Université de Bretagne Occidentale
U.F.R Sciences et Techniques
ED SICMA 0373

THÈSE

Pour obtenir le grade de
DOCTEUR EN SCIENCES

par

Stéphane BAZEILLE

Vision sous-marine monoculaire pour la reconnaissance d'objets

Thèse présentée et soutenue publiquement le 1 Octobre 2008 devant le jury composé de :

Alain TREMEAU	Professeur : Université de Saint-Étienne	*Rapporteur*
Yvan PETILLOT	Maître de conférences : Université d'Heriot Watt	*Rapporteur*
Luc JAULIN	Professeur : Université de Bretagne Occidentale	*Directeur de thèse*
Gilles BUREL	Professeur : Université de Bretagne Occidentale	*Examinateur*
Benoît ZERR	Expert DGA/GESMA	*Examinateur*
Isabelle QUIDU	Maître de conférences : Université de Bretagne Occidentale	*Examinatrice*
Jean-Philippe MALKASSE	Responsable du laboratoire LTSI - Service d'études générales sonar, Thales Underwater Systems S.A.S.	*Invité*

Travaux préparés au sein du laboratoire E^3I^2 - Extraction et Exploitation de l'Information en Environnements Incertains (EA-3876) de l'ENSIETA (École Nationale Supérieure des Ingénieurs des Études et Techniques d'Armement).

Remerciements

Comme le veut la coutume, je voudrais utiliser ces quelques lignes hors sujet que l'on m'a octroyées pour remercier tous ceux qui ont contribué à l'aboutissement de ces travaux de thèse.
Ainsi, pour commencer, je tiens à remercier l'ensemble des personnes de mon entourage (famille, copains, copines et collègues) qui ont partagé ma vie durant ces trois belles années et je citerais notamment : mes parents, mon frère, Anne, Clairette, Régis, Mélisse, Steph, Lucie, Ronan, Vince, Géra, Sandra, Sam et Cédric. Je terminerais cette petite énumération malheureusement non exhaustive par mes collègues de bureau, Isabelle, Fabrice, Lionel, Sandrine, et Laurent.
Dans un deuxième temps, je remercie Ali Khenchaf, directeur du laboratoire E^3I^2 qui m'a accueilli le temps de cette thèse, la région Bretagne qui a financé mes travaux et l'ensemble des collaborateurs scientifiques (chercheurs, enseignants, doctorants) qui ont répondu à mes nombreuses interrogations scientifiques durant ces trois ans. Parmi ceux que je n'ai pas cités plus haut : Andréas, Arunas, Michel, Christophe, Jean-Christophe et Hélène. J'adresse pour finir un remerciement tout particulier à Annick Billon-Coat secrétaire du laboratoire pour le soutien administratif qu'elle a assuré.
Ensuite, j'exprime ma reconnaissance à l'ensemble des personnes composant mon jury de thèse et notamment mes rapporteurs : Alain Trémeau et Yvan Pétillot, mes examinateurs : Gilles Burel et Benoît Zerr, Jean-Philippe Malkasse intervenant industriel et membre invité, et Isabelle Quidu examinatrice mais surtout encadrante que je remercie chaleureusement pour son suivi, son soutien dans mes travaux, et pour ses nombreuses relectures du présent manuscrit.
Enfin, je terminerais en remerciant tout particulièrement mon directeur de thèse Luc Jaulin, qui a su orienter, critiquer, faire progresser mes travaux dans un sens nouveau, et surtout qui m'a passionné par mon sujet et par la recherche grâce à son encadrement presque quotidien.

Pour conclure ce petit aparté, j'ajoute un remerciement général à l'ensemble du personnel de l'ENSIETA, les doctorants et les personnes que j'aurais malencontreusement oubliées dans cette page, et je tiens à dire deux mots sur cette expérience extraordinaire qu'a été pour moi la thèse de doctorat. En effet, à coté de l'expérience de recherche et de l'enrichissement scientifique qui en sont les objectifs fondamentaux, ces trois années de thèse ont été un enrichissement personnel et humain encore bien plus grand, et ce notamment par la grande variété du travail de thèse : documentation, recherche et développement, publication et présentation des travaux, préparation de cours, enseignements et encadrements, auxquels j'ai ajouté dans mon cas une implication forte dans le vie de l'école concrétisée par la présidence de l'association des doctorants pendant plus d'un an. Ces travaux sont le fruit d'un investissement colossal qui restera pour moi une expérience de vie inoubliable. Sur ces derniers mots, j'espère que la lecture de ce manuscrit vous sera plaisante et surtout instructive et je vous souhaite une bonne lecture.

Stéphane BAZEILLE

Vision sous-marine monoculaire pour la reconnaissance d'objets

Stéphane BAZEILLE

Laboratoire E^3I^2-EA 3876,
École Nationale Supérieure des Ingénieurs
des Études et Techniques d'Armement,
2, rue François Verny, 29806 Brest cedex, France.

Mots clés liés au projet : Vision, Sous-marin, Reconnaissance, Robot autonome.

Résumé : Dans le contexte sous-marin, et à l'inverse du capteur sonar qui reste le plus employé à grande distance pour la détection et la classification, la caméra vidéo est efficace à faible portée lors des phases d'approche, de reconnaissance d'objets et d'intervention. Elle dispose en effet d'atouts notables comme une haute résolution, une facilité d'interprétation ou encore un faible coût. Aujourd'hui presque tous les véhicules sous-marins scientifiques, industriels ou militaires en sont équipés. Ils sont actuellement plutôt télé-opérés par un opérateur et on y trouve associés des traitements automatiques que très rarement. Ces traitements automatiques sont pourtant des technologies essentielles pour les développements émergents des robots autonomes sous-marins très prisés aujourd'hui dans un contexte d'expansion des marchés liés à la sécurité et à l'exploitation des ressources maritimes. Ces travaux de thèse visent à apporter les innovations nécessaires, et à promouvoir l'emploi du capteur vidéo dans le domaine sous-marin.

L'étude proposée concerne le développement des traitements automatiques de reconnaissance d'objets en vidéo sous-marine, avec une attention particulière apportée aux objets sous-marins nuisibles (polluant, dangereux, voire létaux). On peut citer par exemple les fûts toxiques, les containers, les batteries à ions lourds, les munitions, les mines, et divers objets manufacturés de grande consommation. Les scènes sous-marines observées sont classiquement plus simples et plus limitées en profondeur d'observation que les scènes urbaines ou que l'intérieur d'un bâtiment. Toutefois, ce contexte présente des difficultés spécifiques telles que les variations d'éclairage et la turbidité de l'eau qui limitent la visibilité et dégradent fortement les images. Ceci a pour conséquence de rendre les traitements difficiles et nécessite donc la création de nouveaux algorithmes de vision robotique. Quatre contributions peuvent résumer l'ensemble de cette étude. D'abord un état de l'art général sur les méthodes de reconnaissance d'objets en vision sous-marine ainsi que sur les méthodes de prétraitement des images sous-marines. Ensuite le développement de deux nouvelles méthodes : une méthode générique et sans *a priori* basée sur la forme adaptée de techniques existantes et une méthode originale basée sur la reconnaissance par la couleur. Enfin ces deux méthodes ont été implémentées et validées sur des données réelles en contexte opérationnel.

Dimension régionale du projet : Ces travaux de thèse financés par la région Bretagne, ont débuté le 1 Octobre 2005 au sein du Laboratoire E^3I^2-EA 3876 de l'ENSIETA à Brest, sous la direction M. Luc JAULIN (Professeur des Universités) et Mlle Isabelle QUIDU (Maître de Conférences). Ils s'ancrent dans une dynamique de compétences relatives à la mer au travers d'un double intérêt à la fois pour la recherche et pour l'industrie locale. Ils bénéficient des données réelles acquises dans le cadre du projet TOPVISION (Test Opérationnels de VIdéos Sous-marines pour l'Identification d'Objets Nuisibles). Ce projet fait partie du programme Techno-Vision lancé en 2005 par le Ministère de la Recherche et le Ministère de la Défense. Il vise à réaliser une campagne d'évaluation d'algorithmes de traitement en vidéo sous-marine pour la détection, la localisation, la caractérisation et l'identification automatiques d'objets nuisibles. Trois des quatre membres de son comité d'organisation appartiennent à la région Bretagne : Thales Underwater Systems (TUS), le Groupe d'Études Sous-Marines de l'Atlantique (GESMA, DGA/DET), et l'ENSIETA. Les principaux domaines abordés par le projet sont la technologie d'observation, le traitement vidéo, la sécurité environnementale, et la sécurité maritime.

Monocular underwater vision for object recognition

Stéphane BAZEILLE

E^3I^2-EA 3876 Laboratory,
École Nationale Supérieure des Ingénieurs
des Études et Techniques d'Armement,
2, rue François Verny, 29806 Brest cedex, France

Keywords : Computer vision, Underwater, Recognition, Autonomous Vehicle.

Abstract : In underwater context, traditional sensing methods like sonar are used at large range for the detection and the classification of objects. For a few years, the sonar sensor has been complemented by a vision sensor more efficient at short range for the approach, the object recognition and the intervention phases. Indeed, the camera has noticeable advantages like for example high resolution, simple interpretation or low cost. Today, it equips nearly any scientific, industrial or military underwater vehicles. Currently, underwater vehicles are rather distance controlled by a manipulator, and automatic processing is really uncommon.
However, automatic processing is an essential technology for the AUV[1] development. These kinds of vehicles are very popular today with the expansion of the market related to the security and the exploitation of the marine resources. This thesis aims at supplying the required advances and promoting the use of video technologies. The proposed study addresses the problem of automatic processing for underwater objects recognition, with a particular focus on harmful underwater objects (polluting, dangerous, or lethal). We can cite for example, toxic drums, containers, batteries, munitions, mines, and various consumers man-made objects. Observed underwater scenes are traditionally simpler but more limited in depth than urban scenes or inside building scenes. Nevertheless, this context presents very specific difficulties like lighting variations and water turbidity which limit visibility and degrade considerably images. As a consequence processing is difficult and require the development of new robotic vision algorithms. Four main contributions have been achieved by this study. First of all, a general literature survey on underwater vision based methods and underwater image pre-processing algorithms. Then, the development of two novel methods : a generic and without prior knowledge method based on the adaptation of existing shape techniques and an original method based on colour recognition. Finally, these two algorithms have been implemented and assessed on real data in operational context.

[1] Autonomous Underwater Vehicle

Table des matières

Avant-propos **3**

1 Introduction générale **11**

2 Vidéos et images sous-marines **15**
2.1 Généralités . . . 15
2.1.1 Qu'est-ce qu'une image ? . . . 15
2.1.2 Qu'est-ce qu'une vidéo ? . . . 15
2.1.3 Caractéristiques d'une image numérique . . . 15
2.2 Acquisition et perception . . . 16
2.2.1 La vision humaine . . . 17
2.2.2 La vision artificielle . . . 19
2.3 Les espaces de représentation des couleurs . . . 21
2.3.1 Les modèles de référence de la CIE : CIE RGB et CIE XYZ . . . 21
2.3.2 Les modèles dérivés d'un système matériel : RGB et YCbCr . . . 23
2.3.3 Les modèles perceptuels : HSL et HSV . . . 24
2.4 Spécificités des images sous-marines . . . 25
2.4.1 Propriétés optiques de la lumière dans la mer . . . 25
2.4.2 Problèmes spécifiques des images sous-marines . . . 26
2.4.3 Particularités des prises de vues sous-marines . . . 29
2.5 Conclusion . . . 31

3 Reconnaissance d'objets sous-marins à l'aide du capteur vidéo **33**
3.1 La reconnaissance d'objets sous-marins . . . 33
3.1.1 Du sonar à la caméra . . . 33
3.1.2 Un exemple de contexte opérationnel . . . 34
3.1.3 Le capteur optique dans le contexte sous-marin . . . 36
3.1.4 L'existant dans le domaine de la vision sous-marine . . . 38
3.2 État de l'art général en reconnaissance d'objets par vision . . . 39
3.2.1 La vision par ordinateur . . . 39
3.2.2 Reconnaissance de formes et analyse de scènes . . . 40
3.2.3 Les méthodes utilisant l'information 3D . . . 41
3.2.4 Les méthodes utilisant l'information 2D . . . 43
3.3 Les méthodes utilisables dans notre contexte . . . 44

3.3.1 La méthode 2D classique orientée forme . . . 45
3.3.2 La méthode 2D cherchant la compatibilité 2D/3D . . . 46
3.3.3 La méthode 2D cherchant la compatibilité de couleur . . . 46
3.4 Conclusion . . . 46

4 Le prétraitement des images sous-marines 49
4.1 Approche globale ou approche séquentielle ? . . . 49
4.1.1 La restauration appliquée aux images sous-marines . . . 49
4.1.2 La méthode adoptée : l'amélioration . . . 53
4.2 État de l'art des méthodes de filtrage . . . 54
4.2.1 Le filtrage spatial . . . 54
4.2.2 Le filtrage fréquentiel . . . 60
4.2.3 Le filtrage multirésolution . . . 65
4.2.4 La correction des couleurs . . . 72
4.3 Le prétraitement des images sous-marines développé . . . 75
4.3.1 Un enchaînement de correction . . . 75
4.3.2 Résultats de l'algorithme de prétraitement . . . 78
4.3.3 Démonstration et quantifications des résultats . . . 78
4.4 Conclusion du chapitre . . . 82

5 La méthode de reconnaissance classique 85
5.1 Une méthode classique mais générique . . . 85
5.2 Détection de changement et prétraitement . . . 86
5.2.1 La détection de changement dans les images . . . 86
5.2.2 Le prétraitement . . . 86
5.3 Segmentation des images . . . 88
5.3.1 Les différentes méthodes : contours ou régions . . . 89
5.3.2 La méthode de détection de contours adoptée . . . 90
5.4 Post-traitement à la segmentation . . . 96
5.4.1 Fermeture des contours . . . 96
5.4.2 Sélection des contours pertinents . . . 98
5.5 La reconnaissance à partir des contours . . . 100
5.5.1 La recherche de primitives . . . 101
5.5.2 Reconnaissance des objets manufacturés . . . 103
5.6 Conclusion du chapitre . . . 105

6 Des méthodes de reconnaissance spécialisées à un type d'objet 109
6.1 Des méthodes pour s'adapter à un contexte difficile . . . 109
6.1.1 A l'opposé des méthodes classiques . . . 109
6.1.2 Intérêt de ces méthodes spécialisées . . . 110
6.2 La reconnaissance par la couleur . . . 110
6.2.1 La couleur dans la reconnaissance d'objets . . . 110
6.2.2 Perception de la couleur d'un objet . . . 111
6.2.3 Principe d'illumination d'un objet . . . 112

6.2.4 Modèle de réflexion de Lambert . . . 112
6.3 Des couleurs compatibles d'un point de vue sous-marin . . . 113
6.3.1 Propagation de la lumière dans l'eau . . . 113
6.3.2 Loi de Beer-Lambert . . . 113
6.3.3 Équivalence des couleurs en milieu sous-marin . . . 114
6.4 Les algorithmes de reconnaissance développés . . . 119
6.4.1 Un modèle simple : la droite de compatibilité . . . 119
6.4.2 Un modèle amélioré : la nappe de compatibilité . . . 123
6.4.3 Comparaison et validation des méthodes . . . 129
6.4.4 Discussions et perspectives : le volume de compatibilité . . . 130
6.5 Conclusion . . . 131

7 Applications des méthodes développées 135
7.1 Deux contextes d'utilisation très différents . . . 135
7.1.1 Description générale du projet TOPVISION . . . 135
7.1.2 Description générale du concours SAUC'E . . . 138
7.2 TOPVISION : reconnaissance d'objets par les contours . . . 140
7.2.1 La solution proposée à partir de l'algorithme développé . . . 140
7.2.2 L'étape évaluée : La détection . . . 140
7.2.3 Les trois autres étapes . . . 141
7.2.4 Illustration des résultats . . . 144
7.2.5 Conclusion et critique de méthode . . . 145
7.3 SAUC'E : reconnaissance par la couleur . . . 147
7.3.1 Quelques résultats en contexte opérationnel . . . 147
7.3.2 Exemple 1 : localisation et contact avec l'objet coloré . . . 148
7.3.3 Exemple 2 : cartographie de la zone de compétition . . . 148
7.3.4 Améliorations, perspectives et conclusion. . . . 151
7.4 Conclusion du chapitre . . . 152

8 Conclusion générale 155

Annexes 161

A Calibration et reconstruction 3D 161

B Le robot SAUCISSE 163

C La webcam : un capteur couleur 165

Glossaire 167

Publications associées 169

Bibliographie 170

On se lasse de tout, excepté d'apprendre.

Virgile.

Chapitre 1

Introduction générale

A la découverte du monde du silence.

Le milieu marin couvre environ 70% de la surface terrestre mais reste encore aujourd'hui très mal connu en raison de sa difficile accessibilité pour l'homme. Son exploration et sa caractérisation font l'objet de nombreux travaux de recherche depuis plusieurs années comme l'atteste l'augmentation croissante du nombre d'ouvrages, de publications, de documentaires ou de reportages télévisuels sur le sujet. Ces études en sont même d'autant plus importantes aujourd'hui avec "la prise de conscience mondiale" sur les problèmes écologiques liés à l'activité humaine.Malgré des progrès incontestables, l'étude du milieu marin nécessite toujours des avancées technologiques importantes notamment en ce qui concerne le développement des véhicules sous-marins et en particulier de leurs capteurs de perception.
Le capteur sonar utilisé depuis près d'un siècle, est le moyen de sondage privilégié de l'environnement marin, il est le plus employé à grande et moyenne distance pour la détection et la classification. A l'inverse, la caméra vidéo est efficace à faible portée lors des phases d'approche, de reconnaissance d'objets et d'intervention grâce à sa haute résolution et sa facilité d'interprétation. En raison de ces avantages ainsi que pour son faible coût, elle équipe aujourd'hui presque tous les véhicules sous-marins scientifiques, industriels ou militaires. Malgré cette utilisation très répandue, elle est plutôt télé-opérée par un opérateur humain et on y trouve associé des traitements automatiques que très rarement (contrairement aux traitements des images sonar qui sont aujourd'hui largement automatisés). Ces traitements automatiques sont pourtant essentiels pour l'émergence des véhicules autonomes sous-marins (AUV[1]).

La recherche en matière de robotique sous-marine est aujourd'hui très florissante notamment parce qu'elle touche de nombreuses applications dans des secteurs variés tels que :

- la sécurité et l'exploitation des ressources maritimes :
  - détection d'objets manufacturés (objets dangereux ou polluants),
  - surveillance du matériel (structure offshore ou pipeline),
  - surveillance des animaux marins (cétacés, poissons, crustacés, mollusques, plancton...),
- ou encore la navigation automatique :
  - détection d'amers,

[1] en anglais AUV signifie Autonomous Underwater Vehicle

  - cartographie,
  - localisation,
  - évitement d'obstacles.
- l'écologie, l'éthologie, la géologie, ou l'archéologie :
  - étude des êtres vivants et du milieu physique dans lequel ils vivent,
  - étude du comportement des animaux dans leur milieu naturel,
  - observation des fonds marins pour l'étude de la Terre,
  - recherche d'épaves,

Ces travaux de thèse visent à apporter les innovations nécessaires pour promouvoir l'emploi de ces technologies essentielles. L'étude proposée concerne le développement des traitements automatiques de reconnaissance d'objets en vidéo sous-marine, avec une attention particulière apportée aux objets sous-marins nuisibles (polluants, dangereux, voire létaux) dans le cadre du projet TOPVISION[2]. On peut citer par exemple les fûts toxiques, les containers, les batteries à ions lourds, les munitions, les mines, et les objets manufacturés divers souvent de grande consommation...
Les scènes sous-marines observées sont classiquement plus simples et plus limitées en profondeur d'observation que les scènes terrestres. Toutefois, ce contexte présente des difficultés spécifiques telles que les variations d'éclairage, la turbidité de l'eau, et la dynamique du véhicule qui limitent la visibilité. Les phénomènes optiques dus aux propriétés particulières de propagation de la lumière dans la mer doivent aussi être prises en compte. Les vidéos à traiter présentent donc des variations spatiales et temporelles importantes. Elles sont constituées d'images très bruitées, peu contrastées, aux couleurs atténuées et présentent souvent un éclairage non uniforme, ce qui rend difficile l'utilisation des traitements classiques et nécessite donc la création de nouveaux algorithmes de vision robotique.

Dans l'objectif de traitement automatique de ces vidéos sous-marines, plusieurs pistes d'exploration sont possibles : corriger ces problèmes spécifiques pour ensuite utiliser les méthodes classiques de reconnaissance d'objets dans les images, utiliser des méthodes plus robustes, ou apporter plus d'informations *a priori*. Dans cette thèse nous avons étudié ces trois approches.
D'abord nous avons testé la correction des problèmes en utilisant une méthode de prétraitement pour pouvoir utiliser les méthodes classiques d'analyse d'image. Les méthodes existantes se concentrant presque uniquement sur la correction de l'éclairage et des couleurs [1][2] et nécessitant la plupart du temps des connaissances *a priori* précises sur les conditions d'acquisition (par exemple la profondeur, la distance objet/caméra, éventuellement la "qualité" de l'eau), nous avons développé notre propre méthode rapide et automatique pour s'affranchir de l'ensemble des défauts précédemment cités. Cet algorithme de prétraitement des images sous-marines est la première contribution de cette thèse.
Les images nettoyées de ces défauts, nous avons travaillé sur la méthode classique de reconnaissance par la forme en milieu perturbé. Cette méthode basée sur un attribut robuste : les contours, est de loin la plus utilisée dans la littérature du domaine (on citera notamment [1]). Malgré tout, les résultats obtenus par les méthodes existantes ne sont pas vraiment satisfaisants nous avons donc amélioré l'étape de segmentation en adaptant les traitements au milieu sous-marin. L'amélioration de cette méthode constitue la deuxième contribution de cette thèse. Elle sera d'ailleurs exploitée sur les données réelles

[2] Certaines informations objets de ce manuscrit ont été obtenus à partir de l'utilisation de données propriété de l'État Français qui ont été fournies par le GESMA (Groupe d'Études Sous-Marines de l'Atlantique) dans le cadre du projet TOPVISION coordonné par Thales Underwater Systems SAS. Ce projet est rattaché au programme Techno-Vision lancé par le Ministère de la Recherche français et le Ministère de la Défense français.

obtenues grâce à la participation au projet TOPVISION visant à réaliser une campagne d'évaluation des algorithmes de reconnaissance d'objets sous-marins à l'aide du capteur vidéo.
Ensuite nous avons développé un autre type de méthodes plus spécialisées mais plus robustes. Assurément, en contexte difficile il est nécessaire d'apporter plus de connaissance, plus d'*a priori* pour améliorer les résultats. La connaissance supplémentaire que nous avons décidé d'apporter est simple et basée sur une constatation générale : on connaît la plupart du temps l'objet que l'on recherche. Supposant connus les objets recherchés, nous avons développé un algorithme recherchant la compatibilité entre la scène représentée dans l'image et l'objet recherché au sens d'un critère. Dans le domaine sous-marin ce critère peut être la forme ou la couleur. La texture ne pouvant être exploitée en raison de la pauvre qualité des images. Par cette approche nous nous attacherons donc seulement à démontrer la compatibilité entre la silhouette 2D et le modèle 3D ou la compatibilité de couleur. Nous avons développé dans cette thèse, la reconnaissance sur compatibilité de couleur qui est la troisième contribution de cette thèse. La compatibilité de forme a été étudiée mais demande de nombreux développements qui sont encore en cours aujourd'hui.
La méthode développée sur la couleur est une méthode originale qui consiste à modéliser la modification subie par la couleur dans l'eau pour pouvoir la retrouver malgré le problème de constance de couleur dû à l'absorption en milieu marin. Elle a aussi été testée et validée sur des données réelles dans le cadre du concours de robotique sous-marine SAUC'E. Les trois principales avancées de cette thèse qui viennent d'être évoquées ont fait l'objet de publications dans des revues et des conférences nationales et internationales. Elles laissent place à de nombreux travaux pouvant donner suite à ce travail. Elles seront chacune développées dans des chapitres indépendants dans ce manuscrit.

Pour résumer les contributions de cette thèse sont :

- un état de l'art sur les méthodes de reconnaissance d'objets, et de prétraitement des images,
- la création d'une méthode de prétraitement automatique permettant d'améliorer la détection de contour,
- le développement d'une méthode de reconnaissance automatique d'objets manufacturés à partir de vidéos sous-marines,
- la conception d'une nouvelle méthode de reconnaissance basée sur la connaissance *a priori* de la couleur,
- et pour finir, l'application de ces méthodes en contexte opérationnel sur des données réelles.

Pour conclure, le plan de ce manuscrit sera le suivant :

- La présente introduction constitue le **chapitre 1**. Elle place cette thèse dans le contexte actuel de la recherche et décrit brièvement les tenants et aboutissants de ce travail.
- Le **chapitre 2** présentera des généralités sur les données images et vidéos à traiter, notamment, le codage des images et des vidéos, le système de vision humain et artificiel, les espaces de représentation des couleurs, et les spécificités des images sous-marines.
- Dans le **chapitre 3**, nous étudierons les différentes méthodes de reconnaissance existantes dans le domaine de la vision par ordinateur puis nous exposerons celles qui sont applicables dans le domaine sous-marin avec les contraintes liées à ce contexte. Pour finir nous détaillerons nos choix quant aux méthodes que nous avons développées.
- Dans le **chapitre 4**, nous exposerons l'algorithme de prétraitement des images sous-marines que

nous avons développé. Ce nouvel algorithme automatique permet de corriger les défauts les plus perturbateurs dus au milieu sous-marin et améliore de façon significative les résultats obtenus par l'utilisation des méthodes classiques de segmentation.

- Le **chapitre 5** détaillera les adaptations et innovations appliquées à la méthode classique pour reconnaître une scène d'intérêt puis pour reconnaître des objets manufacturés dans les images sous-marines. Cette méthode est générique et basée sur la reconnaissance par les contours.
- Le **chapitre 6** présentera quant à lui un autre concept pour la reconnaissance d'objet celui de la recherche d'un attribut connu *a priori*. Suivant ce concept nous présenterons une méthode originale de reconnaissance par la couleur de l'objet. Nous discuterons dans cette partie des perspectives qu'offre cette méthode en contexte difficile.
- Dans le **chapitre 7**, nous présenterons des résultats applicatifs sur les deux méthodes de reconnaissance développées. La première méthode comprenant prétraitement, segmentation, post-traitement et reconnaissance par les contours bénéficiera des données réelles obtenues en mer dans le cadre de la participation au projet TOPVISION. La deuxième méthode basée sur la couleur sera exploitée sur des données réelles obtenues en piscine dans le cadre de concours de robotique sous-marine SAUC'E.
- Enfin nous conclurons dans le **chapitre 8** sur les apports de cette thèse quant au sujet initial en revenant sur l'existant dans la littérature puis nous présenterons les suites nombreuses que nous pourrons donner à ces travaux.

Quelques annexes développant des points précis non intégrés dans les différents chapitres ont été ajoutées à la fin de ce document ainsi qu'un glossaire regroupant les définitions des différents termes techniques et acronymes utilisés.

Chapitre 2

Vidéos et images sous-marines

Ce chapitre est une introduction aux concepts principaux de l'infographie. Il présente de manière succincte un ensemble de notions nécessaires à la compréhension globale du présent document. De nombreuses références permettront au lecteur d'approfondir ces concepts s'il le souhaite. On citera notamment les thèses [3][4] qui font aussi une introduction intéressante aux principaux concepts présentés.

2.1 Généralités

2.1.1 Qu'est-ce qu'une image ?

D'une manière générale, on considère une image comme une représentation, le plus souvent planaire, d'une scène ou d'un objet situé dans un espace à trois dimensions.

2.1.2 Qu'est-ce qu'une vidéo ?

Un flux vidéo est composé d'une succession d'images composant l'illusion de mouvement. Les fréquences standard sont de 24 (Cinéma), 25 (PAL) ou 30 (NTSC) images/seconde.

2.1.3 Caractéristiques d'une image numérique

On désigne sous le terme d'image numérique toute image acquise, créée, traitée ou stockée sous forme binaire. Une image numérique est composée d'un ensemble de points appelés pixels (picture element). Le pixel représente le plus petit élément constitutif d'une image numérique. L'ensemble des pixels est en général contenu dans un tableau à deux dimensions appelé bitmap constituant l'image (dans le cas contraire on parle d'images vectorielles ou géométriques mais ces types d'images ne seront pas abordés dans ce document).

Les deux caractéristiques principales d'une image numérique sont sa définition et sa résolution. On appelle définition le nombre de pixel de l'image, et la résolution le nombre de points par unité de surface (exprimé en Points Par Pouce PPP, en anglais, Dots Per Inch DPI). On peut ajouter une troisième caractéristique qui est son format.

Chaque pixel d'une image numérique est caractérisé par sa couleur codée sur un certain nombre de bits. Il existe plusieurs standards de codage :

- le codage noir et blanc : en stockant un bit pour chaque pixel,
- le codage niveaux de gris : en stockant un octet dans chaque pixel (dans ce cas il est possible de définir 256 dégradés de gris allant du noir au blanc),
- le codage couleur : où chaque pixel est représenté par trois octets pour les trois composantes RGB (voir théorie trichromatique dans la partie suivante), c'est-à-dire au total 24 bits définissant à peu près 16 millions de couleurs. On peut attribuer à une image un canal supplémentaire sur un octet, appelé canal alpha, qui définit le degré de transparence de l'image.

Pour plus de détails sur les images numériques on pourra se reporter par exemple à [5][6][7].

2.2 Acquisition et perception

Nous venons de voir que chaque pixel est caractérisé par sa couleur. Cette couleur ou intensité dépend elle même d'un autre phénomène : la lumière.

La lumière, c'est l'ensemble des rayonnements électromagnétiques visibles (voir Fig. 2.1), c'est-à-dire susceptibles d'être perçus directement par un oeil humain (dont les longueurs d'onde sont comprises entre 400 et 780 nm). Les particules de lumière sont appelées les photons.

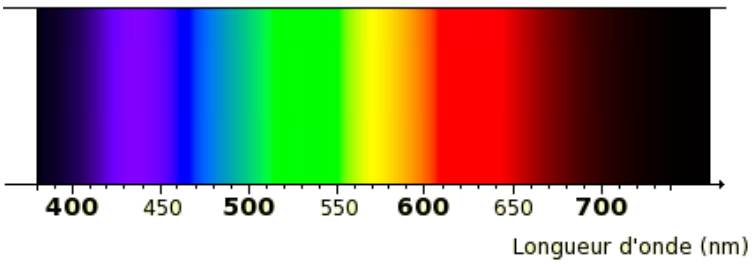


FIG. 2.1 – Spectre de la lumière visible.

La couleur est une notion beaucoup plus subjective, elle n'a pas de réalité physique, elle est une construction de notre cerveau [8][9][10]. Des différences énormes existent dans la perception des couleurs selon les espèces. D'une manière générale, on peut définir la couleur comme une sensation perçue par l'oeil, qui est caractéristique de la longueur d'onde ou du mélange de longueurs d'onde de la radiation lumineuse reçue. Par exemple, un vert "pur" est une radiation monochromatique de longueur d'onde 530 nm, tandis que la lumière blanche est un spectre continu contenant toutes les longueurs d'onde du domaine du visible.

La théorie trichromatique de Young-Helmotz (1801) [11][12], énonce qu'il suffit de trois couleurs de base pour décrire la presque totalité des aspects lumineux. Dans le cas du mélange additif aussi appelé synthèse additive (voir Fig. 2.2), on peut reconstituer toute couleur par une combinaison de

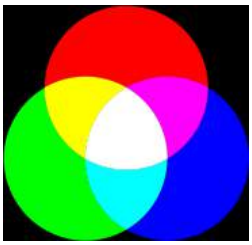


FIG. 2.2 – Synthèse additive des couleurs.

rouge, de vert et de bleu, que l'on appelle les couleurs primaires. Ces couleurs correspondent en fait aux trois longueurs d'ondes auxquelles répondent les trois types de cônes de l'oeil humain.

2.2.1 La vision humaine

L'oeil est l'organe de base de la vision humaine. Il comporte un ensemble d'éléments destinés à recevoir le rayonnement incident, former l'image des objets perçus et traiter les informations recueillies (voir Fig. 2.3).

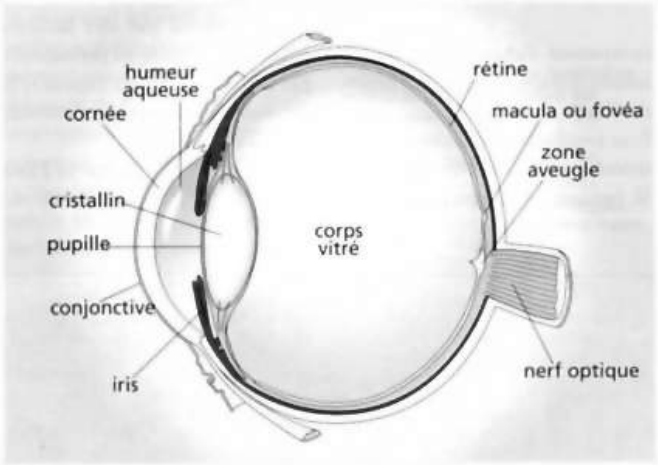


FIG. 2.3 – L'oeil humain. Image originale de [13][14].

Les principaux éléments constitutifs de l'oeil humain sont :

- L'iris : il fonctionne comme un diaphragme en dosant la quantité de lumière qui pénètre dans l'oeil. Son ouverture centrale est la pupille.
- Le cristallin : il fonctionne comme une lentille à focale variable, grâce à sa capacité de modifier sa courbure (phénomène d'accommodation).
- La rétine : c'est sur elle que se forment les images provenant de l'extérieur. La rétine contient deux types de cellules photosensibles : les cônes et les bâtonnets.
- La macula est la zone centrale de la rétine caractérisée par une concentration maximale de cônes. Elle contient en son centre une petite dépression, la fovéa qui est la zone d'acuité maximum de l'oeil. Au pourtour de la macula, les cônes se font plus rares et laissent la place aux bâtonnets qui se retrouvent seuls à la périphérie de la rétine.

Le fonctionnement de l'oeil humain : les rayons lumineux captés par la pupille et réfractés par la cornée et le cristallin forment une image sur la rétine. La rétine transforme ensuite l'énergie de la lumière en message nerveux. La couche supérieure de la rétine est composée exclusivement de photorécepteurs : les cônes et les bâtonnets :

- Les bâtonnets sont responsables de la vision nocturne (vision scotopique) et possèdent un maximum de sensibilité vers 510 nm. Leur sensibilité est liée à un colorant qui blanchit à la lumière du jour, expliquant par là leur insensibilité la journée. Les bâtonnets ne fournissent qu'une réponse photométrique[1] et ne permettent donc pas de déterminer les couleurs.
- Les cônes fournissent une réponse photométrique et chromatique[2], grâce à des pigments dont les maximums d'absorption se situent dans le bleu (450nm), le vert (540nm) ou le rouge (580nm), dont les proportions respectives sont 10%, 60%, et 30% environ. Une lumière, à n'importe quelle longueur d'onde du spectre visible, excite donc plus ou moins les cellules de ces 3 types.

L'oeil ne présente pas la même sensibilité pour toutes les longueurs d'onde. Une source de lumière située vers 660 nm doit être environ 10 fois plus lumineuse qu'une source de 560 nm pour être perçue avec la même intensité. Cette valeur n'est bien sur qu'une moyenne, chaque individu possède sa propre sensibilité chromatique. On peut voir sur la figure 2.4 les courbes de sensibilité normalisées des photorécepteurs de l'oeil humain.

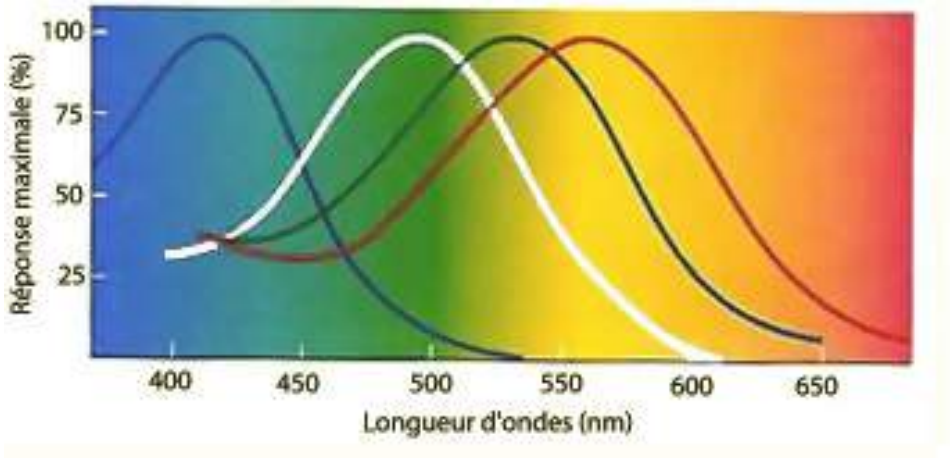
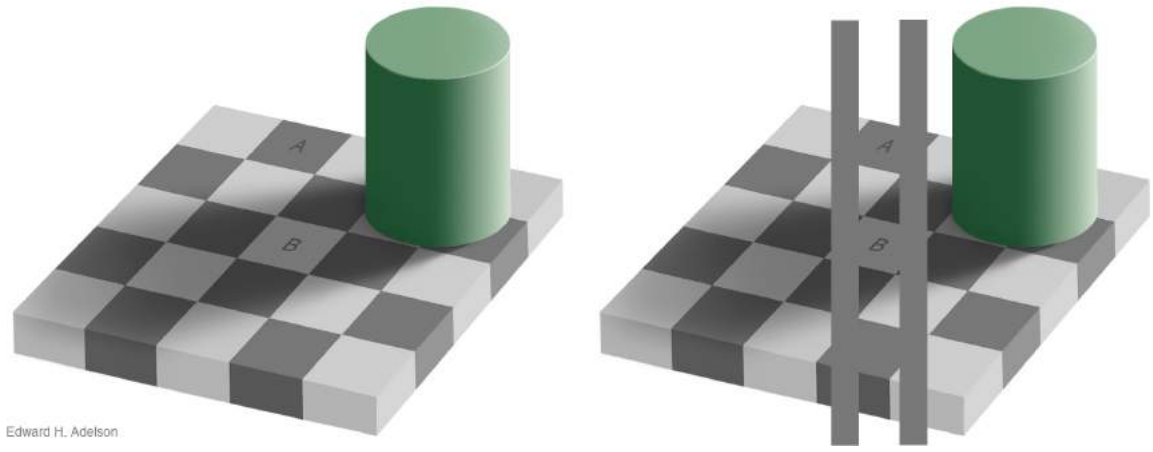


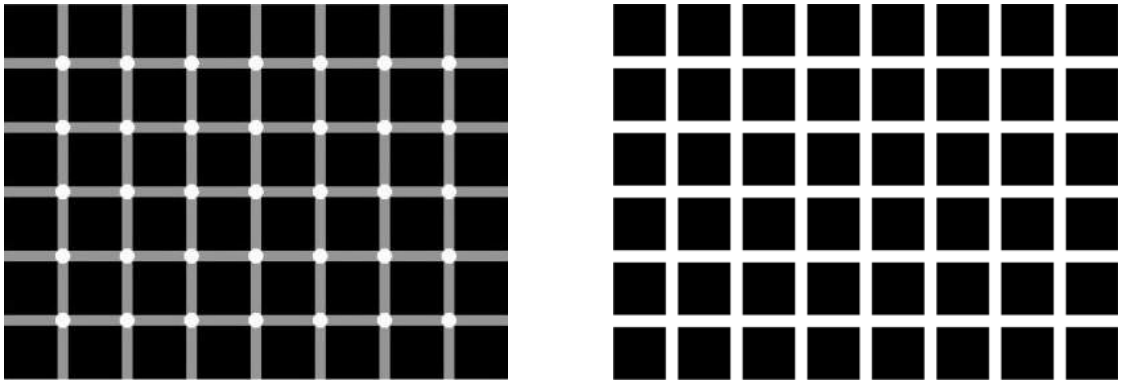
FIG. 2.4 – Sensibilité spectrale normalisée des photorécepteurs de l'oeil humain : bâtonnets (blanc), cônes bleu (bleu), cônes vert (vert) et cônes rouge (rouge). Source `http://lecerveau.mcgill.ca`.

La perception de la couleur de chaque être humain dépend, d'une part du signal couleur parvenant à son cortex visuel (aspect physique et physiologique) et, d'autre part de la façon dont ce signal va être interprété. Les mécanismes neurophysiologiques liés à cette interprétation sont relativement complexes et encore mal connus. Ce que nous savons c'est que notre interprétation d'un signal couleur dépend aussi d'aspects psychologiques et plus précisément de notre connaissance *a priori*, de notre environnement ainsi que de l'apprentissage que nous avons reçu des couleurs que nous percevons. En

[1] Relatif à la mesure des intensités lumineuses.
[2] Relatif à la couleur.



(a)



(b) (c)

FIG. 2.5 – Illusions d'optiques (a) "L'échiquier d'Adelson" (E. H. Adelson, 1995). Les niveaux de gris de la case A et de la case B sont rigoureusement identiques pourtant nous percevons la case B plus claire (b)(c)"La grille d'Hermann" (L. Hermann, 1870). Sur le première image on a l'impression de voir des points noirs alors qu'ils sont tous blancs, sur la deuxième on a l'impression de voir des points gris entre les carrés noirs.

effet le cerveau interprète et corrige les informations qu'il reçoit de l'oeil. Le contexte de l'observation joue aussi un rôle essentiel sur la perception de la couleur (voir Fig. 2.5(a)).

Nous venons de voir le mécanisme général de la perception de la couleur, tout ce processus est très complexe et demeure par conséquent difficilement reproductible par des moyens artificiels. Pour plus de détails on pourra se reporter [15][16]). Malgré cela, l'avantage des systèmes automatiques d'analyse de la couleur est la reproductibilité et l'objectivité des mesures. En effet, chaque individu étant unique, son interprétation de la couleur est également unique.

2.2.2 La vision artificielle

En vision artificielle, l'acquisition d'une image couleur peut se faire avec plusieurs types de caméras : caméras analogiques ou numériques, caméra linéaires ou matricielles... Les capteurs d'image

utilisés en analyse d'image couleur, sont généralement des caméras possédant un élément sensible à la lumière reçue. Nous allons dans cette partie nous attacher aux capteurs standards pour l'acquisition d'images numériques couleur.

Les principaux capteurs analogiques utilisés aujourd'hui pour l'acquisition d'images couleur sont les caméras CCD couleur, pour lesquelles on distingue deux types de technologies : mono-CCD et tri-CCD. Nous aborderons également les capteurs CMOS très utilisés notamment sur les appareils à bas coût. Pour plus de détails sur ces capteurs, on pourra se reporter [17][18][19].

- **Les capteurs CCD** (Dispositifs à Couplage de Charge) captent la lumière sur les petits photosites situés à leur surface, le plus souvent ils sont organisés en matrice. Les capteurs CCD tirent leur nom de la manière dont le nombre de charges est lu après une exposition à la lumière. La photodiode est l'élément optique sensible à la lumière, c'est elle qui capture la lumière transitant par l'objectif et la transforme en impulsion électrique. Ces impulsions électriques sont alors envoyées vers un convertisseur analogique/numérique à l'intérieur duquel sont appliqués à des algorithmes pour numériser l'image.
- **Les capteurs CMOS** (Complementary Metal Oxide Semiconductors) sont des détecteurs semiconducteurs à oxyde de métal complémentaire. Ce sont de minuscules circuits et dispositifs gravés sur des puces de silicium. Chaque cellule élémentaire est composée d'une photodiode et d'une logique d'amplification puis d'obturation. Tout comme les capteur CDD, ils transforment la lumière capturée en impulsion électrique. La fabrication de capteurs d'image CMOS se fait selon le même procédé que pour les puces d'ordinateur.

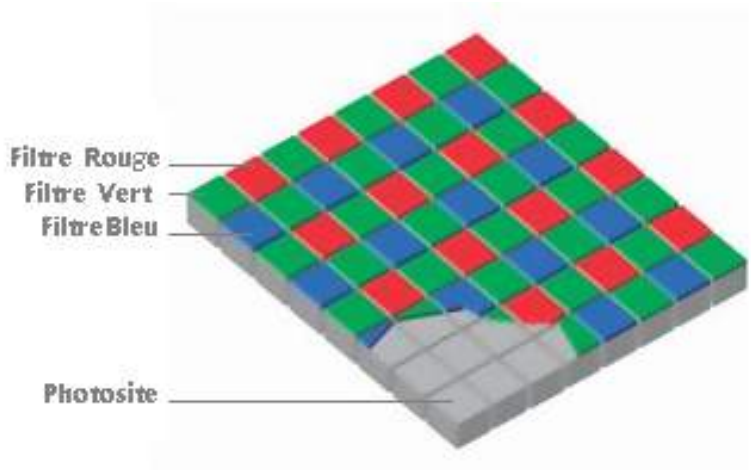


FIG. 2.6 – Matrice de photosites surmontée de la mosaïque Bayer.

Les caméras couleur mono-CCD sont, comme leur nom l'indique, munies d'un seul capteur CCD. Les composantes couleur de la scène sont obtenues en plaçant une mosaïque de filtres colorés, de telle sorte que chaque cellule du capteur CCD ne perçoive qu'une des trois composantes, généralement rouge, verte et bleue. Le filtre le plus utilisé dans les technologies actuelles est le filtre de Bayer, illustré par la figure 2.6. Dans l'image acquise, un pixel est donc représenté par quatre photosites (2 verts, 1 bleu et 1 rouge), il y a plus de vert pour tenir compte du fait que l'oeil humain est plus sensible au vert

qu'aux autres couleurs. Dans le cas des capteurs d'images monochromes la lumière n'est pas filtrée et chaque cellule reçoit le spectre complet. Les limitations du capteur couleur mono-CCD sont donc liées au fait qu'il faille au moins trois cellules pour obtenir l'information couleur complète, d'où une perte de résolution. De plus, les trois cellules utilisées pour définir la couleur d'un point ne sont pas localisées au même endroit, ce qui est à l'origine d'aberrations chromatiques. C'est de cette limitation que sont nées les caméras tri-CCD. En effet, celles-ci sont équipées d'un dispositif à base de prisme, comprenant 3 capteurs CCD. Ainsi, la résolution de l'image couleur n'est pas dégradée par un système de mosaïque et donc chaque point de l'image correspond à 3 cellules CCD.

Comme pour les capteurs CCD, les caméras CMOS sont souvent organisées en lignes et en colonnes et récupèrent l'information de couleur à l'aide d'un filtre. Les avantages de la technologie CMOS par rapport au CCD sont les coûts de fabrication moins importants, la faible consommation d'énergie, et une plus grande rapidité pour obtenir l'image. Par contre, sur un CMOS la sensibilité des photosites est moins grande que celle d'un capteur CCD. On trouve généralement les capteurs CMOS sur les appareils d'entrée de gamme, notamment les caméras USB type webcam (voir annexe C).

Comme on vient de le voir, les technologies CCD et CMOS sont les plus utilisées, cependant on trouve désormais des technologies innovantes comme le Super CCD et le Foveon X3 (CMOS) mais ces technologies ne seront pas présentées ici. L'ensemble des images réelles présentées dans ce document sont des images numériques couleur et monochrome acquises avec l'un des deux types de capteur présentés.

2.3 Les espaces de représentation des couleurs

Nous venons de voir deux capteurs couleurs qui sont l'oeil et la caméra. Dans les deux cas, les images sont codées en RGB (rouge, vert, bleu), il existe cependant plusieurs autres espaces couleurs qui seront plusieurs fois utilisés par la suite.

2.3.1 Les modèles de référence de la CIE : CIE RGB et CIE XYZ

La Commission internationale de l'éclairage (CIE) (en anglais International Commission on Illumination) est une organisation internationale dédiée à la lumière, l'éclairage, la couleur, les espaces de couleur. Cette commission a été créée pour caractériser rationnellement les couleurs telles que les voit le cerveau humain. Pour ce faire elle a mené des expériences systématiques de comparaison de couleurs par de nombreux observateurs de manière à définir un observateur moyen.

La CIE a caractérisé de manière précise le système naturel RGB qui s'appuie sur les couleurs correspondant aux maximums de réponse de chacun des types de cônes. Ces trois couleurs appelées couleurs primaires et notées Rc, Gc et Bc (de longueurs d'onde respectives 700,0 nm, 546,1 nm et 435,8 nm) permettent de former l'ensemble des composantes trichromatiques de tous les stimuli monochromatiques du spectre visible. Les trois courbes suivantes (Fig. 2.7) permettent de connaître dans quelles proportions il faut ajouter les primaires. Par exemple, un stimulus monochromatique défini par une longueur d'onde de 569 nm sera caractérisé par les composantes 0.17, 0.17 et 0 et un stimulus de longueur d'onde 500 nm par -0.07, 0.07 et 0.07.

Deux couleurs peuvent posséder le même caractère chromatique ou chrominance, mais avoir des composantes trichromatiques différentes à cause de leur luminance. Afin d'obtenir des composantes

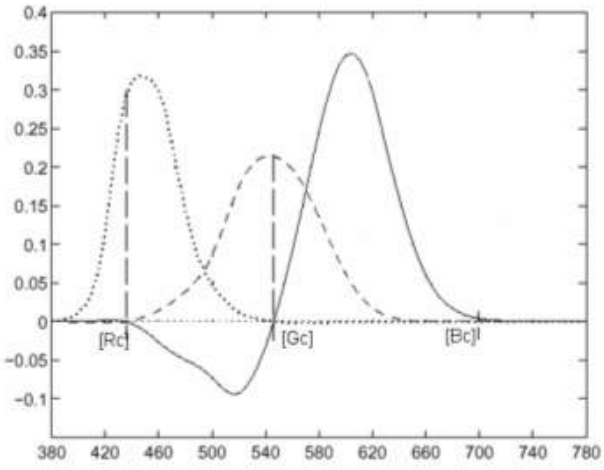
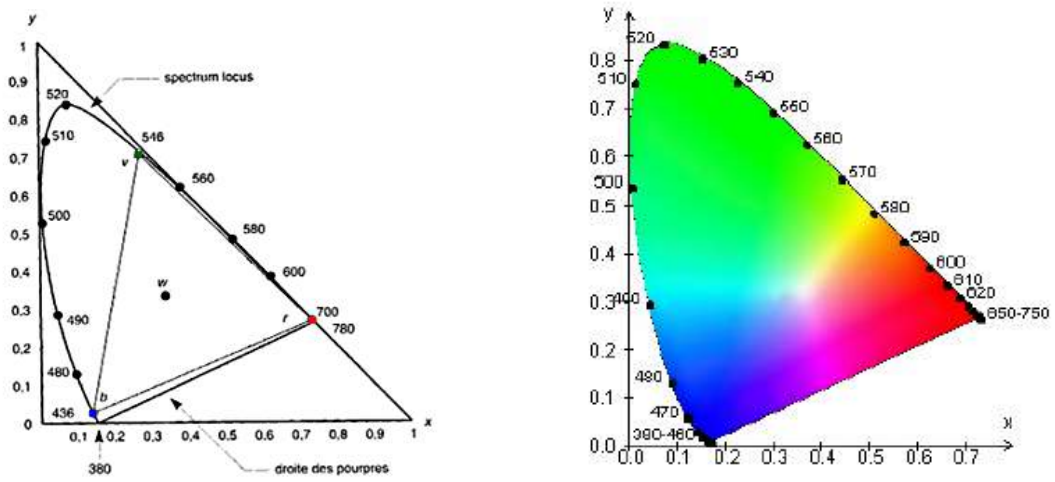


FIG. 2.7 – Fonctions colorimétriques Rc(λ), Gc(λ), Bc(λ) standardisées CIE. Source [20].

qui ne tiennent compte que de la chrominance, on peut normaliser les valeurs des composantes trichromatiques par rapport à la luminance. Ceci est réalisé en divisant chaque composante trichromatique par la somme des trois. Les composantes ainsi obtenues sont appelées les composantes normalisées. L'utilisation des composantes normalisées permet ensuite d'utiliser des coordonnées réduites c'est-à-dire de définir une couleur par seulement deux composantes puisque que la troisième peut être déduite des deux autres.

L'espace colorimétrique CIE RGB est construit à partir des données contenues dans le diagramme des composantes trichromatiques (Fig. 2.7). Il permet de définir le gamut[3]. En utilisant la méthode de normalisation, on peut représenter l'ensemble des couleurs pures observables par un oeil humain dans un système à deux dimensions appelé diagramme de chromaticité xy (voir Fig. 2.8(a)).
Dans ce diagramme, on s'aperçoit que beaucoup de couleurs du spectre visible ne sont donc pas réalisables par synthèse additive avec les primaires CIE RGB choisies (ni avec n'importe quel autre système de primaires issues de sources monochromatiques réelles). Ce sont toutes les couleurs qui ne sont pas dans le triangle à l'intérieur de la figure 2.8(a).
Il n'existe pas de base trichromatique de couleurs visibles qui permette de couvrir l'ensemble des couleurs visibles, il faut avoir recours à un espace dont les primaires sont virtuelles pour couvrir l'ensemble des couleurs visibles plus quelques unes sans réalité physique. Cet espace est l'espace couleur CIE XYZ défini par les primaires virtuelles. C'est un espace dérivé de l'espace RGB, il permet de corriger certains défauts, notamment les triplets décrivant chaque couleur en fonction des primaires ont tous des valeurs positives pour les spectres visibles et la fonction Y représente approximativement la sensibilité de l'oeil humain à la luminosité (composante de luminance). Sur le diagramme de chromaticité xy (Fig. 2.8(b)) on peut voir le gamut de l'oeil autrement dit l'ensemble des couleurs perceptibles par l'oeil.

[3] le gamut représente l'étendue d'un espace de couleur.

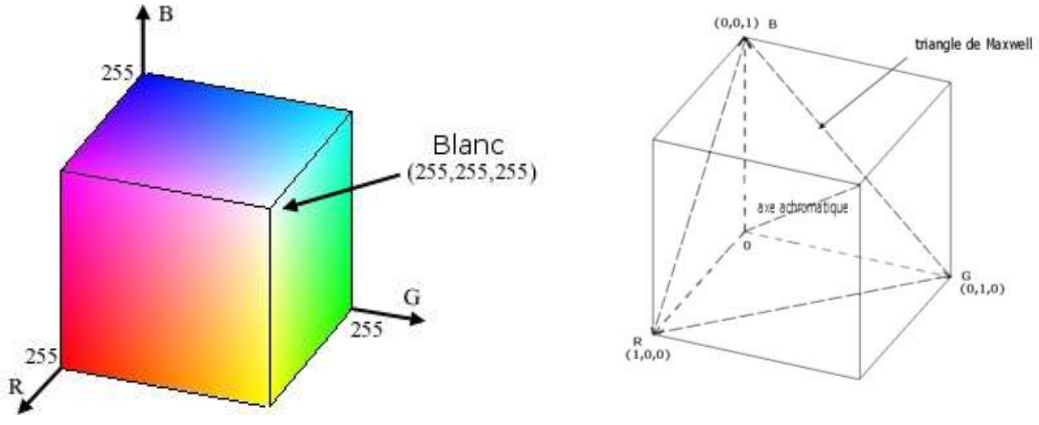


(a) (b)

FIG. 2.8 – Diagramme de chromaticité présentant les espaces CIE RGB et CIE XYZ. Source [20].

2.3.2 Les modèles dérivés d'un système matériel : RGB et YCbCr

L'espace RGB : Pour représenter l'espace RGB, on va faire correspondre aux trois primaires Rc, Gc et Bc, trois vecteurs directeurs qui forment le repère d'un espace vectoriel. Dans cet espace, chaque couleur est représentée par un point qui définit un vecteur couleur. Les coordonnées de ce vecteur sont les composantes trichromatiques Rc, Gc et Bc. Certains de ces points ont des coordonnées négatives puisqu'ils correspondent à des stimuli de couleur que l'on ne peut obtenir par synthèse additive. Les points correspondant à des stimuli de couleur dont les composantes trichromatiques sont positives sont contenus dans un cube, connu sous le nom de cube des couleurs (voir Fig. 2.9(a)).



(a) (b)

FIG. 2.9 – (a) Cube des couleurs représentant l'espace RGB, (b) le triangle de Maxwell représentant l'espace RGB normalisé. Source [20].

L'origine de l'espace vectoriel correspond au noir et le blanc de référence est défini par le mélange

unitaire des trois primaires. La droite passant par le noir de coordonnées (0, 0, 0) et le blanc de coordonnées (1, 1, 1) est appelée axe achromatique. Les points de cette droite d'équation $R = G = B$, représentent des niveaux de gris allant du noir au blanc.

L'espace RGB est l'espace natif. Son utilisation importante s'explique par la dépendance aux matériels : carte d'acquisition, cartes vidéos, caméras, écrans... qui effectuent leurs échanges d'informations uniquement en utilisant des triplets d'entier (R, G, B) codés sur 8 bits (entre 0 et 255). Il est important de noter que l'espace RGB est un standard mais qu'il n'est pas directement reliées à la perception visuelle.

L'espace RGBN : En divisant chaque composante trichromatique par la somme des trois, on obtient les composantes normalisées permettant de s'affranchir de l'information de luminance. La transformation définie par cette normalisation correspond à la projection d'une couleur sur le plan normal à l'axe achromatique, plan d'équation $R + G + B = 1$. Les intersections de ce plan avec le cube des couleurs forment un triangle équilatéral dont les sommets sont les trois primaires. Ce triangle appelé triangle de Maxwell, ou triangle des couleurs représente l'espace RGB normalisé (Fig. 2.9(a)).

L'espace YCbCr : Un autre espace permettant de séparer l'information de luminance et de chrominance est l'espace YCbCr. C'est un espace colorimétrique issu des problèmes de transmission hertzienne. En effet au début, la diffusion télévisuelle se faisait en noir et blanc et elle ne nécessitait que la composante de luminance. Avec l'évolution de la technologie, la télévision s'est vue dotée de la couleur et par conséquent de récepteurs couleur. Néanmoins, il était indispensable de rester compatible avec l'ancien système. Ainsi, les signaux émis par les chaînes de télévision ont séparé l'information de luminance de celle de chrominance. Ce passage s'est effectué en transformant linéairement les composantes trichromatiques RGB. Y représente l'information de luminance ou noir et blanc (qui correspond à la composante Y de l'espace XYZ) et Cb, Cr les informations de chrominance, respectivement le bleu moins la luminance et le rouge moins la luminance. Le récepteur peut ainsi recréer le vert et reproduire une image couleur. Les équations permettant le passage vers YCbCr sont les suivantes :

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{pmatrix} . \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

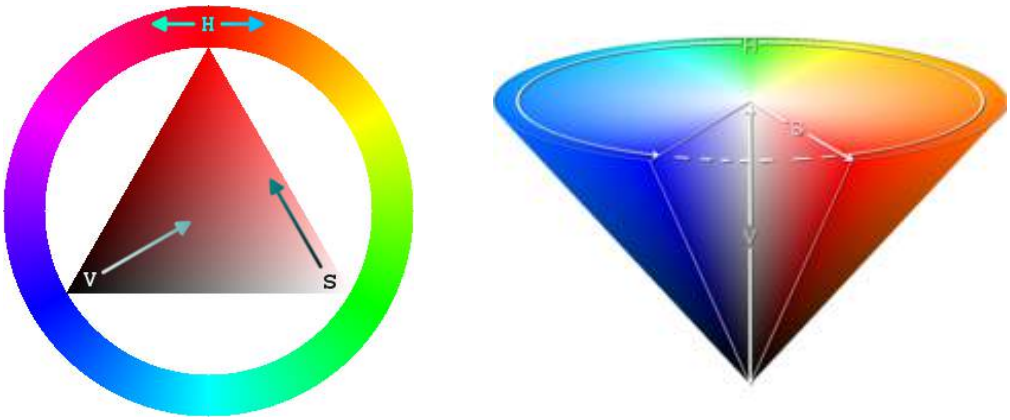
2.3.3 Les modèles perceptuels : HSL et HSV

L'espace HSL : Le modèle HSL (Hue, Saturation, Luminance, ou en français TSL pour teinte, saturation , luminance), s'appuie sur les travaux du peintre A. H. Munsell. C'est un modèle de représentation dit "naturel", c'est-à-dire proche de la perception physiologique de la couleur par l'oeil humain. En effet, le modèle RGB aussi adapté soit-il pour la représentation informatique de la couleur ou bien l'affichage sur les périphériques de sortie, ne permet pas de sélectionner facilement une couleur.

Une sensation visuelle de couleur peut être qualifiée par trois paramètres qui sont la teinte, la saturation et la luminosité, c'est sur ces trois paramètres que l'on va sélectionner une couleur dans l'espace HSL (voir Fig. 2.10(a)).

- la teinte correspond à la perception de la couleur (nuance de couleur ou couleur pure dominante). Elle est mesurée par un angle autour de la roue chromatique.
- la saturation décrit la pureté de la couleur, c'est-à-dire son caractère vif ou terne. Elle est mesurée depuis la couleur la plus neutre au centre de la roue vers les couleurs les plus vives en allant vers les bords.
- la luminosité indique la quantité de lumière de la couleur ou du flux lumineux. Elle est mesurée entre le noir (pas de lumière) et le blanc (lumière maximale).

L'espace HSV Un deuxième modèle beaucoup plus utilisé est le modèle HSV (Hue, Saturation, Value). C'est un dérivé du modèle HSL où on a abandonné la notion de luminosité pour se rapprocher du modèle RGB. Dans ce système la luminosité se fait en deux étapes sur deux axes perpendiculaires : du noir vers la couleur, puis de la couleur vers le blanc.



(a) (b)

FIG. 2.10 – L'Espace HSL et HSV. Source wikipédia.

L'ensemble des images réelles présentées dans les chapitres suivants seront toutes initialement codées dans l'espace RGB. Un complément sur les images numériques couleurs est disponible [21][22][8][5] et sur `http://www.profil-couleur.com/`.

2.4 Spécificités des images sous-marines

Après cette introduction sur la perception et l'acquisition des images puis sur le codage des couleurs, nous allons dans cette partie présenter quelques généralités sur les problèmes de perception visuelle en milieu sous-marin.

2.4.1 Propriétés optiques de la lumière dans la mer

Contrairement à la traversée de l'atmosphère, sous l'eau l'atténuation de la lumière est considérable, les infrarouges sont absorbés dans le premier mètre et peu de lumière atteint les 100 mètres.

Comme dans tous les milieux deux phénomènes contribuent à l'atténuation de la lumière : L'absorption et la diffusion.

- L'absorption de la lumière c'est l'absorption de l'énergie des photons par les constituants l'eau de mer. Les différentes concentrations de ces constituants déterminent la façon dont les différentes longueurs d'onde composant le spectre de la lumière sont absorbées. Du fait de la grande variété des constituants de l'eau de mer, l'analyse théorique de l'interaction entre la lumière et chacun de ces constituants est très difficile. Cela nous conduit à ne considérer que les constituants pour lesquels un effet sur le spectre d'absorption a été mis en évidence expérimentalement. Le premier constituant de l'eau de mer c'est l'eau. L'eau pure est pour la lumière un milieu très absorbant, en particulier pour les infrarouges. Au second rang, du point de vue de l'absorption on trouve les substances organiques en suspension. Par contre, la plupart des substances composant le sel marin ont peu d'effet sur l'absorption de la lumière dans l'eau.
- La diffusion de la lumière est un phénomène physique que l'on rencontre lorsque des inhomogénéités optiques du milieu provoquent des changements aléatoires de la direction des rayons lumineux. Cela est dû par exemple à la présence de particules en suspension ou de particules constituées de molécules d'eau de densité différente du milieu environnant. Pour les grosses particules la diffusion est le résultat de la réflexion, réfraction et diffraction. Lorsque les dimensions des particules sont inférieures à la longueur d'onde considérée, comme c'est fréquemment le cas pour l'eau de mer, l'optique géométrique ne suffit plus à expliquer la diffusion et il faut faire appel à la théorie des ondes électromagnétiques [23].

L'atténuation de la lumière sous l'eau est essentiellement due à l'absorption, la diffusion ayant une moindre importance. Cette atténuation est fonction de la longueur d'onde et des caractéristiques physiques et chimiques de l'eau de mer. Cette propriété optique apparente de l'eau de mer permet de la caractériser et donne des informations sur la densité, les concentrations des matières en suspension et la productivité biologique.

La couleur de la mer : Elle s'étend du bleu au vert. Le bleu caractérise plutôt les mers tropicales et équatoriales, lorsqu'il y a peu de production biologique. A des latitudes plus grandes, la mer est bleu-vert voire verte dans les régions polaires. Les eaux côtières sont en général verdâtres. Dans les régions où la productivité biologique est importante, la chlorophylle contenue dans le phytoplancton absorbe la composante bleue et la lumière se décale vers le vert.

2.4.2 Problèmes spécifiques des images sous-marines

En plus des aberrations lenticulaires et du bruit électronique de la caméra, l'obstacle majeur dans le traitement des images sous-marines résulte de l'atténuation de la lumière dans l'eau. Cette atténuation limite la visibilité à une distance d'environ vingt mètres en eau claire et à moins de cinq mètres en eau turbide. Le processus d'atténuation dépend des phénomènes d'absorption et de diffusion. Ces deux phénomènes étant liés eux-mêmes, aux différents constituants de l'eau, aux différentes matières organiques dissoutes et aux particules en suspension.
De ce fait, l'imagerie sous-marines doit faire face à plusieurs problèmes.
D'abord, l'atténuation rapide de la lumière impose l'installation d'un éclairage artificiel qui induit

FIG. 2.11 – Exemple d'images sous-marines (à noter que les trois images du bas sont extraites des vidéos TOPVISION).

généralement un éclairage non uniforme, c'est-à-dire un effet de spot caractérisé par un éclairage plus important au centre de l'image que sur les bords. La distance variable de la caméra à la scène produit aussi une dominante de couleur en général bleue ou verte car les longueurs d'ondes correspondant aux rouges sont très rapidement absorbés. Ensuite le phénomène de diffusion de la lumière provoque plusieurs effets indésirables (voir Fig. 2.12).
D'une part la composante de diffusion directe provoque un flou dans l'image ce qui atténue les contrastes, et d'autre part la composante de rétrodiffusion est à l'origine d'un éclairage non uniforme. Enfin les particules en suspension très variable en type et en concentration augmentent encore les phénomènes d'absorption et de diffusion, ce qui dégrade les contours, modifie les couleurs et produit des artefacts lumineux. A cause de tous ces problèmes l'imagerie sous-marine impose de travailler sur des images très bruitées (voir Fig. 2.11).

