ACADÉMIE DE MONTPELLIER

UNIVERSITÉ **M**ONTPELLIER **II**

Sciences et **T**echniques du **L**anguedoc

THÈSE

présentée au Laboratoire d'Informatique de Robotique
et de Microélectronique de Montpellier pour
obtenir le diplôme de doctorat

Spécialité	:	**Génie Informatique, Automatique et Traitement du Signal**
Formation Doctorale	:	**Systèmes Automatiques et Microélectroniques**
École Doctorale	:	**Information, Structures, Systèmes**

Fusion multimodale pour la cartographie sous-marine

par

Arnaud MÉLINE

Soutenue le 31 Janvier 2013, devant le jury composé de :

Rapporteurs

M. Pierre DRAP, Chargé de recherche ..... LSIS, UMR CNRS 7296

M. Luc JAULIN, professeur ..... Lab-STICC,ENSTA Bretagne

Président du jury

M. René ZAPATA, professeur ..... LIRMM, Université Montpellier II

Examinateurs

M. Christian BARAT, maitre de conférence ..... I3S, UNSA - CNRS

Directeur de thèse

M. Bruno JOUVENCEL, professeur ..... LIRMM, Université Montpellier II

Co-Directeur de thèse

M. Jean TRIBOULET, maitre de conférence ..... LIRMM, Université de Nîmes

Remerciements

Je souhaite remercier tous les membres du Jury et les encadrants de cette thèse :
- M. René Zapata pour avoir accepté la présidence du jury ;
- M. Pierre Drap et M. Luc Jaulin pour leur temps passé à l'analyse et à la critique de ce manuscrit en tant que rapporteur ainsi que pour les différentes conversations qui ont permis d'enrichir le manuscrit ;
- M. Christian Barat d'avoir pris de son temps pour examiner ma thèse ;
- M. Bruno Jouvencel, directeur de thèse, de m'avoir fait confiance pour la réalisation de cette thèse ;
- M. Jean Triboulet, co-directeur de thèse, pour tout ce qu'il m'a apporté durant ces années passées à travailler ensemble. Son expérience et ses conseils m'ont permis de me motiver dans les moments où le travail n'avançait pas. Je le remercie également pour la grande aide apportée à la réalisation des expérimentations ainsi que sa grande présence et son suivi régulier des travaux de recherche.

Je tiens à remercier les membres de mon comité de suivi de thèse : René Zapata, Lionel Lapierre et David Andreu qui m'ont permis d'avoir un point de vue extérieur au projet de recherche ainsi qu'à leurs conseils et encouragements pour la fin de cette thèse.

Je remercie également Luc Long qui nous à permis d'accéder et de pouvoir réaliser des expérimentations sur des chantiers de fouilles archéologiques sous marins.

Un grand merci à tous les doctorants du département robotique pour toutes ces pauses café et ces bons moments passés ensemble. Merci à Nicolas (qui me suit depuis l'IUT !!! ou alors c'est le contraire, enfin c'est l'histoire de la poule et de l'œuf) pour ces nuits passées à rédiger les articles pour respecter les deadline. Divine ta bonne humeur permanente me surprendra toujours autant, tout comme tes chansons et chants à tout heure. Johann, maintenant que tu est passé à Narbonne tu sais pourquoi on ne craint pas le vent. Je n'oublie pas David, Antoine et plus tard Joris qui ont supporté mes questions "bêtes" de programmation. Et bien sûr Florent mon co-bureau avec qui j'ai passé de très bons moments. Sans oublier les derniers arrivés Adrien, Andrea, Fabien qui ne savent pas encore dans quoi ils se sont lancés.

Je remercie aussi les étudiants en projet et/ou en stage ayant travaillé sur ce projet.

Je remercie également tous mes amis bien qu'ils n'aient toujours pas compris exactement ce que je fais exactement ("Alors ça y est, tu as mis tes poissons en 3D sur ta carte ?"), je ne préfère pas mettre de noms au risque d'oublier quelqu'un, mais un grand merci à vous tous de m'avoir permis de m'évader et de décompresser dans les moments difficiles.

Pour finir je remercie particulièrement mes parents ainsi que ma famille pour leur soutien infaillible durant toutes ces années.

Ce manuscrit n'aurait peut être jamais vu le jour sans le soutien durant les derniers mois d'écriture de Claire, que j'aime de tout mon cœur.

Merci à toutes les personnes que j'ai oublié et je m'en excuse par avance.

Table des matières

Remerciements i

Table des matières iii

I Contexte et problématique 1

1 Introduction 3

2 Conditions expérimentales 7

2.1 Détails des expérimentations 7

2.1.1 Expérimentations en mer 8

2.1.2 Expérimentations en fleuve 8

2.1.3 Expérimentations en étang salé 10

2.1.4 Expérimentations en bassin 11

2.2 La vision et le milieu sous-marin 11

II Vision et calibrage sous-marin 15

3 Principes fondamentaux en vision par ordinateur 17

3.1 Modélisation d'un système de vision 17

3.1.1 Transformation scène - caméra 18
3.1.2 Transformation caméra - image 19
3.1.3 Adaptation du modèle au milieu sous-marin 20
3.2 Géométrie épipolaire et stéréovision 21

4 Calibrage du système de vision **25**
4.1 Méthodes de calibrage 25
4.1.1 Méthodes existantes 26
4.1.2 Discussions et choix de la méthode utilisée 29
4.2 Expérimentations et résultats 31

III Reconstruction 3D locale **37**

5 La reconstruction 3D **41**
5.1 Etat de l'art 41
5.1.1 Reconstruction à partir de la texture 43
5.1.2 Reconstruction à partir de jeux de lumière 44
5.1.3 Reconstruction à partir du mouvement 45
5.1.4 Reconstruction par géométrie épipolaire 47
5.2 Description de notre approche 48

6 Extraction et mise en correspondance de points caractéristiques **51**
6.1 État de l'art 51
6.1.1 Détection des points caractéristiques 51
6.1.2 Mise en correspondance 55
6.1.3 Correction des appariements par géométrie épipolaire 59
6.2 Choix des méthodes utilisées 61
6.2.1 Harris & corrélation 62
6.2.2 SIFT 63
6.3 Expérimentations et résultats 64
6.4 Limitation de SIFT en environnement sous-marin 68
6.4.1 Filtrage des images sous-marines 68
6.4.2 Expérimentations dans différents milieux 69
6.4.3 Amélioration de la robustesse de SIFT 75

7 Création du modèle 3D **77**
7.1 Projection des points dans l'espace 3D 77
7.1.1 Reconstruction à partir d'images stéréoscopiques 78
7.1.2 Reconstruction à partir d'une vidéo 80
7.2 Rendu du modèle 3D 82
7.2.1 Triangulation 82
7.2.2 Plaquage de la texture 84

IV Fusion des différentes cartes **89**

8 Création de la carte acoustique 3D **91**
8.1 Matériels acoustiques sous-marins 91
8.2 Expérimentations et résultats 95

9 Fusion de la carte acoustique globale et du modèle vidéo local **101**
9.1 État de l'art 101
9.1.1 Fusion de données acoustiques et vidéo 101
9.1.2 Recalage de modèles géométriques 3D 104
9.2 Discussions et présentation de notre approche 106
9.3 Expérimentations et résultats 108

V Conclusions **115**

10 Conclusions et Perspectives **117**

Bibliographie **121**

Liste des Publications **131**

Table des figures **133**

Liste des tableaux **137**

Première partie

Contexte et problématique

CHAPITRE

1 Introduction

Depuis toujours, l'Homme explore. Que ce soit sur terre, dans l'air, au-delà de notre planète ou dans l'eau, il cherche continuellement à en apprendre plus sur ce qui l'entoure et veut étudier et comprendre ces nouveaux environnements. Depuis des années, l'eau qui représente plus de 72% de la surface du globe, nous fournit une grande quantité d'informations dans divers domaines. Encore aujourd'hui les biologistes découvrent régulièrement de nouvelles espèces ; des roboticiens s'inspirent des espèces sous-marines pour développer des structures bio-inspirées ; les archéologues trouvent des vestiges du passé nous permettant d'en apprendre plus sur notre histoire.

L'environnement sous-marin est un milieu très vaste et malgré tout ce que nous avons pu trouver, il reste encore très peu exploré. C'est pour cela que, depuis de nombreuses années, des inventeurs et des chercheurs créent de nouveaux moyens d'appréhender les éléments qui les entourent et fabriquent des véhicules, des robots mobiles autonomes, etc.

L'archéologie sous-marine se heurte à de nombreux problèmes comme l'autonomie des plongeurs ou la longueur du temps d'exploration d'une grande zone de recherche. En effet, à l'heure actuelle, beaucoup de travaux de fouilles sont réalisés par des plongeurs et sont limités par les capacités du corps humain à rester sous l'eau. Dans ce contexte d'exploration sous-marine, nous venons utiliser les connaissances du monde de la robotique afin de réduire ces problèmes de temps de prospection.

La robotique a développé un champ spécifique pour le milieu sous-marin. Les contraintes fortes, liées à son environnement, augmentent les difficultés des problèmes pouvant être rencontrés en milieu terrestre. Le travail de ce mémoire se concentre sur la

cartographie de sites archéologiques sous-marins et sur le potentiel d'une approche par fusion multimodale.

L'approche proposée ici utilise un sonar afin d'obtenir une carte de la zone à explorer. Cette carte de faible résolution ne permet pas d'identifier précisément le type d'objets observés mais seulement leur présence. L'utilisation d'un système de vision par caméra permet de solutionner ce problème. En effet, ce dernier possède l'avantage de fournir des modèles 3D texturés des objets d'intérêts. En pratique, un site archéologique est cartographié dans son ensemble par un sonar à balayage. La zone de fouilles ainsi couverte est à minima de plusieurs dizaines de mètres. L'observation de ce profil révèle ensuite des points d'intérêts sur lesquels une information détaillée s'impose. Un système de vision stéréoscopique embarqué sur un véhicule sous-marin autonome (AUV) ou porté par un plongeur permet d'obtenir cette information.

La problématique consiste d'une part à mettre en œuvre une méthode de reconstruction 3D robuste (à partir des données stéréoscopiques) et d'autre part à fusionner cette information avec la carte sonar.

La problématique de reconstruction 3D fait référence à des méthodes de calibrage robustes tenant compte des fortes contraintes liées au milieu (luminosité, turbidité, salinité, température). Elle se complète par des méthodes de recherche de points d'intérêts et des méthodes d'appariements robustes liées à ces mêmes contraintes qu'il faut identifier et quantifier.

La problématique de la fusion multimodale provient essentiellement de l'hétérogénéité des capteurs et particulièrement de leurs résolutions, la résolution du sonar dépendant du choix de la zone de mesure retenue. Le système stéréoscopique permet une résolution élevée en milieu terrestre, or les contraintes liées au milieu font que cette résolution est altérée. Le choix de la méthode de fusion / recalage, en se basant sur des points, doit s'appuyer davantage sur des critères surfaciques pour être plus robuste.

Le problème ainsi posé, le manuscrit permettant d'apporter des éléments de réponse est composée de cinq parties. La première expose le contexte et les différentes expérimentations réalisées pendant la thèse.

Dans la partie deux, nous présentons certains concepts de base en vision par ordinateur ainsi que le modèle de caméra utilisé pour le milieu sous-marin. La détermination des paramètres de ce modèle est ensuite réalisée à l'aide de plusieurs expérimentations et après une étude des méthodes existantes.

La reconstruction de la carte locale 3D par vision est détaillée dans la partie trois. Nous

commençons par un état de l'art qui nous amène à décrire notre approche. Nous nous attardons ensuite sur la partie concernant l'extraction et la mise en correspondance des points dans les images car cette étape est primordiale pour la qualité du modèle 3D. Nous présentons un état de l'art de ces méthodes pour en sélectionner deux. Nous mettons en avant les limitations introduites par un des bruits créé par le milieu sous-marin. Nous proposons un filtre afin d'identifier quel bruit est responsable de cette perte de performance. Une étape de pré filtrage est ensuite étudiée afin d'améliorer les résultats de cette méthode. Une fois l'extraction et l'appariement de points robustes terminés, nous les projetons dans un espace en trois dimensions pour obtenir un nuage de points. Son rendu est ensuite amélioré en y ajoutant une surface ainsi que la texture correspondante.

La partie quatre sera, dans un premier temps, consacrée aux différents types de sonars nous permettant de créer la carte 3D acoustique, puis une fois cette carte obtenue, une bibliographie sur les différentes méthodes existantes pour la fusion de données acoustiques et vidéo sera exposée. Après une discussion sur celle-ci, nous en proposerons une permettant de réaliser la fusion des cartes par recalage géométrique des modèles. Nous finirons par présenter les résultats obtenus avec cette méthode.

Pour finir, la partie cinq expose un bilan du travail réalisé ainsi que des perspectives futures.

L'aspect innovant de cette thèse porte essentiellement sur l'approche proposée pour réaliser la carte multimodale. En effet, l'utilisation d'une carte globale acoustique 3D permet un relevé rapide de la zone complète à cartographier. Lorsque des détails sont nécessaires sur une ou plusieurs parties de cette carte, une reconstruction 3D à partir d'un système stéréoscopique est réalisée. Le modèle obtenu possède une résolution plus élevée et un rendu visuel amélioré par la texture. Un travail important a été d'étudier et de choisir les méthodes les plus appropriée aux contraintes du milieu et à notre système, parmi les nombreuses méthodes existantes en traitement d'images. Afin de vérifier la robustesse des méthodes sélectionnées aux nombreux bruits sous-marins, un filtre modélisant les bruits un à un a été proposé. Cette étude nous a permis d'obtenir des points robustes pour le modèle 3D vidéo. Une autre grande partie du travail effectué a porté sur la mise en œuvre de nombreuses expérimentations afin de tester et de valider les différentes approches proposées.

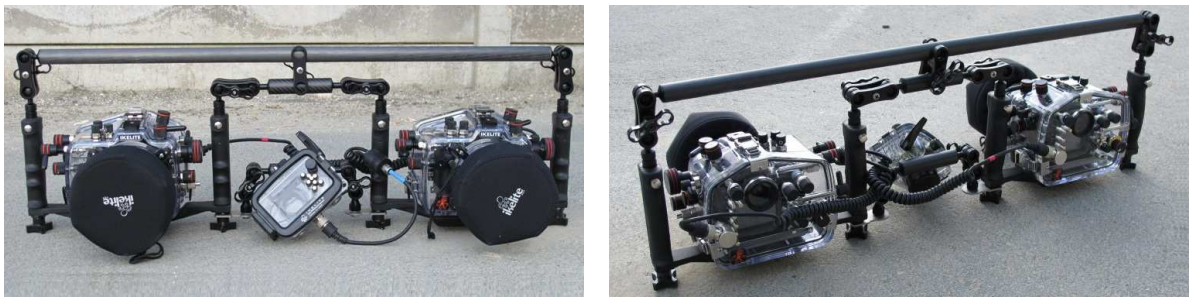
CHAPITRE

2 Conditions expérimentales

2.1 Détails des expérimentations

Nous avons réalisé plusieurs expérimentations en mer, en fleuves, en étang salé, en bassin et hors de l'eau. Le système de vision que nous avons utilisé est composé de deux appareils photos Nikon D7000 embarqués dans des caissons étanches à 40 mètres Ikelite. Le D7000 est équipé d'un capteur à 16.2 millions de pixels avec une grande plage de sensibilité (100 à 25600 ISO). Cette sensibilité très élevée permet d'obtenir de bonnes images avec de mauvaises conditions d'utilisation. Un mode vidéo permet de filmer en haute définition en 1080p à 24 images par seconde. Les deux caissons étanches sont reliés par des barres de carbone pour former la paire stéréoscopique. Une télécommande est utilisée pour permettre la prise de vue synchronisée entre les deux appareils. Cette paire stéréoscopique (fig. 2.1), assemblée par la société plongimage, permet de prendre deux photos d'une même scène dans deux positions différentes.

Lors de toutes nos expérimentations, nous avons immergé un ensemble d'objets de forme et de taille connues tels qu'une statue en forme de poisson, un buste de femme en morceaux et un autre entier, ainsi que deux mires de calibrage. Ces objets ont été numérisés en surface à l'aide d'un scanner 3D Minolta afin d'obtenir leurs modèles 3D. Ils permettent d'avoir une réalité terrain afin de vérifier la validité de notre méthode de reconstruction 3D.



(a) Face (b) Dos

FIGURE 2.1 : Photo du système stéréoscopique.

2.1.1 Expérimentations en mer

Afin de créer une bibliothèque d'images sous-marines, nous avons réalisé les premières expérimentations en mer sur le site des Aresquiers proche de Sète. A chaque plongée, nous avons immergé les objets décrits précédemment et deux mires de calibrage. Deux plongées avec des conditions sous-marines différentes ont été réalisées sur ce site à 6-7 mètres de profondeur. Lors de la première (2.2(a)), la visibilité était de 2-3 mètres et la turbidité de l'eau était faible. La température de l'eau était d'environ 18°C et les conditions d'éclairage étaient bonnes. Lors de la deuxième plongée (2.2(b)), la visibilité était d'environ 1 mètre et la turbidité était plus importante. La température de l'eau avoisinait les 14°C et les conditions d'éclairage étaient moins bonnes. Ces expérimentations ont permis de créer une base de données d'images sous-marines obtenues dans différentes conditions.

M. Luc Long [Long *et al.*, 2009] nous a permis de réaliser des prises de vue d'objets à intérêt archéologique lors d'une campagne en mer. Le site sur lequel nous avons pu travailler est proche des Saintes-Maries-de-la-Mer. Les objets observés se trouvaient à une vingtaine de mètres de profondeur. La température de l'eau était proche de 18°C et les conditions météorologiques étaient bonnes. Cela nous a permis d'obtenir des images sous-marines avec peu de turbidité et avec un éclairage correct. La figure 2.3 représente deux photos stéréoscopiques d'une amphore se trouvant sur ce site.

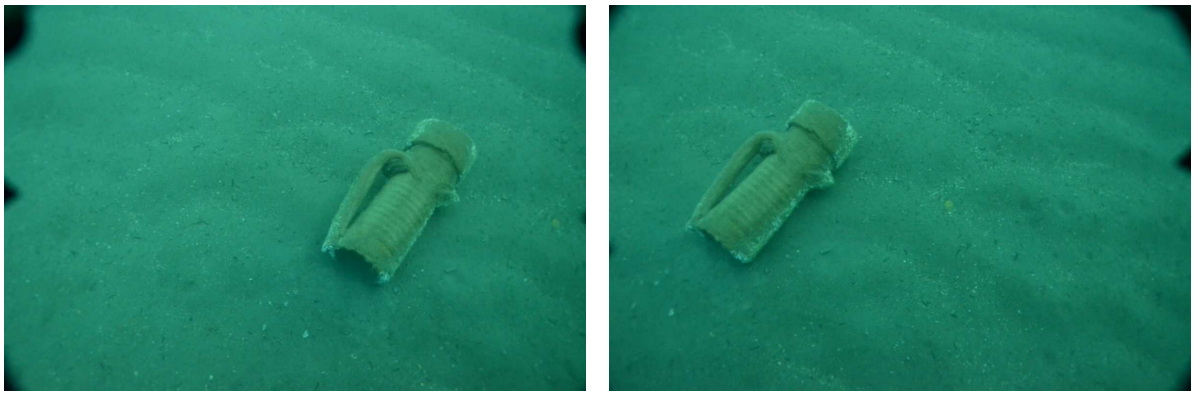
2.1.2 Expérimentations en fleuve

D'autres expérimentations ont pu être réalisées sur le Rhône avec la même équipe d'archéologues sous-marins. L'équipe réalisait la prospection de péniches de l'époque ro-



(a) Première expérimentation (b) Deuxième expérimentation

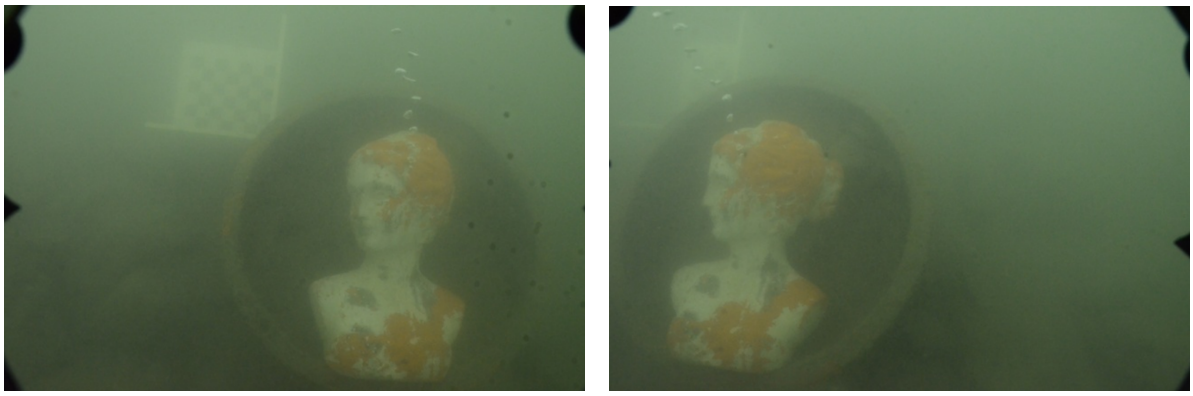
FIGURE 2.2 : Photos de la statue de poisson lors d'expérimentation en mer aux Aresquiers.



(a) Cliché gauche (b) Cliché droit

FIGURE 2.3 : Photos stéréoscopiques d'une amphore lors d'expérimentations en mer aux Saintes-Maries-de-la-mer.

maine et de leurs cargaisons en Arles. Nous n'avons pas pu prendre de photos des objets retrouvés sur ce site mais nous avons immergé la statue du buste de femme. Les photos ont été prises aux environs de sept mètres de profondeur avec des conditions de prises de vues très difficiles, la visibilité n'était que de 40 cm. La turbidité était très importante et les conditions d'éclairage étaient mauvaises. La figure 2.4 permet de visualiser le bruit présent dans les images lors de ces manipulations.



(a) Cliché gauche (b) Cliché droit

FIGURE 2.4 : Photos stéréoscopiques du buste lors d'expérimentations dans le Rhone.

2.1.3 Expérimentations en étang salé

Des expérimentations dans l'étang de Thau ont aussi été réalisées. L'équipe d'archéologues sous-marins nous a prêté une amphore afin que nous puissions avoir un objet archéologique de dimensions connues et faire des relevés de celle-ci. Nous avons immergé cette amphore à environ sept mètres de profondeur. Les conditions de prises de vue n'étaient pas très bonnes. La turbidité était importante mais les conditions de luminosité étaient correctes. La figure 2.5 donne une idée des conditions de prises de vue.

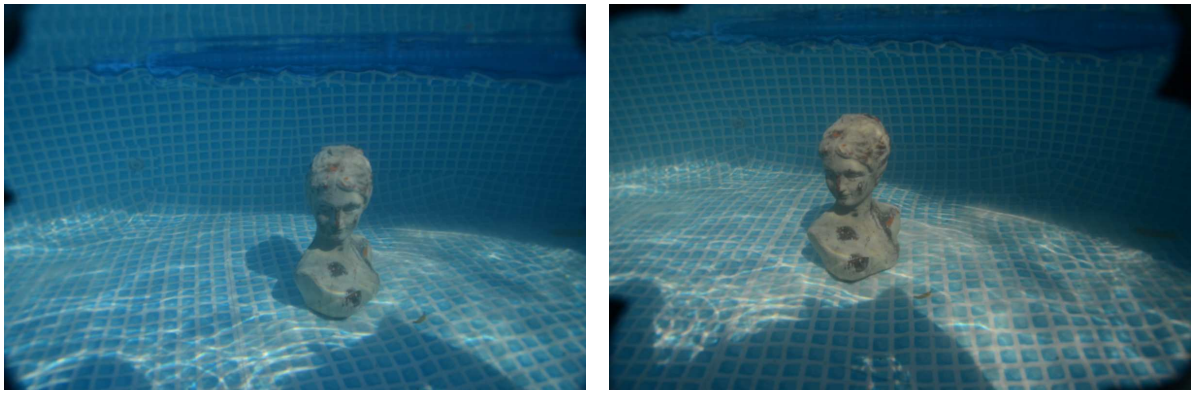


(a) Cliché gauche (b) Cliché droit

FIGURE 2.5 : Photos stéréoscopiques d'une Amphore lors d'expérimentations dans l'étang de Thau.

2.1.4 Expérimentations en bassin

Pour pouvoir tester nos méthodes dans un milieu sous-marin peu perturbé, nous avons réalisé des expérimentations en bassin. La profondeur du bassin était de 2m, la température proche de 20°C et la turbidité était quasi nulle. Les conditions d'éclairage étaient bonnes.



(a) Cliché gauche (b) Cliché droit

FIGURE 2.6 : Photos stéréoscopiques du buste lors d'expérimentations en bassin.

2.2 La vision et le milieu sous-marin

L'utilisation d'images sous-marines est difficile car l'eau introduit d'importantes contraintes. En effet, la qualité des images est fortement dégradée par les effets variables qu'introduit l'eau sur la propagation des signaux. Les principales causes de cette dégradation sont dues à la présence de particules en suspension, aux problèmes d'éclairage ainsi qu'à l'absorption de l'énergie lumineuse.

L'indice optique de l'eau. Tout milieu ou matière est caractérisé par un indice de réfraction. La réfraction représente la déviation que subit le rayon lumineux en passant d'un milieu à un autre. L'indice de réfraction optique de l'eau est un paramètre variable. Il varie en fonction de la température, de la salinité, de la pression (donc de la profondeur) de l'eau ainsi que de la longueur d'ondes du rayon lumineux.

La turbidité et les éléments en suspension. L'eau présente une quantité variable de particules en suspension. Ces particules provoquent des phénomènes aléatoires de réfraction et de réflexion des rayons lumineux lorsqu'elles ont une taille supérieure à $1\mu m$ [Meeteen, 1999]. Chaque fois qu'un rayon lumineux est réfléchi ou réfracté, il perd en intensité. Cela provoque une diminution de l'énergie lumineuse présente sous l'eau.

Intensité lumineuse. Comme nous venons de le voir, la turbidité entraîne une diminution de la luminosité sous l'eau. Son influence n'est pas très importante comparée à l'absorption de l'eau. L'eau limite le spectre de fréquence de la lumière transmise. Elle absorbe progressivement les couleurs de basse fréquence (rouge, jaune) puis celles de haute fréquence (vert, bleu) en fonction de la distance séparant l'objet de la source lumineuse (fig. 2.7). Par exemple, si aucune source lumineuse artificielle est ajoutée, la couleur verte disparaît à 99% à une profondeur de 40 mètres.

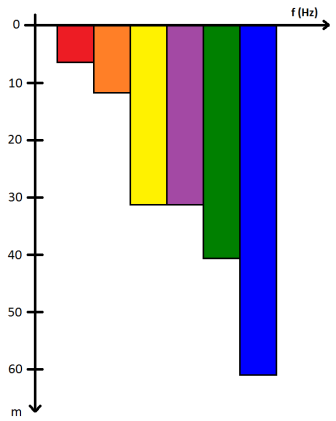


FIGURE 2.7 : Disparition des couleurs en fonction de la distance approximative à partir de la source lumineuse.

Éclairage. L'eau absorbant l'intensité lumineuse, il faut généralement ajouter un éclairage à la scène pour observer toutes les couleurs ou simplement pour voir à de grandes profondeurs. La puissance de la source lumineuse doit être choisie en fonction du milieu. En effet, un sol ou un objet blanc va plus réfléchir la lumière qu'un fond vaseux foncé.

Lorsque la turbidité est importante, l'éclairage doit être utilisé avec précaution car il crée un effet de brouillard sous l'eau.

Enceinte étanche. Les enceintes étanches permettent d'isoler les appareils photos de l'eau en les conservant dans de l'air. De ce fait, les rayons lumineux traversent différents dioptres (eau, hublot, air). La réfraction de ces rayons à travers ces dioptres introduit des erreurs de réfraction. Ces erreurs de réfraction sont proches de celles produites par la distorsion radiale, même si la distorsion et la réfraction sont des phénomènes physiques différents. La distorsion radiale corrige en grande partie la perturbation de la réfraction [Kwon et Lindley, 2000]. Cette approche est fortement dépendante du changement de dioptre eau / hublot. Les enceintes étanches peuvent être réalisées avec différents types de hublots ; plans ou hémisphériques. Les hublots hémisphériques permettent de minimiser les erreurs de réfraction. Cependant, ils introduisent des déformations optiques (aberrations et distorsions), mais ces dernières sont corrigeables à 99% si les hublots sont équipés de correcteurs optiques. Dans notre travail, nous avons utilisé ce type des hublots afin d'intégrer les erreurs de réfraction directement dans la distorsion radiale.

Environnement naturel. Le milieu sous-marin est un environnement qui ne peut être contrôlé. En effet, il est impossible d'éviter aux particules en suspension (voire à un poisson) de passer devant la caméra lors d'une prise de vue. La quantité de particules en suspension est aléatoire et dépend principalement des courants sous-marins. Il faut donc prévoir une grande flexibilité pour travailler dans ce milieu.

Maintenant que les différents types d'images avec lesquelles nous allons travailler ainsi que les différents bruits présents dans le milieu sous-marins ont été présentés, nous allons passer aux principes fondamentaux permettant d'appréhender et de modéliser ces images.

Deuxième partie

Vision et calibrage sous-marin

CHAPITRE

3 Principes fondamentaux en vision par ordinateur

L'environnement dans lequel l'Humain évolue est en trois dimensions, or lorsque nous voyons des objets, ceux-ci créent une image sur notre rétine en deux dimensions. La dimension perdue lors du processus de vision est la profondeur. Elle est reconstruite par notre cerveau principalement grâce à nos deux yeux qui composent un système stéréoscopique. En vision par ordinateur le principe est le même, les objets sont projetés en deux dimensions sur les plans images des appareils photos. Cette section va présenter les notions de bases en géométrie permettant de définir le modèle géométrique d'une caméra. Ce modèle étant composé de nombreux paramètres, il sera possible de les estimer grâce à des méthodes de calibrage. Il faut au minimum deux images de la même scène observée à des positions différentes pour pouvoir estimer la profondeur, comme pour le fonctionnement humain. La relation géométrique entre ces deux images (stéréoscopie) est définie par la géométrie épipolaire.

3.1 Modélisation d'un système de vision

Dans le cadre de ce travail, nous avons considéré le modèle sténopé ou aussi dit trou d'épingle (pinhole en anglais). Il est constitué de quatre repères principaux (fig. 3.1) : le repère monde R_w, le repère caméra R_c placé au niveau du centre optique, le repère rétinien R_r placé au niveau du plan de l'image et le repère image R_i en pixel. Le point M est considéré comme étant dans le repère monde en 3D, et le point m son projeté dans l'image.

Les transformations pour obtenir ce point m peuvent être décrites en trois étapes qui vont passer par les quatre repères. La première T_1 représente le passage du repère R_w vers le repère R_c. T_2 permet de faire la liaison entre le repère R_c et le repère R_r et la dernière T_3 permet d'obtenir les coordonnées du point m dans le repère image (R_i).

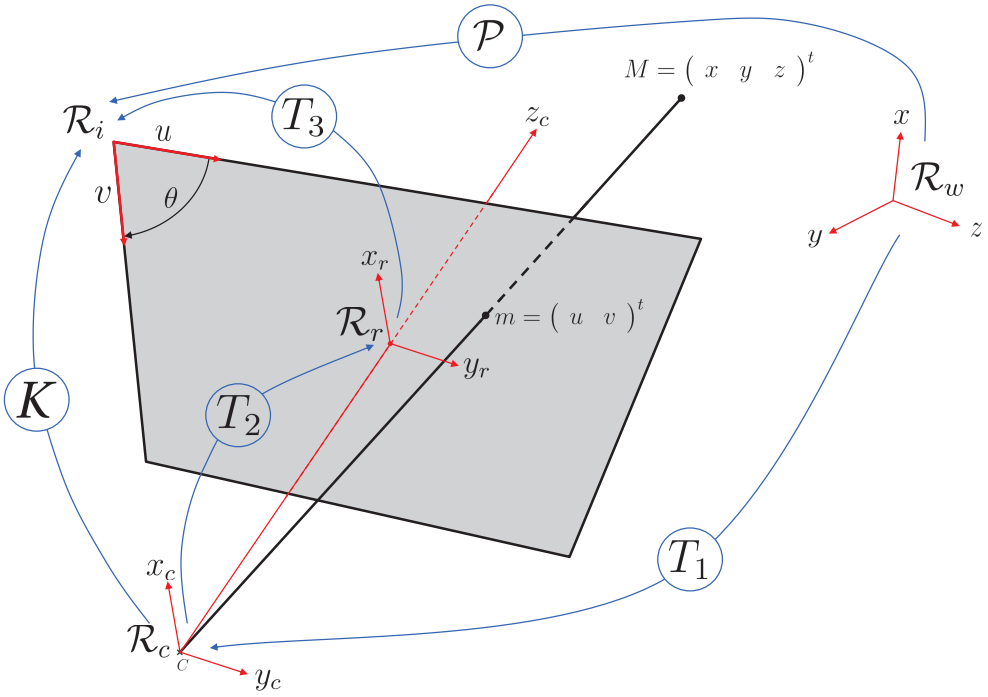


FIGURE 3.1 : Modèle de caméra et changement de repères

3.1.1 Transformation scène - caméra

La transformation scène-caméra est caractérisée par une transformation rigide qui est composée d'une rotation euclidienne et une translation. Pour simplifier le processus de modélisation, nous utilisons la notation matricielle des coordonnées homogènes. Les translations, rotations et transformées perspectives sont alors représentées par des pro-

duits de matrices. Ainsi la transformation entre le repère monde et le repère caméra pour un point $M = (x, y, z, 1)^t$ s'écrit :

$$M_c = T_1 M \tag{3.1}$$

où le point M_c est exprimé dans le repère caméra (R_c) et la matrice T_1 est de la forme :

$$T_1 = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \tag{3.2}$$

avec R matrice de rotation euclidienne de dimension $(3x3)$ et t vecteur de translation de dimension $(3x1)$. Ces paramètres sont appelés *paramètres extrinsèques*.

