



HAL
open science

Reconstruction tridimensionnelle de scènes sous-marines à partir de séquences d'images acquises par des caméras acoustiques

Naouraz Brahim

► **To cite this version:**

Naouraz Brahim. Reconstruction tridimensionnelle de scènes sous-marines à partir de séquences d'images acquises par des caméras acoustiques. Traitement des images [eess.IV]. Télécom Bretagne; Université de Rennes 1, 2014. Français. NNT: . tel-01217501

HAL Id: tel-01217501

<https://hal.science/tel-01217501>

Submitted on 19 Oct 2015

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

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

Sous le sceau de l'Université européenne de Bretagne

Télécom Bretagne

En habilitation conjointe avec l'Université de Rennes 1

Co-tutelle avec l'Université Laval

Ecole Doctorale – MATISSE

Reconstruction tridimensionnelle de scènes sous-marines à partir de séquences d'images acquises par des caméras acoustiques

Thèse de Doctorat

Mention : Traitement du Signal et Télécommunication

Présentée par **Naouraz Brahim**

Département : Image et Traitement de l'Information

Laboratoire : Lab-STICC, CNRS

Directeur de thèse à Télécom Bretagne : Basel Solaiman

Directeur de thèse à l'Université Laval : Sylvie Daniel

Soutenue le 30 Janvier 2014

Jury :

M. Riadh Abdelfattah, Maître de conférences HDR, École Supérieure des Communications de Tunis (Rapporteur)

M. Richard Lepage, Professeur titulaire, École de Technologie Supérieure (Rapporteur)

M. Basel Solaiman, Professeur, Télécom Bretagne (Co-directeur de thèse)

Mme Sylvie Daniel, Professeur titulaire, Université Laval (Co-directrice de thèse)

M. Alfonso Condal, Professeur retraité, Université Laval (Examineur)

M. Didier Guériot, Maître de conférences, Télécom Bretagne (Examineur)

Résumé

Depuis que les études des impacts des changements climatiques ont montré que le milieu marin pourrait être énormément fragilisé par la disparition de certaines espèces de sa faune et de sa flore, ainsi que par le vieillissement rapide de son infrastructure sous-marine, la recherche de systèmes d'observation robustes et continus est classée parmi les sujets de recherche les plus prioritaires des scientifiques. Généralement, l'observation de l'environnement et l'inspection des infrastructures sous-marines se font au moyen des capteurs imageurs tels que les capteurs optiques ou les systèmes acoustiques. Toutefois, ces outils souffrent de certaines limitations lors de leur utilisation. Les caméras optiques fournissent des données caractérisées par une bonne résolution permettant une interprétation facile des scènes observées mais aussi par des problèmes techniques lors de l'acquisition liés aux conditions du milieu marin (e.g. manque de visibilité) empêchant une observation continue du milieu. Les sonars traditionnels produisent aussi des images mais ils n'offrent pas de séquences d'images de haute cadence tels que les capteurs optiques, et leur utilisation est parfois contrainte dans les milieux portuaires et de faible profondeur. C'est pour pallier ces problèmes que les caméras acoustiques ont été conçues. Elles ont la capacité d'acquérir des séquences d'images multi-vues avec une haute cadence et de fonctionner dans des milieux très turbides. Néanmoins, ces caméras ne produisent que des images en 2D où l'élévation de la scène observée est inconnue. Or, une représentation 2D de l'environnement ne peut présenter qu'une partie des informations, elle n'est pas en mesure de représenter "fidèlement" le milieu où le phénomène est observé. Ceci n'est possible qu'à travers une représentation 3D.

L'objectif de cette thèse est donc de développer une approche de reconstruction 3D de scènes sous-marines à partir de séquences d'images acquises par des caméras acoustiques. Pour ce faire, nous nous sommes inspirés du principe de la stéréovision pour une reconstruction 3D à partir de points saillants. Néanmoins, la géométrie et la nature bruitée des images acoustiques ne permettent pas une application directe du principe de la stéréovision. Ainsi nous proposons dans cette thèse, une méthodologie de reconstruction 3D qui répond aux problématiques posées par les images des caméras acoustiques. Elle se base, en première partie, sur la conception d'un processus d'extraction de points saillants **pertinents** sur lesquels, en deuxième partie, va pouvoir s'appuyer la reconstruction 3D de la scène observée. Pour la reconstruction 3D, nous proposons deux approches différentes : une approche curviligne et une approche volumique. Dans ces deux approches, l'algorithme d'optimisation SE-AMC issu de la famille des stratégies d'évolution intervient dans le calcul du mouvement de la caméra entre les images, la détermination de ce mouvement permettant par la suite, l'estimation des informations 3D. La performance de l'approche d'extraction de primitives ainsi que celle des approches de reconstruction 3D ont été évaluées: la première au travers de critères de bonne détection, de répétabilité et de bonne localisation et la deuxième au travers de la comparaison du mouvement et des informations 3D estimés avec des données réelles.

Abstract

According to recent studies, climate change is having a significant impact on our marine environment inducing temperature increases, chemistry changes, ocean circulation influencing both population dynamics and underwater structure stability. Environmental change is thus a growing scientific concern requiring a regular monitoring of the evolution of underwater ecosystems with appropriate studies combined with accurate and relevant detailed information extraction and preservation. Tracking and modeling such changes in a marine environment is one of the current challenges for underwater exploration. The most common technique used to observe underwater environment, relies on vision-based systems either acoustical or optical. Optical cameras are widely used for acquiring images of the seafloor/underwater structures as they can provide information about the physical properties of the image that will enable the description of the observed scene (color, reflection, geometry). However, the range limitation and non-ideal underwater conditions (dark and turbid waters) make acoustic imaging the most reliable means of sight inside the underwater environment. Traditional sonar systems cannot provide an acoustic image sequences like optical cameras.

To overcome those drawbacks, acoustic camera was built. They can produce real time high resolution underwater image sequences, with high refresh rate. Moreover, compared to optical devices, they can acquire acoustic images in turbid, deep and dark water making acoustic camera imaging a reliable means for observing underwater environment. However, although acoustic cameras can provide 2-D resolution of the order of centimeters, they do not resolve the altitude of observed scene. Thus they offer a 2D environment representation which provides incomplete information about the underwater environment. Hence, it would be very interesting to have a system which can provide height information as well as a high resolution. This is the purpose of this thesis where we developed a methodology that enables 3D reconstruction of underwater scenes using sequences of acoustic images.

The proposed methodology is inspired from stereovision techniques that allow 3D information computation from image sequences. It consists of two main steps. In the first step, we propose an approach that enables the extraction of **relevant** salient points from several images. In the second step, two different methods have been proposed (curvilinear approach and volumetric approach) in order to reconstruct the observed scene using images acquired from different viewpoints. The Covariance Matrix Adaptation Evolution Strategy algorithm (SE-AMC) has been used to compute camera movement between images. This movement has been then used to retrieve 3D information. The methodology performances have been evaluated: feature extraction approach has been assessed using criteria of good detection, repeatability and good localization and 3D reconstruction approach has been assessed by comparison between estimated camera movement and 3D information with real data.

Avant-propos

Merci à toutes les personnes qui ont contribué au bon déroulement et à l'aboutissement de cette thèse. Exprimer toute ma gratitude envers ces personnes avec les mots justes me semble aujourd'hui très difficile.

Avant tout, ma thèse a été effectuée entre deux équipes de recherches. La première équipe est celle du Centre de Recherche en Géomatique au Département des Sciences Géomatiques de l'Université Laval (Québec, Canada). C'est au sein de cette équipe que j'ai commencé mes premiers travaux de recherche et où j'ai rencontré des gens très sympathiques et très serviables. La deuxième équipe est celle du Département Image et Traitement de l'Information de Télécom Bretagne (Brest, France). Une équipe qui m'a accueillie au-delà des espérances de tout nouvel arrivant !

Je remercie mes directeurs de thèse, Madame Sylvie DANIEL, Monsieur Basel SOLAIMAN et mon encadrant Monsieur Didier GUÉRIOT de m'avoir encadrée tout au long de cette thèse. Je me considère chanceuse de travailler avec une telle équipe : tous dynamiques, disponibles et jamais à court d'idées. Vous m'avez toujours offert de bonnes conditions pour travailler et j'ai beaucoup appris de vous !

Merci beaucoup Sylvie de m'avoir encadrée depuis mes études de maîtrise. Merci pour ta confiance, ta gentillesse et pour la chance que tu m'as donnée de faire ma thèse en cotutelle avec Télécom Bretagne.

Un grand merci Didier pour ta grande disponibilité et ton suivi sérieux et exigeant, qui m'ont été d'une aide précieuse.

Merci à vous Basel d'avoir accepté de diriger ma thèse à Télécom Bretagne et pour votre générosité et humaine et scientifique.

Je tiens à remercier sincèrement Monsieur Yvan SIMARD, professeur à l'Université du Québec à Rimouski et chercheur chez Pêches et Océans Canada, qui m'a prêté la caméra DIDSON pour faire des acquisitions de données à la piscine profonde de l'Université Laval. Par ailleurs, je remercie Monsieur Régis DUMONT, coordonnateur d'opérations récréatives de l'Université Laval, qui m'a donné la possibilité de réserver la piscine pour faire mes acquisitions de données. Merci à Guy MONTREUIL et Sylvain GAUTIER qui m'ont aidée et accompagnée lors de ces acquisitions.

Je tiens, aussi, à remercier Monsieur Christian GAGNÉ pour sa gentillesse et sa disponibilité pour répondre à mes questions.

Merci à Monsieur Riadh Abdelfattah, Monsieur Alfonso Condal et Monsieur Richard Lepage d'avoir accepté d'évaluer et de juger ma thèse. Je leur remercie, également, pour l'intérêt qu'ils ont porté à mon travail et pour leurs remarques constructives.

Je remercie le Conseil Franco Québécois de Coopération Universitaire (CFQCU), la Mission Universitaire de Tunisie en Amérique du Nord (MUTAN) et l'EGIDE pour leur participation au financement de ma thèse.

Un grand merci à tous mes amis à l'Université Laval et à Télécom Bretagne. Je leur exprime ma profonde gratitude et leur souhaite beaucoup de réussite.

Enfin, j'ai une pensée toute particulière pour mes parents, mes frères, mon cher époux et mon fils. Je suis heureuse de leur dédier ce travail.

*À mon père,
À mon époux, à mon fils et à toute ma famille.*

« À vaincre sans péril, on triomphe sans gloire. »

Corneille (Le Cid)

Table des matières

Résumé.....	i
Abstract.....	iii
Avant-propos.....	v
Table des matières.....	ix
Table des figures.....	xiii
Liste des tableaux.....	xvii
Chapitre 1. Introduction.....	1
1.1 Mise en contexte.....	1
1.1.1 Impact des changements climatiques sur le milieu marin.....	1
1.1.2 Solutions préventives aux changements climatiques.....	2
1.2 Besoins relatifs à la prévention des impacts des changements climatiques.....	5
1.2.1 Besoin d'un nouvel outil d'exploration et de recherche sous-marine.....	5
1.2.2 Besoin d'une représentation tridimensionnelle de l'environnement.....	6
1.2.3 Besoin d'une information précise et conforme à la réalité pour une représentation tridimensionnelle robuste.....	7
1.3 Problématique : exploitation de caméras acoustiques pour la modélisation tridimensionnelle de l'environnement.....	7
1.3.1 Comment peut-on effectuer une reconstruction 3D d'une scène sous-marine à partir d'images issues de caméras acoustiques ?.....	8
1.3.2 Comment peut-on exploiter les informations issues des caméras acoustiques pour obtenir une reconstruction 3D qui soit complète et robuste ?.....	10
1.4 Hypothèses et objectifs de la recherche.....	11
1.4.1 Hypothèses.....	11
1.4.2 Définition de l'objectif général et des objectifs spécifiques.....	12
1.5 Méthodologie.....	13
1.5.1 L'approche méthodologique.....	13
1.5.2 Déroulement de la thèse.....	13
1.5.3 Implémentation et données utilisées.....	15
1.6 Structure de la thèse.....	17
Chapitre 2. État de l'art.....	19
2.1 Introduction.....	19
2.2 Les caméras acoustiques.....	19
2.2.1 Présentation générale et domaines d'application.....	19

2.2.2	Modèle géométrique de la caméra DIDSON.....	20
2.3	La reconstruction 3D multi-vues en vision numérique	21
2.3.1	Méthodologie générale de la reconstruction 3D en vision numérique	22
2.3.2	Extraction, mise en correspondance et recalage de primitives	22
2.3.3	Reconstruction 3D et modélisation 3D	25
2.4	La reconstruction 3D d'une scène sous-marine à partir d'images acoustiques.....	29
2.4.1	Reconstruction 3D multi-capteurs	30
2.4.2	Reconstruction 3D mono-capteurs	30
2.5	Le cadre conceptuel proposé.....	31
2.5.1	Architecture du cadre conceptuel.....	31
2.5.2	Description générale de l'architecture proposée	32
2.6	Synopsis du chapitre 2	35
Chapitre 3.	Extraction de primitives à partir d'images acquises avec des caméras acoustiques	37
3.1	Introduction	37
3.2	Sélection des primitives adaptées à la reconstruction 3D de scène à partir de séquences d'images acoustiques	37
3.2.1	Description des principaux types de primitive et de leurs approches d'extraction	37
3.2.2	Choix de primitives pour la reconstruction 3D.....	45
3.3	Approche proposée	49
3.3.1	Extraction de contours pertinents par analyse multi-échelle	51
3.3.2	Extraction de points saillants.....	56
3.4	Expérimentations et résultats.....	58
3.4.1	Description des jeux de données	58
3.4.2	Description du protocole de test.....	59
3.4.3	Résultats des tests	60
3.4.4	Évaluation quantitative de l'approche	72
3.4.5	Discussion des résultats	79
3.5	La mise en correspondance des points saillants pour une reconstruction 3D.....	81
3.6	Synopsis du chapitre 3	84
Chapitre 4.	Reconstruction 3D de scènes sous-marines à partir de séquences d'images acoustiques.....	87
4.1	Optimisation & algorithmes évolutionnaires	87
4.1.1	Pourquoi se tourner vers les algorithmes évolutionnaires ?.....	87
4.1.2	Que sont les algorithmes évolutionnaires?	89

4.1.3	Les stratégies d'évolution.....	92
4.2	Reconstruction 3D selon une approche curviligne.....	97
4.2.1	Reconstruction à partir de deux images.....	98
4.2.2	Reconstruction à partir de 'n' images.....	102
4.3	Reconstruction 3D selon une approche volumique.....	104
4.3.1	Reconstruction à partir de deux images.....	104
4.3.2	Reconstruction à partir de 'n' images.....	108
4.3.3	Principales différences entre l'approche curviligne et l'approche volumique.....	110
4.4	Contexte d'expérimentation.....	110
4.4.1	Jeux de données exploités.....	110
4.4.2	Mise en œuvre de l'algorithme SE-AMC.....	115
4.5	Application de l'approche curviligne.....	115
4.5.1	Tests de reconstruction 3D multi-vues.....	115
4.5.2	Évaluation de la performance de l'approche curviligne.....	135
4.6	Application de l'approche volumique.....	136
4.6.1	Reconstruction 3D à partir de deux images.....	137
4.6.2	Reconstruction 3D à partir de trois images.....	143
4.6.3	Reconstruction 3D à partir de quatre images.....	151
4.6.4	Impact de la résolution des images sur la précision de la reconstruction 3 D selon l'approche volumique.....	161
4.7	Synopsis du chapitre 4.....	166
Chapitre 5.	Conclusion générale.....	169
5.1	Synthèse de la recherche.....	169
5.1.1	Un cadre conceptuel de reconstruction 3D de scènes.....	169
5.1.2	Une approche d'extraction d'informations pertinentes caractérisant la géométrie de la scène observée.....	170
5.1.3	Une approche d'estimation de la géométrie 3D de la scène.....	172
5.2	Principales observations tirées des approches développées.....	176
5.3	Les contributions de notre travail.....	178
5.4	Perspectives.....	179
5.4.1	Réflexions sur le processus de reconstruction.....	179
5.4.2	Optimisation de la méthodologie de reconstruction.....	180
5.4.3	De l'exportation des scènes reconstruites.....	181
5.4.4	Reconstruction 3D multi-sources.....	181

Bibliographie.....183
Annexe A..... 191
Annexe B..... 193

Table des figures

Figure 1 : Projection d'un point (P) de l'environnement sous-marin sur le plan d'une image DIDSON (placé arbitrairement en élévation, ici l'élévation φ vaut 0). (GESMA, 2008).....	9
Figure 2 : Faisceau d'insonification de la caméra DIDSON.....	10
Figure 4 : Étapes de reconstruction 3D à partir d'un ensemble d'images.....	22
Figure 5 : La géométrie épipolaire. C_1 et C_2 sont les deux caméras, les droites l_{P_1} et l_{P_2} sont des droites épipolaires, P est un point dans l'espace 3D et P_1 et P_2 sont la projection du point P sur deux images, R_1 et R_2 sont les repères respectifs associés aux caméras C_1 et C_2 , d_1 est la droite passant par le centre de C_1 et P_1	24
Figure 6. Architecture du cadre conceptuel.....	32
Figure 7 : Illustration de primitives appariées (P : Points, Profils, S : surfaces).....	35
Figure 8 : Frise de détecteurs de points d'intérêt (Parisot, 2009).....	44
Figure 9 : Exemple de détection de point d'intérêt : (a) Image originale DIDSON d'objets posés au fond d'un bassin; (b) Points d'intérêt détectés par le détecteur de Harris, (c) Points d'intérêt détectés par le détecteur SUZAN; (d) Points d'intérêt détectés par le détecteur SIFT.....	48
Figure 10 : Chaîne de traitements de l'approche proposée d'extraction de points anguleux à partir d'une séquence d'images acoustiques.....	51
Figure 11 : Détection de points anguleux par l'algorithme de Douglas-Peucker (<i>Douglas et Peucker, 1973</i>) a correspond au segment liant les deux extrémités du contour, b est la distance orthogonale entre le segment a et le point de contour c ou sommet le plus éloigné de a. b étant supérieure à la tolérance, ce sommet est conservé comme coin ou point saillant.....	58
Figure 12 : Exemples d'images extraites de différentes séquences d'images acoustiques provenant de la caméra DIDSON.....	59
Figure 13 : Extraction de contours des images de la séquence « Pied de pont ».....	60
Figure 14 : (a) Sélection de contours par mise en correspondance dans l'image i. (b) Sélection de contours par mise en correspondance dans l'image i+1. (c) Mesure $M_{k1 \rightarrow 2}$ des contours de l'image i dans l'image i+1. La barre colorée en vert est le contour 17 en vert dans (a) et (b). La barre colorée en orangée est le contour 34 en orangé dans (a) et (b). La barre colorée en rouge est le contour 60 en rouge dans (a) et (b). La barre colorée en jaune est le contour 26 en jaune dans (a) et (b). Le contour bleu encerclé dans (a) est le contour 46 dans l'histogramme n'ayant pas de correspondant dans l'image (b).....	62
Figure 15 : Influence du seuil de sélection sur la préservation des contours pertinents sur une image i (colonne a) et une image i+1 (colonne b) pour un niveau d'échelle égal à 8. À partir d'un seuil égal à 50%, au niveau des régions marquées par les cercles jaunes, les contours caractérisant la géométrie de la structure commencent à disparaître.....	64
Figure 16 : Mise en correspondance des contours entre une image (i) et une image (i+1). Les régions encerclés en jaune (en c et d) montrent la disparition de faux contours encerclés en rouge (a et b) dans l'image (i) et (i+1) suite à la mise en correspondance.....	65
Figure 17 : Extraction de contours pertinents par une analyse multi-échelle sur l'image i (colonne de gauche) et sur l'image i+1 (colonne de droite).....	67
Figure 18 : Évolution du nombre de contours extraits en fonction de l'échelle d'analyse. E indique l'échelle à partir de laquelle le nombre de contours devient stable.....	68
Figure 19 : Extraction des points saillants à partir des contours extraits du niveau d'échelle 1 toujours présents au niveau d'échelle 8 pour les images i et i+1. (a) et (a') représentent les contours des images i et i+1 qui vont servir à l'extraction des points saillants indiqués sur (b) et (b').....	69

Figure 20 : Extraction de points saillants de la séquence « Épave » : (a et a') contours pertinents extraits des images i et $i+1$ et leurs points saillants (b et b')	70
Figure 21 : Extraction de points saillants de la séquence acquise en bassin intérieur : (a et a') contours pertinents extraits des images i et $i+1$ et leurs points saillants (b et b')	71
Figure 22 : Répétabilité de l'ensemble des points saillants extraits sur les trois jeux de données pour une séquence d'images consécutives. En abscisse, la paire d'images i fait référence aux images $i-1$ et i du jeu de données.....	73
Figure 23 : Extraction de points saillants par l'approche développée (points blancs) et points caractéristiques de référence extraits manuellement (points rouges)	75
Figure 24 : Extraction de points saillants correspondants par corrélation normalisée : (a, a') exemple de points correspondants d'une paire d'images i et $i+1$ extraites de la séquence du « Pied de pont »; (b, b') exemple de points correspondants d'une paire d'images i et $i+1$ extraites de la séquence « Epave»; (c, c') exemple de points correspondants d'une paire d'images i et $i+1$ extraites de la séquence de la piscine.....	83
Figure 25 : Vue d'ensemble d'un algorithme évolutionnaire (Balmisse, 2002)	90
Figure 26 : Représentation de la recherche d'une solution pour une fonction objective $f(x)$ (Alkhatib, 2008). La figure illustre un parcours de recherche dans lequel un optimum local est rencontré mais l'exploration de l'espace des solutions continuant, un meilleur optimum est ensuite rencontré (en l'occurrence l'optimum global).	91
Figure 27 : Pseudo-code associé à l'approche SE-AMC dans une reconstruction 3D multi-images.....	103
Figure 28 : Un pixel de l'image correspond en réalité à un polygone \mathcal{P} (en portée et en azimut) qui induit un volume \mathbf{V} dans la scène. C_1, C_2, C_3 et C_4 sont les quatre arcs de cercle issus des quatre coins de \mathcal{P}	105
Figure 29 : Pseudo-code associé à l'approche SE-AMC dans une reconstruction volumique 3D multi-images.. ..	109
Figure 30 : Reconstruction 3D (exprimées en mètre) calculées par l'approche volumique.....	111
Figure 31 : (a) La scène enregistrée dans la piscine de l'Université Laval. (b) Les positions de la caméra par rapport à la scène lors de l'acquisition des séquences d'images : La séquence 17 (image 1) est acquise à partir de la position 1; Les séquences 21 (image 2) et 24 (image 3) sont acquises à partir de la même position 2; La séquence 8 (image 4) est acquise à partir de la position 3.	112
Figure 32 : Images acquises selon différents angles de vue dans le bassin de l'Université Laval et exploitées lors des expérimentations afin d'estimer les performances des approches de reconstruction : (a) Image 1 (séquence 17, position 1); (b) Image 2 (séquence 21, position 2); (c) Image 3 (séquence 24, position 2); (d) image 4 (séquence 8, position 3).	113
Figure 33 : Position des 20 points saillants extraits de l'image 1 (séquence 17) : 16 points associés à la grille et 4 points associés à la boîte (à droite de l'image).....	114
Figure 34 : Position des points de la boîte et de la grille en (X, Y) (a) et en (X, Y, Z) (b) dans le repère de la piscine.	114
Figure 35 : Évolution de la fonction objective f (a) et des solutions optimales $[x(1), x(2), \dots, x(6)]$ du déplacement de la caméra entre l'image 1 et l'image 2 (b) en fonction du nombre d'itérations.....	116
Figure 36 : Coordonnées 3D calculées par l'approche curviligne (exprimées en mètre).....	119
Figure 37 : Un exemple de reconstruction d'un point de la grille et d'un point de la boîte (coordonnées 3D exprimées en mètre).. ..	120
Figure 38 : Évolution de la fonction objective f (a) et des solutions optimales $[x(1), x(2), \dots, x(12)]$ du déplacement de la caméra entre les images 1 et 2 et les images 1 et 3 (b) en fonction du nombre d'itérations	123

Figure 39 : Coordonnées 3D (exprimées en mètre) calculées par l'approche curviligne..	126
Figure 40 : Un exemple de reconstruction d'un point de la grille et d'un point de la boîte (coordonnées exprimées en mètre).....	127
Figure 41 : Évolution de la fonction objective f (a) et des solutions optimales $[x(1), x(2), \dots, x(18)]$ du déplacement de la caméra entre les images 1 et 2, les images 1 et 3 et les images 1 et 4 (b) en fonction du nombre d'itérations	130
Figure 42 : Coordonnées 3D (exprimées en mètre) calculées par l'approche curviligne.	133
Figure 43 : Un exemple de reconstruction d'un point de la grille et d'un point de la boîte (coordonnées exprimées en mètre).....	134
Figure 44 : Évolution de la fonction objective f (a) et des solutions optimales $[x(1), x(2), \dots, x(6)]$ du déplacement de la caméra entre les images 1 et 2 en fonction du nombre d'itérations.	137
Figure 45 : Coordonnées 3D (exprimées en mètre) calculées par l'approche volumique.	141
Figure 46 : Exemple de reconstruction d'un point de la grille et d'un point de la boîte (coordonnées 3D exprimées en mètre) par l'approche volumique.....	142
Figure 47 : Évolution de la fonction objective f (a) et des solutions optimales $[x(1), x(2), \dots, x(12)]$ du déplacement de la caméra entre les images 1 et 2 et les images 1 et 3 en fonction du nombre d'itérations. .	144
Figure 48 : Coordonnées 3D (exprimées en mètre) calculées par l'approche volumique.).....	148
Figure 49 : Exemple de reconstruction d'un point de la grille et d'un point de la boîte par l'approche volumique (coordonnées 3D exprimées en mètre).....	149
Figure 50 : Coordonnées 3D (exprimées en mètre) d'un point de la grille et d'un point de la boîte calculées par l'approche volumique à partir de deux images et de trois images.	150
Figure 51 : Évolution de la fonction objective f (a) et des solutions optimales $[x(1), x(2), \dots, x(18)]$ du déplacement de la caméra entre les images 1 et 2, les images 1 et 3 et les images 1 et 4 en fonction du nombre d'itérations.	152
Figure 52 : Coordonnées 3D (exprimées en mètre) calculées par l'approche volumique..	157
Figure 53 : Exemple de reconstruction d'un point de la grille et d'un point de la boîte par l'approche volumique (coordonnées 3D exprimées en mètre).	158
Figure 54 : Coordonnées 3D (exprimées en mètre) d'un point de la grille et d'un point de la boîte calculées par l'approche volumique à partir de deux images (a), de trois images (b) et de quatre images (c).....	160
Figure 55 : Coordonnées 3D (exprimées en mètre) calculées par l'approche volumique à partir de deux images avec deux résolutions différentes.	163
Figure 56 : Coordonnées 3D (exprimées en mètre) calculées par l'approche volumique à partir de trois images avec deux résolutions différentes.	164
Figure 57 : Coordonnées 3D (exprimées en mètre) calculées par l'approche volumique à partir de quatre images avec deux résolutions différentes.	165

Liste des tableaux

Tableau 1. Principe des approches de modélisation polygonale 3D	27
Tableau 2 . Moyenne et écart-type du taux de répétabilité pour l'ensemble de points détectés dans les trois séquences	74
Tableau 3. Étude quantitative de la bonne détection des points extraits par l'approche développée (automatiquement);	76
Tableau 4. Évaluation de la localisation des points saillants en comparaison avec les points de référence (distance minimale, distance maximale, moyenne des distances et écart-type).	78
Tableau 5. Évaluation de la localisation des points saillants en comparaison avec les points de référence (distance minimale, distance maximale, moyenne des distances et écart-type).	79
Tableau 6. Coordonnées des positions de la caméra dans le repère de la piscine	113
Tableau 7. Paramètres de la rotation et de la translation de la caméra entre les images 1 et 2.	117
Tableau 8. Les erreurs minimales, maximales et moyennes des coordonnées 3D calculées pour les points de la grille par rapport à leurs coordonnées 3D réelles.	120
Tableau 9. Les erreurs minimales, maximales et moyennes des coordonnées 3D calculées pour les points de la boîte par rapport à leurs coordonnées 3D réelles.	120
Tableau 10. Mouvement calculé de la caméra entre les images, respectivement, (1 et 2) et (1 et 3).	124
Tableau 11. Les erreurs minimales, maximales et moyennes des coordonnées 3D calculées des points de la grille par rapport à leurs coordonnées 3D réelles.	127
Tableau 12. Les erreurs minimales, maximales et moyennes des coordonnées 3D calculées des points de la boîte par rapport à leurs coordonnées 3D réelles.	127
Tableau 13. Mouvement calculé de la caméra entre les images, respectivement, (1 et 2), (1 et 3) et (1 et 4). α , β , γ sont les angles de la rotation selon les axes des X, Y et Z.	131
Tableau 14. L'erreur en translation de la caméra entre les images (1 et 2), (1 et 3) et (1 et 4) selon les trois axes (X, Y et Z).	131
Tableau 15. L'erreur en rotation de la caméra entre les images (1 et 2), (1 et 3) et (1 et 4) selon les trois axes (X, Y et Z).	132
Tableau 16. Les erreurs minimales, maximales et moyennes des coordonnées 3D calculées pour les points de la grille par rapport à leurs coordonnées réelles.	134
Tableau 17. Les erreurs minimales, maximales et moyennes des coordonnées 3D calculées pour les points de la boîte par rapport à leurs coordonnées réelles.	134
Tableau 18. Mouvement de la caméra entre les images 1 et 2 par l'approche volumique (a) et par approche curviligne (a'), calculé dans le repère R1 de l'image 1.	138
Tableau 19. Les intervalles des coordonnées 3D des points de la grille.	139
Tableau 20. Les intervalles des coordonnées 3D des points de la boîte	139
Tableau 21. Mouvement de la caméra entre les images 1 et 2 et les images 1 et 3 par approche volumique (a et b) et par approche curviligne (a' et b'), calculé dans le repère R1 de l'image 1.	144
Tableau 22. Les intervalles des coordonnées 3D des points de la grille.	145
Tableau 23. Les intervalles des coordonnées 3D des points de la boîte	146
Tableau 24. Différence entre les dimensions des intervalles de hauteurs des points de la grille reconstruits à partir de deux images (intervalles 1) et trois images (intervalles 2)	146
Tableau 25. Différence entre les dimensions des intervalles de hauteur des points de la boîte reconstruits à partir de deux images (intervalles 1) et trois images (intervalles 2)	147

Tableau 26. Mouvement de la caméra entre les images 1 et 2, les images 1 et 3 et les images 1 et 4 par approche volumique (a , b et c) et par approche curviligne (a' , b' et c'), calculé dans le repère $R1$ de l'image 1..	152
Tableau 27. Les intervalles des coordonnées 3D des points de la grille	154
Tableau 28. Les intervalles des coordonnées 3D des points de la boîte	154
Tableau 29. Différence entre les dimensions des intervalles de hauteur des points de la grille reconstruits à partir de trois images (intervalles 2) et quatre images (intervalles 3).	158
Tableau 30. Différence entre les dimensions des intervalles de hauteur des points de la grille reconstruits à partir de trois images (intervalles 2) et quatre images (intervalles 3).	159

Chapitre 1. Introduction

1.1 Mise en contexte

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 ces 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 observés actuellement et anticipés dans le futur montrent l'apparition de premiers signes de vulnérabilité dans le milieu marin tels que l'érosion des fonds, la disparition d'espèces, le vieillissement des structures marines (ex. les quais et les digues) etc. La recherche de solutions et de moyens permettant une exploitation durable des ressources et la maintenance des infrastructures maritimes telles que les installations portuaires est devenue ainsi une nécessité primordiale pour assurer la viabilité et la préservation du milieu marin. C'est dans ce contexte qu'est mené actuellement un nombre important de travaux de recherche scientifique dans le domaine maritime (ex. océanographie, acoustique sous-marine, biologie marine, ...). Au travers d'observations et de suivis spatio-temporels du milieu marin et de ses éléments constitutifs, ces travaux contribuent énormément à la conception et au développement d'outils efficaces et innovants rendant le milieu marin plus accessible et permettant la détection et la prévention rapide de différents types de changements spatio-temporels qui peuvent s'y produire.

