
Vers une méthodologie de reconstruction 3D de scènes sous-marines par caméras acoustiques

Problématiques et études préliminaires

Naouraz Brahim^{1,2}, Sylvie Daniel¹, Didier Guériot²,
Basel Solaiman²

1. Université Laval, Département des Sciences Géomatiques
Pavillon Louis Jacques Casault, G1V 0A6, Québec, Canada
naouraz.brahim.1@ulaval.ca, sylvie.daniel@scg.ulaval.ca

2. Institut TELECOM ; TELECOM Bretagne ; UMR CNRS 6285 Lab-STICC/CID
Technopôle Brest Iroise - CS 83818 – F-29238 Brest cedex 3, France
Université Européenne de Bretagne
{didier.gueriot, basel.solaiman}@telecom-bretagne.eu

RÉSUMÉ. Depuis que les études d'impact des changements climatiques ont montré que le milieu marin pourrait être énormément fragilisé par la disparition de certaines espèces parmi la faune et la flore ainsi que par le vieillissement rapide des infrastructures sous-marines, de nouveaux outils d'observation efficaces et robustes deviennent nécessaires. Dans cet article, l'utilisation de caméras acoustiques comme outil novateur d'acquisition de données sous-marines est proposée en compagnie d'un cadre conceptuel qui permet de mettre en œuvre une reconstruction tridimensionnelle pertinente et complète de l'environnement sous-marin à partir de séquences d'images acquises par ces caméras acoustiques. Les différentes données et informations extraites, utilisables et utilisées pour élaborer ce travail ainsi que quelques résultats préliminaires sont abordés.

ABSTRACT. Since climate change impact studies have shown that the marine environment could be greatly weakened by the disappearance of some flora and fauna species and the rapid aging of underwater infrastructure, new, efficient and robust observation tools are required. In this paper we propose acoustic cameras as innovative tools for acquiring underwater data along with a conceptual framework that allows an accurate and complete three-dimensional reconstruction implementation of the underwater environment using image sequences acquired by those acoustic cameras. Examples of information extracted from the acquired acoustic data that participate in such a 3D reconstruction are also provided as well as some preliminary results.

MOTS-CLÉS : observation du milieu marin, caméra acoustique, méthodologie, primitives 3D.

KEYWORDS: marine underwater observation, acoustic camera, methodology, 3D feature.

DOI:10.3166/TS.29.521-551 © 2012 Lavoisier

Extended abstract

The main purpose of the proposed paper is to develop a conceptual framework that allows an accurate and complete three-dimensional reconstruction of underwater scenes using sequences of acoustic images. This three-dimensional reconstruction will be useful for visual search, inspection and survey which are critical in a number of underwater applications in marine sciences, observation, maintenance, repair of undersea structures, etc. Indeed, according to recent studies, climate change is having a significant impact on our marine environment. It relies on temperature increasing, chemistry changes, ocean circulation influencing population dynamics and the underwater structure stability. Furthermore, monitoring the changes in marine environment is one of the current challenges for underwater exploration. In this paper, we use the acoustic camera DIDSON which has several advantages and characteristics that allow her to be a good candidate for underwater three-dimensional reconstruction. Compared to traditional sonar, it provides high resolution and rapid refresh rates. Image updates range between 21 frames/s to 5 frames/s depending on the operating frequency and the maximum range imaged. The camera provides real-time underwater image sequence. In this sequence the same scene appears in different successive images from different point of view. Thus, the proposed research work addressed in this paper aims at developing a method that exploit the sequence image information redundancy to determine the scene elevation. We introduce a method inspired from stereovision techniques consisting of three modules: 1) Feature detection and extraction, 2) Feature matching and 3) 3D reconstruction. These modules are detailed in the paper. Preliminary results are provided based on experiments conducted on diverse data sets.

1. Introduction

Le milieu marin est caractérisé par un potentiel économique et environnemental très important dû essentiellement à ses ressources halieutiques, offshore ainsi qu'au rôle prépondérant qu'il joue comme milieu de transport maritime. Néanmoins, les conditions d'exploration et d'exploitation de ses ressources, qu'elles soient à des fins scientifiques ou économiques, restent toujours difficiles tout particulièrement dans les zones profondes et turbides. De plus, les études récentes sur l'impact des changements climatiques actuels et envisagés dans le futur montrent l'apparition de premiers signes de vulnérabilité de ce milieu marin. Ces signes se présentent entre autres à travers l'érosion des fonds, la disparition d'espèces et le vieillissement des structures marines (ex. les quais et les digues). Il pourrait résulter de ces changements toute une série d'impacts biophysiques et socio-économiques qui, à terme, auraient une incidence sur divers secteurs économiques, notamment la pêche, le transport, le tourisme et les loisirs, de même que sur les collectivités (GIEC, 2007). Il s'avère ainsi primordial et nécessaire de mettre en œuvre des solutions préventives aux impacts des changements climatiques. Ces dernières permettent d'évaluer la vulnérabilité d'une région, d'une zone ou d'une composante du milieu.

Elles se basent sur l'analyse à la fois de l'ampleur des impacts potentiels et des éventuelles mesures d'adaptation à ces impacts. Un facteur important à considérer dans ce genre d'analyse est la vitesse à laquelle on prévoit que les changements vont survenir. Pour évaluer ce facteur, il est nécessaire de disposer de plusieurs mesures d'observation et de modélisation à l'échelle spatio-temporelle du milieu marin (Walther *et al.*, 2002).

Dans le but de répondre à ce besoin, des observatoires ont été construits et répartis dans les mers et les océans (ex. NEPTUNE¹, ESONET²). Les données enregistrées dans le cadre de ces observatoires sont obtenues à partir d'observations automatisées, avec des robots munis d'un ensemble d'outils de mesure et de capteurs notamment des sonars latéraux, des sondeurs multifaisceaux et des capteurs optiques (ex. appareils photographiques, caméras vidéo haute résolution) (Tunnicliffe *et al.*, 2008). Ces instruments de mesure sont adaptés aux conditions du milieu, présentant, par exemple, une résistance aux courants forts, à la pression, etc. Par contre, l'utilisation de certains outils, comme les caméras optiques, indispensables à la transmission de l'image du milieu marin est contraignante et fait face à plusieurs problèmes techniques tels qu'un éclairage qui fait fuir ou, au contraire, attire les organismes ou bien qui modifie d'une certaine manière les phénomènes physiques observés, donnant ainsi une image biaisée du milieu (Jun et Asada, 2007). En plus de ces problèmes, dans certaines conditions météorologiques ou du fait d'un trafic maritime important, la perception de scènes sous-marines à partir de ces capteurs optiques devient une tâche très difficile voire impossible à cause de l'augmentation de la turbidité de l'eau (Belcher *et al.*, 2003 ; Caimi *et al.*, 2010). La recherche d'équipements innovants et adéquats ne souffrant pas de ces inconvénients est un élément-clé pour répondre aux performances demandées à ces observatoires.

Le traitement des données acquises à travers des photographies sous-marines et des relevés de paramètres caractéristiques conduit à la réalisation de cartographies des habitats marins, des fonds etc., et à la production de modèles d'évolution spatio-temporelle du milieu marin. L'information représentée sur ces cartes et ces modèles permet d'évaluer et de prédire les tendances des impacts des changements climatiques, mais reste toujours qualifiée d'information partielle et incomplète puisqu'elle aboutit à une modélisation bidimensionnelle d'un milieu en réalité tridimensionnel. Il s'avère ainsi indispensable, en plus de chercher un outil performant et robuste aux conditions du milieu marin pour l'acquisition des données, de mettre en œuvre une méthodologie permettant de modéliser le milieu dans ses trois dimensions afin d'avoir une vue globale et réelle de ce milieu. Il apparaît également nécessaire que la représentation tridimensionnelle du milieu soit élaborée de manière précise et exacte étant donné son implication ultérieure dans des systèmes d'aide à la décision visant à choisir les solutions préventives aux impacts des changements climatiques à adopter.

1. <http://www.neptunecanada.ca>

2. <http://www.esonet-emso.org>

C'est dans ce contexte que s'inscrit le projet de recherche abordé dans le présent article. Ce projet vise à concevoir et développer une approche permettant la reconstruction tridimensionnelle du milieu marin qui prenne en compte l'imperfection des informations extraites des données collectées. Il propose notamment d'exploiter les caméras acoustiques comme outil d'exploration et d'observation du milieu marin. Ces caméras tirent profit, d'une part, de la capacité qu'ont les ondes acoustiques à se propager dans des milieux turbides et, d'autre part, de l'acquisition de données sous forme de séquences d'images avec une cadence élevée comme c'est le cas pour les caméras optiques (Caimi *et al.*, 2010). Néanmoins, l'exploitation des données de ces caméras fait face à plusieurs problématiques telles que la géométrie de formation des images et leur nature bruitée. Une des principales contributions de cette recherche concerne l'élaboration d'une méthodologie globale de traitement menant à la reconstruction 3D de la scène sous-marine observée. A travers l'élaboration d'une méthodologie globale, nous cherchons à définir les fondements méthodologiques dans lesquels doit s'inscrire une approche de reconstruction 3D d'une scène sous-marine visant à être précise et fiable. Les différents composants de cette méthodologie vont alors représenter les concepts-clés nécessaires pour effectuer cette reconstruction de la scène observée à partir d'images issues de caméras acoustiques. Ils doivent mettre en œuvre des solutions permettant de reconstruire l'environnement marin en tenant compte des problématiques inhérentes à ce milieu et à ces capteurs. Ces problématiques ainsi que la méthodologie globale de traitement proposée constituent l'objet de cet article et vont y être détaillées.

L'article est organisé de la manière suivante. La section 2 se consacre à la description des caractéristiques des caméras acoustiques et de leur exploitation dans le milieu marin. Dans cette section, les problématiques liées à l'utilisation de ces capteurs dans le contexte d'une reconstruction tridimensionnelle sont aussi abordées. Par la suite, un état de l'art des différentes approches de reconstruction tridimensionnelle issues des domaines de la vision numérique et de l'imagerie sonar est proposé. La méthodologie globale de traitement dans laquelle s'inscrit notre démarche de reconstruction 3D du milieu marin est ensuite explicitée et justifiée, cette partie étant suivie de quelques résultats préliminaires avant de conclure.

2. Les caméras acoustiques pour la reconstruction 3D de l'environnement

Les caméras acoustiques ont été initialement conçues dans l'objectif de développer un capteur sonar haute fréquence à des fins d'inspection militaire. Ces caméras se caractérisent essentiellement par la production de séquences d'images avec un taux de rafraîchissement élevé, proche de celui des capteurs optiques. A titre d'exemples, on peut citer les caméras DIDSON (*Dual Frequency Identification Sonar*) proposant une cadence d'images de 4 à 21 images/seconde et deux modes de fonctionnement selon lesquels la portée (distance entre la caméra et la scène) de la caméra varie. En mode haute fréquence (HF), la fréquence du signal acoustique est de 1,8 MHz et la portée de 12 m. En mode basse fréquence (BF), la fréquence du signal est de 1,1 MHz et la portée de 40 m. Les caméras LIMIS (*Limpet Mine Imaging Sonar*) proposent une cadence d'images variant de 5 à 12 images/seconde

avec une fréquence de 2 MHz et une portée de 12 m (Belcher *et al.*, 2003). On peut citer encore les caméras BlueView³ dont la fréquence, la cadence et la portée dépendent de la version de la caméra (ex. BlueView P900 propose une fréquence égale à 900 KHz, une cadence de 10 images/seconde et une portée de 100 m). Les applications de ces caméras sont multiples⁴ : en plus des tâches d'inspection des milieux portuaires et d'identifications d'objets, elles sont également utilisées dans la navigation comme aide à l'évitement d'obstacles pour des robots sous-marins dans des missions de surveillance d'installations sous-marines. Leur utilisation en halieutique est de plus en plus fréquente et ceci pour l'analyse du comportement et de l'observation des bancs de poissons, et pour le suivi de l'évolution de la faune (Moursund *et al.*, 2003).