3.1.2 Transformation caméra - image

La transformation caméra-image est composée de deux étapes fondamentales. La transformation T_2 projette le point M_c sur le plan image, dans le repère rétinien.

$$M_r = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} M_c \tag{3.3}$$

Puis nous effectuons le changement de repère avec T_3, pour passer du repère rétinien aux coordonnées du point en pixel.

$$m = \begin{pmatrix} k_u & r & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} M_r \tag{3.4}$$

Dans ces équations, f représente la distance focale de la caméra, k_u et k_v représentent la taille des pixels puis u_0 et v_0 donnent la position du repère en pixel par rapport au repère rétinien. Le paramètre r aussi appelé " skew " représente le défaut de l'angle qui peut apparaître entre les axes u et v. Généralement ce terme est nul car ce type d'erreur est négligeable pour les capteurs CMOS (Complementary Metal Oxide Semiconductor). Généralement, nous regroupons ces deux transformations en une seule.

$$
\begin{aligned}
K &= \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} k_u & r & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} k_u.f & r & u_0 & 0 \\ 0 & k_v.f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\
&= \begin{pmatrix} \alpha_u & r & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}
\end{aligned} \tag{3.5}
$$

et donc

$$m = KM_c \tag{3.6}$$

Les paramètres α_u, α_u, u_0, v_0 et r représentent les *paramètres intrinsèques* de la caméra. Avec cette dernière forme, nous nous apercevons que la transformation projective de la caméra peut être déterminée sans connaître la valeur de la distance focale. Nous pouvons donc déterminer la projection du point M dans le repère image à l'aide des paramètres extrinsèques et intrinsèques.

$$
\begin{aligned}
m &= \begin{pmatrix} \alpha_u & r & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} M \\
&= K.T_1.M \\
&= P.M
\end{aligned} \tag{3.7}
$$

3.1.3 Adaptation du modèle au milieu sous-marin

Le modèle de caméra définie précédemment est un modèle simplifié et idéal. Il est largement utilisé et suffisant dans la plupart des applications de vision. Or lorsque l'application demande une précision plus importante, il est nécessaire d'employer un modèle plus complexe. Généralement, le modèle sténopé est complété par des termes correspondant à

la déformation de l'image qui est souvent introduite par la conception des optiques utilisés. Cependant, lorsqu'une caméra est introduite dans le milieu sous-marin, des déformations importantes sont ajoutées à l'image. Les hublots sphériques utilisés sur les caissons étanches permettent d'en compenser certaines. La déformation la plus importante est la *distorsion radiale*. Elle entraîne un déplacement des pixels de façon radiale par rapport au modèle linéaire. Le centre de la distorsion de l'optique est différent du centre de distorsion de l'image. Dans notre cas, nous prenons comme hypothèse que les centres de distorsion sont identiques. Cette approximation permet de confondre les deux distorsions et de n'en modéliser qu'une. La distorsion radiale peut être exprimée [Heikkilä et Silvèn, 1997] par :

$$\begin{aligned} \hat{x}_n &= x_0 + x_n\left(k_1 r^2 + k_2 r^4\right) \\ \hat{y}_n &= y_0 + y_n\left(k_1 r^2 + k_2 r^4\right) \end{aligned} \tag{3.8}$$

où k_1 et k_2 sont les coefficients de la distorsion radiale, (x_n, y_n) sont les coordonnées du point exprimées par rapport au centre de l'image (x_0, y_0) et r est la distance du point courant au centre de l'image $r^2 = x_n^2 + y_n^2$.

3.2 Géométrie épipolaire et stéréovision

La stéréovision permet la reconstruction 3D d'un objet observé à partir de deux vues différentes. Elle représente l'approche la plus naturelle puisque le système de vision humain fonctionne de la même façon. Elle permet de lever l'incertitude du choix de la *profondeur* d'un point lors de sa reprojection en 3D. La figure 3.2 schématise ce fonctionnement.

Grâce à la géométrie épipolaire, nous pouvons relier la position des points de l'image droite aux points de l'image gauche dans un système stéréoscopique. La géométrie épipolaire définit les contraintes existantes lorsqu'un point 3D M a pour projection un point m_1 et m_2 respectivement dans les images 1 et 2. Nous définissons comme *plan épipolaire* associé au point M, le plan créé par le point 3D M et les deux centres optiques des caméras (C_1, C_2). L'intersection entre ce plan et les plans images (I_1, I_2) forme respectivement les droites épipolaires d_1 et d_2. Cela crée une relation de dépendance linéaire entre les pro-

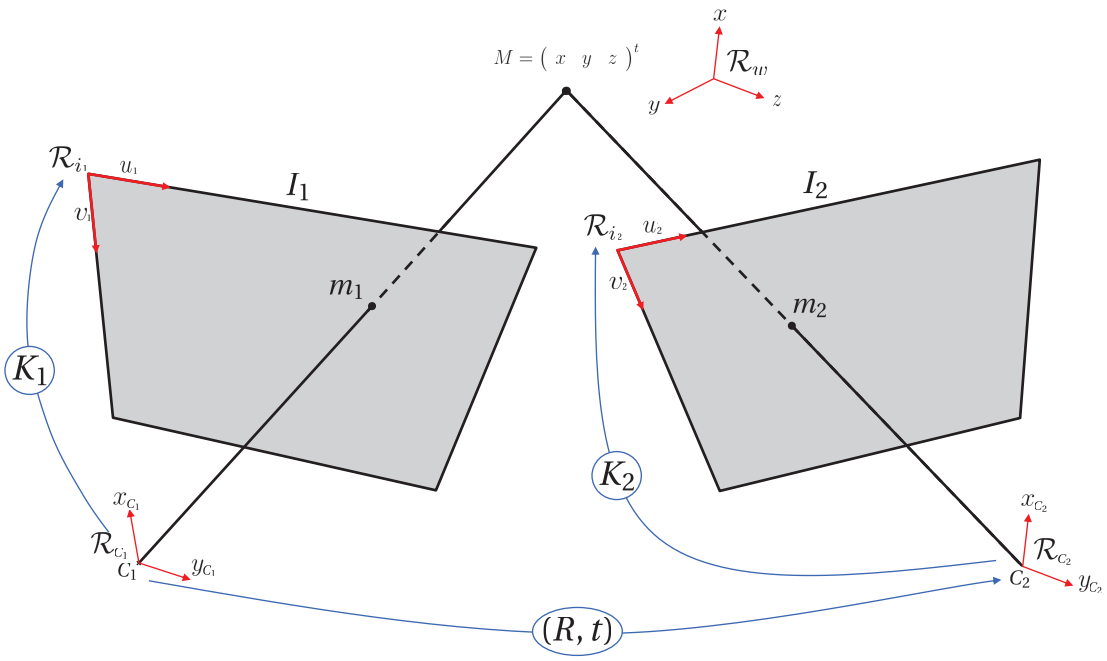


FIGURE 3.2 : Schéma d'un système de stéréovision

jections du point dans les images 1 et 2. la *contrainte épipolaire* est définie par la *matrice fondamentale F* selon l'équation 3.9.

$$m_1^t F m_2 = 0 \tag{3.9}$$

Si nous connaissons les coordonnées d'un des deux points m_1 ou m_2, il est alors possible de définir les droites épipolaires d_1 et d_2 par :

$$d_2 = F m_1 \tag{3.10}$$

$$d_1 = F^t m_2 \tag{3.11}$$

La matrice F dépend des paramètres intrinsèques (K_1, K_2) et de la transformation entre les 2 images :

$$F = K_2^t([t]_x R) K_1^{-1} \tag{3.12}$$

avec t et R représentant respectivement la translation et la rotation entre les deux caméras (C_1 et C_2) et $[t]_x$ matrice antisymétrique associée au vecteur $t = [t_x, t_y, t_z]^t$:

$$[t]_x = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix}$$

Si les paramètres intrinsèques des caméras sont connus, la contrainte épipolaire peut être exprimée dans le repère rétinien. On a alors x_1 et x_2 projetés du point M dans le repère rétinien de l'image I_1 et I_2.

$$x_1 = K_1^{-1} m_1 \tag{3.13}$$

$$x_2^t K_2^t F K_1 x_1 = 0 \tag{3.14}$$

soit

$$x_2^t E x_1 = 0 \tag{3.15}$$

avec :

$$E = K_2^t F K_1 \tag{3.16}$$

La matrice E de l'équation 3.16 est appelée *matrice essentielle*.

Toutes les contraintes précédentes sont très utiles et utilisées en stéréovision. Elles facilitent l'appariement des points entre deux images. Si nous connaissons un point dans la première image, il suffit de chercher son correspondant sur la droite épipolaire dans l'autre image. Réciproquement, elles permettent de détecter des erreurs d'appariements. La matrice fondamentale F est aussi liée au calibrage. Elle peut être évaluée directement à partir des matrices de projection des deux caméras. Dans ce cas, nous cherchons avant tout à calculer la matrice fondamentale et si nécessaire, nous remontons aux propriétés métriques en utilisant d'autres indices visuels (longueurs connues, droites parallèles, etc.).

Les bases des méthodes et modèles utilisés pour représenter les images dans le milieu sous-marin sont maintenant posées. Le chapitre suivant va détailler certaines démarches permettant d'obtenir les paramètres de ces modèles.

CHAPITRE

4 Calibrage du système de vision

Le calibrage d'une caméra est une démarche permettant l'estimation des paramètres intrinsèques et extrinsèques présentés au paragraphe 3.1. On parle alors de *calibrage fort*. Si on estime seulement la géométrie épipolaire de la paire stéréoscopique, on réalise alors un *calibrage faible*. Avec cette méthode, on ne pourra reconstruire un objet en 3D qu'à un facteur d'échelle prêt. La section qui suit va présenter le modèle de caméra qui a été utilisé ainsi que la méthode retenue pour le calibrage.

4.1 Méthodes de calibrage

Pour reconstruire la structure euclidienne d'une scène, la connaissance des paramètres de la ou des caméras est indispensable. Le calibrage permet d'établir la relation entre les dimensions images et les dimensions réelles. Il existe de nombreuses méthodes pour calibrer une caméra. Certaines utilisent des objets de calibrage (mire) de dimension et de forme connues, d'autres utilisent la rigidité de la scène. Ces dernières sont appelées méthode d'auto-calibrage. Ce paragraphe va présenter quelques méthodes de calibrages couramment utilisées regroupées en trois catégories. La première se concentrera sur les méthodes se basant sur le calcul de la matrice fondamentale. Puis une seconde présentera un calibrage fort se basant sur la matrice homographique. La dernière, contrairement aux deux autres qui déterminent tous les paramètres de la caméra, estime seulement les coefficients de la matrice de calibrage.

4.1.1 Méthodes existantes

4.1.1.1 Méthodes utilisant la matrice fondamentale

La matrice fondamentale est une des représentations de la géométrie épipolaire qui correspond à la géométrie entre deux images d'une même scène.

Méthode basée sur les équations de Kruppa. Le calcul des paramètres intrinsèques utilise les équations de Kruppa [Zeller, 1996]. Les équations de Kruppa permettent d'exprimer la rigidité du mouvement entre deux images à partir de la représentation algébrique de la conique absolue.

Lourakis et Deriche [Lourakis et Deriche, 1999] ont proposé une simplification de ces équations. Ils les dérivent en utilisant la décomposition en valeurs singulières de la matrice fondamentale ($F = UDV^t$). Ils obtiennent :

$$\frac{r^2 v_1^T G v_1}{u_2^T G u_2} = \frac{rs v_1^T G v_2}{-u_2^T G u_1} = \frac{s^2 v_2^T G v_2}{u_1^T G u_1} \tag{4.1}$$

Avec F matrice fondamentale, $D = diag(r,s,0)$, r et s valeurs propres de la matrice FF^T, $U = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}$ et $V = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}$ matrices de décomposition en éléments simples de F, $G = KK^T$ et K matrice des paramètres intrinsèques.

L'utilisation de ces équations nécessite une optimisation non-linéaire pour obtenir les paramètres intrinsèques de calibrage. Pour effectuer cette optimisation non-linéaire, il faut en premier lieu avoir un point de départ qui est donné par une solution initiale avec un vecteur de paramètres de taille $(20x1)$ représentant les paramètres de la décomposition en éléments simples.

Méthode de Mendonça et Cipolla. Mendonça et Cipolla [Mendonça et Cipolla, 1999] proposent d'estimer les paramètres intrinsèques à partir des contraintes présentes sur les trois valeurs singulières de la matrice essentielle. La matrice essentielle est équivalente à la matrice fondamentale dans le repère caméra. Elle est composée de paramètres exprimant la position de la caméra. D'après l'équation 3.16 nous pouvons exprimer :

$$F = K^{-t} E K^{-1} \tag{4.2}$$

Avec E matrice essentielle, F matrice fondamentale et $K = K_1 = K_2$ matrices des paramètres intrinsèques des appareils.

A partir de cette relation, ils proposent d'utiliser les paramètres intrinsèques comme arguments et les matrices fondamentales comme paramètres dans une fonction d'optimisation. La fonction utilisée est proportionnelle à la différence des deux valeurs singulières non-nulles de la matrice essentielle.

4.1.1.2 Méthode utilisant la matrice d'homographie

Zhang [Zhang, 2000] propose une méthode basée sur l'observation d'un plan de la scène vu sous différents angles. Il exprime l'homographie entre ce plan et l'image afin d'estimer les paramètres intrinsèques de la caméra. Si $M = (X, Y, Z)$ est un point 3D sur le plan et $m = (u, v)$ son projeté dans l'image, l'homographie H qui les lie s'exprime :

$$s\tilde{m} = H\tilde{M} \tag{4.3}$$

Avec s un facteur d'échelle arbitraire, et $H = K(Rt)$ où K est la matrice des paramètres intrinsèques et (Rt) les paramètres extrinsèques. Si on considère le plan avec pour coordonnées $Z = 0$ on a alors :

$$\begin{aligned} s\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} &= A\begin{pmatrix} r_1 & r_2 & r_3 & t \end{pmatrix}\begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} \\ &= A\begin{pmatrix} r_1 & r_2 & t \end{pmatrix}\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \\ s\tilde{m} &= H\tilde{M} \end{aligned} \tag{4.4}$$

avec r_1, r_2 et r_3 les colonnes de la matrice R tel que : $R = (r_1 r_2 r_3)$.

A partir de l'équation 4.4, deux contraintes sur les paramètres intrinsèques peuvent être exprimées :

$$h_1^t K^{-t} K^{-1} h_2 = 0 \tag{4.5}$$

$$h_1^t K^{-t} K^{-1} h_1 = h_2^t K^{-t} K^{-1} h_2 \tag{4.6}$$

Soit $B = K^{-t}K^{-1}$ une matrice symétrique, on peut alors définir un vecteur de six paramètres composant un vecteur b tel que :

$$b = \begin{pmatrix} B_{11} & B_{12} & B_{22} & B_{13} & B_{23} & B_{33} \end{pmatrix}^t \tag{4.7}$$

L'interprétation géométrique de l'homographie entre le plan et son image permet d'exprimer :

$$h_i^t B h_j = v_{ij}^t b \tag{4.8}$$

avec :

- h_i l'$i^{ème}$ vecteur colonne de la matrice H, $h_i = \begin{pmatrix} h_{i1} & h_{i2} & h_{i3} \end{pmatrix}^t$ et
- $v_{ij} = \begin{pmatrix} h_{i1}h_{j1} & h_{i1}h_{j2} + h_{i2}h_{j1} & h_{i2}h_{j2} & h_{i3}h_{j1} + h_{i1}h_{j3} & h_{i3}h_{j2} + h_{i2}h_{j3} & h_{i3}h_{j3} \end{pmatrix}^t$

Les deux contraintes 4.5 et 4.6 peuvent être réécrites comme deux équations homogènes en b :

$$\begin{pmatrix} v_{12}^T \\ (v_{11} - v_{22})^T \end{pmatrix} b = 0 \tag{4.9}$$

Si n images sont observées, n équations 4.9 seront exprimées et seront mises sous la forme du système suivant :

$$Vb = 0 \tag{4.10}$$

où V est une matrice de taille $(2n \times 6)$. Le vecteur b est estimé en résolvant le système 4.10. Les paramètres intrinsèques seront extraits à partir des coefficients du vecteur b.

Pour estimer les paramètres de distorsion radiale k_1 et k_2 rajoutés au modèle de la caméra, Zhang propose l'algorithme suivant :

1. Calculer les paramètres intrinsèques sans la distorsion (avec la méthode expliquée précédemment)
2. Projeter les points de la mire (dont la géométrie est connue) et calculer k_1 et k_2
3. Effectuer un nouveau calibrage en optimisant tous les paramètres y compris les paramètres de la distorsion radiale dans le calcul.

A la fin de l'algorithme, tous les paramètres de calibrage (intrinsèques et extrinsèques) ainsi que les coefficients de distorsion radiale sont obtenus.

4.1.1.3 Méthode exprimant la matrice de calibrage

Dron [Dron, 1993] propose une méthode pour estimer directement la matrice de calibrage K et non les paramètres individuellement. Le calcul de la matrice K utilise le mouvement de la caméra. Si on prend une séquence de n images, les petits mouvements entre les images de la séquence ne permettent pas d'estimer la matrice K en une seule étape. Une première matrice K_s est estimée, puis la meilleure matrice K_u est cherchée telle que :

$$K = K_s K_u \tag{4.11}$$

Les caractéristiques techniques de la caméra et de la carte d'acquisition des images fournies par le constructeur peuvent être utilisées pour estimer la matrice K_s. Si ces données sont inconnues, elles peuvent être calculées de manière itérative à partir de l'équation 4.11 en remplaçant K_s par la valeur calculée de K à l'itération précédente.

K_u est calculé par une méthode de moindres carrés linéaire à partir des directions de translation pour chaque image de la séquence.

Cette méthode étant itérative en fonction des images de la séquence, elle présente l'avantage de mettre à jour dynamiquement les valeurs de l'estimation. Cela pourrait permettre de prendre en compte les modifications de la distance focale dues au milieu sous-marin.

4.1.2 Discussions et choix de la méthode utilisée

Le choix de la méthode de calibrage est important pour notre application. En effet, les paramètres de la caméra permettent d'exprimer une correspondance entre le repère image (en pixel) et le repère monde (en mètre). Donc les erreurs de calibrage sont faibles, plus la précision du modèle 3D est élevée.

Toutes les méthodes présentées précédemment ont été définies pour des images de surface ; lorsque la caméra est immergée, de nombreux problèmes liés au milieu sous-marin sont introduits et la qualité des images est fortement réduite. La présence de particules en suspension crée des phénomènes de réfraction et de réflexion des rayons lumineux [Meeteen, 1999]. La contrainte la plus importante est la modification de l'indice de réfraction. Cette variation de l'indice de réfraction entraîne une modification de la distance focale apparente et donc un changement des paramètres internes du modèle de la caméra. Cela réduit donc la précision de la reconstruction 3D.

Pour éviter ces problèmes, certains auteurs comme Pessel [Pessel *et al.*, 2003] ont proposé des méthodes d'auto-calibrage qui s'adaptent aux images utilisées. Elles se basent sur la rigidité de la scène pour en extraire des données et les utiliser pour définir la géométrie épipolaire entre deux images. N'ayant aucune information sur la scène observée, nous avons écarté ce type de méthode.

D'autres auteurs comme Lavest et Rives [Lavest *et al.*, 2000] proposent de calibrer la caméra hors de l'eau et d'adapter les paramètres au milieu sous-marin. Pour cela, ils ont défini une relation entre les différents paramètres en surface et le coefficient de réfraction de l'eau. L'inconvénient de cette méthode réside dans le fait que la variation de l'indice optique lors d'une plongée n'est pas prise en compte. Il faudrait donc instrumenter les appareils photos pour analyser l'eau et déterminer le coefficient de réfraction de l'eau en permanence à partir de la profondeur, de la température et de la salinité de l'eau. Les approches utilisant l'estimation de la matrice fondamentale ne représentent pas une solution envisageable. Quelle que soit la méthode utilisée l'estimation de la matrice présente des erreurs. Si on estime les paramètres de la caméra à partir d'un résultat estimé, les erreurs de calculs s'accumulent et la précision des résultats en pâtit.

La dernière solution proposée estimant directement la matrice de calibrage ne correspond pas aux besoins de notre application. La méthode nécessite de connaître le déplacement de la caméra, ce qui n'est pas du tout évident en milieu sous-marin. De nombreux problèmes ne permettent pas d'avoir une bonne estimation du mouvement. Cette erreur d'estimation se répercuterait sur les résultats du calibrage. De plus, elle ne fournit pas les paramètres intrinsèques de la caméra individuellement. Une mesure sur une image demande la connaissance individuelle des paramètres α_u et α_v.

La méthode de calibrage retenue est celle proposée par Zhang [Zhang, 2000]. L'algorithme estime individuellement les paramètres intrinsèques et les paramètres de distorsion radiale. Il présente une flexibilité de mise en œuvre non négligeable pour le milieu sous-marin. Il suffit de disposer d'une mire plane et de réaliser différentes prises de vue de la mire dans différentes positions. De plus, l'algorithme possède une bonne robustesse aux erreurs de détection dans l'image et aux erreurs de caractéristiques géométriques de la mire.

4.2 Expérimentations et résultats

Comme introduit dans la section 4.1.2, la méthode de Zhang [Zhang, 2000] est utilisée pour calibrer les appareils photos du système stéréoscopique. Les paramètres du modèle des appareils photos sont calculés à partir d'au moins deux images et sans connaissance de leurs mouvements. De plus, la distorsion radiale est prise en compte. Elle apparaît dès qu'un appareil photo est immergé. L'optique de certains caissons étanches peut aussi ajouter ce type de déformation.

Le milieu sous-marin introduit une variation de l'indice optique du milieu. La distance focale du système de vision est modifiée lorsque cet indice varie, or elle est directement liée aux paramètres intrinsèques du modèle de la caméra. Cette variation remet donc en cause la validité du calibrage du système vidéo. Pour contourner ce problème, nous avons émis l'hypothèse que les modifications introduites par l'environnement pourraient être comprises dans le modèle de distorsion radiale de la caméra. Afin de justifier ce choix, différentes expérimentations ont été réalisées à partir de photos acquises sous l'eau et hors de l'eau. Pour ces manipulations, les séquences d'images sous-marines utilisées pour le calibrage sont issues des expérimentations réalisées en mer aux Aresquiers.

Après de nombreux calibrages à partir de données sous-marines, nous nous sommes rendu compte que le modèle de distorsion radiale d'ordre 4 n'était pas nécessaire. En effet, le coefficient d'ordre 4 était toujours presque nul (de l'ordre du 10^{-4}) lors de toutes ces expérimentations. Nous avons donc choisi d'utiliser un modèle de distorsion radiale d'ordre 2.

Le premier problème sur lequel nous nous sommes penché est l'influence du caisson étanche sur les paramètres du modèle de la caméra. Pour éliminer cette influence et ne comparer que les variations de paramètres entre l'eau et l'air, nous avons réalisé plusieurs calibrages de l'appareil dans l'air avec et sans le caisson. Le tableau 4.1 regroupe les résultats de calibrage. Nous avons donc montré que le caisson utilisé n'a aucune influence sur les paramètres du modèle et que seul le changement de milieu pouvait influencer les paramètres de distorsion.

Dans le tableau 4.1, α_u, α_v, u_0, v_0 et r sont les paramètres intrinsèques, définis par l'équation 3.5. k_1 représente le coefficient de distorsion radiale d'ordre 2. Le "skew" (r) est négligeable car il est d'ordre 10^{-4}.

Maintenant que nous savons que le caisson étanche n'a aucune influence sur les paramètres de distorsion, nous pouvons étudier l'influence du milieu sous-marin. Pour cela,

	Distance focale		point principal		Skew	Distorsion
	α_u	α_v	u_0	v_0	r	k_1
Avec caisson	3200	3225	1655	1343	0	−0.142
Sans caisson	3248	3248	1643	1297	0	−0.145

TABLE 4.1 : Résultats du calibrage de l'appareil photo en surface.

nous avons réalisé plusieurs calibrages en surface (avec le caisson étanche) et en mer. Le tableau 4.2 regroupe un de ces résultats.

	Distance focale		point principal		Skew	Distorsion
	α_u	α_v	u_0	v_0	r	k_1
Sous-marin	4572	4511	1666	1284	0	0.124
Surface	3248	3248	1643	1297	0	−0.145

TABLE 4.2 : Résultats du calibrage de l'appareil photo en surface et en mer.

Comme nous pouvons le voir dans le tableau 4.2, les points principaux u_0 et v_0 ne sont pas modifiés quelles que soient les conditions. Généralement ces points se situent au milieu de l'image et leurs valeurs sont proches de la moitié de la résolution de l'image. Ici les images utilisées ont une résolution de 3264x2448 donc les valeurs de u_0 et v_0 doivent être voisines de 1632 et 1224. Les valeurs obtenues par notre calibrage correspondent aux résultats attendus.

Lorsque nous immergeons l'appareil photo, les valeurs représentant les longueurs focales (α_u et α_v) sont modifiées par le milieu sous-marin. Ce résultat était prévisible puisque Lavest [Lavest *et al.*, 2000] démontre que les valeurs des distances focales varient d'environ 1.3 entre l'air et l'eau douce. Dans l'eau salée, ce coefficient est généralement plus grand (aux alentours de 1.39 en mer méditerranée). Dans notre cas, nous avons une variation de 1.4 pour α_u et de 1.39 pour α_v.

Le dernier paramètre de calibrage restant est le coefficient de distorsion radiale. Ce dernier subit une grande variation entre l'air (-0.15) et l'eau (0.12). Le changement de signe de ce paramètre signifie que le type de distorsion est modifié. Elle passe d'une distorsion en coussinet à une distorsion en barillet (voir figure 4.1).



(a) Distorsion en coussinet (b) Distorsion en barillet

FIGURE 4.1 : Représentation schématique de la distorsion radiale.

En étudiant les paramètres de calibrage un à un, nous avons pu valider l'hypothèse que nous avions émise sur le fait que les modifications introduites par l'environnement influent sur le modèle de distorsion radiale de la caméra.

Pour vérifier et quantifier la précision de la méthode de calibrage, nous utilisons l'erreur de reprojection. Cette approche consiste à reprojeter les points utilisés pour le calibrage dans l'image en utilisant les résultats du calibrage. Si le système de vision et le calibrage sont parfaits, les points reprojetés auront la même position dans l'image que les points utilisés. Comme un système et un calibrage ne peuvent pas être parfait, le point reprojeté ne se trouve pas sur le points calculé. Plus la distance entre ces points est faible, plus le calibrage sera précis. La figure 4.2 permet de visualiser l'erreur de reprojection de chaque point utilisé pour le calibrage.

Malgré les contraintes de l'environnement, les erreurs de reprojection restent faibles. Le tableau 4.3 regroupe les statistiques des erreurs. Le jeu de données étant proche du zéro, nous pouvons en conclure que le calibrage du système est bon.

Erreur	Minimum	Maximum	Moyenne	Écart-type
Sur x	26.10^{-6}	0.28	56.10^{-4}	0.10
Sur y	22.10^{-6}	0.36	62.10^{-4}	0.11

TABLE 4.3 : Erreur de reprojection du calibrage selon les axes x et y.

A partir de la connaissance des paramètres de calibrage et des dimensions et de la forme de la mire, il est possible de déduire les paramètres extrinsèques des différentes

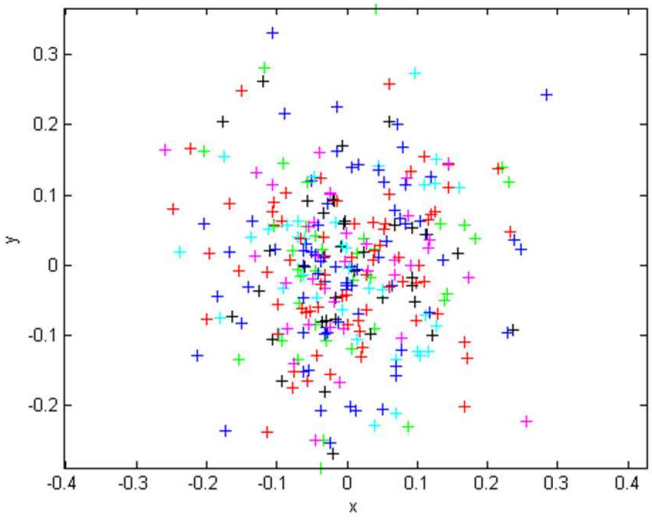


FIGURE 4.2 : Erreur de reprojection du calibrage (en pixel) [Bouguet et Perona, 1998]

prises de vue. La figure 4.3 permet de visualiser les différentes positions de l'appareil photo par rapport à la mire. Les figures 4.2 et 4.3 ont été obtenus à partir de la toolbox "camera calibration" réalisée par J.Y. Bouguet [Bouguet et Perona, 1998].

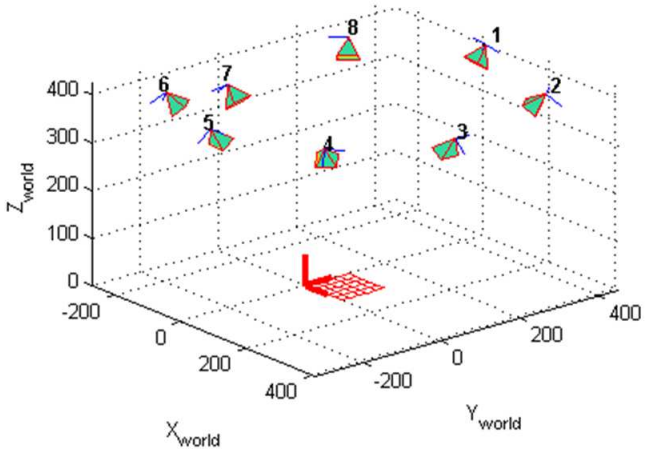


FIGURE 4.3 : Visualisation des positions de l'appareil photo lors du calibrage [Bouguet et Perona, 1998]

Maintenant que nous avons validé le calibrage pour une seule caméra, nous pouvons calibrer l'ensemble de la paire stéréoscopique. Pour cela, nous avons calibré les deux appareils photos (tableau 4.4). Une fois ces résultats obtenus, nous les avons utilisés pour extraire la transformation rigide existant entre les deux appareils photos. L'équation 4.12 présente la matrice de rotation et le vecteur de translation obtenus. La transformation estimée entre les deux appareils photos correspond au système réel.

$$R = \begin{pmatrix} 0.9883 & 0.0087 & 0.1521 \\ 0.0141 & 0.9993 & 0.0346 \\ 0.1517 & 0.0364 & 0.9879 \end{pmatrix}, \quad t = \begin{pmatrix} -334.3 \\ 4.9 \\ 5.0 \end{pmatrix} \tag{4.12}$$

	Distance focale		point principal		Skew	Distorsion
	α_u	α_v	u_0	v_0	r	k_1
Appareil de gauche	1825	1808	1194	884	0	0.24
Appareil de droite	1802	1815	1203	827	0	0.22

TABLE 4.4 : Résultats du calibrage de la paire stéréoscopique en mer.

Ce chapitre a permis d'identifier et de valider les paramètres des modèles proposés pour la représentation des images. Ce travail étant maintenant terminé, le chapitre suivant va introduire les différentes méthodes de reconstruction 3D.

