mine dans une zone précise de la vidéo (entre les images 650 et 875). Cette méthode présente des résultats corrects dans la zone pour laquelle les filtres sont définis. Mesurés sur l'ensemble de la vidéo Topvision003, les résultats sont moyens (*cf.* figure 7.3, page 105). Avec un seuil de 0.01 sur les valeurs du PCE, nous obtenons :

- une probabilité de détection de 47.96%,
- une probabilité de non détection fautive de 52.04%,
- une probabilité de fausse alarme de 0.1%
- et une probabilité de non détection vraie de 99.9%.

La probabilité de non détection fautive élevée montre que cette méthode n'est pas très robuste pour cette application. Cependant la probabilité de fausses alarmes est très faible ce qui montre que cet algorithme est discriminant.

Avec le même seuil, le cylindre est reconnu comme étant une mine Manta à 7.02%. Le temps de calcul sur le PC ISEN (*cf.* annexe E pour les caractéristiques), est de 0.7s par image soit une fréquence de calcul de 1.4 Hz.

Cette méthode présente aussi des inconvénients importants. Tout d'abord, les images de référence sont issues des vidéos réelles. Nous ne pouvons donc obtenir que quelques points de vue des mines et quelques échelles. Or si notre système est intégré à un drone sous marin, il est possible de rencontrer des mines qui seront vues sous d'autres points de vue et échelles que ceux présents dans les vidéos Topvision. Dans ce cas, nous ne serons probablement pas capable de reconnaître ces mines. Le deuxième inconvénient majeur de cette méthode est la nécessité de prétraiter les images de référence. Comme les images de référence sont issues de vidéos réelles, elles sont bruitées et nécessitent des prétraitements. Or il est possible que les prétraitements n'enlèvent pas tout le bruit et que la mine ne soit pas complètement contenue dans l'image. Pour éviter ces inconvénients, nous avons voulu utiliser des images 3D de synthèse des mines.

9.2 Les filtres de corrélation

Nous avons choisi, dans le chapitre sur le contexte de cette thèse (*cf.* chapitre 1), de détecter et reconnaître les mines sphériques, cylindriques, Manta et Rockan. Ces mines sont les plus standards et leurs dimensions sont connues. Nous avons donc créé des modèles 3D de synthèse correspondant à ces mines (*cf.* figure 9.2). Nous avons utilisé Matlab pour obtenir les différents points de vue de ces mines à différentes échelles.

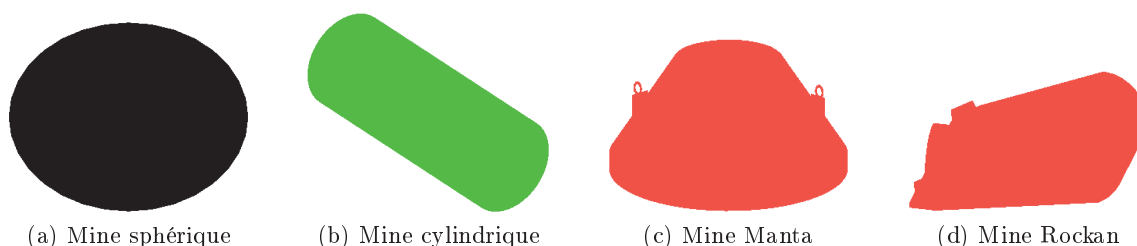


FIGURE 9.2 – Images des modèles de synthèse des mines

9.2.1 Base d'images de référence

Avant d'enregistrer les images de référence qui seront utilisées pour fabriquer les filtres de corrélation, il faut effectuer des choix : comment appliquer la rotation aux modèles pour savoir retrouver l'angle et l'axe de rotation facilement, quels détails sont nécessaires dans les images de référence selon les mines considérées pour avoir une corrélation fiable. Il faut aussi prendre en compte l'effet de la caméra sur les images.

9.2.1.1 Les différentes définitions de la rotation

Une rotation de l'objet peut être définie par une matrice de rotation, par les angles d'Euler, par des vecteurs ou par des quaternions [126]. Nous avons choisi d'utiliser les quaternions. Ils sont équivalents à la représentation vectorielle via la formule :

$$q = \cos\left(\frac{\theta}{2}\right) + \sin\left(\frac{\theta}{2}\right)(in_x + jn_y + kn_z) \quad (9.1)$$

où q est le quaternion, n le vecteur correspondant à l'axe de rotation et θ l'angle de rotation. Pour plus de visibilité, nous avons défini les axes de rotation et fait varier les angles de rotation de -180° à $+180^\circ$. Ensuite nous avons calculé les quaternions correspondants de façon à appliquer la rotation voulue à la mine. Enfin l'image de référence de synthèse est enregistrée.

9.2.1.2 Niveau de détails choisi pour chaque mine

Il est possible avec Matlab d'ajouter à nos modèles de synthèse de la lumière ainsi que de la couleur pour distinguer les différentes surfaces par exemple.

La mine sphérique ne nécessite ni couleur ni lumière car elle ne possède pas de relief ni de sommets à faire ressortir. De plus cette mine possède des symétries qui permettent de n'enregistrer qu'une image par échelle.

Pour la mine cylindrique, nous avons choisi d'utiliser des modèles avec 2 couleurs. Cela permet de distinguer les faces circulaires du corps du cylindre (*cf.* figure 9.3a). L'ajout de lumière n'est pas nécessaire car les modèles ne contiennent pas de détails à faire ressortir. La mine cylindrique possède un axe de symétrie qui permet pour certains axes de rotation de n'enregistrer qu'une image.

Pour la mine Manta, nous avons choisi d'utiliser un modèle bicolore. Ensuite nous avons créé deux bases de références : une avec et une sans lumière (*cf.* figures 9.3b et 9.3c). En effet, la mine Manta possède des crochets sur les côtés qui permettent d'accrocher la mine pour la déposer sur le sol. La présence de lumière permet de faire ressortir ces crochets. Cependant ils ne sont pas visibles sur toutes les vidéos. Il faudra déterminer si la présence de ce détail est importante ou non dans la reconnaissance de la mine. Dans un premier temps, nous utiliserons les images sans lumière.

Pour la mine Rockan, nous avons choisi d'utiliser un modèle bicolore et avec de la lumière (*cf.* figure 9.3d). La lumière permet de distinguer le relief de la mine qui sera détecté dans les vidéos. La mine Rockan ne possède pas d'axe de symétrie. Il faut donc enregistrer toutes les images calculées.

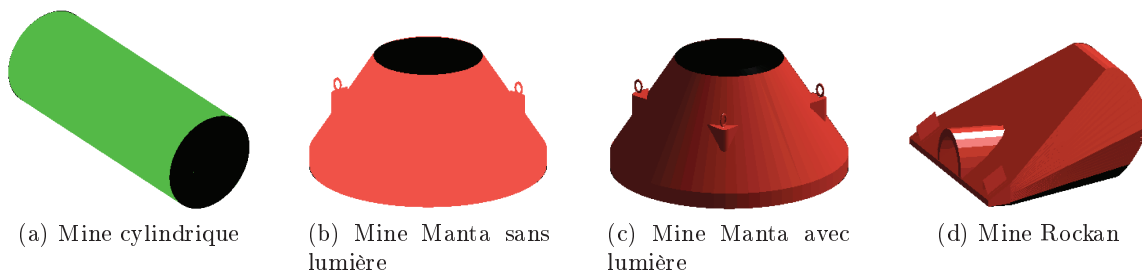


FIGURE 9.3 – Images des modèles de synthèse des mines

9.2.1.3 Influence des caractéristiques de la caméra sur les images de synthèse

Lors de l'enregistrement d'une scène, un point dans l'espace 3D correspond à un point 2D dans l'espace image. La caméra ou l'appareil photo *Cam* est caractérisé par sa distance focale f et son décalage par rapport à l'axe optique c [126]. Le décalage peut être différent selon l'axe x et l'axe y . La distance focale correspond plus précisément au produit kf où k représente le facteur entre la taille du capteur et la taille des pellicules photos de 24 mm x 36 mm.

$$Cam = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \quad (9.2)$$

Les caméras sont aussi définies par leur angle de champ. On relie l'angle de champ à la distance focale (ou au produit kf , appelé dans la formule F) par la formule :

$$\begin{aligned} angle &= 2 * \arctan\left(\frac{L}{2F}\right) \\ F &= \frac{L}{2 * \tan\left(\frac{angle}{2}\right)} \end{aligned} \quad (9.3)$$

où L représente les dimensions du capteur.

9.2.1.4 Enregistrement des images

Pour enregistrer une image sous Matlab, un système d'acquisition des images est mis en place : une caméra et ses caractéristiques sont modélisées et leurs effets appliqués aux images enregistrées. Nous avons appliqué les choix effectués pour notre caméra à la caméra par défaut de Matlab puis nous avons enregistré toutes les images nécessaires. Nous avons enregistré pour chaque mine une matrice d'informations qui nous fournit pour chacune des images de référence l'axe et l'angle de rotation. Nous obtenons un nombre d'images conséquent :

- pour la mine sphérique, nous avons enregistré une seule image par échelle puisque la mine est symétrique. Nous avons utilisé 20 échelles. Nous obtenons donc 20 images de référence de synthèse.
- pour la mine cylindrique, nous avons enregistré 8150 images correspondant à une distance de 1 m à 5.5 m par pas de 0.5 m. Nous avons tenu compte de l'axe de symétrie qui nous permet d'enregistrer une seule image pour chaque face circulaire (petit côté du cylindre).
- pour la mine Manta, nous avons enregistré 9620 images correspondant à une distance de 1 m à 5.5 m par pas de 0.5 m.
- pour la mine Rockan, nous avons enregistré 9620 images correspondant à une distance de 1 m à 5.5 m par pas de 0.5 m.

Nous obtenons ainsi un total de 27410 images. Si chaque image de référence est corrélée à l'image cible, il faudra effectuer 27410 opérations de corrélation pour identifier les objets contenus dans chacune des images cibles. Pour réduire ce nombre d'opérations à effectuer, nous avons décidé d'utiliser des filtres composites. La question qui s'est alors posée est la méthode à utiliser pour assembler ces images. Nous apporterons notre point de vue sur cette question dans la partie suivante.

9.2.2 Fusion des images

Notre étude de la corrélation dans la partie II nous a mené au choix, dans un premier temps, de l'architecture Vanderlugt et des filtres POF. Pour cela, nous avons transformé les images de référence de synthèse en images de contours. Pour fusionner les images et créer nos filtres composites de corrélation, nous devons trouver une méthode de fusion efficace. Le temps de calcul n'est pas une contrainte puisque cette étape se fait en dehors de l'exécution du programme et les filtres de corrélation sont mis en mémoire. La qualité des images de référence fusionnées pour former un filtre est importante pour assurer la qualité de la décision. Nous développons dans cette partie trois méthodes pour fusionner les images.

9.2.2.1 Première approche utilisée

La première méthode utilisée pour fusionner les images est la comparaison des images elles-mêmes par un critère empirique sur le nombre de pixels constituant le contour de l'objet contenu dans l'image de référence de synthèse.

Nous sommes partis du principe selon lequel, le nombre de pixels formant le contour évolue selon l'orientation de la mine : une mine inclinée de 45° n'aura pas le même nombre de pixel qu'une mine verticale ou qu'une mine qui est 2 fois plus petite, par exemple (*cf.* figure 9.4).

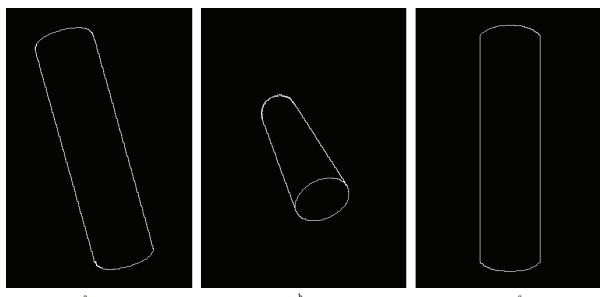


FIGURE 9.4 – Images de contours d'une mine cylindrique. Le contour de l'image a est composé de 1212 pixels, celui de l'image b de 885 pixels et celui de l'image c de 1022 pixels.

Pour créer les différents filtres, nous avons pris la première image de la base à partir de laquelle nous avons créé l'image des contours correspondante. Nous avons alors fait une boucle sur toutes les autres images de la base, transformées en images de contours. Toute image contenant le même nombre de pixels que l'image initiale à $\pm 10\%$ est ajoutée au filtre. Les filtres sont des filtres POF composites. Cela n'est pas suffisant : nous avons obtenu des filtres contenant des mines orientées presque à 90° les unes des autres (*cf.* figure 9.5).

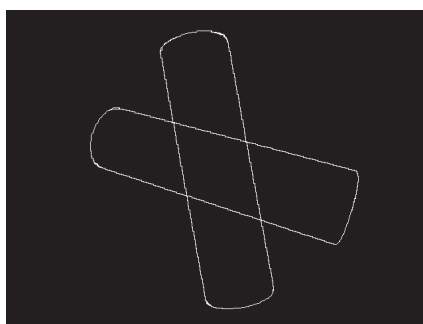


FIGURE 9.5 – Exemple de filtre non satisfaisant

Nous avons alors introduit un critère de distance basé sur la position du pixel le plus en haut à gauche et du pixel le plus en haut à droite de chacune des deux mines. Si ces 2 distances sont supérieures à 10 pixels, les 2 motifs sont considérés comme étant trop différents l'un de l'autre. Cette approche, testée sur un grand nombre de motifs n'est pas très convaincante. En effet, nous obtenons toujours des motifs différents dans les mêmes filtres.

9.2.2.2 Deuxième approche

La première approche n'est, d'après les résultats, pas optimale. Nous nous sommes orientés vers une approche basée sur la corrélation.

Nos images sont enregistrées selon la distance mine - caméra, par pas de 0.5 m. Les images correspondant à la même distance constituent une sous-base. Nous obtenons dix sous-bases pour les mines cylindriques, Manta et Rockan. Nous travaillons ensuite par échelle.

La première image de la sous-base est transformée en filtre POF. Ensuite nous corrélonns ce filtre avec toutes les images de la sous-base. Pour chaque corrélation, nous calculons le PCE du plan de corrélation obtenu. Quand le PCE est supérieur à un seuil (fixé empiriquement),

l'image que nous avons fait corrélérer est ajoutée au filtre composite constitué de l'image filtre et de toutes les images dont le PCE obtenu a été supérieur au seuil. Ensuite nous parcourons toute la sous-base pour créer nos filtres. Cet ensemble d'opérations est effectué pour toutes les sous-bases.

Cette méthode demande un temps de calcul important du fait de toutes ces corrélations, en comparaison au temps de calcul nécessaire pour la première méthode. Cependant, selon la valeur du seuil du PCE pour fusionner deux images, nous sommes capables d'obtenir uniquement des filtres contenant la même orientation de la mine. Plus la valeur du seuil est proche de 1 plus nous aurons de filtres. Plus la valeur du seuil est faible, moins nous aurons de filtres mais nous aurons davantage de mines différentes dans les filtres. Il faut donc trouver un compromis. Nous avons fixé empiriquement le seuil à 0.25. Nous obtenons un nombre important de filtres (*cf.* tableau 9.1), notamment pour la mine Rockan dont les vues diffèrent beaucoup selon les points de vue puisqu'elle ne possède pas de symétrie. Le nombre d'opérations à effectuer est divisé par 3 ce qui est un gain important.

Mine	Nombre d'images	Nombre de filtres
Sphère	20	20
Cylindre	8150	2884
Manta	9620	1234
Rockan	9620	5447
Total	27410	9585

Tableau 9.1 – Nombre de filtres obtenus avec la méthode d'assemblage basée sur la corrélation

En examinant les filtres obtenus, nous avons remarqué que le nombre de filtres n'est pas le même pour toutes les échelles. De plus, plus la distance entre la mine et la caméra augmente plus les images contenues dans un même filtre sont différentes. Il serait donc judicieux d'utiliser plusieurs seuils sur le PCE. Nous avons effectué un test qui nous permet d'obtenir 3845 filtres de mine Manta parmi lesquels il y a peu de filtres non satisfaisants. Il faudrait effectuer davantage de tests pour optimiser les seuils et par conséquent le nombre de filtres. Pour l'instant cette méthode n'a pas été utilisée.

9.2.2.3 Troisième approche

Nous avons ensuite pensé qu'il serait intéressant de regrouper les images par échelle et par point de vue. Cela permettra de connaître rapidement l'orientation de la mine qui a corrélé lors de l'étape de reconnaissance. Notre base de mines est classée selon les axes et les angles de rotation. Pour chaque image, nous calculons la position de l'observateur ainsi que l'orientation de la mine.

Pour cela, nous créons un octaèdre dont nous projetons les sommets sur une sphère, qui modélise l'espace (*cf.* figure 9.6). Pour chaque image, nous regardons quelle est la face correspondante de l'octaèdre. Nous divisons ensuite chaque face de l'octaèdre en 4 triangles. Les sommets sont projetés sur la sphère et la face correspondante à chaque image est mise à jour. Une récurrence est effectuée de façon à obtenir des triangles équivalents à 5° . Pour calculer les faces de l'octaèdre correspondant aux points de vue d'une mine, nous avons eu besoin de calculer l'intersection d'un rayon et d'un triangle. La méthode de calcul utilisée est développée dans l'annexe D.

Cette étape peut être modifiée pour tenir compte des angles de vue sous lesquels l'aspect de la mine ne varie pas beaucoup. Avant de diviser une face de l'octaèdre en 4 triangles, nous faisons corrélérer les images entre elles de façon à arrêter la récurrence sur cette face en cas de mines d'aspects proches.

Lors de l'assemblage des mines entre elles pour créer les filtres de corrélation, il faut tenir compte de l'orientation de la mine ainsi que de la distance de la mine. Cela permet d'obtenir lors de la corrélation les informations d'orientation et de distance.

Cette approche nous permet de réduire la quantité d'images de référence à faire corrélérer avec

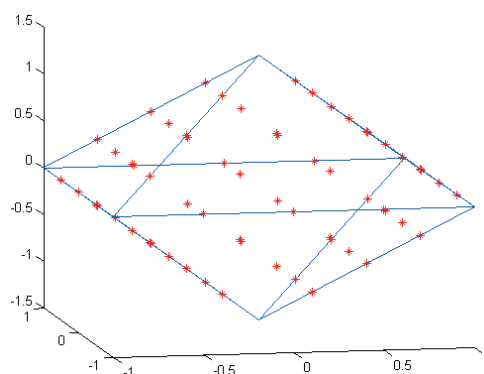


FIGURE 9.6 – Représentation d’un octaèdre après 2 étapes de récurrence, représentant l’espace

une image cible. Le résumé du nombre d’images et de filtres est donné dans le tableau 9.2. La mine sphérique est représentée par une image pour chaque échelle. Par conséquent nous obtenons autant de filtres que d’images. Les mines cylindriques et Manta possèdent des axes de symétrie ce qui permet d’associer plusieurs images. La vérification de l’aspect des mines nous fait obtenir un nombre de filtres précis, 1899 dans le cas de la mine cylindrique et 2440 dans le cas de la mine sphérique. Hormis les faces pour lesquelles les axes de symétrie permettent d’assembler les images, les aspects sont vus comme différents lors de la vérification qui arrête la récurrence. Pour la mine Rockan, nous obtenons 1800 filtres. Comparativement aux mines cylindrique et Manta, c’est peu surtout que cette mine ne présente pas d’axes de symétrie. Ce faible nombre est dû aux formes rectilignes obtenues lors de la transformation de l’image de synthèse en image de contours. Ces formes corrélerent entre elles lors de la vérification de l’aspect. L’aspect est donc considéré comme proche et les mines assemblées.

Sur la figure 9.7 sont représentés les positions de l’observateur avant et après assemblage des images de référence (symboles violets). Avant assemblage des images de référence, l’espace était très décrit par l’observateur. La position de l’observateur correspondant à un filtre est la position de l’observateur de la première image du filtre. Malgré la forte réduction du nombre de positions, l’espace est toujours bien représenté par les filtres de corrélation : il n’y a pas de zones qui ne sont pas représentées par un filtre du fait d’une fusion qui aurait regroupé trop d’images ensemble.

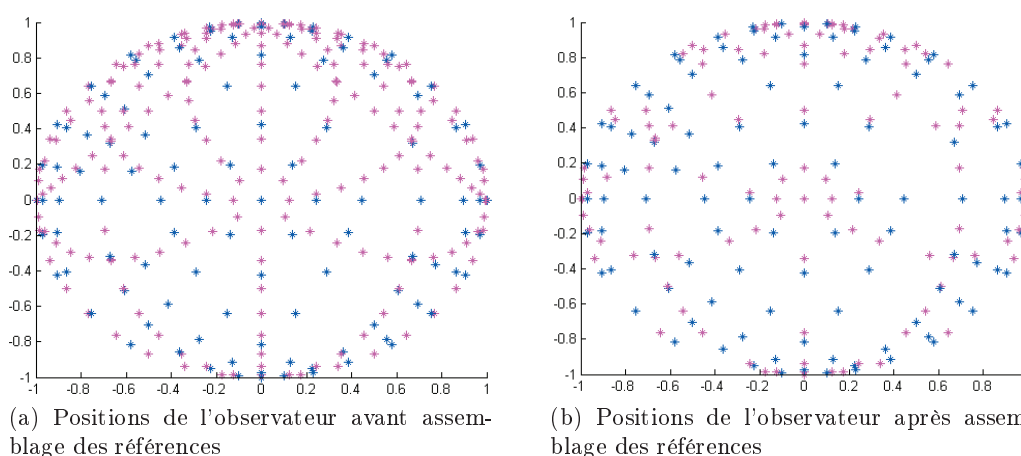


FIGURE 9.7 – Représentation d’un octaèdre et de ses sommets projetés sur une sphère représentant l’espace

Cette méthode, qui regroupe les filtres selon la position de l’observateur et donc l’orientation de la mine présente à la fois des filtres proches de ceux obtenus dans la deuxième approche, présentée précédemment mais aussi des filtres contenant des mines très différentes. Cela est dû aux images correspondant à un sommet de l’octaèdre. Les images sont attribuées à une face de

Mine	Nombre d'images	Nombre de filtres
Sphérique	20	20
Cylindrique	8150	1899
Manta	9620	2440
Rockan	9620	1800
Total	27410	6159

Tableau 9.2 – Tableau résumant le nombre d'images et de filtres obtenus

l'octaèdre alors qu'il faudrait créer un filtre à part.

Nous avons présenté et testé plusieurs méthodes d'assemblage qui présentent leurs avantages et inconvénients. Nous avons dans un premier temps utilisé les filtres obtenus avec la troisième méthode. Une fois les filtres créés, nous pouvons nous intéresser à la partie reconnaissance de l'algorithme qui consiste à comparer l'image cible à ces filtres par une opération de corrélation.

9.3 Corrélation avec tous les filtres

Tout d'abord nous avons appliqué nos filtres de corrélation composites POF aux images Topvision. Pour cela nous avons développé l'algorithme présenté ci dessous.

9.3.1 Algorithme utilisé

Lorsqu'un objet est détecté dans une image, nous appliquons l'algorithme de reconnaissance. L'algorithme de reconnaissance utilisé est présenté sur la figure 9.8.

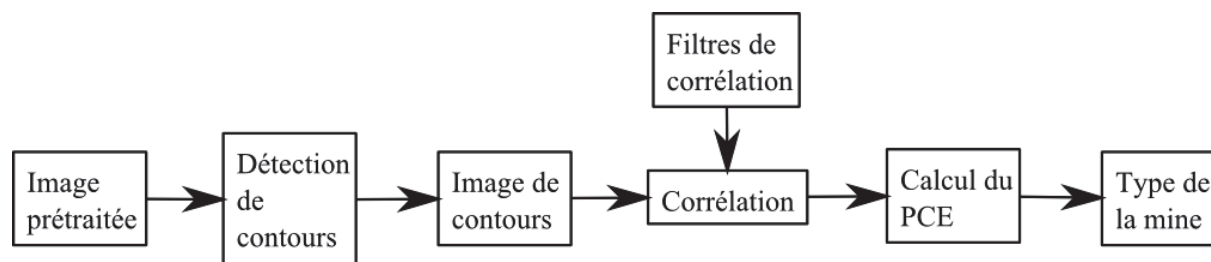


FIGURE 9.8 – Schéma synoptique de l'algorithme utilisé avec la première méthode

Tout d'abord il faut effectuer une étape supplémentaire de prétraitements afin de créer une image de contours qui corrèle avec les filtres POF. Lorsque nous avons développé l'algorithme de détection sur les images Topvision, nous avons choisi d'utiliser, pour les images cibles, les images de contours obtenues à partir de la phase du spectre de l'image. Nous avons utilisé ces images de contours pour la corrélation. Cette méthode donne parfois des résultats non satisfaisants, surtout lorsque l'image de phase de l'image cible donne des contours de faible intensité (ce qui est souvent le cas des images sous marines). Nous avons alors cherché des méthodes de détection de contours plus performantes afin de pallier à ces problèmes. Nous avons trouvé la méthode de segmentation basée sur les contours actifs [127] (*cf.* figure 9.9). L'image est prétraitée de façon à préparer la segmentation par contours actifs. Cette segmentation produit une zone dont nous récupérons les contours. Sur la figure 9.9, nous montrons deux exemples d'images de contours obtenues avec la méthode des contours actifs. En haut, une image qui donne un résultat peu probant et en bas une image qui produit un résultat encourageant. La seconde méthode que nous avons sélectionné est la méthode de détection de contours basée sur l'approche multi-échelle (présentée *cf.* partie 2.2.5). Ces deux méthodes semblent donner des résultats corrects. Nous allons donc comparer les résultats de reconnaissance obtenus avec ces deux méthodes de détection de contours.

Une fois ce prétraitement supplémentaire effectué, nous faisons corréler cette image de contours cible avec tous les filtres de corrélation, comme le montre la figure 9.8. Pour chaque plan de corrél-

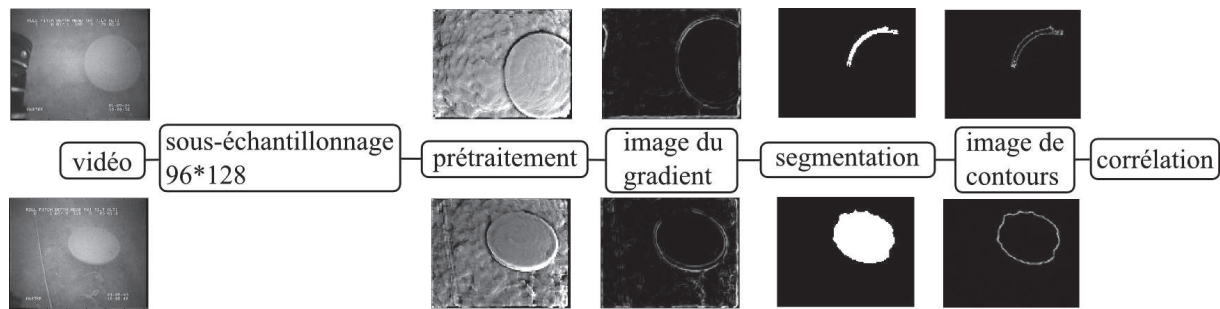


FIGURE 9.9 – Méthode de la détection de contours à partir de la segmentation basée sur les contours actifs

lution obtenu, nous calculons le PCE associé. Ensuite nous identifions le type de la mine comme étant celui du filtre qui a le mieux corrélé. L’algorithme est représenté sur le schéma synoptique de la figure 9.8.

9.3.2 Résultats obtenus

Nous avons comparé les résultats obtenus avec les deux méthodes de détection de contours.

Lors de la détection nous avons défini des probabilités. Nous avons utilisé les mêmes principes avec la reconnaissance :

- lorsqu’une mine est détectée et qu’elle est reconnue par une mine du même type, nous obtenons la probabilité de détection classification P_{dc}
- lorsqu’une mine est détectée et qu’elle est reconnue par une mine d’un autre type, nous obtenons la probabilité type P_{type}
- lorsqu’une mine est détectée au mauvais endroit ou alors qu’il n’y a rien dans l’image, nous obtenons la probabilité de fausses alarmes P_{fac}
- lorsqu’aucune mine n’est pas reconnue alors qu’elle a été détectée, nous obtenons la probabilité de non détection fausse P_{ndfc}
- lorsqu’aucune mine n’est ni détectée ni reconnue et qu’il y a bien une mine, nous obtenons la probabilité de non détection fausse P_{ndfc}
- lorsqu’aucune mine n’est ni détectée ni reconnue et qu’il n’y a rien, nous obtenons la probabilité de non détection vraie P_{ndvc}

Les deux méthodes de détection de contours choisies demandent un temps de calcul important pour des résultats moyens. Le tableau 9.3 présente les résultats obtenus avec la détection à partir des contours actifs. Le tableau 9.4 présente les résultats obtenus avec la détection à partir d’une approche multi-échelle.

Mine	P_{dc}	P_{type}	P_{fac}	P_{ndfc}	P_{ndvc}	Temps de calcul
Mine Manta	25.43%	42.53%	4.83%	21.9%	98.68%	31 s
Mine cylindrique	19.45%	47.16%	17.10%	5.96%	98.90%	33 s
Mine sphérique	17.13%	35.83%	9.38%	24.22%	100%	32 s
Vidéos vide			20.7%		79.30%	33 s

Tableau 9.3 – Résultats obtenus avec la détection de contours basée sur les contours actifs

Les valeurs temporelles ont été calculées à partir du temps total nécessaire à Matlab et au PC ISEN (*cf.* caractéristiques dans l’annexe E) pour traiter les vidéos entièrement et du nombre d’images sur lesquelles on a appliqué l’algorithme de corrélation. Ces valeurs de temps s’expliquent par le fait que nous faisons corrélérer avec les images prétraitées tous les filtres. Leur nombre étant important, le temps de calcul augmente d’autant (environ 30 s par image). De plus les méthodes de détection de contours ne permettant pas d’obtenir tous les contours des mines, il arrive souvent qu’un autre type de mine corrèle mieux que la mine exacte entière. Par la suite,

Mine	Pdc	Ptype	Pfac	Pndfc	Pndvc	Temps de calcul
Mine Manta	21.99%	45.98%	4.83%	21.9%	98.68%	32 s
Mine cylindrique	48.26%	18.35%	17.10%	5.96%	98.9%	29 s
Mine sphérique	2.87%	50.08%	9.38%	24.22%	100%	31 s
Vidéos vide			21.23%		78.77%	31 s

Tableau 9.4 – Résultats obtenus avec la détection de contours basée sur l’approche multi-échelle

nous choisirons la méthode multi-échelle pour la détection de contours car elle est plus simple à mettre en œuvre et plus rapide.

Les taux de correcte identification, Pdc sont faibles (inférieurs à 30% sauf pour la mine cylindrique avec la détection basée sur l’approche multi-échelle). Par conséquent la probabilité type $Ptype$ est élevée puisque nous avons une identification pour chaque image où un objet a été détecté. Le taux de fausses alarmes est plus élevé que dans le cas de la détection seule (cf. tableau 8.1). Cela montre que les corrélations détectent des objets au mauvais endroit de l’image, en plus des fausses détections “héritées” de l’étape de détection. De même les taux de non détections fausses sont un peu plus élevés (cf. tableau 8.1). Ils correspondent au taux de non détections fausses de l’étape de détection augmenté des images dans lesquelles aucun contours n’a été détecté alors qu’un objet est présent dans l’image.

9.3.3 Optimisations apportées à notre algorithme

Comme certaines vues de mines peuvent se confondre (la sphère, les faces circulaires de la mine cylindrique et les faces circulaires de la mine Manta), nous avons ajouté une étape qui nous informe que l’objet présent dans l’image peut être une mine sphérique, une mine Manta ou une mine cylindrique. Nous pensons que la séquence vidéo (approche de la mine) nous permettra de distinguer le type de la mine.

Partant du principe qu’une mine est la même tout au long de la vidéo et surtout qu’elle ne peut pas changer d’une image à l’autre, nous avons décidé d’apporter une amélioration simple à notre algorithme. Nous regardons le type de mine des deux meilleurs PCE pour chaque corrélation. Si le meilleur PCE de l’image i correspond aux meilleurs PCE des images $i - 1$ et $i - 2$ alors le type reste le même. Si les meilleurs PCE ne donnent pas tous le même type de mine, nous regardons les 6 types de mines détectés. Si au moins un type de mine revient sur 2 images alors la mine est de ce type là. Cela correspond “avec les mains” à un algorithme de type algorithme de Viterbi, qui consiste à chercher la meilleure possibilité parmi plusieurs hypothèses retenues à chaque étape [123].

Les taux de reconnaissance présentés dans la partie précédente (cf. tableaux 9.3 et 9.4) étant faibles (inférieurs à 30% sauf dans le cas des mines cylindriques avec l’approche multi-échelle) avec beaucoup d’erreurs sur le type de la mine, les optimisations apportées ne permettent pas d’améliorer les résultats. Les taux de reconnaissance baissent sensiblement pour la mine sphérique, sont comparables pour la mine cylindrique et la mine Manta.

Mine	Pdc	Ptype	Pfac	Pndfc	Pndvc	Temps de calcul
Mine Manta	26.08%	41.89%	4.8%	21.9%	98.68%	31 s
Mine cylindrique	17.67%	48.83%	17.10%	6.07%	98.90%	33 s
Mine sphérique	7.56%	45.17%	9.35%	24.44%	100%	32 s
Vidéos vide			20.7%		79.30%	33 s

Tableau 9.5 – Résultats obtenus avec la détection de contours basée sur les contours actifs

Mine	Pdc	Ptype	Pfac	Pndfc	Pndvc	Temps de calcul
Mine Manta	17.75%	50.22%	4.83%	21.9%	98.68%	32 s
Mine cylindrique	52.68%	13.93%	17.10%	5.96%	98.9%	29 s
Mine sphérique	0.1%	52.81%	9.38%	24.22%	100%	31 s
Vidéos vide			21.23%		78.77%	31 s

Tableau 9.6 – Résultats obtenus avec la détection de contours basée sur l’approche multi-échelle

9.3.4 Distinction entre les objets manufacturés dangereux et les autres objets

Détecter et reconnaître des objets manufacturés est une première étape. Ensuite, il faut distinguer les objets dangereux des objets non dangereux (*a priori* mais nous ne sommes pas à l’abri d’un objet transformé en explosif). Les vidéos vides de Topvision contiennent une vidéo vide et une vidéo contenant des bouts. Il y a très peu de détection et donc de reconnaissance dans ces deux vidéos. Nous pensions donc initialement seuiliser le PCE pour distinguer les deux types d’objets manufacturés. Les vidéos VAMA mettent en scène des objets manufacturés comme une poubelle ou un tonneau, de formes proches de celles des mines recherchées. Les tests montrent que ce raisonnement n’est pas valable.

9.3.5 Conclusion sur cette première méthode

Cette méthode donne des résultats peu encourageants. Le temps de calcul est important, de l’ordre de 30 s par image et les résultats peu satisfaisants. Le taux de reconnaissance est faible. Cela s’explique par le nombre de filtres que l’on fait corréliser avec une image cible. De plus, ce taux vient aussi des mines qui ne sont pas entièrement visibles dans les images ou dont les contours ne sont pas correctement et entièrement détectés.

Nous cherchons une méthode dont la fréquence de calcul approche 2 Hz soit 0.5 s par image. L’algorithme développé nécessite environ 30 s par image sur le PC ISEN (*cf.* caractéristiques dans l’annexe E). Cela s’explique par la corrélation de tous les filtres avec chaque image cible.

Les résultats obtenus sont peu satisfaisants. La probabilité de bonne reconnaissance d’une mine est de l’ordre de 20 %. Cela s’explique par la détection des contours qui ne donne pas toujours de bons résultats. Cela est dû à la qualité de l’image, aux mines qui ne sont pas complètement dans l’image. Comparativement aux filtres qui contiennent les mines entièrement et dont la détection de contours a donné des formes très géométriques, certaines mines corrélaient très bien avec des mines d’un autre type.

Cette étude a montré que la qualité des images est importante pour bien reconnaître un objet. De plus des méthodes rapides comme la corrélation peuvent nécessiter un temps de calcul important.

9.4 Utilisation des données de navigation

Nous venons de présenter une méthode dont les résultats montrent les limites. Pour remédier à ce problème, nous avons utilisé entre autres les informations de navigation. Nous connaissons la distance entre l’objet et la caméra. Nous avons aussi la distance caméra - mine pour chacune des images de référence des filtres. Nous pouvons ainsi sélectionner un sous ensemble de filtres correspondant à l’ordre de grandeur de la distance mesurée entre la caméra et la mine. Cette réduction du nombre de filtres à appliquer aux images cibles va permettre un temps de calcul plus faible. Nous pouvons aussi envisager de ne pas réduire le temps de calcul mais de faire corréliser davantage de filtres, créés avec des critères plus stricts.

9.4.1 Algorithme utilisé

Comme défini précédemment, lorsqu’il n’y a pas de détection, nous n’effectuons pas l’étape de corrélation. Nous avons aussi vu qu’il faut effectuer une étape de détection de contours avant

de corréler les images cibles avec les filtres. Nous avons choisi pour cela, la méthode basée sur l'approche multi-échelle. Si cette approche ne trouve pas de contours dans l'image, nous considérons qu'il n'y avait en fait pas d'objet à détecter dans l'image et étudions l'image suivante.

Lorsqu'un objet est détecté dans l'image et que la détection de contours a trouvé des contours (il peut s'agir des contours de l'objet comme de contours du fond), nous effectuons l'opération de corrélation.



FIGURE 9.10 – Image résultat de la détection de contours basée sur une approche multi-échelle. Vidéo Tovision008 représentant une mine sphérique

Tout d'abord nous essayons de faire corréler les contours trouvés (*cf.* figure 9.10) avec un cercle. Pour cela, nous calculons la largeur et la hauteur de la boîte englobant les contours. Nous calculons le cercle correspondant au plus petit côté (le minimum entre la largeur et la hauteur) (*cf.* figure 9.11).



FIGURE 9.11 – Cercle calculé à partir de l'image de contours

Nous créons un filtre POF avec ce cercle et nous le faisons corréler avec notre image de contours. Si le PCE est supérieur à un seuil fixé empiriquement alors l'objet présent dans l'image est une sphère. Sinon nous faisons corréler l'image de contours avec les filtres sélectionnés. Comme nous connaissons la distance entre l'observateur et l'objet ainsi que la distance fixée entre la caméra et les mines lors de l'enregistrement des images de référence, nous pouvons ainsi sélectionner les filtres dont les distances sont les plus proches de la distance entre la caméra et l'objet à identifier. La caméra ayant été déplacée par pas de 0.5 m lors de l'enregistrement des images et la distance entre l'observateur et l'objet à identifier évoluant de façon continue, nous avons décidé de corréler avec les filtres correspondant aux deux distances les plus proches. Enfin, comme dans

la méthode précédente, le meilleur PCE donne le type de la mine. L'algorithme décrit ici est présenté sur la figure 9.12.

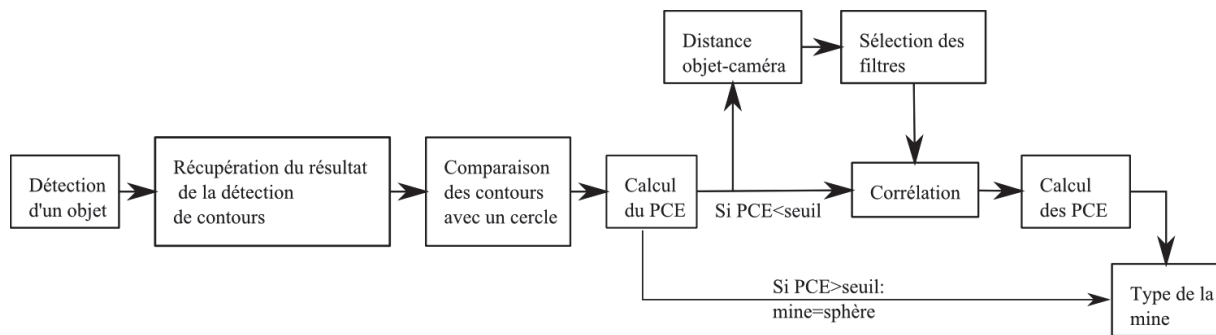


FIGURE 9.12 – Schéma synoptique de la méthode de reconnaissance utilisant les informations de navigation

9.4.2 Résultats obtenus

Nous avons appliqué cette méthode aux vidéos Topvision. Les résultats sont présentés dans le tableau 9.7. Comparativement aux résultats obtenus avec la précédente méthode (*cf.* tableaux 9.4 et 9.6), les probabilités obtenues sont améliorées. Le taux de bonnes reconnaissances augmente pour les mines cylindriques et Manta. La probabilité sur le type de la mine augmente mais cela vient de la baisse de la probabilité de fausses alarmes. Nous identifions des objets dans les images cibles ailleurs qu'à leur place réelle. Le changement des filtres de corrélation permet de reconnaître des objets au bon endroit même si ce n'est pas le bon type de mine. De plus, le temps de calcul a été divisé par 6. Nous n'avons pas encore atteint l'objectif des 2 Hz fixé mais le gain est déjà important.

Mine	Pdc	Ptype	Pfac	Pndvc	Temps de calcul
Mine Manta	22.88%	53.61%	2.04%	98.7%	4.9 s
Mine cylindrique	55.62%	21.74%	10.67%	98.82%	4.6 s
Mine sphérique	1.81%	57%	7%	100%	4.5 s
Vidéos vides			21.2%	78.79%	6 s

Tableau 9.7 – Résultats obtenus sur les vidéos Topvision

Nous avons eu le temps de tester cet algorithme sur quelques vidéos VAMA : les vidéos de sphères et les vidéos vides. Les résultats sont encourageants : la mine sphérique est bien identifiée (50% de mines correspondant à une sphère sur toutes les images de mines présentes dans la vidéo, *cf.* tableau 9.8. Par comparaison, nous avons obtenu 58.1% de détection de mines). Le taux de fausses alarmes est pratiquement nul ce qui est important si le drone doit se servir des résultats de cet algorithme pour se guider. Le taux de non détections fausses est élevé (40%). Nous pouvons constater sur la figure 9.13 que la probabilité de bonne reconnaissance augmente lorsque la caméra se rapproche de la mine pour atteindre presque 100% entre 1 et 2 m de la mine. Passé cette distance, le taux de bonnes reconnaissances chute pour se transformer en taux de probabilité type. Cela s'explique par la courte distance entre la mine et la caméra. A ces courtes distances, la mine est plus grande que la zone imagée par la caméra. Par conséquent, il est difficile d'obtenir une image de contours cible contenant tous les contours de la mine.

De plus, l'analyse du résultat du traitement multi-échelle permet d'éliminer de fausses alarmes dans le cas des vidéos vides (*cf.* tableau 9.9). En effet, le taux de fausses alarmes passe de 4.18% en détection à 1.95% après l'étape de reconnaissance. Le traitement multi-échelle pour obtenir l'image de contours a donc permis d'affiner les résultats de la détection. Nous n'avons pas eu le temps de tester la reconnaissance sur les vidéos de mines Manta, cylindriques et sur les autres objets pour confirmer ces bons résultats.

Reconnaissance	Pdc	Pfac	Ptype	Pndf	Pndv
	50.37%	0.001%	9.08%	40.54%	99.8%

Tableau 9.8 – Résultats de la reconnaissance sur les vidéos de mines sphériques VAMA

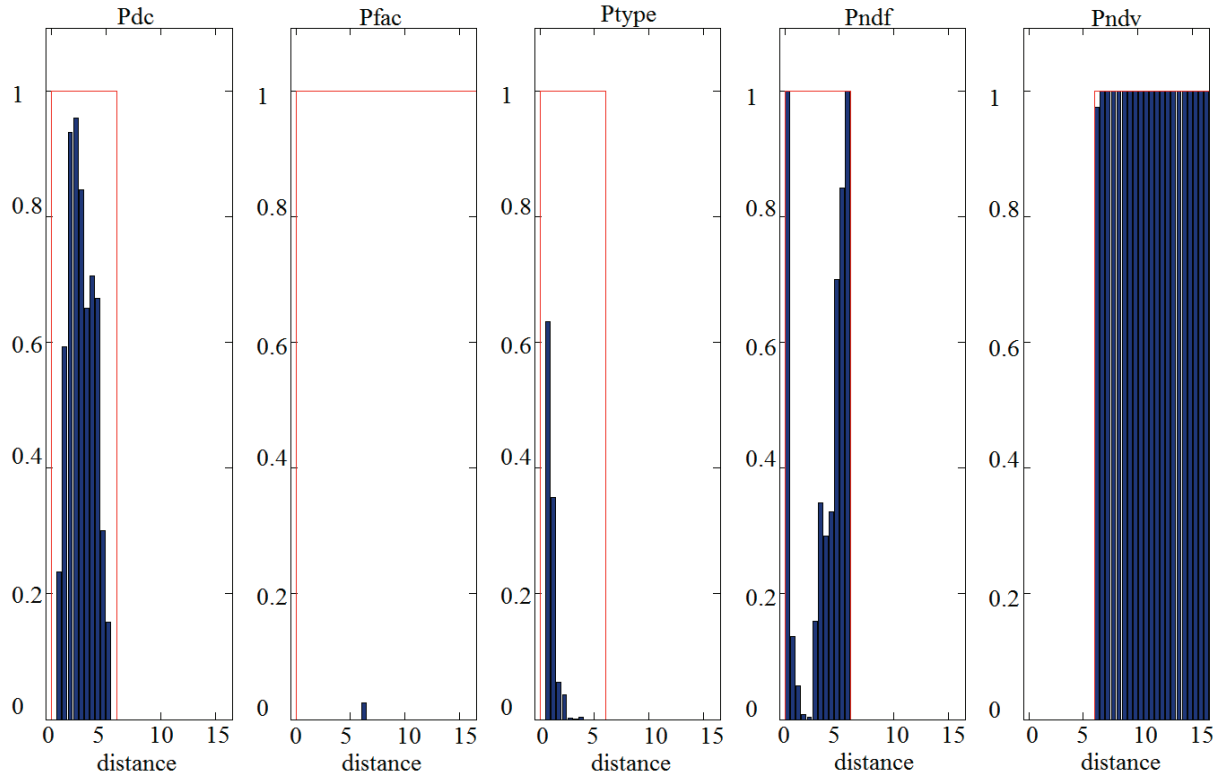


FIGURE 9.13 – Résultats de la reconnaissance

Mine	Pfa	Pndv
Détection	4.18%	95.82%
Reconnaissance	1.95%	98.05%

Tableau 9.9 – Résultats de la détection et de la reconnaissance sur les vidéos vides VAMA

9.5 Conclusion

Dans ce chapitre, nous avons étudié la reconnaissance des mines et la création des filtres de corrélation dans le cas d'utilisation d'images de référence de synthèse.

Nous avons étudié plusieurs méthodes d'assemblage des images de référence. Ces méthodes permettent de diminuer le nombre d'opérations à effectuer. Il peut être intéressant de poursuivre les études menées pour adapter au mieux les méthodes et contraintes pour réduire davantage le nombre de filtres obtenus.

Les tests effectués ont montré que la corrélation d'une image cible avec tous les filtres de la base n'est pas ingénieuse au niveau des résultats et du temps de calcul. Nous avons donc utilisé de nouveau les informations de navigation pour sélectionner les filtres correspondant à la mine. Il est possible d'aller plus loin. En effet, une fois en confiance par rapport au type de la mine, on peut sélectionner une catégorie de filtres, et quelques filtres si on relie l'orientation de la mine aux mouvements du véhicule.

Enfin il est possible de diviser la base de filtres à tester d'un facteur 2 en utilisant les filtres ASPOF, présenté dans la partie 5.3. Cela n'a pas encore été testé par manque de temps.

Conclusion

Conclusion générale

Cette thèse avait pour but de proposer des solutions algorithmiques à la détection, à la classification et à l'identification des mines sous marines par caméra vidéo. Pour traiter ce problème, nous nous sommes intéressés au contexte du problème et de ses contraintes (nécessité de prétraiter les images), ainsi qu'aux différentes méthodes utilisées en reconnaissance optique. Ensuite, nous avons présenté les études menées pour résoudre notre problème.

Tout d'abord, l'étude du milieu sous marin nous a permis d'étudier les problèmes liés à l'utilisation de capteurs optiques. Il en résulte une spécificité des images sous-marines qui nous a amené à introduire des prétraitements. Pour cela, nous avons fait des choix en fonction des images et des informations dont nous disposions.

L'état de l'art des méthodes de détection et de reconnaissance en milieu sous marin nous a fait choisir la corrélation comme base de notre algorithme. Mais avant d'appliquer la corrélation à la détection et la reconnaissance des mines, nous l'avons étudiée en détails, afin de vérifier qu'elle correspondait à nos besoins. Nous l'avons aussi adaptée à nos contraintes.

Les tests effectués ont montré que la corrélation est une méthode correspondant à notre besoin. Nous l'avons appliquée à la détection et à la reconnaissance des mines. Cela n'a pas donné les résultats attendus. Nous avons donc découpé notre algorithme en deux parties distinctes : la détection et la reconnaissance.

Nous avons développé deux méthodes pour détecter des objets dans les images. La première méthode est basée sur la détection de sauts dans l'image spatiale correspondant à la phase du spectre de l'image cible. Cette méthode donne de bons résultats (taux de détection supérieur à 60%) lorsque les images sont contrastées. Pour les images dont les objets ne ressortent pas, nous avons développé une méthode basée sur la soustraction du fond (appris en début de séquence vidéo) et sur la comparaison de l'image initiale et de l'image traitée. Cette méthode a été optimisée en utilisant les informations de navigation qui permettent de réduire les fausses alarmes et les arrêts de détection. Cette méthode a aussi été comparée à une méthode actuellement utilisée dans un drone et développée par TUS et Cybernetix. Nous avons montré que nos résultats sont meilleurs que ceux obtenus avec ce drone.

Une fois l'objet détecté, nous nous sommes intéressés à la reconnaissance par corrélation. Pour cela, nous avons créé des images de synthèse des mines puis les filtres de corrélation correspondants. Dans un premier temps, tous les filtres (soit toutes les images de référence) ont été corrélés aux images cibles. Cela nécessite un temps de calcul important (environ 30 s par image où un objet a été détecté). Nous avons donc réutilisé les informations de navigation. Comme nous connaissons la distance qui nous sépare de la mine, nous pouvons sélectionner les filtres qui correspondent à cette distance et ainsi réduire le nombre de filtres à faire corrélés ainsi que le temps de calcul (environ 5 s par image où un objet a été détecté). Cela a aussi pour effet d'améliorer le taux de bonnes identifications.

Contributions

Dans l'introduction, nous avons annoncé des apports concernant : les architectures et les filtres de corrélation, la détection des mines et l'utilisation des informations de navigation. Nous allons maintenant faire le bilan de ce que nous avons apporté au cours de cette thèse.

Dans cette thèse, nous avons proposé une nouvelle architecture pour le JTC que nous avons

appelée nonlinear fringe-adjusted JTC ainsi que le critère de performance associé qui permettent de faciliter la prise de décision, notamment dans les applications qui nécessitent une décision automatique.

Au vu du nombre de filtres à utiliser pour reconnaître une mine sous marine, nous avons également proposé pour l'architecture de Vanderlugt un nouveau filtre : le filtre ASPOF, pour "*Asymmetric Segmented Phase Only Filter*" qui permet de diviser la base des images de référence par un facteur deux. Le critère de décision (PCE) a dû être adapté et est appelé PCEA "*Peak to Correlation Energy Adapted*". Lors de tests sur les lettres et les visages, ce filtre a montré son intérêt. L'application de ce filtre à l'étude des mines devrait permettre une réduction du temps de calcul nécessaire au traitement des images.

Un autre apport de la thèse concerne les filtres de corrélation pour la reconnaissance 3D (fabrication). En effet, pour créer tous les filtres que nous voulons, nous avons utilisé des modèles de synthèse 3D des mines. Ainsi nous avons pu enregistrer des images 2D correspondant à toutes les vues souhaitées. De plus, les filtres ainsi obtenus ne sont pas perturbés par du bruit, contrairement aux images issues des vidéos réelles.

La détection des mines nous a permis de développer deux nouvelles méthodes : la première basée sur l'image spatiale associée à la phase du spectre de l'image cible, la seconde basée sur la soustraction du fond. La première méthode est nouvelle dans le milieu sous marin. Bien qu'elle ne fonctionne qu'avec des objets contrastés par rapport au fond (c'est une méthode proche d'une détection de contours classique), elle permet d'obtenir de bons résultats. La soustraction du fond d'une image n'est pas nouvelle en soi, mais la manière dont nous avons adapté cette méthode est en rupture avec ce qui se fait actuellement. Nous nous servons de l'image initiale issue de la séquence vidéo et de l'image traitée pour détecter les objets. De plus, nous utilisons les informations de navigation (position et orientation de l'AUV) pour améliorer nos résultats, avec des contraintes "simples" et réalistes.

Nous avons utilisé plusieurs fois ces informations de navigation pour améliorer nos résultats. Elles ont été utilisées dans les prétraitements de façon à appliquer le modèle de Phong aux images. Elles ont aussi été utilisées dans l'étape de reconnaissance pour sélectionner les filtres "*ad hoc*". Cette utilisation permet d'apporter des améliorations non négligeables.

Nous avons divisé l'algorithme permettant de résoudre notre problème en deux parties : la détection d'une cible potentielle et la reconnaissance de celle-ci. La détection permet de limiter le temps de calcul nécessaire pour traiter une image, surtout si elle ne contient pas d'objet. La reconnaissance permet de confirmer la présence d'un objet. L'utilisation des résultats obtenus sur les n dernières images de la séquence vidéo permet d'améliorer la détection et la reconnaissance des objets dans les images à venir de la séquence. L'utilisation des informations de navigation permet d'accroître encore le taux de détection. Il est possible d'aller plus loin en utilisant le JTC (par exemple) pour faire du suivi, lorsque nous avons la certitude de la présence et/ou du type d'un objet.

Perspectives

Outre les pistes d'amélioration proposées dans les différents chapitres, nous avons aussi réfléchi aux suites que nous pourrions donner aux travaux effectués dans cette thèse. Nous avons identifié deux axes de travail : l'utilisation de la polarisation et l'implémentation sur cibles programmables.

Utilisation d'images polarisées

La polarisation est une technique utilisée depuis les années 1960 pour acquérir des images dans le milieu sous marin. Cela permet d'améliorer la qualité des images ainsi que la distance de visibilité. En effet, le milieu sous marin est un milieu diffusant qui atténue la lumière. Comme le montre la figure 9.14, la visibilité et le contraste des images peuvent être faible, il peut y avoir des particules qui réfléchissent la lumière,...



FIGURE 9.14 – Image issue d'une vidéo réelle

Nous allons dans cette partie présenter des généralités sur la polarisation, la problématique entre notre sujet de recherche et l'utilisation de la polarisation ainsi que les tests que nous allons effectuer.

Généralités sur la polarisation

La lumière peut être décrite par une onde plane électromagnétique. Elle possède un champ électrique et un champ magnétique perpendiculaires entre eux et perpendiculaires à la direction de propagation de l'onde. Le champ électrique s'écrit [58] :

$$\vec{E} = E_x \vec{x} + E_y \vec{y} + E_z \vec{z} \quad (9.4)$$

Selon l'expression des termes E_x et E_y , on connaît l'état de polarisation de la lumière.

– La lumière a une polarisation linéaire si

$$\begin{aligned} E_x &= E_{0x} \cos(\omega t - kz) \\ E_y &= E_{0y} \cos(\omega t - kz) \\ &\text{et } E_{0x} = E_{0y} \end{aligned} \quad (9.5)$$

Les polarisations linéaires les plus couramment utilisées sont les polarisations verticales, horizontales et orientées à $+45^\circ$ ou -45° par rapport au plan d'incidence.

- La lumière a une polarisation circulaire si

$$\begin{aligned} E_x &= E_{0x} \cos(\omega t - kz) \\ E_y &= E_{0y} \sin(\omega t - kz) \end{aligned} \quad (9.6)$$

- La lumière a une polarisation elliptique si

$$\begin{aligned} E_x &= E_{0x} \cos(\omega t - kz) \\ E_y &= E_{0y} \cos(\omega t - kz - \varphi) \end{aligned} \quad (9.7)$$

où E_{0x} et E_{0y} représentent l'amplitude maximale de l'onde plane, ω la pulsation, t le temps, k est la nombre d'onde, z la direction de propagation et φ le déphasage entre les composantes E_x et E_y .