Le travail de recherche dont il est question dans cet article s'intéresse spécifiquement aux caméras acoustiques comme outil d'observation du milieu marin. En effet, en plus des caractéristiques présentées dans le paragraphe précédent, les caméras acoustiques se caractérisent par la production de séquences d'images offrant une visualisation d'une même scène sous différents angles de vue. Il devrait donc être possible de les exploiter afin de produire des représentations tridimensionnelles d'une scène observée.

2.1. Nature des images issues des caméras acoustiques

Contrairement aux images optiques où la formation des images suit généralement le modèle sténopé, les images acoustiques ont une géométrie de formation spécifique aux caméras acoustiques. En effet, la formation des images optiques correspond à une projection centrale des rayons optiques sur le plan image et elles sont acquises dans un système de coordonnées cartésiennes. Les images acoustiques quant à elles sont acquises dans un système de coordonnées sphériques. Dans ce système, l'axe des ordonnées correspond à un temps de trajet aller-retour entre l'antenne qui émet et reçoit le signal acoustique et les différents éléments intersectés dans la scène. L'axe des abscisses, quant à lui, correspond à l'angle entre l'axe du capteur (ligne centrale de l'image ou axe des y) et les éléments intersectés dans la scène. Par conséquent, comme illustré dans la figure 1, un point géographique P dans la scène est repéré dans l'image acoustique par le pixel P_0 via sa portée r (distance caméra-scène) et son azimuth θ (l'angle d'insonification). La troisième dimension, qui correspond à l'élévation φ du point par rapport à la caméra est perdue lors de l'acquisition des images (Brahim *et al.*, 2010). Dans la figure 1 l'axe y correspond à la direction d'insonification du sonar tandis que l'axe x est orienté suivant les azimuths positifs (tribord) et l'axe z suivant les élévations positives. Tout point P (r, θ, φ) de l'environnement sous-marin contribue à la valeur du pixel P_0 de coordonnées ($r, \theta, 0$) dans l'image DIDSON qui peut être vue comme appartenant au plan d'élévation nulle ($\varphi=0$).

3. <http://www.blueview.com/2d-Imaging-Sonar.html>

4. <http://www.soundmetrics.com/industries-served>

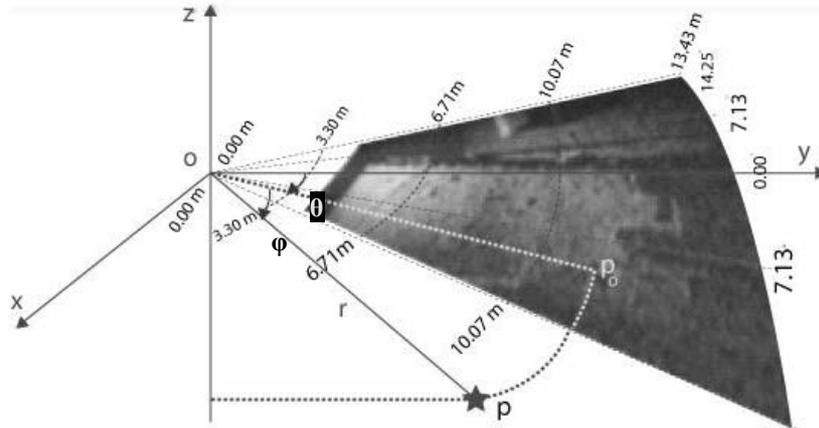


Figure 1. Repère sphérique d'acquisition d'une image par une caméra DIDSON (placée à l'origine O), (Brahim et al., 2010)

Ainsi, selon cette géométrie, chaque pixel dans l'image ne correspond pas à la rétrodiffusion d'un point dans la scène mais plutôt à l'agrégation d'un ensemble de rétrodiffusions retournées par tous les éléments de la scène situés à une même distance du sonar, pour un azimut donné. Idéalement, cet ensemble de rétrodiffusion forme un arc de cercle ; néanmoins, de par les caractéristiques du capteur (ici une caméra DIDSON), il s'agit plutôt d'un morceau de ruban volumique très fin dont les limites sont définies par l'ouverture verticale (10.9° en élévation) et la résolution en distance (Δr qui correspond à la durée d'écoute pour chaque pixel) du capteur mais aussi l'ouverture horizontale $\Delta\theta$ (en azimut) de chaque faisceau (0.3° en basse fréquence ou 0.4° en haute fréquence) (Sekkati et Negahdaripour, 2007). La figure 2 présente la géométrie de chaque faisceau et leurs paramètres caractéristiques impliqués dans la spécification des rubans volumiques associés à chaque pixel de l'image.

Le modèle représentatif de la projection dans l'image acoustique des ondes rétrodiffusées par les éléments de la scène est décrit en section 2.3.1.

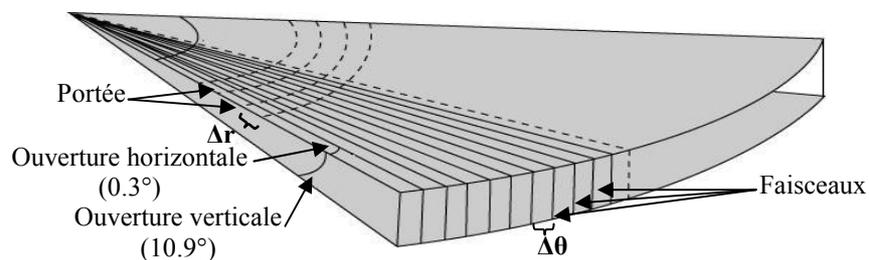


Figure 2. Vue globale des faisceaux d'insonification de la caméra DIDSON

La section suivante décrit le modèle géométrique associé à la formation des images acquises par des caméras de type DIDSON. Compte tenu de la similarité du mode de fonctionnement des différents modèles de caméras acoustiques et des caractéristiques radiométriques et géométriques des images qu'ils permettent d'acquérir (Belcher *et al.*, 2003), un modèle géométrique semblable à celui qui va être présenté ici pourrait être établi pour ces autres modèles.

2.2. Modèle géométrique d'acquisition pour une caméra de type DIDSON

Les schémas précédents illustrent parfaitement le modèle géométrique idéal de la caméra DIDSON. Ainsi, en coordonnées cartésiennes, un point P de la scène observée est représenté par sa position 3D, $P = (x, y, z)^T$, selon les conventions de la figure 1. En coordonnées sphériques dans un repère dont l'origine est la caméra, les composantes de ce point P deviennent : l'angle d'insonification θ correspondant au faisceau auquel le point contribue, la distance r par rapport au sonar ou portée qui est mesurée dans la direction du déplacement du sonar et l'élévation φ . Le modèle de la caméra liant ces deux systèmes de coordonnées est représenté comme suit (Negahdaripour *et al.*, 2005) :

$$\begin{cases} x = r \cdot \sin(\theta) \cdot \cos(\varphi) \\ y = r \cdot \cos(\theta) \cdot \cos(\varphi) \\ z = -r \cdot \sin(\varphi) \end{cases} \quad \begin{cases} r = \sqrt{x^2 + y^2 + z^2} \\ \theta = \arctan(x / y) \\ \varphi = -\arcsin(z / r) \end{cases} \quad (1)$$

Avec un tel modèle, tout point P de la scène se « projette » dans une image bidimensionnelle en un point $P_o(r, \theta)$, l'élévation φ de ce point P étant perdue lors de cette projection dans l'image.

2.3. Problématiques liées à l'exploitation des caméras acoustiques à des fins de reconstruction 3D du milieu marin

Un premier problème rencontré lorsque l'on souhaite exploiter des caméras acoustiques afin de produire des représentations 3D d'une scène sous-marine est le nombre limité de travaux de recherche ayant abordé ce sujet, notamment selon une approche multivue. Les travaux exploitant les images acoustiques concernent essentiellement les sonars latéraux (Lu et Oshima, 2002, Sun *et al.*, 2008) et les sonars frontaux (Zerr et Stage, 1996). Étant donné que les caméras acoustiques sont souvent assimilées à des sonars frontaux, il semble logique de chercher à exploiter les méthodes mises en œuvre dans ce domaine afin de produire des modélisations 3D à partir des images acquises avec les caméras acoustiques. Cependant, ces méthodes s'avèrent peu adéquates car souvent dédiées à une application donnée et à la configuration spécifique d'acquisition du sonar frontal. Ainsi, l'approche proposée par Zerr et Stage (1996) a été développée afin de reconstruire des objets de petites dimensions et de forme convexe et elle s'appuie sur une stratégie contrôlée

d'acquisition d'images (le sonar frontal sectoriel effectue une circumnavigation de l'objet en visant en permanence son centre). Cette approche facilite la procédure de reconstruction des objets mais elle n'est pas efficace dans le contexte de la reconstruction de grandes structures statiques ni dans le cas d'objets en mouvement.

En imagerie optique, les approches de reconstruction à partir d'images multivues sont multiples (ex. la stéréovision, la reconstruction volumétrique, etc.). En revanche, lorsque l'on souhaite appliquer ces approches aux cas des caméras acoustiques, il faut faire face au fait que la géométrie de formation des images est différente. De plus, à cette géométrie d'acquisition différente s'ajoute deux problématiques, l'une liée à la nature très bruitée des images et l'autre liée à la complexité et au caractère évolutif de l'environnement marin.

2.3.1. Reconstruction d'un point 3D à partir d'une paire d'images

Le fait que les caméras acoustiques fournissent des séquences d'images multivues avec une cadence élevée permet d'envisager les approches stéréo et multi-images optiques comme méthodes pour la reconstruction tridimensionnelle d'environnement sous-marin. Cette section est ainsi consacrée à la présentation des différences entre les images acoustiques et les images optiques qu'il faut prendre en considération lors de l'application des méthodes de calcul des coordonnées 3D de la scène. Dans cette section, le mouvement de la caméra entre les deux acquisitions est supposé parfaitement connu, même si en pratique, ce n'est pas le cas lorsque l'on acquiert des données sous-marines.