Troisième partie

Reconstruction 3D locale

Cette partie permet de présenter notre méthode de reconstruction 3D pour la carte locale à partir d'images stéréoscopiques. Après avoir présenté les principales méthodes de reconstruction 3D existantes, nous exposerons les détails de la méthode utilisée. Ensuite nous détaillerons chacune des étapes de la méthode de reconstruction 3D. Lors de cette étude, nous mettrons en avant les problèmes liés au milieu sous-marin. Une solution sera proposée pour améliorer les résultats obtenus par cette méthode.

CHAPITRE

5 La reconstruction 3D

La reconstruction 3D d'objets ou de scènes à partir de séquences d'images est un sujet largement étudié ces dernières années dans la communauté de la vision par ordinateur. De nombreuses méthodes existent, mais nous allons montrer qu'elles ne sont pas toutes applicables au milieu sous-marin. En effet, elles dépendent généralement des connaissances du système de vision et des connaissances a priori de l'environnement.

Dans le paragraphe qui suit, nous allons nous focaliser sur les méthodes de reconstruction 3D employant des ensembles d'images issues d'une caméra ou d'appareils photos en vision active ou passive. La vision passive n'utilise que la lumière présente dans la scène. Dans ces techniques, l'apport de lumière artificielle de type flash ou éclairage uniforme est sollicité juste pour illuminer la scène. Lors de l'emploi de méthodes de vision active, les éclairages sont utilisés pour projeter une lumière structurée formant différentes formes. Toutes les méthodes qui nécessitent d'autres types de capteurs de vision ne seront pas détaillées.

5.1 Etat de l'art

Il existe de nombreux travaux sur la reconstruction 3D dans la littérature. Comme par exemple, Snavely [Snavely *et al.*, 2006] qui a recours à des images récupérées sur internet pour modéliser les bâtiments publics afin de créer une base de données urbaine 3D. Cette méthode n'utilise aucune information sur les appareils ayant pris les photos. Dans le même

esprit, Bartoli [Bartoli et Sturm, 2003] et Barazzetti [Barazzetti, 2011] se servent d'images dont la source est inconnue pour reconstruire des bâtiments. Leurs méthodes s'appuient sur la propriété de planéité des façades des bâtiments pour pouvoir les reconstruire en 3D. Dans un but différent, Sauvée [Sauvée, 2006], Noce [Noce, 2008] et Richa [Richa, 2010] modélisent les mouvements 3D du cœur lors d'opérations chirurgicales. Dans cette application, la scène observée ne peut être exploitée car le cœur est un objet déformable. Par contre les paramètres des caméras sont totalement connus grâce à une étape de calibrage.

En se recentrant sur les reconstructions sous-marines, on retrouve le travail d'Espiau [Espiau, 2002] en 2002 qui propose une méthode pour reconstruire des scènes texturées à partir d'une caméra non calibrée. Pour cela, il propose d'estimer les paramètres de la caméra à partir de la rigidité de la scène. Brandou [Brandou, 2008] travaille sur le même type d'images mais il propose de calibrer la caméra in situ et utilise un bras manipulateur pour déplacer la caméra. Cela lui permet de connaître exactement les conditions de prise de vue et de simplifier la reconstruction.

Kim [Kim et Eustice, 2009] propose une approche SLAM (Simultaneous Localization And Mapping) pour inspecter l'état de la coque d'un navire. Il détermine la position de la caméra en 3D en se servant d'une caméra monoculaire calibrée et une sélection de modèles géométriques, contrairement à Beall [Beall *et al.*, 2010] et Negahdaripour [Negahdaripour et Firoozfam, 2006] qui utilisent une paire stéréo calibrée pour pré-construire la trajectoire 3D. Ensuite, ils optimisent simultanément la trajectoire de la caméra et estiment une reconstruction 3D dense. Dans une même optique, Hogue [Hogue *et al.*, 2006] a développé un système composé d'une paire stéréo et d'une centrale inertielle. En combinant les informations 3D obtenues avec la paire stéréo et les données de la centrale inertielle, il estime la trajectoire du système et crée une carte 3D dense de l'ensemble de la scène.

La difficulté de ces méthodes varie en fonction des contraintes imposées par l'application : type de scènes, connaissance ou non du déplacement de la caméra, des paramètres de calibrage, de la forme de l'objet. Certains auteurs, comme Hartley [Hartley et Zisserman, 2003], Pratt [Pratt, 2001] ou Szeliski [Szeliski, 2010] proposent des méthodes générales pour reconstruire une scène 3D. Hartley et Zisserman [Hartley et Zisserman, 2003] ont proposé une méthode de reconstruction projective qui se décompose en plusieurs phases. Dans un premier temps, ils extraient des primitives dans l'image qui sont généralement des points d'intérêts. Ensuite, les points obtenus sont mis en correspondance. Cela consiste à mettre en relation la même primitive dans différentes images. Une fois tous les points appariés, la

reconstruction projective de la scène est réalisée ce qui permet d'obtenir une scène 3D. Le modèle 3D ainsi trouvé ne conserve que les propriétés projectives de la scène.

Toutes les méthodes de reconstruction présentées précédemment sont répertoriées en quatre catégories :

- Reconstruction à partir de la texture
- Reconstruction à partir de jeux de lumière
- Reconstruction à partir du mouvement
- Reconstruction à partir de la géométrie épipolaire

Nous avons mis volontairement de côté les reconstructions " basées modèles " sachant que nous ne disposons d'aucune connaissance a priori sur les scènes et sur les objets observés.

5.1.1 Reconstruction à partir de la texture

Les méthodes de reconstruction par la texture sont des méthodes de "sculptage de volume" (ou space carving). Elles utilisent généralement un ensemble d'images avec des points de vue différents du même objet ou de la même scène. Le principe de cette méthode est de créer dans un premier temps un volume enveloppant l'objet à reconstruire. Puis, ce volume est érodé successivement afin de se rapprocher du modèle 3D final. La résolution du modèle dépend de la taille des voxels qui sont enlevés par l'érosion. Le rendu obtenu étant constitué de petits voxels, il est amélioré en lissant les voxels La figure 5.1 montre un exemple de résultats obtenus par Kutulakos [Kutulakos et Seitz, 2000].

Pour que ces méthodes fonctionnent, aucune hypothèse sur la scène ni sur l'orientation de l'appareil photo ne doit être faite. Cela en fait un atout majeur, mais les critères d'érosion du modèle sont basés sur la photo-consistance de l'objet. Un voxel est photo-consistant quand, dans toutes les photos considérées, il ne présente aucune variation sur sa luminosité et sa couleur. L'utilisation de ces méthodes nécessite donc une lumière constante et homogène dans les images. En milieu sous-marin, cette hypothèse n'est pas du tout respectée.

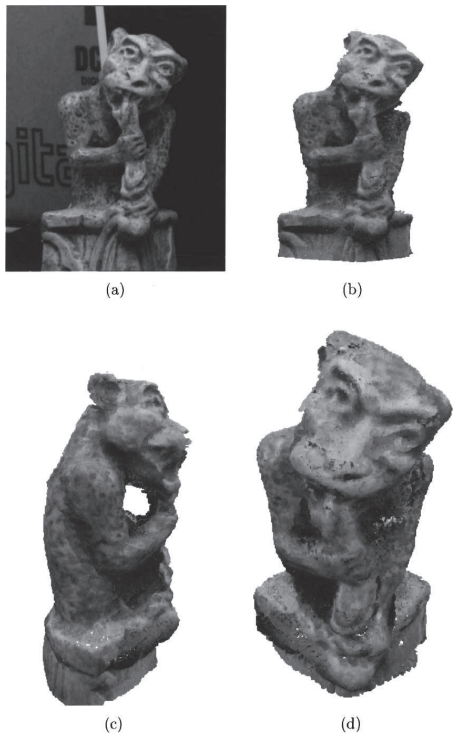


FIGURE 5.1 : Reconstruction d'une sculpture de gargouille [Kutulakos et Seitz, 2000]

5.1.2 Reconstruction à partir de jeux de lumière

Les méthodes de reconstruction à partir de jeux de lumière sont des méthodes qui utilisent ce qu'on appelle la lumière active. Le principe de ces méthodes est de projeter des formes spécifiques (lignes, quadrillage, damier, etc.) sur la scène ou l'objet. La lumière projetée peut être issue d'un projecteur numérique à lumière blanche ou de laser de longueurs d'ondes différentes.

Lorsque la lumière touche l'objet, le motif projeté se déforme et permet le calcul d'informations comme par exemple la distance. En faisant varier la position et/ou le motif dans le temps, nous pouvons extraire plus facilement des points d'intérêt et faciliter la mise en correspondance de ces points dans l'image. Les structures 3D obtenues sont généralement denses et uniformes. Le travail de Bianco [Bianco *et al.*, 2011] nous permet de visualiser les résultats obtenus avec ce type de méthode (fig. 5.2).

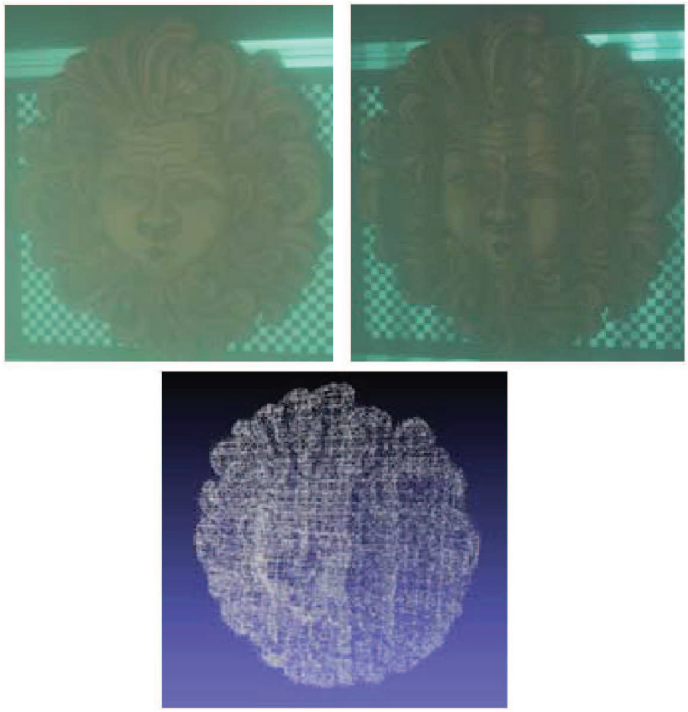


FIGURE 5.2 : Reconstruction d'une sculpture [Bianco *et al.*, 2011]

Dans l'air, ces méthodes sont souvent utilisées et fonctionnent très bien mais en milieu sous-marin, les différentes perturbations rendent leurs utilisations difficiles. L'absorption de couleur ne permet pas d'utiliser les lasers habituels (de couleur rouge). Des lasers de couleur bleue pourraient alors être employés car leur énergie est moins absorbée par l'eau. Un nouveau problème apparaît lorsqu'un laser est utilisé dans un milieu où la turbidité est importante. Les particules en suspension entraîneraient des phénomènes de réflexion et de diffraction du rayon lumineux. Un bruit important apparaîtrait sur l'image, de plus le motif projeté sur l'objet serait peu précis. La précision du modèle serait alors variable en fonction de la turbidité.

5.1.3 Reconstruction à partir du mouvement

Le mouvement réel de l'espace 3D de la caméra peut se projeter en un mouvement 2D dans la séquence d'images. Il forme alors un champ vectoriel de déplacement dans l'image. En mesurant ce champ, nous pouvons estimer la carte de disparité et ainsi retrou-

ver le modèle de l'objet observé. Ces champs de vecteurs peuvent être approximés par des champs de vitesse (flot optique).

La reconstruction à partir du mouvement (shape from motion) consiste à prendre une séquence d'images ou une vidéo d'un objet placé par exemple sur une table tournante (environnement contrôlé). Ensuite, pour chaque image, le mouvement apparent de points de l'image est estimé, ce qui permet de retrouver la structure 3D de l'objet. Pour que cette méthode fonctionne, le mouvement est supposé très faible entre chaque image. La scène observée doit être statique afin que le déplacement estimé ne soit que celui de la caméra et pas des objets. L'avantage de ce genre de méthode vient de la facilité de mettre en correspondance les pixels entre les images avec des méthodes de suivi visuel. L'information sur le mouvement effectué nous permet de prédire les disparités dans l'image suivante. Par contre ces méthodes présentent une grande sensibilité au bruit par rapport aux méthodes stéréos. L'espacement entre deux images consécutives devant être très faible, le calcul de la carte de disparité devient instable et donc aucune structure n'est exploitable. Pollefeys est l'un des précurseurs de ce type de méthode, il propose dans [Pollefeys *et al.*, 2004] une méthode pour reconstruire une scène 3D à partir d'une caméra portée "à la main". La figure 5.3 expose les résultats qu'il a obtenus.



FIGURE 5.3 : Reconstruction 3D d'une méduse antique [Pollefeys *et al.*, 2004]

Les méthodes de reconstruction par le mouvement admettent l'hypothèse que les scènes observées doivent être statiques. Dans le milieu sous-marin, nous ne pouvons garantir cette hypothèse. Les objets en suspension faussent le calcul du mouvement apparent. Comme nous l'avons vu au paragraphe précédent, Beall [Beall *et al.*, 2010] ont essayé de réduire ce phénomène en calculant la trajectoire de la caméra à l'aide de capteurs extéroceptifs. L'ajout de capteurs n'est pas une solution car ils introduisent des problèmes sur la précision de l'estimation du mouvement de la caméra et donc sur la précision de la structure reconstruite.

5.1.4 Reconstruction par géométrie épipolaire

Les méthodes générales de reconstruction 3D par la géométrie épipolaire permettent d'obtenir des structures 3D projectives. Afin de pouvoir réaliser une reconstruction euclidienne (où les angles et distances sont respectés), nous devons ajouter une information supplémentaire. Le calibrage de la caméra permet de connaitre la relation entre les dimensions réelles et les dimensions images. Elle peut s'obtenir à partir de la géométrie de la scène (droites parallèles, dimension d'objet connue, etc.), ou en exploitant des propriétés du déplacement de la caméra (translation, rotation pure, etc.) ou par une mire de calibrage.

Hartley et Zisserman [Hartley et Zisserman, 2003] ont proposé une méthode de reconstruction projective qui se décompose en plusieurs phases. Dans un premier temps, ils extraient des points d'intérêt de deux images d'un même objet (ou scène). Ensuite, les points obtenus sont mis en correspondance. Cela consiste à mettre en relation la même primitive dans différentes images. Une fois tous les points appariés, la reconstruction projective de la scène est réalisée ce qui permet d'obtenir une scène 3D. Le modèle 3D ainsi trouvé ne conserve que les propriétés projectives de la scène.

Les méthodes de reconstruction par géométrie épipolaire sont très nombreuses et varient selon les connaissances du système, de la scène ou d'objets présents dans l'image et l'environnement ne les limite pas. Nous ne pouvons pas généraliser ce type de méthodes. Elles permettent d'obtenir un modèle 3D sans aucune connaissance sur la scène ou l'objet. Cette approche nous paraît être la plus adaptée à notre problématique.

5.2 Description de notre approche

Comme nous l'avons vu au paragraphe 5.1, toutes les méthodes de reconstruction 3D utilisées en extérieur ne sont pas forcément exploitables si nous prenons en compte les perturbations du milieu sous-marin. Pour ces raisons, nous avons retenu les méthodes de reconstruction par la géométrie épipolaire. N'ayant aucune information sur la géométrie et les objets susceptibles d'apparaître dans la scène, ni aucune information sur le mouvement de la caméra, nous avons utilisé la méthode proposée par Harley et Zisserman [Hartley et Zisserman, 2003]. Elle se décompose en plusieurs phases. La figure 5.4 résume les différentes étapes nécessaires pour obtenir le modèle 3D.

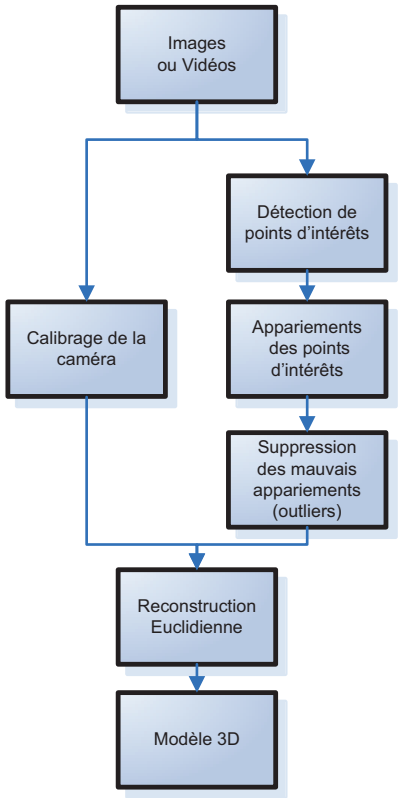


FIGURE 5.4 : Organigramme de la reconstruction 3D

Extraction des points d'intérêts : les points d'intérêts sont des primitives obtenues dans les images. Les différents types de primitives ainsi que les détecteurs de points seront détaillés dans le paragraphe 6.1.1.

Appariement des points : l'appariement consiste à mettre en correspondance les primitives dans plusieurs images qui représentent la même entité. Elles seront expliquées dans la section 6.1.2

Reconstruction Euclidienne : la méthode proosée par Hartley et Zisserman permet de réaliser une reconstruction projective. Seules les propriétés projectives telles que le birapport, l'incidence et la colinéarité sont gardées. La reconstruction euclidienne (ou métrique) n'est possible que si nous apportons des informations complémentaires comme la connaissance du déplacement entre les deux images et le calibrage de la caméra. Si le déplacement entre les deux images n'est pas connu, le modèle obtenu sera euclidien mais non métrique. C'est-à-dire que les formes, angles, etc. seront respectés mais les distances seront représentées à un facteur d'échelle prêt. Pour fixer ce déplacement nous utilisons une paire de caméras stéréoscopiques. Plus de détails seront donnés au paragraphe 7.1.

Calibrage : le calibrage de la caméra permet de mettre en relation les dimensions réelles aux dimensions images. Le calibrage du système stéréoscopique donne le déplacement qu'il y a entre les deux images. Le chapitre 4 expliquera en détail le modèle utilisé ainsi que la technique pour obtenir les paramètres.

Rendu du modèle : afin d'améliorer le rendu visuel de l'objet reconstruit, nous avons réalisé une triangulation de Delauney et plaqué les textures obtenues dans l'image. La section 7.2 détaille la méthode utilisée.

Les conditions sous-marines remettent souvent en cause la validité des modèles et des méthodes. Ainsi la précision de la reconstruction 3D d'un objet est fortement conditionnée par l'estimation des paramètres extrinsèques, intrinsèques et stéréo (paramètres du modèle) de la caméra utilisée. La détermination de ces paramètres passe par une phase de calibrage. Elle s'effectue à l'aide d'une mire par exemple, ou encore grâce à des techniques d'auto-calibrage. Ces dernières nécessitent l'application de mouvements spécifiques à la caméra et une scène rigide.

Une autre grande difficulté dans la reconstruction d'une scène ou d'un objet, est la mise en correspondance des primitives extraites dans la séquence d'images. Ces primitives

doivent être particulièrement robustes aux nombreux bruits et perturbations introduits par le milieu sous-marin.

CHAPITRE

6 Extraction et mise en correspondance de points caractéristiques

L'extraction de points caractéristiques dans les images est une étape permettant de trouver des points ayant un sens géométrique ou mathématique dans une image. La mise en correspondance représente la mise en relation dans deux images du même point de l'espace 3D.

Comme nous l'avons vu précédemment, la détection de points caractéristiques et leurs mises en correspondance sont des étapes essentielles pour faire de la reconstruction 3D. Le choix de ces éléments est crucial pour la qualité du modèle 3D. Les primitives et les appariements doivent être particulièrement robustes à de nombreuses perturbations ou transformations dans l'image, comme le changement d'échelle, d'éclairage ou le bruit dans l'image. La fiabilité de la mise en correspondance va dépendre directement du niveau de précision des informations caractérisant les primitives ; une erreur d'appariement entraînera automatiquement une erreur sur l'estimation d'un point 3D.

6.1 État de l'art

6.1.1 Détection des points caractéristiques

Comme nous l'avons dit précédemment cette étape est très importante pour la qualité du modèle 3D reconstruit. Nous allons commencer par introduire les différents types de primitives existant dans la littérature. Elles se regroupent en trois catégories :

- **Régions :** ces primitives ne correspondent pas à notre application car elles sont sensibles à la distorsion projective qui est présente en vision stéréoscopique. En effet, la transformation projective modifie la taille et la forme d'une même région.
- **Segments :** segments de droite, arcs de cercle, portions de conique, etc. Elles sont généralement utilisées lorsque les images présentent des caractéristiques de scènes polyédriques, c'est-à-dire des scènes où les structures géométriques sont facilement identifiables (ou des scènes très structurées), comme par exemple, pour l'extraction d'objets de formes géométriques connues (mines, caisses, etc.).
- **Points :** points d'intérêts, pixels, éléments de contour, etc. sont des primitives à faible contenu sémantique mais elles sont particulièrement adaptées à la vision stéréoscopique et aux scènes sous-marines. Seulement ce type de primitives est susceptible d'être détecté lorsque les images sont peu structurées.

Le choix des primitives utilisées pour notre application s'est porté sur les points caractéristiques. En effet, nos images ne disposent pas de scènes structurées et ce genre de primitives est bien adapté aux images stéréoscopiques. Les détecteurs de points ont pour but de localiser des points stables dans les images. Contrairement aux contours, les points sont présents dans une grande majorité d'images. De nombreuses méthodes existent mais nous n'en présentons ici que quelques-unes parmi les plus connues.

Détecteur de Moravec.
Le détecteur [Moravec, 1977] considère le voisinage d'un pixel et détermine les changements moyens d'intensité dans ce voisinage. Cette opération se fait à l'aide d'une fenêtre glissante se déplaçant verticalement et horizontalement. Plus précisément, nous considérons la fonction :

$$E\left(x,y\right)=\sum_{u,v} w\left(u,v\right)\left|I\left(x+u,y+v\right)-I\left(x,y\right)\right|^2 \tag{6.1}$$

où $E\left(x,y\right)$ représente la moyenne du changement d'intensité lorsque la fenêtre est déplacée de (x,y), $I(u,v)$ est l'intensité du pixel (u,v) et w spécifie la fenêtre de voisinage autour du pixel (u,v) (de valeur 1 à l'intérieur de la fenêtre et 0 à l'extérieur).

Cette fonction mesure la différence entre une fenêtre centrale et les quatre fenêtres voisines. Si un point ou un coin se trouve dans la zone considérée, la fonction E prendra une valeur élevée dans toutes les directions. Le principe de ce détecteur est donc de chercher les maxima locaux de E en chaque pixel (au-dessus d'un certain seuil).

Opérateur de Beaudet.
En 1978, Beaudet [Beaudet, 1978] cherche à formaliser la détection de coins dans une image et propose le premier détecteur de points. L'opérateur de Beaudet est basé sur l'extraction de maxima locaux à partir des dérivés secondes du signal. Cet opérateur est invariant à la rotation. Il est basé sur le calcul de la matrice Hessienne de $I(x,y)$.

$$K(x,y) = I_{xx}I_{yy} - I_{xy}^2 \tag{6.2}$$

avec : $I_x = \frac{\partial I(x,y)}{\partial x}$, $I_y = \frac{\partial I(x,y)}{\partial y}$, $I_{xx} = \frac{\partial^2 I(x,y)}{\partial x^2}$, $I_{yy} = \frac{\partial^2 I(x,y)}{\partial y^2}$, $I_{xy} = \frac{\partial^2 I(x,y)}{\partial x \partial y}$.

Détecteur de Harris et Stephens.
Harris et Stephen [Harris et Stephens, 1988] se sont aperçus que le détecteur de Moravec fonctionne dans un contexte limité. Ils ont alors identifié les problèmes et en les corrigeant, ils ont proposé un détecteur de coins très populaire : *Le détecteur de Harris*. Ils ont corrigé le caractère discret du déplacement de la fenêtre en considérant le développement de Taylor de la fonction d'intensité I au voisinage du point (u,v). Ils ont remplacé le filtre w qui binarise l'image par un filtre Gaussien. Ces améliorations les ont amenés à proposer le détecteur suivant :

$$K(x,y) = \det(C) - \lambda trace(C)^2 \tag{6.3}$$

où λ est déterminé empiriquement $\lambda = 0.04$, pour maximiser la réponse sur un modèle de coin idéal

La matrice C est la matrice Hessienne de $I(x,y)$ et est définie par :

$$C = \begin{pmatrix} I_{xx} & I_x I_y \\ I_x I_y & I_{yy} \end{pmatrix} \tag{6.4}$$

avec I_x représentant la dérivée de l'image suivant x et I_y celle selon y.

Le terme $trace(C)$ constitue l'information *contour* de l'opérateur, tandis que $\det(C)$ contient l'information *coin*. Les valeurs de $K(x,y)$ sont positives au voisinage d'un coin, négatives au voisinage d'un contour et faibles dans une région d'intensité constante.

SUSAN.
Contrairement aux autres méthodes, S.M. Smith et J.M. Brady [Smith, 1997] ont proposé un détecteur de coins qui n'utilise pas le calcul des dérivées de l'image. L'image est balayée

par un masque de pixels de forme circulaire. On appelle nucleus le centre de ce masque et USAN (Univalue Segment Assimilating Nucleus) l'aire du masque représentée par les pixels ayant un niveau de gris voisin du nucleus. La base du principe du SUSAN vient de l'association à chaque point de l'image d'une zone locale de niveau de gris. En étudiant la taille, le barycentre et les moments de second ordre de l'USAN, il est possible d'en déterminer des coins et des contours dans l'image.

Scale Invariant Feature Transform (SIFT) .
D.G. Lowe propose en 1999 [Lowe, 1999] une nouvelle approche pour détecter et apparier des points dans les images qu'il améliore en 2004 [Lowe, 2004] pour devenir le détecteur SIFT. Elle est composée de quatre étapes principales :

- **Détection des extrema dans l'espace échelle :** La première étape du calcul recherche les points potentiels dans tout l'espace échelle (sur toutes les échelles différentes) et sur toute l'image. La diminution d'échelle est simulée en appliquant successivement des filtres gaussiens qui floutent l'image. Puis la différence de ces images floutées (à chaque échelle) est réalisée. Cette méthode s'appelle une fonction de différences de Gaussienne (DoG). Tous les points candidats sont alors invariants à l'échelle.
- **Localisation des points d'intérêt :** A chaque point candidat, un modèle détaillé est capable de déterminer la position et l'échelle à laquelle il a été extrait. Les points sélectionnés sont ensuite comparés avec leurs huit voisins de leur échelle et les neuf voisins des échelles supérieures et inférieures. Le point est conservé s'il représente toujours un extremum par rapport à tous ces voisins. Tous les points ayant un faible contraste ou se trouvant sur des bords sont ensuite rejetés.
- **Association avec les orientations :** Un histogramme des orientations des gradients locaux est réalisé pour chaque point. L'orientation maximale est attribuée à ce point. Si, dans l'histogramme des orientations, il existe des valeurs à moins de 80% du maximum, un nouveau point est ajouté avec cette orientation. Il est donc possible d'avoir plusieurs candidats ayant la même échelle et les mêmes coordonnées mais avec des orientations de gradients différentes. Pour la suite des calculs, tous les points caractéristiques sont ensuite normalisés par rapport à leur échelle, à leur orientation de gradient et à leur emplacement.
- **Construction des descripteurs des points :** Le gradient local est calculé dans des sous-régions de 4×4 pixel à l'échelle sélectionnée autour de chaque point. Un histo-

gramme à huit barres des orientations des gradients est réalisé pour chaque sous-région. Ce qui donne un vecteur à 128 dimensions ($4 \times 4 \times 8$). Ce vecteur est ensuite normalisé pour le rendre robuste aux changements de luminosité.

Les points détectés sont alors complètement invariants aux translations, rotations, et changements d'échelle. Ils sont peu sensibles aux changements linéaires de luminosité et sont partiellement robustes aux transformations affines. Ils sont aussi fortement liés aux calculs des gradients locaux autour de chaque point caractéristique.

6.1.2 Mise en correspondance

Lorsque nous regardons une scène, la notion de mise en correspondance est implicite. Nos deux yeux et notre cerveau le font en permanence et savent parfaitement interpréter une scène sous différents points de vue. Il est naturel pour l'Homme d'identifier les zones identiques dans plusieurs images. En vision par ordinateur ce travail est étudié depuis plusieurs années par la communauté scientifique mais il reste très difficile et est toujours un sujet de recherche ouvert. La qualité de la mise en correspondance est un aspect très important pour la qualité du modèle 3D. Dans cette section, nous allons passer en revue quelques approches classiques et relativement robustes dans le cas d'images de scènes naturelles inconnues. Nous avons différencié les méthodes d'appariement à proprement dites et les contraintes permettant de simplifier la recherche.

6.1.2.1 Appariement de points caractéristiques

L'appariement de points caractéristiques consiste à chercher quels points 2D dans chaque image correspondent à la projection du même point 3D. Nous avons classé les différentes méthodes existantes en quatre catégories.

Méthode par corrélation.
Cette méthode est certainement une des plus connues car elle est facile à mettre en œuvre et donne de bons résultats dans de nombreuses applications. La principale contrainte à respecter pour pouvoir utiliser ces méthodes est que le mouvement entre les deux images et donc le déplacement entre les deux prises de vue est faible et que les conditions d'éclairages ont sensiblement peu bougé. Si ces deux conditions sont respectées alors il existe une grande similitude entre les valeurs des pixels représentant le même objet de la scène. En résumé, l'image ne s'est presque pas modifiée dans un voisinage proche, donc il suffit

de calculer la ressemblance entre les pixels des deux images. Concrètement, pour mesurer la ressemblance d'un point (u_1, v_1) de l'image I_1 avec un point (u_2, v_2) de l'image I_2, il faut prendre une fenêtre carrée de taille impaire $((2n+1) \times (2n+1))$ autour de chaque point et utiliser des algorithmes de mesure de similarité locale. Il existe différents critères de mesure de similarité :

- CC (*Cross-correlation*) : calcule la somme des produits entre les termes correspondants des fenêtres de corrélation.

$$CC(u,v) = \sum_{i,j} \left(I_1(u+i, v+j) I_2(u+i, v+j) \right) \tag{6.5}$$

- SSD (*Sum of Squared Differences*) : exprime la distance quadratique des intensités sur l'ensemble de la fenêtre de corrélation.

$$SSD(u,v) = \sum_{i,j} \left(I_1(u+i, v+j)^2 - I_2(u+i, v+j)^2 \right) \tag{6.6}$$

- SAD (*Sum of Absolute Differences*) : exprime la somme de la différence des valeurs absolues des intensités sur l'ensemble de la fenêtre de corrélation.

$$SAD(u,v) = \sum_{i,j} \left| I_1(u+i, v+j) - I_2(u+i, v+j) \right| \tag{6.7}$$

Toutes ces méthodes sont sensibles aux différences d'illumination entre les images. Des versions plus élaborées ont été proposées en normalisant par la variance locale des intensités et en calculant les différents critères à partir des intensités filtrées par la moyenne de l'ensemble de la fenêtre de corrélation au lieu de l'intensité seule. Pour la cross-corrélation par exemple la formule de calcul devient :

$$ZNCC(u,v) = \frac{\sum_{i,j} \left(\left(I_1(u+i, v+j) - \overline{I_1} \right) \left(I_2(u+i, v+j) - \overline{I_2} \right) \right)}{\sqrt{\sum_{i,j} \left(\left(I_1(u+i, v+j) - \overline{I_1} \right)^2 \left(I_2(u+i, v+j) - \overline{I_2} \right)^2 \right)}} \tag{6.8}$$

Cela permet de rendre les critères invariants aux variations uniformes de luminosité dans les images et de s'affranchir des différences de gains et d'offsets des caméras.

Méthode par invariants locaux.
Contrairement aux appariements par corrélation qui cherchent les correspondants en utilisant les informations d'intensités, ces méthodes cherchent des caractéristiques invariantes aux rotations, à l'échelle, aux variations d'éclairage, aux transformations projectives, etc. Elles attachent aux points d'intérêts des vecteurs avec les attributs invariants. Ces méthodes sont souvent utilisées pour chercher une image spécifique dans une base de données.

La méthode de SIFT [Lowe, 2004] que nous avons présentée au paragraphe 6.1.1 utilise ce type d'informations. Elle attache à chaque point d'intérêt un vecteur de 128 invariants lié à l'orientation du gradient local. De nombreuses informations peuvent être intégrées à ces vecteurs mais cela au détriment du temps de calcul. Plus les vecteurs sont grands, plus nous avons de chances de trouver le correspondant mais plus nous mettons de temps pour calculer les invariants et pour les comparer.