La polarisation, côté mathématiques

Nous avons précisé dans le chapitre sur les prétraitements (*cf.* chapitre 2) qu'il est possible de décrire mathématiquement l'effet d'un milieu diffusant sur la polarisation de la lumière. Pour cela, on utilise les vecteurs de Stokes et les matrices de Mueller [58]. L'état de la lumière avant et après passage dans le milieu se décrit avec des vecteurs de Stokes composés de l'intensité totale I et de l'état de polarisation décrit par les termes Q, U, V (ces termes sont calculés à partir des polarisations linéaires verticale, horizontale, à $+45^\circ$, à -45° et circulaires). Le milieu est représenté par la matrice de Mueller M . Le système s'écrit :

$$\begin{aligned} A' &= MA \\ \text{avec } A &= \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} \\ \text{et } M &= \begin{pmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{31} & M_{32} & M_{33} & M_{34} \\ M_{41} & M_{42} & M_{43} & M_{44} \end{pmatrix} \end{aligned} \quad (9.8)$$

Calculer l'expression mathématique de la lumière reçue n'est pas toujours possible. En effet, les valeurs de la matrice de Mueller dépendent de la longueur d'onde, des particules présentes (concentration, taille, forme) et de la géométrie de la détection. Par conséquent elles dépendent de l'emplacement géographique et de la turbidité.

La polarisation, côté matériel

Pour contourner le problème des valeurs de la matrice de Mueller, il est possible d'enregistrer des images polarisées. Pour cela, il faut utiliser une caméra polarisée. Une solution simple et efficace consiste à placer un filtre polariseur devant une caméra classique, comme le font Y.Y. Schechner et N. Karpel [32]. Comme la caméra doit être placée dans un boîtier étanche, ils étudient la position du polariseur par rapport au boîtier (à l'intérieur ou à l'extérieur) ainsi que la forme et la composition de la "façade" du boîtier placée devant la caméra. Leur étude [32] montre qu'il est préférable de placer le polariseur à l'extérieur du boîtier dont la "façade" doit être un dôme en verre. En effet, cela permet de limiter les effets sur la polarisation dus au matériau utilisé pour le boîtier (*cf.* figures 9.15 et 9.16).

Avec ce système, ils enregistrent l'image d'intensité I_{\max} et l'image d'intensité I_{\min} . Ces 2 images "extrêmes" sont obtenues pour des orientations orthogonales du polariseur. Ils estiment la lumière rétrodiffusée. A l'aide d'un échantillon de sable supposé blanc, ils compensent simplement les canaux de couleurs afin de retrouver une image aux couleurs plus proches des couleurs réelles. Le système développé et l'algorithme de traitement des images permettent d'améliorer la qualité des images et d'augmenter la portée visuelle [32].

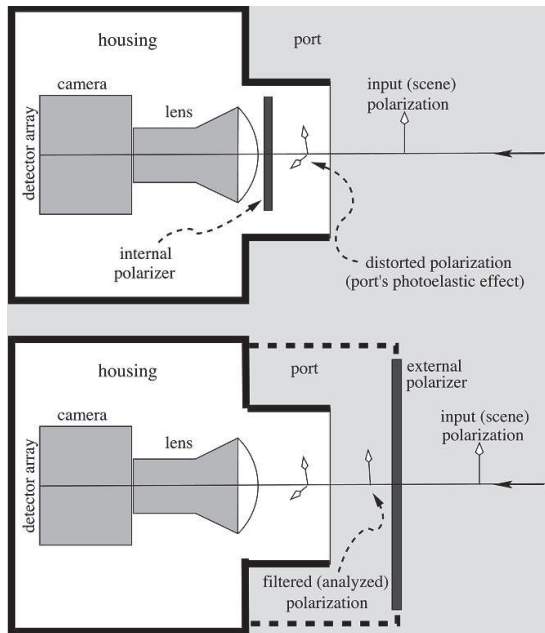


FIGURE 9.15 – Étude de la position du polariseur [32]

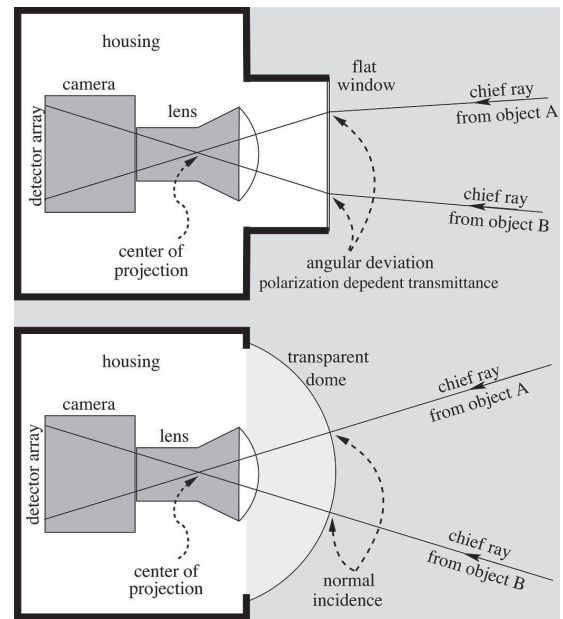


FIGURE 9.16 – Étude de la forme du dôme [32]

Problématique

Nous voulons détecter et reconnaître les mines sous marines. Nous avons vu que la qualité des images est très importante pour obtenir de bons résultats. La polarisation est une piste pour améliorer cette qualité. Des études menées [56,57] ont montré que l'utilisation de la polarisation circulaire permet d'augmenter le contraste et la distance de visibilité.

Nous ne connaissons à priori pas la composition des mines à détecter. Les différents matériaux ne vont pas donner les mêmes résultats sur les images polarisées. En effet, certains matériaux sont plus dépolarisants que d'autres (le plastique dépolarise davantage la lumière que le métal). De plus la peinture dont ils sont recouverts peut influencer la réponse à la lumière polarisée. Enfin l'évolution des mines au cours du temps est à étudier. En effet, si la mine passe beaucoup de temps dans l'eau, son aspect va changer : dépôt de sable, d'algues, de coquillage. Les mines elles même peuvent jouer sur l'efficacité de la polarisation pour traiter notre problème.

Il faut aussi déterminer si notre algorithme peut détecter les objets dans la plage de distance ainsi couverte. En effet, si l'objet est détecté à plus grande distance, nous pouvons acquérir la certitude de la présence de l'objet.

De plus si le contraste est amélioré, il faut étudier dans quelle mesure cela impacte notre algorithme. Les algorithmes de prétraitement et de détection peuvent être à adapter. Par conséquent, les résultats peuvent être modifiés de même que le temps de calcul.

Matériel à disposition et tests à mettre en place

Dans le laboratoire vision de l'ISEN, nous disposons d'un banc d'essai avec du matériel de test pour acquérir des images polarisées, dont un "aquarium". Dans un premier temps, nous allons étudier quelques images d'objets (pièce de monnaie, polystyrène, plastique) afin d'étudier les images polarisées et d'intensité correspondantes. Nous étudierons l'effet de l'eau, de l'eau avec du sel ainsi que du lait. Nous étudierons aussi les signatures polarimétriques des différents matériaux. Nous pourrions commencer à réfléchir aux algorithmes à mettre en place pour détecter et identifier au mieux les objets.

L'ISEN possède aussi un bassin dans lequel nous pourrions étudier notre système dans un volume d'eau plus important, avec une distance caméra - objet plus grande. Ces mesures permettront de nous rapprocher de conditions réelles d'acquisition des images. Il sera donc intéressant de comparer les résultats de nos algorithmes sur les images d'intensité et sur les images polarisées.

Premiers tests

Nous avons commencé par placer une pièce de monnaie dans l'aquarium rempli d'eau claire. Nous avons testé plusieurs polarisations : linéaire, verticale, orientée à 45° et circulaire. Visuellement, les images acquises en polarisation circulaire sont plus contrastées que les autres images acquises. Nous présentons ici l'image d'intensité (*cf.* figure 9.17) ainsi que l'image en degré de polarisation (*cf.* figure 9.18) associée, définie par la formule suivante :

$$DOP = \frac{I_{para} - I_{perp}}{I_{para} + I_{perp}} \quad (9.9)$$

où I_{para} représente l'intensité du faisceau polarisé parallèlement au faisceau incident et I_{perp} l'intensité du faisceau polarisé perpendiculaire au faisceau incident.

Sur l'image d'intensité, figure 9.17, la pièce de monnaie est visible mais les détails se devinent seulement. Elle est masquée par la rétrodiffusion, qui ajoute un voile sur l'image. Sur l'image en degré de polarisation associée, figure 9.18, nous avons réussi à fortement limiter la rétrodiffusion. Ainsi la pièce et ses détails apparaissent nettement.

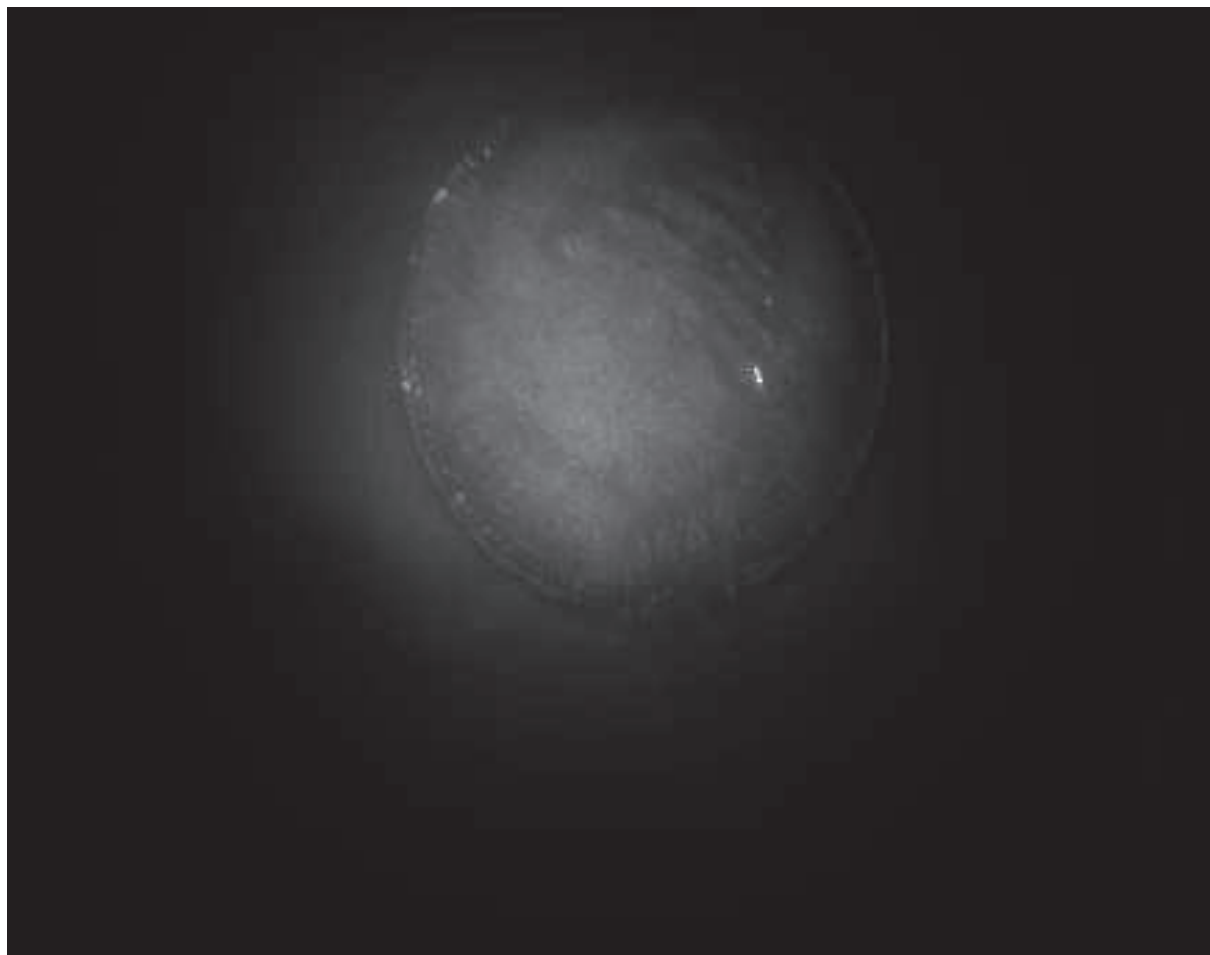


FIGURE 9.17 – Image d'intensité

Nous avons ensuite ajouté un peu de lait écrémé dans l'aquarium. Un volume de 40 mL de lait dans un volume de 10L d'eau permet de créer un milieu plus turbide que l'océan : le coefficient de diffusion typique de l'eau turbide est de 0.04 cm^{-1} et le coefficient de diffusion de notre mélange est de 0.16 cm^{-1} . Sur les figures 9.19 et 9.20, nous présentons respectivement l'image d'intensité totale de la pièce dans cette eau trouble et l'image en degré de polarisation circulaire correspondante. Nous pouvons constater que sur l'image d'intensité la pièce est à peine visible, masquée par la rétrodiffusion tandis que sur l'image en degré de polarisation nous distinguons jusqu'aux détails de la pièce.



FIGURE 9.18 – Image en degré de polarisation

Nous avons effectué un test simple afin de vérifier que la prise de décision après une corrélation est facilitée avec les images polarisées. Pour cela, nous avons créé un cercle de la taille de la pièce. Nous avons créé un filtre de corrélation POF. Nous avons fait corréler ce filtre avec les quatre images présentées ci-dessus. Nous obtenons les plans de corrélation présentés sur la figure 9.21. Nous pouvons remarquer que les images d'intensité (figures 9.21a et 9.21c) ne permettent pas d'identifier la pièce présente dans l'image. Les images en degré de polarisation (figures 9.21b et 9.21d) présentent des pics de corrélation qui permettent de prendre une décision sur la présence de la pièce. Le lait ajouté à l'eau crée une turbidité qui dégrade un peu le plan de corrélation (figure 9.21d) mais le pic est toujours présent et détectable.

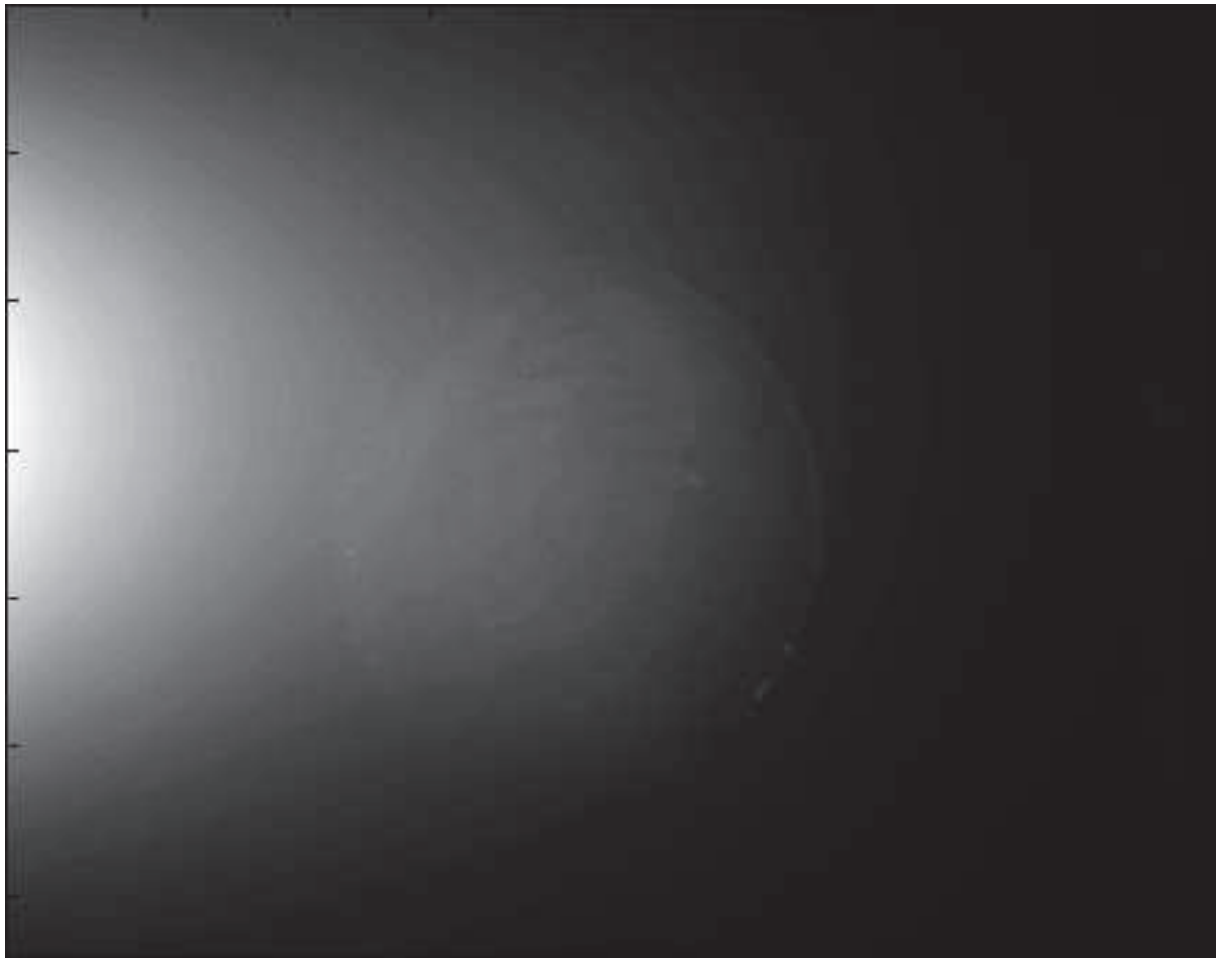


FIGURE 9.19 – Image d'intensité

Implémentation sur cibles programmables

Jusqu'à présent, nous avons travaillé sur Matlab. Il faut cependant rappeler que les travaux effectués dans cette thèse doivent être implantés dans un drone autonome, "sans ordinateur et sans Matlab à disposition". De plus, l'énergie est limitée car les ressources du drone sont aussi utilisées pour la navigation, pour l'utilisation du sonar. Il faut donc envisager d'implanter nos algorithmes sur une cible programmable tout en optimisant la rapidité et la consommation. Nous avons utilisé les fonctions programmées par défaut de Matlab. Il faudra donc les reprogrammer. Dans notre laboratoire, des travaux sont effectués sur les FPGA et sur les GPU [95, 128–130]. Nous envisageons donc de les utiliser pour rendre notre algorithme facilement utilisable sur un drone autonome.

Les cibles programmables

Il existe plusieurs types de cibles programmables qui peuvent se regrouper en trois familles : les circuits à architectures généralistes, les circuits à architectures dédiées et les circuits à architectures programmables [128].

Les circuits à architectures généralistes utilisent des processeurs comme par exemple les CPU, "*Central Processing Unit*" et les GPU, "*Graphical Processing Unit*". Cela leur confère une grande flexibilité pour exécuter efficacement programmes et logiciels. Leur capacité de calcul augmente avec l'augmentation de la fréquence de l'horloge et l'utilisation du parallélisme.

Les circuits à architectures dédiées consistent à concevoir des architectures spécifiques pour des applications bien identifiées en utilisant des circuits intégrés dits ASIC (programmables une seule fois). Ces circuits permettent d'obtenir une rapidité importante. La flexibilité des ASICs

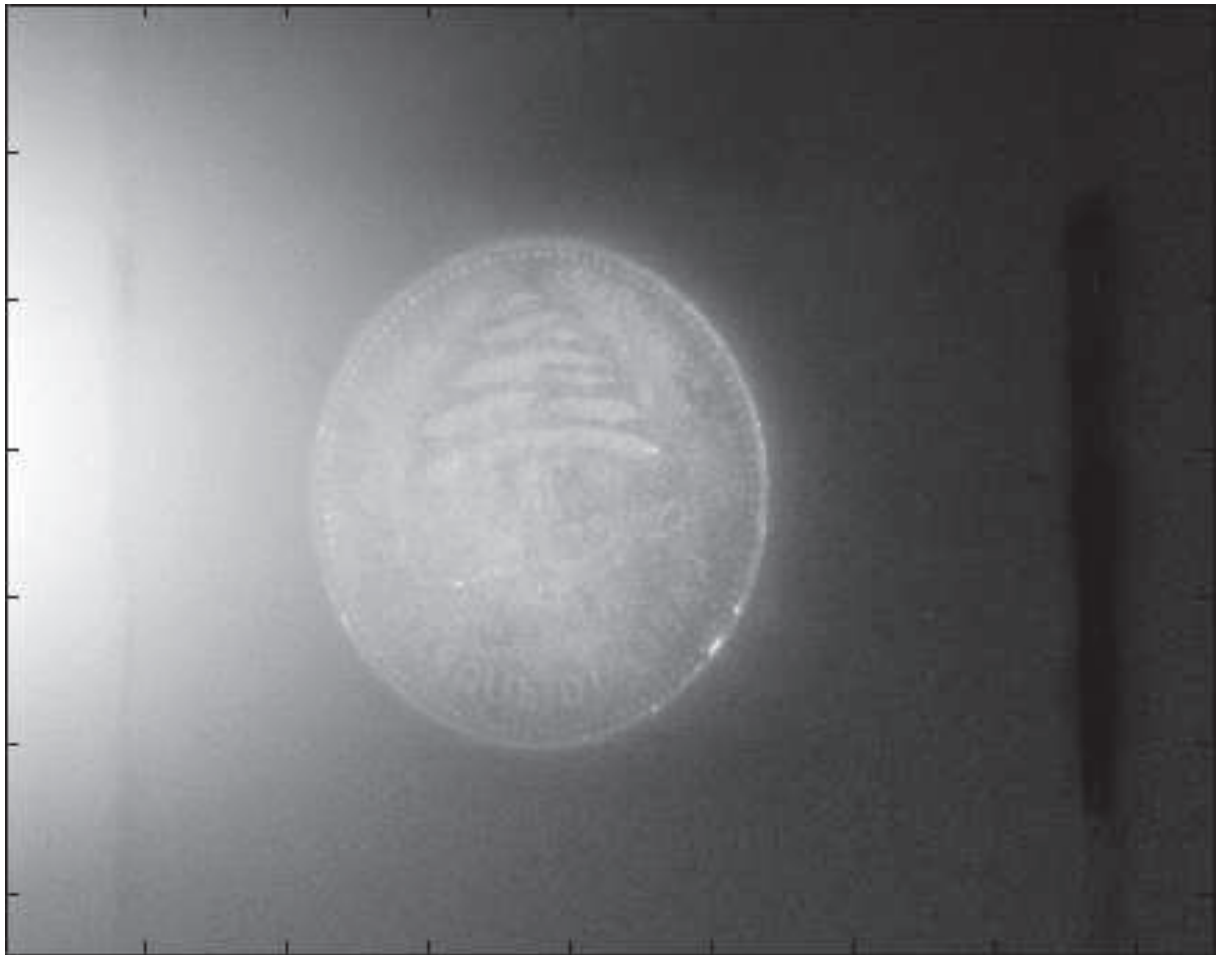


FIGURE 9.20 – Image en degré de polarisation

est très faible puisqu'ils sont développés pour une seule application.

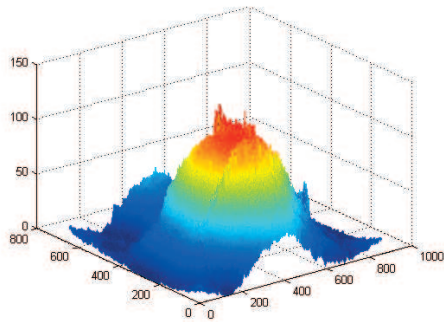
Les circuits à architectures programmables (par exemple les FPGA, "Field Programmable Gate Array") sont une famille intermédiaire. Ces circuits sont composés de portes logiques, de blocs mémoires et d'un réseau d'interconnexion. Par rapport aux ASICs, ils sont reprogrammables mais leurs performances sont moins importantes. De plus, leur consommation statique d'énergie est importante.

La figure 9.22 [128] résume ce que nous venons d'exposer brièvement. Cette figure représente la flexibilité des architectures en fonction de leurs performances. La puissance consommée (axe sur la droite de la figure) évolue de la même manière que la flexibilité. Les ASICs ne sont pas reconfigurables donc pas flexibles mais leur performances sont très bonnes. Les CPU et GPU sont très flexibles mais consomment beaucoup d'énergie. Le coût de développement augmente de façon inverse à la consommation. En effet, les ASICs qui consomment peu et sont spécifiques à une application ont un coût de développement important. Enfin les FPGA représentent un compromis.

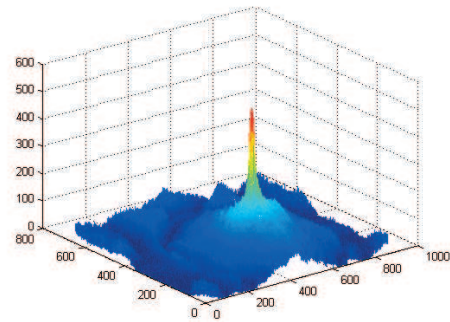
Études faites dans notre laboratoire

Dans notre laboratoire, des travaux ont été effectués sur les FPGA [129, 130] et sur les GPU [95]. Le but est de développer les algorithmes nécessaires pour effectuer la reconnaissance par corrélation.

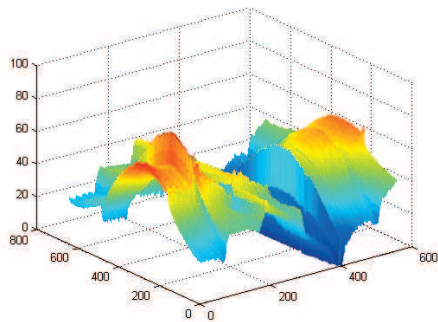
Sur les FPGA, une étude a été faite sur les transformées de Fourier [129, 130]. Un nouvel algorithme pour calculer la transformée de Fourier plus rapide que celles existantes a été développé et validé. Cet algorithme permet, avec le composant utilisé, de calculer 3000 corrélations par seconde sur des images de taille 256 pixels x 256 pixels. C'est équivalent à la cadence d'un



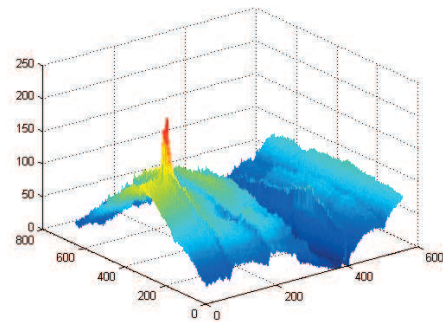
(a) Plan de corrélation obtenu avec l'image d'intensité de la pièce dans l'eau



(b) Plan de corrélation obtenu avec l'image en degré de polarisation de la pièce dans l'eau



(c) Plan de corrélation obtenu avec l'image d'intensité de la pièce dans l'eau et le lait



(d) Plan de corrélation obtenu avec l'image en degré de polarisation de la pièce dans l'eau et le lait

FIGURE 9.21 – Plans de corrélation obtenus après corrélation d'un cercle de la taille de la pièce avec les images d'intensité et en degré de polarisation présentées dans cette partie

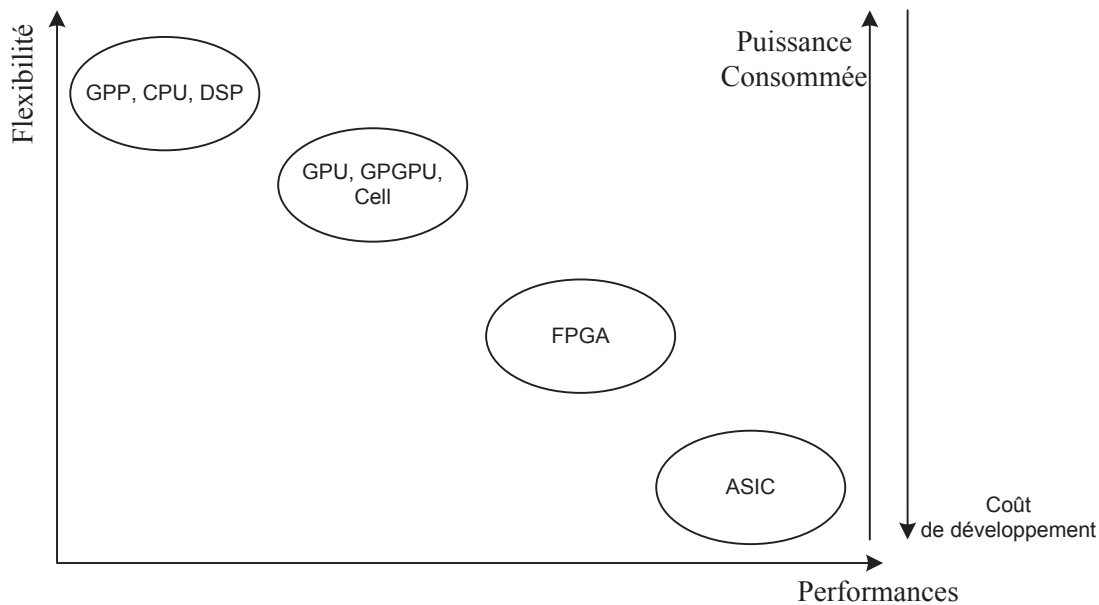


FIGURE 9.22 – Performances des cibles programmables [128]

SLM, soit d'un montage optique [104]. Les filtres de corrélation utilisés sont binaires vu le nombre limité des mémoires embarquées sur le FPGA. Cependant, il est possible d'utiliser des mémoires externes disponibles dans la carte électronique sur laquelle est mis le FPGA pour stocker nos filtres de corrélation dans le cas où la mémoire interne du FPGA ne serait pas suffisante.

Les GPU ont été utilisés pour effectuer la corrélation optique que nous avons utilisée [95]. L'algorithme correspondant a été créé. De plus les performances de la corrélation sur GPU ont été comparées à celles obtenues avec le même algorithme sur CPU. Les résultats montrent que l'utilisation d'une carte graphique permet de réduire les temps de calcul. Les cartes graphiques actuelles ont l'avantage de paralléliser les calculs, ce qui permet de gagner en rapidité.

Problématique

Notre algorithme a été développé sous Matlab. Pour l'utiliser sur un drone sous-marin autonome, il faut tout d'abord choisir une cible programmable.

Le choix de la cible programmable va conditionner l'énergie nécessaire au fonctionnement de la carte électronique responsable de la partie détection et reconnaissance vidéo. Il faut prendre en compte la consommation statique et la consommation dynamique des composants. La consommation statique ne dépend pas de notre algorithme mais du composant lui-même. La consommation dynamique dépend de notre algorithme et de l'utilisation des ressources du composant. Le FPGA a une consommation statique importante. Le GPU est un composant qui ne fonctionne pas de façon autonome. Il nécessite une partie hôte qui peut être le CPU de l'ordinateur ou un micro-contrôleur.

Nos filtres de corrélation sont de taille importante. L'utilisation de mémoires externes pour les stocker peut être nécessaire. Dans ce cas, l'accès aux données demandera un peu plus de temps que si elles étaient stockées dans le composant lui-même.

Quelque soit la cible programmable, il faudra traduire notre algorithme et créer les fonctions nécessaires. En effet, beaucoup de fonctions "de base" sont proposées dans Matlab. Ce n'est pas le cas en langage C, pour le GPU ou en VHDL, pour le FPGA. De plus, la facilité de programmation n'est pas la même.

Annexes

Annexe A

Comparaison entre les filtres ASPOF, AMPOF, MACE, composite et segmenté

A.1 Le principe du filtre AMPOF

Le filtre AMPOF, “*Amplitude Modulated Phase Only Filter*”, présenté par Awwal *et al.* [131], consiste à pondérer le filtre POF de la façon suivante :

$$H(\mu, \nu) = \frac{a|F(\mu, \nu)|\exp(-j\phi(\mu, \nu))}{(b + c|F(\mu, \nu)| + d|F(\mu, \nu)|^2)^m} \quad (\text{A.1})$$

avec F la transformée de Fourier de l’image référence, a , b , c , d et m sont des facteurs permettant de calculer un filtre adapté ($a = b = m = 1$ et $c = d = 0$), POF ($b = d = 0$, $a = c$ et $m = 1$) ou AMPOF (autres valeurs). La division par une fonction de l’amplitude du spectre de l’image référence permet d’obtenir des pics de corrélation étroits. Les résultats obtenus sur la reconnaissance des lettres sans bruit montrent que ce filtre est très discriminant (*cf.* figure 5.14).

A.2 Le filtre MACE

A.2.1 Principe du filtre

Le filtre MACE, “*Minimum average correlation energy*” a pour but de minimiser l’énergie moyenne du plan de corrélation, tout en contraignant la valeur que doit avoir le pic de corrélation [112]. Pour remplir ces conditions, le filtre doit avoir pour expression :

$$H = D^{-1}S(\bar{S}D^{-1}S)^{-1}\bar{c} \quad (\text{A.2})$$

avec D une matrice diagonale de taille $d \times d$ (d étant le nombre de pixels dans l’image) contenant sur la diagonale l’énergie des pixels des spectres des images références. S est une matrice de taille $d \times N$ où N est le nombre d’images du filtre, contenant les N transformées de Fourier des images références sous forme de colonne. c est un vecteur contenant les valeurs contraintes pour les pics de corrélation. (Pour deux images identiques, la valeur du pic de corrélation doit être de 1.)

A.2.2 Résultats obtenus

Tout d’abord nous avons testé la robustesse à la rotation du filtre MACE. Nous pouvons constater sur la figure A.1 que les images comprises entre -20° et $+25^\circ$ (zone B) sont reconnues par le filtre contenant 10 références prises tous les 5° dans cet intervalle d’images. Les images qui ne sont pas contenues dans le filtre (zones A et C) ne sont pas reconnues. Nous pouvons aussi constater que les performances de ce filtre se dégradent lorsque les images cibles sont perturbées par du bruit. Nous avons testé la robustesse au bruit blanc (figure A.2) et au bruit texturé (figure A.3).

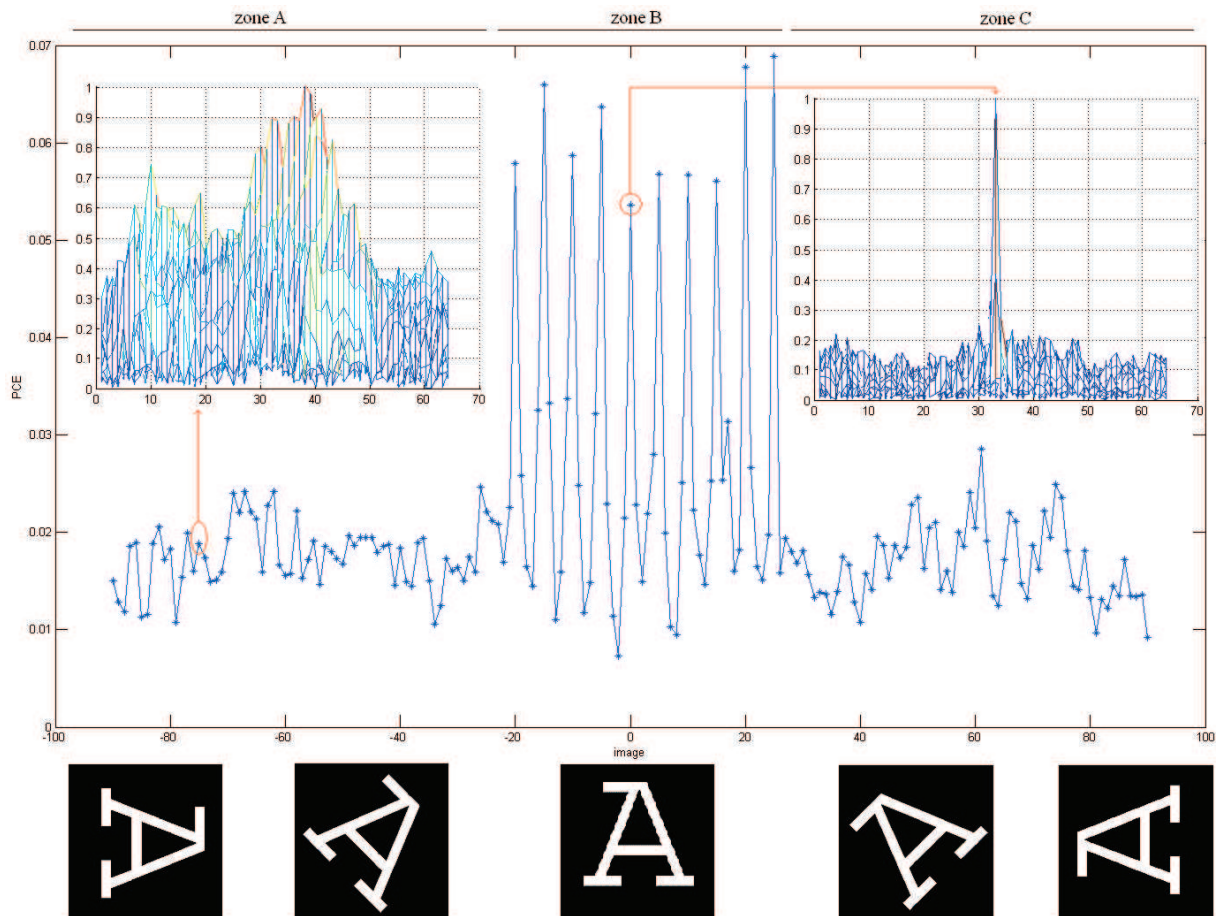


FIGURE A.1 – Reconnaissance de la lettre A par un filtre MACE contenant 10 références

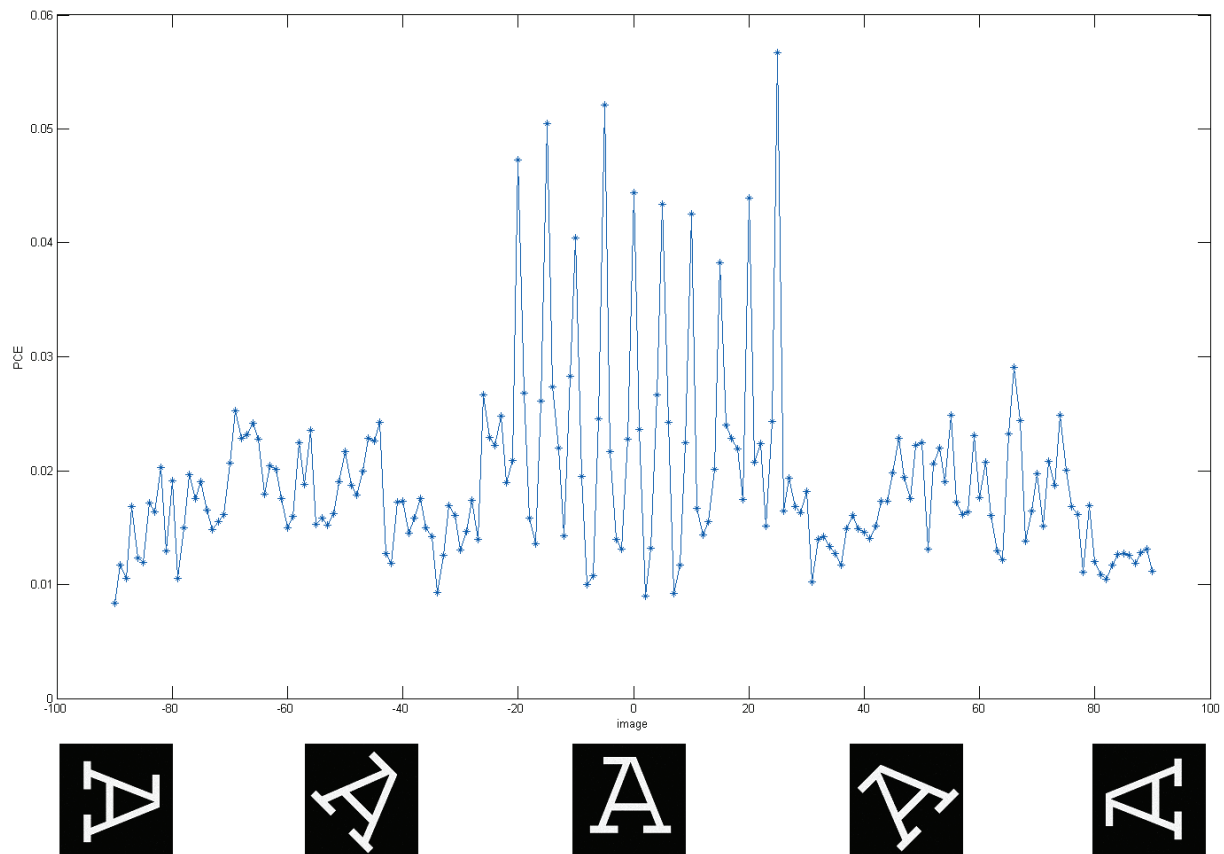


FIGURE A.2 – Reconnaissance de la lettre A bruitée par un bruit blanc par un filtre MACE contenant 10 références

A.3 Robustesse au bruit du filtre ASPOF

Nous avons testé la robustesse du filtre ASPOF sur la reconnaissance d'images bruitées comparativement aux résultats obtenus avec les filtres composite POF, segmenté POF et AMPOF. Pour cela, nous avons utilisé les images sur lesquelles un bruit blanc a été ajouté (*cf.* figure A.4) puis les images sur lesquelles une texture a été ajoutée (*cf.* figureA.5).

Lorsqu'un bruit blanc est ajouté aux images cibles (*cf.* figure A.4), les filtres composite, segmenté et ASPOF continuent de reconnaître la lettre. La séparation entre reconnaissance et non reconnaissance est toujours présente et permet de fixer des seuils pour une décision automatique. Le filtre AMPOF ne les reconnaît que très faiblement. Il s'agit donc d'un filtre très peu robuste.

Lorsqu'un bruit texturé est ajouté aux images (*cf.* figureA.5), le filtre AMPOF ne reconnaît plus les lettres. Le filtre composite ne les reconnaît que très faiblement. Les filtres segmentés et ASPOF continuent de reconnaître les lettres. La décision automatique est toujours applicable pour ces deux filtres.

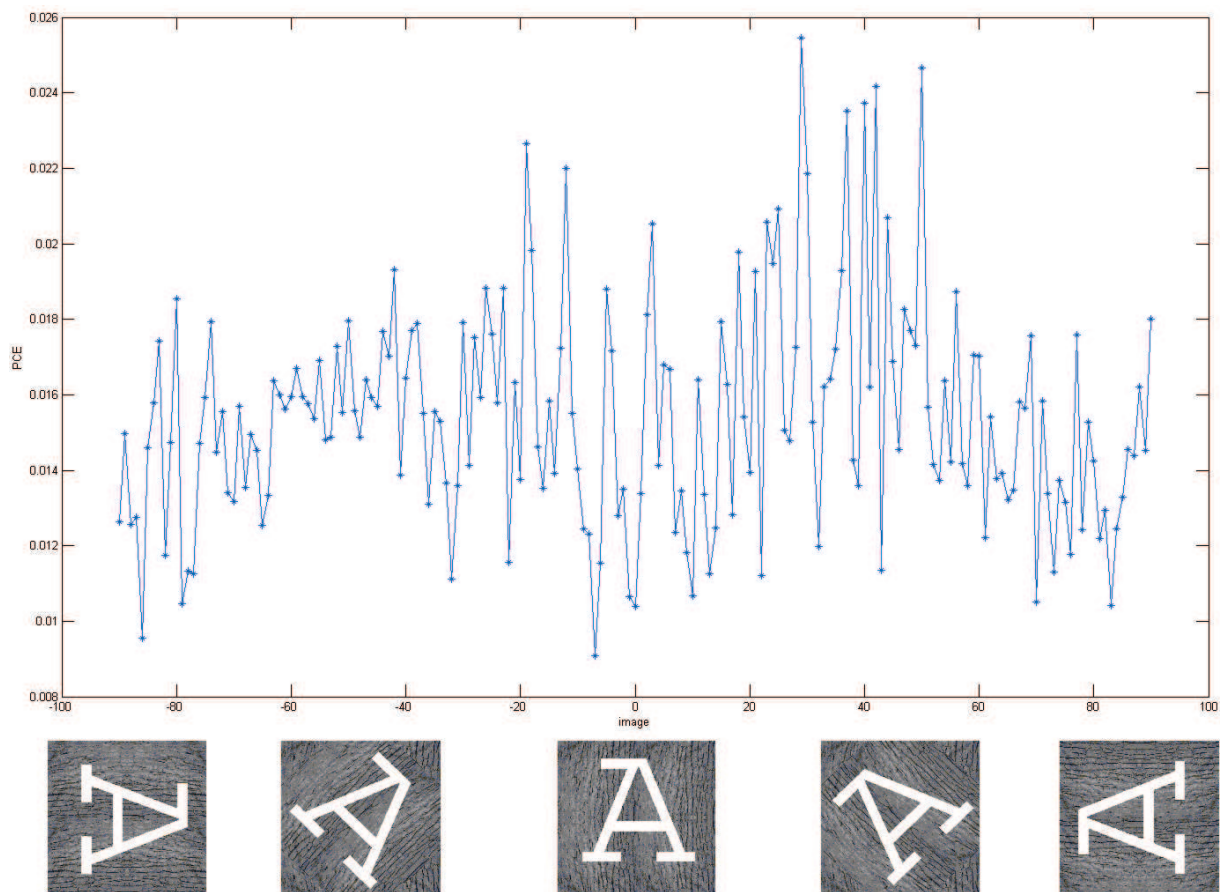


FIGURE A.3 – Reconnaissance de la lettre A bruitée par une texture par un filtre MACE contenant 10 références

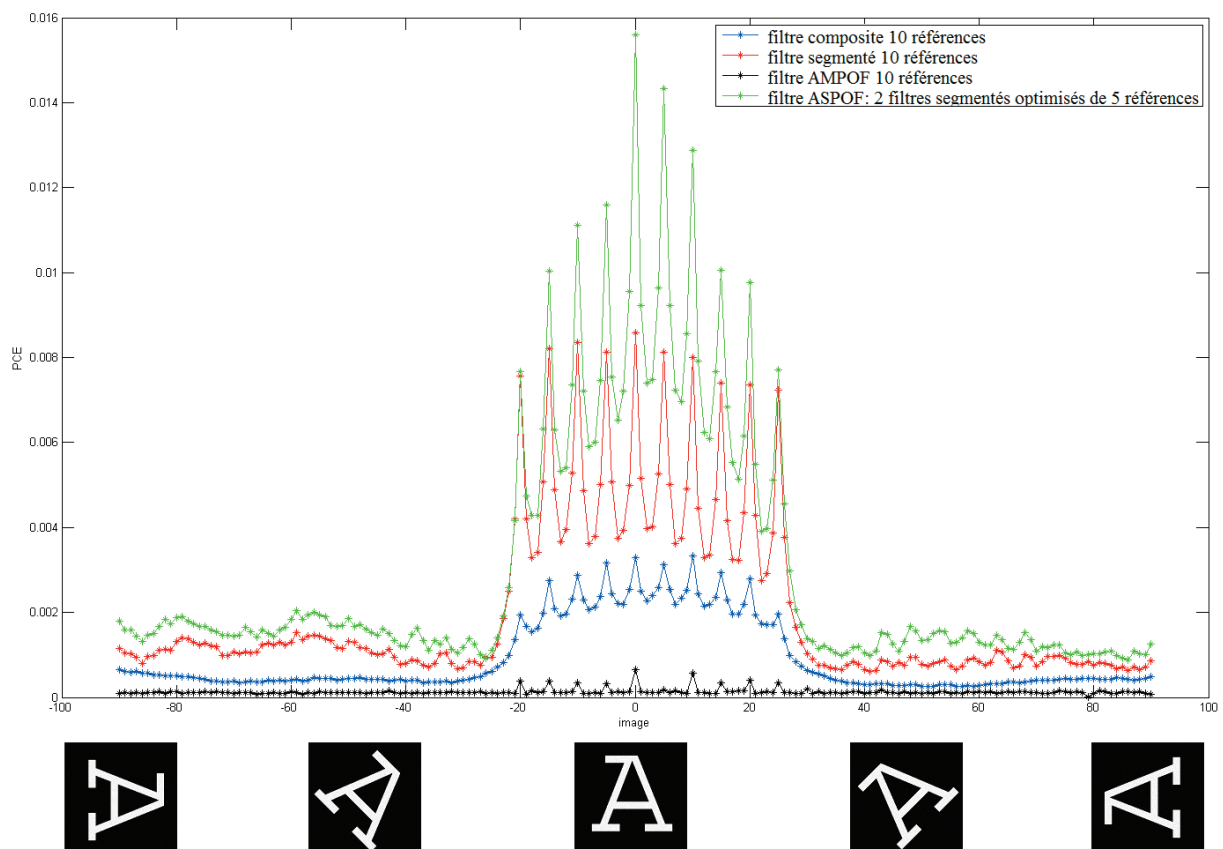


FIGURE A.4 – Comparaison des filtres sur les lettres A perturbées par un bruit blanc additif

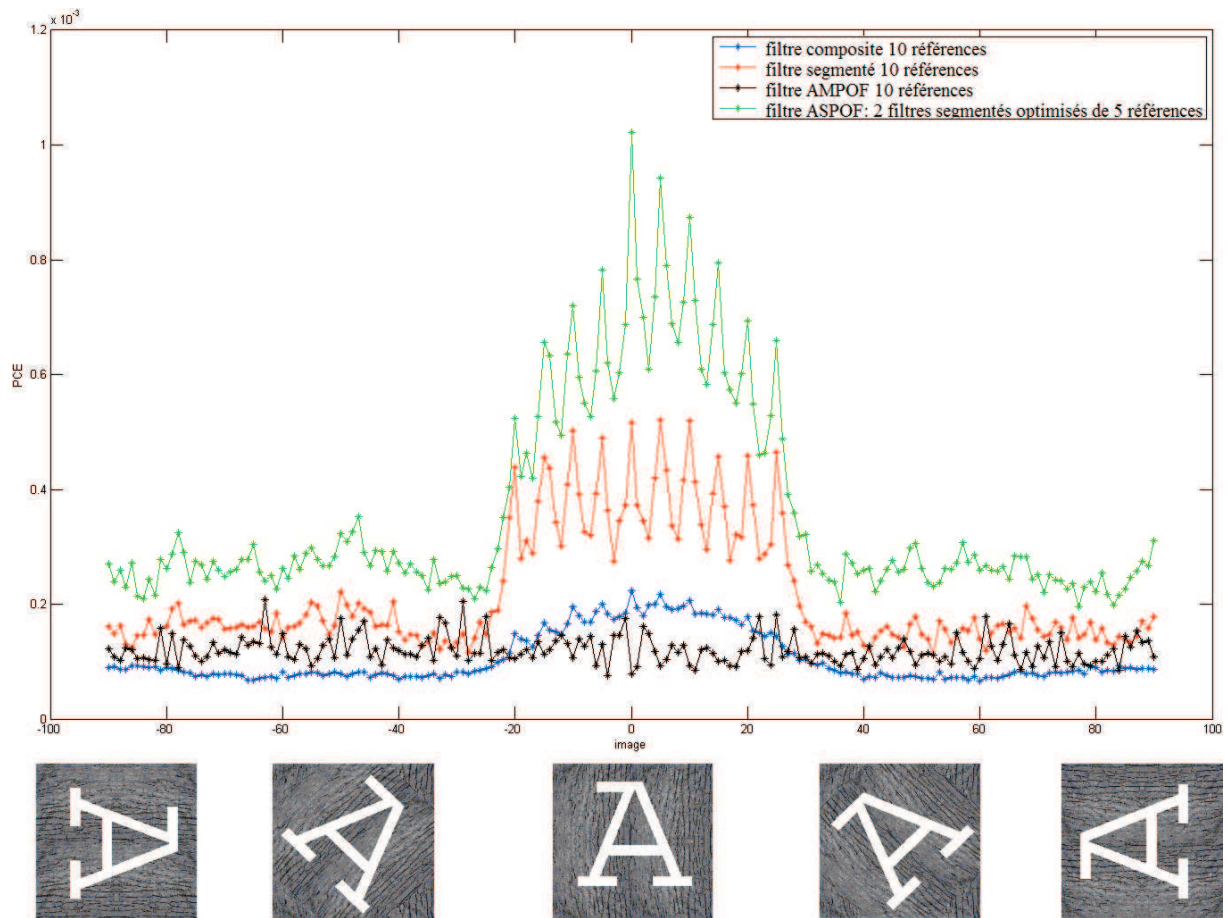


FIGURE A.5 – Comparaison des filtres sur les lettres A perturbées par un bruit texturé

Annexe B

Création des fichiers de navigation pour les vidéos Topvision

Les informations de navigation correspondant aux images Topvision sont affichées sur les images (*cf.* figure B.1). Nous avons à notre disposition la date et l'heure, l'altitude et la profondeur, le cap, le roulis et l'inclinaison. Les améliorations que nous avons apportées à nos algorithmes utilisent ces informations. Nous allons donc les extraire.

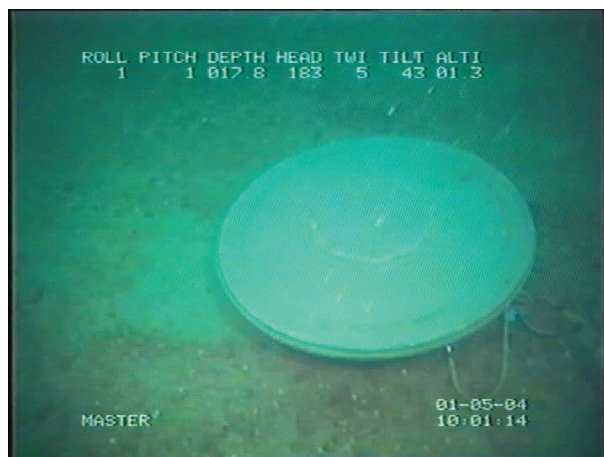


FIGURE B.1 – Image Topvision

Une première méthode consiste à recopier toutes les informations manuellement dans un fichier. Cette méthode est longue et laborieuse. Nous avons préféré nous orienter vers une méthode numérique : la corrélation.

Nous avons donc extrait des images les chiffres de 0 à 9 (*cf.* figure B.2). Nous les avons utilisés comme références pour l'opération de corrélation. Nous avons utilisé des filtres POF.



FIGURE B.2 – Images références

Comme nous connaissons la position exacte, nous avons créé un masque afin de ne garder que cette zone de l'image (*cf.* figure B.3). Ensuite cette image est corrélée avec chaque chiffre. Pour chaque chiffre, nous récupérons les positions et intensité des dix meilleurs pics de corrélation. En effet, si un chiffre apparaît plusieurs fois dans l'image, il y aura autant de pics importants que de fois où apparaît le chiffre étudié (*cf.* figure B.4). Sur la figure B.4a, on voit deux pics se distinguer, correspondant aux 0 présents dans les écritures. Les autres pics correspondent à des parties de chiffres présentes dans les chiffres et dans le zéro. Sur la figure B.4b, on voit un pic

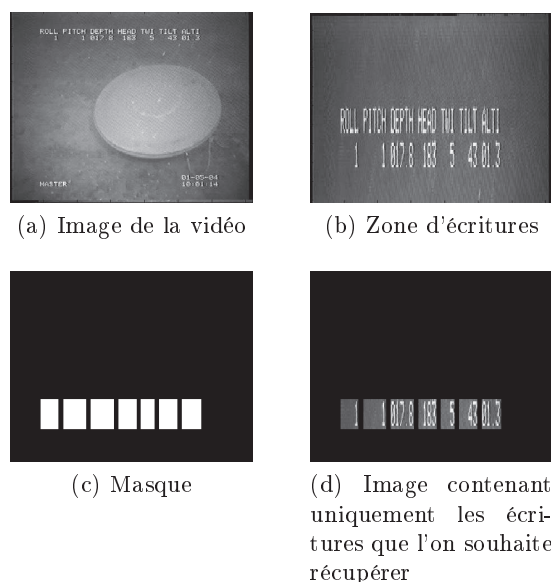


FIGURE B.3 – Les différentes étapes pour extraire les informations souhaitées de l'image

se distinguer, correspondant au 5 présent dans les écritures. Sur la figure B.4c, on voit plusieurs pics qui correspondent aux corrélations entre le chiffre 8 et lui-même mais aussi entre les chiffre 8 et le chiffre 0, très proches.