Pour résumer les images sous marines souffrent des problèmes suivants :

- Visibilité limitée,
- Éclairage non uniforme [24],
- Contrastes faibles,
- Couleurs atténués à dominante bleutée,
- Bruit important.

Plusieurs méthodes de correction "globales" des problèmes précédemment cités existent, on peut notamment citer [25][26].

Une dégradation synthétique provoquée par milieu sous-marin : Pour certains tests sur des algorithmes développés (notamment pour vérifier la robustesse du prétraitement et de la segmentation) une dégradation synthétique avec différents niveaux de sévérité a été ajoutée aux images sous-marines. Cette dégradation synthétique additionnelle a été générée suivant les descriptions de Arredondo [27].

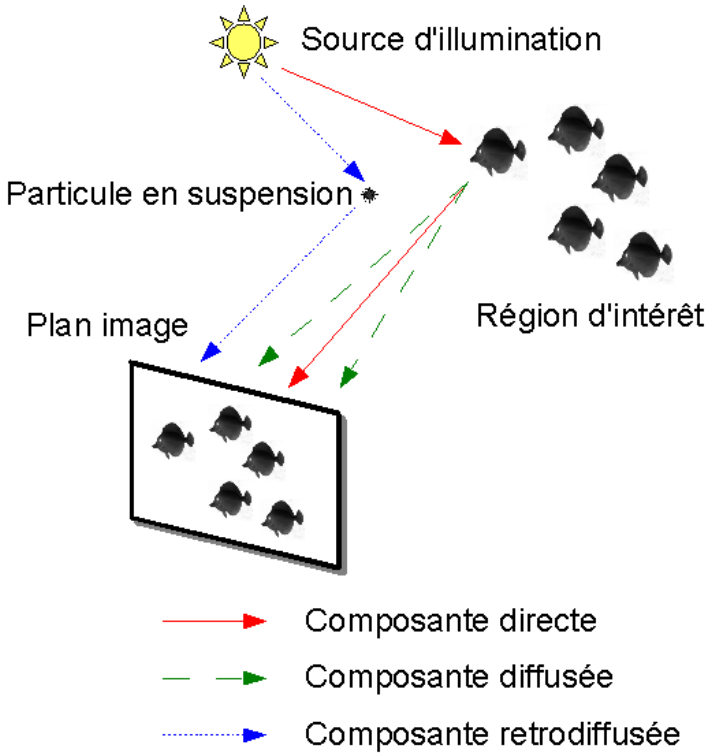


FIG. 2.12 – Mise en évidence des problèmes de diffusion de la lumière dans la mer. On distingue la composante d'éclairage directe, la composante de diffusion ainsi que la composante de rétrodiffusion.

Elle comporte notamment l'application des défauts suivants :

- un éclairage non uniforme par multiplication avec une illumination gaussienne,
- un effet de flou dû à la diffusion modélisée par une convolution [28],
- l'ajout d'un bruit blanc gaussien,
- et une réduction de la gamme de couleur disponibles.

Caractéristiques de prises de vues et problèmes spécifiques des séquences TOPVISION : Une grande partie des images que nous allons traiter ont été extraites de plusieurs séquences vidéos enregistrées au cours de plusieurs plongées dans le cadre du projet TOPVISION. Pour cette raison notre base d'images est très hétérogène notamment en ce qui concerne les couleurs et les conditions d'éclairage.

Nous avons pu observer les problèmes suivants :

- Un effet de moiré ou un effet de bloc. Deux problèmes parasites caractérisés pour le premier par un effet de tramage oblique dans les images, et pour l'autre par un effet de pixellisation (voir Fig. 2.13),
- Un flou : dû à la dynamique du véhicule instable dans le courant,
- Des variations d'éclairage : jour/nuit, naturel/artificiel,
- Deux types de caméra : couleur ou monochrome basse luminance.

Véhicule	*PAP*	SEA TWIN	ACROV
Capteur	Monochrome CCD 1/2	Couleur CCD 1/2	Couleur CDD 1/3
Taille	752 x 582	500 x 582	752 x 582
Focale	6mm	x	2,9mm
Sensibilité	0,05 lux	0,1 lux	0,1 lux
Enregistrement	Numérique	Analogique	Numérique

TAB. 2.1 – Caractéristiques d'acquisition des vidéos TOPVISION.

FIG. 2.13 – Les problèmes spécifiques TOPVISION.

Le tableau 2.1 regroupe quelques caractéristiques d'acquisition des vidéos TOPVISION dont nous disposions. Les informations figurant dans ce tableau ainsi que les images TOPVISION présentées dans ce rapport sont la propriété de l'État Français. Elles ont été fournies par le GESMA (Groupe d'Études Sous-Marines de l'Atlantique) dans le cadre du projet TOPVISION coordonné par Thales Underwater Systems SAS. Ce projet est rattaché au programme Techno-Vision lancé par le Ministère de la Recherche français et le Ministère de la Défense français.

2.4.3 Particularités des prises de vues sous-marines

Comme on a pu le voir dans la partie précédente l'eau agit comme un filtre puissant pour la lumière naturelle comme artificielle. L'atténuation de la lumière dépend principalement de l'épaisseur d'eau entre l'éclairage et le sujet, et entre le sujet et l'optique de prise de vue. Chaque longueurs d'onde est absorbée progressivement, commençant par les infrarouges et les ultraviolets, le rouge, le jaune, le vert et enfin le bleu. Du fait de cette disparition sélective, en descendant en profondeur même les sujets le plus multicolores paraissent souvent peu colorés alors qu'en réalité la faune et la flore sous-marines présentent souvent des couleurs intenses.

Le matériel : éclairage, optique. La première nécessité pour continuer à prendre des photos alors que l'on descend un peu en profondeur est donc d'apporter une source d'éclairage artificielle pour compenser le manque de lumière. L'usage des flashs intégrés aux appareils photo numériques est à proscrire puisque qu'il amplifie l'effet de rétrodiffusion et ajoute ainsi un effet de neige marine très néfaste à l'image. Pour les prises de vues sous-marines l'éclairage artificiel doit donc être exclusivement externe et astucieusement placé, on recommande un angle d'environ 45° entre l'axe optique et l'axe du flash avec une portée sur le sujet d'un à cinq mètres en fonction des conditions de visibilité.
L'éclairage artificiel externe peut être de deux types : flashs ou phares.

- Les flashs sont de loin les plus utilisés, en raison de leur plus forte puissance et de leur portée plus importante. Ils sont fixés à l'appareil par un bras articulé, et déclenchés par le flash intégré à l'appareil photographique. Aujourd'hui, ils sont plutôt électroniques et utilisés avec un éclairage type "lumière du jour" (température de couleur 5500K)[4] c'est-à-dire qu'ils émettent approximativement la même quantité d'énergie dans toutes les longueurs d'onde.
- Les phares sont en général plus encombrants et posent le problème de l'autonomie, ils sont utilisés par grande profondeur et surtout pour la vidéo sous-marine. Les plus communs sont les phares halogène (3200K) mais deux autres types sont très utilisés notamment pour leur autonomie, leur faible encombrement, et leur spectre type "lumière du jour" : les phares HID[5] et les phares à LED[6].

On peut noter qu'un éclairage artificiel à basse température de couleur (type halogène par exemple) permet de diminuer l'effet bleutée. Si l'on dispose de sources dont la température est plus élevée on peut utiliser des filtres correcteurs teinte chaude (orange par exemple) pour atténuer un peu la dominante bleu (ce filtrage à cependant pour conséquence de baisser encore l'intensité lumineuse).
A noter que d'autres types de filtres beaucoup plus chers à l'achat permettent aussi de corriger les réflections parasites de la lumière, ce sont les filtres polarisants, leur utilisation reste cependant très marginale. D'une manière générale, on préfère aujourd'hui un éclairage externe avec un spectre type "lumière du jour" suivi d'une correction des couleurs par traitement d'image.

Les conditions de prises de vues. En plus de cet appauvrissement de la palette colorimétrique dû à l'absorption, un autre phénomène produit un flou dans l'image et un affaiblissement des contrastes, c'est la diffusion. Lors des prises de vues sous-marines, il est primordial de travailler au plus prêt du sujet pour limiter les phénomènes perturbateurs. C'est pourquoi on privilégie en général des optiques large angle ou très large angle autorisant la proximité avec le sujet, tout en préservant une largeur de champ assez importante. L'optique grand angle est utilisée pour les photos dites "portrait rapproché" et le très grand angle pour les photos dites "d'ambiance".

Les réglages spécifiques des appareils photo numériques. En plus du matériel et des conditions de prise de vues spécifiques, en environnement sous-marin quelques réglages sont indispensables pour obtenir des photos nettes, bien exposées et avec de belles couleurs. Certains fabricants ont inclus dans leurs modes "scène" un mode spécial "prise de vue sous-marine" cependant lorsque ce mode n'existe pas, ces réglages sont à faire manuellement. Le réglage le plus important est la balance des blancs qui consiste à régler l'appareil pour adapter la colorimétrie aux conditions de luminosité et

[4] La température de couleur permet de déterminer la couleur d'une source de lumière. Elle se mesure en Kelvins.
[5] High Intensity Discharge
[6] Light Emitting Diode

éviter une dominante colorée. En d'autres termes rééquilibrer la lumière en restituant au mieux un blanc proche de sa valeur de référence. Généralement, on recommande le mode dit "ensoleillé" en eaux peu profondes (5 à 10 mètres) avec ou sans éclairage artificiel, et le mode dit "nuageux" pour les eaux plus profondes avec éclairage artificiel. On peut aussi appliquer quelques réglages spécifiques comme une saturation et un contraste plus élevés pour raviver les couleurs, un autofocus en mode "continu" pour que l'appareil puisse s'adapter aux mouvements constants de l'eau, et enfin un réglage de la distance de mise au point en mode "standard" pour les photos d'ambiance et en mode "macro" pour les photos très rapprochées. En cas de sous-exposition on peut même compenser un peu la lumière avec des paramètres de luminosité négatif.

Pour plus de détails, on renvoie le lecteur sur l'ouvrage très complet sur la photographie sous-marine de Guillen [29].

2.5 Conclusion

Nous avons exposé dans ce chapitre quelques généralités sur les images, sur le processus de vision ainsi que sur les particularités, des données à traiter et des prises de vues sous-marines. Nous avons notamment présenté le système d'acquisition numérique des images, les différents types de codage de l'information de couleur et les propriétés optiques de la lumière dans la mer (phénomènes d'absorption et de diffusion). Nous avons ensuite montré que les images sous-marines comportent des défauts très spécifiques et c'est d'ailleurs pourquoi les algorithmes standard de traitement d'image ou de vision par ordinateur dans le domaine aérien sont très difficilement exploitables. L'ensemble de ces problèmes peuvent cependant être atténués par un traitement matériel ou logiciel. Nous avons détaillé à la fin de ce chapitre quelques réglages matériel notamment au niveau de l'éclairage artificiel, de l'optique, et des conditions de prises de vues permettant d'atténuer les défauts des images en milieu sous-marin. Malgré cela, il est presque toujours indispensable de faire un prétraitement logiciel, excepté dans le cas où l'on utilise des capteurs spécifiquement développés (par exemple [26]) qui offrent une qualité d'image incomparable.
Pour conclure, il est clair que certains traitements matériel, quelques réglages particuliers ou quelques efforts sur la prise de vue permettent d'améliorer nettement la qualité des images. Dans notre cas, en raison de contraintes opérationnelles fortes, nous disposions de données, obtenues par des caméras "standard", analogiques ou numériques, dans des conditions de prises de vues inconnues, assez variables voire parfois médiocres, nous avons donc développé un prétraitement logiciel des images permettant d'améliorer leurs qualités avant d'effectuer les traitements de reconnaissance.

Chapitre 3

Reconnaissance d'objets sous-marins à l'aide du capteur vidéo

Dans ce chapitre nous allons présenter une étude des différentes méthodes de reconnaissance d'objets existantes en vision. Nous commencerons par évoquer les méthodes utilisées dans le domaine sous-marin puis nous élargirons à celles utilisées dans le domaine terrestre beaucoup plus étudié. Nous expliquerons à la suite de cette étude générale les méthodes que nous avons choisi de développer dans cette thèse, c'est-à-dire dans un contexte opérationnel et restreint à la vision monoculaire sans connaissances *a priori* sur les conditions de prises de vues. A noter que les paragraphes 3.1.2, 3.1.3 et 3.1.4 ont été inspirés de l'étude générale de Arnold-Bos [30].

3.1 La reconnaissance d'objets sous-marins

Aujourd'hui le moyen de sondage privilégié de l'environnement sous-marin est le capteur acoustique de type sonar. En effet il est utilisé depuis près d'un siècle[1], alors que le capteur optique n'est utilisé que depuis une dizaine d'années [31]. Son utilisation beaucoup plus tardive dans le milieu sous-marin est essentiellement due à la faible qualité des images (problèmes d'éclairage, de turbidité, de bruit important...) et à la complexité des traitements. Ces traitements complexes sont en plus limités par les contraintes des robots autonomes : stockage physique, énergie, capacité de calcul.

3.1.1 Du sonar à la caméra

Le sonar (SOund Navigation And Ranging) utilise les propriétés très favorables de la propagation du son dans l'eau pour détecter et localiser des objets [32] mais également pour mesurer la profondeur, cartographier le fond des océans ou encore pour l'aide à la navigation [33]. On entend par objets : des navires, des sous-marins ou des mines en ce qui concerne le domaine militaire, des bancs de poissons,

[1] on parle ici du sonar actif c'est-à-dire émettant une impulsion sonore et écoutant son écho. A l'inverse, le sonar passif écoute le son ambiant sans émettre.

	Avantages	Inconvénients
Sonar	Longue portée Traitements automatiques opérationnels	Faible résolution Identification difficile
Caméra	Haute résolution Facilité d'interprétation Faible coût	Très courte portée Complexité d'exploitation automatique Données volumineuses et de qualité variable

TAB. 3.1 – Le sonar et la caméra : Deux capteurs complémentaires pour la perception de l'environnement en milieu sous-marin.

des épaves, ou divers objets polluants ou perdus en mer en ce qui concerne le domaine civil (voir Fig. 3.1).

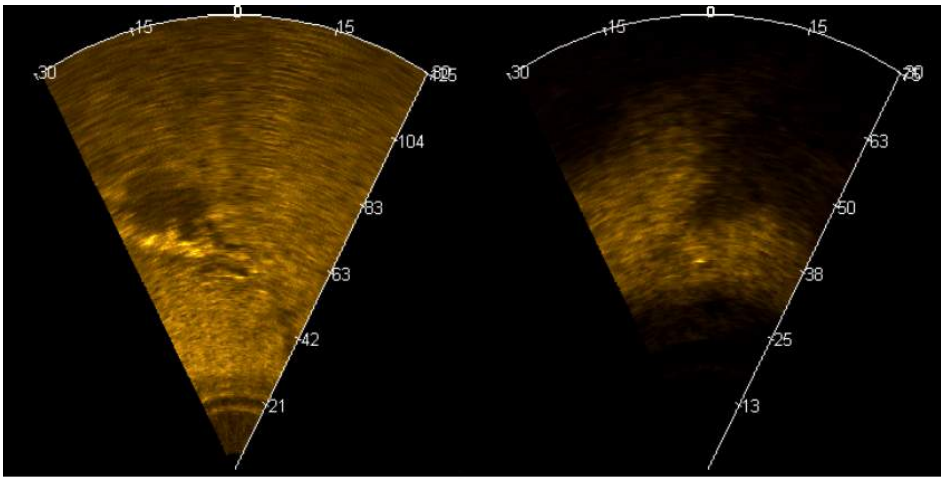
Dans le contexte de reconnaissance d'objets sous-marins, les capteurs acoustiques sont efficaces à grande distance pour la détection puis à moyenne distance pour la classification. Les capteurs optiques sont quant à eux efficaces à courte portée pour les phases d'approche, la reconnaissance d'objets et l'intervention. Les deux capteurs ont des propriétés singulières et répondent chacun à des problèmes très différents (voir Tab. 3.1). Par cette complémentarité et du fait que l'on dispose aujourd'hui des capacité informatiques nécessaires à leur exploitation, l'utilisation de la vidéo est de plus en plus courante. Les atouts et faiblesses des deux capteurs sont répertoriés dans le tableau 3.1.

3.1.2 Un exemple de contexte opérationnel

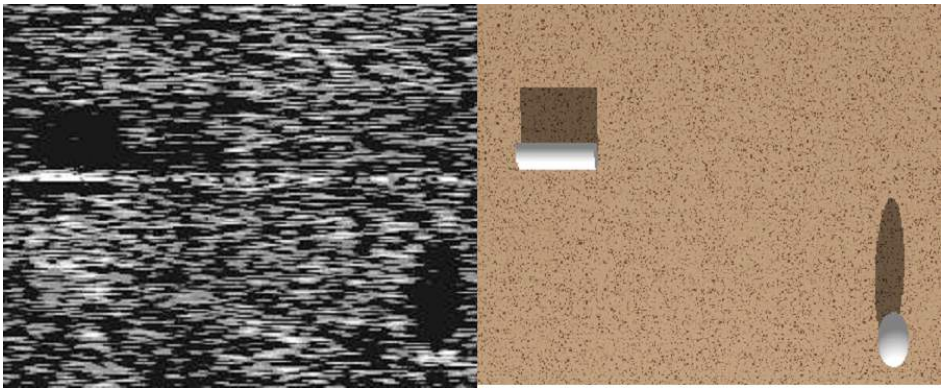
Pour donner un exemple d'utilisation de la reconnaissance d'objet prenons le cas d'une campagne de déminage ou de dépollution. Dans ce contexte, on envoie des navires pour quadriller la zone à explorer, les objets sont d'abord détectés par un sonar détecteur avec une portée de l'ordre de plusieurs centaines de mètres. Les objets détectés, une phase d'approche est amorcée dans le but de vérifier la détection puis reconnaître l'objet à l'aide d'un sonar classificateur ayant une meilleure résolution mais une portée plus faible. On peut voir sur la figure 3.2 les deux types de sonar. Après cette seconde phase, s'il y a eu reconnaissance, l'objet est détruit ou repêché le souvent grâce à l'intervention de plongeurs.
Ce scénario classique est rarement respecté car malgré la résolution plus fine du sonar classificateur, l'étape de reconnaissance est parfois impossible ou peu sûre, et c'est dans ce cas qu'intervient la caméra vidéo. Dans la pratique, un ROV[2] ou un AUV (muni d'un sonar et d'une caméra) est envoyé sur l'objet pour s'en approcher dans un premier temps puis pour l'identifier voire éventuellement pour intervenir ou préparer l'intervention.
Les objets potentiels à reconnaître sont les suivants :

- Dans le contexte de déminage on peut trouver quatre grands types de mine :
  - la mine Manta : de forme tronconique avec un diamètre de 98cm à la base et de 47cm au sommet pour une hauteur de 37cm,
  - la mine Rockan de forme très particulière que l'on peut simplifier par un polyèdre à 6 faces (4 faces rectangulaires et deux faces en forme de trapèze) 102 cm x 80 cm x 38,5 cm.

[2] en anglais Remote Operated Vehicle littéralement véhicule commandé à distance désigne un petit sous-marin téléguidé.



(a)



(b)

FIG. 3.1 – Exemple d'images sonar (a) Images obtenues par un sonar frontal de détection, à droite une épave et à gauche une mine (Source GESMA, référence DEVITOBS'06) (b) Image d'un sonar frontal et image de synthèse correspondante (Source TUS).

  - la mine cylindrique de longueur standard 2 mètres pour un diamètre de 50 cm.
  - la mine sphérique d'un diamètre de 70 cm à 1 m. (en général entre deux eaux rattachée à une ancre par un orin).
- Dans le contexte de dépollution, on peut trouver des objets manufacturés de formes très diverses :
  - polyèdre, parallélépipède, cube, cylindre, sphère, tore... pour les formes simples.

Tous ces objets manufacturés ont un point commun intéressant et assez simple pour notre problème : ils ont des formes régulières et font apparaître des droites, des cercles, des ellipses et souvent des parallélismes et des symétries. A noter que dans le contexte sous-marin les objets peuvent être partiellement enfouis notamment par un effet d'impact lors de la chute de l'objet ou par enfouissement naturel dans le sédiment dû au courant sous-marin par exemple.

Dans notre étude, nous nous attacherons donc à un contexte opérationnel comme celui-ci avec

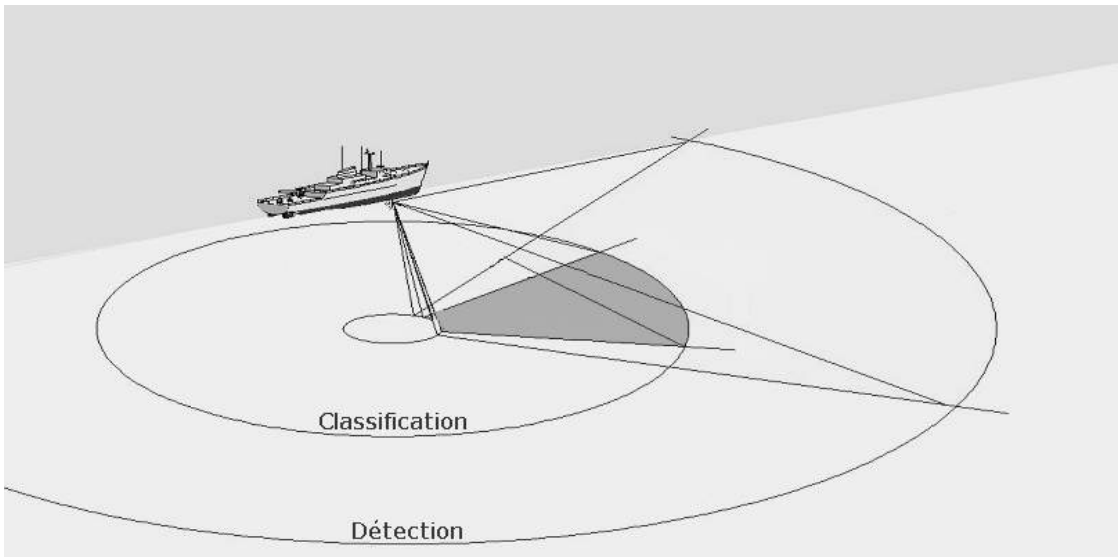


FIG. 3.2 – Les deux types de sonar : sonars détecteur et classificateur (Source TUS).

des objets répondant aux critères que nous avons évoqués. Puisque nous traitons le cas général, nous considérerons aussi les cas où nous disposons des données *a priori* suivantes :

- Caractéristiques des objets recherchés (modèle 3D, couleur, dimensions)
- Géoréférencement du robot (profondeur, position, attitude)
- Prises de vue (position et orientation, ou référence métrique donnée par deux pointeurs laser par exemple)

L'objectif final étant bien évidemment de limiter au maximum les informations *a priori* pour rester dans des contraintes opérationnelles réalistes.

3.1.3 Le capteur optique dans le contexte sous-marin

Les contraintes liées à l'utilisation du capteur optique dans le contexte sous-marin

- Faible portée visuelle ,
- Conditions d'éclairage variables (éclairage artificiel ou naturel),
- Turbidité plus ou moins importante (particules ou sédiments en suspension) parfois aggravée par le passage du robot près du fond,
- Couleurs atténuées (nuance bleutée plus ou moins prononcée dû au filtrage bleu-vert naturel causé par le milieu),
- Non stabilité des séquences (véhicule en mouvement, possibilité de courant sous-marin),
- Fond naturel inconnu, position des objets inconnue, et possibilité d'occultation et d'enfouissement,

A cela, il faut ajouter quelques contraintes liées à l'utilisation des robots autonomes :

- Espace de stockage limité,
- Puissance de calcul limitée.

Pour respecter le contexte opérationnel, on vise aussi l'implémentation d'algorithmes génériques qui permettent le cas échéant d'ajouter facilement des nouveaux objets à la base de reconnaissance,

et surtout des algorithmes rapides voire temps réel pour obtenir un comportement satisfaisant. On considérera d'ailleurs en particulier le cas de la vision monoculaire avec une caméra standard non calibrée, de caractéristiques inconnues et avec une orientation quelconque. Ce qui représente le cas le plus courant. Dans un souci de complétude de l'étude menée, nous discuterons cependant d'utilisations moins contraignantes notamment la possibilité de stéréovision, de différents types de caméra, de calibration... Pour résumer, les problèmes sont ceux liés à la robotique terrestre avec une dimension supplémentaire (l'altitude) et un milieu où la propagation de la lumière est un phénomène très complexe. On dispose donc d'images de faible qualité avec du bruit, des faibles contrastes, des couleurs atténuées et une portée visuelle très limitée.

Le choix du capteur optique à utiliser

Pour avoir une revue complète et détaillée des différents capteurs optiques utilisés actuellement en imagerie sous-marine on pourra se reporter à la publication de Jaffe [34]. On en distingue trois types :

- Les systèmes passifs : Les caméras conventionnelles en sont le premier exemple. Ces systèmes captent la lumière provenant d'une source extérieure au système imageur (lampe ou soleil par exemple). Les rayons transitent par un système de lentilles et sont récupérés sur le capteur optique plat. Les capteurs utilisés sont essentiellement CDD et CMOS conventionnels (voir chapitre 2) mais on utilise également des capteurs intensifiés par exemple SIT[3] et ISIT[4], ou encore ICCD[5], afin de gagner en sensibilité et diminuer les problèmes causés par l'atténuation. Ces capteurs intensifiés sont monochromes et plus onéreux que les caméras conventionnelles.
- Les systèmes semi-actifs : Ces systèmes sont beaucoup plus chers. L'éclairage standard est remplacé par un faisceau laser pulsé qui éclaire la scène (en anglais Range Gated). Connaissant la distance approximative de l'objet et sachant que le laser est synchronisé avec la caméra, on peut rejeter les rayons ayant eu un trajet trop court (c'est-à-dire rétrodiffusés par l'eau). L'image obtenue par ces systèmes est une image d'intensité correspondant à la distance supposée de l'objet [35].
- Les systèmes actifs : Ces systèmes utilisent aussi un faisceau laser mais tiennent cette fois explicitement compte du temps de parcours de la lumière entre le système et l'objet afin d'établir une carte de profondeur de la scène [36]. Ce faisceau laser est balayé sur la zone à couvrir.

Le choix de la configuration à utiliser

Dans le cas le plus courant où l'on utilise des caméras conventionnelles, on peut ensuite choisir d'utiliser une ou plusieurs caméras. En effet, avec une seule image prise par une caméra la composante de profondeur est perdue, l'image étant la projection dans un plan de la scène 3D. Le choix du nombre de caméra dépend donc de l'utilisation de la 3D pour reconnaître les objets présents dans la scène.
Dans le cas où l'on veut retrouver la composante de profondeur il faut disposer de deux images de la même scène prise d'un point de vue différent tout en connaissant parfaitement la transformation géométrique permettant de passer d'un point de vue à l'autre. On qualifie ce système de stéréoscopique (plus de détails dans [37]). On a donc le choix entre trois configurations possibles : les deux premières permettant de retrouver la composante de profondeur et donc de travailler en 3D, la dernière étant purement 2D :

[3] Silicon Intensifier Target
[4] Intensified IST
[5] Intensifier CDD

- Vision stéréo binoculaire. C'est le cas le plus simple pour retrouver la 3D. On utilise un banc rigide sur lequel sont placées deux caméras. On détermine la position relative entre les deux points de vue par une procédure de calibration (voir annexe A). Connaissant cette position, il reste à réaliser pour chaque couple d'images un appariement des pixels puis à reconstruire la scène 3D par une procédure d'inversion (voir annexe A). A noter que la reconstruction impose des images de bonne qualité pour pouvoir retrouver la structure tridimensionnelle. Dans le cadre sous-marin on peut trouver quelques exemples dans [38][39][40].
- Vision stéréo monoculaire. Ce système consiste à prendre deux vues d'une même scène avec la même caméra que l'on aura bougée entre les deux points de vue. En plus du problème de l'appariement des images qui est en commun avec la vision binoculaire s'ajoute le problème de l'estimation de la transformation géométrique entre les deux points de vue qui est cette fois inconnue. Pour retrouver la rotation et la translation, on peut appliquer différentes méthodes notamment celles utilisant des informations sur la scène ou sur le déplacement de la caméra. Dans le domaine sous-marin on peut citer par exemple [41][42].
- Vision monoculaire pure. Dans ce cadre, on suppose que l'on néglige la composante 3D. En effet les contraintes imposées par les systèmes stéréoscopiques sont importantes et dans de nombreux cas on dispose juste d'une caméra conventionnelle sans *a priori* sur la scène ou sur les mouvements de la caméra, ce qui nous empêche de retrouver la composante de profondeur. Dans ce cas on va donc reconnaître les objets à partir des images 2D. De nombreuses méthodes de reconnaissance utilisent ce type de vision (pour le milieu sous-marin voir [43][44][45][46][1]).

A noter que le choix de l'angle de vue est également important dans le cadre de la reconnaissance. En effet, en forçant un angle de vue on limite la complexité de problème. Dans le contexte sous-marin, on peut par exemple supposer une vue de dessus avec des objets posés sur un fond supposé plat ce qui simplifie grandement le nombre de formes sous lequel on peut voir l'objet.

3.1.4 L'existant dans le domaine de la vision sous-marine

Beaucoup de travaux ont été effectués dans le domaine de la reconnaissance d'objets en imagerie dans un milieu relativement contrôlé, c'est-à-dire un environnement uniforme, des positions bien définies et des objets avec une forme géométrique simple. Dans le milieu naturel et en particulier le milieu sous-marin, la reconnaissance de cibles est essentiellement faite à partir d'images sonar. Très peu de contributions existent dans le domaine de l'imagerie optique. On trouve cependant des contributions dans des domaines voisins tels que l'assemblage de photos du fond [47][48][49][50], la navigation des AUV [24][51][52][53], le suivi de pipelines, ou de structures sur le fond [54][43][44][45][46][55], l'identification de plancton ou de poissons [56][57]. On trouve aussi des contributions sur la détection de scènes d'intérêt dans des vidéos [58][38][59][60]. Enfin dans des domaines plus lointains comme dans le domaine terrestre, on peut citer quelques contributions intéressantes liées à notre problématique [61][62].

Le domaine étudié le plus proche est la détection d'objets manufacturés en vidéos sous-marines. Dans ce domaine, la thèse de Olmos [1][63] fait référence. En effet ce travail est à notre connaissance le seul travail réellement comparable. Olmos s'intéresse à la détection d'objets manufacturés dans des vidéos sous-marines dans des conditions un peu plus particulières, c'est-à-dire des objets posés sur le

fond avec une prise de vue proche de la verticale (conditions de survol) et disposant de quelques informations sur les conditions de prises de vues comme par exemple la focale ou la profondeur. Elle utilise un processus automatique de restauration par filtrage inverse, un système d'extraction de contours multirésolution avec une recherche automatique de l'échelle optimale, un algorithme d'extraction d'attributs basé sur l'attention visuelle et le groupement visuel et enfin une méthode de classification de paramètres aboutissant à la détection ou non d'un objet manufacturé. Les résultats obtenus sont intéressants, d'autant que c'est une méthode complète et automatique testée sur des données réelles. Cette chaîne de traitement globale n'a pas été re-implémentée mais cette méthode sera souvent notre base de comparaison dans le suite de ce manuscrit. On basera donc la comparaison sur les résultats présentés dans la thèse. Pour finir, ce qu'il faut retenir de la littérature dans le domaine de la reconnaissance d'objets, c'est qu'elle impose souvent des hypothèses sur l'environnement car elle est très sensible aux variations dans les conditions d'acquisition.

3.2 État de l'art général en reconnaissance d'objets par vision

3.2.1 La vision par ordinateur

> **Qu'est-ce que la vision ?** L'espace qui nous entoure a une structure tridimensionnelle (3D). Lorsque l'on demande à une personne de décrire ce qu'elle voit, elle n'éprouve aucune difficulté à nommer les objets qui l'entourent : téléphone, table, livre... Et pourtant l'information qui est réellement disponible sur la rétine de ses yeux n'est, ni plus ni moins, une collection de points (environ un million !). En chaque point ou pixel (picture element) il y a tout simplement une information qui donne une indication quant à la quantité de lumière et la couleur qui proviennent de l'espace environnant et qui ont été projetées à cet endroit de la rétine. Le téléphone, la table ou le livre n'existent pas sur la rétine. Guidé à la fois par l'information codée dans l'image (ou la rétine) et par ses propres connaissances, le processus visuel construit des percepts. Le téléphone ou le livre sont les réponses finales, résultant d'un processus d'interprétation qui fait partie intégrante du système de vision. De plus, il n'y a pas de correspondance terme à terme entre l'information sensorielle (la lumière et la couleur) et la réponse finale (des objets 3D). Le système de vision doit fournir les connaissances nécessaires afin de permettre une interprétation non ambiguë.
>
> Vision par ordinateur, Outils fondamentaux
> R. Horaud et O. Monga,
> Editions Hermès, 1995.

La vision humaine est extrêmement complexe et incroyablement performante. Ce court paragraphe tiré du livre de R. Horaud et O. Monga [37] suffit à le comprendre. Les systèmes de vision à développer ne cherchent donc pas à comprendre ou à reproduire la vision humaine mais à construire un modèle algorithmique qui, vu de l'extérieur, possède des propriétés semblables.

Dans ce rapport nous nous intéresserons à une des tâches principales du système de vision, la reconnaissance des objets qui composent une scène et ce, appliqué au contexte sous-marin [64][53][65]. Concrètement, la reconnaissance consiste à comparer ou mettre en correspondance des indices visuels

bi ou tridimensionnels extraits des images avec les indices des objets à reconnaître extraits des modèles (voir Fig 3.3). L'ensemble des attributs des modèles constitue la base de connaissance (2D ou 3D). La constitution de cette base peut et doit en général se faire avant utilisation de l'algorithme de reconnaissance. A partir de cette base souvent réduite on va pouvoir généraliser la reconnaissance en appliquant un algorithme de classification. Cet algorithme permet l'identification des objets dans les images malgré les changements de prises de vue, les éventuels occultations et les changements des conditions d'acquisition. Dans notre contexte de reconnaissance par vision, on s'intéressera essentiellement à la partie traitement d'image c'est-à-dire l'extraction d'attributs de l'image. La partie apprentissage/classification/décision n'a pas été développée puisqu'elle avait déjà été étudiée et implémentée par Olmos [1]. Pour notre étude, on se contentera donc d'une prise de décision simple et empirique sans apprentissage ni classification. Une des perspectives de ce travail sera bien évidemment de valider les méthodes développées en évaluant et comparant les résultats après l'étape finale de classification.

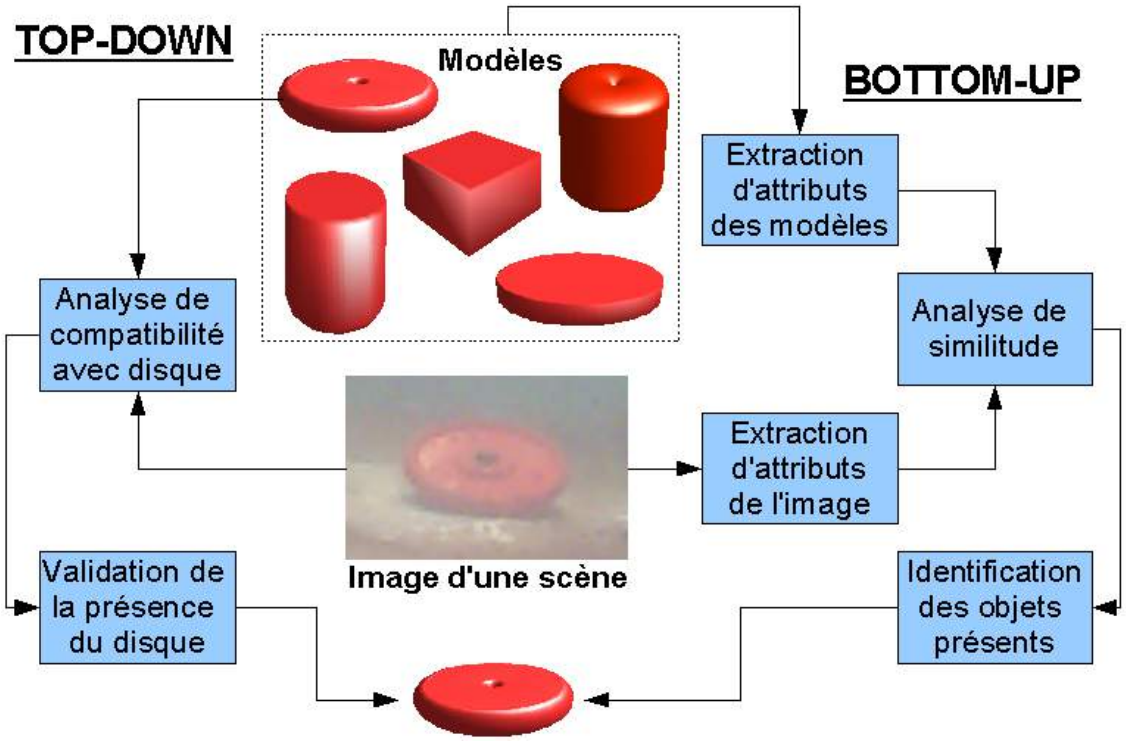


FIG. 3.3 – Les algorithmes de reconnaissance.

3.2.2 Reconnaissance de formes et analyse de scènes

En reconnaissance de formes et analyse de scènes, on distingue plusieurs méthodes répertoriées par Battle [61]. On peut en extraire trois grandes approches qui diffèrent notamment sur l'utilisation de l'information *a priori* :

– L'approche "**Top-down**" : On fait l'hypothèse que l'image contient un certain objet, puis on vérifie la présence de l'objet supposé. On doit appliquer cette technique autant de fois qu'il y a

d'objet mais l'hypothèse de base augmente la robustesse de l'algorithme.
- L'approche "**Bottom-up**" : Dans ce cas, la scène est segmentée utilisant des méthodes générales sans aucun *a priori* sur une quelconque présence. Ensuite des caractéristiques sont extraites de l'image segmentée. Enfin elles sont comparées à celles des modèles puis classifiées.
- L'approche "**Hybride**" : C'est une combinaison des deux méthodes précédentes. Plusieurs techniques sont possibles. La plus commune segmente l'image par des méthodes non spécialisées et ensuite les objets sont identifiés à partir de procédures spécialisées à chaque modèle.

Les deux premières approches sont illustrées sur le schéma 3.3.
Dans notre contexte, puisque nous cherchons à reconnaître des objets dans des séquences d'images, on peut alors décomposer encore un peu plus le processus de reconnaissance et on obtient quatre étapes :
- la détection qui spécifie l'absence ou la présence d'un objet dans une image,
- la localisation qui spécifie la place de l'objet dans l'image présentant un objet (cette étape parfois fastidieuse peut être éventuellement supprimée au prix d'une extraction d'attributs n'appartenant pas à l'objet mais au fond par exemple),
- la caractérisation qui est une mesure d'attributs sur l'objet (par exemple attributs de forme, couleur, texture, dimensions),
- la classification qui analyse les similitudes entre les attributs extraits de l'image et ceux contenus dans la base puis les classifient pour aboutir à l'identification de l'objet.

Les attributs extraits (des images ou des modèles) peuvent être de deux types, 3D ou 2D, et dépendent des connaissances *a priori* disponibles. Soit on dispose du modèle 3D des objets et des informations nécessaires pour reconstruire un modèle 3D de l'objet, on alors parlera de méthode 3D. Soit on ne dispose pas du modèle *a priori* ou des informations nécessaires à la reconstruction, et dans ce cas on parlera de méthode 2D. Les méthodes 3D vont extraire des attributs 3D des images pour reconnaître l'objet. Les méthodes 2D vont elles extraire des attributs 2D. En fonction de ces attributs extraits, on obtient donc deux types de méthodes assez différentes même si elles respectent toutes les deux l'architecture Fig. 3.3.

3.2.3 Les méthodes utilisant l'information 3D

L'utilisation de l'information 3D est naturelle quand elle est disponible afin de faciliter la reconnaissance. Pour pouvoir apparier des vues 3D avec un modèle 3D, il faut reconstruire l'objet et pour cela on distingue encore deux approches "feature-based" et "model-based". Dans le premier cas, la reconstruction 3D est basée sur la poursuite sur la séquence d'images de caractéristiques géométriques. Dans le deuxième cas, un modèle global est utilisé pour aider à la reconstruction, imposant des contraintes globales supplémentaires pour plus de robustesse. Dans les deux cas la mise en correspondance se fait sur des primitives géométriques : sommets, arêtes, faces (obtenues par recherche de points remarquables ou de droites). On peut citer deux méthodes très utilisées : 3DPO[66], PREMIO [67] et on pourra avoir une bonne revue des différentes méthodes de reconnaissance utilisant l'information 3D dans [68].

Le processus classique pour la reconnaissance d'objets sous-marins par une méthode 3D est le suivant [37] :

1. Prétraitement de l'image.
   - Approche déconvolution : filtrage inverse [1],
   - Approche rehaussement : débruitage, rehaussement de contraste, uniformisation de l'éclairage,
2. Extraction de points caractéristiques.
   - Points de contour (Sobel, Canny, Deriche),
   - Points d'intérêt et coins (Susan, Css, Harris),
   - Points invariants à l'espace échelle.
3. Mise en correspondance de points.
   - Méthode de corrélation,
   - Méthode de suivi de points (points d'intérêt),
   - Méthode de caractérisation de points par des invariants locaux.
4. Reconstruction de l'objet.
   - Sans information supplémentaire : reconstruction d'un modèle 3D à un facteur d'échelle près.
   - Avec des connaissances a priori sur la scène : reconstruction d'un modèle métrique.
5. Classification et reconnaissance à partir du modèle ou des attributs métriques extraits.
   - Méthode Bayésienne,
   - Réseau de neurones,
   - SVM : Support Vector Machine,

Toute méthode de reconstruction nécessite la calibration de la caméra (voir annexe A). Dans notre cas les hypothèses nécessaires à l'autocalibration[6] [69] de la caméra sont :
- Scène rigide,
- Paramètres internes de la caméra constants (ni zoom, ni autofocus).

Nous devons pratiquer l'autocalibration puisque nous ne disposons pas de la caméra mais seulement des séquences. Sous ces hypothèses permettant la calibration de la caméra nous pouvons faire une reconstruction euclidienne à un facteur d'échelle près. Pour déterminer ce facteur d'échelle, il est nécessaire de disposer de connaissances *a priori* sur la scène ou sur le mouvement de la caméra (pour plus d'explications sur les méthodes de calibration en milieu sous-marin, nous renvoyons le lecteur à [38][40]). On peut supposer que l'on pourrait sans trop de difficulté disposer d'une estimation des mouvements de caméra, ou d'une information métrique : soit la distance entre la scène et la caméra (sondeur, sonar), soit une référence métrique dans la scène (par pointeur laser par exemple). Cependant cette méthode, même si elle paraît théoriquement possible, semble très compliquée à réaliser techniquement. Elle nécessite de consacrer une grande partie de l'étude à la reconstruction de modèle 3D en environnement non contrôlé. En effet, dans notre contexte, elle est peu robuste vu la qualité médiocre des images et le manque de connaissance sur les conditions d'acquisition.
Un important travail de documentation a été fait sur les techniques de calibration et autocalibration ainsi que sur les techniques de reconstruction 3D multivues [70] et monovue [41][42][71][72][73][74][75][76]. La méthode 3D robuste aux changements de points de vue et aux occultations partielles sous entend par contre que l'on sait déjà que l'objet est présent dans l'image et que l'on cherche à le reconnaître. Cette méthode n'a pas été appliquée dans le contexte qui nous intéressait cependant nous

[6] par autocalibration nous désignons une méthode permettant de calibrer automatiquement une caméra avec les images disponibles sans utiliser des objets spécifiques de calibrage comme par exemple une mire.

pensons qu'elle est utilisable dans le contexte sous-marin mais en se plaçant par exemple dans le cas monoimage et utilisant les méthodes de suivi [77].

Le problème principal des méthodes de reconnaissance 3D est la lourdeur algorithmique. De plus, elles sont en général très spécialisées à un type d'objet (prenons l'exemple classique du contrôle qualité dans l'industrie) et donc peu généralisables (la base d'objet est typiquement très petite si l'on veut avoir des résultats robustes et assez rapides). La réussite des méthodes 3D dépend de la précision dans la détection des contours. Dans notre contexte, cette exigence n'est en général pas respectée contrairement aux environnements très contrôlés.

3.2.4 Les méthodes utilisant l'information 2D

Les méthodes 2D vont généralement beaucoup plus vite car elles utilisent des attributs très simples. Elles sont moins spécialisées et donc beaucoup plus facilement extensibles. Cependant, elles sont souvent plus délicates pour l'étape de classification et de décision car aucun changement de points de vue n'est pris en compte. Cette méthode a pour but la reconnaissance d'un objet 3D utilisant comme modèle des attributs basés sur une ou plusieurs vues 2D de l'objet ou son modèle 3D. On peut qualifier cette méthode de "basée apparence". On pourra trouver une bonne description de cette méthode dans [78]. On peut cependant distinguer d'ores et déjà deux types approches dans ces méthodes basées apparence dépendant encore une fois des données *a priori* disponibles : soit on dispose d'un modèle 3D des objets et dans ce cas on peut utiliser l'information 3D du modèle et essayer de l'apparier avec une vue 2D, soit on ne dispose que de vues 2D de l'objet et dans ce cas on va tenter de mettre en correspondance des attributs extraits et ceux de la base de reconnaissance.
L'appariement entre le modèle 3D et une vue 2D (appariement 3D/2D) se fait à partir de primitives géométriques : sommets, arêtes, faces (obtenues par recherche de points remarquables ou de droites). Il est compliqué et impose en général des hypothèses fortes pour être robuste. Cette approche est donc difficilement applicable dans la version proposée dans la littérature [79] vue la qualité des images et les objets que l'on recherche. Malgré cela, cette méthode reste intéressante et sera rediscutée dans le chapitre 6. On peut citer deux exemples couramment utilisés dans la littérature effectuant ce type de travail : ACRONYM [80], SCERPO [79].
Généralement, on ne dispose pas de l'information 3D, la méthode la plus classique consiste donc en l'appariement d'attributs 2D/2D. On entend par attributs 2D à la fois les attributs de forme mais aussi d'autres attributs comme la couleur, la texture... On peut citer notamment les méthodes utilisant les histogrammes de couleurs [81], les histogrammes de couleurs, de forme et de texture (le système SEEMORE [82]), le système HYPER [83], le système SPIKENET [84], les méthodes d'agrégation de Viola [85], la silhouette (courbure des contours) [86], également différentes méthodes utilisant les ondelettes et la transformée de Fourier [87] ou la transformée de Hough généralisée [88].

Supposant une approche "Bottom-Up", le processus classique pour la reconnaissance d'objets sous-marins par méthode 2D est le suivant [89][90] :

1. Prétraitement de l'image.
   - Approche déconvolution : filtrage inverse...[1]
   - Approche rehaussement : débruitage, rehaussement de contraste, uniformisation de l'éclairage et correction des couleurs.

A priori	*A posteriori*	Avantages	Inconvénients
3D	3D	Très robuste	Système complexe (reconstruction) Base petite et peu extensible
3D	2D	Robuste Rapide	Hypothèses fortes Sensible aux bruits
2D	2D	Simple Rapide Extensible	Sensible aux points de vue Sensible à l'environnement (fond, éclairage) Sensible à la qualité des données

TAB. 3.2 – Les différents types de méthodes en reconnaissance d'objet.

2. Segmentation de l'image.
   - Approche frontière : méthodes dérivatives classiques suivies d'une fermeture de contour, méthodes globales des contours actifs, méthodes markoviennes.
   - Approche région : méthodes markoviennes, méthodes structurales.
3. Extraction d'attributs.
   - Forme : Présence de formes particulières (coins, cercles, d'ellipses, de droites...), attributs des contours (nombre, longueur, régularité, courbure...), descripteurs de forme
   - Texture,
   - Couleur.
4. Classification et reconnaissance de l'objet.
   - Appariement 2D/2D (méthode Bayésienne, réseau de neurones, SVM)
   - Appariement 2D/3D.

3.3 Les méthodes utilisables dans notre contexte

Dans le contexte opérationnel que nous avons présenté précédemment, on focalisera notre étude sur la détection et la reconnaissance des objets manufacturés dans les séquences. On pourra éventuellement supposer que la détection des objets potentiels dans la séquence a été effectuée. En effet il existe de nombreuses méthodes de détection automatique de changements en vidéos sous-marines [59][91][60]. On se place dans un cadre d'étude de vision monoculaire non géoréférencée, avec une scène rigide inconnue, une caméra non calibrée et de caractéristiques inconnues, aucune contrainte sur l'angle de prise de vue, et une approche plutôt image par image. L'aspect temporel de la séquence sera utilisé dans un deuxième temps pour améliorer les performances des algorithmes.

Les méthodes utilisant un modèle 3D des objets connus *a priori* sont plus robustes mais spécialisées pour un type de problème et peu généralisables à d'autres objets. La base des objets à reconnaître est généralement très petite (typiquement moins de cinq objets) et difficilement extensible. Les objets doivent être géométriquement simples et faciles à décrire, et la réussite de l'identification dépend de la bonne détectabilité des attributs utilisés, c'est-à-dire souvent des contours [92][93]. Les méthodes basées seulement sur des attributs 2D sont plus rapides car elles utilisent des attributs plus simples. Elles sont également moins spécialisées, donc beaucoup plus facilement extensibles à d'autres objets. Par contre, elles ne prennent en général pas en compte du changement de points de vue, et pour un même objet, il faut parfois autant d'images dans la base que de points de vue possibles (voir tableau

récapitulatif des méthodes de reconnaissance 3.2). A noter une méthode intéressante indépendante du changement de point de vue : SIFT[94].

Le choix de la méthode à utiliser dépend de la qualité l'information *a priori* et de l'environnement d'acquisition. Durant cette thèse nous avons donc seulement travaillé sur les méthodes utilisant l'information 2D mais avec les deux approches proposées c'est-à-dire avec un appariement 3D/2D et 2D/2D. Nous présenterons également deux méthodes différentes en appariement 2D/2D.

3.3.1 La méthode 2D classique orientée forme

D'abord nous avons testé la méthode utilisée classiquement en vidéo sous-marine c'est-à-dire la méthode 2D/2D basée sur les contours avec une approche "Bottom-up". En effet le contour est l'un des seuls attributs réellement exploitable pour la reconnaissance d'objet en milieu perturbé. Cette méthode 2D/2D est séquentielle et nécessite plusieurs étapes comme nous avons pu le voir dans la partie précédente.
La première est le prétraitement des images. Il a pour but de s'affranchir de la variabilité des conditions d'acquisition dans l'objectif global de traitement automatique. Différents prétraitements ont été étudiés. Le choix s'est porté sur une approche basée sur le rehaussement et le lissage. C'est-à-dire que nous avons décidé de traiter chaque problème séquentiellement : d'abord uniformiser l'éclairage, débruiter, rehausser les contrastes, et enfin corriger les couleurs. L'algorithme proposé est un algorithme rapide et automatique qui ne nécessite aucune connaissance *a priori* et aucun réglage de paramètre par l'utilisateur. La méthode globale de correction basée sur la déconvolution a été abandonnée car elle nécessite des connaissances *a priori* (par exemple la distance à la scène, la qualité de l'eau, la profondeur...) et qu'il est très difficile de déterminer les paramètres d'un modèle physique de diffusion de la lumière dans l'eau.
Ensuite nous avons appliqué à ces images prétraitées un algorithme de segmentation pour en extraire les contours. Nous avons testé plusieurs détecteurs et différents post-traitements pour fermer les contours et sélectionner les plus pertinents, autrement dit les plus marqués. Cette étape a permis d'obtenir de façon automatique et robuste les contours pertinents dans les images.
Pour finir, une étape d'extraction d'attributs de forme a été appliquée. L'étape de classification n'a pas été développée puisque que nous nous sommes arrêtés à quelques attributs simples avec une prise de décision empirique. En perspectives, nous prévoyons évidemment d'effectuer cette classification de manière à obtenir une validation effective des résultats. La partie la plus importante du travail sur cette méthode classique est basée sur le prétraitement et la segmentation des images sous-marines. Grâce au prétraitement, la méthode de segmentation développée est assez robuste au bruit, aux problèmes d'éclairage et aux couleurs atténuées. En plus, elle est générique et sans *a priori*. Cependant, puisque c'est une méthode 2D, elle est dépendante du point de vue et des occultations. Elle est aussi sensible au fond malgré les efforts sur la segmentation et la sélection des contours pertinents. Enfin comme la recherche est passive c'est-à-dire que l'on adopte une stratégie partant de l'image brute et ne faisant aucune hypothèse sur la présence ou non d'un objet, il est parfois difficile de conclure sur le type d'objet présent.

3.3.2 La méthode 2D cherchant la compatibilité 2D/3D

Dans un deuxième temps nous avons voulu tester d'autres méthodes plus "spécialisées" notamment des méthodes utilisant la connaissance *a priori* de l'objet recherché, c'est-à-dire que supposant une approche "Top-down". En effet pour améliorer la qualité de la reconnaissance nous allons rechercher dans l'image l'objet que nous voulons détecter. Puisque la reconstruction est trop difficile et comme nous disposons des modèles 3D des objets, nous avons cherché à développer des méthodes d'appariement 2D/3D c'est-à-dire des méthodes cherchant à prouver la compatibilité entre un modèle 3D et une image 2D d'un objet. Cette méthode qui a été étudiée n'a pas été développée aujourd'hui, nous y reviendrons dans le chapitre 6 sur les méthodes spécialisées à un type d'objet. Cependant, il apparaît clairement que cette méthode est théoriquement applicable dans notre contexte de reconnaissance. Le travail est par contre très complexe vu la qualité des images.

3.3.3 La méthode 2D cherchant la compatibilité de couleur

A la suite des problèmes sur la méthode précédente et sachant qu'en plus du modèle 3D *a priori* des objets, sa couleur était connue, nous avons essayé d'utiliser la couleur comme paramètre discriminant. En effet la couleur est un attribut simple, robuste aux changements d'échelles, aux changements de points de vue, et aux occultations. La méthode développée suit donc le même concept que la méthode précédente c'est-à-dire de vérifier la compatibilité entre les couleurs d'une image et celle d'un objet connu *a priori*. Cette méthode doit par contre faire face à des problèmes très spécifiques dans le domaine sous-marin. En effet les couleurs sont atténuées et modifiées selon la distance en un mot, nous avons un problème de constance de couleur. La littérature utilise en général la forme, plus robuste à cause des problèmes cités précédemment. Cependant la couleur est un attribut très utilisé dans le domaine terrestre ou il est très robuste et souvent discriminant [95][96]. Dans le milieu sous-marin, la couleur a été utilisée comme un paramètre additionnel [57] mais jamais comme un paramètre autosuffisant. Cette dernière méthode constitue une contribution originale dans le domaine sous-marin.

3.4 Conclusion

Dans ce chapitre, nous avons discuté des méthodes de reconnaissance d'objets en vision. Nous avons présenté les méthodes existantes aujourd'hui dans la littérature et discuté des différentes contraintes liées au contexte opérationnel ainsi que des différentes configurations possibles pour les systèmes de vision. Dans le contexte de notre étude, c'est-à-dire en vision monoculaire non géoréférencée avec une scène rigide inconnue, une caméra non calibrée de caractéristiques inconnues, et aucune contrainte sur l'angle de la prise de vue, nous avons détaillé les méthodes utilisables. Ces méthodes sont au nombre de trois : la méthode de reconnaissance basée sur les contours très utilisée dans la littérature, et deux méthodes basées sur un concept de compatibilité : la comptabilité entre le modèle 3D et l'image 2D et compatibilité de couleur. Ces deux dernières méthodes supposent une connaissance *a priori* sur les objets recherchés. Parmi ces trois méthodes que nous avons étudiées, nous avons seulement implémenté deux d'entre elles. La troisième est l'une des perspectives de cette thèse. Les deux méthodes que nous avons développées sont présentées sur le schéma 3.4. Elles ont fait l'objet de trois contributions détaillées dans les trois prochains chapitres. Ces deux méthodes ont été validées

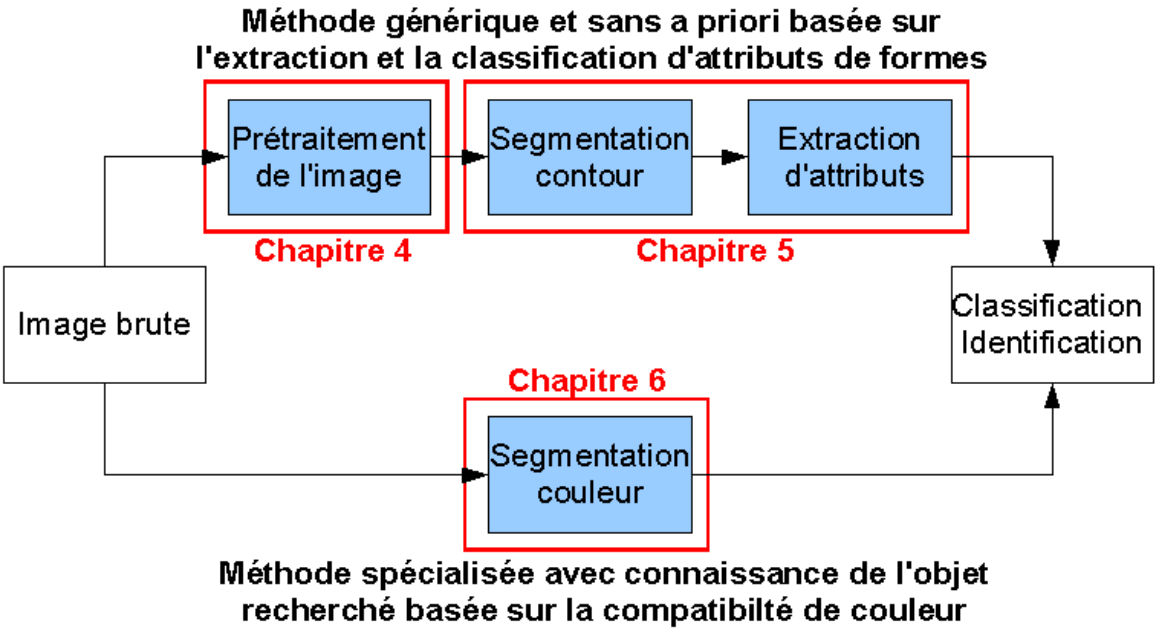


FIG. 3.4 – Les contributions de la thèse.

sur des données réelles, les résultats seront présentés dans le chapitre 8.

Chapitre 4

Le prétraitement des images sous-marines

Dans la plupart des cas, les méthodes classiques de segmentation sont inexploitables avec les images sous-marines en raison notamment du bruit important et des faibles contrastes. Pour pouvoir les utiliser il est nécessaire de prétraiter les images pour éliminer les phénomènes perturbateurs dus au contexte sous-marin. Dans ce chapitre nous allons exposer deux approches de prétraitement : la restauration et le rehaussement [97].

4.1 Approche globale ou approche séquentielle ?

Deux méthodes existent pour prétraiter les images, c'est-à-dire pour corriger leurs défauts [90]. Une méthode globale : la restauration et une méthode séquentielle : l'amélioration. Dans le premier cas, on s'attache à corriger des défauts connus par un algorithme d'inversion, dans le deuxième cas on tente d'atténuer les défauts et d'améliorer la qualité de l'image.

4.1.1 La restauration appliquée aux images sous-marines

La restauration d'image a pour objectif principal d'atténuer voire de supprimer dans le cas le plus favorable, les dégradations qu'a subies une image. Le plus souvent, les dégradations considérées sont celles qui apparaissent au moment de l'acquisition de l'image, c'est-à-dire au moment de sa création. La restauration d'image suit un principe très simple : modéliser la dégradation et appliquer la dégradation inverse à l'image dégradée pour obtenir l'image originale.

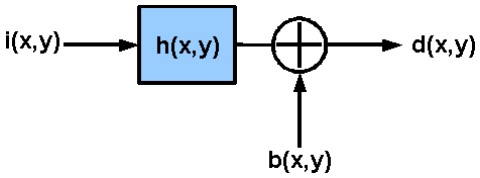
Les dégradations (ou bruit) peuvent être de plusieurs types :
- bruit lié au contexte d'acquisition (un effet de bougé ou des conditions d'éclairage modifiées),
- bruit lié au capteur (une distorsion géométrique ou un flou de mise au point),
- bruit lié à l'échantillonnage (un effet de moiré ou un bruit de quantification "poivre et sel"),
- bruit lié à la nature de la scène (par exemple l'environnement sous-marin qui cause de nombreuses dégradations spécifiques telles que les particules en suspension, l'atténuation ou la diffusion).

Les sources de dégradations étant souvent très difficiles à modéliser, on préfère en général se ramener à une modélisation de leurs effets dans l'image. On peut classer ces effets en trois catégories :

- les dégradations additives : Elles proviennent de l'addition d'un signal bruit à l'image idéale. C'est la modélisation généralement adoptée pour l'intervention d'un bruit blanc dans une image.
- les dégradations multiplicatives : La dégradation de l'image idéale se fait par multiplication. L'exemple le plus classique que l'on peut citer pour illustrer ce type de dégradation est celui de l'éclairage d'une scène qui multiplié avec la réflectance des objets produit l'image observée.
- les dégradations convolutives : Elles agissent sur l'image idéale comme un filtre convolutif, pour donner l'image dégradée. De nombreuses dégradations peuvent se modéliser de cette façon, comme les effets d'une mauvaise mise au point, d'un bougé...

La modélisation du problème de restauration

D'une façon générale on peut représenter une modification quelconque de l'image par le modèle suivant (voir Fig. 4.1.1) :



Ce modèle peut être résumé par l'équation suivante :

$$d(x,y) = i(x,y) \otimes h(x,y) + b(x,y) \tag{4.1}$$

où i(x,y) est l'image originale, b(x,y) un bruit additif, d(x,y) l'image dégradée et h(x,y) la réponse impulsionnelle du filtre linéaire invariant en translation[1].

L'objectif de la restauration est donc de retrouver l'image i(x,y) à partir de d(x,y). Selon les méthodes on doit ou non disposer d'informations supplémentaires sur la dégradation. Soit des informations *a priori* c'est-à-dire des connaissances du processus physique qui provoque la dégradation ; soit des informations *a posteriori* c'est-à-dire des connaissances déduites de l'image dégradée. La qualité de restauration dépendant fortement de la quantité d'informations à disposition.

Le filtrage inverse

Les méthodes de déconvolution (ou filtrage inverse) traitent en général le problème par l'intermédiaire de représentations fréquentielles. En effet les dégradations que subissent les images ont des effets qu'il est plus facile de modéliser sur une représentation fréquentielle que sur l'image elle même. Comme nous avons pu le voir une image dégradée peut être approximativement représentée par l'équation 4.1, où h est l'opérateur de distorsion aussi appelé PSF (Point Spread Function). C'est la convolution de cet opérateur avec l'image qui crée la distorsion.
Tout le problème de la restauration est donc de déconvoluer l'image bruitée par la PSF qui décrit exactement la distorsion. La qualité de la restauration dépend bien évidemment de la précision de

[1] invariant en translation signifiant qu'il s'agit d'un filtre dont l'action est indépendante de la position spatiale de la zone de l'image

cette PSF [98].
En passant dans le domaine fréquentiel ou domaine de Fourier, l'équation 4.1 devient :

$$D(u,v) = I(u,v).H(u,v) + B(u,v) \tag{4.2}$$

En l'absence de bruit on définit donc la déconvolution par l'équation 4.3, dans le cas contraire par l'équation 4.4 :

$$I(u,v) = \frac{D(u,v)}{H(u,v)} \tag{4.3}$$

$$I(u,v) = \frac{D(u,v)}{H(u,v)} - \frac{B(u,v)}{H(u,v)} \tag{4.4}$$

Dans la pratique la présence de bruit dans le processus de dégradation introduit une dérive importante sur le résultat de la restauration. Lorsque le bruit devient prépondérant dans l'image de sortie, il est donc important d'adapter la fonction de restauration pour obtenir de bons résultats. On utilise alors le filtre de Wiener qui repose sur l'utilisation d'une mesure objective de la qualité de la restauration obtenue par moindres carrés. Ce filtre est destiné à minimiser l'erreur liée au processus de restauration, à savoir la différence entre l'image restaurée et l'image originale, au sens des moindres carrés.

Le filtrage inverse pour prétraiter les images sous-marines

Le modèle de formation des images sous-marines étant connu (modèle de Jaffe et McGlammery [28][99]) et modélisable sous la forme d'une convolution, le filtrage inverse semble la méthode de prétraitement la plus évidente. Dans sa thèse Olmos [1] a proposé une méthode de restauration des images sous-marines basée sur le modèle précédent qui permet de générer des images sous-marines connaissant les caractéristiques du milieu. Ces paramètres physiques et la configuration de la scène n'étant pas toujours connus, Olmos a proposé une méthode permettant d'estimer les paramètres nécessaires à la déconvolution par une procédure d'optimisation au sens d'un critère d'évaluation du flou (ces paramètres doivent cependant être initialisés dans des intervalles de valeurs raisonnables pour que l'estimation converge). La méthode de déconvolution proposée est la suivante :

$$d(x,y) = TF^{-1}[TF(i(x,y)).OTF(x,y)], \tag{4.5}$$

avec TF et TF^{-1} la transformée de Fourier (voir 4.2.2) et transformée de Fourier inverse, $OTF = \frac{1}{K}.\exp(R_c.f.c)$. L'OTF (Optical Transfert Function) est la transformée de Fourier de la PSF. Dans cette équation K est une constante [0.2,0.9], c correspond au coefficient d'atténuation de l'intensité lumineuse dans l'eau [0.05,0.5], R_c est la profondeur, et f la distance focale de la caméra.