Méthode par suivi visuel.
Au lieu d'essayer d'apparier directement deux ensembles de points entre deux images, il peut être intéressant d'essayer de suivre ces points tout au long de la séquence d'images en utilisant les informations temporelles. On réalise alors du " tracking " ou suivi de caractéristiques. Le but de ce type de méthode est de suivre le plus grand nombre de primitives entre des images successives représentant la même scène. Le type de primitives à suivre n'a pas d'importance, on peut travailler avec des points, des coins, des droites, des contours, des formes, etc. La littérature dans ce domaine étant abondante, nous ne présentons ici que quelques méthodes de référence. Nous laissons le choix au lecteur de s'en rapprocher s'il souhaite plus d'informations.

Tomasi et Kanade [Tomasi et Kanade, 1991] ont proposé une approche de suivi de points en se basant sur un calcul de mesures par corrélation pour déterminer la transformation entre les deux images. Ils sont partis du principe que, dans une séquence vidéo, le déplacement entre deux images successives et les variations de luminosité sont faibles. Shi et tomasi [Shi et Tomasi, 1994] ont travaillé sur la qualité intrinsèque des points à suivre. Pour faciliter leur suivi, les points doivent avoir certaines caractéristiques, comme la robustesse aux transformations affines ou aux changements de luminosité par exemple. Ces travaux sont à l'origine de la méthode de suivi très populaire KLT (Kanade, Lucas, Tomasi) [Lucas et Kanade, 1981; Tomasi et Kanade, 1991; Shi et Tomasi, 1994]. L'inconvénient de ce type de méthodes réside dans le fait que les hypothèses de fonctionnement (faible mou-

vement et changement infime entre les formes des objets) ne sont pas toujours respectées. Lorsque le mouvement est trop important ou quand la déformation de l'objet est trop grande, ces méthodes ne fonctionnent plus. Pour éviter ce type de problèmes, certains auteurs ont proposé de remettre la primitive à jour en respectant certaines conditions pour éviter les dérives temporelles par exemple.

Jurie [Jurie et Dhome, 2002] a proposé une méthode de suivi de zone d'intérêt en temps réel, robuste aux rotations, aux translations, à l'échelle et aux variations de luminosité. Il se base sur le calcul d'une matrice d'interactions dans laquelle vont être stockées toutes les modifications possibles du patch sélectionné pour le suivi. Ce calcul se fait hors-ligne lors de l'initialisation de l'algorithme. Le suivi se décompose en plusieurs étapes ; la première consiste à prédire la translation, la rotation et l'échelle à l'instant t ; puis il faut calculer la différence entre le motif de référence et le motif " vu " à la position prédite. Pour finir, cette différence est multipliée par la matrice d'interaction et donne la correction à apporter à la prédiction pour avoir le mouvement complet entre l'instant $t-1$ et t.

Méthode dense.
Les méthodes denses se décomposent en deux catégories différentes ; les méthodes locales et les méthodes globales. Dans les méthodes locales, nous retrouvons principalement les méthodes que nous avons vues précédemment. Elles ne prennent plus en compte une seule primitive mais un ensemble de primitives. Dans les méthodes que nous avons vues précédemment, nous n'utilisions qu'un point ou une fenêtre autour d'un point alors qu'ici les méthodes utilisent par exemple des nuages de points ou des ensembles de patchs. Les critères de choix sont légèrement plus complexes mais les principes restent les mêmes.

Les méthodes globales permettent de palier certains inconvénients des méthodes locales. Elles augmentent souvent la robustesse des zones locales où par exemple, il y aurait occultation d'une primitive ou une zone de texture uniforme. Cependant ces méthodes sont beaucoup plus complexes et sont souvent formulées sous forme de problèmes d'optimisation et rajoutent donc le risque de " tomber " dans des extrema locaux.

6.1.2.2 Contraintes géométriques

Afin d'améliorer les méthodes décrites à la section 6.1.2.1 qui n'utilisent que des informations photométriques de l'image, nous pouvons ajouter d'autres contraintes mais cette fois-ci de types géométriques.

Ordonnancement.
La contrainte d'ordonnancement indique que l'ordre des pixels est préservé sur la ligne épipolaire correspondante [Koch *et al.*, 2000]. On entend par là que, si on connaît un point p_1 et son correspondant p_1^* dans l'autre image, alors on sait que tous les points de la ligne épipolaire se trouvant à droite de p_1 sont à droite sur la ligne épipolaire de p_1^* dans l'autre image.

Unicité.
Un point 3D dont la projection est un pixel sur une image a forcément une projection sous forme de pixels qui lui correspond sur une autre image. Cette contrainte n'est pas valable lors d'occultation. La correspondance entre les points doit être bidirectionnelle, c'est-à-dire que si un point p_1 a pour correspondant un point p_1^* alors le point p_1^* doit avoir pour correspondant le point p_1. Cette contrainte permet de vérifier les mauvais appariements et les occultations.

Mouvement global.
Lorsque le mouvement global dans l'image est connu, nous pouvons l'utiliser pour réduire la zone de recherche des points correspondants. Il est inutile de chercher le point correspondant dans toute l'image mais il suffit de le chercher dans une zone beaucoup plus petite où nous savons qu'il doit se trouver. Cette contrainte est valable s'il n'y a pas d'occultations et permet donc de vérifier les appariements et de détecter les occultations.

Géométrie épipolaire.
Comme nous l'avons vu dans la section 3.2, la géométrie épipolaire permet de réduire la zone de recherche du point correspondant à sa ligne épipolaire dans l'autre image. En réalité les erreurs d'estimations de la matrice fondamentale modifient cette limitation à un bandeau et plus à une ligne. Comme pour les autres méthodes géométriques, elle n'est valable que lorsqu'il n'y a pas d'occultations. Cette contrainte nous permet aussi de vérifier si les appariements sont corrects et s'il n y a pas occultations.

6.1.3 Correction des appariements par géométrie épipolaire

Toutes les méthodes présentées précédemment ont des limitations et peuvent fournir de faux appariements ("outliers"). L'utilisation de ce type de points crée généralement des

points aberrants dans le modèle 3D. Afin d'éviter leurs utilisations, il existe des algorithmes de vérifications comme les Moindres Carrés Médians (LMEDS Least Median of squares) ou RANSAC (Random Sample Consensus) qui permettent de classifier les couples de points appariés en bons ou mauvais appariements (inliers et outliers). Ces méthodes donnent des résultats assez proches, mais la méthode de RANSAC permet d'estimer simultanément la matrice fondamentale. Dans ce paragraphe, nous allons présenter une méthode de vérifications utilisant l'algorithme de vérifications de RANSAC.

Fischler et Bolles [Fischler et Bolles, 1981] ont proposé en 1981 une méthode d'estimation robuste de modèles : l'algorithme RANSAC. L'idée de cet algorithme est de définir un modèle et d'identifier ses paramètres avec N sous-ensembles (tirages) aléatoires de n données. A partir de ces N estimations du modèle, une classification des données est possible. La validation des données est basée sur le calcul d'une distance critique du modèle estimé. Nous considérons qu'une estimation robuste du modèle est atteinte pour le modèle maximisant le nombre de données valides.

Dans notre application, cet algorithme est utilisé pour calculer la matrice fondamentale définissant la géométrie épipolaire de la scène. Le modèle de calcul utilisé pour déterminer la matrice fondamentale est la méthode des huit points présentée par Hartley [Hartley, 1997]. Il améliore la méthode des huit points proposée par H.C. Longuet-Higgins en 1981 [Longuet-Higgins, 1981] en normalisant les points avant l'estimation. Comme son nom l'indique cette méthode ne nécessite que huit paires de points appariés pour estimer les paramètres de la matrice fondamentale. L'algorithme que nous avons utilisé pour corriger les appariements et avoir une estimation robuste de la matrice fondamentale est le suivant :

1. Sélection au hasard de huit points et estimation de la matrice fondamentale
2. Calcul pour chaque point de l'erreur de reprojection (la distance euclidienne entre le point projeté dans la deuxième image et la droite épipolaire correspondante). Si la distance est supérieure à un seuil d'acceptation, le point est rejeté. Tous les points respectant la distance sont regroupés dans un sous-ensemble solution.
3. Répétition des étapes 1 et 2 N fois. Le nombre N est choisi suffisamment grand pour s'assurer qu'il existe au moins un des sous-ensembles composé uniquement de bons appariements (avec une probabilité de 0.99).
4. Choix du sous-ensemble qui a le plus grand nombre d'éléments et calcul de la matrice fondamentale à partir de ces points.

Le sous-ensemble ayant le plus grand nombre de points est donc composé avec une probabilité de 0.99 de bons appariements et quasiment tous les mauvais appariements ont été supprimés.

6.2 Choix des méthodes utilisées

La détection et l'appariement de points caractéristiques sont des étapes essentielles pour obtenir la reconstruction 3D d'un objet. La qualité du modèle 3D est liée à la précision de ces éléments. Comme nous l'avons indiqué précédemment au paragraphe 6.1, il existe de nombreuses méthodes pour réaliser ces étapes mais elles ne possèdent pas les mêmes caractéristiques donc elles n'auront pas les mêmes performances sur des images sous-marines. Nous rappelons que la robustesse des méthodes est un aspect très important dans notre travail par le fait que nous travaillons dans l'eau qui est un milieu fortement contraint. Il faut que nous soyons capables d'obtenir des résultats quelles que soient les conditions de prises de vues.

Pour cela, nous avons choisi d'utiliser des primitives de type *points*. Le bruit présent dans le milieu sous-marin réduit fortement les informations utiles dans l'image. De ce fait, les images utilisées sont peu structurées. Les primitives *points* sont les plus efficaces dans ces conditions. Les primitives *segments* sont difficiles à utiliser à cause du manque de structure dans l'image. Le système de prises de vues que nous utilisons favorise aussi l'utilisation de points. En effet, la géométrie épipolaire permet de réduire la complexité de certaines méthodes avec l'utilisation de points. Les primitives *régions* sont sensibles aux changements que subit l'image (transformation projective) qui modifient la taille et la forme de la région. Ce manque de robustesse par rapport au système que nous utilisons ne nous permet pas de les utiliser.

L'environnement sous-marin entraîne une forte présence de bruits dans les images. Mais ces bruits ne sont ni constants ni prévisibles. De ce fait, il ne faut pas simplement se focaliser sur la robustesse aux bruits élevés mais il faut aussi s'assurer que la méthode que nous utilisons fonctionne lorsque le bruit est très faible. Afin d'obtenir un modèle 3D le plus dense et précis possible, nous devons retenir une méthode robuste dont les points détectés sont caractéristiques de la géométrie de l'objet à reconstruire. Pour satisfaire ces contraintes, nous avons choisi d'utiliser deux méthodes de détections et d'appariements de points présentées au paragraphe 6.1.

La première méthode retenue est celle de Harris et Stephens [Harris et Stephens, 1988]. Sa construction s'appuie sur des informations de coins et de contours, elle nous permet d'obtenir des points proches des coins de l'objet et donc de trouver des points caractérisant sa géométrie. Afin de réaliser l'appariement entre ces points, nous avons utilisé une méthode par corrélation car toutes les conditions pour avoir un résultat robuste sont présentes (faibles déplacements, faibles variations de la luminosité).

Le milieu sous-marin nécessitant une grande robustesse, la deuxième méthode mise en place est la méthode de SIFT [Lowe, 2004]. Celle-ci est reconnue comme étant une des plus robustes et fonctionnant le mieux, par le monde de la vision par ordinateur. Elle est robuste aux translations, aux changements d'échelles, aux rotations et est peu sensible aux variations linéaires de la luminosité et aux transformations affines. Mise à part la robustesse de cette méthode, le choix des points caractéristiques de cette méthode évite les coins et contours. Nous n'aurons donc aucune information sur les caractéristiques géométriques de l'objet mais des points plus répartis sur les volumes des objets. Ils viendront parfaitement compléter les points de Harris qui eux se trouvent sur les coins. Avec l'utilisation de ces deux méthodes, le modèle 3D possèdera des points sur les contours et coins grâce à Harris et des points sur la surface de l'objet grâce à SIFT.

6.2.1 Harris & corrélation

Par rapport aux contraintes imposées par l'environnement et l'application, les méthodes de détection de points de Harris et d'appariement par corrélation ont été sélectionnées. En effet, Harris permet d'obtenir des points caractéristiques de la géométrie de l'objet à reconstruire et la corrélation semble être une méthode robuste aux conditions sous-marines.

Pour le calcul de la corrélation nous avons utilisé une méthode qui permet de rendre le critère invariant aux variations uniformes de luminosité dans l'image et de s'affranchir de gains et offsets de la caméra. Le critère utilisé est celui défini par l'équation 6.8.

La recherche des points à apparier se fait de manière globale, c'est-à-dire que nous avons pour un point de l'image 1 tous les points de l'image 2 comme candidats potentiels et inversement. Les appariements sont validés lorsque le couple de points correspond dans les deux sens. Cette méthode de recherche de candidat est très longue car il faut calculer la corrélation entre tous les points des deux images. Afin d'améliorer les performances et de diminuer le temps de calcul, nous réduisons le nombre de candidats possibles en gui-

dant les appariements avec les contraintes géométriques présentées à la section 6.1.2.2. Toutes les contraintes ne peuvent pas être utilisées pour guider les appariements puisque certaines nécessitent la connaissance de la matrice fondamentale ou d'un certain nombre de couples déjà appariés. Celles qui ne sont pas utilisées pour guider les appariements sont utilisées dans une étape de corrections des appariements.

Notre application comprend deux systèmes de captures d'images : la paire stéréoscopique et la vidéo. Lorsque nous utilisons les images issues de la paire stéréoscopique, nous avons connaissance de la matrice fondamentale grâce à une étape de calibrage réalisé précédemment. Les contraintes d'ordonnancement, de géométrie épipolaire et d'unicité peuvent alors être utilisées pour guider l'appariement. Les points candidats ne seront pas cherchés dans toute l'image mais seulement dans une zone spécifique pour chaque point. Une fois tous les couples appariés, une étape de correction est ajoutée afin de vérifier si, malgré le guidage, il ne subsiste pas certaines erreurs. Pour cela nous utilisons dans un premier temps l'algorithme de RANSAC puis nous terminons avec une vérification par contrainte du mouvement global de l'image.

Lorsque deux images successives sont issues de la vidéo, la matrice fondamentale n'est pas déterminée par le calibrage et les méthodes de guidage par géométrie épipolaire et ordonnancement ne peuvent être appliquées. L'appariement se fait donc de manière globale en respectant la contrainte d'unicité. L'étape de vérification est cette fois-ci plus importante. Nous commençons par vérifier les couples de points avec la méthode de RANSAC afin d'obtenir la matrice fondamentale. Nous utilisons ensuite les contraintes d'ordonnancement et de mouvement global de l'image. Certaines étapes de correction sont redondantes avec les étapes de guidage mais cela permet de limiter au maximum le nombre de faux appariements.

6.2.2 SIFT

La méthode de SIFT extrait des points d'intérêt robustes aux translations, aux changements d'échelle et aux rotations. Elle est peu sensible aux variations linéaires de la luminosité et aux transformations affines. A chaque point d'intérêt est associé un vecteur d'invariants locaux à 128 dimensions qui correspond à l'orientation du gradient local dans une zone de 4×4 autour du point. Ces vecteurs sont ensuite utilisés pour associer un point de l'image 1 à son correspondant dans l'image 2. La mise en correspondance est réalisée en calculant de manière globale la distance euclidienne entre les vecteurs d'invariants. Lors

de la sélection des points caractéristiques, la méthode supprime automatiquement tous les points se trouvant sur des bords et sur des contours. Cela permet d'obtenir des points répartis sur la surface de l'objet. Ils seront donc forcément différents de ceux obtenus par la méthode de Harris et permettront d'augmenter le nombre de points utilisés pour le modèle 3D.

Comme pour la méthode utilisant Harris et la corrélation, nous corrigeons et/ou aidons l'appariement en fonction du système de capture d'images. Si les photos sont issues de la paire stéréoscopique, grâce à l'étape de calibrage, nous connaissons les paramètres de la matrice fondamentale. Nous pouvons donc aider les appariements avec les contraintes géométriques d'unicité, d'ordonnancement et de géométrie épipolaire. Une fois les couples points obtenus, nous utilisons la contrainte sur le mouvement global de l'image pour réduire le nombre d'outliers. Pour finir nous appliquons une dernière correction par l'algorithme de RANSAC.

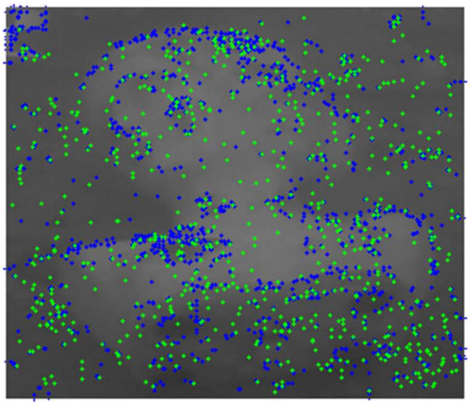
Lorsque les images sont issues d'une vidéo, la matrice fondamentale n'étant pas connue, nous ne pouvons plus guider les appariements avec la contrainte d'ordre et de géométrie épipolaire. Seule la contrainte d'unicité est utilisée. La correction des appariements se fait alors en premier avec RANSAC pour obtenir la matrice fondamentale. Nous pouvons ensuite appliquer l'ordonnancement et pour finir la correction par mouvement global.

Toutes ces corrections nous permettent d'avoir le moins de mauvais appariements et donc de réduire le nombre de points aberrants du modèle 3D.

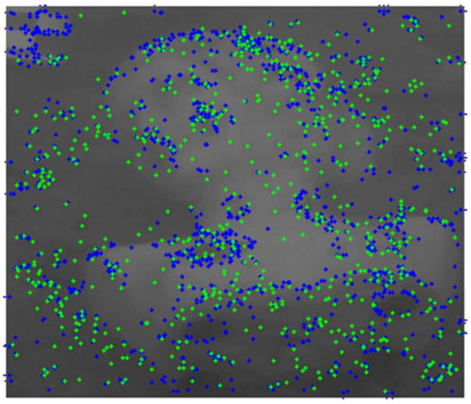
6.3 Expérimentations et résultats

Comme nous l'avons vu précédemment, la détection et la mise en correspondance de points d'intérêts sont des étapes essentielles pour la reconstruction 3D. Nous avons appliqué les deux méthodes sélectionnées sur une paire d'images du buste prises lors des expérimentations en mer aux Aresquiers. La figure 6.1 montre les points détectés sur les deux images avec les deux méthodes. Nous retrouvons les points de Harris en bleu et les points de la méthode SIFT en vert. Les figures 6.2 et 6.3 permettent de visualiser le mouvement estimé entre les deux images, à partir des points appariés avant et après la suppression des erreurs de mise en correspondance. Sur la figure 6.3(b), nous pouvons voir qu'après la correction des appariements, il reste encore quelques outliers. Ce phénomène est dû à la méthode de correction par géométrie épipolaire (RANSAC) définie à la section 6.1.3 qui

garantit une probabilité de 0.99 de bons appariements. La figure 6.2(b) ne présente pas d'outlier dans cet exemple mais la probabilité de faux appariement reste la même.

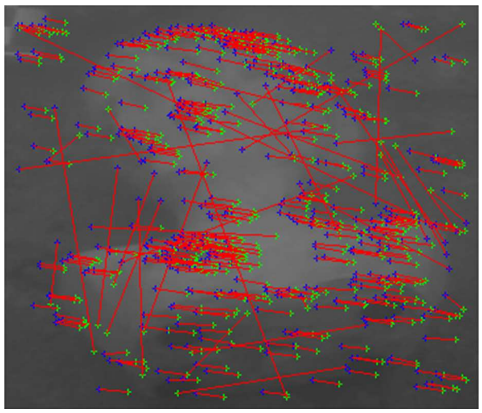


(a) Image de gauche

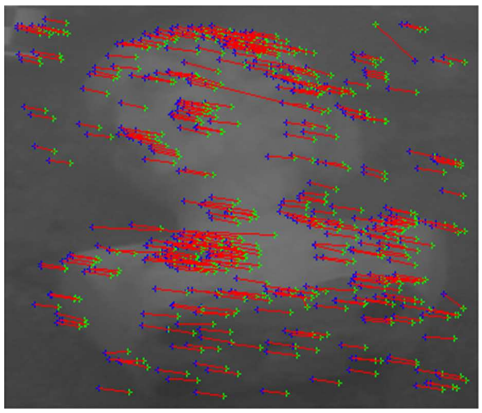


(b) Image de droite

FIGURE 6.1 : Détection des points caractéristiques sur une image prise en mer aux Aresquiers.

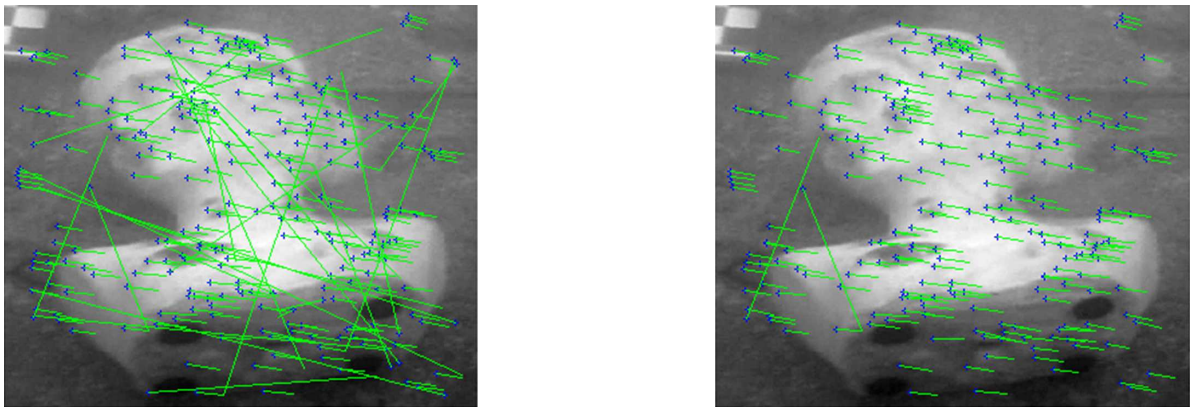


(a) Mouvement avant correction



(b) Mouvement après correction

FIGURE 6.2 : Estimation de mouvement global à partir des appariements des points de Harris.



(a) Mouvement avant correction (b) Mouvement après correction

FIGURE 6.3 : Estimation de mouvement global à partir des appariements des points de SIFT.

Nous avons comparé le nombre de points détectés et appariés pour les deux approches. Le tableau 6.3 résume les résultats obtenus à partir de quatre jeux d'images. Les trois premières paires d'images ont été sélectionnées afin d'avoir un faible déplacement entre les deux images. La dernière paire d'images possède un déplacement plus important. Les seuils de détection des méthodes ont été ajustés pour obtenir le même nombre de points caractéristiques. Pour chaque paire, nous avons comparé le nombre de points détectés et le nombre de points appariés avant et après la correction des appariements.

Ces expérimentations montrent que la méthode de Harris & corrélation donne de meilleurs résultats que celle de SIFT. Le nombre de points robustes détectés et appariés avec Harris est aux alentours de 34% alors que SIFT n'en trouve que 17%. Ces résultats sont en contradictions avec ceux que nous trouvons généralement sur des images en surface. Dans le cas de la quatrième paire d'images, les résultats de Harris sont légèrement inférieurs par rapport aux trois premières car le déplacement entre les deux images est plus grand. En effet, nous atteignons une des limitations de l'utilisation de la méthode de corrélation.

Pour les deux approches, la méthode de correction des appariements permet de supprimer environ 4% de faux points. Ce nombre n'est pas très grand mais il permet de rendre la méthode plus robuste. De plus, les faux appariements ont une importance non négli-

	Harris			SIFT		
	Points détectés	Points appariés	Inliers	Points détectés	Points appariés	Inliers
Paire 1	902 I1 902 I2	342 38%	305 34%	902 I1 1122 I2	198 22%	162 18%
Paire 2	797 I1 797 I2	278 35%	244 30%	797 I1 894 I2	167 21%	146 18%
Paire 3	894 I1 894 I2	339 38%	310 35%	894 I1 999 I2	196 22%	148 17%
Paire 4	1148 I1 1148 I2	344 30%	291 26%	1505 I1 1148 I2	229 20%	186 16%

TABLE 6.1 : Comparaison du nombre de points appariés et du nombre de bon appariement obtenu avec les deux méthodes dans un environnement sous-marin.

geable dans l'estimation du mouvement global des images. Le tableau 6.3 regroupe les différents déplacements estimés à partir des points appariés.

		Mouvement estimé avant correction				Mouvement estimé après correction			
		Min.	Max.	Moy.	Ecart type	Min.	Max.	Moy.	Ecart type
Paire 1	sur x	−584	420	100	120	101	149	129	14
	sur y	−464	482	42	78	46	58	51	2
Paire 2	sur x	−584	535	98	120	106	132	119	7
	sur y	−471	425	34	77	30	44	38	3
Paire 3	sur x	−129	282	−58	42	−76	−50	−64	6
	sur y	−171	155	3	24	−4	9	3	4
Paire 4	sur x	−613	554	236	236	321	415	368	33
	sur y	−472	424	93	134	129	149	139	7

TABLE 6.2 : Estimation du mouvement à partir des points de Harris (en pixel).

Comme nous pouvons le voir dans le tableau 6.3, l'estimation du mouvement à partir des points corrigés est largement plus précise. En effet, même si les valeurs moyennes restent proches, l'écart type est beaucoup plus faible.

6.4 Limitation de SIFT en environnement sous-marin

Beaucoup d'études sur les résultats fournis par SIFT en surface ont été réalisées dans la littérature mais très peu pour des images sous-marines [Oliver *et al.*, 2010], [Sarafraz *et al.*, 2010]. Dans cette section, nous allons montrer à travers différentes expérimentations pourquoi la méthode de SIFT n'est pas robuste aux bruits sous-marins et quel bruit en particulier la perturbe.

6.4.1 Filtrage des images sous-marines

Afin de déterminer quel bruit met SIFT en échec, nous proposons d'ajouter le bruit présent en mer sur des images de surface. Il existe différentes méthodes de filtrage dans la littérature comme le filtre proposé par Jaffe [Jaffe, 1995] ou encore celui de Hou [Hou, 2009] mais tous utilisent un filtre global. Tous les bruits pris en compte par ces méthodes sont inclus directement dans un même et unique filtre. Ces approches ne sont pas envisageables dans notre cas car elles ne permettent pas de définir le ou les bruits responsables de cette baisse de performance. Nous avons donc créé différents filtres modélisant les trois principales causes de bruits sous-marins.

Le premier phénomène que nous avons pris en compte est souvent le plus visible lorsque nous regardons une image sous-marine avec un éclairage naturel. Nous avons l'impression que les images sont vertes et ont perdu des couleurs. Cet effet est dû à l'absorption de l'énergie lumineuse de l'eau qui entraine la disparition de certaines couleurs. En effet, les couleurs disparaissent en fonction de la distance séparant la source de lumière de l'objet. Lors d'un éclairage naturel, la source de lumière est représentée par la surface de l'eau. Les couleurs qui disparaissent en premier sont celles dont la longueur d'ondes est la plus grande, comme le rouge ou l'orange (aux alentours de cinq mètres). Le bleu est celle qui se propage le mieux car elle a la longueur d'ondes la plus faible. Afin de modéliser ce phénomène, nous avons choisi de créer un filtre réduisant l'amplitude de chaque canal de couleurs d'une image RVB (Rouge-Vert-Bleu) en fonction de l'atténuation du milieu. C'est-à-dire que, si, par exemple, l'absorption de la couleur rouge est de 50%, l'intensité de tous les pixels du plan image rouge sera diminuée de 50%.

Comme cela a été indiqué dans la section 2.2, les particules en suspension ont une influence très importante sur la qualité des images. Elles provoquent des phénomènes de réflexion et de réfraction des rayons lumineux [Meeteen, 1999] qui entrainent ce bruit élevé sur les images. Pour modéliser ce phénomène, nous nous sommes appuyé sur des tra-

vaux réalisés dans le domaine de la physique appliquée. Certains auteurs comme Bandyopadhyay [Bandyopadhyay *et al.*, 2005] ou Adrian [Adrian et Westerweel, 2010] utilisent le bruit généré par les particules sur une image pour en définir certaines caractéristiques. Ils illuminent les particules et prennent une photo à l'intérieur du milieu. Ils analysent ensuite le bruit présent dans l'image pour déterminer la vitesse de déplacement des particules, la densité, etc. Le bruit généré dans l'image par l'illumination des particules en suspension est un bruit de speckle. Cela nous amène à modéliser le phénomène de turbidité dans le milieu sous-marin par ce type de bruit.

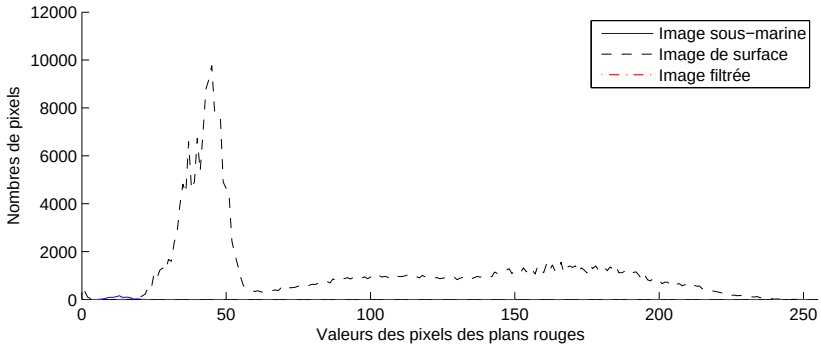
De plus, les particules en suspension créent une composante de diffusion directe qui provoque un effet de flou et diminue les composantes dynamiques des images [Bazeille, 2008]. De nombreux filtres existent pour modéliser ce type d'effet [Schettini et Corchs, 2010], nous avons utilisé un filtre Gaussien passe-bas.

6.4.2 Expérimentations dans différents milieux

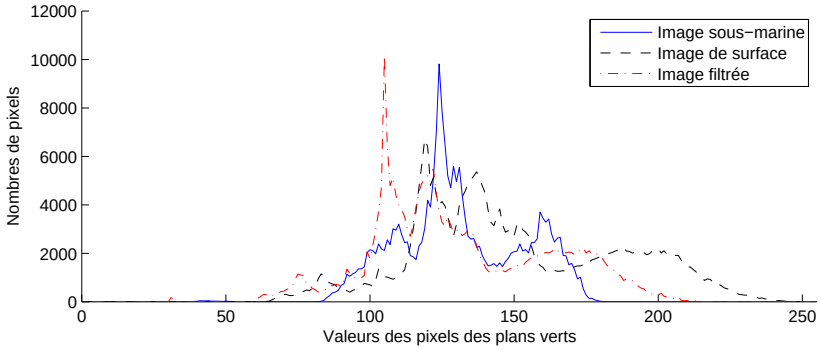
Comme nous venons de le montrer, la méthode de SIFT ne fonctionne pas bien sous l'eau. Nous proposons de chercher quelle caractéristique du milieu sous-marin pose problème. Nous utilisons des images de surface sur lesquelles nous ajoutons les filtres proposés précédemment pour simuler les conditions sous-marines. Les deux méthodes de détections et d'appariements sont testées avec, dans un premier temps un seul filtre, puis tous les filtres sont ajoutés un par un pour arriver à une image sous-marine simulée.

Le premier filtre utilisé pour tester ces méthodes simule le phénomène d'absorption de l'intensité lumineuse (présenté à la section 2.2). Pour déterminer les proportions de l'atténuation de couleurs, nous avons utilisé des histogrammes de couleurs pour chaque plan de l'image. La figure 6.4 représente les histogrammes de couleurs d'une image sous-marine, d'une image de surface et d'une image filtrée.

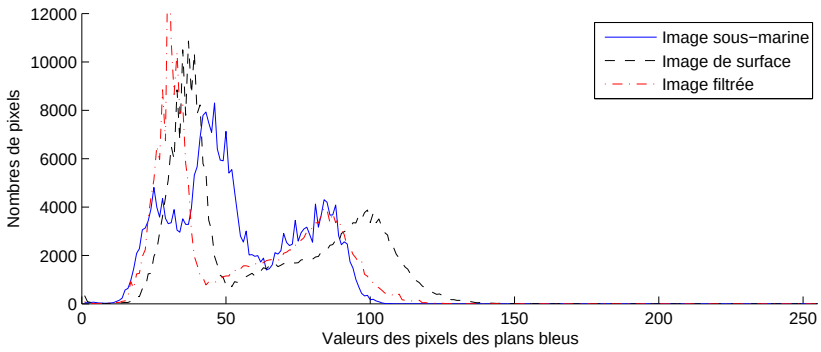
Le plan rouge a complètement disparu dans l'image sous-marine. L'intensité des plans vert et bleu a diminué respectivement de 15% et 13% par rapport à l'image de surface. Nous avons testé plusieurs images et cette tendance s'est confirmée dans tous les cas. Les résultats obtenus sont en corrélation avec la théorie. En effet avec ces conditions sous-marines (salinité, profondeur, température, distance de la source de lumière), le rouge disparaît totalement aux environs de 5 mètres et les couleurs verte et bleue sont moins atténuées.