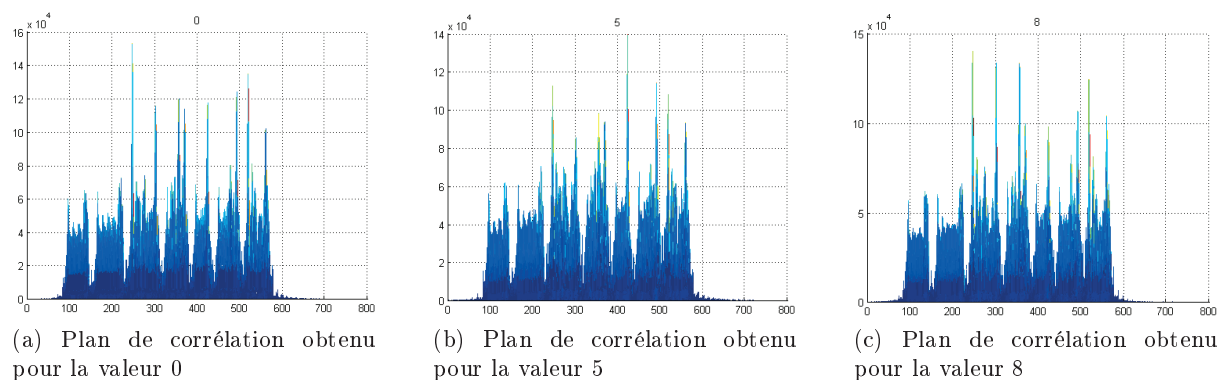


FIGURE B.4 – Plans de corrélation pour les chiffres 0, 5 et 8

Nous obtenons une fois toutes les corrélations effectuées cent pics triés selon la valeur du chiffre référence (*cf.* table B.1). Ils sont triés par valeurs croissantes de leur position horizontale dans l'image, dans la table B.2 (les références sont centrées donc le pic de corrélation indique la position du chiffre dans l'image cible).

numéro de colonne	1	2	3	4	5	6	7	8	9	10	11	12	...	100
chiffre	0	0	0	0	0	0	0	0	0	0	1	1	...	9
position	247	518	492	356	424	302	370	520	560	249	245	260	...	249

Tableau B.1 – Table regroupant les 10 meilleurs pics de corrélation pour chaque chiffre

Ensuite on prend la première position horizontale. On regarde dans notre matrice s'il y a des pics proches (à une distance de 6 pixels, fixées après étude des positions des différents pics). Sur notre exemple, dans les deux premiers cas, nous n'avons qu'un pic notable il s'agit donc de la valeur que l'on enregistre. Pour le troisième cas, il s'agit de comparer les intensités des pics

numéro de colonne	1	2	3	4	5	...	18	19	20	21	22	...	100
chiffre	1	1	1	8	0		9	8	0	9	4	...	3
position	138	220	245	246	247		247	248	249	249	256	...	560

Tableau B.2 – Table classant les pics de corrélation selon leur position

référéncés dans les colonnes 3 à 21. On enregistre les informations contenues dans la colonne référénçant le pic d'intensité maximale parmi les pics étudiés (*cf.* table B.3). Dans l'exemple présenté ici, nous obtenons 1 pour le roulis, 1 pour le tangage (pitch), 0178 pour la profondeur, 183 pour le cap, 5 pour le twi, 43 pour l'inclinaison (tilt), et 0113 pour l'altitude. Sur cette dernière valeur, il y a un chiffre de trop.

numéro de colonne	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
chiffre	1	1	0	1	7	8	1	8	3	5	4	3	0	1	1	3
position	138	220	247	260	275	301	342	355	369	423	478	492	518	524	532	560

Tableau B.3 – Table regroupant les pics de corrélation d'intensité maximale

Nous avons effectué une vérification manuelle, pour compenser les points et les signes “-” que nous n'avons pas inclus dans notre programme. Les informations sont enregistrées dans un fichier excel, dans lequel Matlab ira récupérer les informations que nous souhaitons utiliser.

Annexe C

Détails des résultats de la détection

C.1 Résultats sur la mine sphérique

C.1.1 Méthode basée sur la phase

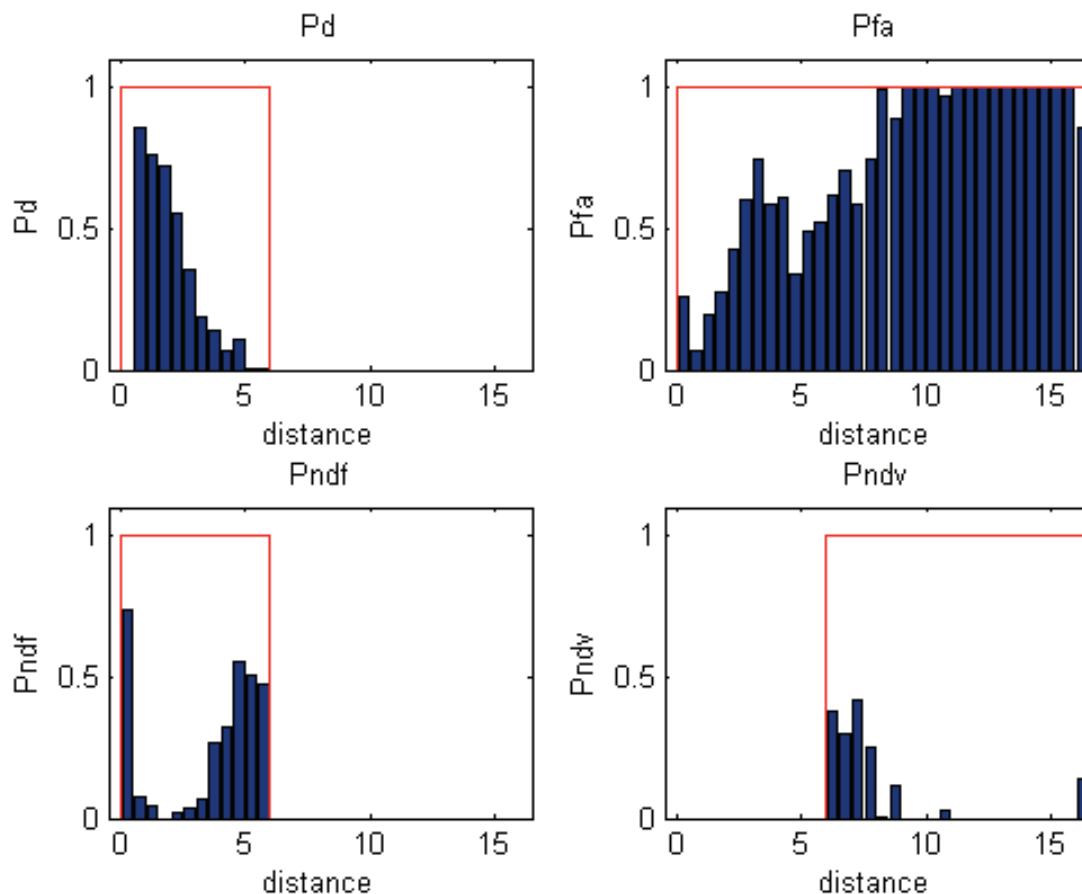


FIGURE C.1 – Résultats obtenus avec la méthode basée sur la phase, développée pour les images Topvision. $Pd=31.45\%$, $Pfa=69.94\%$, $Pndf=25.95\%$ et $Pndv=7.10\%$

Sur la figure C.1 sont représentés les résultats obtenus sous forme de quatre graphes de résultats. Nous avons tout d'abord représenté la probabilité de détection Pd . Le rectangle rouge représente la zone dans laquelle nous devons détecter un objet. La distance maximale de détection a été fixée à 6 m; Nous devons détecter l'objet jusqu'à ce que la distance entre l'objet et la caméra soit nulle. La probabilité maximale dans cette zone est de 1 (c'est à dire que les objets sont correctement détectés dans toutes les images). Ensuite nous avons représenté la probabilité de fausses alarmes, Pfa . Comme l'algorithme basé sur la soustraction du fond permet plusieurs détections par images, il est possible d'obtenir des fausses alarmes tout au long de la vidéo.

Nous avons aussi représenté la probabilité de non détections fausses, $Pndf$. Cette probabilité ne se calcule que dans les images contenant un objet ce qui explique le rectangle rouge entre 0 et 6 m. Enfin il y a la probabilité de non détections vraies, $Pndv$. Cette probabilité se calcule dans les images ne contenant pas d'objets soit les images prises pour une distance objet-caméra supérieure à 6 m.

Le taux de détection augmente lorsque l'on se rapproche de la mine. Cependant la probabilité de détection n'atteint pas 1 et est nulle lorsque la caméra est très près de la mine. Cette probabilité nulle s'explique par l'absence de contours dans les images. En effet, à cette distance, la mine occupe toute l'image. Le taux de fausses alarmes est très élevé. Il y a détection d'un objet dans presque toutes les images, ce qui n'est pas le cas.

C.1.2 Méthode basée sur la soustraction du fond

Nous pouvons remarquer le faible taux de fausses alarmes obtenu dans les deux tests. De plus le taux de détection augmente lorsque la caméra se rapproche de la mine. Ce taux a tendance à diminuer lorsque la caméra est trop proche de la mine. Nous pouvons aussi noter l'amélioration de la détection après l'optimisation de la détection et l'utilisation des données de navigation. Cette amélioration était attendue puisque les optimisations ont été développées à partir de l'analyse des résultats obtenus sur ces vidéos. Le taux de non détections vraies est supérieur à 99%. Comparativement aux résultats obtenus avec la méthode basée sur la phase, il n'y a pas de détection d'objets dans les images vides.

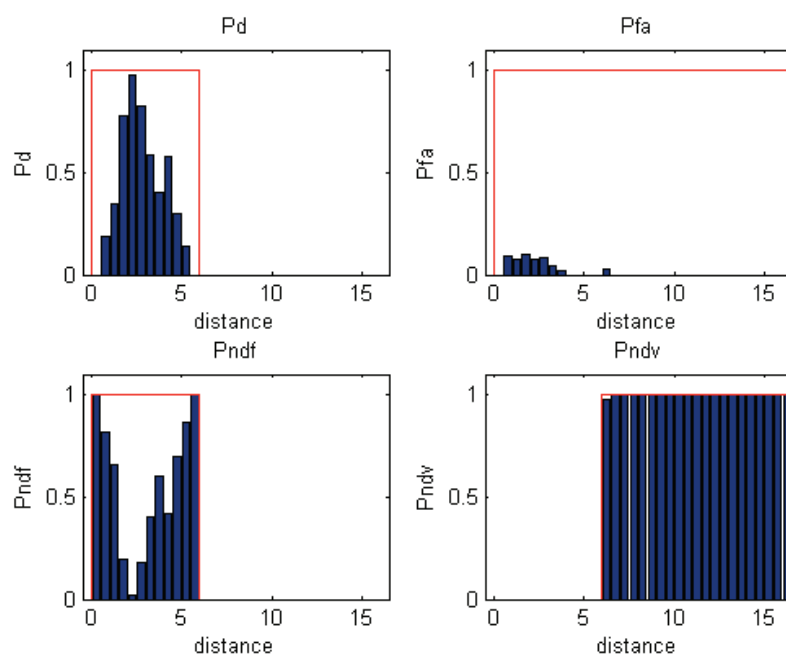


FIGURE C.2 – Résultats bruts de la détection pour la mine sphérique. $Pd=43.23\%$, $Pfa=2.10\%$, $Pndf=56.41\%$ et $Pndv=99.4\%$

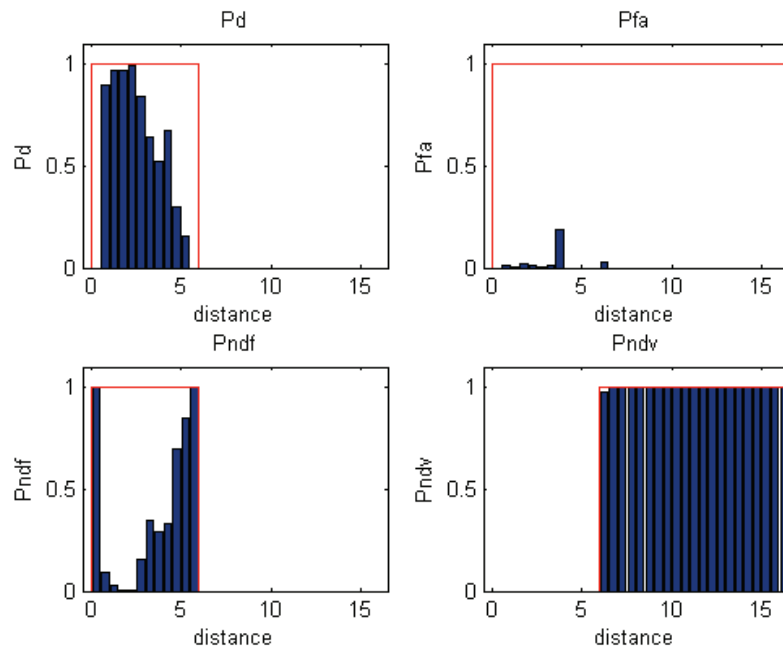


FIGURE C.3 – Résultats optimisés de la détection pour la mine sphérique. $P_d=58.10\%$, $P_{fa}=1.13\%$, $P_{ndf}=39.19\%$ et $P_{ndv}=99.8\%$

C.1.3 Algorithme de Cybernetix

Nous pouvons remarquer comme sur nos résultats que la probabilité de détection augmente lorsque l'on se rapproche de la mine. Notre algorithme basé sur la soustraction du fond optimisé permet d'obtenir un meilleur taux de détection et un taux de fausses alarmes plus faible.

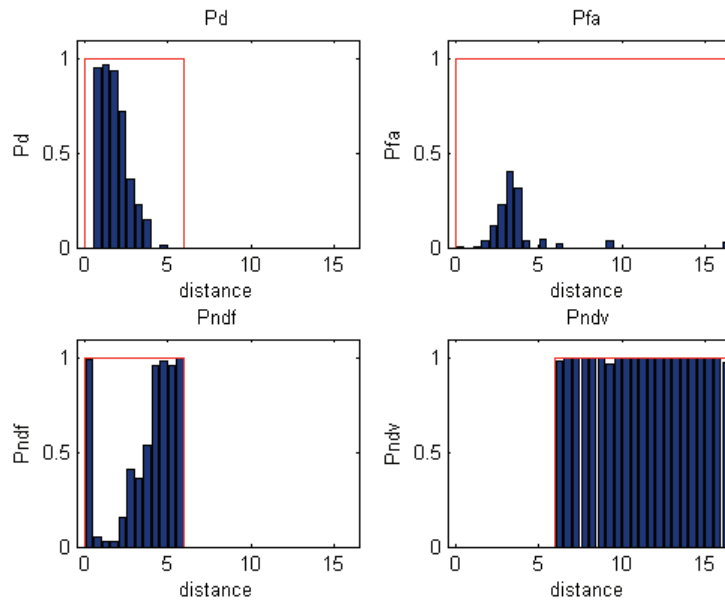


FIGURE C.4 – Résultats Cybernetix sur les vidéos de sphère VAMA. $P_d=36.05\%$, $P_{fa}=3.83\%$, $P_{ndf}=54.04\%$ et $P_{ndv}=99.64\%$

C.2 Résultats sur la mine Manta

C.2.1 Méthode basée sur la phase

Le taux de fausses alarmes est très élevé, surtout à grande distance de la mine où l'algorithme ne devrait rien détecter.

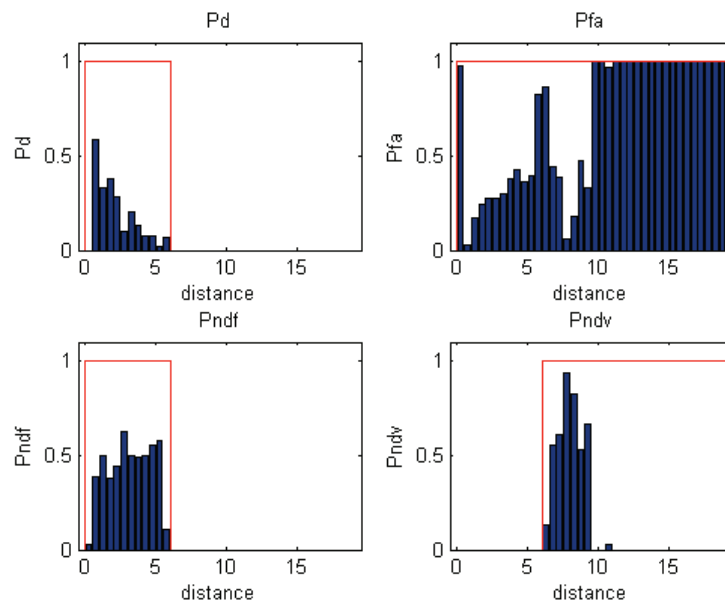


FIGURE C.5 – Résultats obtenus avec l'algorithme basé sur la phase. $P_d=18.87\%$, $P_{fa}=71.64\%$, $P_{ndf}=42.34\%$ et $P_{ndv}=14.74\%$

C.2.2 Méthode basée sur la soustraction du fond

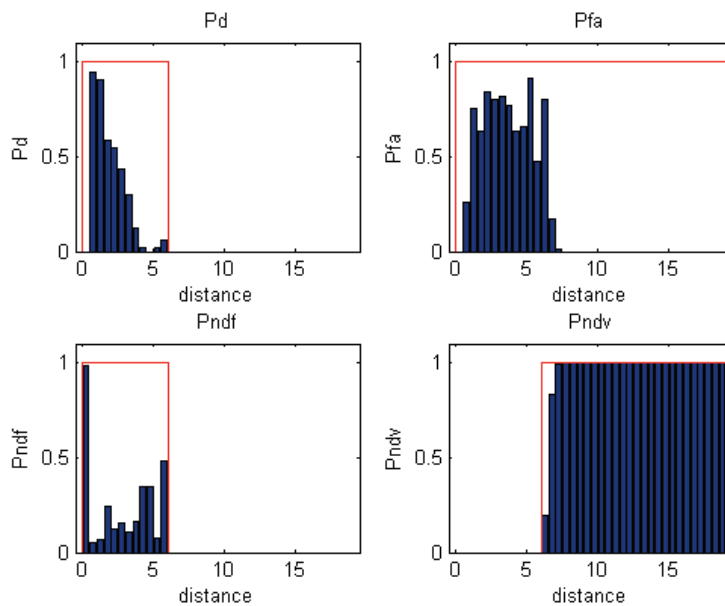


FIGURE C.6 – Résultats bruts de la détection pour la mine Manta. $P_d=32.86\%$, $P_{fa}=20.83\%$, $P_{ndf}=26.19\%$ et $P_{ndv}=96.61\%$

Le taux de fausses alarmes est élevé, à cause entre autres du sol très clair (illuminé par le soleil) comparativement à l'image du fond. Les résultats sont similaires à ceux obtenus avec la mine sphérique par leur forme : le taux de détection augmente lorsque l'on se rapproche de la mine et il n'y a pas de fausses alarmes lorsqu'on est situé très loin de la mine.

Le taux de bonnes détection chute fortement après application des optimisations développées à partir des vidéos de sphères. On peut noter la baisse du taux de fausses alarmes et la hausse du taux de non détections vraies.

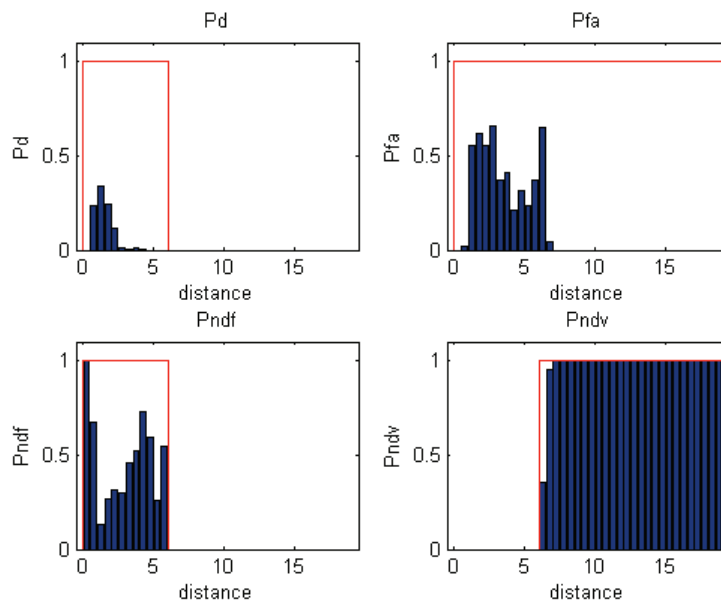


FIGURE C.7 – Résultats obtenus avec l’algorithme optimisé. $Pd=7.98\%$, $Pfa=12.21\%$, $Pndf=48.27\%$ et $Pndv=97.61\%$

C.2.3 Algorithme de Cybernetix

Ces résultats présentent un faible taux de fausses alarmes. Le taux de détection est cependant inférieur à celui obtenu avec l’algorithme basé sur la soustraction du fond, non optimisé.

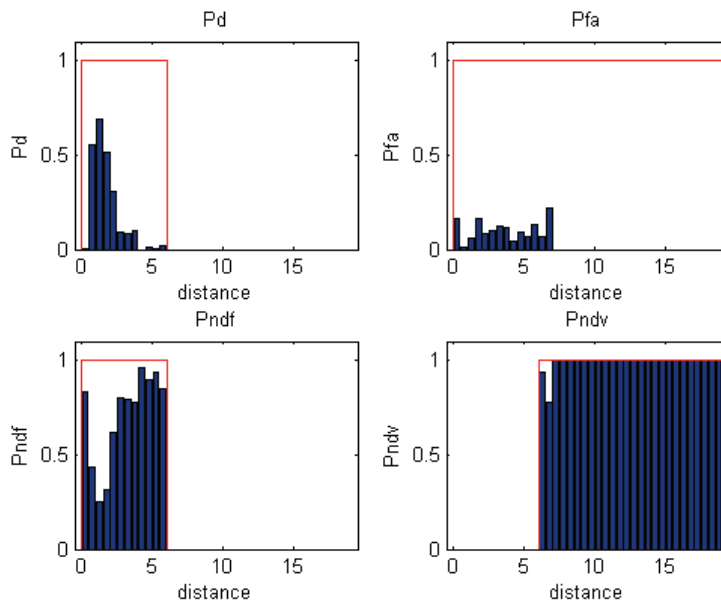


FIGURE C.8 – Résultats Cybernetix sur les vidéos de Manta VAMA. $Pd=19.78\%$, $Pfa=3.54\%$, $Pndf=70.51\%$ et $Pndv=99.02\%$

C.3 Résultats sur la mine cylindrique

C.3.1 Méthode basée sur la phase

Sur ces résultats, nous pouvons remarquer que la détection est présente sur toute la plage de distances où la mine est présente, avec une amélioration lorsque l’on se rapproche de la mine. Le taux de fausses alarmes est dans ce cas aussi relativement élevé.

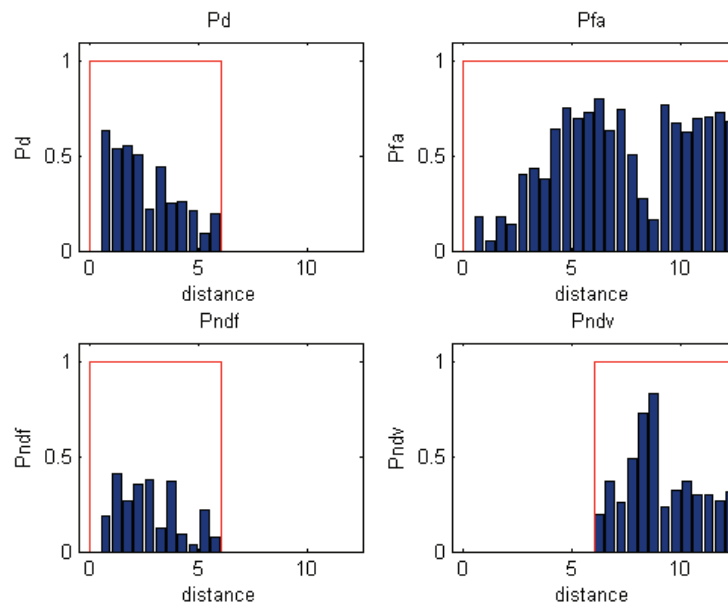


FIGURE C.9 – Résultats obtenus avec l'algorithme basé sur la phase. $Pd=32.42\%$, $Pfa=46.72\%$, $Pndf=20.96\%$ et $Pndv=33.21\%$

C.3.2 Méthode basée sur la soustraction du fond

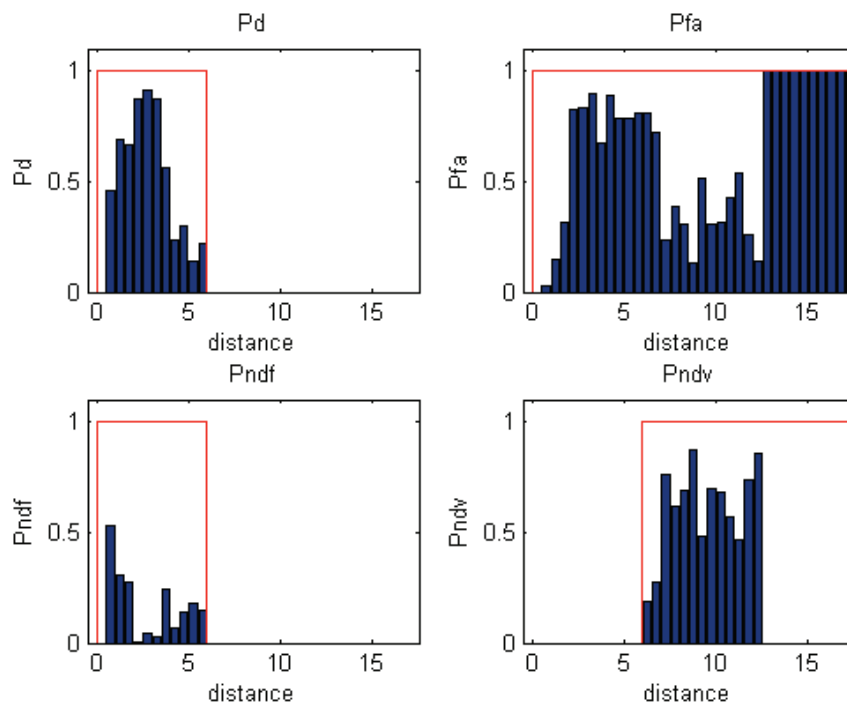


FIGURE C.10 – Résultats bruts de la détection pour la mine cylindrique. $Pd=49.41\%$, $Pfa=62.37\%$, $Pndf=16.48\%$ et $Pndv=31.61\%$

Sur ces figures, nous pouvons constater qu'il y a détection d'un objet dans toutes les images loin de la mine. Le taux de fausses alarmes est aussi très élevé. Ce taux est dû à la différence entre l'image du fond utilisée et l'image cible.

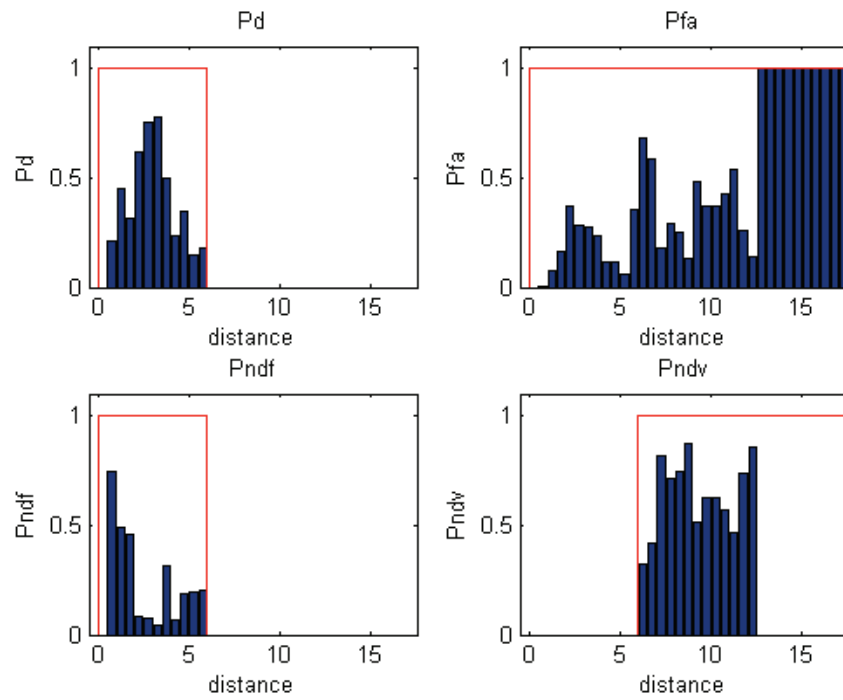


FIGURE C.11 – Résultats obtenus avec l’algorithme optimisé. $Pd=37.69\%$, $Pfa=48\%$, $Pndf=23.87\%$ et $Pndv=33.17\%$

C.3.3 Algorithme de Cybernetix

L’algorithme de Cybernetix permet d’obtenir un taux de fausses alarmes faibles et un taux de non détections vraies élevé, comparativement aux résultats obtenus avec nos algorithmes.

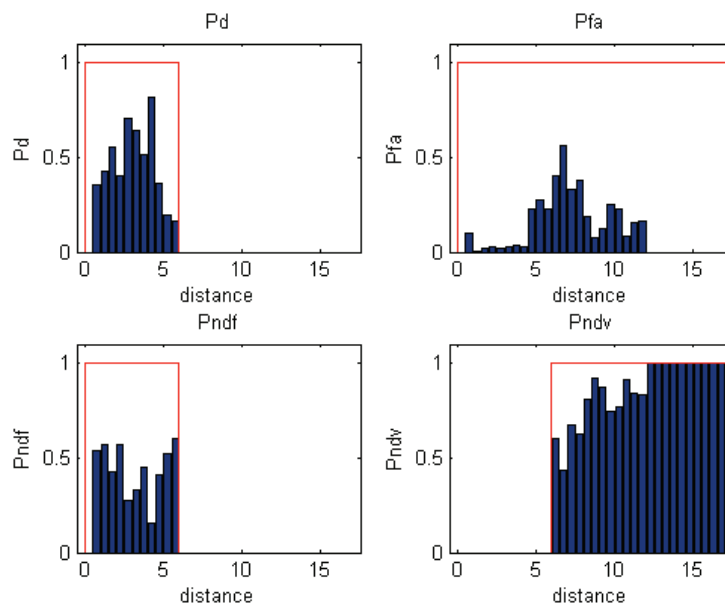


FIGURE C.12 – Résultats Cybernetix sur les vidéos de cylindre VAMA. $Pd=42.91\%$, $Pfa=10.68\%$, $Pndf=40.42\%$ et $Pndv=84.20\%$

C.4 Résultats sur les autres objets

C.4.1 Méthode basée sur la phase

Sur ces images, notre algorithme détecte beaucoup de choses mais rarement ce qui nous intéresse, comme le montrent le taux de détections et le taux de non détections vraies.

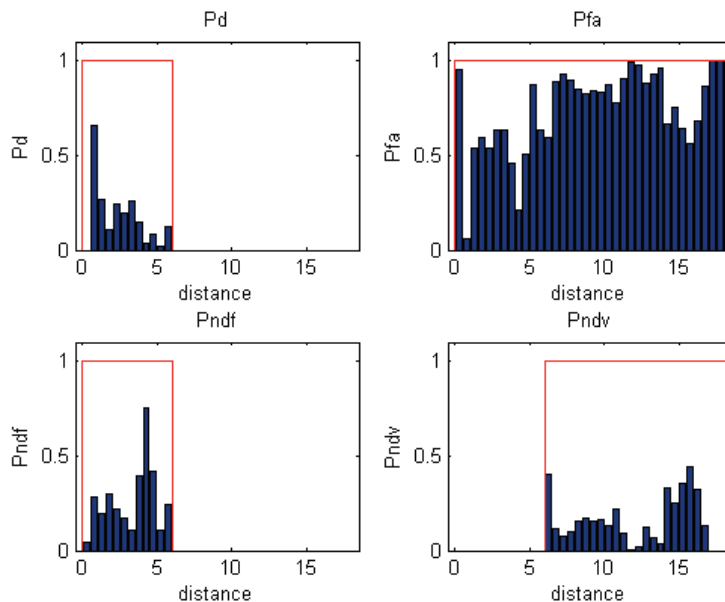


FIGURE C.13 – Résultats obtenus avec l'algorithme basé sur la phase. $Pd=17.92\%$, $Pfa=71.15\%$, $Pndf=26.90\%$ et $Pndv=14.35\%$

C.4.2 Méthode basée sur la soustraction du fond

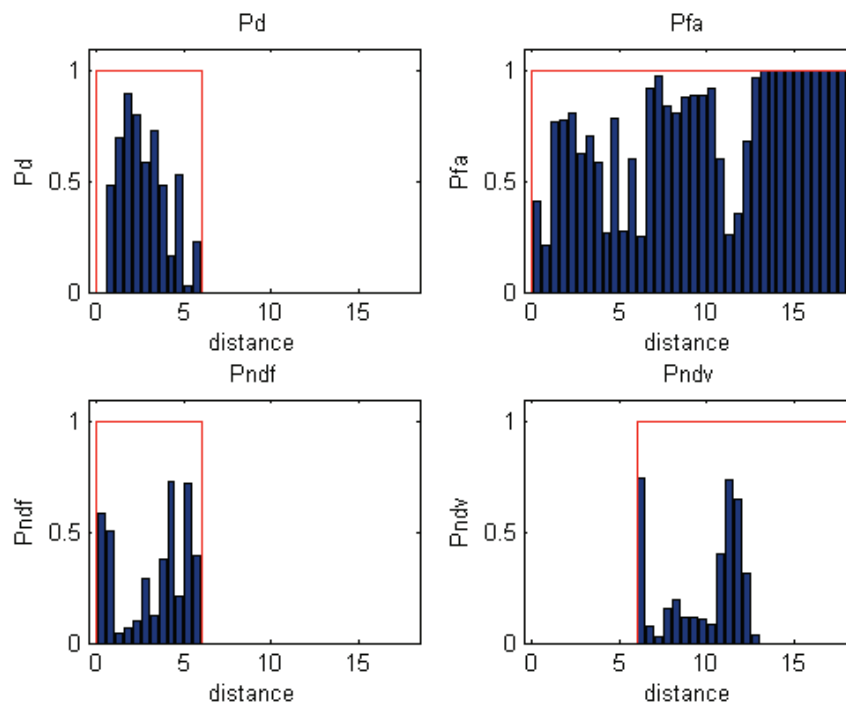


FIGURE C.14 – Résultats bruts de la détection pour les autres objets. $Pd=46.82\%$, $Pfa=71.94\%$, $Pndf=34.72\%$ et $Pndv=13.96\%$

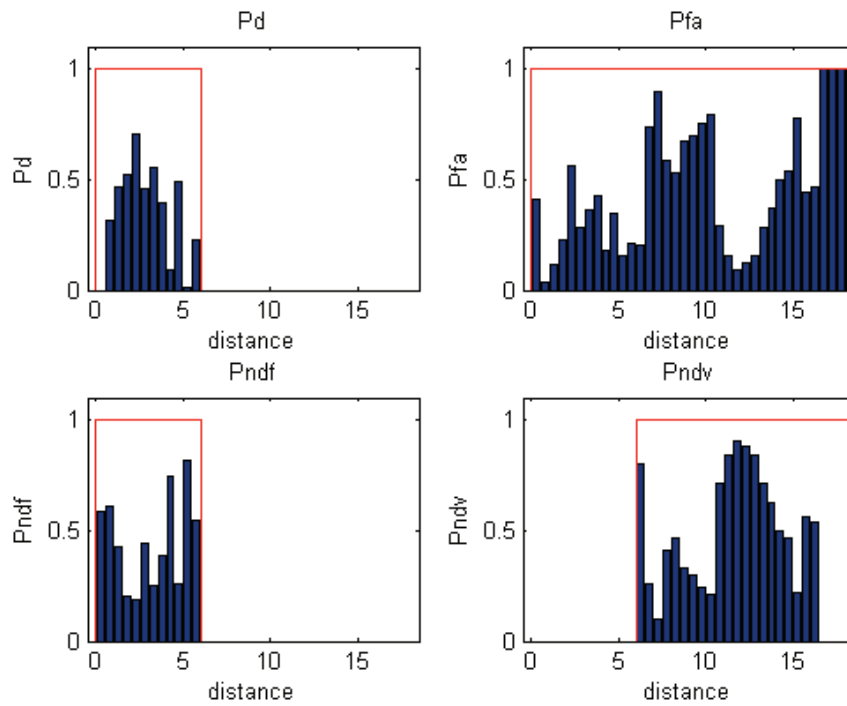


FIGURE C.15 – Résultats obtenus avec l'algorithme optimisé. $Pd=35.35\%$, $Pfa=44.57\%$, $Pndf=45.43\%$ et $Pndv=40.52\%$

Bien que le taux de bonnes détections baisse entre l'algorithme initial et l'algorithme optimisé, le taux de fausses alarmes baisse et le taux de non détections vraies augmente.

C.4.3 Algorithme de Cybernetix

Dans ce cas, l'algorithme de Cybernetix ne détecte pas grand chose avec des taux de détection et de fausses alarmes inférieurs à 20%.

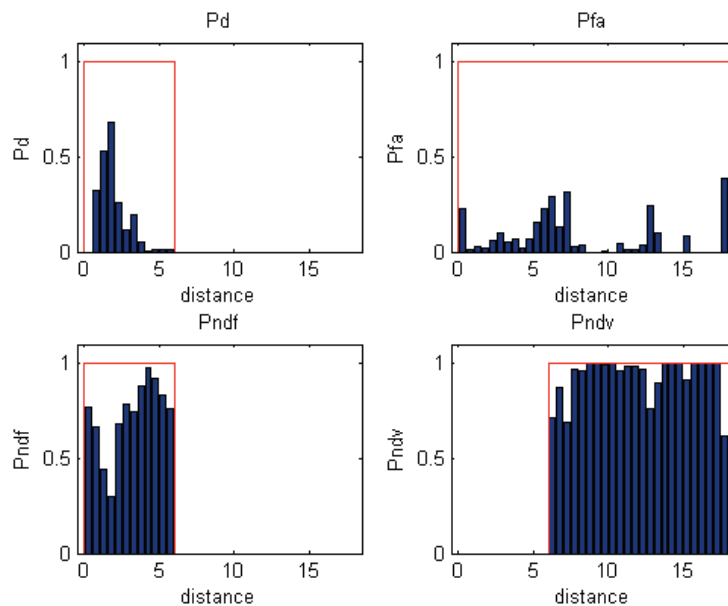


FIGURE C.16 – Résultats Cybernetix sur les vidéos d'objets VAMA. $Pd=18.30\%$, $Pfa=9.63\%$, $Pndf=73.02\%$ et $Pndv=82.55\%$

C.5 Résultats sur les vidéos vides

C.5.1 Méthode basée sur la phase

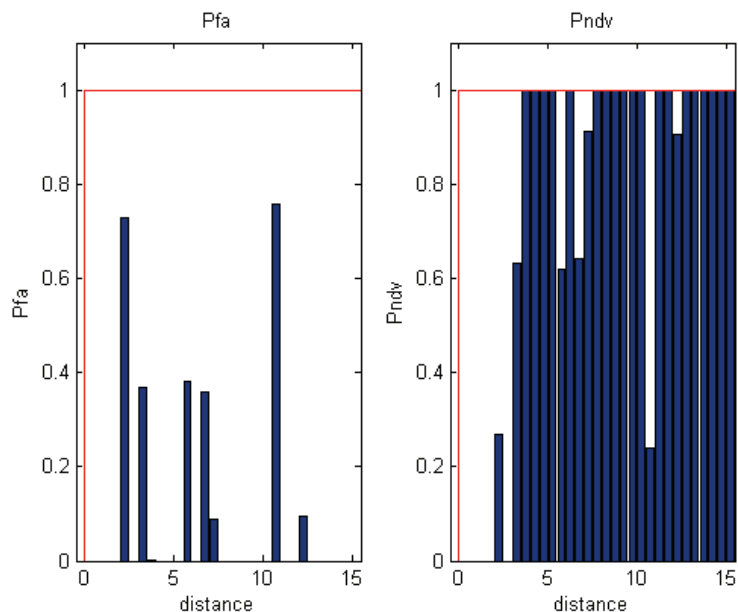


FIGURE C.17 – Résultats obtenus avec l'algorithme basé sur la phase. $Pfa=8.97\%$ et $Pndv=74.90\%$

Les fausses alarmes détectées dans les vidéos vides correspondent en partie à des objets que nous n'avons pas classés dans la catégorie "autres objets". En effet, il s'agit d'objet complètement recouverts par les algues ou d'un bloc de béton.

C.5.2 Méthode basée sur la soustraction du fond

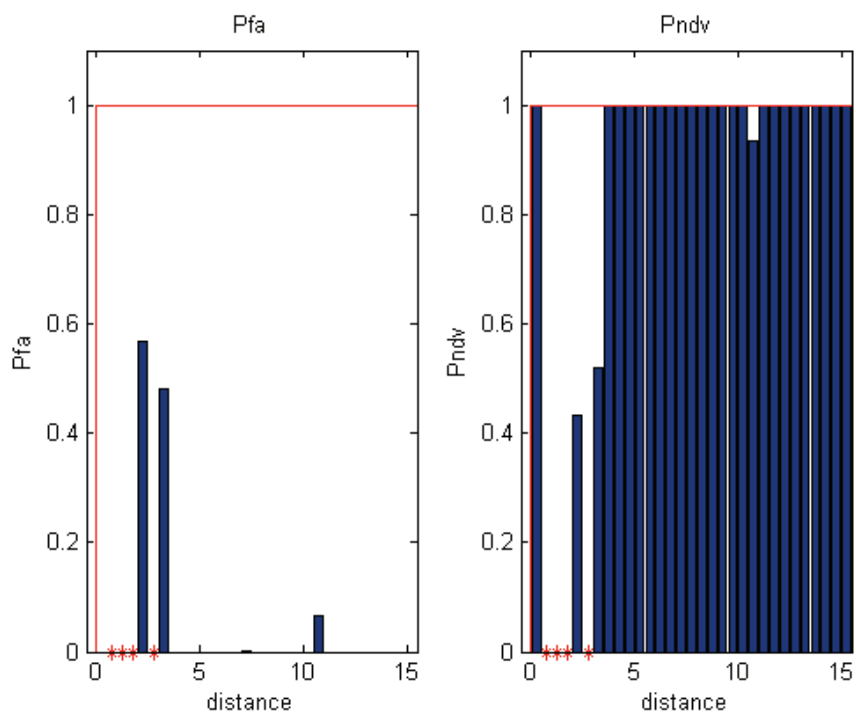


FIGURE C.18 – Résultats bruts de la détection pour les vidéos vides. $Pfa=4.13\%$ et $Pndv=95.87\%$

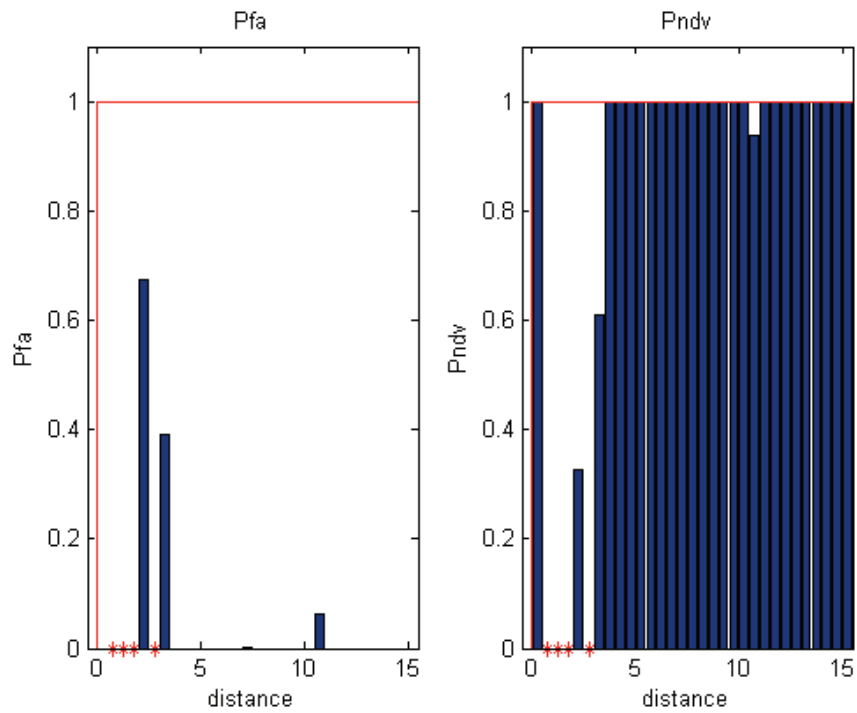


FIGURE C.19 – Résultats optimisés de la détection pour les vidéos vides. $P_{fa}=4.18\%$ et $P_{ndv}=95.82\%$

L'optimisation de l'algorithme n'a pas apporté d'amélioration significative aux résultats. Les astérisques rouges correspondent aux distances pour lesquelles nous n'avons pas d'images.

C.5.3 Algorithme de Cybernetix

L'algorithme de Cybernetix détecte davantage d'objets ou de fausses alarmes dans les vidéos, comparativement à notre algorithme basé sur la soustraction du fond.

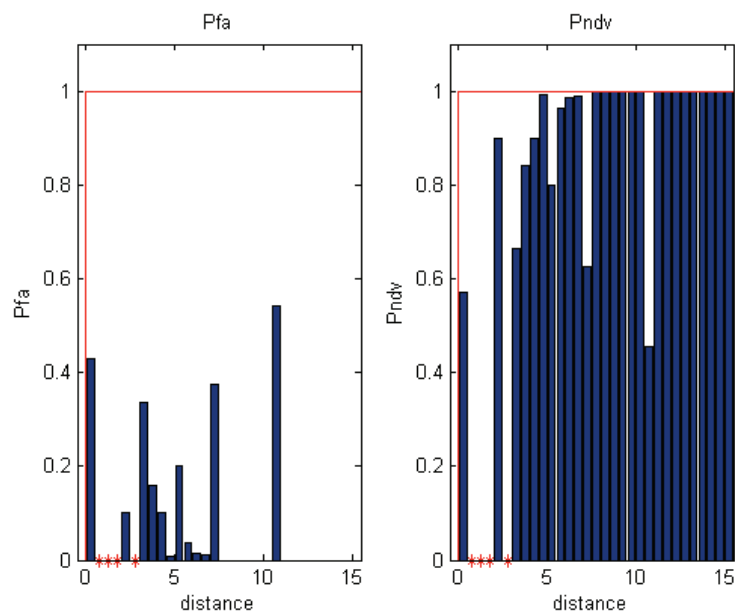


FIGURE C.20 – Résultats Cybernetix sur les vidéos vides VAMA. $P_{fa}=8.57\%$ et $P_{ndv}=91.43\%$

Annexe D

Intersection d'un rayon et d'un triangle

Un rayon peut être représenté par l'équation [132] :

$$P(t) = S + tV \quad (\text{D.1})$$

où S est le point d'origine du rayon et V son vecteur directeur. Le facteur t de l'équation du rayon se calcule avec la formule :

$$t = -\frac{L.S}{L.V} \quad (\text{D.2})$$

Un triangle est défini par ses trois sommets : P_0 , P_1 et P_2 . A l'aide de ces sommets et de la normale N au triangle, on peut calculer un vecteur à 4 dimensions : $L = \langle N, -N.P_0 \rangle$, qui représente le plan contenant le triangle. Si le produit $L.V$ est nul alors il n'y a pas d'intersection.

Le point P , obtenu en intégrant le facteur t dans l'équation, est le point d'intersection entre le rayon et le plan du triangle. Il reste à trouver si ce point se situe à l'intérieur du triangle. Cela s'effectue en calculant les poids de l'équation : $P = w_0P_0 + w_1P_1 + w_2P_2$ où $w_0 + w_1 + w_2 = 1$. w_1 et w_2 sont obtenus avec la formule :

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \frac{1}{Q_1^2Q_2^2 - (Q_1Q_2)^2} \begin{pmatrix} Q_2^2 & -Q_1Q_2 \\ -Q_1Q_2 & Q_1^2 \end{pmatrix} \begin{pmatrix} RQ_1 \\ RQ_2 \end{pmatrix} \quad (\text{D.3})$$

avec $Q_1 = P_1 - P_0$
avec $Q_2 = P_2 - P_0$
avec $R = P - P_0$

Le point P est à l'intérieur du triangle si les conditions suivantes sont vérifiées :

$$\begin{aligned} w_0 &\leq 0 \\ w_1 + w_2 &\leq 1 \end{aligned} \quad (\text{D.4})$$

Annexe E

Caractéristiques de PCs et logiciels Matlab utilisés

E.1 PC ISEN

Les temps de calcul effectués sur Matlab et le PC de l'ISEN correspondent à :

- la version R2007a de Matlab
- à un PC dont le processeur Intel Core 2 Quad CPU a une fréquence maximale de fonctionnement de 2.66 GHz.

E.2 PC Thales

Les temps de calcul effectués sur Matlab et le PC de Thales correspondent à :

- la version R2010b de Matlab
- à un PC dont le processeur Pentium Dual Core CPU a une fréquence maximale de fonctionnement de 2.6 GHz.

Annexe F

Mes publications

Liste de mes publications

1. I. Leonard, A. Arnold-Bos, et A. Alfalou, "Interest of correlation-based automatic target recognition in underwater optical images : theoretical justification and first results," in *Proc. SPIE*, vol. 7678, 2010
2. I. Leonard, A. Alfalou, et M. Alam, "Nonlinear fringe adjusted JTC based face tracking using an adaptive decision criterion," in *Proc. SPIE*, vol. 8134, 2011
3. I. Leonard, A. Alfalou, et C. Brosseau, "Spectral optimized asymmetric segmented phase-only correlation filter," in *Applied Optics*, vol.51, p.2638-2650, 2012
4. I. Leonard, A. Alfalou, et C. Brosseau, *Face recognition : Methods, Applications and Technology*, 2012, ch. Face recognition based on composite correlation filters : Analysis of their performances.
5. I. Leonard, A. Alfalou, M.S. Alam, et A. Arnold-Bos, "Adaptive nonlinear fringe-adjusted joint transform correlator," in *Optical Engineering*, vol. 51, p.098201-1-098201-14, 2012
6. I. Leonard, A. Arnold-Bos, A. Alfalou, et N. Mandelert, "Improvement of automatic man-made object detection in underwater videos by use of navigational information," in ICoURS'12, octobre 2012
7. A. Alfalou, C. Brosseau, B.-E. Benkelfat, S. Qasmi et I. Leonard, "Towards all-numerical implementation of correlation," in *Proc. SPIE*, vol. 8398, 2012

Interest of correlation-based automatic target recognition in underwater optical images: theoretical justification and first results

I. Leonard^a, A. Arnold-Bos^b and A. Alfalou^a

^aISEN, 20 rue Cuirasse Bretagne, CS 42807, 29228 BREST CEDEX 2, FRANCE;

^bTHALES UNDERWATER SYSTEMS, Route de Sainte Anne du Portzic, CS 43814, 29238 BREST CEDEX 3, FRANCE

ABSTRACT

In this paper, we explore the use of optical correlation-based recognition to identify and position underwater man-made objects (*e.g.* mines). Correlation techniques can be defined as a simple comparison between an observed image (image to recognize) and a reference image; they can be achieved extremely fast. The result of this comparison is a more or less intense correlation peak, depending on the resemblance degree between the observed image and a reference image coming from a database. However, to perform a good correlation decision, we should compare our observed image with a huge database of references, covering all the appearances of objects we search. Introducing all the appearances of objects can influence speed and/or recognition quality. To overcome this limitation, we propose to use composite filter techniques, which allow the fusion of several references and drastically reduce the number of needed comparisons to identify observed images. These recent techniques have not yet been exploited in the underwater context. In addition, they allow for integrating some preprocessing directly in the correlation filter manufacturing step to enhance the visibility of objects. Applying all the preprocessing in one step reduces the processing by avoiding unnecessary Fourier transforms and their inverse operation. We want to obtain filters that are independent from all noises and contrast problems found in underwater videos. To achieve this and to create a database containing all scales and viewpoints, we use as references 3D computer-generated images.

Keywords: Automatic target recognition, underwater optical images, fast correlation-based methods, composite filter.

1. INTRODUCTION

In underwater mine warfare, unmanned underwater vehicles (UUVs) are used as a complement to divers, to detect mines and eventually destroy them. Currently, Remotely Operated Vehicles (ROVs) are used. They typically have a sonar and a camera. They require the constant attention of an operator and they have a short range since they are tethered to a mothership. Thus the planned evolution to ROVs are Autonomous Underwater Vehicles (AUVs). In a typical mine hunting scenario, the mothership uses its sonars to study an area in order to detect and classify mines. Once mines have been detected, ROVs are launched to perform an identification and eventually destroy the mine. During the transit to the target, the vehicles are guided by their sonar. When close to the mine, video can be used. Because of the reduced visibility in the underwater medium, the range at which video is usable is limited to a few meters only. Video is used to detect, localize and identify the underwater object. The specificities of the underwater medium such as light absorption and scattering lead to images with noise, no contrast and unusable color information. Depending on the location and the recording environment (like water turbidity), these effects can be more or less accentuated in the images. For these reasons, and because correlation is sensitive to these problems, images need preprocessing. Our studies concern a system that is able to perform detection, localization and identification at same time while preprocessing images. Moreover, we add a constraint on the processing time: we want a real-time process.

Further author information: Send correspondence to I. Leonard
E-mail: isabelle.leonard@isen.fr

Research on pattern recognition has been carried out for various purposes like handwritten, fingerprint or face recognition, to name only a few well-known ones. Recognition is a difficult problem since there are many objects to identify and lots of points of view of them and imaging conditions can modify their appearance. In this article, we compare a few methods, selected according to criteria like the processing time and the number of required operations, the idea being to provide elements justifying the choice behind our own approach.

Our paper is organized as follows. First, we present the setting, the studied targets and the context of underwater optical imaging. Then we expose the state of the art and discuss our choice of using optical correlation-based methods to recognize underwater objects. We explain our integrated preprocessing features to improve target images without increasing processing time, our filter manufacturing and our decision of trying the use of computer-generated images instead of underwater images. Finally we present first results obtained with data acquired at sea that seem to validate our approach.

2. CONTEXT OF STUDY

2.1 Use of video in the underwater environment

Video is used in several applications. Some are briefly developed here.

A first category is visual servoing. For some biological surveys applications, it is required to follow fishes. The fish must be kept in the center of the image recorded by the camera.¹ This application needs to have a robot that is able to adapt to the speed of the fishes and to their orientation changes. Cables and pipes following is another example.²

A last example is station keeping:¹ the robot modifies its position depending on the chosen feature displacement, which is measured in comparison with its initial position.

Target recognition is the second application of video, as in the case with mine warfare, but also the retrieval of lost objects, such as flight recorders.

2.2 Underwater image

In underwater imaging, visibility is a major problem. It varies from about 30 m in very clear water to about 0.75 m in very turbid water like a harbor water.

The received image I_R is the addition of the attenuated image I_A , the forward scattered image I_{FS} and the backscattered component I_{BS} :³

$$I_R = I_A + I_{FS} + I_{BS} \quad (1)$$

The light attenuation increases exponentially with distance (according to the Beer-Lambert law) and limits visibility. The backscattered component corresponds to the light that has been reflected by particles towards the video camera (see Figure 1). This hides objects present in the scene. The forward scattered component is the light reflected by the object that has been diffused by particles on its way back to the camera, which produces blurred images (see Figure 1).