1.1.1 Impact des changements climatiques sur le milieu marin

Le Groupe d'experts Intergouvernemental sur l'Évolution du Climat (GIEC) a évalué les impacts actuels et à venir du changement climatique (GIEC, 2011¹). Il est principalement envisagé que la variation du niveau de la mer et le changement de la température entraîneront un accroissement de l'étendue et de la gravité des tempêtes avec des conséquences de plus en plus graves sur l'écosystème marin. Sur le plan physiographique, une élévation rapide du niveau de la mer aurait principalement pour effet d'accélérer les changements littoraux qui se produisent actuellement dans les zones côtières. L'érosion et le recul des plages, l'érosion des falaises et la migration vers les terres des îles-barrières se poursuivraient, mais de façon plus rapide et généralisée (Walker et Barrie, 2004). L'augmentation progressive de la température de l'eau de mer entraînerait un changement radical au niveau du fonctionnement de l'écosystème: changement de la direction des courants d'eau profonds et surfaciques, migration des espèces marines à la recherche de milieux adéquats, disparition d'espèces et apparition d'autres etc. (Com. Europ, 2007). L'accroissement des ondes de

¹ <http://www.ledevoir.com/environnement/actualites-sur-l-environnement/336525/rapport-du-groupe-d-experts-intergouvernemental-sur-l-evolution-du-climat-meteo-extremes-climatiques-a-prevoir>

tempêtes provoquant des inondations et des perturbations au niveau de l'hydrodynamique font également partie des préoccupations actuelles. 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 les pêches, le transport, le tourisme et les loisirs, de même que sur les collectivités (GIEC, 2011, Com. presse, 2011²). En effet, ces impacts toucheront particulièrement, d'une part, la faune et la flore de l'écosystème et, d'autre part, les infrastructures côtières telles que les ports, les ponts, les plates-formes pétrolières etc. Dans les pays ayant de grands lacs tels que le Canada, les impacts des changements climatiques sur le milieu marin toucheront aussi les écosystèmes lacustres puisqu'ils sont interconnectés via des rivières et des cours d'eau. Par conséquent, de lourdes conséquences économiques en découleront, puisque ces milieux se caractérisent par une activité économique importante. En effet, la variation du niveau d'eau dans les lacs (qu'elle soit une baisse ou une augmentation) risque d'affecter différents types d'infrastructures qui exploitent ou reposent sur cette ressource. À titre d'exemple, on peut citer les ouvrages de prise d'eau municipaux et industriels qui ont été conçus en fonction de l'échelle traditionnelle de variation du niveau des lacs.

L'étude des impacts des changements climatiques actuels et futurs cités ci-dessus a fait naître un grand intérêt pour la recherche de solutions visant la protection et la restauration des ressources du milieu marin et ceci dans la stratégie du développement durable actuel. Ces solutions sont plutôt des solutions préventives vis-à-vis de l'évolution des composantes de l'écosystème marin associées à des mesures d'adaptation aux changements climatiques actuels.

1.1.2 Solutions préventives aux changements climatiques

Les solutions préventives permettent d'évaluer la vulnérabilité d'une région, d'une zone, d'une composante du milieu, etc. 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).

Après un recensement de la littérature, on constate que l'observation et la modélisation des fonds et de la colonne d'eau océanique ont reposé pendant de longues années sur une instrumentation s'appuyant sur des navires de surface et fonctionnant pendant des durées relativement courtes (de quelques semaines à quelques mois) (Delaney *et al*, 2003). Mais ce type d'observation ne suffit plus. En effet, la modélisation de certains phénomènes nécessite l'acquisition de données précises sur de longues périodes (plusieurs années).

² <http://www.ledevoir.com/motcle/groupe-d-experts-intergouvernemental-sur-l-evolution-du-climat-giec/>

De plus, la détection précoce et fiable de tout changement spatio-temporel du milieu implique la mise en place d'équipements de surveillance en continu (Rapport SIVOM, 2002).

Dans le but de répondre à ce besoin, des observatoires ont été construits et repartis dans les mers et les océans (ex. NEPTUNE³, ESONET⁴). Les données enregistrées dans le contexte de ces observatoires sont obtenues, entre autres, à partir d'observations robotisées, 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 (appareils photographiques, caméras vidéo haute résolution) (Tunncliffe *et al*, 2008). Ces instruments de mesure sont adaptés aux conditions du milieu telles que la résistance aux courants forts, à la pression, etc. Par contre, l'utilisation de certains outils comme les caméras optiques est contraignante dans les grandes profondeurs, où la visibilité est nulle, puisqu'elle nécessite alors l'ajout d'une source lumineuse. La recherche d'équipements innovants et adéquats est un élément-clé pour répondre aux performances demandées à ces observatoires.

Le traitement de ces données aboutit à la production de représentations cartographiques (ex. cartographie des habitats marins, des fonds, etc.) et à la production de modèles d'évolution prédictifs qui se basent sur les données archivées et actuelles pour prévoir l'évolution spatio-temporelle du milieu marin et en conséquence déterminer les tendances des impacts des changements climatiques. À titre d'exemple, on cite le projet mis en place par le Syndicat Intercommunal à Vocation Multiple (SIVOM⁵) du littoral des Maures (Sud de la France) depuis Janvier 2010, qui en s'appuyant sur des relevés de paramètres caractéristiques et des photographies au sein de l'observatoire marin, cherche à évaluer l'évolution de l'herbier de posidonie en développant un modèle de cette évolution. Ce modèle est exploité comme source d'information par un système de surveillance du milieu marin qui à travers la dynamique évolutive des herbiers permet de déterminer la qualité du milieu, d'identifier les sources de nuisances ayant un impact au niveau des fonds et de suivre, secteur par secteur, la dynamique du milieu. Ce système de surveillance utilise l'herbier de posidonie comme bio-indicateur de la qualité globale du milieu marin (Rapport SIVOM, 2002). Des systèmes d'information couplés à des systèmes de gestion de base de données (SGBD) tels que les Systèmes d'Information Géographique (SIG) sont généralement utilisés pour l'analyse et la visualisation des données spatiales, permettant ainsi de faciliter les prises de décision quant aux risques environnementaux. Parmi ces systèmes développés, on cite

³ <http://www.neptunecanada.ca>

⁴ <http://www.esonet-emso.org>

⁵ <http://www.observatoire-marin.com/>

le système mondial d'information sur les pêches (Fisheries Global Information System FIGIS)⁶ et le système d'information biogéographique sur l'océan (Ocean Biogeographic Information System OBIS)⁷.

Les impacts des changements climatiques ne sont pas uniquement évalués à partir de l'observation continue de l'environnement marin. Ces impacts peuvent également être estimés à partir de l'observation de divers types d'infrastructures maritimes. En effet, ils se traduisent par l'apparition de signes d'usure, de détérioration et par la présence d'affouillements sous les caissons composant les quais, au niveau des pieds des ponts et des bases des plates-formes pétrolières. Ces effets compromettent la stabilité de ces infrastructures et affectent le bon déroulement des activités maritimes correspondantes. Plusieurs études et efforts ont été investis afin de prévoir d'une façon fiable la durée de vie de ces infrastructures mais il s'avère qu'il est extrêmement difficile de prévoir exactement la date d'apparition des premiers signes de vieillissement (Ferreira *et al*, 2004; Ferreira et Jalali, 2004; Khatri et Sirivivatnanon, 2004) ou de défectuosité. L'explosion de la plate forme pétrolière Deepwater Horizon dans le golfe de Mexique le 20 Avril 2010 en est une très bonne illustration. Suite à cette explosion, l'équivalent de 5000 barils de pétrole (800 000 litres) a été déversé chaque jour dans le golfe jusqu'au 19 Septembre 2010. Cette catastrophe aurait probablement pu être évitée si des outils d'inspection continue avaient été mis en place, l'origine de l'explosion étant due à une fuite dans une conduite pétrolière (Com. Presse, 2011)⁸. C'est pour cette raison que l'inspection et le contrôle continu des infrastructures maritimes sont devenus des préoccupations prioritaires impliquant une gestion régulière des contrôles et des rétroactions adaptées.

Ainsi dans le cadre d'une stratégie de développement durable, détecter, identifier et cartographier de manière automatisée les signes d'usure, les défectuosités et donc les zones à risque des infrastructures maritimes, notamment portuaires, devraient permettre d'alimenter régulièrement des systèmes d'aide à la décision qui décriront l'état courant de ces infrastructures, et permettront de déterminer les mesures éventuelles de maintenance et d'adaptation à mettre en œuvre. Ce type de systèmes automatiques s'intègre parmi les approches de prévention des risques et constitue une approche novatrice. En effet, jusqu'à présent les approches développées qui se basent, notamment sur l'utilisation de capteurs imageurs ne sont pas en mesure de prendre en compte simultanément et d'une façon fiable tous les éléments nécessaires pour une inspection et un contrôle complet (la détection des signes d'usures, l'identification, la localisation, etc.) et continu (Guo *et al*, 2003; Reed *et al*, 2000) des infrastructures.

⁶ <http://www.fao.org/fishery/figis/2/en>

⁷ <http://www.iobis.org/>

⁸ http://www.lemonde.fr/planete/article/2011/09/14/deepwater-horizon-les-autorites-federales-accusent-bp-et-des-sous-traitants_1572408_3244.html

1.2 Besoins relatifs à la prévention des impacts des changements climatiques

Comme vu au paragraphe précédent, les besoins en termes de solutions et moyens permettant la prévention des impacts des changements climatiques et la préservation des ressources marines figurent parmi les préoccupations majeures de la recherche scientifique. Ces besoins se résument en trois principaux points. Le premier concerne les outils traditionnellement utilisés pour l'observation de la faune et de la flore sous-marines et l'inspection des infrastructures maritimes (i.e. caméras optiques) dont les performances sont limitées dans le milieu marin. Il y a donc un besoin de nouveaux outils d'exploration et de recherche sous-marine. Le deuxième concerne la modélisation tridimensionnelle du milieu dans l'objectif de surveiller et de détecter des changements affectant le milieu sous-marin et les infrastructures maritimes. Il y a un besoin de méthodes adaptées pour réaliser ce suivi de manière quasi continue et répondre aux spécifications des approches de prévision des impacts des changements climatiques et de préservation des ressources sous-marines. Un troisième besoin s'ajoute aux deux précédemment cités; il concerne la qualité de l'information utilisée pour la modélisation. En effet, les solutions préventives des impacts des changements climatiques nécessitent une modélisation précise et fidèle de la réalité de la scène observée afin que les décideurs puissent prendre des décisions éclairées et adéquates quant aux actions à mettre en œuvre pour faire face aux risques environnementaux. Les paragraphes suivants vont aborder ces trois besoins en détail.

1.2.1 Besoin d'un nouvel outil d'exploration et de recherche sous-marine

Comme dans le cas des observatoires sous-marins, l'observation et le contrôle des infrastructures maritimes se font généralement au moyen de capteurs optiques (Tunncliffe *et al*, 2008). Ces capteurs sont souvent installés sur des véhicules filoguidés, autonomes ou manipulés par des plongeurs. Les caméras optiques sous-marines actuelles sont de haute technologie, facile à manipuler et offrent des images à haute résolution. Cependant, leur utilisation fait toujours face à différents problèmes technologiques limitatifs tels que : un champ de vue restreint, 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), un stockage et transfert des images lents entraînant de faibles cadences d'acquisition. 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). Il est donc indispensable de trouver d'autres outils ne souffrant pas de ces inconvénients.

D'après la littérature concernant les nouveaux outils d'exploration et de recherche sous-marine, la tendance actuelle semble s'orienter vers des systèmes qui combinent à la fois des technologies acoustiques et optiques.

Ces systèmes exploitent, d'une part, la capacité qu'ont les ondes acoustiques à se propager dans des milieux turbides comme c'est le cas avec les systèmes sonar et, d'autre part, 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). C'est dans ce contexte que les caméras acoustiques ont été conçues telles que les caméras DIDSON⁹ (Dual Frequency Identification Sonar) et Blueview¹⁰. Celles-ci présentent des caractéristiques très performantes en termes de cadence d'acquisition et de résolution d'images. Par exemple, la résolution en azimuth des images de la caméra DIDSON opérant en mode haute fréquence avec une portée de 10 mètres, est de 2 cm au début de l'enregistrement (c'est-à-dire à une distance de 4.5 mètres du sonar) et de 5 cm en fin d'image ou de portée (c'est-à-dire à une distance de 10 mètres du sonar). Selon des recherches récentes, ces caméras ont montré un grand potentiel dans la détection et l'observation de la faune marine ainsi que dans l'inspection des infrastructures portuaires. Par exemple, la caméra DIDSON a été utilisée par Burwen *et al* (2007) pour estimer la quantité de saumon dans les rivières lors de leur migration ainsi que par Jun et Asada (2007) pour détecter et suivre le mouvement du poisson Ayu dans l'archipel japonais. Dans le cadre de l'inspection des infrastructures, une reconstruction d'une mosaïque d'images DIDSON a permis à Asada (2008) de visualiser un quai de 150 mètres de large sur 11 mètres de haut dans le port d'Otaru au Japon. Cette mosaïque d'images a proposé une vue globale du quai permettant alors d'évaluer son état global et de détecter des signes d'usure éventuels. Aucune reconstruction tridimensionnelle du quai n'a néanmoins été réalisée puisque cette mosaïque résultait de la fusion, après recalage, des différentes images collectées. D'après ces travaux, les caméras acoustiques constituent ainsi un candidat pertinent par rapport aux caméras optiques pour l'observation et la modélisation des phénomènes sous-marins (Caimi *et al*, 2010).

1.2.2 Besoin d'une représentation tridimensionnelle de l'environnement

La représentation tridimensionnelle de l'environnement présente un véritable intérêt dans le contexte de mise en œuvre de solutions préventives des changements climatiques. En effet, une représentation 2D, qui ne peut présenter synthétiquement qu'une partie des données ou des informations, n'est pas en mesure de représenter fidèlement le milieu où le phénomène est observé. Par contre, il est possible d'obtenir explicitement à partir d'une représentation 3D plusieurs informations pertinentes permettant l'analyse et le suivi des problèmes environnementaux (ex. inondations, propagations de pollutions aériennes ou de liquides) et de sécurité (ex. modélisation de feux ou d'accidents liés à des risques naturels) (Pomnon, 2009). Dans le contexte de l'environnement marin, la représentation tridimensionnelle a toujours été considérée comme un moyen pertinent pour représenter l'espace et la dynamique physique du milieu marin (Cugier *et al*, 2001). Elle permet de déduire et de calculer la position ou la profondeur des fissures présentes sur des structures sous-marines, la taille ou la densité d'herbiers présents sur le fond marin, à une période donnée, ou bien encore

⁹ <http://www.soundmetrics.com/industries-served>

¹⁰ <http://www.blueview.com/2d-Imaging-Sonar.html>

d'estimer et de représenter les agrégations de poissons (Carette, 2008). On assiste ainsi ces dernières années, au développement de divers modèles tridimensionnels permettant la modélisation de l'hydrodynamique, des sédiments, de la matière en suspension, de l'écologie, etc. tels que le modèle hydrosédimentaire SIAM 3D développé par Cugier et Le Hir (2002), le modèle 3D de la matière en suspension de Cugier et Le Hir (2000) (Cugier *et al*, 2001), le modèle représentant l'agrégation de poissons (Carette, 2008), etc.

1.2.3 Besoin d'une information précise et conforme à la réalité pour une représentation tridimensionnelle robuste

Une représentation tridimensionnelle fidèle à la réalité de la scène observée résulte en partie de l'acquisition d'une information précise, complète et non ambiguë. La prise en compte de la qualité de l'information est primordiale tout particulièrement dans le cadre de l'observation de l'environnement. En effet, une information présentant un certain degré d'imperfection influence la qualité des résultats obtenus en introduisant des erreurs au niveau de la géo-localisation des objets (ex. fissures dans les structures), du suivi de l'évolution spatiale d'un phénomène (ex. l'érosion des fonds, la dynamique des poissons), etc. Ces imperfections peuvent aussi s'accumuler et être à l'origine de mauvaises prises de décision, celles-ci s'appuyant sur des éléments de connaissances erronés ou insuffisamment représentatifs des conditions réelles.

Les données issues du milieu marin sont de nature imparfaite tout particulièrement quand il s'agit de données permettant de définir les frontières des zones d'herbiers ou de classer les types de sédiments sur le fond. Selon Wright et Goodchild (1997), la position et les frontières des objets en milieu marin tendent à être plus imprécises et floues que celles déterminées dans un environnement terrestre. En effet, contrairement au milieu terrestre, il n'existe pas de points géodésiques ni de points de référence équivalents dans les océans. En conséquence, les données marines sont souvent acquises par rapport à la position d'objets qui sont continuellement en mouvement, comme des navires, et qui ont par conséquent des positions imprécises et incertaines (Carette, 2008). Il est vrai que dans le contexte d'une reconstruction tridimensionnelle à partir d'images acoustiques, le positionnement de la scène reconstruite sera initialement « relatif ». Néanmoins, il est inévitable de passer au positionnement « absolu » quand il s'agit de l'exploitation des informations concernant la position exacte de la scène dans le milieu marin. Dans ce cas, la notion globale d'incertitude doit être prise en compte pour que les résultats soient utiles lors la phase décisionnelle.

1.3 Problématique : exploitation de caméras acoustiques pour la modélisation tridimensionnelle de l'environnement

Comme vu dans la partie précédente (Paragraphe 1.2.1), les caméras acoustiques se présentent comme une technologie d'imagerie très prometteuse pour l'exploration du milieu sous-marin, que ce soit à des fins

d'inspection des infrastructures (Asada, 2008), ou pour l'observation et le contrôle de l'évolution de la faune et de la flore présentes dans cet environnement (Jun et Asada, 2007). Les caméras acoustiques produisent des séquences d'images offrant une visualisation d'une même scène sous différents angles de vue. Par analogie avec d'autres capteurs, notamment optiques, fournissant de multiples points de vue consécutifs sur une même scène, et permettant de reconstruire cette scène en 3D, il devrait donc être possible d'exploiter ces séquences d'images acoustiques afin de produire des représentations tridimensionnelles de la scène. Comme décrit au paragraphe (1.2.2), les représentations tridimensionnelles offrent un moyen d'observation pertinent duquel diverses informations sur la position des objets, leur hauteur, leur forme, etc. peuvent être déduites. De plus, il a été souligné que de telles modélisations 3D deviennent incontournables pour les systèmes visant l'adaptation aux impacts des changements climatiques ce qui explique la diversité des modèles océaniques 3D recensés dans la littérature (ex. modèle 3D de la sédimentation, des agrégations des poissons, etc.). Ces divers modèles permettent d'évaluer l'évolution du milieu au fil du temps. Néanmoins, on ne trouve pas de modèles de représentations tridimensionnelles de structures, de la faune et de la flore marines, développés à partir d'images bidimensionnelles de caméras acoustiques. Les paragraphes suivants vont détailler les problématiques inhérentes à la mise en œuvre de représentations tridimensionnelles précises et robustes à partir de données issues de caméras acoustiques.

1.3.1 Comment peut-on effectuer une reconstruction 3D d'une scène sous-marine à partir d'images issues de caméras acoustiques ?

En imagerie sonar, peu de travaux de recherche ont abordé la reconstruction 3D à partir d'images multi-vues issues de caméras acoustiques. Les travaux exploitant les images acoustiques concernent essentiellement les sonars latéraux (Sun *et al*, 2008; Lu et Oshima, 2002) 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, celles-ci 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 formes convexes et 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 un contexte de reconstruction de grandes structures statiques ni dans le cas d'objets en mouvement.

En imagerie optique, les approches de reconstruction à partir d'images multi-vues sont multiples (ex. stéréovision, reconstructions volumétriques, etc.). Par contre pour appliquer ces approches aux cas des

caméras acoustiques, il faut faire face à deux problématiques liées essentiellement à la géométrie de formation et à la nature bruitée des images.

1.3.1.1 La géométrie de formation des images

Contrairement aux images optiques où la formation des images suit généralement le modèle sténopé (Hartley et Zisserman, 2004), les images issues de caméras acoustiques ont une géométrie de formation spécifique (cf. Figure 1). 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 présenté dans la Figure 1, l'axe des abscisses (X) 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 ordonnées (Y), quant à lui, correspond à l'angle entre l'axe du capteur (ligne centrale de l'image) et les éléments rétrodiffusants. Par conséquent, un point géographique dans la scène est repéré dans l'image acoustique par la portée (distance caméra-scène) et l'angle d'insonification (azimut). Schématiquement, 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 azimut θ (l'angle d'insonification). La troisième dimension, qui correspond à l'élévation ϕ du point par rapport à la caméra est de ce fait, perdue lors de l'acquisition des images (GESMA, 2009, Negahdaripour, 2007).

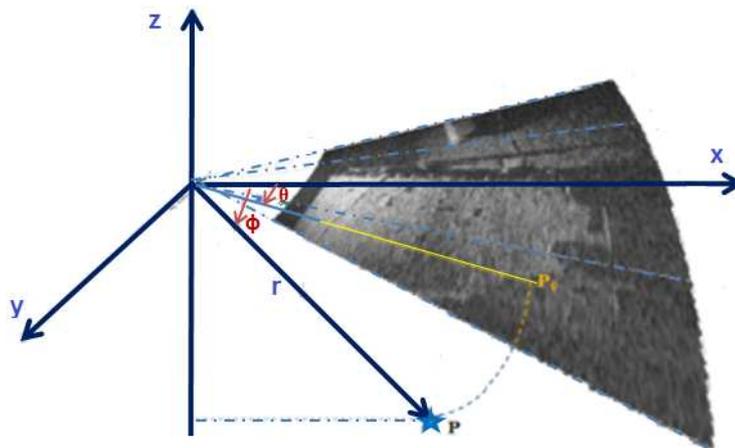


Figure 1 : Projection d'un point (P) de l'environnement sous-marin sur le plan d'une image DIDSON (placé arbitrairement en élévation, ici l'élévation ϕ vaut 0). L'image est acquise dans un système de coordonnées sphériques centrée sur la caméra acoustique. L'axe Y représente l'axe de l'azimut du sonar, l'axe X est le deuxième axe du plan image, perpendiculaire à l'axe Y et orienté suivant les portées positives et l'axe Z est l'axe perpendiculaire au plan image, orienté suivant les élévations positives (GESMA, 2008).

De plus, 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. Cet ensemble de rétrodiffusion se trouve sous forme d'un arc de cercle voire un ruban très fin si on prend en compte l'ouverture verticale ($10,9^\circ$ dans le cas de la DIDSON) et l'ouverture horizontale du faisceau ($0,3^\circ$ en basse fréquence ou $0,4^\circ$ en haute fréquence dans le cas de la DIDSON) (cf. Figure 2).

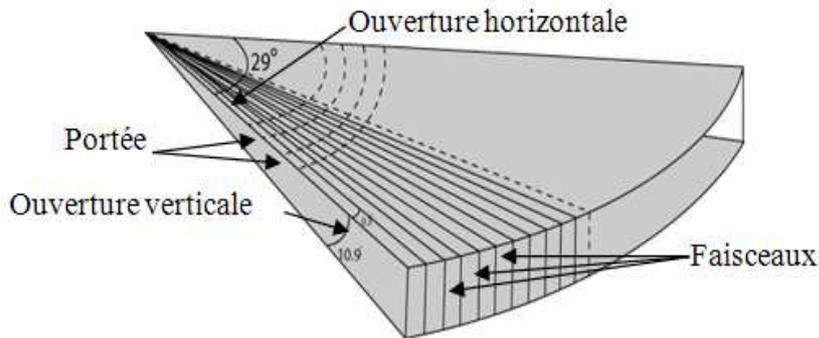


Figure 2 : Faisceau d'insonification de la caméra DIDSON.

1.3.1.2 La nature bruitée et l'inhomogénéité radiométrique des images

Les caractéristiques radiométriques des images acoustiques font de la reconstruction 3D une problématique relativement difficile à résoudre. En effet, la reconstruction s'appuie essentiellement sur les informations enregistrées au sein des images. Une étape d'extraction de ces informations est ainsi nécessaire. Néanmoins, les images sonar 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 d'une part de la nature hétérogène du milieu marin et des phénomènes d'absorption et de diffusion des ondes acoustiques dans l'eau et d'autre part de leur interaction avec la scène. 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 de surbrillance sur les images causant des ambiguïtés lors de la recherche d'une même information dans des images différentes. Par conséquent, deux pixels représentant une même région dans la scène peuvent avoir des caractéristiques radiométriques différentes sur des images différentes.

1.3.2 Comment peut-on exploiter les informations issues des caméras acoustiques pour obtenir une reconstruction 3D qui soit complète et robuste ?

Les caméras acoustiques produisent des séquences d'images où la scène sous-marine peut être visualisée sous différents angles de vue. Ces séquences sont ainsi riches en informations spatiales, décrivant la

géométrie de la scène, et en informations multi-temporelles dérivant de l'acquisition continue des images. Ces deux dimensions induisent donc redondances et complémentarités dans les données mises à disposition pour la reconstruction 3D.

- La redondance est observée sur un ensemble d'images dont l'acquisition est temporellement très proche. Dans ce contexte, à un même point de la scène va correspondre une série d'informations associées à chacune des positions d'acquisition.
- La complémentarité est une caractéristique très importante lors de la recherche d'une reconstruction complète, fiable et fidèle. Elle provient en partie, de la multiplicité de points de vue de la même scène au travers d'une séquence d'images. Cette multiplicité introduit de fait une complémentarité spatiale des images entre elles permettant d'obtenir en final une vue globale des objets présents dans la scène.
- Par ailleurs, plusieurs algorithmes ou traitements avec leurs spécificités, forces et faiblesses extraient à partir d'une seule image des informations différentes mais de même nature (par exemple, utilisation de différents détecteurs de contours sur une image). L'exploitation de la complémentarité des attributs ainsi extraits peut contribuer à l'amélioration (robustesse ou/et précision) des informations finales extraites. Cette complémentarité des attributs n'est évidemment pas exclusive avec la complémentarité spatiale. L'exploitation de la redondance, de la complémentarité spatiale et de celle des attributs, associées à l'aspect multi-temporel des séquences d'images acoustiques permet d'améliorer la qualité de l'information spatiale produite afin de fournir une réponse adaptée aux besoins opérationnels pour l'analyse de la situation observée. Elle permet notamment d'améliorer la robustesse de l'analyse vis-à-vis de son environnement et l'acuité des informations transmises. Le problème auquel on fait face alors concerne la méthodologie à adopter afin de pouvoir exploiter cette redondance, complémentarité et multi-temporalité des différentes séquences d'images à disposition pour la reconstruction 3D de la scène sous-marine observée.

1.4 Hypothèses et objectifs de la recherche

1.4.1 Hypothèses

Afin d'apporter une solution à la problématique énoncée dans le cadre de ce projet de recherche, les hypothèses suivantes ont été établies :

- 1) La représentation tridimensionnelle d'une scène sous-marine offre un moyen d'observation pertinent, fidèle de la réalité observée et duquel on peut extraire diverses informations sur l'état de

l'environnement (ex. localisation des fissures sur les structures, des herbiers sur le fond, etc.). Celle-ci devient incontournable pour les solutions de prévention des impacts des changements climatiques.