(a) Histogrammes des rouges



(b) Histogrammes des verts



(c) Histogrammes des bleus

FIGURE 6.4 : Histogrammes des plans rouge (a), vert (b) et bleu (c) d'images sous-marines, de surfaces et filtrées.

Ces valeurs sont utilisées pour ajuster le filtre d'atténuation des couleurs pour toutes les images suivantes. La figure 6.5 représente une image sous-marine (6.5(a)) et une image de

surface filtrées (6.5(b)) avec les paramètres déterminés précédemment. Les couleurs entre les deux images sont sensiblement les mêmes.



(a) Image sous-marine

(b) Image de surface filtrée

FIGURE 6.5 : Image sous-marine et résultats du filtrage de couleur sur une image de surface.

Ensuite, nous avons identifié l'influence du bruit de speckle et du flou dans les images sous-marines. Ces deux bruits ont été estimés en même temps pour deux raisons. La première est que ces bruits sont liés physiquement, la turbidité ajoute un effet de flou et de speckle. Deuxièmement, la mesure et l'identification exacte des noyaux de bruits dans des images ne sont pas une chose évidente à réaliser avec précision. C'est pour ces raisons que nous proposons de calculer la variance locale dans les images, pour identifier simultanément les paramètres du filtre de speckle et du filtre Gaussien. Nous avons analysé l'histogramme et calculé les moyennes de la variance locale des images sous-marines. Les paramètres de chaque filtre ont été ajustés afin d'obtenir les mêmes valeurs de variance locale entre les images sous-marines et les images filtrées. La figure 6.6 montre l'histogramme de variance locale d'une image de surface, d'une image sous-marine et de l'image filtrée. Les paramètres des deux filtres estimés ont été appliqués sur une image de surface.

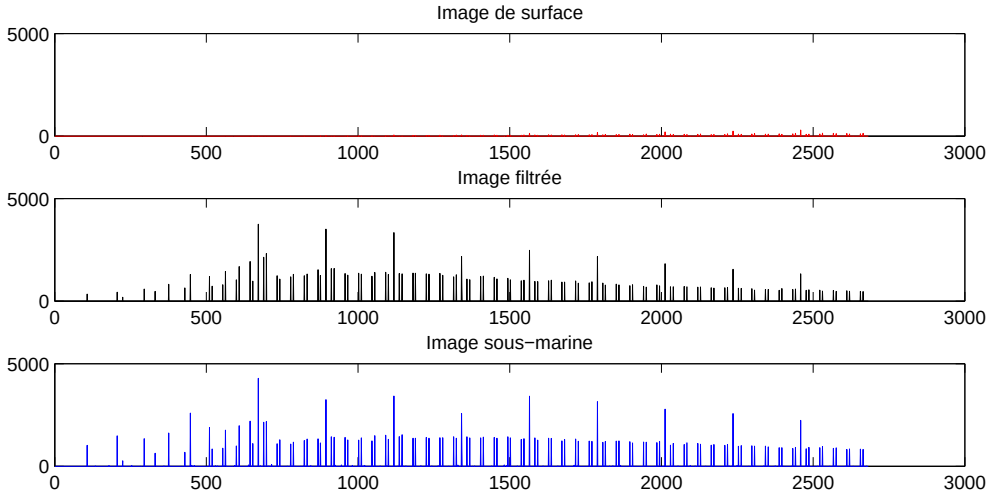


FIGURE 6.6 : Histogrammes des variances locales d'images sous-marines, de surfaces et filtrées.

La figure 6.7 représente une image sous-marine (6.7(a)) et une image de surface filtrée (6.7(b)) avec les paramètres déterminés précédemment.



(a) Image sous-marine



(b) Image de surface filtrée

FIGURE 6.7 : Image sous-marine et résultats du filtrage de speckle et de flou sur une image de surface.

Maintenant que les paramètres de tous les filtres sont identifiés, leurs effets sont testés sur les deux méthodes de détections et d'appariement de points. Pour commencer, nous avons utilisé une paire d'images dont le déplacement entre les images correspondait à celui du tableau 6.3. Le tableau 6.3 résume les résultats obtenus pour les deux méthodes avec les différents filtres.

	Harris		SIFT	
	Points détectés	Inliers	Points détectés	Inliers
Aucun	617 I1 617 I2	154 25%	625 I1 617 I2	212 35%
Couleur	629 I1 629 I2	143 22%	692 I1 629 I2	223 36%
Flou	631 I1 631 I2	140 22%	631 I1 666 I2	227 36%
Speckle	581 I1 581 I2	139 24%	581 I1 593 I2	36 6%
Speckle & Flou	631 I1 631 I2	152 24%	631 I1 641 I2	91 15%
Tous	583 I1 583 I2	153 26%	608 I1 583 I2	84 15%

TABLE 6.3 : Comparaison des nombres de points des deux méthodes avec différents filtres appliqués sur une image de surface.

Comme ces expérimentations le montrent, SIFT fonctionne mieux que Harris dans des conditions identiques à celles de surface (changement de couleurs ou effet de flou). Lorsque le bruit de speckle est introduit dans l'image, le nombre de bons appariements diminue de façon dramatique avec seulement 6% de points appariés. Cette détérioration est due à la partie appariement de SIFT. Les descripteurs permettant l'appariement sont calculés à partir du gradient local autour du point d'intérêt, or le bruit de speckle introduit une granularité aléatoire dans les images. Le calcul du gradient se retrouve complètement perturbé par ce type de bruit.

Nous avons ensuite testé différentes combinaisons de filtres mais à chaque fois que le bruit de speckle était présent, les performances de SIFT étaient mauvaises. A l'opposé, les résultats de la méthode de Harris sont quasiment constants quels que soient les filtres utilisés.

Afin de vérifier si cette tendance était générale, nous avons réalisé les mêmes tests sur un groupe d'une vingtaine d'images avec des objets et des fonds différents. Le minimum, le maximum et la moyenne du taux de bons appariements des deux méthodes sont regroupés dans le tableau 6.4.

	Harris Inliers			SIFT Inliers		
	Min.	Max.	Moy.	Min.	Max.	Moy.
Aucun	22.3	27.6	24.8	34.4	37.3	35.5
Couleur	22.7	26.6	24.1	33.6	38.2	35.7
Flou	21.4	27.0	24.5	34.8	38.9	36.4
Speckle	21.3	26.2	23.9	5.6	7.0	6.2
Speckle & Flou	23.1	26.7	24.6	13.8	15.3	14.5
Tous	24.1	27.0	25.7	12.1	15.0	14.1

TABLE 6.4 : Statistiques du taux de bons appariements obtenus avec les deux méthodes sur 20 images.

Maintenant que nous avons mis en évidence la faiblesse des descripteurs de SIFT au bruit de speckle, nous proposons de vérifier les deux algorithmes sur des images prises en bassin. En effet, elles sont soumises au même modèle de bruits que les images en mer mais la turbidité est quasiment nulle en basin. Les caractéristiques du milieu nous permettent de tester les deux méthodes dans une eau avec très peu de bruit de speckle. Nous avons testé les méthodes sur une vingtaine d'images de chaque type de milieu (surface, bassin, mer et filtré) et réalisé des statistiques des résultats. Le tableau 6.4.2 regroupe les statistiques du taux de bons appariements des images prises dans les différents milieux.

	Harris Inliers			SIFT Inliers		
	Min.	Max.	Moy.	Min.	Max.	Moy.
Surface	22.3	27.6	24.8	34.4	37.3	35.5
Filtré	24.1	27.0	25.7	12.1	15.0	14.1
Mer	21.4	28.1	25.1	11.9	16.7	15.2
Bassin	23.8	28.5	26.1	26.2	30.7	28.4

TABLE 6.5 : Statistiques du taux de bons appariements des deux méthodes dans différents milieux (surface, filtré, mer et bassin).

Comme nous pouvions le prévoir, le taux de bons appariements de SIFT est supérieur à celui de Harris pour des images en surface et en bassin. La méthode de SIFT possède de

meilleurs résultats pour des images en bassin que pour celles en mer. Ce phénomène est dû au faible nombre de particules en suspension dans le bassin et donc du faible bruit de speckle.

Ces résultats confirment notre idée que le bruit de speckle met en échec l'appariement de SIFT. Nous pouvons conclure que le descripteur de SIFT n'est pas robuste au bruit de speckle et donc qu'il n'est pas le plus adapté aux images sous-marines, contrairement à la méthode de Harris/corrélation qui est robuste quels que soient les bruits présents dans l'eau.

6.4.3 Amélioration de la robustesse de SIFT

Afin d'améliorer la robustesse de SIFT, nous proposons de réaliser un préfiltrage pour réduire le bruit de speckle présent dans les images sous-marines. Nous retrouvons dans la littérature des méthodes de préfiltrage global des images. Par exemple Bazeille [Bazeille, 2008] a proposé une méthode pour corriger l'ensemble des bruits présents en milieu sous-marin. Les traitements réalisés sur l'image améliorent les résultats qualitatifs de l'image mais la précision de détection sur l'image est réduite. Pour cette raison nous avons essayé de réduire au maximum le bruit de speckle par un seul filtre.

Le bruit de speckle étant courant dans les images sonars ou échographiques, de nombreuses méthodes de filtrages existent. Les comparaisons entre les filtres les plus courants tels que les filtres de Lee [Lee, 1980], de Frost [Frost *et al.*, 1982], de Kuan [Kuan *et al.*, 1985], le filtre anisotropique [Perona et Malik, 1990; Grieg *et al.*, 1992] et le filtre à base de transformées en ondelettes [Mallat, 2000] ont montré que le filtre anisotropique est celui donnant les meilleurs résultats. Il permet d'augmenter l'homogénéité de l'image tout en préservant les contours et les détails [Tahiri Alaoui *et al.*, 2005].

Nous avons testé les deux méthodes de détections et d'appariements avec différentes valeurs de filtrage et dressé des statistiques dans le tableau 6.4.3.

Ces résultats montrent que le pré filtrage anisotropique améliore les résultats de SIFT. L'image est légèrement lissée et la granularité dans l'image est atténuée donc les détecteurs de SIFT sont moins perturbés par le bruit (le gradient local est plus précis). Lorsque le filtre est appliqué sur les images, la méthode de SIFT obtient de meilleurs résultats que Harris, mais ces résultats restent inférieurs à ceux obtenus par Harris sans pré filtrage. Ce résultat peut se justifier par le fait que le filtre anisotropique permet de simplifier graduellement l'image et de se débarrasser d'oscillations indésirables mais il engendre de légères dégra-

	Harris Inliers			SIFT Inliers		
	Min.	Max.	Moy.	Min.	Max.	Moy.
Mer	21.4	28.1	25.1	12.9	16.7	15.2
Filtrage faible	17.2	23.4	20.6	12.1	16.2	14.8
Filtrage moyen	12.4	17.1	14.9	15.9	18.9	17.4
Filtrage fort	7.9	13.7	10.8	18.8	21.2	20.1

TABLE 6.6 : Statistiques du taux de bons appariements des deux méthodes après différentes valeurs du filtrage anisotropique

dations de la structure. L'utilisation d'une méthode de pré filtrage anisotropique n'est pas envisageable pour notre application car elle réduit la précision de la sélection des points dans l'image.

Ce chapitre a permis la détection et la mise en correspondance robuste de points dans des images sous-marines très bruitées. Ces points étant obtenus, le chapitre suivant va présenter les méthodes de projection vers la 3D pour obtenir le modèle final 3D.

CHAPITRE

7 Création du modèle 3D

La projection des points en 3D et l'amélioration du rendu visuel sont les dernières étapes à réaliser pour obtenir le modèle 3D local. Le but de cette partie est d'utiliser les points caractéristiques 2D extraits des images pour les projeter dans un espace 3D. Le nuage de points ainsi obtenu est ensuite traité pour y ajouter une surface et une texture.

La projection des points en 3D peut se faire de différentes manières. Si nous utilisons des photos successives issues d'une vidéo, les contraintes géométriques du système ne nous permettent pas de reconstruire le nuage de points avec les distances métriques exactes mais seulement à une homothétie, à un facteur d'échelle prêt. Cela n'est pas le cas avec les images stéréoscopiques. Elles nous permettent d'ajouter une contrainte géométrique sur la transformation rigide entre les deux images et de ce fait d'obtenir un nuage de points avec le bon facteur d'échelle (métrique).

7.1 Projection des points dans l'espace 3D

Comme nous l'avons vu précédemment, deux types de capteurs différents sont utilisés : une caméra et une paire stéréoscopique. La méthode de reconstruction est quasiment identique pour les deux systèmes. Seule la connaissance de la géométrie épipolaire les différencie. La méthode utilisant la paire stéréoscopique sera présentée en premier. Les étapes de calibrage des appareils photos et du système stéréo permettent de connaître tous les paramètres de la géométrie épipolaire, contrairement à la caméra qui ne permet

pas d'obtenir ces informations. La géométrie épipolaire sera estimée à partir de la *matrice essentielle (E)* définie à la section 3.2.

7.1.1 Reconstruction à partir d'images stéréoscopiques

Grâce au chapitre 4 et au chapitre 6, les paramètres de la géométrie épipolaire et les appariements des points caractéristiques 2D sont connus. A partir de ces informations, nous pouvons effectuer la reconstruction 3D de ces points dans le repère caméra. Nous avons le choix entre plusieurs approches :

- L'approche linéaire qui utilise directement l'expression des rayons visuels issus des deux caméras pour obtenir le point 3D. Nous avons alors le système suivant :

$$\begin{cases} (p_{11} - p_{31}u)\,x + (p_{12} - p_{32}u)\,y + (p_{13} - p_{33}u)\,z + (p_{14} - u) = 0 \\ (p_{21} - p_{31}v)\,x + (p_{22} - p_{32}v)\,y + (p_{23} - p_{33}v)\,z + (p_{24} - v) = 0 \\ (p'_{11} - p'_{31}u')\,x + (p'_{12} - p'_{32}u')\,y + (p'_{13} - p'_{33}u')\,z + (p'_{14} - u') = 0 \\ (p'_{21} - p'_{31}v')\,x + (p'_{22} - p'_{32}v')\,y + (p'_{23} - p'_{33}v')\,z + (p'_{24} - v') = 0 \end{cases} \tag{7.1}$$

Où : (u, v) et (u', v') sont les coordonnées des points dans les images et $p_{i,j}$ et $p'_{i,j}$ représentent les éléments des matrices de projection des deux caméras définies par l'équation (3.7). Nous choisissons de placer le repère monde (définie en 3.1) sur le repère de la caméra gauche de notre système stéréo. Pour cela nous fixons la transformation entre les deux repères (T) comme identité et donc nous avons P :

$$\begin{aligned} P &= K.T \\ &= K \begin{pmatrix} R_p & t_p \\ 0 & 1 \end{pmatrix} \\ &= K \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{aligned} \tag{7.2}$$

où K est la matrice de calibrage de l'appareil photo. La transformation P' sera de la forme :

$$P' = K \begin{pmatrix} R_{p'} & t_{p'} \\ 0 & 1 \end{pmatrix} \tag{7.3}$$

où $R_{p'}$ et $t_{p'}$ représentent respectivement la rotation et la translation rigide existant entre les deux appareils photos du système stéréo.

On peut réécrire le système 7.1 sous la forme :

$$A.X = B \tag{7.4}$$

avec :

$$X = (x, y, z)^t \tag{7.5}$$

$$A = \begin{pmatrix} p_{11} - p_{31}u & p_{12} - p_{32}u & p_{13} - p_{33}u \\ p_{21} - p_{31}v & p_{22} - p_{32}v & p_{23} - p_{33}v \\ p'_{11} - p'_{31}u' & p'_{12} - p'_{32}u' & p'_{13} - p'_{33}u' \\ p'_{21} - p'_{31}v' & p'_{22} - p'_{32}v' & p'_{23} - p'_{33}v' \end{pmatrix} \tag{7.6}$$

$$B = \begin{pmatrix} p_{14} - u \\ p_{24} - v \\ p'_{14} - u' \\ p'_{24} - v' \end{pmatrix} \tag{7.7}$$

Il suffit de résoudre ce système par un calcul de pseudo-inverse pour obtenir les coordonnées du point 3D. Pour obtenir le nuage de points complet de la scène observée, ce calcul doit être effectué pour tous les couples de points appariés.

Cette méthode présente l'avantage d'une résolution simple et rapide mais si l'une des données d'appariements ou de calibrage présente une erreur, elle se répercutera sur les coordonnées des points 3D.

- L'approche non-linéaire consiste à minimiser l'erreur de reprojection par une méthode numérique. Cette méthode permet de réduire l'influence des mauvais appariements sur la position des points 3D. L'erreur de reprojection consiste à projeter le point 3D estimé $\hat{X}$ dans les plans images en $\hat{x}$ et $\hat{x'}$. Les points satisfont donc la géométrie épipolaire. La distance euclidienne entre x et $\hat{x}$ et x' et $\hat{x'}$ est calculée et minimisée pour définir le nouveau point 3D $\hat{X}$.

L'erreur de reprojection est minimisée en utilisant une méthode numérique de minimisation telle que Levenberg-Marquardt. La fonction de minimisation est :

$$f\left(x,x'\right)=\sqrt{d(x,\hat{x})^2+d\left(x',\hat{x}'\right)^2} \tag{7.8}$$

où $d\left(x,\hat{x}\right)$ est la distance euclidienne entre le point détecté et le point reprojeté.

La méthode non linéaire permet de corriger des erreurs sur l'appariement des points si ceux-ci ne se trouvent pas sur la droite épipolaire correspondante. Les couples de points que nous utilisons sont vérifiés par géométrie épipolaire (algorithme de RANSAC). Tous les couples ont normalement leur correspondant sur la droite épipolaire correspondante. L'utilisation de l'approche non-linéaire n'apporte pas de robustesse supplémentaire à la méthode linéaire. De ce fait, nous avons utilisé l'approche linéaire.

7.1.2 Reconstruction à partir d'une vidéo

Dans l'approche que nous utilisons pour la reconstruction à partir des images stéréoscopiques, nous avons besoin de connaître la transformation rigide liant les deux caméras or cette transformation est inconnue avec l'utilisation d'images issues d'une vidéo. La section 4 nous permet d'exprimer les paramètres intrinsèques de la caméra. La méthode de RANSAC estime les valeurs de la matrice fondamentale (cf. section 6.1.3). Ces deux résultats permettent d'exprimer la matrice essentielle (eq. 3.16). Hartley [Hartley, 1992] propose une méthode pour estimer cette transformation à partir de la matrice essentielle. La translation trouvée sera estimée à un facteur d'échelle prêt. On a vu précédemment que la matrice essentielle s'exprime :

$$E=K^tFK \tag{7.9}$$

où K est la matrice de calibrage de la caméra et F la matrice fondamentale.

En réalisant la décomposition en éléments simples de E, Hartley montre qu'il est possible d'obtenir deux matrices de rotation et deux vecteurs de translation.

$$R_1 = U_E W V_E^t \tag{7.10}$$
$$R_2 = U_E W^t V_E^t \tag{7.11}$$
$$t_1 = U_{E,3} \tag{7.12}$$
$$t_2 = -U_{E,3} \tag{7.13}$$

avec : - $SVD(E) = U_E S_E V_E^t$,

- $W = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
- U_3 la troisième colonne de la matrice U.

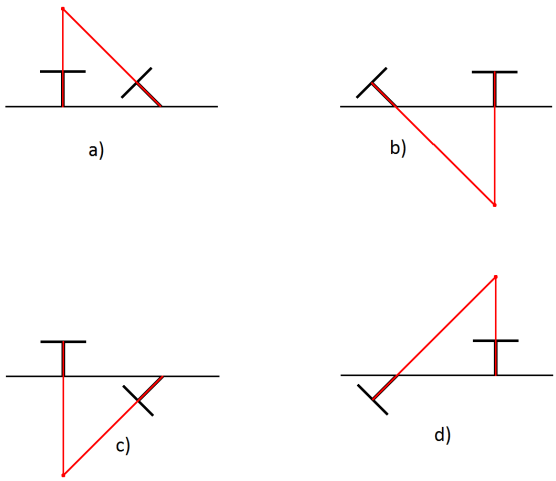


FIGURE 7.1 : Les quatre solutions possibles pour l'estimation de la transformation. *Seule la solution (a) possède les points reconstruits devant les deux caméras.*

Il existe donc quatre transformations possibles comme solutions. La figure 7.1 permet de visualiser les quatre orientations possibles des caméras. Afin de déterminer celle qui correspond à un système réel, nous devons vérifier si tous les points reconstruits se trouvent devant les caméras. Pour ce faire, nous reconstruisons les points en 3D avec la méthode précédente en utilisant chacune des transformations. Nous vérifions quelles transformations permettent d'avoir les points devant la caméra de gauche. Il ne reste alors que deux transformations possibles. Nous exprimons ensuite les points reconstruits dans le repère de la caméra de droite et nous gardons la transformation dont les points se retrouvent devant la caméra.

La transformation entre les deux prises de vue étant estimée, le nuage de points 3D est calculé à partir des points appariés précédemment. Le nuage obtenu sera Euclidien mais non métrique, c'est-à-dire qu'il sera reconstruit à un rapport d'homothétie prêt.

7.2 Rendu du modèle 3D

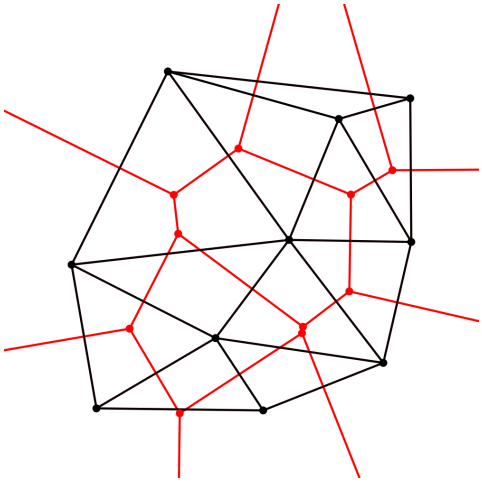
A la vue des nuages de points obtenus précédemment, nous avons du mal à identifier les objets qu'ils représentent. Pour améliorer le rendu 3D, nous ajoutons la surface et la texture aux modèles.

7.2.1 Triangulation

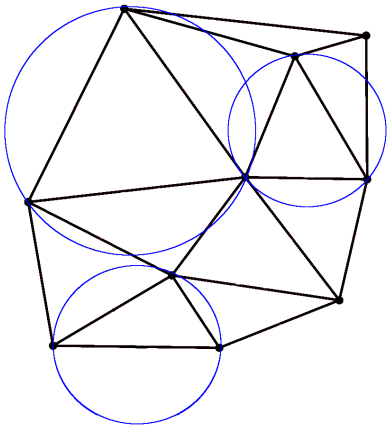
Le but de cette partie est d'obtenir le maillage de la surface représentant le mieux l'objet ou la scène reconstruite, en reliant tous les points du modèle par des arêtes. Il n'est pas évident de le retrouver à partir de points 3D. La triangulation fournira souvent l'enveloppe convexe de l'objet et non la surface désirée. Pour contourner ce problème, nous avons contraint la surface 3D à partir des triangles 2D. Le maillage est calculé à partir des points 2D puis il est projeté sur les points 3D. Cette approche permet de respecter la géométrie de l'objet et d'être sûr de n'avoir que des triangles visibles lors du plaquage des textures.

La triangulation de Delaunay 2D [Delaunay, 1934] est effectuée à partir des points détectés dans l'image de gauche. Cette image est choisie comme image de référence car le repère monde est positionné sur cette caméra. La triangulation de Delaunay permet de maximiser les petits angles et ainsi d'éviter les triangles allongés. Elle est obtenue en reliant par une arête les points voisins des régions correspondantes dans le diagramme de Voronoï (voir figure 7.2(a)). Un ensemble de points maillés est dit de Delaunay si la triangulation est telle qu'aucun triangle ne contient un autre point à l'intérieur de son cercle circonscrit. La figure 7.2(b) permet de visualiser cette condition. Les triangles ainsi obtenus sont projetés sur la structure 3D.

La triangulation que nous venons de présenter permet d'obtenir le maillage des nuages de points présentés précédemment. La figure 7.3 montre les résultats obtenus à partir des photos stéréoscopiques de la figure 2.3. Il est difficile pour un œil non avisé d'identifier une quelconque forme sur ce modèle. Sans reconnaître l'objet observé, il est impossible d'envisager qu'un archéologue puisse réaliser des mesures métriques. C'est principalement pour ces raisons que nous avons plaqué la texture sur la surface 3D dans le paragraphe suivant.

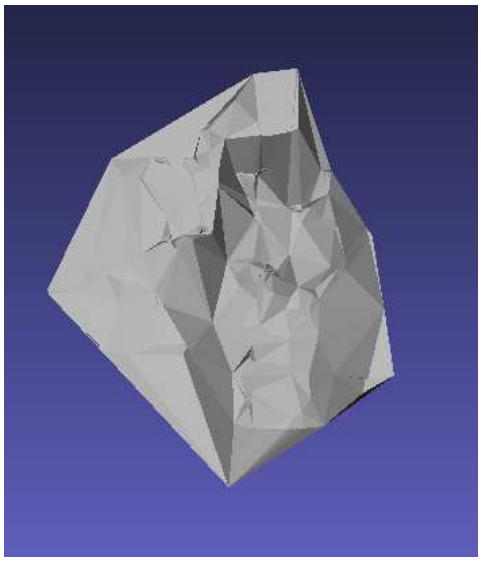


(a) Relation dual entre le diagramme de Voronoï (en rouge) et la triangulation de Delaunay (en noir)

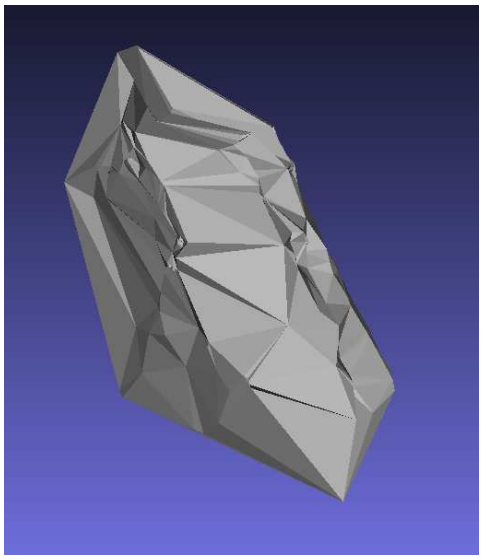


(b) Visualisation des conditions sur les cercles circonscrits

FIGURE 7.2 : Triangulation de Delaunay.



(a) Image de face



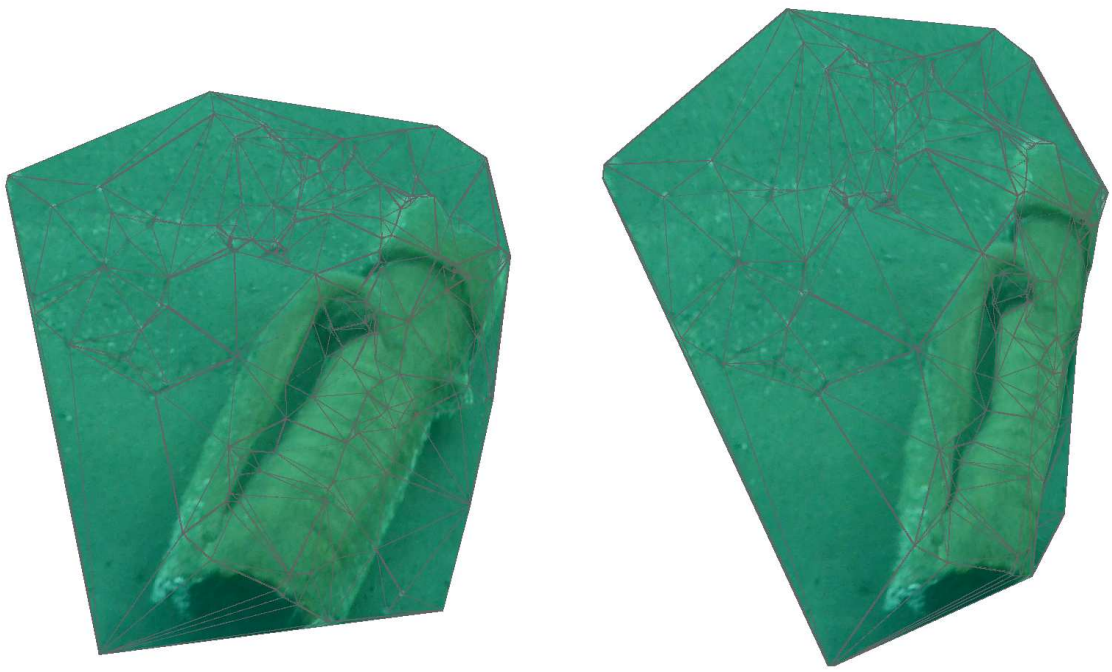
(b) Image de profil

FIGURE 7.3 : Surface 3D d'un col d'amphore des Saintes-Maries-de-la-mer.

7.2.2 Plaquage de la texture

Pour terminer le rendu visuel du modèle et donner un aspect photo réaliste à l'objet reconstruit, la texture est plaquée sur les triangles obtenus précédemment. La texture utilisée est extraite de l'image de référence choisie pour le maillage. Chaque texture d'un triangle 2D est plaquée sur le triangle 3D correspondant.

Les méthodes décrites précédemment permettent d'obtenir les résultats présentés par la figure 7.4. Nous pouvons ainsi visualiser la reconstruction 3D de l'amphore de la figure 2.3. Le plaquage de texture apporte une information très utile pour identifier l'objet observé. Contrairement à la figure 7.3, nous pouvons déterminer le type d'objet présent dans l'image.

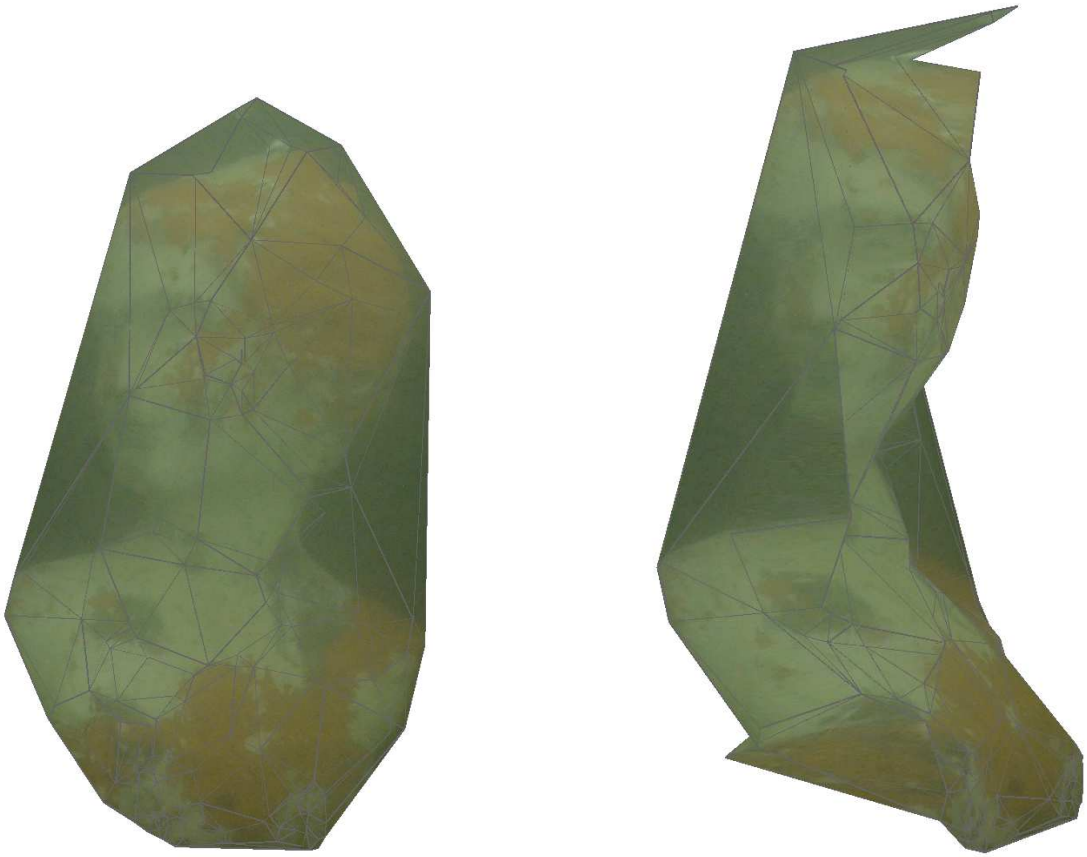


(a) Image de face (b) Image de profil

FIGURE 7.4 : Modèle 3D texturé d'un col d'amphore des Saintes-Maries-de-la-mer.