Ces différences résident premièrement au niveau du modèle de la caméra : le modèle sténopé de formation des images optiques est totalement différent du modèle des caméras acoustiques. Ce modèle simplifie l'ensemble des lentilles qui composent le système optique par un point où convergent tous les rayons lumineux pour aller se projeter sur le plan image par des droites sécantes à l'origine du repère de la caméra (cf. figure 3a). À partir de cette simplification, le modèle de la caméra est approché par un modèle projectif homographe simple qui est le modèle sténopé. Par une projection perspective, il est ainsi possible de calculer les coordonnées tridimensionnelles des points dans les images. L'utilisation d'une paire d'images permet de procéder par triangulation pour remonter à la détermination de la profondeur d'un point par rapport à la caméra en exploitant ses points homologues détectés dans les deux images. La triangulation consiste à déterminer l'intersection dans l'espace des deux droites projectives qui passent par les points homologues et le centre de projection de la caméra (Hartley et Zisserman, 2004 ; Rabaud, 2005). Ce modèle est différent du modèle des caméras acoustiques présenté et décrit dans la section 2.2 où l'ensemble des rétrodiffusions qui forment un pixel de l'image provient des régions de la scène se trouvant dans un arc de cercle ou un ruban très fin (cf. figure 3b). De ce fait, deux points homologues issus de deux images différentes ne permettent pas de procéder à une triangulation au sens du modèle sténopé. Dans le cas des images acoustiques, il est nécessaire de résoudre au mieux le système d'équations relatif à l'intersection de deux arcs de cercle dans la scène afin de déterminer l'élévation qui permettra de positionner un point de l'image dans l'espace 3D de la scène observée.

Ces différences résident deuxièmement au niveau des informations inconnues pour positionner en 3D les éléments de la scène. Ainsi, dans le cas des images acoustiques, la donnée d'élévation associée aux pixels des images représente l'information inconnue pour repérer ce pixel dans le monde réel. Dans le cas des images optiques, c'est l'information de profondeur (*i.e.* distance caméra-scène) qui est l'inconnue à lever et qui correspond à la portée pour les images acoustiques.

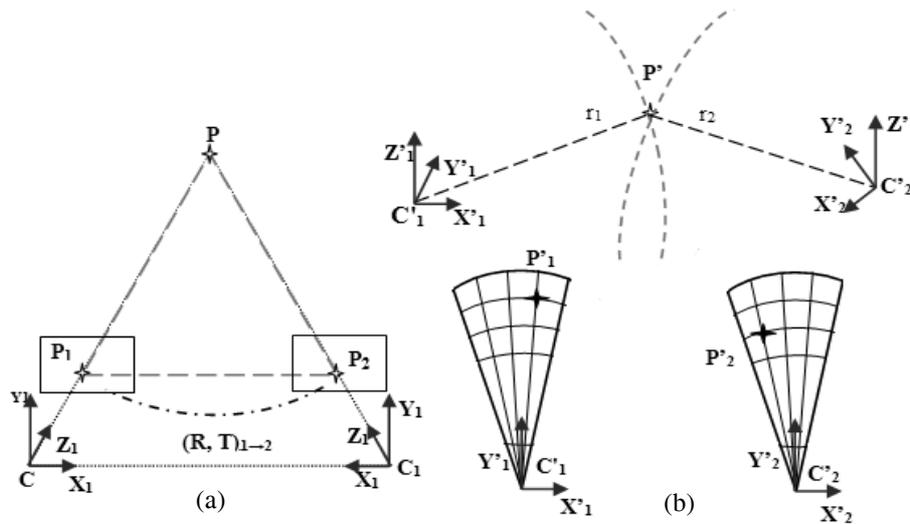


Figure 3. (a) Triangulation en imagerie optique. P_1 et P_2 sont les deux points homologues du point P . (b) Géométrie de projection en imagerie acoustique : P'_1 et P'_2 sont les deux points homologues du point P'

2.3.2. Validation du modèle géométrique des caméras acoustiques

Dans l'optique d'exploiter les images des caméras acoustiques pour une reconstruction 3D précise de l'environnement, il est important d'en savoir davantage sur l'exactitude du modèle géométrique idéal introduit dans la section 2.2. L'un des composants des caméras acoustiques intervenant dans le processus d'acquisition des images correspond aux lentilles acoustiques. L'utilisation de ces lentilles induit l'apparition de distorsions lors de la formation des images. Negahdaripour (2005) a procédé à la calibration de la caméra DIDSON en plaçant dans la scène une grille ayant une forte réflectivité et dont les coordonnées de chaque point sont connues. L'évaluation des erreurs dites erreurs de rectification entre les coordonnées de ces points et les coordonnées estimées à partir du modèle de la caméra a montré que la moyenne et la variance des erreurs résiduelles est de l'ordre de 0,13 cm pour la moyenne et de 3,99 cm² pour la variance dans le cas de la portée de 0.0003deg pour la moyenne et de 0.946 deg² pour la variance dans le cas de l'angle d'insonification (azimut) (Negahdaripour, 2005). En fonction de la précision visée par l'application, la prise en compte de ces erreurs dues aux lentilles peut s'avérer nécessaire.

2.3.3. *Nature bruitée et inhomogénéité radiométrique des images acoustiques*

Les caractéristiques radiométriques des images acoustiques font de la reconstruction 3D une problématique relativement difficile à résoudre. En effet, la reconstruction s'appuie exclusivement sur les informations enregistrées au sein des images. Une étape d'extraction de ces informations est ainsi nécessaire. Néanmoins, les images acoustiques sont de nature très bruitée, elles présentent un fort niveau de chatoiement (*i.e. speckle* en anglais) et une inhomogénéité radiométrique (Chailloux, 2007). L'apparition du speckle résulte entre autres, de la nature hétérogène du milieu marin et des phénomènes d'absorption et de diffusion des ondes acoustiques dans l'eau. L'inhomogénéité radiométrique est due à la variation de l'amplitude du signal rétrodiffusé entre le début et la fin de la portée compte tenu de l'absorption du signal par le milieu au cours de son trajet. Cette inhomogénéité se traduit par l'apparition de zones d'intensité très différente sur les images causant des ambiguïtés lors de la recherche d'une même information dans plusieurs images. En effet, deux pixels représentant une même région dans la scène peuvent avoir des caractéristiques radiométriques différentes sur des images différentes.

2.3.4. *Nature complexe de l'environnement sous-marin*

Le milieu marin est un environnement de travail complexe compte tenu notamment de son caractère incertain, évolutif et dynamique. Le caractère incertain et évolutif du milieu marin s'observe tout particulièrement au niveau de la définition des frontières ou des formes sur les images (Wright et Goodchild, 1997). Celles-ci peuvent apparaître floues et variables dans le temps, d'une image à une autre (ex. les sédiments, les herbiers, etc.). En conséquence, des ambiguïtés et des conflits peuvent être engendrés lors de l'interprétation des informations (ex. la localisation des objets, la délimitation des régions, etc.). De plus, la turbidité et la densité de la matière en suspension dans l'eau sont parmi les principales causes qui engendrent l'apparition du bruit dans les images acoustiques puisqu'elles constituent des obstacles à la propagation du signal émis vers les objets cibles. Et en fonction de la nature du fond, des objets, de la faune et de la flore, le signal réfléchi peut subir plusieurs réfractions et donc se propager selon des trajets multiples engendrant ainsi des phénomènes d'interférences. À ces incertitudes et imprécisions s'ajoute un autre élément de complexité : la variation de l'information suivant l'angle de prise de vue. Cette multiplicité d'informations partielles peut même conduire sans précaution à des conclusions conflictuelles. L'exploitation de la redondance et de la complémentarité observées dans une série d'images permet alors de confirmer ou/et rejeter une information et aboutir à une représentation fiable de la scène. Pour ce faire, les techniques de la fusion offrent des outils permettant d'une part, d'exploiter la redondance des informations lors de la manipulation d'une séquence d'images et, d'autre part, de modéliser l'imperfection des données manipulées et des traitements mis en jeu.

3. État de l'art des approches de reconstruction 3D en vision numérique et en imagerie sonar

3.1. Reconstruction 3D en vision numérique

La reconstruction 3D d'une scène consiste à déterminer la géométrie tridimensionnelle de celle-ci à partir d'une ou plusieurs images 2D. Pour ce faire, de nombreuses méthodes ont été développées dans le domaine de la vision par ordinateur. Dans le cadre du présent projet, on s'intéresse à la reconstruction à partir de plusieurs images à l'aide de méthodes issues de la vision passive. En effet, il existe deux grandes catégories de vision : la vision active et la vision passive. Au niveau de la vision optique active, les objets sont reconstruits à l'aide d'un éclairage contrôlé illuminant la scène d'une manière particulière (Roussel, 1999) contrairement à la vision passive qui n'utilise aucune source supplémentaire pour caractériser la scène. Les seules données disponibles sont les images de la scène issues de la caméra optique ce qui correspond au contexte du présent travail.

Étant donné que la caméra acoustique est à la fois l'émetteur et le récepteur des signaux exploités pour la formation de l'image, elle est classée dans la catégorie des capteurs de vision active. Néanmoins, les approches proposées dans la catégorie de la vision optique active ne répondent pas au contexte de mise en œuvre de ces caméras. En effet, les caméras acoustiques ne sont pas combinées avec d'autres outils (ex. système de projection d'un motif structurant (Roussel, 1999)) ce qui limite les informations géométriques pouvant être exploitées lors de la reconstruction de la scène observée au niveau de la variation des intensités des images acquises. Seules les images acoustiques sont exploitées dans le processus de reconstruction 3D, comme c'est le cas en vision passive pour les caméras optiques.

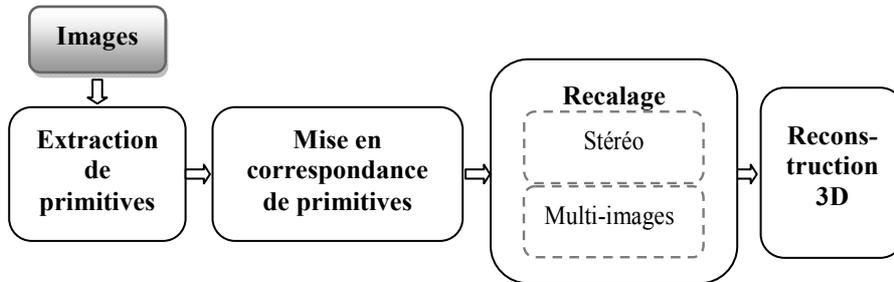


Figure 4. Principales étapes impliquées dans la reconstruction 3D de scènes selon une approche stéréoscopique ou multi-image

En vision numérique selon le nombre d'images utilisées, les approches de reconstruction peuvent être classées en deux catégories qui sont, respectivement, la stéréovision lorsqu'une paire d'images est utilisée et la multi-stéréovision lorsque plus de deux images sont utilisées. Pour ces deux catégories, la démarche générale

de la reconstruction 3D d'une scène est similaire et s'appuie sur les étapes indiquées à la figure 4.

L'extraction de primitives vise à localiser dans des images différentes les projections d'une même entité de la scène. La qualité de la reconstruction 3D va dépendre de la qualité de l'appariement entre ces projections identifiées sur les images d'où l'importance de l'étape d'extraction des primitives qui s'avère être un problème délicat. En fait, de nombreuses difficultés peuvent entraîner des résultats erronés lors de l'extraction de primitives telles que, le bruit présent dans les images et les occlusions (zones de la première image contenant des pixels sans correspondant dans la deuxième image). Ces primitives peuvent appartenir à différentes catégories telles que les primitives ponctuelles, curvilignes et surfaciques (Rosten *et al.*, 2010).

Le recalage est l'étape qui permet de calculer la transformation géométrique inter-images en s'appuyant sur l'ensemble de primitives homologues identifiées lors de la mise en correspondance (Rabaud, 2005). En fonction des primitives exploitées, on distingue le recalage iconique qui s'appuie sur les primitives pixels et le recalage symbolique qui s'appuie sur les primitives extraites à partir d'une étape de segmentation (par exemple, des primitives curvilignes ou surfaciques). Selon le nombre d'images utilisées, on distingue deux catégories de recalage à savoir le recalage par stéréovision (2 images) et recalage par multi-image (plus de 2 images). Le recalage est aussi classifié selon le type de transformation utilisée. Celle-ci peut être rigide ou déformable (Hartley et Zisserman, 2004 ; Mouragnon *et al.*, 2006 ; Rodriguez *et al.*, 2009).