Cette méthode donne de bons résultats et permet d'améliorer nettement la qualité visuelle des images ainsi que la qualité des résultats de segmentation utilisant les opérateurs standards (l'amélioration de la segmentation a été démontrée à l'aide d'un critère statistique évaluant les probabilités de bonnes et fausses détections par rapport aux vérités terrain). Malgré cela, cette méthode présente deux défauts majeurs : d'abord elle requiert des informations *a priori* (coefficients d'atténuation, profondeur lors de la prise de vue et focale de la caméra) ou bien des intervalles estimés de ces paramètres d'initialisation pour l'algorithme d'auto-évaluation (de l'initialisation correcte de ces intervalles dépend la précision de l'auto-évaluation). Ensuite elle reste limitée au traitement du phénomène de diffusion (caractérisé par un flou dans l'image), et elle ne prend pas en compte l'atténuation.

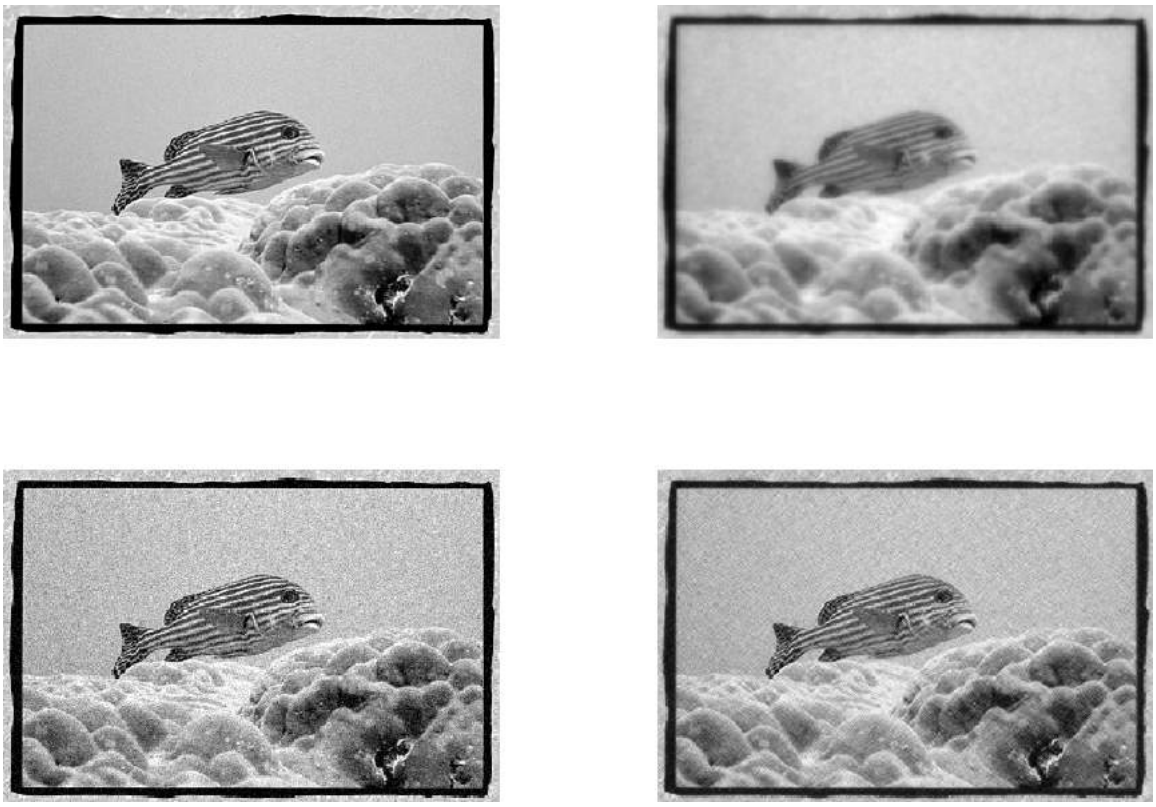


FIG. 4.1 – Filtrage inverse : (a) image originale, (b) image dégradée par la PSF de Olmos, (c) image après déconvolution standard, (d) image après déconvolution Wiener, les deux déconvolutions utilisent la PSF de Olmos (Image P. Hince).

Conclusion

Les méthodes de restauration sont très performantes lorsque l'on connaît avec précision la dégradation qu'a subie l'image (c'est-à-dire dans un contexte très particulier comme un effet de bougé par exemple). Dans notre cas, pour prétraiter des images sous-marines le modèle de dégradation globale est difficile à utiliser car les paramètres sont nombreux, souvent inconnus et surtout difficilement estimables car ils sont très changeants entre les différentes séquences d'images. Le modèle approché et évaluant automatiquement ces paramètres proposé par Olmos [1] n'a pas été applicable dans le cadre de notre étude puisque que nos images d'entrée étaient très hétérogènes et que l'on ne disposait d'aucune information *a priori* sur toutes ces images pour initialiser correctement le processus d'auto-évaluation. Du fait de cette mise en oeuvre délicate et sachant que ce modèle approché ne traitait pas la totalité des défauts constatés (uniquement le flou dû à la diffusion), nous avons préféré à la restauration la méthode de rehaussement. Cette méthode est d'ailleurs souvent préférée dans le domaine de la vision sous-marine.

FIG. 4.2 – Filtrage inverse de Olmos sur des images avec une dégradation inconnue : Images originales à droite, images après déconvolution Wiener à gauche. On distingue des contrastes plus marqués mais davantage de bruit après déconvolution.

4.1.2 La méthode adoptée : l'amélioration

Contrairement à la restauration qui est un processus objectif corrigeant les altérations de signal que l'on peut reconnaître et quantifier, l'amélioration d'image s'attache à la qualité de l'image qui est une notion beaucoup plus subjective. Les défauts que l'on cherche à corriger sont généralement inconnus et on utilise généralement des techniques heuristiques[2]. La stratégie adoptée a donc été d'appliquer différentes méthodes de filtrage pour corriger un à un les défauts présentés dans le chapitre 2. Cette correction a été optimisée pour améliorer les résultats de segmentation.
Étant donnée la littérature très riche sur les méthodes de filtrage améliorant la qualité des images, nous allons dans la partie suivante en présenter un éventail assez représentatif avant d'expliquer les traitements choisis pour prétraiter nos images sous-marines. Pour plus d'informations sur tous ces algorithmes de filtrage on renvoie le lecteur à quelques ouvrages de référence par exemple [89][90]. Mise à part la restauration, il n'existe à notre connaissance aucun algorithme dédié spécifiquement aux

[2] une méthode heuristique ou méthode approximative est un algorithme qui fournit une solution réalisable, pas nécessairement optimale, pour un problème d'optimisation NP-difficile.

images sous-marines. Nous allons présenter dans la partie suivante des diverses méthodes de traitement d'images permettant chacune de corriger ou d'atténuer des défauts dus au milieu sous-marin.

4.2 État de l'art des méthodes de filtrage

De nombreuses méthodes de filtrage existent dans le domaine de la vision, parmi lesquelles on peut distinguer trois types :

- Le filtrage spatial,
- Le filtrage fréquentiel,
- Le filtrage multirésolution.

Dans le cas général, toutes ces techniques sont indépendantes de la couleur, et traitent soit des images monochromes, soit les trois plans couleur : rouge, vert, bleu, indépendamment. Dans la dernière partie de cet état de l'art nous présenterons le cas particulier de quelques algorithmes appliqués à la correction des couleurs.

4.2.1 Le filtrage spatial

Le filtrage spatial regroupe toutes les fonctions qui traitent l'image dans le domaine spatial c'est-à-dire au niveau du pixel élémentaire. Il est utilisé pour le rehaussement de contraste, la détection de contour, l'élimination de bruit... Nous verrons dans cette partie, les méthodes basées sur l'histogramme, et les opérations basées sur la convolution.

Filtrage par l'histogramme

Les méthodes de filtrage par l'histogramme dites aussi méthodes ponctuelles traitent les pixels de l'image de façon indépendante de leurs voisinages (quelques exemples sont présentés sur la figure 4.3). La seule information prise en compte est la valeur des niveaux de gris, quelle que soit la position des pixels dans l'image. Ainsi tous les pixels qui ont le même niveau de gris dans l'image d'origine ont le même niveau de gris dans l'image améliorée. Parmi ces méthodes basées sur la modification ou la spécification d'histogramme on distingue :

- la translation d'histogramme (voir paragraphe "égalisation des couleurs"),
- le recadrage d'histogramme (aussi appelé expansion de dynamique ou normalisation),
- la modification de dynamique (saturation des valeurs aux bornes),
- et les méthodes basées sur l'égalisation d'histogramme.

L'expansion de dynamique aussi appelée normalisation est une méthode élémentaire de rehaussement, elle permet d'améliorer le contraste dans l'image en étirant la palette d'intensité de l'image de manière à utiliser au mieux l'échelle de niveau de gris disponible. Soient $f(x, y)$ l'image de départ et $f'(x, y)$ l'image après transformation. Soient $[f_{min}, f_{max}]$ l'intervalle des intensités présentes dans l'image et $[min, max]$ l'intervalle disponible. L'expansion de dynamique correspond à la transformation linéaire suivante (éq. 4.6) :

$$f'(x, y) = (f(x, y) - f_{min})(\frac{max - min}{f_{max} - f_{min}}) + min \tag{4.6}$$

Le problème de cette méthode de recadrage est qu'il suffit qu'un pixel soit excentré pour obtenir une transformation non représentative de la réalité. C'est pourquoi une approche plus robuste consiste à prendre d'abord un histogramme de l'image et à sélectionner f_{min} et f_{max} comme par exemple 5% et 95% de l'histogramme. Ceci prévient des éventuels défauts dus aux pixels excentrés. Le cas échéant, pour pallier aux problèmes des pixels hors limites et sachant que la normalisation se fait entre 0 et 1 on applique la transformation linéaire suivante saturant les valeurs excentrées aux bornes de normalisation (voir éq. 4.7).

$$f'(x,y) = \begin{cases} \frac{f(x,y)-f_{min}}{f_{max}-f_{min}} & \text{si } 0 < f(x,y) < 1 \\ 0 & \text{si } 0 > f(x,y) \\ 1 & \text{si } 1 < f(x,y) \end{cases} \tag{4.7}$$

L'égalisation d'histogramme est la méthode de spécification d'histogramme la plus utilisée. Elle consiste à rendre l'histogramme de l'image le plus plat possible. L'image étant associée à une modélisation probabiliste où l'histogramme représente la distribution des niveaux de gris. On applique une transformation permettant d'obtenir une densité uniforme c'est-à-dire une image après égalisation où tous les niveaux de gris sont équiprobables. Cette méthode permet souvent de renforcer le contraste sur des détails de l'image.
Remarque : En raison de son action globale, l'égalisation d'histogramme est mal adaptée aux images sous-marines présentant souvent un éclairage non uniforme.

L'égalisation locale d'histogramme utilise une fenêtre glissante pour rendre locale l'action de l'égalisation d'histogramme. La stratégie consiste à définir un voisinage $n \times n$, puis calculer l'histogramme de cette zone et ensuite appliquer la fonction d'égalisation d'histogramme. On remplace l'intensité du pixel situé au centre de la fenêtre par la nouvelle intensité calculée après égalisation. Un inconvénient majeur de cette méthode est évidemment le temps de calcul. Un second inconvénient est que cette méthode a tendance à amplifier le bruit dans les zones peu contrastées. La méthode suivante permet de pallier à ce problème.

L'égalisation d'histogramme adaptative à contraste limité aussi appelée CLAHE (Contrast Limited Adaptative Histogram Equalization) est une amélioration de la méthode exposée ci-dessus [24]. Plutôt que d'utiliser une fenêtre glissante, on divise l'image en une mosaïque d'imagette de taille $n \times n$ et on calcule l'histogramme de chaque imagette. Lors de l'égalisation on calcule un seuil haut H défini comme un multiple de la moyenne de l'histogramme original de l'imagette. Toutes les valeurs de l'histogramme qui ont plus de H pixels sont tronquées à cette valeur puis le surplus de pixel est redistribué uniformément sur toutes les valeurs non nulles de l'histogramme original de l'imagette. On peut ensuite égaliser cet histogramme. Afin d'éviter les effets de bords entre les imagettes on réalise aussi une interpolation bilinéaire des histogrammes des imagettes voisines à l'imagette courante.

Filtrage linéaire standard (la convolution)

L'ensemble des méthodes précédentes n'utilise que l'information de niveau de gris portée par le pixel pour le modifier. Ainsi deux pixels de même niveau de gris mais appartenant à des régions différentes vont voir leur niveau de gris modifié de la même façon. A l'évidence cela n'est pas très satisfaisant dans un grand nombre de cas. Comme on le comprend, la limitation essentielle de ces

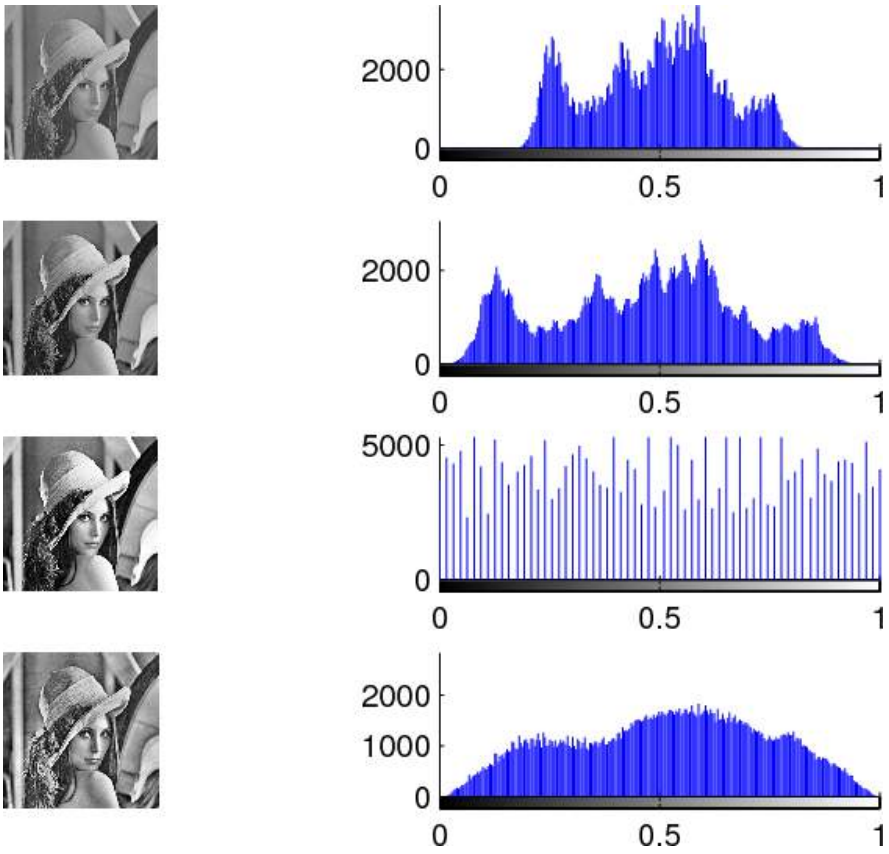


FIG. 4.3 – Les méthodes de filtrage par l'histogramme : Image originale, image après expansion de dynamique, image après égalisation d'histogramme et image après égalisation d'histogramme adaptative.

méthodes est liée au fait qu'elles ne tiennent pas compte de l'information contenue dans le voisinage du pixel à traiter. La méthode suivante permet de pallier à ce problème.

La convolution spatiale peut se "généraliser" comme une moyenne pondérée entre le pixel central et ses voisins. Une fenêtre de coefficients, le noyau, est multiplié par chaque pixel et ses voisins dans une région limitée. Les résultats sont additionnés et la valeur finale prend la place du pixel central. Cette opération est faite pour tous les pixels de l'image. (éq. 4.8).

$$f_c(x,y) = f(x,y) \otimes h(x,y) = \sum_{j=-T}^{T} \sum_{k=-T}^{T} f(x+j, y+k).h(j+T, k+T) \tag{4.8}$$

avec h la matrice de convolution de taille 2T+1 × 2T+1.

Le produit de convolution peut être utilisé à la fois pour l'élimination du bruit, pour l'amélioration de la "qualité" des images et pour la détection de contours. Nous présentons dans le paragraphe suivant quelques masques de convolution très utilisés (résultats sur la figure 4.4) :

– Moyenneur et Gaussien (passe bas). *Élimination du bruit.*

$$H = \begin{pmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{pmatrix} \qquad H = \begin{pmatrix} 1/16 & 2/16 & 1/16 \\ 2/16 & 4/16 & 2/16 \\ 1/16 & 2/16 & 1/16 \end{pmatrix}$$

– Laplacien et Gradient horizontal et vertical de Sobel (passe-haut). *Détection de contour.*

$$H = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix} \quad H_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \quad H_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}$$

En ce qui concerne l'élimination du bruit, ces filtres ne sont pas très bons car ils dégradent considérablement les contours et rendent l'image floue. Les plus utilisés sont le filtre "moyenneur" et le filtre "gaussien" sur de petites fenêtres 3×3 et 5×5 parce qu'avec ces tailles on obtient un bon compromis qualité/coût de calcul.

Le filtrage spatial pour la correction de l'éclairage

Le filtre que nous allons présenter dans cette partie est un filtre par soustraction de l'arrière plan. Cette méthode (appelée en anglais "background substraction") est utilisée pour corriger les images présentant un éclairage non uniforme (cas typique en imagerie sous-marine). Elle consiste à estimer l'illumination dans l'image en effectuant un filtrage passe-bas puis à diviser l'image originale par l'image basse fréquence pour supprimer les défauts d'illumination (voir éq. 4.9)

$$g(x,y) = \frac{f(x,y)}{f_s(x,y)}.\delta \tag{4.9}$$

où f est l'image originale, f_s l'image lissée et δ une constante de normalisation permettant de rétablir la luminance générale de l'image.

Pour obtenir l'image passe bas, le lissage doit être très fort c'est-à-dire un masque de convolution d'une taille à peu près équivalente à 1/8ème de l'image. En raison du temps de calcul très important on effectue ce traitement dans le domaine de Fourier de manière à transformer la convolution en multiplication. (voir paragraphe "transformée de Fourier" 4.2.2).
On peut encore améliorer ce résultat en utilisant la méthode de [25] basée sur la fusion des résultats d'un filtrage par division et par soustraction, car ils apportent des informations complémentaires (voir Fig. 4.5)

L'algorithme du filtrage par fusion des résultats de division et soustraction est le suivant.
– Calcul de l'image basse fréquence (convolution spatiale ou multiplication fréquentielle)
– Division de l'image originale par l'image basse fréquence.
– Soustraction de l'image originale par l'image basse fréquence.
– Fusion des deux images précédentes par l'équation 4.10 :

$$I_{res} = (F * S + (1/F) * D)/(F + 1/F). \tag{4.10}$$

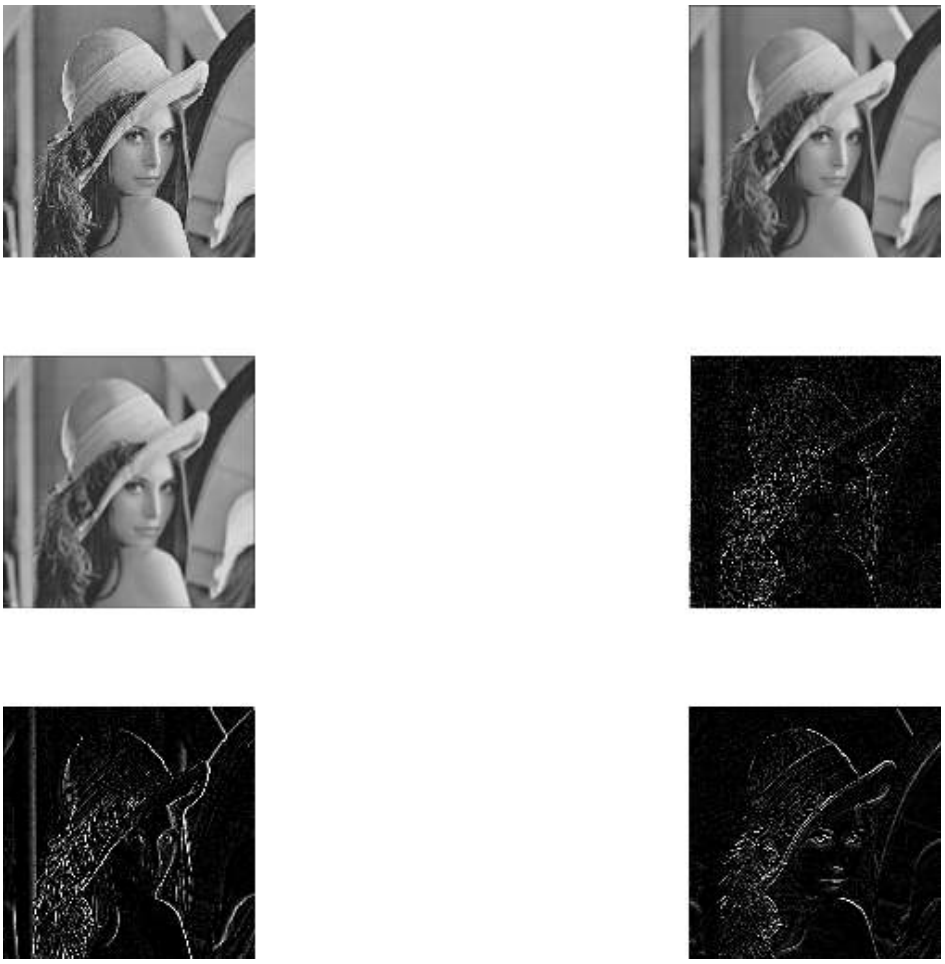


FIG. 4.4 – Quelques exemples de convolution : Image originale, image après filtrage gaussien 7×7, image après filtrage moyenneur 7×7, après filtrage laplacien et image après filtrage de Sobel dans les directions horizontale et verticale (dans le sens classique de lecture).

où F est l'image basse fréquence, D l'image après division par l'image basse fréquence et S l'image après soustraction par l'image basse fréquence.

Filtrage non linéaire : médian et anisotropique

Le filtre non linéaire le plus courant est sans nul doute le filtre médian. Son principe est très simple et il donne des résultats très intéressants. Il est basé sur des considérations statistiques robustes qui préconisent l'utilisation de la médiane à la place de la moyenne en situation bruitée. Cette dernière étant en effet très sensible à la présence de valeur erronée. Lors de l'application de ce filtre on affecte donc à chaque pixel la valeur médiane dans une zone de voisinage (par exemple 5 x 5).

Un autre filtre non linéaire courant est le filtrage anisotropique. Ce filtre est présenté en détails ici car il a été utilisé dans l'algorithme de prétraitement développé. Le problème d'un filtrage gaussien

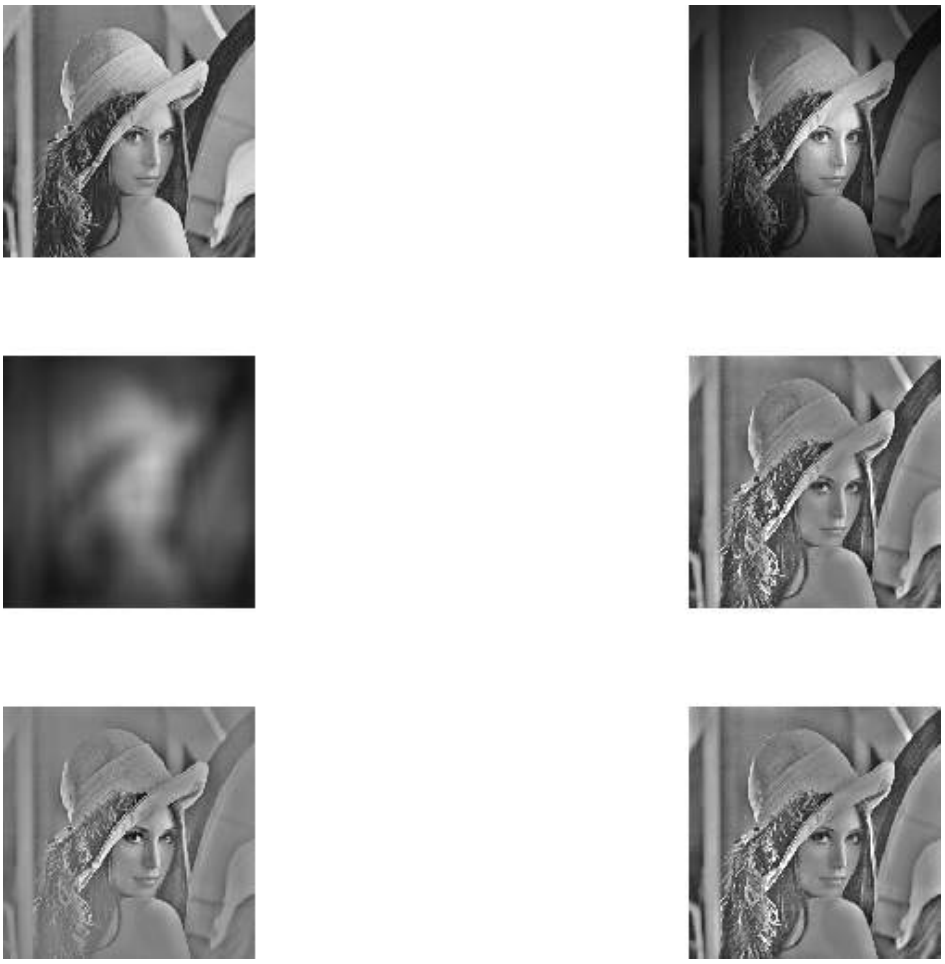


FIG. 4.5 – Filtrage par soustraction de l'arrière plan : Image originale, image dégradée par une illumination gaussienne, image passe-bas, filtrage par division, filtrage par soustraction, fusion des résultats de division et soustraction.

(équivalent du filtrage par diffusion isotropique) est le lissage de toute l'image. Bien que le lieu des contours soit conservé, ces derniers deviennent flous. Il est donc nécessaire de limiter ou interdire la diffusion dans les zones de l'image représentant des contours. Pour cela, un coefficient (noté C) contrôlant la diffusion doit varier en fonction de la position dans l'image. On parle alors de diffusion anisotropique [100][101][102]. Le filtrage anisotropique est un processus itératif. A chaque itération, les contours sont détectés par calcul du gradient de l'image, et ensuite pour chaque pixel un coefficient de diffusion dépendant de cette valeur du gradient est calculé. Pour de faibles valeurs de gradient, on considère que l'on est dans une zone homogène de l'image, on autorise alors la diffusion avec un coefficient élevé. Sur un contour caractérisé par un fort gradient, la diffusion est limitée par un coefficient faible. Les résultats sont présentés sur la figure 4.6. Dans le cas des images couleur le coefficient de diffusion dépend de la norme du maximum des gradients des trois plans couleur, et la diffusion est effectuée sur ces trois plans ou sur le plan de luminance.

Une boucle de l'algorithme de diffusion anisotropique peut être décomposée comme suit :
Pour chaque pixel :

1. Calcul des différences avec les plus proches voisins dans les quatre directions Nord Sud Est Ouest. Calcul du coefficient de diffusion à partir des résultats précédents. Il existe plusieurs possibilités pour ces calculs. La plus utilisée est la suivante :

$$\begin{aligned} \nabla_N I_{i,j} &= I_{i-1,j} - I_{i,j}, & c_{N_{i,j}} &= g(|\nabla_N I_{i,j}|) \\ \nabla_S I_{i,j} &= I_{i+1,j} - I_{i,j}, & c_{S_{i,j}} &= g(|\nabla_S I_{i,j}|) \\ \nabla_E I_{i,j} &= I_{i,j+1} - I_{i,j}, & c_{E_{i,j}} &= g(|\nabla_E I_{i,j}|) \\ \nabla_W I_{i,j} &= I_{i,j-1} - I_{i,j}, & c_{W_{i,j}} &= g(|\nabla_W I_{i,j}|) \end{aligned}$$

où la fonction g est définie par : $g(\nabla I) = e^{(-(\|\frac{\nabla I}{K}\|)^2)}$ avec K fixé à 0.1.
Cette fonction de diffusion privilégie de forts contrastes. Les contours sont même accentués car la valeur de la dérivée au point d'inflexion augmente en valeur absolue.

2. Modification de la valeur du pixel par (Eq.4.11)

$$I_{i,j} = I_{i,j} + \lambda[c_N.\nabla_N I + c_S.\nabla_S I + c_E.\nabla_E I + c_W.\nabla_W I]_{i,j} \tag{4.11}$$

En général λ est $0 \leq \lambda \leq 1/4$. En ce qui concerne le choix de K, il dépend beaucoup du but recherché, de l'image et du nombre d'itérations choisi. Mais on remarque que si on veut vraiment respecter des contours d'amplitude A, il vaut mieux un K de l'ordre de A/2. Dans notre chaîne de prétraitement globale, l'algorithme de filtrage anisotropique a été utilisé pour compléter l'algorithme de débruitage par ondelettes, et le rehaussement de contraste.

Conclusion

Les méthodes de filtrage spatial sont de loin les plus utilisées en traitement d'image. Elles permettent la correction de nombreux problèmes. Nous avons présenté dans cette partie un algorithme de soustraction de l'arrière plan, de filtrage anisotropique, et de traitement par l'histogramme que l'on peut utiliser pour le traitement des images sous-marines. Cependant des méthodes "plus globales" permettent d'obtenir des résultats meilleurs ou équivalents avec des temps de calcul moindres, notamment les méthodes de filtrage fréquentiel et les méthodes utilisant les ondelettes. Nous présenterons ces méthodes dans les deux parties suivantes.

4.2.2 Le filtrage fréquentiel

Contrairement aux méthodes précédentes qui traitent les pixels un par un en tenant compte ou non de leur voisinage, il existe des méthodes dites globales qui traitent tous les pixels en même temps. C'est le cas des méthodes de filtrage fréquentiel utilisant la transformée de Fourier. L'approche consiste à calculer la transformée de Fourier pour ensuite agir de façon directe sur les fréquences de l'image. La modification globale que subit l'image est directement liée aux fréquences que l'on a choisi de modifier et au type de modification que l'on a choisi d'appliquer (atténuation ou seuillage).
Dans l'espace de Fourier on peut donc réaliser des opérations de filtrage moins coûteuses que les opérations de convolution spatiales standard. Le temps de calcul des filtres spatiaux est assez rapide

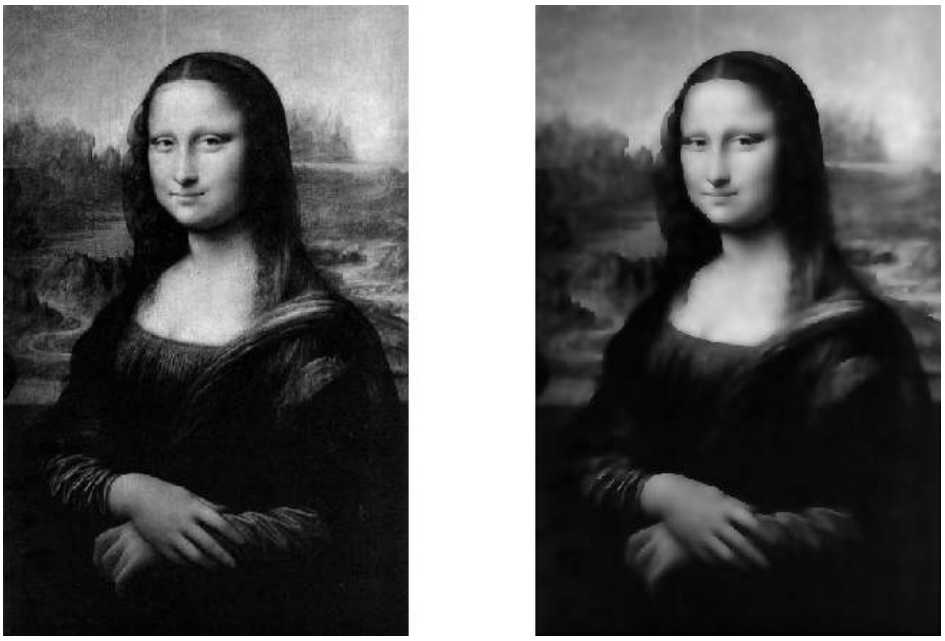


FIG. 4.6 – Image originale et image après filtrage anisotropique avec 10 itérations et $\lambda = 0.25$, $K = 0.1$.

lorsque le noyau de convolution est petit. Lorsqu'on augmente la taille du noyau, on arrive rapidement au point où le temps de calcul dans le domaine des fréquences est plus rapide. Dans ce cas, il est plus simple de passer au domaine fréquentiel et d'effectuer des multiplications au lieu de calculer les convolutions dans le domaine spatial. La procédure de filtrage va impliquer le calcul de la transformée de Fourier de l'image et du noyau puis la multiplication de chaque pixel par son pixel correspondant dans le noyau. L'image filtrée dans le domaine spatial est ensuite récupérée par application de la transformée de Fourier inverse sur le produit précédent.

La transformée de Fourier d'une image

L'analyse de Fourier (ou analyse spectrale) permet de décomposer un signal en une somme de sinus et de cosinus de différentes fréquences. Le nombre de coefficients non nuls dans la transformée dépend de la régularité du signal, c'est-à-dire la continuité du signal et ses dérivées.

Soit une fonction $f(x, y)$ définie sur deux dimensions avec $0 \leq x \leq M-1$, $0 \leq y \leq N-1$, alors on peut définir la transformée de Fourier discrète (DFT[3]) de $f(x, y)$ par la fonction complexe :

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y).\exp[-j(\frac{(2\pi)}{M})ux].\exp[-j(\frac{(2\pi)}{N})vy]$$

u=0, 1, ..., M-1 et v=0, 1, ..., N-1.

$F(u, v)$ est appelé spectre de $f(x, y)$, $|F(u, v)|$ spectre d'amplitude et $\arg(F(u, v))$ le spectre de phase.

[3] Discrete Fourier Transform.

La transformée de Fourier admet une inverse (IDFT[4]) définie par

$$f(x,y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v).\exp[j(\frac{(2\pi)}{M})ux].\exp[j(\frac{(2\pi)}{N})vy]$$

$$\text{x=0, 1, ..., M-1 et y=0, 1 ... N-1.}$$

Pour le calcul de la transformée de Fourier discrète et de son inverse on utilise l'algorithme FFT[5] qui réduit le nombre de calcul pour N points de $2N^2$ à $2N\log_2(N)$.

Spectre d'amplitude et spectre de phase : La transformée de Fourier d'une image est à valeurs complexes. Pour la représenter on utilise le spectre d'amplitude et le spectre de phase qui correspondent respectivement au module et à l'argument des valeurs complexes. Les valeurs que contient le spectre d'amplitude ont en général une dynamique très importante c'est pourquoi on les représente souvent par l'expression $\log(1+|F(u,v)|)$. Le spectre de phase quand à lui contient une information très importante mais très difficilement interprétable. De ce fait, on travaille en général de manière à conserver la phase. Ce spectre est souvent absent des représentations fréquentielles.

Filtrage passe-haut, passe-bas, passe-bande

Comme nous avons pu le voir, une image bidimensionnelle est constituée d'une série de composantes fréquentielles correspondant aux variations des niveaux de gris entre les pixels selon leurs distances. Les variations rapides représentent les composantes de hautes fréquences, et les transitions douces les composantes de basses fréquences. Les amplitudes les plus élevées sont concentrées au milieu du spectre et correspondent aux basses fréquences (niveau général de l'illumination, par exemple), tandis qu'à la périphérie on trouve les hautes fréquences.

Dans le cas du filtrage passe-bas, on doit supprimer ou atténuer les hautes fréquences et conserver les basses et moyennes fréquences. Ce type de filtrage est généralement utilisé pour diminuer le bruit, ou pour estimer l'illumination dans l'image. Les filtres passe-haut, à l'inverse, atténuent les composantes de basse fréquence de l'image et permettent notamment d'accentuer les détails et le contraste. Pour finir, le filtre passe-bande permet d'obtenir une bande de fréquence limitée de l'image. Les résultats de ces différents filtrages fréquentiels sont affichés sur la figure 4.7.

Le filtrage fréquentiel pour la suppression de l'effet de *moiré*

Cet algorithme est présenté ici car il a été ajouté au prétraitement développé dans le cadre du projet TOPVISION pour corriger un défaut très spécifique mais très gênant présent sur un certains nombres de vidéos à traiter : le *moiré*. Le *moiré* est un effet de tramage oblique sur l'image. Il est considéré comme un phénomène d'"aliasing" c'est-à-dire comme une distorsion numérique se produisant lors du processus de numérisation. On peut atténuer ce phénomène par filtrage fréquentiel, car l'effet de texture du tramage oblique apparaît sous la forme de pics de hautes fréquences dans la transformée de Fourier [103]. La suppression de l'effet de *moiré* consiste donc à détecter puis atténuer ces pics dans la transformée de Fourier de l'image (voir Fig. 4.8)

[4] Inverse Discrete Fourier Transform.
[5] Fast Fourier Transform.

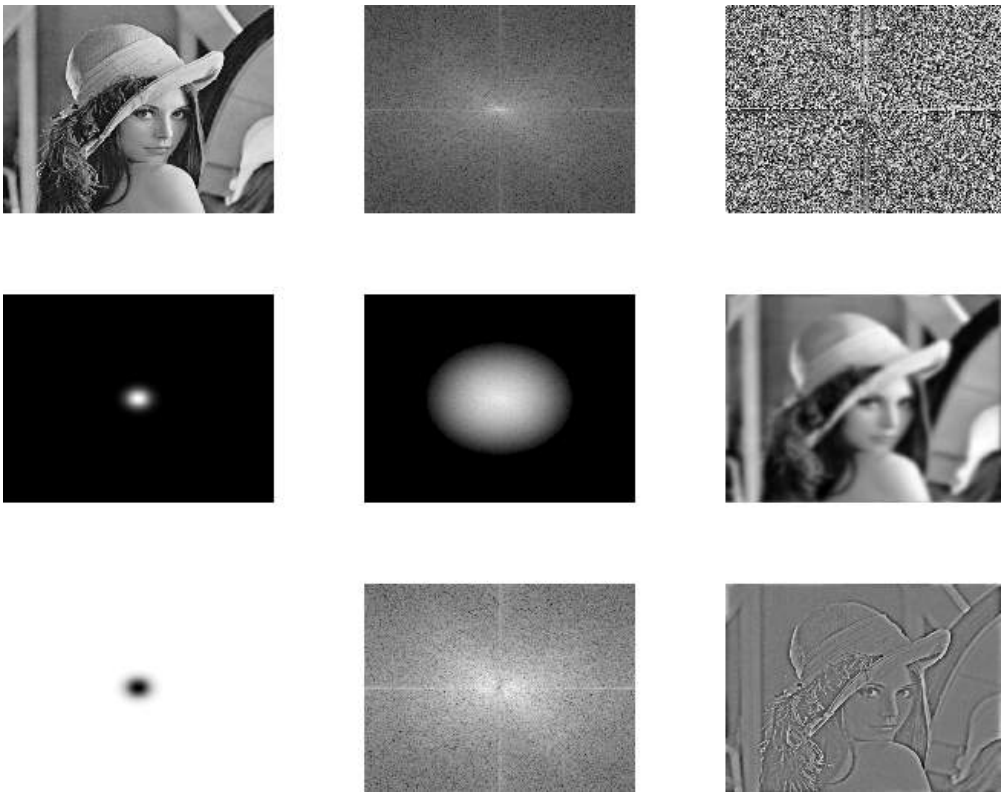


FIG. 4.7 – Filtrage fréquentiel : Image originale, spectre d'amplitude, spectre de phase, filtre passe-bas, spectre d'amplitude après filtrage passe-bas, image basse fréquence (lissée), filtre passe-haut, spectre d'amplitude après filtrage passe-haut, image hautes fréquences (contours).

Détails de l'algorithme de *démoirage* :

- Calcul de la transformée de Fourier,
- Sauvegarde d'une partie du spectre basse fréquence. Il y a toujours un pic important dans les basses fréquences,
- Identification des pics dans la partie non sauvegardée du spectre d'amplitude,
- Remplacement des fréquences anormalement élevées par une valeur plus faible (la valeur médiane dans une zone de voisinage par exemple),
- Reconstruction de l'image à partir de la transformée filtrée.

Le filtrage fréquentiel pour la correction de l'éclairage

Les méthodes de filtrage précédentes permettent d'extraire certaines fréquences du spectre de l'image de départ. Cette approche laisse supposer que les fréquences que l'on garde et celles que l'on supprime se sont superposées par addition. On parle alors de perturbation additive. Cependant la situation de perturbation multiplicative est beaucoup plus courante puisque l'image d'une scène quelconque peut se décomposer selon la réflectance et l'illumination, ces deux composantes étant combinées par multiplication : l'intensité lumineuse perçue en un point est le résultat du produit du facteur de

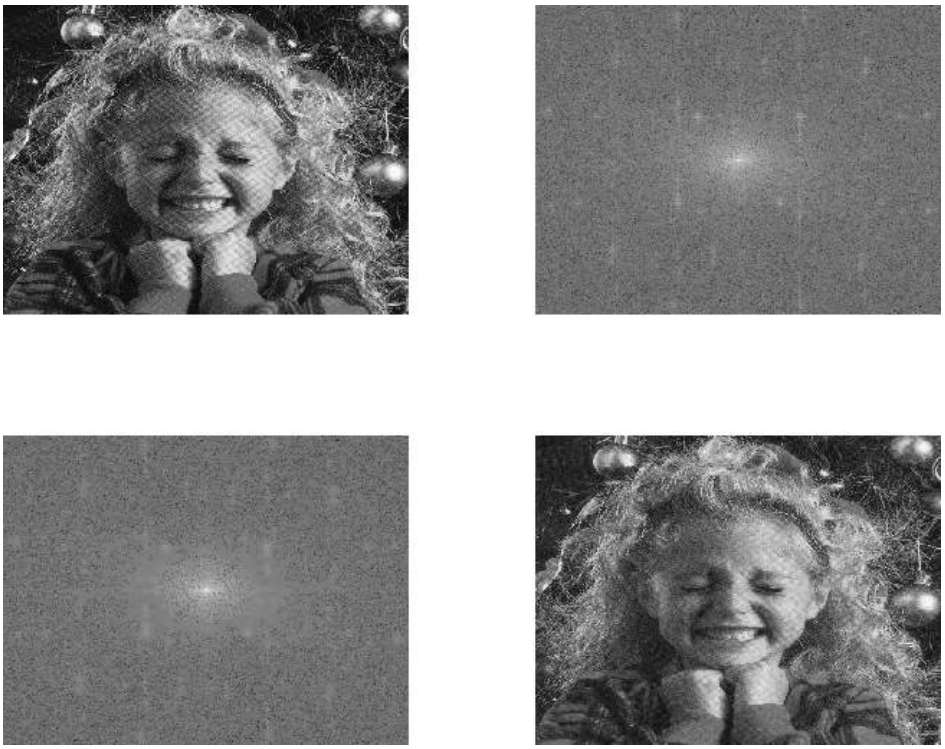


FIG. 4.8 – Suppression de l'effet de *moiré* : Image originale, spectre d'amplitude, spectre d'amplitude filtré par atténuation des pics de fréquences, image démoirée.

réflectance de ce point (facteur qui est propre au point considéré dans la scène) par l'intensité d'illumination en ce point. Si cette illumination n'est pas uniforme les filtres précédents sont inutiles car non adaptés. On utilise dans ce cas le filtrage homomorphique [104].

En considérant le modèle d'illumination-réflectance, on suppose l'image comme une fonction constituée du produit de l'illumination par la réflectance (éq. 4.12).

$$g(x,y) = i(x,y).r(x,y) \tag{4.12}$$

où $g(x,y)$ est l'image provenant de la caméra, $i(x,y)$ le facteur multiplicatif d'illumination, et $r(x,y)$ la fonction de réflectance, c'est-à-dire l'image idéale en l'absence de nuance d'illumination.

Le filtrage homomorphique est un filtrage fréquentiel permettant de supprimer les effets de non uniformité de l'éclairage dans les images [24]. Considérant ce modèle, on peut supposer que la composante d'illumination subit des variations faibles, et qu'elle représente donc les basses fréquences sur la transformée de

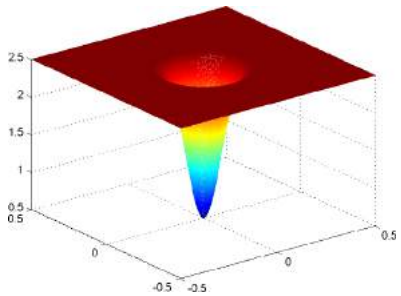


FIG. 4.9 – Réponse fréquentielle du filtre passe-haut choisi pour le filtrage homomorphique dans le cas des images sous-marines.

Fourier alors que la composante de réflectance subit des variations plus importantes qui sont associées aux moyennes et hautes fréquences. L'objectif de ce filtrage est donc de supprimer les basses fréquences pour atténuer les irrégularités d'illumination et accentuer les hautes fréquences de manière à augmenter les contrastes (voir Fig. 4.10). L'algorithme peut se décomposer de la manière suivante :

- Séparation des composantes d'illumination et de réflectance en prenant le log de l'image (Eq.4.13). Le logarithme convertit l'effet multiplicatif en un effet additif. Le logarithme n'étant pas défini en zéro on ajoute une petite valeur à l'ensemble des valeurs si nécessaire .

$$g(x,y) = \ln\big(f(x,y)\big) = \ln\big(i(x,y).r(x,y)\big) = \ln\big(i(x,y)\big) + \ln\big(r(x,y)\big) \tag{4.13}$$

- Calcul de la transformée de Fourier de la log-image (Eq.4.14)

$$G(w_x, w_y) = I(w_x, w_y) + R(w_x, w_y) \tag{4.14}$$

- Multiplication du spectre d'amplitude par un filtre passe-haut. Ce filtre atténue les basses fréquences (illumination) et accentue les hautes fréquences (réflectance) c'est-à-dire les contours des objets dans l'image (Eq.4.16).
  La réponse fréquentielle du filtre est affichée sur la figure 4.9. On pourra voir les effets de ce filtre sur des images sous-marines dans la partie suivante présentant la chaîne de prétraitement complète. Ce filtre à été spécifiquement développé pour être indépendant au "type" d'image.

$$S(w_x, w_y) = H(w_x, w_y).I(w_x, w_y) + H(w_x, w_y).R(w_x, w_y) \tag{4.15}$$

$$\text{avec,} \quad H(w_x, w_y) = (r_H - r_L).(1 - \exp(-(\frac{w_x^2 + w_y^2}{2\delta_w^2}))) + r_L \tag{4.16}$$

  où r_H et r_L sont les coefficients maximum et minimum et δ_w un facteur permettant de contrôler la fréquence de coupure.
- Calcul de la transformée de Fourier inverse pour revenir dans le domaine spatial et passage à l'exponentielle pour récupérer l'image filtrée.

Conclusion

En ce qui concerne les images sous-marines, on peut noter l'utilisation très intéressante du filtrage fréquentiel pour l'uniformisation de l'éclairage, le rehaussement de contraste, et également le démoirage (ou détramage d'image). Ces algorithmes sont très performants et donnent des résultats généralement plus intéressants que le filtrage spatial.
Pour conclure maintenant sur l'analyse de Fourier, on peut dire qu'elle bien adaptée à l'analyse de comportements harmoniques surtout apparents dans les zones bien texturées des images, mais elle n'est pas appropriée à l'analyse de formes bien localisées spatialement. Pour avoir la possibilité de conserver une analyse spectrale tout en obtenant un pouvoir de localisation spatiale il faut adopter une approche espace fréquence en choisissant des fonctions à la fois oscillantes et de support limité. Nous allons voir ce type de fonction dans la partie suivante.

4.2.3 Le filtrage multirésolution

La transformée de Fourier analyse le "contenu fréquentiel" d'un signal. Malheureusement elle fait perdre toute notion de localisation. C'est une représentation globale du signal. La transformée de

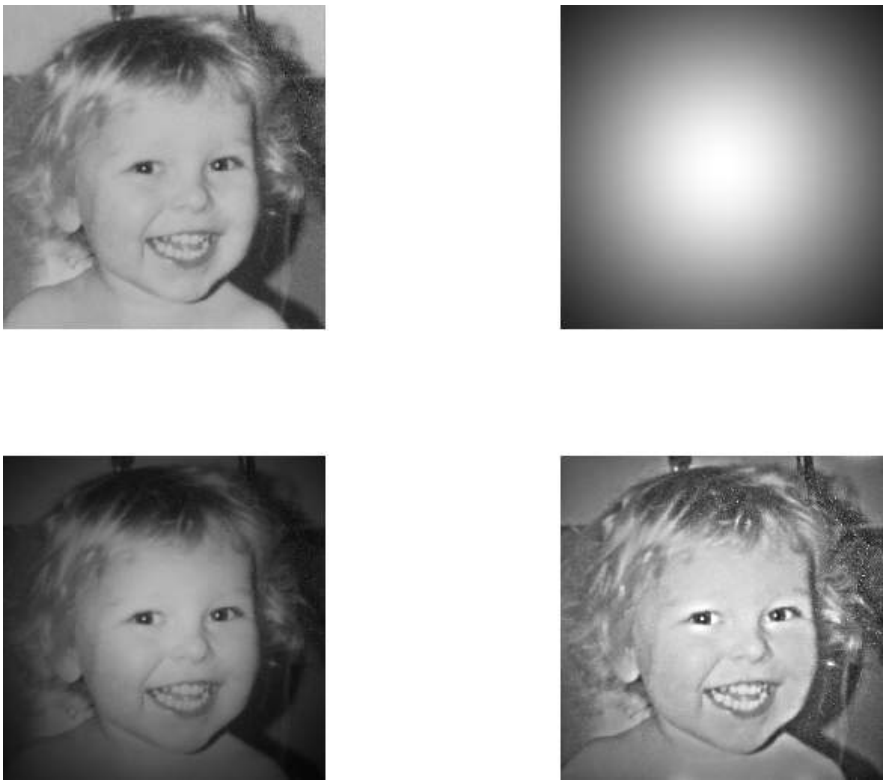


FIG. 4.10 – Filtrage homomorphique : Image originale, modèle d'illumination gaussienne, image avec illumination gaussienne, image corrigée par un filtrage homomorphique.

Fourier peut être vue comme une représentation à base de sinusoïdes. Ces sinusoïdes sont très bien localisées en fréquence, mais pas en espace, car leur support est infini.

La transformée en ondelettes.

La transformée en ondelettes [105] remplace la sinusoïde de la transformée de Fourier par une famille de translations et dilatations d'une même fonction, l'ondelette. La transformée en ondelettes est une méthode de traitement du signal adaptée aux signaux non périodiques. Elle permet une analyse temps-fréquence que ne permet pas une analyse par Fourier. Les paramètres de translation et de dilatation sont les deux arguments de la transformée en ondelettes (voir Fig. 4.13). Cette analyse sera appelée espace-fréquence pour les images.

Le principe de l'algorithme consiste à diviser en quatre l'image à chaque itération : trois blocs concernant les détails de l'image (les hautes fréquences), et le quatrième correspondant aux informations les plus importantes pour l'oeil (les basses fréquences), qui sert de base pour la prochaine itération. Les algorithmes d'ondelettes ont deux composantes, une fonction d'ondelette et une fonction d'échelle (dénommées parfois filtre passe-haut et filtre passe-bas). Il existe un algorithme rapide pour le calcul de la transformée en ondelettes (FWT pour "Fast Wavelet Transform"). Celui-ci effectue une transformée discrète par des convolutions circulaires calculées par FFT (voir Fig. 4.11, 4.12).

Avec,

- $2 \downarrow 1$ l'opérateur de sous échantillonnage en colonne (conservation des colonnes d'indice pair).
- $1 \downarrow 2$ l'opérateur de sous échantillonnage en ligne.
- $1 \uparrow 2$ l'opérateur de sur échantillonnage en colonne (insertion de zéros).
- $2 \uparrow 1$ l'opérateur de sur échantillonnage en ligne
- HiD l'opérateur de convolution par un filtre passe-haut.
- LoD l'opérateur de convolution par un filtre passe-bas.

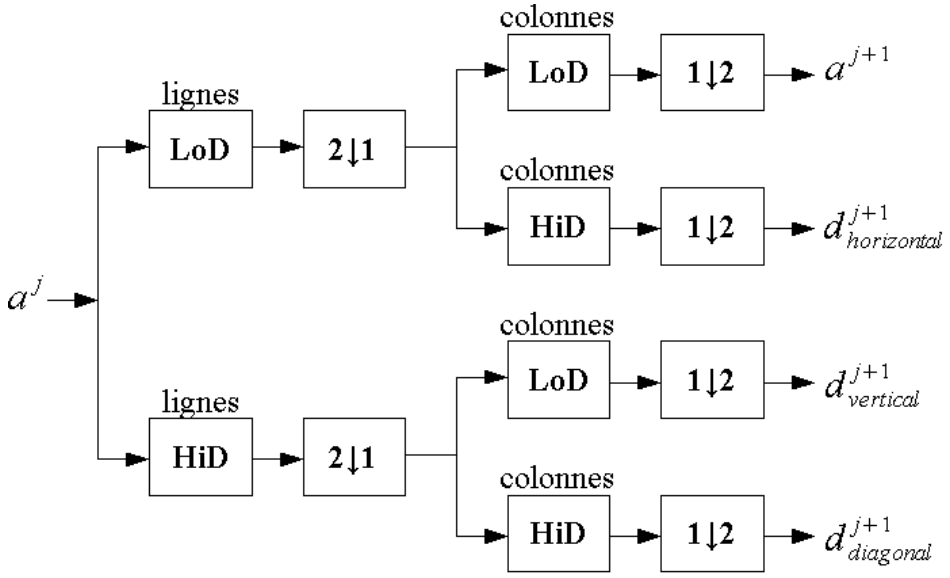


FIG. 4.11 – Principe de décomposition en ondelettes.

Remarque sur l'invariance en translation : La transformée discrète souffre d'un inconvénient connu : ce n'est pas une transformée invariante par translation dans le temps. Ce qui signifie que même dans le cas d'une extension périodique d'un signal la transformée d'une version translatée n'est pas en général la version translatée de la transformée. Pour restaurer l'invariance par translation, l'idée est d'effectuer la moyenne des DWT[6] de tous les signaux translatés-périodisés issus du signal. Cette propriété est intéressante dans plusieurs domaines d'applications notamment le débruitage, car on fait la moyenne de plusieurs signaux débruités ce qui atténue les artefacts. Le débruitage utilisant l'invariance en translation donne les meilleurs résultats mais il est très long en temps de calcul.

Remarque sur les filtres en quadrature : Les filtres miroirs en quadrature (QMF) sont utilisés pour reconstruire des signaux à partir de composantes basses fréquences et hautes fréquences, bien que ces composantes ne respectent pas le théorème de Shannon : le sous-échantillonnage induit un repliement spectral. Toutefois, la forme complémentaire des filtres passe-bas et passe-haut garantit une reconstruction exacte.

[6] Discrete Wavelet Transform

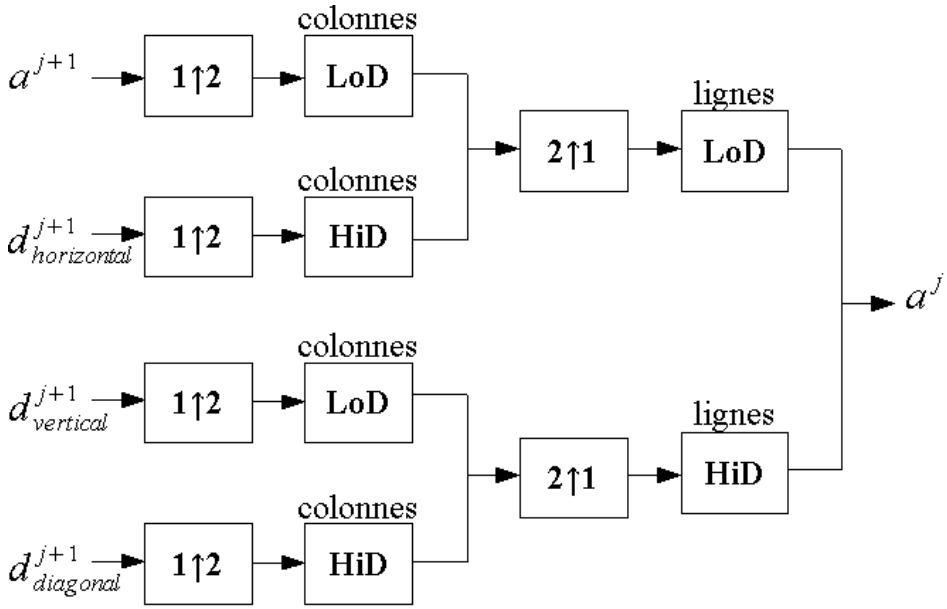


FIG. 4.12 – Principe de reconstruction à partir des coefficients d'ondelettes.

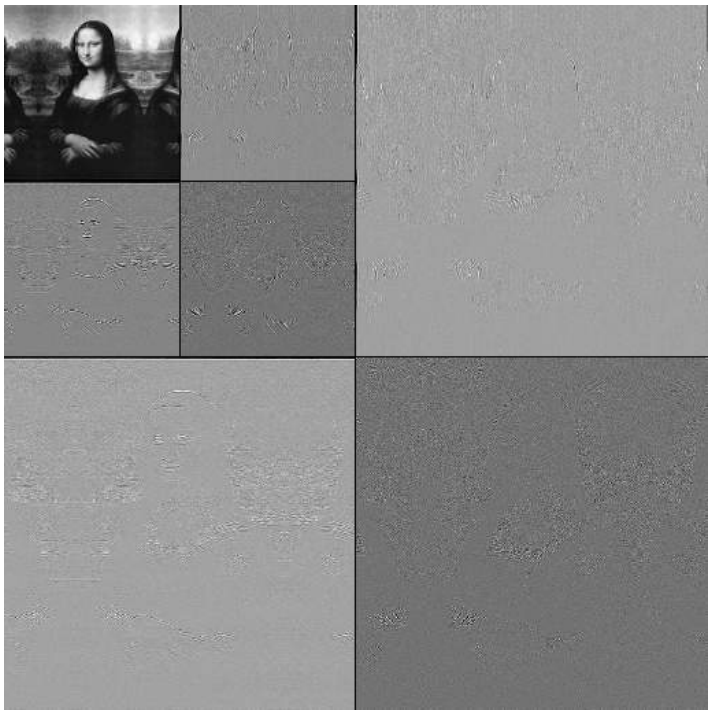


FIG. 4.13 – Exemple d'image décomposée en ondelettes sur 2 niveaux.

Débruitage par ondelettes

Très souvent l'essentiel de l'information d'un signal se situe dans ses singularités et ses structures irrégulières. Pour le traitement d'images, par exemple, les contours sont caractérisés par la discontinuité

de l'intensité. La transformée en ondelettes permet de réaliser une analyse des structures locales d'un signal avec un zoom qui dépend de l'échelle considérée.

Nous avons vu qu'il était possible de réaliser une décomposition en ondelettes d'une image puis de reconstruire cette image à partir de ses coefficients d'ondelettes. L'intérêt de cette technique est donc de modifier ces coefficients pour améliorer l'image initiale.

Les coefficients d'ondelettes marquent les discontinuités qui interviennent dans l'image. Lors du débruitage (ou "denoising") on ne garde que les coefficients les plus grands et on met les autres à zéro puis on reconstruit l'image. Le bruit correspond en général à des détails faibles donc il est éliminé par ce seuillage des coefficients d'ondelettes [106][107][108]. Nous obtenons alors une image plus "lisse" donc débruitée.

Détails de l'algorithme de débruitage :
- Redimensionnement de l'image de manière à obtenir une taille en puissance de 2,
- Décomposition rapide en ondelettes,
- Estimation du seuil sur la première décomposition en ondelettes,
- Seuillage des coefficients (seuillage strict ou souple),
- Reconstruction de l'image à partir des coefficients seuillés.

L'estimation du seuil peut se faire via le seuillage robuste de Donoho et Johnstone[108]. Si l'on considère un bruit blanc gaussien additif de variance σ^2, il est en effet montré que l'amplitude du bruit a une très forte probabilité d'être juste en dessous du seuil $T = \sigma\sqrt{2.\log(N)}$ où N est le nombre d'échantillons. Il suffit donc pour estimer le seuil de calculer la variance du bruit, et on utilise pour cela l'estimateur robuste présenté par l'équation 4.17 sur les coefficients d'ondelettes à l'échelle la plus fine.

$$\sigma^2 = (M/0.6745)^2. \tag{4.17}$$

Cette équation peut être démontrée grâce à la proposition suivante : si M est la valeur de la médiane des valeurs absolues de N variables aléatoires gaussiennes indépendantes de moyenne nulle et d'écart type σ, alors on a $E\{M\} = 0.6745\sigma$.

Le seuillage strict est celui qui est le plus "intuitif". On se fixe un seuil T>0. On ne conserve que les coefficients d'ondelettes supérieurs à T et on met à zéro les autres (Eq. 4.18) :

$$f(x) = \begin{cases} 0 & \text{si } |x| < T \\ x & \text{si } |x| \geq T \end{cases} \tag{4.18}$$

Dans le cas du seuillage souple, on met toujours à zéro les coefficients inférieurs à un seuil T. Par contre, pour ceux supérieurs à T, on atténue l'amplitude des coefficients par la valeur du seuil afin de s'assurer d'avoir enlevé l'effet du bruit même pour les forts coefficients (Eq. 4.19).

$$f(x) = \begin{cases} 0 & \text{si } |x| < T \\ x - sign(x) * T & \text{si } |x| \geq T \end{cases} \tag{4.19}$$

Pour éviter les artefacts on peut effectuer le débruitage sur plusieurs ondelettes (la famille des ondelettes de Daubechies par exemple) puis moyenner les résultats. Les résultats sont meilleurs en

général mais le temps de calcul est de ce fait plus important.

De nombreuses ondelettes différentes existent avec chacune des propriétés particulières. On peut citer par exemple l'ondelette de Haar, les ondelettes de Daubechies, les Symlets..., le choix de l'ondelette dépend beaucoup du but recherché. Dans notre cas le choix de l'ondelette a été fait de manière empirique. Nous nous sommes davantage attardés sur le choix de la méthode de seuillage.

Quantification des résultats et PNSR : Le PSNR acronyme de Peak Signal to Noise Ratio est une mesure de distorsion utilisée en image numérique tout particulièrement en compression d'image. Il s'agit de quantifier la performance des codeurs en mesurant la qualité de reconstruction de l'image compressée par rapport à l'image originale.

Le PSNR est défini comme suit :

$$PSNR = 10 \cdot \log_{10} \frac{d^2}{EQM} \tag{4.20}$$

où d est la dynamique du signal. Dans le cas standard d'une image où un pixel est codé sur 8 bits par composante, d = 255, et EQM est l'erreur quadratique moyenne et est définie pour 2 images I et J de taille $m \times n$ par :

$$EQM = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} ||I(i,j) - J(i,j)||^2 \tag{4.21}$$

Les valeurs typiques de PSNR pour des images de bonne qualité varient entre 30 et 40 dB.

A noter que si le PSNR est utile pour mesurer la proximité de l'image compressée par rapport à l'originale au niveau du signal, il ne prend pas en compte la qualité visuelle de reconstruction et ne peut être considéré comme une mesure objective de la qualité visuelle d'une image. La figure 4.14 présente les résultats de PSNR pour chaque méthode de débruitage.

Le débruitage par ondelettes : la méthode de Selesnick [109]

Le filtrage multirésolution dispose d'avantages importants pour le débruitage. La méthode de débruitage de Selesnick utilise une base d' ondelettes orthogonales presque symétriques avec un seuillage exploitant des dépendances inter-échelle [109][110]. Cette méthode de débruitage ne suppose pas l'indépendance des coefficients comme les méthodes standard type Donoho [111], car les coefficients d'ondelettes dans les images naturelles montrent des dépendances importantes.

L'algorithme peut se décomposer comme suit :
D'abord on définit la fonction $(g)_+$ telle que :

$$(g)_+ = \begin{cases} 0 & \text{si } g < 0 \\ g & \text{sinon} \end{cases}$$

- Décomposition multirésolution de l'image corrompue par un bruit blanc gaussien en utilisant la transformée en ondelettes avec la base Farras (banc de filtres quasi symétrique pour la reconstruction 2D orthogonale parfaite).
- Estimation de la variance du bruit σ_n^2 sur le "subband"[7] HH utilisant l'équation 4.17.
- Pour chaque "subband" de chaque niveau de décomposition excepté pour l'image basses fréquences (passe-bas).

  a) Calcul de la variance du signal par l'équation 4.22.

$$\sigma = \sqrt{(\sigma_y^2 - \sigma_n^2)_+}, \quad \text{ou } \sigma_y^2 = \frac{1}{M} \sum_{y_i \in N(k)} y_i^2 \tag{4.22}$$

  avec M la taille du voisinage N(k).

  b) Modification des coefficients d'ondelettes suivant l'équation 4.23.

$$y_1 = \frac{(\sqrt{y_1^2 + y_2^2} - \frac{\sqrt{3}\sigma_n^2}{\sigma})_+}{\sqrt{y_1^2 + y_2^2}} . y_1 \tag{4.23}$$

  où y_1 représente le fils et y_2 le coefficient parent.

- Inversion de la décomposition multirésolution pour reconstruire l'image débruitée.

Un deuxième débruitage par ondelettes : la méthode de Kovesi [112]

De nombreuses méthodes temps fréquence sont utilisées pour débruiter. Cependant ces méthodes présentent deux inconvénients majeurs : la non invariance en translation et non la préservation de la phase qui déforment les contours.

Kovesi propose un débruitage multi-échelle basé sur les ondelettes logarithmiques de Gabor complexes et non orthogonales. Il utilise des ondelettes qui permettent de calculer l'amplitude et la phase du signal à plusieurs échelles. Ensuite seuls les coefficients de l'amplitude sont filtrés par rapport à une estimation de la contribution du bruit à chaque échelle. La phase est donc préservée. Le bruit est assumé blanc gaussien. Les coefficients correspondants au bruit sont en majorité aux premières échelles.