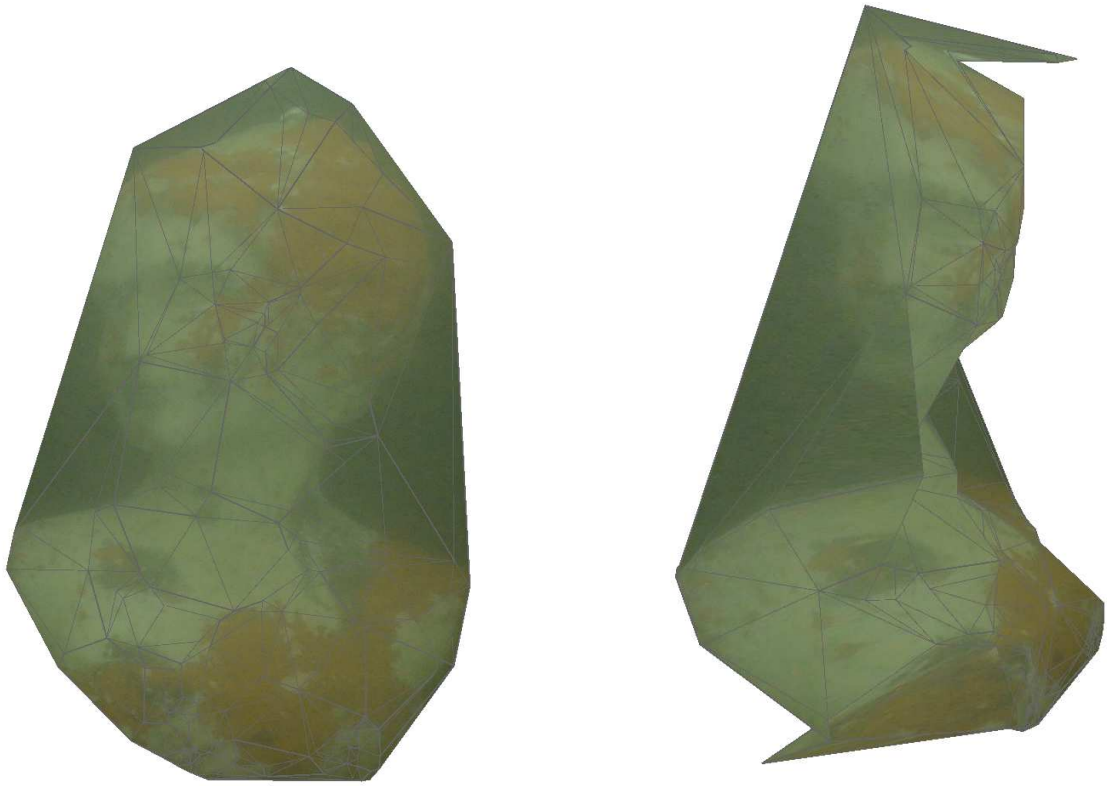
Nous avons ensuite travaillé avec les photos du buste de la figure 2.4. La figure 7.5 représente le modèle 3D texturé métrique du buste. Il a été créé en utilisant la méthode de

reconstruction stéréoscopique. La figure 7.6 est issue des mêmes images mais la reconstruction a été obtenue à partir de la méthode non-calibrée, c'est-à-dire la méthode où la transformation rigide entre les deux caméras n'est pas connue. Les résultats sont visuellement similaires et nous ne sommes pas capables de dire quel modèle a le bon facteur d'échelle.



(a) Image de face (b) Image de profil

FIGURE 7.5 : Modèle 3D texturé métrique du buste immergé dans le Rhône.



(a) Image de face (b) Image de profil

FIGURE 7.6 : Modèle 3D texturé non métrique du buste immergé dans le Rhône.

Les modèles 3D obtenus ressemblent visuellement aux objets, mais pour les valider, nous devons vérifier la taille des objets. Pour cela, nous allons mesurer certaines distances caractéristiques de nos modèles 3D et les comparer soit au modèle numérisé [1] soit aux valeurs mesurées sur l'objet. Le tableau 7.1 regroupe ces résultats. Dans des conditions de prises de vues difficiles (expérimentations dans le Rhône) nous arrivons à avoir un modèle 3D métrique et texturé avec seulement 3% d'erreur de mesure. Lorsque les conditions de prise de vue sont meilleures (expérimentations en étang salé) nous obtenons une erreur de seulement 0.5% à 1% (tableau 7.2).

1. Le modèle numérisé a été scanné avec un scaner Minolta Vivid 9I de précision 3D xyz de 0.1 mm



(a) Image de face (b) Image de profil

FIGURE 7.7 : Modèle 3D texturé d'une amphore immergée dans l'étang de Thau.

Point	Modèle scanné (mm)	Modèle reconstruit (mm)	Erreurs (mm)	Erreurs en %
Nez - Œil	44.5	45.2	0.7	1.6 %
Cou - Tête	199.3	204.7	5.4	2.7 %
Épaules/Épaules	189.9	195.5	5.6	2.9 %

TABLE 7.1 : Mesure de certaines distances caractéristiques du buste des expérimentations dans le Rhône.

Point	Objet réel (mm)	Modèle reconstruit (mm)	Erreurs (mm)	Erreurs en %
Anse - Col	104	104.8	0.8	0.7 %
Anse/Anse	236	237.7	1.7	0.7 %
Pied - Col	767	775.2	8.2	1.0 %
Concrétion/Concrétion	186	187.0	1.0	0.5 %

TABLE 7.2 : Mesure de certaines distances caractéristiques de l'amphore des expérimentations en étang salé.

Ce chapitre a permis de créer les modèles 3D texturés et précis d'objets archéologiques sous-marins à partir d'images issues d'une paire stéréoscopique. Maintenant que ces modèles 3D sont obtenus, il faut passer à la création de la carte acoustique sur laquelle ils seront fusionnés.

Quatrième partie

Fusion des différentes cartes

CHAPITRE

8 Création de la carte acoustique 3D

8.1 Matériels acoustiques sous-marins

En environnement sous-marin, les capteurs acoustiques sont largement utilisés pour de multiples applications. Dans ce paragraphe, nous allons nous focaliser sur les technologies de sonars servant pour la cartographie.

Les sonars créent des ondes acoustiques qui sont transmises par l'eau. Si un objet se trouve sur sa trajectoire, l'onde est réfléchie et retourne vers le sonar. Il est alors possible de mesurer l'intensité du signal et de déterminer la distance à laquelle se trouve l'objet. L'excellente propagation du son dans l'eau permet aux ondes acoustiques de parcourir plusieurs centaines de mètres en ne subissant qu'une légère perte de puissance. Cependant, il existe un compromis entre distance de détection et résolution ; plus les fréquences acoustiques sont basses, plus la distance de détection sera grande mais cela se fera aux dépens de la résolution qui s'améliore en fonction de l'augmentation de la fréquence du signal. L'avantage principal de ce type de technologie est qu'elle n'est pas ou très peu sensible aux perturbations de type particules en suspensions.

Dans le paragraphe qui suit, nous allons présenter les principales technologies de sonar permettant de réaliser la cartographie du fond marin. Elle se décompose en quatre catégories :

Sonars mono-faisceau.
Utilisé depuis les années 1920, le sonar mono-faisceau est la plus vieille et la plus basique

des technologies de sonar. Comme son nom l'indique, il est composé d'un seul transducteur qui permet d'envoyer un signal étroit (5° à 15°). Il sert généralement pour mesurer la profondeur sous un bateau ou pour détecter un objet dans une zone précise. Cependant, il permet de réaliser des analyses plus complexes du signal de retour, comme la force de retour de l'écho, ou l'étude du second retour et il peut donner des informations sur le type de fond (sable, roche, etc.). La résolution verticale varie en fonction des fréquences acoustiques utilisées, mais elles sont généralement comprises entre 5cm et 50cm. La résolution horizontale dépend de nombreux paramètres comme l'angle de mesure, la profondeur, etc. mais elle est généralement voisine du mètre. Ce type de sonar n'est pas le plus recommandé pour cartographier une zone complète car l'utilisation d'un seul transpondeur rendrait la tâche très longue.

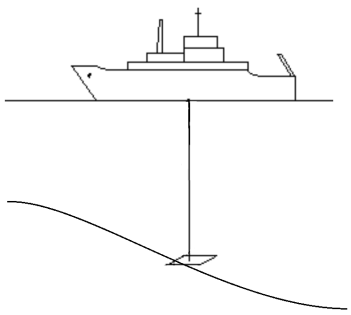


FIGURE 8.1 : Représentation schématique d'un relevé avec un sonar mono faisceau.

Sonars à balayage.
Ce capteur est composé d'une partie mécanique qui permet d'orienter la mesure sur différents angles. Généralement, le secteur couvert peut être de quelques degrés à une rotation complète (360°). Cela en fait un capteur très intéressant pour la détection d'obstacles. S'il est utilisé en "down looking", il permet de scanner le fond pour réaliser une carte. L'inconvénient de ce type de montage est qu'il faut prendre en compte le temps de balayage mécanique pour créer la carte. En effet, même s'il permet de couvrir une zone de 180°, la carte peut comporter des trous si la vitesse de déplacement est trop grande.

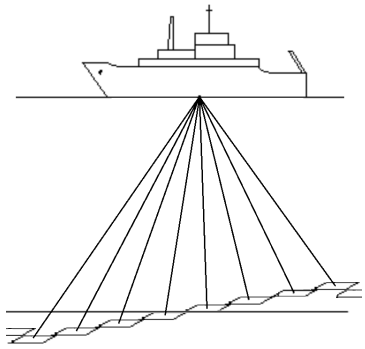


FIGURE 8.2 : Représentation schématique d'un relevé avec un sonar à balayage.

Sonars multifaisceaux.
Pour compenser ces problèmes de temps de balayage et pour répondre aux besoins de plus en plus grands de cartographie des fonds marins, le sonar multifaisceaux a fait son apparition dans les années 1970. Comme son nom l'indique, il est composé de plusieurs émetteurs/récepteurs acoustiques. Pour ceux ayant les résolutions les plus élevées, les transducteurs sont placés à quelques dixièmes de degré les uns des autres et peuvent couvrir un angle de 160°. En plaçant ces sonars perpendiculairement au sol, ils permettent de réaliser plusieurs mesures de profondeur en un seul tir acoustique. Dans des applications de cartographie, cela permet d'obtenir la représentation d'une tranche (voir figure 8.3). Au fur et à mesure que le bateau avance, d'autres tranches sont scannées et la carte du fond se dessine. Ce type de technologie est particulièrement adapté pour réaliser une carte du fond marin.

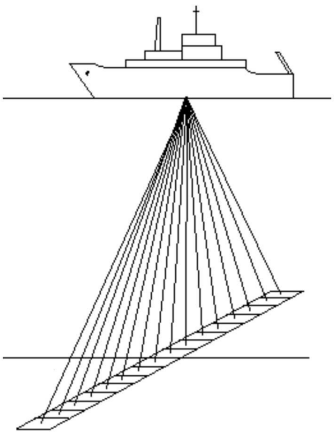


FIGURE 8.3 : Représentation schématique d'un relevé avec un sonar multifaisceaux.

Si le sonar est utilisé en position avant (fig. 8.4), il est généralement appelé sonar "Forward-Looking". Il permet de réaliser des images acoustiques de grande résolution. On le retrouve depuis peu sous le nom de "caméras acoustiques".

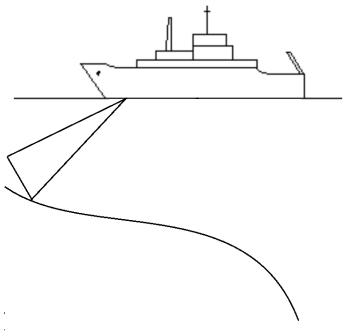


FIGURE 8.4 : Représentation schématique d'un relevé avec un sonar "Forward Looking".

Sonars latéraux.
Le sonar latéral est l'un des capteurs acoustiques les plus connus. Généralement les sonars latéraux sont montés directement sur ce qu'on appelle un poisson tracté (fig. 8.5). Il est composé de deux émetteurs/récepteurs mais ils sont placés de part et d'autre du sonar. En plaçant le sonar le plus proche du fond, l'onde acoustique est émise avec une incidence rasante et intercepte le fond sur une bande très fine. L'écho reçu au cours du temps sera la représentation de la présence des objets et obstacles présents sur le fond. Ce signal, enregistré latéralement à la direction d'avancée du sonar, est juxtaposé aux signaux successifs déjà obtenus par le sonar pendant son avancée, constituant ainsi, ligne après ligne, une "image acoustique du fond".

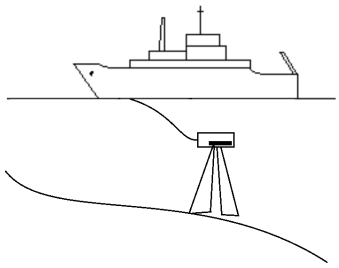


FIGURE 8.5 : Représentation schématique d'un relevé avec un sonar latéral.

8.2 Expérimentations et résultats

Pour réaliser la carte acoustique globale du chantier de fouille archéologique, nous avons utilisé un *sonar multifaisceaux* "837 DeltaT Profiling" de chez Imagenex (fig. 8.6). Il fonctionne à une fréquence de 260 kHertz et permet de réaliser une cartographie 3D du fond marin pour des profondeurs allant de 5 à 100 mètres. La résolution fournie par le constructeur est de 0.2% de la profondeur, c'est-à-dire 1 centimètre d'erreur pour une profondeur de 5 mètres et 20 centimètres pour la distance de détection maximum. Le sonar utilise 480 faisceaux sur 120 degrés d'angle de vue. La résolution angulaire est donc de 0.25 degré. La faible consommation énergétique de ce sonar permet de l'utiliser directement sur un engin sous-marin autonome.

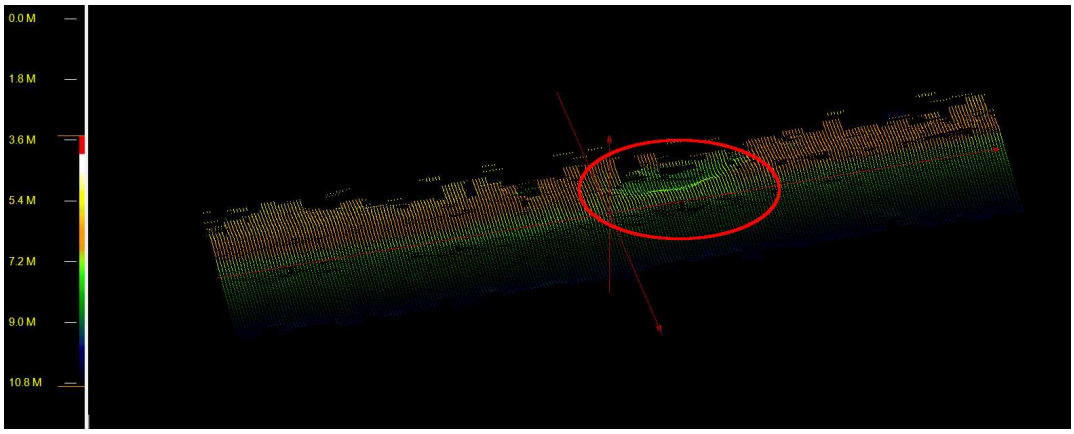


FIGURE 8.6 : Sonar multifaisceaux DeltaT Profiling

Les premières expérimentations réalisées à l'aide de ce sonar ont été effectuées dans le Rhône en Arles sur le même chantier de fouille archéologique que pour les expérimentations avec la paire stéréoscopique (voir 2.1.2). Nous avons fait plusieurs passages au-dessus de l'épave d'une péniche romaine se trouvant à environ 7 mètres de profondeur. Ces expérimentations ont permis de prendre en main le sonar et les logiciels associés. La carte 3D fournie peut être affichée en temps réel ou hors ligne.



(a) Premier passage



(b) Deuxième passage

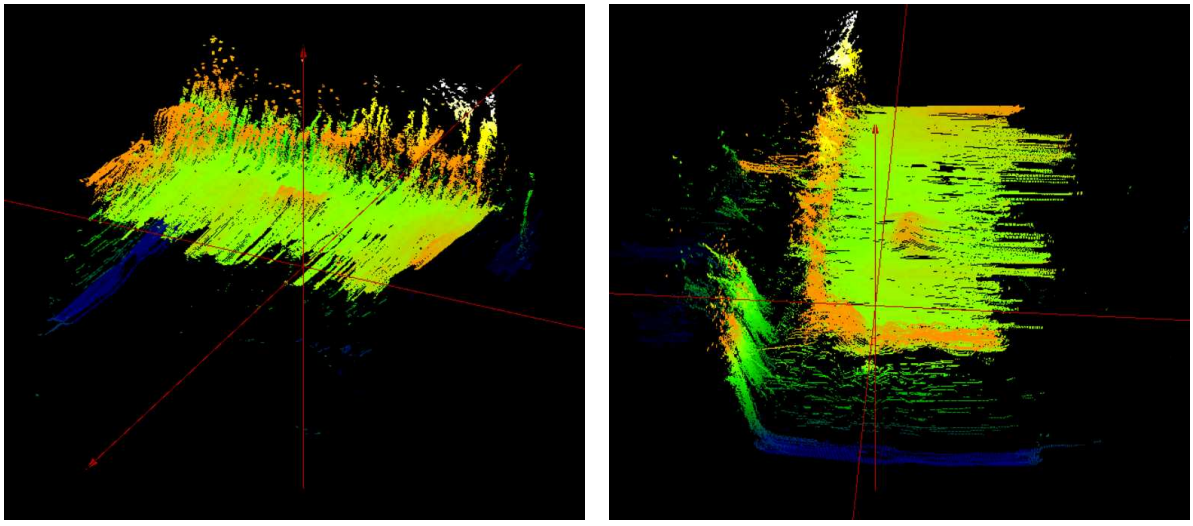
FIGURE 8.7 : Carte 3D sonar de l'épave romaine en Arles.

Lors de ce relevé, à deux reprises, nous apercevons l'épave sur la carte acoustique 3D. L'épave est identifiée par le cercle rouge sur les figures 8.7(a) et 8.7(b).

Nous pouvons observer que la taille de l'épave n'est pas la même sur les deux figures. Ce problème de métrique est introduit lors de la reconstruction en 3D de la carte. En effet, le logiciel met bout à bout toutes les "tranches" fournies par le sonar. La difficulté ici est de connaître la distance (ou le "pas") séparant les tranches à chaque instant. Si, lors du relevé, la vitesse de déplacement du sonar est constante, le pas sera identique pour toute la carte. En pratique, même pour un système autonome, cette contrainte est très difficile à respecter. Le "pas" varie alors légèrement entre chaque "tranche".

Dans notre cas, le bateau tractant le sonar naviguait à une vitesse de 3 noeuds. Cependant, cette donnée est imprécise puisque une erreur est ajoutée par les variations de courant dans le Rhône. La puissance du courant n'est forcément pas la même lorsque le bateau réalise un relevé sonar en descendant ou en remontant le fleuve. La vitesse exacte du bateau étant inconnue, la distance entre deux "tranches" ne peut pas être estimée. Pour résoudre, ce problème plusieurs solutions s'offrent à nous. La première consiste à utiliser une centrale inertielle. Elle permet d'estimer la vitesse du sonar et d'en déduire le déplacement, donc le "pas". Une autre solution consiste à utiliser un GPS qui fournit la position du sonar en permanence. Ces deux solutions nécessitent l'instrumentation du sonar. Une dernière solution repose sur la connaissance de la taille d'un objet de référence sur la carte, mais elle implique de prendre pour hypothèse que la vitesse de déplacement du sonar est constante.

Pour obtenir un jeu de données où les données globales (sonar) et locales (vidéo) correspondaient, nous avons réalisé un relevé sonar de la zone où nous avions immergé une amphore dans l'étang de Thau (voir section 2.1.3). L'amphore était posée sur le pont à l'arrière d'une épave de bateau. La figure 8.8 permet de visualiser la carte sonar de l'épave où nous voyons apparaître un objet (l'amphore) en plein milieu du pont. Pour corriger la métrique de la carte 3D, nous avons instrumenté le sonar en y ajoutant une centrale inertielle.



(a) Vue de profil (b) Vue de dos

FIGURE 8.8 : Carte 3D sonar de l'épave de l'étang de Thau avec une amphore sur le pont.

Les cartes sonar 3D utilisées sont brutes, nous entendons par là qu'il n'y a aucun traitement pour améliorer la précision ou le rendu. Par exemple, un filtre est couramment appliqué pour boucher les trous de la carte en ajoutant des points au nuage.

La figure 8.8 met en avant un autre phénomène qui est généralement corrigé. Nous pouvons voir sur les bords de la carte que toutes les tranches ne se terminent pas en même temps. En effet, lorsque nous réalisons le relevé sonar, le capteur ne doit subir aucune rotation ni translation autre que le mouvement d'avance pour obtenir une carte précise. Lorsqu'il subit une perturbation (une vague, une accélération de l'engin le tractant, etc.) celui-ci réalise un relevé dans une nouvelle position et le repère de référence n'est plus respecté. Pour corriger ce phénomène, il suffit d'utiliser une centrale inertielle qui fournit les rotations et les translations subies par le sonar et, ainsi, les perturbations peuvent être compensées. Dans notre application, nous avons utilisé une carte brute et donc moins précise car la carte sonar (carte globale) n'a pas besoin d'une résolution importante. Il faut pouvoir déterminer la présence d'un objet mais pas forcément sa nature ni sa forme car c'est le rôle de la carte locale.

Les logiciels fournis par le fabricant du sonar ne permettant pas d'obtenir la carte 3D dans un format non propriétaire, nous avons extrait la carte à partir des trames fournies par le sonar. La figure 8.9 représente la même carte que la figure 8.8 extraite à partir des trames et visualisée sous Meshlab [MeshLab].

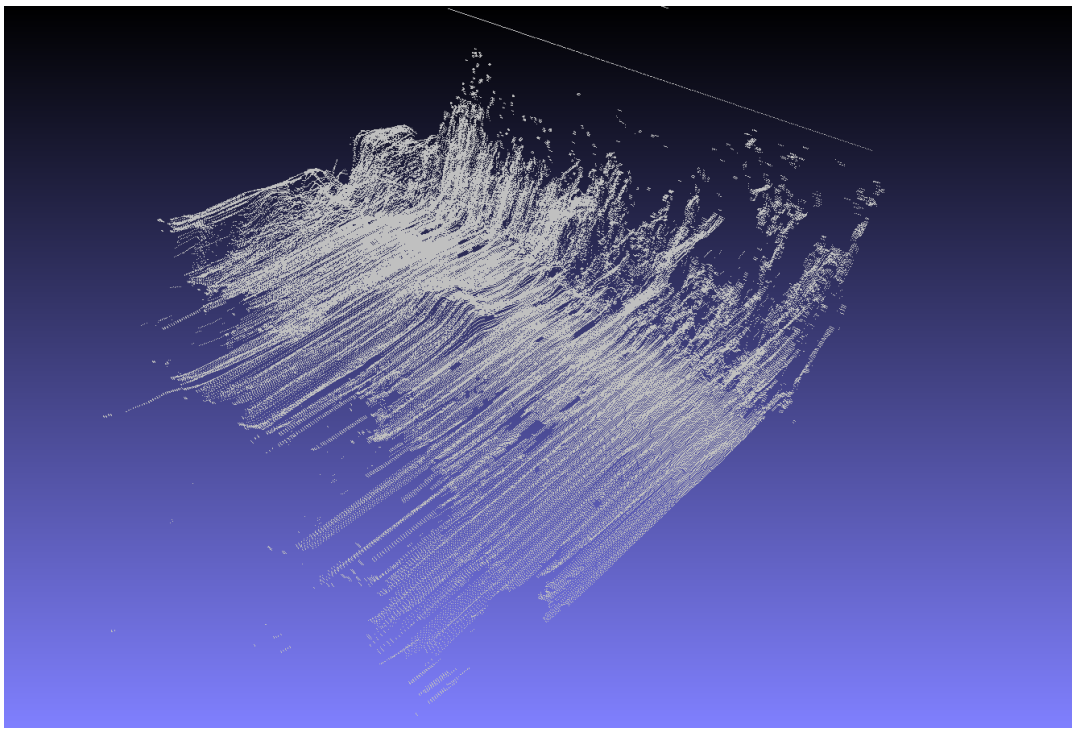


FIGURE 8.9 : Carte 3D sonar de l'épave de l'étang de Thau visualisée sous MeshLab.

CHAPITRE

9 Fusion de la carte acoustique globale et du modèle vidéo local

La fusion consiste à faire de deux cartes, une carte unique plus fiable et plus complète. Dans la section qui suit, nous allons présenter quelques approches utilisées pour fusionner des données acoustiques et vidéo en milieu sous-marin. La difficulté principale de ce travail est la multi-modalité des données. En effet, les deux cartes utilisées ne possèdent pas la même résolution, la même précision de données, le même type d'informations, ni le même échantillonnage. Toutes ces différences font qu'il existe un grand nombre de méthodes de fusion.

9.1 État de l'art

Dans cette section, nous allons présenter différentes méthodes existant dans la littérature pour réaliser la fusion de données acoustiques avec des données vidéo. Nous présenterons deux méthodes de recalages de modèle géométrique 3D.

9.1.1 Fusion de données acoustiques et vidéo

La fusion multimodale est un domaine très actif dans la communauté scientifique de la robotique mobile comme dans la communauté sous-marine. De nombreuses études ont été réalisées pour fusionner divers types d'informations. En robotique mobile, des capteurs de vision couplés avec des LIDAR (Laser Imaging Detection and Ranging) sont généralement utilisés, mais cette technologie est difficilement applicable en milieu sous-marin

à cause des différents bruits qu'il introduit (voir section 2.2). Le capteur le plus souvent utilisé en environnement sous-marin est le sonar car il n'est que très peu sensible aux perturbations sous-marines. Nous allons présenter différentes approches de fusion utilisant les deux capteurs les plus répandus en milieu sous-marin : les sonars et les systèmes visuels.

La nature complémentaire de ces deux capteurs rend leur utilisation simultanée très intéressante mais cela pose des problèmes de fusion de données. En effet, le capteur acoustique et le capteur visuel fournissent des informations à des niveaux d'échelle différents et ne sont pas sensibles aux mêmes perturbations. Une caméra optique est capable de donner des images haute résolution tandis que les capteurs sonars fournissent généralement des cartes bathymétriques de résolution beaucoup plus faible. Il est intéressant de souligner que, pour construire une carte bathymétrique, la majorité des sonars a besoin d'enregistrer des données supplémentaires comme leur position ou les mouvements qu'ils subissent. Le problème ne vient pas de l'incapacité des sonars à produire des données de haute résolution, mais de la limitation de la précision de navigation.

Pour s'affranchir de ce problème, Drap propose [Drap *et al.*, 2011] une approche utilisant un sonar fixé à un trépied équipé d'un système pan-tilt. Le système est posé sur le fond pour réaliser une image sonar 3D haute résolution d'une partie de la scène, puis il est déplacé pour finir les parties de la scène non scannées. Ces différentes cartes acoustiques haute résolution sont assemblées en utilisant un algorithme ICP puis les informations visuelles y sont fusionnées. L'avantage majeur de ce système est qu'il donne une carte haute résolution quelles que soient les conditions sous-marines, mais il impose un temps de numérisation de la scène important. Cela peut poser des problèmes d'autonomie des plongeurs ou des engins autonomes.

Singh [Singh *et al.*, 2000] a proposé l'un des premiers travaux visant à fusionner des données provenant des modalités complémentaires (optique et acoustique) pour obtenir des représentations plus précises des fonds marins. La méthode présentée combine des informations de haute résolution obtenues à partir d'une méthode de photomosaïques avec des cartes acoustiques de faible résolution. Les données de cette carte sont interpolées pour correspondre à la résolution de la mosaïque photo. Des points de correspondances sont identifiés manuellement sur les deux cartes et la carte photo est modifiée afin de coller à la carte acoustique.

Deux ans plus tard, il [Singh *et al.*, 2002] décrit une méthode pour fusionner des données vidéo 3D avec des données issues d'un sonar mono-faisceau. Le modèle 3D vidéo est

reconstruit à un facteur d'échelle prêt à partir d'une méthode de "Structure from Motion" (SFM). La SFM se base sur le mouvement de la caméra pour obtenir la structure 3D, or si ce mouvement n'est pas connu parfaitement et que l'erreur s'accumule, le modèle 3D accumulera ces erreurs. Les auteurs proposent d'ajouter de nouveaux capteurs pour connaitre parfaitement le mouvement et ainsi réduire les erreurs de reconstruction. Ils ont alors intégré ces informations à la méthode SFM, ce qui a grandement amélioré les résultats.

Williams [Williams et Mahon, 2004] présente un algorithme capable de reconstruire un environnement naturel en utilisant des données vidéo et acoustiques. Il a développé une méthode de SLAM (Simultaneous Localization And Mapping)afin de pouvoir déplacer un engin autonome dans un environnement non structuré en utilisant les caractéristiques naturelles du milieu. Il utilise un sonar à balayage fixé sur le robot et crée la carte acoustique du fond lorsque le robot se déplace. La caméra est fixée juste à côté du sonar et la transformation rigide entre les deux capteurs est connue. Cela permet à l'auteur de projeter la carte acoustique dans l'image 2D et de rechercher des objets à fort contraste visuel comme amer. Une fois les amers identifiés, ils sont suivis directement dans l'image. L'utilisation des deux types d'informations permet d'améliorer les résultats du SLAM au niveau cartographie et localisation. La carte 3D est générée par le sonar et les données visuelles sont projetées dessus. La figure 9.1 permet de visualiser la carte obtenue.

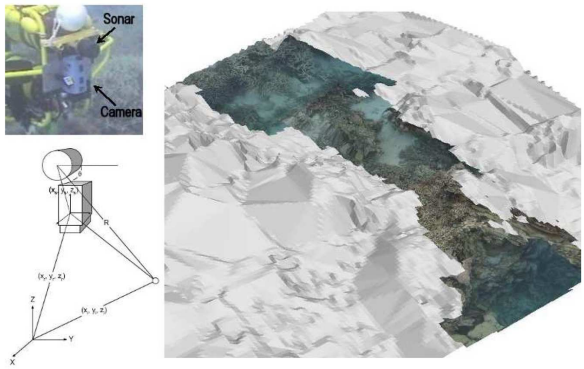


FIGURE 9.1 : Résultats de fusion de données acoustiques et visuelles obtenus par Williams [Williams et Mahon, 2004].

Fusiello [Fusiello *et al.*, 1999; Fusiello et Murino, 2000] fusionne des données acoustiques et visuelles afin d'améliorer la compréhension des scènes sous-marines pour les

pilotes de ROV (Remotly Operated Vehicles). Pour créer un modèle 3D acoustique précis, une caméra acoustique haute résolution est utilisée. La fusion entre la vidéo et le modèle 3D acoustique est réalisée à l'aide de caractéristiques géométriques et du modèle de l'objet observé. Dans [Fusiello et Murino, 2004], cette méthode est améliorée pour être utilisée sur des plates-formes pétrolières. Après plusieurs étapes de pré-filtrage, le modèle acoustique est recalé sur un modèle de référence (déjà connu) à l'aide d'un algorithme d'IPC. Les données visuelles sont recalées sur cet objet de référence à l'aide d'une méthode d'estimation de la pose développée par Lowe [Lowe, 1991]. Une fois les deux jeux de données alignées sur le modèle de référence, la transformation rigide entre les deux modalités (acoustiques et visuelles) peut être estimée. Cette méthode permet de contourner le problème de la fusion multimodale acoustique/vision mais nécessite la connaissance du modèle de l'objet observé.

Dans une autre optique, Negahdaripour et Sekkati [Sekkati et Negahdaripour, 2006; Negahdaripour *et al.*, 2007] ont proposé une nouvelle approche de reconstruction 3D basée sur le déploiement de caméras optiques et acoustiques en configuration stéréo. L'objectif est d'utiliser le chevauchement des vues de l'image stéréo opti-acoustique, afin d'améliorer la précision de la reconstruction 3D dans des conditions de turbidités élevées. La configuration stéréo leur permet d'utiliser la géométrie épipolaire qui facilite un grand nombre d'opérations comme la mise en correspondance des points sonars avec les points acoustiques. Les capteurs de la paire stéréo étant différents, ils ont été obligés d'adapter la géométrie épipolaire à des modèles de projections asymétriques. La problématique de la fusion des données est directement intégrée à la géométrie épipolaire et au calibrage du système stéréo.