3.2. Reconstruction 3D en imagerie sonar

Les premiers travaux qui ont mené aux travaux actuels de reconstruction 3D du fond marin et des objets par imagerie sonar étaient destinés à la navigation sous-marine. Les méthodes développées permettaient principalement de repositionner les robots sous-marins dans un environnement, en utilisant souvent conjointement plusieurs sources d'information (ex. des sonars, des cartes issues de modèles numériques de terrain). Avec l'amélioration de la qualité des images sonar, les méthodes de construction de cartes essaient de qualifier plus finement les points de référence de la navigation en introduisant des informations concernant le relief local. La problématique de reconstruction 3D du milieu sous-marin se retrouve également dans des applications de classification et de reconnaissance des objets sous-marins. A partir des informations présentes dans les images, l'objectif consiste à déterminer la classe des objets et leur identité à partir de la reconstruction de leur forme. Un état de l'art des travaux de reconstruction 3D à partir d'images sonar est présenté dans cette partie. Celle-ci est articulée suivant deux sections : la reconstruction mono-image et la reconstruction multi-image.

3.2.1. Reconstruction 3D d'une scène à partir d'une seule image acoustique (Shape From Shading)

L'approche *Shape From Shading* (SFS) est une méthode classique de reconstruction 3D en imagerie sonar à partir d'une seule image. Elle permet de calculer la variation des pentes locales des surfaces de la scène et ceci en se basant sur le fait que l'intensité acoustique réverbérée est modulée par l'angle d'insonification du sonar. Les sections suivantes décrivent les différentes approches de SFS recensées dans la littérature dans le domaine de l'imagerie sonar.

3.2.1.1. Approches SFS mono-capteur

Les approches SFS utilisées dans le contexte de l'imagerie acoustique sous-marine se répartissent selon les 3 catégories suivantes :

- les approches par minimisation : elles estiment le relief en s'appuyant sur la minimisation d'une fonction d'énergie composée de deux contrastes. Brooks et Horn (1985) utilisent un contraste de luminosité (*Brightness*) et un contraste de lissage de surface (*Smoothness*). Le contraste de luminosité exige que le relief reconstruit ait la même luminosité que l'image utilisée. Le contraste de lissage correspond à la construction d'une surface lisse. Zheng et Chellapa (1991) ont introduit le contraste de gradient à la place du lissage. La fonction d'énergie représente la variation des pentes des surfaces au niveau des images ;

- les approches par propagation : elles se basent sur la propagation de l'information de relief calculée au niveau d'un ensemble de points caractéristiques de la surface pour estimer le relief global de l'image (Dura *et al.*, 2004) ;

- les approches linéaires : elles s'appuient sur la linéarisation du modèle de réflexion. Dura *et al.* (2004) ont établi une relation entre l'intensité des images et la forme du fond *via* une transformation de Fourier. Comme il s'agit d'une relation non linéaire difficile à résoudre, ils ont procédé à la linéarisation de cette fonction en se basant sur des solutions proposées par Pentland (1990).

Dura *et al.* (2004) ont fait une comparaison en termes de précision et de similarité du relief reconstruit entre les approches linéaires et les approches de propagation. Leur étude a montré que les approches linéaires sont plus performantes quand il s'agit de reconstruire des surfaces isotropiques à partir d'images très bruitées. Les surfaces isotropiques se caractérisent par l'invariance de leur réflexion en fonction de la direction de propagation des ondes réfléchies. Par contre, les approches de propagation sont plus robustes pour les surfaces anisotropiques telles que les fonds sableux où il y a une dominance des zones d'ombre due à une grande variation de relief.

3.2.1.2. Les approches SFS multicapteurs

La majorité des travaux portant sur l'application du SFS sur des images sonar, recensés dans la littérature, visent à reconstruire le relief du fond marin ou le modèle numérique de terrain (MNT). Généralement l'application du SFS est combinée avec des mesures de bathymétrie. Langer et Hebert (1991) ont exploité l'intensité des

images sonar latéral, la géométrie d'acquisition du sonar et des données connues d'élévation (*i.e.* données bathymétriques) pour la reconstruction du relief. La technique utilisée correspond à l'approche de propagation du SFS. Cette technique prend en compte le modèle de réflexion de Lambert et considère que la surface du fond est constituée par un ensemble de segments au niveau desquels la pente de la surface est calculée. La fonction de propagation propage le calcul du relief sur la totalité des surfaces.

Sintes et Guériot (2005) utilisent une approche par minimisation pour calculer le relief. En effet, ils minimisent l'erreur quadratique entre le relief estimé par SFS et le relief reconstruit à partir de données interférométriques. Les travaux de Li et Pai (1991) et Johnson et Hebert (1996) ont aussi combiné des données bathymétriques et les mesures de pente calculées par SFS afin de reconstruire le relief. Selon Johnson et Hebert (1996), il est très difficile de reconstruire la surface tridimensionnelle en n'impliquant que l'information de rétrodiffusion des images. En effet, les données de relief résultantes sont locales et partielles et doivent être combinées avec d'autres informations pour augmenter leur robustesse. Les approches proposées dans ce contexte sont ainsi très limitées. Le plus souvent, faute de données bathymétriques disponibles à des échelles compatibles, ces approches ne sont pas validées à partir des données réelles mais simplement à l'aide d'images simulées. De plus, aucune analyse exhaustive mettant en jeu différents types de fond n'a été réalisée (Dura *et al.*, 2004). Quant aux objets reposant sur le fond ou sur des structures, leur reconstruction par SFS nécessite l'extraction de certains attributs complémentaires, par exemple les caractéristiques des ombres portées, ce qui conduit à des informations curvilignes telles que des profils. En effet, l'étendue de l'ombre dans le sens de la portée est proportionnelle à la hauteur de l'objet ou de l'obstacle qui l'a générée. En conséquence, l'introduction de cette information augmente la précision de la reconstruction des objets.

La méthode du Shape From Shading ne fournit que des informations locales concernant les variations de relief. Selon Sintes et Guériot (2005) travailler avec un relief issu uniquement de cette méthode est très difficile et imprécis. Par conséquent, reconstruire totalement une scène à partir d'une seule image ne semble pas envisageable. Le recours à des méthodes impliquant plusieurs images semble être une solution plus appropriée pour une représentation globale et fiable de la scène.

3.2.2. Reconstruction 3D à partir de plusieurs images acoustiques

La majorité des travaux de reconstruction 3D multi-image combinent les données de sonars imageurs, des cartes issues de modèles numériques de terrain et aussi des données issues de sonars interférométriques. Mais avec l'amélioration progressive de la qualité des images sonar tout particulièrement celles issues des caméras acoustiques, on assiste au développement de certaines approches basées uniquement sur les informations issues des images sonar. Dans les paragraphes qui suivent, quelques travaux de reconstruction 3D mono et multicapteur sont présentés.

3.2.2.1. Reconstruction 3D multicapteur

La reconstruction 3D à partir d'images sonar multivues exploite différentes approches. Certaines de ces approches dérivent du domaine de la stéréovision et d'autres impliquent l'utilisation de différentes sources d'informations (ex. données interférométriques). Parmi les travaux dont le principe dérive de la stéréovision, on peut citer ceux de Lu et Oshima (2002). Ils ont développé l'approche « Stereo-Like Vision » pour la reconstruction du fond marin à partir d'images sonar latéral. Cette approche se base sur la combinaison des données GPS et des données ponctuelles d'intensité issues de la méthode du MGPM (*i.e. Multi Step Gray Level Projective Matching*). La méthode du MGPM est une méthode de recalage qui s'appuie sur l'étude des histogrammes des intensités (*i.e. profil de distribution des niveaux de gris*). Les zones ayant un même profil correspondent à des zones identiques. Le principe de la stéréovision est largement utilisé lors de la combinaison d'images issues respectivement d'un sonar frontal et d'une caméra optique sous-marine. La combinaison entre ces données acquises simultanément permet d'appliquer la contrainte épipolaire et la triangulation stéréoscopique entre les deux types de données (*i.e. Optic-Acoustic Stereo*) (Negahdaripour, 2005 ; Caimi *et al.*, 2010). De telles méthodes exploitent des données complémentaires pour la reconstruction 3D à partir d'images sonar mais ne sont réellement utilisées que dans des contextes particuliers à cause des contraintes d'acquisition requises (ex. conditions météorologiques optimales pour l'exploitation des caméras optiques, une bonne synchronisation entre les différents capteurs utilisés, etc.). Pour éviter l'utilisation conjointe de multiples capteurs, d'autres approches ont été proposées.

3.2.2.2. Reconstruction 3D monocapteur

Sekkati et Negahdaripour (2007) proposent une approche de reconstruction 3D monocapteur. Ils se basent sur une approximation de la planéité de la surface de la scène et de la stationnarité du mouvement du sonar pour estimer le mouvement 3D du sonar (rotation et translation de la caméra) à partir des images DIDSON. En effet, ces approximations leur ont permis de simplifier le problème et de développer un modèle d'homographie sonar pour calculer les paramètres de transformation entre les images. L'extraction de l'information 3D d'une séquence d'images exploite aussi les algorithmes d'estimation de mouvement tels que le filtre de Kalman (Victor and Sentieiro, 1994) ou le flot optique pour l'extraction et le suivi de points d'intérêt (Zhen *et al.*, 2009). L'estimation de la position 3D de ces points se base sur le développement d'un modèle prenant en compte la géométrie et le déplacement du sonar d'une image à l'autre (Trucco et Curletto 2003).

Dans le cas spécifique de la reconstruction 3D d'objets reposant sur le fond marin, la méthode la plus couramment utilisée est la reconstruction volumétrique à partir de la fusion d'images multivues (Zerr et Stage, 1996 ; Sun *et al.*, 2008). Dans ce cas, la procédure d'acquisition de ces images est généralement contrôlée. En effet, le capteur réalise une circumnavigation de l'objet en visant son centre et en gardant la même altitude par rapport au fond. La reconstruction 3D est déduite à partir de la combinaison de deux fonctions complémentaires issues de l'exploitation des informations des échos (fonction de la réflexion) et des ombres (fonction

d'élévation) (Zerr et Stage, 1996 ; Sun *et al.*, 2008). Les paramètres de la fonction de réflexion sont déterminés en appliquant aux échos une méthode inspirée de la tomographie. La fonction d'élévation des objets est établie en fonction de la longueur des ombres, de la hauteur du sonar et de sa portée (Zerr et Stage 1996).

Les travaux de reconstruction impliquant plusieurs images sonar ne sont pas nombreux. La majorité des exemples cités dans le paragraphe précédent sont récents et utilisent des images acquises par des sonars frontaux. En effet, les sonars frontaux permettent l'acquisition d'images multivues d'une même scène, d'une manière plus souple que d'autres sonars tels que les sonars latéraux pour lesquels une configuration d'acquisition particulière est requise (ex. une circumnavigation de l'objet).