Conclusion

Les résultats de l'ensemble des méthodes de débruitage présentées sont rassemblés dans la figure 4.14. Pour chaque méthode on a calculé le PNSR à partir de l'image originale. En ce qui concerne le temps de calcul, toutes les méthodes présentées précédemment sont à peu près équivalentes et très rapides excepté le débruitage invariant en translation, le débruitage avec moyennage et le débruitage par la méthode de Kovesi. Pour conclure, l'utilisation des ondelettes est aujourd'hui quasiment incontournable pour les problèmes de débruitage et de compression d'image. Nous avons donc opté pour ce type de méthode pour supprimer le bruit toujours très présent dans les images sous-marines. Les

[7] Ce terme pourrait être traduit en français par sous-image. Une sous-image pouvant être un des niveaux de détails : horizontal (LH), vertical (HL) ou diagonal (HH) ou le niveau d'approximation (LL).

FIG. 4.14 – Débruitage par seuillage des coefficients d'ondelettes sur 4 échelles, de gauche à droite et de haut en bas : Image originale, image bruitée par un bruit blanc gaussien, débruitage invariant en translation et seuillage strict, débruitage invariant en translation et seuillage souple, débruitage standard et seuillage strict, débruitage standard et seuillage souple, débruitage standard et seuillage multi-échelle, débruitage par la méthode de Selesnick, débruitage standard et seuillage strict avec moyennage sur 10 ondelettes (aussi appelé "à diversité augmentée"), débruitage standard et seuillage souple avec moyennage sur 10 ondelettes, débruitage par la méthode de Kovesi avec seuillage strict, débruitage par la méthode de Kovesi avec seuillage souple. Pour l'ensemble des méthodes, on a utilisé l'ondelette de 'Daubechies 8' excepté pour le débruitage par moyennage où on a utilisé les ondelettes de Daubechies 2 à 20 et pour les méthodes de Selesnick et Kovesi qui disposent de leur propre base d'ondelettes .

deux méthodes que nous avons présentées sont automatiques c'est-à-dire que les paramètres sont auto-ajustés avec un avantage énorme en temps de calcul pour la méthode de Selesnick. Éventuellement pour compléter cette étude on pourra investiguer les nouveaux algorithmes basés sur les curvelets qui donnent aussi de bons résultats [113] [114].

4.2.4 La correction des couleurs

En imagerie sous-marine les couleurs sont rarement bien équilibrées. On remarque souvent une dominante bleue ou verte et le plan rouge est souvent très atténué. Dans cette partie nous allons présenter deux algorithmes de correction des couleurs car dans le prétraitement que nous avons développé

FIG. 4.15 – Illustrations des deux algorithmes de correction des couleurs : Image originale, couleurs corrigées par compensation de l'absorption, couleurs corrigées par égalisation des moyennes. Les coefficients appliqués pour la compensation de l'absorption $K_R = 0.4, K_G = 0.06, K_B = 0.05, z = 1$ et $K_R = 0.8, K_G = 0.1, K_B = 0.8, z = 1$.

nous avons voulu être le plus complet possible donc nous avons essayé de corriger aussi les couleurs qui sont fortement modifiées. Les méthodes que nous présentons sont des méthodes simples permettant à la base d'améliorer surtout la qualité visuelle des images. On a cependant montré par la suite que cette correction des couleurs permet aussi d'améliorer légèrement les contrastes et par conséquent la détection de contour.

La compensation de l'absorption

L'atténuation de la lumière sous l'eau est essentiellement due à l'absorption, la diffusion ayant une moindre importance. Cette atténuation est fonction de la longueur d'onde et des caractéristiques physiques et chimiques de l'eau de mer.

Pour un éclairement naturel de la mer par le soleil, on estime que l'intensité lumineuse à une longueur d'onde donnée varie en fonction de la profondeur suivant la loi :

$$I_\lambda(z) = I_{0,\lambda}.\exp(-k_\lambda.z)$$

où $I_{0,\lambda}$ est l'intensité de la lumière à la surface pour la longueur d'onde λ, k_λ est le coefficient d'atténuation verticale de l'eau de mer (Fig. 4.16) et z la distance parcourue par la lumière dans l'eau.

La correction des couleurs par compensation de l'absorption se fait indépendamment pour chaque plan RGB utilisant les formules :

$$\begin{cases} R = R/e^{-K_R.z} \\ G = G/e^{-K_G.z} \\ B = B/e^{-K_B.z} \end{cases} \quad (4.24)$$

Dans le cas de l'image présentée on peut par exemple utiliser les coefficients d'atténuation suivant $K_R = 0.40$, $K_G = 0.06$, $K_B = 0.05$, et $z = 1$ (eau "claire").

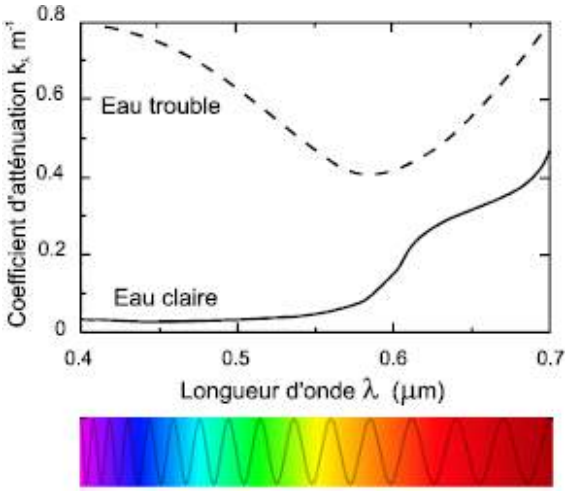


FIG. 4.16 – Coefficients d'atténuation approximatifs en fonction de la longueur d'onde pour une eau claire et pour une eau trouble. Source [115].

A noter que ce traitement avec les coefficients proposés ne fonctionne pas toujours correctement, dans certains cas son action est très bonne et permet de retrouver les vraies couleurs notamment en restaurant le rouge très absorbé (voir correction de la première image Fig. 4.15). Au contraire, pour d'autres images la compensation ne fonctionne pas avec les coefficient "standard" et augmente encore la dominante de couleur (voir correction de la deuxième image Fig. 4.15). Dans ce cas par exemple il faut appliquer d'autres coefficients. Il serait donc nécessaire de disposer d'une part de davantage de courbes d'atténuation et d'autre part d'une estimation automatique des coefficient appropriés basée sur une estimation de la couleur dominante par exemple ou sur une optimisation des résultats de correction. On pourrait bien évidemment obtenir de meilleurs résultats de correction disposant de connaissances *a priori* sur la qualité de l'eau.

Égalisation des moyennes

Une deuxième méthode de compensation colorimétrique consiste à supprimer les dominantes de couleurs sans tenir compte du phénomène d'absorption de la lumière dans la mer. Cette méthode consiste simplement à recentrer les moyennes des trois plans rouge, vert, bleu de manière à supprimer ou atténuer la couleur prépondérante par rapport aux deux autres. On parle pour ce traitement de décalage d'histogramme : On calcule la moyenne des trois plans RGB puis on décale l'histogramme de chaque plan de manière à ce que la moyenne de chaque plan RGB soit égale à la moyenne des trois plans précédemment calculés (si certaines valeurs débordent de l'intervalle autorisé, elles sont évidemment saturées).

L'algorithme proposé est une simple translation linéaire de trois histogrammes RGB de manière à égaliser leurs moyennes. Il peut donc être appliqué sur toutes les images sous-marines sans avoir à estimer la couleur dominante (Fig. 4.15) et il permet de nettement améliorer le rendu visuel au prix malheureusement de la création de fausses couleurs. Cet algorithme n'est pas à proprement parlé un algorithme de correction des couleurs on parlera dans ce cas contrairement au cas précédent d'amélioration des couleurs.

Conclusion

Le filtrage des couleurs est une partie moins importante dans le prétraitement des images sous-marines. L'étude d'éventuelles corrections colorimétriques a été faite dans un souci de complétude. L'utilisation de l'une ou l'autre des méthodes de correction doit dépendre de l'algorithme utilisé *a posteriori*. En effet la création de fausses couleurs par le deuxième algorithme n'est pas préjudiciable pour la détection de contour bien au contraire d'ailleurs (nous le démontrerons dans le chapitre suivant) mais elle le serait pour une détection par la couleur. Comme pour les méthodes précédentes, le choix de la méthode se fera dans l'optique d'un traitement rapide et automatique. A noter que certaines techniques beaucoup plus complexes donnent de bien meilleurs résultats que celles présentées cependant elles nécessitent des informations *a priori* importantes notamment la thèse de Ahlen [2] et [116][117], ou un temps de calcul trop excessif [57][118].

4.3 Le prétraitement des images sous-marines développé

Comme nous l'avons expliqué dans la première partie de ce chapitre, nous avons opté pour une méthode d'amélioration c'est-à-dire la correction un à un des défauts dans l'image. Ces défauts étant le bruit important, l'illumination non uniforme, des contrastes faibles et des couleurs atténuées... Dans la partie précédente, nous avons détaillé un certain nombre d'algorithmes de filtrage permettant la correction de ces problèmes. Le choix de l'algorithme a été en grande partie déterminé par le contexte de l'étude et les données à traiter. Sur ces données très hétérogènes, aucun profil de bruit particulier ne pouvait être extrait, des problèmes diverses étaient présents et aucune connaissance *a priori* sur les conditions de prise de vue n'était disponible. De plus des contraintes de rapidité, et d'automatisme des traitements étaient demandées. Puisqu'à part la restauration, aucun algorithme global et spécialisé aux images sous-marines n'existait, nous avons utilisé des algorithmes conventionnels en traitement d'images. Dans cette partie nous allons présenter l'algorithme d'amélioration développé à partir de ces méthodes. Cet algorithme qui ne nécessite ni paramétrage manuel ni information *a priori*, permet d'atténuer les défauts précédemment cités et d'améliorer la qualité des images. L'éclairage, le bruit, les contrastes puis les couleurs sont corrigés séquentiellement. Le développement de cet algorithme a fait l'objet d'une publication en revue [119].

4.3.1 Un enchaînement de correction

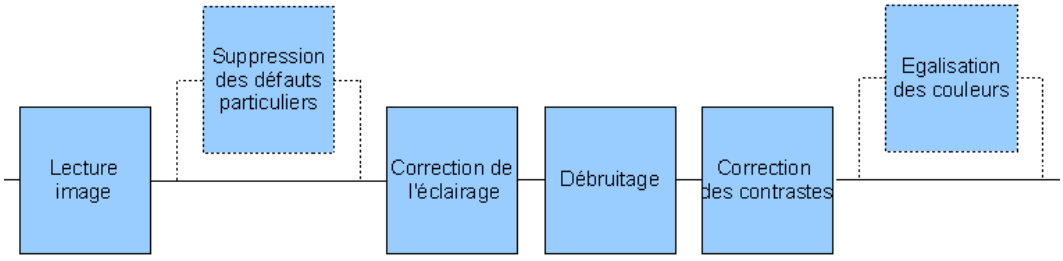


FIG. 4.17 – Organigramme de la méthode de prétraitement développée. Deux étapes sont facultatives, la suppression des défauts particuliers (comme par exemple l'effet de moiré, l'effet de bloc, ou la surimpression de texte), et la correction des couleurs appliquée uniquement sur les images couleur.

La chaîne de prétraitement permet d'atténuer les défauts les plus gênants et d'améliorer de façon significative la qualité des images à la fois d'un point de vue visuel (critère subjectif mais toujours important) et aussi bien évidemment, d'un point de vue plus objectif, l'amélioration des résultats de segmentation. Elle sera d'ailleurs quantifiée, validée sur un critère basé sur le gradient. Cette chaîne de prétraitement est aussi automatique, seule la première étape concernant la suppression des éventuels défauts très spécifiques est sélectionnée par l'utilisateur. L'algorithme développé peut être décomposé en 9 étapes (voir Fig. 4.18) :

1. **Suppression des défauts particuliers.** La première étape de cette chaîne de prétraitement consiste à supprimer des défauts très spécifiques qui dégradent considérablement nos résultats, il s'agit soit d'un effet de trame, soit d'un effet de compression.
   Cas du *moiré* : le *moiré* n'est pas une dégradation due au milieu sous-marin mais il dégrade les traitements suivants lorsqu'il est présent puisque les algorithmes de rehaussement de contraste amplifient gravement le phénomène. De nombreuses vidéos TOPVISION qui constituent la base de test pour les algorithmes de prétraitement souffrent de ce problème. L'effet de *moiré* est supprimé par filtrage fréquentiel (voir chapitre 4.2.2). Ce traitement pourrait être combiné avec le filtrage homomorphique puisque ce sont tout deux des filtrages fréquentiels, cependant nous avons tenu à garder cette étape indépendante du reste de la chaîne puisque seules les vidéos TOPVISION souffrent de ce phénomène.
   Dans l'autre cas : l'effet de "bloc" dû à la compression d'image est corrigé par un filtrage médian à la place de l'algorithme de démoirage.
   Dans la plupart des cas, aucun de ces problèmes n'apparaît et cette étape est donc négligée.
2. **Redimensionnement et extension symétrique de l'image.** Cette transformation permet d'obtenir une image carrée ayant une taille en puissance de deux. Le redimensionnement prévient des effets de bords et accélère le calcul de la transformée de Fourier et de la transformée en ondelettes.
3. **Conversion de l'image RGB dans l'espace colorimétrique YCbCr.** Cette deuxième étape permet d'accélérer les traitements ultérieurs en réduisant le prétraitement au plan de luminance (voir chapitre 2).
4. **Filtrage homomorphique.** Ce traitement supprime les défauts de non uniformité de l'illumination et rehausse les contrastes dans l'image. D'autres algorithmes existent pour corriger ces défauts mais le filtrage homomorphique a été préféré pour sa rapidité de calcul et pour son double effet de renforcement des contours. Il fonctionne avec des paramètres constants pré-calculés de façon empirique (voir filtre Fig. 4.9).
5. **Débruitage par ondelettes par la méthode Selesnick.** Cet algorithme utilise un seuillage multi-échelle des coefficients d'ondelettes pour atténuer le bruit d'acquisition. Le bruit est toujours présent dans les images naturelles et il est nettement amplifié par l'étape précédente (voir chapitre 4.2.3). Cette méthode de débruitage évalue automatiquement ses paramètres de seuillage. Elle offre également un très bon compromis qualité/rapidité.
6. **Filtrage anisotropique.** Le filtrage anisotropique est utilisé pour compléter le débruitage et le rehaussement de contraste. En effet il permet de simplifier graduellement l'image et de se débarrasser d'oscillations indésirables au prix d'une légère dégradation de la structure. L'image est lissée dans les zones homogènes mais les contours sont préservés et même rehaussés (voir

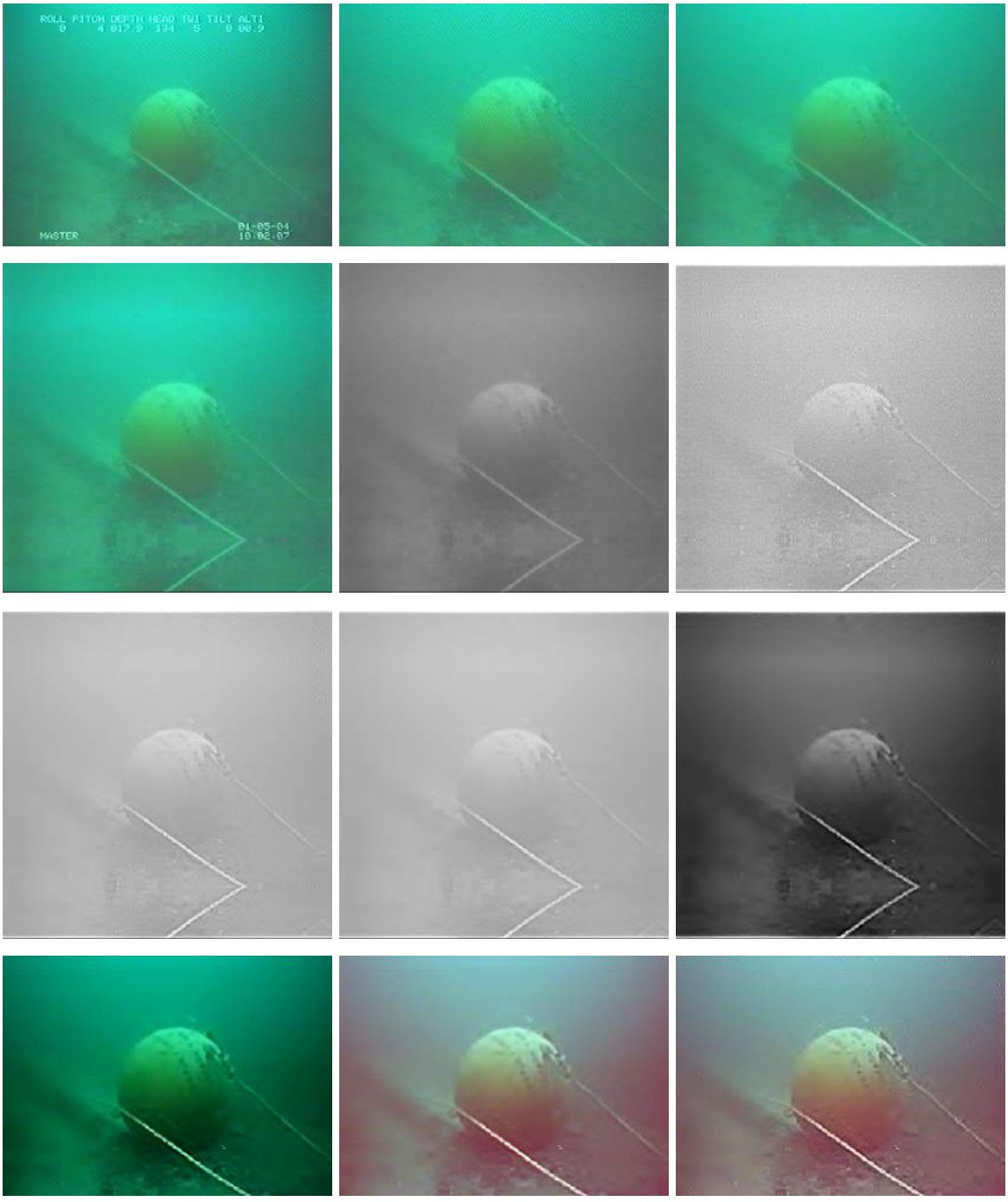


FIG. 4.18 – Le prétraitement d'une image souffrant d'un effet de moiré en 11 étapes (la dernière image correspond au résultat du prétraitement sans le filtrage anistropique) : 1) Image originale 2) Image sans texte 3) Suppression du *moiré* 4) Redimentionnement et extension symétrique 5) Conversion d'espace couleur RGB à YCbCr 6) Filtrage homomorphique 7) Débruitage par ondelettes 8) Filtrage anisotropique 9) Correction des contrastes et expansion de la dynamique 10) Conversion dans l'espace RGB et découpage de l'extension symétrique 11) Égalisation des moyennes de couleurs.

chapitre 4.2.1). Ce filtrage a lui aussi des paramètres constants pré-ajustés. Il privilégie les forts contrastes, et de manière à être rapide il est utilisé sur peu d'itération avec une diffusion rapide.

7. **Correction des contrastes.** Cette étape comprend la suppression des éventuelles valeurs aberrantes c'est-à-dire en dehors de l'intervalle des valeurs d'intensités autorisées apparu lors du filtrage homomorphique, et ensuite une expansion de la dynamique de manière à utiliser l'échelle complète des niveaux de gris disponibles.
8. **Conversion en RGB et découpage de l'extension symétrique.** Le plan de luminance ayant été traité, l'image couleur est retrouvée à partir des composantes chromatiques Cb, Cr puis convertie en RGB, ensuite on supprime l'extension symétrique ajoutée dans l'étape 2.
9. **Égalisation des couleurs.** Enfin on traite les couleurs. La suppression de dominante de couleurs se fait par égalisation des moyennes RGB. Cette étape très simple améliore la qualité visuelle de l'image ainsi que les contrastes dans l'image. A noter que cet algorithme ne restitue pas la fidélité des couleurs et génère des fausses couleurs. Mais comme nous l'avons dit précédemment, ceci n'est pas préjudiciable pour les traitements que nous effectuerons par la suite.

L'ordre dans l'enchaînement des algorithmes est bien évidemment très important notamment pour les trois algorithmes les plus essentiels (étape 4, 5 et 6). Le filtrage homomorphique est pratiqué en premier car il amplifie de manière significative le bruit dans les images, il doit donc donc être suivi d'un débruitage. Quant au filtrage anisotropique, il est utilisé en fin de chaîne car son effet sur une image bruitée est nettement diminué. Pour conclure, la contribution principale de ce prétraitement réside dans le choix, dans l'enchaînement astucieux des différents algorithmes, et dans l'automatisation.

4.3.2 Résultats de l'algorithme de prétraitement

Les résultats du prétraitement sur des images naturelles sont présentés sur la figure 4.19.

4.3.3 Démonstration et quantifications des résultats

La quantification des résultats est une étape très difficile. En effet pour valider un algorithme comme celui-ci, il faut trouver un critère robuste permettant de noter la qualité globale de l'image pour évaluer l'apport réel de l'algorithme. Des critères de ce type existent par exemple le PSNR ou le SSIM[8] cependant, ils nécessitent de connaître la vérité terrain associée à l'image dégradée. Dans notre cas, puisque nous ne disposions pas d'une telle information, il n'existait aucun critère de qualité globale utilisable dans la littérature, d'ailleurs des recherches sur ce sujet pourront faire l'objet de travaux futurs.

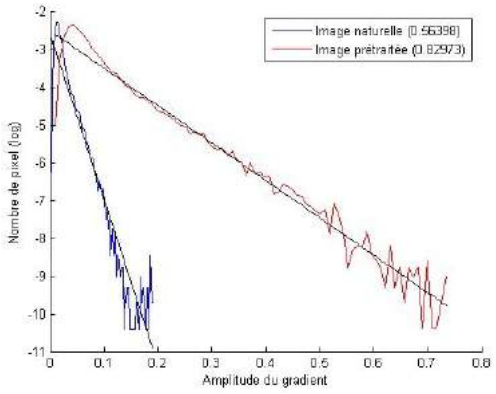


FIG. 4.22 – Courbes $\ln \overline{h}(i)$ et leurs approximations linéaires $a + b.i$.

Nous sommes donc partis du constat qu'un algorithme de prétraitement précède la plupart du temps un algorithme de segmentation (par exemple une détection

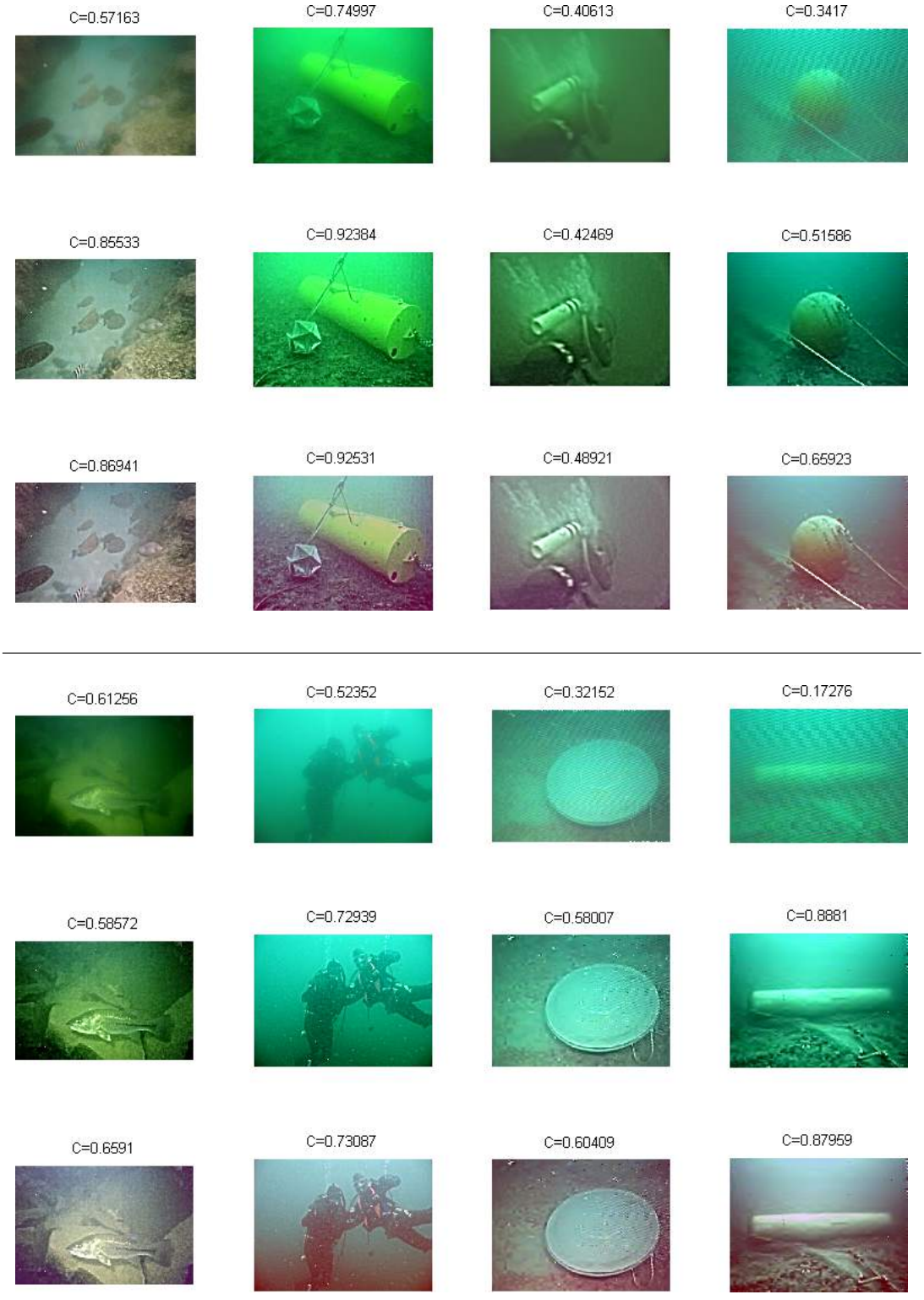
[8] Structural SIMilarity : Mesure de similarité entre deux images numériques en terme de qualité visuelle

FIG. 4.19 – Couple d'images naturelles diverses avant (gauche) et après pré-traitement (droite). Les quatre dernières images (sur les deux dernières lignes) sont extraites des séquences TOPVISION.

de contours ou une extraction de points d'intérêts). De manière à illustrer les qualités de notre algorithme de rehaussement nous allons donc présenter ses effets sur le calcul du gradient en utilisant le critère d'Arnold-Bos [120].
D'après Pratt [121], une image débruitée et bien contrastée est caractérisée par un histogramme du gradient "ressemblant" à une exponentielle décroissante (excepté pour les faibles gradients). En effet si l'histogramme du gradient respecte cette loi exponentielle, les contours sont plus facilement séparés du bruit. A noter que ce critère n'est valide que pour des images assez hétérogènes de par leurs régions et leurs couleurs. La figure 4.3 présente les histogrammes d'amplitude du gradient pour les huit images de la figure 4.19 avant et après l'application de l'algorithme de prétraitement.

On peut voir à partir de ces histogrammes que les valeurs du gradient sont plus importantes après prétraitement, et que la courbe est plus proche d'une exponentielle. Les contours sont donc plus marqués et le seuillage pour les dissocier du bruit est donc moins difficile. Pour démontrer la qualité des résultats nous avons aussi utilisé le critère de robustesse de [120] basé sur la proposition de Pratt [121]. Le but de ce critère est d'approximer la courbe des gradients par une exponentielle et d'attribuer une note entre 0 et 1 en fonction de la distance entre ces deux courbes. Meilleure est la note, meilleur est le rehaussement.

Le calcul de ce critère se fait par la méthode suivante :
– Calcul de l'histogramme normalisé du gradient $h(i)$ de l'image (Fig. 4.3).



TAB. 4.1 – Calcul du critère de contraste avant et après prétraitement et avec ou sans correction des couleurs. Les lignes 1 et 4 regroupent les huit images initiales, les lignes 2 et 5 les images prétraitées et les lignes 3 et 6 les images prétraitées avec correction des couleurs.

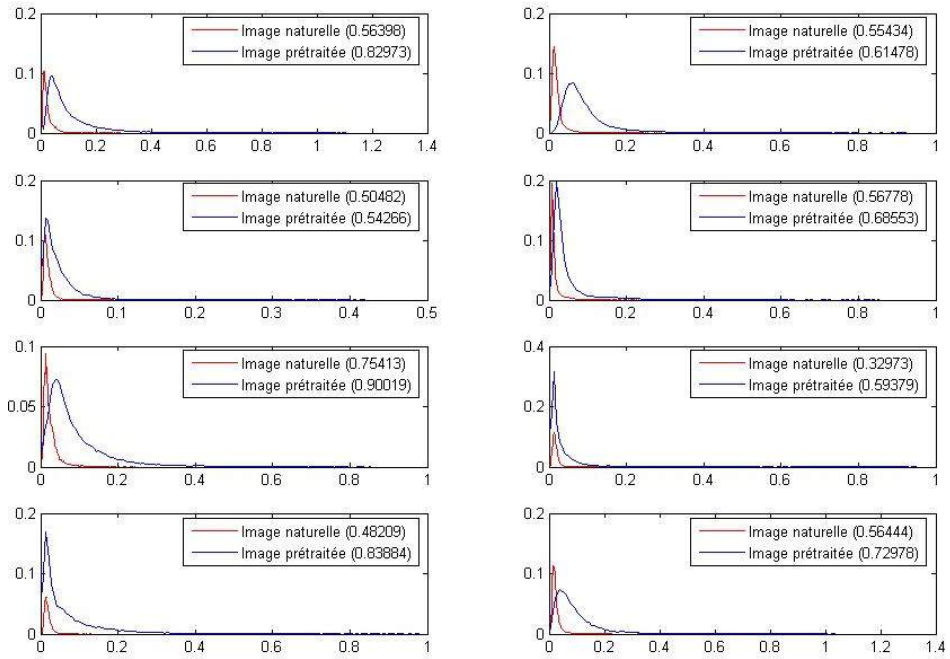


FIG. 4.20 – Histogrammes du gradient des huit images précédentes (voir Fig. 4.19) avant prétraitement (rouge) et après prétraitement (bleu). L'axe des abscisses correspond à l'amplitude de gradient, et l'axe des ordonnées au nombre de pixels (cette valeur est normalisée par la somme totale des pixels). La valeur numérique affichée représente la valeur du critère calculée d'après l'Eq. 4.25.

- Régression linéaire sur la partie décroissante de $h(i)$ suivant l'équation $\ln \overline{h}(i) = a + b.i$.
- Calcul du critère par l'équation (Eq. 4.25).

$$c = \exp(-var_i(\ln h(i) - \ln \overline{h}(i)). \tag{4.25}$$

Analyse des résultats. Le tableau 4.1 présente les résultats du critère de contraste sur les images originales, sur les images prétraitées sans correction des couleurs et sur les images prétraitées avec correction des couleurs. L'analyse de ces résultats montre assez clairement que l'algorithme de prétraitement améliore nettement la valeur du critère. De même que l'étape de correction des couleurs rehausse aussi le critère dans la plupart des cas. A la vue des résultats on peut être surpris par certaines valeurs du critère *a priori* en contradiction avec notre perception visuelle du contraste. Plus clairement certaines images visuellement plus contrastées peuvent avoir une valeur du critère plus faible. Ceci est dû à son calcul basé sur une proposition empirique vis-à-vis des contrastes locaux dans l'image (le gradient) alors que notre perception visuelle apporte une information plus globale sur le contraste. Aussi, comme nous l'avons précisé dans la partie précédente, ce critère n'est vraiment cohérent que pour des images assez hétérogènes de par leurs couleurs et leurs régions. Malgré ces critiques, ce critère permet de quantifier et de valider les résultats du prétraitement.

Conclusion sur le prétraitement développé. Nous avons eu pour le développement de ce prétraitement une approche pragmatique, dont l'originalité réside dans le choix et l'enchaînement des algorithmes et surtout dans l'automatisation. L'automatisation obtenue par une solution empirique

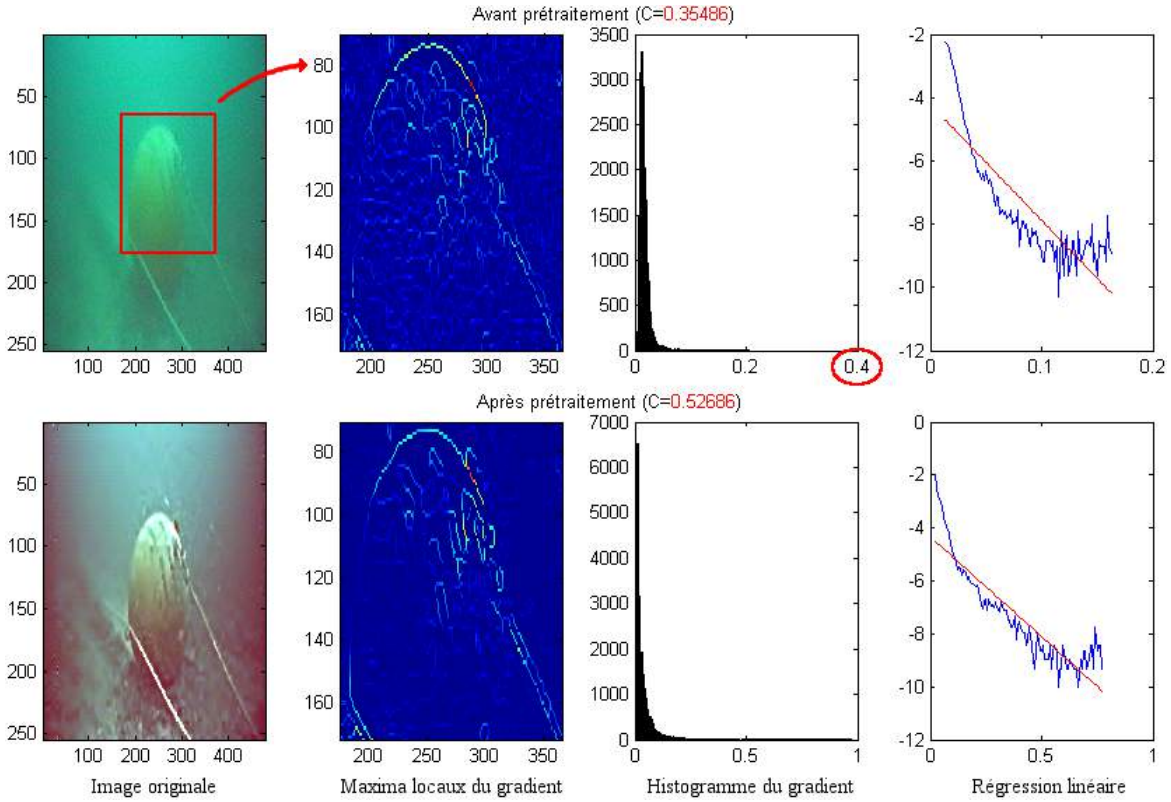


FIG. 4.21 – Calcul de critère de contraste sur une image naturelle et une image prétraitée.

rend l'algorithme robuste à un grand nombre d'image. L'algorithme en Matlab est rapide et pourrait être amélioré par une traduction en langage C. A titre d'exemple sur un ordinateur 3Ghz une image couleur 512×512 sans phénomène de *moiré* est traitée en 1.5 seconde, et il faut à peu près 1 minute pour traiter 25 images constituant une seconde de vidéo. A noter que pour une vidéo on peut ne pas traiter 25 images par seconde. Une dizaine d'images traitées pour une seconde de vidéo paraissent largement suffisantes. Les résultats montrent que la qualité des images est nettement améliorée quasiment dans tous les cas.

4.4 Conclusion du chapitre

L'obstacle majeur dans le traitement des images sous-marines résulte des phénomènes d'absorption et de diffusion dus aux propriétés optiques particulières de la lumière dans l'eau. Ces deux phénomènes auxquels s'ajoute le problème de turbidité, impose de travailler sur des images très bruitées, avec souvent une illumination non uniforme, des contrastes faibles, des couleurs atténuées. Pour s'affranchir de la grande variabilité des conditions d'acquisition ainsi que de l'ensemble de défauts mettant en échec la plupart des méthodes de traitement d'image classiques, l'étude d'une étape de prétraitement était nécessaire. Dans ce chapitre nous avons présenté deux approches de prétraitement possibles et nous avons montré que l'algorithme de rehaussement était le plus approprié puisqu'il ne nécessitait aucune connaissance *a priori* et qu'il permettait d'être plus robuste aux différents problèmes présents

sur les séquences a traiter. Nous avons étudié un grand nombre de méthodes de filtrage utilisables et permettant de corriger certains défauts précédemment cités. A partir de cette étude nous avons développé une "nouvelle méthode" de prétraitement des images sous-marines préparant la segmentation. L'algorithme proposé est une chaîne de prétraitement permettant de corriger les défauts des images sous-marines et d'améliorer la "qualité visuelle" des images. Cette chaîne comprend notamment une correction de l'éclairage, un débruitage, un rehaussement de contraste et une correction des couleurs. L'algorithme proposé est automatique et ne nécessite aucune connaissance *a priori* sur les conditions d'acquisition ni aucun paramétrage manuel de l'utilisateur c'est-à-dire que tous les paramètres utilisés sont auto-évalués ou ont été pré-ajustés de manière empirique. En l'état, la méthode automatique et rapide ici proposée permet déjà d'envisager le traitement temps réel des images sous-marines par une traduction en langage C. Nous avons montré que ce prétraitement améliore de façon importante la qualité visuelle et plus particulièrement les contrastes et les contours des objets dans l'image. Nous avons quantifié les améliorations du prétraitement avant une étape de détection de contour grâce à un critère basé sur l'histogramme de distribution du gradient. Une validation plus "rigoureuse" des résultats de prétraitement sera faite en fin de chaîne sur les résultats de reconnaissance.
D'autres améliorations pourraient encore être apportées pour perfectionner les résultats, on pourrait notamment étudier l'apport des méthodes basées sur les "curvelets" qui permettent le débruitage et rehaussement de contrastes, [113, 114] ainsi que, les méthodes de déconvolution qui donnent aujourd'hui de très bons résultats à condition d'inclure des informations *a priori* [98]. Grâce à l'utilisation de cet algorithme, l'étape de segmentation par les méthodes classiques est nettement facilitée, on peut notamment de manière plus robuste effectuer une détection de contour pour la reconnaissance de forme ou une détection de points d'intérêts pour faire du suivi dans les séquences.

Chapitre 5

La méthode de reconnaissance classique

Nous avons distingué dans le chapitre 3, deux types de méthodes pour la reconnaissance d'objets, les méthodes faisant usage de la 2D et les méthodes faisant usage d'un modèle 3D. Dans ce chapitre nous allons détailler la première méthode, la plus classique, qui base sa reconnaissance sur l'apparence 2D des objets autrement dit sur la forme de leurs silhouettes. Cette méthode consiste à détecter les contours, à en extraire des caractéristiques et enfin à reconnaître l'objet en comparant les informations extraites et les informations contenues dans la base d'apprentissage.

5.1 Une méthode classique mais générique

Dans le domaine terrestre où la littérature en matière de reconnaissance d'objets est très riche, il est classique d'utiliser des méthodes basées sur des vues 2D pour reconnaître des objets, on citera par exemple [122]. Ces méthodes sont généralement découpées en trois étapes qui sont : la segmentation (terme que nous définirons dans le paragraphe suivant), l'extraction d'attributs et la classification de ces attributs qui aboutit à la reconnaissance des entités présentes dans la scène. Les attributs classiquement utilisés étant basés sur la forme, la couleur et la texture. Au contraire, dans la littérature consacrée à l'imagerie sous-marine, l'utilisation des contours (ou attributs de forme) est toujours privilégiée au détriment de la couleur ou la texture qui sont dans ce contexte difficilement exploitables. En effet les contours sont une caractéristique assez stable en imagerie sous-marine [1][46].
Malgré son utilisation souvent privilégiée la méthode de reconnaissance par les contours mérite de nombreuses adaptations pour pouvoir être utilisée en milieu sous-marin, en raison de la qualité souvent mauvaise des images. Une de ces adaptations a été présentée dans le chapitre précédent, il s'agit du prétraitement qui vise à s'affranchir des défauts les plus gênants liés au contexte sous-marin. La deuxième adaptation concerne la segmentation qui doit assurer une extraction d'attributs robuste dans un environnement non contrôlé, c'est-à-dire avec des objets parfois partiellement occultés, noyés dans un fond très hétérogène (sable, roche...), et sous différents points de vue et à différentes distances.
Nous avons choisi d'implémenter cette méthode dans le cadre du projet TOPVISION[1] (voir chapitre

[1] projet centré sur la reconnaissance d'objets manufacturés dans des vidéos sous-marines.

7). Nous avons montré dans le chapitre 3 que cette méthode est générique, peu spécialisée, facilement extensible et qu'il est possible de l'utiliser dans un contexte plutôt défavorable comme celui-ci avec quelques adaptations spécifiques. La figure 5.1 présente l'organigramme de la méthode. Elle est composée d'un certain nombre d'étapes : prétraitement, détection de contours, extraction de paramètres sur les contours (recherche de formes particulières), et enfin reconnaissance à partir des paramètres extraits.

Dans ce chapitre, nous nous focaliserons essentiellement sur les problèmes de segmentation ou d'extraction de l'apparence 2D ainsi que sur la recherche de paramètres discriminants (permettant ensuite une prise de décision correcte). L'étape finale de classification ne sera pas abordée dans ce rapport. Concernant les paramètres discriminants, nous nous sommes arrêtés à des paramètres simples, caractéristiques des objets manufacturés que sont les droites, les cercles ou les ellipses.

5.2 Détection de changement et prétraitement

Avant d'effectuer la segmentation des images, deux algorithmes sont appliqués, d'abord une détection des changements dans l'image puis le prétraitement de celle-ci par l'algorithme présenté dans le chapitre précédent.

5.2.1 La détection de changement dans les images

La détection de changement dans les images a un rôle important dans l'optimisation du temps de calcul lorsque l'on traite des séquences d'images. C'est pourquoi, dans l'objectif d'un traitement temps réel des données, il évidemment primordial de ne pas traiter chaque image indépendamment des autres sans prendre en compte la cohérence de la séquence. Dans notre contexte, les mouvements des véhicules sous-marins sont supposés assez lents, plusieurs images successives dans les séquences sont donc souvent très ressemblantes voire presque identiques. On peut donc traiter la première image et ensuite négliger le traitement des suivantes si aucun changement n'est signalé. Les algorithmes de détection de changement sont nombreux et variés. Ils procèdent notamment par soustraction de l'arrière plan, par histogrammes couleurs ou par méthodes statistiques (on citera par exemple ces quelques publications [59][60][58][123]). Cette étape de l'algorithme, indispensable pour accélérer le traitement, n'a pas encore été implémentée et reste toujours à l'étude. Cependant, nous penchons aujourd'hui sur des méthodes statistiques très rapides applicables directement sur les trois canaux RGB. Ces méthodes statistiques pourraient d'ailleurs être utilisées pour estimer le bruit pour le prétraitement.

5.2.2 Le prétraitement

Lorsque des changements ont été détectés dans la séquence, l'image doit être segmentée. On passe alors à la deuxième étape de notre algorithme : le prétraitement de l'image. En effet comme on a pu le voir dans le chapitre précédent, les images sous-marines souffrent de défauts très particuliers qui mettent en échec la plupart des méthodes usuelles de traitement d'image. Le prétraitement, c'est-à-dire la correction d'un certain nombre de ces défauts ou l'amélioration de la qualité des images est donc nécessaire avant leur utilisation. Il a été démontré dans le chapitre précédent, que l'algorithme

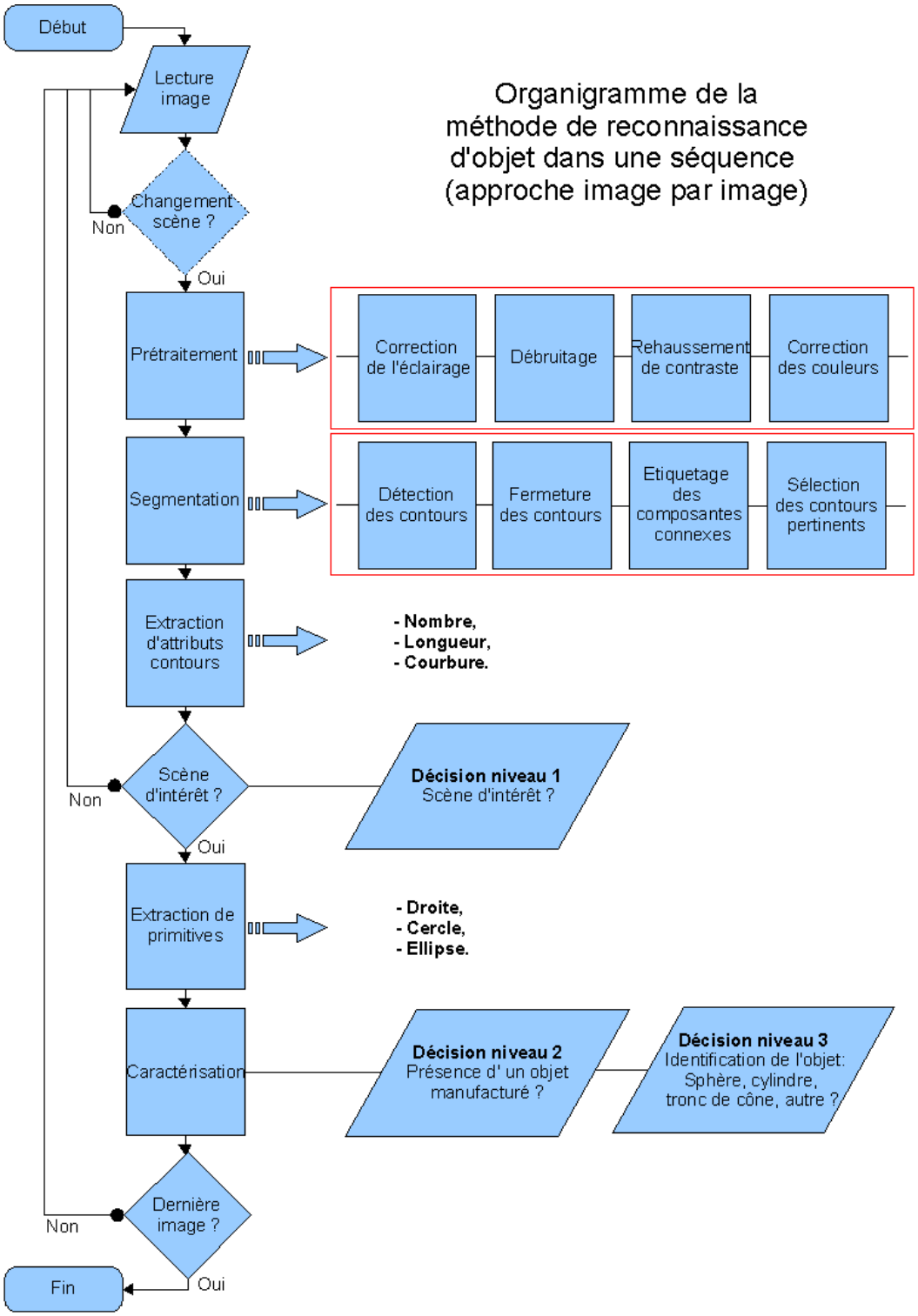


FIG. 5.1 – Organigramme de la méthode proposée pour la reconnaissance automatique d'objets manufacturés dans des séquences d'images sous-marines (Projet TOPVISION)

développé améliore de façon significative les contrastes dans l'image (en renforçant la ressemblance entre les pixels appartenant à une même région, ou en accentuant la dissemblance entre les pixels appartenant à des régions différentes). En améliorant les contrastes, il améliore la qualité et la robustesse de la segmentation et facilite ensuite la sélection des contours pertinents. Le prétraitement a été longuement développé dans le chapitre 4, nous avons montré son l'efficacité sur la détection de contour, à l'aide d'un critère basé sur le gradient. Ses effets sur la segmentation de l'image sont présentés sur la figure 5.2 et la figure 4.21 du chapitre précédent. On peut constater que l'histogramme du gradient est plus allongé (étalé sur un plus grand nombre de valeurs), on a donc des contours plus marqués, mieux localisés et mieux séparés du bruit.

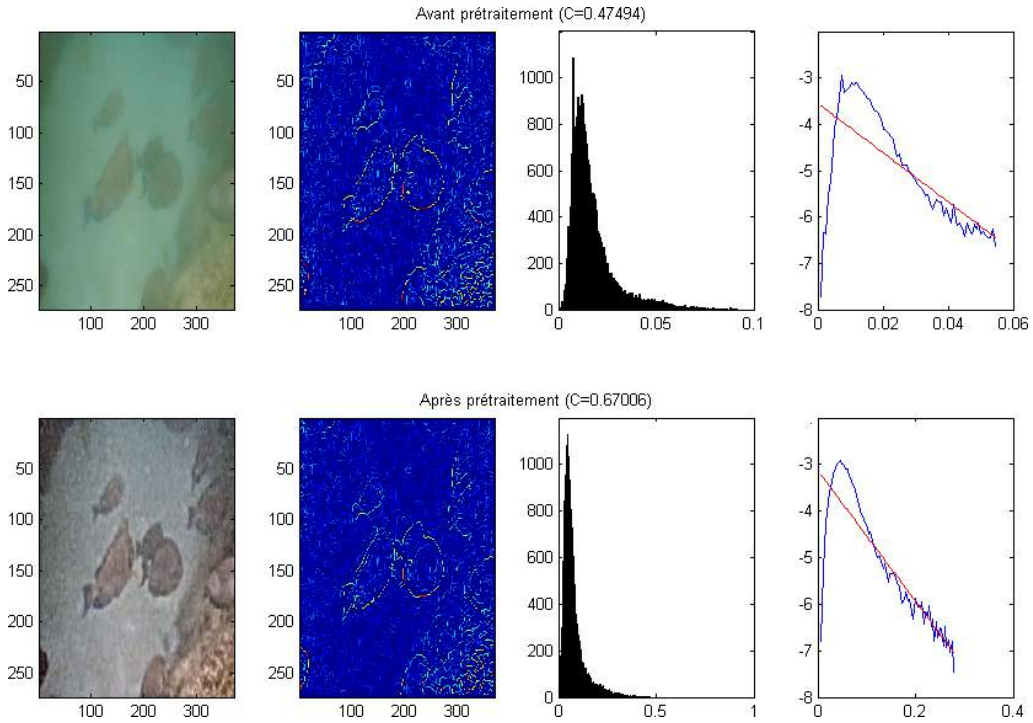


FIG. 5.2 – Démonstration de l'efficacité du prétraitement. Comparaison entre une image brute et une image prétraitée. Colonne 1 : les images, 2 : les maxima locaux du gradient, 3 : l'histogramme des maxima locaux du gradient, 4 : la courbe du logarithme de l'histogramme normalisé, ainsi que sa régression linéaire (le critère de robustesse correspond à la distance entre la courbe et la droite).

5.3 Segmentation des images

L'objectif de la segmentation d'image est d'établir une description compacte et représentative, plus exploitable que l'ensemble de ses pixels. Elle doit aboutir à l'extraction d'indices visuels pertinents sur l'ensemble des entités qui composent la scène. Pour obtenir cette description compacte, on va tenter de partitionner l'image en régions homogènes par rapport à un ou plusieurs critères.
L'étape de segmentation est une étape primordiale, car c'est à partir de l'image segmentée que les mesures sont effectuées pour l'extraction des paramètres discriminants en vue de la classification. Il

est donc important de bien choisir la méthode à utiliser, parmi les nombreuses proposées dans la littérature (très abondante dans ce domaine). On citera par exemple des ouvrages de références tels que [89][121][124][125].

5.3.1 Les différentes méthodes : contours ou régions

La notion d'indice visuel n'est pas simple à définir. On peut présenter un indice visuel comme une information perceptible directement à partir de la visualisation de l'image (contours, régions homogènes au sens des niveaux de gris ou de la texture, formes élémentaires par exemple). Ces indices visuels correspondent à des phénomènes photométriques et géométriques remarquables.
En matière de segmentation d'image, on distingue deux grands types d'approche : l'approche "frontière" et l'approche "région".

- L'approche "frontière" recherche les variations d'intensité ou les discontinuités entre les propriétés de deux ensembles connexes de points. Elle regroupe les méthodes de détection de contours et notamment les méthodes dérivatives, surfaciques, variationnelles, markoviennes ou morphologiques.
- L'approche "région" recherche quant à elle des groupements de points ayant des propriétés communes. Elle regroupe les méthodes de multiseuillage ou "clustering", les méthodes de fusion ou division de régions ainsi que les méthodes dites d'agrégation de points.

Les méthodes "frontière" ne conduisent pas directement à une segmentation de l'image, des post-traitements sont souvent nécessaires car les contours sont rarement connexes. Contrairement aux méthodes "région" qui aboutissent directement à une partition de l'image, chaque pixel étant affecté à une région unique.
Dans notre cas, vis à vis des résultats de l'étude bibliographique, nous avons opté pour une segmentation de type "frontière". La segmentation "région" étant plus adaptée aux images bien contrastées, texturées et aux couleurs marquées. Nous nous intéresserons plus particulièrement aux méthodes dérivatives de détection de contours. Ces approches sont les plus immédiates pour détecter et localiser les variations d'intensité. Bien évidemment des post-traitements seront appliqués pour fermer les contours.

La détection de contour : Le but de la détection de contour est comme on l'a vu de repérer les points de l'image qui correspondent à une variation brutale d'intensité. En ne gardant que ces pixels, on va pouvoir réduire la quantité d'information contenue dans l'image en éliminant des informations non pertinentes mais en préservant les propriétés structurelles importantes. Parmi les méthode existantes on peut extraire deux catégories, la première méthode recherche les extremums de la dérivée première par exemple les maximums locaux du gradient. La seconde recherche les annulations de la dérivée seconde par exemple les annulations du laplacien.

Nous avons opté pour les méthodes utilisant la dérivée première, les plus utilisées. Cette catégorie regroupe toutes les méthodes utilisant un filtrage par différents masques de convolution (voir chapitre précédent). On peut citer par exemple les masques de Roberts, Prewitt, Sobel... Ainsi que la méthode optimale de Canny [126], et sa version modifiée par Deriche [127].

5.3.2 La méthode de détection de contours adoptée

De manière intuitive, nous avons d'abord décidé de tester des méthodes conventionnelles dans le cas particulier que avions à traiter. Notre approche s'est donc basée sur la réutilisation de méthodes de segmentation appliquées en environnement contrôlé que l'on a adaptées au contexte spécifique qu'est le milieu sous-marin. On pourra notamment avoir un exemple global de cette méthode dans [1][128].

La segmentation par détection de contours : De nombreuses méthodes de détection de contours existent dans la littérature, elles ont été évoquées dans la partie précédente. Dans notre cas, nous avons opté pour la méthode de loin la plus utilisée et une des plus performantes : la méthode de Canny. En effet cette méthode est optimisée au sens de trois critères : une bonne détection (c'est-à-dire un faible taux d'erreur), une bonne localisation et une réponse unique. Elle peut être découpée en quatre différentes étapes (voir Fig. 5.3)

1. Lissage de l'image par un filtre gaussien pour réduire le bruit et supprimer les détails et textures inutiles. Une forte valeur de σ induit un fort lissage.

   $f_s(x,y) = f(x,y) \otimes h_\sigma(x,y)$ avec h_σ un filtre gaussien défini par :

   $h_\sigma = \frac{1}{2\pi\sigma^2}.\exp\big(-(\frac{x^2+y^2}{2\sigma^2})\big)$ où σ est le facteur l'aplatissement.

2. Calcul de la norme $\rho(x,y)$ et de l'orientation du gradient $\theta(x,y)$ pour chaque pixel de l'image (le gradient étant la grandeur vectorielle caractérisant la variation d'intensité). Généralement, la norme et l'orientation du gradient sont obtenues à partir des dérivées partielles, elles-même obtenues par une approximation par différences finies. Un exemple d'affichage du vecteur gradient pour des pixels de contour est présenté Fig. 5.5.

3. Extraction des maximums locaux de la norme du gradient, en comparant la norme du gradient (en chaque point) à celle de ses deux voisins situés dans la direction du gradient à une distance unitaire de part et d'autre.
   Soit G_M la norme du gradient du point courant $M(x,y)$, et G_{M_1} et G_{M_2} les normes des gradients en M_1 en M_2. Le point M est un maximum local c'est à dire un point de contour si $G_M > G_{M_1}$ et $G_M > G_{M_2}$ (voir Fig. 5.4)

4. Enfin sélection des contours par un double seuillage par hystérésis des maximums locaux de la norme du gradient. Ce seuillage a pour but principal de préserver les contours tout en supprimant le bruit. A l'issue de cette étape on obtient une image binarisée (image segmentée) où tous les pixels à 1 représentent un pixel de contour.

Nous avons ainsi adopté la méthode proposée par Canny, cependant cette méthode peut utiliser différents algorithmes de calcul du gradient plus ou moins performants. Nous allons dans la partie suivante en présenter quelques uns.

Différentes méthodes de calcul du gradient : Toujours avec la méthode de Canny, nous avons testé plusieurs opérateurs de calcul du gradient : l'opérateur original de Canny, l'opérateur de Canny modifié par Fleck [129], et l'opérateur de Sobel. Ces trois opérateurs ont été testés dans deux contextes différents : le calcul du gradient niveau de gris (ou gradient marginal), et le calcul du gradient couleur (ou gradient vectoriel) de Di Zenzo [130]. Les opérateurs de calcul du gradient utilisés sont définis sur la figure 5.6.

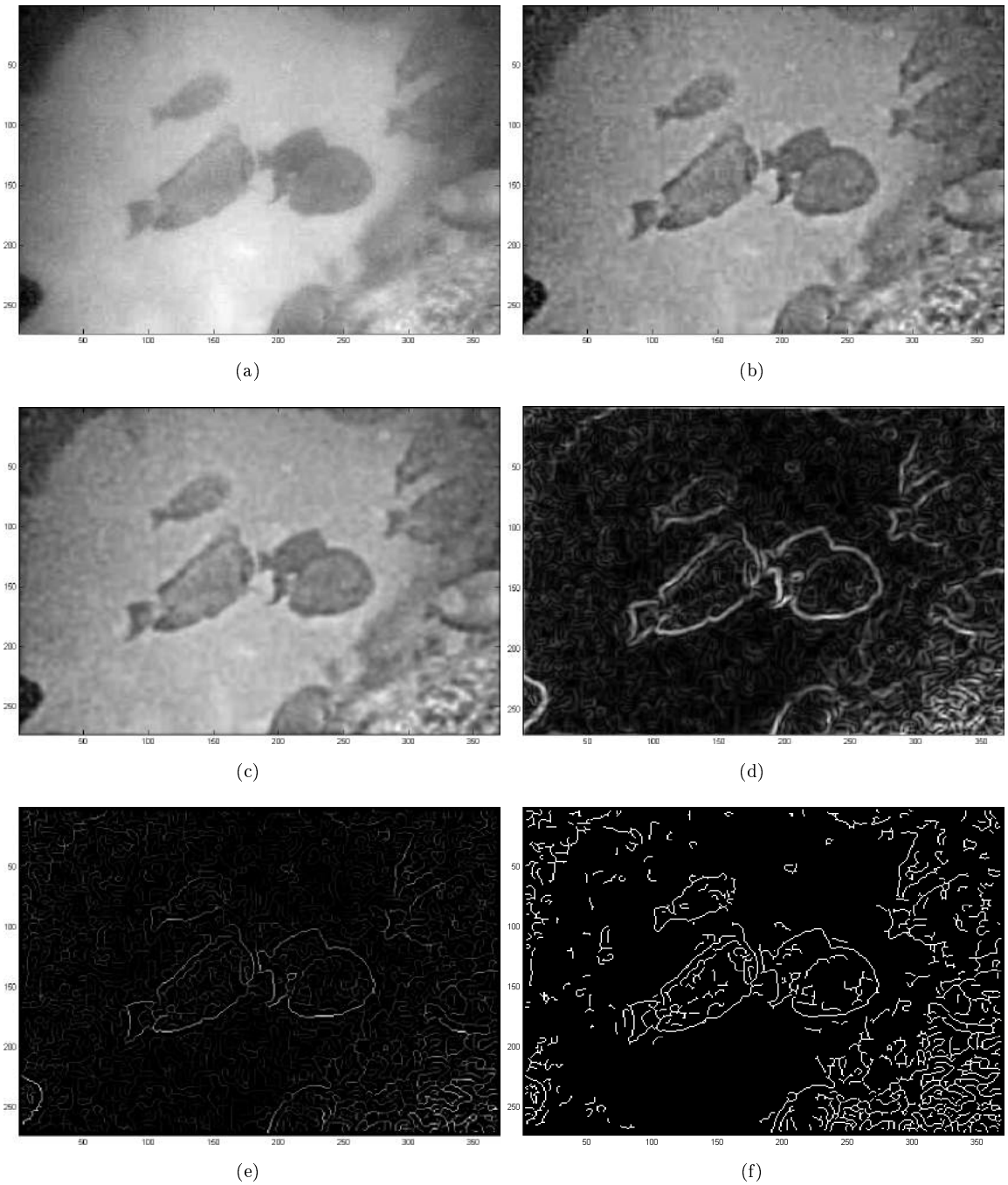


FIG. 5.3 – Les différentes étapes de la détection de contour, (a) Image initiale (b) Image après prétraitement (c) Image lissée par un filtre gaussien de taille 7 x 7 et une déviation fixée à 1 (d) Image de la norme du gradient par les filtres de Canny (e) Image des maxima locaux de la norme du gradient (f) Image des contours obtenus après seuillage par hystérésis.

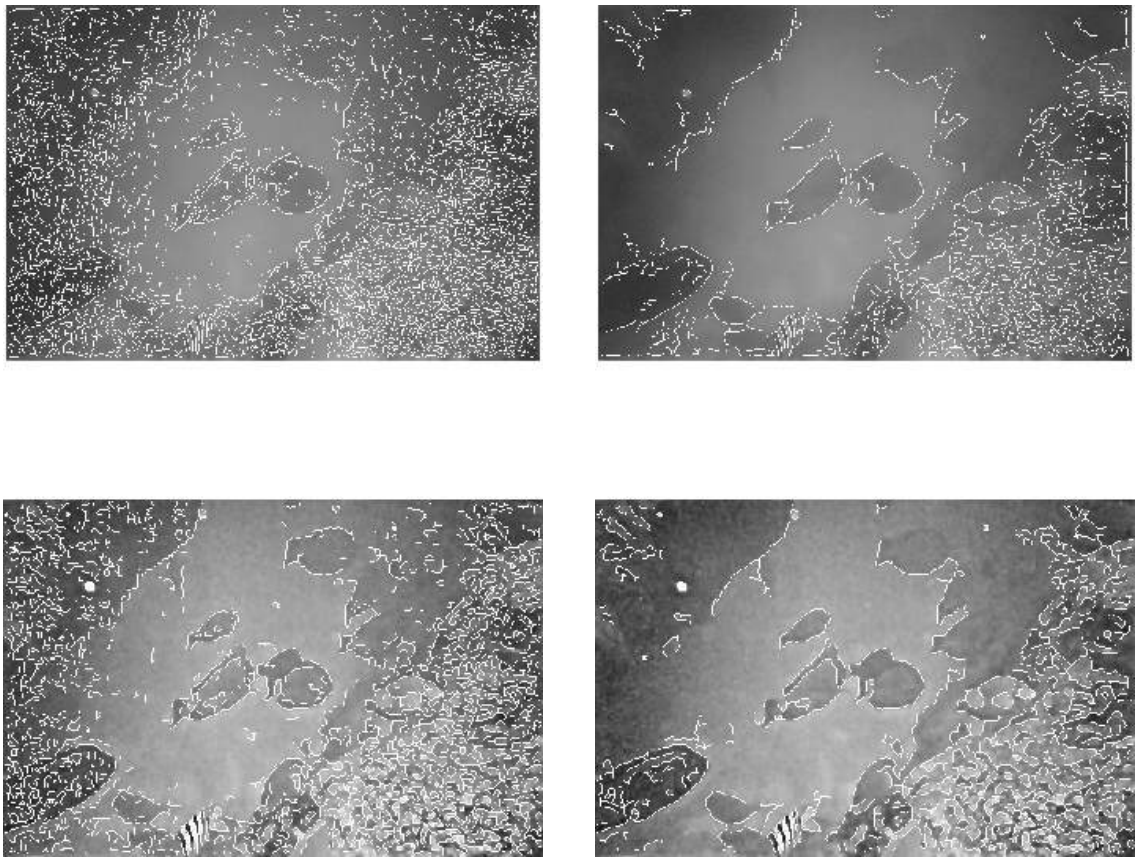


FIG. 5.4 – Contours obtenus après seuillage des maxima locaux du gradient avec deux différents seuils : seuil fort à droite, et seuil plus faible à gauche. En haut sur l'image non prétraitée, et en bas sur l'image prétraitée.

Cas du gradient marginal : Le gradient marginal (en opposition au gradient vectoriel) est calculé à partir de l'image monochrome (ou image de luminance). La norme et l'orientation du gradient sont obtenues à partir des images de gradient horizontal et vertical g_x et g_y (voir équation 5.2) qui sont elles mêmes obtenues par convolution de l'image originale avec les masques précédents.

$$\rho(x,y) = \sqrt{g_x^2(x,y) + g_y^2(x,y)} \tag{5.1}$$

$$\theta(x,y) = \arctan\left(\frac{g_x(x,y)}{g_y(x,y)}\right) \quad \text{avec } \theta = 0 \text{ si } g_y(x,y) = 0$$

$$\text{avec,} \begin{cases} g_x = H \qquad g_y = V \qquad \text{dans le cas de Sobel et Canny original} \\ \\ g_x = H + \frac{D1+D2}{2} \qquad g_y = V + \frac{D1+D2}{2} \quad \text{dans le cas de Fleck} \end{cases} \tag{5.2}$$

En sortie des opérateurs de calcul du gradient on obtient deux images de même taille que l'image initiale, l'image de la norme du gradient et l'image de l'orientation du gradient.