Une étude plus récente réalisée par Roman [Roman *et al.*, 2010] propose une nouvelle approche qui consiste à fusionner des données sonars haute résolution avec un modèle 3D visuel obtenu par reconstruction par lumière structurée. Cette méthode est très robuste dans des mauvaises conditions de visibilité mais elle n'apporte pas d'informations qualitatives.

9.1.2 Recalage de modèles géométriques 3D

Le recalage de modèles géométriques 3D consiste à prendre deux nuages de points en 3D représentant le même objet et à trouver la transformation correcte pour les aligner afin

de ne faire plus qu'un seul modèle plus complet. Dans cette section, nous allons présenter les deux méthodes utilisées pour réaliser le recalage des deux cartes.

Méthode directe.
La première méthode mise en place est une méthode de recalage directe [Sabata et Aggarwal, 1991]. Il est possible de déterminer directement les paramètres de la transformation reliant les deux jeux de données en résolvant le système suivant :

$$A = R.B + t \tag{9.1}$$

Où A et B sont les coordonnées de deux points identiques dans les deux modèles 3D, R est la matrice de rotation et t le vecteur de translation. Ce système compte 12 inconnues ; 9 pour la matrice de rotation et 3 pour le vecteur de translation. Lorsque nous prenons une paire de points identiques dans les deux nuages, nous obtenons trois équations indépendantes. Donc, à partir de la connaissance de quatre paires de points, nous pouvons obtenir la transformation. Des contraintes sont ajoutées pour respecter la taille du nuage de point. Cette méthode n'utilise que quatre points sur l'ensemble du jeu de données, ce qui rend les résultats entièrement dépendants de la sélection des points. Si une erreur est faite sur une paire de points, la transformation trouvée présentera une forte erreur.

Algorithme ICP (Iterative Closest Point).
L'algorithme ICP (Iterative Closest Point) introduit simultanément par Besl [Besl et McKay, 1992] et Zhang [Zhang, 1992, 1994] présente une solution itérative très connue répondant au problème de recalage de modèle 3D.Cette méthode est générale et permet de recaler tous les types de primitives géométriques (nuage de points, segments, courbes 2D, ensemble de points triangulés, surface 3D, etc.). A partir d'un recalage approximatif entre deux modèles, l'algorithme estime une solution plus fine en réduisant itérativement l'erreur entre les modèles. La méthode consiste à trouver un ensemble de correspondances entre les deux nuages de points et à estimer la transformation rigide les séparant. L'algorithme itératif passe d'une itération à une autre en minimisant la distance euclidienne entre les points du premier modèle et leurs plus proches voisins dans le second modèle. Cet algorithme est reconnu comme étant robuste pour l'estimation de la transformation rigide entre les deux jeux de données. Cependant, on lui reproche souvent de converger vers un minimum local. La qualité et la précision du recalage final dépendent fortement de la transformation initiale choisie. La version de Besl [Besl et McKay, 1992] nécessite que

les deux modèles soient homologues (le modèle 2 doit être un sous-ensemble du modèle 1). La méthode de Zhang est plus tolérante aux erreurs de mesure et aux occultations. L'algorithme ICP de base se décompose de la façon suivante :

1. **Sélection des points :** pour chaque point du modèle 1, on cherche son point le plus proche dans le modèle 2.
2. **Définition de la métrique d'erreur :** calcul de la somme de la distance euclidienne entre chaque point ayant un correspondant.
3. **Estimation de la transformation :** calcul de la transformation minimisant le critère d'erreur
4. **Calcul du critère d'arrêt :** application de la transformation estimée sur le modèle 2 et calcul du critère d'erreur. S'il est supérieur au seuil désiré, on retourne à 1, sinon arrêt.

Cette méthode a connu un nombre impressionnant d'améliorations et d'adaptations à des problèmes spécifiques. Tous les niveaux de l'algorithme ont vu une modification proposée. Par exemple, certains auteurs ont choisi de ne pas utiliser tous les points de la structure mais seulement un sous-échantillonnage uniforme ou un sous-échantillonnage aléatoire. Le critère de la métrique d'erreurs s'est vu aussi apporter de nombreuses modifications. Johnson [Johnson et Kang, 1997] a proposé d'ajouter la couleur RGB aux informations de distance. D'autres auteurs ont proposé d'ajouter des informations sur la structure intrinsèque des modèles (normales, courbures principales, etc.), ou encore sur des éléments invariants aux transformations rigides (la courbure, le moment d'ordre 2, etc.).

9.2 Discussions et présentation de notre approche

Comme nous venons de le voir dans la section 9.1, il existe de nombreuses méthodes pour mettre en correspondance des données vidéo avec des données sonar.

L'approche proposée par Singh [Singh *et al.*, 2000] utilisant une carte acoustique de faible résolution et des données visuelles n'est pas envisageable pour plusieurs raisons. La première est que la précision de la carte acoustique est très faible ; de plus, l'interpolation des données ajoute encore plus d'incertitudes à la carte. La méthode manuelle d'alignement des modèles n'est pas précise du tout car elle est entièrement dépendante du choix des points.

L'idée de corriger la reconstruction par le mouvement (SFM) à partir de nouveaux capteurs [Singh *et al.*, 2002] était intéressante mais la fiabilité des données utilisées et la faible résolution du modèle acoustique ne permettent pas d'avoir un point de vue objectif sur cette approche. De plus la méthode de fusion n'est pas détaillée dans l'article. Si nous supposons que Singh utilise le recalage manuel présenté précédemment, le modèle reste toujours aussi imprécis.

L'approche de Williams [Williams et Mahon, 2004] est intéressante sur le fait que l'utilisation des données est double. Cependant, il utilise un système spécifique et la connaissance de la position de la caméra par rapport au sonar est indispensable, or nous n'avons aucunement ce type d'information dans notre application. Pour les mêmes raisons, l'approche utilisant un système stéréo opti-acoustique de Sekkati et Negahdaripour [Sekkati et Negahdaripour, 2006; Negahdaripour *et al.*, 2007] ne peut être retenue.

Les résultats obtenus par Fusiello [Fusiello *et al.*, 1999; Fusiello et Murino, 2000, 2004] nécessitent la connaissance du modèle CAO (Conception Assistée par Ordinateur) de l'objet ou de la scène observée or, dans notre application, nous n'avons aucune connaissance a priori de la scène, ni des objets.

La carte acoustique 3D que nous utilisons est une carte brute, sans aucun traitement pour améliorer la précision ni le rendu. Nous savons donc, qu'en plus de posséder une faible résolution, les données présentes sur cette carte comporteront forcément des erreurs. Tout objet détecté n'aura pas une résolution assez importante pour que nous puissions avoir un a priori sur sa forme. A l'extrême opposé, la carte locale vidéo fournit des informations fiables sur la forme géométrique et la taille des objets. Une approche probabiliste basée sur une fusion de croyance n'est donc pas envisageable. La croyance en la carte acoustique étant très faible par rapport à la carte vidéo, seules les données vidéo seraient retenues.

Nous nous sommes appuyé sur ces conclusions pour choisir une approche par recalage de modèles géométriques. Pour réaliser la mise en correspondance de nos deux cartes, un algorithme d'ICP a été utilisé. Comme nous l'avons vu précédemment, il est basé sur un calcul itératif de minimisation de distances entre les deux modèles. Cette approche nous permet de nous affranchir des problèmes d'échantillonnage entre les cartes. En effet, l'algorithme cherchera la meilleure transformation entre les deux nuages de points. L'inconvénient majeur de ce type d'approche est que la convergence vers la solution globale n'est pas assurée. L'initialisation de l'algorithme est cruciale pour la précision de l'alignement des modèles. Dans la littérature, cette tâche est souvent décomposée en deux parties. La

première réalise un alignement approximatif des deux nuages de points. Ensuite, l'ICP sert à affiner le recalage des deux modèles et ainsi à augmenter la précision. Lorsque la position de la carte vidéo sur la carte sonar sera trouvée, les données sonar non cohérentes ne seront plus utilisées.

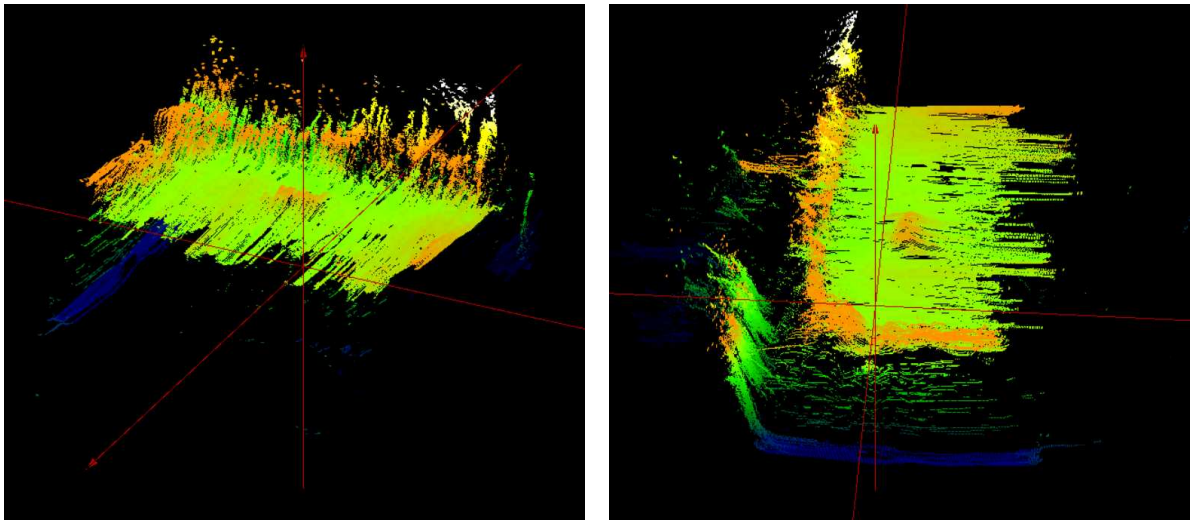
L'alignement approximatif des deux cartes a été réalisé à partir d'une méthode linéaire directe. Elle consiste à sélectionner manuellement au moins quatre points de correspondances sur chaque carte et à calculer la transformation existant entre les deux jeux de points. Une fois la transformation obtenue, elle est utilisée comme valeur d'initialisation de l'algorithme ICP.

9.3 Expérimentations et résultats

Comme introduit dans la section 9.2, la méthode de recalage se décompose en deux parties. La première consiste à trouver un alignement approximatif entre les deux cartes et la seconde permet un alignement plus fin à l'aide de la méthode ICP. Nous obtenons alors une carte acoustique complète de la zone avec le modèle 3D vidéo d'objets d'intérêts venant le compléter.

Pour ces expérimentations, nous avons utilisé les données obtenues pendant les manipulations dans l'étang de Thau. Lors de cette campagne, nous avons immergé une amphore sur le pont arrière d'une épave de bateau. Dans un premier temps, la carte sonar globale de la zone (fig. 9.2) nous a permis d'identifier un objet d'intérêt sur le pont de l'épave. Ensuite, nous sommes redescendu sur zone avec le système stéréoscopique pour réaliser des clichés de cet objet. Une fois le modèle 3D texturé de l'objet obtenu (fig. 9.3), le recalage des deux données a pu être réalisée. Afin de tester notre approche, nous avons réalisé deux cartes différentes à partir de deux modèles vidéo différents de l'amphore (Fig. 9.3 et fig. 9.4).

Pour cela, nous avons d'abord effectué le recalage approximatif des deux cartes en sélectionnant manuellement quatre points sur chaque carte. Ensuite, l'algorithme ICP a utilisé 100 points pour affiner le recalage. Le tableau 9.1 regroupe les erreurs d'alignements



(a) Vue de profil (b) Vue de dos

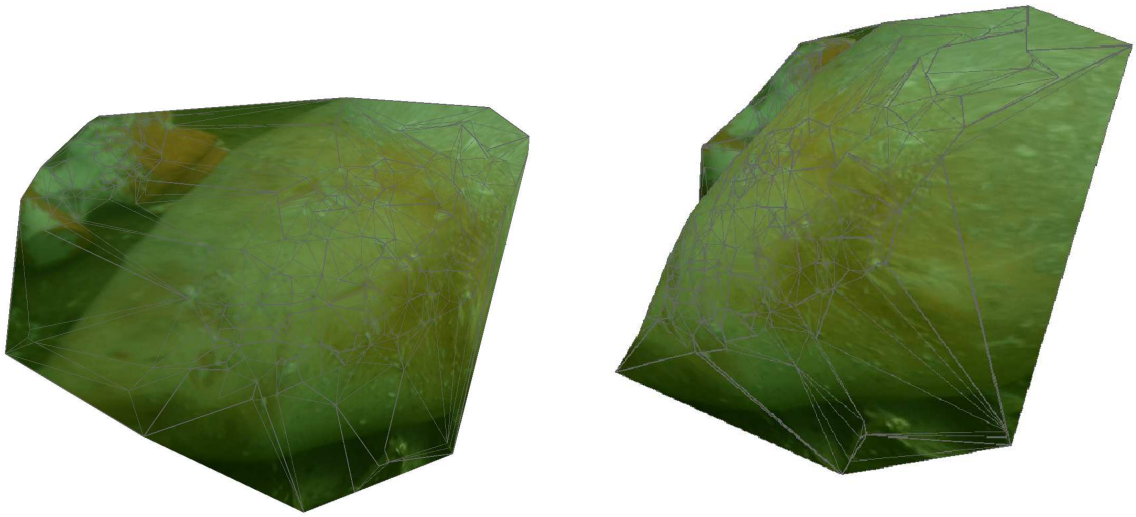
FIGURE 9.2 : Carte 3D sonar de l'épave de l'étang de Thau avec une amphore sur le pont.



(a) Modèle de face (b) Modèle de profil

FIGURE 9.3 : Premier modèle 3D texturé de l'amphore immergée dans l'étang de Thau.

(critère d'erreur de l'ICP) obtenues lors des deux étapes de recalage. Les valeurs obtenues paraissent importantes de prime abord mais il faut prendre en compte la taille des modèles (plusieurs mètres pour la carte sonar et quasiment un mètre pour l'amphore) ainsi que la différence d'échantillonnage qui entraîne forcément une erreur. Nous travaillons avec une carte acoustique brute, ce qui implique que la carte elle-même possède un grand nombre



(a) Modèle de face (b) Modèle de profil

FIGURE 9.4 : Deuxième modèle 3D texturé de l'amphore immergée dans l'étang de Thau.

d'erreurs, mais cela n'empêche pas d'avoir une position correcte du modèle vidéo. Avec 1 cm d'erreur d'alignement par point sur un modèle global de 7 mètres par 5 mètres, l'erreur est inférieure à 1%. De plus, l'échantillonnage de la carte sonar est de 1.5 cm, donc avec une erreur d'environ 1 cm, nous avons un décalage inférieur à un pas d'échantillonnage. L'apport de la méthode ICP par rapport au recalage manuel est important et permet dans notre cas de diviser l'erreur par trois.

	Erreur d'alignement manuel (cm)		Erreur d'alignement par ICP (cm)	
	Globale	moy. par point	Globale	moy. par point
Carte 1	294	2.94	97	0.97
Carte 2	316	3.16	123	1.23

TABLE 9.1 : Mesure des erreurs de recalage des modèles.

Les figures 9.5, 9.6, 9.7 et 9.8, 9.9, 9.10 permettent de visualiser les cartes finales obtenues. Les deux cartes réalisées correspondent aux conditions expérimentales que nous avions mises en place.

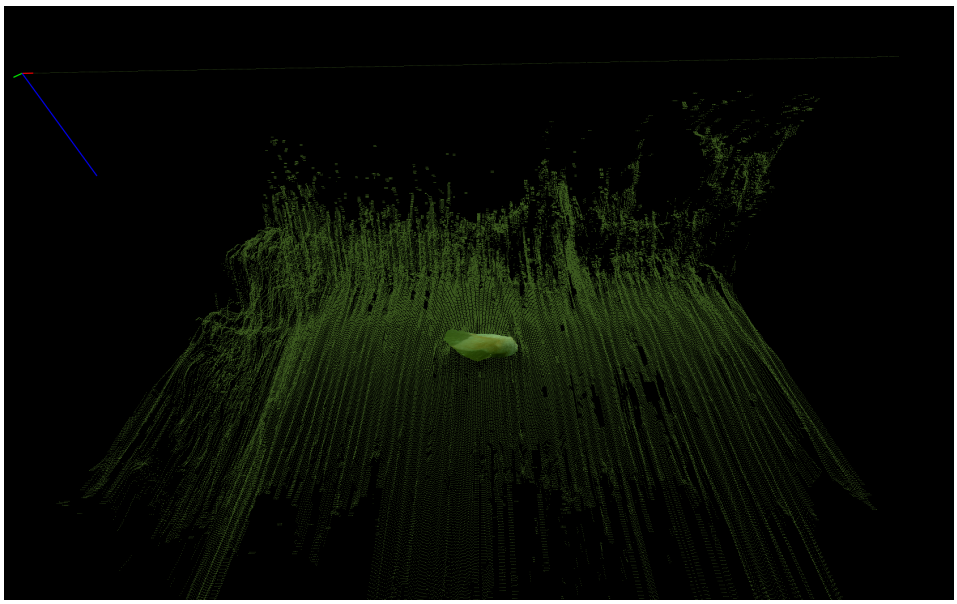


FIGURE 9.5 : Vue globale de la première carte 3D finale de l'épave de l'étang de Thau et de l'amphore.

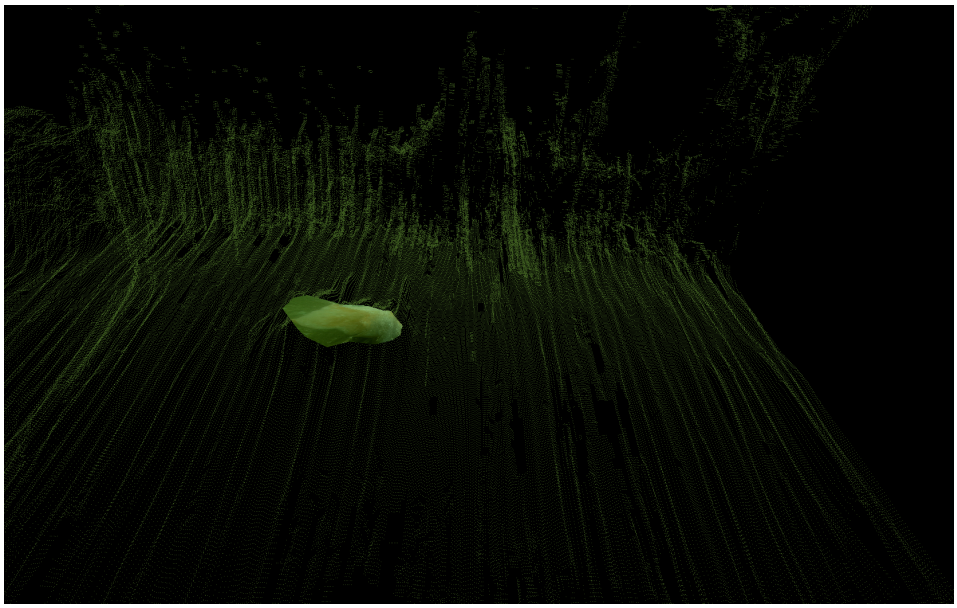
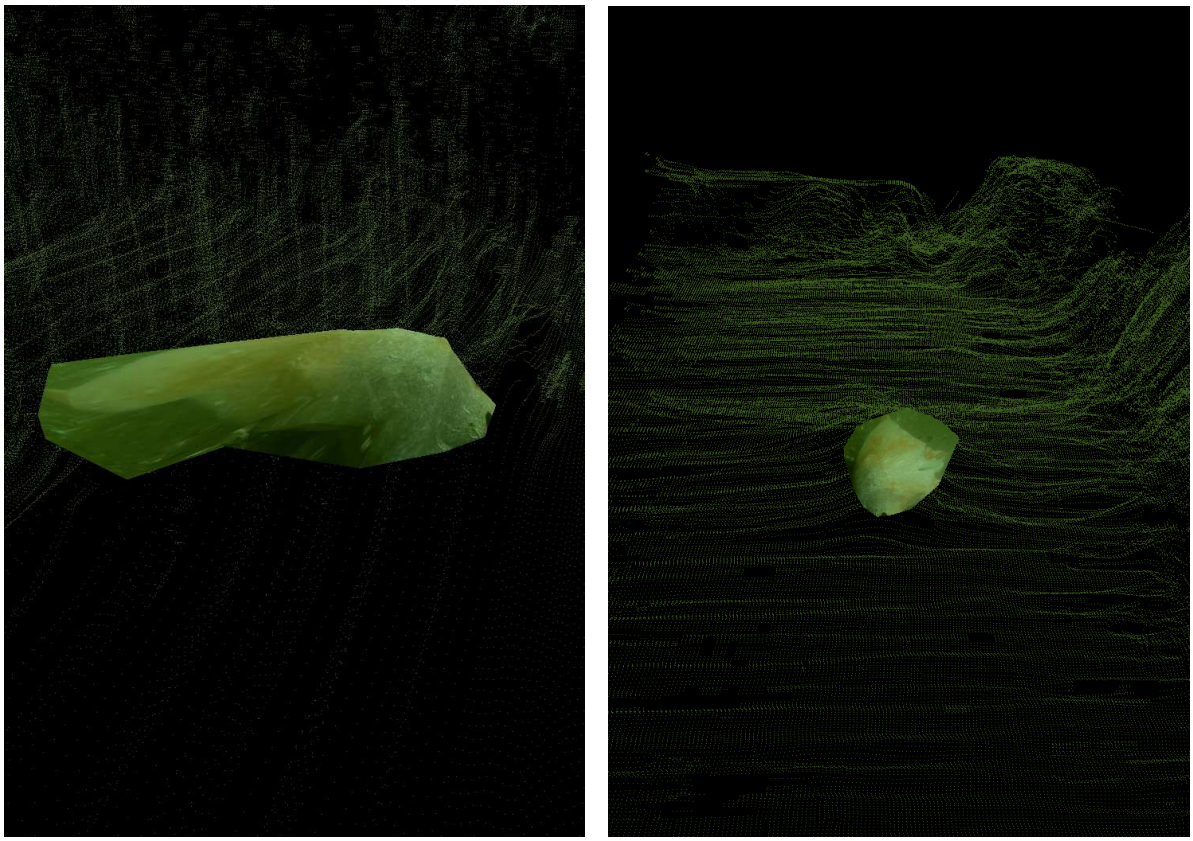


FIGURE 9.6 : Vue de la première carte 3D finale de l'épave de l'étang de Thau et de l'amphore.



(a) Vue de profil (b) Vue de face

FIGURE 9.7 : Vue zoomée de la première carte 3D finale de l'épave de l'étang de Thau et de l'amphore.

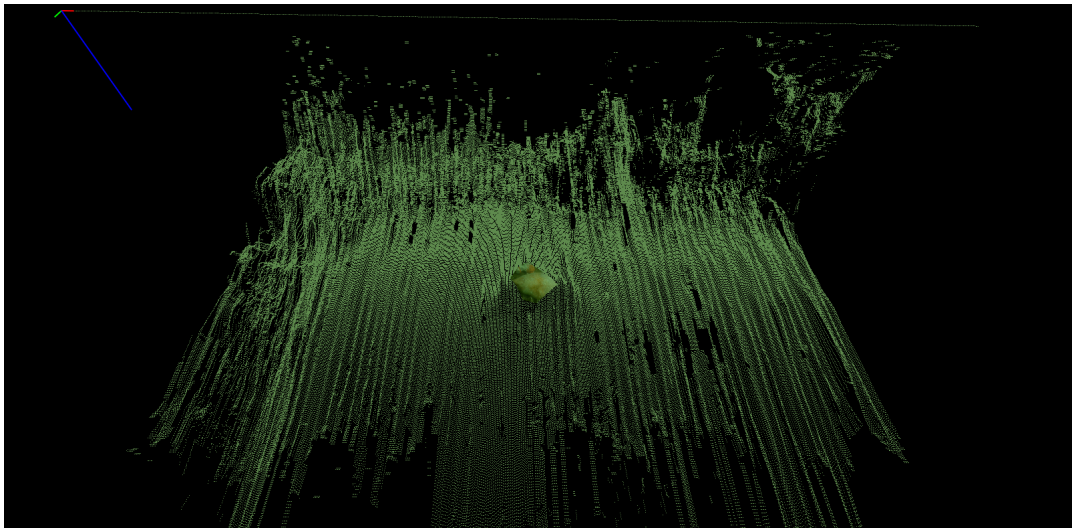


FIGURE 9.8 : Vue globale de la deuxième carte 3D finale de l'épave de l'étang de Thau et de l'amphore.

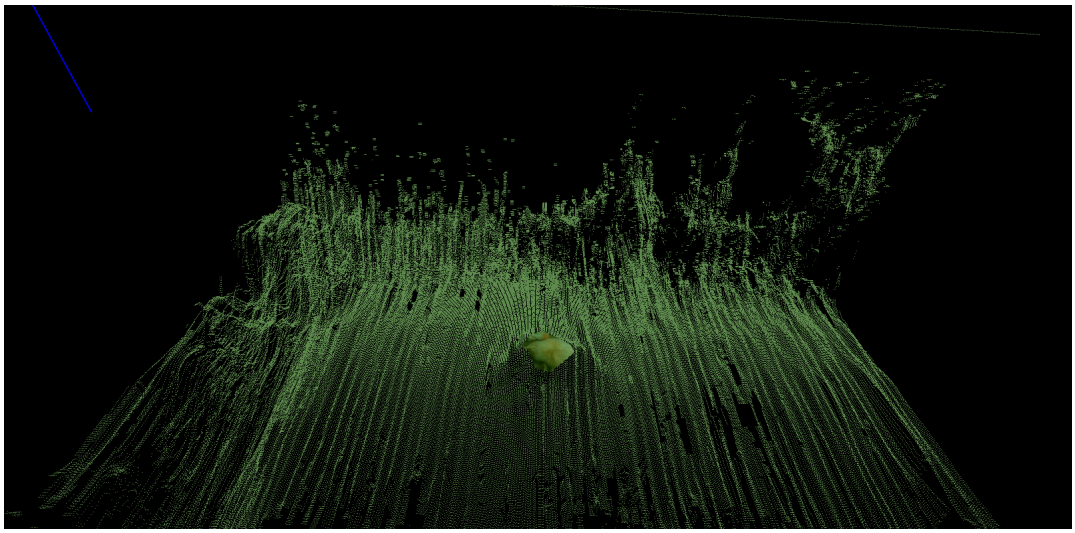
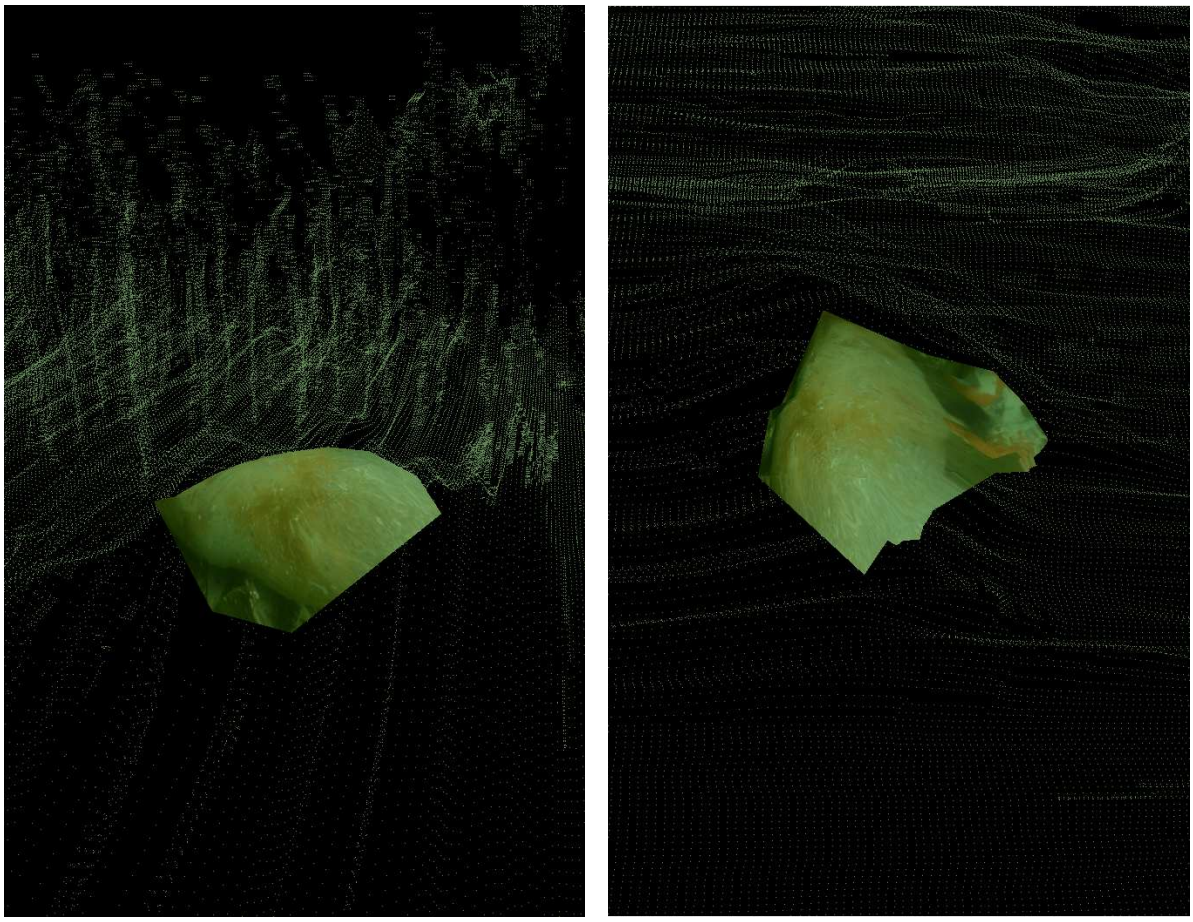


FIGURE 9.9 : Vue de la deuxième carte 3D finale de l'épave de l'étang de Thau et de l'amphore.



(a) Vue de profil (b) Vue de face

FIGURE 9.10 : Vue zoomée de la deuxième carte 3D finale de l'épave de l'étang de Thau et de l'amphore.

Cinquième partie

Conclusions

CHAPITRE

10 Conclusions et Perspectives

Ce manuscrit présente les travaux réalisés dans un contexte de cartographie sous-marine multimodale. Notre approche de cartographie est basée sur l'utilisation de deux types de capteurs pour mieux répondre aux fortes contraintes de l'environnement sous-marin qui augmentent les difficultés des problèmes pouvant être rencontrés en milieu terrestre. Les développements proposés sont organisés autour d'un capteur acoustique qui fournit une carte globale à faible résolution de la zone à cartographier et d'un capteur visuel stéréoscopique qui vient ajouter du détail à cette carte.

Le travail exposé dans ce document est articulé autour de trois problématiques. Tout d'abord, il a fallu prendre en compte les contraintes de l'environnement sous-marin pour réaliser la reconstruction 3D à partir des données visuelles. Nous avons ensuite réalisé la carte sonar 3D globale. La dernière étape consistait à réaliser la fusion des données 3D visuelles sur la carte acoustique.

Afin de réaliser la reconstruction 3D à partir des données visuelles stéréoscopiques, nous avons dû calibrer les appareils photos en prenant en compte les perturbations introduites par le milieu sous-marin. Pour réaliser un calibrage robuste, nous avons proposé un modèle et une procédure de calibrage permettant d'intégrer les problèmes dus au milieu sous-marin. Une série d'expérimentations ont permis de valider notre approche.

Après avoir exposé les différentes options s'offrant à nous pour réaliser le modèle 3D visuel, la méthodologie a été décomposée en plusieurs parties. Les étapes les plus importantes pour la précision de la structure 3D sont la détection de points d'intérêts dans les images stéréoscopiques, ainsi que la mise en correspondance de ces points entre les deux

images stéréo. Nous avons comparé la robustesse de deux méthodes de détections et d'appariement de points couramment utilisées en surface : la méthode de Harris&Correlation et la méthode de SIFT. Nous avons prouvé que la méthode de SIFT, étant reconnue comme étant la plus performante en surface, obtenait de moins bonnes performances en milieu sous-marin que celle de Harris. A partir de ce constat, nous avons cherché à comprendre et à identifier quels bruits du milieu sous-marin pouvaient réduire les performances de SIFT à ce point. Nous avons réalisé plusieurs expérimentations dans différents milieux (en surface, en mer, en étang salé, en fleuve et en bassin) pour tester la méthode de SIFT dans différentes conditions. Nous avons créé ensuite un filtre représentant les principaux bruits sous-marins afin de pouvoir les ajouter un à un sur des images de surface. A la suite de toutes ces expérimentations, nous avons pu constater que la méthode de SIFT n'était pas robuste au bruit introduit par les particules en suspension : le bruit de speckle. Pour améliorer les résultats de cette méthode en réduisant le bruit, nous avons proposé d'utiliser un filtre anisotropique. Comme nous pouvions l'attendre, les résultats de SIFT se sont améliorés mais même avec un filtrage important, ils restaient inférieurs à ceux obtenus avec Harris&corrélation. Afin de corriger les méthodes de mise en correspondance qui donnent parfois de faux appariements, nous avons utilisé plusieurs caractéristiques du système de vision ainsi que l'algorithme de RANSAC. Une fois les points robustes extraits et appariés, nous les avons projetés dans un espace 3D pour obtenir un nuage de points de la structure. Le rendu visuel a été amélioré en ajoutant une surface ainsi que la texture au nuage de points.