- 2) Les caméras acoustiques sont de bons candidats dans le cadre d'applications visant l'observation, l'inspection ou le contrôle quasi en continu de scènes ou d'infrastructures sous-marines. Elles permettent d'acquérir des séquences d'images dans des milieux profonds, turbides où la visibilité est très faible voire nulle (Belcher *et al*, 2003). Ces images peuvent être exploitées dans le cadre d'études visant l'observation de l'environnement (Caimi *et al*, 2010). Plusieurs travaux ont déjà montré qu'il était possible de visualiser les structures sous-marines, ou de suivre et de dénombrer des poissons à partir de ces séquences d'images (Jun et Asada, 2007; Burwen *et al*, 2007, Caimi *et al*, 2010). Toutefois elles n'ont pas été exploitées jusqu'à présent afin d'élaborer des représentations tridimensionnelles ni pour effectuer un suivi des changements intervenant dans l'environnement marin.

1.4.2 Définition de l'objectif général et des objectifs spécifiques

L'objectif général de ce travail est de proposer une méthodologie permettant la reconstruction 3D complète et fidèle de l'environnement marin à partir d'images issues de caméras acoustiques. Cette méthodologie permettra de mettre en œuvre des solutions visant l'observation de la faune et de la flore et l'inspection des structures dans un contexte de prévention face aux changements climatiques.

Le problème fondamental auquel on est confronté pour atteindre cet objectif consiste à élaborer une approche qui prenne en compte la géométrie de formation d'images spécifique aux caméras acoustiques et le caractère évolutif du milieu marin afin d'aboutir à une reconstruction complète de l'environnement. Si le problème est résolu et l'objectif est atteint, les hypothèses de départ seront validées.

Pour atteindre cet objectif, plusieurs objectifs spécifiques structurant les étapes de la méthodologie, devront être réalisés. Ces objectifs spécifiques s'énoncent comme suit:

- 1) Proposer un cadre conceptuel permettant la mise en œuvre des différentes étapes d'une méthodologie de reconstruction tridimensionnelle complète d'une scène observée par des caméras acoustiques.
- 2) Proposer des approches permettant l'extraction d'informations pertinentes pour la reconstruction 3D de scènes sous-marines à partir des images acoustiques.
- 3) Proposer une approche permettant de calculer l'élévation de la scène pour aboutir à une scène tridimensionnelle reconstruite conforme à la réalité.

1.5 Méthodologie

La méthodologie proposée pour atteindre les objectifs et sous-objectifs énoncés précédemment a combiné deux approches méthodologiques : une approche conceptuelle et une approche hypothético-déductive. En suivant ces approches, la réalisation de la thèse s'est répartie sur sept étapes. Ces approches ainsi que les étapes de déroulement de la thèse sont décrites ci-dessous.

1.5.1 L'approche méthodologique

Afin de faire une collecte de connaissances, des théories et des travaux se rapportant à la thématique de la recherche proposée, nous avons commencé la thèse par une recherche conceptuelle. Cette dernière s'est appuyé essentiellement sur une revue de littérature sur les caméras acoustiques, la reconstruction 3D en vision numérique et en imagerie acoustique. Il a ainsi été possible d'identifier les problématiques relatives à l'utilisation des caméras acoustiques pour la reconstruction tridimensionnelle de l'environnement sous-marin et les différentes approches possibles pour y apporter une solution. Cette revue de littérature, présentée dans le chapitre 2, a conduit à la proposition d'un cadre conceptuel. Ce dernier présente les idées directrices et les bases théoriques de la solution proposée. Les principaux composants de cette solution, leur organisation structurant le flux de traitement des données depuis la séquence d'images acoustiques jusqu'à la représentation 3D de la scène y sont présentés.

Une fois la recherche conceptuelle complétée, la démarche méthodologique a consisté à appliquer une méthode hypothético-déductive. En effet, la revue de littérature a permis d'identifier un ensemble d'hypothèses concernant les théories et les approches envisageables pour répondre à la problématique énoncée ci-avant (la qualité des images acoustiques, la géométrie de formation des images et l'exploitation des informations acoustiques pour la reconstruction de l'environnement). Ces hypothèses ont été inspirées de diverses disciplines (vision numérique, photogrammétrie, traitement d'images sonar et optique, approches d'optimisation, etc.). Au fur et à mesure de la compréhension des problématiques identifiées lors de la recherche conceptuelle, nous avons été à même de déduire et d'extraire de toutes ces disciplines les connaissances nécessaires pour mettre en œuvre une méthodologie permettant d'atteindre l'objectif de la recherche.

1.5.2 Déroulement de la thèse

La méthodologie globale appliquée pour atteindre les objectifs de la recherche est présentée dans la Figure 3. En incluant les étapes correspondant à l'étude bibliographique et l'élaboration du cadre conceptuel, elle est composée des 7 étapes suivantes :

1ère étape : Recherche bibliographique générale qui a abouti à l'identification du contexte et des besoins qui ont conduit au développement du présent projet de recherche.

2ème étape : Élaboration de la problématique de recherche et formulation des objectifs de recherche.

3ème étape : Revue de littérature approfondie qui a mis en évidence la multitude des approches possibles pour répondre aux objectifs de recherche. Des approches théoriques ont été identifiées tout particulièrement dans le domaine de la vision numérique. Ces approches ont été classées comme solutions potentielles pour la résolution de la problématique énoncée dans l'étape précédente.

4ème étape : Mise en place progressive d'un cadre conceptuel visant à cerner la démarche méthodologique à appliquer. Ce cadre conceptuel s'est enrichi tout au long de la recherche, au fur et à mesure des lectures théoriques. Il correspond à la structuration progressive d'une première proposition théorique d'une méthodologie. Cette étape s'est concrétisée par la rédaction d'un article publié dans la revue *Traitement du Signal* (Brahim *et al*, 2012).

Ces premières étapes (la mise en contexte, les besoins, la problématique, les objectifs et le cadre conceptuel) ont été validées par le comité de thèse lors de l'examen doctoral (Août 2010).

5ème étape : Développement d'une méthode de reconstruction 3D multi-vues à partir d'un ensemble d'images acquises par des caméras acoustiques. Les traitements développés peuvent être regroupés en deux catégories. La première correspond aux traitements de bas niveau. En effet, l'utilisation directe des images pour la reconstruction n'est pas possible parce que les images sont bruitées et contiennent une quantité importante et redondante d'information. Une première série de traitements est donc requise afin d'extraire un ensemble de primitives caractérisant la scène observée sur les images. La deuxième catégorie correspond aux traitements de haut niveau permettant d'exploiter les primitives afin de passer d'une représentation bidimensionnelle à une représentation tridimensionnelle de la scène. Ces deux catégories de traitements ont été développées sur deux périodes successives, on les a ainsi classées en deux sous-étapes :

- 1^{ère} sous-étape : Développement des traitements de bas niveau;
- 2^{ème} sous-étape : Développement des traitements de haut niveau;

Entre ces deux sous-étapes, nous avons réalisé des expérimentations dans un bassin intérieur au cours desquelles des acquisitions de données ont été effectuées. Ces données ont été utilisées par la suite pour appliquer et tester les traitements développés lors de la 5ème étape. Compte tenu de notre avancement dans

la recherche, les acquisitions ont été spécifiées en fonction des approches mises en œuvre dans ces traitements. Les expérimentations sont décrites dans la 6^{ème} étape.

6ème étape : Réalisation d'expérimentations en piscine afin de collecter des données représentant des objets de différentes tailles, observés selon différents points de vue, tout en connaissant leurs positions et orientations. Ces acquisitions de séquences d'images acoustiques DIDSON ont été réalisées à la piscine profonde de l'Université Laval (Décembre 2010).

7ème étape: Dans chacune des catégories de traitement, un ensemble de tests a été réalisé. Ces tests ont permis de tester les approches et de les améliorer d'une façon itérative.

8ème étape: Évaluation finale de la méthodologie globale de la reconstruction 3D. Cette étape a conclu le travail conduisant à l'élaboration d'une synthèse des difficultés rencontrées et des perspectives possibles de cette recherche. Elle a débouché sur la rédaction de la thèse.

1.5.3 Implémentation et données utilisées

L'implémentation des algorithmes a été réalisée dans l'environnement Matlab. Comme décrit dans la 5^{ème} étape, ces algorithmes sont classés en deux catégories : des algorithmes de bas niveau permettant d'extraire les primitives des images (ex. extraction de contours, filtrage gaussiens, etc.) et des algorithmes de haut niveau pour la reconstruction 3D (ex. l'utilisation d'algorithmes d'optimisation). Au fur et à mesure de la mise en œuvre des étapes des algorithmes, le travail et les résultats ont été publiés dans des articles et présentés dans des conférences : Oceans (Oceans'10 et Oceans'11), CHC 2010, Taima 2011. Un article a été également publié dans la revue *Traitement du Signal* (Brahim *et al*, 2012).

Les algorithmes développés sont appliqués sur des séquences d'images acoustiques issues d'une caméra DIDSON. Ces séquences représentent différents contextes et sont acquises dans différents milieux. Elles correspondent à des séquences acquises avec différentes fréquences en milieu marin, des séquences d'images simulées et des séquences acquises lors des expérimentations dans un bassin intérieur (6^{ème} étape).

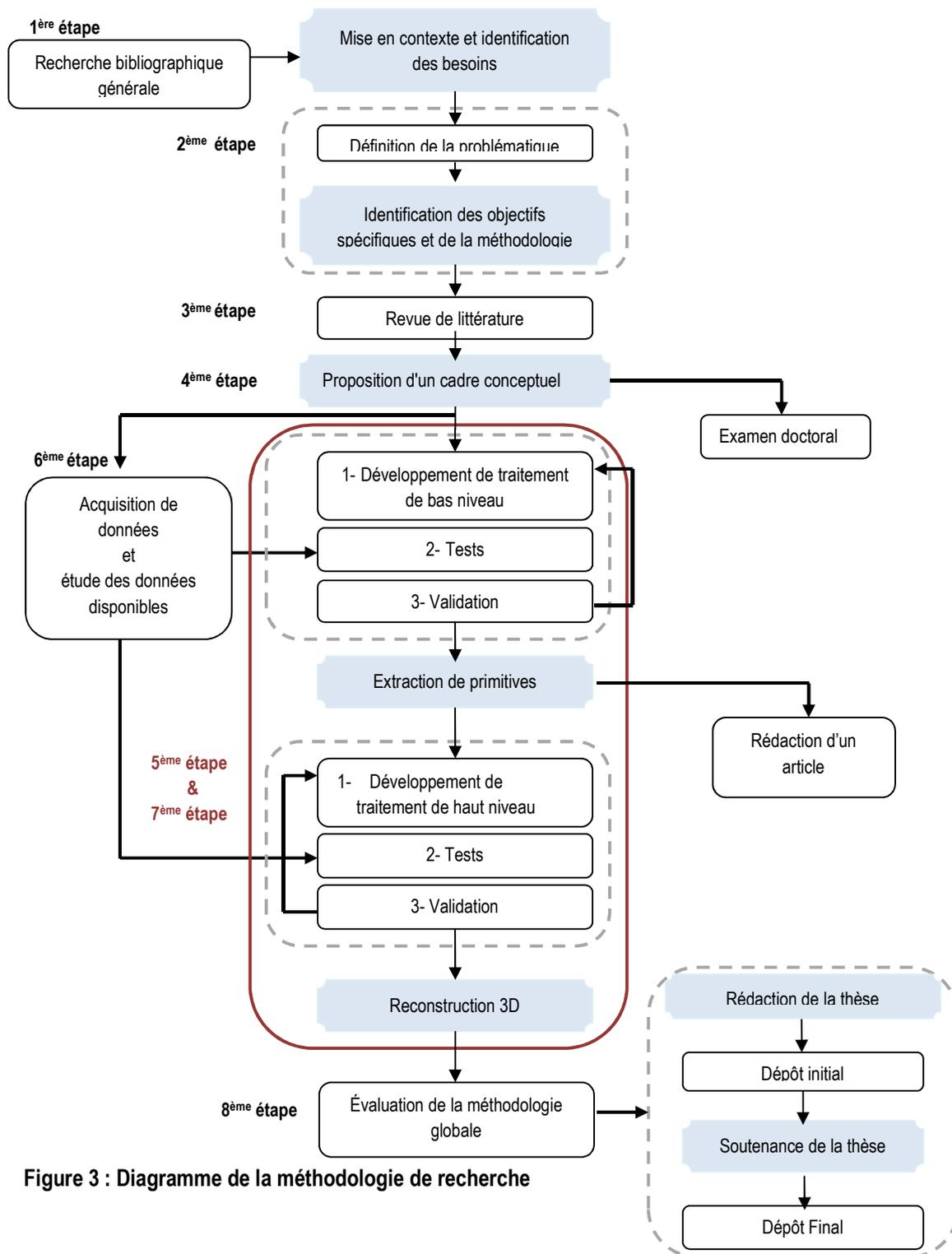


Figure 3 : Diagramme de la méthodologie de recherche

1.6 Structure de la thèse

L'articulation des différents chapitres de la thèse va suivre les principales étapes de la méthodologie énoncées ci-dessus. Ainsi, le prochain chapitre va porter en premier lieu sur la revue de littérature qui se concentre tout particulièrement sur les différentes approches de reconstruction 3D recensées en vision numérique et en imagerie sonar et en deuxième lieu sur le cadre conceptuel proposé. Le chapitre 3 porte sur les primitives et leur extraction. Quant au chapitre 4, il est consacré à la présentation des approches développées pour l'estimation de l'élévation associée à chaque pixel. Dans ces deux chapitres, des expérimentations et des résultats d'analyses vont être présentés. Le chapitre 5 propose un ensemble de conclusions et de perspectives quant au travail de recherche réalisé.

Chapitre 2. État de l'art

2.1 Introduction

Ce chapitre va être consacré en premier lieu à la présentation générale des caméras acoustiques, à leurs domaines d'application, ainsi qu'à leur modèle de formation des images. Cette description est nécessaire afin d'identifier les approches et la méthodologie de reconstruction 3D qui seront les plus adaptées aux images issues de caméras acoustiques. La suite du chapitre est consacrée à l'état de l'art présentant les différentes approches de reconstruction tridimensionnelle issues des domaines de la vision numérique et de l'imagerie sonar. Nous concluons ce chapitre par la proposition d'un cadre conceptuel dans lequel s'inscrit notre démarche de reconstruction 3D du milieu marin.

La partie concernant le cadre conceptuel a fait l'objet d'un article publié dans la revue *Traitement du Signal* et intitulé « Vers une méthodologie de reconstruction 3D de scènes sous-marines par caméras acoustiques : Problématiques et études préliminaires » (Brahim *et al*, 2012).

2.2 Les caméras acoustiques

2.2.1 Présentation générale et domaines d'application

Les caméras acoustiques ont été conçues dans l'objectif de développer un capteur sonar haute fréquence à des fins d'inspection militaire (Belcher *et al*, 2003). Ces caméras se caractérisent essentiellement par la production de séquences d'images avec un taux de rafraîchissement élevé, proche de celui des capteurs optiques. Parmi ces caméras, on peut citer les caméras DIDSON (Dual Frequency Identification Sonar), LIMIS (Limpet Mine Imaging Sonar) (Belcher *et al*, 2003) ou encore BlueView¹¹. Les applications de ces caméras sont multiples : en plus des tâches d'inspection des milieux portuaires et d'identification d'objets (Belcher *et al*, 2007), elles ont été 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'observation et l'analyse du comportement des bancs de poissons, ainsi que pour le suivi de l'évolution de la faune (Belcher *et al*, 2007).

Dans notre travail, on s'intéresse aux caméras acoustiques comme outil d'observation du milieu marin. Étant donné que les caméras acoustiques se ressemblent au niveau de leur mode de fonctionnement ainsi qu'au niveau des caractéristiques radiométriques et géométriques de leurs images (Belcher *et al*, 2003), nous allons restreindre la description de leur modèle de projection des images à celui de la caméra DIDSON.

¹¹ <http://www.blueview.com/2d-Imaging-Sonar.html>

2.2.2 Modèle géométrique de la caméra DIDSON

Selon le modèle géométrique de la caméra DIDSON, un point P de la scène, en coordonnées cartésiennes, est défini par ses coordonnées 3D selon l'axe des X, l'axe des Y et l'axe des Z, $P = (x, y, z)^T$. En coordonnées sphériques dans un repère dont l'origine est la caméra, les composantes de ce point P deviennent :

- 1) l'angle d'insonification θ c'est-à-dire le faisceau (par rapport à l'axe de pointage du sonar) auquel le point P contribue;
- 2) la portée r , c'est-à-dire la distance entre le point P et le sonar
- 3) l'élévation φ .

Le modèle mathématique de la caméra liant ces deux systèmes de coordonnées est représenté comme suit (Negahdaripour *et al*, 2005) :

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

Avec un tel modèle, tout point P de la scène se projette dans une image bidimensionnelle en un point $P_0(r, \theta)$. Le point P_0 dans l'image est représenté par un pixel dont les dimensions dépendent de sa position par rapport à la portée (r) et par rapport à l'angle (θ). La projection de ce pixel dans la scène est une région à laquelle appartient le point P. Quant à l'élévation φ de ce point P, elle est perdue lors de sa projection dans l'image. Ce sont les images bidimensionnelles qui forment les données brutes qui vont être utilisées dans le présent travail pour remonter à l'élévation de chaque pixel.

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 ci-dessus. Les lentilles acoustiques font partie des composants des caméras acoustiques et interviennent dans le processus d'acquisition des images. L'utilisation de ces lentilles induit l'apparition de distorsions lors de la formation des images. Pour évaluer les imperfections engendrées par ces distorsions, 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. Les expérimentations ont été réalisées dans un bassin contrôlé afin d'imager la scène avec précision. 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 cm pour la variance dans le cas de la portée et $0,0003^\circ$ pour la moyenne et de $0,946^\circ$ pour la variance dans le cas de l'angle d'insonification (azimut). L'erreur au niveau de la portée est importante par rapport à celle au niveau de l'azimut. En fonction de la précision visée par l'application, la prise en compte de ces erreurs dues aux lentilles peut s'avérer nécessaire. Ces erreurs évaluées par Negahdaripour (2005) sont issues uniquement des distorsions des lentilles de la caméra DIDSON. Les imperfections engendrées par les caractéristiques physiques de l'environnement marin, notamment la turbidité de l'eau ne n'ont pas été évaluées.

2.3 La reconstruction 3D multi-vues en vision numérique

La reconstruction 3D d'une scène à l'aide d'approches de vision numérique 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. Parmi les méthodes proposées dans ce domaine, on s'intéresse à celles qui appartiennent à la vision passive. En effet, il existe deux grandes catégories de vision : la vision active et la vision passive. Malgré l'appartenance des caméras acoustiques à la catégorie de la vision active où les objets sont reconstruits à l'aide d'un éclairage contrôlé illuminant la scène d'une manière particulière (Roussel, 1999), les approches proposées dans cette catégorie ne répondent pas au contexte de mise en œuvre de ces caméras. En effet, les caméras acoustiques n'enregistrent pas l'information d'une façon passive comme c'est le cas pour les caméras optiques. Elles sont aussi la source de l'énergie contrôlée pour « illuminer » la scène. Elles ne sont pas combinées avec d'autres outils supplémentaires (ex. système de projection d'un motif structurant) (Ben Amor, 2006) ce qui limite les informations géométriques mises ainsi en évidence et ne facilitent pas leur exploitation lors de la reconstruction de la scène observée du fait principalement de la variation des intensités des images acquises. On se trouve ainsi dans le contexte de 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.

Dans le contexte de la reconstruction exploitant plusieurs images en vision passive, on distingue deux grandes catégories d'approches : stéréo (deux images) et multi-images (plus de deux images). Les paragraphes suivants vont s'intéresser à ces approches et les techniques qui y seront citées concernent des images optiques acquises par des capteurs sous-marins ou terrestres. Néanmoins, elles peuvent s'avérer intéressantes dans le contexte de la reconstruction 3D à partir d'images issues de caméras acoustiques.

2.3.1 Méthodologie générale de la reconstruction 3D en vision numérique

Que la reconstruction 3D s'opère selon une approche « stéréo » lorsqu'une paire d'images est utilisée ou « multi-images » lorsque plus de deux images sont utilisées, le principe général est similaire et s'appuie sur les étapes suivantes (cf. Figure 4) :

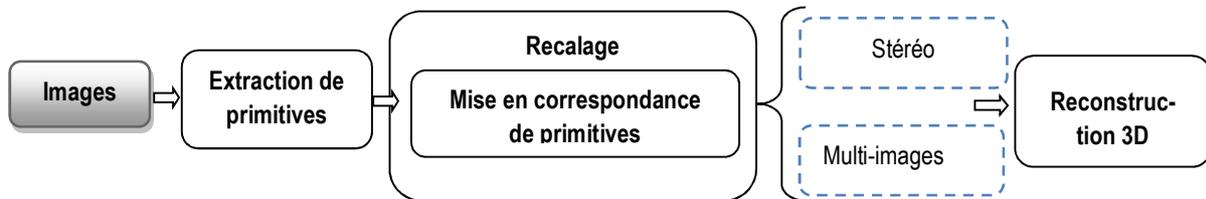


Figure 4 : Étapes de reconstruction 3D à partir d'un ensemble d'images.

L'extraction de primitives vise à localiser dans les différentes images, les projections d'une même entité (ou caractéristique) de la scène. La qualité de la reconstruction 3D va dépendre de la qualité de l'appariement entre ces projections identifiées sur les images. Ainsi, l'étape d'extraction des primitives est cruciale. De plus, elle s'avère également être un problème délicat car de nombreuses difficultés peuvent entraîner des résultats erronés comme par exemple, le bruit présent dans les images et les occlusions (zones de la première image contenant des pixels sans correspondance dans la seconde image). Le recalage des images correspond au calcul de la transformation inter-images liant les entités mises en correspondance. Quant à l'étape finale, la reconstruction 3D, celle-ci correspond au calcul du modèle 3D de la scène obtenue à partir des informations extraites des différentes images et mises en correspondance. Ces approches sont décrites dans les paragraphes suivants.

2.3.2 Extraction, mise en correspondance et recalage de primitives

On peut définir plusieurs familles de méthodes d'extraction de primitives dans le cadre de la définition des méthodes de recalage. Elles s'appuient sur l'extraction de primitives visibles dans les images telles que les pixels individuels, les contours, les régions ou les points caractéristiques :

- *Primitives pixels* : elles correspondent à l'information associée aux pixels (l'intensité) de l'image. Ces primitives sont exploitées tout particulièrement lors du recalage iconique. C'est une catégorie de méthodes de recalage qui n'exploite que l'intensité des images et une mesure de similarité entre les images pour le calcul de la transformation inter-image (Rominger et Martin, 2010).
- *Primitives ponctuelles ou points anguleux (points d'intérêt)* : elles caractérisent les zones ponctuelles de forte variation d'intensité, d'entropie, etc. Leur extraction peut se faire selon deux approches à

savoir à partir des contours ou tout simplement par l'application de détecteurs de points (Rosten *et al*, 2010).

- *Primitives curvilignes ou contours* : ces primitives sont souvent des segments de droite ou des courbes extraits selon une approche de segmentation par contours et ceci à travers l'application de filtres (ex. Canny, Sobel, etc.) (Maitre, 2003, Chapitre 2).
- *Primitives surfaciques ou régions* : elles regroupent les pixels connexes ayant des caractéristiques communes telles que l'intensité, l'entropie. Ces caractéristiques sont généralement extraites selon une approche de segmentation par régions (Gonzalez et Woods, 2008, Chapitre 10).

2.3.2.1 Mise en correspondance de primitives

La mise en correspondance correspond à l'étape d'appariement de primitives caractéristiques, extraites d'images différentes. Elle permet d'identifier les primitives homologues (ex. points identiques dans deux images prises à partir d'angles de vue différents) afin d'estimer la transformation permettant de recalibrer les images. Les approches de mise en correspondance peuvent se baser sur des mesures de corrélation, d'entropie, sur la somme des différences au carré (« Sum of Square Differences » en anglais), (Rabaud, 2005), etc.

2.3.2.2 Recalage : calcul de la géométrie inter-image

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. 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 (ex. primitives surfaciques, curvilignes). Selon le nombre d'images utilisées, on distingue deux catégories de recalage à savoir le recalage par stéréovision (2 images) et le recalage par multi-images (plus de 2 images). Une présentation de ces deux catégories est proposée dans les paragraphes suivants.

2.3.2.2.1 Recalage par stéréovision

La stéréovision caractérise la technique permettant de reconstruire l'information tridimensionnelle d'une zone à partir d'un couple d'images acquises selon des angles de vue différents. Généralement, en imagerie optique, la caméra est modélisée selon le modèle sténopé (« pinhole » en anglais). 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. À partir de cette simplification, le modèle sténopé correspond à un modèle projectif homographique simple qui permet de calculer la position spatiale de points à partir des coordonnées de leurs projections dans deux vues différentes afin de reconstruire la structure tridimensionnelle de la scène (Trucco et Verri, 1998, Brown *et al*, 2003).

Dans ce contexte de la stéréovision, le recalage consiste à se placer dans un modèle particulier qui décrit les relations géométriques entre les deux images. Cette géométrie dite épipolaire est illustrée dans la Figure 5. Soit P un point 3D de l'environnement observé par deux caméras C_1 et C_2 ayant chacune son repère propre R_1 et R_2 . P_1 et P_2 sont les points image de P dans les plans images de ces deux caméras. La transformation entre les deux repères est définie par la translation T et la rotation R . Le plan construit par P et ses points images associés P_1 et P_2 est le plan épipolaire. Ces paramètres épipolaires (P_1 , P_2 , les repères et le plan épipolaire) servent à définir deux matrices importantes qui sont la matrice essentielle et la matrice fondamentale et qui aboutissent au calcul de l'élévation des points de la scène à partir de leurs points image. Ces matrices sont définies dans (Trucco et Verri, 1998).

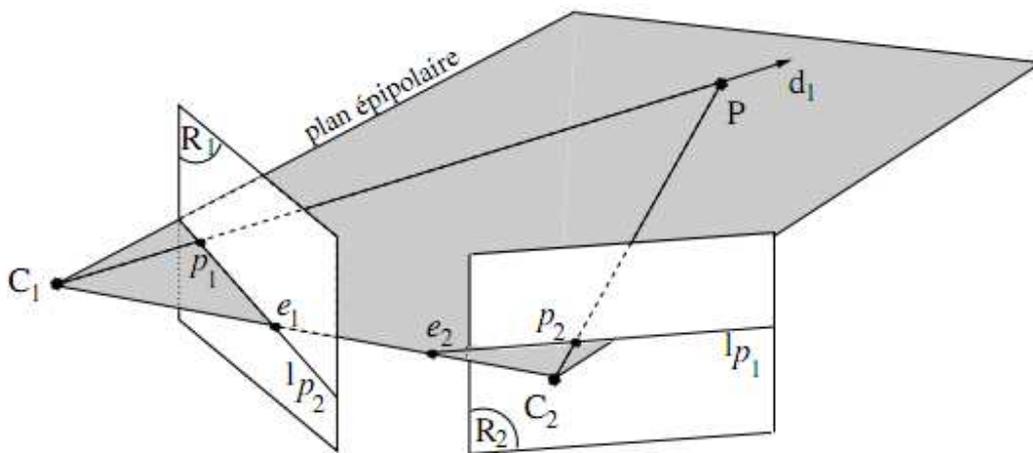


Figure 5 : La géométrie épipolaire. C_1 et C_2 sont les deux caméras, les droites l_{P_1} et l_{P_2} sont des droites épipolaires, P est un point dans l'espace 3D et P_1 de P_2 sont la projection du point P sur deux images, R_1 et R_2 sont les repères respectifs associés aux caméras C_1 et C_2 , d_1 est la droite passant par le centre de C_1 et P_1

2.3.2.2.2 Recalage par multi-stéréo et multi-images

Contrairement à la stéréovision, le recalage multi-images s'appuie sur l'utilisation de plus de deux images pour la reconstruction 3D. On peut concevoir la multi-stéréo comme une généralisation de la stéréovision en traitant une séquence de n images comme une séquence de $(n-1)$ paires stéréo. Les approches de la stéréovision seront appliquées autant de fois que de nombre de paires d'images. On peut aussi concevoir le recalage multi-images comme une prise en considération de toutes les images en même temps. On parle ainsi d'un « bloc d'images » où les traitements qui y sont appliqués sont spécifiques et différents de ceux de la multi-stéréo. Ces images peuvent appartenir à une séquence d'images prise par une seule caméra ou bien à un ensemble d'images prises par des caméras différentes.

Généralement les approches de recalage multi-images déterminent à la fois les transformations géométriques entre les images et la structure 3D de la scène. On distingue deux ensembles d'approches. Le premier ensemble s'appuie sur des reconstructions incrémentales en se basant sur l'extraction de primitives ponctuelles. Ces méthodes utilisent généralement des algorithmes d'optimisation (ex. Levenberg-Marquedt, Gauss-Newton, la descente du gradient) (Mouragnon *et al*, 2006) dont l'objectif est la minimisation de l'erreur de reprojection des points tridimensionnels afin de déterminer la transformation inter-image entre l'ensemble de points homologues et aussi la position 3D dans l'espace correspondant. La résolution du problème par des algorithmes d'optimisation nécessite d'une part de choisir la méthode d'optimisation qui minimise le risque de calculer des solutions sous-optimales, et d'autre part de ne pas aboutir à un résultat en un temps prohibitif, le temps de calcul étant le principal inconvénient de ces méthodes tout particulièrement quand il s'agit de traiter une longue séquence d'images. En fait, à chaque nouvelle image de la séquence les points homologues sont identifiés et intégrés dans la procédure d'optimisation. La qualité des résultats finaux dépend aussi des paramètres d'initialisation des algorithmes utilisés (pose initiale de la caméra et estimation des points 3D) (Hartley et Zisserman, 2004, Mouragnon *et al*, 2006). Le deuxième ensemble d'approches s'appuie sur l'estimation du mouvement apparent dans les images ou flux optique pour déterminer l'information tridimensionnelle. En faisant l'hypothèse que l'intensité des objets est conservée au cours du changement de point de vue, le calcul du flux optique peut servir à l'estimation de la transformation géométrique des images (Rodriguez *et al*, 2009).