Cas particulier du gradient vectoriel de Di Zenzo : Contrairement au gradient marginal ou unidimensionnel présenté dans le paragraphe précédent, le gradient vectoriel ou gradient multicom-

(a) (b)

FIG. 5.5 – Exemple d'affichage du vecteur gradient pour quelques pixels de contour d'une image.

$$\begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$$
(a) Sobel H

$$\begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}$$
(b) Sobel V

$$\begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}$$
(c) Canny H

$$\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$$
(d) Canny V

$$\begin{pmatrix} -1 & 0 & 1 \end{pmatrix}$$
(e) Fleck H

$$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$
(f) Fleck V

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
(g) Fleck D1

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}$$
(h) Fleck D2

FIG. 5.6 – Quelques opérateurs de calcul du gradient.

posante est calculé à partir des trois composantes RGB, en recherchant la direction pour laquelle les variations sont les plus élevées. Ce calcul s'effectue par minimisation d'un critère de distance caractérisant la différence vectorielle dans l'espace colorimétrique.
Les approches proposées d'abord par Di Zenzo[130], puis Lee et Cok et Sapiro [131][132] sont les approches les plus souvent utilisées dans les méthodes de détermination de contour multicomposante.

La procédure du calcul du gradient vectoriel est la suivante :

- Calcul des images de gradient horizontal et vertical des trois plans RGB : $g_x^R, g_y^R, g_x^G, g_y^G, g_x^B, g_y^B$ par Fleck. On obtient les informations d'amplitude et d'orientation par composante.
- Calcul pour chaque pixel de l'image du gradient vectoriel associé (a b c), via les trois coefficients suivants :

$$a = (g_x^R)^2 + (g_x^G)^2 + (g_x^B)^2 \tag{5.3}$$
$$b = g_x^R . g_y^R + g_x^G . g_y^G + g_x^B . g_y^B \tag{5.4}$$
$$c = (g_y^R)^2 + (g_y^G)^2 + (g_y^B)^2 \tag{5.5}$$

Ensuite on obtient l'amplitude et l'orientation du gradient vectoriel par les formules de Sapiro :

$$\rho(x,y) = \sqrt[4]{(a-c)^2+4b^2} \tag{5.6}$$

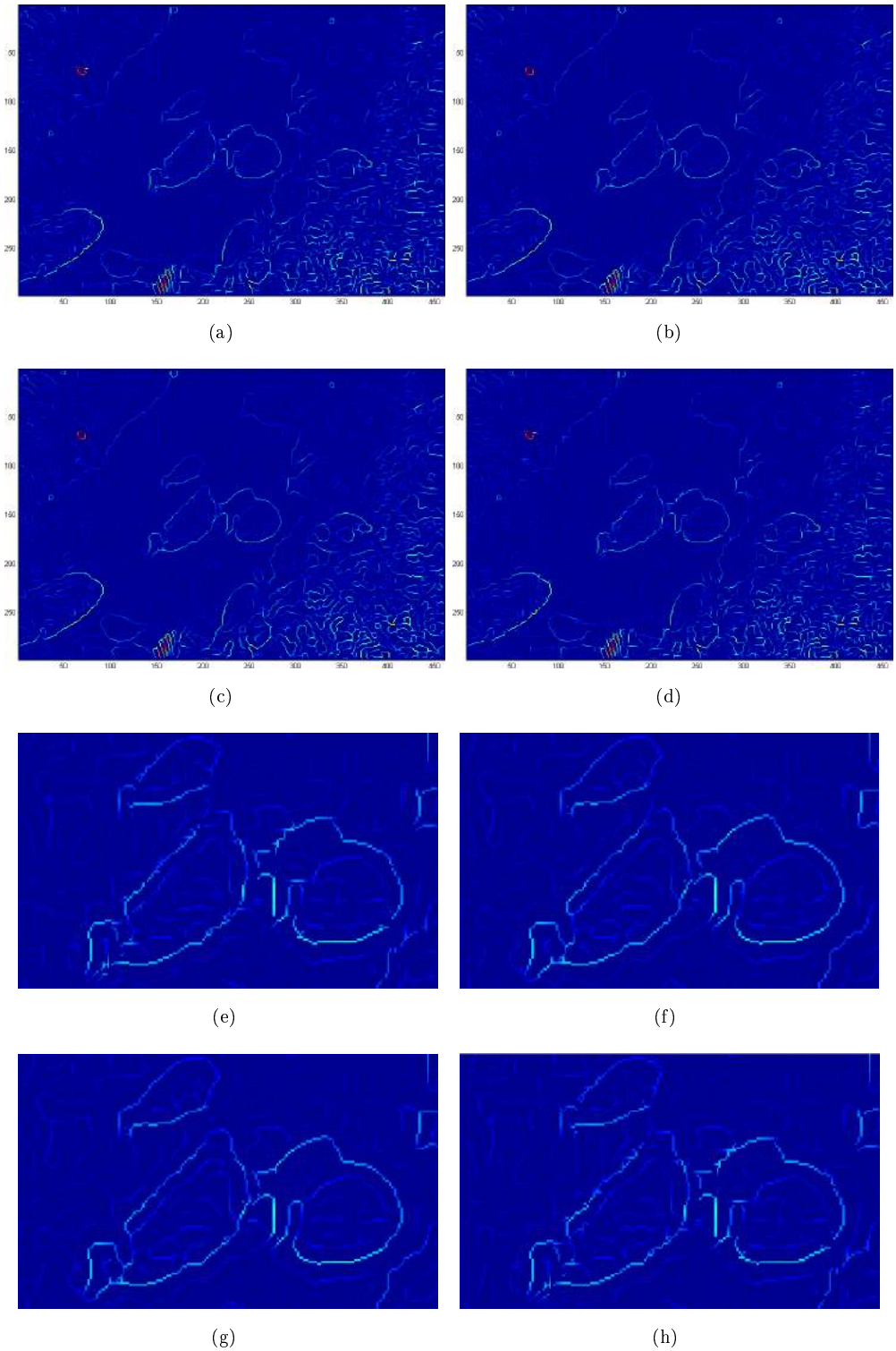
$$\theta(x,y) = \begin{cases} \arctan\left(\frac{-a+c+\sqrt{(a-c)^2+4b^2}}{2b}\right) & \text{si } b \neq 0 \\ \frac{\pi}{2} & \text{si } b = 0 \end{cases} \tag{5.7}$$

En sortie en obtient comme pour le gradient marginal les images de norme et d'orientation du gradient.

Conclusion et comparaison de résultats de calcul du gradient Nous avons testé plusieurs méthodes de calcul du gradient pour la détection de contours par la méthode de Canny. Nous ne présenterons que les quatre meilleurs résultats (voir Fig. 5.7) : le calcul du gradient marginal par l'opérateur de Sobel, et de Fleck ainsi que le gradient vectoriel par les deux mêmes opérateurs. Le calcul du gradient par la méthode de Canny n'a pas été retenu car nous lui avons préféré sa version améliorée par Fleck. Parmi ces méthodes de détection de contours implémentées, deux se distinguent par leurs résultats : celle utilisant le gradient monochrome et l'opérateur de Fleck et celle utilisant le gradient couleur et l'opérateur de Sobel. Ces deux méthodes génèrent des images de gradient un peu plus riches et donc des contours plus continus. Sur notre base d'images tests (base très diversifiée d'environ 50 images sous-marines provenant de TOPVISION, SAUC'E et WWW) les contours sont moins morcelés, mieux localisés et mieux séparés du bruit. Les deux méthodes sont à peu près équivalentes sur le résultat de détection de contour. Cependant on peut observer que l'image des maximums locaux obtenue par le calcul du gradient vectoriel apporte un peu plus d'information que celle obtenue par un gradient marginal et ce, sur une gamme de valeurs un peu plus large, donc *a priori* plus facile à seuiller. Quoi qu'il en soit ces méthodes sont peu distinguables et les résultats peuvent fluctuer d'une image à l'autre. On utilisera par la suite les opérateurs de Fleck sur l'image de luminance pour le calcul du gradient pour une question de rapidité.

A partir de ce calcul, on extrait les maximums locaux de la norme du gradient, on obtient donc une image qui associe à chaque point de contour une valeur caractérisant sa "force", c'est-à-dire l'importance de la transition entre les deux régions de par et d'autre du contour. Cette étape doit être suivie d'un seuillage pour terminer la segmentation de l'image. Le réglage du seuil résulte d'un compromis pour extraire au mieux les contours "vrais" du bruit dans les maxima locaux du gradient. Dans le cas de la méthode de Canny on utilise un double seuillage par hystérésis.

Le seuillage par hystérésis : En raison du bruit, les opérateurs de détection de contours font que, quand un seuil unique est utilisé sur l'image des maximums locaux pour obtenir les contours, il y a un risque de création de lacunes.
Le seuillage par hystérésis avec deux valeurs de seuil : seuil haut noté s_h et seuil bas s_b ($s_h > s_b$) est employé pour limiter ce défaut. Les maximums locaux dont la valeur est supérieure à s_h sont conservés, ainsi que tous ceux dont la valeur est supérieure à s_b et qui appartiennent à une composante connexe comprenant au moins une valeur supérieure à s_h. Ce seuillage permet de combler les trous et donc de réduire le risque de coupure d'un contour. Le seuillage par hystérésis permet un prolongement des



(a) (b) (c) (d) (e) (f) (g) (h)

FIG. 5.7 – Quelques résultats de détection de contour sur une image complexe avec différents méthodes de calcul du gradient : a) gradient marginal et Sobel, b) gradient marginal et Fleck, c) gradient vectoriel Sobel, d) gradient vectoriel Fleck e) f) g) h) zoom sur la partie centrale des images précédentes.

contours détectés par le premier seuil. L'algorithme de seuillage est le suivant :

$$f(x,y) = \begin{cases} 1 & \text{si } \rho(x,y) > S_h \\ & \text{ou} \\ & \text{si } S_h > \rho(x,y) > S_b \text{ et s'ils sont connectés à un pixel} > S_h \\ & \text{par une chaîne de pixels} > S_b \\ 0 & \text{sinon} \end{cases}$$

Remarque : Le seuillage est évidemment une étape décisive, cependant cette étape résulte d'un compromis entre le bruit et les contours. Le problème principal de ce seuillage est le choix du seuil qui reste toujours très subjectif. Dans notre cas nous nous sommes attachés à trouver un seuillage robuste et non un seuillage optimal. On entend par seuillage robuste, le fait que la quantité de pixels de contours ne varie pas trop avec le seuil. Cela permet d'avoir un réglage choisi largement et valable pour plusieurs images. Les seuils du seuillage par hystérésis sont donc relatifs au gradient le plus fort, techniquement ils ont été fixés à 10% et 20% du gradient maximum. Le seuil haut est volontairement choisi assez bas pour garder un maximum de contours puisque les contours en imagerie sous-marines sont en général difficilement détectables.

5.4 Post-traitement à la segmentation

Les contours détectés dans l'image, il s'agit maintenant de les fermer puis de sélectionner ceux d'entre eux qui sont pertinents pour la reconnaissance. La fermeture doit permettre de parfaire la segmentation. Elle est nécessaire car comme on a pu le voir précédemment les contours détectés par des méthodes "frontière" sont rarement connexes et présentent des lacunes ayant différentes origines. Ensuite la sélection des contours doit permettre de supprimer les contours isolés ou trop petits (correspondant à des détails insignifiants pour la reconnaissance de forme) et de garder les contours structurels de l'image.

5.4.1 Fermeture des contours

La fermeture des contours a été initiée par le seuillage par hystérésis. Cependant il reste des lacunes et beaucoup de petits segments isolés subsistent. Nous avons donc opté pour une autre étape de fermeture simple et rapide : la fermeture par morphologie mathématique.

Fermeture par morphologie mathématique : La morphologie mathématique traite les images binaires et fait appel à la théorie des ensembles. Elle utilise un élément de géométrie et de taille connues appelé élément structurant (par exemple un carré de 3 x 3 pixels). Une image binaire contient un certain nombre de pixels connexes codés à 1 que l'on peut définir comme les objets d'intérêts, le fond étant codé à 0. La processus de morphologie mathématique consiste à déplacer l'élément structurant de façon a ce que son centre passe par toutes les positions de l'image binaire à traiter. Pour chacune de ces positions on effectue un calcul relatif à l'union ou l'intersection de l'élément structurant et les pixels de l'image. On obtient ainsi une nouvelle image. Les opérateurs que nous avons utilisés sont : la fermeture (dilatation suivie d'une érosion) et l'opérateur de squelettisation ou d'amincissement. Pour plus de détails sur ces deux méthodes nous renvoyons le lecteur aux références [133][134].

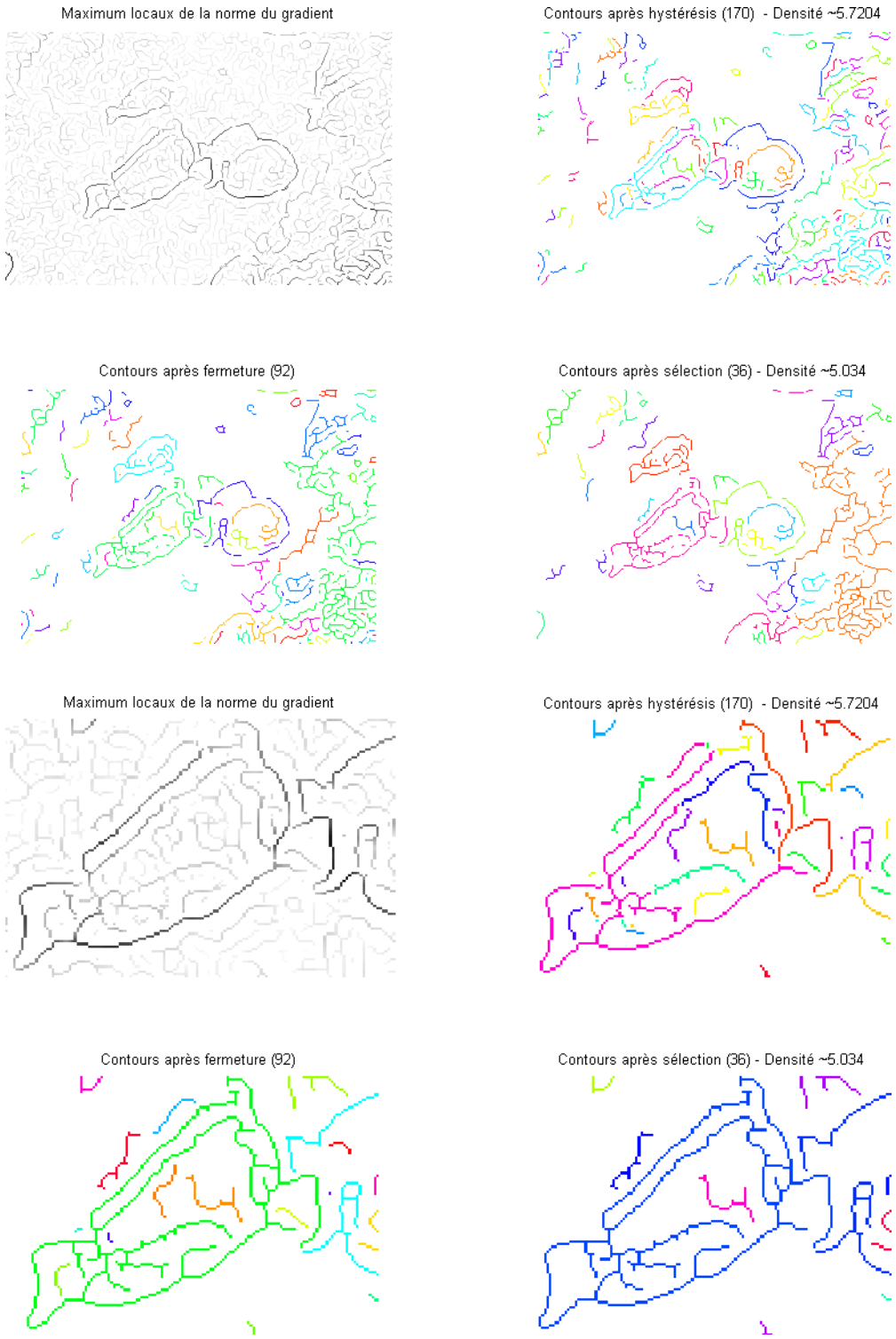


FIG. 5.8 – Fermeture et sélection des contours pertinents. Les quatre images du bas sont un zoom sur la partie centrale des images du haut. Les informations affichées sont : le nombre de contour et le pourcentage de pixels de contour dans l'image.

L'utilisation de la fermeture permet de combler les petites lacunes restantes et de rattacher les contours isolés. En effet deux contours situés à une distance inférieure à l'élément structurant vont être fusionnés. Quant à la squelettisation, elle va permettre d'amincir tous les trous dans les objets comblés par la fermeture. Les résultats de ces deux étapes conjuguées sont présentés Fig. 5.8.
D'autre approches de fermeture sont utilisables et plus robustes que la morphologie mathématique. On peut distinguer par exemple [90] :

- les méthodes exploitant les techniques d'exploitation de graphe,
- les méthodes d'inspiration neurobiologique,
- les méthodes exploitant les contours déformables.

Notamment la première approche procédant à une recherche du meilleur chemin (à partir de chaque extrémité de contour, il y a élaboration d'un chemin qui suit la meilleure ligne de crête dans l'image de la norme du gradient) pourrait permettre de parfaire les résultats de fermeture, elle est cependant beaucoup plus gourmande en temps de calcul.

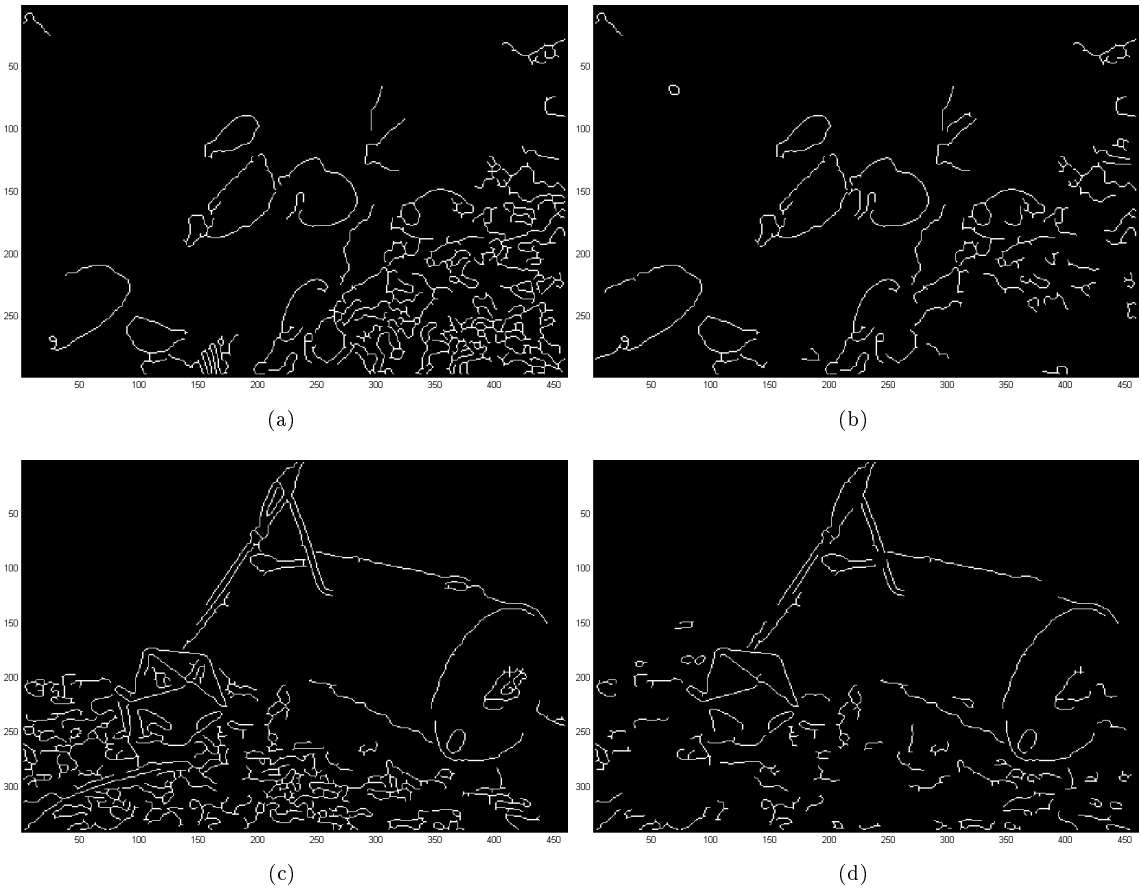
5.4.2 Sélection des contours pertinents

Après l'étape de fermeture, on dispose d'une image binaire où tous les pixels correspondant à un contour ont la valeur 1 et les autres la valeur 0. On doit alors procéder à l'étiquetage des composantes connexes c'est-à-dire des contours à proprement parler (en effet à l'étape précédente aucun lien entre les pixels n'existe). Il s'agit d'attribuer à chaque composante 8-connexe (8 voisins) une étiquette permettant donc de séparer les objets. A l'issue de cette étape on travaille donc sur liste de contours, constitués eux-mêmes d'une liste de pixels. On peut de cette manière faire des mesures sur ces contours. L'étiquetage terminé, on va tenter de sélectionner les contours intéressants dans l'image des contours. Cette sélection s'applique en deux phases :

- Suppression des contours dans les zones à fortes densités,
- Sélection des contours les plus longs.

1. D'abord on va effectuer une suppression des contours dans les zones à fortes densités c'est-à-dire où le taux de pixels de contour est trop important. En effet ces zones texturées sont difficiles à exploiter et souvent à l'origine d'erreurs dans l'étape de classification (voir Fig. 5.9).

   Statistiquement les contours ne représentent dans une image pas plus de 5% des pixels. On peut donc effectuer une suppression de densité lorsque celle-ci est trop importante. Dans les images segmentées on obtient souvent un grand nombre de contours dans certaines zones. Ce phénomène est du en général à la texture du fond sous-marin comme par exemple des mousses, des algues, des rochers, des coraux... Ces contours souvent longs et recourbés ne sont pas supprimés lors de la sélection basée sur un critère de longueur et induisent en erreur les algorithmes de recherche de formes particulières appliqués sur les contours pertinents sélectionnés. Puisque nous n'avons pas exploité le critère basé sur la courbure, il est donc nécessaire de dégarnir ces zones ou les contours sont très concentrés qui ne correspondent généralement pas à des objets (même si elles y correspondent, elles sont de toute façon inexploitables). Nous avons pour cela opté pour une méthode de morphologie mathématique. Dans une fenêtre glissante qui va parcourir toute l'image, on ne garde le pixel central que si le pourcentage de contour présents dans la



(a) (b) (c) (d)

FIG. 5.9 – Suppression des zones à forte densité en contour sur deux images.

fenêtre est inférieur à un seuil de densité donné. D'autres techniques pourraient être utilisées pour supprimer ces zones texturées par exemple [135][136].

2. Ensuite sur cette liste de contours restants et relabellisés après la suppression dans les zones à fortes densités, on va essayer de sélectionner les contours pertinents. La première sélection va donc être de supprimer les contours inférieurs à une certaine longueur (généralement une dizaine de pixels). En effet les petits contours non raccordés après fermeture sont considérés comme insignifiants pour nos traitements. Puis pour terminer notre sélection nous allons sélectionner la moitié des plus grands contours, c'est-à-dire ceux dont la longueur est supérieure à la taille médiane des contours. Ce choix nous a paru encore une fois le plus intuitif et le plus robuste puisqu'il permet de garder les contours les plus longs quel que soit le nombre de contours détectés dans l'image. Éventuellement, pour parfaire la sélection de contours pertinents, on peut appliquer une vérification sur un critère de couleur, basée sur un principe simple utilisé dans [128]. En effet, contrairement aux contours correspondants au fond, les contours des objets ont des pixels de couleurs différentes de part et d'autre (l'objet et le fond).

Remarque : Dans Olmos [1], la sélection des contours pertinents se fait par filtrage multirésolution et sélection automatique de la meilleure échelle par optimisation. Cette technique est assez intéressante

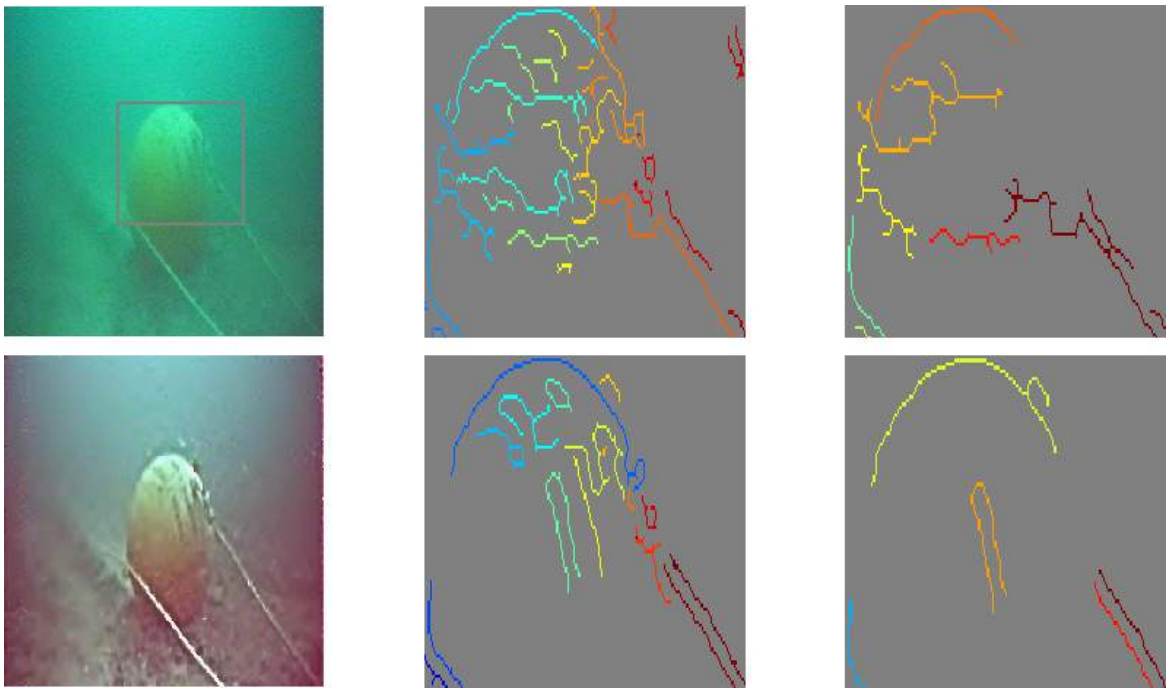


FIG. 5.10 – Segmentation et post-traitement avec ou sans prétraitement.

mais le calcul est complexe et lourd en temps, et les résultats de segmentation sont à notre avis peu améliorés. C'est pourquoi, nous avons dans notre cas privilégié, le traitement unique (en opposition au traitement multirésolution) de chaque image avec des seuils soit prédéfinis, soit automatiques sans chercher à optimiser un critère ou à converger vers le seuil optimal.

Après le deux étapes précédentes qui sont le prétraitement et la segmentation on dispose d'un ensemble des composantes connexes qui correspondent aux contours les plus pertinents (on peut voir sur la figure 5.10 les résultats de segmentation de post-traitement). A partir de cette image on va tenter de reconnaître les objets manufacturés par extraction d'attributs et classification.

5.5 La reconnaissance à partir des contours

A ce niveau du traitement on dispose de la scène sous la forme d'une image segmentée dans laquelle nous devons reconnaître les entités présentes. Nous avons dans notre schéma 5.11 appelé cette étape "extraction de paramètres et décision". Dans le cas général, les objets manufacturés sont caractérisés par des contours longs, lisses, et peu nombreux, alors que des objets naturels font plutôt apparaître des contours tortueux et irréguliers, souvent nombreux et courts. De plus, dans notre contexte, on s'intéresse à des objets manufacturés aux formes élémentaires c'est-à-dire composés de droites, cercles et ellipses. Sur ces simples considérations nous avons décidé de baser notre décision de reconnaissance sur deux critères simples : le nombre de contours et la présence de droite ou cercle/ellipse dans les contours. D'autres critères comme la courbure seront sans doute ajoutés par la suite pour améliorer la robustesse.

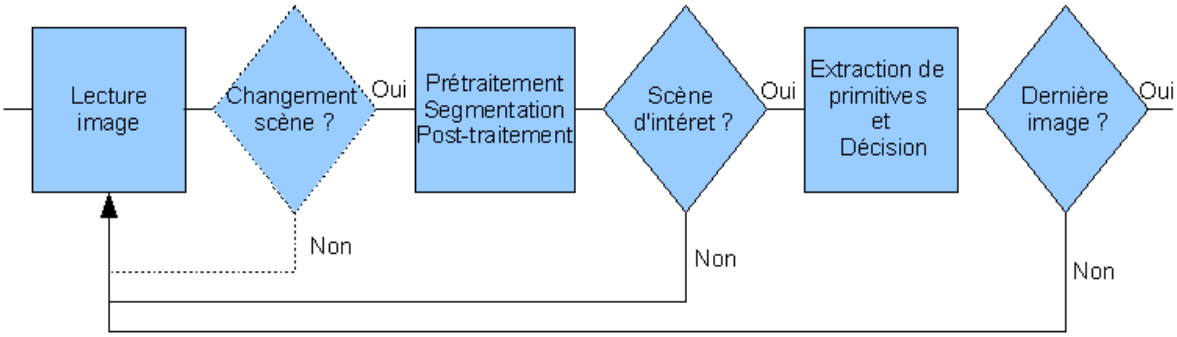


FIG. 5.11 – Les différentes étapes du processus de reconnaissance.

5.5.1 La recherche de primitives

Pour la recherche de formes particulières dans nos images nous avons utilisé la transformée de Hough (inventée par Hough en 1962 et brevetée par IBM) qui est une technique standard parmi les plus simple en reconnaissance de formes (ou vision artificielle). C'est une méthode classique utilisée dans sa version originale pour détecter les droites mais qui a été améliorée pour détecter d'autres formes. Aujourd'hui la transformée telle qu'on l'utilise pour l'extraction de primitives a été développée par Duda [137]. Elle permet la détection de droites, de cercles ou d'ellipses de façon traditionnelle, mais elle peut aussi être étendue à la description d'objets plus complexes (voir Transformée de Hough généralisée [88]).
Le principe général de la transformée de Hough est d'établir une projection entre l'espace de l'image et un espace de paramètres représentatifs de la forme recherchée. Dans cet espace on va accumuler, des données confirmant sa présence. La transformée de Hough permet d'avoir une bonne robustesse contre les discontinuités et les éléments manquants des contours.

La transformée de Hough : Dans ce paragraphe nous allons développer l'aspect théorique de la méthode dans son application la plus simple c'est-à-dire pour reconnaître les droites, ensuite nous développerons les modifications à apporter pour reconnaître d'autre formes comme par exemple des cercles ou des ellipses.

Cas de la recherche de droite : Il faut considérer que toute droite d'un plan peut être défini par une représentation polaire

$$r = x * \cos\theta + y * \sin\theta. \tag{5.8}$$

Chaque droite est donc caractérisée par un vecteur de coordonnées qui contient : l'angle du vecteur noté θ et sa norme notée r (la norme représente la longueur du segment perpendiculaire à la droite et passant par l'origine). Dans l'espace de Hough ou espace des paramètres (r, θ), on peut associer à chaque point (x, y) d'un contour une courbe unique contenant chaque valeur possible pour les paramètres. Si deux courbes associées à deux points se coupent, l'endroit où elles se coupent dans l'espace de Hough correspond aux paramètres d'une droite qui relie ces deux points. A partir de ces représentations dans l'espace de Hough, c'est-à-dire après l'étape d'accumulation ou de vote, il est nécessaire de trouver les maximums locaux dans l'accumulateur, et d'appliquer un seuillage. Ensuite il est aisé de sélectionner les segments correspondant à des contours (voir Fig. 5.12). Le seuillage des maximums

locaux est un élément essentiel qui peut accroître la vitesse et augmenter la précision de la détection.

Cas de la recherche de cercle : Il faut dans le cas du cercle considérer l'équation suivante

$$(x-a)^2 + (y-b)^2 = r^2. \tag{5.9}$$

Nous avons donc 3 paramètres à identifier : l'espace des paramètres est un espace à trois dimensions (a, b, r) définissant les coordonnées du centre et le rayon du cercle [138]. On va appliquer le même calcul que pour la droite mais on va le répéter pour chaque valeur de rayon que l'on suppose "plausible" en évitant les rayons trop petits ou trop grands (la détermination du juste pas entre les rayons est très délicate). Le recherche de cercle dans une image est beaucoup plus lourde en temps de calcul que la recherche de droites a cause de ce bouclage qui fait que l'on obtient autant d'accumulateurs que de rayons recherchés. Pour finir, on retient le cercle ayant un pourcentage de couverture par les pixels de contour le plus important, mais aussi supérieur à un seuil minimum défini par l'utilisateur (voir Fig. 5.12).

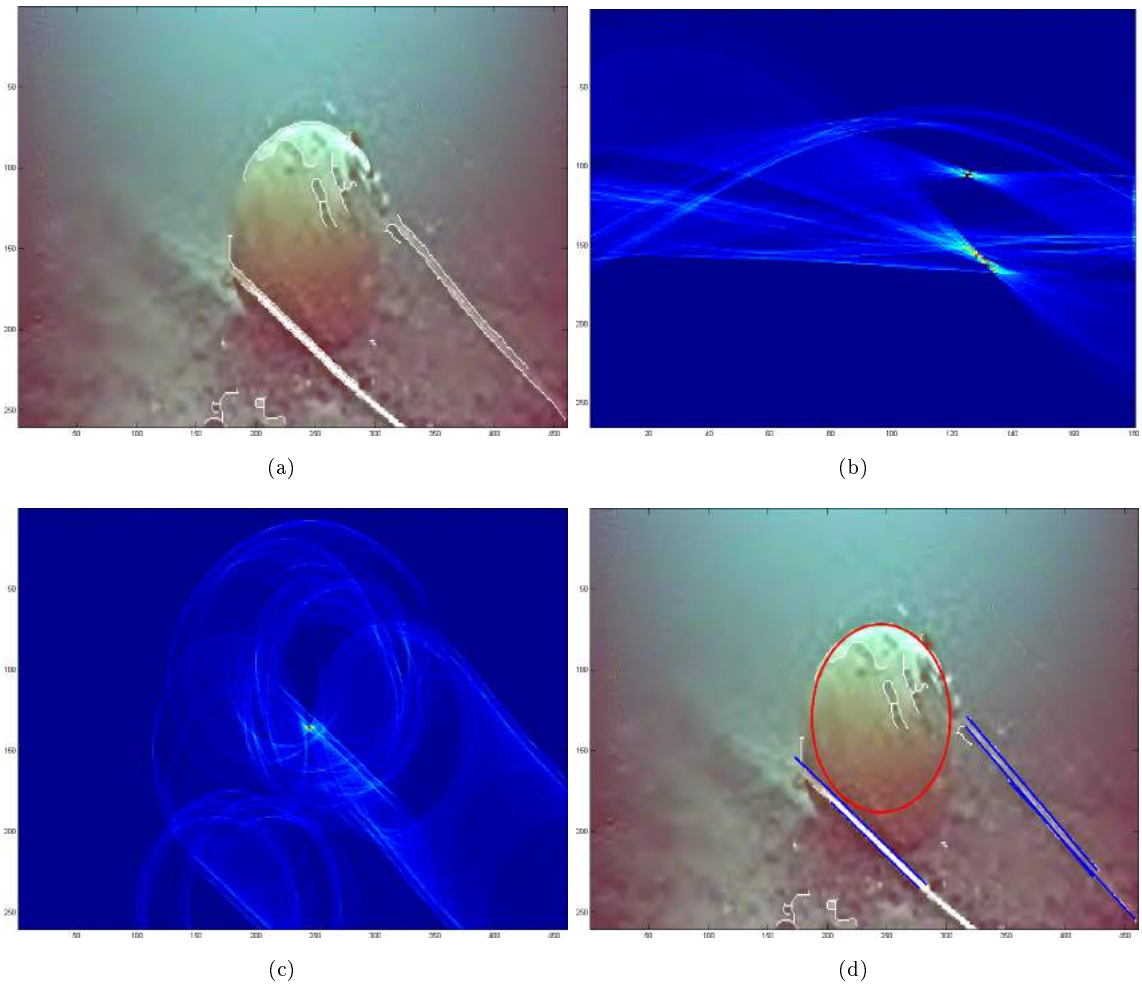
Cas de la recherche d'ellipse : Dans le cas de l'ellipse il faut déterminer 5 paramètres car on doit considérer l'équation suivante :

$$a * (x-p)^2 + 2b * (x-p)(y-q) + c(y-q)^2 = 1. \tag{5.10}$$

L'application directe de la transformée de Hough aux cas des ellipses demande donc une projection dans un espace de Hough à 5 dimensions. Cette approche n'est pas vraiment utilisable car l'espace de paramètres est trop grand pour avoir des accumulations significatives. Il faut donc utiliser une approche permettant de traiter un problème de dimension inférieure. On peut citer notamment la méthode [139] qui découpe le problème en deux passes : d'abord la recherche des centres d'ellipses possibles (espace à 2 dimensions), ensuite la recherche des trois autres paramètres. Nous n'allons pas développer davantage cette méthode, puisque nous ne l'avons pas implémentée. Nous renvoyons le lecteur intéressé sur cette référence bibliographique [140]. Ce qui est par contre nécessaire de préciser c'est que le temps de calcul pour la recherche d'ellipse est encore plus long que la recherche de cercles.

Quelques remarques sur la transformée de Hough : Dans l'objectif de traitements temps réel, il est indispensable d'utiliser des méthodes d'optimisation pour accélérer la transformée de Hough. La transformée consiste en effet en un parcours de l'ensemble des points de contour et un test de l'ensemble des formes possibles. Ceci rend les calculs très lourds à la fois en temps et en ressource. On pourrait donc aussi améliorer la méthode standard par l'application des méthodes plus rapides :

- la transformée de Hough probabiliste [141] qui au lieu de traiter tous les pixels de l'image applique les calculs à une proportion de pixels dans l'image. Kiryati [142] propose entre 10% et 20% pour que l'histogramme soit comparable à celui obtenu par la méthode traditionnelle.
- la transformée par tirage aléatoire [143], version améliorée de la méthode précédente.
- la transformée hiérarchique [144]. Comme on l'a dit dans la partie précédente le choix de quantification dans l'espace de paramètres (c'est-à-dire la taille du ou des accumulateurs) est très difficile à définir. La transformée de Hough hiérarchique consiste à commencer les calculs avec un pas de quantification large puis à affiner ces pas pour les accumulateurs ayant reçu le meilleur nombre de votes.



(a) (b)

(c) (d)

FIG. 5.12 – Détection de formes sur les contours (a) image originale, (b) accumulateur pour la recherche de droite, (c) accumulateur pour la recherche de cercles pour le rayon le plus approprié, (d) image originale avec les droites et les cercles détectés.

5.5.2 Reconnaissance des objets manufacturés

Pour résumer, l'algorithme de reconnaissance développé à partir des techniques classiques est le suivant :

1. détection de contours,
2. fermeture des contours,
3. étiquetage des composantes connexes,
4. sélection de contours pertinents,
5. détection de droite (recherche de pics dans l'orientation du gradient, et transformée de Hough),
6. détection de cercle (transformée de Hough).

Les résultats obtenus sont les suivants (les images correspondantes sont présentées sur la figure 5.13).

- **Présence d'un objet : 0**, Nombre de contour : 37,
  Pic dans l'orientation du gradient : 0, Nombre de droite : 0, Nombre de cercle : 0.
- **Présence d'un objet : 1**, Nombre de contour : 36,
  Pic dans l'orientation du gradient : 0, Nombre de droite : 2, Nombre de cercle : 0.
- **Présence d'un objet : 1**, Nombre de contour : 58,
  Pic dans l'orientation du gradient : 0, Nombre de droite : 2, Nombre de cercle : 0.
- **Présence d'un objet : 1**, Nombre de contour : 12,
  Pic dans l'orientation du gradient : 1, Nombre de droite : 5, Nombre de cercle : 1.
- **Présence d'un objet : 1**, Nombre de contour : 27,
  Pic dans l'orientation du gradient : 0, Nombre de droite : 1, Nombre de cercle : 1.
- **Présence d'un objet : 1**, Nombre de contour : 24,
  Pic dans l'orientation du gradient : 1, Nombre de droite : 4, Nombre de cercle : 0.

D'autres résultats sur des séquences d'images sont présentés sur la figure 5.14. On obtient des temps de calcul très raisonnables d'environ 10 à 20 secondes par image 512*512 sur un PC 3Gh sous Matlab pour un traitement complet de l'image c'est-à-dire jusqu'à la détection des formes particulières (plus des trois quarts du temps est consacré à la détection de cercle). Le temps moyen pour le prétraitement et la segmentation est d'à peu près 3 secondes. Comme nous l'avons vu la recherche de formes particulières ne se fait que si le nombre de contours n'est pas trop élevé pour limiter le temps de calcul et les fausses alertes (des résultats chiffrés et plus détaillés sont présentés dans le chapitre 7). Nous n'avons pas effectué de recherche d'ellipses trop lourde (bien que les cercles soient projetés dans un plan sous la forme d'ellipse), nous avons préféré la détection de cercle par transformée de Hough, avec un minimum de couverture à 20% environ pour accepter des ellipses. L'application d'une telle technique permet d'obtenir la détection d'ellipse dans des temps plus acceptables au prix de quelques compromis. La détection d'ellipse à la place de notre détection de cercle est une extension possible de notre programme qui permettrait de nettement améliorer les détections tout en diminuant les fausses alarmes. On pourra notamment tester les méthodes optimisées de [145][146].

Les améliorations citées précédemment comme l'utilisation d'un critère de courbure et de régularité sur les contours et l'utilisation de la détection d'ellipse de manière performante permettront facilement de parfaire ces résultats déjà prometteurs. Une meilleure utilisation de la direction du gradient pourra aussi améliorer la robustesse de l'algorithme. Pour l'instant nous nous sommes focalisés essentiellement sur la présence ou non d'un objet manufacturé pour respecter les contraintes liées au projet TOPVISION (voir chapitre 7). A noter que l'algorithme actuel permet déjà d'aller plus loin qu'une simple détection notamment nous avons extrait des paramètres pertinents nécessaires pour l'étape de caractérisation. En effet puisque l'on recherche en majorité des objets tels que des cylindres, des cônes ou troncs de cône, des boules, des parallélépipèdes, ou des tores, les cercles et les droites sont de paramètres déterminants parfois même autosuffisants dans certains cas particuliers.

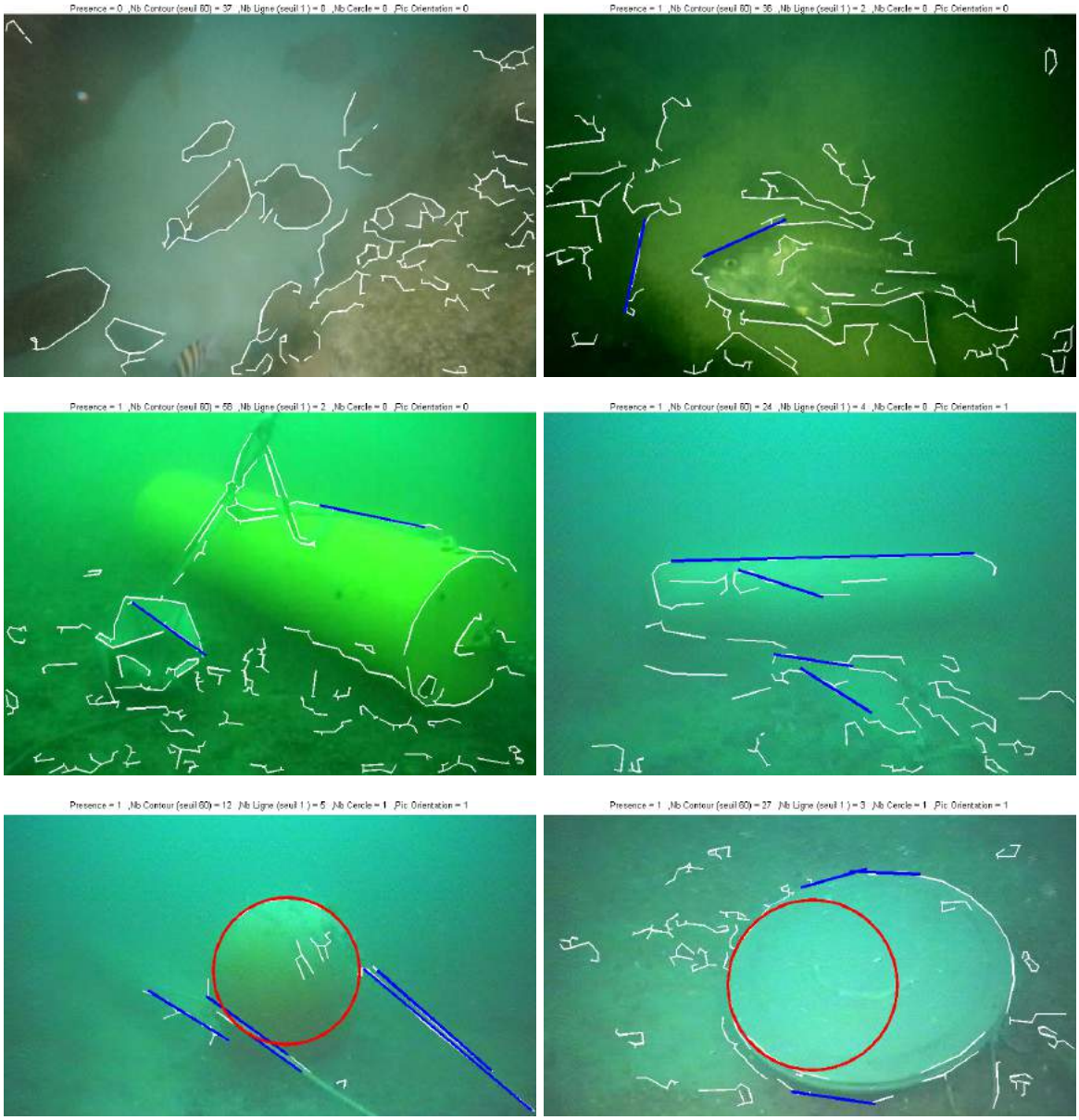


FIG. 5.13 – Résultats de reconnaissance d'objets manufacturés sur 6 scènes d'intérêt, en blanc les contours pertinents polygonalisés, en bleu les droites et en rouge les cercles détectés.

5.6 Conclusion du chapitre

Dans ce chapitre nous avons présenté une méthode classique de reconnaissance d'objets basée sur les contours. Cette méthode automatique, générique et sans *a priori*, permet aujourd'hui de détecter des objets manufacturés dans des images sous-marines. Elle nécessite encore quelques développements pour aboutir à une reconnaissance complète des objets. Nous avons proposé des techniques pour poursuivre sur cette reconnaissance mais ces travaux sont encore en cours aujourd'hui. L'essentiel de la contribution de ce chapitre provient de l'étape de segmentation que nous avons automatisée et adap-

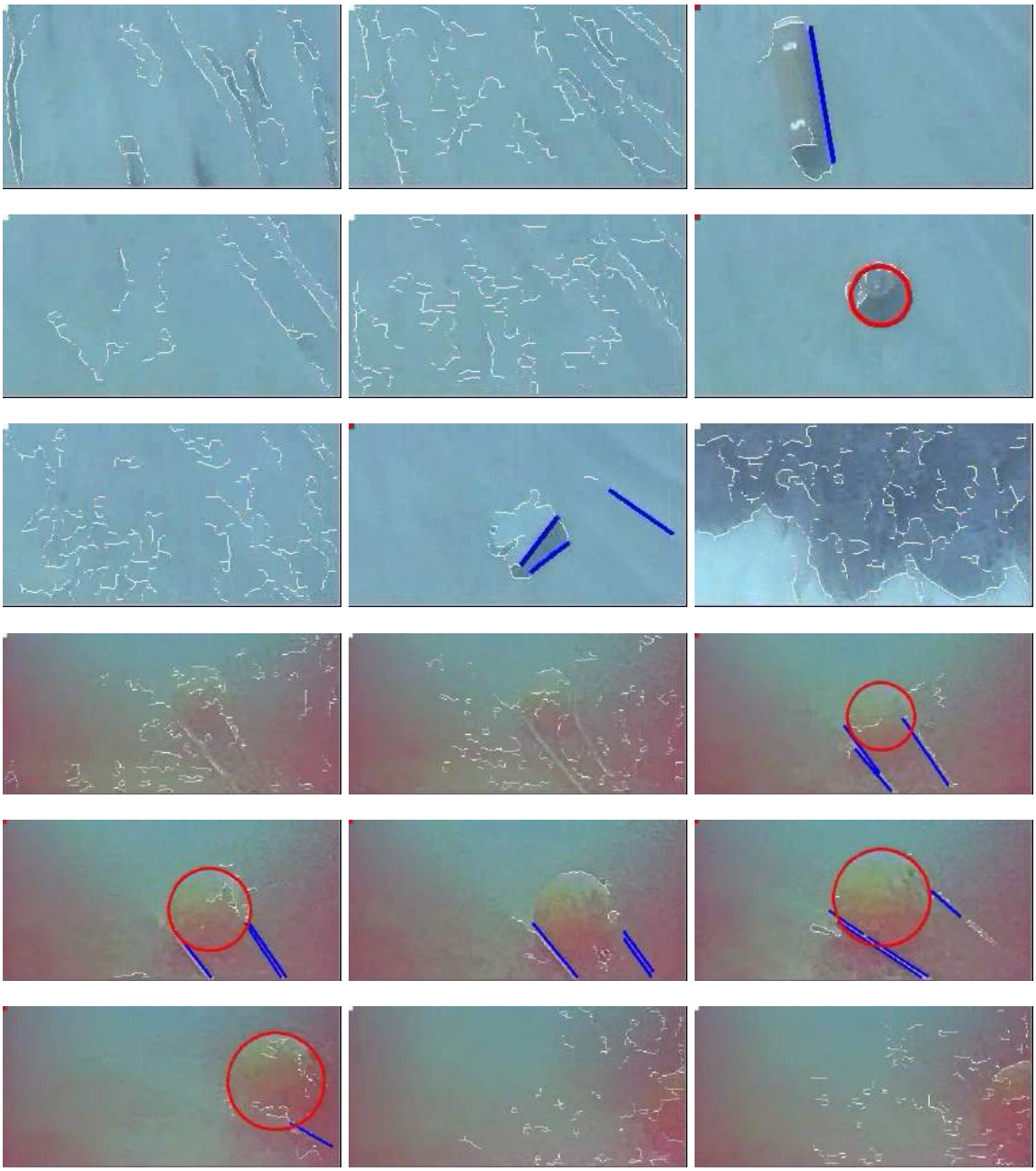


FIG. 5.14 – Résultats de reconnaissance d'objets manufacturés sur 2 séquences d'images.

tée aux conditions difficiles liées au contexte (prétraitement, détection de contours, post-traitement). En effet l'application d'un extracteur de contour tel quel sur les images dont nous disposons donne des résultats impossibles à exploiter. Cette méthode se veut universelle et permet donc à la suite du prétraitement présenté dans le chapitre précédent d'obtenir des contours pertinents dans les images sous-marines. A la suite de ce développement sur la segmentation, nous avons développé une méthode de détection d'objets manufacturés répondant à la problématique TOPVISION. Cette méthode de

détection robuste et sans ajustement de paramètres est basée sur une décision prise en deux temps : d'abord une détection de scène d'intérêt puis le cas échéant une détection de l'objet manufacturé présent. Les résultats présentés montrent que l'utilisation de la méthode classique est possible dans ce contexte moyennant quelques adaptations spécifiques. Aucune donnée *a priori* (ou données d'apprentissage) n'est aujourd'hui nécessaire contrairement à la plupart des méthodes de détection d'objets standard [85]. Par rapport à la thèse de Olmos [1] nous ne travaillons pas uniquement sur des objets posés sur le fond, nous sommes donc plus robuste aux différents types de séquences et nous caractérisons de manière plus précise les objets grâce à l'extraction de primitives de forme. Les perspectives d'amélioration de l'algorithme présenté sont nombreuses comme nous l'avons vu dans les paragraphes précédents. L'une d'entre elle que nous avons peu développée, est l'utilisation de la cohérence de la séquence ou le suivi d'objet dans la séquence qui permettrait d'approfondir la robustesse en allant vérifier dans le temps si les détections étaient bien fondées. En effet l'aspect temporel des séquences traitées est une information très importante, on peut donc facilement faire un "lissage" des résultats supposant qu'une même détection apparaît sur un certain nombre d'images consécutives puisque les mouvements de la caméra sont plutôt lents. Une autre information que nous avons négligée est la connaissance *a priori* des objets recherchés. En effet pour cette méthode nous avons tenté de rester le plus générique possible. Dans le chapitre suivant nous allons étudier des méthodes utilisant cet *a priori* fort de connaissance des objets recherchés en travaillant donc sur des méthodes beaucoup plus spécialisées.

Chapitre 6

Des méthodes de reconnaissance spécialisées à un type d'objet

Dans le chapitre précédent, nous avons présenté une adaptation des méthodes classiques de segmentation au milieu sous-marin. Ces méthodes peuvent être utilisées cependant l'aboutissement de la reconnaissance est généralement difficile. C'est pourquoi nous avons exploré de nouvelles méthodes plus spécialisées mais avec des résultats plus facilement interprétables. Cette spécialisation qui est en fait un apport d'informations supplémentaires permet d'améliorer les résultats en milieu difficile. La majeure partie de ce chapitre concernera l'apport de la couleur.

6.1 Des méthodes pour s'adapter à un contexte difficile

6.1.1 A l'opposé des méthodes classiques

La méthode classique et générique que nous avons présentée dans le chapitre précédent donne de bons résultats quant à la segmentation en contexte difficile vu la qualité médiocre des images. Cependant, il est toujours délicat de conclure sur l'étape de reconnaissance. En effet, puisque la qualité des images est faible, on dispose de peu d'informations exploitables avec un problème d'appariement très complexe dû à la prise de vue quelconque. Seules les caractéristiques de formes sont utilisables car la texture et la couleur sont inefficaces et peu robustes dans leurs utilisations classiques. Et même en ce qui concerne les contours, ils sont en général fragmentés ou incomplets. On ne peut donc pas espérer par l'utilisation de telles méthodes obtenir des résultats équivalents à ceux obtenus en milieu aérien par exemple. Pour résoudre cette difficulté nous avons exploré une nouvelle approche moins générique mais plus adaptée au contexte perturbé. Nous sommes partis du constat suivant : on cherche en général à reconnaître des caractéristiques dans les images et ensuite on cherche à reconnaître les objets d'après ces caractéristiques. Dans le cas où les caractéristiques sont peu nombreuses, cette approche n'est pas satisfaisante puisqu'il est souvent compliqué de conclure sur la reconnaissance. C'est pourquoi nous avons décidé de spécialiser les méthodes et donc de rechercher la présence ou non d'un objet particulier (approche "Top-Down" voir chapitre 3).

6.1.2 Intérêt de ces méthodes spécialisées

Abandonner la généricité en spécialisant les méthodes à un type d'objet particulier permet d'éviter l'étape de classification fastidieuse, et donc de prendre la décision sur la classe de l'objet directement s'il y a eu détection. Cependant, cela oblige à appliquer l'algorithme pour chaque type d'objet que l'on cherche à reconnaître. L'apport d'une information *a priori* précise est aussi nécessaire pour combler le manque d'information que l'on pourra extraire. Les connaissances *a priori* nécessaires peuvent être par exemple la forme, la texture, la couleur. Les paramètres les plus robustes et autosuffisants dans le domaine terrestre étant le modèle 3D et la couleur car ils sont invariants vis-à-vis de la distance, et des changements d'orientation.
La texture ou la forme n'étant pas facilement identifiables dans les images sous-marines, nous avons développé une méthode basée sur la couleur seule. Elle consiste à rechercher un objet sous-marin par sa couleur sachant que l'on sait caractériser la déformation des couleurs dans l'eau (la déformation de la couleur comprenant sa modification par l'illumination, par l'eau). La couleur devient ainsi un paramètre robuste et efficace. Cette méthode permet d'améliorer la qualité de la reconnaissance et d'accélérer le processus lorsque l'on cherche des objets dont la couleur est connue précisément *a priori*.

6.2 La reconnaissance par la couleur

6.2.1 La couleur dans la reconnaissance d'objets

Les objets du monde réel sont très variés en aspect, taille, couleur, position dans l'espace, etc... Le système de reconnaissance doit donc présenter une très grande adaptabilité. Aujourd'hui, de nombreuses méthodes de reconnaissance existent et permettent de s'adapter à un bon nombre de configurations, nous les avons énumérées dans le chapitre 3 (pour plus de détails, on se rapportera aux ouvrages suivants [37][89]). Cependant on peut remarquer qu'aucune de ces méthodes n'est robuste ni générique. Elles sont très souvent développées pour une application particulière. Le développement de ces méthodes de reconnaissance est aujourd'hui encore un sujet très ouvert comme l'atteste le nombre de publications annuel sur ce sujet.

Toute méthode utilise différents critères pour identifier les objets qui composent une scène par exemple les dimensions, la forme, la texture ou bien encore la couleur... La couleur est un critère très pertinent pour la reconnaissance d'objet dans les images. Elle est utilisée très souvent dans la littérature pour le domaine aérien (terrestre) [147][148][95][149][150]. Les méthodes étant en général basées sur les changements d'espaces couleurs, les histogrammes de couleurs ou la normalisation des couleurs. Malgré cela, le critère de couleur souffre de plusieurs phénomènes perturbateurs. Alors que la perception de la couleur par le système visuel humain est peu affectée par le type d'éclairage (naturel, incandescent, fluorescent, halogène...), la vision artificielle est très influencée par les conditions d'illumination. C'est ce qu'on appelle le problème de constance de couleur (en anglais "color constancy"). De nombreuses publications ont traité de ce sujet [151][152][96][153][154].

En milieu sous-marin, à cause des propriétés optiques particulières de propagation de la lumière dans l'eau (notamment les phénomènes d'absorption et de diffusion [155]), les couleurs sont fortement atténuées et modifiées ce qui accentue encore le problème de constance de couleur. Et c'est pour

cette raison que la littérature liée à la reconnaissance d'objets sous-marins n'utilise en général que des méthodes basées sur la forme [156][1]. Il existe tout de même quelques techniques pour réduire le problème de constance de couleur (par exemple [157][57]), mais en raison de ces difficultés supplémentaires la couleur dans l'eau est un critère rarement utilisé, et toujours comme un critère secondaire, jamais comme un critère autosuffisant (on peut citer [96][158][159][160]). Malgré cela, la couleur reste un critère très pertinent pour la reconnaissance d'objet voire même le plus pertinent dans certains contextes. En effet, elle offre une liste d'avantages significatifs comme par exemple sa simplicité de calcul, et sa robustesse face aux occlusions partielles, aux rotations, aux changements d'échelle ou de résolution. Nous avons donc décidé de l'appliquer dans le domaine sous-marin en prenant en compte le problème de constance de couleur évoqué. Deux approches permettent de prendre en considération ce problème. Soit on corrige et rehausse les couleurs (voir [161][162][2]) puis on utilise les méthodes standards, soit on développe une méthode d'identification prenant en compte directement le phénomène. Nous avons dans notre cas opté pour la seconde solution.

6.2.2 Perception de la couleur d'un objet

La perception colorée d'un objet est liée à différentes composantes physiques, physiologiques et psychologiques :

- la composition spectrale de la lumière qui éclaire l'objet coloré,
- la réflectance[1] spectrale ou communément la couleur de l'objet observé,
- la propagation de la lumière dans le milieu,
- la perception physiologique de l'oeil de l'observateur,
- l'interprétation faite du stimulus visuel par l'observateur.

Nous nous contenterons, dans le cadre de cette étude, d'étudier les phénomènes purement physiques :

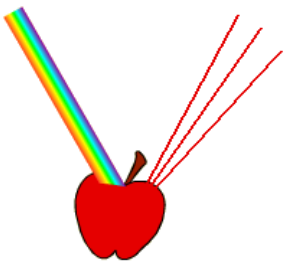


FIG. 6.1 – Absorption sélective et diffusion de la couleur par un objet.

La première composante citée est donc l'éclairage : toute source lumineuse peut être caractérisée par sa répartition spectrale, c'est-à-dire l'énergie émise par intervalle de longueur d'onde. C'est la nature de cette répartition qui induit la couleur apparente de la source. Le même objet paraîtra de couleurs différentes selon qu'il est éclairé par le soleil, une lampe à incandescence ou un spot lumineux, de couleur verte par exemple.
La deuxième composante est la réflectance spectrale de l'objet. En effet la perception de la couleur d'un objet, dépend principalement des propriétés de sa surface (on considère des surfaces Lambertiennes[2]). L'absorption de la lumière est caractérisée par la fraction d'énergie absorbée en fonction de la longueur d'onde. L'objet éclairé réémet une partie de la lumière qui l'éclaire. Par exemple, un point rouge pur (monochromatique) absorbe toutes les couleurs et ne diffuse que le rouge. La couleur d'un point est donc le résultat d'une soustraction (une absorption sélective) de couleurs à la lumière blanche ; on

[1] Reflectance : Fraction de lumière incidente réémise par réflection.
[2] La luminance d'une surface dite "Lambertienne" est isotrope, c'est à dire que les rayons lumineux incidents sont diffusés de sorte que la brillance apparente de la surface soit la même quel que soit l'angle de vue de l'observeur.

parle de "synthèse" soustractive des couleurs (voir Fig. 6.1).
La troisième composante physique prise en compte est la propagation de la lumière dans le milieu (par exemple l'air ou l'eau).

6.2.3 Principe d'illumination d'un objet

Nous venons de voir que lorsqu'un objet reçoit de la lumière, il l'absorbe et en réfléchit une partie. Cependant, la couleur d'un objet n'est jamais uniforme, elle dépend en effet de différents paramètres qui contrôlent ce phénomène de réflexion. Ces paramètres sont :

- la position et la direction de l'objet,
- la position et la direction de l'observateur,
- la position et la direction de la source lumineuse,
- les propriétés d'absorption de l'objet,
- la couleur et l'intensité de la lumière générée par la source.

Deux phénomènes de réflexion permettent de prendre en compte ces paramètres : la réflexion diffuse et la réflexion spéculaire. Comme nous l'avons dit nous considérons des objets avec une surface Lambertienne, on considérera donc la composante diffuse selon le modèle de réflexion de Lambert.

6.2.4 Modèle de réflexion de Lambert

Lorsqu'un rayon lumineux rencontre un point d'un objet, il est réfléchi et plusieurs nouveaux rayons se forment. Leurs couleurs et intensités dépendent des propriétés d'absorption de l'objet et de l'angle d'incidence qui se forme entre le rayon incident et la normale à l'objet (voir Fig. 6.2). Ce phénomène est appelé : la réflexion diffuse.

Considérons θ l'angle formé par le rayon incident et une normale à l'objet (angle d'incidence), on peut écrire que lorsque θ augmente, l'intensité lumineuse baisse et lorsque θ baisse, l'intensité lumineuse augmente. On remarque donc que l'intensité lumineuse est proportionnelle à l'angle θ. La relation entre ces deux valeurs est appelée : loi de Lambert. On constate que l'intensité est nulle quand θ est supérieur à 90°, et est à son maximum quand θ est nul. La loi de Lambert s'écrit de la manière suivante :

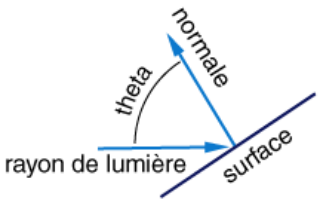


FIG. 6.2 – La réflexion diffuse

$$I_r = I_i.S.cos(\theta). \tag{6.1}$$

avec I_r l'intensité de la lumière reçue, I_i l'intensité de la lumière incidente, S une constante liée à l'objet qui représente la fraction de lumière diffuse (pouvant varier entre 0 et 1) et θ l'angle d'incidence.

6.3 Des couleurs compatibles d'un point de vue sous-marin

6.3.1 Propagation de la lumière dans l'eau

Comme nous l'avons vu précédemment, la lumière est une radiation électromagnétique. Quand elle se propage dans l'eau à une vitesse d'environ $2.2 \times 10^8 ms^{-1}$, son intensité décroît exponentiellement au fur et à mesure que la distance à la source augmente (voir [155][163] pour plus de détails sur la propagation de lumière dans l'eau). Cette perte d'intensité appelée atténuation a deux causes principales :

- l'absorption qui implique une conversion de l'énergie électromagnétique sous forme de chaleur ou d'énergie chimique par exemple. Les principaux éléments absorbants sont :
  - les algues (phytoplancton),
  - les particules organiques et inorganiques en suspension,
  - les substances dissoutes,
  - et l'eau elle-même,
- la diffusion qui provoque des changements aléatoires de la direction des rayons lumineux (dus notamment à la présence de particules en suspension). La diffusion par les très petites particules est généralement à faibles angles. On ne considère pas la rétrodiffusion.

Les radiations électromagnétiques sont caractérisées par leurs longueurs d'onde et leurs intensités. La diffusion de la lumière est largement indépendante de la longueur d'onde mais l'absorption ne l'est pas.

6.3.2 Loi de Beer-Lambert

La loi de Beer-Lambert est une relation empirique reliant l'absorption de la lumière aux propriétés des milieux dans lesquels elle passe (voir [164][165][166][160]).

La loi de Beer-Lambert établit une proportionnalité entre la concentration d'une entité chimique en solution, l'absorbance de celle-ci et la longueur du trajet parcouru par la lumière dans la solution (voir Fig. 6.3).
Soit un rayonnement électromagnétique de longueur d'onde λ traversant un milieu transparent, l'intensité de ce rayonnement subit une diminution exponentielle en fonction de la distance parcourue et de la densité d'absorbants dans ce milieu.