Le deuxième point important de ce manuscrit était d'extraire une carte 3D acoustique à partir d'un sonar. Cette carte sert à fournir une information sur l'ensemble de la zone à cartographier. Sa résolution et sa précision ne sont pas importantes car elle doit nous permettre d'identifier un objet susceptible d'être intéressant. Nous n'avons pas besoin de connaître le type ni la dimension exacte de l'objet puisque ce sera le rôle du modèle 3D vidéo. Nous avons donc fait le choix d'utiliser la carte acoustique brute.

La dernière partie de ce travail consistait à fusionner les deux modèles (acoustique et vidéo) que nous avons obtenus. Après avoir présenté différentes méthodes utilisées par la communauté, nous avons opté pour un recalage géométrique des deux cartes. Cet alignement se décompose en deux étapes. Nous réalisons dans un premier temps un alignement approximatif manuel des deux modèles, en sélectionnant quatre points identiques dans chaque modèle. Ensuite, une méthode d'ICP permet d'affiner ce résultat. Les résultats ob-

tenus sont concluants et l'erreur moyenne d'alignement entre les points des deux modèles est inférieure au pas de la grille du sonar.

La continuité logique de ce travail serait d'arriver à enrichir et à augmenter la résolution du modèle 3D vidéo. Pour le réaliser, nous n'avons utilisé qu'une seule paire d'images ; il serait intéressant de se servir de toutes les images d'un objet pour enrichir ce modèle. Nous pourrions aussi utiliser les points robustes déjà définis pour lancer une nouvelle recherche locale autour de ces points.

Nous avons présenté dans ce document deux méthodes de reconstruction 3D, l'une à partir des images stéréoscopiques et l'autre à partir d'une séquence vidéo. Le milieu sous-marin ne nous permet pas d'avoir des résolutions aussi importantes qu'en surface. Alors pourquoi ne pas se servir d'un système vidéo-stéréoscopique afin d'utiliser la redondance d'informations ? La résolution du modèle 3D serait alors grandement améliorée, de plus, le problème de facteur d'échelle présent dans les méthodes de reconstruction à partir de la vidéo pourrait être levé grâce aux informations supplémentaires fournies par le système stéréo.

Lorsque nous voulons appliquer notre méthode sur un site de fouille archéologique complet, où les modèles locaux sont plus nombreux, nous nous apercevons rapidement que la méthode de recalage approximative manuelle est un frein important. En effet, lorsque des dizaines d'objets (donc de modèles vidéo) sont présents sur le site de fouille, leur recalage par sélection manuelle n'est plus concevable. Pour résoudre se problème, plusieurs solutions sont envisageables. La méthode de recalage fin impose cette première étape de recalage approximatif afin de s'initialiser et pour éviter de tomber dans une solution locale, or il serait possible de modifier cette méthode de minimisation afin d'obtenir une solution globale. Ainsi le recalage manuel ne serait plus nécessaire. Une autre solution consisterait à ajouter des capteurs extéroceptifs comme une centrale inertielle ou un loch-doppler afin de connaître le déplacement du système stéréoscopique. De ces informations, la position par rapport à la carte globale pourrait être déterminée et fournir une position approximative pour initialiser l'algorithme de recalage fin. Généralement, les robots sous-marins possèdent ce type de capteurs, donc s'ils portent le système stéréoscopique toutes ces données seraient accessibles.

En utilisant une meute de robots sous-marins autonomes, il serait possible d'automatiser la cartographie complète de la zone. Certains robots spécialisés dans le relevé sonar créeraient la carte acoustique complète de la zone à cartographier. Ils pourraient au fur et à mesure de la création cette carte déterminer les zones locales à détailler et les commu-

niquer à d'autres robots spécialisés dans le relevé stéréoscopique. Ainsi tout un chantier archéologique ou toute une zone sous-marine pourrait être cartographiée.

Bibliographie

R.J. Adrian et J. Westerweel : *Particle Image Velocimetry*. Cambridge Aerospace Series. Cambridge University Press, 2010. ISBN 9780521440080. URL http://books.google.co.in/books?id=jbDl2-yHbooC. Cité page 69.

R. Bandyopadhyay, A. S. Gittings, S. S. Suh, P. K. Dixon et D. J. Durian : Speckle-visibility spectroscopy : A tool to study time-varying dynamics. *Review of Scientific Instruments*, 76(9):093110, 2005. URL http://link.aip.org/link/?RSI/76/093110/1. Cité page 69.

L. Barazzetti : A trip to rome : physical replicas of historical objects created in a fully automated way from photos. *In Proceedings of the First international conference on Human-computer interaction, tourism and cultural heritage*, HCITOCH'10, pages 63–72, Berlin, Heidelberg, 2011. Springer-Verlag. ISBN 978-3-642-18347-8. URL http://portal.acm.org/citation.cfm?id=1946206.1946213. Cité page 42.

A. Bartoli et P. Sturm : Constrained structure and motion from multiple uncalibrated views of a piecewise planar scene. *International Journal on Computer Vision (IJCV)*, 52(1):45–64, April 2003. Cité page 42.

S. Bazeille : *Vision sous-marine monoculaire pour la reconnaissance d'objets*. Thèse de doctorat, ENSTA Bretagne, 2008. Cité pages 69 and 75.

C. Beall, B.J. Lawrence, V. Ila et F. Dellaert : 3d reconstruction of underwater structures.

In Proceedings of the 2010 IEEE/RSJ international conference on Intelligent robots and systems, IROS'10, Taipei, december 2010. IEEE Press. Cité pages 42 and 47.

P. Beaudet : Rotationally invariant image operators. *In 4th International Conference on Pattern Recogition (ICPR)*, pages 579–583, Kyoto, Japan, November 1978. Cité page 53.

P.J. Besl et N.D. McKay : A method for registration of 3d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(2):239–256, 1992. Cité page 105.

G. Bianco, A. Gallo, F. Bruno et M. Muzzupappa : A comparison between active and passive techniques for underwater 3d applications. *International Journal for Photogrammetry and Remote Sensing (ISPRS)*, XXXVIII(5-W16):7, Mars 2011. Cité pages 44, 45, and 133.

J. Y. Bouguet et P. Perona : Camera calibration from points and lines in dual space geometry. Rapport technique, California Institute of Technologie, 1998. Cité pages 34 and 133.

V. Brandou : *Stéréovision locale et reconstruction 3D/4D*. Thèse de doctorat, Université de Nice, Décembre 2008. Cité page 42.

B. Delaunay : Sur la sphère vide. *Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk*, 7:793–800, 1934. Cité page 82.

P. Drap, D. Merad, J.M. Bo&iuml ;, W. Boubguira, A. Mahiddine, B. Chemisky, E. Seguin, F. Alcala et O. Bianchimani : Rov-3d, 3d underwater survey combining optical and acoustic sensor. *In VAST11 : The 12th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage*, pages 177–184, Prato, Italy, Novembre 2011. Eurographics Association. ISBN 978-3-905674-34-7. URL http://diglib.eg.org/EG/DL/WS/VAST/VAST11/177-184.pdf. Cité page 102.

L. Dron : Dynamic camera self-calibration from controlled motion sequences. *In Conference on Computer Vision and Pattern Recognition*, pages p. 501–506. IEEE Computer Society Press, 1993. Cité page 29.

F. X. Espiau : *Métrologie 3D par vision active sur des objets naturels sous marin*. Thèse de doctorat, Université de Nice, Février 2002. Cité page 42.

M. A. Fischler et R. C. Bolles : Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. *Com-*

munication of the ACM, 24(6):381–395, juin 1981. ISSN 0001-0782. URL http://dx.doi.org/10.1145/358669.358692. Cité page 60.

V.S. Frost, J.A. Stiles, K.S. Shanmungan et J.C. Holtzman : A model for radar images and it 's application to adaptive digital filtering of multiplicative noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(2):157–166, March 1982. Cité page 75.

A. Fusiello, R. Giannitrapani, V. Isaia et V. Murino : Virtual environment modeling by integrated optical and acoustic sensing. *In Second International Conference on 3-D Digital Imaging and Modeling*, pages 437–446, 1999. Cité pages 103 and 107.

A. Fusiello et V. Murino : Calibration of an optical-acoustic sensor for underwater applications, 2000. Cité pages 103 and 107.

A. Fusiello et V. Murino : Augmented scene modeling and visualization by optical and acoustic sensor integration. *IEEE Transactions on Visualization and Computer Graphics*, 10(6):625–635, 2004. Cité pages 104 and 107.

G. Grieg, R. Kubler, R. Kikinis et F. A. Jolesz : Nonlinear anisotropic filtering of mri data. *IEEE Transactions on Medical Imaging*, 11(2):221–232, June 1992. Cité page 75.

C. Harris et M. Stephens : A combined corner and edge detection. *In The Fourth Alvey Vision Conference*, pages 147–151, 1988. Cité pages 53 and 62.

R. Hartley : Estimation of relative camera positions for uncalibrated cameras. *In Proceedings of the Second European Conference on Computer Vision*, ECCV '92, pages 579–587, London, UK, UK, 1992. Springer-Verlag. ISBN 3-540-55426-2. URL http://dl.acm.org/citation.cfm?id=645305.648678. Cité page 80.

R. Hartley : In defence of the 8-points algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580–593, June 1997. Cité page 60.

R. Hartley et A. Zisserman : *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, second edition édition, 2003. ISBN 0521540518. Cité pages 42, 47, and 48.

J. Heikkilä et O. Silvèn : A four-step calibration procedure with implicit image correction. *In Computer Vision and Pattern Recognition (CVPR97)*, pages 1106–1111, San Juan, Porto Rico, 1997. Cité page 21.

A. Hogue, A. German, J. Zacher et M. Jenkin : Underwater 3d mapping : Experiences and lessons learned. *Computer and Robot Vision, Canadian Conference*, 0:24, 2006. Cité page 42.

W. Hou : A simple underwater imaging model. *Optics letters*, 34(14):2688–2690, 2009. URL http://cat.inist.fr/?aModele=afficheN&cpsidt=21960232. Cité page 68.

J. S. Jaffe : Monte carlo modeling of underwater-image formation : validity of the linear and small-angle approximations. *Applied Optics*, 34(24):5413–5421, Aug 1995. URL http://ao.osa.org/abstract.cfm?URI=ao-34-24-5413. Cité page 68.

A. E. Johnson et S. B. Kang : Registration and integration of 3d data. *In International Conference on Recent Advances in 3D Digital Imaging and Modeling (3DIM97).*, 1997. Cité page 106.

F. Jurie et M. Dhome : Real time tracking of 3d objects : an efficient and robust approach. *Pattern Recognition*, 35(2):317 – 328, Feb 2002. ISSN 0031-3203. URL http://hal.inria.fr/inria-00548266. Cité page 58.

A. Kim et R. Eustice : Pose-graph visual slam with geometric model selection for autonomous underwater ship hull inspection. *In Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems*, IROS'09, pages 1559–1565, Piscataway, NJ, USA, 2009. IEEE Press. ISBN 978-1-4244-3803-7. URL http://portal.acm.org/citation.cfm?id=1733343.1733623. Cité page 42.

R. Koch, M. Pollefeys et L. V. Gool : Realistic surface reconstruction of 3d scenes from uncalibrated image sequences. *Journal of Visualization and Computer Animation*, 11(3):115–127, 2000. Cité page 59.

D.T. Kuan, A.A. Sawchuk, T.C. Strand et Chavel P. : Adaptive noise srnoothing filter for images wi th signal-dependent noise. *IEEE Transactions on Acoustics Speech and Signal Processing*, 7(2):165–1 77, March 1985. Cité page 75.

K. N. Kutulakos et S. M. Seitz : A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):199–218, 2000. Cité pages 43, 44, and 133.

Y. H. Kwon et S. Lindley : Applicability of four localized-calibration methods in underwater motion analysis. *In 18 International Symposium on Biomechanics in Sports*, Hong Kong, chine, 25 -30 Juin 2000. Cité page 13.

J. M. Lavest, G. Rives et J. T. Lapresté : Underwater camera calibration. *In* David Vernon, éditeur : *6th European Conference on Computer Vision (ECCV)*, volume Vol. 2 de *Lecture Notes in Computer Science*, pages 654–668, Dublin, Ireland, June 26 - July 1 2000. Springer. Cité pages 30 and 32.

J. S. Lee : Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(2):165–168, March 1980. Cité page 75.

L. Long, P. Picard et Musée de l'Arles antique : *César, le Rhône pour mémoire : vingt ans de fouilles archéologiques dans le fleuve à Arles*. Actes Sud, 2009. ISBN 9782742786107. URL http://books.google.fr/books?id=LvuDQgAACAAJ. Cité page 8.

H. C. Longuet-Higgins : A computer algorithm for reconstructing a scene from two projections. *Nature*, 293(10):133–135, sept 1981. Cité page 60.

M. Lourakis et R. Deriche : Camera self-calibration using the singular value decomposition of the fundamental matrix : from point correspondences to 3d measurements. Rapport technique 3748, INRIA, August 1999. Cité page 26.

D. G. Lowe : Object recognition from local scale-invariant features. *In International Conference on Computer Vision*, pages 1150–1157, Corfu, Greece, 1999. Cité page 54.

D. G. Lowe : Distinctive image features from scale-invariant keypoints. *International Journal in Computer Vision (IJCV)*, 60(2):91–110, 2004. ISSN 0920-5691. Cité pages 54, 57, and 62.

D.G. Lowe : Fitting parameterized three-dimensional models to images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5):441–450, 1991. Cité page 104.

B.D. Lucas et T. Kanade : An iterative image registration technique with an application to stereo vision. *In Proceedings of the 7th international joint conference on Artificial intelligence*, volume 2 de *IJCAI'81*, pages 674–679, San Francisco, CA, USA, 1981. Morgan Kaufmann Publishers Inc. URL http://dl.acm.org/citation.cfm?id=1623264.1623280. Cité page 57.

S. Mallat : Une exploration des signaux en ondelettes. Rapport technique, Editions de l'Ecole Polytechnique, 2000. Cité page 75.

G. H. Meeteen : *The measurement, intrumentation, and sansors handbook*, chapitre 61 : Refractive Index Measurement. CRC Press LLC - Springer-Verlag, 1999. Cité pages 12, 29, and 68.

P. Mendonça et R. Cipolla : A simple technique for self-calibration. *In Conference on Computer Vision and Pattern Recognition*, volume Vol. I, pages p. 500–505, Fort Collings, Colorado, June 1999. Cité page 26.

MeshLab : *Visual Computing Lab - ISTI - CNR*. URL http://meshlab.sourceforge.net. Cité page 98.

H. Moravec : Towards automatic visual obstacle avoidance. *In International Conferences on Artificial Intelligence*, page 584, Cambridge, 1977. Cité page 52.

S. Negahdaripour et P. Firoozfam : An rov stereovision system for ship-hull inspection. *IEEE Journal of Oceanic Engineering*, 31(3):551–564, July 2006. Cité page 42.

S. Negahdaripour, H. Sekkati et H. Pirsiavash : Opti-acoustic stereo imaging, system calibration and 3-d reconstruction. *In Computer Vision and Pattern Recognition, 2007. CVPR '07.*, pages 1 –8, june 2007. Cité pages 104 and 107.

A. Noce : *Vision et reconstruction 3D basée sur la texture : application à la robotique médicale*. Thèse de doctorat, Université de Montpellier 2, 2008. Cité page 42.

K. Oliver, W. Hou et S. Wang : Image feature detection and matching in underwater conditions. *In Ocean Sensing and Monitoring II - SPIE 7678*, 2010. Cité page 68.

P. Perona et J. Malik : Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990. Cité page 75.

N. Pessel, J. Opderbecke et M. J. Aldon : Camera self-calibration in underwater environment. *In International Conference on Computer Graphics, Visualization and Computer Vision*, 2003. Cité page 30.

M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops et R. Koch : Visual modeling with a hand-held camera. *International Journal of Computer Vision (IJCV)*, 59 (3):207–232, 2004. Cité pages 46 and 134.

W. K. Pratt : *Digital Image Processing : PIKS Inside*. John Wiley & Sons, Inc., New York, NY, USA, 2001. ISBN 0471374075. Cité page 42.

R. Richa : *Suivi 3d robuste pour la chirurgie cardiaque robotisée*. Thèse de doctorat, Université Montpellier II - Sciences et Techniques du Languedoc, 2010. Cité page 42.

C. Roman, G. Inglis et J. Rutter : Application of structured light imaging for high resolution mapping of underwater archaeological sites. *In OCEANS 2010 IEEE - Sydney*, pages 1 –9, may 2010. Cité page 104.

B. Sabata et J.K. Aggarwal : Estimation of motion from a pair of range images : A review. *Computer Vision Graphics and Images Processing : Image Understanding*, 54(3):309–324, 1991. Cité page 105.

A. Sarafraz, S. Negahdaripour et Y.Y. Schechner : Performance assessment in solving the correspondence problem in underwater stereo imagery. *In OCEANS'10*, Seattle Etats-Unis, 2010. MTS/IEEE. Cité page 68.

M. Sauvée : *Contribution à l'aide aux gestes pour la chirurgie cardiaque à coeur battant. Guidage échographique par asservissement prédictif non linéaire*. Thèse de doctorat, Université de Montpellier 2, Décembre 2006. Cité page 42.

R. Schettini et S. Corchs : Underwater image processing : state of the art of restoration and image enhancement methods. *EURASIP Journal on Advances in Signal Processing*, pages 141–147, January 2010. ISSN 1110-8657. URL http://dx.doi.org/10.1155/2010/746052. Cité page 69.

H. Sekkati et S. Negahdaripour : Direct and indirect 3-d reconstruction from opti-acoustic stereo imaging. *In 3D Data Processing, Visualization, and Transmission*, pages 615 –622, june 2006. Cité pages 104 and 107.

J. Shi et C. Tomasi : Good features to track. *In IEEE Conférence on Computer Vision and Pattern Récognition*, pages 593–600, Seattle, 1994. Cité page 57.

H. Singh, C. Roman, L. Whitcomb et D. Yoerger : Advances in fusion of high resolution underwater optical and acoustic data. *In Underwater Technology, 2000.*, pages 206 –211, 2000. Cité pages 102 and 106.

H. Singh, G. Salgian, R. Eustice et R. Mandelbaum : Sensor fusion of structure-from-motion, bathymetric 3d, and beacon-based navigation modalities. *In International Conference on Robotics and Automation, 2002. ICRA '02.*, volume 4, pages 4024 – 4031 vol.4, 2002. Cité pages 102 and 107.

J. M. Smith, S. M. Brady : Susana new approach to low level image processing. *International Journal in Computer Vision (IJCV)*, 23(1):45–78, 1997. Cité page 53.

N. Snavely, S. M. Seitz et R. Szeliski : Photo tourism : exploring photo collections in 3d. *ACM Transaction on Graphics*, 25:835–846, July 2006. ISSN 0730-0301. URL http://doi.acm.org/10.1145/1141911.1141964. Cité page 41.

R. Szeliski : *Computer Vision : Algorithms and Applications*. 24 march édition, 2010. Cité page 42.

M. Tahiri Alaoui, S. M. Farssi, A. Touzani, K. Mortalla et P. Bunel : Caractérisation de texture : application sur les images échographiques du rein. *Revue Africaine de la recherche en informatique et mathématiques appliquées (ARIMA)*, pages 1–24, Nomvembre 2005. Cité page 75.

C. Tomasi et T. Kanade : Détection and tracking of point features. Rapport technique CS-91-132, Carnegie Mellon University, 1991. Cité page 57.

S. Williams et I. Mahon : Simultaneous localisation and mapping on the great barrier reef. *In IEEE International Conference on Robotics and Automation, 2004.ICRA '04.*, volume 2, pages 1771 – 1776 Vol.2, 26-may 1, 2004. Cité pages 103, 107, and 134.

C. Zeller : *Calibration Projective, Affine et Euclidienne en vision par ordinateur et application à la perception tridimensionnelle*. Thèse de doctorat, Ecole Polytechnique, Février 1996. Cité page 26.

Z. Zhang : Iterative point matching for registration of free-form curves. Rapport technique 1658, INRIA, mars 1992. Cité page 105.

Z. Zhang : Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(2):119–152, 1994. Cité page 105.

Z. Zhang : A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1330–1334, 2000. ISSN 0162-8828. Cité pages 27, 30, and 31.

Liste des Publications

Conférences internationales

1. A. Méline, J. Triboulet and B. Jouvencel, "Comparative study of two reconstruction methods for underwater archaeology" in *IEEE/RSJ International conference on Intelligent Robots and Systems*, Vilamoura, Algarve, oct. 2012 (IROS12).
2. A. Méline, J. Triboulet and B. Jouvencel, "A camcorder for 3D underwater reconstruction of archeological objects", *Proceedings of MTS/IEEE OCEAN10 Seattle*, Seattle, Washington, September 20-23, 2010.

Communications nationales

1. A. Méline, J. Triboulet et B. Jouvencel, "Fusion multimodale pour la cartographie sous-marine", *GdR Rob - Perception sous-marine*, Paris, 22 Novembre, 2012.
2. A. Méline, J. Triboulet et B. Jouvencel, "De la vidéo d'objets archéologiques sous-marins aux modèles 3D", *JJCR11*, La rochelle, 18 Octobre, 2011.
3. A. Méline, J. Triboulet et B. Jouvencel, "Fusion multimodale pour la cartographie sous-marine", *GdR ISIS - Adéquation Analyse et Architecture*, Paris, 6 Octobre, 2011.
4. A. Méline, J. Triboulet et B. Jouvencel, "Reconstruction 3D d'objets archéologiques sous-marins", *Doctiss11*, Montpellier, 19 Mai, 2011.

5. A. Méline, J. Triboulet et B. Jouvencel, "Reconstruction en 3 dimensions d'objets archéologiques sous-marins", *JJCR10*, Paris, 8 Novembre, 2010.

Table des figures

2.1 Photo du système stéréoscopique. 8
2.2 Photos de la statue de poisson lors d'expérimentation en mer aux Aresquiers. . . 9
2.3 Photos stéréoscopiques d'une amphore lors d'expérimentations en mer aux Saintes-Maries-de-la-mer. 9
2.4 Photos stéréoscopiques du buste lors d'expérimentations dans le Rhone. 10
2.5 Photos stéréoscopiques d'une Amphore lors d'expérimentations dans l'étang de Thau. 10
2.6 Photos stéréoscopiques du buste lors d'expérimentations en bassin. 11
2.7 Disparition des couleurs en fonction de la distance approximative à partir de la source lumineuse. 12

3.1 Modèle de caméra et changement de repères 18
3.2 Schéma d'un système de stéréovision 22

4.1 Représentation schématique de la distorsion radiale. 33
4.2 Erreur de reprojection du calibrage (en pixel) [Bouguet et Perona, 1998] 34
4.3 Visualisation des positions de l'appareil photo lors du calibrage [Bouguet et Perona, 1998] 34

5.1 Reconstruction d'une sculpture de gargouille [Kutulakos et Seitz, 2000] 44
5.2 Reconstruction d'une sculpture [Bianco *et al.*, 2011] 45

5.3 Reconstruction 3D d'une méduse antique [Pollefeys *et al.*, 2004] 46
5.4 Organigramme de la reconstruction 3D 48

6.1 Détection des points caractéristiques sur une image prise en mer aux Aresquiers. 65
6.2 Estimation de mouvement global à partir des appariements des points de Harris. 65
6.3 Estimation de mouvement global à partir des appariements des points de SIFT. 66
6.4 Histogrammes des plans rouge (a), vert (b) et bleu (c) d'images sous-marines, de surfaces et filtrées. 70
6.5 Image sous-marine et résultats du filtrage de couleur sur une image de surface. 71
6.6 Histogrammes des variances locales d'images sous-marines, de surfaces et filtrées. 72
6.7 Image sous-marine et résultats du filtrage de speckle et de flou sur une image de surface. 72

7.1 Les quatre solutions possibles pour l'estimation de la transformation. *Seule la solution (a) possède les points reconstruits devant les deux caméras.* 81
7.2 Triangulation de Delaunay. 83
7.3 Surface 3D d'un col d'amphore des Saintes-Maries-de-la-mer. 83
7.4 Modèle 3D texturé d'un col d'amphore des Saintes-Maries-de-la-mer. 84
7.5 Modèle 3D texturé métrique du buste immergé dans le Rhône. 85
7.6 Modèle 3D texturé non métrique du buste immergé dans le Rhône. 86
7.7 Modèle 3D texturé d'une amphore immergée dans l'étang de Thau. 87

8.1 Représentation schématique d'un relevé avec un sonar mono faisceau. 92
8.2 Représentation schématique d'un relevé avec un sonar à balayage. 93
8.3 Représentation schématique d'un relevé avec un sonar multifaisceaux. 93
8.4 Représentation schématique d'un relevé avec un sonar "Forward Looking". . . . 94
8.5 Représentation schématique d'un relevé avec un sonar latéral. 94
8.6 Sonar multifaisceaux DeltaT Profiling 95
8.7 Carte 3D sonar de l'épave romaine en Arles. 96
8.8 Carte 3D sonar de l'épave de l'étang de Thau avec une amphore sur le pont. . . . 97
8.9 Carte 3D sonar de l'épave de l'étang de Thau visualisée sous MeshLab. 99

9.1 Résultats de fusion de données acoustiques et visuelles obtenus par Williams [Williams et Mahon, 2004]. 103
9.2 Carte 3D sonar de l'épave de l'étang de Thau avec une amphore sur le pont. . . . 109

9.3 Premier modèle 3D texturé de l'amphore immergée dans l'étang de Thau. 109
9.4 Deuxième modèle 3D texturé de l'amphore immergée dans l'étang de Thau. . . 110
9.5 Vue globale de la première carte 3D finale de l'épave de l'étang de Thau et de l'amphore. 111
9.6 Vue de la première carte 3D finale de l'épave de l'étang de Thau et de l'amphore. 111
9.7 Vue zoomée de la première carte 3D finale de l'épave de l'étang de Thau et de l'amphore. 112
9.8 Vue globale de la deuxième carte 3D finale de l'épave de l'étang de Thau et de l'amphore. 113
9.9 Vue de la deuxième carte 3D finale de l'épave de l'étang de Thau et de l'amphore. 113
9.10 Vue zoomée de la deuxième carte 3D finale de l'épave de l'étang de Thau et de l'amphore. 114

Liste des tableaux

4.1 Résultats du calibrage de l'appareil photo en surface. 32
4.2 Résultats du calibrage de l'appareil photo en surface et en mer. 32
4.3 Erreur de reprojection du calibrage selon les axes x et y. 33
4.4 Résultats du calibrage de la paire stéréoscopique en mer. 35

6.1 Comparaison du nombre de points appariés et du nombre de bon appariement obtenu avec les deux méthodes dans un environnement sous-marin. 67
6.2 Estimation du mouvement à partir des points de Harris (en pixel). 67
6.3 Comparaison des nombres de points des deux méthodes avec différents filtres appliqués sur une image de surface. 73
6.4 Statistiques du taux de bons appariements obtenus avec les deux méthodes sur 20 images. 74
6.5 Statistiques du taux de bons appariements des deux méthodes dans différents milieux (surface, filtré, mer et bassin). 74
6.6 Statistiques du taux de bons appariements des deux méthodes après différentes valeurs du filtrage anisotropique 76

7.1 Mesure de certaines distances caractéristiques du buste des expérimentations dans le Rhône. 87
7.2 Mesure de certaines distances caractéristiques de l'amphore des expérimentations en étang salé. 87

9.1 Mesure des erreurs de recalage des modèles. 110

Résumé

Le but de ce travail est d'analyser des scènes sous-marines naturelles et en particulier cartographier des environnements sous-marins en 3D. Il existe aujourd'hui de nombreuses techniques pour résoudre ce problème. L'originalité de ce travail se trouve dans la fusion de deux cartes obtenues avec des capteurs de différentes résolutions. Dans un premier temps, un engin autonome (ou un bateau) analyse les fonds marins avec un sonar multifaisceaux et crée une première carte globale de la zone. Cette carte est ensuite décomposée en petites cellules représentant une mosaïque du fond marin. Une deuxième analyse est ensuite réalisée sur certaines cellules particulières à l'aide d'un second capteur avec une résolution plus élevée. Cela permettra d'obtenir une carte détaillée 3D de la cellule. Un véhicule autonome sous-marin ou un plongeur muni d'un système de vision stéréoscopique effectuera cette acquisition.

Ce projet se décompose en deux parties, la première s'intéressera à la reconstruction 3D de scènes sous-marines en milieu contraint à l'aide d'une paire stéréoscopique. La deuxième partie de l'étude portera sur l'aspect multimodal. Dans notre cas, nous utilisons cette méthode pour obtenir des reconstructions précises d'objets d'intérêts archéologiques (statues, amphores, etc.) détectés sur la carte globale.

La première partie du travail concerne la reconstruction 3D de la scène sous-marine. Même si aujourd'hui le monde de la vision a permis de mieux appréhender ce type d'image, l'étude de scène sous-marine naturelle pose encore de nombreux problèmes. Nous avons pris en compte les bruits sous-marins lors de la création du modèle 3D vidéo ainsi que lors de la calibration des appareils photos. Une étude de robustesse à ces bruits a été réalisée sur deux méthodes de détections et d'appariements de points d'intérêt. Cela a permis d'obtenir des points caractéristiques précis et robustes pour le modèle 3D. La géométrie épipolaire nous a permis de projeter ces points en 3D. La texture a été ajoutée sur les surfaces obtenues par triangulation de Delaunay.

La deuxième partie consiste à fusionner le modèle 3D obtenu précédemment et la carte acoustique. Dans un premier temps, afin d'aligner les deux modèles 3D (le modèle vidéo et le modèle acoustique), nous appliquons un recalage approximatif en sélectionnant manuellement quelques paires de points équivalents sur les deux nuages de points. Pour augmenter la précision de ce dernier, nous utilisons un algorithme ICP (Iterative Closest Points).

Dans ce travail nous avons créé une carte 3D sous-marine multimodale réalisée à l'aide de modèles 3D " vidéo " et d'une carte acoustique globale.

Mots clefs : *Vision sous-marine, Reconstruction 3D, Fusion / Recalage 3D, cartographie*

Abstract

This work aims to analyze natural underwater scenes and it focuses on mapping underwater environment in 3D. Today, many methods exist to solve this problem. The originality of this work lies in the fusion of two maps obtained from sensors of different resolutions. Initially, an autonomous vehicle (or boat) analyzes the seabed with multibeam sonar and creates a first global map of the area. This map is then divided into small cells representing a mosaic of the seabed. A second analysis is then performed on some particular cells using a second sensor with a higher resolution. This will provide a detailed map of the 3D cell. An autonomous underwater vehicle (AUV) or a diver with a stereoscopic vision system will make this acquisition.

This project is divided into two parts; the first one focuses on the 3D reconstruction of underwater scenes in constrained environment using a stereoscopic pair. The second part investigates the multimodal aspect. In our study, we want to use this method to obtain accurate reconstructions of archaeological objects (statues, amphorae, etc.) detected on the global map.

The first part of the work relates the 3D reconstruction of the underwater scene. Even if today the vision community has led to a better understanding of this type of images, the study of natural underwater scenes still poses many problems. We have taken into account the underwater noise during the creation of the 3D video model and during the calibration of cameras. A study of the noise robustness was performed on two methods of detection and matching of features points. This resulted into obtaining accurate and robust feature points for the 3D model. Epipolar geometry allowed us to project these points in 3D. The texture was added to the surfaces obtained by Delaunay triangulation.

The second part consists of fusing the 3D model obtained previously with the acoustic map. To align the two 3D models (video and acoustic model), we use a first approximated registration by selecting manually few points on each cloud. To increase the accuracy of this registration, we use an algorithm ICP (Iterative Closest Point).

In this work we created a 3D underwater multimodal map performed using 3D video model and an acoustic global map.

Keywords: *Underwater vision, 3D reconstruction, Fusion / 3D registration, mapping*

LIRMM — 161, rue Ada — 34095 Montpellier cedex 5 — France