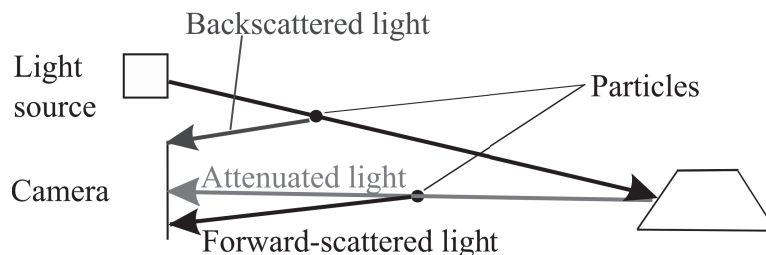


Figure 1. Diagram of the light propagation in an underwater medium

It is possible to add artificial lighting in the scene; this increases the visibility range but it produces a non-uniform lighting. Video detects also particles on which light is reflected, called marine snow.¹

These consequences of the underwater medium are annoying, especially backscattering that is the most important effect. Thus underwater images need preprocessing.

2.3 Studied targets

The studied targets are underwater mines (Figure 2). There are many kinds of mines. In this work we focus on three of them: spherical mines (Figure 2-a), cylindrical mines (Figure 2-b) and Manta mines (Figure 2-c) (the Manta mine is a truncated cone).

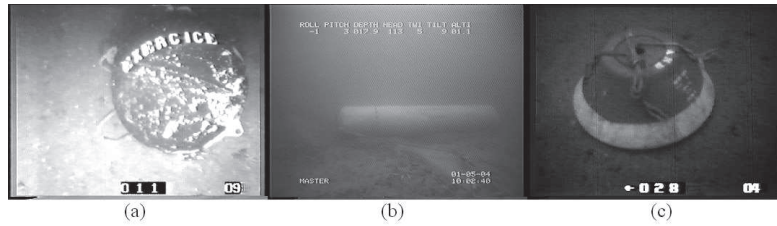


Figure 2. Grayscale images representing three studied targets that come from underwater videos: a spherical mine (a), a cylindrical mine (b) and a Manta mine(c)

3. IMAGE RECOGNITION METHODS: A BRIEF STATE OF ART

Pattern recognition is a problem which has received a great deal of attention in the literature. The purpose of this paper is not to give an exhaustive review of all existing methods. In this paragraph, we expose a few methods selected according to criteria like short time processing and the ability to provide the location and/or the orientation of the mine. First we expose methods that have already been used for underwater target recognition purposes. Then we expose some other methods used in other fields, which could have the potential of being adapted for the underwater environment.

Some methods applied on underwater images are based on contours analysis. Thales's module for underwater mine detection preprocesses images before applying segmentation to extract contours.⁴ These contours are compared with lines and ellipses, to keep only man-made object contours. The center of contours determines the mine location. To identify a mine, they compare the mine shape with reference shapes and they correlate this information with navigational information, in order to obtain a measure of match quality. We do not use this method because it needs contours extraction and processing of these contours: in the underwater environment, the contours may be too faintly visible for the match to be good.

In their article,⁵ Olmos *et al.* suggest and validate a method that consists in distinguishing man-made objects from natural objects in underwater videos. Therefore, they extract edges with a Canny detector and a measure of uniformity which consists in comparing the uniformity of the intensities of each side of the contour. Assuming that man-made objects have straight edges, they use lengths, regularity and the number of found contours to classify natural and man-made objects. As the previous method, this one needs a Canny edge detector. This detector may not be able to find all the interesting contours in very noisy images.

Bazeille *et al.* identify known objects with color analysis.⁶ They base their studies on the Beer-Lambert law and showed that a color has a set of compatible colors in the underwater medium. Their algorithm has a training step to find the compatible color range of a known object (through a logarithmic transform and principal analysis component). Then it analyzes image pixels to search objects. Since we have color and grayscale images, we cannot use exclusively color-based methods.

Underwater imaging is also used for tracking. Trucco et Plakas made a review of some recent papers.² We can note that developed methods are essentially applied to mosaics, cables and pipes following or positioning. These methods are based on Hough transforms, on corner detection, and on correlation-based algorithms. Hough transform-based methods need another processing to analyze results that may complicate a system. Hough transforms are essentially based on contours, which, as we said above, may be too faintly visible in the image. Corner detectors "lock" on any feature of the image where the gradient is strong in any direction. Any feature such as text painted on the object will be caught by the corner detector. Since the presence of such features is unpredictable, it makes corner detectors hard to use for image recognition.

All existing pattern recognition methods have not yet been applied to underwater target recognition. Therefore, we look after some other methods from terrestrial applications.

In a recognition application for vehicles in terrestrial images, Olson *et al.* proposed an approach based on the use of both edge maps of images and references.⁷ They compare the position and the gradient of pixels from the input image to the position and the gradient of pixels from references. The decision is based on a threshold on the number of corresponding pixels. This method has two drawbacks: it needs a good edge detection method and it does comparison on a per-pixel basis.

In another approach, Cole *et al.* proposed a method adapted to Lego pieces recognition.⁸ Therefore, they have a database divided in classes. Each class represents one Lego piece and contains an average image and a cluster image. Each image to identify is compared to average images with a normalized cross correlation and a threshold. Then the image is compared to clusters with which there might be a match to obtain the name of the Lego piece. This method needs two comparisons: the first one with a huge database and the second one with a reduce database that leads to the decision.

In the Ref. 9, Schiele and Crowley use color histogram matching to recognize objects. The first step is to calculate the color image histograms. Then they compare them to reference histograms with a χ^2 test. Different tests⁹ prove that this method is robust to orientation change, scale change, partial occlusion and viewing point change. This method is a simple, robust, and has a short processing time. But it depends on color and light intensity, that are underwater image problems.

SpikeNet is a real-time method, developed by Thorpe *et al.*^{10,11} They have developed a very fast program that simulates neuron network with very few parameters, to recognize faces. The first version of their algorithm has four steps: a step to know at which state the image pixel correspond to (by analogy with the retina which have two states of its cells: on and off), two steps to recognize contour orientations and then specific features that the program had learnt and a final step to analyze results. Following this version, they proposed an optimized version that is able to recognize all kinds of objects if learnt specific features contain oriented structures. This method needs a good contour visibility and objects from which we can extract specific features. In our case, this means specific features for each mine, each orientation and each scale and as a consequence a huge database of reference features.

In the same category, Viola and Jones used the AdaBoost classifier to rapidly detect faces in grayscale images.¹² They select some classifiers that they assemble in a cascade. Thus they achieve a low error rate on their training database. The drawback to this approach is that AdaBoost, like neural-network-based methods, is of the black box kind: it is hard to debug.

4. THEORETICAL JUSTIFICATION: WHY CORRELATION?

4.1 Choice

As shown in part “context of study” (part 2, p. 2), in the water medium, color is attenuated and objects’ color changes on each video. Moreover, in our studies, we have color images as well as grayscale images. Thus color methods⁶ cannot be used. Schiele and Crowley’s method⁹ is not able to provide orientation information. We do not use methods based on edge maps⁷ and/or binary images.⁵ Indeed, noise present in images affects object contours. These are not clear in all images, which can perturb recognition results.

For these reasons, we chose the correlation method. It compares an image to a reference image. The presence of a correlation peak indicates the resemblance and therefore the mine’s orientation with respect to the video camera. The correlation peak location also gives the location of the mine in the image. This method mainly recognizes contours that can be obtained from grayscale images. Recognition can be obtained even if entire contours are not visible in the studied image.

4.2 Correlation

Correlation is a signal and image processing method that compares a target image with a reference image. The result of this comparison is a more or less intense correlation peak, depending on the resemblance degree between these two images. Mathematically, correlation can be written:

$$c(x_0, y_0) = h(x_0, y_0) *^c s(x_0, y_0) \tag{2}$$

$$c(x_0, y_0) = \int \int_{-\infty}^{+\infty} h^*(x, y) s(x + x_0, y + y_0) dx dy \quad (3)$$

where “c” is the result of the correlation operation, “*c” is the correlation product, “*” is the complex conjugate operator, “h” is a filter or reference, “s” is the image to analyze, “(x₀, y₀)” are spatial coordinates and “(x,y)” are integration variables. The correlation operation can be expressed with a Fourier transform:

$$C(\mu, \nu) = H^*(\mu, \nu) \cdot S(\mu, \nu) \quad (4)$$

where μ and ν are the coordinates in the frequency plane, C , S and H^* are the respective Fourier transforms of functions c , s and h^* .

The classical matched filter¹³ has been modified by introducing information in order to obtain some robustness to noise or orientation change for instance.¹⁴ Here we develop only three filters: classical matched filter,¹³ phase only filter (POF)¹⁵ and optimal trade-off filter (OT filter).¹⁶ Other filters have been developed, in the literature, like the binary phase only filter¹⁷ and the inverse filter.¹⁸ These filters, expressed in the Fourier plane, may then replace “ H^* ” in equation 4.

The most known filter is the classical matched filter (F_{CMF}) defined in the Fourier plane as:¹³

$$F_{CMF}(\mu, \nu) = \frac{\alpha S^*(\mu, \nu)}{B(\mu, \nu)} \quad (5)$$

where “ $S^*(\mu, \nu)$ ” denotes the complex conjugate of the reference, “B” the spectral density of the background and “ α ” is a constant. This filter is robust but has a low discriminating power.¹⁴

The phase only filter (F_{POF}) is defined in the Fourier plane as:¹⁵

$$F_{POF}(\mu, \nu) = \frac{S^*(\mu, \nu)}{|S^*(\mu, \nu)|} \quad (6)$$

where “ $|S^*(\mu, \nu)|$ ” is the module of the reference spectrum. This filter gives a sharp correlation peak, it is very discriminative¹⁵ but also noise sensitive.

The optimal trade-off filter (F_{OT}) is defined in the Fourier plane as:¹⁶

$$F_{OT}(\mu, \nu) = \frac{S^*(\mu, \nu)}{\alpha B(\mu, \nu) + (1 - \alpha) |S^*(\mu, \nu)|^2} \quad (7)$$

From now on, in this article, we will use only the POF filter. Since we need a discriminative filter. We don't use the OT filter because in order to obtain good performances we have to know the spectral density of the background which must be very close to the one in the actual images.¹⁹ In underwater applications, it is very difficult to know the spectral density of the background because of the ocean floor, turbidity and all noises that perturb an image and vary throughout the video.

Correlation is extremely fast. The fast Fourier transform takes $O(n \log(n))$ arithmetical operations²⁰ and each image multiplication takes $O(n^2)$ arithmetical operations, where n is one of the two size of the image to analyze (here, we suppose that image is a square image).

To perform a good correlation decision, we should compare our image with a huge database of references, covering all the appearances of mines. Using the filter defined previously, this operation takes a lot of time (proportional to the number of references). To overcome this problem, we use at first the classical composite filter techniques¹⁴ and the POF technique. In future works, we plan to use an optimized version of the composite filter called the segmented composite filter.²¹ The classical composite filter technique used in this work consists in linearly adding several references (Figure 3): first each reference image is Fourier transformed. Then we add all the reference spectra to obtain a composite filter in the Fourier domain.

One problem we can meet when using this technique, is a saturation problem. The values stored in the filter are quantized in a fixed amount of values, *e.g.* 255 for 8-bits images. If there are too many references, this

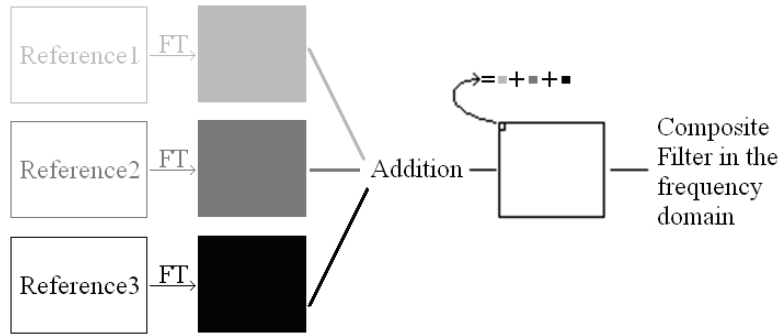


Figure 3. Composite filter technique

maximum value will be reached and adding another reference to the filter will not allow to recognize it in an image. To overcome this problem, we can use the segmented filter technique. In this latest method, the fusion of different references is performed according to a segmentation criterion. This criterion consists in dividing the Fourier plane into separate areas: a first area is common to the different spectra, where the different images have similar spectra (these areas being more or less important depending on the degree of resemblance between the reference images), and other areas are specific to the reference image spectra (for each area, specific information to the reference image spectrum is introduced).

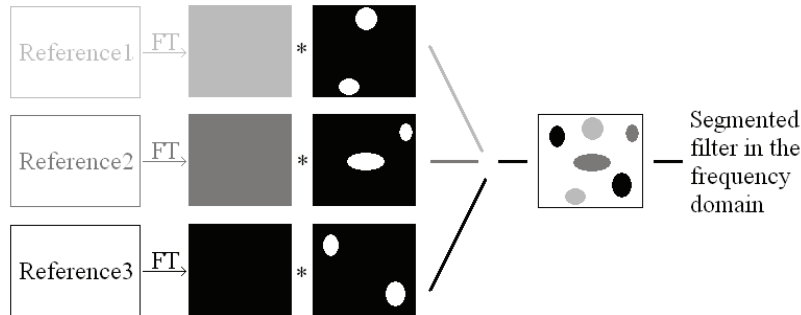


Figure 4. Segmented filter technique

The correlation method has some drawbacks, which can be annoying like the non-robustness to noise, rotation and scale modifications or the fact that it considers only one point of view. Objects can be seen with any point of view and they can have any scale, depending on distance between the object and the video camera. Therefore, we have a huge database of references. Composite filter techniques allow reducing the database size and therefore the number of needed operations. The result of the correlation indicates the resemblance degree between the object and the reference image thus we obtain orientation information. The location of the correlation peak gives the location of the object in the image. The method also allows for inserting information in filters. Indeed, in the algorithm, we multiply the target image spectrum and the filter spectrum in the Fourier domain. The “context of study” section of this paper (part 2, p. 2) shows that underwater images cannot be used in the raw state and need preprocessing. Many of preprocessing we need have to be performed in the Fourier domain. It is possible to integrate them directly into filters and reduce the whole processing time by avoiding several Fourier transforms. Moreover, the correlation method has some advantages like its small computation time and robustness to translation.

Optical correlation is cheap and efficient technique. It is also possible to implement correlation using optical hardware, by using either the Vander Lugt¹³ or the Weaver and Goodman setup.²² A hardware implementation has the advantage of being extremely fast compared to a digital implementation. However, it is less reconfigurable and also more expensive, thus not desirable when used aboard one-shot mine destruction vehicles. As a result, only a digital implementation of correlation has been considered.

5. CORRELATION APPLIED TO UNDERWATER MINE RECOGNITION

Our algorithm is divided in several steps (Figure 4).

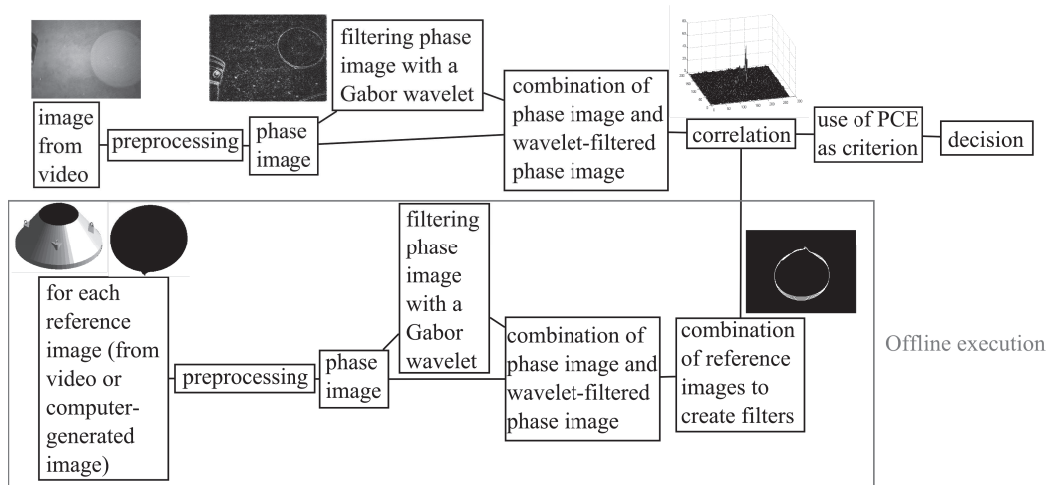


Figure 5. Our algorithm diagram

We have an offline part that consists in preprocessing reference images and compositing them to create filters. The second part of our algorithm consists in preprocessing images. Then we perform correlation by multiplying filters and the preprocessed image. We apply the PCE criterion (part 5.5, p. 9) to decide if searched mines are in the studied image.

5.1 Preprocessing

To save processing time, images are resized to suppress navigational information. To suppress the moiré effect, we suppress the high frequencies corresponding specifically to the moiré pattern in the spectrum as proposed by Sidorov and Kokaram.²³ The moiré effect is due to the use of analog cameras; it can be avoided by the use of digital cameras.

To enhance the contrast of the images, we implement the preprocessing algorithm described in Ref. 24. From the initial image, we apply a low pass filter to create a low pass image. This low pass image will be used twice. Firstly, it will be subtracted to the initial image to suppress backscattered light in the additive model. Secondly, it will divide the initial image in order to suppress the illumination component in the illumination-reflectance model. After clipping values, final image is the result of the division or the result of the subtraction function of the gradient of each image.

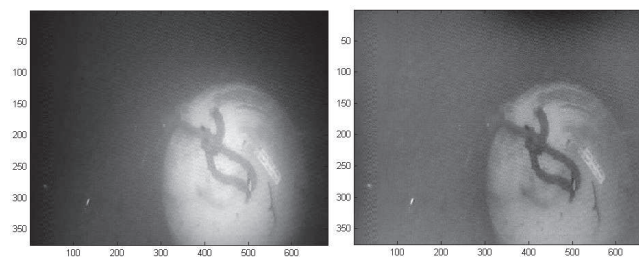


Figure 6. before (left) and after (right) contrast enhancement

5.2 Phase images and filtering with a Gabor wavelet

To efficiently recognize objects, we work on grayscale images where the information is related to contours. However, contrarily to typical contour-based methods, we do not threshold a gradient, but use the phase of the

image. Edges correspond to sharp phase change while information contained by the amplitude are not interesting in our case.²⁵ Using the phase is a way to avoid any thresholding: indeed, thresholding is a problem when the signal to noise ratio for the edge strength is low, as is the case for underwater images. To achieve this, we work on the Fourier domain. It is the processed image, once processed in the Fourier plane, that can be expressed as: $I = r.e^{i\theta}$. The amplitude information is contained by r and the phase information by the exponential function. To keep only the phase information, all values are divided by their amplitude: $I = e^{i\theta}$.

Since there is still some noise, we use a band pass created with a Gabor wavelet: $f(x) = \cos(\frac{2\pi x}{\lambda})e^{-\frac{x^2}{\sigma^2}}$, where “ λ ” is the wavelength and “ σ ” is the standard deviation of the Gaussian exponential function. This operation remove some noise but the edges are not as straight as in the phase image. To solve this problem we combine these two images by addition.

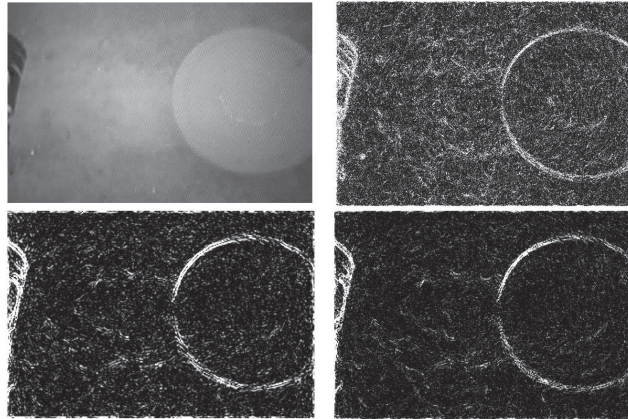


Figure 7. initial image (top left), phase image (top right), phase and wavelet image (bottom left), combination of phase image and wavelet image (bottom right)

On Figure 7, there are images of different steps of our preprocessing. We can notice the apparition of contours in the phase image. The wavelet image filters some noise but edges are not as sharp as in the phase image. The combined image has advantages of the phase image and the wavelet image: we can see quite sharp edges and less important noise.

5.3 Introduction of preprocessing in the filter step

Suppressing the moiré effect, normalizing the amplitude, filtering by a wavelet, is all done in the Fourier domain. All these operations could be done one after the other before computing the Fourier transforms of the images. But, since the Fourier transform is linear, it is much more efficient to do them in one step directly after computing the Fourier transform of the image, which is already needed to do the correlation.

5.4 Sector filter

Composite filter techniques allow for reducing the database size by grouping several reference images in a single filter. However, as we want to know the mine orientation, we use sector filters. This approach means that we divide the reference database into so-called sectors: one sector contains several reference images that represent the mine under orientations that are close to each other, and with one scale. There is one filter per sector. Thus, the presence of the correlation peak will indicate the presence of the mine, its orientation and its approximate location.

To create filters, we automatically assemble close reference images obtained by computing edges of reference images. The criteria used to assemble two references are the size and the orientation of the mine shape. For each image in the reference database, we compare the mine size and its orientation to reference mines that have already been studied. If the values of the criteria are close to the criteria of an already processed image, we add this image to the corresponding filter. If it does not correspond, we create a new filter. This way, it is possible to use a huge reference database. These filters are created offline and once for all.

5.5 Decision with the PCE criterion

To measure the performance of our correlation algorithm, we need a method to evaluate the resemblance. We have chosen the peak to correlation energy (PCE) criterion:

$$\text{PCE} = \frac{\text{correlation peak energy}}{\text{correlation plane energy}} \quad (8)$$

This criterion evaluates energy contained in the correlation peak, in comparison to energy contained in the correlation plane.

There are other existing criteria like the signal to noise ratio or the Horner efficiency.¹⁸

PCE criterion varies depending on each correlation plane. It needs some modifications. Indeed, the PCE criterion is not a normalized criterion. Thus PCE values can strongly vary from one video to another and according to the implemented algorithm. To compare different methods, some normalization is needed. This could be done with the normalized cross correlation that use as normalization tools the local mean and the standard deviation of the image and of each reference image.

6. IMAGES FROM VIDEOS AND COMPUTER GENERATED IMAGES

It is not advisable to use reference images extracted from videos, since those images will be corrupted by noise and not necessarily visible from all orientations and angles. Also, the background is also present, which is a problem when trying to correlate the reference with an observed image. As a result, we decided to use computer-generated images as references. This choice is guided by the fact that computer-generated images are independent from noise and provide us with all scales and all viewpoints we want. Thus, we obtain an extensible database: for each mine we are able to create images for each scale and each point of view.

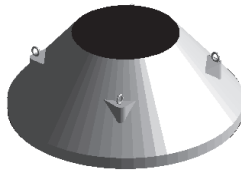


Figure 8. A computer-generated image of a Manta mine

The use of several colors and light gives more details when we perform the contour image. With more details, it is easier to discriminate objects when these details are to be detected in videos. We first create a computer-generated images based database representing the Manta mine. For future works, we will generate database for the spherical and the cylindrical mines.

7. FIRST RESULTS

7.1 Results obtained with filters created with images from videos

As a first, quick experiment, we created filters with images from a video sequence to recognize the Manta mine from image 650 to image 850. On Figure 9, we can notice that filters do indeed recognize these images. Images around image 1700 have PCE values indicating some degree of recognition as well. This result could be expected since the appearance of the mine in this place is close to the appearance of the mine contained in filters. The black line represents the comparison of Manta based filters and a video of a cylinder mine. Obtained values of PCE means that the mine is not recognized as a Manta mine.

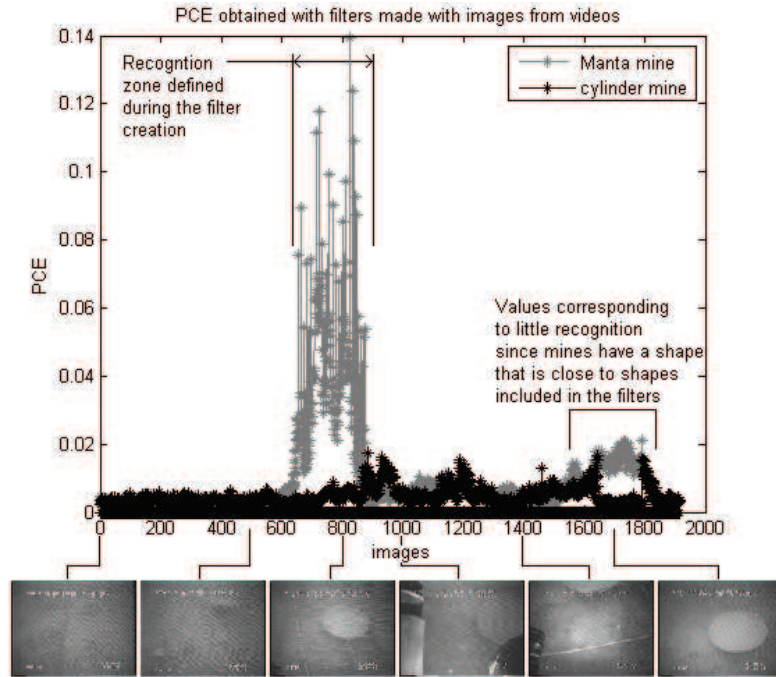


Figure 9. Results obtained with filters created with images from videos

7.2 Results obtained with filters based on computer-generated images

We created five filters with computer-generated images that represent the main mine appearances we can see in the studied video. Each filter contains five reference images. Between two reference images there is an angle difference of 2° . Figure 10 shows that the mine is recognized almost everywhere it is present in the video. However, in images located from 1200 to 1400, there is no mine but we see the video camera and it is recognized as a mine by the system. The cylinder mine is not recognized as a Manta mine. This is explained by the fact that the Manta mine is mainly view as a circle or a cone and there is no view of the circular side of the cylinder in the video used to obtain these results.

7.3 Results

Our algorithm has been run using Matlab, version R2007a. The computer we use is a Intel Core 2 Quad CPU cadenced at 2.66GHz.

Table 1. Results

Method	Video images based sector filters	Computer-generated images based sector filters
Number of filters	5	5
Execution time	0.7s per image	0.4s per image
Recognition rate	50.4%	79.5%
No detection rate	49.6%	20.5%
Error rate	6%	15%
False detection rate	0%	2.9%

The recognition rate and the no detection rate correspond to the number of images containing a mine that have been correctly or not identified in comparison to the total number of images containing a mine. The error rate corresponds to images containing another type of mine identified as a manta mine in comparison to the

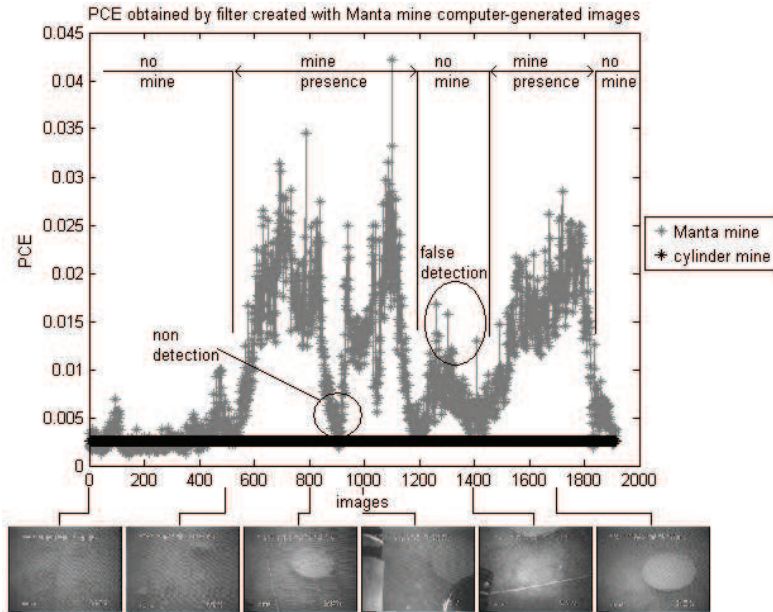


Figure 10. Results obtained with filters created with computer-generated images

number of studied images. The false detection rate corresponds to images without mines that have been identified as an image containing a mine, in comparison to the number of images that does not contain a mine.

In these examples, the decision threshold was manually set at 0.01 to achieve a good tradeoff between the recognition rate and the false alarm rate. In Table 1, the error rate has been obtained using a video where the circular side of the cylindrical mine is presented. That is not the case of the video used to record the Figures 9 and 10. These results are encouraging: they validate the use of the correlation method and composite filter techniques to automatic target recognition in the underwater medium. Moreover, the use of computer-generated images to create filters provides good recognition rates. Error rates are low since there are few images representing the spherical side of the cylinder mine. Seen from some point of views, spherical mines, Manta mines and cylindrical mines have almost identical appearances and this can produce errors.

8. CONCLUSION

Our aim is to realize an algorithm that is able to detect, identify and locate underwater mines with a short processing time. Orientation information is important too. With these criteria, we chose the correlation method, in comparison with other methods selecting on these criterion.

To obtain a short processing time we have integrated preprocessing that need a Fourier transform in the filter step. This integration avoids several Fourier transforms and their inverse. To correlate the images with a huge database in a short execution time, we use composite filter techniques which drastically reduce the size of the database. First results seem to validate the use of this technique. Moreover, the use of computer-generated images to create filters is validated by recognition of the mine in videos.

To optimize our filters and recognition, we have several opportunities like the use of a segmentation criterion to composite our filters instead of adding reference images. We will look after a decision criterion since the PCE criterion provides us a decision but without normalization, we cannot compare different methods.

ACKNOWLEDGMENTS

The raw mine images contained in this publication are derived from data property of the French State that have been provided by the GESMA (Groupe d'Etudes Sous-Marines de l'Atlantique) within TOPVISION project coordinated by Thales Underwater Systems SAS. This project is related to Techno-Vision Programme launched

by French Ministry of Research and French Ministry of Defense. More detailed information can be found on <http://topvision.gesma.fr> site.

This work is supported in part by the Regional Council of Brittany.

REFERENCES

1. R. Marks, M. Lee, and S. Rock, "Using visual sensing for control of an underwater robotic vehicle," in *Proceedings of 1ARP second workshop on mobile robots for subsea environments*, 1994.
2. E. Trucco and K. Plakas, "Video tracking: a concise survey," *IEEE Journal of Oceanic Engineering* **31**(2), pp. 520–529, 2006.
3. B. McGlamery, "A computer model for underwater camera systems," in *SPIE Ocean Optics VI*, **28**, 1979.
4. A. Arnold-Bos and N. Mandelert, "Sonar and video perception for an autonomous mine disposal vehicle," in *UDT Europe*, 2008.
5. A. Olmos and E. Trucco, "Detecting man-made objects in unconstrained subsea videos," in *BMVC*, 2002.
6. S. Bazeille, I. Quidu, and L. Jaulin, "Identification of underwater man made object using color," in *Proceedings of the Institute of Acoustics 29*, 2007.
7. C. Olson and D. Huttenlocher, "Automatic target recognition by matching oriented edge pixels," *IEEE Transactions on image processing* **6**(1), pp. 103–113, 1997.
8. L. Cole, D. Austin, and L. Cole, "Visual object recognition using template matching," in *Australian conference on robotics and automation*, 2004.
9. B. Schiele and J. Crowley, "Object recognition using multidimensional receptive field histograms and its robustness to view point changes," in *1995 Rosenon workshop on computational vision*, 1995.
10. S. Thorpe, "Ultra-rapid scene categorization with a wave of spikes," in *BMVC*, 2002.
11. A. Delorme, J. Gautrais, R. Van Rullen, and S. Thorpe, "Spikenet: a simulator for modeling large networks of integrated and fire neurons," *Neurocomputing* **26-27**, pp. 989–996, 1999.
12. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings IEEE Conference on computer vision and pattern recognition*, 2001.
13. A. Vander Lugt, "Signal detection by complex filtering," *IEEE Transactions on information theory*, pp. 139–145, 1964.
14. B. Vijaya Kumar, "Tutorial survey of composite filter designs for optical correlator," *Applied Optics* **31**, pp. 4773–4801, 1992.
15. J. Horner and P. Gianino, "Phase-only matched filter," *Applied Optics* **23**(6), pp. 812–816, 1984.
16. P. Refregier, "Optimal trade-off filters for noise robustness, sharpness of the correlation peak and horner efficiency," *Optics Letters* **16**(11), pp. 829–831, 1991.
17. J. Horner, B. Javidi, and J. Wang, "Analysis of the binary phase-only filter," *Optics Communications* **91**, pp. 189–192, 1992.
18. B. Vijaya Kumar and L. Hassebrook, "Performance measures for correlation filters," *Applied Optics* **29**(20), pp. 2997–3006, 1990.
19. A. Grunnet Jepsen, S. Tonda, and V. Laude, "Convolution-kernel-based optimal trade-off filters for optical pattern recognition," *Applied Optics* **35**(20), pp. 3874–3879, 1996.
20. J. Cooley and J. Tukey, "An algorithm for the machine calculation of complex fourier series," *Mathematics of computation* **19**(90), pp. 297–301, 1965.
21. A. Alfalou, G. Keryer, and J. de Bougrenet, "Optical implementation of segmented composite filtering," *Applied Optics* **38**(29), pp. 6129–6135, 1999.
22. C. Weaver and J. Goodman, "A technique for optically conciolving two functions," *Applied Optics* **5**(7), pp. 1248–1249, 1966.
23. D. Sidorov and A. Kokaram, "Suppression of moiré patterns via spectral analysis," in *Proceedings of SPIE in visual communications and image processing*, 2002.
24. A. Arnold-Bos, J. Malkasse, and G. Kervern, "Towards a model-free denoising of underwater optical images," in *Proceedings of the IEEE conference on Ocean (Europe)*, 2005.
25. A. Oppenheim and J. Lim, "The importance of phase in signals," in *Proceedings of the IEEE*, **69**(5), pp. 529–541, 1981.

Nonlinear Fringe-adjusted JTC Based Face Tracking Using an Adaptive Decision Criterion

I. Leonard^a, A. Alfalou^{a*}, M. S. Alam^b

^a ISEN Brest, Département Vision, L@bISEN, 20 rue Cuirassé Bretagne, CS 42807, 29228 Brest Cedex 2, France

^b Department of Electrical and Computer Engineering, EEB 75, University of South Alabama 6001 USA South Dr., Mobile, AL 36688-0002

ABSTRACT

In this paper, we propose a new technique for rotation invariant recognition and tracking of the face of a target person in a given scene. We propose an optimized method for face tracking based on the Fringe-adjusted JTC architecture. To validate our approach, we used the PHPID data base containing 1 faces with various in-plane rotations. To enhance the robustness of the proposed method, we used a three-step optimization technique by: (1) utilizing the fringe-adjusted filter (H_{FAF}) in the Fourier plane, (2) adding nonlinearity in the Fourier plane after applying the H_{FAF} filter, and (3) using a new decision criterion in the correlation plane by considering the correlation peak energy and five largest peaks outside the highest correlation peak. Several tests were made to reduce the number of reference images needed for fast tracking while ensuring robust discrimination and efficient of the desired target.

Keywords: correlation, nonlinear fringe-adjusted JTC, face recognition.

1. INTRODUCTION

Correlation is a robust and efficient pattern recognition technique widely studied in the literature [1]. Correlation based discrimination measures the degree of similarity between a target image (image to be recognized) and a reference image (belonging to a known database). Two families of architectures are used by most researchers to implement the correlation technique: (1) the VanderLugt Correlator (VLC) [2] and (2) the Joint Transform Correlator (JTC) [3].

In this paper, we utilize the JTC-based correlation architecture for the proposed face tracking system as it is more suitable for near real time operation. In addition, the proposed system must be robust to changes in in-plane rotation of the face while ensuring accurate discrimination between multiple targets. Our objective is to track the face of person in a given scene.

To enhance the performance of the proposed tracking system, we used an optimized version of the JTC architecture, namely the fringe-adjusted JTC [4]. In addition, nonlinearity is introduced in the Fourier plane of the fringe-adjusted JTC correlator. Special attention is given to establishing an optimized correlation based decision criterion an adaptive version of the enhanced peak-to-correlation energy is proposed and validated. Finally, test results are presented to verify the effectiveness of the proposed technique.

1. ANALYTICAL MODELING

The JTC correlation architecture has been described in detail in the literature, but we briefly review the principle of JTC

--

* ayman.al-falou@isen.fr

architecture in this paper. For more information, we invite the reader to consult the many works proposed and validated in this area by several researchers and groups working in this field, such as [1-4].

In a JTC, the known reference image " r " and the unknown input scene " i " (image to be recognized) are displayed side-by-side in the input plane of the correlator. This plane can be represented mathematically by function " f " described in Eq. (1). To simplify its mathematical representation, we consider only a shift distance " x_0 " along the x axis between the target image and the reference image.

$$f(x, y) = i(x + x_0, y) + r(x - x_0, y) \quad (1)$$

After applying Fourier transform to Eq. (1) i.e., the input plane, we can record the intensity or the joint power spectrum (JPS) using a square law device such as a CCD detector array. In a classical JTC [3], an inverse Fourier transform of the JPS yields the correlation output in the output plane. This correlation output contains two important items: (1) a central autocorrelation peak, and (2) two crosscorrelation peaks corresponding to the correlation between image " i " and image reference " r ". These two crosscorrelation peaks are located on either side of the central zero-order peak at a distance proportional to " x_0 ".

To optimize this basic architecture, Alam et al. [4] proposed to introduce a real-valued filter, fringe-adjusted filter (H_{FAF}), in the Fourier plane of the correlator, defined as

$$H_{FAF}(\mu, \nu) = \frac{B(\mu, \nu)}{A(\mu, \nu) + |R(\mu, \nu)|^2} \quad (2)$$

where " R " denotes the intensity corresponding to the Fourier transform of the reference image, " B " is a constant or a function used to control the gain, and " A " is a function used for avoiding the pole problem and/or reducing the effects of noise. Before discussing the results obtained with this fringe-adjusted JTC architecture, we present the application considered in this paper. Our application consists in recognizing and tracking a face that changes according to different rotation angles as shown in Fig. 1. The system proposed in this paper should allow robust tracking and discrimination of a face turning in different rotation angles ranging from -90° to $+90^\circ$.

Figure 2 presents the results obtained with fringe-adjusted JTC using a single reference image to construct the filter H_{FAF} (the face of the subject presented in Fig. 1 without rotation). The reference image is introduced on one side of the input plane. The background is defined as information outside a square around the face in the reference image considered. Then we display the target image on the other side of the input plane. The target faces are obtained from the database presented in Fig. 1, one at a time starting from -90° to $+90^\circ$ in a sequential manner.



Figure 1. Examples of facial rotation (-90° to $+90^\circ$). Subject number (1), Base PHPID [6].

We initialize the system with a known reference image, i.e., the subject's face (number 1) to be tracked so that initial position of subject's face is known. By comparing this reference face with the target image, we ascertain the presence or absence of the subject (number 1) in the target image and the position of the subject's face in the target image. Thereafter, only the information around the position of the face is selected in this target image.

To do this, we multiply the target image with a filter equal to "1" around the position of the face found and "0" elsewhere. Afterwards, we introduce the preprocessed image as the reference image in the fringe-adjusted JTC input plane. Then a new target image (subject number 2) is introduced in the input plane and the process is repeated until all images from the database are processed.

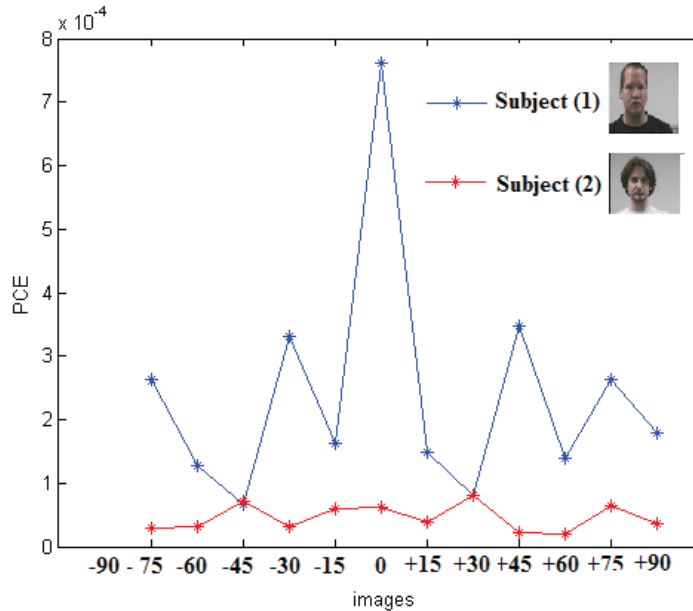


Figure 2. Results obtained with a classical FJTC using only one reference image position without any rotation in the H_{FAF} Filter to define the background noise.

With the various tests performed by using this algorithm, we have identified a tolerance rotation angle equal to 15° between the target image and reference image in which the correlation remains robust. Beyond this value between the reference and target images, the correlation becomes less robust and may lose track of the subject's face. With a tolerance angle equal to 15° , the tracking problem of a person with in-plane rotation (between -90 and $+90^\circ$), consists of correlating "13" pairs of images: $(-90, -75)$, $(-75, -60)$, ..., $(-15, -0)$, $(0, +15)$... $(+60, +75)$, and $(+75, +90)$.

The tracking results of subject number (1) are shown in Figure (2) (blue curve). The PCE criterion was used for decision making in the correlation plane [7]. Figure 2 also shows the results when subject number 2 i.e., undesirable subject is present in the input scene (false alarm: red curve). The results depicted in Fig. 2 show that the proposed fringe-adjusted JTC yields good results. However, for some cases, the values of the two cases are too close for a making robust system.

To optimize the performance of the fringe-adjusted JTC system, we propose to utilize an adaptive fringe-adjusted filter (H_{FAF}) in the Fourier plane [8]. In this technique, the background noise is calculated for each reference. For this purpose and after finding the position of head in the target image at time t_0 , we multiply the target image with a filter equal which has a value of 0 around the position of the face and "1" elsewhere. One can also replace the 0 value with the average of the values around the 0 value zone.

The target image obtained at time t_0 , becomes the reference image at time t_1 in the proposed tracking system. Thereafter, we introduce the preprocessed image as the background noise in the fringe-adjusted filter formulation. The results obtained after the first iteration is shown in Fig. 3 which clearly shows that the adaptive fringe-adjusted JTC improved the discrimination performance when compared to the first case.

To further increase the performance of the proposed system, we propose to incorporate a decision criterion in the correlation plane. Moreover, we applied the k -th power nonlinearity to the fringe-adjusted JPS, where k is a real and positive number. The optimization procedure is described in the following section.

2. TEST RESULTS

To increase the performance of our system we carried out two other optimizations:

- 1) Enhanced PCE decision criterion (P): we propose to take into account the correlation peak energy as well as the five highest peaks excluding the highest correlation peak. The Enhanced PCE decision (P) criterion is defined as

$$P = \frac{\left(\sum_{i=i_0-t}^{i=i_0+t} \sum_{j=j_0-t}^{j=j_0+t} |C(i,j)|^2 \right) - \left(\sum_{k=1}^5 \left(\sum_{i=i_k-t}^{i=i_k+t} \sum_{j=j_k-t}^{j=j_k+t} |C(i,j)|^2 \right) / 5 \right)}{\sum_{i=1}^{i=N} \sum_{j=1}^{j=M} |C(i,j)|^2} \quad (3)$$

where (i_0, j_0) denote the position of the correlation peak, $C(i, j)$ denotes the value of the correlation plane at point (i, j) , t is set to the number of neighboring pixels used around the considered peak, (i_k, j_k) are the positions of five highest peaks excluding the correlation peak, and (N, M) denote the size (in pixels) of the considered correlation plane.

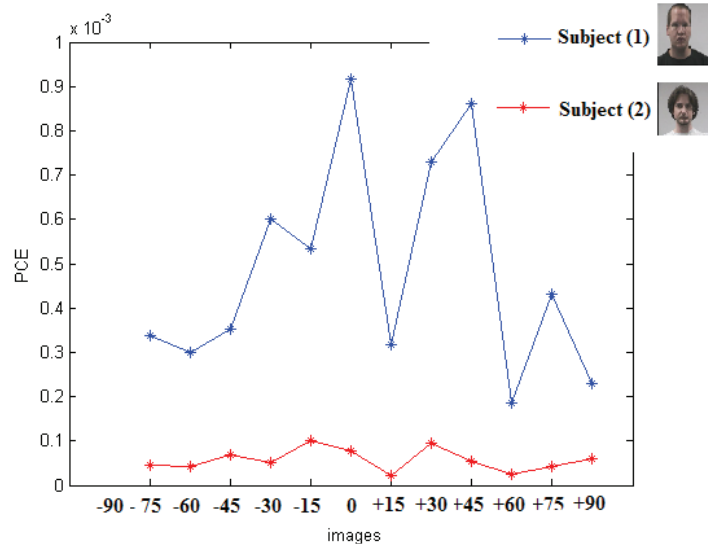


Figure 3. Results obtained using the adaptive fringe-adjusted JTC.

- 2) The second optimization performed incorporated in the fringe-adjusted filter involves adding nonlinearity in the Fourier plane of the correlator [5] such that

$$F_{NL} = F^k \quad (4)$$

where F represents the filtered fringe-adjusted JPS. After several tests in our laboratory, we set the value of k to 0.75.

Figure 4 shows the different test results obtained with our tracking system based on an optimized fringe-adjusted JTC. By comparing the results presented in Figs. 1 and 4, it is evident that the proposed technique yields better discrimination. Indeed the difference between the two curves (red and blue) is larger.

3. CONCLUSION

In this paper we proposed and validated the first version of a robust tracking system based on the fringe-adjusted JTC architecture. By choosing a tolerance rotation angle of 15° (rotation between the target image and the reference image), we considerably reduced the number of correlations necessary and therefore increased the processing speed of the system. Moreover, the adaptive decision criterion used in the output plane minimized the number of false correlation peaks in the correlation plane thus reducing the false alarms. Finally, the addition of nonlinearity in the Fourier plane makes it possible to adapt or find a compromise between the robustness and discrimination of our tracking system by changing the value of parameter k .

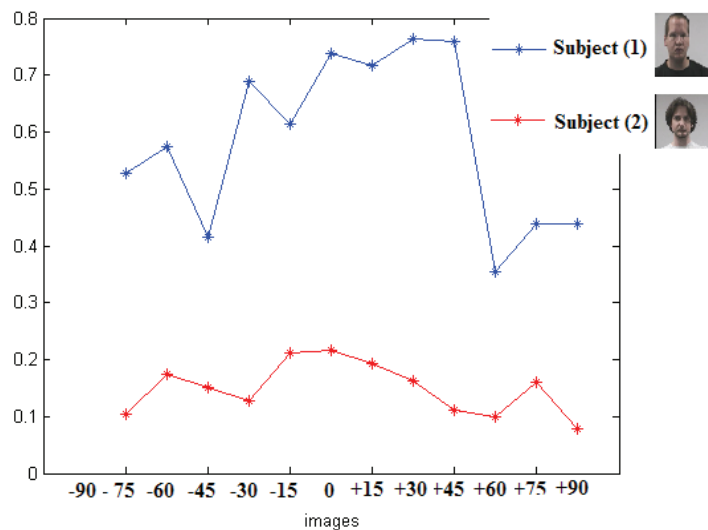


Figure 4. Results obtained with a non-linearity in the Fourier plane of the FA-JTC and the new criterion

REFERENCES

- [1] Alfalou A. and Brosseau C., "Understanding Correlation Techniques for Face Recognition: From Basics to Applications," *Face Recognition*, Milos Oravec (Ed.), ISBN: 978-953-307-060-5, INTECH. Available from: <http://sciyo.com/articles/show/title/understanding-correlation-techniques-for-face-recognition-from-basics-to-applications>.
- [2] VanderLugt V., "Signal detection by complex spatial filtering," *IEEE Trans. Info. Theory*, IT-10, 139-145 (1964).

- [3] Weaver C.S., Goodman J.W., "A Technique for Optically Convolver Two Functions," *Applied Optics*, 5, 1248-1249 (1966).
- [4] Alam M.S., Karim M.A., "Fringe-adjusted joint transform correlation," *Applied Optics*, 32, 4344-4350 (1993).
- [5] Javidi B., "Nonlinear joint power spectrum based optical correlation," *Applied Optics*, 28, 2358-2367 (1989).
- [6] Gourier N., Hall D. and Crowley J. L., "Estimating Face Orientation from Robust Detection of Salient Facial Features," *Proc. of Pointing 2004, ICPR, International Workshop on Visual Observation of Deictic Gestures (2004)*.
- [7] Horner J.L., "Metrics for assessing pattern-recognition performance," *Applied Optics*, 31, 165-166 (1992).
- [8] Alam M.S. and Bal A., "Dynamic target tracking with fringe-adjusted joint transform correlation and template matching," *Applied Optics*, 43, 4874-4881 (2004).

Spectral optimized asymmetric segmented phase-only correlation filter

I. Leonard,¹ A. Alfalou,^{1,3} and C. Brosseau^{2,*}

¹Institut Supérieur d'Electronique et du Numérique (ISEN) Brest, Département Vision, Laboratoire de Recherche de l'ISEN, 20 rue Cuirassé Bretagne, CS 42807, 29228 Brest Cedex 2, France

²Université Européenne de Bretagne, Université de Brest, Laboratoire en Sciences et Techniques de l'Information, de la Communication, et de la Connaissance, CS 93837, 6 avenue Le Gorgeu, 29238 Brest Cedex 3, France

³e-mail: ayman.al-falou@isen.fr

*Corresponding author: brosseau@univ-brest.fr

Received 26 September 2011; revised 21 January 2012; accepted 27 January 2012;
posted 2 February 2012 (Doc. ID 155301); published 8 May 2012

We suggest a new type of optimized composite filter, i.e., the asymmetric segmented phase-only filter (ASPOF), for improving the effectiveness of a VanderLugt correlator (VLC) when used for face identification. Basically, it consists in merging several reference images after application of a specific spectral optimization method. After segmentation of the spectral filter plane to several areas, each area is assigned to a single winner reference according to a new optimized criterion. The point of the paper is to show that this method offers a significant performance improvement on standard composite filters for face identification. We first briefly revisit composite filters [adapted, phase-only, inverse, compromise optimal, segmented, minimum average correlation energy, optimal trade-off maximum average correlation, and amplitude-modulated phase-only (AMPOF)], which are tools of choice for face recognition based on correlation techniques, and compare their performances with those of the ASPOF. We illustrate some of the drawbacks of current filters for several binary and grayscale image identifications. Next, we describe the optimization steps and introduce the ASPOF that can overcome these technical issues to improve the quality and the reliability of the correlation-based decision. We derive performance measures, i.e., PCE values and receiver operating characteristic curves, to confirm consistency of the results. We numerically find that this filter increases the recognition rate and decreases the false alarm rate. The results show that the discrimination of the ASPOF is comparable to that of the AMPOF, but the ASPOF is more robust than the trade-off maximum average correlation height against rotation and various types of noise sources. Our method has several features that make it amenable to experimental implementation using a VLC. © 2012 Optical Society of America

OCIS codes: 070.0070, 100.3008, 100.5010, 100.5090.

1. Introduction

Intense interest in optical correlation techniques over a prolonged period has focused substantially on the filter designs. These techniques represent a powerful tool for a wide range of applications like target tracking, identification, and classification. These applications require real-time processing, must be

robust to several perturbations, e.g., face rotation, and should be insensitive to noise. The optical correlation technique is a good candidate for very fast recognition systems. However, the principle of correlation based on a comparison between the target image and a correlation filter makes this method not robust to rotation. One of the techniques used to overcome this problem is the composite filter, i.e., it contains information coming from multiple images. Based on the composite technique, the overall goal of this line of research is to develop and optimize a

1559-128X/12/142638-13\$15.00/0

© 2012 Optical Society of America

novel filter for the VanderLugt correlator (VLC) used in face recognition [1] that is able to fulfill the above requirements and can handle a large database. In this study, we use different faces with several viewpoints from the Pointing Head Pose Image Database (PHPID) [2]. The results will show the impact of using the fusion criterion permitted to assign a single reference to each filter pixel. Moreover, thanks to the symmetry property of the spectrum, the number of real references in the composite filter is increased. This has improved significantly the decision making process.

Following this brief introduction, we have divided the rest of the paper as follows. A general overview of optical correlation methods is given in Section 2, providing technical details for the system under consideration and a connection to earlier ideas. In Section 3, we present a number of relevant examples to illustrate our strategy for improving the effectiveness of the VLC using the asymmetric segmented phase-only filter (ASPOF). This face recognition algorithm was compared to a whole set of composite correlation filters: all systems were trained and tested on the same images. Our conclusion is that the increase of performance obtained with the ASPOF indicates that we were able to propose and validate a new composite correlation filter, allowing us to increase considerably the number of references to be incorporated in one filter. The receiver operating characteristic (ROC) curves allow us to distinguish between the good recognition and the true nonrecognition values based on an optimized peak-to-correlation energy (PCE) criterion. Finally, we draw our conclusions in Section 4.

2. Some Preliminary Considerations and Relation to Previous Work

The optical correlation method can be traced back to the pioneering research in the 1960s [1,3]. This method consists of comparing a target image and a reference image that comes from a database. This comparison leads to correlation peak in the correlation plane. This correlation peak is more or less intense, depending on the degree of resemblance between the target and reference images. To optimize recognition, several filters created from the reference image have been studied [4–6]. However, the correlation approach, which needs to compare the target image with a large number of reference images in order to increase its robustness, is time consuming. To reduce the processing time, several composite filters have been developed [6–11].

Over the past decade, there has been a resurgence of interest, driven by recognition and identification applications [12–16], of the correlation methods. For example, Alsamman and Alam [16] demonstrated the good performances of the correlation method compared to all numerical ones based on the independent component model. Another significant example in this area of research is the work by Romdhani *et al.* [17], which compared face recognition algorithms

with respect to those based on correlation. These references are far from a complete list of important advances, but fortunately the interested reader can easily trace the historical evolution of these ideas with Kumar’s review paper [7], Yu and Jutamulia’s book [6], Tribillon’s book [18], and the chapter by Alfalou and Brosseau containing an extensive bibliography in [4]. According to the above considerations, our goal is to propose and validate a new robust and discriminating composite correlation filter that is able to include a large set of training images into this filter, which allows one to deal with the diversity of facial expressions as input information. To proceed further, we begin with a brief overview of the VLC method. Next, we wish to specify the definition of standard performance measures that will be used to compare filters.

A. VLC

In this study, we consider only the VLC scheme, and we focus on the design of a new type of optimized composite filter. The VLC was preferred to the joint transform correlator [19–22] because it allows exploitation of all the finite bandwidth in the input plane [23]. Figure (1) illustrates a simplified schematic of the VLC [1,4]. It is based on the multiplication between the spectrum, obtained by Fourier transforming the target image [Fig. 2(a)], with a correlation filter made from the Fourier transform (FT) of a reference image and placed in the Fourier plane. The correlation plane is obtained from FT^{-1} of the product of both spectra. This results in a more or less intense central correlation peak depending on the degree of similarity between the target object and the image reference. Many approaches for designing the correlation filter can be found in the literature according to the specific objects that need to be recognized [4,6,18,24].

B. Performance Measures for Filter Design

In order to quantify the correlation plane by the correlation of a filter and a test image, different correlation plane performance measures were used [25,26]. We choose the PCE criterion [25,26], which is defined by

$$PCE = \frac{\text{(Energy in the correlation peak)}}{\text{(overall in the correlation plane)}}. \quad (1)$$

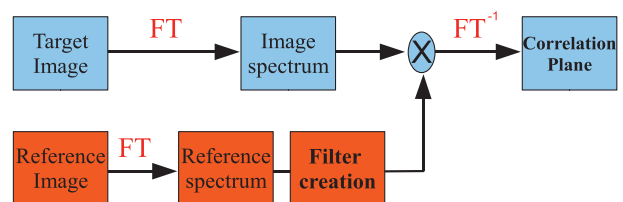


Fig. 1. (Color online) Schematic of the principle of the VLC scheme.



Fig. 2. Illustrating the saturation effect: (a) three 8 bit grayscale images. (b) Image obtained by a classical linear combination of the three images shown in (a). (c) Image obtained using an optimized merging (spectral segmentation).