2.3.3 Reconstruction 3D et modélisation 3D

2.3.3.1 La reconstruction 3D

La reconstruction 3D des objets consiste à déterminer les positions dans l'espace des primitives de la scène à partir des images de cette dernière. Dans les paragraphes précédents portant sur le recalage en stéréovision et le recalage multi-images, nous avons abordé la notion de la reconstruction 3D puisque généralement le calcul de la transformation géométrique et l'estimation de la position des primitives dans l'espace sont combinés. En stéréovision, on distingue 3 types de reconstruction : reconstruction projective (les points sont définis dans un repère projectif), reconstruction affine (les points sont définis dans un repère affine) et reconstruction euclidienne (les points sont définis dans un repère euclidien) (Hartley et Zisserman, 2004). Quant à la reconstruction multi-images, on y trouve tout particulièrement l'approche *Structure From Motion* (SFM) (Trucco et Verri, 1998, Hartley et Zisserman, 2004). Comme décrit dans le paragraphe portant sur le recalage multi-images, la position 3D des points est estimée essentiellement par l'utilisation d'algorithmes d'optimisation minimisant l'erreur de projection des points dans l'espace. En SFM, on distingue des approches s'appuyant sur la décomposition de la séquence en un ensemble de paires d'image stéréo (Hartley et

Zisserman, 2004), ou bien en un ensemble de triplets d'images impliquant ainsi l'utilisation du tenseur trifocal (Hartley et Zisserman, 2004) ou encore en bloc d'images (Mouragnon *et al*, 2006).

Ayant obtenu des points en 3D dans la reconstruction 3D, la modélisation 3D permet de visualiser la forme de l'objet reconstruit à partir de ces points. Les approches de modélisation 3D sont abordées dans les paragraphes qui suivent.

2.3.3.2 *La modélisation 3D*

Disposant des informations tridimensionnelles issues de l'étape de la reconstruction 3D, l'étape de modélisation permet de visualiser la géométrie 3D des objets, d'où l'appellation « modélisation géométrique 3D ».

Ce sont les primitives géométriques utilisées qui différencient une approche d'une autre d'où l'existence de cinq familles de modélisation : modélisation filaire, modélisation surfacique, modélisation volumétrique, modélisation hybride (surfacique et volumétrique) et modélisation paramétrique.

2.3.3.3 *Modélisation filaire (ou par fils de fer)*

Comme son nom l'indique, la modélisation par fil de fer s'appuie sur les sommets, les lignes et les arêtes pour la représentation des objets. En fait, dans ce cas de modélisation, les objets sont décrits par des segments de droite ou arête dont l'assemblage donne l'ossature de l'objet. Les surfaces des objets sont alors représentées et limitées par des contours formés par ces arêtes.

C'est une méthode simple et utilisée par la plupart des logiciels de modélisation 3D. Comparativement à d'autres méthodes telle que la modélisation surfacique, elle est connue pour être la méthode la moins complexe et la plus rapide en temps de calcul, notamment pour manipuler, modifier ou stocker les formes géométriques. Par contre, elle peut être ambiguë du fait que deux observateurs du même modèle peuvent avoir des interprétations différentes puisqu'elle ne permet pas la distinction entre des facettes pleines (représentant la surface de l'objet) et des facettes vides qui ne correspondent pas à la surface de l'objet (Ramos, 2003).

2.3.3.4 *Modélisation surfacique*

La modélisation surfacique décrit les objets en fonction de leurs surfaces (Pouliot *et al*, 2006; Ramos, 2003). C'est une approche très répandue ayant plusieurs avantages dont en particulier la description exacte de l'enveloppe de l'objet et la bonne visualisation de la forme. Trois approches caractérisent la modélisation surfacique : la modélisation polygonale, la B-Rep (« boundary representation » en anglais) et les NURBS (« Non Uniform Rational Basic Spline » en anglais).

- La modélisation polygonale se sert de polygones pour représenter les objets. L'exploitation de polygones dans cette approche est avantageuse : la description est simple, les calculs sont rapides et la modélisation d'objets complexes est facilitée. En effet toute surface peut être représentée par un polygone (surface plane) ou par la collation d'un ensemble de polygones constituant un maillage. La représentation polygonale est donc plus adaptée aux surfaces planes. En effet, plus la forme de l'objet est complexe, plus le nombre de facettes représentant les surfaces augmente, induisant un temps de calcul et une taille de stockage plus importants (Koussa, 2011). Parmi les opérations de modélisation polygonale, on peut citer : l'extrusion, la coupe, la soudure et la révolution (cf. Tableau 1). Ces approches sont utilisées tout particulièrement en infographie 3D.

Tableau 1¹². Principe des approches de modélisation polygonale 3D

Approches	Description
Extrusion	consiste à surélever une ou plusieurs faces (adjacentes ou non) ou un profil 2D le long d'une trajectoire et à créer les faces venant combler le vide occasionné par le déplacement de l'élément de départ
Coupe	consiste à créer des arêtes sur un maillage, ainsi que les points correspondant aux intersections des arêtes déjà existantes et celles nouvellement créées
Soudure	consiste à souder les sommets entre eux, pour simplifier un maillage, boucher des trous ou faire converger des arêtes
Révolution	consiste à faire tourner un profil 2D autour d'un axe 3D : on obtient ainsi un volume de révolution.

- La B-Rep ou « Modélisation par frontière surfacique ou par bord (Ramos, 2003)» représente les objets par leurs limites et les décrit de manière explicite en termes de nœud, arêtes et faces. Elle présente un grand avantage puisqu'elle inclut à la fois des informations de géométrie et de topologie. La géométrie décrit la forme et la position de chaque nœud, arête et face. La topologie exprime les relations de connexion entre ces trois éléments descriptifs de l'objet. A chaque primitive correspond ainsi deux attributs, soit sa géométrie et sa topologie. En plus, la modélisation B-Rep offre la possibilité de calculer le volume des objets. Néanmoins elle nécessite un volume de stockage très important.

- Les NURBS correspondent à l'interpolation d'un ensemble de courbes à partir de différents points de contrôle. Il s'agit de la création d'un maillage permettant de subdiviser la surface d'un objet en un ensemble de sous-surfaces dépendant de la complexité de la courbe (Koussa, 2011). Ces dernières sont définies par

¹² http://fr.wikipedia.org/wiki/Mod%C3%A9lisation_3D

NURBS et sont faciles à modifier via la manipulation de points de contrôle. Ce type de modélisation est très précis car les NURBS s'appuient sur une formulation mathématique des objets à représenter.

2.3.3.5 Modélisation volumétrique

Contrairement à la modélisation surfacique, la modélisation volumétrique reconstruit les objets en utilisant des volumes élémentaires spécifiques (ex. voxels). Cinq méthodes exploitent ce type de modélisation : CSG (« Constructive Solid Geometry » en anglais), RBSO (« Regularized Boolean Set Operation » en anglais), SPR (« Spatial Partitioning Representation » en anglais), par Octree et par éléments tétraédriques.

- La modélisation **CSG** est une approche qui s'appuie sur le rassemblement d'objets élémentaires ou solides pour approcher au mieux la forme 3D de l'objet en question. Les primitives utilisées dans la modélisation CSG correspondent à des solides de forme connue tels que des sphères et des cubes, reliés par des opérations booléennes (intersection, union et différence). Il est facile de calculer la surface de l'objet reconstruit étant donné qu'il existe des formules de calcul de surface associées à chaque primitive élémentaire. Cette approche rencontre des difficultés lors de la construction d'objets naturels car les primitives CSG et les opérations booléennes sont parfois insuffisantes pour décrire la forme exacte de l'objet réel.

- La modélisation **RBSO** repose sur les mêmes principes que la modélisation CSG en combinant des objets simples afin de reconstruire un objet complexe. Par contre, elle s'appuie sur des opérations booléennes (union, intersection et différence) dites régularisées qui ne peuvent générer que des objets solides, alors qu'un opérateur booléen ordinaire peut engendrer des objets non solides tels que surfaces, points, etc.

- La modélisation **SPR** ne rassemble pas les objets élémentaires pour reconstruire un objet complexe mais elle décompose le volume de ce dernier en un ensemble de solides élémentaires (ex. voxels) adjacents, qui ne se chevauchent pas. La modélisation SPR permet de calculer le volume de l'objet et de distinguer son intérieur, de son extérieur (Koussa, 2011 et Ramos 2003).

- La modélisation **par octree** est une variante de la modélisation SPR puisqu'elle s'appuie sur la subdivision récursive de l'espace en voxels de tailles variables tout en utilisant une structure d'arbre. La représentation en voxels offre une flexibilité de reconstruction mais demande un espace de stockage important pour des scènes volumineuses ou complexes.

- La modélisation **par éléments tétraédriques** décompose l'espace en plusieurs éléments tétraédriques. L'utilisation de ces éléments est très avantageuse par rapport aux approches qui utilisent les voxels. En effet, les éléments tétraédriques permettent de lisser les surfaces et donc d'éviter une

représentation « en forme d'escalier ». Ils fournissent ainsi une représentation plus réaliste mais avec un grand nombre d'éléments (Follet, 2003).

2.3.3.6 Modélisation hybride et modélisation paramétrique

Dans le but de combiner les avantages et de réduire les limitations des méthodes précédentes, la modélisation surfacique et la modélisation volumétrique ont été combinées. L'exemple le plus répandu d'une telle méthode hybride est celui de la combinaison des modélisations B-Rep et CSG. La méthode résultante est une approche très avantageuse par rapport à la forme de l'objet. Celle-ci est plus précise et proche de la forme réelle de l'objet. Par contre, cette approche nécessite la maintenance de la cohérence entre les primitives combinées issues de chaque approche et un temps de calcul important. Quant à la modélisation paramétrique, elle s'appuie sur la caractérisation de la géométrie de l'objet par un ensemble de paramètres tels que la longueur, la largeur, les angles, etc.

Dans le présent travail, notre approche prend fin à l'étape de la reconstruction 3D où on produit des informations tridimensionnelles. Ces dernières correspondent à des coordonnées tridimensionnelles de points qui pourront être par la suite exploitées afin de modéliser la scène et la visualiser. C'est une étape cruciale lors de l'observation de la scène sous-marine dans le contexte global de l'évaluation de l'impact des changements climatiques sur le milieu. En fait, la visualisation de la scène permet de simplifier la prise de décision quant à l'état de la scène et aussi permet de déduire plus facilement le changement spatio-temporel qui s'est passé. En conséquence, introduire cette partie sur la modélisation 3D est important afin de comprendre comment les informations 3D qui seront produites dans le présent travail vont être exploitées.

2.4 La reconstruction 3D d'une scène sous-marine à partir d'images acoustiques

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 leur environnement. La problématique de reconstruction 3D du milieu sous-marin est répandue aussi 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 ou leur identité en s'aidant de la reconstruction de leur forme.

Ces applications se réalisent souvent en utilisant conjointement plusieurs sources d'informations. En fait, la majorité des travaux de reconstruction 3D combinent les données de sonars imageurs, des cartes issues de modèles numériques de terrain mais 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, on présente quelques travaux de reconstruction 3D mono et multi-capteurs.

2.4.1 Reconstruction 3D multi-capteurs

La reconstruction 3D à partir d'images sonar multi-vues 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 » en anglais). 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 » en anglais) (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.

2.4.2 Reconstruction 3D mono-capteurs

Negahdari et Sekkati (2007) proposent une approche de reconstruction 3D mono-capteur. 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 permettent 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 flux optique pour l'extraction et le suivi de points d'intérêt. 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 multi-vues (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 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 similaire à 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 ; Sun *et al*, 2008).

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 multi-vues 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).

2.5 Le cadre conceptuel proposé

Au travers de l'élaboration d'un cadre conceptuel global, 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 complète en fournissant une représentation réelle de la scène, précise et fiable.

Les différents composants de ce cadre représentent les concepts-clés nécessaires pour effectuer cette reconstruction de la scène observée à partir d'images issues de caméras acoustiques. Ces concepts s'inspirent de l'ensemble des approches recensées dans la revue de littérature. Ils mettent en œuvre des solutions permettant de reconstruire l'environnement marin en tenant compte des problématiques inhérentes à ce milieu et à ces capteurs, comme cela a été détaillé dans le premier chapitre (i.e. géométrie particulière de formation des images, nature bruitée et inhomogène des images, environnement incertain et évolutif).

2.5.1 Architecture du cadre conceptuel

Suite à l'étude des différentes approches proposées en vision numérique et imagerie sonar et des caractéristiques des caméras acoustiques, nous avons procédé à l'identification des différents composants devant intervenir dans la reconstruction 3D 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. 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 6. Celle-ci propose également une représentation de l'architecture adoptée.

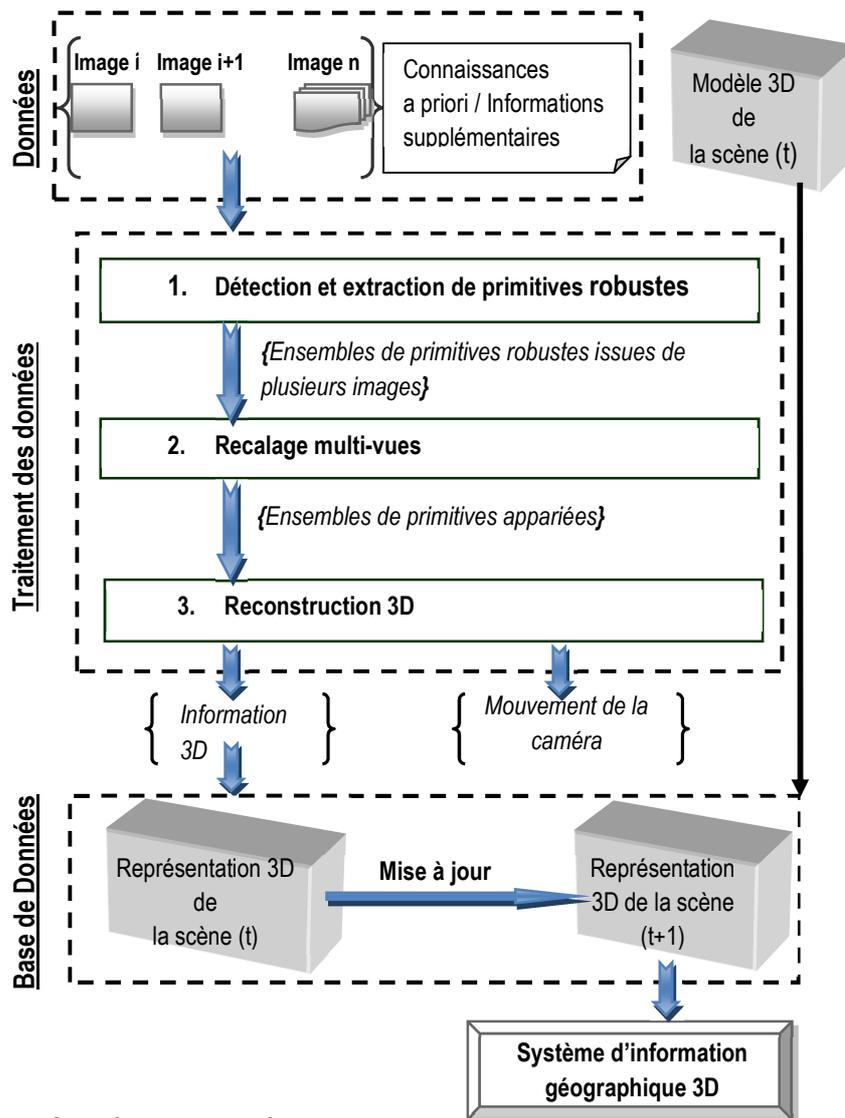


Figure 6. Architecture du cadre conceptuel

2.5.2 Description générale de l'architecture proposée

Nous proposons une architecture séquentielle qui, à partir d'un ensemble de séquences d'images et de connaissances *a priori*, permet de calculer des caractéristiques tridimensionnelles intervenant dans la reconstruction 3D de la scène observée et puis de mettre à jour la représentation courante avec ces nouvelles caractéristiques. Le choix d'une telle architecture repose sur le contexte d'acquisition qui aboutit à la production de séquences d'images. Le traitement et l'extraction d'informations depuis les images sont donc réalisés au fur et à mesure de leur acquisition.

L'architecture proposée est constituée principalement de trois 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. Ces trois modules sont les suivants :

Le premier module « Détection et extraction de primitives » regroupe l'ensemble des prétraitements à appliquer aux images acquises avant de pouvoir procéder à leur appariement. Il s'agit d'algorithmes qui permettent d'extraire un ensemble de primitives représentant les points saillants de la scène. Le résultat fourni par ce premier module consiste en des ensembles de primitives robustes issues de plusieurs images.

Le deuxième module effectue le recalage multi-vues des primitives. C'est une étape délicate étant donné qu'elle peut être affectée par le bruit des images. A l'issue de ce deuxième module, les primitives extraites de chaque image se trouvent appariées.

Le troisième module réalise la reconstruction 3D de la scène à partir des primitives 2D recalées. À ce niveau de l'architecture, l'attribut « élévation » des primitives est déterminé. Les informations tridimensionnelles calculées permettent de mettre en œuvre un modèle 3D partiel de la scène à un temps (t+1) qui servira à la mise à jour du modèle 3D complet construit au temps (t), pour obtenir le modèle 3D complet de la scène à l'instant (t+1).

Une description plus détaillée de chacun de ces modules est présentée dans les paragraphes suivants.

2.5.2.1 Module 1 : Détection et extraction de primitives robustes

L'image acoustique contient une multitude de primitives de différentes natures. La première étape intervenant dans l'architecture du cadre conceptuel 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 et calcul de la normale à ces régions

Lorsque l'on dispose de connaissances *a priori* sur les caractéristiques des images acoustiques et sur les problématiques liées au contexte du milieu, l'exploitation des primitives ponctuelles s'avèrent être une solution adéquate. En effet, ces primitives sont non seulement descriptives de la géométrie des objets mais aussi ils sont plus robustes aux bruits des images acoustiques, faiblement affectés par le changement de l'angle d'acquisition des images par rapport aux contours et aux régions. De plus, lors de l'étape de mise en correspondance, il peut être coûteux en termes de temps de calcul de comparer des régions ou des contours. D'autre part, les approches ponctuelles robustes réduisent l'impact du bruit dans les images et sont faiblement

affectées par le changement de l'angle de vue de la caméra et par les perturbations du milieu. En conséquence, 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 légèrement différents et surtout ayant la même position géographique dans la scène. De plus, la manipulation de contours ou des régions lors de leur extraction puis de leur mise en correspondance (en supportant des déformations éventuellement importantes) est très coûteuse en termes de temps de calcul ce qui constitue un inconvénient important.

2.5.2.2 Module 2 : Recalage multi-vues

Composant fondamental dans l'architecture du cadre conceptuel, le recalage est utilisé pour appairer les primitives extraites des images acquises selon différents points de vue (cf. Figure 7). Les techniques d'appariement de primitives peuvent être divisées en deux catégories (Nillius et Eklundh, 2002). La première catégorie se base sur la correspondance par bloc ou fenêtre (« block matching » en anglais) et la deuxième catégorie se base sur les caractéristiques des primitives. 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 rotation d'où la nécessité de prendre en considération les impacts de la rotation et l'orientation des fenêtres de corrélation en fonction de la prise de vue. La sensibilité aux distorsions géométriques demeure tout de même. Quant à la deuxième catégorie, elle concerne la mise en correspondance des ensembles de points saillants. Cette catégorie est plus avantageuse particulièrement dans le contexte de déformations et de distorsions des images. Les approches de recalage par mise en correspondance reposent sur une étape de détection de points et la définition d'un critère de similarité. Ce dernier permet d'identifier les points identiques. Cette catégorie de recalage est utilisée, particulièrement, dans les travaux exploitant le mouvement de la caméra pour calculer la structure tridimensionnelle d'un objet (i.e. « structure from motion » en anglais) (Nister *et al*, 2003 et 2006), comme c'est le cas dans l'architecture proposée dans le présent travail.

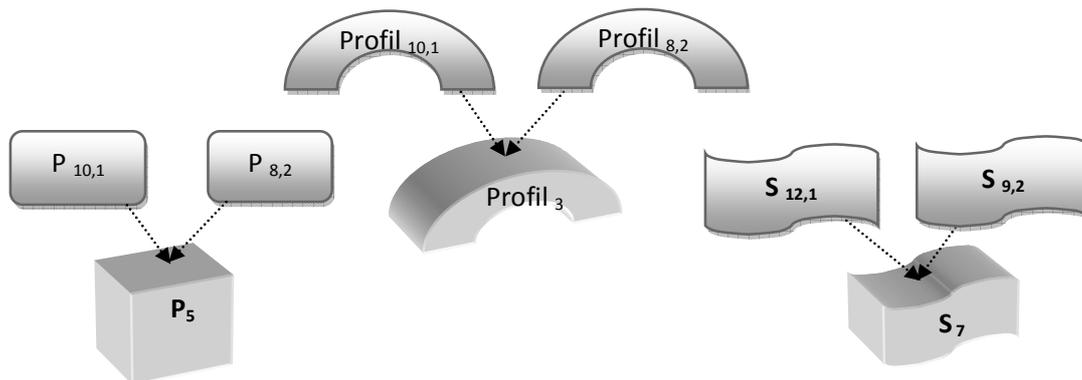


Figure 7 : Illustration de primitives appariées (P : Points, Profils, S : surfaces). Chaque primitive est référencée par un indice et le numéro de l'image de laquelle elle a été extraite (ex. $P_{10,1}$: Point 10 de l'ensemble de points extraits de l'image 1). Les primitives appariées sont référencées dans leur ensemble (P_5 est la paire de points numéro 5 dans l'ensemble des points présents dans l'environnement sous-marin).

2.5.2.3 Module 3 : Reconstruction 3D

L'obtention d'un ensemble de primitives appariées et robustes permet de procéder à l'étape de la reconstruction 3D. À ce niveau, on calcule les coordonnées 3D de ces primitives.

Dans la littérature, on a vu que les approches de reconstruction peuvent être classées en deux catégories : les approches stéréo et les approches multi-images. Dans le contexte d'une reconstruction 3D de scène sous-marine qui soit précise et fiable, et quand il s'agit de la manipulation d'images bruitées, une hypothèse s'impose : une approche multi-images s'avère plus pertinente qu'une approche par stéréovision, l'exploitation d'une paire d'images se révélant insuffisante en termes de précision au niveau du calcul de l'élévation de la scène. En effet, en appliquant le modèle des caméras acoustiques entre un ensemble de primitives correspondantes, la résolution du problème revient à faire face à un système d'équations représentant l'intersection de deux arcs de cercle dans l'espace. De plus chaque pixel dans l'image est associé à une région dans la scène où une multitude de points peuvent y être des candidats pour la reconstruction 3D. L'imprécision associée au choix précis du point de la scène qui a contribué à la formation de l'image bidimensionnelle peut induire des imperfections lors de la reconstruction de l'objet. Cette géométrie induit des imprécisions lors de la localisation des primitives et du calcul de leurs caractéristiques 3D.

2.6 Synopsis du chapitre 2

Ce chapitre a présenté les caractéristiques des caméras acoustiques, leurs applications et leur modèle géométrique. Le fait que les caméras acoustiques fournissent des séquences d'images multi-vues nous a conduits à aborder dans la revue de littérature en plus de l'étude des approches de reconstruction 3D en imagerie sonar, l'étude des approches stéréo et multi-images optiques. Celles-ci ont été présentées en second lieu dans ce chapitre. Le travail qui sera proposé va clairement s'inspirer des axes de recherche ainsi dégagés

par cette revue, revue qui a aussi montré la nécessité de prendre en compte les différences de nature entre images optiques et images acoustiques. 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. Deuxièmement, 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 alors que dans le cas des images optiques, c'est l'information de profondeur (i.e. distance caméra-scène) qui est l'inconnue à lever.

S'inspirant des démarches classiques de reconstruction tridimensionnelle à partir d'un ensemble d'images en y intégrant les différences fondamentales entre images optiques et acoustiques, nous avons conclu ce chapitre par la proposition d'un cadre conceptuel présentant la démarche méthodologique possible pour notre application.

La première étape de cette démarche méthodologique correspondant à l'extraction de primitives des images sera présentée au chapitre 3. Les primitives sélectionnées pour la reconstruction 3D, l'approche proposée, les données utilisées, les tests et les résultats seront aussi détaillés.

Chapitre 3. **Extraction de primitives à partir d'images acquises avec des caméras acoustiques**

3.1 Introduction

Ce chapitre décrit l'approche proposée afin d'extraire les primitives nécessaires pour la reconstruction des scènes observées dans une séquence d'images acoustiques. En effet, dans une démarche méthodologique de reconstruction tridimensionnelle, l'une des premières étapes importantes consiste à extraire les primitives descriptives de la géométrie de l'objet, sachant que les performances de cette étape de détection de primitives vont grandement conditionner les résultats finaux de reconstruction 3D.

Comme présenté dans le cadre conceptuel (cf. Paragraphe 2.5), il existe différents types de primitives telles que les points, les contours et les régions. Chacune d'entre elles caractérise certaines propriétés géométriques particulières de la scène observée et requiert une approche d'extraction appropriées. La première partie de ce chapitre est consacrée à la description de ces différentes primitives afin d'en déduire celles qui sont adéquates pour une reconstruction 3D d'objets à partir d'images acoustiques. Dans la deuxième partie de ce chapitre, l'approche d'extraction de primitives proposée est présentée ainsi qu'un ensemble d'expérimentations qui ont été menées afin d'en évaluer les performances et les résultats obtenus.

3.2 Sélection des primitives adaptées à la reconstruction 3D de scène à partir de séquences d'images acoustiques

3.2.1 Description des principaux types de primitive et de leurs approches d'extraction

Il est pratiquement impossible de passer en revue de manière exhaustive toutes les primitives et leurs approches d'extraction étant donné la popularité d'un tel sujet dans les domaines du traitement d'images et de la vision par ordinateur. Cette section cible principalement les primitives les plus répandues et utilisées ainsi que les approches développées afin de les extraire des images. En ordre décroissant de dimension dans l'image, ces primitives consistent en : 1) des régions 2) des contours, et 3) des points d'intérêt. Toutes ces primitives vont être décrites; néanmoins, les approches d'extraction de points d'intérêt vont être particulièrement développées dans cette première partie afin de mieux introduire dans la suite du chapitre, la contribution de l'approche proposée dans le cadre de cette recherche.

3.2.1.1 Les primitives de type « régions »

Les régions d'une image constituent l'ensemble des surfaces regroupant des pixels spatialement voisins présentant une certaine homogénéité relative à un critère de similarité prédéfini. Ce critère de similarité recouvre des réalités sémantiques de complexité très différentes (par exemple, simplement basé sur l'intensité des pixels ou bien reposant sur une analyse de texture). De ce fait, même si elle est associée à un pixel, cette mesure de similarité prend en compte un contexte spatial plus étendu.

Une segmentation par région permet d'agréger les pixels adjacents similaires selon un certain critère de similarité formant ainsi des régions distinctes (Sebari et He, 2007), contrairement à la segmentation par contours, qui détecte les transitions entre les régions de l'image. Afin d'extraire ces régions, plusieurs approches de segmentation par régions ont été développées, ces approches se distinguant essentiellement par le critère de similarité utilisé.

Parmi les méthodes de segmentation par régions recensées dans la littérature, certaines opèrent par un simple seuillage de l'histogramme de l'image (Gonzalez et Ozguner, 2000). D'autres procèdent selon une approche de regroupement (i.e. « clustering » en anglais). Il s'agit de partitionner les pixels de l'image en un ensemble de groupes (ou classes) par l'optimisation d'une fonction objective. Les pixels ayant des caractéristiques similaires sont groupés dans une même classe. La méthode des K-means est l'algorithme le plus utilisé dans le contexte de cette segmentation (Niemistö *et al*, 2007). Cet algorithme est très populaire car extrêmement rapide en pratique. En termes de performance, cet algorithme ne garantit pas un optimum global. La qualité de la solution dépend grandement des ensembles initiaux (Maître, 2003). Contrairement à d'autres méthodes hiérarchiques qui créent une structure en « arbre de régions ¹³ » pour décrire les groupements (Roux, 2006), la méthode des K-means ne crée qu'un seul niveau de régions. Elle prévoit initialement une attribution binaire des éléments à une région ou à un groupe. Selon Daniel (1998), les frontières qui séparent les différents groupes ont tendance à être vagues, ce qui diminue la précision du résultat de cet algorithme. Une version floue des K-means appelée algorithme des Fuzzy C-means (Pham et Prince, 1998) a été développée. Cet algorithme offre une répartition des groupes plus précise et plus flexible en termes de classification des pixels. Il souffre du même inconvénient que les K-means, le nombre de régions devant être fixé au préalable. Dans le cas des images acoustiques, la connaissance des trois classes principales d'information relatives à ces images à savoir écho, ombre et réverbération, aide très souvent à spécifier ce nombre de classes. En fait, ces trois classes diffèrent par l'intensité de la réflexion des ondes acoustiques par la scène. La classe 'écho' contient les pixels dont l'intensité est plus élevée par rapport aux autres pixels, indiquant ainsi la présence d'un fort réflecteur tel qu'un objet sur le fond marin ou une structure.

¹³ Résulte d'une hiérarchie de segmentations; l'image est représentée par un arbre de régions qui décrit l'image avec un certain niveau de détail.