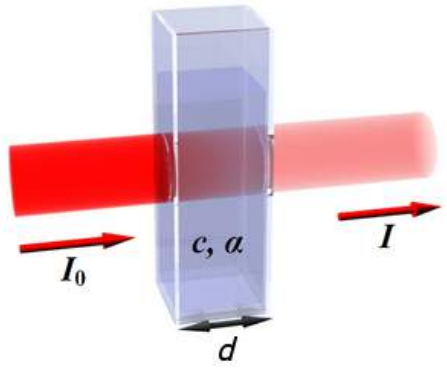


FIG. 6.3 – Atténuation de la lumière traversant un milieu d'épaisseur d, de constante d'absorption α et une concentration d'absorbants c.

$$I_{\lambda,d} = I_{\lambda,0}.e^{\alpha_\lambda.c.d} = I_{\lambda,0}.e^{-k_\lambda.d} \tag{6.2}$$

où $I_{\lambda,0}$ est l'intensité de la lumière incidente, $I_{\lambda,d}$ l'intensité de la lumière sortante après un trajet optique de longueur d, c la densité de l'entité absorbante, α_λ le coefficient d'absorption. On note k_λ le coefficient d'atténuation équivalent.

6.3.3 Équivalence des couleurs en milieu sous-marin

On a pu voir dans la partie précédente que, dans l'eau, la lumière subit plusieurs modifications au cours de son trajet entre la source (émetteur) et l'objet puis entre l'objet et la caméra (récepteur). On peut représenter l'ensemble de ces modifications par le schéma 6.4 et l'équation 6.3 correspondante.

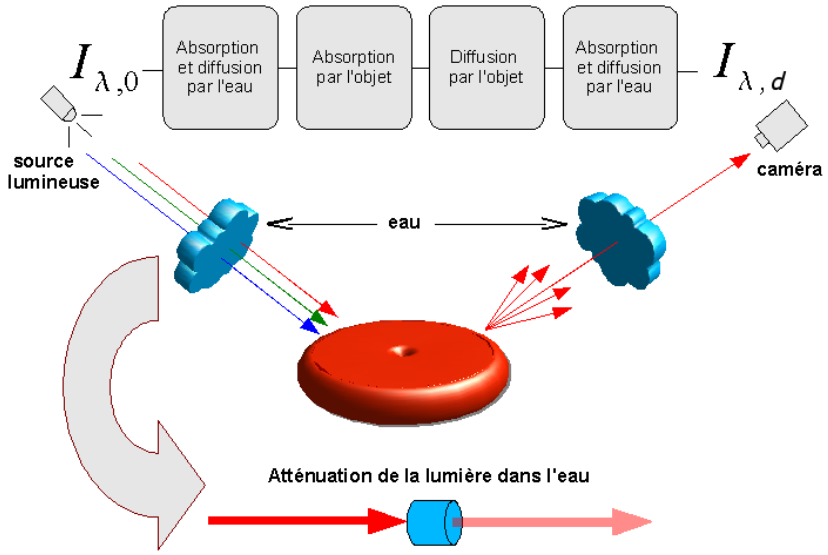


FIG. 6.4 – Modèle représentant les différentes transformations subies par la lumière au cours de son trajet entre la source et la caméra.

$$\begin{pmatrix} I_{R,d} \\ I_{G,d} \\ I_{B,d} \end{pmatrix} = \begin{pmatrix} e^{-k_R.d_2} & 0 & 0 \\ 0 & e^{-k_G.d_2} & 0 \\ 0 & 0 & e^{-k_B.d_2} \end{pmatrix} . \begin{pmatrix} S.\cos(\theta) & 0 & 0 \\ 0 & S.\cos(\theta) & 0 \\ 0 & 0 & S.\cos(\theta) \end{pmatrix}$$

$$. \begin{pmatrix} a_R & 0 & 0 \\ 0 & a_G & 0 \\ 0 & 0 & a_B \end{pmatrix} . \begin{pmatrix} e^{-k_R.d_1} & 0 & 0 \\ 0 & e^{-k_G.d_1} & 0 \\ 0 & 0 & e^{-k_B.d_1} \end{pmatrix} . \begin{pmatrix} I_{R,0} \\ I_{G,0} \\ I_{B,0} \end{pmatrix} \quad (6.3)$$

Ce qui revient à l'équation 6.4 où l'on suppose $I_{\lambda,0} = 1$, où l'on définit un nouveau terme ℓ regroupant le phénomène de diffusion $S.\cos(\theta)$ de la lumière et l'intensité lumineuse incidente $(I_{\lambda,0})$. Nous appellerons ce terme la luminance. Pour finir on note d la distance $d_1 + d_2$, et a_λ les coefficients d'absorption de l'objet.

$$I_{\lambda,d} = \ell.a_\lambda.e^{-k_\lambda.d} \quad (6.4)$$

A partir de l'équation précédente, et supposant connu $\overline{y_\lambda}$ et k_λ, on peut définir la fonction f :

$$f \left\| \begin{array}{l} \mathbb{R}^2 \to \mathbb{R}^3 \\ d, \ell \to \left\{ \begin{array}{l} y_R = \ell.\overline{y_R}.e^{-k_R.d} \\ y_G = \ell.\overline{y_G}.e^{-k_G.d} \\ y_B = \ell.\overline{y_B}.e^{-k_B.d} \end{array} \right. \end{array} \right.$$

Définition 1 (Couleurs compatibles) *Dans un milieu sous-marin, si l'on considère un objet coloré ponctuel éclairé par une source ponctuelle de lumière blanche alors on définit comme couleurs compatibles, toutes les couleurs de cet objet coloré perceptibles sous différentes intensités d'éclairage et à différentes distances.*

Théorème 1 (Compatibilité des couleurs sous-marines) *Soit un milieu sous-marin caractérisé par le vecteur d'absorption $k(k_R, k_G, k_B)$, et un objet coloré ponctuel de couleur $\overline{y}(\overline{y}_R, \overline{y}_G, \overline{y}_B)$ éclairé par une source ponctuelle de lumière blanche alors toute couleur $y(y_R, y_G, y_B)$ est compatible avec la couleur $\overline{y}$ si* $\exists\, d, \ell\ :\ \boldsymbol{y} = f_{k,\overline{y}}(d, \ell)$.

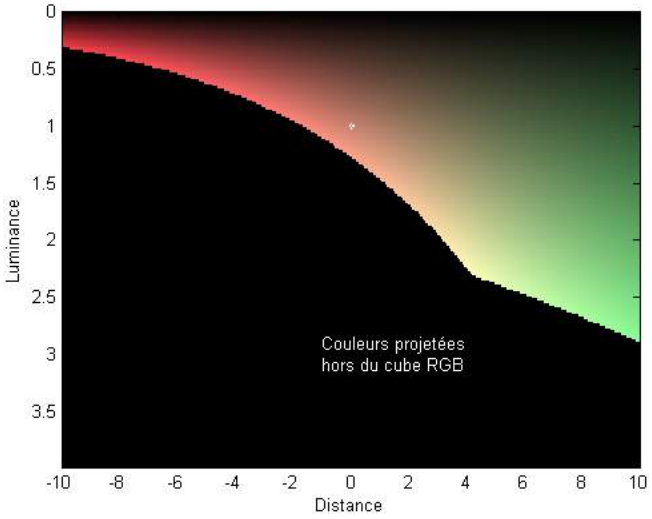


FIG. 6.5 – Ensemble des couleurs compatibles avec la couleur $\overline{y}(200, 130, 110)$ dans un milieu sous-marin $k(0.142, 0.039, 0.076)$ avec $d \in [-10, 10]$, et $\ell \in [0, 4]$. La couleur nominale $\overline{y}$ est supposée à $d = 0$ et $\ell = 1$, elle est entouré par un carré blanc. Les distances négatives ont un sens uniquement dans le cas où la couleur nominale de l'objet $\overline{y}$ est une couleur perçue sous l'eau à une distance non nulle.

En supposant le modèle précédent, caractérisé par l'équation 6.4, on décide d'étudier deux situations dans l'eau.

1. D'abord, on considère une lumière ponctuelle blanche d'intensité ℓ_a, qui éclaire un objet ponctuel d'absorption (a_R, a_G, a_B). Une caméra observe ce point et on note (y_R, y_G, y_B) le triplet RGB correspondant à la couleur perçue. La distance totale parcourue par la lumière (partant de l'émetteur, en passant par le point et jusqu'à la caméra) est noté d_a.
2. Ensuite, seule l'intensité de la lumière et la position des objets changent. On note ℓ_b cette nouvelle intensité, d_b la nouvelle distance ℓ_b et (z_R, z_G, z_B) la nouvelle couleur perçue.

On note aussi (k_R, k_G, k_B) les trois coefficients d'atténuation des trois longueurs d'onde rouge, vert et bleu, supposés connus. On suppose pour les calculs suivants que $y_R, y_G, y_B, z_R, z_G, z_B, \ell_a, \ell_b > 0$ on ne considère donc pas les pixels noirs "purs" correspondant à l'absence totale de couleurs ainsi que les intensités lumineuses nulles correspondant à l'absence d'éclairage. Les deux couleurs (y_R, y_G, y_B) et (z_R, z_G, z_B) perçues lors de ces deux expériences sont deux couleurs émanant d'un même objet, on

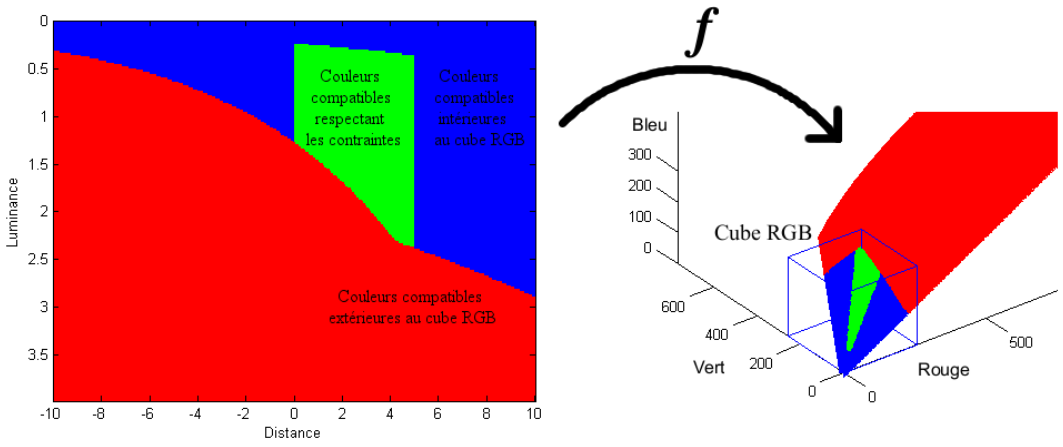


FIG. 6.6 – Illustration des couleurs compatibles dans l'espace d, ℓ et dans l'espace RGB. Les contraintes de distance et de luminance permettant de limiter les couleurs compatibles seront détaillées dans le paragraphe 6.11

considère donc que ces deux couleurs sont compatibles d'un point de vue sous-marin.
On obtient donc le système suivant :

$$\begin{cases} y_R = \ell_a . a_R . e^{-k_R . d_a} \\ y_G = \ell_a . a_G . e^{-k_G . d_a} \\ y_B = \ell_a . a_B . e^{-k_B . d_a} \\ z_R = \ell_b . a_R . e^{-k_R . d_b} \\ z_G = \ell_b . a_G . e^{-k_G . d_b} \\ z_B = \ell_b . a_B . e^{-k_B . d_b} \end{cases}$$

Simplifions un peu le système précédent. Éliminons d'abord les variables (a_R, a_G, a_B) par division des équations, on obtient :

$$\begin{cases} \frac{y_R}{z_R} = \frac{\ell_a}{\ell_b} . e^{k_R . (d_b - d_a)} \\ \frac{y_G}{z_G} = \frac{\ell_a}{\ell_b} . e^{k_G . (d_b - d_a)} \\ \frac{y_B}{z_B} = \frac{\ell_a}{\ell_b} . e^{k_B . (d_b - d_a)} \end{cases}$$

De la même manière, éliminons le rapport d'intensité $\frac{\ell_a}{\ell_b}$, on obtient :

$$\begin{cases} \frac{y_R . z_G}{z_R . y_G} = e^{(k_R - k_G) . (d_b - d_a)} \\ \frac{y_R . z_B}{z_R . y_B} = e^{(k_R - k_B) . (d_b - d_a)} \end{cases} \Longrightarrow \begin{cases} \log(\frac{y_R . z_G}{z_R . y_G}) = (k_R - k_G) . (d_b - d_a) \\ \log(\frac{y_R . z_B}{z_R . y_B}) = (k_R - k_B) . (d_b - d_a) \end{cases}$$

Enfin, éliminons le terme $(d_b - d_a)$, on obtient :

$$(k_R - k_B) . \log(\frac{y_R . z_G}{z_R . y_G}) - (k_R - k_G) . \log(\frac{y_R . z_B}{z_R . y_B}) = 0. \tag{6.5}$$

Si on note K la constante $(\frac{k_R - k_G}{k_R - k_B})$, on a :

$$(1 - K) . \log(y_R) - \log(y_G) + K . \log(y_B) = (1 - K) . \log(z_R) - \log(z_G) + K . \log(z_B). \tag{6.6}$$

Définition 2 (Teinte sous-marine) *Pour toute couleur $x(x_R, x_G, x_B)$, dans un milieu sous-marin caractérisé par le vecteur d'absorption $k(k_R, k_G, k_B)$ on définit la teinte sous-marine $h_K(x)$ par :*

$$h_K(\mathbf{x}) = (1-K).\log(x_R) - \log(x_G) + K.\log(x_B) \qquad \text{avec } K = (\frac{k_R - k_G}{k_R - k_B}).$$

Théorème 2 (Compatibilité des couleurs sous-marines) *Deux couleurs $y(y_R, y_G, y_B)$ et $z(z_R, z_G, z_B)$ sont compatibles d'un point de vue sous-marin si* $\exists\ K : h_K(\boldsymbol{y}) = h_K(\boldsymbol{z})$.

D'après le théorème 2 deux couleurs sont compatibles si il existe K tel que :

$$(1-K).\log(y_R) - \log(y_G) + K.\log(y_B) = (1-K).\log(z_R) - \log(z_G) + K.\log(z_B).$$

En isolant K on peut écrire :

$$K = \frac{\log(z_R) - \log(y_R) - \log(z_G) + \log(y_G)}{\log(z_R) - \log(y_R) - \log(z_B) + \log(y_B)}. \tag{6.7}$$

Définition 3 (Fonction de compatibilité) *Soit deux couleurs $y(y_R, y_G, y_B)$ et $z(z_R, z_G, z_B)$ on peut définir la fonction de compatibilité sous-marine ψ tel que :* $\psi(\boldsymbol{y},\boldsymbol{z}) = \frac{\log(z_R)-\log(y_R)-\log(z_G)+\log(y_G)}{\log(z_R)-\log(y_R)-\log(z_B)+\log(y_B)}$.

Corollaire 1 *Une couleur $\boldsymbol{z}$ est compatible avec la couleur donnée $\boldsymbol{y}$ si* $\exists\ K : \psi(\boldsymbol{y},\boldsymbol{z}) = K$.

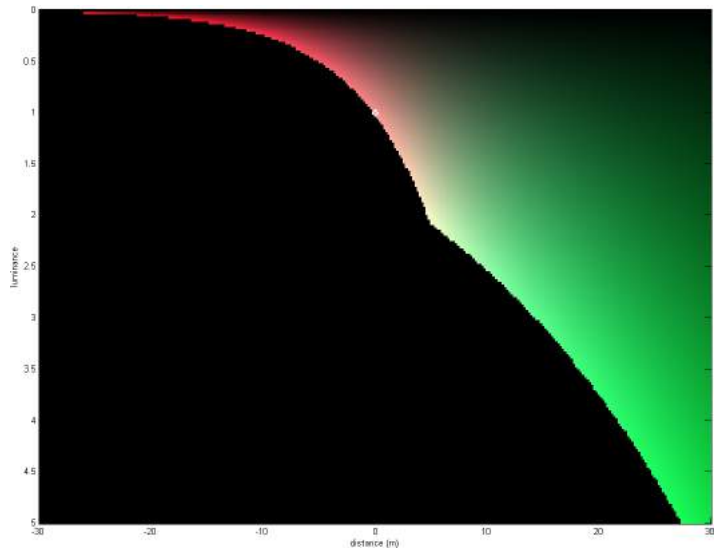
Et, de manière plus robuste puisque K peut ne pas être connu avec précision, pour limiter les non détections et le risque de fausse alerte, on peut supposer que l'on peut encadrer K par un intervalle $[K_{min}, K_{max}]$.

$$p \begin{pmatrix} d \\ l \end{pmatrix} \underset{g}{\overset{f}{\rightleftarrows}} y \begin{pmatrix} y_R \\ y_G \\ y_B \end{pmatrix} \xrightarrow{\psi} K$$

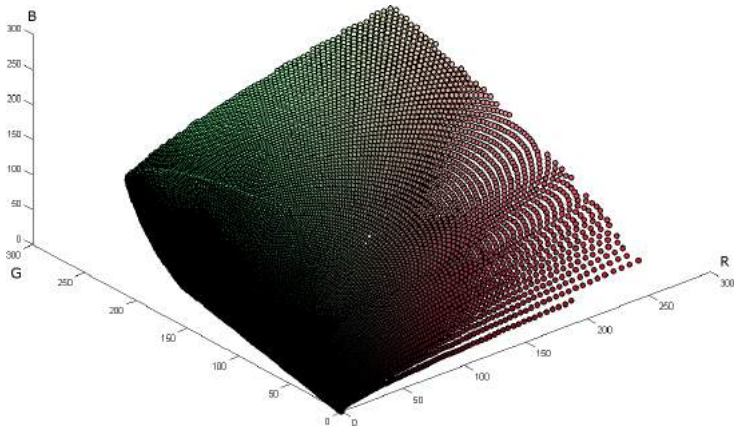
Théorème 3 (Compatibilité robuste des couleurs sous-marines) *Une couleur $y(z_R, y_G, y_B)$ est compatible avec la couleur donnée $\overline{y}(\overline{y}_R, \overline{y}_G, \overline{y}_B)$ dans un milieu sous-marin caractérisé par l'intervalle $[K_{min}, K_{max}]$ si :* $\psi(\overline{y}, y) \in [K_{min}, K_{max}]$.

L'équation 6.6 est la **condition de compatibilité de couleur**. Autrement dit, lorsque l'on connaît une couleur, toutes les couleurs ayant la même teinte, représentent la même couleur source sous différentes distances, positions et intensités lumineuses. Ce travail sur la compatibilité de couleur est une des principales contributions de cette thèse.

Deux couleurs **y** et **z** sont donc compatibles c'est-à-dire qu'elles proviennent d'un même objet si le théorème 3 est vérifié. A partir de cette équation et connaissant la couleur d'un objet, on peut donc retrouver toutes les couleurs que pourra avoir cet objet à différentes distances, sous différentes positions et sous différentes intensités d'éclairage. Supposant K connu, l'ensemble des couleurs compatibles appartiennent à une surface dans l'espace couleur RGB (voir Fig. 6.7(b)). De manière à illustrer cet ensemble de couleurs dites compatibles, nous avons utilisé deux représentations. D'abord sur la figure 6.7(a) nous avons représenté la modification des couleurs en fonction de la distance et de la luminance. Pour ce cas d'étude, le pixel correspondant à la couleur recherchée est entouré par un rond blanc. On peut observer la modification de la couleur en fonction de la distance suivant l'axe des abscisses et en fonction de luminance suivant l'axe des ordonnées.
Une autre représentation est présentée sur la figure 6.8, dans ce cas nous avons à partir de l'équation



(a)



(b)

FIG. 6.7 – (a) Représentation de l'ensemble des couleurs compatibles avec la couleur (248, 149, 144) avec $k_R = 0.156, k_G = 0.045, k_B = 0.040$ dans l'espace distance-luminance (faisant varier la distance et le alpha définis dans l'équation 6.4), (b) Représentation de l'ensemble précédent dans l'espace RGB.

6.5 généré tous les triplets RGB possibles répondant au théorème 2.

Pour reconnaître un objet par sa couleur en milieu sous-marin, nous venons de voir que nous utilisons la loi de Beer-Lambert. Pour utiliser cette loi, nous avons fait l'hypothèse que les coefficients d'absorption des longueurs d'onde rouge, vert et bleu dans l'eau étaient connus ou que l'on disposait des moyens nécessaires pour les estimer (plusieurs images d'un objet immergé quelconque à différentes distances par exemple).
Dans le cas où les coefficients sont inconnus, la première partie de notre système d'identification va donc être leurs estimations. On appellera cette étape la calibration du système. A chaque fois que l'on changera de milieu, c'est-à-dire que l'eau sera différente, il faudra recommencer cette étape.

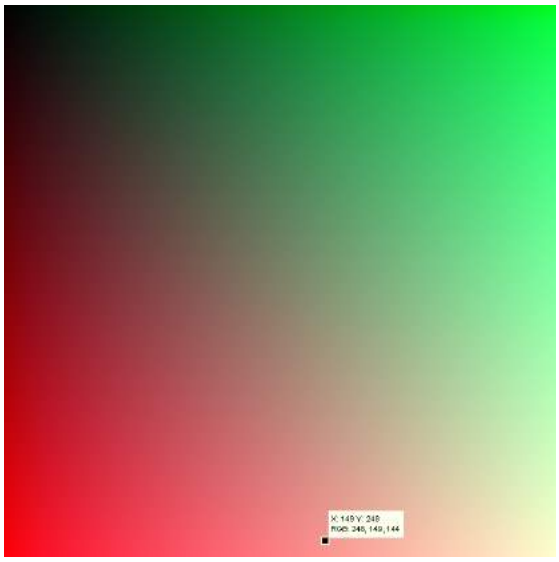


FIG. 6.8 – Ensemble des couleurs compatibles avec la couleur (248, 149, 144) appartenant au cube RGB. Avec $k_R = 0.156, k_G = 0.045, k_B = 0.040$, nous avons calculé B pour chaque valeur de R et G.

Lorsque l'on parle de reconnaissance, on suppose évidemment que chaque objet recherché a une couleur unique. Dans le cas inverse, on ne peut reconnaître complètement l'objet, et il faut utiliser un autre critère (la forme par exemple) pour réellement discriminer l'objet. Par la suite, on confondra parfois les termes "segmentation" et "reconnaissance".

6.4 Les algorithmes de reconnaissance développés

Dans ce rapport, nous allons présenter deux algorithmes de segmentation développés. Tout deux, considèrent que l'on recherche dans des images sous marines un objet de couleur connue *a priori*. Ils ont été développés dans la continuité l'un de l'autre c'est-à-dire que le deuxième algorithme est basé sur le même principe que le premier mais permet de pallier les problèmes de celui-ci (c'est ce deuxième algorithme qui utilise la nappe d'équivalence décrite dans le paragraphe 6.3.3).
Contrairement aux méthodes usuelles qui cherchent à prétraiter les images pour obtenir la constance de couleur, nous avons construit un nouveau modèle permettant de prédire la modifications des couleurs par le milieu sous-marin. En d'autres termes, cette approche permet de calculer l'ensemble des couleurs que pourrait avoir l'objet recherché à différentes distances, dans différentes orientations et sous différentes illuminations. Nous avons appelé cet ensemble de couleurs, "des couleurs compatibles". La segmentation des images en deux classes (couleurs compatibles ou non) permet ensuite de reconnaître les objets par seuillage sur le nombre de couleurs compatibles détectées.

6.4.1 Un modèle simple : la droite de compatibilité

Le premier algorithme considère uniquement l'atténuation de la couleur par la distance (Loi de Beer-Lambert Eq. 6.3). Il nécessite de disposer de plusieurs images (3 au minimum) d'un objet de calibration à différentes distances (ou différentes profondeurs), dans des conditions d'éclairage "constantes"

et dans le "même milieu aquatique" que celui où seront effectués les recherches de l'objet. A noter que cette méthode a fait l'objet d'une publication dans une conférence internationale en Septembre 2007 [167].

Principe de la méthode : Dans le contexte sous-marin, on considère un point lumineux dont la couleur dans l'espace RGB est $\overline{y}(\overline{y}_R, \overline{y}_G, \overline{y}_B)$. Selon la loi de Beer-Lambert (voir paragraphe 6.3.2) la couleur de ce point lumineux reçue à une distance d est donnée par $y(y_R, y_G, y_B)$ avec,

$$y_\lambda = \overline{y}_\lambda . e^{-k_\lambda . d} \quad \text{avec } \lambda \in \{R, G, B\} \tag{6.8}$$

Cette équation devient une droite dans le domaine logarithmique. On a donc :

$$\ln y_\lambda = \ln \overline{y}_\lambda - k_\lambda . d \tag{6.9}$$

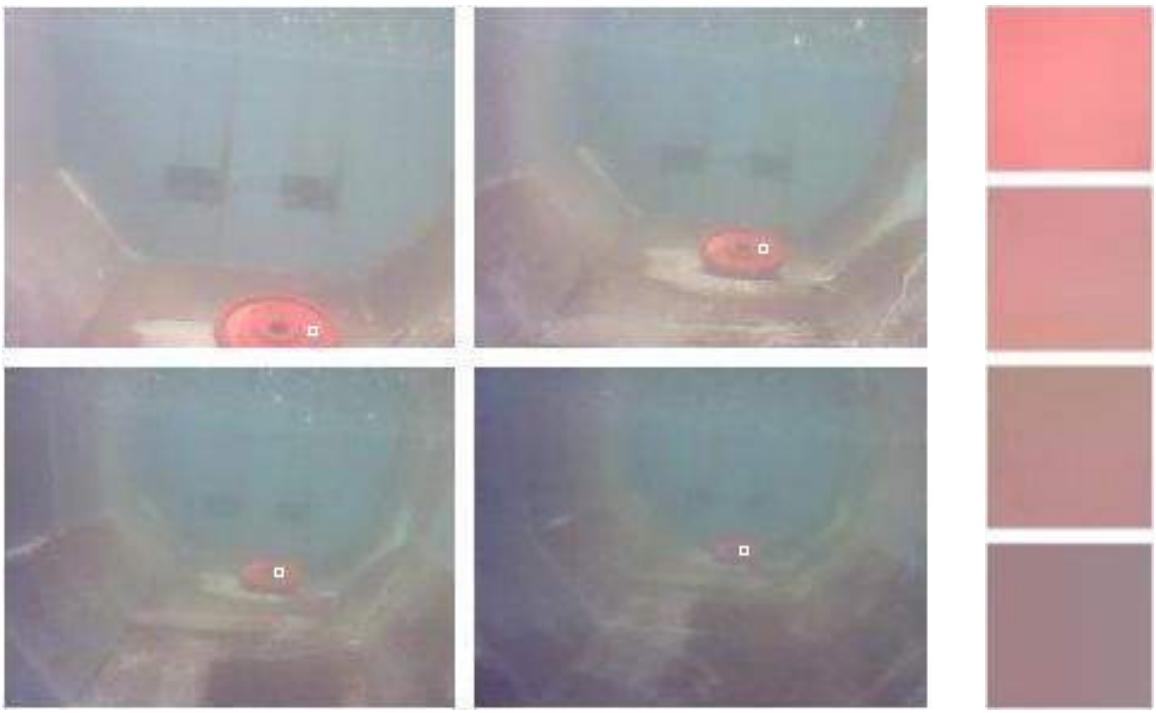
Un pixel y est compatible (du point de vue de la distance) avec un pixel $\overline{y}$ s'il appartient à la droite définie par l'équation paramétrique suivante :

$$\begin{cases} \ln y_R = -k_R . d + \ln \overline{y}_R \\ \ln y_G = -k_G . d + \ln \overline{y}_G \\ \ln y_B = -k_B . d + \ln \overline{y}_B \end{cases} \tag{6.10}$$

La segmentation de l'image par cet algorithme va donc être basée sur ce critère. On va partager les pixels de l'image en deux classes : la classe des couleurs compatibles (pixels qui appartiennent à la droite de compatibilité ou à son voisinage) et les autres.
Dans le pratique on va projeter chaque pixel dans le plan perpendiculaire à la droite de compatibilité appelé plan discriminant, puis, dans ce plan, on va vérifier s'il appartient ou non à la zone de projection de la couleur de l'objet recherché.

Calibration du processus d'identification : Dans le cas où l'on ne dispose pas des coefficients d'atténuation de la lumière propres au milieu sous-marin dans lequel nous allons rechercher l'objet, ce qui se passe d'ailleurs la plupart du temps, il est d'abord nécessaire d'estimer la droite de compatibilité précédemment définie. Pour cela, il suffit de prendre quelques photos de l'objet recherché à différentes distances et sous un même éclairage (sans forcément connaître ces distances). A noter que sur les images présentées les conditions d'éclairage n'étaient pas très stables. Grâce à ces images, dites de calibration (voir Fig. 6.9(a)), on dispose de plusieurs nuages de pixels correspondant à la même couleur, à différents éloignements, comme on peut le voir sur les figures 6.9(b) et 6.10. On peut alors calculer la droite de compatibilité en appliquant une ACP sur ces pixels d'initialisation passés dans l'espace logarithmique. La droite de compatibilité ou droite d'atténuation est définie par l'axe principal de l'ACP.
De manière à être plus stable vis-à-vis des variations d'intensité et d'éclairage [65], les pixels sont tous convertis de l'espace RGB vers l'espace RGB normalisé (voir chapitre 2) avant d'être passés dans le domaine logarithmique.

On peut résumer le calcul de la droite de compatibilité comme suit :
- Récupération des pixels d'initialisation dans une matrice,
- Conversion de ces pixels RGB vers l'espace RGB normalisé,
- Passage dans le domaine logarithmique,



(a) (b)

FIG. 6.9 – (a) Images de calibration à des distances de 2, 3, 4 et 5 mètres prises dans le bassin de l'ENSIETA, (b) Sous-images extraites de l'objet sur chacune des images de calibration. L'ensemble des pixels de ces imagettes constitue ce que l'on appellera l'échantillon d'initialisation.

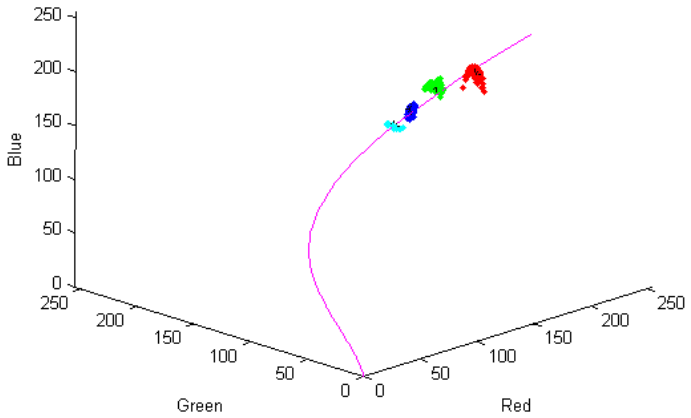


FIG. 6.10 – Pixels d'initialisation à différentes distances Fig. 6.9(b) et la courbe d'atténuation approximée.

– Soustraction de la moyenne sur chaque dimension,
– Calcul de la matrice de covariance,
– Calcul des vecteurs propres et des valeurs propres.

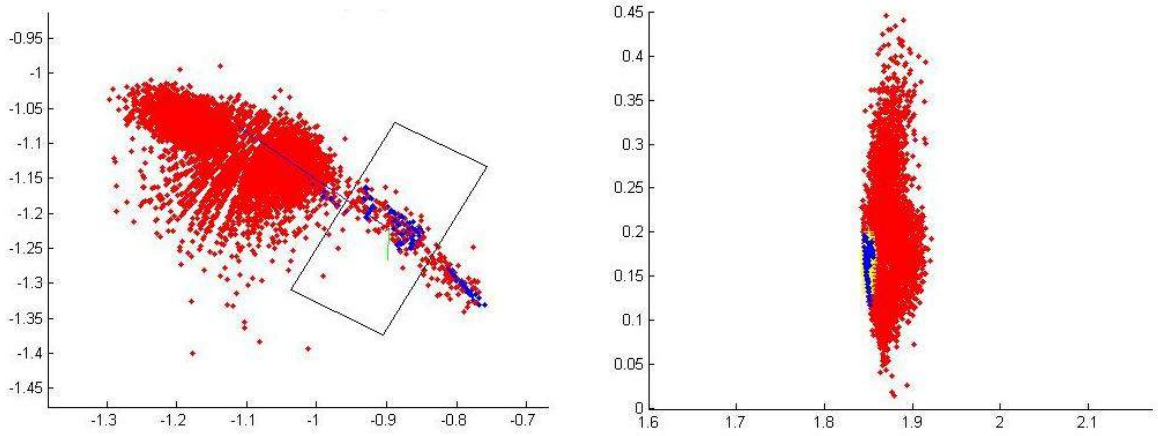
} ACP

Lorsque l'on dispose de cette droite de compatibilité, on peut ensuite calculer la matrice de passage de l'espace log RGB normalisé vers l'espace discriminant défini comme étant le plan perpendiculaire à la droite de compatibilité. Dans ce plan, on projette ensuite quelques pixels correspondants à la couleur de l'objet, puis on les approxime par une ellipse via une autre ACP. C'est cette ellipse, c'est-à-dire la zone de projection de la couleur de l'objet, qui nous permet ensuite de segmenter l'image en deux classes.

Pour résumer la méthode de calibration peut être décomposée de la manière suivante :

- A partir des pixels de calibration.
  - Calcul de la droite de compatibilité.
  - Calcul de la matrice de transformation vers l'espace discriminant (plan perpendiculaire à la droite de compatibilité)
- A partir de quelques pixels sélectionnés sur l'objet recherché dans une image.
  - Approximation de la zone de projection de la couleur de l'objet dans l'espace discriminant.

Segmentation de l'image et décision : Pour classifier l'ensemble des pixels d'une image quelconque dans laquelle on recherche l'objet, et décider s'ils appartiennent à la classe objet ou non, il suffit ensuite de les projeter dans le plan perpendiculaire à la droite d'atténuation précédemment calculée (voir Fig. 6.11(a) et 6.11(b)) aussi appelé plan discriminant. Dans cet espace, chaque pixel est affecté à la classe objet s'il appartient à l'ellipse. La segmentation se fait donc selon l'appartenance à la zone de projection. On peut voir les résultats de cette segmentation sur les figure Fig. 6.12 et 6.13(b). Pour



(a) (b)

FIG. 6.11 – (a) Nuage d'initialisation (bleu), et ses axes principaux (vert), autres pixels des images Fig.6.9(a) (rouge), (b) Projection des pixels précédents dans le plan discriminant. L'ellipse des couleurs compatibles est signalée en jaune.

finir, on décide si l'objet appartient à l'image ou non, en fonction du "regroupement" et du nombre des pixels de couleur compatible qu'elle contient.

En résumé, on décide si une image contient l'objet recherché en appliquant la méthode suivante :

- Pour chaque pixel de l'image.
  - Projection dans l'espace discriminant,
  - Vérification de l'appartenance à l'ellipse et affectation à une classe.
  - Prise de décision en fonction du nombre et du regroupement des couleurs compatibles détectés.

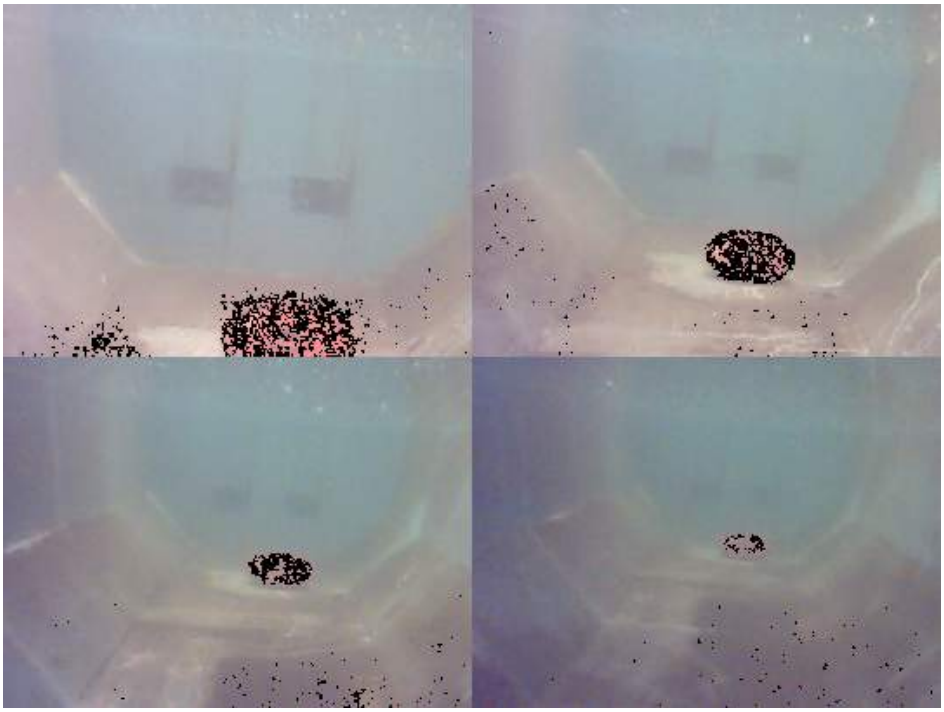
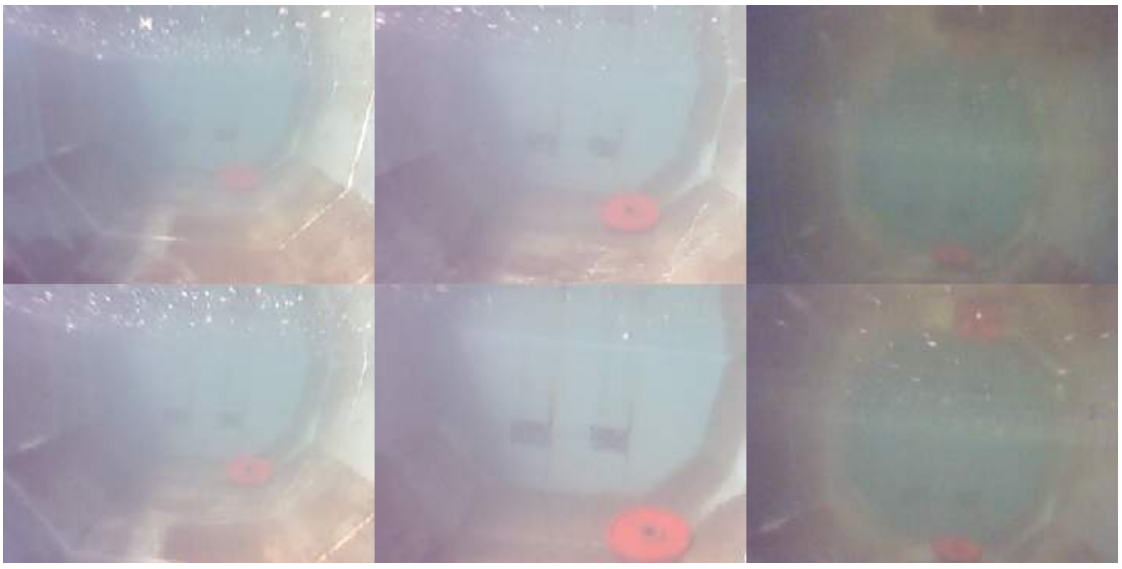


FIG. 6.12 – Détection des pixels des couleurs compatibles avec l'objet sur les images de calibration.

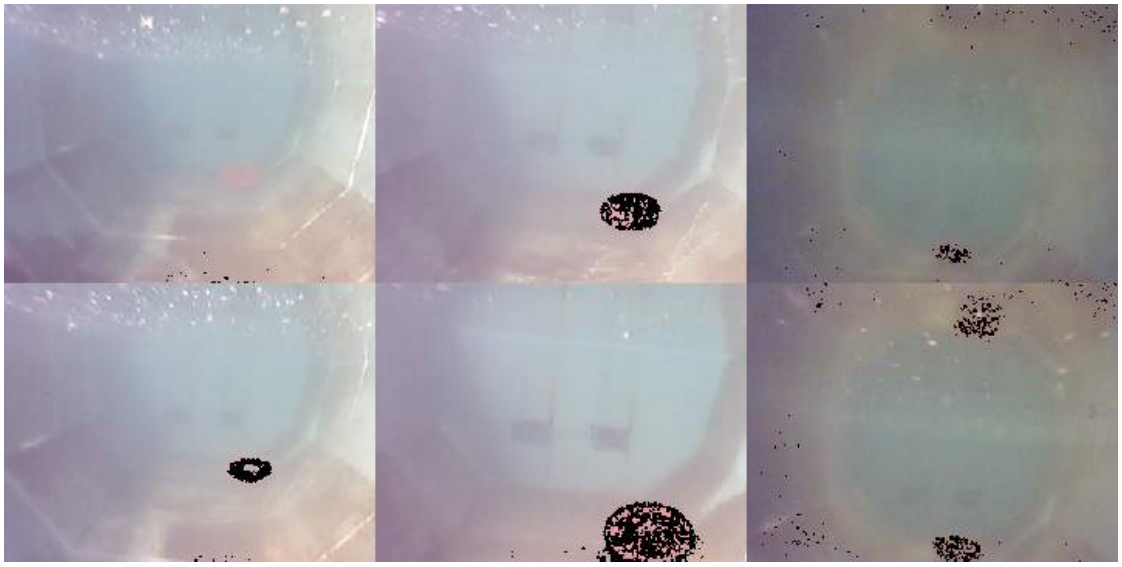
Résultats et critiques de la méthode : Les résultats présentés Fig. 6.13(b) montrent que l'objet est bien détecté, de même que la réflexion de l'objet à la surface de l'eau dans la dernière image. Évidemment quelques fausses alarmes apparaissent mais elles restent très isolées et sont donc facilement supprimables par post-traitement (en utilisant des algorithmes de morphologie mathématique par exemple) ou par l'utilisation de la cohérence de la séquence c'est-à-dire un suivi de l'objet. Plus qu'une méthode de reconnaissance, cette méthode peut être utilisée comme une méthode de segmentation d'image permettant de récupérer l'enveloppe englobante d'un objet connaissant *a priori* sa couleur. Vu les résultats assez encourageants de cette méthode, les travaux de recherche ont continué et le modèle a été affiné pour gérer les différences d'illumination et les différentes orientations de l'objet. La partie suivante présentera l'utilisation de ce modèle amélioré pour la reconnaissance d'objet.

6.4.2 Un modèle amélioré : la nappe de compatibilité

Un modèle plus réaliste : Pour ce deuxième algorithme en plus de considérer l'atténuation de la couleur par la distance, on considère aussi les différentes orientations de l'objet ainsi que les différentes illuminations. En ce qui concerne les hypothèses de départ cet algorithme comme le précédent impose de connaître les coefficients d'atténuation de la lumière dans l'eau. Cette méthode a fait l'objet d'un



(a)



(b)

FIG. 6.13 – Reconnaissance d'un objet par détection des couleurs compatibles sur plusieurs images extraites de vidéos enregistrées dans le bassin de l'ENSIETA.

article soumis à la revue IEEE Oceanic Engineering en Mai 2008.

Estimation des paramètres : Pour ce deuxième algorithme, l'étape de calibration peut se faire de deux façons. Soit juste en estimant K à partir du théorème 2 soit en estimant les paramètres d'absorption de l'eau [168] : k_R, k_G, k_B. Dans le cas où l'on applique des contraintes sur la nappe (voir suite) il est nécessaire d'avoir une estimation précise des coefficients. Nous allons donc expliquer la méthode pour les estimer. Pour cela, il est nécessaire de disposer de plusieurs images (3 au minimum)

d'un même objet à différentes distances connues et ce, sous les mêmes hypothèses que pour la méthode précédente, c'est-à-dire dans des conditions d'éclairage "constantes" et dans le "même milieu aquatique" que celui des recherches.

Cette hypothèse un peu plus contraignante d'un point de vue opérationnel permet de nettement améliorer les résultats de reconnaissance. Un exemple de calcul des coefficients d'absorption dans trois différents bassins à partir des images Fig. 6.15 est présentée ci-après :

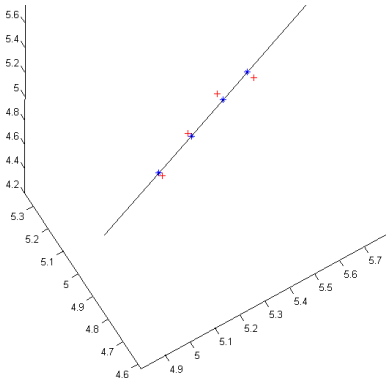


FIG. 6.14 – Droite de régression dans l'espace log RGB obtenue à partir de 4 pixels sélectionnés sur la figure 6.15(a) pour le calcul des coefficients d'absorption. En rouge les pixels initiaux, en bleu les pixels projeté sur la droite.

Bassin ENSIETA :
$k_R = 0.156, k_G = 0.045, k_B = 0.040.$

Bassin GESMA :
$k_R = 0.241, k_G = 0.059, k_B = 0.011.$

Bassin QINETIQ :
$k_R = 0.110, k_G = 0.043, k_B = 0.025.$

A partir de ces différentes images, on dispose de plusieurs triplets RGB correspondant tous à la même couleur, celle de l'objet ayant servi à la calibration. En fonction de la modification de la couleur et connaissant les distances entre chaque image, on peut calculer les coefficients d'absorption du milieu. On peut voir par exemple, 3 échantillons d'images de calibration sur la Fig. 6.15. Le calcul des coefficients d'absorption se fait par une régression linéaire sur plusieurs pixels appartenant à l'objet et sélectionnés dans les images de calibration. Ensuite chaque pixel est projeté orthogonalement sur la droite estimée par la régression. A partir de ces pixels corrigés, on calcule les coefficients d'absorption pour chaque composante grâce à l'équation 6.9. Avant la sélection des pixels de couleur, les images sont lissées (lissage anisotropique par exemple) pour atténuer le bruit d'acquisition qui pourrait dégrader les résultats.

Segmentation de l'image et décision : Connaissant *a priori* la couleur d'un objet, on va donc essayer de retrouver toutes les couleurs que pourra avoir cet objet à différentes distances, sous différentes positions et sous différentes intensités d'éclairage. Comme on a pu le voir dans la partie théorique 6.3.3, les couleurs $\mathbf{y}$ et $\mathbf{z}$ sont compatibles si le théorème 2 est vérifiée c'est-à-dire si :

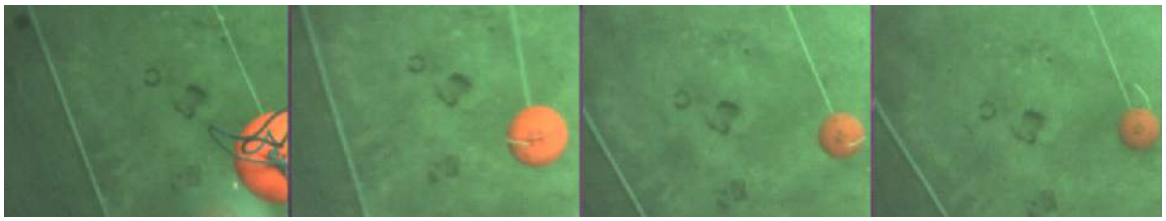
$$\exists\ K : h_K(\mathbf{y}) = h_K(\mathbf{z}).$$

D'un point de vue informatique et dans le contexte d'identification d'objet par la couleur, supposons maintenant que l'on recherche dans une image toutes les couleurs compatibles $y(y_R, y_G, y_B)$ avec une couleur $\overline{y}(\overline{y}_R, \overline{y}_G, \overline{y}_B)$ donnée et connaissant la valeur de K. Alors, une couleur $y(y_R, y_G, y_B)$ est compatible avec la couleur $\overline{y}(\overline{y}_R, \overline{y}_G, \overline{y}_B)$ si elle appartient à la nappe de compatibilité c'est-à-dire si

$$K_{min} < \psi(\overline{y}, y) < K_{max}.$$

(a)

(b)



(c)

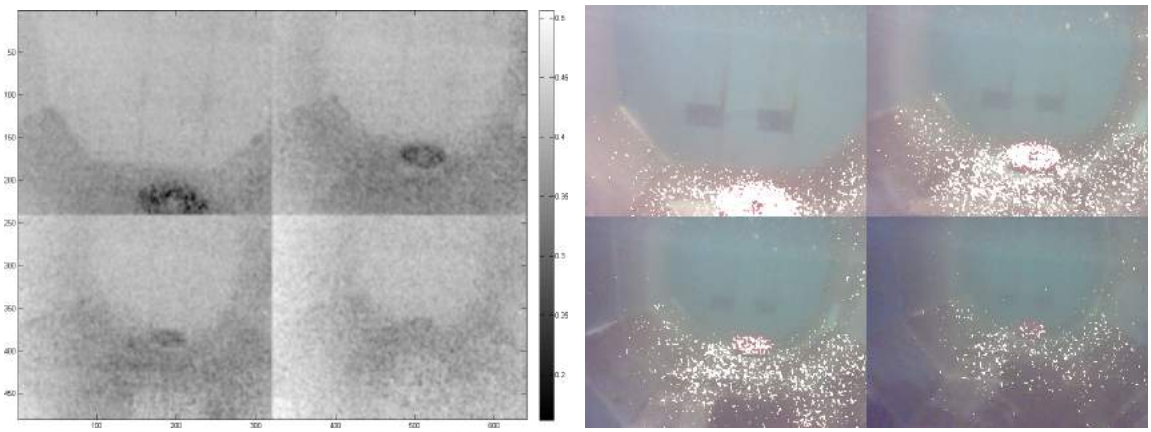
FIG. 6.15 – Différentes images de calibration avec différents objets (a) un poids rouge à 2, 3, 4 et 5m (Bassin ENSIETA) (b) une boule jaune à 1, 2, 3, et 4m (Bassin GESMA) (c) une boule orange à 2, 3, 4, 5m (Bassin QinetiQ Concours SAUC'E 2007).

La fonction $|\psi(\mathbf{z},\mathbf{y}) - K|$ pour la couleur y=(248, 149, 144) est présentée sur la figure 6.16(a) ainsi que les résultats sur les images 6.9(a) après seuillage sur la figure 6.16(b).

Après le seuillage du critère de compatibilité, un objet est identifié dans une image s'il y a des pixels aux couleurs compatibles en nombre suffisant et s'il sont assez regroupés spatialement. De la manière la plus simple on peut par exemple appliquer des opérateurs de morphologie mathématique puis effectuer un seuillage sur le pourcentage de pixels compatibles détectés dans l'image. On peut éventuellement utiliser en plus la cohérence temporelle de la séquence pour améliorer les résultats en diminuant les fausses alarmes.

Contraintes de compatibilité : Pour réduire cet ensemble des couleurs compatibles, nous avons aussi ajouté des contraintes sur la valeur minimale des pixels considérés et sur la distance d'atténuation. Nous avons d'abord négligé les pixels trop sombres qui généraient beaucoup de fausses alarmes. En effet, les couleurs noires sont compatibles avec n'importe quelle autre couleur pourvu qu'on s'en éloigne suffisamment.
Les pixels sont négligés si leurs trois composantes sont inférieures au seuil de couleur minimale, autrement dit, une couleur est acceptée si au moins une de ses composantes dépasse ce seuil.

Nous avons également ajouté une deuxième équation pour limiter la distance étant donné que la



(a) (b)

FIG. 6.16 – (a) Image de distance dans l'espace des couleurs compatibles (mesure de compatibilité). Représentation de la fonction $|\psi(\mathbf{z},\mathbf{y}) - K|$ pour la couleur $\mathbf{y}$=(248, 149, 144). (b) Résultat de segmentation par seuillage de la fonction précédente.

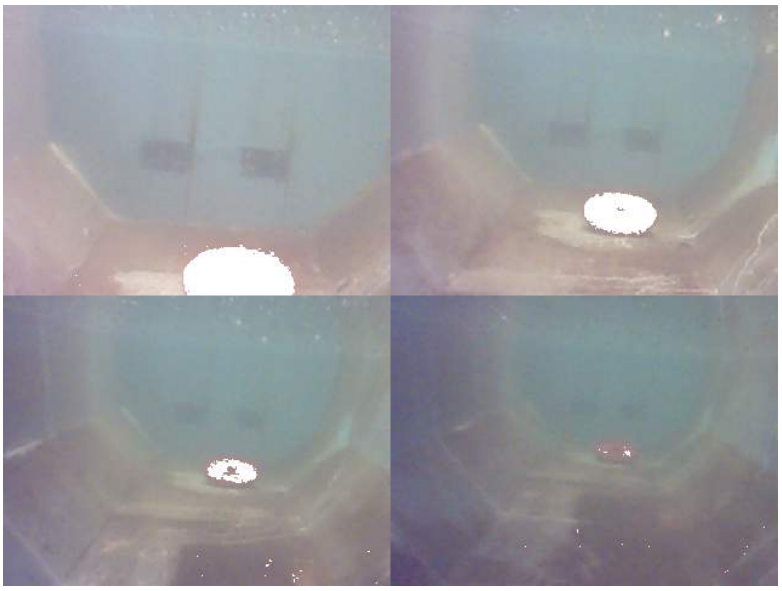
visibilité est très réduite sous l'eau. Cette deuxième contrainte nous permet de supprimer de nombreuses fausses alarmes en limitant considérablement le nombre de couleurs compatibles.
Son équation obtenue à partir de l'équation 6.5 est la suivante :

$$\frac{1}{k_G - k_R} . \log\left(\frac{y_R . \overline{y}_G}{\overline{y}_R . y_G}\right) < d_{max} \tag{6.11}$$

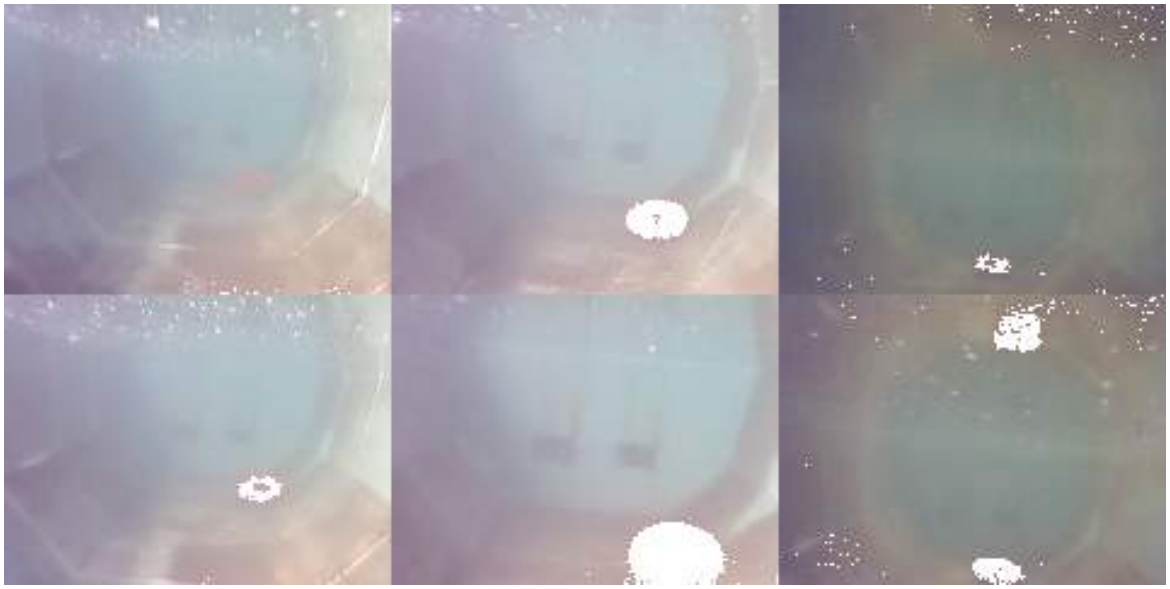
Les résultats de seuillage sous contraintes sont présentés sur les figures 6.17(a) et 6.17(b). L'effet des contraintes sur les nappes de compatibilité est affiché sur les figures 6.18 et 6.19. Sur la figure 6.18, le pixel correspondant à la couleur recherchée est entouré par un cercle blanc. On suppose pour ce pixel la distance de 0 et un facteur α à 1. L'ensemble des couleurs compatibles respectant les contraintes de longueur et de couleur minimale sont incluses dans la zone blanche entourant le cercle blanc précédemment décrit. Les couleurs noires situées dans la partie inférieure gauche sont des pixels saturés c'est-à-dire dépassant l'intervalle de valeurs acceptées. On peut voir également sur la figure 6.19 des nappes globales et restreintes par les contraintes ici présentées. Les éventuels débordements d'intervalle sont évidemment saturés c'est-à-dire remplacés par la couleur existante la plus proche.

Enfin, pour visualiser de manière plus "interactive" l'effet des paramètres de luminance et de distance de la méthode présentée nous avons développé une interface graphique. Cette interface permet de visualiser trois équations différentes de compatibilité de couleur :

- la première équation modélise les couleurs compatibles du point de vue de la luminance (Fig. 6.20),
- la deuxième équation modélise les couleurs compatibles du point de vue de la distance (Fig. 6.21),
- la troisième équation modélise les couleurs compatibles du point de vue de la distance et luminance combinées c'est-à-dire l'équation nappe de compatibilité (voir Fig. 6.22).



(a)



(b)

FIG. 6.17 – Autres résultats de segmentation par seuillage de la fonction précédente avec application des contraintes.

Améliorations et limites de la méthode : Cette méthode de reconnaissance est malheureusement limitée aux couleurs claires, en effet les couleurs trop sombres c'est-à-dire proches du noir sont compatibles avec un ensemble important d'autres couleurs du cube RGB. Une critique pourrait aussi être la nécessité de l'étape de calibration. Cependant cette étape est incontournable si l'on veut utiliser la couleur dans le milieu sous-marin, car contrairement au domaine terrestre, il n'y a pas de constance de couleur suivant la distance. Les améliorations pour l'utilisation de cette méthode pourrait être

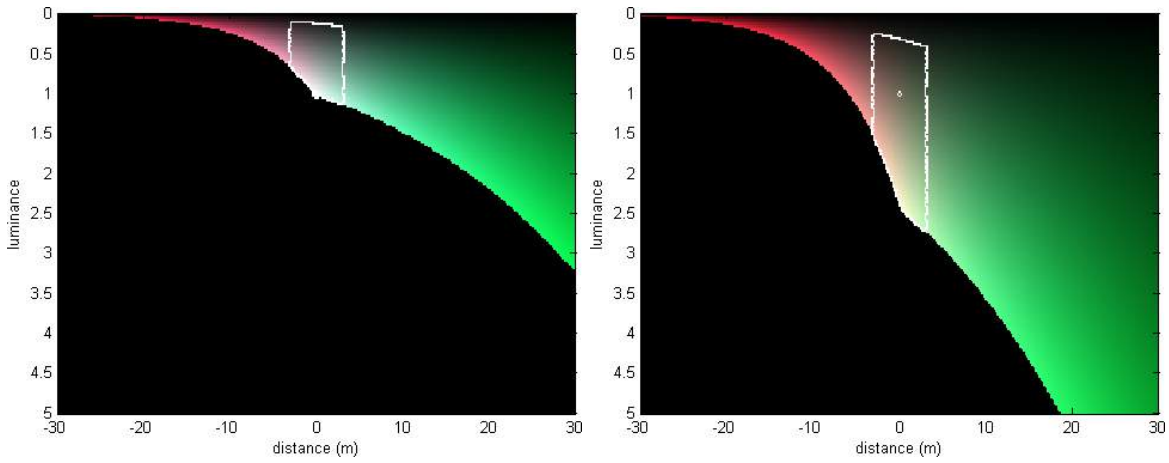
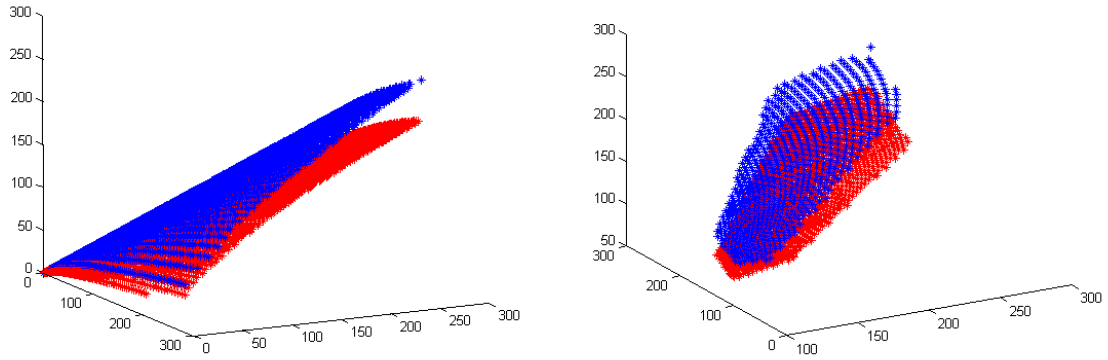


FIG. 6.18 – Nappe des couleurs compatibles avec les couleurs (255, 255, 255) et (110, 106, 88) avec une contrainte de distance fixée à 3 mètres.



(a) (b)

FIG. 6.19 – Nappes de compatibilité globale et restreinte pour les couleurs (255, 255, 255) en bleu et (110, 106, 86) en rouge.

par exemple l'utilisation de la cohérence de la séquence pour diminuer les fausses alarmes et même augmenter les taux de reconnaissance.

6.4.3 Comparaison et validation des méthodes

Pour évaluer les résultats présentés nous avons voulu comparer notre algorithme au seuillage sur la composante de teinte dans l'espace HSV (en français, teinte, saturation, luminance) décrit dans le chapitre 2. Cet espace est en effet plus robuste face au problème de constance de couleur que l'espace RGB et donc souvent utilisé dans la segmentation couleur. On trouvera dans [169] une étude globale intéressante sur la segmentation couleur dans l'espace HSV, et des applications pour la reconnaissance d'objets dans les articles suivants [95][147][170][171].
Nous présentons sur les figures 6.23 la composante de teinte des images de calibration (voir Fig. 6.12)

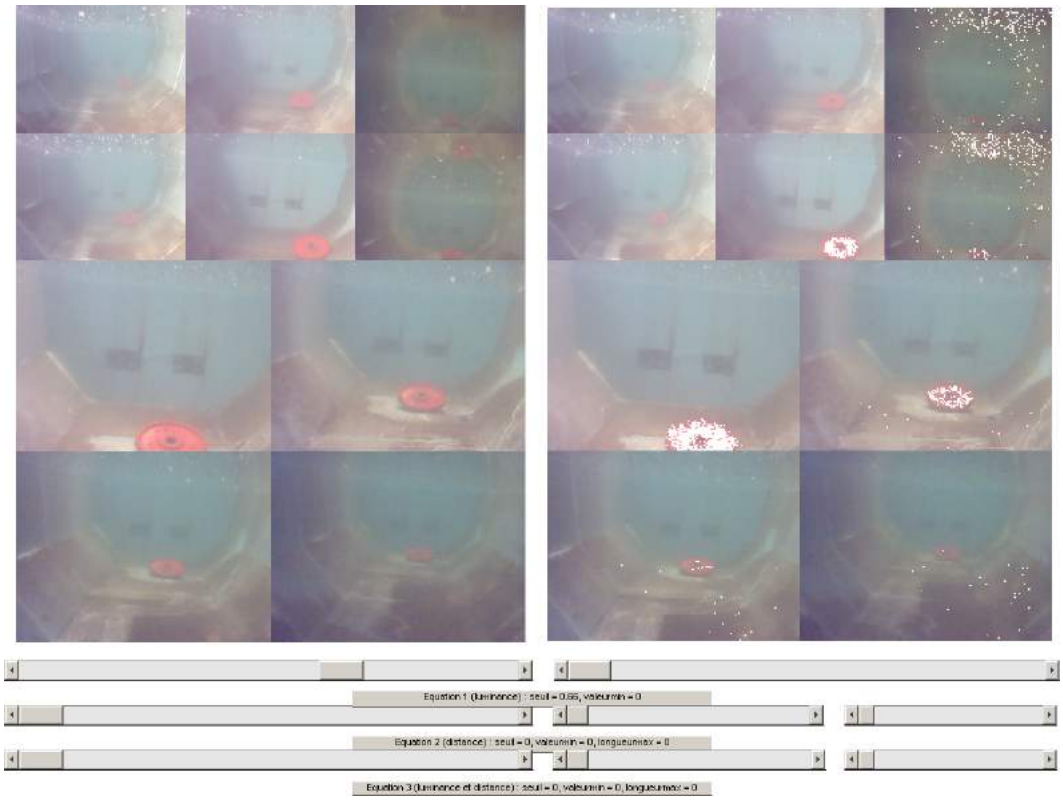


FIG. 6.20 – Compatibilité de couleur d'un point de vue de la luminance seulement.

et l'image de distance dans "l'espace de teinte". Cette dernière image étant l'équivalent de l'image de distance dans l'espace des couleurs compatibles présentée sur la figure 6.16(a). Le résultat de seuillage dans cet espace de teinte est présenté sur la figure 6.24.
Pour comparer objectivement les résultats nous nous sommes attachés à effectuer le seuillage sur la composante de teinte de manière à obtenir sensiblement les mêmes niveaux de détection sur l'objet. Nous pouvons constater que le seuillage sur la teinte fait apparaître beaucoup de fausses alarmes. Les résultats de segmentation utilisant la nappe de compatibilité non contrainte sont déjà meilleurs que ceux obtenus par ce seuillage sur la teinte. En appliquant les contraintes, les fausses alarmes sont très faibles et la qualité des détections est clairement meilleure. Nous avons présenté dans ce chapitre les résultats sur une seule base d'images test, d'autre résultats sur des séquences en contexte opérationnels seront présentés dans le chapitre suivant. La validation des résultats sur une base d'images plus conséquente, et avec une comparaison à davantage de méthodes est en cours aujourd'hui.