In other words, the PCE must be interpreted as the weight of the output signal (correlation peak) in the correlation plane relative to noise. In practice, one should take into account the width of the correlation peak in the calculation of the PCE. In what follows, PCE will be evaluated as

$$\text{PCE} = \frac{\sum_{x=x_0-t}^{x=x_0+t} \sum_{y=y_0-t}^{y=y_0+t} |C(x,y)|^2}{\sum_{x=1}^{x=N} \sum_{y=1}^{y=M} |C(x,y)|^2}, \quad (2)$$

where (x_0, y_0) denote the position of the correlation peak, $C(x,y)$ denotes the value of the correlation plane at point (x,y) , t is set to the number of neighboring pixels used, and (N, M) denote the size (in pixels) of the correlation plane. A distinct advantage of the PCE is that it permits evaluation of the relative importance of the correlation peak with reference to the noise of the correlation plane.

Among standard measures for parametrizing the correlator performances, the ROC curve, which allows the characterizing of the binary classifiers, was selected [27]. The ROC curve can be represented by plotting the fraction of true positives out of the positive true positive rate (TPR, or sensitivity) versus the fraction of false positives out of the negative false positive rate (FPR, or $1 - \text{specificity}$). We find it can be more useful to consider adapted ROC curves [27]. We fix a weight according to the relative position of the PCE with respect to threshold. We shall consider ten zones separating the threshold and the maximum PCE values obtained for a target face and affect the weight = 0.1 for the first zone, 0.2 to the second zone, and so on.

3. ASPOF: A Spectral Optimized Filter

The preceding analysis demonstrates the impact of a reduced number of correlations to obtain a reliable decision by using a composite filter. However, this filter needs to be optimized so that it can permit a merging of a large number of reference images within a given filter. For that purpose we suggest a new composite filter, the ASPOF. The ASPOF is an optimized version of the segmented correlation filter [8]. Our algorithm has been extensively tested by comparing its performance with those of several composite filters [28].

A. Segmented Composite Filter

The main idea behind the segmented composite filter (SPOF) is that the high saturation regions of the reference images are suppressed. Briefly stated, this is

achieved through two steps [8]. First, a segmentation of the spectral plane of the correlation filter is realized into several independent regions. Second, each region is assigned to a single reference. This assignment is done according a specific energy criterion

$$\frac{E_{u,v}^l}{\sum_{i,j} E_{i,j}^l} > \frac{E_{u,v}^k}{\sum_{i,j} E_{i,j}^k}, \quad (3)$$

where $E_{u,v}^l$ denotes the energy on the pixel (u,v) of the reference image (l) in the spectral domain. This criterion compares the energy (normalized by the total energy of the spectrum) for each frequency of a given reference with the corresponding energies of another reference. Assignment of a region to one of the two references is done according Eq. (3). Hence, the SPOF contains frequencies with the largest energy.

B. ASPOF Algorithm

A serious limitation of the SPOF is that the regions to which the winning references are assigned become smaller as the similarity between the reference images is increased. This effect eventually leads to isolated pixels. This issue can be critical since this paper deals with training sets of reference images bearing a strong resemblance, i.e., database obtained by rotating images within the range $(-90^\circ, +90^\circ)$, and we want to increase the number of reference images in the filter. Our optimizing method from the six-step scheme that defines the ASPOF follows.

(i) We first separate the reference image base $\{\mathcal{R}_{i=1 \dots N}\}$, where N denotes the number of used references, into two subclasses: one, $\{\mathcal{R}_i^1\}$, which contains the reference images with $i = 1, 3, 5, \dots, N-1$, and the other one, $\{\mathcal{R}_i^2\}$, deals with the other indices $i = 2, 4, 6, \dots, N$ (Fig. 3).

(ii) Next, in stage (ii), two segmented filters are constructed, using an similar analysis method to [29], for the two subclasses, H_{SPOF}^1 and H_{SPOF}^2 .

(iii) Selection and assignment optimized criterion used for segmentation of the Fourier plane for both filters are realized as follows:

$$H_{\text{SPOF}}^1 = \begin{cases} \text{Spect}_{u,v}^l & \text{if } \frac{E_{u,v}^l}{\sum_{i,j} E_{i,j}^l} \geq \alpha \frac{E_{u,v}^k}{\sum_{i,j} E_{i,j}^k} \\ \text{with } l \in [0, N-1] & \\ \text{Not affected} & \text{otherwise} \end{cases} \quad (4)$$

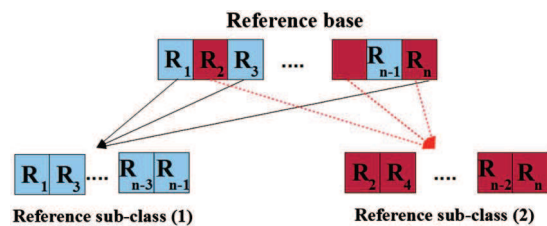


Fig. 3. (Color online) Technique used to classify the reference images in two subclasses.

Here $\text{Spect}^l_{u,v}$ is the spectrum of the reference R_l at position (u, v) and $E^l_{u,v}$ is the energy of the reference R_l at the position (u, v) . The constant a (here set to 1.2) is a pixel assignment coefficient. Basically, it means that the pixel of the filter is assigned to a given reference if and only if its energy relative to spectral position (u, v) is larger than the energies of all other references multiplied a times. Otherwise, it is not assigned. In Eq. (4), the usefulness of the coefficient a is motivated by the fact that sometimes there exist areas of the spectral plane that are very similar. Thus, by introducing this coefficient a , these areas are not assigned to any spectrum according to the segmentation criterion [Eq. (4)]. These areas will be assigned according to the neighborhoods of each spectrum (Fig. 4). A similar operation for the second filter H^2_{SPOF} is applied using the spectra of the second subclass and considering again Eq. (4).

(iv) In this procedure, unclassified pixels, i.e., those that do not satisfy Eq. (4), are assigned to one of the two spectra by looking at their closest neighbors (Fig. 4) in order to avoid isolated pixels. An isolated pixel represents a pixel of a spectrum l that is surrounded by pixels of the spectrum k . Isolated pixels are detrimental to the segmented filter's performance, the effect being more and more important as the number of references that define the filter is increased. Faced with several options for the purpose of decreasing this effect, we chose the following scenario. The assignment of pixels that do not satisfy Eq. (4) is realized according to the schematic illustration depicted in Fig. 4. This figure shows different situations; e.g., in Figs. 4(a) and 4(b) the considered pixel is surrounded by pixels of only one spectrum.

Hence, the pixel is assigned to this spectrum. In another situation, Fig. 4(c) illustrates the case of a pixel surrounded by pixels belonging to the two spectra. In this case, the algorithm searches for the immediate neighborhood (3×3 pixels) of the pixel, and the rule is to assign this pixel to the spectrum that has the largest number of pixels. Figure 4(d) represents an example in the case that the surrounding environment of the considered pixel contains exactly

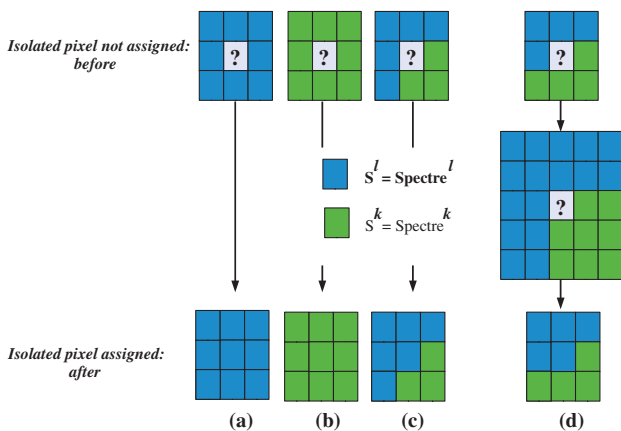


Fig. 4. (Color online) Optimized assignment procedure for isolated pixels.

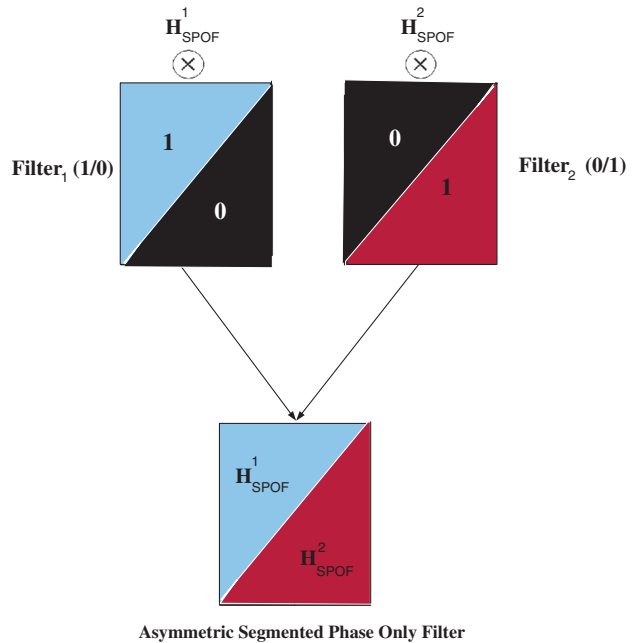


Fig. 5. (Color online) Merging technique based on the Fourier plane symmetry property.

the same number of pixels for the two spectra. In this case, the algorithm searches for the spectrum that has the largest number of pixels in an enlarged neighborhood, i.e., 5×5 pixels.

(v) Next, the assignment step is repeated. Interestingly, we observed that a single iteration is necessary for letters.

(vi) Next, H^1_{SPOF} is multiplied with a symmetric filter [denoted as filter₁(1/0) in Fig. 5], and H^2_{SPOF} is multiplied with the corresponding filter₂(0/1) in Fig. 5. Finally, we sum the two filters in the Fourier plane. Overall, the ASPOF can be described by $H_{\text{ASPOF}} = p_1 H^1_{\text{SPOF}} \text{filter}_1 + p_2 H^2_{\text{SPOF}} \text{filter}_2$, with p_1 and p_2 being two coefficients. In the following we will study the effect of these two numbers on the performance of the ASPOF.

C. ASPOF Parameterization

Two kinds of parameter were considered. On the one hand, to illustrate how t in Eq. (2), i.e., $(2t + 1)^2$ pixels are required for the calculation of the PCE, can impact the correlation performances, we consider the letter A (with rotations between -90° and 90°) and compare the performances of the phase-only filter (POF), SPOF, amplitude-modulated phase-only filter (AMPOF), and ASPOF. The fabrication of these filters was realized using with ten references, i.e., binary images of the letter A and ten rotation angles. Figure 6 compares the PCE resulting from correlating the letters A and V with the four composite filters. Taking $t = 1$ pixel [Fig. 6(a)], the ASPOF is found to be very discriminating. One way to increase the robustness of the ASPOF is to increase t . It can be seen that the ASPOF shows the best compromise between robustness and discrimination factor when

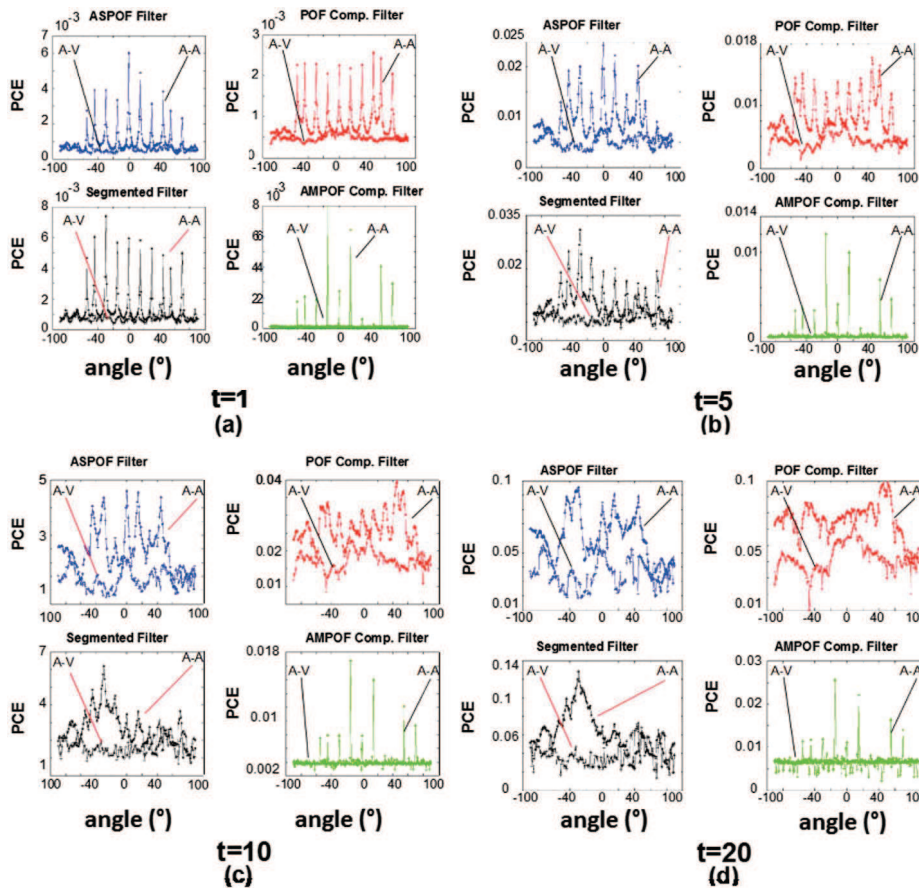


Fig. 6. (Color online) PCEs obtained with ten-reference POF, SPOF, AMPOF, and ASPOF. (a) $t = 1$, (b) $t = 5$, (c) $t = 10$, and (d) $t = 20$.

$t = 10$. For t larger than 10, the discrimination quality decreases [Fig. 6(d)]. In the following, t is set to 10. On the other hand, we further show how p_1 and p_2 can influence the performances of the ASPOF. Figure 7(a) shows the PCE values with $p_1 = p_2 = 1$, where we choose the same illustrating example as above. By a trial and error method, we found that the best performances were obtained by taking $p_1 = 2$ and $p_2 = 1.5$ [Fig. 7(b)].

D. Composite Filters Used for Comparison with the ASPOF

In this subsection, we will review various composite filters in order to select some of them [POF, SPOF, AMPOF, and trade-off maximum average correlation height (OT-MACH)] that will be compared with the ASPOF in order to demonstrate its robustness and its good discrimination. We first study the effect of the background. We considered different classical composite filters (the full list is shown in Table 1). A detailed comparative study of the discrimination and robustness performances for standard composite correlation filters employed for face recognition is presented in the supplementary material [28].

As expected, the adapted composite filter is robust against rotation and not discriminating. Moreover,

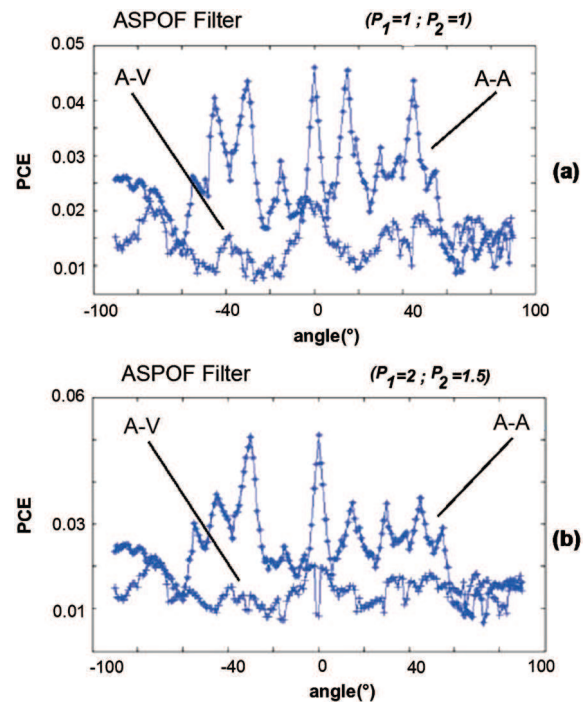


Fig. 7. (Color online) PCEs obtained with a ten-reference ASPOF and $t = 10$ pixels. (a) $p_1 = p_2 = 1$, (b) $p_1 = 2$ and $p_2 = 1.5$.

Table 1. Different Composite Filters Used in This Study^a

Composite Filter	Notation
Adapted filter	H_{compAd}
POF	$H_{\text{compPOF}}^1, H_{\text{compPOF}}^2$
Binary POF	$H_{\text{compBPOF}}^1, H_{\text{compBPOF}}^2$
Inverse filter	H_{compIF}
Compromise optimal filter	H_{compOT}
Segmented filter	$H_{\text{SPOF}}^1, H_{\text{SPOF}}^2$
Segmented binary filter	$H_{\text{BSPOF}}^1, H_{\text{BSPOF}}^2$
MACE filter	H_{compMACE}
AMPOF	$H_{\text{compAMPOF}}$

^a H_{compAd} denotes the adapted composite filter. It is realized by considering a linear combination of reference images and using the adapted filter definition. $H_{\text{Comp_POF}}$ is the POF composite filter. We tested two different schemes for realizing the composite POF filter. In the first scheme (H_{compPOF}^1), a linear combination of reference images was used to create the POF. The second scheme (H_{compPOF}^2) involves performing the POF for each reference and using the linear combination of these POFs. H_{compBPOF}^1 and H_{compBPOF}^2 are the binarized versions of the filters H_{compPOF}^1 and H_{compPOF}^2 , respectively. The composite inverse filter H_{compIF} is the inverse filter of the linear combination of reference images. The optimal composite filter H_{compOT} is realized by linearly combining reference images. H_{compSPOF} denotes the segmented filter realized by doing segmentation and assignment with the energy criterion. The calculation of filter H_{compSPOF}^2 is done by replacing the energy with the square of the real part of the different references spectra to be merged. H_{compMACE} is the composite filter of the minimum average correlation energy (MACE) filter. $H_{\text{compAMPOF}}$ is the composite version of the AMPOF.

its low discriminating character is more and more visible as the number of references is increased. Our findings suggest that the composite POF is quite robust but not very discriminating; see Figs. 8 and 9. For obtaining these figures, we have correlated different rotations (from -90° to 90°) of the A and V letters with different composite POF filters using different rotations of the letter A as reference images. As seen in Fig. 8, the different PCE data show that the correlation values have an average PCE value larger than 3×10^{-3} . It is important to point out that the PCE values obtained from the correlation of the letter V correlated with the same filters have an average value less than 10^{-3} and support the conclusions reported in [30]. In addition, a composite POF shows robustness to noise, especially when the noise is clearly identified. However, we find that the energy contained in the correlation peak decreases significantly, i.e., the PCE is decreased by a factor of 3 when using a POF containing three references by contrast with a POF realized with a single reference (Fig. 8). This supports also the conclusions made in [30]. Figure (8) shows that, for an 11-reference POF, the PCE is decreased by an order of magnitude that renders unreliable the decision on the letter identification. Beyond 11 references, the weakness of the magnitude of the PCE makes the recognition of the images forming the filter very difficult. This is partly due to the saturation. Indeed, when the number of training images is large, saturation occurs because

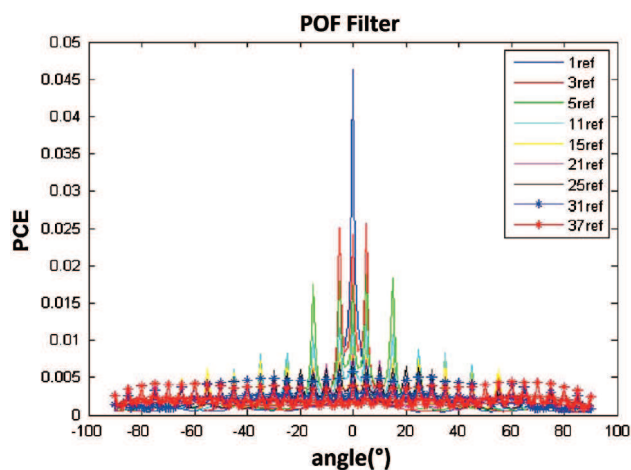


Fig. 8. (Color online) PCEs obtained with the composite POF. The colors shown in the key denote the different filters as a function of the number of references used.

the correlation filter is pixelated, i.e., since each pixel is encoded with a fixed number of bits, increasing the latter has the effect of slowing down the filtering process and thus increasing the required memory space. To overcome this saturation limitation, the SPOF was proposed and validated in [8]. To further show the interest in using a segmented filter with respect to the saturation problem that affects the classical composite filter, we show in Fig. 2(b) the 8 bit image of the sum (without segmentation) of the three spectra corresponding to the reference images [Fig. 2(a)]. Figure 2(c) shows the corresponding sum with segmentation [8]. Our calculations clearly indicate that the image with segmentation shows significantly less saturation than that obtained without segmentation.

For comparison, we performed simulations for a binarized composite POF [28]. Overall, the behavior of this kind of filter is very similar to that of a POF, albeit with smaller performances. We should notice that this conclusion is similar to that noted by Mohammed *et al.* in [31]. The low performances of the binarized POF are confirmed by a comparison between the results obtained by Mohammed *et al.* [31] and those reported by Rahman *et al.* [30].

Overall, we found that the composite POF has a good robustness against rotation but has a low discrimination. To look at the discrimination of composite filters in more detail, we shall consider the inverse composite filter. Such a filter shows a strong discriminating ability and a low robustness against small changes of the target image with respect to the reference image [28]. Figure 10 presents the correlation results obtained for different rotations of the letter A and several inverse filters. This leads to the conclusion that this inverse composite filter is not well adapted for the composite architecture.

Next, testing the compromise optimal filter, we observed that its composite version is robust to noise when the latter is clearly identified [28]. We also observed that the optimal trade-off (OT) filter is not robust to image rotation when the images are noisy,

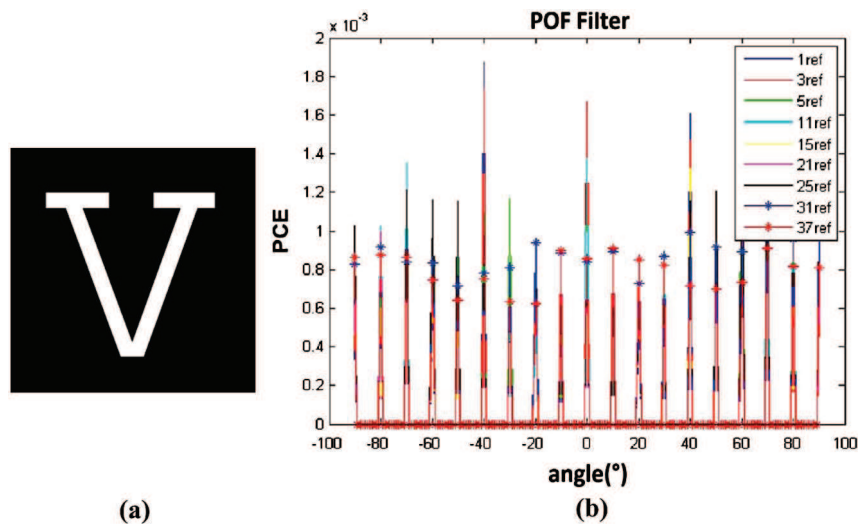


Fig. 9. (Color online) Discrimination results: (a) the target image is the letter V. (b) PCEs obtained with a filter fabricated with reference images of letter A. The colors shown in the key denote the different filters as a function of the number of references used.

especially if the noise cannot be explicitly evaluated; see Fig. 11 (different versions of this filter were considered in [28]).

The MACE filter was also studied [11,32]. Its correlation performance is good when the target image is similar to one of the reference images; see Region B in Fig. 12. Nevertheless, the robustness of this filter is very poor since the correlation peak disappears in Regions A and C (Fig. 12). In [33], Iftikharuddin *et al.* proposed and validated an optimized MACE filter, the amplitude coupled MACE filter, that is robust to rotation. However, like the classical MACE filter, the optimized version is also sensitive to noise, especially to structured noise.

Next, the AMPOF [34,35] was considered because it is high discriminating performance. Awwal *et al.* [35,36] suggested an optimization of the POF filter based on the following idea: the more the spectrum of the Fourier plane can be flattened, the sharper will be the correlation peak. These authors introduced

the AMPOF (H_{AMPOF}). As shown in our tests [28], the AMPOF is not robust to rotation and is weakly robust to noise. However, the AMPOF is found to have significantly superior correlation discrimination capability. In [37], Iftikharuddin *et al.* studied a discretized version of the AMPOF. The resulting filter has good recognition performances of a noisy object. In addition, the good discrimination performances of the AMPOF led Iftikharuddin *et al.* [38] to introduce an amplitude-modulated inverse filter (AMIF) that is robust to Gaussian white noise. But, the AMIF is very sensitive to face rotation. Thus, we considered this AMPOF in order to compare it with our new ASPOF. This is to highlight the good discriminatory performance of our ASPOF. Another filter that was considered in this study is the OT-MACH filter [39,40]. From these different comparisons, the POF, AMPOF, and the MACH composite filters were selected for the comparison of correlation performances with our optimized ASPOF.

E. Performance Criterion Adapted to the ASPOF

Here we will modify the definition of the PCE, i.e., Eq. (2), since we assumed that only a single correlation peak is considered and that the filter that led to the correlation peak contains the total energy. These assumptions should be reconsidered for the ASPOF. Only half of the energy of the output plane is associated with each filter since the ASPOF is composed of two distinct filters. On the other hand, each filter leads to a correlation peak. Hence, for various situations, two correlation peaks can appear in the output plane [see, e.g., Fig. 13(a)]. To circumvent the problem connected to the energy lowering of the correlation peak with the ASPOF, i.e., only half the filter plane is used, the correlation peak magnitude could be multiplied by a factor of 2. However, this “trick” is insufficient. In fact the asymmetry feature of this filter leads to a correlation between the target image and a part of our filter, while there is intercorrelation

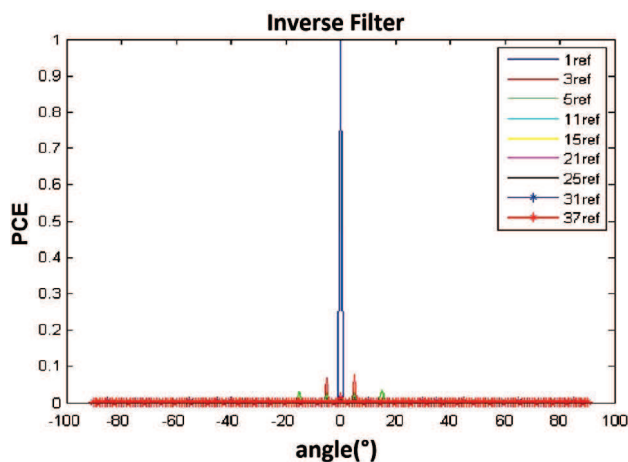


Fig. 10. (Color online) PCEs obtained with the inverse composite filter. The colors shown in the key denote the different filters against the number of references used.

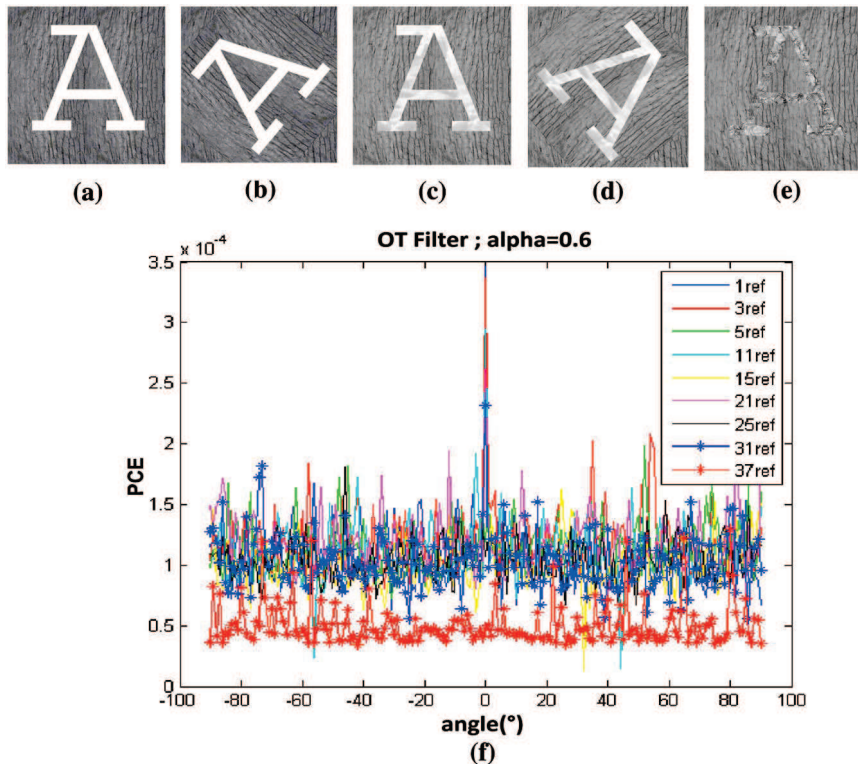


Fig. 11. (Color online) (a) Illustrating the letter A with additive background noise. (b) Same as in (a) with a rotation angle of -50° . (c) Illustrating the letter A with structured noise. (d) Same as in (c) with a rotation angle of 50° . (e) Illustrating the letter A for a weak contrast. (f) PCEs obtained with the OT composite filter taking $\alpha = 0.6$. The colors shown in the key denote the different filters versus the number of references used.

with the second part. This can be observed in Fig. 13, which shows three correlation planes of the target image defined by the letter A.

As is apparent in Fig. 13, no correlation peaks are visible with the composite filters and AMPOF. However, it should be noted that the correlation plane

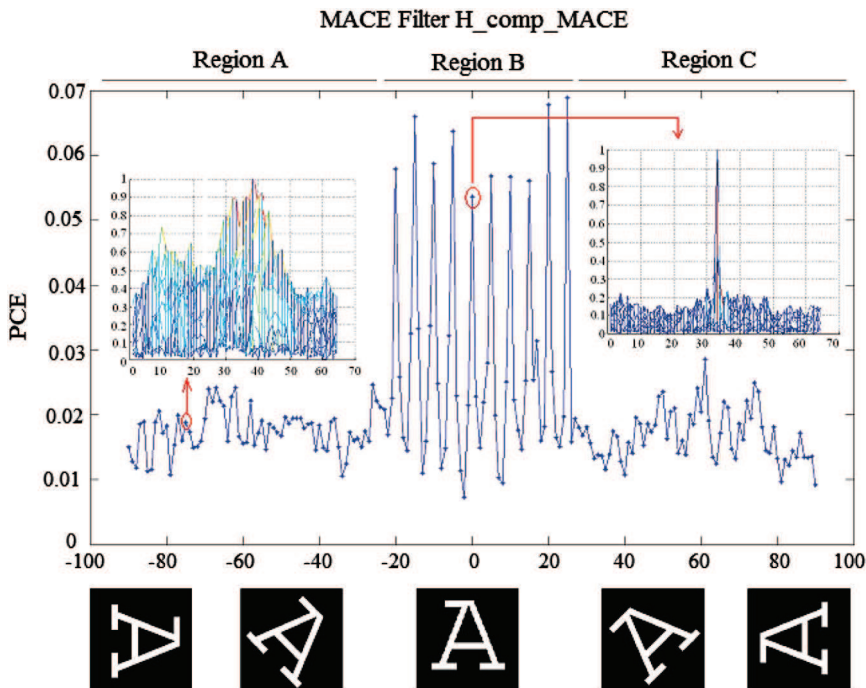


Fig. 12. (Color online) PCEs obtained with a ten-reference MACE when the target images are noiseless. Several examples of the rotated letter A are shown at the bottom of this figure. The inset shows two correlation planes. (Right) Autocorrelation obtained without rotation. (Left) Intercorrelation obtained with the letter A oriented at -75° .

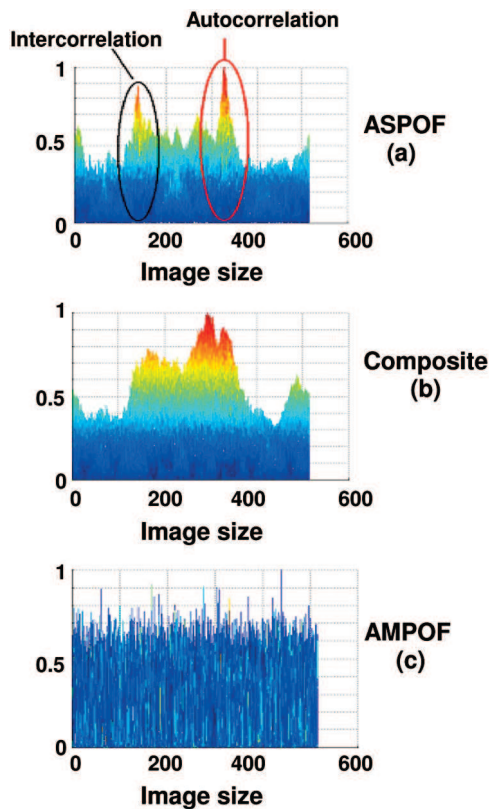


Fig. 13. (Color online) Output normalized Fourier planes obtained by correlating of the same target image (letter A with a rotation angle of -75°) with (a) ASPOF, (b) composite filter, and (c) AMPOF.

displayed in Fig. 13(a) shows two peaks: a first peak arises from the autocorrelation between the target image and the part of the ASPOF that contains the reference image that is similar to the target image, and the second peak can be identified as due to the intercorrelation between the target image and the other part of the ASPOF. Overall, the autocorrelation plane is noised by the intercorrelation plane. This effect has the effect of decreasing the PCE values and consequently the identification performances of the ASPOF. In addition, it may happen that the target image is similar to two reference images, with one of them being in the first part of the ASPOF and the other one being in the second part of the ASPOF. To address this difficulty, we defined a different criterion, hereinafter denoted as the adapted peak-to-correlation energy (PCEA), which can be written as

$$\text{PCEA} = \frac{\sum_{x=x_0-t}^{x=x_0+t} \sum_{y=y_0-t}^{y=y_0+t} |2C(x,y)|^2}{\sum_{x=1}^{x=N} \sum_{y=1}^{y=M} |C(x,y)|^2 + 3 \left(\sum_{x=x_0-t}^{x=x_0+t} \sum_{y=y_0-t}^{y=y_0+t} |C(x,y)|^2 \right)}. \quad (5)$$

To further investigate this point, i.e., to separate the autocorrelation plane from the intercorrelation one, we have added a specific phase term to the two parts of the ASPOF. This phase is chosen so that the correlation can be placed at a given point in the output plane. The respective correlation plane performances of the different simulations are in very good agreement with those gauged with the PCEA performance metric.

4. Illustrating the ASPOF Performances

A. Robustness against Face Rotation

We considered the behavior of the ten-reference H_{ASPOF} for the problem of identifying the letter A with a rotation angle ranging from -20° to 25° . Then, each letter A of the database (obtained by rotating the A image over $(-90^\circ, 90^\circ)$ in increments of 1° counterclockwise) was correlated with the ten-reference H_{ASPOF} . As shown by the green line in Fig. 14(a), the letter A has been identified over $(-20^\circ, 25^\circ)$. To emphasize the benefit of the optimization stage concerning the isolated pixels [Stage (iv) of the algorithm] on the filter behavior, it is instructive to contrast the results obtained for the H_{ASPOF} without taking into account this step in our algorithm as shown by the green line in Fig. 14(b). In the latter case, it should be noted that the PCE values are decreased by a factor of 2. It is evident that this optimization stage allows us to decrease the influence of the isolated pixels in the classical segmented filter H_{SPOF} .

For comparison, we also plot in Figs. 14(a) and 14(b) the correlation performances obtained with H_{POF}^1 (blue line), H_{SPOF} (red line), and H_{AMPOF} (black line) in the case of a noise-free target image. Figures 14(a) and 14(b) show that H_{ASPOF} is characterized by the largest PCE values in the correlation region A, including those of the AMPOF. In addition, H_{ASPOF} presents the most significant contrast between the correlation region A and the no-correlation regions B and C. This shows that H_{ASPOF} is robust in the correlation region and discriminating in the no-correlation regions.

B. Robustness to Rotation and Noise

As we next demonstrate, the conclusions are robust since the ASPOF remains effective when noise is contained in the target image. The noise analysis procedure was identical to that performed previously. Figure 14(c) corresponds to a centered white noise of variance set to 0.1, and Fig. 14(d) considers a

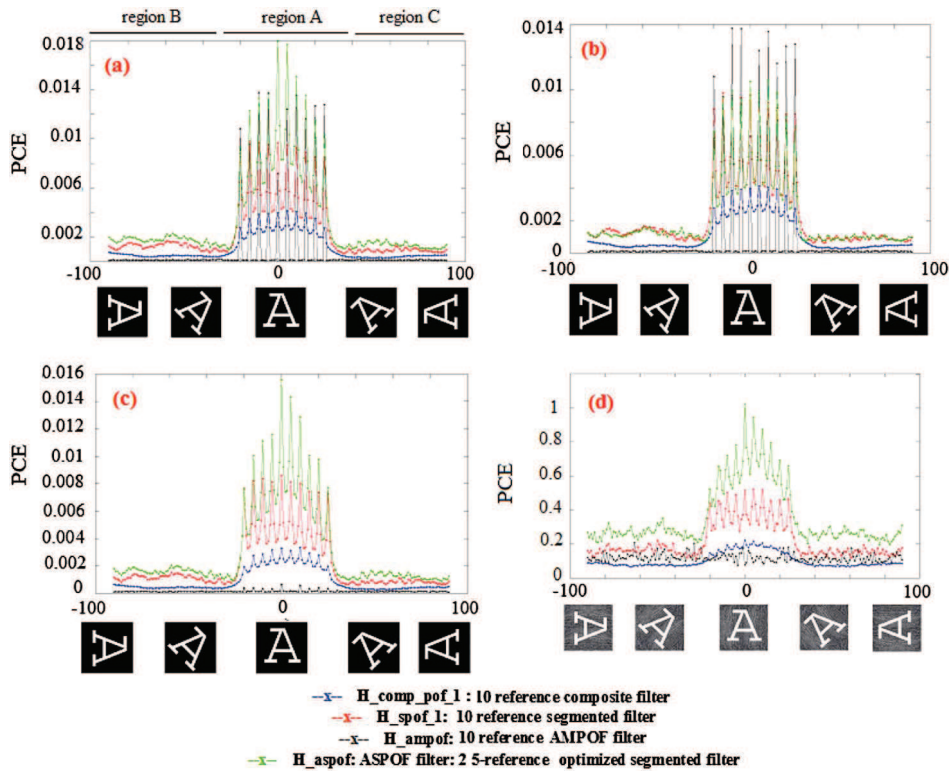


Fig. 14. (Color online) Comparison between the different correlations of letter A (we consider rotation angles ranging between -90° and 90°) with the ten-reference composite filters: POF (blue line), segmented (red line), AMPOF (black line), and ASPOF (green line). (a) PCEs obtained using the optimization stage concerning the isolated pixels, (b) PCEs obtained without the optimization stage concerning the isolated pixels. (c) and (d) represent the PCEs obtained with noised target images.

structured background noise. For each case, H_{ASPOF} is a robust noise filter in comparison with the composite filters that were considered previously. Different tests involving other types of noise (not shown) are consistent with this observation.

C. ROC Curves

The ROC curves in Fig. 15 demonstrate the effectiveness of the ASPOF for image recognition. The data-

base consisted of ten reference images of the letter A (binary image, 512×512 pixels) taking a range for angle of rotation from -60° to 60° , i.e., -55° , -45° , -30° , -15° , 0° , 15° , 30° , 45° , 55° , 60° . The correlation planes are quantified using Eqs. (2) and (5) (we choose $t = 10$ pixels), and the recognition performance of composite filters for this specific set of reference images are studied and compared (supplemental material [28]). The merging of the two

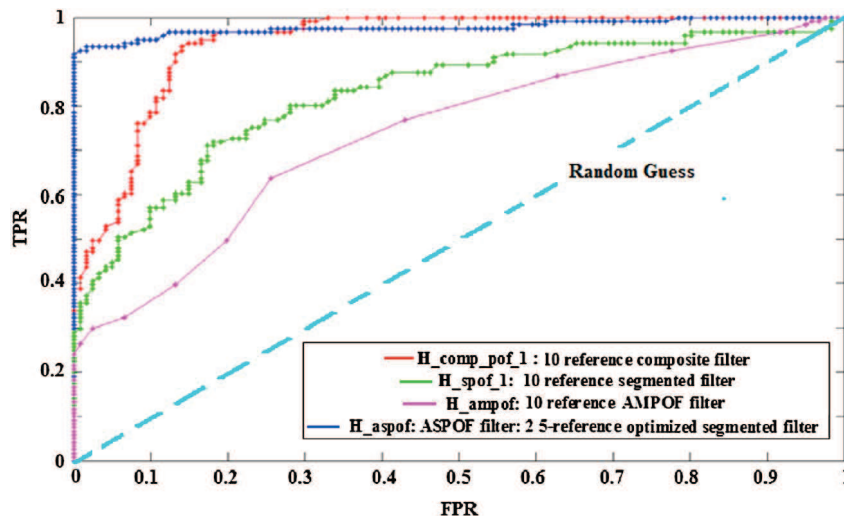


Fig. 15. (Color online) ROC curves obtained with ten-reference composite filters: POF (red), SPOF (green), AMPOF (purple), and ASPOF (navy blue). The sky-blue line shows the random guess.

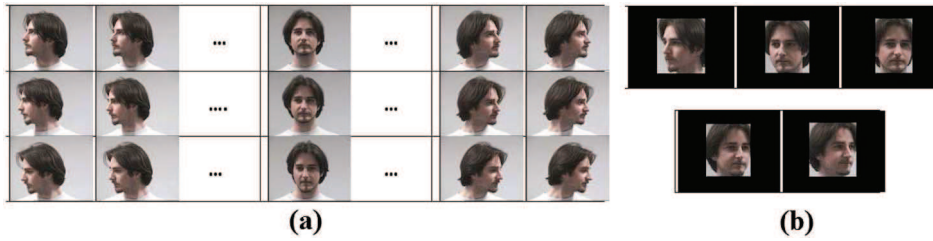


Fig. 16. (Color online) Example of face images from the PHPID that were captured under variable angle conditions, i.e., from top to bottom, +15°, 0°, +15°; from left to right, -90°, 0°, +90°. (a) Target faces; (b) references.

components of the ASPOF, i.e., $H_{ASPOF} = aH_{SPOF}^1 + bH_{SPOF}^2$, was also optimized by choosing $a = 2$ and $b = 1.5$. Figure 15 shows the ROC curves obtained with the capital letters A and V (by rotating the V letter in increments of 1° counterclockwise from -90° to 90°) for different composite filters. As can be seen in the ROC curves of Fig. 15, the AMPOF gives the closest results to a random guess. This reflects the strong discrimination and the low robustness to rotation of the AMPOF. From Fig. 15, we can see that the overall performance of the segmented filter is better than that of the AMPOF, but remains poor

because of the isolated pixels. The POF is able to achieve better performance than the AMPOF; however, because of the saturation problem only a weak difference between the good decisions and the false alarms is observed (Fig. 14). This is in contrast with the different performance points that can be achieved with the ASPOF. On one hand, the results of our tests clearly show the good recognition performance of the ASPOF, i.e., TPR = 0.92 corresponding to FPR = 0. On the other hand, it presents good separation between good decisions and false alarms (Fig. 14).

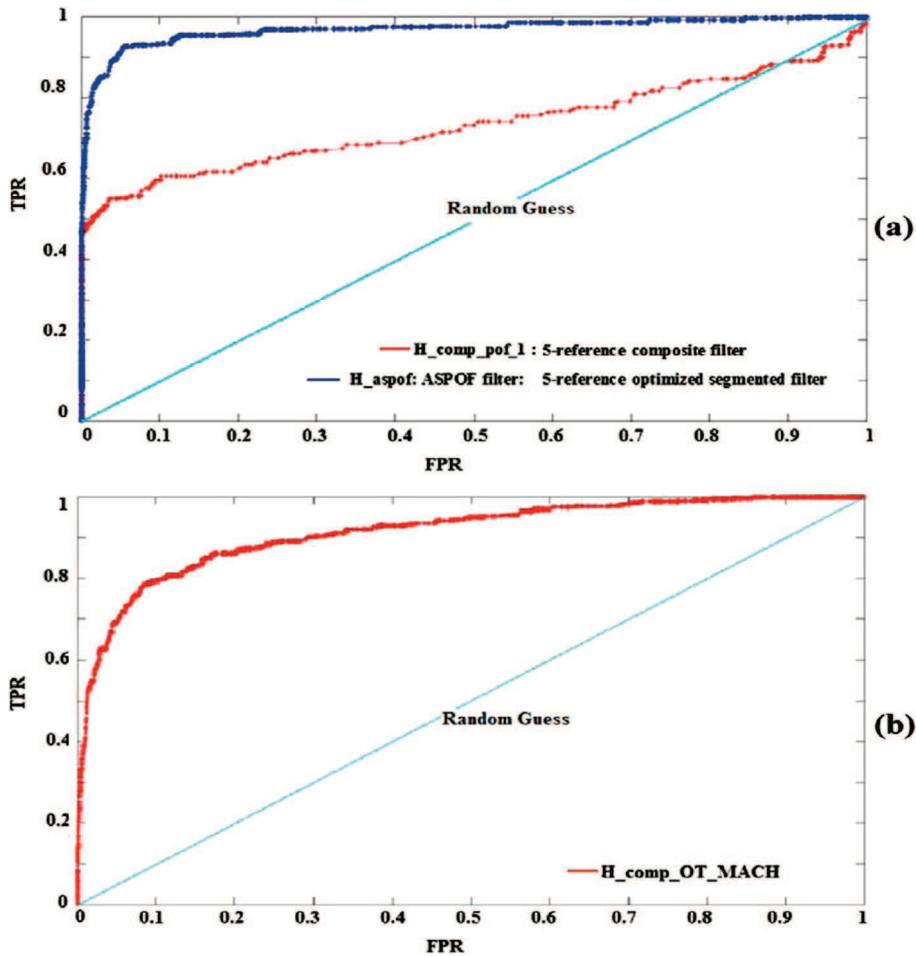


Fig. 17. (Color online) (a) ROC curves obtained by correlating faces of a given subject, e.g., Fig. 9(a), with six other individuals with five-reference ASPOF (navy blue) and POF (red) composite filters. The sky-blue line shows the random guess. (b) ROC curve obtained with an OT-MACH.

D. Application to Face Recognition

The final point we wish to mention here is that the above ideas can be extended to the problem of grayscale image identification. For this purpose we considered six different subjects from the PHPID [2]; an example is given in Fig. 16(a) with several training images per person corresponding to different face orientations. As above, it is possible to define a five-reference ASPOF for a given subject (with variable rotation angles, i.e., from top to bottom, -15° , 0° , $+15^\circ$; from left to right, $+90^\circ$, 0° , -90°) in the following manner: three references were placed in the first half of the ASPOF, and the remaining two references were positioned in the other half part of the ASPOF.

In this example, we choose $t = 1$ pixel, $a = 3$, and $b = 1.5$. The values of these parameters are chosen in relation with the considered application. The five-reference images correspond to -45° , -30° , -15° , 15° , and 45° rotation angles [Fig. 16(b)]. ASPOF training set correlations in Fig. 17(a) were observed to produce good recognition performances in comparison to those of the POF. It is likely that these differences originate from the saturation problem, which is more pronounced for the case of faces. The recognition performance of the ASPOF was also compared to that of the OT-MACH filter with a background noise. The parameters for the OT-MACH are $\alpha = 0$, 25 , $\beta = 1$, and $\gamma = 0, 1$ [28, 39, 40]. The results in Fig. 17(b) show that the recognition performances are significantly larger with the ASPOF. The result [Fig. 17(b)] of our method demonstrates the good recognition performance of the ASPOF, i.e., $\text{TPR} = 0.69$ corresponding to $\text{FPR} = 0$ to be compared with $\text{TPR} = 0.33$ and $\text{TPR} = 0.47$ for the OT-MACH and POF, respectively).

5. Summary

In summary, we have developed a face recognition system based on optical correlation for use in identification and classification. The system will be able to detect, identify, and track various targets with robustness and discrimination. The method has two main stages: increase the number of reference images in order to consider as much as possible target image changes, and use composite filters. The main advantage of these filters is to merge the information of many reference images in a single filter, thus reducing the number of correlations to make a reliable decision. However, the robustness and discrimination performances decrease as the number of reference images is increased, i.e., saturation. To overcome this limitation, we have presented herein a fairly thorough analysis of an optimized composite filter, i.e., the ASPOF, for improving the effectiveness of a VLC when used for face identification. The principle of our filter consists in a specific segmentation of the filter's Fourier plane in several regions. Next, each region was assigned to a single reference image. This optimization stage allowed us to increase significantly the number of reference images of the segmented filter. However, the main drawback of this solution is to generate isolated pixels that yield to

low overall performances of this kind of filter. Next, we have proposed a new type of optimized composite filter, i.e., the ASPOF. We found useful to introduce a new segmentation criterion to characterize the correlation plane of this filter. Each of the stages has various parameters that must be optimized in order to increase the robustness and discriminating ability while increasing the number of reference images contained in the filter. We have validated this technique and applied it to several binary and grayscale image identifications. Overall, our results are encouraging and demonstrate that the ASPOF is characterized by a good compromise between robustness and discrimination.

Furthermore, we believe that the simplicity of the technique provides sufficient appeal from the experimental viewpoint. Of equal or potentially more importance will be to optically or numerically implement it. We believe that it would be interesting to explore this topic further—we discuss very briefly below how this might be done. There are two key directions for improving the results presented here. First, a careful optimization of the merging method is required to maximize the use of the filter's Fourier plane bandwidth. Second, we expect that these general results will apply in three-dimensional face recognition application as well. Work addressing these avenues has been initiated.

The authors acknowledge the partial support of the Conseil Régional de Bretagne and thank A. Arnold-Bos (Thales Underwater Systems) for helpful discussions. They also acknowledge S. Quasmi for her help with the simulations. Lab-STICC is Unité Mixte de Recherche CNRS 3192.

References

1. A. Vanderlugt, "Signal detection by complex spatial filtering," *IEEE Trans. Inf. Theory* **10**, 139–145 (1964).
2. Pointing Head Pose Image (PHPID), http://www.ecse.rpi.edu/~cvtl/database/other_Face_Databases.htm.
3. C. S. Weaver and J. W. Goodman, "A technique for optically convolving two functions," *Appl. Opt.* **5**, 1248–1249 (1966).
4. A. Alfalou and C. Brosseau, "Understanding correlation techniques for face recognition: from basics to applications," in *Face Recognition*, M. Oravec, ed. (In-Tech, 2010).
5. J. L. Horner and P. D. Gianino, "Phase-only matched filtering," *Appl. Opt.* **23**, 812–816 (1984).
6. F. T. S. Yu and S. Jutamulia, *Optical Pattern Recognition* (Cambridge University 1998).
7. B. V. K. V. Kumar, "Tutorial survey of composite filter designs for optical correlators," *Appl. Opt.* **31**, 4773–4801 (1992).
8. A. Alfalou, G. Keryer, and J. L. de Bougrenet de la Tocnaye, "Optical implementation of segmented composite filtering," *Appl. Opt.* **38**, 6129–6135 (1999).
9. H. J. Caulfield and W. T. Maloney, "Improved discrimination in optical character recognition," *Appl. Opt.* **8**, 2354–2356 (1969).
10. C. F. Hester and D. Casasent, "Multivariate technique for multiclass pattern recognition," *Appl. Opt.* **19**, 1758–1761 (1980).
11. A. Mahalanobis, B. V. K. V. Kumar, and D. Casasent, "Minimum average correlation energy filters," *Appl. Opt.* **26**, 3633–3640 (1987).
12. A. Alfalou and C. Brosseau, "Robust and discriminating method for face recognition based on correlation technique and independent component analysis model," *Opt. Lett.* **36**, 645–647 (2011).

13. M. Elbouz, A. Alfalou, and C. Brosseau, "Fuzzy logic and optical correlation-based face recognition method for patient monitoring application in home video surveillance," *Opt. Eng.* **50**, 067003 (2011).
14. I. Leonard, A. Arnold-Bos, and A. Alfalou, "Interest of correlation-based automatic target recognition in underwater optical images: theoretical justification and first results," *Proc. SPIE* **7678**, 76780O (2010).
15. V. H. Diaz-Ramirez, "Constrained composite filter for intra-class distortion invariant object recognition," *Opt. Lasers Eng.* **48**, 1153–1160 (2010).
16. A. Alsamman and M. S. Alam, "Comparative study of face recognition techniques that use joint transform correlation and principal component analysis," *Appl. Opt.* **44**, 688–692 (2005).
17. S. Romdhani, J. Ho, T. Vetter, and D. J. Kriegman, "Face recognition using 3-D models: pose and illumination," *Proc. IEEE* **94**, 1977–1999 (2006).
18. J. L. Tribillon, *Corrélation Optique* (Teknéa 1999).
19. B. Javidi and C.-J. Kuo, "Joint transform image correlation using a binary spatial light modulator at the Fourier plane," *Appl. Opt.* **27**, 663–665 (1988).
20. B. Javidi, "Nonlinear joint power spectrum based optical correlation," *Appl. Opt.* **28**, 2358–2367 (1989).
21. M. S. Alam and M. A. Karim, "Fringe-adjusted joint transform correlation," *Appl. Opt.* **32**, 4344–4350 (1993).
22. M. S. Alam and S. F. Goh, "Three-dimensional fringe-adjusted joint transform correlation," *Appl. Opt.* **43**, 3680–3685 (2004).
23. G. Keryer, J. L. de Bougrenet de la Tocnaye, and A. Alfalou, "Performance comparison of ferroelectric liquid-crystal-technology-based coherent optical multichannel correlators," *Appl. Opt.* **36**, 3043–3055 (1997).
24. A. Alfalou, "Implementation of optical multichannel correlators: application to pattern recognition," Ph.D. thesis (Université de Rennes 1—Ecole Nationale Supérieure des Télécommunications Bretagne, 1999).
25. B. V. K. V. Kumar and L. Hassebrook, "Performance measures for correlation filters," *Appl. Opt.* **29**, 2997–3006 (1990).
26. J. L. Horner, "Metrics for assessing pattern-recognition performance," *Appl. Opt.* **31**, 165–166 (1992).
27. J. P. Egan, *Signal Detection Theory and ROC Analysis*, Academic Press Series in Cognition and Perception (Academic, 1975).
28. I. Leonard, A. Alfalou, and C. Brosseau, "Face recognition based on composite correlation filters: analysis of their performances," in *Face Recognition: Methods, Applications and Technology*, A. Quaglia and C. M. Epifano, eds. (Nova, 2012).
29. A. Pe'er, D. Wang, A. W. Lohmann, and A. A. Friesem, "Achromatic optical correlation," *Opt. Lett.* **25**, 776–778 (2000).
30. M. Rahman, A. A. S. Awwal, and K. S. Gudmundsson, "Composite filter for search time reduction for 3D model based object recognition," *Proc. SPIE* **5201**, 97–107 (2003).
31. T. Mohammad, M. Rahman, A. A. S. Awwal, and K. S. Gudmundsson, "Optical pattern recognition of three dimensional images using composite binary phase only filters," *Proc. SPIE* **5556**, 146–157 (2004).
32. D. Casasent and G. Ravichandran, "Advanced distortion-invariant MACE filters," *Appl. Opt.* **31**, 1109–1116 (1992).
33. K. M. Iftekharuddin, F. Ahmad, and M. A. Karim, "Rotation invariant target recognition using amplitude coupled minimum average correlation energy filter," *Opt. Eng.* **35**, 1009–1014 (1996).
34. A. A. S. Awwal, "What can we learn from the shape of a correlation peak for position estimation?" *Appl. Opt.* **49**, B40–B50 (2010).
35. A. A. S. Awwal, M. A. Karim, and S. R. Jahan, "Improved correlation discrimination using an amplitude-modulated phase-only filter," *Appl. Opt.* **29**, 233–236 (1990).
36. A. A. S. Awwal, W. A. McClay, W. S. Fergusson, J. V. Candy, J. T. Salmon, and P. J. Wegner, "Detection and tracking of the back reflection of KDP images in the presence or absence of a phase mask," *Appl. Opt.* **45**, 3038–3048 (2006).
37. K. M. Iftekharuddin, M. A. Karim, P. W. Elloe, and A. A. S. Awwal, "Discretized amplitude modulated phase only filter," *Opt. Laser Technol.* **28**, 93–100 (1996).
38. K. M. Iftekharuddin, M. A. Karim, and A. A. S. Awwal, "Optimization of amplitude modulated inverse filter," *Math. Comput. Model.* **24**, 103–112 (1996).
39. A. Mahalanobis, B. K. V. V. Kumar, S. Song, S. R. F. Sims, and J. F. Epperson, "Unconstrained correlation filters," *Appl. Opt.* **33**, 3751–3759 (1994).
40. H. Zhou and T.-H. Chao, "MACH filter synthesizing for detecting targets in cluttered environment for grayscale optical correlator," *Proc. SPIE* **3715**, 394 (1999).