4. Description de la méthodologie de traitement proposée

4.1. Fondement de la méthodologie proposée

Suite à l'étude de différentes approches proposées en vision numérique et en imagerie sonar et des caractéristiques des caméras acoustiques, nous avons procédé à l'identification des différents composants devant intervenir dans la reconstruction 3D (qui inclut le calcul du mouvement de la caméra entre les images et les coordonnées 3D de la scène) et au design de l'architecture la plus adéquate permettant de combiner les informations issues de ces composants tout en prenant en compte les incertitudes, imprécisions et ambiguïtés qui les caractérisent.

La particularité de la méthodologie proposée est l'adaptation de la stratégie de reconstruction 3D largement utilisée en vision numérique (section 3.1) à l'imagerie acoustique. En fait, les spécificités de cette stratégie relativement à l'imagerie acoustique sont capturées par les différents composants de la méthodologie. En effet, compte tenu des différences existant entre les images respectivement acoustiques et optiques, il est nécessaire d'adapter voire de proposer de nouvelles approches et algorithmes qui non seulement répondent aux objectifs de chaque composant mais aussi qui tiennent compte des caractéristiques des caméras acoustiques (ex. géométrie de formation, nature bruitée et caractéristiques des capteurs). Par conséquent, l'originalité de la méthodologie proposée repose sur le fait qu'elle a été confrontée aux problématiques inhérentes aux images acoustiques et notamment à l'exploitation de séquences d'images acquises avec des caméras. Même si le flot de traitements semble classique dans un contexte de vision optique, leur enchaînement et les composants impliqués dans l'architecture proposée ont été spécifiés sur la base des travaux se rapportant au sujet abordé recensés dans la littérature et en tenant compte des problématiques inhérentes au milieu sous-marin et aux caméras acoustiques.

Ces composants, pour lesquels la nature des données d'entrée, les traitements impliqués et le résultat recherché ont été spécifiés, sont illustrés à la figure 5. Celle-ci propose également une représentation de l'architecture adoptée.

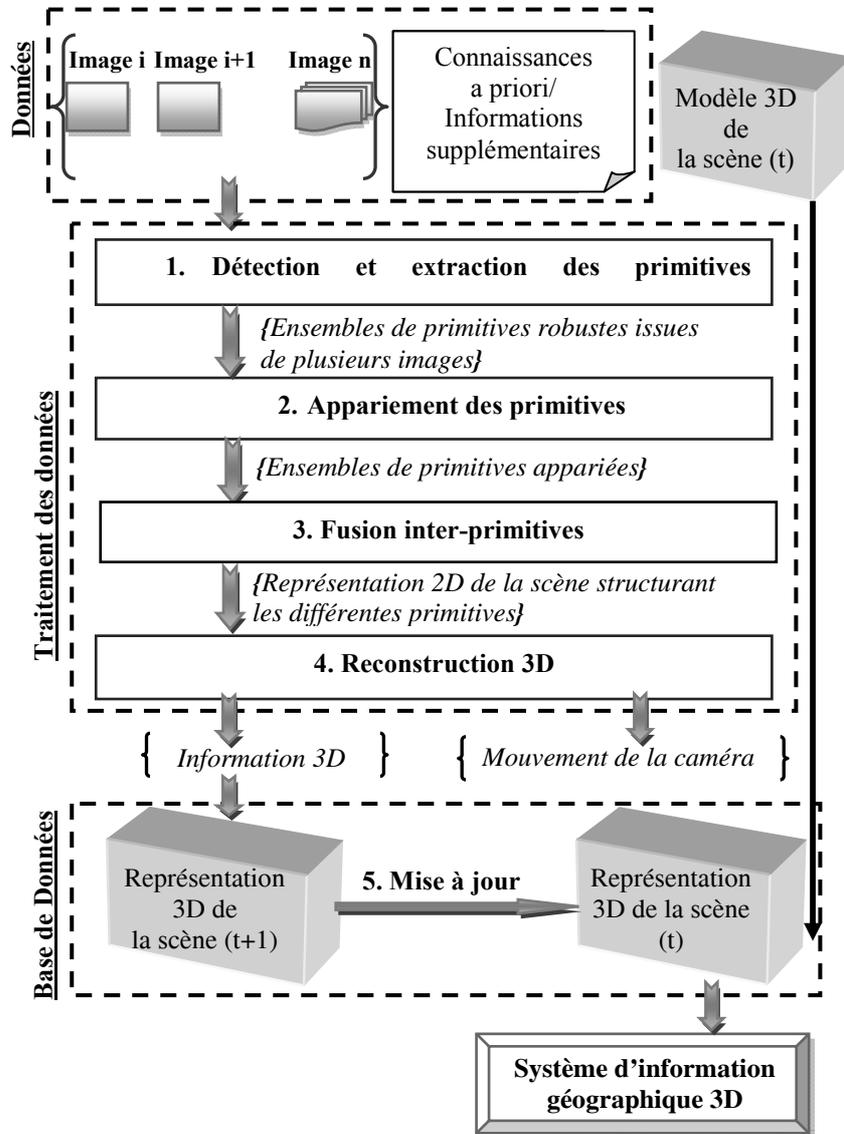


Figure 5. Architecture de la méthodologie proposée

4.2. Description générale de l'architecture proposée

Nous proposons une architecture séquentielle qui, à partir d'une séquence d'images et de connaissances *a priori* (ex. modèle de la caméra, cadence des images,

etc.), permet de calculer des caractéristiques tridimensionnelles intervenant dans la reconstruction 3D de la scène observée et/ou sa mise à jour. Le choix d'une telle architecture repose sur le contexte d'acquisition des images qui sont enregistrées sous forme de séquences. Le traitement et l'extraction d'information des images sont donc réalisés au fur et à mesure de leur acquisition. Cette architecture met l'accent sur la fiabilité et la robustesse de chaque traitement individuellement en y associant un module de fusion et en y exploitant une diversité d'information complémentaire ou redondante. L'approche se veut simple mais efficace, les mécanismes de raisonnement étant inclus implicitement dans l'articulation des composants de l'architecture et explicitement au sein du module de fusion.

L'architecture est constituée principalement de quatre modules dont le résultat permet d'alimenter une base de données descriptive du modèle 3D de la scène observée. Une telle base de données peut être exploitée au sein d'un système d'information géographique ou d'aide à la décision. Une description détaillée de ces modules est présentée dans les sections suivantes.

4.2.1. Module 1 : détection et extraction de primitives robustes

L'image acoustique contient une multitude de primitives de différentes natures. La première étape de l'architecture de la reconstruction 3D consiste à extraire ces primitives qui comprennent les points, contours etc., présents dans l'image acoustique. Pour les extraire trois approches peuvent être appliquées :

- approche ponctuelle : extraction de points saillants (ex. coins, points de contour) ;
- approche curviligne : extraction de contours (les zones de forte variation d'intensité) ;
- approche surfacique : extraction de régions d'intensité homogène.

Dans le contexte d'une reconstruction tridimensionnelle, l'exploitation des approches ponctuelles s'avère une solution adéquate. À travers les approches ponctuelles, il est possible d'identifier et d'extraire les coins, les points de courbure de contours d'un même objet présent dans plusieurs images successives acquises selon des angles de vue différents. Ces coins détectés dans plusieurs images sont d'autant plus intéressants qu'ils sont le reflet de caractéristiques géométriques saillantes de la scène et de ce fait, chaque série de coins extraits de plusieurs images correspond à une position géographique unique dans la scène. Pour ce faire, l'idée proposée se compose de trois phases. La première consiste à extraire des contours dans l'objectif d'identifier les emplacements potentiels des points représentatifs de la géométrie des objets. En effet, les contours étant provoqués par de forts contrastes géométriquement consistants, de tels points saillants doivent s'y retrouver. Par ailleurs, une telle démarche permettra de restreindre l'espace de recherche des primitives ponctuelles. La deuxième phase concerne l'élimination des faux contours. Elle est nécessaire pour minimiser l'effet du bruit présent dans les images acoustiques qui provoque l'extraction de contours ne reflétant pas la géométrie de la scène. Néanmoins, de par leur nature, ces contours non pertinents ne se retrouvent

pas par exemple, sur plusieurs images successives de la même scène ou bien présentent une robustesse moindre à une analyse multi-échelle d'une seule image acoustique. Finalement, ayant à disposition uniquement les contours pertinents de la scène, les points saillants peuvent être extraits lors de la troisième phase.

4.2.2. Module 2 : appariement des primitives

Composant fondamental dans l'architecture de la méthodologie proposée, l'appariement des primitives est utilisé pour appairer les primitives extraites des images acquises selon différents points de vue. Les techniques d'appariement de primitives peuvent être divisées en deux catégories (Nillius et Eklundh, 2002). La première catégorie dite appariement iconique se base sur une ressemblance au niveau pixel, des sous-images ou fenêtres issues des images (*block matching* en anglais). En revanche, la deuxième catégorie ou appariement symbolique, opère non pas sur les intensités des pixels des images mais sur les propriétés des primitives extraites de ces images.

Ainsi, les méthodes de correspondance par bloc s'appuient essentiellement sur la corrélation croisée de l'image avec une image de référence (Nillius et Eklundh, 2002). Elles sont les plus utilisées à cause de leur robustesse au bruit. Par contre, elles atteignent leurs limites en présence de distorsions géométriques ou de rotations d'où la nécessité de prendre en considération l'orientation des fenêtres de corrélation en fonction de la prise de vue.

Quant à la deuxième catégorie, elle concerne la mise en correspondance d'ensembles de primitives symboliques telles que les points saillants. Cette catégorie est plus avantageuse particulièrement dans un contexte où déformations et distorsions entre les images sont envisageables (typiquement, une observation multiple par une caméra acoustique selon différents points de vue). Certaines approches d'appariement reposant sur la mise en correspondance de points s'appuient d'une part, sur une étape de détection de points saillants (c'est le cas dans l'architecture proposée dans le présent travail), et, d'autre part, sur un critère de similarité caractérisant ces points, critère qui fait souvent intervenir le voisinage de ces points (méthode SIFT (Lowe, 2004), par exemple).

4.2.3. Module 3 : fusion inter-primitives

Les étapes précédentes fournissent des ensembles de primitives de même type, appariées (un ensemble d'agrégats de points saillants, un autre ensemble d'agrégats de contours...). Le rôle de cette troisième étape est d'établir des liens ou des relations sémantiques entre les primitives appariées de même type (lien homogène) mais aussi de types différents (liens hétérogènes). L'idée maîtresse de cette étape est de permettre de replacer les différentes primitives extraites, de tout type, dans un référentiel commun, à travers par exemple des relations spatiales décrivant les positions relatives des entités les unes par rapport aux autres.

Ainsi le module 3 de l'architecture par l'établissement de relations homogènes ou hétérogènes entre primitives permet d'aboutir à une représentation multiprimitive de la scène étudiée.

En effet, en premier lieu, l'approche consiste essentiellement à identifier les relations topologiques entre primitives de la scène. Ce processus de fusion regroupe des primitives de tout type, tout en assurant la cohérence géométrique (les coordonnées géométriques dans la scène) et la logique spatiale des primitives (ex. les coins sont aux extrémités des contours, etc.). La connaissance de la sémantique de ces contraintes caractérisant les relations spatiales inter-primitives, est essentielle à exploiter dans l'objectif d'aboutir à une représentation multiprimitive de la scène pour en déduire une reconstruction 3D ultérieure. En second lieu, les informations associées aux primitives de même nature issues d'images différentes, sont par essence, incomplètes et souvent insuffisantes pour caractériser complètement la scène. Dans ce cas, la fusion proposée exploite la complémentarité des informations portées par les primitives, et permet de combiner des primitives différentes afin d'obtenir une représentation riche et plus complète de la scène.