6.4.4 Discussions et perspectives : le volume de compatibilité

Il peut être parfois difficile de connaître avec précision les coefficients d'absorption K_λ pour les trois longueurs d'onde rouge, vert et bleu, ne disposant pas des données dans les conditions adéquates par exemple. Par contre il semble assez aisé d'obtenir des intervalles plus ou moins précis pour ces coefficients. Dans ce cas, comme nous l'avons montré par le théorème 2 la méthode reste utilisable. La nappe de compatibilité devient alors un volume de compatibilité. Cette méthode utilisant ces

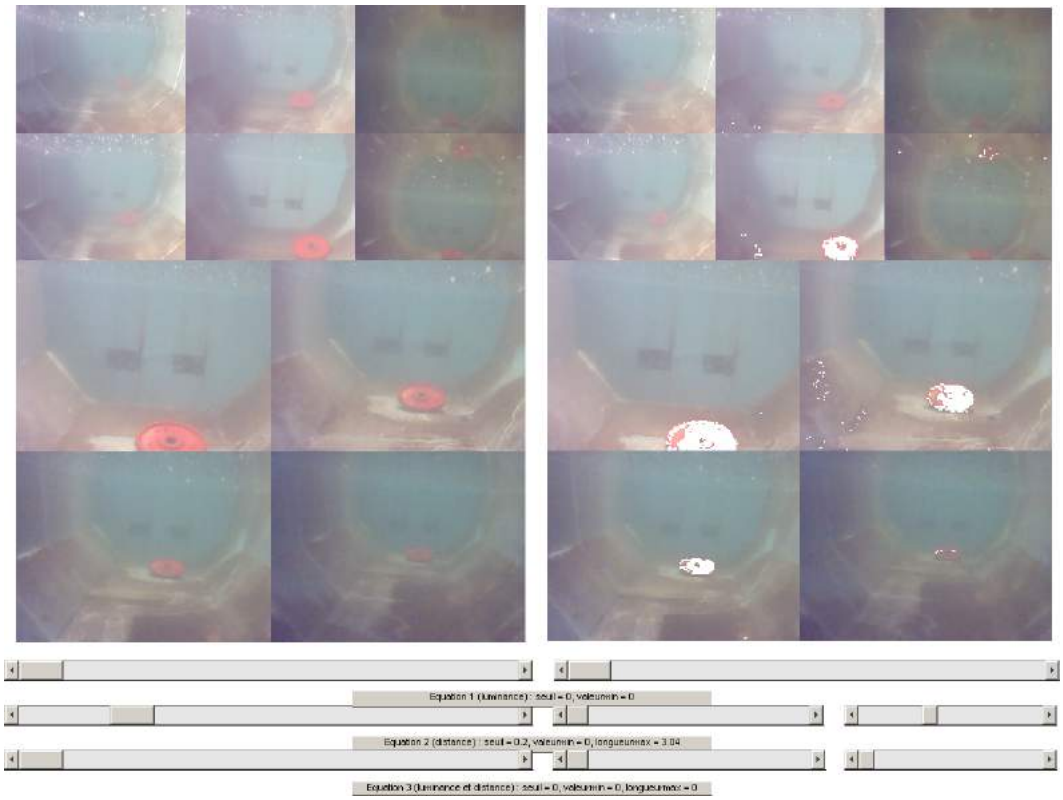


FIG. 6.21 – Compatibilité de couleur d'un point de vue de la distance seulement.

approximations par intervalles est à l'étude aujourd'hui. En effet, elle permettra une plus grande liberté d'utilisation puisque l'on s'affranchit encore d'une information *a priori*. Évidemment la perte de cette connaissance *a priori* fait diminuer la précision de l'algorithme et impose donc un travail sur les contraintes plus approfondi. La liberté d'utilisation que nous pourrions obtenir en validant ce modèle serait un point important pour la réutilisation potentielle de ce concept de compatibilité dans d'autres domaines.

6.5 Conclusion

Nous avons donc présenté dans ce chapitre une nouvelle méthode de reconnaissance basée sur la couleur en milieu sous-marin. La méthode classique présentée dans le chapitre 5 est difficile à appliquer dans le domaine sous-marin. En effet il est souvent difficile de conclure sur l'étape de reconnaissance de manière robuste vu la qualité médiocre des images et du fait que la méthode générique est sans *a priori*. C'est pourquoi, nous avons décidé d'apporter de la connaissance *a priori* sur l'objet à rechercher, de manière à améliorer les performances de reconnaissance. Au prix de la perte de la généricité, nous augmentons considérablement la robustesse de la décision parce que nous savons ce que nous cherchons. Ce type de méthode que nous avons appelé méthode spécialisée utilise donc une autre approche dite "Top-down" (voir chapitre 3), qui consiste à vérifier la compatibilité entre la scène et un attribut de l'objet connu *a priori*. L' attribut que nous avons présenté pour illustrer ce type de méthode est la

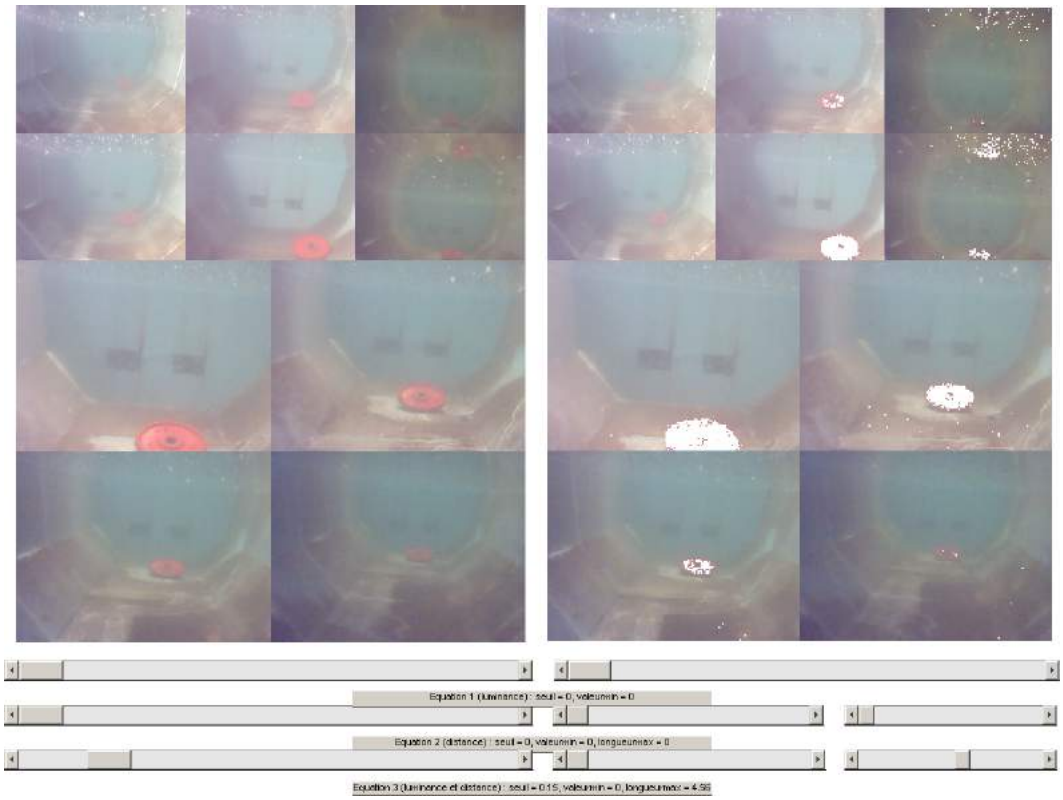
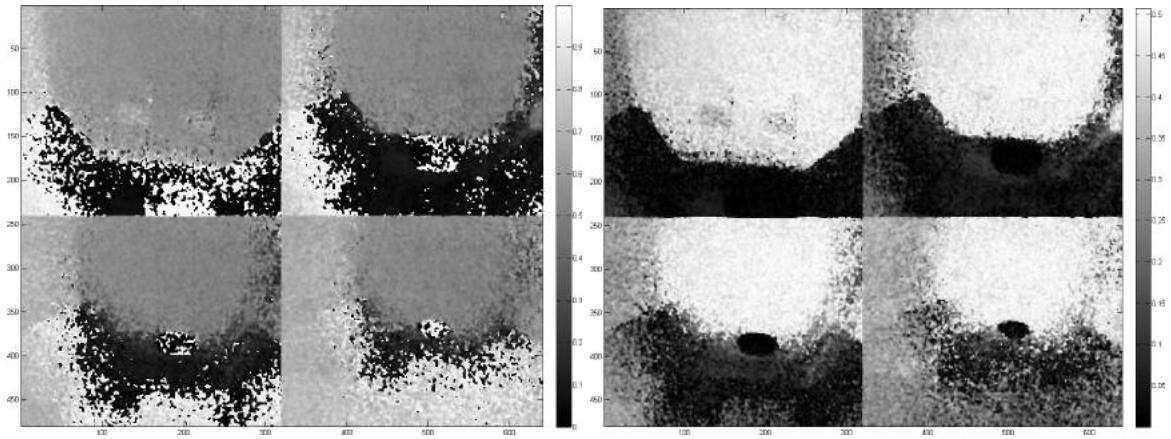


FIG. 6.22 – Compatibilité de couleur d'un point de vue de la luminance et de la distance.



(a) (b)

FIG. 6.23 – (a) Représentation de la composante de teinte dans l'espace HSV, (b) Image de distance avec la teinte de l'objet dans l'espace de teinte.

couleur. Mais une des perspectives de ce travail sera d'élargir l'utilisation de ce type de méthode à d'autres attributs comme la forme en essayant de retrouver la compatibilité entre la représentation 2D de l'objet et son modèle 3D par exemple. Des méthodes de ce type existent dans la littérature (lié au

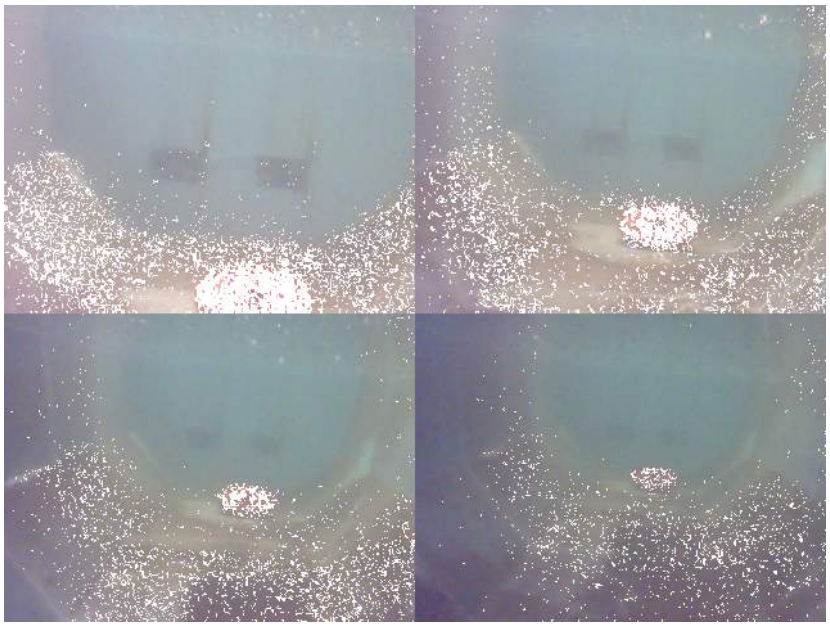


FIG. 6.24 – Résultat de segmentation par seuillage sur un critère de distance dans l'espace de teinte.

domaine terrestre) mais sous des hypothèses fortes notamment l'utilisation des primitives "segment" qui induit une bonne détectabilité donc des conditions favorables en terme de bruit [79]. Bien évidemment une représentation des objets sous forme de segments doit être raisonnable. Cette méthode est à l'étude aujourd'hui et nous privilégions une utilisation des points d'intérêts et des méthodes par intervalles [172][38][173]. Les travaux se poursuivent également sur la détection des couleurs non prises en compte à l'heure actuelle et sur l'utilisation du volume de compatibilité avec des intervalles de valeurs pour les coefficients d'absorption.

Concernant la méthode présentée, basée sur la couleur, nous avons montré qu'elle est utilisable comme critère unique pour la reconnaissance d'objet sans prétraitement des images c'est-à-dire directement sur les données brutes. Dans le contexte sous-marin puisqu'il n'y a pas invariance des couleurs suivant la distance, nous avons développé un modèle permettant de prévoir la déformation des couleurs. Ce modèle original développé pour le milieu sous-marin prend en compte la distance, les changements d'illumination et de position de l'objet. Les résultats présentés sont prometteurs et une validation plus large et plus concrète est en cours aujourd'hui. Cependant, d'autres résultats dans le cadre du concours SAUC'E seront présentés dans le chapitre 7. Ces résultats en contexte opérationnel montrent que, malgré son utilisation presque nulle dans la littérature, la couleur est un attribut exploitable et très avantageux en milieu sous-marin, notamment en terme de rapidité et de robustesse. C'est même un attribut autosuffisant pour la reconnaissance d'objets dans des conditions favorables. Il pourra d'ailleurs être utilisé dans bien d'autres situations notamment pour épurer les images en supprimant un bon nombre de couleurs non compatibles avec ce que l'on cherche (le fond : algues par exemple) et donc simplifier les traitements ultérieurs. Il peut également servir pour l'estimation de la distance avec des objets ou des obstacles dans le cadre du déplacement d'un robot dans une piscine uniquement à l'aide du capteur vidéo par exemple. Deux applications dans le cadre du concours SAUC'E seront présentées dans le chapitre 7.

Chapitre 7

Applications des méthodes développées

Dans ce chapitre des résultats applicatifs des deux méthodes de reconnaissance développées dans les chapitres 5 et 6 vont être présentés. Ces résultats sur des données réelles ont été obtenus suite à la participation à deux projets. L'un sur la vision sous-marine, l'autre sur la robotique sous marine. Ces projets ont permis de tester, illustrer et par conséquent valider expérimentalement les algorithmes.

7.1 Deux contextes d'utilisation très différents

D'abord, nous allons présenter une utilisation de la méthode classique (chapitre 5) dans un contexte de reconnaissance des objets sous-marins nuisibles, sur des données réelles en mer obtenues dans le cadre de la participation au projet TOPVISION. Le deuxième algorithme basé sur une méthode spécialisée sur la couleur sera lui aussi illustré sur des données réelles mais prises cette fois dans un contexte beaucoup plus favorable c'est-à-dire en piscine lors du concours de robotique sous-marine SAUC'E.

7.1.1 Description générale du projet TOPVISION

Le projet TOPVISION (Tests OPérationnels de VIdéos Sous-marines pour l'Identification d'Objets Nuisibles) a été présenté dans le cadre de l'appel à propositions Techno-Vision du Ministère de la Recherche et du Ministère de la Défense. Son objectif était de mener une campagne d'évaluation des algorithmes de traitement de séquences vidéos pour la détection, la localisation, la caractérisation et l'identification automatiques d'objets sous-marins nuisibles (manufacturés, polluants, dangereux, voire létaux) par caméra mobile c'est-à-dire portée par un AUV. Il s'est déroulé sur un peu plus de trois ans.

Contexte opérationnel : Aujourd'hui, la plupart des véhicules sous-marins sont équipés de deux types de capteurs pour la reconnaissance d'objets les uns acoustiques, les autres optiques. Lorsque l'on recherche des objets, on utilise d'abord un sonar pour repérer la présence d'un objet à longue distance puis après un phase d'approche pour tenter de discriminer le type de l'objet présent. Dans le cas, ou le capteur acoustique n'a pas permis d'identifier avec certitude l'objet détecté, on utilise le capteur

optique. Son rôle est de reconnaître l'objet en obtenant son type et ses caractéristiques détaillées puis éventuellement s'il doit y avoir intervention d'obtenir le meilleur positionnement relativement à l'objet. Sa portée est très courte, pas plus de 20 mètres dans des conditions très favorables et seulement quelques mètres dans une eau très turbide.

Spécificités du projet :
- Scénario simple : 1 ou 2 objets,
- Taille des objets de 10 cm à 1 m,
- Variation d'éclairage (jour/nuit, naturel/artificiel),
- Turbidité de l'eau variable,
- Dynamique du véhicule (non stabilité dans le courant),
- Deux types de caméra (couleur ou basse luminance monochrome),
- Texture du fond variable (l'objet peut aussi être entre deux eaux).

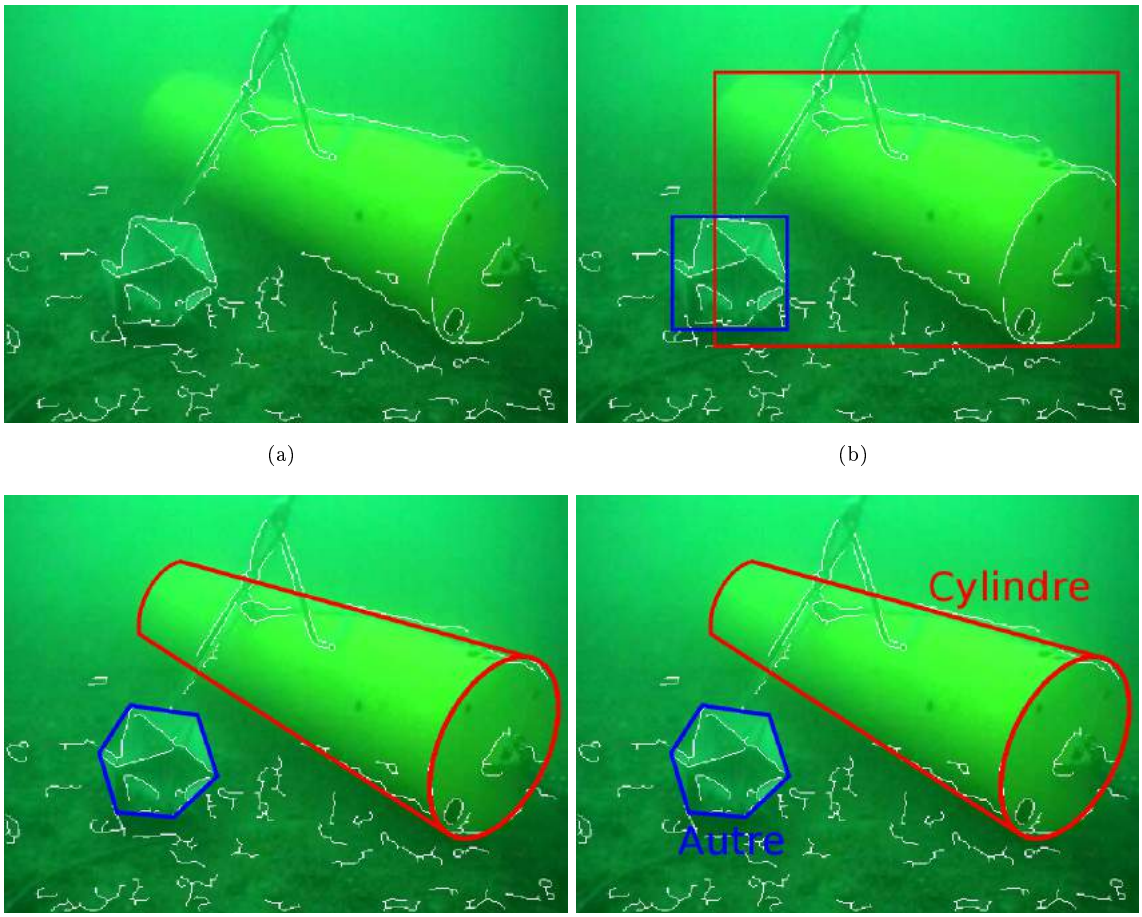
Problématique et résultats attendus : La vidéo est classiquement téléopérée par un opérateur humain, l'automatisation de cette tâche est un des enjeux de ce projet. En raison du bas débit du lien acoustique, on ne peut faire un traitement déporté sur un bâtiment principal. Pour respecter les contraintes industrielles réalistes, il est donc aussi demandé de développer des algorithmes rapides et peu gourmands en ressources matérielles. Les propositions de fonctions à évaluer sont les suivantes : (Illustration Fig. 7.1)

1. Détection de l'objet,
   $\Longrightarrow$ Présence ou non d'un ou plusieurs objets,
2. Localisation,
   $\Longrightarrow$ Mesure du lieu de l'objet (1 pixel ou une région dans l'image),
3. Caractérisation,
   $\Longrightarrow$ Mesure d'attributs de l'objet (contour apparent ou polygone englobant),
4. Identification,
   $\Longrightarrow$ Désignation dans une base de données.

A partir de cette problématique, les résultats attendus pour ce projet sont : la création et le partage d'une base de données réelles de vidéos sous-marines ainsi que le développement d'une plate-forme d'évaluation d 'algorithmes. Et dans un contexte plus technique, la réalisation d'une campagne d'évaluation d'algorithmes entre différents spécialistes scientifiques avec communication et publication des résultats. Cet état de l'art doit apporter les éléments nécessaires pour dynamiser le domaine de recherche dans les traitements vidéos automatiques sur les robots autonomes sous-marins.
Un site internet d'hébergement (`http://www.topvision.gesma.fr`) de la structure d'évaluation incluant la présentation du projet, la mise à disposition de données de test, des outils logiciels de visualisation et d'évaluation ainsi que les résultats finaux a été réalisé en 2008 et sera maintenu pour une durée de cinq ans.

Caractéristiques de la base de données

Nature des données : Les données (données de test et données d'évaluation) peuvent être de deux types : des vidéos sous-marines enregistrées sous un format numérique ou des vidéos existantes



(a) (b)

(c) (d)

FIG. 7.1 – Les étapes de reconnaissance proposées pour TOPVISION, (a) Détection, (b) Localisation (c) Caractérisation (d) Identification

converties. Ces vidéos sont complétées par des fichiers de description générale comportant un descriptif de la vidéo (durée, hauteur, largeur...) et des conditions d'acquisition utiles pour leur traitement (par exemple couleur ou noir et blanc, éclairage artificiel ou naturel, type de codec...). On dispose également pour ces vidéos de fichiers de vérités terrain annotés manuellement par un opérateur humain. Ces vérités terrain constituent la base de résultats attendus, image par image pour les quatre niveaux de reconnaissance : détection, localisation, caractérisation et identification c'est-à-dire le nombre d'objets présents et pour chaque objet : sa nature, sa localisation en coordonnées image et ses caractéristiques mesurées. Avant chaque traitement sur l'un des quatre niveaux on dispose également d'un fichier de vérité terrain partielle contenant les résultats de l'étape précédente. Lorsque l'on utilise la détection, une information *a priori* importante est donnée dans la vérité terrain partielle, c'est la zone utile de l'image c'est-à-dire sans effet de bord et sans surimpression de texte.

Méthode de collecte des données : Les vidéos ont été obtenues de plusieurs manières : certaines en bassin et d'autres réalisées en mer notamment dans la rade de Brest ou dans ses alentours.

Deux scénarios ont été appliqués : des séquences d'approche, et des séquences d'évitement autour de l'objet prenant l'objet sous plusieurs points de vue. Les objets présents ont généralement des formes simples : des sphères, des cylindres, dans certains cas des formes plus furtives de type troncs de cône ou parallélépipèdes. Certaines vidéos ne présentent pas d'objet. Les données de test ainsi que le données d'évaluation forment une base d'environ trente vidéos annotées.

Définition et mise en oeuvre des tests

Métrique de l'évaluation : Pour évaluer les algorithmes candidats, des critères de performances ont été développés pour chaque niveau de la reconnaissance. Ces critères sont calculés pour chaque image à partir des vérités terrain et des fichiers de sortie des algorithmes.
Description des métriques :

- Etape 1 : Détection de présence,
- Etape 2 : Distance de localisation (0 si le pixel appartient à l'objet, la distance entre le pixel et le contour de l'objet le plus proche sinon),
- Etape 3 : Taux de caractérisation (pourcentage entre intersection et union du polygone englobant de la vérité terrain et du résultat algorithmique),
- Etape 4 : Identification du type de l'objet.

Infrastructure d'évaluation : Le processus d'évaluation est simple. Il consiste d'abord à générer les données d'entrée (extraction des images et création des vérités terrain partielles). Puis l'algorithme candidat est appliqué pour chaque image de la séquence. Cet algorithme peut lire la description générale et la vérité terrain partielle pour utiliser les informations *a priori* nécessaires et doit ensuite écrire un fichier résultat (du même type que la vérité terrain). Après le traitement de toutes les images d'une séquence, un compte-rendu d'évaluation est généré. Il comporte le temps de calcul global et des statistiques sur l'ensemble des métriques, par exemple les taux de bonnes détections et de fausses alarmes pour l'étape de détection. Des résultats de ce type sont présentés par la suite. Le temps de calcul est un critère important d'un point de vue opérationnel mais il n'a pas été pris en compte directement dans l'évaluation des algorithmes.

7.1.2 Description générale du concours SAUC'E

FIG. 7.2 – Logo du concours SAUC'E

Le concours SAUC'E (Student Autonomous Underwater Challenge - Europe) est un concours de robotique sous-marine sponsorisé par le Ministère de la Défense Anglaise (Research Acquisition Organisation et DSTL). Ce concours a lieu chaque année. Sa troisième édition a eu lieu à Brest en 2008. Il est réservé à des équipes d'étudiants des différentes universités et écoles européennes (un tiers de scientifiques encadrants). Son but est de promouvoir les technologies autonomes en robotique sous-marine. De nombreuses informations concernant ce concours peuvent être trouvées sur `http://www.dstl.gov.uk/news_events/competitions/sauce/index.php`.

Les objectifs du concours : Le concours SAUC'E a été élaboré pour permettre des avancées scientifiques dans le domaine de la robotique sous-marine en faisant concourir des équipes multidisciplinaire pour réaliser une mission autonome en environnement sous-marin. Il permet d'encourager les étudiants à s'intéresser aux techniques de la robotique, et aussi aux enseignants-chercheurs et ingénieurs encadrants de tester sur un cas réel des algorithmes et des solutions technologiques. Il s'agit de construire un robot complètement autonome capable de se déplacer seul, de s'immerger et de réagir à son environnement. L'épreuve finale est une mission réaliste incluant les points suivants :

1. Se déplacer du point de départ et s'immerger,
2. Franchir un cadre immergé sans le toucher (taille du cadre 3m x 2,5m)
3. Larguer un marqueur au dessus d'une cible circulaire posé sur le fond. Cette cible de 1 mètre de diamètre étant marquée d'une croix noire et signalée par clignotement lumineux blanc en son centre.
4. Localiser une cible rouge "entre deux eaux" et la toucher. Un leurre un peu plus petit et de couleur différente est aussi présent dans le bassin.
5. Faire surface dans un hexagone blanc flottant (3m x 3m) situé entre un plot et un pneu.
6. Produire une cartographie de la zone de compétition (30m x 20m). C'est-à-dire obtenir les coordonnées de chaque objet : le cadre, la cible, la boule rouge, le leurre, le pneu, le plot, et la zone de sortie flottante.
7. Fournir un "compte rendu" sur l'ensemble de la mission détaillant son fonctionnement.

Les dimensions de la zone de compétition sont connues et délimitées par des lignes noires sur le fond de la piscine. Les caractéristiques précises de tous les objets présents sont elles aussi connues excepté bien sûr leur place dans la zone de compétition qui change systématiquement à chaque début de manche. La seule contrainte est que la mission doit être réalisée en une seule fois sans contrôle ou communication avec un opérateur. Il existe aussi des contraintes matérielles sur le robot comme par exemple la taille et le poids (maximum 2m x 1m x 1m et 70kg), ou la non utilisation d'un système Doppler... On doit évidemment aussi prendre en compte les contraintes des robots autonomes en terme de stockage physique, de puissance énergétique, et de complexité calculatoire.

La compétition : Elle est réalisée sur plusieurs jours incluant des séances d'essais et de qualifications (l'épreuve de qualification consistant seulement à franchir le cadre immergé) et la compétition à proprement parler a lieu le dernier jour. Elle consiste à effectuer la mission globale présentée précédemment. Plusieurs essais sont autorisés durant la finale mais durant un temps limité (avec interdiction d'apprendre le terrain). L'évaluation est faite sur une présentation orale et écrite du système développé et sur son évaluation pratique en autonomie, différents points étant attribués en fonction de la réussite des différentes actions demandées lors de la mission.

Participation de l'ENSIETA : L'ENSIETA a participé à ce concours en 2007 et en 2008. Les résultats présentés dans ce chapitre constitue l'intégralité de la partie vision embarquée pour le concours 2007.

7.2 TOPVISION : reconnaissance d'objets par les contours

7.2.1 La solution proposée à partir de l'algorithme développé

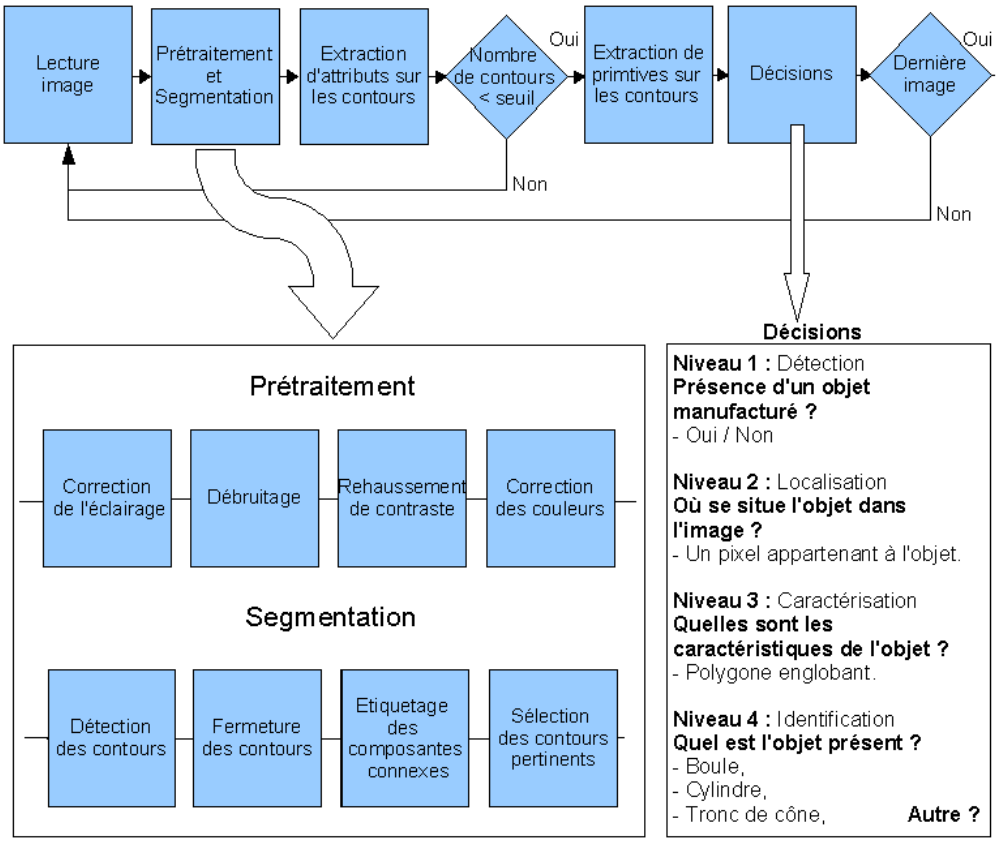


FIG. 7.3 – Organigramme de l'algorithme utilisé pour traitement des séquences TOPVISION.

Pour répondre aux contraintes liées à l'évaluation des algorithmes dans le cadre du projet TOPVISION, nous avons proposé une méthode utilisant l'algorithme développé dans le chapitre 5. L'étape de la reconnaissance que nous avons évaluée est l'étape de détection. Dans certains cas, cette étape validée, nous pouvons aussi valider la localisation et la caractérisation notamment lorsque des primitives de type cercles et ellipses sont détectées. Cependant cette partie n'a pas été évaluée puisque les résultats ne sont pas obtenus dans tous les cas.

7.2.2 L'étape évaluée : La détection

Par cette première étape du processus de reconnaissance, on doit décider de la présence ou non d'un ou plusieurs objets dans l'image. De nombreuses techniques existent pour faire de la détection d'objets dans les séquences. On peut par exemple en supposant qu'il n'y pas de "leurre" utiliser une simple détection de changement [59][91][60]. Cependant, ces méthodes rapides ne nous satisfaisaient pas puisqu'elles sont trop sensibles aux fausses alarmes. Nous avons donc décidé pour l'évaluation de la détection d'objets manufacturés d'utiliser l'algorithme complet de reconnaissance par les contours

présenté précédemment, c'est-à-dire en allant jusqu'à la recherche de primitives. En effet en recherchant des primitives sur les scènes d'intérêts détectées, on améliore nettement la robustesse de l'algorithme (un poisson ou un rocher ne pouvant être confondus avec un objet manufacturé). Un organigramme de cette méthode est présentée sur la figure. 7.3.
La méthode utilisée peut se résumer en deux phases :

1. La première consiste à prétraiter et à segmenter chaque image puis à calculer des caractéristiques sur les contours extraits. A partir de ces caractéristiques, nous sommes en mesure de décider de la présence ou non d'une scène d'intérêt. La discrimination entre une scène d'intérêt et une scène à négliger se fait par un seuillage empirique vis-à-vis d'une considération simple sur le nombre de contours. D'après un constat simple en effet, une image naturelle est caractérisée par un grand nombre de petits contours, tandis que les images d'objets manufacturés contiennent des zones d'intensités différentes caractérisées par des contours peu nombreux, marqués et longs. Ce seuillage permet d'interrompre le traitement pour les images qui ne contiennent *a priori* rien d'intéressant (on peut d'ailleurs voir sur la figure 7.4 le résultat de la segmentation sur une scène "vide"). On va donc pouvoir négliger un bon nombre d'image dans la séquence ce qui a pour conséquence d'accélérer nettement le traitement. Cette première étape de l'algorithme de détection est très rapide mais très sensible à la présence d'un rocher, d'une trace dans le sable, d'une plante ou d'un poisson par exemple.
2. La deuxième phase consiste à décider de la présence ou non d'un objet sachant que l'on analyse une scène d'intérêt. Cet algorithme en deux phases permet d'améliorer les performances et la robustesse. Dans le cas où une scène d'intérêt est présente, on recherche des primitives telles que des droites, des cercles, ou des ellipses caractéristiques des objets manufacturés dans l'image de contour. Nous avons utilisé trois critères discriminants pour la recherche de ces primitives : d'abord l'orientation du gradient (en effet la présence d'une direction privilégiée dans l'orientation du gradient révèle en général la présence de droites), puis l'espace de Hough pour les droites et pour les cercles. Dans ces trois représentations, on va donc rechercher des pics et en extraire des primitives telles que des droites, cercles ou ellipses. La décision de présence se fait donc sur cette deuxième étape du traitement à partir des primitives extraites : la présence de deux droites ou d'un cercle valide la détection d'un objet manufacturé.

Quelques exemples de résultats obtenus en détection sont présentés sur les figures 7.5 et 7.6.

7.2.3 Les trois autres étapes

Localisation et Caractérisation : A partir des résultats de l'étape précédente, et dans le cas où la détection est positive, nous sommes déjà allés beaucoup plus loin qu'une simple détection au sens propre du terme puisque nous avons extrait des primitives caractéristiques de l'objet. On peut donc dans certains cas répondre directement aux deux étapes suivantes du processus de reconnaissance c'est-à-dire la localisation et la caractérisation. Dans le cas où un cercle est détecté, on peut directement renvoyer le centre du cercle comme résultat de la localisation et un polygone englobant constitué d'un certain nombre de points du cercle pour le résultat de caractérisation. Quand aucun cercle n'est détecté on peut renvoyer un point appartenant à la droite la plus grande (appartenant *a priori* à un objet) comme résultat de la localisation. Dans le cas contraire (aucun cercle et aucune droite) aucun résultat n'est possible. La validation de ces deux étapes a été testée mais non validée

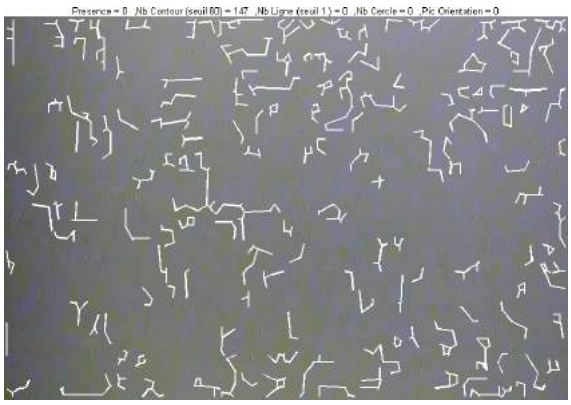


FIG. 7.4 – Résultats de segmentation sur une scène “vide”.

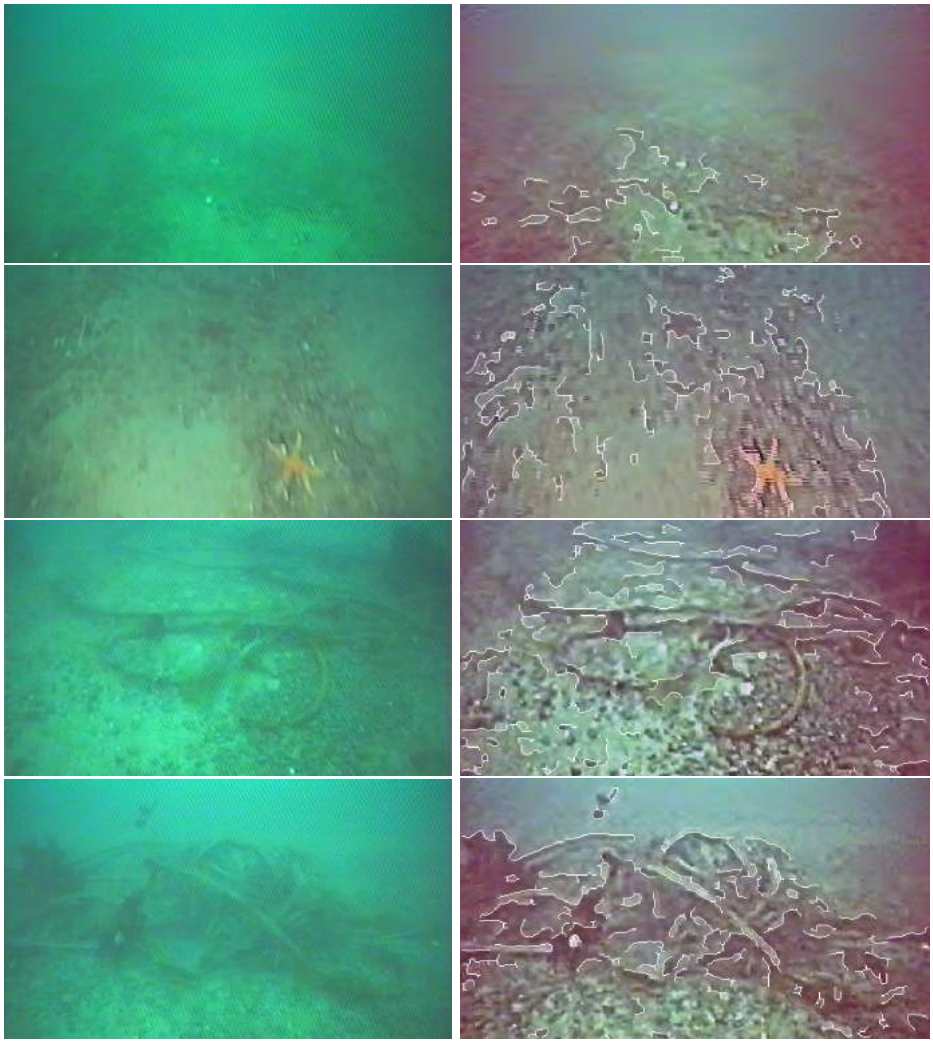


FIG. 7.5 – Exemple de non détection (scène vide, étoile de mer, bouts).

FIG. 7.6 – Exemples de détections d'objets (sphère, cylindre).

sur les données d'évaluation.

Identification : Le dernier niveau du processus de reconnaissance consiste à classer l'ensemble des paramètres extraits pour pouvoir prendre une décision sur le type d'objet présent. La méthode développée est basée simplement sur deux paramètres très simples (le nombre de droites et de cercles). La décision est donc difficile. Il est cependant possible de décider de la présence d'une sphère lorsque seul un cercle est trouvé ou un cylindre lorsque que l'on a un cercle et deux droites et éventuellement un tronc de cône lorsque l'on a deux cercles presque concentriques. Dans notre cas, les paramètres ex-

traits n'étant pas en nombre suffisant, nous n'avons pu valider l'identification avec assez de robustesse.

Améliorations et perspectives : Plusieurs améliorations pourraient être apportées à l'algorithme ici présenté. La première amélioration importante serait d'ajouter la recherche de la primitive de type ellipse. D'une part parce qu'elle est pour l'instant approximée par un cercle et que l'on a donc un taux de détection plus faible. D'autre part parce que cette primitive sera un critère supplémentaire qui nous permettra de différencier avec beaucoup plus de robustesse la sphère, le cylindre et le tronc de cône.
Une deuxième amélioration serait d'ajouter encore d'autres attributs comme la couleur, ou des attributs sur les contours. En effet la caractérisation est basée pour l'instant uniquement sur des attributs géométriques.
En qui concerne la couleur, en plus d'un paramètre supplémentaire pour l'étape d'identification, elle permettra d'améliorer la recherche du polygone englobant en effectuant une segmentation couleur (contour actif à partir du point résultat de la localisation par exemple). Cette technique permettrait également de fermer les contours autour des régions de couleur ou de texture uniforme et de supprimer les contours liés au fond ou au bruit ce qui aura pour but d'améliorer la qualité des attributs et de diminuer les fausses alarmes.
Et en ce qui concerne les attributs sur les contours comme la courbure utilisée par [1], ou bien les descripteurs de forme parmi les nombreux existants [122], ils permettraient eux aussi de parfaire les résultats. Malheureusement, dans le cas d'une telle augmentation de paramètres, une autre évolution serait obligatoire. C'est-à-dire qu'il faudra mettre en place un classifieur qui s'attachera à retrouver l'objet à partir des paramètres d'entrée. De nombreux classifieurs existent dans la littérature on pourra notamment se référer à [174], on pourra aussi se référer à [175][93][136] pour les méthodes de reconnaissance basées sur les contours.
Enfin, et pour conclure on pourrait nettement optimiser la localisation qui a été un peu mise de côté puisque l'on suppose qu'il n'y a qu'un objet dans l'image. Un travail sur le regroupement des contours permettrait de détecter le nombre d'objets présents dans la scène. On pourra notamment tester plusieurs classifieurs non supervisés comme par exemple la classification ascendante hiérarchique par lien simple ou bien des classifieurs supervisées classiques avec une méthode d'optimisation permettant de décider du nombre de classe le plus approprié (on peut supposer que l'on aura au maximum 3 classes dans une image). On peut citer quelques ouvrages référençant ces méthodes [174][176]. Grâce à ce regroupement de pixels ou "clustering", on pourra ensuite calculer un rectangle englobant autour du ou des objets présents avec "les noyaux et les rayons" de chaque classe. Ce rectangle permettrait de simplifier l'étape de caractérisation en restreignant la zone de recherche des primitives. Pour appliquer une telle méthode, il faudra par contre prendre en compte les problèmes suivants : le premier est qu'un objet fait en général apparaître des contours sur sa silhouette mais rien à l'intérieur de cette silhouette, le second provient de l'étape de la segmentation qui a supprimé les zones de forte densité en points de contours pour éviter les erreurs dues aux objets texturés.

7.2.4 Illustration des résultats

Pour obtenir de bonnes performances de détection, l'idée est d'aller plus loin dans le processus de reconnaissance c'est-à-dire jusqu'à la caractérisation puis de revenir en arrière pour conclure avec certitude sur la détection. En effet comment prendre une décision sur la présence d'un objet sans l'avoir

	UTT		I3S		E3I2	
Séquence	DT	FA	DT	FA	DT	FA
TopvisionVide	0.0000	0.2099	0.0000	0.2000	**0.0000**	**0.0000**
Topvision009	0.6586	0.0000	0.5650	0.0000	**0.6963**	**0.0000**
ACrov0951	**0.6911**	**0.0816**	0.5769	0.0000	0.1003	0.0000
ACrov0954	**0.5410**	**0.0167**	0.5094	0.0000	0.0485	0.0000
ACrov1044	0.0000	0.0000	0.3492	0.0149	**0.6835**	**0.0000**
ACrov1318	**0.9125**	**0.0788**	0.0314	0.0677	0.7363	0.0039
ACrov1429	**1.0000**	**0.4125**	0.0680	0.1095	x	x
ACrov1511	0.7653	0.0000	0.9238	0.0517	**1.0000**	**0.0911**
ACrov0947	x	x	**0.1728**	**0.0000**	0.0968	0.0000
ACrov0940	0.6187	0.0000	0.9104	0.0000	**0.9883**	**0.0115**
ACrov0955	0.6387	0.0000	**0.9240**	**0.0000**	0.2668	0.0000

TAB. 7.1 – Résultats des 3 laboratoires ayant participé aux évaluations (DT pour Détections, et FA pour Fausses Alarmes). Les deux résultats non chiffrés sur dus à un bug lors du traitement de la vidéo par les évaluateurs.

localisé ou extrait quelques unes de ses caractéristiques ? De même comment localiser les objets sans avoir recherché leurs caractéristiques ? Il faudrait dans ce cas supposer que toute modification dans la scène (détection de changement, ou diminution brutale du nombre de contours) est une détection. Ce qui revient à la présence de scène d'intérêt que l'on a évoqué précédemment. Autrement dit, y-a-t'il une entité présente dans la scène : objet manufacturé, rocher, poisson...
Les résultats obtenus grâce à cette stratégie sont affichés Fig. 7.7. La participation effective à l'évaluation a été de seulement trois participants (UTT, I3S et ENSIETA). Ces résultats de l'évaluation finale sont répertoriés dans le tableau 7.1 et seront publiés dans un article durant l'année 2008.

7.2.5 Conclusion et critique de méthode

Résumés rapides des algorithmes concurrents :

- UTT : Méthode statistique de détection de scène d'intérêt. Elle est basée sur une estimation du "fond" (turbidité, éclairage) dans les premières secondes de la séquence et ensuite une détection est validée à tout changement notable dans les pixels de l'image.
- I3S : C'est un processus automatique basé sur l'attention visuelle. Le processus d'attention visuelle permet de trouver la zone la plus saillante de l'image. Ce processus est combiné à un classifieur statistique qui permet de donner une probabilité de présence d'un objet manufacturé dans cette zone.

Pour résumer ce qui est détaillé précédemment, notre algorithme de détection est basé sur l'analyse des contours. La détection est validée si les contours sont peu nombreux et présentant des formes particulières (droites ou cercles).

Une première remarque à faire est que les trois algorithmes sont très différents dans leur approches et donc sur leurs critères de décision. Cependant, ils sont tous sans *a priori*, automatique et générique.

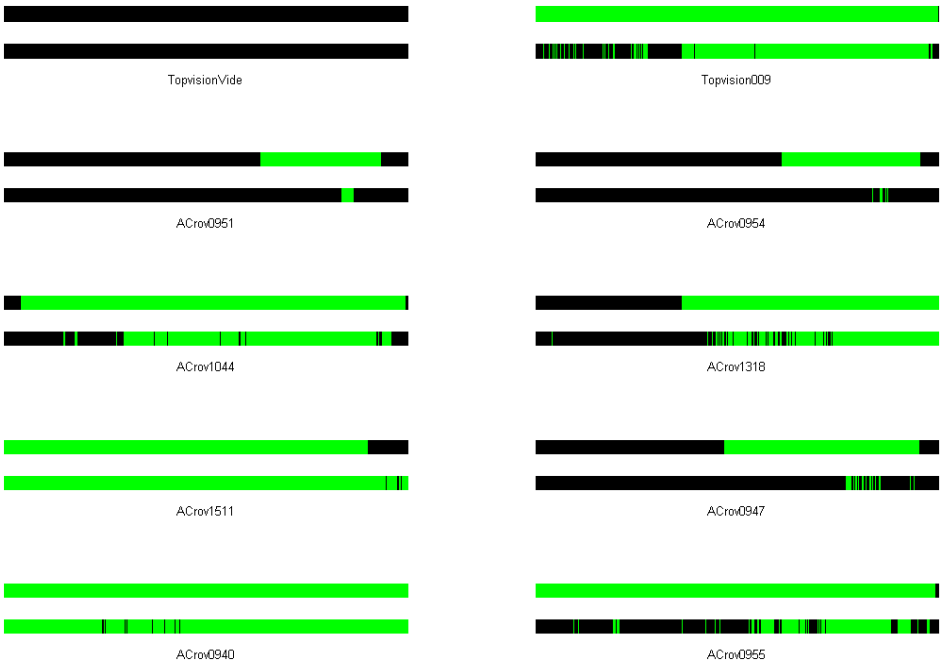


FIG. 7.7 – Vérités terrain (en haut) et résultats de l'algorithme présenté (en bas) en détection, sur les 10 vidéos d'évaluation. En noir la non détection en vert la détection.

Notre méthode donne des résultats assez intéressants sur les vidéos d'évaluation. Des résultats obtenus en détection sur des images de "vrais négatifs" sont présentés sur la figure 7.5. On peut voir deux images représentant le fond, et deux images représentant des objets "bouts". Les contours extraits de ces quatre images ne répondent pas aux critères "objets manufacturés" que nous considérons. D'autres résultats obtenus cette fois sur des images présentant un objet manufacturé sont présentés sur la figure 7.6. On peut se rendre compte que les contours sont beaucoup moins nombreux et qu'ils présentent des formes particulières (droites et/ou cercles). On peut voir sur la dernière image que l'objet n'est malheureusement pas détecté puisque ses contours sont moins marqués que ceux des algues sur le fond (ils n'ont donc pas été retenus). Ce qu'il faut noter quant à notre algorithme c'est qu'il ne tient pas en compte de la séquence. Il est aussi générique et sans *a priori*, automatique et robuste aux différentes qualités des images.

Passons maintenant à l'analyse des résultats d'évaluation présentées dans le Tab. 7.1. Ce qui ressort assez nettement à la vue des résultats, c'est que les trois algorithmes évalués fournissent des résultats assez proches. On peut voir en effet que les meilleurs scores (en gras) pour chaque vidéo sont assez partagés, avec cependant un léger avantage pour notre algorithme. Par rapport aux autres algorithmes on peut noter un taux de fausses alarmes généralement plus faible justifié par le fait l'on recherche directement des formes particulières sur les contours (il est donc normal que sur la séquence vide par exemple rien ne soit détecté puisque rien n'est présent en tout cas aucune forme ressemblant à ce que l'on recherche). Une critique viendrait par contre du temps de calcul puisque notre approche est plus lente que les autres, de l'ordre de 15 secondes par image en Matlab dans le cas où l'on va au bout de l'algorithme.
Si l'on veut maintenant analyser les défaillances de notre algorithme elles sont en grande partie dues

à deux problèmes : d'abord le manque de contraste (même après le prétraitement) et le processus de seuillage automatique va faire que l'image d'un objet manufacturé sera négligée puisque trop de contours sont présents. En effet, sur une image ne contenant rien (voir Fig. 7.4) on va obtenir un grand nombre de contours. Ensuite, ce sont les contours de l'objet qui peuvent être moins marqués que le fond donc négligés lors de la sélection des contours pertinents ou bien trop découpés. Dans les deux cas la recherche des formes particulières échoue et la détection d'objet est nulle.

7.3 SAUC'E : reconnaissance par la couleur

En 2007, le concours s'est déroulé du 12 au 15 juillet au QinetiQ Haslar Ocean Basin de Gosport (120m x 60m x 5,5m), ville portuaire de la Baie de Portsmouth (UK). Six équipes y ont participé et l'ENSIETA a terminé 3^{e} en remportant en plus le prix de l'innovation en contrôle. Contrairement au projet TOPVISION en mer, les conditions de visibilité et la qualité des images étaient très bonnes puisque nous étions dans un bassin. De plus, toutes les images étaient acquises dans les mêmes conditions (éclairage naturel, eau claire) par une webcam étanche (voir annexe B). Même la dynamique du véhicule n'était pas un problème puisque les mouvements restaient extrêmement lents. Dans ce contexte et sachant que les caractéristiques des cibles était parfaitement connues nous avons basé notre système de vision sur la méthode de reconnaissance par la couleur développée dans le chapitre 6. Cette méthode est en effet très rapide, peu coûteuse en ressources matérielles et facile à mettre en oeuvre.

7.3.1 Quelques résultats en contexte opérationnel

Les résultats présentés dans cette partie ont été obtenus en contexte opérationnel sur le robot sous-marin de l'ENSIETA préparé pour le concours SAUC'E 2007.

Dans l'objectif de ce concours, la partie vision devait permettre de reconnaître les cibles pour pouvoir effectuer les missions. L'algorithme utilisé a été présenté au chapitre 6.
Le programme développé et mis en place sur le robot dispose en entrée de deux fichiers texte : le premier fichier contenant les paramètres de calibration c'est-à-dire les coefficients d'absorption de l'eau dans le bassin ainsi que les paramètres de tolérance (contrainte sur la distance et sur "les couleurs foncées"), et le second fichier contenant les couleurs de chaque objet à rechercher (notamment le cadre rouge, une sphère orange et une sphère beige).
L'algorithme de vision est lancé au début de la mission et doit traiter en temps réel l'ensemble des images acquises par les deux caméras présentes sur le robot (caméra frontale et ventrale). Dès qu'une couleur est trouvée avec un nombre suffisant de pixels compatibles, une détection est signalée et une trame de texte est renvoyée. Cette trame de texte présentée ci-après est en même temps enregistrée dans un fichier utile ensuite pour la cartographie de la zone de jeu. Cette trame contient :

- les coordonnées en x et y du pixel médian parmi les pixels compatibles,
- les trois coordonnées RGB de la couleur détectée,
- le nom de l'image résultat enregistrée (Fig. 7.9),
- et le nombre de pixels compatibles détectés dans l'image.

Un exemple de trame résultat :

```
...
057 212 228 093 046 img_000171_228_93_46.ppm 0000000487
065 201 228 093 046 img_000174_228_93_46.ppm 0000000652
185 078 110 106 086 img_000178_110_106_86.ppm 0000000946
...
```

En fonction de l'objet détecté un comportement peut être ensuite associé en fonction de l'épreuve à exécuter.

7.3.2 Exemple 1 : localisation et contact avec l'objet coloré

Prenons un exemple pour illustrer notre utilisation d'algorithme de couleur : la quatrième action de la mission à effectuer dans le cadre du concours SAUC'E fut de localiser la boule rouge et de faire contact avec elle. La place n'étant pas connue à l'avance, il faut la chercher en explorant la zone de compétition dans sa plus grande partie. Dès que la boule a été détectée, le point médian suffit comme information pour l'asservissement visuel de la position du robot de manière à aller toucher la boule rouge. On peut notamment voir quelques images de détection d'une boule de couleur sur les figures 7.8 et 7.9. Un leurre c'est-à-dire une deuxième boule plus petite de couleur différente (beige) est présente dans la piscine mais on ne peut bien sûr pas la confondre par l'utilisation de ce critère couleur. La couleur nous permet de détecter l'ensemble des objets de la zone de compétition. Cependant elle ne nous permet pas de différencier les deux objets posés sur le fond (le pneu et la cible tous les deux de même couleurs). Il faut pour cela rajouter un critère comme par exemple l'utilisation de la lumière clignotante blanche placée au centre la cible ou la forme des contours (détection de la croix).

7.3.3 Exemple 2 : cartographie de la zone de compétition

En supposant que chaque objet soit bien différencié, on dispose à l'issue de la mission d'un relevé d'images présentant les objets vus à différentes positions. Pour chacune de ces images on connaît les coordonnées d'un point caractéristique de l'objet et on dispose des positions du sous-marin dans le repère du bassin ainsi que des paramètres de la caméra (focale, taille des pixels ...). A partir de plusieurs vues d'un même objet à des positions différentes, on va pouvoir retrouver sa position dans le repère du bassin (voir Fig. 7.11).

L'algorithme est le suivant :

- Pour chaque image de l'objet
  - Calcul de l'équation de la droite entre le centre optique et le point caractéristique de l'objet dans le repère caméra,
  - Transformation de l'équation de cette droite du repère caméra au repère bassin,
  - Appariement aléatoire des droites par algorithme de type RANSAC[1][177] avec estimation de leur point d'intersection par moindres carrés,
  - Calcul du point d'intersection médian sur l'échantillon validé.

[1] RANSAC est une méthode itérative utilisée pour estimer des paramètres d'un modèle mathématique à partir d'un échantillon de données contenant des valeurs erronées.

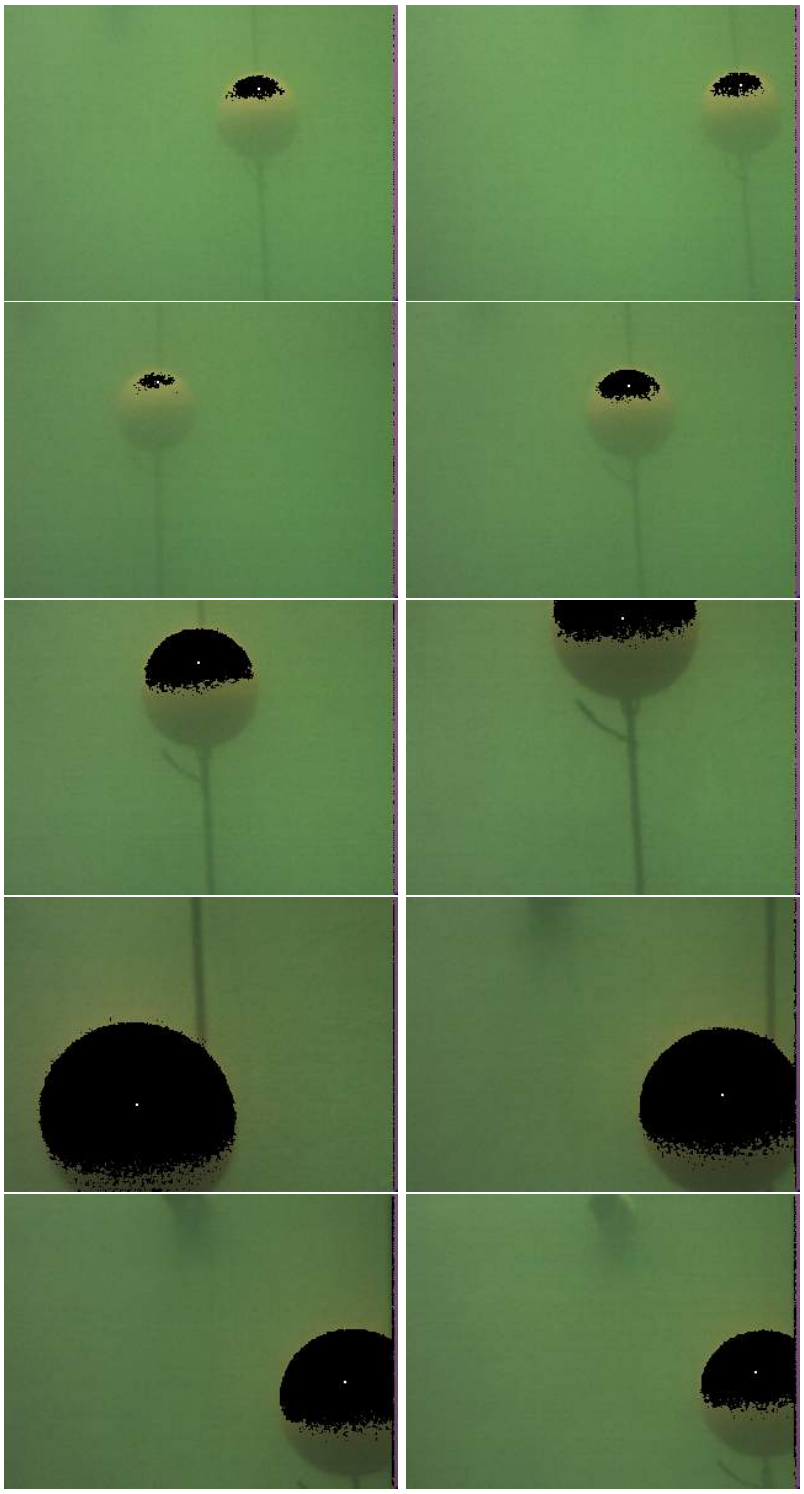


FIG. 7.8 – Détection et localisation de l'objet sphérique de couleur dans le bassin d'essai du GESMA (Juin 2007).

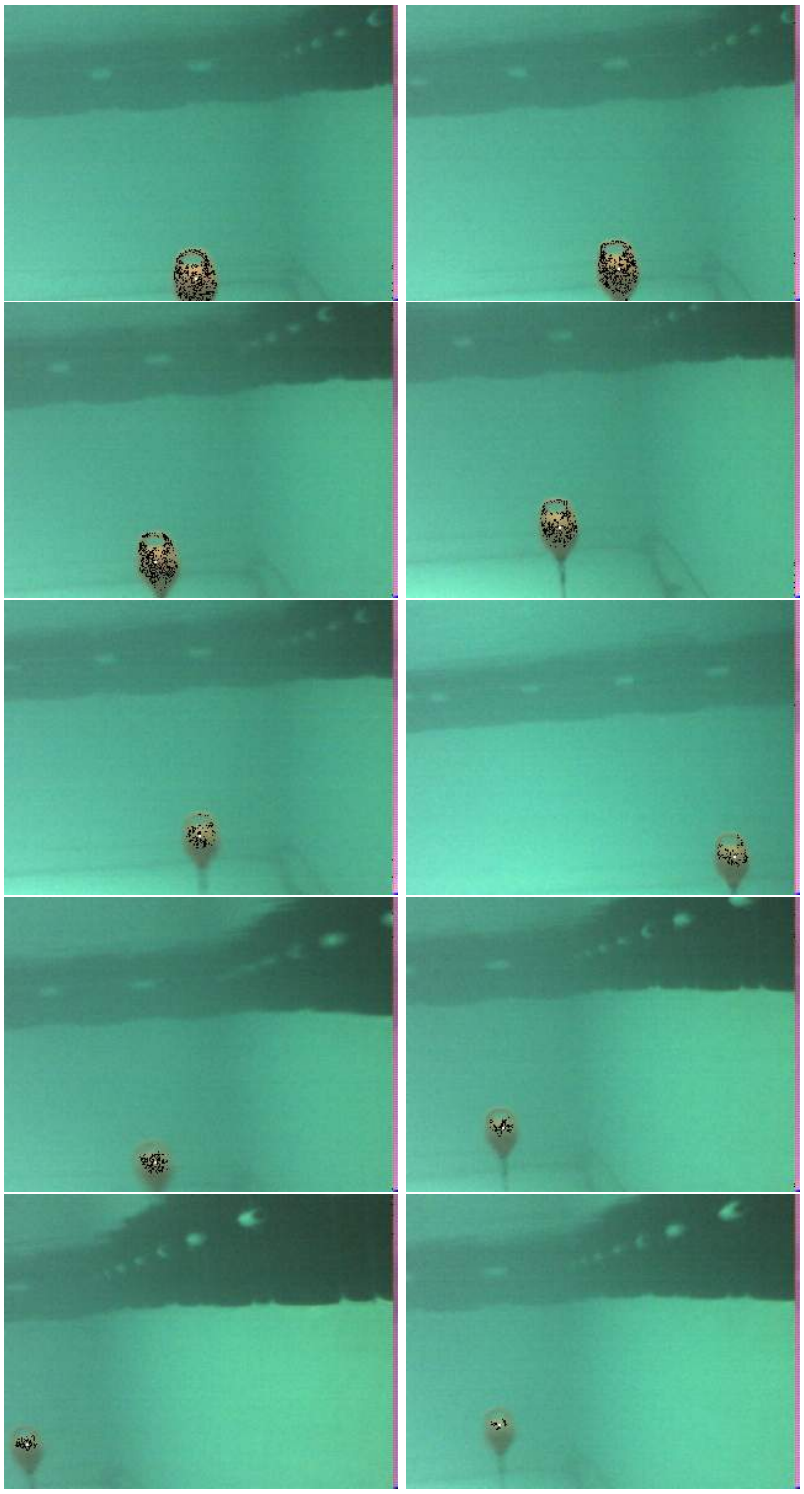


FIG. 7.9 – Détection et localisation de l'objet sphérique de couleur lors du concours SAUC'E 2007 dans le bassin QinetiQ.