La classe ombre regroupe quant à elle les pixels d'intensité plus faible qui se trouvent généralement à l'arrière des objets vis-à-vis de la direction d'insonification et qui correspondent à une absence de signal rétrodiffusé. Finalement, la classe réverbération caractérise les pixels ayant une intensité moyenne qui représentent en général le fond (Brahim *et al*, 2009).

Des méthodes de segmentation par régions ont été proposées utilisant des champs de Markov (Mignotte, 1998) et des modèles statistiques du bruit. Ces méthodes se distinguent des méthodes classiquement utilisées (citées dans le paragraphe précédent) par la prise en compte des interactions locales entre chaque pixel et ses pixels voisins pour définir les différentes régions de l'image. Ceci a permis d'obtenir des segmentations avec des régions homogènes avec une bonne localisation des frontières. Néanmoins, elles souffrent de deux inconvénients qui sont la complexité algorithmique et un temps de calcul important. En fait, l'utilisation du modèle de Markov se base sur la représentation de l'image par un champ de Markov ce qui nécessite le recours à des méthodes itératives telles que le modèle de Gibbs pour estimer la loi du champ. De plus, ces méthodes s'appuient sur le calcul de solutions bayésiennes en ayant recours à des estimateurs et à la mise en œuvre d'algorithmes itératifs tels que le recuit simulé et ou l'algorithme ICM (« Iterated Conditional Modes » en anglais). Finalement, il est à noter que leur utilisation se heurte aussi au choix initial de leurs paramètres souvent obtenus par d'autres algorithmes de segmentation qui peuvent engendrer des charges de calcul supplémentaires (Alkama et Berkani, 2010). D'autres méthodes ont été proposées afin d'améliorer le temps de calcul et la performance de la segmentation markovienne telle une approche multi-résolution effectuant une discrétisation spatiale des observations à différents niveaux de résolution pour aboutir à une représentation pyramidale des données (Boukerroui, 2000), ou encore une approche markovienne hiérarchique multi-modèle avec un modèle markovien différencié selon l'échelle de représentation des données (Collet *et al*, 1998).

3.2.1.2 Les primitives de type « contours »

Les contours dans une image correspondent généralement à des changements brusques de propriétés géométriques et physiques dans la scène observée. En effet, ces propriétés étant par exemple, l'orientation d'une surface ou sa texture, leur changement se traduit par une discontinuité au sein de la fonction représentative de l'intensité d'une image. Plus explicitement, à tout pixel de coordonnées (x, y) dans une image I est associée une intensité $I(x, y)$. Un contour dans cette image correspond à un ensemble de pixels où sont localisées les très fortes variations de $I(x, y)$. Soit \vec{G} , le gradient de $I(x, y)$:

$$\vec{G} = \vec{\nabla} I = \begin{pmatrix} \frac{\partial I(x,y)}{\partial x} \\ \frac{\partial I(x,y)}{\partial y} \end{pmatrix} \quad (2)$$

On calcule pour chaque pixel de l'image :

1. le module du gradient de l'intensité:

$$\|\vec{G}\| = |\nabla I(x, y)| = \left[\left(\frac{\partial I(x, y)}{\partial x} \right)^2 + \left(\frac{\partial I(x, y)}{\partial y} \right)^2 \right]^{1/2} \quad (3)$$

2. la direction du gradient :

$$\vec{g} = \tan^{-1} \left(\frac{\frac{\partial I(x, y)}{\partial y}}{\frac{\partial I(x, y)}{\partial x}} \right) \quad (4)$$

Un contour est défini comme le lieu de forte variation du module du gradient (les maxima) (équation 3) dans la direction \vec{g} du gradient (équation 4) (Canny, 1986).

En partant de ces fonctions, plusieurs détecteurs de contours ont été développés. On retrouve parmi ceux-ci les opérateurs classiques (des opérateurs locaux) tels que le filtre de Sobel (Vincent et Folorunso, 2009) ou de Prewitt (Juneja et Sandho, 2009) ou le Laplacien (dérivée d'ordre 2 à deux dimensions de l'image) (Wang, 2007) qui sont souvent impliqués dans plusieurs techniques de détection de contours. Ils s'appuient sur une dérivation du premier ou du deuxième ordre de la fonction d'intensité des images, suivie par une recherche de maxima locaux (dans le cas du premier ordre) ou de passage par zéro (dans le cas du second ordre). Ces opérateurs sont simples et rapides. Toutefois, ils ne donnent pas souvent de bons résultats surtout quand il s'agit de la manipulation d'images réelles où les changements d'intensité sont rarement nets et abrupts. En effet, les contours détectés par ce type d'opérateurs ne sont pas très bien localisés et souvent épais (Maitre, 2003, Chapitre 2).

Pour pallier ces limitations, des filtres ont été conçus dont les plus connus sont les filtres respectivement de Canny, Deriche, Shen et Castan (Laligant *et al*, 1998). Parmi ces filtres, le filtre de Canny (Canny, 1986) est le plus répandu et le plus robuste. Il se caractérise par : 1) une bonne détection, 2) une bonne localisation 3) la non-multiplicité des réponses, ce qui signifie la limitation du nombre de maxima locaux détectés en réponse à un seul contour. Le filtre de Deriche est une variante du filtre de Canny. En fait, la dérivée du filtre de Deriche est la solution exacte de l'équation de Canny. Les deux filtres présentent des avantages et des inconvénients dans leur utilisation pour la détection de contours. Le filtre de Canny est un filtre optimal largement utilisé que soit pour des filtrages 1D ou 2D. L'inconvénient de ce filtre est que la complexité du calcul croît linéairement avec le nombre de coefficients significatifs de la réponse impulsionnelle et donc avec la diminution de la résolution. Deriche a proposé une extension récursive du filtre de Canny conduisant à une complexité de calcul indépendante de la résolution. La résolution est, en fait, définie par un seul coefficient. L'inconvénient majeur du filtre de Deriche est que l'infinité de la réponse impulsionnelle impose pour des raisons de stabilité

quatre parcours de l'image (gauche vers droite, droite vers gauche, haut vers bas et bas vers haut) et donc un temps de traitement considérable. Cet inconvénient empêche l'utilisation du filtre de Deriche pour des signaux temps réel à échantillonnage temporel et favorise l'utilisation du filtre de Canny. Ceci explique pourquoi le filtre de Canny est largement utilisé par rapport au filtre de Deriche (Demigny *et al*, 2001).

Une autre catégorie d'opérateurs de détection de contours correspond aux opérateurs de morphologie mathématique. Ces opérateurs offrent des outils non linéaires de détection de contours tels le Gradient par dilatation, le Gradient par érosion, le Laplacien morphologique (qui résulte de la différence entre Gradient par dilatation et Gradient par érosion), etc. Le filtrage morphologique fait appel à deux opérateurs qui sont l'érosion et la dilatation (Coustya *et al*, 2013). D'autres opérateurs morphologiques peuvent être aussi construits en combinant les deux précédents. L'érosion est une opération qui vise à réduire la masse d'une région alors que la dilatation est l'opération contraire (i.e. augmentation de la masse). Pour ce faire, le filtre morphologique utilise une forme de référence dite élément structurant. C'est un masque dont la taille et la forme sont déterminées par l'utilisateur. En parcourant tous les pixels de l'image, l'intensité du pixel central de l'élément structurant est comparée à celles des pixels qui sont inclus dans l'élément structurant. Ce pixel central prendra ainsi soit la valeur minimale (lors d'une érosion) soit la valeur maximale (lors d'une dilatation) des pixels de l'élément structurant. En parcourant toute l'image, un ensemble de pixels qui représentent les régions ayant une variation d'intensité par rapport à leurs voisinages se forme. Le chaînage de ces pixels détectés qu'ils soient par érosion ou dilatation, correspond aux contours de l'image. Quand il s'agit d'image bruitée, on peut avoir un ensemble de segments de contours éparpillés dans l'image. C'est pour cette raison, que généralement les approches morphologiques sont utilisées pour détecter et supprimer le bruit impulsionnel (c'est-à-dire les points isolés issus de la présence d'un bruit dans l'image) (Wenbin, 2006).

3.2.1.3 Les primitives de type « points d'intérêt »

Selon Moravec (1977), un point d'intérêt est une région de l'image où l'intensité varie dans au moins deux directions et dont la dimension ne dépasse pas un pixel. Cette variation de l'intensité dans plusieurs directions permet de détecter des points supplémentaires par rapport à des points de contours où l'intensité varie seulement selon une seule direction (cf. Paragraphe 3.2.1.2). En effet, ces changements d'intensité caractérisent les coins, les jonctions et les points de forte variation de texture, de courbure et de bruit (ex. le bruit impulsionnel).

Depuis les années 70, plusieurs détecteurs de points d'intérêt ont été développés. On distingue deux catégories principales :

- 1) ceux qui se basent sur une première extraction de contours afin d'en déduire les points de forte courbure.

- 2) ceux qui se basent sur la fonction d'intensité de l'image pour en extraire les points de forte variation d'intensité.

Dans la suite, nous présentons les principaux travaux portant sur ces deux catégories de détecteurs de point. Dans la deuxième catégorie, nous allons présenter les travaux dans leur ordre chronologique afin de suivre leur amélioration au fil du temps.

3.2.1.3.1 Méthodes basées sur les contours

Le principe de ces méthodes correspond à la recherche des points de courbure maximale le long des contours extraits, des extrémités de contours et des points d'inflexion et/ou d'intersection entre deux contours (Rosten *et al*, 2010).

Plusieurs approches ont été développées dans ce contexte. Les premières approches ont déterminé les points de courbure en calculant les angles entre les extrémités de différents segments de contours (Rosenfeld et Weszka, 1975). Les coins correspondent selon cette démarche, aux points pour lesquels les valeurs d'angle forment des minima locaux. Le calcul des tangentes aux contours combiné au calcul de la longueur de ces contours permet aussi de retrouver les points isolés (Beus et Tiu, 1987). Les splines ont également été utilisées afin de déterminer les points de courbure (Medioni et Yasumoto, 1987). Par exemple, Medioni et Yasumoto (1987) approximent les contours extraits par des B-splines dont les fortes valeurs de ses dérivées secondes indiquent les points de haute courbure (Rosten *et al*, 2010).

Les coins peuvent être ainsi détectés en appliquant un filtrage sur les contours. Ils s'identifient à des points présentant une forte variation du gradient ou à des points dont la courbure a diminué rapidement vers le plus proche minima local (He et Yung, 2004). L'utilisation d'un filtrage multi-échelle représente une autre alternative pour la détection des coins. Ils correspondent alors aux points ayant une position stable et détectée à travers toutes les échelles (Lee *et al*, 1995). La méthode CSS (« Curvature Scale Space » en anglais) est une technique où l'échelle de filtrage est choisie en fonction de la longueur des contours. Elle définit les coins comme les points dont la courbure est très supérieure à celle du prochain point ayant une courbure maximale locale dans le contour parcouru (Rosten *et al*, 2010). Ces approches en particulier les méthodes CSS et B-splines sont de complexité algorithmique élevée impliquant ainsi un temps de calcul important. Toutefois, les points d'intérêt issus de ces approches sont couramment utilisées en reconnaissance d'objets car ils caractérisent la forme de ces objets; en effet, ces points ne se situent que sur les contours de ces objets (Merad, 2004).

3.2.1.3.2 Méthodes basées sur l'intensité des images

Ces méthodes de détection de points opèrent directement sur l'intensité des images sans passer par la détection des contours. Ce sont les méthodes les plus répandues et les plus développées (Deriche et Giraudon, 1991, Mikolajczyk et Schmid, 2004, Rosten *et al*, 2010). La majorité de ces approches effectue un calcul des dérivées du premier ou deuxième ordre puis un calcul de courbure en chaque pixel de l'image par rapport à son voisinage pour la détection de points d'intérêt.

La Figure 8 présente d'une façon chronologique un ensemble de détecteurs de points d'intérêt. Ainsi, le premier détecteur de point d'intérêt a été développé par Beaudet en 1976 (Beaudet, 1976). Son approche est très simple. Elle utilise la dérivée de second ordre de l'image suivant l'équation suivante :

$$DET = I_{xx}I_{yy} - I_{xy}^2 \quad (5)$$

où $I(x, y)$ est l'intensité de l'image, I_{xx} , I_{yy} et I_{xy} représentent ses dérivées seconde respectivement selon l'axe des x , l'axe des y et selon les deux axes x et y (Beaudet, 1978). En ne s'appuyant que sur le calcul de la dérivée seconde des images, ce détecteur est sensible au bruit dans les images. Moravec (1977) propose par la suite d'utiliser une estimation de la fonction d'autocorrélation f en chaque pixel, pour déterminer la variation de l'intensité dans les huit directions d'un voisinage W carré (de dimension 3×3 , 5×5 , etc.). Son approche permet de détecter les coins des objets dans l'image en y associant les maxima locaux de la fonction f en chaque pixel. f est représentée par l'équation (6):

$$f(x, y, \delta x, \delta y) = \sum_W (I(x_k, y_k) - I(x_k + \delta x, y_k + \delta y))^2 \quad (6)$$

La fonction $f(x, y, \delta x, \delta y)$ correspond à la somme au carré de la différence entre les intensités de deux sous-images de I de taille W , l'une centrée en (x, y) et l'autre décalée de $(\delta x, \delta y)$ car centrée en $(x + \delta x, y + \delta y)$. $I(x_k, y_k)$ représente l'intensité des pixels de la sous-image centrée sur (x, y) dans le voisinage (W).

L'approche de Moravec est une approche simple. Néanmoins elle souffre d'un certain nombre de limitations. Celles-ci concernent sa non invariance par rotation due au voisinage rectangulaire utilisé. Dans ce contexte, la réponse du détecteur est anisotropique et très sensible au bruit. Il peut ainsi détecter et identifier des pixels isolés issus du bruit comme points d'intérêt.

Plusieurs algorithmes visant à améliorer le principe des approches proposées par Beaudet (1976) et Moravec (1977) ont été développés. On peut citer notamment le détecteur de Harris et Stephens (1988) qui est basé sur le calcul de la fonction d'autocorrélation et l'utilisation d'un filtre gaussien au lieu d'un filtre binaire (cas du

détecteur de Moravec) ainsi que le développement de Taylor de la fonction d'intensité pour la détection du changement d'intensité dans les différentes directions du voisinage d'un pixel. Le détecteur de Harris est la version améliorée du détecteur de Harris et Stephens. L'amélioration réside dans l'utilisation d'une fonction de lissage gaussien pour le calcul des dérivées de l'image (Harris et Stephens, 1988). Le détecteur de Harris reste parmi les algorithmes les plus utilisés en détection de points d'intérêt compte tenu de la qualité de ses résultats et la facilité de son implémentation.

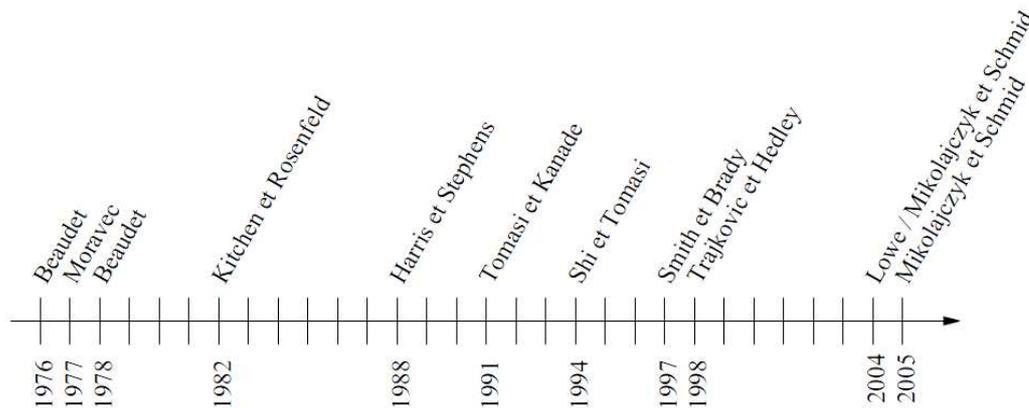


Figure 8 : Frise de détecteurs de points d'intérêt (Parisot, 2009)

L'inconvénient principal des détecteurs de points cités dans les paragraphes précédents est leur instabilité face au bruit dans les images. Ceci est dû au fait qu'ils se basent sur un calcul local des dérivées.

Smith et Brady (1997) ont proposé un détecteur de coins et de points de contours, appelé SUZAN (« Smallest Univalued Segment Assimilating Nucleus » en anglais), qui ne nécessite pas de calcul des dérivées de l'image pour assurer la détection des points d'intérêt, ce qui le rend plus robuste au bruit. Il détecte les points d'intérêt en utilisant un masque circulaire ayant un centre dit « Noyau ». SUZAN parcourt l'image en comparant le niveau de gris de chaque pixel contenu dans le masque avec celui du noyau. Une zone dite « USAN » (« Univalued Segment Assimilating Nucleus » en anglais) est alors définie comme la zone de même niveau de gris que le noyau. À partir de la taille, du barycentre et des moments de second ordre de «USAN», on peut déduire à la fois les contours et les coins de l'image (Smith et Brady, 1997). Cette méthode est réputée pour être moins sensible au bruit et donne des résultats satisfaisants quant au nombre de points d'intérêt détectés.

D'autres détecteurs tels que SIFT (« Scale Invariant Feature Transform » en anglais), algorithme proposé par Lowe (2004), ont connu un grand succès depuis ces dernières années. SIFT détecte les points de forte variation d'intensité (les extrema) dans une image analysée selon différentes échelles dans une pyramide gaussienne. Un pixel est retenu comme étant un maximum local (respectivement minimum local) s'il a une valeur maximale (respectivement minimale) parmi ses 8 voisins directs et ses 18 voisins des niveaux

supérieurs et inférieurs de la pyramide. Une description plus détaillée de l'algorithme SIFT est présentée dans (Lowe, 2004). La sélection des points à travers une pyramide gaussienne des différentes échelles, produit des points invariants aux changements d'échelle et de rotation (Lowe, 2004). À chaque point d'intérêt sélectionné est associé un descripteur. Ce descripteur contient des propriétés caractéristiques de chaque point telles que l'orientation, qui vont servir par la suite à l'appariement de points issus d'images différentes.

L'algorithme SIFT est connu pour sa robustesse; néanmoins il présente aussi des inconvénients tout particulièrement au niveau de la vitesse de calcul (Bauer *et al*, 2007). Afin d'y remédier, plusieurs autres algorithmes partant du même principe que celui de SIFT ont été développés tels que SURF (« Speeded Up Robust Features » en anglais) (Bay *et al*, 2006), FAST (« Features from Accelerated Segment Test » en anglais) (Rosten *et Drummond*, 2005), les Ferns (Ozuysal *et al*, 2007, Oshin *et al*, 2008), etc.

Il y a de nombreux détecteurs de points d'intérêt recensés dans la littérature (Busque, 2001, Grand-brochier, 2009); ceux cités dans les paragraphes précédents sont les plus connus. Ces détecteurs ont le même objectif à savoir la détection de points d'intérêt dans des images mais ils diffèrent tout particulièrement au niveau du temps de calcul, de la répétabilité vis-à-vis des changements de points de vue, du changement d'échelle, du bruit et des variations d'illumination.

Ces détecteurs de points détectent tout particulièrement les pixels de l'image où une grande variation d'intensité dans plusieurs directions est observée. En conséquence, les pixels ou points détectés peuvent caractériser tout aussi bien les coins des objets que la présence de bruit dans les images. Ceci peut constituer un inconvénient pour certaines applications dont notamment la reconstruction 3D des objets. En effet, une telle reconstruction ne peut s'opérer correctement qu'à partir de points d'intérêt représentant effectivement certaines caractéristiques géométriques des objets à reconstruire.

3.2.2 Choix de primitives pour la reconstruction 3D

Suite au recensement dans la littérature des principales primitives pouvant être extraites des images et en tenant compte des caractéristiques des images acoustiques, notamment leur niveau de bruit, les informations ponctuelles se sont révélées comme le type de primitive le plus adéquat pour la mise en œuvre de l'approche de reconstruction 3D visée dans le cadre de cette recherche. Les paragraphes suivants présentent les raisons pour lesquelles les informations ponctuelles ont été choisies comme primitives et indiquent pourquoi les détecteurs classiques (Harris, SIFT, ...) ne sont pas suffisants pour répondre à notre problème de reconstruction 3D.

3.2.2.1 Comparaison du potentiel des différents types de primitive pour la caractérisation des objets en imagerie acoustique

En se basant sur nos connaissances relatives aux caractéristiques des images acoustiques et aux primitives les plus utilisées pour la caractérisation des objets (les régions, les contours et les points), les remarques suivantes peuvent être faites :

- Les contours sont parmi les primitives les plus utilisées pour caractériser la forme des objets en imagerie (acoustiques et optiques). En effet, dans le cas des images sonar, les contours caractérisent généralement les frontières entre des zones de rétrodiffusion différentes telles que les arêtes des objets et les débuts et fins des zones d'ombre. Toutefois, le nombre et la forme des contours extraits ne sont pas invariants au sein d'une séquence d'images sonar à cause du bruit et de la variation d'angle d'acquisition des images. De plus, un même contour observé sur plusieurs images peut provenir de lieux géométriques différents (ex. les contours des ombres) ainsi que de fausses détections (Brahim *et al* (b), 2011, Brahim *et al*, 2010). Il est rare aussi de parvenir à détecter complètement un contour fermé caractérisant les différentes surfaces de l'objet. Par ailleurs, il s'avère difficile d'exploiter uniquement des segments de contours ouverts pour déduire et extraire la forme caractéristique des éléments de la scène. En revanche, ces segments de contours sont clairement une source d'information importante exploitable pour déduire la forme des objets.

- Les régions caractérisent les surfaces de l'objet, surfaces ayant des propriétés similaires. La problématique de l'utilisation des régions dans la reconstruction 3D des images sonar réside au niveau du temps de calcul nécessaire pour la manipulation des surfaces dans les différentes étapes de la reconstruction (ex. extraction des régions caractéristiques, appariement, etc.). De plus, les exemples recensés dans la littérature portant sur la reconstruction 3D surfacique en imagerie acoustique consistent principalement en l'application d'approches de type « Shape From Shading » et montrent que ce type de reconstruction nécessite l'exploitation d'informations supplémentaires pour aboutir à une reconstruction précise (Johnson et Hebert, 1996, Dura *et al*, 2004). En fait, le « Shape From Shading » est une approche qui à partir d'une seule image, cherche à estimer la variation des pentes locales des surfaces de la scène, en se basant sur le fait que l'intensité acoustique réverbérée est modulée par l'angle d'insonification du sonar par rapport à cette surface (Dura *et al*, 2004). Selon Johnson et Hebert (1996), c'est un problème complexe à résoudre car il nécessite l'intégration de plusieurs connaissances *a priori* pour contraindre l'espace de calcul des variations d'inclinaison (par exemple, l'altitude du sonar ou encore des mesures bathymétriques grossières). Ainsi, généralement, des informations ponctuelles constituant des points d'ancrage sont exploitées en combinaison avec les informations d'inclinaison pour aboutir à une reconstruction 3D précise (Dura *et al*, 2004).

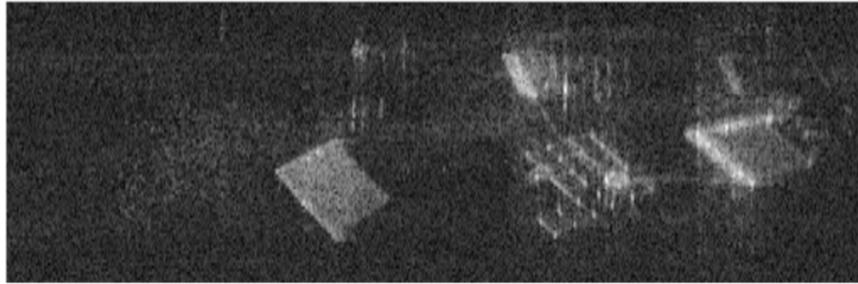
- Dans le contexte de l'imagerie acoustique, exploiter des informations ponctuelles pour caractériser la géométrie des objets semble donc plus pertinent qu'exploiter contours et régions. En fait, ces informations

ponctuelles représentent les caractéristiques géométriques les plus élémentaires d'un objet en reflétant principalement ses points anguleux tels que par exemple, les huit coins d'une boîte rectangulaire dans une scène. Ces points représentent partiellement mais fidèlement la forme de celui-ci en marquant les points de rupture dans sa géométrie. En effet, hors cas d'occlusion, chaque point de rupture se retrouve matérialisé de manière unique par un point (ou pixel) d'intérêt, sur chaque image successive. Ainsi, même si la localisation de ces points d'intérêt dans les images d'une séquence dépend logiquement de la position de la caméra dans la scène, l'ensemble des points d'intérêt associés à un point de rupture et détectés sur plusieurs images, correspond toujours à une même coordonnée géographique (celle du point saillant) dans la scène. Cette association sur plusieurs images d'une même séquence, entre point d'intérêt (dans l'image) et point saillant (dans la scène) est un avantage crucial de ce type de primitives.

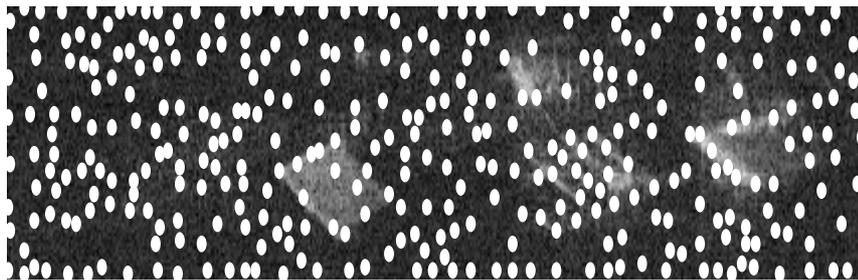
3.2.2.2 Illustration du comportement de détecteurs standards de points d'intérêt pour la construction de primitives ponctuelles

Bien qu'il existe plusieurs détecteurs de points connus pour leur robustesse et leur performance en imagerie optique, il n'est pas évident d'avoir les mêmes performances en imagerie acoustique. En effet, les détecteurs tels que SIFT, SUZAN et Harris détectent tous les points ayant une grande variation d'intensité par rapport à leur voisinage. Toutefois, les images acoustiques sont des images caractérisées par la présence d'un bruit ayant une structure granulaire. Ce bruit correspond à un bruit impulsionnel représenté par des pixels brillants et sombres éparpillés d'une façon aléatoire sur les images. Il dégrade la qualité des images et constitue un inconvénient majeur pour la détection et l'extraction de points anguleux dans les images acoustiques. En fait, les détecteurs de points traditionnellement utilisés en imagerie optique vont identifier les pixels résultant du bruit comme des points de forte variation d'intensité donc des points d'intérêt.

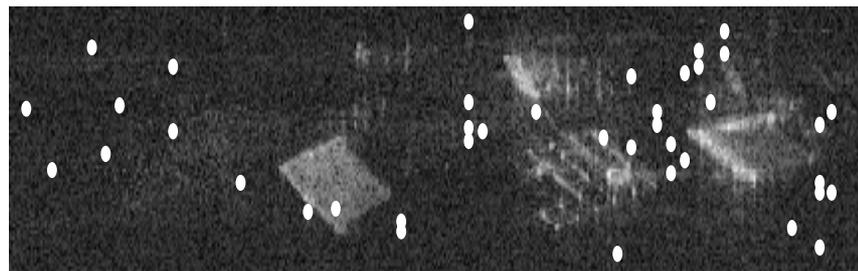
Afin de comparer la qualité des résultats de ces détecteurs et d'évaluer leur performance dans la détection de points caractéristiques de la géométrie des objets à partir d'images acoustiques, nous avons réalisé plusieurs tests en modifiant d'un test à l'autre les paramètres des détecteurs. Les paramètres modifiés sont en particulier les seuils spécifiques pour la sélection des maxima parmi les points d'intérêt détectés. Ces seuils sont différents étant donné que chacun de ces détecteurs s'appuie sur un principe différent (cf. Paragraphe précédent). La Figure 9 représente des exemples de résultat d'application respectivement du détecteur de Harris (Figure 9.b), de l'algorithme SUZAN (Figure 9.c) et du SIFT (Figure 9.d) sur une image acoustique DIDSON (Figure 9.a). Les résultats proposés sont ceux qu'on a jugé les meilleurs.



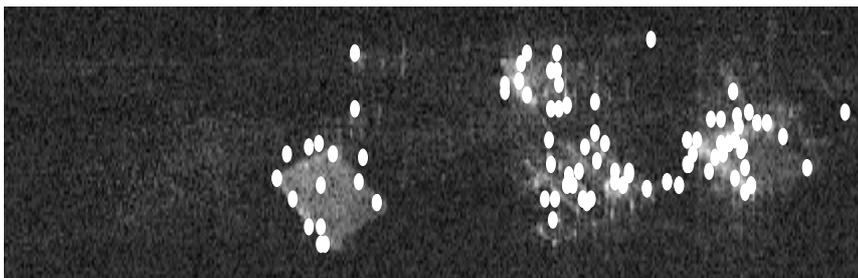
(a)



(b)



(c)



(d)

Figure 9 : Exemple de détection de point d'intérêt : (a) Image originale DIDSON d'objets posés au fond d'un bassin; (b) Points d'intérêt détectés par le détecteur de Harris, (c) Points d'intérêt détectés par le détecteur SUZAN; (d) Points d'intérêt détectés par le détecteur SIFT

La Figure 9.a représente une image DIDSON acquise dans une piscine au fond de laquelle se trouvent des objets de différentes tailles et formes. Comme indiqués sur la Figure 9.b, c et d, les points détectés par les détecteurs de Harris, SUZAN et SIFT sont différents au niveau de leur nombre et de leur localisation. Une simple comparaison visuelle permet de déduire que les résultats de l'algorithme SIFT sont nettement meilleurs

que ceux obtenus avec le détecteur de Harris et SUZAN, pour ce qui concerne la caractérisation des objets dans la scène. Toutefois, on peut constater que globalement les points détectés par ces méthodes ne représentent pas seulement la géométrie des objets mais aussi des pixels provenant du bruit ou d'artefacts de propagation.