Chapter

FACE RECOGNITION BASED ON COMPOSITE CORRELATION FILTERS: ANALYSIS OF THEIR PERFORMANCES

I. Leonard¹, A. Alfalou^{1,} and C. Brosseau²*

¹ISEN Brest, Département Optoélectronique, L@bISEN,
20 rue Cuirassé Bretagne, Brest Cedex 2, France

²Université Européenne de Bretagne, Université de Brest,
Lab-STICC, Brest Cedex 3, France

ABSTRACT

This chapter complements our paper: "Spectral optimized asymmetric segmented phase-only correlation filter ASPOF filter" published in *Applied Optics* (2012).

1. INTRODUCTION

Intense interest in optical correlation techniques over a prolonged period has focused substantially on the filter designs for optical correlators and, in particular, on their important role in imaging systems using coherent light because of their unique and quite specific features. These techniques represent a powerful tool for target tracking and identification [1].

In particular, the field of face recognition has matured and enabled various technologically important applications including classification, access control, biometrics, and security systems. However, with security (e.g. fight terrorism) and privacy (e.g. home access) requirements, there is a need to improve on existing techniques in order to fully satisfy these requirements.

In parallel with experimental progress, the theory and simulation of face recognition techniques has advanced greatly, allowing, for example, for modeling of the attendant

* E-mail: ayman.al-falou@isen.fr.

variability in imaging parameters such as sensor noise, viewing distance, emotion recognition facial expressions, head tilt, scale and rotation of the face in the image plane, and illumination. An ideal real-time recognition system should handle all these problems.

It is within this perspective that we undertake this study. On one hand, we make use of a Vander Lugt correlator (VLC) [2]. On the other hand, we try to optimize correlation filters by considering two points. Firstly, the training base which serves to qualify these filters should contain a large number of reference images from different viewpoints. Secondly, it should correspond to the requirement for real-time functionality. For that specific purpose, our tests are based on composite filters.

The objectives of this chapter are first to give a basic description of the performances of standard composite filters for binary and grayscale images and introduce newly designed ASPOF (asymmetric segmented phase-only filter), and second to examine robustness to noise (especially background noise).

This paper deals with the effect of rotation and background noise problems on the correlation filtering performance. We shall not treat the deeper problem of lighting problems. Phong [3] described methods that are useful to overcome the lighting issue in terms of laboratory observables.

Adapted playgrounds for testing our numerical schemes are binary and grayscale image databases. Each binary image has black background with a white object (letter) on it with dimension 512 x 512 pixels. Without loss of generality, our first tests are based on the capital letters A and V because it is easy to rotate them with a given rotation angle (procedures for other letters are similar).

Next simulations were performed to illustrate how this algorithm can identify a face with grayscale images from the Pointing Head Pose Image Database (PHPID) [4] which is often used to test face recognition algorithms. In this study, we present comprehensive simulation tests using images of five individuals with 39 different images captured for each individual.

We pay special attention to adapting ROC curves for different phase only filters (POFs), for two reasons. Firstly, POFs based correlators and their implementations have been largely studied in the literature, see e.g. [1, 5].

In addition, optoelectronics devices, i.e. spatial light modulators (SLMs) allow implementing optically POFs in a simple manner. Secondly, numerical implementation of correlation have been considered as an alternative to all-optical methods because they show a good compromise between their performance and their simplicity.

High speed and low power numerical processors, e.g. field programmable gate array (FPGA) [6] provide a viable solution to the problem of optical implementation of POFs. Such numerical procedure allows one to reduce the memory size (by decreasing the number of reference images included in the composite filter) and does not consider the amplitude information which can be rapidly varying.

Face identification and underwater mine detection with background noise are two areas for which the FPGA has demonstrated significant performance improvement, such as image registration and feature tracking.

Following this brief introduction, we have divided the rest of the paper as follows: a general overview of the optical correlation methods is given in Sec. 2. Then, in Sec. 3, we review a series of correlation filters, which are next compared in Sec. 4.

2. SOME PRELIMINARY CONSIDERATIONS AND RELATION TO PREVIOUS WORK

The subject of correlation methods is long and quite a story. Here we will review various aspects of the problem discussed in the literature which relate to this paper. The modern study of optical correlation can be traced back to the pioneering research in the 1960s [2, 7]. In what became a classic paper, Vander Lugt presented a description of the coherent matched filter system, i.e. the VLC [2]. Basically, this method is based on the comparison between a target image and a reference image. This technique consists in multiplying an input signal (spectrum of image to be recognized) by a correlation filter, originating from a training base (i.e. reference base), in the Fourier domain. The result is a correlation peak (located at the center of the output plane i.e. correlation plane) more or less intense, depending on the degree of similarity between the target image and reference image. Correlation is perceived like a filtering which aims to extract the relevant information in order to recognize a pattern in a complex scene.

However, this approach requires considerable correlation data and is difficult to realize in real time. This led to the concept of POF (carried out from a single reference) whose purpose is to decide if a particular object is present or not, in the scene. To have a reliable decision about the presence, or not, of an object in a given scene, we must correlate the latter with several correlation filters taking into account the possible modifications of the target object, e.g. in-plane rotation and scale. Perhaps more problematic is the fact that a simple decision based on the presence, or not, of a correlation peak is insufficient. Thus, use of adequate performance criteria such as those developed in [8-9] is necessary.

During the 1970s and 1980s correlation techniques developed at a rapid pace. A plethora of advanced composite filters [10-12], and more general multi-correlation approaches [13] have been introduced. A good source for such results is the book of Yu [14]. However, experimental state of the art shows that optical correlation techniques almost found themselves in oblivion in the late 1990s for many reasons. While numerous schemes for realizing all-optical correlation methods have been proposed [13-15], up to now, they all face technical challenge to implement, notably those using spatial light modulators (SLMs) [16] because these methods are very sensitive to even small changes in the reference image. In addition, they usually require a lot of correlation data and are difficult to realize in real time.

Over the last decade, there has been a resurgence of interest, driven by recognition and identification applications [17-22], of the correlation methods. For example, Alam *et al.* [22] demonstrated the good performances of the correlation method compared to all numerical ones based on the independent component model. Another significant example in this area of research is the work by Romdhani *et al.* [23], which compared face recognition algorithms with respect to those based on correlation. Other recent efforts include the review by Sinha *et al.* [24] dealing with the current understanding regarding how humans recognize faces. Riedel *et al.* [25] have used the minimum average correlation energy (MACE) and unconstrained MACE filters in conjunction with two correlation plane performance measures to determine the effectiveness of correlation filtering in relation to facial recognition login access control.

Wavelets provide another efficient biometric approach for facial recognition with correlation filters [26]. A photorefractive Wiener-like correlation filter was also introduced by Khoury *et al.* [27] to increase the performance and robustness of the technique of correlation

filtering. Their correlation results showed that for high levels of noise this filter has a peak-to-noise ratio that is larger than that of the POF while still preserving a correlation peak that is almost as high as that of the POF. Another optimization approach in the design of correlation filters was addressed to deal with the ability to suppress clutter and noise, an easy detection of the correlation peak, and distortion tolerance [28]. The resulting maximum average correlation height (MACH) filter exhibit superior distortion tolerance while retaining the attractive features of their predecessors such as the minimum average correlation energy filter and the minimum variance synthetic discriminant function filter.

A variant of the MACH filter was also developed in [29]. Pe'er and co-workers [30] presented a new apochromatic correlator, in which the scaling error has three zero crossings, thus leading to significant improvement in performance. These references are far from a complete list of important advances, but fortunately the interested reader can easily trace the historical evolution of these ideas with Vijaya Kumar's review paper, Yu's book, and the chapter of Alfalou and Brosseau containing an extensive bibliography [1, 14-15, 31]. As mentioned above, we have a dual goal which is first to introduce standard correlation filters, and second to compare their performances.

3. A BRIEF OVERVIEW OF CORRELATION FILTERS

First we present the most common correlation filters. We turn attention to the general merits and drawbacks of composite filters. This discussion is simply a brief review and tabulation of the technical details for the basic composite filters. For that purpose we consider a scene s containing a single or several objects o with noise b . The input scene is written as $s(x, y) = o(x, y) + b(x, y)$. Let its two-dimensional FT be denoted by $S(\mu, \nu) = \rho(\mu, \nu) \exp[i\theta(\mu, \nu)]$. In the Fourier plane of the optical set-up, the scene spectrum is multiplied by a filter $H(\mu, \nu)$, where μ and ν denote the spatial frequencies coordinates.

Many approaches for designing filters to be used with optical correlators can be found in the literature according to the specific objects that need to be recognized. Some have been proposed to address hardware limitations; others were suggested to optimize a merit function. Attempts will be made throughout to use a consistent notation.

3.1. Adapted Filter (Ad)

The Ad filter [2] has for main purpose to optimize the SNR and reads

$$H_{ad}(\mu, \nu) = \alpha \frac{R^*(\mu, \nu)}{\Gamma_b(\mu, \nu)} \quad (1)$$

where α denotes a constant, $R^*(\mu, \nu)$ is the complex conjugate of the spectrum of the reference image ($R(\mu, \nu) = \rho_0(\mu, \nu) \exp(i\theta_0(\mu, \nu))$), and $\Gamma_b(\mu, \nu)$ represents the spectral density of the

input noise. If we assume that the noise is white and unit spectral density, we obtain $H_{Ad}(\mu, \nu) = \alpha R^*(\mu, \nu)$. A main advantage of this filter is the increase of the SNR especially when white noise is present. The drawback of this filter is that it leads to broad correlation peaks in the correlation plane. Since the output plane is scanned for this peak, and its location indicates the position of the target in the input scene, we can conclude that the target is poorly localized. In addition, its discriminating ability is weak.

3.2. Phase-Only Filter (POF)

The phase is of paramount importance for optical processing with coherent light [32]. For example, Horner and Gianino [33] suggested a correlation filter which depends only on the phase of a reference image (with which the scene is compared). Without loss of generality, this POF is readily expressible as

$$H_{POF}(\mu, \nu) = \frac{R^*(\mu, \nu)}{|R(\mu, \nu)|} = \exp(-i\theta_0(\mu, \nu)) \quad (2)$$

The main feature of the POF is to increase the optical efficiency η . It is worthy to note that Eq. (2) depends only the phase of the reference. Besides the ability to get very narrow correlation peaks, POF have another feature that Ad filters lack: the capacity for discriminating objects. Because POF use only the reference's phase, they can be useful as edge detector. However, as is well known the POF is very sensitive to even small changes in rotation, scale and noise contained in the target image [34].

3.3. Binary Phase-Only Filter (BPOF)

We consider next the binarized version of the phase-only filter [35], or alternatively defined as a two-phase filter where the only allowed values are 1 and -1 such as

$$H_{BPOF} = \begin{cases} 1 & \text{if the real part of POF filter} \geq 0 \\ -1 & \text{otherwise} \end{cases} \quad (3)$$

Other definitions of BPOF were also considered by Vijaya Kumar [36]. Generally, BPOF have weaker performances than POF. It is helpful in certain applications for which the size of the filter should be small and also for optical implementation. Like POF, BPOF is very sensitive to rotation, scale, and noise in the target images.

3.4. Inverse Filter (IF)

IF [37-38] is defined as the ratio of POF by the magnitude of the reference image spectrum, and can be expressed as

$$H_{IF}(\mu, \nu) = \frac{R^*(\mu, \nu)}{|R(\mu, \nu)|^2} = \frac{\exp(-i\theta_0(\mu, \nu))}{\rho_0(\mu, \nu)} \quad (4)$$

The main advantage of this filter is to minimize the correlation peak width, or in other words, to maximize the PCE. It has the desirable property of being very discriminating. Despite this, an IF has a number of drawbacks. It is very sensitive to deformation and noise contained in the target image with respect to the reference image.

3.5. Compromise Optimal Filter (OT)

To realize a good correlation, the filter should be discriminating and robust. A filter showing a trade-off between these two properties was suggested in [39]. The OT filter is conveniently written out as

$$H_{OT}(\mu, \nu) = \frac{R^*(\mu, \nu)}{\alpha |R(\mu, \nu)|^2 + (1-\alpha) |R(\mu, \nu)|^2} = \frac{\rho_0 \exp(-i\theta_0)}{\alpha |\rho_0|^2 + (1-\alpha) |\rho_0|^2} \quad (5)$$

where α denotes a discrimination and robustness degree. If α is set to zero, Eq. (5) yields the inverse filter, while the adapted filter is recovered when α is equal to one.

3.6. Classical Composite Filter (COMP)

In general, taking a decision based on a single correlation obtained by comparing the target image with only one filter, i.e. single reference, does not allow getting a reliable identification [31]. To alleviate the problems associated with this drawback, multi-correlation approaches have been suggested. One way to realize multi-correlation within the VLC configuration is by employed the classical composite filter (COMP). The basic idea consists in merging several references by linearly combining them such as

$$H_{Comp}(\mu, \nu) = \sum_{i=1}^M R_i(\mu, \nu) \quad (6)$$

where $R_i(\mu, \nu)$ denotes each reference spectrum. Observe that a weighing factor can be used in some cases for specific purpose [13].

3.7. Segmented Composite Filter (SPOF)

For the purpose of reducing the number of correlation requested to take a reliable decision, the number of references in the filter should be increased. However, increasing the latter has for effect to induce a local saturation phenomenon in a classical composite filter [5]. This can be remedied by use of a recently proposed spectral multiplexing method [5]. This

method consists in suppressing the high saturation regions of the reference images. Briefly stated, this is achieved through two steps [5]. First, a segmentation of the spectral plane of the correlation filter is realized into several independent regions. Second, each region is assigned to a single reference. This assignment is done according a specific energy criterion

$$\frac{E_{u,v}^l}{\sum_{i,j} E_{i,j}^l} > \frac{E_{u,v}^k}{\sum_{i,j} E_{i,j}^k} \quad (7)$$

This criterion compares the energy (normalized by the total energy of the spectrum) for each frequency of a given reference with the corresponding energies of another reference.

Assignment of a region to one of the two references is done according Eq. (7). Hence, the SPOF contains frequencies with the largest energy.

3.8. Minimum Average Correlation Energy (MACE)

For good location accuracy in the correlation plane and discrimination, we need to design filters capable of producing sharp correlation peaks. One method [12] to realize such filters is to minimize the average correlation plane energy that results from the training images, while constraining the value at the correlation origin to certain prespecified values.

This leads to the MACE filter which can be expressed in the following compact form:

$$H_{MACE} = D^{-1} S (S^+ D^{-1} S)^{-1} c^+ \quad (8)$$

where D is a diagonal matrix of size $d \times d$, (d is the number of pixels in the image) containing the average correlation energies of the training images across its diagonals; S is a matrix of size $N \times d$ where N is the number of training images and $+$ is the notation for complex conjugate.

The columns of the matrix S represent the Discrete Fourier coefficients for a particular training image. The column vector c of size N contains the correlation peak constraint values for a series of training images. These values are normally set to 1 for images of the same class [14]. A MACE filter produces outputs that exhibit sharp correlation peaks and ease the peak detection process.

However, there is no noise tolerance built into these filters. In addition, it appears that these filters are more sensitive to intraclass variations than other composite filters [40].

3.9. Amplitude-Modulated Phase-Only Filter (AMPOF)

Awwal *et al.* [21, 28] suggested an optimization of the POF filter based on the following idea: if the correlation plane of the POF spreads large, it yields a correlation peak described by a Dirac function. One way to realize this has been put forward in Ref. [21, 28], where the authors suggested the amplitude-modulated phase-only filter (H_{AMPOF})

$$H_{AMPOF}(\mu, \nu) = \frac{D \exp(-j\phi(\mu, \nu))}{|F(\mu, \nu)| + a} \quad (9)$$

where $F(\mu, \nu) = |F(\mu, \nu)| \exp(j\phi(\mu, \nu))$ is the reference image spectrum, μ and ν denote the spatial frequencies, D is a parameter within the range $[0, 1]$, and the factor a ($a \leq D$) appearing in the denominator is useful in overcoming the indeterminate condition and ensuring that the gain is less than unity. It can be a constant or a function of μ and ν , and thus can be used to either suppress noise or bandlimit the filter or both.

3.10. Optimal Trade-off MACH (OT MACH)

Another optimization approach in the design of correlation filters was addressed to deal with the ability to suppress clutter and noise, an easy detection of the correlation peak, and distortion tolerance [28].

The resulting maximum average correlation height (MACH) filter exhibit superior distortion tolerance while retaining the attractive features of their predecessors such as the minimum average correlation energy filter and the minimum variance synthetic discriminant function filter. A variant of the MACH filter was also developed in [29], i.e. the optimal trade-off MACH filter which can be written as

$$H_{OTMACH} = \frac{m_x^*}{\alpha C + \beta D_x + \gamma S_x} \quad (10)$$

where m_x is the average of the training image vectors, C is the diagonal power spectral density matrix of the additive input noise, D_x is the diagonal average power spectral density of the training images, S_x denotes the similarity matrix of the training images, and α , β , and γ are three numerical coefficients.

3.11. Asymmetric Segmented Phase Only Filter (ASPOF)

The last filter which is presented in this chapter is the ASPOF. See, e.g. [41], for its definition. The reference image database is divided in two sub-databases (with reference to Figure1).

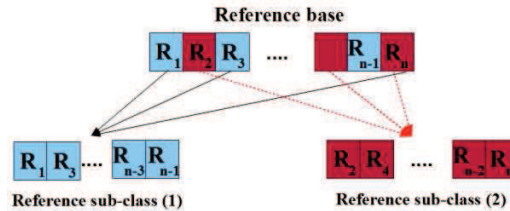


Figure 1. Technique used to separate the reference images into 2 sub-classes [41].

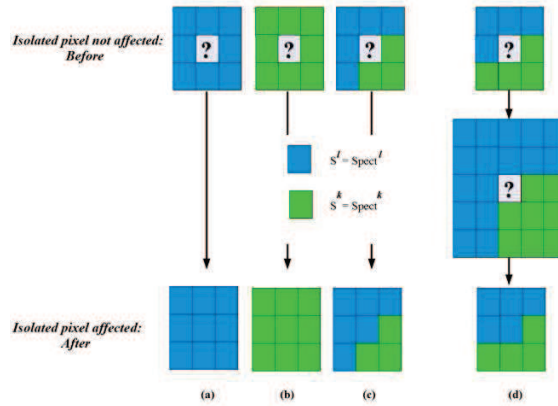


Figure 2. Illustrating the optimized assignment procedure for isolated pixels [41].

A SPOF is constructed from each of these databases according to the criterion defined by Eq.11.

$$\frac{E_{\mu,\theta}^l}{\sum_{i,j} E_{\mu,\theta}^l} \wedge \frac{E_{\mu,\theta}^k}{\sum_{i,j} E_{\mu,\theta}^k} \quad (11)$$

Pixels which are not assigned using Eq.(11) are further assigned to the majority reference in the pixel's neighborhood (see Figure2).

4. COMPARATIVE STUDY OF COMPOSITE CORRELATIONS FILTERS WITH BINARY IMAGES

Much research has been devoted to discovering new composite filters with higher efficiencies. An extensive review of composite filters has been found to be given by, where much can be found about distortion-invariant optical pattern recognition.

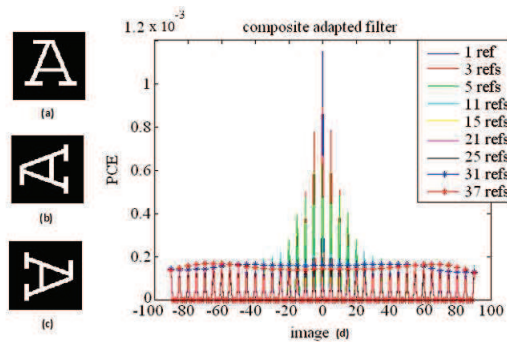


Figure 3. (Color online) Binary image for the uppercase letter A in the English alphabet: (a) standard, (b) 90° counterclockwise rotation, (c) the same as a 90° clockwise rotation, (d) PCEs obtained with the composite adapted filter. The colors shown in the inset denote the different composite adapted filters depending on the number of references used.

Table 1. Illustrating the Different Composite Filters Used

Composite filter	Notation	Equation
Adapted filter	H_{compAd}	(1)
Phase-only filter	$H_{compPOF}^1, H_{compPOF}^2$	(2)
Binary phase-only filter	$H_{compBPOF}^1, H_{compBPOF}^2$	(3)
Inverse filter	H_{compIF}	(4)
Compromise optimal filter	H_{compOT}	(5)
Segmented filter	H_{SPOF}^1, H_{SPOF}^2	(7)
Segmented binary filter	H_{BSPOF}^1, H_{BSPOF}^2	(7)
Minimum average correlation energy filter	$H_{compMACE}$	(8)
Amplitude modulated phase-only filter	$H_{compAMPOF}$	(9)
Optimal trade off MACH	$H_{compOTMACH}$	(10)
Asymmetric segmented phase only filter	$H_{compASPOF}$	(11)

H_{compAd} denotes the Adapted composite filter. This later is realized by considering a linear combination of reference images, and then using the adapted filter definition (Eq. (1)). $H_{comp-POF}$ is the POF composite filter. We tested two different schemes for realizing the Composite POF filter. In the first scheme ($H_{compPOF}^1$) we used a linear combination of reference images to create the POF, i.e. Eq. (2). The second scheme ($H_{compPOF}^2$) involves performing the POF, via Eq. (2), for each reference, and then using the linear combination of these POFs. $H_{compBPOF}^1$ and $H_{compBPOF}^2$ are the binarized versions of the filters $H_{compPOF}^1$ and $H_{compPOF}^2$ obtained from Eq. (3), respectively. The composite inverse filter H_{compIF} is the inverse filter (Eq. (4)) of the linear combination of reference images. The optimal composite filter H_{compOT} is realized by linearly combining reference images (Eq. (5)). $H_{compSPOF}^1$ denotes the segmented filter realized by doing segmentation and assignment with the energy criterion (Eq. (7)). The calculation of filter $H_{compSPOF}^2$ is done by replacing the energy in Eq. (7) with the square of the real part of the different references spectra to be merged. $H_{compMACE}$ is the composite filter of the MACE filter developed in Eq. (8). $H_{compAMPOF}$ is the composite version of AMPOF (Eq. (9)). $H_{compOTMACH}$ is the composite version of OTMACH (Eq.(10)). $H_{compASPOF}$ is the ASPOF (Eq.(11)). [41].

In particular, there are many other facets of composite filters not mentioned in section (3). In a general context, it is instructive to compare the performance of a selection of composite filters described in section (3). To aid the reader of this section, we briefly recap the filters characteristics and some of our terminology in Table 1. The main goal of this section is to identify the parameters which introduce limitations in the performances of these composites filters with and without noise in the input plane.

The binary (black and white) images from Figure 3 (a)-(c) were chosen for testing the composite filters because they are easier to process and analyze than gray level images, and the letter base can be digitized under controlled conditions, i.e. easy to process morphological operation and addition of input noise. Each image has black background with a white object (letter) on it with dimension 512×512 pixels. Here, we will limit ourselves to a data-base by rotating the A image (Figure 3(a)) in increments of 1° counter clockwise to get 181 images.

We now compare in a systematic way the performances of the composite filters of Table I for the data-base displayed in Figure 1(a)-(c). In performing this comparison a normalization of the correlation planes was realized. An illustration of the effects of the number of reference images (typically ranging from 1 to 37) employed to realize the composite filter on rotation of the input image will also be given.

4.1. Adapted Composite Filter

We start our discussion by considering the adapted composite filter in the Fourier plane of the VLC. Figure 3 (d) shows PCE by introducing every image of the data-base of 181 images, one by one, in the entrance plane. Each curve of Figure 1 (d) has a specific color which depends on the number of references used to realize the adapted filter, e.g. the red one considers a 3-reference filter (-5° , 0° , and 5°).

As expected the adapted composite filter is robust against rotation. It is also worthy to observe that the energy contained in the correlation peak decreases as the number of references chosen to realize the filter is raised. This decrease is detrimental to the usefulness of this type of composite filter. Its low discriminating character is more and more visible as the number of references is increased. This is consistent with previous studies [13].

4.2. Composite POF

Figure 4 (a) shows the PCE results for the composite POF $H_{compPOF}^1$. As described previously, $H_{compPOF}^1$ is realized by considering a linear combination of reference images (ranging from 1 to 37) to create a composite image.

The input images are then correlated with this filter. We find that the energy contained in the correlation peak decreases significantly, i.e. the PCE is decreased by a factor of 3 when using a POF containing 3 references by contrast with a POF realized with a single reference. For a 11-reference POF, the PCE is decreased by an order of magnitude which renders unreliable the decision on the letter identification. For 3 references only the images forming the filter are recognized.

However, beyond 11 references, the weakness of the magnitude of the PCE makes the recognition of the images forming the filter very difficult. Figure 4 (b) shows the correlation obtained with filter which is obtained by linearly combining the different POFs of different reference images.

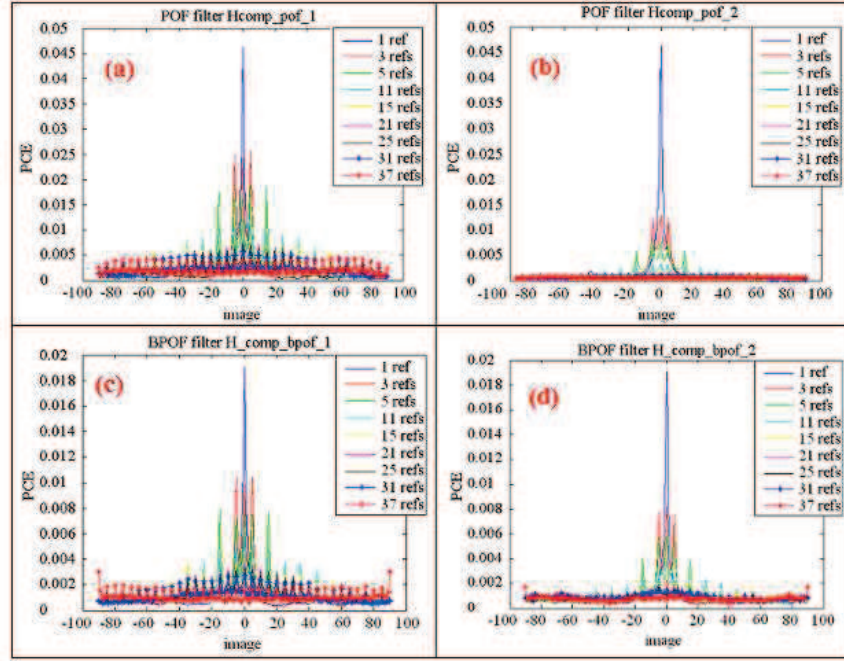


Figure 4. (Color online) (a) PCEs obtained with the POF composite filter. The colors shown in the inset denote the different filters depending on the number of references used. (b) Same as in (a) for filter $H_{comp_pof_2}$. (c) and (d) Same as in (a) and (b) for Binary filter.

The magnitude of the PCE decreases with raising the number of reference images of the filter. From the point of view of recognition application it appears that the saturation problem is more serious than that obtained with filter $H_{comp_POF}^1$, i.e. it is difficult to recognize a letter with a filter composed of more than 5 reference images even if the letter to be recognized belongs to the set of reference images. Thus, the overall performance for letter identification using this correlation technique decrease by employing filter $H_{comp_POF}^2$.

From the combined observations above, an especially meaningful feature emerges: to get a reliable decision, a 3-reference POF should be used. One of the distinctive features shown in Figure 4 (a) is that this filter allows one to recognize the letter A only taking a range for angle of rotation from -7° to 7° . Recognition of the full base requires fabricating at least 12 POFs, each having 3 references. Hence, this procedure cannot permit significant reduction in the time of decision since other phenomena can also affect the target image, e.g. scale.

4.3. Composite Binary POF

Binarized POF in the Fourier domain (Eq. (3)) is an alternative to POF. Figure 4 (c) (resp. Figure 4 (d)) shows PCE results obtained by binarization of $H_{comp_POF}^1$ (resp. $H_{comp_POF}^2$). Our calculations shown in these two graphs can be discussed in the same way as was done for Figs. 2 (a) and (b). At the same time, a comparison between Figs. 4 (a) and (b) and Figs. 4 (c)

and (d) indicates a decrease of the PCE values. This is reminiscent of the noise induced by the binarization protocol.

4.4. Inverse Composite Filter

It has been known for a while that the inverse filter shows a strong discriminating ability and a low robustness against small changes of the target image with respect to the reference image. In practice, is realized by defining the inverse filter of the linear combination of different references.

In Figure 5 we plot the corresponding PCE values for the letter base A correlated with filter and different numbers of references. These results are consistent with our previous observation of the PCE decrease as the number of references is raised. We also check that these simulations are consistent with the above mentioned characteristics of the inverse filter. Indeed, correlation vanishes even when the target image is identical to one of the reference images used to realize the filter.

Up to now, our results show that filter has the best performance among the selected composite filters studied so far. To orient the subsequent discussion, we show the good discriminating ability of the composite POF, with parameters chosen for comparison with the above-described data. Our previous calculations suggest that the more discriminating efficiency of the filter is associated with the weaker false alarm rate. For that specific purpose, the letter V base (Figure6 (a)), i.e. constituted by 19 images obtained by rotating the V every 10° , was correlated with filter realized with reference images of letter A. Although the letter V has a great similarity with the letter A with a 180° rotation, it is easily seen that the different composite POFs do not recognize V as being an A since no false alarm can be detected (Figure6 (b)). Nominal values of PCE are less than 0.002, ca. over 20 times less than the maximum value seen in Figure 4 (c). This is a clear indication of the good discriminating ability of the POF.

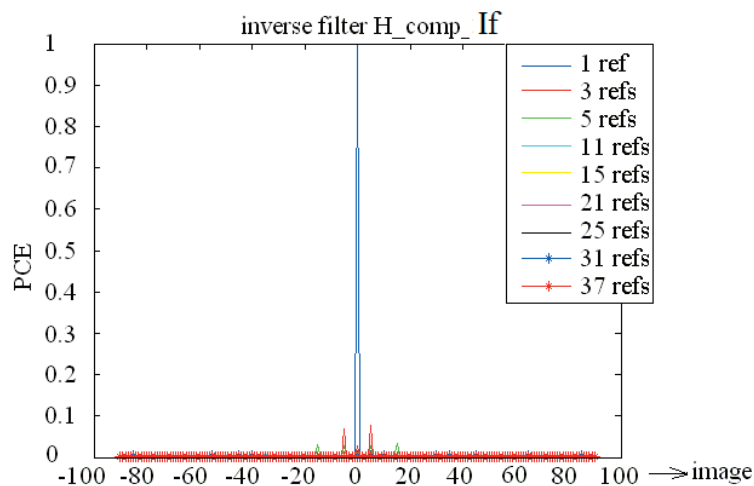


Figure 5. (Color online) PCEs obtained with the inverse composite filter. The colors shown in the inset denote the different adapted filters depending on the number of references used [41].

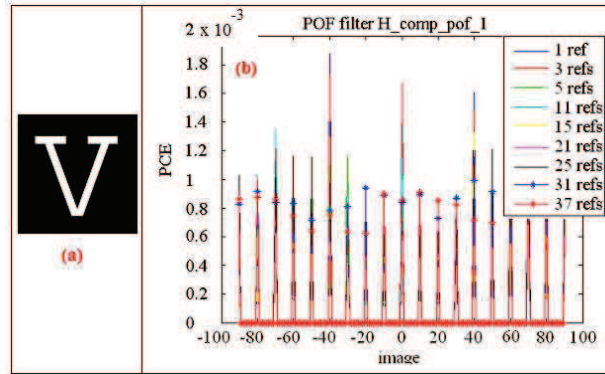


Figure 6. (Color online) Discrimination tests: (a) The target image considered is the letter V. (b) PCEs obtained with a filter realized with reference images of letter A. The colors shown in the inset denote the different adapted filters depending on the number of references used [41].

4.5. Robustness against Noise

In realistic object recognition situations, some degree of noise is unavoidable. A second series of calculations was conducted in which standard noise types were added to the target image. In this section, we shall mainly consider the compromise optimal filter (OT) H_{compOT} since it represents a useful trade-off between adapted and inverse filters. Its tolerance to noise is also remarkable. Throughout this section, our calculations will be compared with results obtained with filter $H_{compPOF}^1$. At a first look at the performance of the OT filter with noise, we consider the special case of background noise, i.e. the black background is replaced by the gray texture shown in Figure 7 (a). Figure 7 (b) shows the uppercase letter A with rotation (-45°).

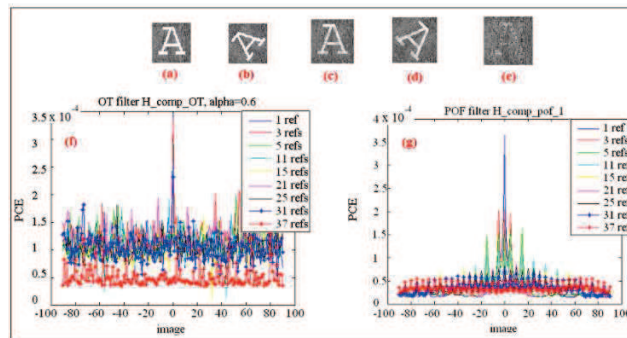


Figure 7. (Color online) (a) Illustrating the letter A with additive background structured noise. (b) Same as in (a) with a rotation angle of -50° . (c) Illustrating the letter with structured noise. (d) Same as in (c) with a rotation angle of 50° . (e) Illustrating the letter A for a weak contrast. (f) PCEs obtained with the OT composite filter taking $\alpha=0.6$. The colors shown in the inset denote the different adapted filters depending on the number of references used. (g) Corresponding PCEs for a POF. The colors shown in the inset denote the different adapted filters depending on the number of references used.

Figure 7 (f) shows that the filter OT can recognize this letter only for a noisy image oriented at 0° . The results indicate that PCE decreases as α is increased. If α is set to zero, this filter cannot recognize any letter.


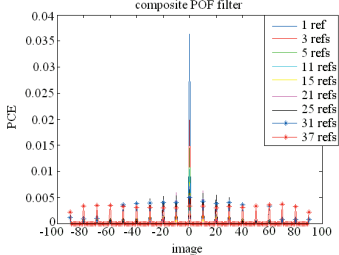

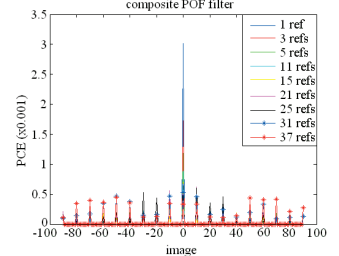

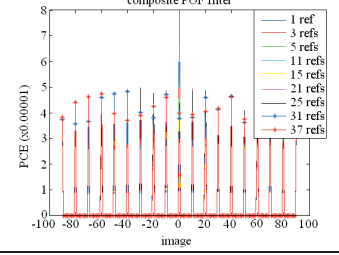
We also observe that the filter OT is not robust to image rotation when the images are noisy, especially if the noise cannot be explicitly evaluated.

One of the reasons we will not pursue the characterization of this filter stems from the fact that the input noise cannot be always determined in a real scene. We now exemplify the effect of background noise (applying an analysis similar to that above) by evaluating the

performance of the POF ($H^1_{compPOF}$). A noise was also added in the white part of the letter (with reference to Figs. 7 (c) and (d)). As illustrated in Figure 7 (g), POF is more robust to background noise than filter OT. As mentioned previously, this is consistent with the good discrimination ability of the composite POF filters. One interesting result is that the performances of composite filters decrease when the input image is weakly contrasted with respect to background, as evidenced in Figure 7 (e).

In another set of calculations, we considered the case of a Gaussian white noise on the composite POF for which the expectation value can be 0 or 1, and its variance can be set to 0, 0.1 and 1 (Table 2).

Table 2. Calculated correlation results (third row) obtained with different composite POFs. The first row considers the numerical characteristics of the white Gaussian noise used. The second row shows a typical realization of the noisy images

White centered Gaussian noise, with variance set to 0.1		
White centered Gaussian noise, with variance set to 1		
White centered Gaussian noise, with expectation value set to 1 and variance equal to 0.1		

Examples of noisy images are shown in the second row of Table II. Insight is gained by observing in the third row of Table 2 how the correlation results vary for different composite POFs realized with noise free reference images. As was evidenced for the standard POF, composite POFs show robustness to noise, i.e. we were able to identify noisy images using filter $H_{compPOF}^1$. However, it is apparent that only noisy images which have been rotated with similar angles to the reference images have been identified.

As we have seen so far, the compromise optimal filter is robust to noise when the latter is clearly identified. However, the performance of POF is better when the characteristics of noise are unknown. It is also important to point out that the performance of both composite filters decrease when the number of reference images forming the filter is increased. It should be emphasized once more that this effect is more likely when the images are noisy.

4.6. Optimized Composite Filters

Next, we are interested in the design of an asymmetric segmented composite phase-only filter whose performance against rotation will be compared to the MACE filter, POF, SPOF and AMPOF. To illustrate the basic idea, let us consider composite filters which are constructed by using 10 reference images obtained by rotating the target image by 0° , -5° , $+5^\circ$, -10° , $+10^\circ$, -20° , $+20^\circ$, $+25^\circ$, respectively.

To begin our analysis, we consider the composite filter MACE. Figure 8 presents correlation results of the letter base A (data-base obtained by rotating the A image in increments of 1° within the range $(-90^\circ, +90^\circ)$) with a composite MACE filter containing 10 reference images (0° , -5° , $+5^\circ$, -10° , $+10^\circ$, -20° , $+20^\circ$, $+25^\circ$). Here, the basic purpose is to recognize the letter A even when it is rotated with an angle ranging from -20° to 25° . In the angular dependence of the PCE value shown in Figure 8, we can distinguish three regions exhibiting distinct correlation characteristics (referred to as A, B and C, respectively). One notices in Figure 8 that if we restrict ourselves to region B only, correlation appears when the target image is similar to one of the reference images (Figure 8). No correlation is observed in regions A and C of Figure 8. The MACE composite filter is weakly robust to structured noise. Another example is shown in Figure 9 (a) when a centered Gaussian noise of variance 0.1 is added to the input image. This figure shows the sensitivity of the MACE composite filter against this type of noise. In fact, it gives lower PCE values even with a low noise level.

Figure 9 (b) shows the results for the filter MACE with a structured background noise. With reference again to Figure 9 (b) no correlation were observed even in the angular region ranging from -20° to 25° suggesting the poor correlation performances of filter MACE. We have also confirmed that the MACE composite filter is very sensitive to noise, and especially to structured noise. For this reason, we will not pursue the study of this filter in the remainder of this paper. The preceding analysis prompted us to study the composite filter performances based on different optimized versions of the POF, i.e. filters H_{SPOF} , $H_{compAMPOF}$ and $H_{compASPOF}$. Here we reinvestigate the identification problem of letter A in the angular region ranging from -20° to 25° by considering a 10-reference composite filter. Furthermore, we shall compare these results with those obtained using the classic composite filter $H_{compPOF}^1$.

Parenthetically, there are similarities between the PCE calculations obtained for filters and $H_{compPOF}^2$ with those based on filters and $H_{compPOF}^1$.

Our illustrative correlation calculations for filter and the letter base A (data-base obtained by rotating the A image in increments of 1° within the range $(-90^\circ, +90^\circ)$) are given in Figure 10 (a) and (b). Shown in this figure are the PCEs for the composite POF (blue curve), the segmented composite POF (red curve), the composite AMPOF (black curve) and the composite ASPOF (green curve). We first note, in Figure 10 (a) that the PCE values for the composite ASPOF are larger than the corresponding values when the optimization stage (see Figure 2) has been applied to the filter. When the optimization stage has not been applied, the ASPOF PCE values are similar to the SPOF PCE values, see Figure 10 (b) [41]. Otherwise, even if the PCEs for the composite AMPOF are larger than those for the two other filters, there is a range of rotation angle, i.e. region A, for which the segmented filter shows correlation. Also apparent is that the PCE values calculated for the segmented filter H_{SPOF}^1 are larger than the corresponding values of the classical composite POF in the correlation region A.

In this region A, we observe large variations of the PCE values, but all the correlation values are larger than the PCE values obtained in the no-correlation regions B and C. Maximal PCE values correspond to auto-correlation of the 10 reference images. Outside the A region correlation deteriorates rapidly. From these simulations, we concluded that it is difficult to identify the letter for the three filters considered. The PCE results show significant dependence on the rotation of the target image with respect to the reference images for composite AMPOF.

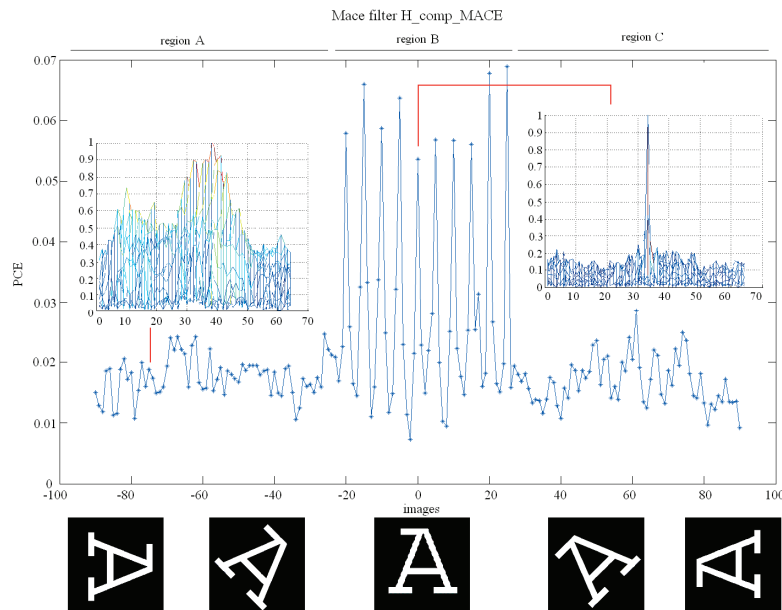


Figure 8. (Color online) PCEs obtained with a 10-reference MACE when the target images are noise free. Several examples of the rotated letter A are illustrated at the bottom of the figure. The insets show two correlation planes: (right) autocorrelation obtained without rotation, (left) inter-correlation obtained with the letter A oriented at -75° [41].

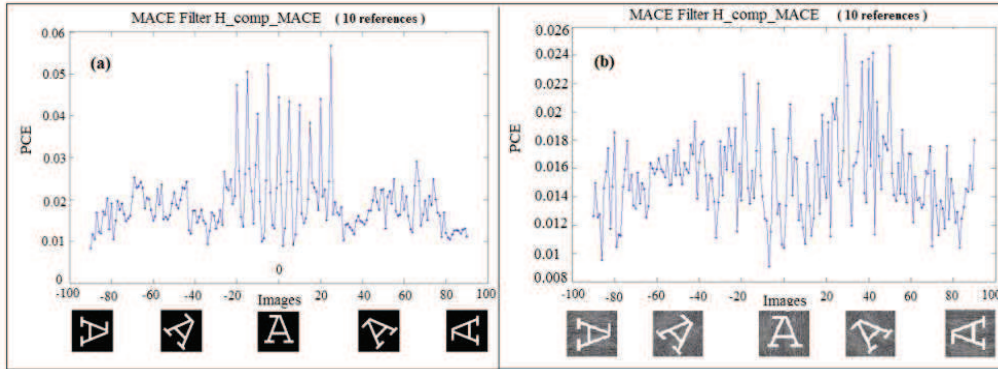


Figure 9. (Color online) (a) PCEs obtained with Mac composite filter and additive Gaussian centered noise of variance 0.1. (b) Same as in (a) with additive background structured noise.

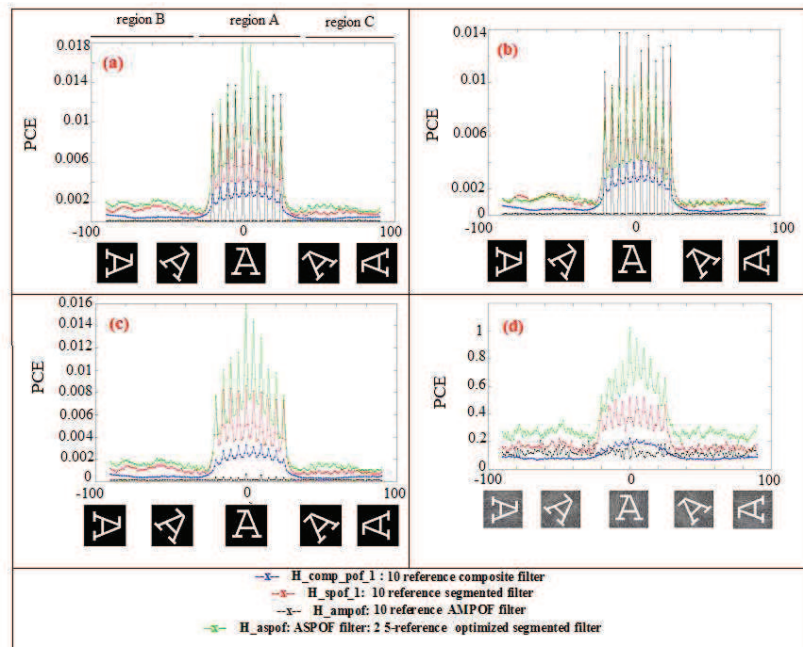


Figure 10. (Color online) Comparison between the different correlations of letter A (we consider rotation angles ranging between -90° and 90°) with the 10-reference composite filters: POF (blue line), Segmented (red line), AMPOF (black line), and ASPOF (green line). (a) PCEs obtained using the optimization stage concerning the isolated pixels, (b) PCEs obtained without the optimization stage concerning the isolated pixels. (c) and (d) represent the PCEs obtained with noised target images.

Having discussed image rotation dependence without noise of the composite filters response we now determine the impact of noise. For this purpose, we applied two types of noise to the target image, either background structured noise (Figure 7 (a)), or a centered white noise with variance set to 0.01. Interestingly, one can see in Figs. 10 (c) and (d) the results of the PCE calculations which show the good performance of asymmetric segmented

filter $H_{COMPASPOF}$. Even when noise is present, the ASPOF yields correlation in region A. By contrast, there is no correlation in the A region with the AMPOF composite filter.

However, identification of the full letter data-base requires an increase of the reference images. This leads to the decay of the segmented filter's performance. Interestingly, Figure 11 indicates that the segmented filter's performance is very sensitive to the number of references forming the filter. We also studied the effect of binarization on the performances of the segmented composite filter. In fact, this binarization can be an effective solution to reduce the memory size to store these filters without altering the efficiency of the decision.

To further show the interest in using a segmented filter with respect to the saturation problem which affects the classical composite filter, we show in Figure 12 (b) the 8-bit image of the sum (without segmentation) of the three spectra corresponding to the reference images. Figure 12 (c) shows the corresponding sum with segmentation. Our calculations clearly indicate that the image with segmentation shows significantly less saturation than that obtained without segmentation.

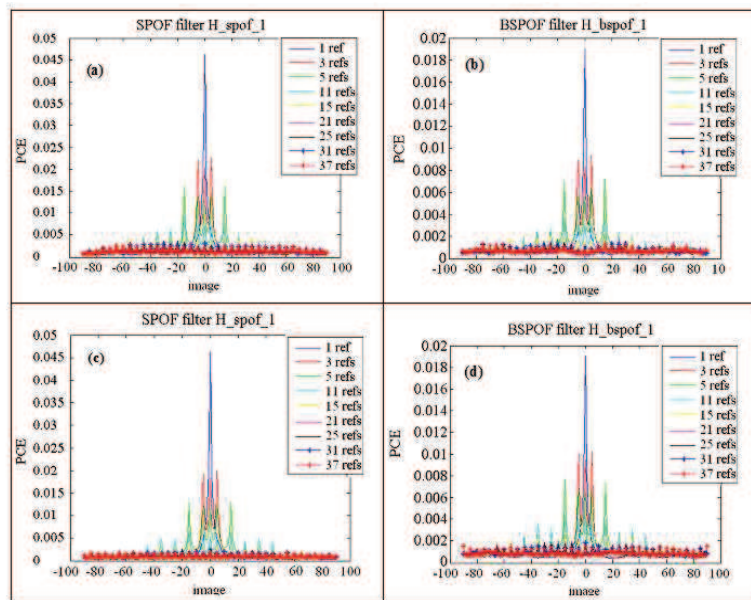


Figure 11. (Color online) PCEs obtained with a segmented composite filter : (a) using the energy criterion, (b) using the segmented binarized filter, (c) using filter the real part criterion, (d) corresponding binarized filter to (c).

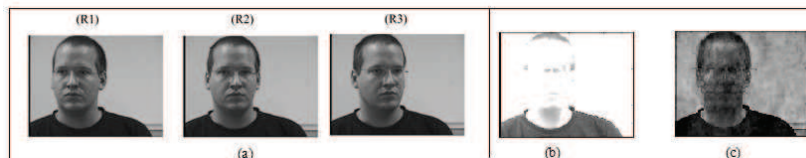


Figure 12. Illustrating the saturation effect: (a) three 8-bit grey scale images. (b) Image obtained by a classical linear combination of the three images shown in (a). (c) Image obtained using an optimized merging (spectral segmentation).

CONCLUSION

We now conclude with a brief discussion of the robustness of the ASPOF. In Figure 13, we have represented the ROC curves obtained with filters (Composite-POF, SPOF, AMPOF and ASPOF) containing each 10 references (from -60° to $+60^\circ$).

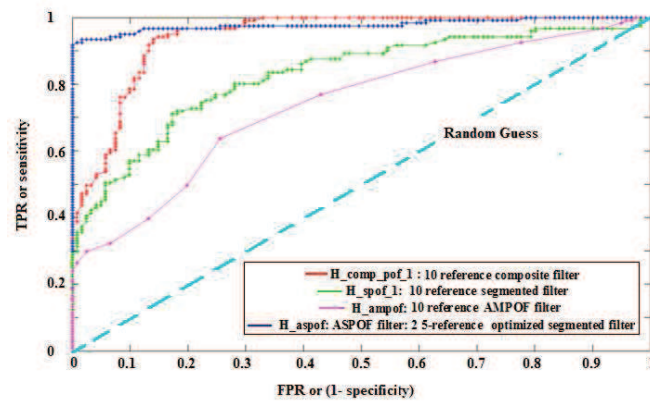


Figure 13. (Color online) ROC curves obtained with 10-reference composite filters: POF (red), SPOF (green), AMPOF (purple) and ASPOF (navy blue). The sky-blue line shows the random guess [41].

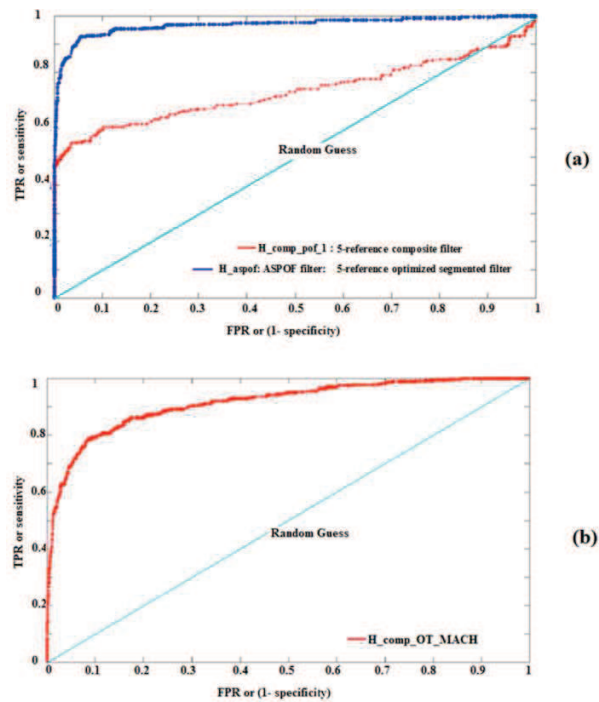


Figure 14. (Color online) (a) ROC curves obtained by correlating faces of a given subject, e. g. Figure 12 (a), with 6 other individuals with 5-reference ASPOF (navy blue) and POF (red) composite filters. The sky-blue line shows the random guess. (b) ROC curve obtained with an OT MACH.

We can see that the ASPOF filter is effective for image recognition. The true recognition rate is equal to 92% when the false alarm rate is set to 0% .

We also compared the ROC curves obtained with the ASPOF, POF, and OT MACH filters for the face recognition application (with reference to Figure 14). We fabricated 5-reference composite filters. For the ASPOF, we used a 2-reference SPOF and a 3-reference SPOF to compute the ASPOF. The reference images correspond to -45° , -30° , -15° , $+15^\circ$ and $+45^\circ$ rotation angles. The ASPOF produces better correlation performances than the POF filter (Figure 14 (a)). We also compared these results with the ROC curve of the OT MACH (Figure 14 (b)). The distance between the two curves is shorter than the distance between the ROC curves of the ASPOF and POF filters but the ASPOF still indicates better performances.

ACKNOWLEDGMENTS

The authors acknowledge the partial support of the Conseil Régional de Bretagne and thank A. Arnold-Bos (Thales Underwater Systems) for helpful discussions. They also acknowledge S. Quasmi for her help with the simulations. Lab-STICC is Unité Mixte de Recherche CNRS 6285.

REFERENCES AND LINKS

- [1] Alfalou and C. Brosseau "Understanding Correlation Techniques for Face Recognition: From Basics to Applications," in Face Recognition, Milos Oravec (Ed.), ISBN: 978-953-307-060-5, In-Tech (2010).
- [2] VanderLugt, "Signal detection by complex spatial filtering," *IEEE Trans. Inf. Theor.* 10, 139-145 (1964).
- [3] B.T. Phong, "Illumination for computer generated pictures", *Communications of the ACM*, 18, no.6, (1975).
- [4] Pointing Head Pose Image (PHPID), http://www.ecse.rpi.edu/~cvrl/database/other_Face_Databases.htm.
- [5] Alfalou, G. Keryer, and J. L. de Bougrenet de la Tocnaye, "Optical implementation of segmented composite filtering," *Appl. Opt.* 38, 6129-6135 (1999).
- [6] Y. Ouerhani, M. Jridi, and A. Alfalou, "Implementation techniques of high-order FFT into low-cost FPGA," presented at the Fifty Fourth IEEE International Midwest Symposium on Circuits and Systems, Yonsei University, Seoul, Korea, 7-10 Aug. 2011.
- [7] S. Weaver and J. W. Goodman, "A technique for optically convolving two functions," *Appl. Opt.* 5, 1248-1249 (1966).
- [8] V. K. Vijaya Kumar and L. Hassebrook, "Performance measures for correlation filters," *Appl. Opt.* 29, 2997-3006 (1990).
- [9] J. L. Horner, "Metrics for assessing pattern-recognition performance," *Appl. Opt.* 31, 165-166 (1992).
- [10] H. J. Caufield and W. T. Maloney, "Improved discrimination in optical character recognition," *Appl. Opt.* 8, 2354-2356 (1969).