4.2.4. Module 4 : reconstruction 3D

L'obtention d'un ensemble de primitives 2D appariées et dotées de relations topologiques créant ainsi une représentation robuste multiprimitive de la scène étudiée, permet de procéder à l'étape de reconstruction 3D. En s'appuyant sur la représentation ainsi construite, il s'agit de calculer l'élévation manquante (φ) de ces primitives dans la scène ainsi que d'estimer le mouvement de la caméra entre les différentes acquisitions. L'estimation du mouvement de la caméra correspond au calcul de deux paramètres : la translation (T) et la rotation (R).

Dans le cadre de cet article, en première approche, on travaillera uniquement avec des primitives de type point saillant, primitives robustes, extraites à partir de contours et tirant profit de la redondance d'observation de la scène entre plusieurs images. L'algorithme d'extraction utilisé est décrit dans la section suivante.

L'algorithme de reconstruction qui est envisagé dans cette première approche repose sur les hypothèses suivantes :

- le déplacement du capteur entre l'acquisition de l'image I_1 et l'acquisition de l'image I_2 est modélisé par la composition d'une translation T (T_x, T_y, T_z) et d'une rotation R définie par trois angles ($\alpha_x, \alpha_y, \alpha_z$) qui correspondent à des angles de roulis, tangage et lacet. Cette transformation permet donc de passer du repère de l'image I_1 (O_1, X_1, Y_1, Z_1) au repère de la deuxième image I_2 (O_2, X_2, Y_2, Z_2) ;
- selon le modèle géométrique des caméras acoustiques présenté dans la section 2.2, les coordonnées de deux primitives correspondantes (P_1) et (P_2) s'expriment dans leurs repères respectifs d'acquisition en coordonnées polaires et cartésiennes comme suit :

$$P_1 = \begin{pmatrix} x_P^1 \\ y_P^1 \\ z_P^1 \end{pmatrix} = \begin{pmatrix} [r_1 \sin(\theta_1)] \cos(\varphi_1) \\ [r_1 \cos(\theta_1)] \cos(\varphi_1) \\ [-r_1] \sin(\varphi_1) \end{pmatrix} = \begin{pmatrix} a_1 \cos(\varphi_1) \\ b_1 \cos(\varphi_1) \\ c_1 \sin(\varphi_1) \end{pmatrix} \quad (2)$$

$$P_2 = \begin{pmatrix} x_P^2 \\ y_P^2 \\ z_P^2 \end{pmatrix} = \begin{pmatrix} [r_2 \sin(\theta_2)] \cos(\varphi_2) \\ [r_2 \cos(\theta_2)] \cos(\varphi_2) \\ [-r_2] \sin(\varphi_2) \end{pmatrix} = \begin{pmatrix} a_2 \cos(\varphi_2) \\ b_2 \cos(\varphi_2) \\ c_2 \sin(\varphi_2) \end{pmatrix} \quad (3)$$

Avec ces hypothèses, les coordonnées P'_1 de la primitive P_1 , dans la scène, exprimées dans le repère d'acquisition de I_2 peuvent s'écrire *via* l'équation (4)

$$P'_1 = R.P_1 + T \quad (4)$$

Sachant que les primitives (P'_1) et (P_2) représentent le même point de la scène puisque (P_1) et (P_2) sont appariées, l'objectif de la reconstruction est de retrouver la translation T , la rotation R et les élévations φ_1 et φ_2 qui minimisent la distance $\|P'_1 - P_2\|$. Idéalement, ces inconnues devraient être la solution du système d'équations (5), ce qui correspond globalement à la recherche de l'intersection de deux arcs de cercle dans l'espace.

$$\left\{ \begin{array}{l} a_2 \cos(\varphi_2) = a_1 \cos(\varphi_1)R_{1,1} + b_1 \cos(\varphi_1)R_{1,2} + c_1 \sin(\varphi_1)R_{1,3} + T_x \\ b_2 \cos(\varphi_2) = a_1 \cos(\varphi_1)R_{2,1} + b_1 \cos(\varphi_1)R_{2,2} + c_1 \sin(\varphi_1)R_{2,3} + T_y \\ c_2 \sin(\varphi_2) = a_1 \cos(\varphi_1)R_{3,1} + b_1 \cos(\varphi_1)R_{3,2} + c_1 \sin(\varphi_1)R_{3,3} + T_z \end{array} \right\} \quad (5)$$

Les paramètres (a_1, b_1, c_1) et (a_2, b_2, c_2) sont connus, ils proviennent respectivement des équations (2) et (3), ($R_{1,1}, \dots, R_{3,3}$) sont les paramètres inconnus de la matrice de rotation R et (T_x, T_y, T_z) sont les paramètres inconnus du vecteur de la translation.

Il est évident qu'il est impossible de résoudre ce système avec seulement deux primitives appariées. Seule l'exploitation conjointe d'un certain nombre d'autres primitives issues des mêmes images et donc partageant les mêmes inconnues R et T , permettra de disposer de suffisamment d'équations pour le nombre d'inconnues. Si le nombre de primitives appariées entre deux images n'est pas suffisant, il est possible de considérer des primitives appariées sur plus de deux images, en supposant toutefois que le mouvement du capteur durant ces acquisitions soit uniforme.

Par ailleurs, comme déjà évoqué, la résolution de la caméra en portée et en azimut conduit à ne pas considérer des arcs de cercle mais des morceaux de ruban comme lieux géométriques possibles des primitives dans la scène. Ceci introduit donc une description plus complexe des coordonnées des primitives dans la scène et

de ce fait, une complexité accrue dans l'expression de leurs caractéristiques 3D (principalement le volume des coordonnées possibles).

L'élévation (plus exactement le volume des coordonnées possibles) calculée dans ce module pour chaque primitive, est ajoutée comme un attribut supplémentaire à la représentation des entités de la scène (par rapport à une description strictement image qui ne contient qu'azimut et distance). La représentation 3D ainsi obtenue représente les éléments de la scène en trois dimensions (ex. les objets, le fond, etc.) en y incluant leurs relations topologiques. Cette représentation est ensuite complétée et/ou mise à jour au fil des nouvelles images acquises et des primitives appariées, primitives communes ou non avec celles déjà incluses dans la représentation.

La résolution du système d'équations présenté dans cette section et la représentation 3D de la scène seront abordées dans des travaux futurs.

5. Résultats préliminaires

Les résultats préliminaires présentés dans cette section concernent tout particulièrement le premier module de l'architecture présentée dans la section précédente, avec une implémentation robuste du processus d'extraction de points saillants. C'est une étape cruciale puisqu'elle fournit les primitives ponctuelles qui sont utilisées dans les modules suivants.

5.1. Données utilisées

Les données utilisées pour tester l'approche développée sont un ensemble d'images acquises dans deux milieux différents à savoir : un bassin intérieur et le milieu marin réel. Le contenu de la scène observée varie également pour chacun des jeux de données exploités, tels que décrits ci-dessous :

- jeu de données n° 1 : une séquence d'images acquise en mode haute fréquence représentant plusieurs objets localisés sur le fond d'un bassin intérieur (piscine de l'Université Laval, Québec) d'une profondeur de 3m (figure 6a) ;
- jeu de données n° 2 : une séquence d'images représentant un pied de pont acquise en mode haute fréquence dans le milieu marin (figure 6b) ;
- jeu de données n° 3 : une séquence d'images représentant une épave acquise en mode basse fréquence dans le milieu marin (figure 6c).

Le mouvement de la caméra entre les images exploitées dans cette partie n'est pas connu. C'est lors de l'étape de la reconstruction 3D qui n'est pas présentée dans cet article que le mouvement de la caméra entre les images est calculé. Ce calcul permet par la suite d'identifier la nature du mouvement qu'il soit de type translation horizontale et/ou rotation entre les différentes images.

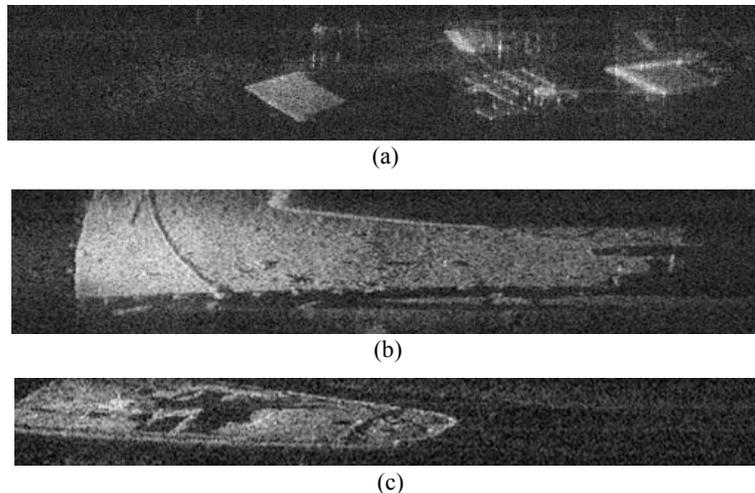


Figure 6. Exemples d'images extraites de séquences acquises avec la caméra DIDSON : a) séquence des objets en bassin ; b) séquence d'un pied de pont en milieu marin ; c) séquence d'une épave en milieu marin

5.2. Approche d'extraction de points caractéristiques des contours

La méthode vise en premier lieu à identifier les contours les plus pertinents présents dans les images puis à extraire de ces contours un ensemble de points caractérisant les zones de forte courbure des contours. Les différentes étapes intervenant dans cette approche sont les suivantes :

- 1. Détection de contours : le filtre Canny (Grigorescu *et al.*, 2004) a été utilisé pour extraire les contours des objets. Malgré la robustesse au bruit de cette technique, des faux contours sont aussi détectés parmi les contours pertinents ;
- 2. Mise en correspondance des contours afin d'éliminer les faux contours : pour éliminer les faux contours nous avons développé une nouvelle approche qui se base sur l'hypothèse d'un faible déplacement de la caméra entre deux images successives. Cette approche vérifie la présence de deux contours supposés homologues dans les images. Une recherche est menée dans l'image 2 dans un voisinage restreint autour de la position qu'occupent les points du contour C1 dans l'image 1. Cette recherche permet de déterminer le pourcentage des pixels représentant C1 dans l'image 2. Les contours détectés sur une seule image et ceux représentés par un très faible nombre de points sont ainsi éliminés ;
- 3. Analyse multi-échelle pour l'extraction des contours pertinents : cette étape vient renforcer l'étape précédente. En effet, ces deux étapes sont liées. Une pyramide gaussienne d'images (Lakshmi et Rakshit, 2010) est construite conduisant

à une représentation multi-échelle de l'image originale. A chaque niveau d'échelle, pour lequel le niveau de bruit est réduit suite à l'application d'un filtre gaussien, on procède à une mise en correspondance des contours détectés entre deux images consécutives en utilisant l'approche décrite à l'étape 2. L'approche de mise en correspondance est effectuée jusqu'au niveau d'échelle pour lequel le nombre de contours pertinents est stable avant de décroître rapidement pour les échelles suivantes. La persistance des contours à travers différents niveaux d'échelle est significative de la robustesse de ces contours ;

– 4. Extraction des points saillants : une fois les contours pertinents détectés, les points saillants peuvent être extraits. Ces points saillants représentent les points de haute courbure des contours.