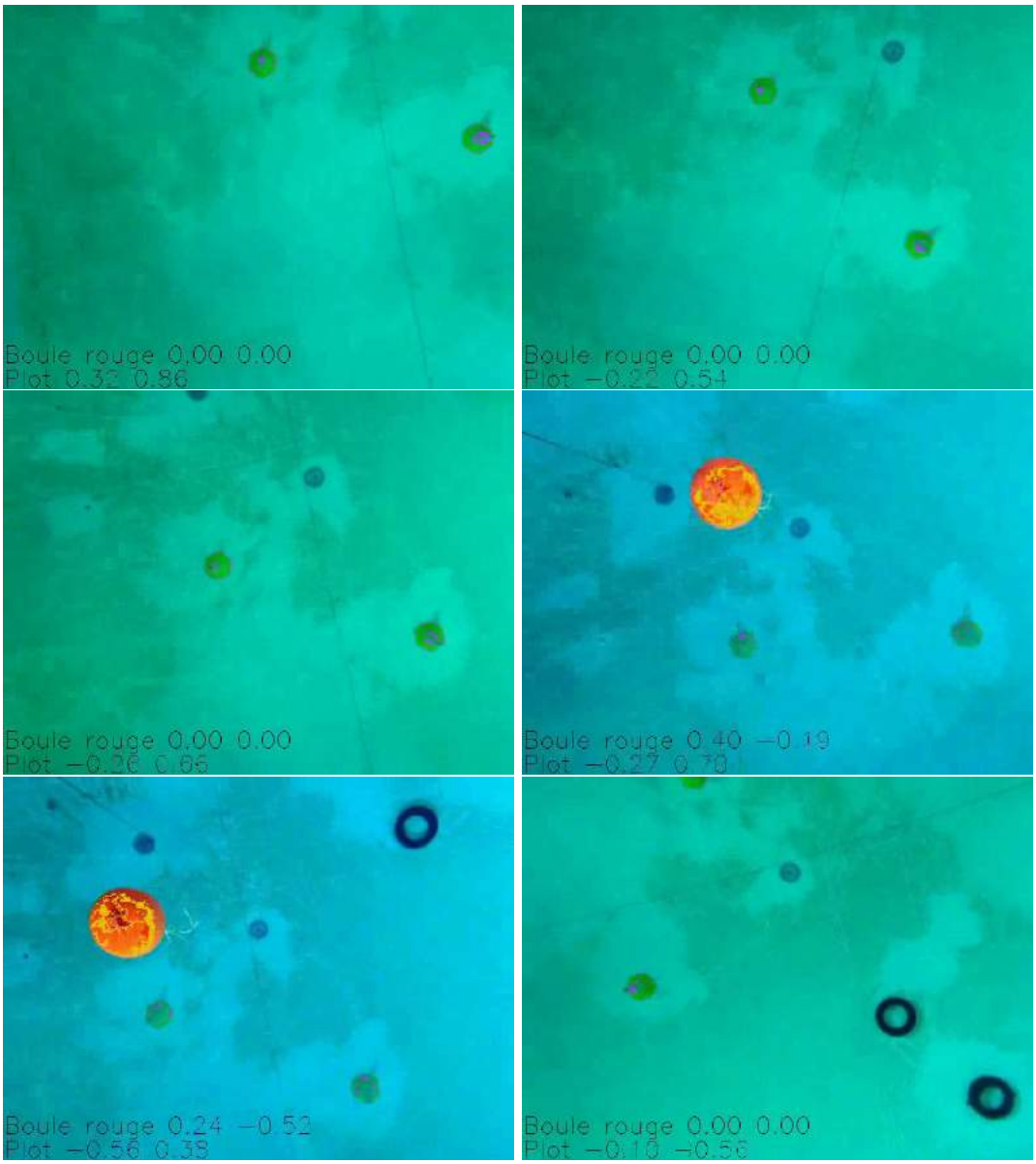
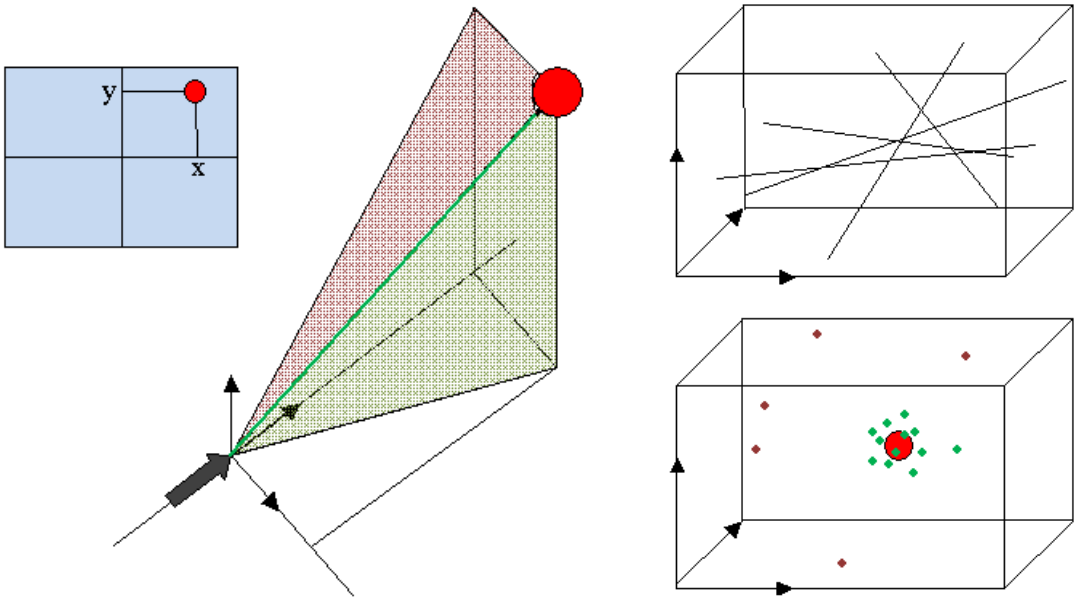


FIG. 7.10 – Détection et localisation d'un plot et d'une boule rouge sur une séquence d'image lors du concours SAUC'E 2008 dans le bassin IFREMER. Pixels compatibles avec les plots en violet et avec la boule en orange.

7.3.4 Améliorations, perspectives et conclusion.

Le concours SAUC'E nous a permis de valider la méthode de reconnaissance basée sur la couleur développée dans le chapitre 6. Cette méthode a d'ailleurs été réutilisée pour le concours SAUC'E 2008. En ce qui concerne les perspectives, une première idée serait d'exploiter la forme notamment pour des épreuves comme le franchissement du cadre immergé ou pour la recherche de la cible posée sur le fond sur laquelle il faut effectuer un largage.



(a) (b)

FIG. 7.11 – (a) Droite à estimer pour une vue de l'objet (b) ensemble des droites calculées pour chaque vue de l'objet et la position estimée de l'objet.

Une autre perspective importante sera d'utiliser la couleur afin d'estimer la distance. En effet grâce à la méthode développée nous pouvons estimer la distance à l'objet connaissant sa couleur nominale à une distance donnée. Cette estimation pourrait rendre le robot autonome uniquement grâce avec son système de vision en cas de défaillance du système de localisation par exemple. Pour se faire il faut pouvoir évaluer la distance non pas aux objets mais aux bords de la piscine (la détection d'objet devient alors détection d'obstacle). Ensuite il est possible d'explorer la majeure partie de la piscine et de réaliser dans des conditions favorables des missions à effectuer.

En conclusion sur cette utilisation du capteur vidéo, nous pouvons affirmer qu'il apporte une information très riche et exploitable comme nous l'avons montré dans cette partie. Il peut même en améliorant la robustesse des algorithmes se suffire à lui même comme unique capteur pour presque toutes les actions demandées pour réussir la mission.

7.4 Conclusion du chapitre

Dans ce chapitre, nous avons présenté des résultats applicatifs sur les deux méthodes de reconnaissance développées dans les chapitres 5 et 6. Ces résultats en contexte opérationnel ont été obtenus suite à la participation à deux projets TOPVISION et SAUC'E. Pour le premier projet nous avons choisi d'utiliser la méthode classique basée sur les contours. Dans ce contexte, en effet nous n'avions pas d'informations très précises sur les objets à rechercher et les conditions d'acquisition étaient très variables. Cette méthode générique et assez robuste à la grande variabilité des séquences nous a permis de valider l'étape de détection d'objet lors du projet TOPVISION. Notre méthode se démarque par

des temps de calculs plus importants mais des fausses alarmes nettement moins nombreuses. Nous avons expliqué dans ce chapitre que cette méthode pourrait être grandement améliorée que ce soit au niveau du temps de calcul ou de la précision des résultats. L'approche couleur développée dans le chapitre 6 n'a pas pu être appliquée dans le cadre de ce projet parce que nous ne disposions pas des informations *a priori* nécessaires et que la qualité des vidéos était trop moyenne pour l'exploitation des couleurs. L'algorithme devait en plus être automatique.
Cette seconde méthode basée sur l'exploitation unique de la couleur a été quant à elle testée lors du concours de robotique sous-marine SAUC'E auquel l'ENSIETA a participé pour la première fois en 2007. L'approche couleur était très appropriée à ce contexte puisque le but du concours était d'effectuer des missions en autonomie mais avec des connaissances importantes sur l'environnement qui se veut à la base très favorable. Sont notamment connus : la délimitation de la zone de jeu, l'ensemble des objets présents ainsi que les conditions d'acquisition. Dans ce contexte nous avons démontré la faisabilité d'une détection par couleur. En effet c'est un critère simple mais qui représente une information très riche même en milieu sous-marin. Cette méthode fut d'ailleurs optimisée et réutilisée pour le concours SAUC'E 2008, complétée par une méthode basée sur les contours mais ce avec une contrainte de traitement temps réel.
Pour conclure, ce chapitre a permis d'illustrer et surtout de valider les algorithmes présentés dans les chapitres 4, 5 et 6. Les conditions réelles prouvent que les méthodes développées dans cette thèse sont des solutions applicables opérationnellement.

Chapitre 8

Conclusion générale

De ses cinq sens, celui que l'homme utilise le plus est la vue.

La vision par ordinateur est la science qui étudie ce que les machines "voient". En temps que discipline scientifique, elle donne les fondements théoriques pour construire des systèmes artificiels qui obtiennent des informations à partir des images. En d'autres termes ils permettent à une machine de comprendre ce qu'elle voit lorsqu'on la connecte à une ou plusieurs caméras. En terme de discipline technique, la vision par ordinateur touche un large panel d'applications telles que :

- la commande de processus (robots industriels ou véhicules autonomes),
- la détection d'évènements (surveillance, comptage),
- l'organisation de l'information (indexation d'images ou de séquences d'images),
- la modélisation d'objet ou de l'environnement (cartographie, reconnaissance),
- l'interface homme/machine (écran tactile).

Les problèmes posés par la vision par ordinateur sont multiples et encore loin d'être résolus car ils sont souvent très complexes. C'est un sujet de recherche qui passionnent les chercheurs depuis plus de trente ans. Aujourd'hui, on confie aux systèmes de vision des tâches simples comme la reconstruction de scène, la détection d'événement, le suivi ou la reconnaissance d'objet, la restauration d'image... Dans le cadre de cette thèse nous avons concentré notre travail sur une petite partie du vaste domaine de la vision par ordinateur : l'étude des traitements automatiques de reconnaissance d'objets en vision sous-marine monoculaire.

La reconnaissance d'objets est une tâche qui peut paraître très simple pour un opérateur humain. Effectivement il est capable de s'adapter à une grande variété de situations comme par exemple le changement de fond, les variations d'illuminations, les occultations, ou les changements de points de vue. Il fait également facilement la différence entre un objet manufacturé et un objet naturel par exemple. A l'inverse, la reconnaissance d'objet est une tâche très complexe à réaliser par une machine. En effet, durant le processus de perception, l'homme utilise une large quantité d'informations : deux collections de points disponibles sur les rétines, et une quantité très importante de connaissances. Ces informations sont traitées avec une rapidité et une robustesse incroyables et avec une qualité telle qu'il peut même la plupart du temps "reconnaître" un objet qu'il n'a jamais vu. L'homme est donc sans nul doute le plus parfait des systèmes de reconnaissance.

Sans aller jusqu'a l'idée utopique de le copier, on doit donc essayer d'imiter quelques unes de ses facul-

tés. En reconnaissance d'objet par vision artificielle on va donc essayer d'imiter cette adaptation aux différentes conditions d'acquisition. Dans notre contexte, c'est-à-dire en milieu sous-marin, on doit en plus prendre en compte les difficultés supplémentaires comme la portée visuelle limitée, les contrastes faibles, le bruit important ou l'atténuation des couleurs. Ces problèmes font qu'aujourd'hui toutes les techniques usuelles sont inutilisables ou alors très peu fiables dans le contexte sous-marin. D'ailleurs c'est pour cela que l'on n'utilise la caméra que pour la télé-opération.
La reconnaissance d'objet en vision sous-marine monoculaire est un problème très délicat, dont l'étude est aujourd'hui indispensable avec l'émergence des véhicules sous-marins autonomes. Assurément la caméra vidéo est le capteur efficace à faible portée lors des phases d'approche, de reconnaissance d'objets et d'intervention grâce à sa haute résolution et sa facilité d'interprétation. Les enjeux de cette thèse étaient donc de développer de nouveaux algorithmes de vision sous-marine monoculaire pour la reconnaissance d'objet. Dans notre contexte, on a même restreint le problème aux objets manufacturés dans un contexte opérationnel de déminage ou de dépollution (fûts toxiques, containers, batteries à ions lourds, munitions, mines, et objets manufacturés divers souvent de grande consommation). Cette étude, même si elle se voulait générale, a été orientée par les données dont nous disposions dans ce contexte opérationnel particulier.

Les objectifs à atteindre étaient les suivants :
- produire un état de l'art sur les méthodes de correction permettant de s'affranchir des problèmes particuliers liés à l'imagerie sous-marine,
- produire un état de l'art sur les approches de reconnaissances et les différentes configurations utilisées pour l'identification d'objet,
- développer des algorithmes automatiques de reconnaissance répondant aux contraintes du projet TOPVISION,

De manière à apporter une contribution plus originale et profitant du développement au laboratoire d'un robot sous-marin en vue de la participation à des concours de robotique, nous avons été amené à considérer deux objectifs supplémentaires :
- développer des méthodes basées sur un autre critère que les contours,
- appliquer les méthodes développées aux données réelles acquises dans le cadre du concours de robotique sous-marine SAUC'E.

Les différentes études menées durant cette thèse ont ainsi abouti à cinq contributions :
- un état de l'art complet sur la reconnaissance d'objet en vidéo sous-marine ainsi que sur le prétraitement des images sous-marines,
- une méthode de prétraitement automatique des images sous-marines sans réglage manuel et sans *a priori* [178][119],
- une méthode de détection automatique des objets manufacturés par vision sous-marine monoculaire basée sur l'attribut classiquement utilisé : les contours [128],[1]
- une méthode originale de reconnaissance d'objets par leur couleur connue *a priori*, [167][2]
- l'application de ces deux algorithmes sur des données réelles au sein du projet TOPVISION et du concours SAUC'E.

[1] Une publication présentant les résultats d'évaluation du projet TOPVISION sera soumise d'ici la fin de l'année 2008.
[2] Un deuxième article sur ce sujet a été soumis en Mai 2008 à la revue IEEE Oceanic Engineering.

Pour commencer ces travaux nous sommes partis de l'existant. Le principal travail réalisé sur ce sujet de recherche est la thèse de Olmos [1] publié en Décembre 2002. Cette thèse aborde la problématique de détection d'objets manufacturés dans des vidéos sous-marines. Elle présente une technique automatique de restauration par filtrage inverse, un système d'extraction de contour multirésolution avec une recherche automatique de l'échelle optimale, un algorithme d'extraction d'attributs basé sur l'attention visuelle et le groupement visuel, et une méthode de classification d'attributs aboutissant à la détection ou non d'un objet manufacturé. Les résultats obtenus sont intéressants, notamment le développement d'une méthode complète et automatique testée sur des données réelles. Cependant cette méthode n'est pas vraiment applicable à notre contexte. Pour s'en convaincre, on peut citer deux raisons majeures. D'une part nous ne pouvions réutiliser la méthode de restauration car elle ne prend en compte uniquement la composante de diffusion directe c'est-à-dire le flou dans l'image. Une telle correction est trop légère par rapport aux nombreux problèmes présents dans nos images et elle suppose en plus quelques connaissances *a priori* sur le milieu. D'autre part, l'extraction de contour multirésolution avec convergence automatique vers l'échelle optimale est lourde en temps de calcul et ne peut pas apporter des résultats robustes dans notre contexte tant le fond est varié et parfois prédominant dans les images que nous avons à traiter.
Partant de ce constat, après une étude bibliographique complète sur les méthodes de reconnaissance d'objet et de prétraitement des images, nous avons donc poursuivi par le développement d'un nouvel algorithme de prétraitement permettant de corriger les défauts présents dans nos images et préparant de manière robuste la segmentation par détection de contour. L'algorithme développé est séquentiel et basé sur le rehaussement et le débruitage. Il permet de corriger l'éclairage non uniforme, de rehausser les contrastes et les couleurs, de débruiter l'image et ce, de manière totalement automatique sans aucun paramétrage de l'utilisateur. De plus, cet algorithme est rapide et a été testé et validé sur un grand nombre d'images très diverses. Son bénéfice sur la détection de contour a été démontré grâce à un critère basé sur l'histogramme du gradient.

Ensuite le problème de reconnaissance de l'objet à proprement dit dans les images débruitées a été découpé en deux phases : d'abord, une phase de détection de scène d'intérêt puis, une phase de reconnaissance dans le cas où un objet est présent. Cette méthode n'a pas abouti au développement complet d'un système de reconnaissance général même si nous avons dans certains cas précis, obtenu cette reconnaissance, notamment pour les objets sphériques ou cylindriques. Dans de nombreux autres cas nous ne pouvons conclure sur la reconnaissance étant donné que l'on ne dispose d'aucune connaissance *a priori* et que les paramètres extraits sont insuffisants. Nous avons essentiellement concentré nos efforts sur la segmentation, la sélection automatique des contours et l'extraction d'attributs robustes. La méthode de reconnaissance développée est donc utilisée aujourd'hui comme une méthode de détection des objets manufacturés dans des vidéos sous-marines ce qui correspond à la première étape de l'algorithme de reconnaissance dans le cadre du projet TOPVISION. Dans ce cadre, les attributs extraits servent seulement à confirmer la présence d'un objet. Il est nécessaire d'extraire d'autres attributs et d'appliquer un algorithme de classification pour aboutir à une méthode de reconnaissance complète, ce qui est d'ailleurs une des perspectives de ce travail. Quoi qu'il en soit, la méthode développée est automatique et sans *a priori* sur les objets, c'est-à-dire que ne n'avons pas de modèles des objets seulement des considérations simples de formes comme la présence de droites, de cercles et d'ellipses. Elle permet de segmenter l'image prétraitée avec une meilleure qualité de contours et de

sélectionner automatiquement dans les images les contours les plus pertinents (qui serviront ensuite à reconnaître l'objet). Elle a été validée lors du séminaire d'évaluation du projet TOPVISION avec des performances satisfaisantes.

Cette méthode, plutôt classique au vu de l'existant était très difficile à appliquer puisque sans aucun *a priori* (approche type "bottom up", image, segmentation, extraction d'attributs, classification, reconnaissance) et avec des images de qualité médiocre. Dans un deuxième temps, nous nous sommes donc tournés vers des approches apportant de l'*a priori* pour améliorer les résultats. Ce type d'approches que l'on nomme dans la littérature "top-down" fait l'hypothèse que l'image contient l'objet connu *a priori* que l'on cherche, et s'attache donc seulement à vérifier sa présence. Avec cette approche plus simple et plus robuste que la précédente, deux connaissances sur l'objet étaient exploitables : la forme et la couleur. Dans notre cas, puisque la couleur était un attribut très peu exploité dans le domaine sous-marin nous avons opté pour le développement d'une méthode basée sur la couleur. La méthode basée sur la forme a été étudiée et présente des aspects très prometteurs dans un contexte réaliste mais son développement est encore incomplet et entre dans les perspectives de ces travaux de thèse.
Alors qu'elle est très peu exploitée dans le milieu sous-marin, la couleur est un attribut pertinent souvent utilisé dans le contexte terrestre, notamment pour sa simplicité et sa robustesse face aux changements d'échelle, aux changements de prises de vue, aux occultations et à la distance. Cet attribut souffre par contre du problème de constance dû à l'éclairage et à la réflexion des objets. Dans le domaine sous-marin, en plus de ce problème, la couleur n'est pas invariante à la distance à cause du phénomène d'absorption. Pour faire face à ces problèmes, nous avons modélisé la déformation des couleurs dans l'eau due à l'éclairage et à la distance et recherché dans les images toutes les couleurs déformées provenant de la couleur de l'objet connue *a priori*. La méthode obtenue est robuste et très rapide. Elle nécessite néanmoins l'estimation préalable des coefficients d'absorption de l'eau. Ceux-ci peuvent être obtenus à partir de plusieurs images de calibration grâce à une méthode que nous avons développée. Cette méthode de reconnaissance sur la couleur a été testée dans des conditions réelles lors d'un concours de robotique sous-marine réunissant des étudiants européens. Les résultats prouvent que la couleur est un attribut utilisable et fiable dans l'eau.

Pour conclure sur les perspectives de travail, reprenons l'état de l'art sur les méthodes de reconnaissance d'objets. Nous avons montré que plusieurs approches sont théoriquement possibles respectant nos contraintes opérationnelles. L'utilisation de chaque méthode dépendant à chaque fois de l'information *a priori* disponible :

- disposant d'un modèle 3D de l'objet, on peut utiliser :
  - l'approche 3D avec reconstruction d'un modèle en stéréoscopie monoculaire,
  - l'approche 2D par compatibilité 2D/3D,
- disposant d'une image de l'objet, on s'orientera vers :
  - l'approche forme par classification à partir d'attributs sur les contours,
  - l'approche couleur avec reconnaissance par compatibilité de couleur,
- disposant d'indice sur la forme des objets manufacturés,
  - l'approche forme par recherche de forme particulières dans les images.

Quelques unes de ces approches ont été développées et rappelées ci-dessus. Les autres restent à l'état de perspectives. Parmi celles-ci, la partie 3D avec reconstruction n'a pas été implémentée car ce type de méthode se fait en général dans un environnement très contrôlé pour assurer de bons résultats. Cependant cette piste est théoriquement possible. On pourrait éventuellement faciliter le développement de cette méthode de reconnaissance en ajoutant soit un *a priori* sur le déplacement du robot, soit une deuxième caméra pour obtenir la stéréoscopie binoculaire. Une autre perspective envisageable à court terme serait de faire de la reconnaissance par compatibilité entre une image 2D et le modèle 3D de l'objet connu *a priori*. Cette méthode a été discutée dans le chapitre 6 mais n'a pas pu être véritablement validée pour les raisons citées précédemment.

Une autre suite possible à ces travaux serait d'améliorer les deux principales contributions de cette thèse. Pour l'approche classique, on peut compléter l'extraction d'attributs de la méthode de reconnaissance par les contours par l'ajout d'attributs tels que la régularité, la courbure des contours ou différents descripteurs de formes, puis, ajouter une étape de localisation de l'objet dans les images pour pouvoir extraire des attributs provenant uniquement de l'objet. Pour cela, l'utilisation de l'information de couleur, des méthodes de regroupement de contours, et l'utilisation des contours actifs pour détourer l'objet seraient envisageables. Enfin cette approche classique pourrait être améliorée en implémentant une méthode de classification pour finaliser le système de reconnaissance. Pour finir, concernant l'approche couleur on pourrait améliorer la méthode en utilisant des intervalles d'approximations des différents coefficients pour relâcher la contrainte sur l'information *a priori* nécessaire.

En conclusion ce qui ressort très bien de cette étude, c'est que le capteur vidéo donne une information très riche mais encore difficile à exploiter en contexte sous-marin notamment avec une approche classique. Cependant nous avons montré que les potentialités de ce capteur sont nombreuses. L'innovation apportée par la seconde approche basée sur la couleur en est la preuve. Grâce à elle, le capteur vidéo peut s'autosuffire notamment dans un contexte favorable comme le concours SAUC'E, pour des applications "simples" comme par exemple la navigation visuelle (détection de cible ou d'obstacle), la cartographie d'une zone, la recherche d'objet et l'intervention sur cet objet, le suivi ou le contrôle de commande, ou encore le SLAM (localisation et la construction de cartes).

Annexe A

Calibration et reconstruction 3D

Calibration

La caméra est l'outil géométrique qui fait le lien entre une scène 3D et son image 2D. Le processus de saisie d'une image à l'aide d'une caméra peut être décrit par un modèle géométrique assez simple. Ce modèle est caractérisé par un certain nombre de paramètres qui sont estimés lors d'une étape dite d'étalonnage ou de calibration. Il peut être décomposé par deux transformations (voir schéma ci-dessous) :

- une projection qui transforme un point de l'espace 3D dans le repère caméra en un point dans le repère 2D image.
- une transformation du repère 3D caméra au repère 3D scène.

Système de coordonnées scène (3D)	$\Longleftrightarrow$	Matrice des paramètres extrinsèques	$\Longleftrightarrow$	Système de coordonnées caméra (3D)	$\Longleftrightarrow$	Matrices des paramètres intrinsèques	$\Longleftrightarrow$	Système de coordonnées image (2D)

La première transformation est un changement de repère entre celui de la caméra et celui de l'image. Cette transformation est formalisée par la matrice des paramètres intrinsèques. Elle prend en compte les propriétés physiques du système et les caractéristiques technologiques du capteur. La seconde transformation est un changement de repère entre celui de la scène et celui de la caméra. Elle est représentée par la matrice des paramètres extrinsèques. Cette transformation possède 6 degrés de liberté : trois rotations et trois translations.
La relation algébrique qui lie l'espace 3D scène et l'espace 2D image peut être décrite de la manière suivante :

$$\underbrace{\begin{pmatrix} u \\ v \\ s \end{pmatrix}_{R_{\text{image}}}}_{\text{point image}} = \underbrace{\begin{pmatrix} fm_x & w & p_x \\ 0 & fm_y & p_y \\ 0 & 0 & f \end{pmatrix}}_{\text{paramètres intrinsèques}} \underbrace{\begin{pmatrix} R & t \end{pmatrix}}_{\text{paramètres extrinsèques}} \underbrace{\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}_{R_{\text{scène}}}}_{\text{point scène}} \tag{A.1}$$

avec, f la distance focale, m_x et m_y les facteurs d'échelle, w le biais d'orientation, p_x et p_y les coordonnées du point principal, R une matrice de rotation (3 x 3) et t un vecteur de translation (3 x 1).

La calibration consiste donc à estimer en même temps les paramètres intrinsèques et extrinsèques à partir d'un objet étalon dont la position est connu avec précision. Pour simplifier on considère en général des pixels carrés et aucun biais d'orientation. L'objet étalon que l'on utilise couramment est une mire plane constituée de carrés noirs sur fond blanc. Elle doit fournir des points facilement détectable par traitement d'image pour ensuite effectuer une estimation robuste des différents paramètres. A noter que la calibration d'une caméra consiste en réalité en la calibration de l'ensemble objectif optique, caméra convertisseur analogique/digital. Si l'on change l'ouverture, la mise au point, la focale, ou les paramètres du convertisseur analogique/digital, il faut alors recalibrer l'ensemble du système.

Reconstruction 3D

L'utilisation de la géométrie projective permet de décrire sous une forme algébrique simple la relation qui lie les espaces 3D scène et 2D image. Utilisant cette relation, et à partir de plusieurs images d'une même scène on va donc pouvoir en faire une reconstruction 3D. Cette reconstruction va être basée sur l'appariement de point dans les différentes images. A partir de deux points image (2D) d'un même point dans la scène (3D), un algorithme de triangulation permet de retrouver la composante de profondeur.

La méthode classique pour la reconstruction est la suivante :

1. Prétraitement de l'image et extraction de points caractéristiques.
   - Points de contour (Sobel, Canny, Deriche),
   - Points d'intérêts et coins (Susan, Css, Harris),
   - Points invariants à l'espace échelle.
2. Mise en correspondance des points caractéristiques.
   - Méthode de corrélation,
   - Méthode de suivi de points (points d'intérêts),
   - Méthode de caractérisation de points par des invariants locaux.
3. Reconstruction 3D à un facteur d'échelle près.
   - Calibration de la caméra.
   - Estimation de la pose pour chaque image.
   - Reconstruction 3D par triangulation.

La reconstruction 3D est en générale faite à un facteur d'échelle près, il est nécessaire de disposer d'une information métrique sur la scène (par exemple la taille réelle d'un objet ou bien la distance entre la caméra et un élément de la scène) pour lever cette indétermination et reconstruire un modèle métrique.

Annexe B

Le robot SAUCISSE

SAUCISSE (pour SAUC'Interval Super Submarine Of Ensieta) est le robot sous-marin de l'ENSIETA ayant participé au concours SAUC'E en 2007 et 2008. La deuxième édition du concours s'est déroulée en Juillet 2007 à Gosport dans la baie de Postsmouth (UK). Elle a réuni six équipes et SAUCISSE s'est hissé sur la troisième marche du podium et a également remporté le prix "innovation et contrôle". En 2008 pour sa troisième édition le concours aura lieu à Brest.
La première version du robot sous-marin SAUCISSE est présentée sur la figure B.1. Sa version 2008 a subie quelques améliorations tout en gardant la même architecture mécanique, électronique et informatique.

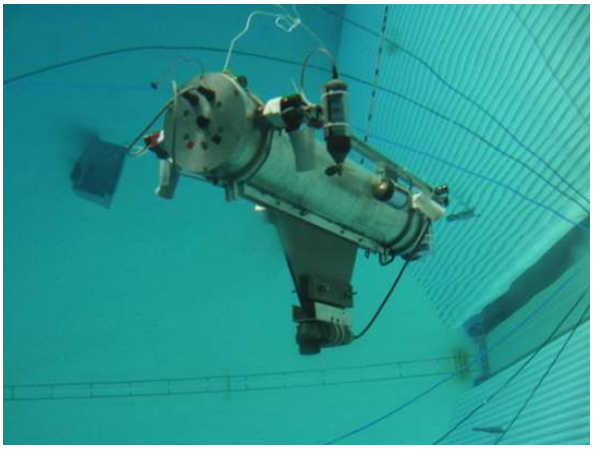


FIG. B.1 – Vue du robot SAUCISSE 2007 en plongée.

Dans sa version 2008 l'équipement du robot est le suivant : En ce qui concerne l'architecture externe il est composé de :

– Un tube d'aluminium d'environ 70cm avec une quille lestée et deux barres de fixation sur les cotés. Ce tube est fermé par deux tapes en aluminium sur lesquelles sont fixés les connecteurs qui permettent la communication avec les périphériques extérieurs. Il contient bien évidemment l'ensemble de l'électronique embarquée ainsi que les batteries.

- Un sonar compact permet la localisation du sous-marin dans la piscine par détection des murs.
- Un capteur de pression détermine la profondeur du sous-marin.
- Un dispositif d'arrêt d'urgence sous la forme d'un aimant permet de commander un interrupteur de mise en marche du robot.
- Une antenne WiFi autorise la communication en surface et lors des phases de plongée jusqu'à 1m, avec un PC extérieur ou une console de type PSP.
- Deux webcams (une frontale et une ventrale) repèrent les différents objets ou obstacles dans la piscine.
- Trois propulseurs (un vertical et deux horizontaux) permettent le déplacement du robot.

En ce qui concerne l'électronique embarquée elle est constitué de :

- Un PC embarqué (PC104) alimenté en tension continue 5V, processeur 1.44Ghz, sous Linux.
- Un boiter LabJack UE9 permettant de générer des signaux PWM pour commander les moteurs,
- Un routeur WiFi autorisant la communication entre le PC embarqué et un PC extérieur.
- Une centrale d'attitude utilisée pour l'aide à la navigation.
- Trois cartes de puissance permettant de délivrer les tensions requises aux moteurs.

On peut voir sur la figure B.2 un organigramme des entrées sorties du robot.

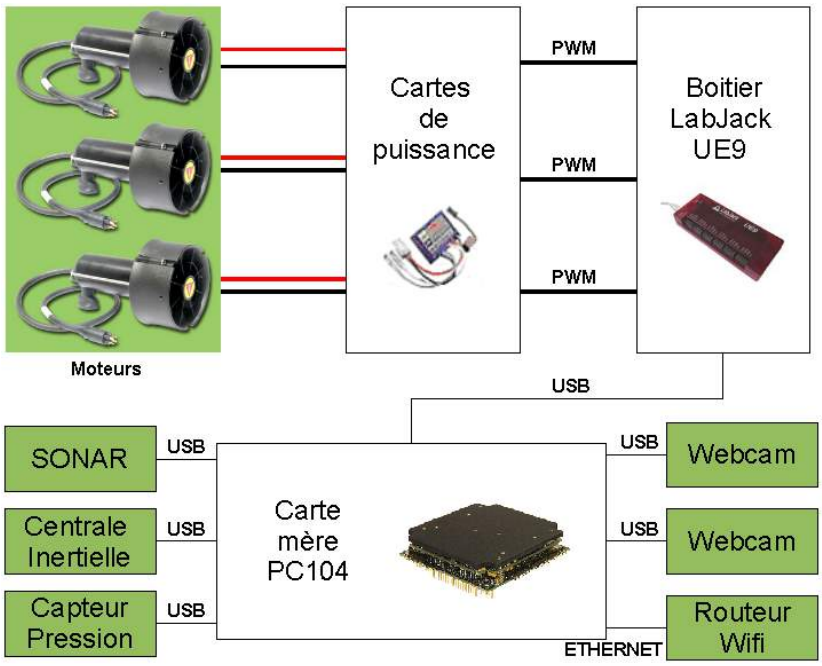


FIG. B.2 – Organigramme des entrées-sorties.

Annexe C

La webcam : un capteur couleur

La caméra est l'un des éléments principaux de la chaîne d'acquisition d'une image, il en existe de nombreux types. Dans notre cas nous nous intéresserons seulement aux caméras de type webcam qui se connectent à l'ordinateur via un port USB. Ce type de caméra a été choisi essentiellement pour son faible coût et pour sa facilité d'utilisation en vue d'applications robotiques.
Les webcams possèdent un capteur, qui peut être CCD, ou CMOS. Pour améliorer l'image, les webcams sont souvent couplées à un système logiciel d'interpolation, ayant pour but d'afficher une image détaillée à partir d'une image de faible qualité en créant des pixels intermédiaires dont la couleur est calculée par comparaison aux pixels adjacents. La résolution d'une webcam est en général assez faible comparée aux caméras numériques standards. Elle peut atteindre 640 x 480 points mais en général elle est de 320 x 240. Naturellement, son capteur est sensible à l'ensemble du spectre de la lumière visible, et c'est grâce à un filtre de Bayer que l'on obtient une image couleur (voir chapitre 2).

FIG. C.1 – Une webcam haute résolution de la marque Logitech.

FIG. C.2 – Webcam haute résolution étanche.

Sur le robot sous-marin SAUCISSE, nous avons utilisé en 2007 une webcam bas de gamme de résolution 320 x 240 et sans autofocus (voir Fig. 7.8 et 7.9). Cependant vu la qualité des images et étant donné les difficultés liées à l'imagerie sous-marine nous avons opté en 2008 pour une webcam de haute qualité avec une haute résolution. Cette nouvelle webcam présentée sur la figure C.1 doit permettre de largement améliorer les résultats. En effet, sa résolution est de 800 x 600 (environ 2 millions de pixels) et peut atteindre jusqu'a 1600 x 1200 par extrapolation à 8 millions de pixels. Cette webcam utilise en plus la technologie RightLight qui compense le manque de lumière et dispose d'une optique Carl Zeiss avec mise au point automatique. Enfin elle est dotée aussi d'un champ de vision élargi. Elle devrait permettre dans le cadre du concours SAUC'E 2008 d'améliorer les résultats de détection (voir Fig. C.2).

Glossaire

ACP : Analyse en Composantes Principales.
Analogique : Se dit de données représentés par la variation continue d'une grandeur physique, en opposition à numérique où on a une suite de valeurs discrètes.
AUV : Autonomous Underwater Vehicle.
CCD : Charge-Coupled Device.
CIE : Commission Internationale de l'Éclairage.
CMOS : Complementary metal oxide semi-conductor.
Chromatique : Relatif à la couleur.
DFT : Discrete Fourier Transform.
DGA/DET : Délégation générale pour l'Armement / Direction de l'Expertise Technique.
DSTL : Defense Science and Technology Laboratory.
ENSIETA : École Nationale Supérieure des Ingénieurs des Etudes des Techniques d'Armement.
E3I2 : Extraction et Exploitation de l'Information en Environnements Incertains.
FFT : Fast Fourier Transform.
Gamut : Gamme de couleur qu'un périphérique peut reproduire. Espace de toutes les couleurs possibles par synthèse additive ; dépend du blanc de référence et des primaires utilisés par le moniteur.
GESMA : Groupes d'études sous-marines de l'Atlantique.
HSV : Hue Saturation Value.
Irradiance : Flux de radiation arrivant sur une surface par unité d'air.
Luminance : Intensité d'une source de lumière. Notion qui traduit la sensation de luminosité indépendamment de la couleur.
Lambertien : Propriété des surfaces qui sont purement diffuses.
MATLAB : Matrix Laboratory.
OTF : Optical Transfert Function.
PC : Personal Computer.
PSF : Point Spread Function.
PSNR : Peak Signal to Noise Ratio, mesure de distorsion entre deux images numériques.
RANSAC : RANdom SAmple Consensus.
Réflexion diffuse : Émission égale dans toutes les directions par une surface.
Réflectance : Fraction de la lumière incidente réémise par réflexion.
RGB : Red Green Blue.
ROV : Remote Operated Vehicle.
Saturation : Intensité de la teinte (vive ou fade).

SAUC'E : Student Autonomous Underwater Challenge - Europe.
SAUC'ISSE : SAUC' Interval Super Submarine of Ensieta.
SLAM : Simultaneous Localization And Mapping.
SONAR : SOund Navigation And Ranging.
Spéculaire : Réflexion de la lumière avec une préférence directionnelle.
SSIM : Structural SIMilarity, mesure de similarité entre deux images numériques en terme de qualité visuelle.
Teinte : Forme pure d'une couleur, c'est-à-dire sans adjonction de noir ou de blanc qui permettent d'obtenir ses nuances.
TOPVISION : Test Opérationnels de VIdéos Sous-marines pour l'Identification des Objets Nuisibles.
TUS : Thales Underwater Systems.
USB : Universal Serial Bus.
WiFi : Wireless Fidelity.

Publications

- S. Bazeille, I. Quidu and L. Jaulin. *Using light absorption for the identification of underwater coloured objects.* IEEE Journal of Oceanic Engineering. 2008. [Soumis].
- S. Bazeille, I. Quidu and L. Jaulin. *Identification of underwater man-made object using a colour criterion.* Proceedings of the Institute of Acoustics, Vol. 29. Pt6. 2007.
- S. Bazeille, I. Quidu and L. Jaulin and J-P. Malkasse. *Une méthode de prétraitement automatique pour le débruitage des images sous marines.* Revue Traitement du Signal, numéro spécial "Caractérisation du milieu marin". 2007.
- S. Bazeille, I. Quidu, L. Jaulin and J.P Malkasse. *Automatic underwater image pre-preprocessing.* Proceedings of the Sea Tech Week "Caractérisation du milieu marin", 2006.
- S. Helan, J-C. Burie, T. Bouwmans and S. Bazeille. *Object detection in underwater images.* Proceedings of the Sea Tech Week "Caractérisation du milieu marin". 2006.

Bibliographie

[1] A. T. Olmos Antillon. *Detecting underwater man-made objects in unconstrained video image.* PhD thesis, Heriot Watt University, Department of Computing and Electrical Engineering, 2002.

[2] J. Ahlen. *Color correction of underwater images using spectral data.* PhD thesis, Uppsala University, Centre for Image Analysis, 2005.

[3] N. Vandenbroucke. *Segmentation d'images couleur par classification de pixels dans les espaces d'attributs colorimétriques adaptés. Application à l'analyse d'images de football.* PhD thesis, Université de Lille, 2000.

[4] L. Busin. *Sélection automatique d'espaces couleur pour la segmentation d'images. Application au contrôle qualité de motifs de verres serigraphiés.* PhD thesis, Université de Lille, 2006.

[5] A. Tremeau, C. Fernandez-Maloigne, and P. Bonton. *Image numérique couleur : De l'acquisition au traitement.* Editions Dunod, 2005.

[6] R. Bouillot. *Cours de traitement numérique de l'image.* Editions Dunod, 2005.

[7] R. Bouillot. *Cours de photographie numérique : Principes, acquisition et stockage.* Editions Dunod, 2006.

[8] J.-P. Couwenbergh. *L'indispensable pour maîtriser la couleur.* Marabout Informatique, 1992.

[9] G. Niemetzky. *Gestion de la couleur - Calibrage et profils ICC.* Eyrolles, 2004.

[10] Bruce Fraser and Chris Murphy. *Gestion des couleurs.* Peachpit Press., 2003.

[11] P. D. Sherman. *Colour vision in the nineteenth century : The Young-Helmholtz-Maxwell theory.* Hilger. Bristol., 1981.

[12] L. Zuppiroli and M.-N. Bussac. *Traité des couleurs.* Presses Polytechniques et Universitaires Romandes, 2001.

[13] C. Le Moine, C. Pedersen, and B. Zana. La perception visuelle, un système de haute technologie. *Revue TDC (Textes et Documents pour la Classe), N. 817*, 817, 2001.

[14] B. Zana and C. Le Moine. http ://www.cndp.fr/revuetdc/817-41419.htm.

[15] J.-M. Jolion. *Les systèmes de vision.* Lavoisier, 2000.

[16] H. Bernard, F. Helt, and M. Sintas. *L'image numérique et le cinéma.* Eyrolles, 2000.

[17] P. Martinez and A. Klotz. *Le Guide Pratique de l'Astronomie CCD.* Editions ADAGIO, 1994.

[18] G. Asch. *Les capteurs en instrumentation industrielle.* Editions Dunod, 2003.

[19] G. Asch. *Acquisition de données : Du capteur à l'ordinateur.* Editions Dunod, 2003.

[20] P. Courtellemont. http ://perso.univ-lr.fr/pcourtel/espadon/site_web/ch5/page5-2.htm.

[21] M. Bret. *Images de synthèse.* Dunod Informatique, 1988.

[22] G. J. Klinker, S. A. Shafer, and T. Kanade. A physical approach to color image understanding. *International Journal of Computer Vision*, 4(1), 1990.

[23] A Ishimaru. *Electromagnetic wave propagation, radiation, and scattering.* Prentice Hall Englewood Cliffs, 1991.

[24] R. Garcia, T. Nicosevici, and X. Cufi. On the way to solve lighting problems in underwater imaging. *IEEE OCEANS*, 2002.

[25] A. Arnold-Bos, J. P. Malkasse, and G. Kervern. Towards a model-free denoising of underwater optical image. *IEEE OCEANS*, 2005.

[26] Y. Y. Schechner and N. Karpel. Clear underwater vision. *Computer Vision and Pattern Recognition*, 2004.

[27] M. Arredondo and K. Lebart. A methodology for the systematic assessment of underwater video processing algorithms. *IEEE OCEANS*, 2005.

[28] J. S. Jaffe. Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, 15(2), 1990.

[29] A. Guillen I. Guillen. *La photo numérique sous-marines - Guide expert.* Editions VM, 2005.

[30] A. Arnold-Bos. Reconnaissance et classification automatiques d'objets en vidéo sous-marine. *Master Thesis, Rapport interne TUS SAS 04/B/EGS/RG/112-AA*, 2004.

[31] D. M. Kocak and F. M. Caimi. The current art of underwater imaging, with a glimpse of the past and vision of the future. *MTS Journal*, 39, 2005.

[32] S. Reed, Y. Petillot, and J. Bell. An automatic approach to the detection and extraction of mine features in sidescan sonar. *IEEE Journal of Oceanic Engineering*, 28(1) :90–105, 2003.

[33] Y. Petillot, I. Tena Ruiz, and D. M. Lane. Underwater vehicle obstacle avoidance and path planning using a multi-beam forward looking sonar. *IEEE Journal of Oceanic Engineering*, 26(2), 2001.

[34] J. S. Jaffe, J. McLean, M. P. Strand, and K. D. Moore. Underwater optical imaging : Status and prospects. *Oceanography*, 2001.

[35] H. Duo-Min and G. G. Seet. Underwater vision enhancement in turbid water by range-gated imaging system. *Proceedings Lasers and Electro-Optics*, 2001.

[36] E. A. McLean, H. R. Burris, and M. P. Strand. *Short-pulse range-gated optical imaging in turbid water.* Applied Optics, 1995.

[37] R. Horaud and O. Monga. *Vision par ordinateur, Outils fondamentaux.* Editions Hermès, deuxième édition revue et augmentée, 1995.

[38] F. X. Espiau. *Métrologie 3D par vision active sur des objets naturels sous marins.* PhD thesis, Université de Nice - Sophia Antipolis, 2002.

[39] S. Negahdaripour and H. Madjidi. Stereovision imaging on submersible platforms for 3-d mapping of benthic habitats and sea-floor structures. *IEEE Journal of Oceanic Engineering*, 28(4) :625–650, 2003.

[40] N. Pessel. *Auto-calibrage d'une caméra en milieu sous marin.* PhD thesis, Université de Montpellier II, 2003.

[41] K. Plakas and E. Trucco. Uncalibrated vision for 3-d underwater applications. *IEEE OCEANS*, 1998.

[42] K. Plakas and E. Trucco. Uncalibrated computer vision techniques in underwater environment. *Underwater Applications of Image Processing*, 11, 1998.

[43] S. M. Zanoli and P. Zingaretti. Underwater imaging system to support rov guidance. *IEEE OCEANS*, 1998.

[44] G. L. Foresti, S. Gentilli, and M. Zampato. A vision-based system for autonomous underwater vehicule navigation. *IEEE OCEANS*, 1998.

[45] G. L. Foresti. Visual inspection of the sea bottom structures by an autonomous underwater vehicle. *IEEE Transactions on Systems, Man, and Cybernetics*, 31(5), 2001.

[46] G. L. Foresti and S. Gentili. A hierarchical classification system for object recognition in underwater environments. *IEEE Journal of Oceanic Engineering*, 27(1), 2002.

[47] R. Marks, S. Rock, and M. Lee. Using visual sensing for control of an underwater robotic vehicle. *Proceedings of IARP Second Workshop on Mobile Robots*, 1994.

[48] R. Marks, H. Wang, M. Lee, and S. Rock. Automatic visual station keeping of an underwater robot. *IEEE OCEANS*, 1994.

[49] R. Marks, S. Rock, and M. Lee. Real time video mosaicking of the ocean floor. *IEEE Journal of Oceanic Engineering*, 20(3), 1995.

[50] O. Pizarro and H. Singh. Toward large-area mosaicing for underwater scientific applications. *IEEE Journal of Oceanic Engineering*, 28(4) :651–672, 2003.

[51] S. D. Fleischer and S. M. Rock. Experimental validation of a real time vision sensor and navigation system for intelligent underwater vehicles. *IEEE International Conference on Intelligent Vehicles (ICIV)*, 1998.

[52] T. Rees. Augmented state kalman filtering for auv navigation. *Research report*, 2002.

[53] J. Sattar, P. Giguere, G. Dudek, and C. Prahacs. A visual servoing system for an aquatic swimming robot. *International Conference on Intelligent Robots and Systems*, 2005.

[54] G. Tascani, P. Zingaretti, and G. Conte. Real-time inspection by submarine images. *Journal of Electronic Imaging*, 5(4), 1996.

[55] J. Guo, S.-W Cheng, C.-Y Ying, and T.-C Liu. Image registration for the underwater for the underwater inspection using a maximum a posteriori technique. *IEEE Journal of Oceanic Engineering*, 28(1), 2003.

[56] X. Tang and al. Automatic plankton image recognition. *Artificial Intelligence Review*, 1998.

[57] M. Chambah, D. Semani, A. Renouf, P. Courtellemont, and A. Rizzi. Underwater color constancy : enhancement of automatic live fish recognition. *Proceedings of the IS&T/SPIE 16th Annual Symposium on Electronic Imaging*, 2004.

[58] K. Lebart, E. Trucco, and D. M. Lane. Real-time automatic sea-floor change detection from video. *IEEE OCEANS*, 2000.

[59] D. R. Edgington, K. A. Salamy, M. Risi, R. E. Sherlock, D. Walther, and C. Koch. Automated event detection in underwater video. *IEEE OCEANS*, 2003.

[60] D. Walther, D. R. Edgington, and C. Koch. Detection and tracking of objects in underwater video. *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2004.

[61] J. Batlle, A. Casals, J. Freixenet, and J. Marti. A review on strategies for recognizing natural objects in colour images of outdoor scenes. *Image and Vision Computing*, 2000.

[62] X. Yang, Q. Sun, and Q. Tian. Content-based video identification : a survey. *IEEE International Conference on Information Technology (ICIT)*, 2003.

[63] A. T. Olmos Antillon and Emanuelle Trucco. Detecting man-made objects in unconstrained subsea videos. *Proceedings of the British Machine Vision Conference (BMVC)*, 2002.

[64] G. Dudek and al. A visually guided swimming robot. *International Conference on Intelligent Robots and Systems*, 2005.

[65] J. Sattar and G. Dudek. On the performance of color tracking algorithms for underwater robots under varying lighting and visibility. *IEEE International Conference on Robotics and Automation (ICRA)*, 2006.

[66] R. Horaud and P. Bolles. A three dimensional part orientation system. *International Journal of Robotics Research*, 5(3) :3–26, 1986.

[67] O. Camps, L. Shapiro, and R. Haralick. *Image prediction for computer vision.* Elsevier Science, 1993.

[68] P. J. Besl and R. C. Jain. Three dimensionnal object recognition. *Computing survey*, 1985.

[69] R. Hu and Q. Ji. Camera self calibration from ellipse correspondences. *IEEE International Conference on Robotics and Automation (ICRA)*, 2001.

[70] M. Pollefeys. *Self-calibration and metric 3D reconstruction from uncalibrated image sequences.* PhD thesis, Katholieke Universiteit Leuven, Belgium, 1999.

[71] A. Crimisi and A. Zisserman. A new approach to to obtain height measurements from video. *Proceedings of SPIE*, 1999.

[72] A. Crimisi, I. Reid, and A. Zisserman. Single view metrology. *International Journal of Computer Vision*, 40(2), 2000.

[73] P. Strum and S. Maybank. A method for interactive 3d reconstruction of piecewise planar objects from single images. *British Machine Vision Conference (BMVC)*, 1999.

[74] H. Foroosh, M. Balci, and X. Cao. Self-calibrated reconstruction of partially viewed symmetric objects. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005.

[75] M. Prasad, A. Zisserman, and A. Fitzgibbon. Single view reconstruction of curved surfaces. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[76] P. F. Sturm. *Vision 3D non calibrée : contributions à la reconstruction projective et études des mouvements critiques pour l'auto-calibrage.* PhD thesis, Institut National Polytechnique de Grenoble, 1997.

[77] C. Barat and M. J. Rendas. Exploiting natural contours for automatic sonar-to-video calibration. *IEEE OCEANS*, 2005.

[78] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31(8) :983–1001, 1998.

[79] D. G. Lowe. Three dimensional object recognition from single two-dimentionnal image. *Artificial Intelligence*, 1987.

[80] R. A. Brooks. Symbolic reasonings among 3d models and 2d images. *Artificial Intelligence*, 1981.

[81] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1) :11–32, 1991.

[82] B. W. Mel. Seemore : Combining color, shape and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Computation*, 1997.

[83] N. Ayache and O. Faugeras. Hyper : a new approach for the recognition and positioning of two dimensional objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 8(1) :44–54, 1986.

[84] A. Delorme, J. Gautrais, and S. Thorpe R. VanRullen. Spikenet : A simulator for modeling large networks of integrate and fire neurons. *Neurocomputing*, 1991.

[85] P. Viola and M. Jones. Robust real time object detection. *International Journal of Computer Vision*, 2002.

[86] F. Mokhtarian, S. Abbasi, and J. Kittler. Robusts and efficient shape indexing through curvature scale space. *Proceedings of the British Machine Vision Conference (BMVC)*, 1996.

[87] T. M. Quang and W. W. Boles. Recognition of 2d objects contour using the wavelet transform zero-crossing representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(8) :910–916, 1995.

[88] D. H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern recognition*, 13(2) :111–122, 1981.

[89] J. P. Cocquerez and S. Philipp. *Analyse d'images : filtrage et segmentation*. Editions Masson, 1998.

[90] H. Maitre. *Traitement des images*. Editions Hermès, 2003.

[91] K. Lebart, C. Smith, E. Trucco, and D. M. Lane. Automatic indexing of underwater survey video : Algorithm and benchmarking method. *IEEE Journal of Oceanic Engineering*, 28(4), 2003.

[92] O. Carmichael and M. Herbert. Object recognition by a cascade of edge probes. *Proceedings of the British Machine Vision Conference (BMVC)*, 2002.

[93] K. Mikolajczyk, A. Zisserman, and C. Schmid. Shape recognition with edge-based features. *Proceedings of the British Machine Vision Conference (BMVC)*, 2003.

[94] D. G. Lowe. Object recognition from local scale-invariant features. *Proceedings of the International Conference on Computer Vision*, 1999.

[95] T. Gevers and A. W. M. Smeulders. Color based object recognition. *Lecture Notes In Computer Science*, 1997.

[96] S. Obdrzalek, J. Matas, and O. Chum. On the interaction between object recognition and colour constancy. *IEEE International Workshop on Color and Photometric Methods in Computer Vision*, 2003.

[97] A. Bovik. *Handbook of image and video processing*. Academic Press, 2000.

[98] Z. Liu, Y. Yu, K. Zhang, and H. Huang. Underwater image transmission and blurred image restoration. *Journal of Optical Engineering*, 40(6), 2001.

[99] B. L. McGlamery. A computer model for underwater camera systems. *Proceedings of SPIE Ocean Optics*, 1979.

[100] P. Perona and J.Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 12(7), 1990.

[101] G. W. Wei. Generalized perona-malik equation for image restoration. *IEEE Signal Processing Letters*, 1999.

[102] M. J. Black, G. Sapiro, D. Marimont, and D. Heeger. Robust anisotropic diffusion. *IEEE Transactions on Image Processing*, 7(3), 1998.

[103] D. N. Sidorov and A. C. Kokaram. Suppression of moiré patterns via spectral analysis. *Proceedings of SPIE Visual Communications and Image Processing*, 2002.

[104] W. H. Liao and D. Y. Li. Homomorphic processing techniques for near-infrared images. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2003.

[105] S. Mallat. *Une exploration des signaux en ondelettes*. Editions de l'Ecole Polytechnique, 2000.

[106] A. Buades, B. Coll, and J. M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling and Simulation (SIAM interdisciplinary journal)*, 2005.

[107] F. Luisier, T. Blu, B. Forster, and M. Unser. Which wavelet bases are the best for images denoising ? *Proceedings of SPIE*, 2005.

[108] D. L. Donoho and I. M. Johnstone. Threshold selection for wavelet shrinkage of noisy data. *Proceedings of 16th Annual International Conference of the IEEE on Engineering Advances*, 1994.

[109] L. Sendur and I. W. Selesnick. Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. *IEEE Transactions on Signal Processing,*, 50(11), 2002.

[110] A. F. Abdelmour and I. W. Selesnick. Symmetric nearly orthogonal and orthogonal nearly symmetric wavelets. *Research Report*, 2003.

[111] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81(3), 1994.

[112] P. Kovesi. Phase preserving denoising of images. *Proceedings of the Australian Pattern Recognition Society Conference*, 1999.

[113] J. L. Starck, F. Murtagh, E. J. Candés, and D. L. Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11(6), 2002.

[114] J. L. Starck, E. J. Candés, and D. L. Donoho. Gray and color image contrast enhancement by the curvelet transform. *IEEE Transactions on Image Processing*, 12(6), 2003.

[115] O. Le Calvé. http ://isitv.univ-tln.fr/˜lecalve/oceano/fiches/fiche3e.htm.

[116] T. Y. Fissenko, V. T. Fissenko, K. S. Rogatchev, and G. A. Sushchev. An interpolation method for color image reconstruction in deep underwater observation. *Proceedings of DS Rozhdestvensky Optical Society*, 2003.

[117] F. Gasparini and R. Schettini. Color balancing of digital photos using simple image statistics. *Pattern recognition*, 2003.

[118] A. Rizzi, C. Gatta, and D. Marini. From retinex to ace : Issues in digital images unsupervised color equalization. *Journal of Electronic Imaging*, 13, 2004.

[119] S. Bazeille, I. Quidu, L. Jaulin, and J.-P Malkasse. Une méthode de pré-traitement automatique pour le débruitage des images sous marines. *Traitement du signal, numéro spécial Caractérisation du milieu marin*, 2006.

[120] A. Arnold-Bos, J. P. Malkasse, and G. Kervern. A preprocessing framework for automatic underwater images denoising. *European Conference on Propagation and Systems*, 2005.

[121] W. K. Pratt. *Digital Image Processing.* John Wiley & Sons, 1991.

[122] M. Kunt. *Reconnaissance des formes et analyse de scènes.* Presses Polytechniques et Universitaires Romandes, 2000.

[123] Z. Qi and J. R. Cooperstock. Automated change detection in an undersea environment using a statistical background model. *IEEE OCEANS*, 2007.

[124] S. E. Umbaugh. *Computer imaging.* CRC Press, 2002.

[125] J. C. Russ. *Image processing handbook (The).* CRC Press, 2002.

[126] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 8(6), 1986.

[127] R. Deriche. Using canny's criteria to derive a recursively implemented optimal edge detector. *Internationnal Journal of Computer Vision*, 1(2), 1987.

[128] S. Helan, J.-C Burie, T. Bouwmans, and S. Bazeille. Object detection in underwater image. *Sea Tech Week Caractérisation du milieu marin*, 2006.

[129] L. Fleck. Some defects in finite-difference edge finders. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 14(3), 1992.

[130] S. Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics, and Image Processing*, 33(1), 1986.

[131] H. C. Lee and D. R. Cok. Detecting boundaries in a vector field. *IEEE Transactions on Signal Processing*, 39(5), 1991.

[132] G. Sapiro. Color snakes. *Computer Vision and Image Understanding*, 1997.

[133] J. Serra. *Image Analysis and Mathematical Morphology Vol.I.* Academic Press, 1982.

[134] J. Serra. *Image Analysis and Mathématical Morphology Vol.II Theoretical Advances.* Academic Press, 1988.

[135] C. Grigorescu, N. Petkov, and M. Westenbreg. Contour and boundary detection by surround suppression of textures edges. *Image and Vision Computing*, 2004.

[136] J. Malik, S. Belongie, J. Shi, and T. Leug. Textons, contours and region : cue integration in image segmentation. *International Conference on Computer Vision (ICCV)*, 1999.

[137] R. O. Duda and P. E. Hart. Use of the hough transform to detect lines and curves in pictures. *Comm. ACM*, 15, 1972.

[138] C. Kimme, D. H Ballard, and J. Sklansky. Finding circles by an array accumulators. *Comm. ACM*, 1975.

[139] H. K. Yuen, J. Illingworth, and J. V. Kittler. Detecting partially occluded ellipses using the hough transform. *Image and Vision Computing*, 7(1) :31–37, 1989.

[140] S. Tsuji and F. Matsumoto. Detection of ellipses by a modified hough transformation. *IEEE Transactions on Computers*, 27(8) :777–781, 1979.

[141] H. Kalvianen, P. Hirvonen, L. Xu, and O. Oja. Comparisons of probabilistic and non probabilistic hough transforms. *European Conference on Computer Vision*, 1994.

[142] N. Kiryati, Y. Eldar, and A. Bruckstein. A probabilistic hough transform. *Pattern recognition*, 24(4) :303–316, 1991.

[143] L. Xu and E. Oja. Randomized hough transform (rht) basic mechanisms, algorithms, and computational complexities. *Computer Vision, Graphics, and Image Processing*, 57(2) :131–154, 1993.

[144] L. S. Davis. Hierarchical generalized hough transforms and line segment based generalized hough transforms. *Pattern recognition*, 15(4) :277–285, 1982.

[145] A. S. Agado and M. S. Nixon. A new hough transform mapping for ellipse detection. *Research Journal Image, Speech and Intelligent Systems*, 1995.

[146] Y. Xie and Q. Ji. A new efficient ellipse detection method. *International Conference on Pattern Recognition (ICPR)*, 2002.

[147] V. Kravtchenko. *Tracking color objects in real time.* PhD thesis, University of British Columbia, 1999.

[148] A. Lehuger, P. Lechat, N. Laurent, and P. Pérez. Maximisation du pouvoir discriminant des espaces colorimétriques pour le suivi d'objets. *Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle*, 2006.

[149] E. Wharton, K. Panetta, and S. Agaian. Human visual system based multi-histogram equalization for non uniform illumination and shadow correction. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2007.

[150] M. S. Drew, J. Wei, and Z-N. Li. Illumination-invariant color object recognition via compressed chromaticity histograms of normalized images. *International Conference on Computer Vision (ICCV)*, 1998.

[151] B. V. Funt, K. Barnard, and L. Martin. Is color constancy good enough ? *5th European Conference on Computer Vision*, 1998.

[152] S. Wesolkowski. Shading and highlight invariant color image segmentation. *Proceedings of SPIE*, 2000.

[153] J. Angulo and B. Marcotegui. Sur l'influence des conditions d'éclairage dans la segmentation morphologique couleur par lpe. *COmpression et REprésentation des Signaux Audiovisuels*, 2005.

[154] B. Bascle, O. Bernier, and V. Lemaire. Illumination-invariant color image correction. *Lecture notes in Computer Science*, 2006.

[155] S. Q. Duntley. *Light in the sea.* J. Opt. Soc. Am., 1963.

[156] P. Zingaretti and S. M. Zanoli. Robust real-time detection of an underwater pipeline. *Engineering Applications of Artificial Intelligence*, 1998.

[157] J.Ronald, V. Zaneveld, and W. S. Pegau. Robust underwater visibility parameter. *Optics Express*, 11(23), 2003.

[158] Y. Rzhanov, A. Mamaenko, and M. Yoklavich. Uvsd : Software for detection of color underwater features. *IEEE OCEANS*, 2005.

[159] D. Semani, M. Chambah, and P. Courtellemont. Processing of underwater color images applied to live aquarium videos. *International Journal of Robotics and Automation*, 2005.

[160] A. Yamashita, M. Fujii, and T. Kaneko. Color registration of underwater of images for underwater sensing with consideration of light attenuation. *IEEE International Conference on Robotics and Automation (ICRA)*, 2007.

[161] L. A. Torres-Mendez and G. Dudek. Color correction of underwater images for aquatic robot inspection. *Lectures notes in computer science*, 2005.

[162] L. A. Torres-Mendez and G. Dudek. A statistical learning-based method for color correction of underwater images. *Computer Science, Advances in Artificial Intelligence Theory*, 17, 2005.

[163] N. G. Jerlov. *Marine optics*. Elsevier oceanography series, 1976.

[164] R. H. Stavn. Lambert-beer law in ocean waters. *Applied Optics,*, 27, 1988.

[165] H. R. Gordon. Can the lambert-beer law be applied to the diffuse attenuation coefficient of ocean water. *Limnology and Oceanography*, 34(8), 1989.

[166] J. Ahlen, E. Bengtsson, and T. Lindell. Color correction of underwater images based on estimation of diffuse attenuation coefficients. *PICS Conference*, 2003.

[167] S. Bazeille, I. Quidu, and L. Jaulin. Identification of underwater man-made object using colour. *Proceedings of the Institute of Acoustics*, 2007.

[168] D. R. Mishra, S. Narumalani, D. Rundquist, and M. Lawson. Characterizing the vertical diffuse attenuation coefficient for downwelling irradiance in coastal waters. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(1), 2005.

[169] T. Carron. *Segmentation d'images couleur dans la base Teinte-Luminance-Saturation, approche numérique et symbolique*. PhD thesis, Université de Savoie, 1995.

[170] S. Sural, G. Qian, and S. Pramanik. Segmentation and histogram generation using the hsv color space for image retrieval. *IEEE International Conference on Image Processing*, 2002.

[171] S. Lefevre and N. Vincent. Apport de l'espace teinte-saturation-luminance pour la segmentation spatiale et temporelle. *Revue Traitement du Signal*, 23(1), 2006.

[172] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter. *Applied interval analysis*. Springer, 2001.

[173] B. Telle. *Méthode ensembliste pour une reconstruction tridimensionnelle garantie par stéréovision*. PhD thesis, Université de Montpellier II, 2003.

[174] R. O. Duda, P. E. Hart, and D. G. Strok. *Pattern classification, second edition*. John Wiley & Sons, 2001.

[175] L. Issen. Using edge statistics for object recognition. *Computer Sciences, Spring*, 2006.

[176] C. M. Bishop. *Neural network for pattern recognition*. Oxford University Press, 1995.

[177] M. A. Fischler and R. C. Bolles. Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 1981.

[178] S. Bazeille, I. Quidu, L. Jaulin, and J.-P Malkasse. Automatic underwater image pre-preprocessing. *Sea Tech Week Caractérisation du milieu marin*, 2006.

Vision sous-marine monoculaire pour la reconnaissance d'objets.

Mots clés : Vision, Sous-marin, Reconnaissance, Robot autonome.

Résumé : Dans le contexte sous-marin, et à l'inverse du capteur sonar qui reste le plus employé à grande distance pour la détection et la classification, la caméra vidéo est efficace à faible portée lors des phases d'approche, de reconnaissance d'objets et d'intervention. Elle dispose en effet d'atouts notables comme une haute résolution, une facilité d'interprétation ou encore un faible coût. Aujourd'hui presque tous les véhicules sous-marins scientifiques, industriels ou militaires en sont équipés. Ils sont actuellement plutôt télé-opérés par un opérateur et on y trouve associés des traitements automatiques que très rarement. Ces traitements automatiques sont pourtant des technologies essentielles pour les développements émergents des robots autonomes sous-marins très prisés aujourd'hui dans un contexte d'expansion des marchés liés à la sécurité et à l'exploitation des ressources maritimes. Ces travaux de thèse visent à apporter les innovations nécessaires, et à promouvoir l'emploi du capteur vidéo dans le domaine sous-marin. L'étude proposée concerne le développement des traitements automatiques de reconnaissance d'objets en vidéo sous-marine, avec une attention particulière apportée aux objets manufacturés. Les scènes sous-marines observées sont classiquement plus simples et plus limitées en profondeur d'observation que les scènes urbaines ou que l'intérieur d'un bâtiment. Toutefois, ce contexte présente des difficultés spécifiques telles que les variations d'éclairage et la turbidité de l'eau qui limitent la visibilité et dégradent fortement les images. Ceci a pour conséquence de rendre les traitements difficiles et nécessite donc la création de nouveaux algorithmes de vision robotique.

Monocular underwater vision for object recognition.

Keywords : Computer vision, Underwater, Recognition, Autonomous vehicle.

Abstract : In underwater context, traditional sensing methods like sonar are used at large range for detection and classification of objects. For a few years, the sonar sensor has been complemented by a vision sensor more efficient at short range for the approach, the object recognition and the intervention phases. Indeed, the camera has noticeable advantages like for example high resolution, simple interpretation or low cost. Today, it equips nearly any scientific, industrial or military underwater vehicles. Currently, underwater vehicles are rather distance controlled by a manipulator, and automatic processing is really uncommon. However, automatic processing is an essential technology for the AUV development. These kinds of vehicles are very popular today with the expansion of the market related to the security and the exploitation of the marine resources. This thesis aims at supplying the required advances and promoting the use of video technologies. The proposed study addresses the problem of automatic processing for underwater objects recognition, with a particular focus on man-made objects. Observed underwater scenes are traditionally simpler but more limited in depth than urban scenes or inside building scenes. Nevertheless, this context presents very specific difficulties like lighting variations and water turbidity which limit visibility and degrade considerably images. As a consequence processing is difficult and require the development of new robotic vision algorithms.