Il s'avère ainsi nécessaire de proposer une approche spécifique aux images acoustiques permettant de détecter et d'extraire des points qui représentent la géométrie des objets et qui peuvent être retrouvés dans des images acquises selon différents points de vue.

3.3 Approche proposée

Cette partie présente le problème de la caractérisation robuste de la forme des objets représentés dans une séquence d'images multi-vues acquises avec des caméras acoustiques en procédant à l'extraction des points caractéristiques de leur géométrie. Pour extraire de manière robuste de tels points représentatifs de la géométrie des objets, la solution que nous proposons s'appuie non seulement sur la redondance d'observation de la scène sur plusieurs images mais aussi sur la localisation des primitives recherchées dans l'image. En fait, contrairement aux détecteurs habituellement utilisés en imagerie optique dont notamment SIFT et SUZAN où la détection des points résulte d'une analyse de l'image afin d'en extraire une information locale caractérisant l'intensité d'un pixel par rapport à son voisinage, l'approche que l'on propose extrait les points par le biais d'une analyse globale des points de contours des objets de la scène.

L'approche proposée s'articule autour des trois grandes étapes suivantes dont les deux premières s'exécutent au sein d'une décomposition multi-échelle:

- 1) Extraction des contours.
- 2) Élimination de faux contours par analyse multi-images. Les faux contours sont les contours qui sont issus de la présence du bruit dans les images et ne représentent pas la géométrie des objets.
- 3) Extraction des points anguleux par détection de points de haute courbure au sein des contours. Les points anguleux sont les coins des objets qui caractérisent leur forme.

Le recours à l'extraction de contours (étape 1) a pour but d'identifier les emplacements potentiels des points représentatifs de la géométrie des objets. Une telle démarche permet donc de restreindre l'espace de recherche des primitives ponctuelles. La deuxième étape portant sur l'élimination des faux contours, est ensuite cruciale pour minimiser l'effet du bruit présent dans les images. Elle s'appuie sur l'appariement des

contours issus de deux images consécutives permettant en conséquence d'éliminer ceux provenant du bruit (c'est-à-dire les contours qui ne se retrouvent pas d'une image à une autre).

Par ailleurs, l'image est analysée à travers différentes échelles d'une pyramide gaussienne. Cette dernière correspond à un filtrage de l'image par une gaussienne à différentes résolutions. Ainsi, pour chaque échelle, les étapes 1 et 2 précédentes sont appliquées aux images jusqu'à l'élimination de tous les faux contours.

L'exploitation de la cadence d'acquisition des caméras acoustiques et de la redondance d'observation résultante est un élément clé de cette approche, puisqu'afin d'éliminer les faux contours, on s'appuie sur des paires d'images successives. Ainsi l'approche proposée tire tout particulièrement profit des caractéristiques intrinsèques des caméras acoustiques rendant l'approche proposée spécifique à leur mode d'acquisition.

Le diagramme illustré à la Figure 10 présente les différents traitements constitutifs de cette approche. Ceux-ci sont détaillés dans les paragraphes qui suivent.

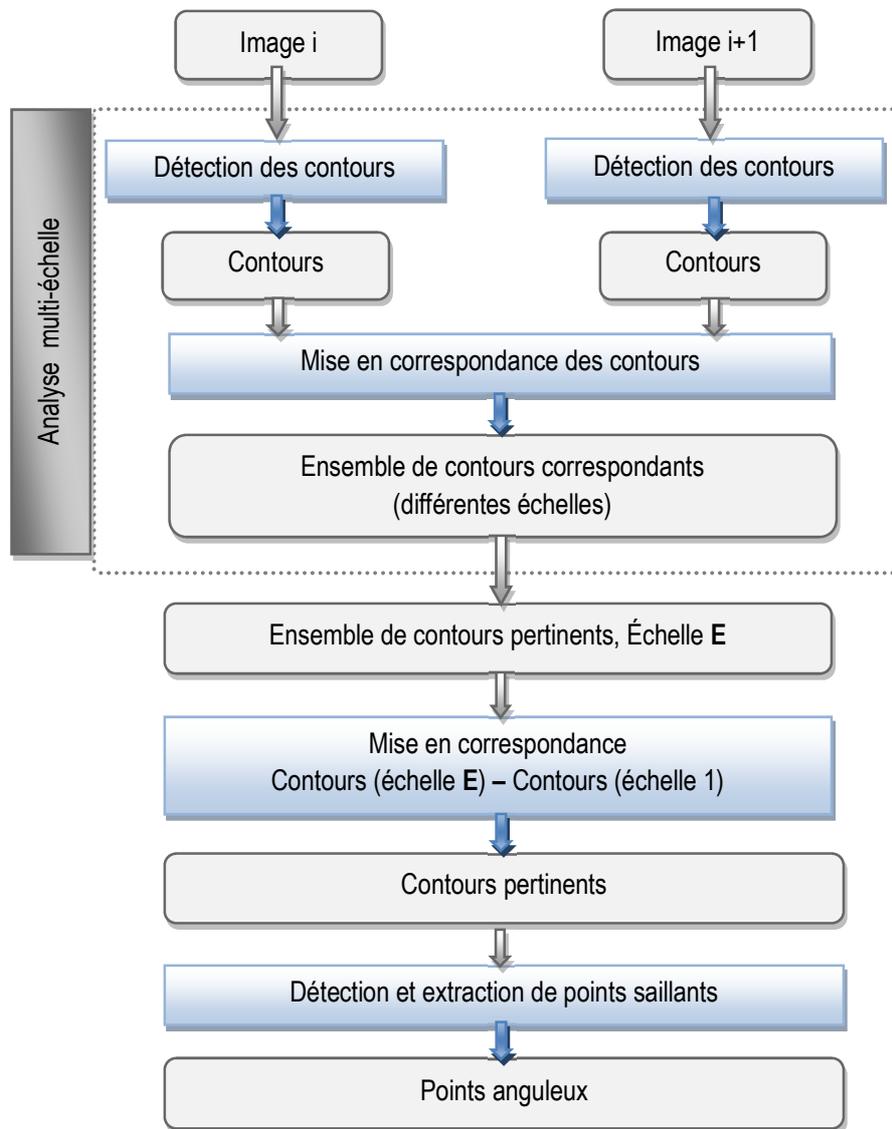


Figure 10 : Chaîne de traitements de l'approche proposée d'extraction de points anguleux à partir d'une séquence d'images acoustiques

3.3.1 Extraction de contours pertinents par analyse multi-échelle

Les pixels représentant des caractéristiques géométriques saillantes des objets se trouvent *a priori* sur des contours que l'on va chercher à extraire des images. Néanmoins, ces contours ne doivent être issus que de la forme des objets de la scène. C'est pourquoi, l'approche proposée consiste en trois étapes :

- une première étape de détection de tous les contours présents dans les images;

- une seconde étape de mise en correspondance de ces contours entre images successives afin de réduire les faux contours issus du bruit des images.
- une troisième étape d'analyse des contours à travers une pyramide gaussienne d'images filtrées afin d'identifier les contours pertinents qui représentent la géométrie des objets de la scène.

3.3.1.1 Détection de contours

Comme décrit précédemment (paragraphe 3.2.1.2), les contours caractérisent une variation brusque des intensités des pixels dans une image. Cette variation peut s'étendre de manière graduelle sur plusieurs pixels adjacents. La détection de ces variations et leur localisation ont fait l'objet de nombreuses publications. Les approches s'appuyant sur le calcul du gradient des intensités (ex. Laplacien, Sobel) (Kant, 2013) sont des méthodes classiques dont les performances chutent rapidement si l'image est bruitée comme c'est le cas pour les images sonar. Le filtre de Canny (Canny, 1986) qui réalise une détection de gradient suite à un filtrage a été spécifiquement développé dans le but de fournir un opérateur robuste au bruit et performant. C'est cet opérateur que nous avons exploité afin d'extraire les contours des objets d'intérêt de nos images acoustiques.

Le filtre de Canny a donc pour objectif d'extraire des arêtes de l'image, les contours étant ensuite construits en reliant les arêtes locales connexes. Le principe d'application de ce filtre se décompose en trois étapes : un filtrage passe-bas (gaussien) pour réduire le bruit dans l'image, un rehaussement des arêtes via le calcul du gradient de l'image, un seuillage par hystérésis afin de ne retenir que les arêtes susceptibles de former un contour continu. Ce seuillage compare l'intensité du gradient en chaque pixel avec deux seuils (un seuil haut et un seuil bas) pour retenir ou non le pixel en tant qu'arête. Ainsi, pour chaque pixel, si l'intensité de son gradient est :

- Inférieure au seuil bas : le pixel est rejeté comme arête;
- Supérieure au seuil haut : le pixel est considéré comme une arête;
- Comprise entre le seuil bas et le seuil haut : le pixel est accepté comme arête uniquement s'il est connecté à un pixel défini au préalable comme une arête.

Lors de l'application du filtre de Canny, nous avons choisi un seuil haut égal à 0.25 et un seuil bas égal à 0.10. Quant à la résolution du filtre gaussien (sigma), elle est égale à 1. Le choix de ces valeurs a été déterminé empiriquement de sorte à conserver au maximum les contours qui caractérisent la forme des objets tout en réduisant la détection des contours issus du bruit des images (faux contours). En fait, un seuil trop bas peut conduire à la détection de faux contours. Inversement, un seuil trop haut peut empêcher la détection de certains contours pertinents. Le choix empirique de ces valeurs est dû au fait qu'il n'existe pas actuellement de méthodes génériques pour déterminer des seuils produisant des résultats satisfaisants sur tous les types

d'images. Les expérimentations ont montré que ces seuils choisis sont efficaces pour la détection de tous les contours pertinents qui caractérisent la géométrie de la scène.

Trois critères de qualité définissent le filtre de Canny: 1) une détection forte même pour les contours des objets n'ayant pas un fort contraste en termes d'intensité avec leur voisinage dans l'image; 2) une localisation précise de tous les contours; 3) une détection unique du même contour (Canny, 1986).

Chaque arête détectée par le filtre de Canny a un ou deux voisins dans l'image sauf aux points de jonction où le nombre de voisins est supérieur à deux. Une arête ayant un seul voisin correspond à la terminaison d'un contour. Ce dernier peut être ouvert lorsque chacune de ses extrémités est une terminaison ou une arête membre d'une jonction. Un contour est fermé lorsque le premier et le dernier point sont connexes. En s'appuyant sur ces points qui définissent la création d'un contour à partir d'une carte d'arêtes, il est possible de passer d'un ensemble d'arêtes à un ensemble de contours caractérisant les zones de variation d'intensité dans l'image.

3.3.1.2 Mise en correspondance des contours et élimination des faux contours

Parmi les contours détectés par l'algorithme de Canny, certains résultent de la présence du bruit dans l'image. Ces « faux contours » nuisent donc à la caractérisation souhaitée de la forme des objets. Pour éliminer ces contours, nous avons élaboré un algorithme d'appariement multi-images qui se base sur deux hypothèses: 1) il y a un faible déplacement de la caméra entre deux prises de vue consécutives au sein d'une séquence et donc entre les deux images considérées; 2) un contour pertinent (représentatif de la géométrie d'un objet) va être présent avec une forme très similaire, dans plusieurs images consécutives, alors que les contours provenant du bruit (ou faux contours) ne se répètent pas, *a fortiori* avec la même forme, d'image en image. Cet algorithme d'appariement permet d'identifier les contours pertinents en recensant les contours qui se retrouvent de manière similaire sur des images successives et donc de rejeter les faux contours.

3.3.1.2.1 Définition d'une mesure de similarité inter-contours.

L'algorithme d'appariement développé repose donc sur un critère de similarité entre contours provenant de deux images consécutives i et $i+1$. Ces contours sont considérés comme correspondants s'ils représentent la même discontinuité dans les deux images tout en étant situés aux mêmes endroits à un voisinage près, dans ces images. Le faible déplacement de la caméra est pris en compte par l'introduction de ce voisinage. Ainsi, le correspondant du contour C_k^1 de l'ensemble C_1 des contours détectés sur l'image i est recherché dans l'ensemble C_2 des contours issus de l'image $i+1$ dans le voisinage qu'aurait C_k^1 sur cette image $i+1$. Pour quantifier la présence de C_k^1 dans les contours C_2 de l'image $i+1$, la mesure $M_k^{1 \rightarrow 2}$ calcule le nombre de points de C_k^1 appartenant à des contours de C_2 à l'emplacement qu'aurait le contour C_k^1 dans l'image $i+1$ à un voisinage η près. L'équation (9) définit une telle mesure normalisée par la longueur du contour C_k^1 :

$$M_k^{1 \rightarrow 2} = \frac{\sum_{p \in C_k^1} \mathbb{1}[\eta(p) \in C_2]}{l(C_k^1)} \quad (7)$$

- $l(C_k^1)$: longueur du contour C_k^1 en nombre de pixels,
- p : pixels dans l'image i appartenant au contour C_k^1 ,
- $\eta(p)$: pixels appartenant au voisinage du pixel p dans l'image $i+1$,
- $\mathbb{1}[\eta(p) \in C_2] = 1$ si dans le voisinage de p dans l'image $i+1$, il existe un pixel de contours appartenant à C_2 , 0 sinon.

Tout contour C_k^1 de l'image i représenté par un très faible nombre de points $M_k^{1 \rightarrow 2}$ dans l'image $i+1$ correspond alors à un contour provenant d'un phénomène non redondant entre les deux images et peut donc être éliminé. Ceci permet de définir $M_k^{1 \rightarrow 2}$ comme une mesure de redondance d'un même contour sur deux images consécutives.

La taille du voisinage $\eta(p)$ est fonction de la cadence d'acquisition des images acoustiques et du déplacement de la caméra. En effet, ces deux paramètres permettent d'estimer de combien de pixels les objets se déplacent entre deux images. Dans le présent contexte, le décalage en pixels entre deux images successives varie entre un et trois pixels maximum. En s'appuyant sur ce déplacement, la dimension du voisinage à choisir pour retrouver les mêmes pixels dans deux images successives avec un déplacement dans une direction unique correspond soit à 3×3 (pour un décalage de 1 pixel) ou 7×7 (pour un décalage de 3 pixels). Néanmoins, avec les séquences d'images utilisées dans les expérimentations, il s'est avéré qu'un voisinage d'une dimension de 3×3 est adéquat pour le calcul du nombre de points $M_k^{1 \rightarrow 2}$.

3.3.1.2.2 Élimination des faux contours

Ayant la mesure $M_k^{1 \rightarrow 2}$ indiquant le nombre de points qui représente chaque contour C_k^1 d'une image i dans l'image $i+1$, nous pouvons estimer un seuil qui permettra par la suite de sélectionner les contours pertinents et d'éliminer les faux contours. En fait, la mesure $M_k^{1 \rightarrow 2}$ permet de détecter la présence d'un contour de l'image i même s'il a été décomposé en un ensemble de contours dans l'image $i+1$. Elle fournit ainsi une information robuste sur la redondance des contours d'une image à une autre. Un faux contour est ainsi représenté par une valeur faible de $M_k^{1 \rightarrow 2}$ contrairement à un contour pertinent.

La représentation sous forme d'un diagramme des valeurs de la mesure $M_k^{1 \rightarrow 2}$ donne alors une vue globale sur l'ensemble des contours détectés. On distingue deux classes de contours : 1) les contours d'une image i qui sont représentés par une forte valeur de $M_k^{1 \rightarrow 2}$ dans l'image $i+1$: les contours pertinents; 2) les contours

d'une image i qui sont représentés par une faible valeur de $M_k^{1 \rightarrow 2}$ dans l'image $i+1$: les contours issus du bruit.

Le seuil défini pour la suppression des contours issus du bruit provient de l'observation des valeurs prises par la mesure $M_k^{1 \rightarrow 2}$ en fonction des contours. Les contours dont $M_k^{1 \rightarrow 2}$ est supérieure à 40 % du nombre total de points qui représente le contour C_k^1 d'une image i dans l'image $i+1$ sont tous des contours pertinents d'où la définition d'un seuil de sélection égale à 40 %. En fait, l'objectif final de la définition de ce seuil est de contribuer à la sélection de points saillants les plus pertinents et les plus représentatifs de la géométrie de la scène. Certes, l'intégration d'un seuil dans la sélection de contours pertinents aura des conséquences au niveau du nombre de points saillants détectés. En fait, certains points qui représentent les détails d'un objet dans la scène seront éliminés. Ceci peut affecter la reconstruction tridimensionnelle si on cherche à représenter fidèlement l'objet. Par contre, si la reconstruction vise à représenter la forme globale des objets, les points issus des contours pertinents qui caractérisent sa géométrie globale seront suffisants. En fait, sans l'utilisation de ce seuil de sélection de contours pertinents, de faux contours seront conservés et on se retrouve face à des points saillants issus de la présence de bruit comme dans le cas des points détectés par le détecteur de Harris, SIFT, etc. (cf. Paragraphe 3.2.2.2). Ceci peut, aussi, nuire à la méthodologie de reconstruction puisque il faut trouver des solutions efficaces pour ne sélectionner que les points saillants caractérisant la forme de l'objet.

3.3.1.3 *Sélection des contours par analyse multi-échelle*

L'analyse multi-échelle combine les étapes citées dans les deux sous-sections précédentes, la détection des contours et leur mise en correspondance, à différentes échelles de filtrage gaussien de l'image (pyramide gaussienne). Ces contours sont suivis sur plusieurs échelles jusqu'à l'atteinte d'une échelle de filtrage à partir de laquelle les contours pertinents peuvent être extraits. Une description de l'évolution des contours et de l'identification de l'échelle de filtrage est effectuée dans les sections qui suivent

3.3.1.3.1 *Suivi de l'évolution des contours*

L'image est filtrée à plusieurs échelles. Les contours sont par la suite détectés, mis en correspondance et sélectionnés selon la méthode décrite dans le paragraphe précédent. En plus de ces étapes, les contours détectés sont comptés par l'algorithme. Ayant le nombre des contours pour chaque échelle de filtrage, un diagramme représentant la variation de leur nombre en fonction de l'échelle de filtrage est établi. L'observation de ce diagramme permet de voir l'évolution du nombre de contours détectés le long de l'analyse multi-échelle.

3.3.1.3.2 *Sélection de l'échelle de filtrage*

L'échelle de filtrage E correspond à l'échelle au niveau de laquelle le nombre de contours détectés devient stable. Cette stabilité indique que tous les contours issus du bruit ont été éliminés et seuls les contours

caractérisant la géométrie des objets résistent au filtrage gaussien répétitif d'une échelle à une autre. Cette stabilité peut être observée sur le diagramme qui représente l'évolution du nombre de contours en fonction de l'échelle de filtrage (paragraphe précédent). Elle est aussi détectée par l'algorithme dès que la variation du nombre de contours devient très faible voire invariable d'une échelle à une autre (ex. 5 contours restant d'un ensemble initial de 36 contours à partir d'une échelle de l'ordre de 8). Des expérimentations ont montré que lorsqu'à deux échelles successives le nombre de contours est le même, les contours de ces échelles sont, en fait, les contours pertinents. De ce fait, un seuil de stabilité égal à 1 a été fixé et l'équation permettant de remonter à l'échelle de sélection est la suivante :

$$\text{Nombre de contours à l'échelle } (i+1) - \text{Nombre de contours à l'échelle } i = < 1 \quad (8)$$

Ainsi, l'algorithme sélectionne automatiquement l'échelle à partir de laquelle la stabilité du nombre de contours apparaît.

L'échelle **E** n'est pas spécifique à la détection d'un contour pertinent précis dans l'ensemble des contours détectés mais correspond plutôt à un seuil de détection de tous les contours pertinents associés aux objets dans l'image.

3.3.1.3.3 Extraction de contours pertinents

Ayant identifié les contours pertinents, une mise en correspondance finale entre les contours de l'échelle **E** et ceux de l'échelle 1 (i.e. l'image originale) est nécessaire. En fait, à cause du lissage gaussien de l'image, les contours pertinents sélectionnés à l'échelle **E** se trouvent déformés. La mise en correspondance avec ceux de l'échelle 1 permet ainsi de retrouver des contours précis caractérisant la forme des objets. Aucun faux contour n'est détecté à partir de l'échelle 1 puisque cette mise en correspondance permet d'extraire uniquement les contours ayant des correspondants détectés au niveau de l'échelle **E**. De ces contours pertinents ainsi extraits de l'image à l'échelle 1 seront extraits les points saillants. À l'échelle 1, l'image est filtrée par un filtre gaussien de niveau 1 ce qui réduit le bruit au niveau des contours. En conséquence, les points saillants détectés représentent uniquement les coins et les points d'inflexion des contours pertinents. Ces points sont les seuls points robustes qui seront utilisés dans les prochaines étapes de la reconstruction 3D.

3.3.2 Extraction de points saillants

L'étape précédente a permis d'extraire des contours représentatifs de la géométrie des objets. L'extraction des points dits saillants ou anguleux des objets va donc consister à déterminer les pixels de courbure maximale de ces contours mais aussi les pixels correspondant au point d'intersection entre ces contours. Les pixels extraits vont donc correspondre à des points saillants de la géométrie des objets de la scène (principalement des coins ou des sommets). Pour ce faire, nous avons appliqué l'approche proposée par (Douglas et Peucker,

1973). Cet algorithme se retrouve cité à maintes reprises dans la littérature (Weidner et Forstner, 1995; Alharthy et Bethel, 2002; Alharthy et Bethel, 2004) et récemment dans des travaux portant sur la manipulation d'informations ponctuelles issues de capteurs LiDAR (Light Detection And Ranging) (Tarsha Kurdi, 2008). L'algorithme de Douglas-Peucker est un algorithme récursif qui a prouvé son efficacité dans la simplification ou le lissage d'éléments non linéaires en formes rectilignes (Lahamy, 2008, He et Ye, 2013).

Pour la détection de points saillants des contours, l'algorithme de Douglas-Peucker s'initialise en joignant deux points non consécutifs du contour par une corde (a) (cf. Figure 11) et fait intervenir un seuil S qui permet de détecter les sommets susceptibles d'être des coins d'objets qui se trouvent entre ces deux points. Pour ce faire, pour chacun des sommets, il calcule la distance orthogonale (distance b, Figure 11) ou écart angulaire qui le sépare de la corde. Si la distance orthogonale b est supérieure au seuil de tolérance S, alors ce sommet est identifié comme un « coin ». La valeur du seuil fixée dans cette thèse est égale à deux pixels. Le point qui se trouve au-delà de deux pixels par rapport à la corde est ainsi considéré comme un coin de contour. Ce seuil est le seuil minimum (deux pixels) qu'on peut fixer pour la détection de points de courbure. Le choix de ce seuil minimum a été fait dans l'objectif de détecter tout point de contour présentant une courbure par rapport à son voisinage et d'extraire ainsi les points qui caractérisent la morphologie du contour.

L'algorithme répète cette opération récursivement jusqu'à ce qu'il ne soit plus possible de trouver de sommets vérifiant $b > S$. A l'issue de ce processus, à chaque contour détecté est associée une liste de points saillants identifiés comme les coins des objets.

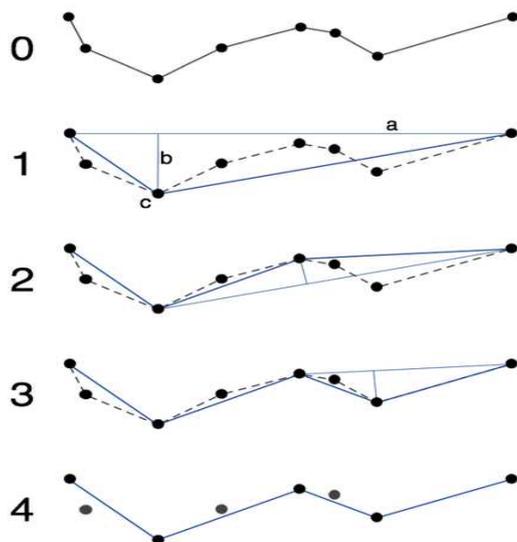


Figure 11 : Détection de points anguleux par l’algorithme de Douglas-Peucker (*Douglas et Peucker, 1973*) a correspond au segment liant les deux extrémités du contour, b est la distance orthogonale entre le segment a et le point de contour c ou sommet le plus éloigné de a. b étant supérieure à la tolérance, ce sommet est conservé comme coin ou point saillant.

3.4 Expérimentations et résultats

Cette section est consacrée à la présentation des expérimentations réalisées pour tester et valider l’approche d’extraction de points saillants. Les résultats affichés dans la première partie de cette section concernent la séquence « Pied de pont ». Ils représentent en détail les différentes étapes de la méthodologie. Quant à la deuxième partie, elle est consacrée aux résultats des expérimentations utilisant les séquences « Épave » et « Piscine ».

3.4.1 Description des jeux de données

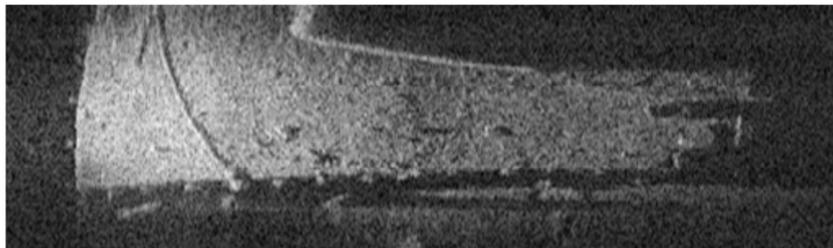
Les données exploitées dans cette partie sont des séquences d’images acquises avec une caméra acoustique DIDSON dans :

1. un bassin intérieur (la piscine profonde de l’Université Laval) : les séquences ont été acquises en mode haute fréquence (1,8 MHz) avec une portée de 5 m. La Figure 12.a illustre la scène imagée. Elle représente un ensemble d’objets (i.e. boîtes, grille) de différentes tailles déposés sur le fond de la piscine (profondeur = 3 m).
2. un milieu marin : on dispose de deux catégories de séquences d’images. La première catégorie regroupe les séquences acquises en mode haute fréquence (1,8 MHz) avec une portée de 9 m et représente une structure sous-marine (la base d’un pied de pont) (cf. Figure 12.b). La deuxième

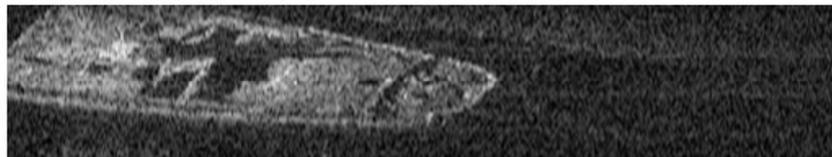
catégorie correspond aux séquences acquises en mode basse fréquence (1,0 MHz) avec une portée de 18 m et représente une épave sur le fond marin (Figure 12.c)



(a) Séquence: objets en piscine



(b) Séquence: pied de pont en milieu marin



(c) Séquence: épave en milieu marin

Figure 12 : Exemples d'images extraites de différentes séquences d'images acoustiques provenant de la caméra DIDSON

3.4.2 Description du protocole de test

L'objectif principal des tests est d'évaluer l'efficacité de l'approche proposée pour extraire les points anguleux de la scène observée sur les séquences. Son efficacité est dépendante de l'efficacité de ses différentes étapes constitutives. Par conséquent, les expérimentations ont été articulées autour de chacune de ces étapes et de la validation de leurs performances. Plus précisément, elles portent sur :

1. L'extraction des contours;
2. La sélection de contours pertinents qui représentent uniquement les objets de la scène;
3. L'extraction de points anguleux.

Pour ce qui est de l'extraction de points anguleux (étape 3), l'efficacité de leur extraction concerne non seulement la bonne sélection de points représentatifs d'objets dans la scène mais aussi l'aptitude à retrouver ces points dans plusieurs images de la séquence, images acquises selon différents points de vue. De ce fait,

les tests ont été appliqués sur un ensemble d'images successives. Ayant des séquences d'images différentes acquises dans des milieux et avec des contextes variés, les tests ont été repartis en fonction des séquences.

Étant donné que l'enchaînement des étapes intervenant dans l'approche proposée se fait de façon séquentielle, la qualité des résultats fournis par chaque étape dépend de la qualité des résultats fournis par l'étape précédente. Par conséquent, les tests et les expérimentations réalisés pour la mise en œuvre de l'approche d'extraction de points saillants se sont faits de manière itérative jusqu'à l'obtention de résultats jugés satisfaisants selon certains critères d'évaluation. Ces critères d'évaluation sont la qualité de détection des points, leur localisation et la répétabilité de ces points.