5.3. Exemple de résultats obtenus

Les figures suivantes présentent les résultats obtenus par application de l'approche d'extraction des points caractéristiques proposée sur les jeux de données décrits en section 5.1. Ces résultats préliminaires sont satisfaisants étant donné que les points détectés sont représentatifs de la forme des objets. De plus, ils ont été identifiés sur plusieurs images successives ou décalées en temps de la séquence. Ceci confirme la possibilité de procéder à la reconstruction 3D stéréo aussi bien que multi-image en s'appuyant sur de tels points caractéristiques.

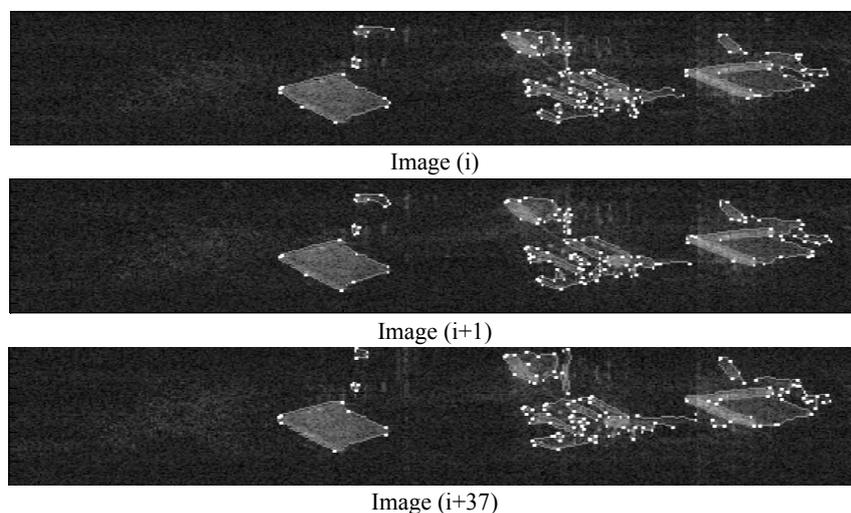


Figure 7. Extraction de points saillants sur plusieurs images consécutives extraites de la séquence représentant des objets dans un bassin intérieur

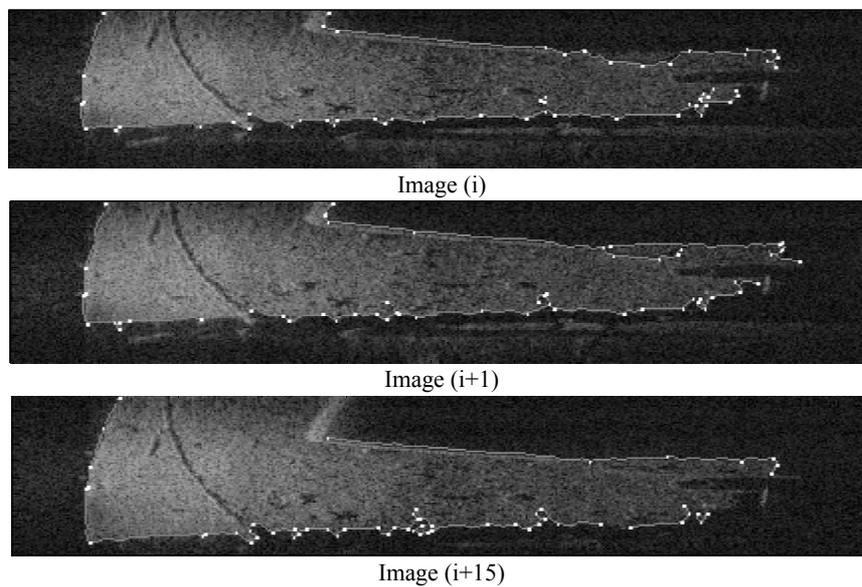


Figure 8. Extraction de points saillants sur plusieurs images consécutives extraites de la séquence représentant un pied de pont en milieu marin

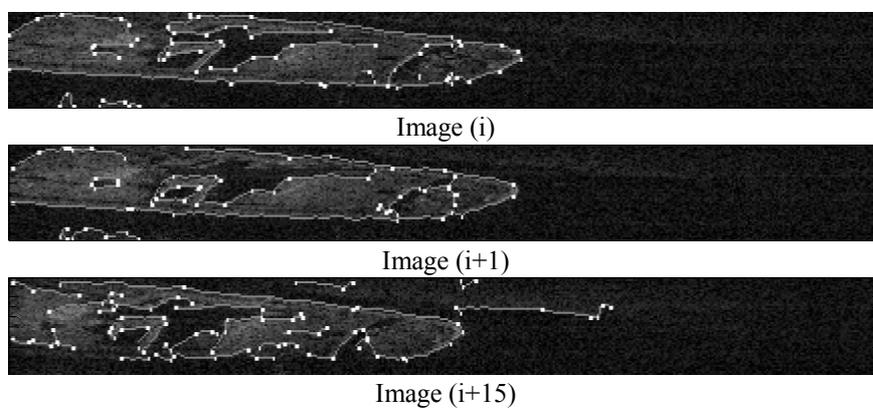


Figure 9. Extraction de points saillants sur plusieurs images consécutives extraites de la séquence représentant une épave en milieu marin

5.4. Évaluation quantitative des résultats présentés

Afin d'évaluer la performance de l'approche de détection de points proposée, les trois critères suivants ont été calculés : le taux de répétabilité, le taux de bonne

détection et la bonne localisation des points. Les résultats sont illustrés dans le tableau 1.

5.4.1. Le taux de répétabilité

Le taux de répétabilité, exprimé en pourcentage, correspond au rapport du nombre de points répétables dans une image sur le nombre de points détectés dans une image de référence. Dans l'objectif d'une reconstruction 3D multivue, ce taux permet d'extraire les images ayant un taux de répétabilité de points important. Selon les résultats obtenus et affichés dans le tableau 1, au fur et à mesure que les images sont espacées dans la séquence le taux de répétabilité diminue. En effet, lorsque deux images successives sont utilisées, il y a plus de points qui se répètent (pied de pont : 70 %, épave : 53 %, piscine et cube : 88 %), que lorsque les deux images sont espacées par 15 ou 30 enregistrements. La diminution de ce pourcentage est essentiellement due au changement du point de vue. Néanmoins, même avec ce changement de point de vue, les points qui caractérisent la partie de la scène commune entre les images sont détectés.

Le tableau 1 présente le taux de répétabilité pour trois configurations différentes : images successives τ_1 (%); %, images espacées de 15 images dans la séquence τ_2 (%); images espacées de 30 images dans la séquence τ_3 (%).

5.4.2. Bonne détection

Le taux de bonne détection quantifie la pertinence des points détectés en construisant le rapport du nombre de points détectés caractérisant la géométrie de la scène sur le nombre total de points détectés. Afin de calculer ce critère, les points caractéristiques de la géométrie de la scène ont été extraits manuellement par un opérateur, formant ainsi un ensemble de points de référence. Le tableau 1 rassemble différents résultats relatifs à ce critère de bonne détection. Ainsi pour chaque jeu d'images, on retrouve le nombre de points extraits respectivement manuellement (R) et automatiquement (A), le nombre de points détectés simultanément par ces deux approches (VP), le nombre de points du jeu de référence n'ayant pas été détectés automatiquement (VN) et le nombre de points détectés automatiquement n'ayant pas de correspondant dans le jeu de référence (FA).

Les résultats obtenus mettent en évidence que l'approche proposée se caractérise par une bonne détection de points caractéristiques de la géométrie de la scène, puisque sur l'ensemble des points extraits manuellement, plus de 80 % ont été détectés par l'approche d'extraction automatique. Les points supplémentaires détectés par l'approche automatique caractérisent essentiellement les zones d'ombre (tout particulièrement dans la séquence de la piscine). Il est logique que ces points supplémentaires ne fassent pas partie des jeux de référence puisqu'ils ne sont pas utilisés dans l'approche de reconstruction 3D proposée. Afin de ne conserver parmi les points détectés automatiquement que les points du jeu de référence, il conviendrait d'analyser le voisinage sur les images, des points détectés automatiquement pour vérifier la présence d'ombre. Néanmoins, ces points de contours d'ombre détectés automatiquement, caractérisent la géométrie des ombres

provoquées par certains éléments de la scène et donc des profils de relief de ces éléments. Ces points détectés associés aux ombres portées pourraient alors servir de support à de nouvelles primitives curvilignes renseignant sur certains profils d'éléments de la scène.

Tableau 1. Évaluation du taux de répétabilité, de la bonne détection et de la localisation des points pertinents extraits selon l'approche proposée

Image	1) Taux de répétabilité (%)			2) Détection (R et A en Points, VF, VN, FA en %)				
	τ_1	τ_2	τ_3	R	A	VP	VN	FA
Piscine	86	81	77	55	122	42	5.45	58
Pied de pont	70	52	19	49	57	71	18	29
Épave	53	47	30	40	65	47	22.5	53

5.4.3. Bonne localisation

Ayant les coordonnées images des points extraits manuellement, nous avons procédé à la comparaison de leurs coordonnées avec celles des points extraits automatiquement. Les erreurs estimées en portée et en azimut sont données dans le tableau 2 suivant.

Tableau 2. Évaluation de l'erreur de la localisation des points extraits automatiquement en comparaison avec les points extraits manuellement

	Portée [min, max] [moyenne, écart type] [cm]	Azimut [min, max] [moyenne, écart-type] [degré]
Pied de pont	[0.0, 7.0], [2.2, 2.1]	[0.0, 0.06], [0.0, 0.0]
Épave	[0.0, 13.0], [6.2, 4.8]	[0.0, 1.2], [0.1, 0.0]
Piscine	[0.0, 5.0], [2.0, 1.6]	[0.0, 0.604], [0.0, 0.0]

Selon les résultats affichés dans le tableau 2 les moyennes des erreurs sont égales à zéro en azimut et de l'ordre de 2 [cm] en portée pour les deux séquences « Pied de pont » et « Piscine », ceci nous permet de juger que l'algorithme extrait les points avec une bonne localisation. Quant à la séquence « Épave » les points sont extraits avec une moyenne de 0,1 degré en azimut et de 6,2 cm en portée. Ceci est dû probablement au mode d'acquisition de cette séquence (basse fréquence) qui fait que l'algorithme extrait les points des images avec moins de précision que des images acquises en haute fréquence telles que celles des séquences « Pied de pont » et « Piscine ». L'évaluation des erreurs issues de l'étape d'extraction de points est importante. Elles seront prises en considération lors de l'évaluation finale des erreurs de la reconstruction 3D de la scène.

6. Conclusion

La reconstruction tridimensionnelle du milieu marin est d'une grande importance dans le contexte de son exploitation pour observer les impacts des changements climatiques sur ce milieu. Néanmoins, pour ce faire il faut disposer d'un ensemble de données précises, détaillées et qui soient acquises d'une façon continue afin de pouvoir procéder à la reconstruction de l'environnement dans ses trois dimensions. Nous avons proposé dans cet article d'exploiter les caméras acoustiques compte tenu de leur adéquation vis-à-vis de la problématique abordée : elles permettent d'acquérir des images dans les milieux profonds et turbides et offrent des séquences d'images de haute cadence.