-
- [11] F. Hester and D. Casasent, "Multivariate technique for multiclass pattern recognition," *Appl. Opt.* 19, 1758-1761 (1980).
 - [12] Mahalanobis, B. V. K. Vijaya Kumar, and D. Casasent, "Minimum average correlation energy filters," *Appl. Opt.* 26, 3633-3640 (1987).
 - [13] Alfalou, "Implementation of Optical Multichannel Correlators: Application to Pattern Recognition," *PhD Thesis*, Université de Rennes 1-ENST Bretagne, Rennes-France (1999).
 - [14] T. S. Yu, S. Jutamulia "Optical Pattern Recognition," Cambridge University Press (1998).
 - [15] J. L. Tribillon, "Corrélation Optique," Edition Teknéa, Toulouse, (1999).
 - [16] Keryer, J. L. de Bougrenet de la Tocnaye, and A. Alfalou, "Performance comparison of ferroelectric liquid-crystal-technology-based coherent optical multichannel correlators," *Appl. Opt.* 36, 3043-3055 (1997).
 - [17] Alfalou and C. Brosseau, "Robust and discriminating method for face recognition based on correlation technique and independent component analysis model," *Opt. Lett.* 36, 645-647 (2011).
 - [18] M. Elbouz, A. Alfalou, and C. Brosseau, "Fuzzy logic and optical correlation-based face recognition method for patient monitoring application in home video surveillance," *Opt. Eng.* 50, 067003(1)-067003(13) (2011).
 - [19] Leonard, A. Arnold-Bos, and A. Alfalou, "Interest of correlation-based automatic target recognition in underwater optical images: theoretical justification and first results," *Proc. SPIE* 7678, 76780O (2010).
 - [20] V. H. Diaz-Ramirez, "Constrained composite filter for intraclass distortion invariant object recognition," *Opt. Lasers Eng.* 48, 1153 (2010).
 - [21] A. S. Awwal, "What can we learn from the shape of a correlation peak for position estimation?," *Appl. Opt.* 49, B40-B50 (2010).
 - [22] Alsamman and M. S. Alam, "Comparative study of face recognition techniques that use joint transform correlation and principal component analysis," *Appl. Opt.* 44, 688-692 (2005).
 - [23] S. Romdhani, J. Ho, T. Vetter, and D. J. Kriegman, "Face recognition using 3-D models: pose and illumination," in *Proceedings of the IEEE* 94, 1977-1999 (2006).
 - [24] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russel, "Face recognition by humans: nineteen results all computer vision researchers should know about", in *Proceedings of the IEEE* 94, 1948-1962 (2006).
 - [25] D. E. Riedel, W. Liu, and R. Tjahyadi "Correlation filters for facial recognition login access control", in *Advances in Multimedia Information Processing*, 3331/2005, 385-393 (2005).
 - [26] P. Hennings, J. Thornton, J. Kovacevic, and B. V. K. Vijaya Kumar, "Wavelet packet correlation methods in biometrics," *Appl. Opt.* 44, 637-646 (2005).
 - [27] J. Khoury, P. D. Gianino, and C. L. Woods, "Wiener like correlation filters," *Appl. Opt.* 39, 231-237, 2000.
 - [28] Mahalanobis, B. K. V. Vijaya Kumar, S. Song, S. R. F. Sims, and J. F. Epperson, "Unconstrained correlation filters," *Appl. Opt.* 33, 3751-3759 (1994).
 - [29] Zhou and T.- H. Chao, "MACH filter synthesizing for detecting targets in cluttered environment for grayscale optical correlator," *Proc. SPIE* 3715, 394 (1999).

-
- [30] Pe'er, D. Wang, A. W. Lohmann, and A. A. Friesem, "Achromatic optical correlation," *Opt. Lett.* 25, 776-778 (2000).
 - [31] V. K. Vijaya Kumar, "Tutorial survey of composite filter designs for optical correlators," *Appl. Opt.* 31, 4773-4801 (1992).
 - [32] V. Oppenheim and J.S. Lim, "The importance of phase in signals," *Proc. of IEEE* 69, 529-541 (1981).
 - [33] J. L. Horner and P. D. Gianino, "Phase-only matched filtering," *Appl. Opt.* 23, 812-816 (1984).
 - [34] J. Ding, M. Itoh, and T. Yatagai, "Design of optimal phase-only filters by direct iterative search" *Opt. Comm.* 118, 90-101 (1995).
 - [35] J.L. Horner, B. Javidi, and J. Wang, "Analysis of the binary phase-only filter," *Opt. Comm.* 91, 189-192 (1992).
 - [36] V. K. Vijaya Kumar, "Partial information filters," *Digital signal processing* 4, 147-153 (1994).
 - [37] R. C. Gonzalez and P. Wintz, "*Digital Image Processing*," Addison-Wesley (1987).
 - [38] Inbar and E. Marom, "Matched, phase-only, or inverse filtering with joint-transform correlators," *Opt. Lett.* 18, 1657-1659 (1993).
 - [39] P. Refrégier, "Optimal trade-off filters for noise robustness, sharpness of the correlation peak, and Horner efficiency," *Opt. Lett.* 16, 829-831 (1991).
 - [40] Casasent and G. Ravichandran, "Advanced distortion-invariant MACE filters," *Appl. Opt.* 31, 1109-1116 (1992).
 - [41] Leonard, A. Alfalou and C. Brosseau, "Spectral optimized asymmetric segmented phase only correlation filter", *Appl. Opt.* (2012)

Adaptive nonlinear fringe-adjusted joint transform correlator

Isabelle Leonard

Ayman Alfalou

ISEN Brest

Département Vision, L@bISEN, 20 rue Cuirassé
Bretagne

CS 42807, 29228 Brest Cedex 2, France

E-mail: ayman.al-falou@isen.fr

Mohammad S. Alam

University of South Alabama

Department of Electrical and Computer
Engineering

SH 4122, 150 Jaguar Drive

Mobile, Alabama 36688-0002

Andreas Arnold-Bos

Thales Underwater Systems

Route de Sainte Anne du Portzic

CS 43814, 29238 Brest Cedex 3, France

Abstract. An optimized technique, based on the fringe-adjusted joint transform correlator architecture, is proposed and validated for rotation invariant recognition and tracking of a target in an unknown input scene. To enhance the robustness of the proposed technique, we used a three-step optimization. First, we utilized the fringe-adjusted filter (H_{FAF}) in the Fourier plane, then we added nonlinear processing in the Fourier plane, and, finally, we used a new decision criterion in the correlation plane by considering the correlation peak energy and the highest peaks outside the desired correlation peak. Several tests were conducted to reduce the number of reference images needed for fast tracking, while ensuring robust discrimination and efficient tracking of the desired target. Test results, obtained using the pointing head pose image database, confirm robust performance of the proposed method for face recognition and tracking applications. Thereafter, we also tested the proposed technique for a challenging application such as underwater mine detection and excellent results were obtained. © 2012 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: [10.1117/1.OE.51.9.098201](https://doi.org/10.1117/1.OE.51.9.098201)]

Subject terms: correlation; nonlinear fringe-adjusted joint transform correlator; face recognition; underwater mine recognition.

Paper 120308P received Mar. 3, 2012; revised manuscript received Jul. 6, 2012; accepted for publication Jul. 13, 2012; published online Sep. 14, 2012.

1 Introduction

Correlation is a robust and efficient pattern recognition technique, which was widely studied in the literature.¹ Correlation based discrimination measures the degree of similarity between a target image (unknown image to be recognized) and a reference image (belonging to a known database). Two families of optical architectures are used by most researchers to implement the correlation technique which are the Vander-Lugt correlator (VLC),² and the joint transform correlator (JTC).³ Unlike the JTC, VLCs require the implementation of a complex-valued filter. The optical implementation of the JTC architecture does not need to align the optical setup, especially in the Fourier domain between the filter and the image target spectrum (VLC). Moreover, the reference images can be updated in near real time.

In this paper, we utilized the JTC-based correlation architecture to realize a near real-time face tracking system. The proposed system must be robust to changes in in-plane rotation of the target, while ensuring accurate discrimination between multiple targets. The main objective, of this paper, is to accurately recognize and track the desired target, such as the face of person in a given scene or to recognize underwater mines from video sequences.

To enhance the performance of the proposed tracking system, we used an optimized version of the JTC architecture, namely the fringe-adjusted JTC (FJTC).⁴ This fringe-adjusted JTC provides excellent correlation discrimination, but the robustness of FJTC may not be sufficient for some applications. Consequently, we added a nonlinearity function to the fringe-adjusted JTC architecture to enhance the

robustness and discrimination capability. In addition, special attention is given to generate an optimized correlation output by applying a decision criterion in the correlation plane. The peak to correlation energy (PCE) criterion is not very suitable for automatic decision making.^{5,6} The presence of autocorrelation peak, and the presence of two correlation peaks, can perturb the decision criterion. To address this limitation, we used an adaptive FJTC architecture to suppress the zero-order autocorrelation peak. Since the correlation plane is symmetric, we work on one half of the correlation plane and modify the criterion to obtain values included in the interval [0 1]. Test results are presented to verify the effectiveness of the proposed technique.

2 Analysis

The JTC technique is one of the most widely used correlation methods reported in the literature.¹ This technique has been adapted and used in many recognition and/or tracking applications.^{1-4,7,8} In this work, we have used this JTC method for a very the demanding application of underwater mines recognition. The underwater mines recognition application requires an adaptive JTC method to overcome the various problems caused by the specific conditions of underwater imaging. We begin this section by reviewing the principles of several main JTC architectures. The principle of the JTC correlation architecture has been described in detail in the literature.^{1-4,7,8} In the classical JTC, the known reference image, r , and the unknown input scene, i , (image to be recognized) are displayed side-by-side in the input plane of the correlator, which are separated by a distance, $2x_0$, along the x -axis between the target image and the reference image, to form the input joint image, expressed as

$$f(x, y) = i(x + x_0, y) + r(x - x_0, y). \quad (1)$$

After applying Fourier transform to Eq. (1) such as the input plane, we can record the intensity or the joint power spectrum (JPS), F , using a square law device such as a CCD detector array. In a classical JTC, an inverse Fourier transform of the JPS yields the correlation output in the output plane.³ This correlation output contains two important terms which are a central autocorrelation peak, and two cross-correlation peaks corresponding to the crosscorrelation between the unknown input scene, i , and the known reference image, r . These two cross-correlation peaks are located on either side of the central zero-order peak separated by a distance, $4x_0$.

The classical JTC provides broad correlation peaks and strong side lobes which can complicate the decision making. Several JTC architectures have been developed to solve this problem such as the binary JTC, nonzero-order JTC, and fringe-adjusted JTC.^{4,7,9} In a binary JTC,⁷ the JPS is binarized using a suitable threshold before applying the inverse Fourier transform step. The binarized JPS is defined as

$$F_{\text{binary}}(\mu, \vartheta) = \begin{cases} +1 & \text{if } F(\mu, \vartheta) \geq \text{threshold} \\ -1 & \text{if } F(\mu, \vartheta) \leq \text{threshold} \end{cases} \quad (2)$$

This architecture provides narrow correlation peaks. However, it is sensitive to noise, rotation and scale variations of the target, and the impact of other artifacts. Moreover, the binarization step requires on-line processing. Another interesting JTC architecture is the nonlinear JTC where a nonlinear threshold is applied to the JPS before applying the inverse Fourier transform to yield the correlation output, expressed as:³

$$F_{\text{NL}} = F^k. \quad (3)$$

In Eq. (3), the user sets the value of k depending on the application. If k is set to 1, we obtain the classical JTC. The nonlinear JTC provides superior discrimination compared to the classical JTC. It is more efficient when the value of k is lower than 0.5.^{10,11}

The last optimization we present, herein, is based on the fringe-adjusted JTC. Alam et al.^{4,10,11} proposed to introduce a real-valued filter, called fringe-adjusted filter (H_{FAF}), in the Fourier plane of the correlator, defined as:

$$H_{\text{FAF}}(\mu, \vartheta) = \frac{B(\mu, \vartheta)}{A(\mu, \vartheta) + |R(\mu, \vartheta)|^2}, \quad (4)$$

where R , denotes the intensity corresponding to the Fourier transform of the known reference image, B , is a constant or a function used to control the gain, and, A , is a function used for avoiding the pole problem and/or reducing the effects of noise. The FJTC is a very efficient architecture, very simple to implement and very convenient for many applications, especially when the noise and background are known or can be estimated. Moreover, many optimized versions of this architecture are proposed and validated.¹ However, the FJTC architecture, based on spectral filtering, makes it very sensitive to noise changes which are a very common problem in underwater images. Examples of noise in underwater images are presented in Fig. 1. Thus, to overcome this problem and to make an efficient architecture, we need to update the noise function used in the spectral filter for each image, which may become a bottleneck for real-time applications. Thus, it is necessary to make optimizations to obtain a method such that it is less sensitive to noise encountered in underwater imagery. To achieve this objective, we introduced a special image preprocessing step and added nonlinearity to the JTC architecture to make it more robust, while ensuring excellent discrimination.

Another class of architecture, called the multi object JTC or nonzero-order JTC, has been found to be very useful for enhancing the correlation discrimination.^{9,11} In this architecture, the reference-only and the input-scene-only power spectra are subtracted from the JPS in the Fourier domain, given by:

$$\begin{aligned} |F_{\text{NZ}}(\mu, \vartheta)|^2 &= |F(\mu, \vartheta)|^2 - |R(\mu, \vartheta)|^2 - |I(\mu, \vartheta)|^2 \\ &= |R(\mu, \vartheta)|^2 - |I(\mu, \vartheta)|^2 \\ &\quad + 2|R(\mu, \vartheta)||I(\mu, \vartheta)| \cos(\phi_i(\mu, \vartheta) - \phi_r(\mu, \vartheta) \\ &\quad + 2ux_0) - |R(\mu, \vartheta)|^2 + |I(\mu, \vartheta)|^2 \\ &= 2|R(\mu, \vartheta)||I(\mu, \vartheta)| \cos(\phi_i(\mu, \vartheta) \\ &\quad - \phi_r(\mu, \vartheta) + 2ux_0), \end{aligned} \quad (5)$$

where I represents the input-scene-only power spectrum. In this technique, the zero-order term and the autocorrelation term, among similar input scene targets or nontarget objects, are automatically eliminated while leaving the desired cross-correlation peaks in the output plane.

In this paper, we propose and validate a nonzero-order fringe-adjusted JTC architecture (NFJTC) to ensure robust discrimination while simplifying the decision making. It may be mentioned, that the nonlinearity increases the robustness of the JTC architecture. Thus, we applied a special

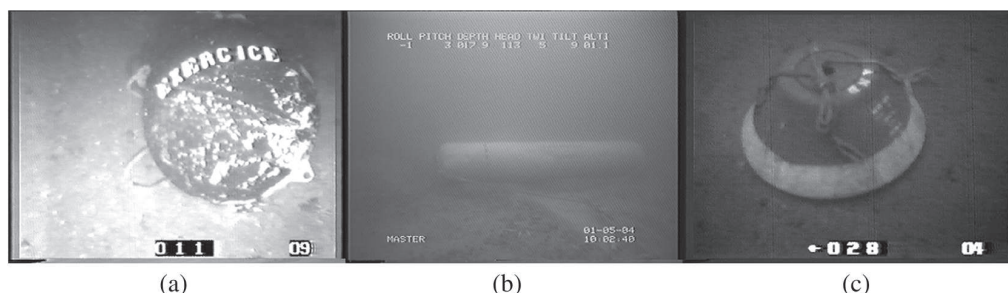


Fig. 1 Grayscale images representing three studied targets that come from underwater videos: a spherical mine (a), a cylindrical mine (b) and a Manta mine (c).

nonlinear function to the NFJTC to increase the robustness of the architecture.

To investigate the performance of the proposed technique, let us consider two practical applications. The first application involves recognizing and tracking a human face that changes according to different rotation angles. The system proposed, in this paper, allows robust tracking and discrimination of a face turning in different rotation angles ranging from -90 deg to $+90$ deg. Facial images contain information as well as noise. In general, the noise is known, and is quite similar, from one image to another in the sequence. Once our system is validated, we applied it for underwater mine detection and recognition.¹² The main problem, in this application, is the poor visibility which can limit the visibility to only a few meters or less in turbid waters.

3 Challenges of Underwater Imaging

A generic underwater image formation model is shown in Fig. 2.¹³ In Fig. 2, light source (1) transmits light (2) towards the object (6). On the trajectory to and from the object, light is intercepted by undesired underwater particles (3), and light rays (2), (5), (7) and (9) become exponentially dimmer due to direct attenuation. Part of the light is also reflected towards the camera by medium (3), resulting in the backscattered component (4), which corresponds to the image of the water volume. Finally, some light is slightly deviated by medium (3) on its way back to the camera, resulting in the forward scattered component (8) which adds blurring to the image. Underwater images are very noisy images. Moreover, for underwater imaging, this noise changes from one image to another. This limits the efficiency of image recording medium and requires a noise model to reduce its impact. Detection and recognition can be significantly perturbed by this noise component.

To address this challenge, we introduced a three-stage preprocessing step as shown in Fig. 3. The first stage removes the moiré effect which is generated due to the use of analog cameras. It is modeled by a sine function following the procedure outlined in Ref. 14. Weak contrast often reduces the efficiency of recognition methods. In order to improve the underwater image contrast, we adapted the preprocessing technique presented in Ref. 15. The overall algorithm involved in the implementation of the second stage is summarized in Fig. 4. The output image with enhanced contrast is calculated as:

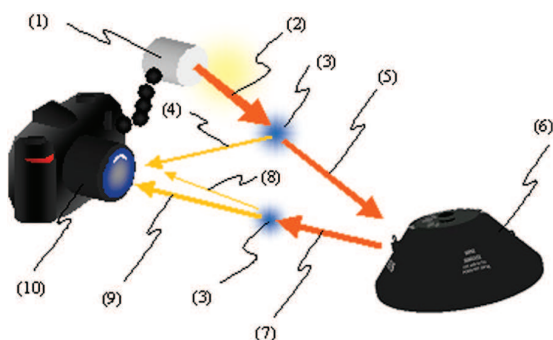


Fig. 2 Diagram explaining the light propagation in an underwater medium.

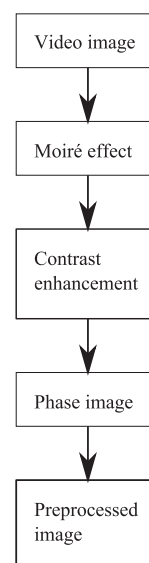


Fig. 3 Preprocessing steps to improve image quality.

$$I = \frac{I_{LP}I_s + \frac{I_d}{I_{LP}}}{I_{LP} + \frac{1}{I_{LP}}}, \quad (6)$$

where I_{LP} , represents the original image filtered with a FGaussian low-pass filter, I_s represents the result of the subtraction of the I_{LP} from the original image, and I_d represents the result of the division of the original image by I_{LP} . Details of this algorithm can be found in Ref. 15. Finally, in the third stage, we calculate the phase image and we reduce noise with a band pass filter created with a Gaussian wavelet as depicted in Fig. 5.¹²

4 Analytical Modeling

To investigate the robustness of the proposed technique, we selected several existing algorithms and compared their performance by using three images as shown in Figs. 6 to 8. The first two images came from one video sequence acquired by a diver and represents an upside down Manta mine (MN 103). These two image frames are closely located in the video (8 frames apart) where the mine-like characteristic did not change significantly. These two images were used to test the robustness of the proposed and alternate architectures. Since these two images were shot within a small time-frame and are similar, so the output of the filter should also be similar. The discrimination of our system is tested by applying JTC architectures on a mine image and the Lena image as shown in Fig. 8.

At first, we compared the classical JTC and the nonzero-order JTC on the manta mine image as shown in Fig. 6. We used image frame 1082 of the test video as the reference image and as the target image and the results are shown in Fig. 9. From Fig. 9, it is evident that the suppression of the autocorrelation peak simplifies the decision making. Moreover, this suppression reduces the effect of noise in the correlation plane.

Next, we investigated the performance of the nonzero-order JTC for the above mentioned images as shown in Figs. 6 to 8. We used PCE as the decision criterion in the nonzero-order JTC correlation plane, defined as:⁵

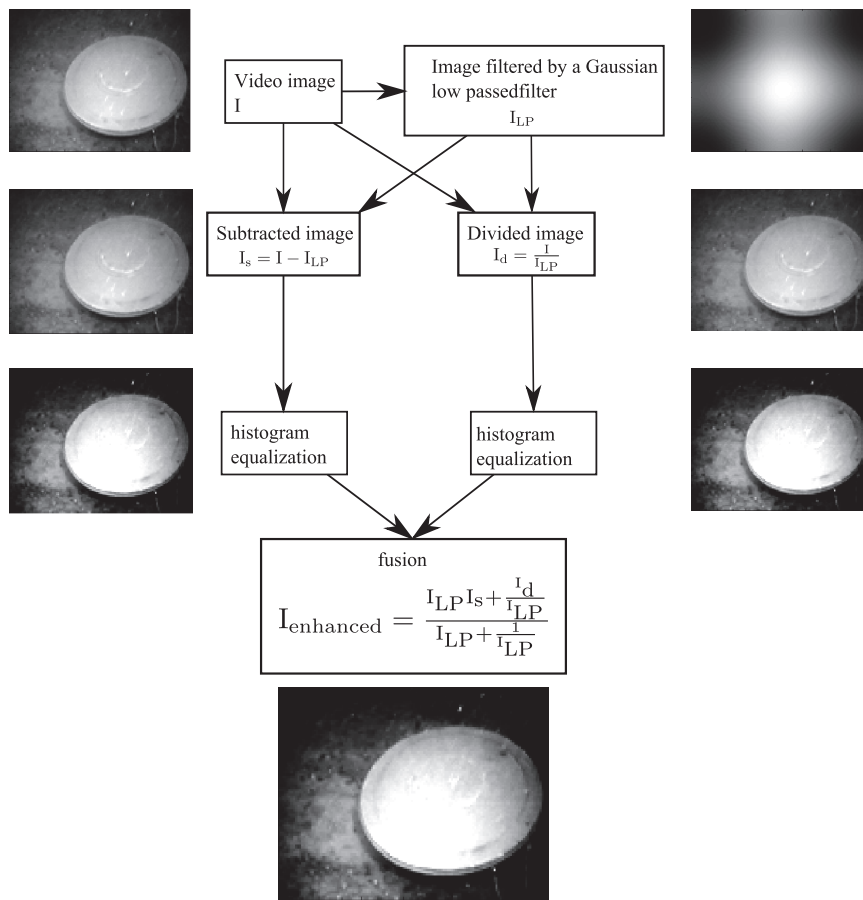


Fig. 4 Image enhancement algorithm.¹⁵

$$PCE = \frac{\sum_{x=x_0-t}^{x=x_0+t} \sum_{y=y_0-t}^{y=y_0+t} |C(x, y)|^2}{\sum_{x=1}^N \sum_{y=1}^M |C(x, y)|^2}, \quad (7)$$

where (x_0, y_0) denote the position of the correlation peak, $C(x, y)$ denotes the value of the correlation plane at point (x, y) , t is set to the number of neighboring pixels used, and (N, M) denote the size (in pixels) of the correlation plane. The corresponding results are shown in Fig. 10.

Figure 9 shows that the nonzero-order JTC architecture yields helpful results for decision making. Figure 10 shows that the fringe adjusted JTC architecture observes these constraints on PCE values better than the binary

JTC architecture. The nonlinear JTC architecture is a robust one, such that, PCE values are close between Figs. 9 and 10 as shown in the left columns. However, it does not provide robust discrimination as shown in Fig. 10. PCE values are close for the cases of (i) two mines on the left, and (ii) one mine and the Lena.

In this paper, our idea is to combine the attractive features, of fringe-adjusted JTC and nonlinear JTC, in order to obtain a robust architecture for correlation discrimination. In the proposed architecture, called nonlinear fringe-adjusted JTC (NNJTC), the JPS is modified to obtain the nonlinear fringe-adjusted JPS. The mathematical steps needed to obtain this nonlinear fringe-adjusted JPS are summarized in Eq. (8):

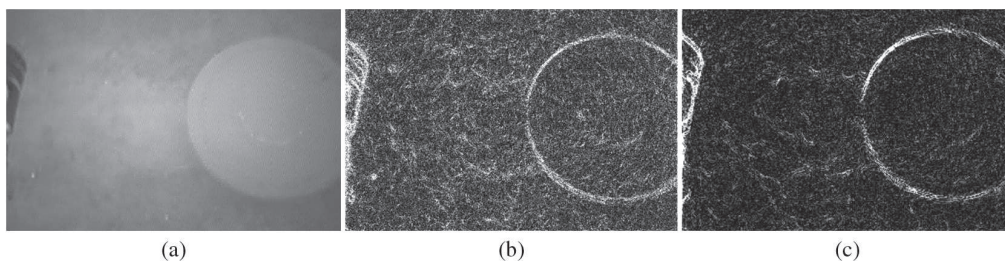


Fig. 5 Initial image, phase image, phase and wavelet image.



Fig. 6 Manta mine, image no. 1082.



Fig. 7 Manta mine, image no. 1090.

$$\begin{aligned}
 f(x, y) &= i(x + x_0, y) + r(x - x_0, y)FF(\mu, \vartheta) \\
 &= |I(\mu, \vartheta)| \exp(\phi_i(\mu, \vartheta)) \exp(jux_0) + \dots \\
 &\quad + |R(\mu, \vartheta)| \exp(\phi_r(\mu, \vartheta)) \exp(-jux_0)G(\mu, \vartheta) \\
 &= |F(\mu, \vartheta)|^2 = |I(\mu, \vartheta)|^2 + |R(\mu, \vartheta)|^2 \\
 &\quad + 2|I(\mu, \vartheta)||R(\mu, \vartheta)| \cos(\phi_i(\mu, \vartheta)) \dots \\
 &\quad - \phi_r(\mu, \vartheta) + 2ux_0|F_{NZ}(\mu, \vartheta) \\
 &= 2|I(\mu, \vartheta)||R(\mu, \vartheta)| \cos(\phi_i(\mu, \vartheta) - \phi_r(\mu, \vartheta) \\
 &\quad + 2ux_0)|G1(\mu, \vartheta) \\
 &= F_{NZ} \cdot H_{FAF}(\mu, \vartheta)G1(\mu, \vartheta) \\
 &= F_{NZ} \frac{B(\mu, \vartheta)}{A(\mu, \vartheta) + |R(\mu, \vartheta)|^2} G_{NN-FAF}(\mu, \vartheta) \\
 &= (G1(\mu, \vartheta))^k, \tag{8}
 \end{aligned}$$

where G , represents the classical JPS, F_{NZ} represents non-zero JPS, H_{FAF} represents the FAF filter, $G1$ is the nonzero-order fringe-adjusted JPS, G_{NN-FAF} is the nonlinear and the

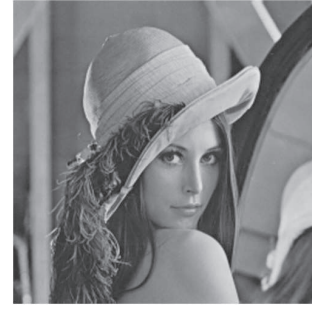


Fig. 8 Lena.

nonzero-order fringe-adjusted JPS and k denotes the degree of nonlinearity. The value of k has been selected, after several tests, to provide the best compromise between robustness and discrimination. Before applying the NNJTC, we investigated the performance of the binary JTC and fringe-adjusted JTC for underwater images discussed in the previous section.

5 Binary JTC and Fringe-Adjusted JTC Performance Comparison

To compare the performance of nonzero-order binary JTC and the nonzero-order fringe-adjusted JTC, we consider the Lena image, Fig. 8, as the known reference image as well as the unknown target image, which is a scenario of perfect match between the reference and the target. To mimic the effects of underwater imaging, we modified the target image with underwater-like noise. At first, we added blur as if the image is perturbed by forward scattered light. Then we added nonuniform luminosity, which is a problem when the underwater robot uses artificial light. The next step was to combine the effects of blur and nonuniform light into the final target image. For the last test, we manually removed the background of the reference image. The target image is the Lena image corrupted with blur and nonuniform light.

The correlation output results are presented in Table 1. It is evident that the nonzero-order, fringe-adjusted JTC yields better results compared to the binary JTC. The FJTC is robust against different type of noise, such as underwater noise depicted in Table 1. According to these results and for underwater videos, the FJTC can provide better results compared to the binary JTC.

6 Test Results

6.1 Face Recognition

At first, we applied the fringe-adjusted JTC for face recognition and tracking application. The objective is to detect and track a person via face recognition in a video sequence as shown in Fig. 11. Figure 12 presents the results obtained with fringe-adjusted JTC where only one reference image is used to construct the fringe-adjusted filter H_{FAF} . It should be noted that the face of the subject shown in Fig. 11 without rotation. The reference image is introduced on one side of the input plane. The background is defined as information outside a square around the face in the reference image considered. Then we display the target image on the other side of

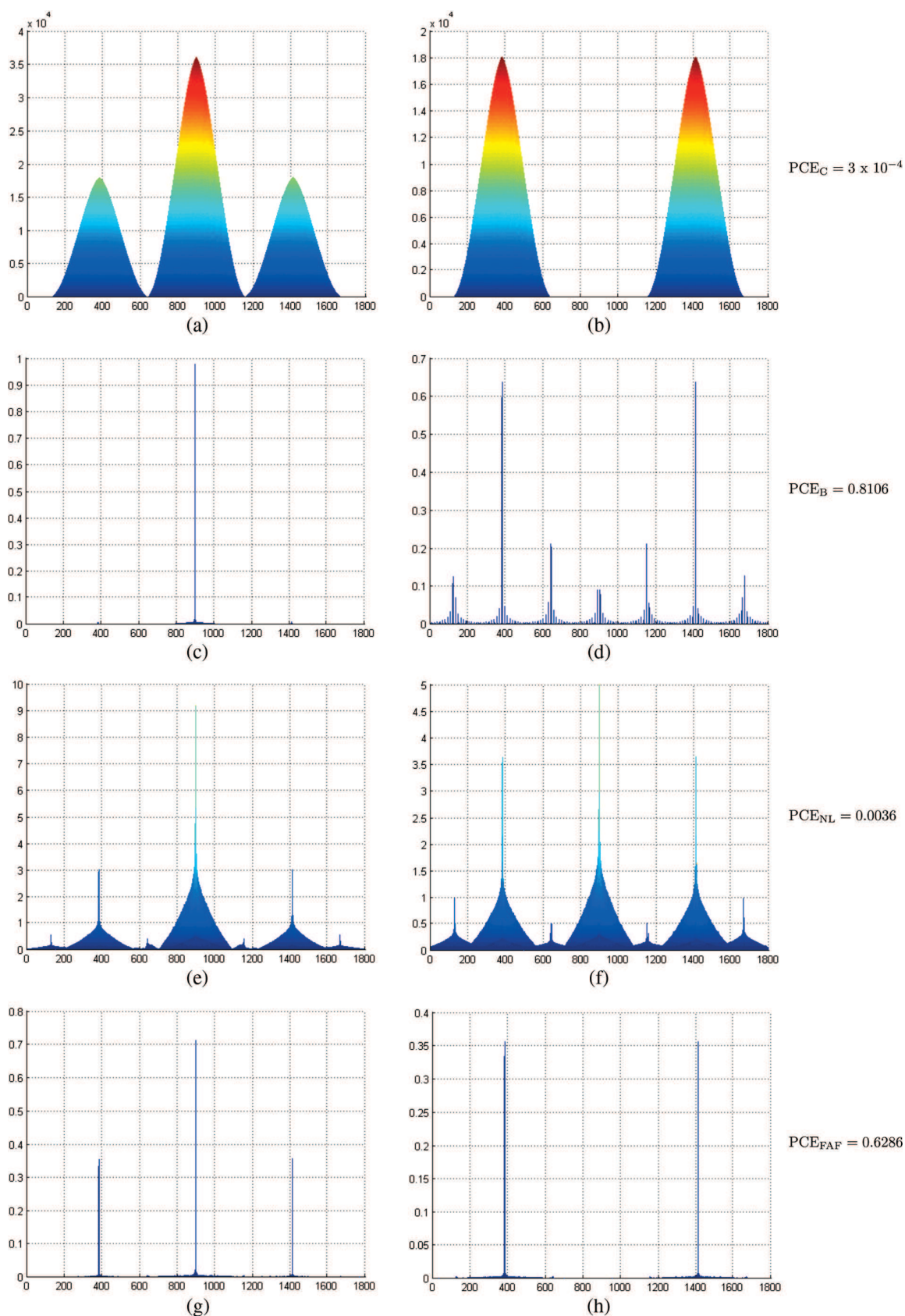


Fig. 9 Comparison of JTCs with zero order (left column) and nonzero JTCs (right column). (a) and (b): classical JTC. (c) and (d): binary JTC. (e) and (f): nonlinear JTC. (g) and (h): fringe adjusted JTC.

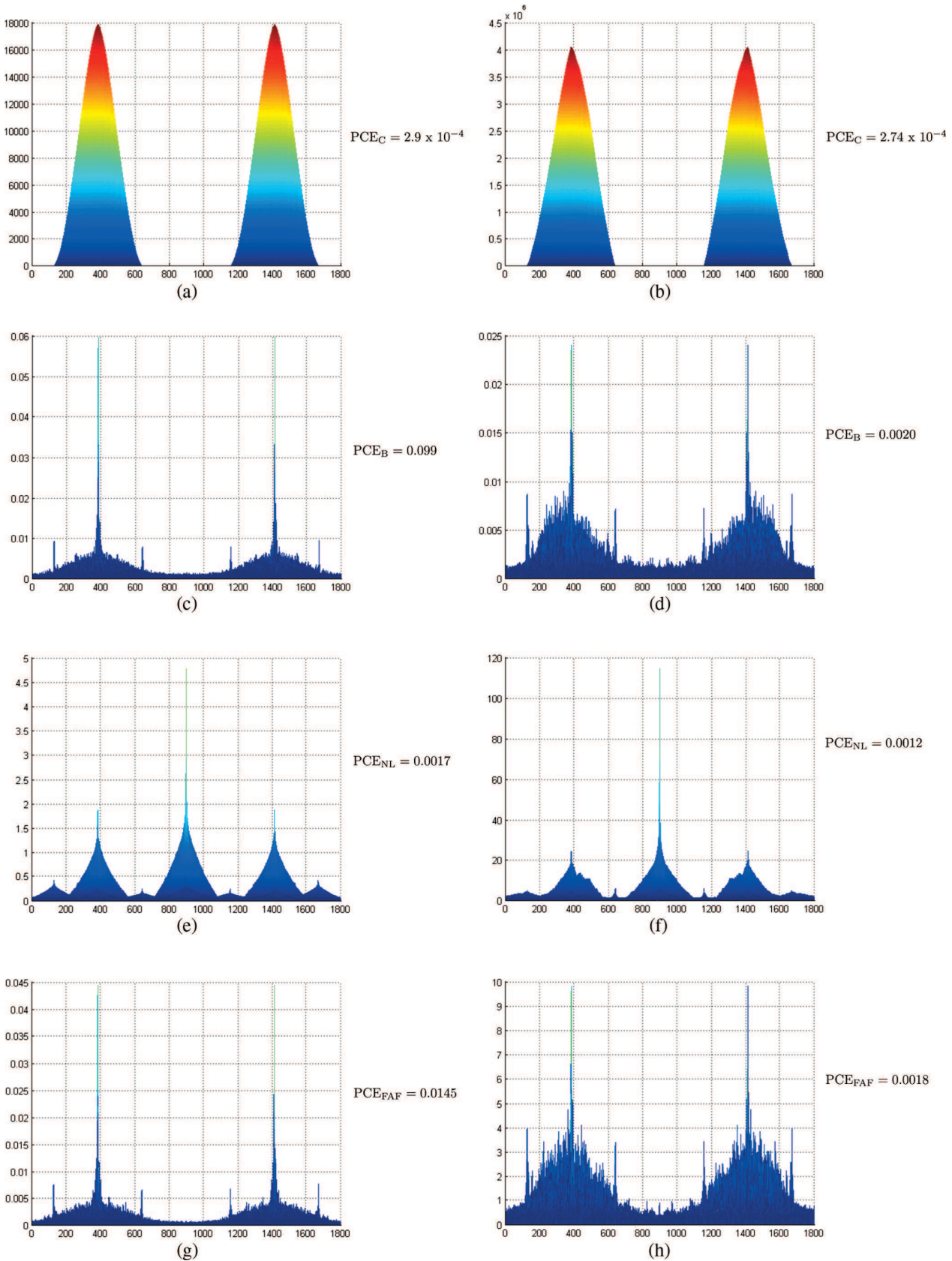


Fig. 10 Results of the robustness and the discrimination tests. Left column: comparison of the 2 mine images. Right column: comparison of the mine image and Lena image. (a) and (b): classical JTC. (c) and (d): binary JTC. (e) and (f): nonlinear JTC. (g) and (h): fringe adjusted JTC.

Table 1 Comparison of the binary JTC and the fringe adjusted JTC on images with underwater noises. a: Lena-Lena. b: Lena-blurred Lena. c: Lena-Lena with non uniform light. d: Lena-Lena with blurred and nonuniform light. e: Modified Lena- Lena with blurred and nonuniform light.

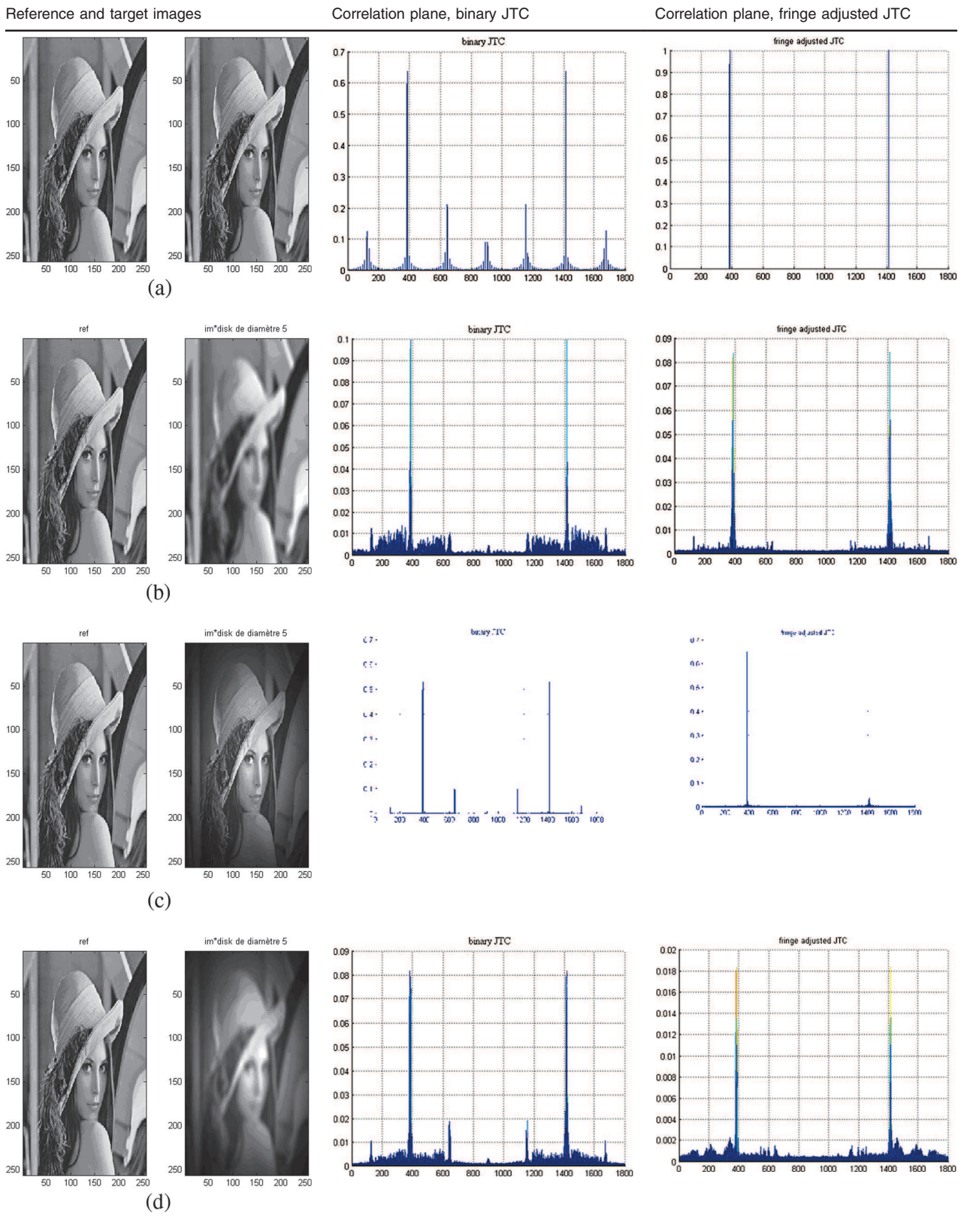
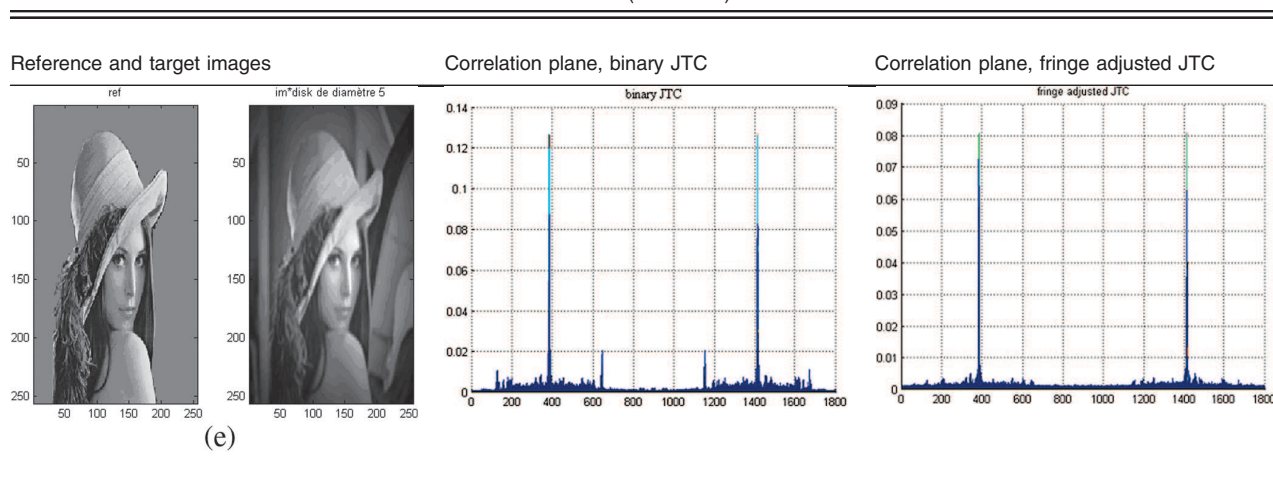


Table 1 (Continued).

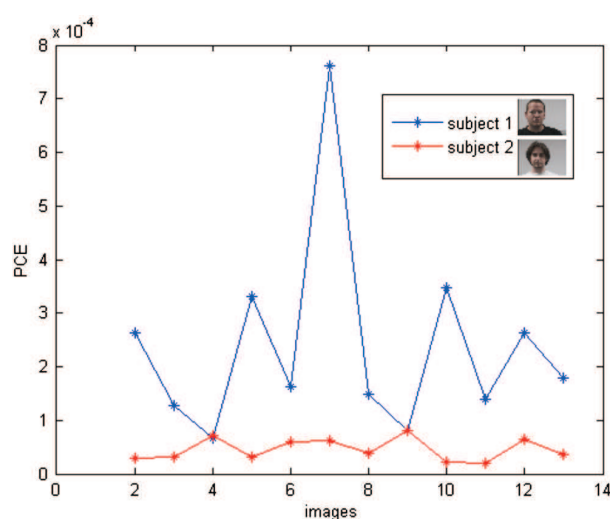
Fig. 11 Examples of facial rotation (-90 deg to $+90$ deg). Subject number (1), Base PHPID.¹⁶

the input plane. The target faces are obtained from the database presented in Fig. 11, one at a time starting from -90 deg to $+90$ deg in a sequential manner.¹⁶

We initialize the system with a known reference image, such as the subject's face (number 1), to be tracked so that initial position of subject's face is known. By comparing this reference face with the target image, we ascertain the presence or absence of the subject (number 1) in the target image and the position of the subject's face in the target image. Thereafter, only the information around the position of the face is selected in this target image. To achieve this, we multiply the target image, with a filter equal to 1, around the position of the face found and 0 elsewhere. Afterwards, we introduce the preprocessed image as the reference image in the fringe-adjusted JTC input plane. Then a new target image, subject number 2, is introduced in the input plane and the process is repeated until all images from the database are processed. After various tests performed by using this algorithm, we have identified a tolerance rotation angle equal to 15 deg between the target image and reference image in which the correlation remains robust. Beyond this value between the reference and target images, the correlation becomes less robust and may lose track of the subject's face. With a tolerance angle equal to 15 deg, the tracking problem of a person, with in-plane rotation between -90 and $+90$ deg, consists of correlating 13 pairs of images: $(-90, -75)$, $(-75, 60)$, ..., $(-15, -0)$, $(0, +15)$, ..., $(+60, +75)$, and $(+75, +90)$.

The tracking results of subject number (1) are shown in Fig. 12 (blue curve). The PCE criterion was used for decision

making in the correlation plane.⁵ Figure 12 also shows the results when subject number 2, such as an undesirable subject present in the input scene (false alarm: red curve). The results, depicted in Fig. 12, show that the proposed fringe-adjusted JTC yields good results. However, for some cases, the values of the two cases are too close for making a robust system. A robust system is a system where the values of

Fig. 12 Results obtained with a FA-JTC using only one reference image position without any rotation in the H_{FAF} Filter to define the background noise.

recognition are close to 1. A discriminated system is a system where the false alarms are close to 0, with significant difference between true and false alarms.

To optimize the performance of the fringe-adjusted JTC system, we propose to utilize an adaptive fringe-adjusted filter (H_{FAF}) in the Fourier plane.¹⁷ In this technique, the background noise is calculated for each reference. For this purpose and after finding the position of head in the target image at time t_0 , we multiply the target image with a filter that has a value of 0 around the position of the face and 1 elsewhere. One can also replace the 0 value with the average of the values around the 0 value zone.

The target image obtained at time, t_0 , becomes the reference image at time, t_1 , in the proposed tracking system. Thereafter, we introduce the preprocessed image as the background noise in the fringe-adjusted filter formulation. The results obtained after the first iteration are shown in

$$P = \frac{\left(\sum_{i=i_0-t}^{i_0+t} \sum_{j=j_0-t}^{j_0+t} |C(i, j)|^2 \right) - \left\{ \sum_{k=1}^5 \frac{\left[\left(\sum_{i=i_k-t}^{i_k+t} \sum_{j=j_k-t}^{j_k+t} |C(i, j)|^2 \right) \right]}{5} \right\}}{\sum_{i=i_0-t}^{i_0+t} \sum_{j=j_0-t}^{j_0+t} |C(i, j)|^2}, \tag{9}$$

where (i_0, j_0) denote the position of the correlation peak, $C(i, j)$ denotes the value of the correlation plane at point (i, j) , t is set to the number of neighboring pixels used around the considered peak, (i_k, j_k) are the positions of five highest peaks excluding the correlation peak.

Step 2: Apply additional nonlinearity in the Fourier plane of FJTC similar to Eq. (3).¹⁰ After many trial runs in our laboratory, we set the value of k to 0.75.

Figure 14 shows the different test results obtained with our tracking system based on the optimized fringe-adjusted JTC. By comparing the results presented in Figs. 12 and 13, it is evident that the proposed technique yields better discrimination. Indeed, the difference between the two curves (red and blue) is much larger. The robustness has also improved as shown by the shorter range of values of the blue curve compared to the previous tests.

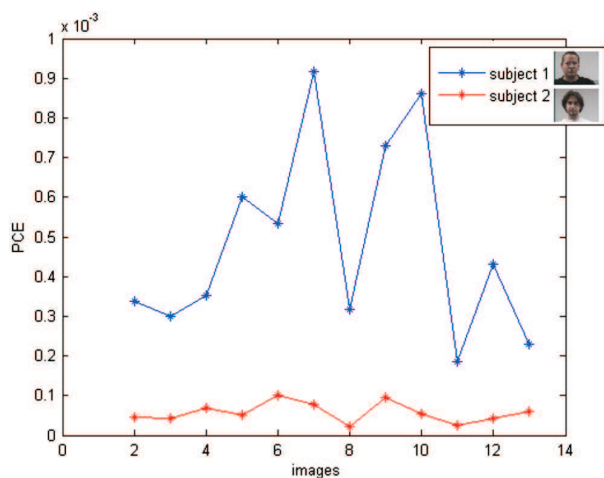


Fig. 13 Results obtained using the adaptive fringe-adjusted JTC.

Fig. 13 which clearly shows that the adaptive fringe-adjusted JTC improves discrimination performance considerably, when compared to the first case. Figure 13 shows a system that has a better robustness and discriminatory, but we think that we can find a system and a criterion that increase robustness and discriminatory.

To further increase the performance of the proposed system, we propose to incorporate a decision criterion in the correlation plane which involves the following optimizations steps:

Step 1: Enhanced PCE decision criterion (P): we propose to take into account the correlation peak energy as well as the five highest peaks excluding the highest correlation peak. The Enhanced PCE decision (P) criterion is defined as:

For noisy images, we searched for a robust criterion that provides values close to 1 in case of recognition and values close to 0 in case of false alarm. An idea is to compare the highest peak of the correlation plane to other high peaks. Consequently, we remove the autocorrelation peak in order to compute a criterion only with inter-correlation peaks and noise. Since the correlation plane is symmetric, we can work on one half of the correlation plane. Our objective is to find the maximum value, M , of the new correlation plane. We set a first threshold, M_{-3dB} to the value of the correlation peak M divided by $\sqrt{2}$, and n this value corresponds to 3 dB and to 50% of the correlation peak energy. We compute the cumulative sum (S1) of the pixels around the correlation peak and with values higher than M_{-3dB} . Next we compute the sum (S2) of all the pixels excluding the highest correlation peak with values higher than $0.75 \times M_{-3dB}$, as shown in Fig. 15. Finally, we introduce a distance metric defined as:

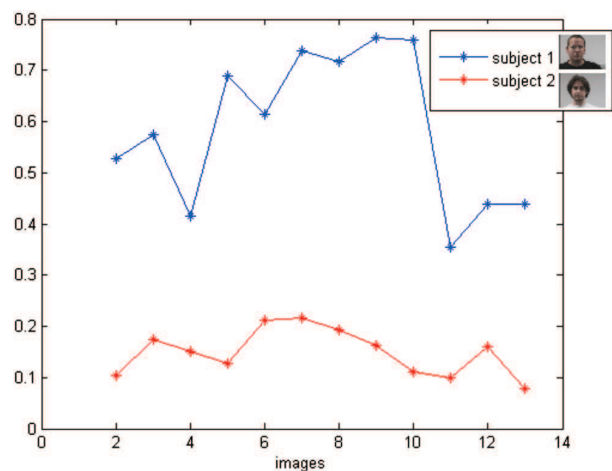


Fig. 14 Results obtained with a nonlinearity in the Fourier plane of the FA-JTC and the new criterion.

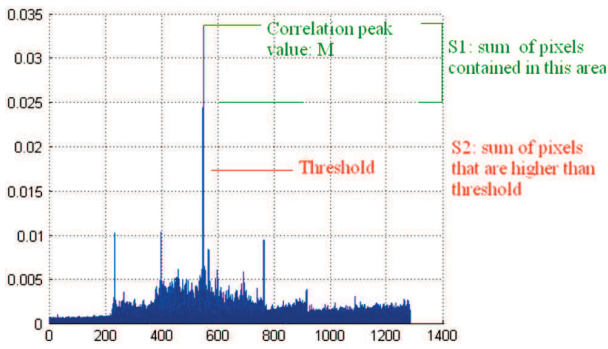


Fig. 15 Computation of our new criterion.

$$C = \frac{S1 - S2}{S1} \tag{10}$$

The distance criterion yields values in the interval [0 1]. In case of recognition, we obtain a high correlation peak with negligible noise, therefore, the value S2 will be very low. In case of false alarm, the correlation plane will be very noisy and the highest correlation peak will be surrounded by high peaks and the value S2 will be very close to the value S1.

We applied this criterion to the previous test and the corresponding results are shown in Fig. 16. This criterion provides recognition values close to 1 and false alarm values close to 0. Thus, the robustness and discrimination ability of the proposed technique is very high. Using the aforementioned procedure, users can set a threshold for automatic decision making.

6.2 Underwater Mine Recognition

In the previous section, we validated the proposed architecture and criterions for face recognition applications. In this section, we applied the proposed technique for underwater mine detection. As there are a lot of images, we used composite filter techniques to fuse close references and reduce the size of the database.¹² For this application, we selected five filters to recognize mines in a specific region of an underwater video sequence. As explained earlier, underwater

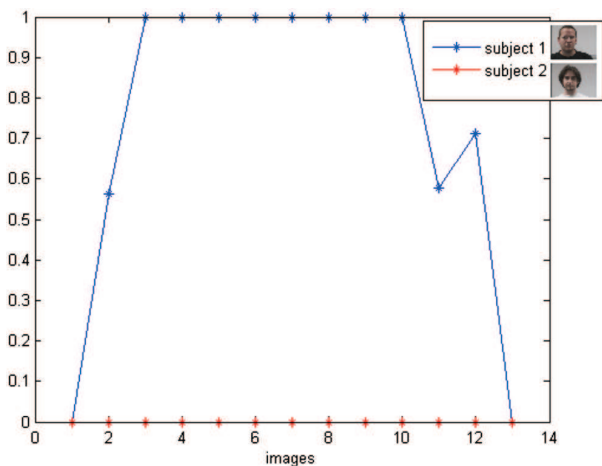


Fig. 16 Results obtained with a nonlinearity in the Fourier plane of the FA-JTC and the last criterion.

images are perturbed by the medium and need preprocessing. Without preprocessing, correlation results will not provide desired output since target related contours will not stand out. The preprocessing step consists of resizing the image to reduce computation time. Then, we enhance the contrast of images using the procedure described in Ref. 14.

To obtain the contour image, we chose to compute the phase of the image. In fact, edges correspond to sharp phase changes.¹⁸ To achieve this, we compute the Fourier transform of the image where the image can be expressed as $\text{Spetrum} = \rho \cdot e^{i\theta}$. The amplitude information is contained in ρ and the phase information is contained in the exponential function. To keep only the phase information, all values are divided by their amplitude ρ yielding $\text{Spetrum} = e^{i\theta}$. Next, we use a band pass filter, constructed with a Gaussian wavelet, to reduce the remaining noise. The results of this preprocessing step are shown in Fig. 5.

From this contour image database, we conducted five composite filters for mine recognition in the mine-presence zones as illustrated in Fig. 17. Each filter is constructed with five reference contour images. With these five composite filters and preprocessed images, we tested the performance of the nonzero-order fringe adjusted JTC and the nonzero-order nonlinear fringe adjusted JTC. We set the nonlinearity factor k to 0.85 and corresponding correlation outputs are shown in Table 2. From Table 2, it is obvious that the nonzero-order fringe adjusted JTC provides a high false alarm rate while the nonzero-order nonlinear fringe adjusted JTC provides moderate recognition rate.

To enhance the decision performance further by increasing the recognition rate and decreasing the false alarm rate, we decided to add an optimization step. We assume that for 10 images before the present target image if we have five correlations, it means that a mine should be detected in the present target image. Knowing that the mine is fixed

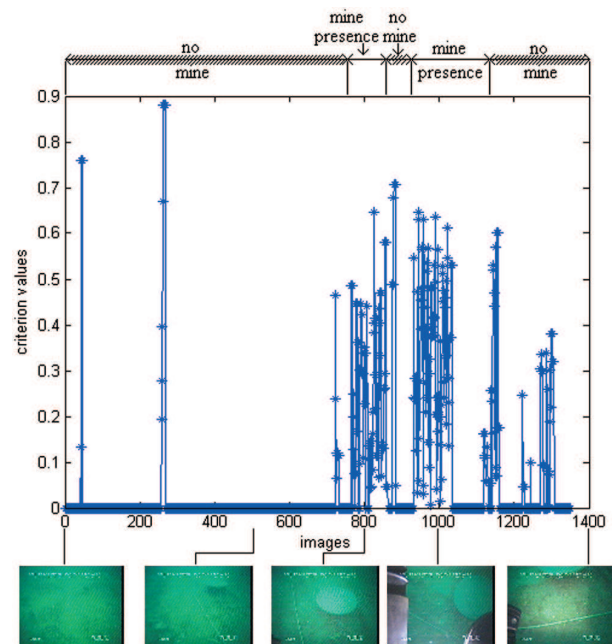


Fig. 17 Results of the optimized nonlinear fringe-adjusted JTC, $k = 0.85$.

Table 2 Results of the different JTCs for mine recognition.

Method	Recognition (%)	False alarm (%)
Fringe adjusted JTC	93.10	56.44
Nonlinear fringe adjusted JTC, $k = 0.85$	52.07	16.29

Table 3 Results of the nonlinear fringe-adjusted JTC with optimization.