3.4.3 Résultats des tests

3.4.3.1 Tests sur les séquences acquises en milieu marin

3.4.3.1.1 Séquence « Pied de pont »

a) Détection de contours

La figure suivante (cf. Figure 13) illustre le résultat de l'application du filtre de Canny pour la détection des arêtes et leur chainage pour la détection de contours sur deux images successives de la séquence « Pied de pont » :

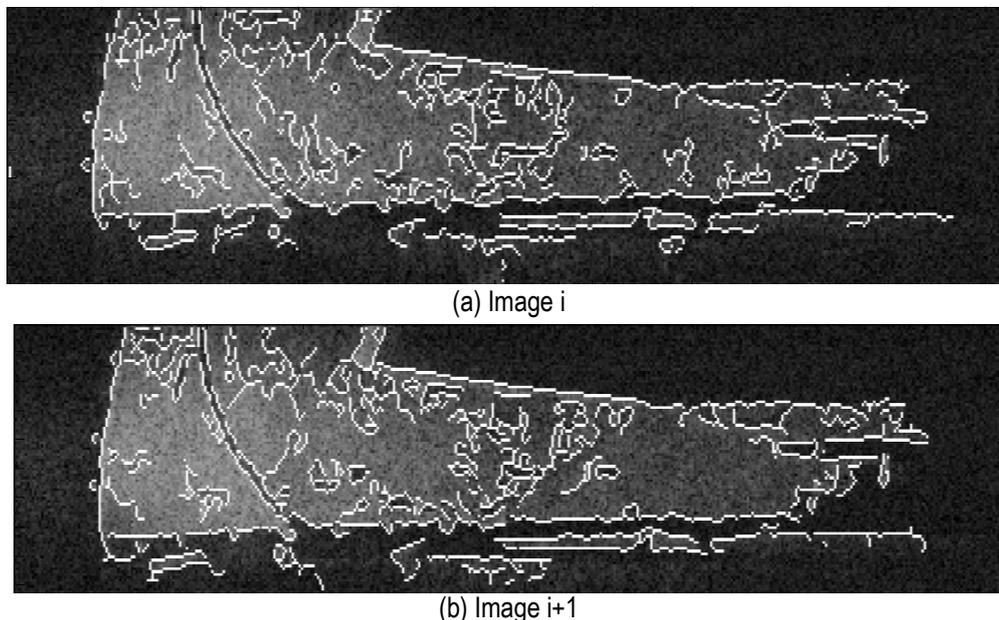


Figure 13 : Extraction de contours des images de la séquence « Pied de pont ».

On distingue sur ces images la présence de différents contours. Il est difficile de séparer ceux qui appartiennent effectivement à la structure de ceux qui résultent du bruit.

b) Sélection des contours pertinents sur des images successives

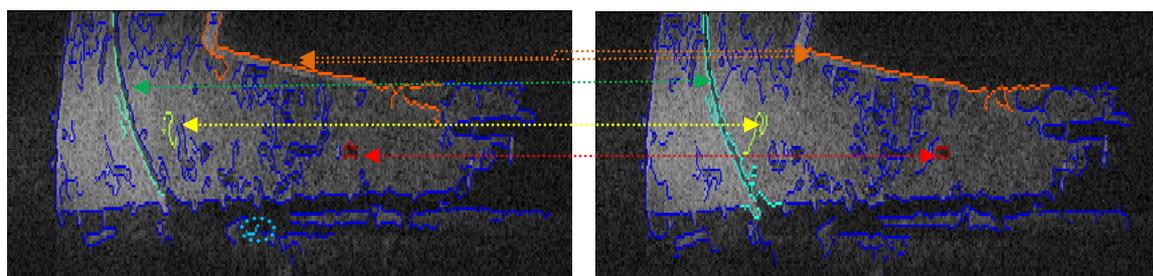
La mise en correspondance des contours a été effectuée sur les contours extraits lors de l'étape précédente (extraction des arêtes et chaînage) sur différentes paires d'images successives. Dans ce paragraphe est présenté l'exemple des deux images successives (image i et image $i+1$) de la Figure 13. Ces tests ont pour objectif d'évaluer l'efficacité de cette approche dans l'élimination de faux contours.

Ainsi, pour l'exemple de la paire d'images utilisée, un histogramme va illustrer les valeurs prises par la mesure $M_k^{1 \rightarrow 2}$ calculée pour chaque contour de l'image i dans l'image $i+1$ (cf. Figure 14). Cet histogramme indique clairement que les contours pertinents c'est-à-dire représentatifs de la géométrie de la scène ont une valeur de $M_k^{1 \rightarrow 2}$ élevée alors que ceux issus de la présence de bruit ont une valeur de $M_k^{1 \rightarrow 2}$ faible.

La visualisation de la mesure $M_k^{1 \rightarrow 2}$ des contours de l'image i dans l'image $i+1$ par l'historgramme de la Figure 14 montre une variation de $M_k^{1 \rightarrow 2}$ de 0 % à 100 %. Nous montrons quelques exemples de contours ayant différentes valeurs de $M_k^{1 \rightarrow 2}$:

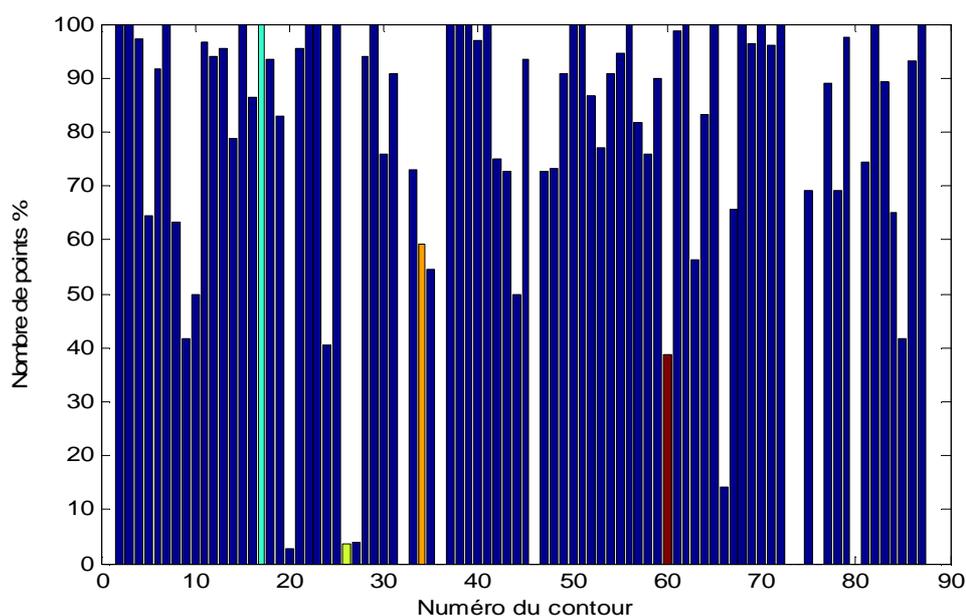
- Le contour numéro 17 dans l'historgramme a un correspondant dans l'image $i+1$ avec 100 % de points. Ce contour est coloré en vert dans la Figure 14.a, b et c.
- Le contour numéro 34 dans l'historgramme a un correspondant dans l'image $i+1$ avec 60 % de points. Ce contour est coloré en orangé dans la Figure 14.a, b et c.
- Le contour numéro 60 dans l'historgramme a un correspondant dans l'image $i+1$ avec 60 % de points. Ce contour est coloré en rouge dans la Figure 14.a, b et c.
- Le contour numéro 26 dans l'historgramme a un correspondant dans l'image $i+1$ avec 4 % de points. Ce contour est coloré en jaune dans la Figure 14.a, b et c.
- Le contour numéro 46 dans l'historgramme n'a pas de correspondant dans l'image $i+1$. Il apparaît logiquement avec 0 % de points dans l'historgramme. Ce contour est mis en évidence dans l'image i (cf. Figure 14.a) avec la couleur bleu clair (et encerclé en bleu). On remarque qu'il s'agit d'un petit contour qui n'appartient pas à la structure.

Les contours illustrés en Figure 14 regroupent à la fois des contours pertinents d'où des points saillants seront extraits dans les prochaines étapes pour la reconstruction de la scène tels que le contour 34, des contours qui représentent la présence d'un objet dans la scène (toutefois ceux-ci ne fournissent pas de points saillants tels que le contour 17) et des faux contours tels que le contour 46.



(a) image i

(b) image i+1



(c) Histogramme des contours de l'image i dans l'image i+1

Figure 14 : (a) Sélection de contours par mise en correspondance dans l'image i. (b) Sélection de contours par mise en correspondance dans l'image i+1. (c) Mesure $M_k^{1 \rightarrow 2}$ des contours de l'image i dans l'image i+1. La barre colorée en vert est le contour 17 en vert dans (a) et (b). La barre colorée en orangée est le contour 34 en orangé dans (a) et (b). La barre colorée en rouge est le contour 60 en rouge dans (a) et (b). La barre colorée en jaune est le contour 26 en jaune dans (a) et (b). Le contour bleu encerclé dans (a) est le contour 46 dans l'historgramme n'ayant pas de correspondant dans l'image (b)

Afin de réduire le nombre de faux contours, des tests ont été réalisés dans l'objectif de définir un seuil de sélection de contours pertinents. Plusieurs valeurs de sélection (ex. 10, 25, 30, 40, etc.) ont été testées. L'analyse des résultats et de l'historgramme des contours met en évidence que les contours pertinents ont une valeur de $M_k^{1 \rightarrow 2}$ supérieure à 40 %. Ceux ayant une valeur de $M_k^{1 \rightarrow 2}$ inférieure à 40 % représentent, généralement, des contours issus du bruit c'est-à-dire des contours dont seulement certains points ont la même localisation sur plusieurs images, d'où leur faible valeur de $M_k^{1 \rightarrow 2}$ d'une image à une autre. Ils ne sont

pas nombreux par rapport aux contours pertinents, particulièrement, dans le cas de contours correspondants entre deux images successives. Ils sont représentés par des valeurs de $M_k^{1 \rightarrow 2}$ variant de 5 à 25 % environ.

Ce seuil de sélection est appliqué lors de chaque étape d'une analyse multi-échelle des images (cf. description dans le paragraphe 3.3.1.3) avant de compter le nombre de contours pertinents restant dans l'image. La Figure 15 illustre l'effet de différents seuils de sélection sur la détection des contours à un niveau d'échelle égal à 8 de l'analyse multi-échelle. Suite à l'analyse de ces résultats, le seuil est fixé à une valeur égale à 40 %. Au-delà de ce seuil, on remarque la disparition d'un certain nombre de contours pertinents (au niveau des zones encadrées en jaune dans la Figure 15). De plus, cette analyse a montré que quel que soit le niveau de l'échelle, les contours pertinents qui représentent la forme de l'objet sont toujours présents avec une valeur M supérieure à 40 %. En fait, les contours pertinents résistent au filtrage multi-échelle donc ils gardent, généralement, la même valeur de M jusqu'à la suppression de tous les faux contours présents dans l'image. Donc, ce seuil de sélection est valable pour toutes les échelles.

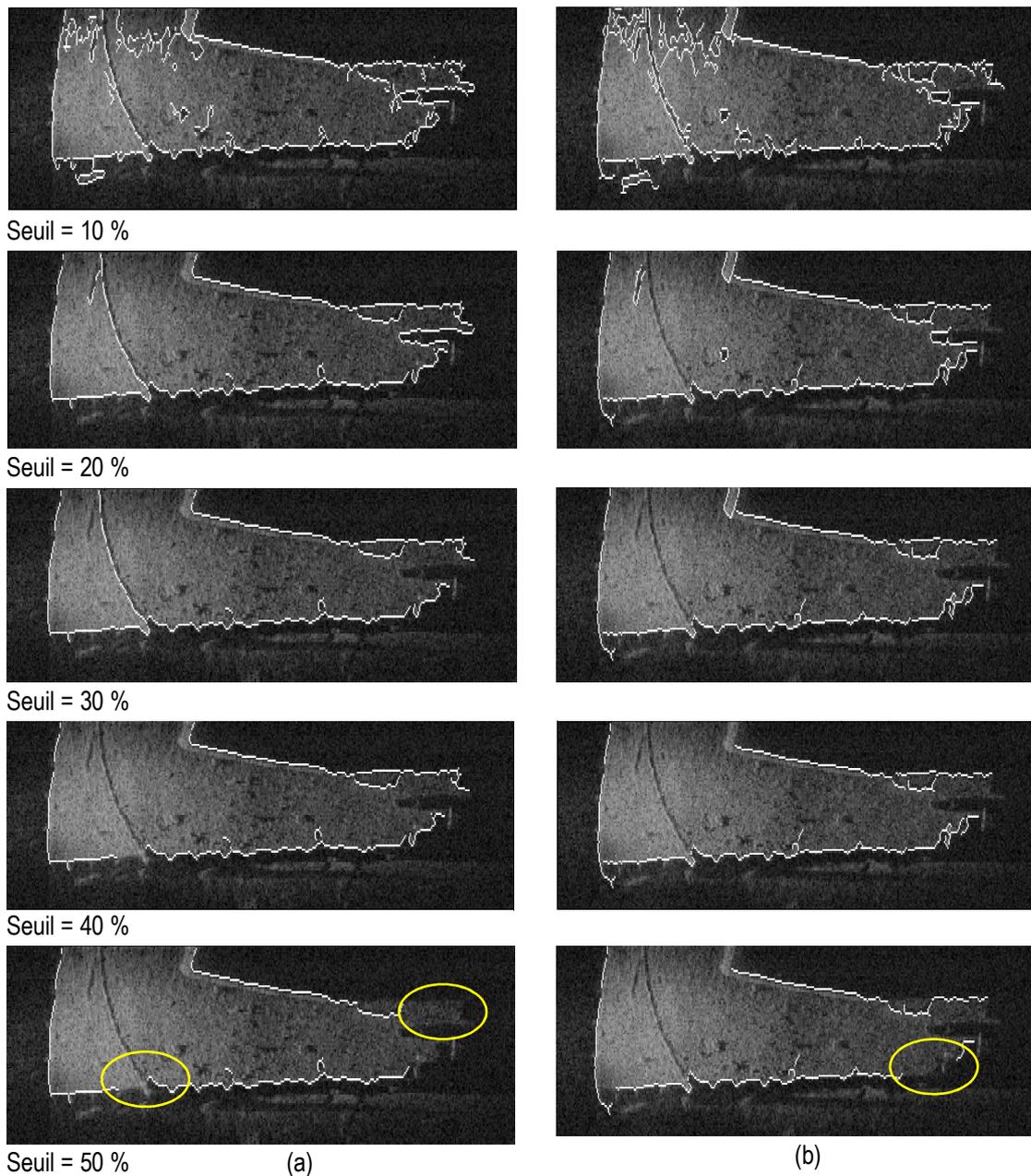


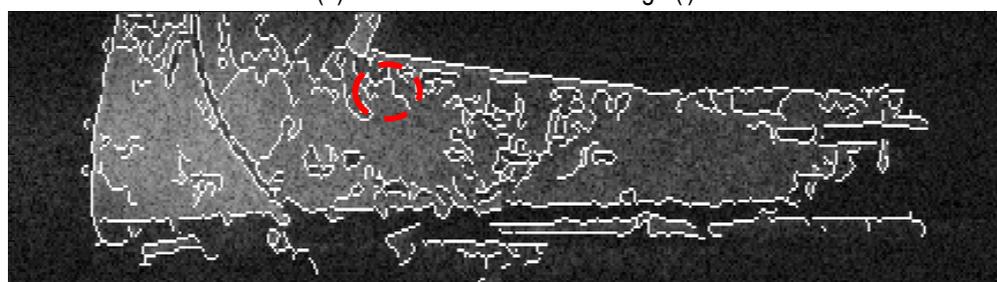
Figure 15 : Influence du seuil de sélection sur la préservation des contours pertinents sur une image i (colonne a) et une image $i+1$ (colonne b) pour un niveau d'échelle égal à 8. À partir d'un seuil égal à 50%, au niveau des régions marquées par les cercles jaunes, les contours caractérisant la géométrie de la structure commencent à disparaître.

La Figure 16 montre le résultat de l'application d'une mise en correspondance des contours détectés et l'application d'un seuil de 40 % à un niveau 1 du filtrage de deux images successives. Dans les deux premières images, Figure 16.a et Figure 16.b, on distingue tous les contours détectés par le filtre de Canny.

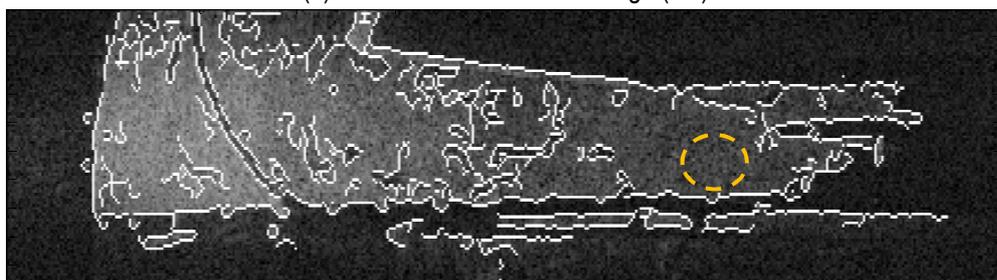
Les régions encadrées en rouge illustrent des exemples de faux contours qui ont été supprimés suite à la mise en correspondance entre les deux images.



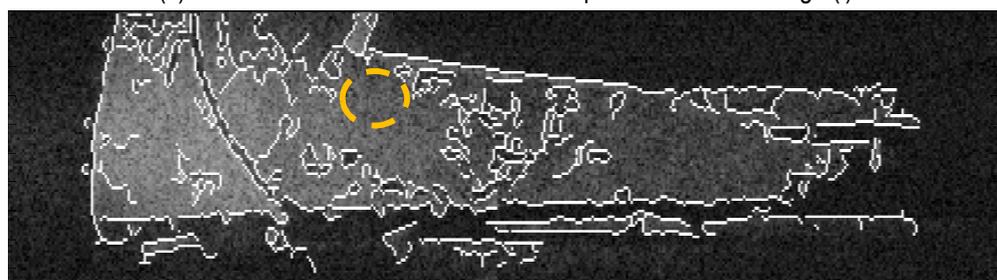
(a) Contours extraits dans l'image (i)



(b) Contours extraits dans l'image (i+1)



(c) Résultat de la sélection des contours pertinents dans l'image (i)



(d) Résultat de la sélection des contours pertinents dans l'image (i+1)

Figure 16 : Mise en correspondance des contours entre une image (i) et une image (i+1). Les régions encadrées en jaune (en c et d) montrent la disparition de faux contours encadrés en rouge (a et b) dans l'image (i) et (i+1) suite à la mise en correspondance.

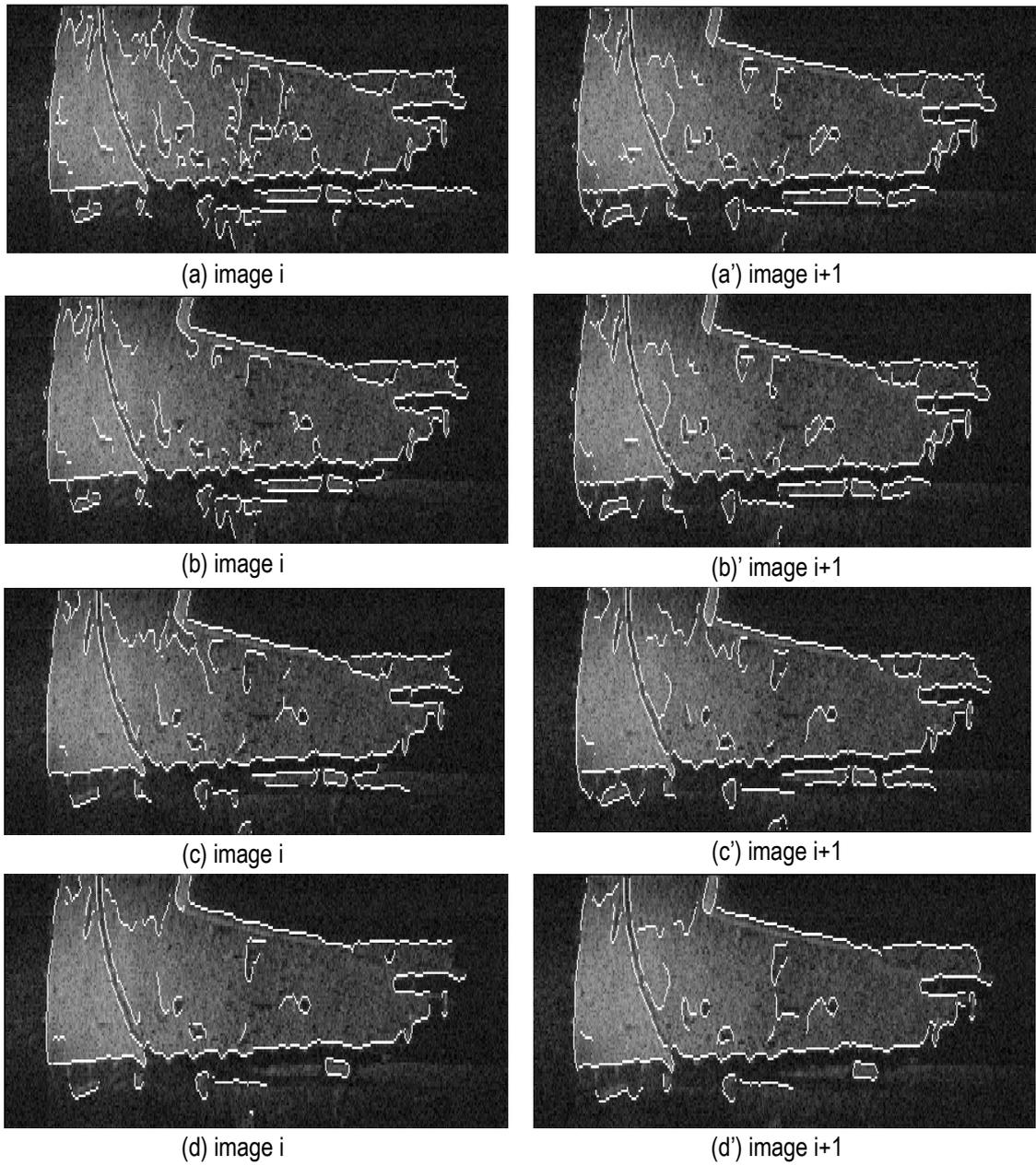
c) Analyse multi-échelle

Cette analyse multi-échelle consiste en l'application de l'algorithme suivant pour chaque échelle de l'analyse:

1. l'image est filtrée par l'application du filtre gaussien adapté à l'échelle considérée

2. les contours sont détectés par le filtre de Canny et reliés par une étape de chaînage
3. les faux contours sont éliminés par l'étape de mise en correspondance

La Figure 17 montre l'évolution des contours en appliquant ces étapes aux différentes échelles de filtrage de la pyramide gaussienne.



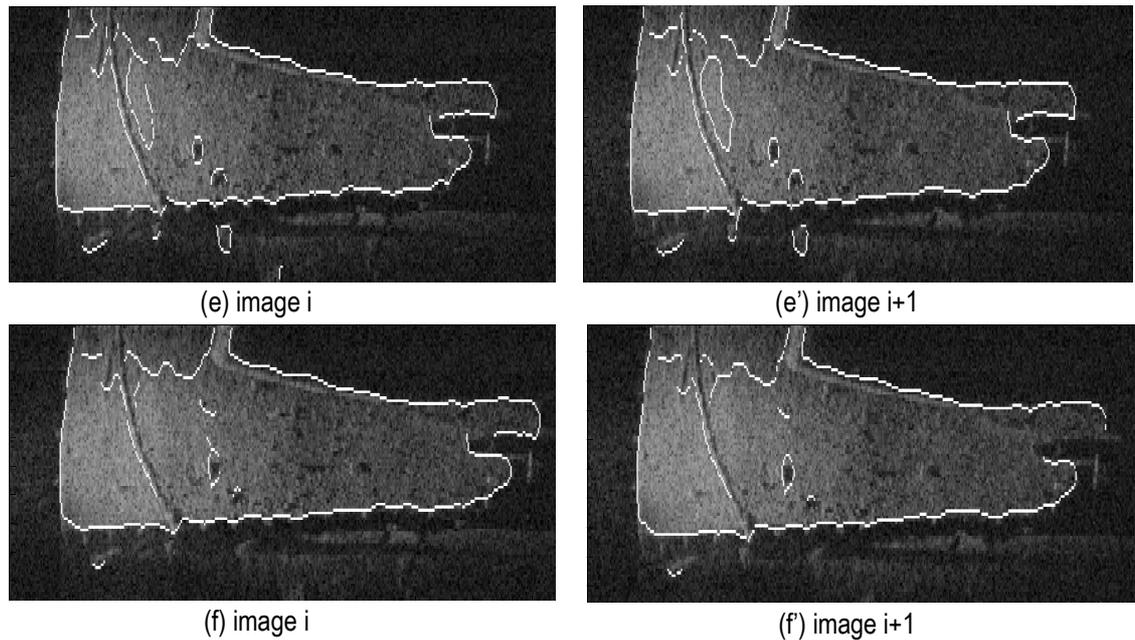


Figure 17 : Extraction de contours pertinents par une analyse multi-échelle sur l'image i (colonne de gauche) et sur l'image $i+1$ (colonne de droite).

Au niveau de l'échelle 8, on remarque que les contours qui restent sur l'image sont des contours qui représentent la structure de la scène. À partir de cette échelle, les contours restent les mêmes indiquant qu'il s'agit de contours pertinents de l'image. Ceci est observé aussi en suivant l'évolution du nombre de contours à travers l'analyse multi-échelle dans la Figure 18. La Figure 18 illustre bien la diminution de nombre de contours au fur et à mesure que l'échelle de filtrage augmente jusqu'à sa stabilisation à la 8ème échelle, échelle que l'on note **E**.

L'objectif de cette étape n'est pas de réaliser une recherche exhaustive de tous les contours mais essentiellement d'aboutir à la sélection de contours pertinents qui nous permettront par la suite de reconstruire la scène de l'image.

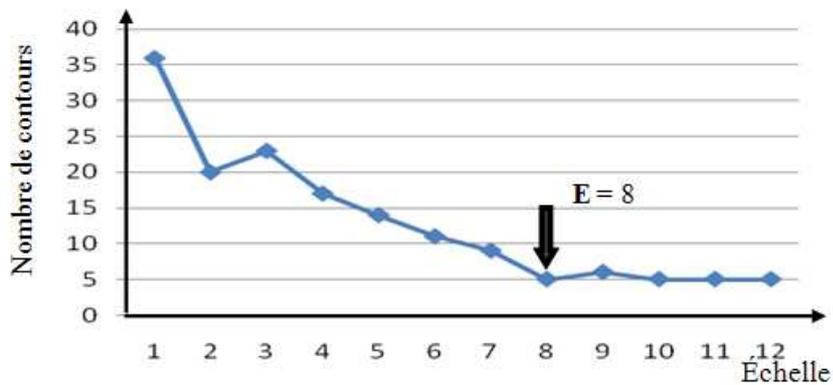


Figure 18 : Évolution du nombre de contours extraits en fonction de l'échelle d'analyse. E indique l'échelle à partir de laquelle le nombre de contours devient stable.

d) Extraction de points saillants

Les contours les plus pertinents ont été détectés à l'échelle **E**; néanmoins, ceux-ci ont une forme très lissée (cf. Figure 17.f). Une dernière étape de mise en correspondance est donc nécessaire pour retrouver dans l'ensemble des contours de l'échelle 1, ceux qui subsistent toujours à l'échelle **E**. Cette étape permet d'obtenir des contours robustes (car issus de l'échelle **E**) et précis en termes de localisation (car issus de l'échelle 1, avec un filtrage de niveau 1). Ces derniers sont bien représentatifs de la forme des objets et donc constituent un support robuste pour l'extraction des points anguleux décrivant la géométrie des objets. Pour ce faire, nous avons utilisé l'algorithme de Douglas-Peucker afin d'extraire de tels points saillants. La Figure 19 montre les points saillants extraits depuis les images i et $i+1$ après application de toutes les étapes de l'approche proposée.

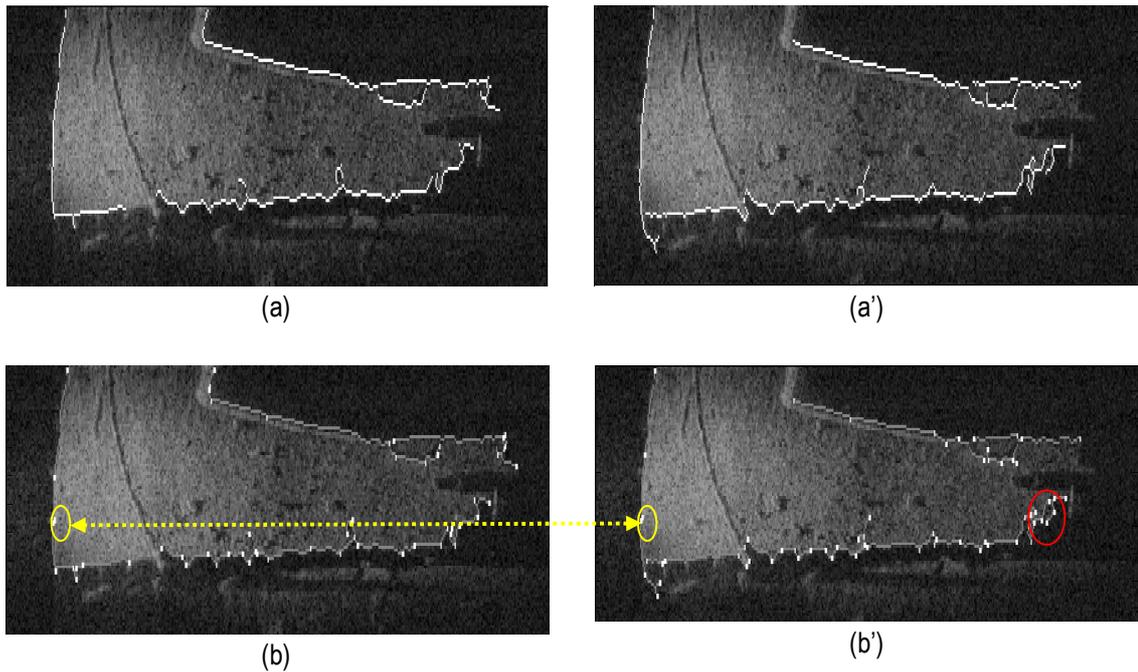


Figure 19 : Extraction des points saillants à partir des contours extraits du niveau d'échelle 1 toujours présents au niveau d'échelle 8 pour les images i et $i+1$. (a) et (a') représentent les contours des images i et $i+1$ qui vont servir à l'extraction des points saillants indiqués sur (b) et (b').