Cet article a été spécifiquement consacré à la description de la méthodologie globale de traitement que nous avons proposée afin de guider la conception d'une solution de reconstruction tridimensionnelle de l'environnement marin qui soit précise, robuste et représentative de la réalité observée. Les composants de cette méthodologie consistent en une détection et extraction de primitives robustes, un recalage multivue, une fusion inter-primitives, une reconstruction 3D et une mise à jour du modèle 3D. Aux traitements de reconstruction s'ajoutent les approches de fusion et les informations supplémentaires qui vont permettre d'aboutir à une reconstruction complète de la scène.

Les premiers travaux effectués correspondent à la première étape de l'architecture proposée. Elle concerne l'extraction de points caractéristiques de la scène. C'est une étape assez délicate puisqu'elle fournit les points nécessaires pour effectuer le recalage inter-images et la reconstruction 3D. Des résultats préliminaires ont été obtenus sur 3 jeux de données différents démontrant la performance et la robustesse de la méthode. Néanmoins, dans l'approche appliquée, la problématique liée à la présence de bruit dans les images n'a pas pu être totalement surmontée. Ce bruit affecte les informations extraites et les résultats obtenus et rend difficile la manipulation des images des caméras acoustiques.

Nos travaux futurs vont aborder le développement et la mise en œuvre d'approches de recalage des images et de la reconstruction 3D. Ces approches devront tenir compte de la géométrie spécifique des images des caméras acoustiques afin de reconstruire la scène d'une façon pertinente.

Remerciements

Ce travail de recherche a été réalisé avec l'appui financier du Conseil Franco-Québécois de Coopération Universitaire (CFQCU).

Bibliographie

- Belcher E. O., Barone J. R., Gallagher D. G., Honaker R. E. (2003). Acoustic lens camera and underwater display combine to provide efficient and effective hull and berth inspections. *Proc. Oceans'2003*, 22-26 September, San Diego, USA. p. 22-26.

- Brahim N., Guériot D., Daniel S., Solaiman B. (2010). 3D reconstruction of underwater scenes using image sequences from acoustic camera. *OCEANS 2010 IEEE - Sydney*, 24-27 May, Sydney, Australia.
- Brooks M. J., Horn B. K. P. (1985). Shape and source from shading. *IJCAI'85 Proceedings of the 9th international joint conference on Artificial intelligence*, vol. 2, p. 932-936.
- Caimi F. M., Kocak D. M., Dalgleish F., Watson J. (2010, April). Underwater imaging and optics- recent advances. *IEEE Oceanic Engineering Society Newsletter*, p. 21-29.
- Chailloux C. (2007). *Recalage d'images sonar par appariement de régions. Application à la génération d'une mosaïque*. Thèse de doctorat, ENST Bretagne, Université de Rennes.
- Dura E., Bell J., Lane D. (2004). Reconstruction of textured seafloors from side-scan sonar images. *IEE Proc., Radar, Sonar Navig.*, p. 114-126.
- GIEC (2007). *Bilan 2007 des changements climatiques*. Rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat.
- Grigorescu C., Petkov N., Westenberg M. A. (2004). Contour and boundary detection improved by surround suppression of texture edges. *Image and Vision Computing*, vol. 22 n° 8, p. 609-622.
- Hartley R., Zisserman A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge, UK, Cambridge University Press.
- Johnson A.E., Hebert M. (1996). Seafloor Map Generation for Autonomous Underwater Vehicle Navigation. *Auranomous Robors*, p. 145-168.
- Jun H., Asada A. (2007). Acoustique counting method of upstream juvenile Ayu *Plecoglossus altivelis* by using DIDSON. *Symposium on Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies*, vol. 1 et 2, p. 459-462.
- Lakshmi A., Rakshit S., (2010). Gaussian Restoration pyramid: Application of image restoration to Laplacian pyramid compression. *Advance Computing Confrence (IACC)*. 19-20 February, p. 66-71.
- Langer D. Hebert M. (1991). Building Qualitative Elevation Maps From Side Scan Sonar Data For Autonomous Underwater Navigation. *Proceedings of the International Conference on Robotics and Automation*, vol. 3, p. 2478-2483.
- Li R., Pai S. (1991). Improvement of Bathymetric Data Bases by Shape from Shading Technique using Side-Scan Sonar Images. *Proc. MTS/IEEE Oceans 1991*, p. 320-324.
- Lowe G.D. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *Inter. Journal of Computer Vision*, vol. 60, n° 2, p. 91-110.
- Lu Y., Oshima M. (2002). 3-D reconstruction of the underwater acoustic images using the GPS positioning and the image matching technology. *Proc. MTS/IEEE Oceans 2002*, 29-31 October, vol. 4, p. 2273-2278.
- Mouragnon E., Lhuillier M., Dhome M., Dekeysery F., Saydy P. (2006). 3D reconstruction of complex structures with bundle adjustment: an incremental approach. *Proc. of the IEEE International Conference on Robotics and Automation*. 15-19 May, p. 3055-3061.

- Moursund R.A., Carlson T.J., Peters R. D. (2003). A fisheries application of a dual-frequency identification sonar acoustic camera. *ICES Journal of Marine Science*, vol. 60, n° 3, p. 678-683.
- Negahdaripour S. (2005). Calibration of DIDSON Forward-Scan Acoustic Video Camera. *Proc. MTS/IEEE Oceans 2005*. 17-23 September, vol. 2, p. 1287-1294.
- Nillius P., Eklundh J.-O. (2002). *Fast Block Matching with Normalized Cross-Correlation using Walsh Transforms*. Report ISRN, September.
- Pentand A.P. (1990). Linear shape from shading. *Intl. J. Computer Vision*, vol 4, n° 2, p. 153-162.
- Ramos F. (2003). *Modélisation et Validation d'un Système d'Information Géographique 3D opérationnel*. Thèse de doctorat, Université de Marne-La-Vallée.
- Rabaud C. (2005). *Une nouvelle approche de mise en correspondance stéréoscopique dense par méthodes possibilistes*. Thèse de doctorat, Université de Montpellier.
- Rodriguez S. A., Frémont V., Bonnifait P. (2009). An Experiment of a 3D Real-Time Robust Visual Odometry for Intelligent Vehicles. *Proc. IEEE Conference on Intelligent Transportation Systems (ITSC 09)*, 4-7 October, p. 226-231.
- Rosten E., Porter R., Drummond T. (2010). Faster and better: A machine learning approach to corner detection. *Pattern Analysis and Machine Intelligence*, vol. 32, n° 1, p. 105-119.
- Roussel D. (1999). *Reconstruction de Courbes et de Surfaces 3d en Stéréo-Acquisition*. Thèse de doctorat de l'Université Paris XI.
- Sekkati H., Negahdaripour S. (2007). 3-D Motion Estimation for Positioning from 2-D Acoustic Video Imagery. *Pattern Recognition and Image Analysis 2007*, vol. 4478/2007, p. 80-88.
- Sintes C., Gueriot D. (2005). *Etude Shape From Shading*. Rapport.
- Sun N., Shim T., Cao M. (2008). 3D reconstruction of seafloor from sonar images based on the multi-sensor method. *IEEE Inter. conf. Multisensor Fusion and Integration for Intelligent Systems*, 20-22 August, p. 573-577.
- Trucco A., Curletto S. (2003). Extraction of 3-D Information from Sonar Image Sequences. *IEEE Transactions on Systems, Man and Cybernetics*, vol. 33, n° 4, p. 687-699.
- Tunnicliffe V., Barnes C. R., Dewey R. (2008). Major advances in cabled ocean observatories (VENUS and NEPTUNE Canada) in coastal and deep sea settings. *Proc. US/EU-Baltic International Symposium 2008*, 27-29 May.
- Victor J. S., Sentieiro J. (1994). The Role of Vision for Underwater Vehicles. *IEEE International Symposium on Autonomous Underwater Vehicle Technology*. 19-20 July, p. 28 -35.
- Walther G.R., Post E., Convey P., Menzel A., Parmesan C., Beebee T.J., Fromentin J.M., Hoegh-Guldberg O., Bairlein F. (2002). Ecological responses to recent climate change. *Nature*, vol. 416, p. 389-95.

- Wright D. J., Goodchild M. F. (1997). Data from the deep: implications for the GIS community. *International Journal of Geographical Information Science*, vol. 11, n° 6, 1st September, p. 523-528.
- Zerr A., Stage B. (1996). Three-dimensional reconstruction of underwater objects from a sequence of sonar images (1996). *Pro. International conference on Image processing*, 25-27 September, vol. 3, p. 927-930.
- Zhen C., Ming L., Shao-feng J., Shui-gen W., Yu W. (2009). A New Non-Linear Algorithm For 3D Reconstruction From Straight-line Optical Flow. *International Asia Conference on Informatics in Control, Automation and Robotics*, 1-2 February. p. 311-316.
- Zheng Q., Chellappa R. (1991, July). Estimation of Illuminant Direction, Albedo, and Shape From Shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, p. 680-702.

Naouraz Brahim est ingénieur en Halieutique de l'Institut National Agronomique de Tunisie. Suite à l'obtention de son M.Sc. en Sciences géomatiques de l'Université Laval (Canada), elle poursuit actuellement ses études en doctorat en cotutelle entre l'Université Laval (Département des Sciences Géomatiques) et Télécom Bretagne (Département Image et Traitement de l'Information, France).

Sylvie Daniel est professeure agrégée au Département des sciences géomatiques de l'Université Laval (Canada). Elle compte plus de 16 années d'expérience dans le domaine de l'imagerie géospatiale, durant lesquelles elle a dirigé et géré des projets de R&D d'envergure internationale et obtenu des octrois de recherche auprès d'organismes de financement compétitifs. Elle est membre de l'IEEE, membre du Centre de Recherche en Géomatique et chercheuse régulière du Réseau de Centres d'Excellence GEOIDE. Ses intérêts de recherche incluent l'acquisition de données, le traitement d'image et l'intelligence artificielle, la fusion de données et la réalité augmentée.

Didié Guériot (Senior Member IEEE) est Maître de Conférences à Télécom Bretagne (École nationale supérieure des télécommunications), Brest (France) depuis 2005 après 13 années de R&D sur l'imagerie sonar haute résolution au sein du Groupe d'études sous-marines de l'Atlantique (GESMA) à Brest (France) et le traitement de données issues de sondeurs multi-faisceaux chez RESON à Santa Barbara (Californie). Docteur de l'Université de Haute Alsace en Electronique, ses centres d'intérêt vont du traitement automatique des données sonar à l'intelligence artificielle, appliqués à la fusion multicapteur en vue de l'interprétation et la reconstruction de scènes sous-marines.

Basel Solaiman a obtenu son diplôme d'ingénieur en télécommunications de l'École nationale supérieure des télécommunications de Bretagne (TELECOM Bretagne) en 1983, Brest, France. Il a ensuite obtenu son Doctorat et HDR de l'Université de Rennes I respectivement en 1989 et 1997. De 1984 à 1985, il a occupé le poste d'assistant de recherche au sein du Groupe de Communication du Centre d'Études et de Recherche Scientifique, Damas, Syrie. En 1992, il a rejoint le Département Image et Traitement de l'Information (ITI) à Télécom Bretagne, Brest, France. Actuellement, il est professeur à Telecom Bretagne/Institut Telecom, chef du département ITI et professeur visiteur dans plusieurs universités à l'échelle mondiale. Ses intérêts de recherche portent sur l'information et fusion de données, reconnaissance des formes, image et interprétation de scènes et traitement du signal.