Method	Recognition (%)	False alarm (%)
Optimized nonlinear fringe adjusted JTC, $k = 0.85$	65.52	10.13

and knowing the navigation parameters, a mine, which is detected in one image frame, is highly unlikely to disappear in the following image frame. Our choice of 10 images has been made after several tests in order to find the right compromise between robustness and computing time. Using this concept, our optimization consist in considering only the last 10 raw decisions, coming from last 10 images, for a given present target image. Using last 10 nonoptimized results leads us to decrease the false alarm rate. In fact, if we use the optimized results, the detector never stops detecting after five detections, even if the five detections are false. However if we use our raw data, this optimization leads to a stop of detection, five images after the last image where a mine has been detected. On the test results, we can see that the recognition rate has improved while the false alarm rate has decreased, as illustrated in Table 3. In Fig. 17, we see regions where a mine is supposed to be detected and regions where the algorithm is supposed to find nothing. The results presented Table 3 and Fig. 17 shows that we managed to increase the recognition rate to 65.52 percent and we decrease the false alarm rate to 10.13 percent.

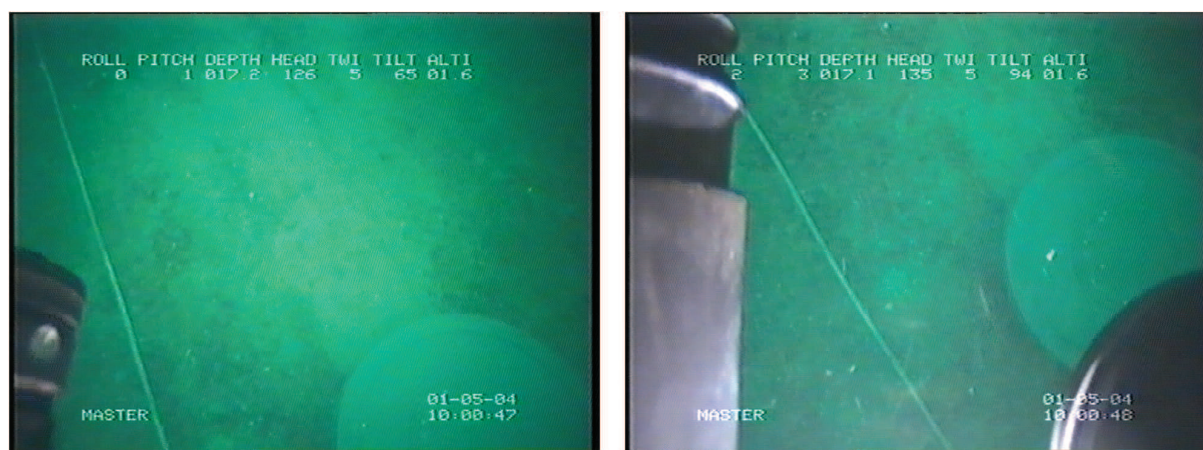
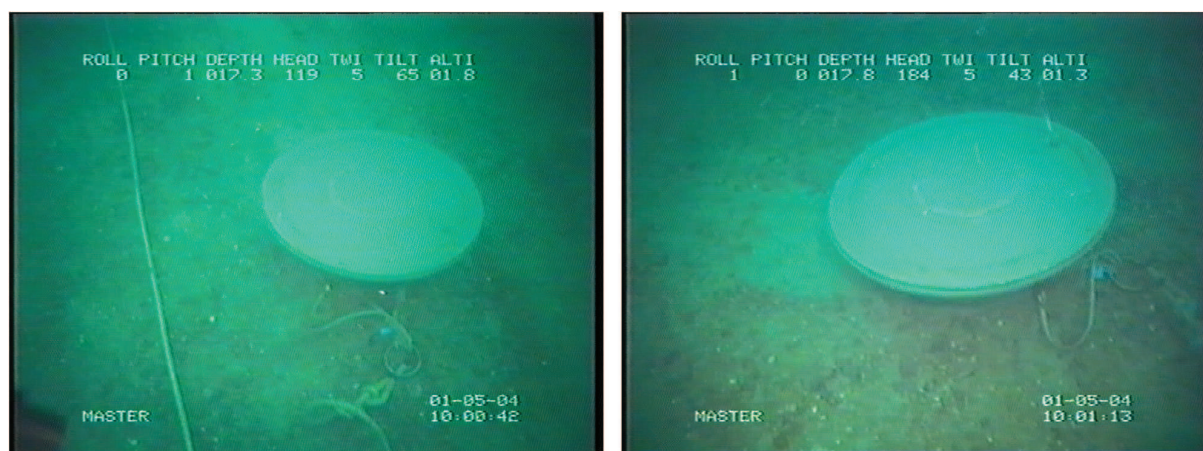
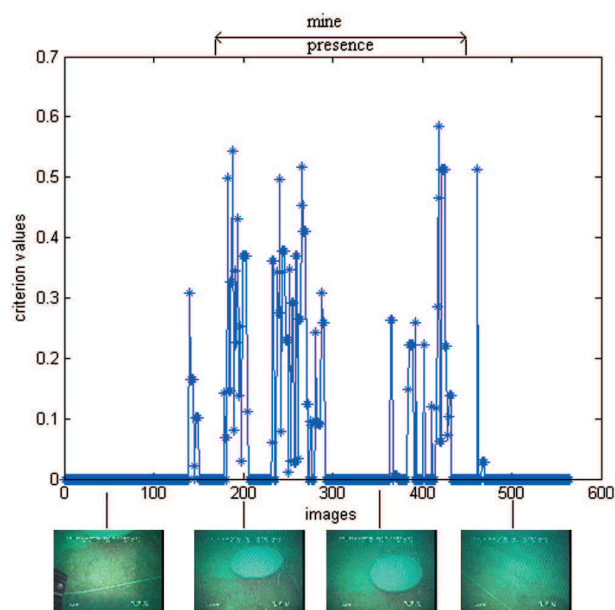
**Fig. 18** Images with presence of the camera support.**Fig. 19** Difference between mine size in the 2 video sequences.

Table 4 Robustness results of the different JTCs.

Method	Recognition (%)	False alarm (%)	True to false alarm ratio (%)
Fringe-adjusted JTC	88.52	59.93	1.48
Nonlinear fringe-adjusted JTC, $k = 0.85$	34.07	12.33	2.76
Optimized nonlinear fringe-adjusted JTC, $k = 0.85$	41.11	5.48	7.5

Nevertheless, our optimized algorithm detected some objects, in the last region of the video sequence, where there is no mine. This can be explained by the presence of the camera support in target images as shown in Fig. 18. We did not add a mask to suppress this man-made object because there are images with no camera support and images with camera support, but it did not occupy the same proportion of the image as shown in Fig. 18, however, this constraint can be avoided when recording the images.

To show the robust performance of our proposed approach, several tests were conducted and the results were compared with nonoptimized methods. To do this, we selected a video sequence where the same mine is recorded, but the distance between the viewer and the mine is shorter in this case compared to the previous case shown in Fig. 19. Results obtained, in terms of recognition and false alarm probabilities, are summarized in Table 4. The visualization of the robustness of the optimized NNJTC is shown in Fig. 20. The nonzero-order fringe adjusted JTC still provides a high false alarm rate while the nonzero-order nonlinear fringe adjusted JTC presents good recognition rate. The

**Fig. 20** Results of the optimized nonlinear fringe-adjusted JTC, $k = 0.85$.

optimization allows an increase of the recognition rate and a decrease of the false alarm rate, when compared to the nonlinear fringe adjusted JTC.

7 Conclusion

In this paper, we proposed and validated the first version of a robust recognition and tracking system based on the fringe adjusted JTC architecture. In the proposed technique, we introduced nonlinearity to the fringe adjusted JTC, and also used an adaptive correction criterion in the correlation plane.

On the face recognition application, we show that a rotation angle of 15 deg between the target image and the reference image does not disable the recognition ability of the system, while considerably reducing the number of correlations necessary and, therefore, increasing the processing speed of the system. Moreover, the adaptive decision criterion, used in the output plane, minimized the number of false correlation peaks in the correlation plane thus reducing the false alarms. Finally, the addition of nonlinearity in the Fourier plane makes it possible to adapt or find a compromise between the robustness and discrimination of our tracking system by adjusting the degree of nonlinearity (k). For underwater mine recognition application, the proposed NNJTC architecture yields promising results.

Acknowledgments

This paper complements our paper "Nonlinear fringe-adjusted JTC-based face tracking using an adaptive decision criterion".¹⁹ The raw mine images used in this publication were provided by the GESMA (Groupe d'Etudes Sous Marines de l'Atlantique) under the TOPVISION project coordinated by Thales Underwater Systems SAS. This project is related to Techno-Vision Program, which was launched by the French Ministry of Defense. Detailed information can be found in the website <http://topvision.gesma.fr>. This work is also partly supported by the Regional Council of Brittany.

References

1. A. Alfalou and C. Brosseau, "Understanding correlation techniques for face recognition: from basics to applications," *Face Recognition*, M. Oravec, Ed., ISBN: 978-953-307-060-5, INTECH. Available from: <http://sciy.com/articles/show/title/understanding-correlation-techniques-for-face-recognition-from-basics-to-applications>.
2. V. VanderLugt, "Signal detection by complex spatial filtering," *IEEE Trans. Info. Theory* **10**(2), 139–145 (1964).
3. C. S. Weaver and J. W. Goodman, "A technique for optically convolving two functions," *Appl. Opt.* **5**(7), 1248–1249 (1966).
4. M. S. Alam and M. A. Karim, "Fringe-adjusted joint transform correlation," *Appl. Opt.* **32**(23), 4344–4350 (1993).
5. J. L. Horner, "Metrics for assessing pattern-recognition performance," *Appl. Opt.* **31**(2), 165–166 (1992).
6. I. Léonard, A. Alfalou, and C. Brosseau, "Spectral optimized asymmetric segmented phase-only correlation filter," *Appl. Opt.* **51**(14), 2638–2650 (2012).
7. B. Javidi and C. J. Kuo, "Joint transform image correlation using a binary spatial light modulator at the Fourier plane," *Appl. Opt.* **27**(4), 663–665 (1988).
8. L. Guibert et al., "On-board optical joint transform correlator for real-time road sign recognition," *Opt. Eng.* **34**(1), 135–143 (1995).
9. C. T. Li, S. Yin, and F. T. S. Yu, "Nonzero-order joint transform correlator," *Opt. Eng.* **37**(1), 58–65 (1998).
10. M. S. Alam, "Fractional power fringe-adjusted joint transform correlation," *J. Opt. Eng.* **34**(11), 3208–3216 (1995).
11. M. S. Alam and M. A. Karim, "Multiple target detection using a modified fringe-adjusted joint transform correlator," *J. Opt. Eng.* **33**(5), 1610–1617 (1994).
12. I. Leonard, A. Arnold-Bos, and A. Alfalou, "Interest of correlation-based automatic target recognition in underwater optical images: theoretical justification and first results," *Proc. SPIE* **7678**, 767800 (2010).
13. S. Q. Duntley, "Light in the sea," *JOSA*, **53**(2), 214–233 (1963).

14. D. N. Sidorov and A. C. Kokaram, "Suppression of moiré patterns via spectral analysis," *Proc. SPIE*, **4671**, 895 (2002).
15. A. Arnold-Bos, J.-P. Malkasse, and G. Kervern, "Towards a model-free denoising of underwater optical images," in *Proc. of the IEEE conference on Ocean (Europe)*, IEEE, Europe (2005).
16. N. Gourier, D. Hall, and J. L. Crowley, "Estimating face orientation from robust detection of salient facial features," in *Proc. of Pointing 2004, ICPR, International Workshop on Visual Observation of Deictic Gestures*, Cambridge, United Kingdom (2004).
17. M. S. Alam and A. Bal, "Dynamic target tracking with fringe-adjusted joint transform correlation and template matching," *Appl. Opt.* **43**(25), 4874–4881 (2004).
18. A. Oppenheim and J. Lim, "The importance of phase in signals," *Proc. IEEE* **69**(5), 529–541 (1981).
19. I. Léonard, A. Alfalou, and M. S. Alam, "Nonlinear fringe-adjusted JTC-based face tracking using an adaptive decision criterion," *Proc. SPIE* **8134**, 81340M (2011).

Isabelle Leonard received the engineering diploma in optronic from the French engineering school Polytech'Paris-Sud in 2009. Since 2009, she is a PhD student at Institut Supérieur d'Electronique et du numérique (ISEN). Her research interests deal with image processing and underwater medium. She has published over 7 refereed journal articles or conference papers.



Ayman Alfalou received his PhD in telecommunications and signal processing from École nationale supérieure des télécommunications de Bretagne (ENSTB)-France and the University of Rennes 1 in 1999. He held a one-year post-doctoral position at ENSTB during which he designed and realized a compact and high-rate optical correlator. Since 2000, he has been a professor of telecommunications and signal processing at ISEN-Brest. At ISEN, he founded the Vision-L@ISEN. His research interests deal with optical engineering, optical information processing, signal and image processing, telecommunications and optoelectronics. He has supervised several PhD, MSc, and engineering school students. He has published over 110 refereed

journal articles or conference papers and special sessions. He is a senior member of SPIE, OSA, IEEE, and a member of IoP.



Mohammad S. Alam is a professor and Chair at the University of South Alabama (USA). His research interests include ultra fast computing architectures and algorithms, image processing, pattern recognition and tracking, biometric recognition, infrared imaging systems, and smart energy management and control. He is the author or co-author of over 475 refereed journals, conference publications, or project reports, 15 book chapters, and a book on IPTV (IEC Press).

He received numerous excellences in research, teaching and service awards. He served or serves as the PI or Co-PI of many research projects totaling over \$14M, supported by NSF, FAA, DoE, ARO, AFOSR, SMDC, NASA, WPAFB, BP and ITT industry. He has organized and chaired many international conferences and serves as a Guest Editor for several professional journals. He is a Fellow of OSA, SPIE, IoP, IET, a Life Fellow of the Bangladesh Computer Society (BCS), and of the Institution of Engineers Bangladesh (IEB), and a Senior Member of IEEE.

Andreas Arnold-Bos graduated from École nationale supérieure de l'aéronautique et de l'espace (SUPAERO) [(now Institut Supérieur de l'aéronautique et de l'espace (ISAE)], Toulouse, France, in 2004, with the engineering degree of aerospace engineering, as well as a master's degree in Signal and Image Processing. From 2004 to 2007 he worked towards a PhD thesis on radar engineering at the École Nationale Supérieure des Ingénieurs des Études et Techniques d'Armement in Brest, France. In 2007, he joined Thales Underwater Systems as a full-time R & D engineer while simultaneously pursuing his PhD thesis, which he successfully defended in 2010. His past and current interests involve radar and sonar engineering, underwater optics and autonomous systems.

Improvement of automatic man-made object detection in underwater videos by using of navigational information

Isabelle LEONARD⁽¹⁾, Andreas ARNOLD-BOS⁽²⁾, Ayman ALFALOU⁽¹⁾,
Nicolas MANDELERT⁽²⁾

⁽¹⁾ Département Vision L@bISEN, ISEN, 20 rue Cuirassé Bretagne, 29228 BREST CEDEX 2, FRANCE

E-mail: {isabelle.leonard, ayman.al-falou}@isen.fr

⁽²⁾ General Sonar Studies, Thales Underwater Systems, Route de Sainte Anne du Portzic, CS 43814, 29238 BREST CEDEX 3, FRANCE

E-mail: {Andreas.Arnold-Bos, nicolas.mandelert}@fr.thalesgroup.com

SHORT ABSTRACT: *In order to detect man-made objects in underwater video, we propose and validate a novel method based on background subtraction methods. The other main contribution is the introduction of a priori elements deduced from positioning sensors. These elements allow to enhance the visibility of underwater objects thanks to the calculation of the sun position in relation to the vehicle position, the detection with the distance from the object and the post-processing with constraints defined on the vehicle movements. These constraints allow to reject false detections and to better know the position of the detected object. We tested our algorithm on data acquired at sea and show that we improve detection results and decrease false alarm rate, comparing to our former work. Both algorithms have been applied on the same videos. We still have to increase the true detection rate while reducing processing time i.e. processing time should be close to video rate.*

Keywords: *underwater images, mine detection, background subtraction, use of navigational information.*

1 INTRODUCTION

Underwater mines represent an important threat. This threat is generally addressed using a system with four steps: a detection step, a classification step, an identification step, and a neutralization step [1]. Nowadays, the trend is to design autonomous systems (Autonomous underwater vehicles, AUV) to avoid the involvement of clearance divers. Some of these AUVs are equipped with optical video camera besides sonars and positioning sensors. In case of identification mission, detection and guidance are done by sonar. When the vehicle is close to the mine, the video camera is activated. However, video images are affected by the underwater medium. Scattering and absorption cause images with weak contrast: objects are difficult to distinguish on the ocean floor. Besides, real time preprocessing and detection algorithms are necessary to improve identification results and closed-loop vehicle guidance.

In this article, our vehicle is supposed to be able to automatically identify a mine using a video camera. For that, we assume that the vehicle knows approximately the position of the target object, using sonar navigation data. Thus, first of all, our system has to detect this object and position the mine exactly.

We present a novel method based on background subtraction and an adaptation of our detection method to use navigational information. Videos are analyzed image per image. One of the contributions of this article is the use of navigational information in order to improve true detection rate. Another key point of our method is to consider the temporal aspect, *i.e.* the link between successive

images.

First we will detail our problematic and present some performance criteria. Then we will detail our proposed algorithm and our first experimental results. Finally, we will compare the performances of our algorithm with a state-of-art algorithm.

2 PROBLEM STATEMENT

This work takes place in the underwater mine detection domain. This underwater medium affects the light used to illuminate the target scene through absorption and scattering phenomena. The visibility range reaches only a few meters. This limits the use of video cameras in underwater applications. Moreover, underwater images have a weak contrast, require preprocessing steps and restraint the efficiency of edge detection methods. Accordingly, we are interested in background subtraction methods less sensitive to underwater medium than edge detection methods.

Furthermore, we must take processing time constraint into account. In fact, our system should be embedded on an AUV. This vehicle guides itself depending on the results obtained with the detection and identification steps. Thus, the processing time should be close to the video recording rate.

Our preliminary tests, using experimental data acquired at sea, showed the good algorithm efficiency.

3 HOW TO MEASURE ALGORITHM PERFORMANCES?

First of all, we have to define some parameters to test the detection algorithm performances. The first parameter is the definition of mine presence zones. The other parameters are used to calculate the detection probabilities. To compute these probabilities, we create for each experimental image an annotation file, containing among others the viewer-object distance and the position of the region of interest located around the target object.

3.1 Definition of mine presence zones

In this article we work on underwater videos. We apply our algorithm on the video, *a posteriori*, not during the video acquisition. But in order to guide the underwater vehicle when the algorithm is embedded on it, we will have to define mine presence zone. We can ask a human operator to select zones where there is an object. But this solution is not a robust one and especially not a repetitive method. Indeed, the presence zone detection differs in a random movement, at the beginning and the end of the zone, depending on the operator and the video. Some people wait for the entire object when others need few millimeters to detect it. Therefore, we are looking for a subjective method. For that, we carefully investigated our videos and chose a distance criterion. We empirically fixed a maximum viewer-object distance to 6 meters. According to visibility and turbidity, this distance is a satisfactory compromise.

3.2 Detection probabilities

Our new algorithm is able to detect several objects in an image. For that, we have to distinguish true and false detections. Moreover, we wish to test detection performances of our algorithm and compare them with another one (our former detection algorithm). To solve this problem (distinguish true and false detections), we defined several probabilities, summarized in figure 1.

		Reality	
		Mine	Nothing
Results	Mine	P _{tp} true positive	P _{fp} false positive
	Nothing	P _{fn} false negative	P _{tn} true negative

Fig. 1: Definition of different used probabilities

These probabilities are defined as follows:

- if our algorithm detects an object with correct location, we have a true positive detection, noted P_{tp} (cf. figure 2)

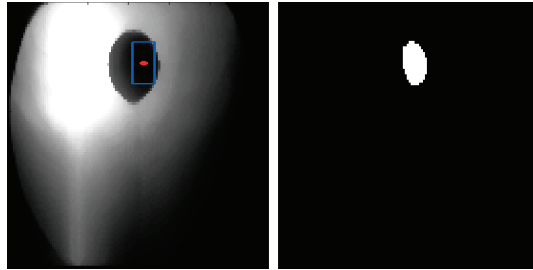
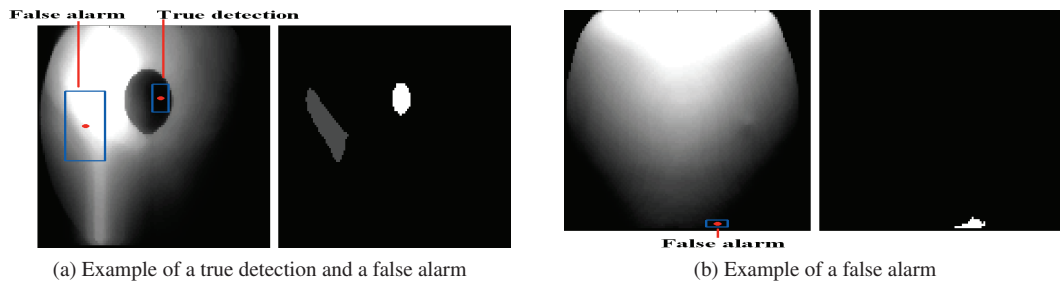


Fig. 2: Example of a true positive detection

- if our algorithm detects an object with wrong location (cf. figure 3a) or in an empty image (cf. figure 3b), we have a false alarm or a false positive detection, noted P_{fp} (cf. figure 3)



(a) Example of a true detection and a false alarm

(b) Example of a false alarm

Fig. 3: Example of false alarm detection

- if we miss a detection in an image with an object (cf. figure 4), we have a false negative detection noted P_{fn}

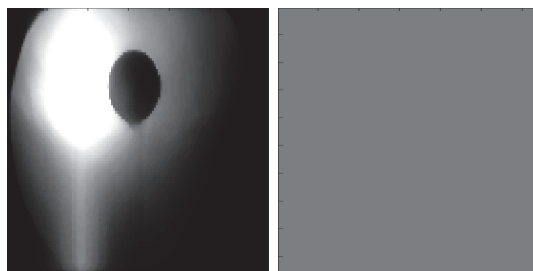


Fig. 4: Example of a missed detection

- if no object is detected in an empty image (cf. figure 5), we have a true negative detection, noted P_{tn}

As we define the mine presence-zone parameter according to viewer-object distance, we present our detection results according to the same viewer object distance too. However, we cannot calculate probabilities for each single distance. For that, we group distances by intervals equal to 0.5 meter: we group all our detection results for a given video with a viewer-object distance from 0 meter to 0.5 meter and so on. Thanks to annotation files, we know the viewer-object distance for each image. Thus we know how many images and how many true objects correspond to each distance interval. For each distance interval, each probability is defined as:

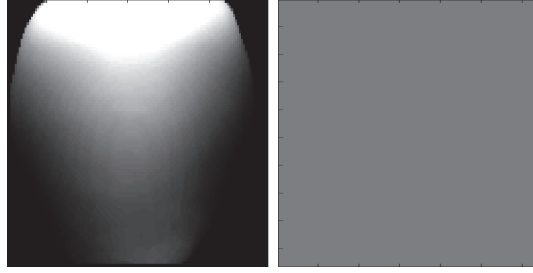


Fig. 5: Example of a true negative detection

$$\frac{\text{number of detection objects in the considered distance interval corresponding to the probability definition}}{\text{number of all objects in the distance interval}} \quad (1)$$

4 ALGORITHM

Underwater images have a poor contrast caused by light absorption, which increases with water turbidity. To increase the mine detection rate, we have to preprocess images. We presented our preprocessing in a previous article [2]. To limit the moiré effect and the processing time, we resize the images. Then we apply the edge enhancement proposed by Arnold Bos *et al.* [3, 4]. Finally, we use Phong's model [5] to reduce the sun reflection on the sea floor.

According to Phong [5], the received intensity I_r is the combination of the ambient light I_a (a constant), the scattered light I_d and the specular light I_s . In the underwater realm, the absorption balances this combination:

$$I_r = e^{-c \cdot z} (I_a + I_d + I_s) \quad (2)$$

where c represents the absorption coefficient and z the distance between the object and the viewer. The specular intensity depends on the viewer and light source positions. The scattered light depends on the light source position [5]. The equation 2 can also be written as:

$$I_r = e^{-c \cdot z} (I_a + (-\vec{L}\vec{N})I_e + (\vec{R}\vec{O})I_e) \quad (3)$$

where I_e is the emitted intensity, the vector \vec{L} represents the source-object vector, the vector \vec{N} is the vector perpendicular to the object, the vector \vec{R} represents the reflected light and the vector \vec{O} represents the object-viewer vector. These vectors are explained on figure 6.

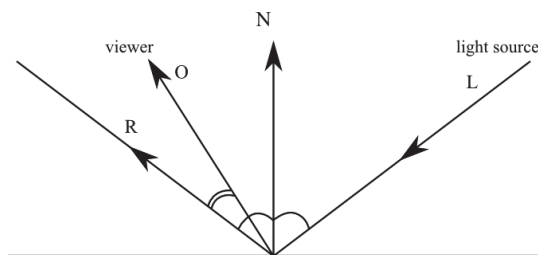


Fig. 6: Definition of angles and vectors used to calculate the received intensity

Navigational information contain the AUV position and time of video recording. Thus we know the viewer position and we can calculate the sun position thanks to Reda and Andreas algorithm [6].

On figures 7a and 8a, we present preprocessed images.

In our previous publication [2], we used the phase of the image spectrum to detect mine edges. We obtained good results. However, this method has a very low detection rate especially when we have images with weak contrast, as shown on table 1. The detection probability (table 1, 4th column) is low (below than 35%) while the false alarm rate (table 1, 5th column) is very high (above 45%).

Mine	Number of studied images	Number of images with a mine	Ptp	Pfp	Pfn	Ptn
Manta	25205	18275	18.87%	71.64%	42.34%	14.74%
Cylinder	49251	37564	32.42%	46.72%	20.96%	33.21%
Sphere	11376	7919	31.45%	69.94%	25.95%	7.10%
Other objects	13905	10222	17.92%	71.15%	26.90%	14.35%
Empty videos	17389	0		8.97%		74.90%

Table 1: Results obtained with the method proposed in [2]

To improve these rates, we are interested in other kinds of methods. Edge detection methods are not always the most efficient methods when the contrast is limited. Region subtraction and especially background subtraction methods can solve this problem. This is not a new method in the underwater domain [7]. Moreover, these methods have demonstrated their good detection performances in other domains, *e.g.* Edginton *et al.* [7] proposed and validated a new system based on these methods to detect animals. Thus, we adapted our algorithm based on these methods to detect correctly most of the true objects: we increase the true detection rate. Based on the background subtraction methods, we decided to use several images to create correctly a background image. In fact, we observe that the detection is more robust when we create the background image using mean images instead of using only one image to create the background. For that, we select the first hundred images of the video (where there is no object). We preprocessed these images and create the background image by averaging the images.

We decided to compare preprocessed images (*cf.* figures 7a and 8a) to preprocessed images with background subtraction (*cf.* figures 7b and 8b). Then we compared both images and looked for corresponding high intensity and very low intensity zones. Thresholds have been fixed empirically. Figures 7c and 8c show the results of the different steps.

Figures 7a and 8a present the preprocessed images. The images with background subtraction are shown in figures 7b and 8b. The comparison results are on figure 7c and 8c. On these examples, mines are clearly visible. On figure 7c, we obtain only the spherical mine. However, on figure 8c, the mine is truly detected but other zones are also presented, corresponding to difference between the sea floor and the images used to create the background, at the beginning of a given video. Thus there are some false alarm detections (*cf.* figure 8c).

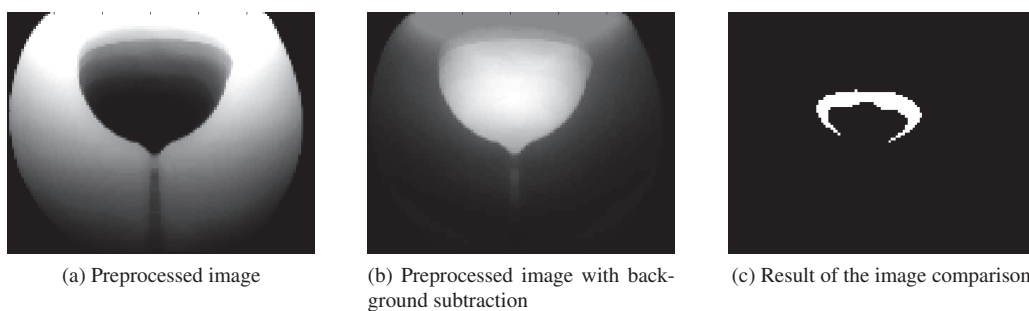


Fig. 7: Result of our algorithm on spherical mine image

The results obtained with this method are presented on table 2. Detection probability (table 2, 4th column) is higher than detection probability previously obtained (*cf.* table 1), above 30%. The false alarm rate (table 2, 5th column) decreases. The true negative detection rate (Ptn, table 2, last column) is higher than 90% for mine Manta and spherical mine. The different results presented on table 2 show a good improvement compared to table 1, but the developed method is not the best one since all the probabilities are not close to the optimal probabilities.

It can be noted that we also improved the processing time. Our algorithm has been run using Matlab, version R2007a, using an Intel Core 2 Quad CPU cadenced at 2.66Ghz. The processing time is 0.08s per image, which is close to the video rate.

Further analysis of our results shows that detection can be changed between two consecutive images,

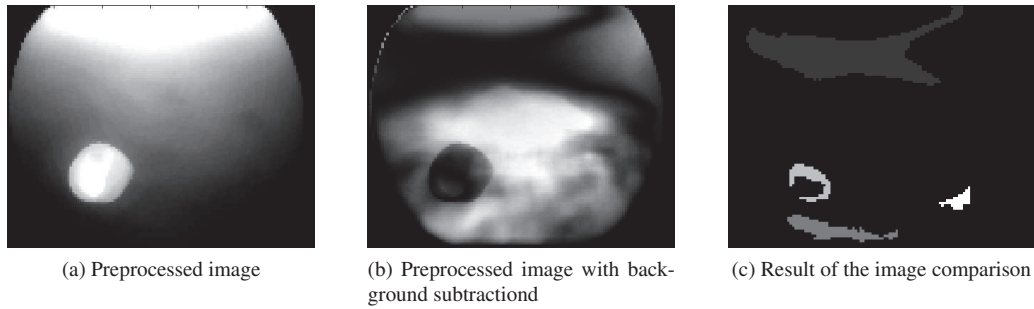


Fig. 8: Result of our algorithm on Manta mine image

Mine	Number of studied images	Number of images with a mine	Ptp	Pfp	Pfn	Ptn
Manta	25205	18275	32.86%	20.83%	26.19%	96.61%
Cylinder	49251	37564	49.41%	62.37%	16.48%	31.61%
Sphere	11376	7919	43.23%	2.10%	56.41%	99.40%
Other objects	13905	10222	46.82%	71.94%	34.72%	13.96%
Empty videos	17389	0		4.13%		95.87%

Table 2: Results obtained with our proposed method

when neither the underwater vehicle moves nor the mine. So we thought that we could improve the detection probabilities thanks to use of a special constraint application. The constraints we apply come from navigational information concerning the vehicle movement data: position (x, y and z) and orientation (tilt, roll and head). Mines are not supposed to move. The vehicle has no abrupt movements. Moreover, when the mine has been detected in the center of the considered image, it is not likely to disappear. Consequently, detection stops can be avoided by navigational information and detection position analysis.

These results need a comparison with other results obtained with a state-of-art algorithm. We chose an algorithm developed by Cybernetix and Thales [8]. This algorithm preprocesses the image and segments it to obtain contour images. More details are provided in [8]. The results obtained with state-of-art algorithm are presented in table 3.

Mine	Number of studied images	Number of images with a mine	Ptp	Pfp	Pfn	Ptn
Manta	25205	18275	19.78%	3.54%	70.51%	99.02%
Cylinder	49251	37564	42.91%	10.68%	40.42%	84.20%
Sphere	11376	7919	36.05%	3.83%	54.04%	99.64%
Other objects	13905	10222	18.30%	9.63%	73.02%	82.55%
Empty videos	17389	0		8.57%		91.43%

Table 3: Results obtained with a state-of-art method

We particularly worked on videos containing spherical mines. Consequently, our detection probability (Ptp, 4th column in table 2) and true negative detection probability (Ptn, last column of table 2) are higher than reference algorithm probabilities (table 3, columns 4 and 7) and false alarm rate and false non detection rate are lower. Our algorithm is more competitive on empty videos. For objects on the sea floor, our background image is not well optimized but we reach a false non detection rate lower than the reference algorithm. On the videos used in this work, we detect a lot of false objects but these false detections can be filtered out during the identification step.

5 CONCLUSION

In this article, we present a novel method based on background subtraction, on comparison and on the use of navigational information. Our algorithm works in three steps. First we preprocessed

our images. Besides classical preprocessing using only information present on the image, we use Phong's model and the sun position to limit the light effects. Then we detect objects. To do this we use the background subtraction algorithm. Navigation information are necessary to learn background when the distance is sufficiently high to be sure of the absence of the object (Object position has been indicated by the sonar detection). Finally, we increase the detection rate and decrease false alarm rate with post-processing. The use of navigational information and fusion of it with detection results form a novel result improvement method. Knowing the vehicle's motion, we can define constraints on the detected object in images.

We tested our algorithm on data acquired at sea and show that we improve detection results and decrease false alarm rate, comparing to our former works. All algorithms have been applied on the same videos. Detection results obtained with the presented algorithm are higher than those obtained with the algorithm based on the phase of the image spectrum and with the reference algorithm. False alarms on empty videos are reduced. False negative probabilities have been decreased between the proposed algorithm and the reference algorithm. In a word, we detect more underwater mines than the reference algorithm.

There still are some improvements to bring on the background image, especially when objects are on the sea floor. Then further work will consist in identifying the detected objects. This second step will help to reduce the false alarm rate.

References

- [1] Ocean Studies Board National Research Council, *Oceanography and mine warfare*. National Academy Press, 2000.
- [2] I. Leonard, A. Arnold-Bos, and A. Alfalou, "Interest of correlation-based automatic target recognition in underwater optical images: theoretical justification and first results," in *Proc. SPIE*, vol. 7678, 2010.
- [3] A. Arnold-Bos, J. Malkasse, and G. Kervern, "Towards a model-free denoising of underwater optical images," in *Proceedings of the IEEE conference on Ocean (Europe)*, 2005.
- [4] A. Arnold-Bos, J.-P. Malkasse, and G. Kervern, "A pre-processing framework for automatic underwater images denoising," in *Proceedings of the European Conference on Propagation and Systems*, 2005.
- [5] B. Phong, "Illumination for computer generated pictures," *Communications of the ACM*, vol. 18, pp. 311–317, June 1975.
- [6] I. Reda and A. Andreas, "Solar position algorithm for solar radiation applications," *Solar Energy*, vol. 76, pp. 577–589, 2004.
- [7] D. Edgington, D. Cline, D. Davis, I. Kerkez, and J. Mariette, "Detection, tracking, and classifying animals in underwater video," in *MTS/IEEE Oceans 2006 Conference Proceedings*, 2006.
- [8] N. Mandelert and A. Arnold-Bos, "Joint sonar and video sensing for a fire-and-forget underwater mine disposal munition," in *Proc. 3rd conference on Maritime Systems and Technology*, 2008.

Bibliographie

- [1] J. Needham, *Science and civilisation in China*, U. of Cambridge, Editeur. University of Cambridge, 1986.
- [2] T. Hooton, "Hunting the assassin in the deep : mine countermeasures in the 21st century," *Military technology*, vol. March, pp. 66–69, 2012.
- [3] L. C. A. Donahoe, "Les mines vont-elles couler l'US Navy ?" *Défense et sécurité internationale*, vol. Hors série n.14, 2010.
- [4] "Guerre des mines : à l'heure de la robotique," *Mer et marine*, vol. 1, p. 51, 2010.
- [5] [En ligne] Adresse : www.meretmarine.net
- [6] S. Nitschke, "Countermine technologies for naval expeditionary forces : key challenges for detection and clearance of sea, littoral and surf zone mines," *Naval Forces*, vol. VI, pp. 36–42, 2008.
- [7] Ocean Studies Board National Research Council, *Oceanography and mine warfare*. National Academy Press, 2000.
- [8] N. Mandelert et A. Arnold-Bos, "Joint sonar and video sensing for a fire-and-forget underwater mine disposal munition," in *Proc. 3rd conference on maritime systems and technology*, 2008.
- [9] S. Duntley, "Light in the sea," *JOSA*, vol. 53, pp. 214–233, 1963.
- [10] W. Wells, "Loss of resolution in water as a result of multiple small angle scattering," *Journal of the optical society of America*, vol. 59, pp. 686–691, 1969.
- [11] J. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE Journal of Oceanic Engineering*, vol. 15, pp. 101–110, 1990.
- [12] B. McGlamery, "Computer analysis and simulation of underwater camera system performance," University of California, San Diego, Scripps Institution of Oceanography, Visibility Laboratory, Rapport technique, 1975.
- [13] G. Mie, "Beiträge zur optik trüber medien, speziell kolloidaler metallösungen," *Annalen der Physik*, vol. 330, pp. 377–445, 1908.
- [14] [En ligne] Adresse : <http://fr.wikipedia.org/wiki/>
- [15] R. Marks, M. Lee, et S. Rock, "Using visual sensing for control of an underwater robotic vehicle," in *Proceedings of IARP Second Workshop on Mobile Robots for Subsea Environments*, 1994.
- [16] F. Sogandares et E. Fry, "Absorption spectrum (340-640nm) of pure water. i. photothermal measurements," *Applied Optics*, vol. 36, pp. 8699–8709, 1997.
- [17] A. Morel, B. Gentili, H. Claustre, M. Babin, A. Bricaud, J. Ras, et F. Tièche, "Optical properties of the "clearest" natural waters," *Limnol. Oceanogr.*, vol. 52, pp. 217–229, 2007.
- [18] R. Smith et K. Baker, "Optical properties of the clearest natural waters (200-800 nm)," *Applied Optics*, vol. 20, pp. 177–184, 1981.
- [19] M. Bouvet, *Traitement des signaux pour les systèmes sonar*, Masson, Editeur. Masson, 1992.
- [20] X. Lurton, *Acoustique sous-marine. Présentation et applications*, Ifremer, Editeur. Ifremer, 1998.

- [21] M. Ainslie et J. McColm, "A simplified formula for viscous and chemical absorption in sea water," *J. Acoust. Soc. Am.*, vol. 103, pp. 1671–1672, 1998.
- [22] J. D. Jackson, *Classical Electrodynamics*. John Wiley and sons, Inc., 1975, ch. Chapter 7 : Plane electromagnetic waves and wave propagation, pp. 314–316.
- [23] R. Eustice, O. Pizarro, et H. Singh, "Visually augmented navigation in an unstructured environment using a delayed state history," in *ICRa*, 2004.
- [24] P. Blondel, *The handbook of sidescan sonar*, Springer, Editeur. Springer, 2009.
- [25] F. Jean et F. Mosca, "Frontal sonar," U.S. Brevet US 2009/0 231 957 A1, 2009.
- [26] E. Belcher, W. Hanot, et J. Burch, "Dual frequency identification sonar (didson)," in *Proceedings of the 2002 International Symposium on Underwater Technology*, 2002.
- [27] G. Fiocco et G. Grams, "Observations of the upper atmosphere by optical radar in alaska and sweden during the summer 1964," *Tellus*, vol. 18, pp. 34–38, 1966.
- [28] [En ligne] Adresse : <http://www.arete.com>
- [29] D. Neumann, "Range gated imaging system," U.S. Brevet 3 380 358, 1968.
- [30] Y. Schechner, E. Namer, et S. Shwartz, "System and method for dehazing," U.S. Brevet US2011/0 043 603 A1, 2011.
- [31] Y. Schechner et N. Karpel, "Enhanced underwater imaging," U.S. Brevet US 7,804,518 B2, 2010.
- [32] Y. Schechner et N. Karpel, "Recovery of underwater visibility and structure by polarization analysis," *IEEE Journal of Oceanic Engineering*, vol. 30, pp. 570–587, 2005.
- [33] Y. Schechner et N. Karpel, "Clear underwater vision," *Proc. computer vision and pattern recognition*, vol. I, pp. 536–543, 2004.
- [34] N. Karpel et Y. Schechner, "Portable polarimetric underwater imaging system with a linear response," in *Proceedings of SPIE*, 2004.
- [35] [En ligne] Adresse : www.netmarine.net
- [36] [En ligne] Adresse : <http://www.gican.asso.fr/sites/default/files/11-265.jpg>
- [37] [En ligne] Adresse : www.defense.gouv.fr
- [38] [En ligne] Adresse : www.eca-robotics.com
- [39] [En ligne] Adresse : <http://auvac.org/configurations/view/192>
- [40] [En ligne] Adresse : <http://topvision.gesma.fr>
- [41] [En ligne] Adresse : <http://www.saabgroup.com/en/>
- [42] [En ligne] Adresse : <http://www.kongsberg.com/>
- [43] [En ligne] Adresse : <http://www.atlas-elektronik.com/>
- [44] [En ligne] Adresse : <http://www.lockheedmartin.com/>
- [45] A. Olmos, "Detecting underwater man-made objects in unconstrained video images," Thèse de doctorat, Heriot Watt University, 2002.
- [46] S. Bazeille, I. Quidu, L. Jaulin, et J. Malkasse, "Automatic underwater image preprocessing," in *CMM'06 - Caractérisation du milieu marin*, 2006.
- [47] S. Helan, J. Burie, T. Bouwmans, et S. Bazeille, "Object detection in underwater images," in *CMM'06 - Caractérisation du milieu marin*, 2006.
- [48] W. Hou, D. Gray, A. Weidemann, G. Fournier, et J. Forand, "Automated underwater image restoration and retrieval of related optical properties," in *IGARS 2007*, 2007.
- [49] F. Fan, K. Yang, M. Xia, W. Li, B. Fu, et W. Zhang, "Comparative study on several blind deconvolution algorithms applied to underwater image restoration," *Optical Review*, vol. 17, pp. 123–129, 2010.
- [50] L. Chao et M. Wang, "Removal of water scattering," in *2nd International conference on computer engineering and technology*, 2010.

- [51] A. Arnold-Bos, J. Malkasse, et G. Kervern, "Towards a model-free denoising of underwater optical images," in *Proceedings of the IEEE conference on Ocean (Europe)*, 2005.
- [52] A. Rizzi, C. Gatta, et D. Marini, "From retinex to automatic color equalization : issues in developing a new algorithm for unsupervised color equalization," *Journal of Electronic Imaging*, vol. 13, pp. 75–84, 2004.
- [53] M. Chambah, D. Semani, A. Renouf, P. Courtellement, et A. Rizzi, "Underwater color constancy : enhancement of automatic live fish recognition," in *Proc. of SPIE*, 2004.
- [54] L. Torres-Mendez et G. Dudek, "Color correction of underwater images for aquatic robot inspection," in *Energy minimization methods in computer vision and pattern recognition*. Springer Berlin / Heidelberg, 2005.
- [55] K. Iqbal, R. Salam, A. Osman, et A. Talib, "Underwater image enhancement using an integrated colour model," *IAENG International journal of computer science*, vol. 34, pp. 529–534, 2007.
- [56] G. Gilbert et J. Pernicka, "Improvement of underwater visibility by reduction of backscatter with a circular polarization technique," *Applied Optics*, vol. 6, pp. 741–746, 1967.
- [57] G. Lewis, D. Jordan, et P. Roberts, "Backscattering target detection in a turbid medium by polarization discrimination," *Applied Optics*, vol. 38, pp. 3937–3944, 1999.
- [58] C. Brosseau, *Fundamentals of polarized light a statistical optics approach*, J. Wiley et Sons, Editeurs. Wiley Interscience publication, 1998.
- [59] R. Garcia, T. Nicosevici, et X. Cufi, "On the way to solve lighting problems in underwater imaging," in *Proceedings of the IEEE Oceans*, 2002.
- [60] A. Arnold-Bos, J.-P. Malkasse, et G. Kervern, "A pre-processing framework for automatic underwater images denoising," in *Proceedings of the European Conference on Propagation and Systems*, 2005.
- [61] D. Sidorov et A. Kokaram, "Suppression of moiré patterns via spectral analysis," in *Proceedings of SPIE in visual communications and image processing*, 2002.
- [62] I. Leonard, A. Arnold-Bos, et A. Alfalou, "Interest of correlation-based automatic target recognition in underwater optical images : theoretical justification and first results," in *Proceedings of SPIE*, 2010.
- [63] B. Phong, "Illumination for computer generated pictures," *Communications of the ACM*, vol. 18, n. 6, pp. 311–317, June 1975.
- [64] J. Tyler, "The secchi disc," *Limnology and oceanography*, vol. 13, pp. 1–6, 1968.
- [65] I. Reda et A. Andreas, "Solar position algorithm for solar radiation applications," *Solar Energy*, vol. 76, pp. 577–589, 2004.
- [66] J. Canny, "A computational approach to edge detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. PAMI-8, pp. 679–698, 1986.
- [67] A. Oppenheim et J. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 69, pp. 529–541, 1981.
- [68] E. Adelson, C. Anderson, J. Bergen, P. Burt, et J. Ogden, "Pyramid methods in image processing," *RCA Engineer*, vol. 29-6, pp. 33–41, 1984.
- [69] J. Kuo et K. Chen, "The application of wavelets correlator for ships wake detection in sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, pp. 1506–1511, 2003.
- [70] S. Bazeille, I. Quidu, et L. Jaulin, "Identification of underwater man-made object using color," *Proceedings of the Institute of Acoustics*, vol. 29, pp. 45–52, 2007.
- [71] A. Balasuriya et T. Ura, "Underwater cable following by twin-burger2," in *Proceedings of the 2001 IEEE International conference on robotics and automation*, 2001.
- [72] A. Rova, G. Mori, et L. Dill, "One fish, two fish, butterfly, trumpeter : recognizing fish in underwater video," in *IAPR conference on machine vision applications*, 2007.

- [73] R. Prim, "Shortest connection networks and some generalizations," *The Bell system technical journal*, vol. 36, pp. 1389–1401, 1957.
- [74] G. Foresti, "Detecting elliptical structures in underwater images," *Journal of Electronic Imaging*, vol. 11, pp. 59–68, 2002.
- [75] S. Di Zenzo, "A note on the gradient of a multi-image," *Computer Vision, Graphics and Image Processing*, vol. 33, pp. 116–125, 1986.
- [76] D. Kocak, N. da Vitoria Lobo, et E. Widder, "Computer vision techniques for quantifying, tracking and identifying bioluminescent plankton," *IEEE Journal of Oceanic Engineering*, vol. 24, pp. 81–95, 1999.
- [77] C. Barat et R. Phlypo, "A fully automated method to detect and segment a manufactured object in an underwater color image," *EURASIP Journal on advances in signal processing*, vol. 2010, pp. 1–10, 2010.
- [78] L. Itti, C. Koch, et E. Niebur, "A model of saliency based visual attention for rapid scene analysis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 20, pp. 1254–1259, 1998.
- [79] M. Kass, A. Witkin, et D. Terzopoulos, "Snakes : active contour models," *International journal of computer vision*, vol. 1, pp. 312–331, 1988.
- [80] P. Correia, P. Lau, P. Fonseca, et A. Campos, "Underwater video analysis for norway lobster stock quantification using multiple visual attention features," in *15th European signal processing conference*, 2007.
- [81] T. van Valkenburg van Haarst et K. Scholte, "Polynomial background estimation using visible light video streams for robust automatic detection in a maritime environment," in *Proceedings of SPIE*, vol. 7482, 2009.
- [82] D. Edgindton, D. Cline, D. Davis, I. Kerkez, et J. Mariette, "Detection, tracking, and classifying animals in underwater video," in *MTS/IEEE Oceans 2006 Conference Proceedings*, 2006.
- [83] E. Trucco et K. Plakas, "Video tracking : a concise survey," *IEEE Journal of Oceanic Engineering*, vol. 31, pp. 520–529, 2006.
- [84] L. Wang, S. Hu, et X. Zhang, "Detecting and tracking of small moving target under the background of sea level," in *ICSP2008 Proceedings*, 2008.
- [85] R. Eustice, O. Pizarro, et H. Singh, "Visually augmented navigation for autonomous underwater vehicles," *IEEE Journal of Oceanic Engineering*, vol. 33, pp. 103–122, 2008.
- [86] G. Welch et G. Bishop. (2002) An introduction to the kalman filter.
- [87] G. Le Brun, B. Le Jeune, J. Cariou, et J. Lotrian, "Analysis of polarization signature of immersed targets," *SPIE*, vol. 1747, pp. 128–139, 1992.
- [88] L. Wolff, "Polarization vision : a new sensory approach to image understanding," *Image and vision computing*, vol. 15, pp. 81–93, 1997.
- [89] L. Cole, D. Austin, et L. Cole, "Visual object recognition using template matching," in *Australian conference on robotics and automation*, 2004.
- [90] B. Schiele et J. Crowley, "Object recognition using multidimensional receptive field histograms and its robustness to view point changes," in *1995 Rosenon workshop on computational vision*, 1995.
- [91] S. Thorpe, "Ultra rapid scene categorization with a wave of spikes," in *BMVC*, 2002.
- [92] A. Delorme, J. Gautrais, R. Van Rullen, et S. Thorpe, "Spikenet : a simulator for modeling large networks of integrated and fire neurons," *Neurocomputing*, vol. 26-27, pp. 989–996, 1999.
- [93] P. Viola et M. Jones, "Rapid object detection using a boosted cascade of simple features," in *PROceedings IEEE Conference on computer vision and pattern recognition*, 2001.

- [94] A. Alfalou et C. Brosseau, *Face recognition*. In-Tech, 2010, ch. Understanding correlation techniques for face recognition : from basics to applications, pp. 353–380.
- [95] Y. Ouerhani, M. Jridi, et A. Alfalou, “Fast face recognition approach using a graphical processing unit "gpu",” in *IEEE International conference on imaging systems and techniques*, 2010.
- [96] I. Leonard, A. Alfalou, et C. Brosseau, “Spectral optimized asymmetric segmented phase-only filter,” *Applied Optics*, vol. 51, n. 14, pp. 2638–2650, 2012.
- [97] I. Leonard, A. Alfalou, et C. Brosseau, *Face recognition : methods, applications and technology*. Nova publishers, 2012, ch. Face recognition based on composite correlation filters : analysis of their performances.
- [98] I. Leonard, A. Alfalou, et M. Alam, “Nonlinear fringe adjusted jtc based face tracking using an adaptive decision criterion,” in *Proc. SPIE*, vol. 8134, 2011.
- [99] J. Tribillon, *Traitement optique de l'information et reconnaissance des formes par voie optique*, Teknea, Editeur. Teknea, 1998.
- [100] A. VanderLugt, “Signal detection by complex spatial filtering,” *IEEE Transactions on Information Theory*, vol. 10, pp. 139–145, 1964.
- [101] C. Weaver et J. Goodman, “A technique for optically convolving two functions,” *Applied Optics*, vol. 5, pp. 1248–1249, 1966.
- [102] K. Rice, T. Taha, A. Chowdhury, A. Awwal, et D. Woodard, “Design and acceleration of phase only filter based optical pattern recognition for fingerprint identification,” *Optical Engineering*, vol. 48, pp. 117 206–1 – 117 206–11, 2009.
- [103] J. Horner, “Metrics for assessing pattern-recognition performance,” *Applied Optics*, vol. 31, pp. 165–166, 1992.
- [104] A. Alfalou, “Implantation optique de corrélateurs multivoies appliqués à la reconnaissance des formes,” Thèse de doctorat, Université de Rennes, 1999.
- [105] P. Katz, A. Alfalou, C. Brosseau, et M. Alam, *Face recognition : methods, applications and technology*. Nova publishers, 2012, ch. Correlation and independent component analysis based approaches for biometric recognition.
- [106] J. Horner et P. Gianino, “Phase-only matched filtering,” *Applied Optics*, vol. 23, pp. 812–816, 1984.
- [107] B. Javidi, S. Odeh, et Y. Chen, “Rotation and scale sensitivities of the binary phase-only filter,” *Optics Communications*, vol. 65, pp. 233–238, 1988.
- [108] J. Horner, B. Javidi, et J. Wang, “Analysis of the binary phase only filter,” *Optics Communications*, vol. 91, pp. 189–192, 1992.
- [109] R. Gonzalez et P. Wintz, *Digital Image Processing*. Addison Wesley, 1987.
- [110] P. Refregier, “Optimal trade-off filters for noise robustness, sharpness of the correlation peak and horner efficiency,” *Optics Letters*, vol. 16, pp. 829–831, 1991.
- [111] C. Hester et D. Casasent, “Multivariant technique for multiclass pattern recognition,” *Applied Optics*, vol. 19, pp. 1758–1761, 1980.
- [112] B. V. Kumar, “Tutorial survey of composite filter designs for optical correlators,” *Applied Optics*, vol. 31, pp. 4773–4801, 1992.
- [113] A. Alfalou, G. Keryer, et J. de Bougrenet de la Tocnaye, “Optical implementation of segmented composite filtering,” *Applied Optics*, vol. 38, pp. 6129–6134, 1999.
- [114] J. Lewis, “Fast normalized cross-correlation,” *Vision interface*, vol. 10, pp. 120–123, 1995.
- [115] N. Gourier, D. Hall, et J. Crowley, “Estimating face orientation from robust detection of salient facial features,” in *Proc. of Pointing 2004, ICPR, International workshop on visual observation of deictic gestures*, 2004.
- [116] C.-T. Li, S. Yin, et F. T. Yu, “Nonzero order joint transform correlator,” *Optical Engineering*, vol. 37, pp. 58–65, 1998.

- [117] B. Javidi et C. Kuo, "Joint transform image correlation using a binary spatial light modulator at the fourier plane," *Applied Optics*, vol. 27, pp. 663–665, 1988.
- [118] B. Javidi, "Nonlinear joint power spectrum based optical correlation," *Applied Optics*, vol. 28, pp. 2358–2367, 1989.
- [119] M. Alam et M. Karim, "Fringe-adjusted joint transform correlation," *Applied Optics*, vol. 32, pp. 4344–4350, 1993.
- [120] A. Bal et M. Alam, "Dynamic target tracking with fringe-adjusted joint transform correlation and template matching," *Applied Optics*, vol. 43, pp. 4874–4881, 2004.
- [121] S. Tonissen et R. Evans, "Performance of dynamic programming techniques for track-before-detect," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, pp. 1440–1451, 1996.
- [122] S. Davey, M. Rutten, et B. Cheung, "A comparison of detection performance for several track-before-detect algorithms," *EURASIP Journal on advances in signal processing*, vol. 2008, pp. 1–10, 2008.
- [123] L. Rabiner et B. Juang, "An introduction to hidden markov models," *IEEE ASSP Magazine*, vol. 3, pp. 4–16, 1986.
- [124] R. Kalman, "A new approach to linear filtering and prediction problems," *Journal of basic engineering*, vol. 82, pp. 35–45, 1960.
- [125] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 257–286, 1989.
- [126] R. Horaud et O. Monga, *Vision par ordinateur : outils fondamentaux*, E. Hermes, Editeur. Editions Hermes, 1993.
- [127] T. Chan, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, pp. 266–277, 2001.
- [128] Y. Ouerhani, "Contribution à la définition, à l'optimisation et à l'implantation d'ips de traitement du signal et des données en temps réel sur des cibles programmables," Thèse de doctorat, Université de Bretagne Occidentale, 2012.
- [129] Y. Ouerhani, M. Jridi, et A. Alfalou, "Implementation techniques of high-order fft into low-cost fpga," in *IEEE 54th international midwest symposium on circuits and systems*, 2011.
- [130] Y. Ouerhani, M. Jridi, et A. Alfalou, "Area-delay efficient fft architecture using parallel processing and new memory sharing technique," *Journal of circuits, systems and computers*, vol. 21, 2012.
- [131] A. Awwal, "What can we learn from the shape of a correlation peak for position estimation?" *Applied Optics*, vol. 49, pp. B40–B50, 2009.
- [132] E. Lengyel, *Mathematics for 3D game programming and computer graphics*, D. Pallai, Editeur. Charles River Media, Inc., 2004.