Les points extraits des images sont de deux catégories : la première catégorie qui regroupe les points saillants qui sont répétitifs dans l'image i et l'image $i+1$ (cf. Figure 19, point encerclé en jaune) et qui représentent effectivement les points de déformation de la géométrie de l'objet en marquant ses coins et ses zones d'inflexion. La deuxième catégorie regroupe des points qui sont très proches et concentrés dans la même région de l'image (cf. Figure 19.b', point encerclé en rouge) et qui représentent la présence de zones d'ombre et des objets déposés sur la structure. Ces derniers points ne sont pas directement représentatifs et de manière robuste, de la géométrie de la structure et devraient ainsi être éliminés. Avant de procéder à l'étape de la reconstruction 3D de la scène à partir de l'ensemble de points détectés, une étape de mise en correspondance permettra de sélectionner les points saillants les plus pertinents, notamment les points répétitifs d'une image à une autre.

3.4.3.1.2 Séquence « Épave »

Contrairement à la séquence « Pied de pont », la séquence « Épave » est acquise en mode basse fréquence. Ceci a certainement influencé les détails enregistrés au niveau des images. Les étapes de la méthodologie ont été appliquées aux images de la séquence « Épave » et les résultats sont illustrés à la Figure 20.

Tous les contours pertinents qui ont été extraits des images sont détectés automatiquement par l'algorithme à une échelle de filtrage $E = 6$. À ce niveau de filtrage, les contours détectés illustrent la géométrie de l'épave et

ils ont été ainsi mis en correspondance avec les contours détectés au niveau de la première échelle. Un exemple de résultat est présenté dans la Figure 20 où sont indiqués les contours pertinents extraits d'une paire d'images de la séquence « Épave » (cf. Figure 20, a et a') et les points saillants détectés sur ces contours (cf. Figure 20, b et b').

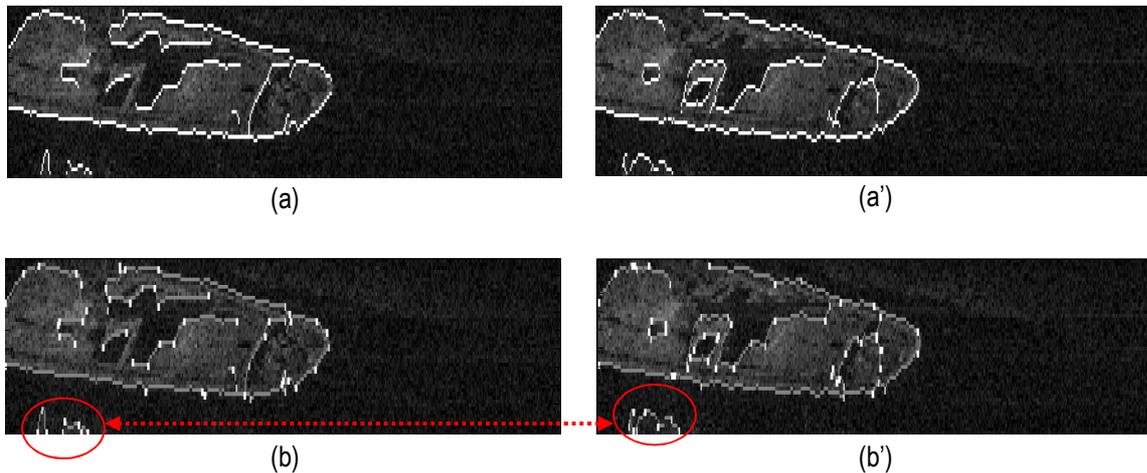


Figure 20 : Extraction de points saillants de la séquence « Épave » : (a et a') contours pertinents extraits des images i et $i+1$ et leurs points saillants (b et b')

Les points saillants détectés représentent non seulement la géométrie de l'épave dans l'image mais on remarque aussi la détection de points non pertinents (cf. Figure 20. b et b', points encadrés en rouge). Ces derniers n'appartiennent pas à l'objet principal (qui est l'épave) mais caractérisent néanmoins la présence d'une variation d'intensité localisée dans les images successives, probablement un autre élément de la scène imagée (cailloux, débris, relief, ...).

3.4.3.2 Tests sur les séquences acquises en piscine

Ayant des données acquises dans la piscine de l'Université Laval, la méthodologie d'extraction de points saillants a été appliquée à un ensemble d'images. La Figure 21 illustre les contours pertinents extraits d'une paire d'images et les points saillants détectés sur ces contours. Pour cet ensemble de données, les contours pertinents sont détectés à une échelle de filtrage $E = 3$. En fait, les séquences sont acquises dans un milieu contrôlé, « la piscine », où les images enregistrées sont moins bruitées que celles acquises en milieu naturel.

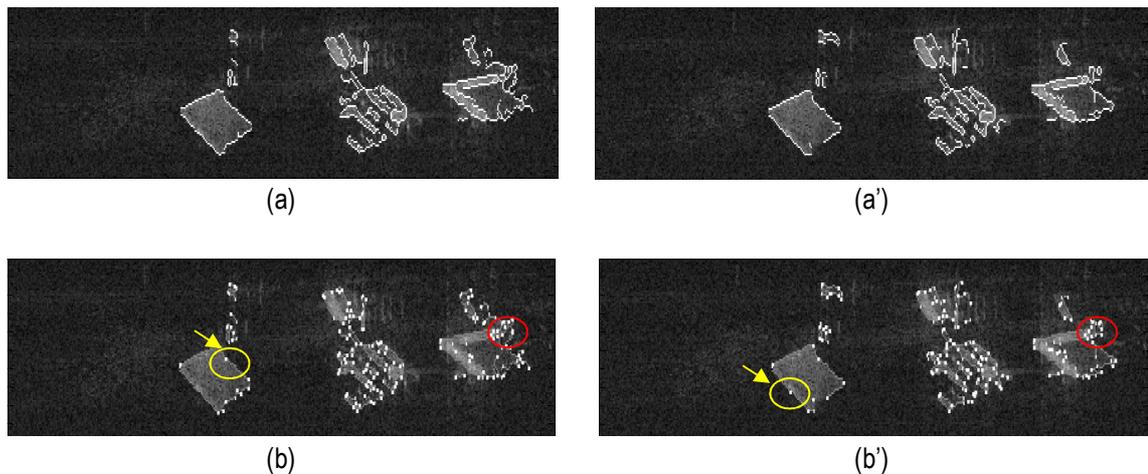


Figure 21 : Extraction de points saillants de la séquence acquise en bassin intérieur : (a et a') contours pertinents extraits des images i et $i+1$ et leurs points saillants (b et b').

La détection des contours au niveau d'une échelle égale à 3 s'explique notamment par un niveau de bruit moins élevé dans les images de la séquence « Piscine » comparativement aux séquences d'images précédentes (séquence « Pied de pont » où $E = 8$ et séquence « Épave » où $E = 6$).

Comme dans les paragraphes précédents les points saillants détectés caractérisent la géométrie de l'objet. Cependant, on remarque la présence de régions où il y a une forte détection de points saillants telles que les régions encadrées en rouge dans la Figure 21.b et b' ainsi qu'au niveau de la grille. Ces régions regroupent de nombreux points saillants dont certains ne sont pas pertinents pour la reconstruction 3D.

Un autre type de points est aussi observé, ce sont les points qui se trouvent au milieu de contours rectilignes tels les points encadrés en jaune dans la Figure 21.b et b'. Ces points constituent des faux coins. La détection de ces points est liée au seuil S de l'algorithme Douglas-Peucker qui a été fixé à une valeur égale à deux pixels (Paragraphe 3.3.2). En fait, en augmentant la valeur du seuil vers une valeur qui permet d'éliminer ces faux coins, on constate que des points saillants pertinents sont aussi éliminés. Donc, dans l'objectif d'extraire le maximum de points saillants pertinents qui caractérisent au mieux la géométrie de la scène et par la suite de les exploiter pour la reconstruction 3D, la valeur du seuil a été maintenue à deux pixels. Quant à la détection de ce type de faux coins, elle n'affectera pas l'étape de la reconstruction 3D. En fait, il existe des solutions qui peuvent être appliquées pour l'élimination de ces faux coins détectés au milieu de contours rectilignes et des régions formant des amas de points resserrés et condensés. En fait, les amas de points pourraient être éliminés puisque il serait difficile d'y identifier et d'y sélectionner de manière robuste les points pertinents représentatifs de la géométrie de la scène de ceux non pertinents. Ceci est réalisable par des approches de segmentation et de sélection de régions d'intérêt. Quant aux points détectés sur les contours rectilignes, une analyse de la géométrie des contours permettra de les identifier et par la suite de les éliminer.

L'analyse géométrique des contours correspond à l'étude de la forme des contours qui permet d'identifier s'il est rectiligne, circulaire, etc. Pour ce faire, les approches de reconnaissance de formes peuvent être appliquées. Finalement, l'étape de la mise en correspondance, qui identifie les paires de points correspondants entre deux images, permettra aussi de pallier le problème de la présence de points non pertinents et la sélection des points saillants pertinents pour la reconstruction 3D.

3.4.3.3 Principales connaissances tirées des expérimentations

En comparant les résultats issus de différentes séquences, on remarque que les faux contours sont plus nombreux dans les séquences acquises en milieu marin que dans celles acquises en piscine. En effet, dans le milieu marin, il y a de nombreux facteurs propices à l'apparition de bruit dans les images acoustiques dont notamment la turbidité de l'eau, la présence de matière en suspension et d'objets flottants qui n'existent pas dans un bassin fermé. Même si dans ce dernier il peut y avoir certaines réflexions multiples des ondes à cause de la matière réfléchissante du fond et aussi à cause de la faible profondeur du bassin, celles-ci ne semblent pas affecter autant la qualité des résultats puisque les traitements appliqués aux images acoustiques acquises en bassin ont permis la détection d'un faible nombre de faux contours comparativement à ceux détectés dans les images acquises en milieu marin.

Afin d'évaluer la pertinence des résultats de l'approche de détection de points saillants des images acoustiques, une évaluation quantitative s'avère nécessaire. Cette dernière est présentée dans les prochains paragraphes.

3.4.4 Évaluation quantitative de l'approche

Afin d'évaluer les performances de l'approche de détection de points saillants proposée dans ce travail, nous avons défini trois critères relatifs aux points saillants extraits: leur répétabilité, leur bonne détection et leur bonne localisation.

3.4.4.1 La répétabilité

La répétabilité est un critère d'évaluation important tout particulièrement dans le contexte d'une reconstruction 3D multi-vues. Dans ce contexte, elle quantifie l'accumulation, pour un même point saillant, de détection dans une suite d'images, acquises sous différents points de vue. En cas de forte répétabilité, un point saillant est donc considéré comme représentatif (partiellement) de la géométrie de la scène observée. Le déplacement du capteur entre l'acquisition des images considérées induit un déplacement des points saillants extraits entre ces images. La définition d'un critère de répétabilité doit prendre en compte ce décalage en considérant un certain voisinage (ϑ) dans lequel un même point peut se trouver d'une image à l'autre.

Afin de calculer la répétabilité, nous avons utilisé la formule définie par Schmid *et al* (2000). Soit deux ensembles de points détectés $\{d_1\}$ et $\{d_i\}$ sur deux images l_1 et l_i avec n_1 le nombre de points du premier ensemble $\{d_1\}$ et n_i le nombre de points du deuxième ensemble $\{d_i\}$. Un point P de l'ensemble $\{d_1\}$ (image l_1) est dit répétable dans l'ensemble $\{d_i\}$ s'il existe un point de l'ensemble $\{d_i\}$ dans un voisinage ϑ autour de la position qu'aurait P dans l'image l_i . La mesure de répétabilité τ_i de l'ensemble $\{d_1\}$ dans l'ensemble $\{d_i\}$ est donc définie comme le ratio du nombre de points répétables de l'ensemble $\{d_1\}$ dans l'ensemble $\{d_i\}$ sur la dimension de l'ensemble ayant le plus faible nombre de points détectés entre l'image l_1 et l'image l_i (cf. Équation 9). Sachant qu'un point P_1 d'une image ne peut avoir qu'un seul point P_i correspondant dans la deuxième image, l'équation de la répétabilité est la suivante :

$$\tau_i = \frac{R_i(\vartheta)}{\min(n_1, n_i)} \quad (9)$$

Avec $R_i(\vartheta) = \{P_1 \in \{d_1\} \mid \exists P_i \in \{d_i\}, dist(P_1, P_i) < \vartheta\}$ où $dist(P_1, P_i)$ mesure sur l'image l_i , la distance entre la position qu'aurait P_1 sur l'image l_i et la position de P_i

Cette mesure de répétabilité a été calculée pour chaque jeu de données (Pied de pont, Épave et Piscine), sur une séquence de 11 images consécutives. La Figure 22 montre la variation du taux de répétabilité pour chaque jeu de données. Cette illustration est intéressante pour la reconstruction 3D multi-images. Elle permet de mettre en évidence les images où les points saillants ont une répétabilité élevée. Par exemple, dans les séquences d'images « Épave » et « Piscine », à partir de l'image 7 jusqu'à l'image 10 le taux de répétabilité des points est élevé. Plusieurs paires de points correspondants peuvent être, par conséquent, extraites de ces images et exploitées pour la reconstruction 3D. La moyenne et l'écart-type de ce taux de répétabilité sont présentés dans le Tableau 2.

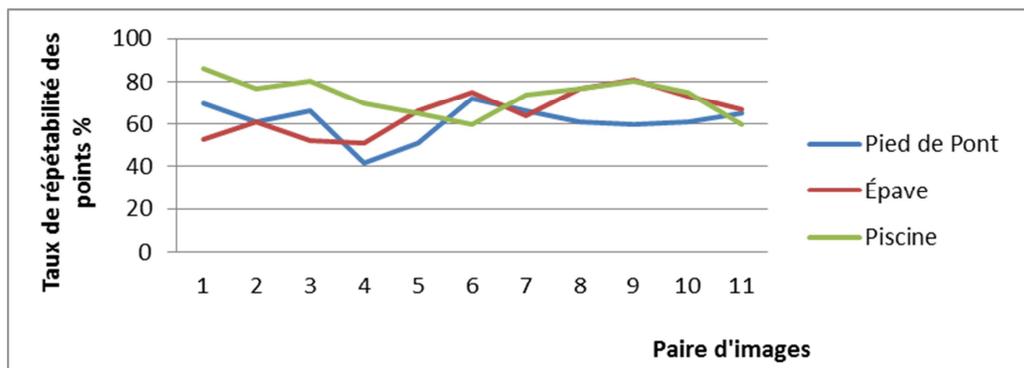


Figure 22 : Répétabilité de l'ensemble des points saillants extraits sur les trois jeux de données pour une séquence d'images consécutives. En abscisse, la paire d'images i fait référence aux images $i-1$ et i du jeu de données.

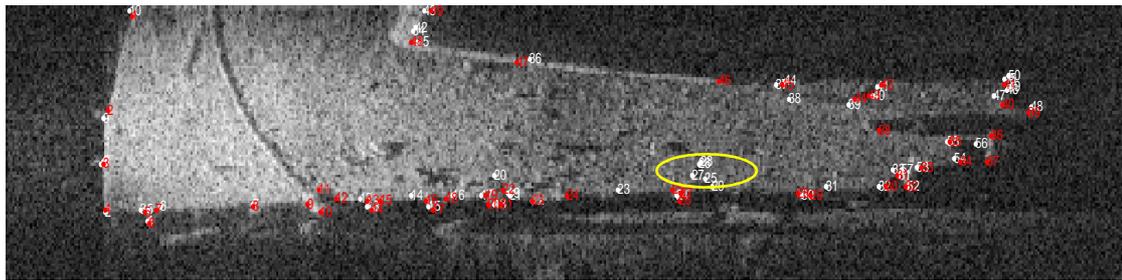
Tableau 2 . Moyenne et écart-type du taux de répétabilité pour l'ensemble de points détectés dans les trois séquences

	Moyenne %	Écart-type
Pied de pont	61,36	8,53
Épave	65,45	10,45
Piscine	73,09	8,45

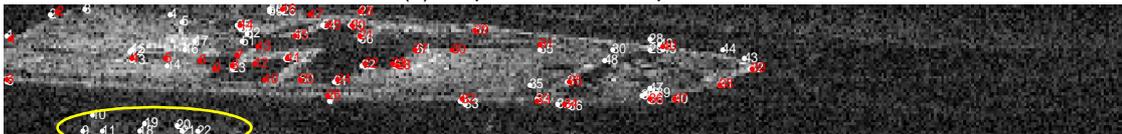
Selon les résultats présentés dans le Tableau 2, la moyenne de la répétabilité, enregistrée pour l'ensemble des points des différentes séquences d'images, est supérieure à 60 %. Bien que cette moyenne soit élevée puisque plus de 60 % de points en moyenne se répètent le long d'une séquence de 11 images, cela signifie qu'environ 40 % de la totalité des points ne se répètent pas d'une image à une autre. Ces 40 % de points non-répétables n'impactent pas la reconstruction 3D puisqu'ils sont dus à des détections fluctuantes au fil des images mais la géométrie sous-jacente reste présente sur toute la séquence d'images. Ces détections fluctuantes peuvent être engendrées par le changement de la position de la caméra par rapport à la scène et/ou l'impact du lissage des contours par le filtrage multi-échelle qui peut varier d'une image à une autre.

3.4.4.2 Bonne détection

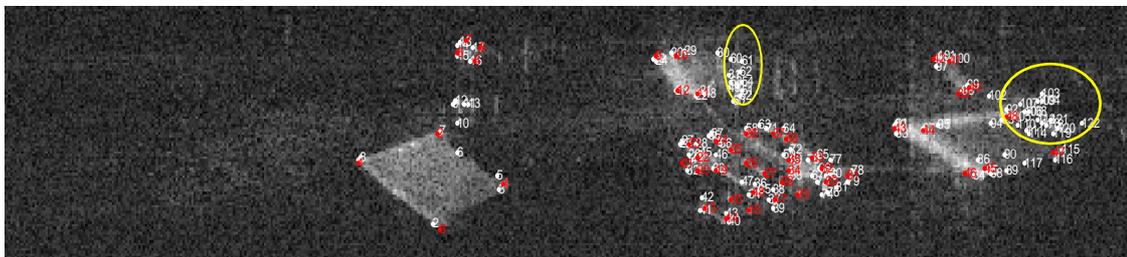
La bonne détection des points correspond à la détection de points qui caractérisent uniquement la géométrie de la scène. Pour vérifier ce critère, nous avons procédé à l'extraction de points de référence. Ces points sont choisis de manière à ce qu'ils soient représentatifs de la géométrie de la scène. Leur utilisation dans une reconstruction 3D permet d'obtenir une représentation 3D complète de la scène et fidèle à sa géométrie. Nous avons élaboré une étude comparative entre ces points que nous avons extraits manuellement (les points de référence) et les points identifiés par l'approche développée. Selon le nombre de points de référence détectés et ceux non détectés, on peut évaluer la performance de l'approche de détection de points proposée. Un point de référence n'est pas donc nécessairement détecté par l'algorithme. La Figure 23 montre les deux ensembles de points : les points rouges correspondent aux points de référence et les points blancs sont les points extraits automatiquement par notre approche.



(a) Séquence « Pied de pont »



(b) Séquence « Epave »



(c) Séquence « Piscine »

Figure 23 : Extraction de points saillants par l'approche développée (points blancs) et points caractéristiques de référence extraits manuellement (points rouges)

Afin de tester la fiabilité de l'algorithme de détection, nous présentons dans le Tableau 3 une étude comparative entre les points saillants extraits de manière automatique et ceux de référence extraits manuellement. Cette étude est basée sur l'évaluation des critères suivants :

1. Détection des « Vrais positifs » : ce sont les points extraits automatiquement et qui correspondent aux points de référence.
2. Détection des « Faux négatifs » : ce sont les points qui ne sont pas détectés automatiquement alors qu'ils sont bien présents en points de référence.
3. Détection des « Fausses alarmes » : ce sont les points détectés automatiquement alors qu'ils ne sont pas des points de référence (cf. Tableau 3).

Tableau 3. Étude quantitative de la bonne détection des points extraits par l'approche développée (automatiquement);

Image	Nombre de points de référence (R)	Nombre de points détectés par l'approche (A)	Ratio du nombre de vrais positifs par rapport à R (%)	Ratio du nombre de vrais positifs par rapport à A (%)	Ratio du nombre de faux négatifs par rapport à R (%)	Ratio du nombre de fausses alarmes par rapport à A (%)
Pied de pont	49	57	81,63	70,17	18,63	29,82
Épave	40	65	77,5	47,69	22,5	52,30
Piscine	48	122	91,67	36,06	8,33	63,93

Selon les résultats du Tableau 3, pour les trois jeux de données, le nombre de points extraits automatiquement rassemble à la fois des points pertinents qui correspondent aux points de référence et des points qui sont des fausses alarmes. En fait, les points de référence sélectionnés dans la présente étape correspondent uniquement aux points qui caractérisent la géométrie des objets de la scène (ex. les coins et les points de déformation de la forme). Donc à travers l'évaluation du pourcentage des fausses alarmes pour chaque jeu de données, on déduit la robustesse de l'algorithme dans la détection des points saillants pertinents. Le ratio des fausses alarmes est calculé en évaluant en termes de pourcentage le nombre points détectés automatiquement comme points saillants alors qu'ils ne sont pas des points de référence par rapport à l'ensemble de points détectés automatiquement. Par exemple, pour la séquence « Pied de pont » sur l'ensemble de 57 points détectés automatiquement, 17 points n'ont pas de points correspondants dans l'ensemble de points de référence ce qui correspond à 29,82 % de points détectés automatiquement.

En observant les résultats du Tableau 3, on remarque qu'au niveau des séquences « Piscine » et « Épave », le nombre de fausses alarmes est supérieur au nombre de vrais positifs. Elles sont respectivement de l'ordre de 63,93 % et 52,30% des points extraits automatiquement. Pour la séquence « Pied de pont », le nombre de fausses alarmes est bien inférieur au nombre de vrais positifs puisqu'elles représentent 29,82% des points extraits automatiquement. Ces valeurs nous informent que l'algorithme de détection présente une sensibilité élevée aux fausses alarmes dans les séquences « Piscine » et « Épave » par rapport à la séquence « Pied de pont ». Néanmoins, en étudiant la localisation des points extraits automatiquement dans les images, on constate qu'une forte concentration des fausses alarmes se trouve dans les zones d'ombre provoquées par les objets telles que les zones encerclées dans les images (a et c) de la Figure 23. Une partie des fausses

alarmes est également associée à des variations locales du relief du fond de la piscine ou à des points de contours jugés inutiles pour la reconstruction des objets tels que ceux détectés sur l'image (cf. Figure 23.b). Une solution possible pour réduire ces points, est de procéder à la segmentation des images telle que proposé dans (Brahim *et al*, 2008, Brahim, 2009). Brahim *et al* (2008, 2009) ont proposé une approche de détection d'objets sur des infrastructures sous-marines dans le contexte de l'évaluation du potentiel des caméras acoustiques dans l'inspection des infrastructures sous-marines. L'approche proposée se base sur l'algorithme de segmentation Fuzzy C-means. Elle a permis de séparer les zones d'ombre des zones d'écho d'objets ce qui offre la possibilité d'extraire uniquement les régions d'intérêt pour la reconstruction 3D et d'éviter le problème de points saillants supplémentaires. De telles approches peuvent être exploitées pour réduire les fausses alarmes.

D'autre part en évaluant le pourcentage des vrais positifs et des faux négatifs par rapport aux points de référence, on déduit que l'algorithme détecte un grand nombre de points de référence (séquence « Pied de pont » 81,63%, séquence « Épave » 77,5 %, et séquence « Piscine » 91,66 % des points de référence). En fait, pour les trois jeux de données, le nombre de faux négatifs est faible. Ils sont de l'ordre de 18,63 % pour la séquence « Pied de Pont », 22,5 % pour la séquence « Épave » et 8,33 % pour la séquence « Piscine ». L'algorithme de détection détecte ainsi la majorité des points de référence pour les trois séquences. Il reconnaît ainsi les points saillants pertinents pour la reconstruction 3D ce qui constitue un avantage pour l'approche proposée.

Toutefois, l'inconvénient de cet algorithme est le nombre élevé de fausses alarmes détectées parmi les points extraits automatiquement. En fait, l'approche de détection de points saillants s'appuie sur un seuil de sélection de contours (Paragraphe 1.3.1.2) fixé dans les expérimentations réalisées dans la section précédente à 40 %. Selon les résultats de l'évaluation de la bonne détection de l'algorithme, cette valeur a engendré la détection de fausses alarmes. Or, la modification de ce seuil peut engendrer deux résultats : si la valeur du seuil est supérieure à 40 %, la sélection de points saillants par l'algorithme sera plus stricte, le nombre de fausses alarmes diminuera mais aussi le nombre de points saillants puisque certains contours vont être éliminés; si la valeur du seuil est inférieure à 40%, le nombre de fausses alarmes augmentera et l'algorithme détectera plus de vrais positifs parmi les points. Néanmoins, une reconstruction 3D fiable nécessite la sélection et l'utilisation de points saillants robustes. Pour ce faire, l'utilisation d'un seuil supérieur à 40 % qui permet d'éliminer le maximum de fausses alarmes et de détecter un ensemble de points saillants qui soit robustes et caractéristiques de la géométrie de la scène s'avère plus adéquat. Par contre, si l'augmentation de ce seuil engendre non seulement l'élimination de fausses alarmes mais aussi la disparition de quelques points saillants caractéristiques de la géométrie de la scène alors choisir un seuil égal à 40% ou moins et envisager par la

suite des approches de sélection et d'extraction de points saillants parmi l'ensemble de tous les points détectés serait une alternative pertinente.

3.4.4.3 Bonne localisation

Afin d'évaluer la précision de la localisation des points saillants extraits par l'approche proposée, nous avons utilisé les points identifiés dans le paragraphe précédent comme « vrai positif ». Ils appartiennent simultanément à l'ensemble des points saillants extraits automatiquement et à l'ensemble des points de référence. Ces points de référence et les points saillants sont extraits de la même image de sorte que l'évaluation de l'erreur de la localisation s'effectue en comparant la localisation de chaque paire de points homologues. Ces points sont répartis sur l'ensemble de l'image: on trouve ainsi des points en début, au milieu et en fin de portée. Le nombre de points ainsi sélectionnés correspond respectivement à 40 points (séquence de « Pied de pont »), 31 points (séquence « Épave ») et 44 points (séquence « Piscine »).

L'erreur de localisation consiste à mesurer pour chaque paire de points, l'erreur en portée et en azimut entre le point de référence et le point extrait automatiquement. L'erreur est initialement calculée en pixel. Elle est affichée dans le Tableau 4.

Tableau 4. Évaluation de la localisation des points saillants en comparaison avec les points de référence (distance minimale, distance maximale, moyenne des distances et écart-type).

Images	En portée (pixel)				En azimut (pixel)			
	min	max	moy	sigma	min	max	moy	sigma
Pied de pont	0,0	2,0	0,82	0,71	0,0	4,0	1,82	1,03
Épave	0,0	3,0	0,70	0,73	0,0	4,0	0,70	0,93
Piscine	0,0	3,0	0,96	0,72	0,0	6,0	1,66	1,19

Les erreurs maximales et moyennes les plus élevées pour les trois séquences sont observées en particulier au niveau de l'azimut. En fait, en azimut, l'erreur maximale est de l'ordre de 4 pixels pour la séquence « Pied de pont », 4 pixels pour la séquence « Épave » et 6 pixels pour la séquence « Piscine » avec une erreur moyenne respectivement de l'ordre de 1,82, 0,70 et 1,66 pixels. Ces valeurs sont presque le double de l'erreur calculée en portée. En fait en portée, l'erreur moyenne pour les trois séquences est inférieure à un pixel. L'erreur maximale est de l'ordre de 2 pixels pour la séquence « Pied de pont » et 3 pixels pour les deux autres séquences. Ceci nous mène à conclure que les erreurs en termes de localisation des points saillants sont concentrées au niveau de l'azimut. Afin d'estimer l'impact que ces erreurs vont avoir sur la précision de la reconstruction 3D de la scène, il s'avère ainsi important de les évaluer dans la scène. Pour ce faire, les points sont extraits en coordonnées images puis transformés en coordonnées sphériques, permettant d'exprimer

l'erreur en portée en une distance métrique (exprimée en centimètres) et l'erreur en azimut en un écart en angle (exprimé en degrés). La connaissance de cette imprécision servira d'ailleurs de connaissance *a priori* associée à la reconstruction 3D de la scène. En fait, l'intégration de cette imprécision en portée et en azimut dans l'algorithme de reconstruction 3D permet de cerner la taille de la région à laquelle appartient un point saillant dans la scène. Le tableau ci-dessous affiche les erreurs calculées en portée (cm) et en azimut (degrés) à partir des coordonnées de points de référence et celles des points extraits automatiquement.

Tableau 5. Évaluation de la localisation des points saillants en comparaison avec les points de référence (distance minimale, distance maximale, moyenne des distances et écart-type).

Images	En portée (cm)				En azimut (°)			
	min	max	moy	sigma	min	max	moy	sigma
Pied de pont	0,0	6,99	3,19	1,81	0,0	1,2	0,54	0,30
Épave	0,0	13,99	2,48	3,28	0,4	2,8	0,82	0,95
Piscine	0,0	5,86	0,96	0,72	0,0	1,8	0,49	0,35

Le Tableau 5 montre que la séquence « Épave » se distingue par une erreur moyenne élevée en azimut (0,82°) par rapport aux deux autres séquences par contre c'est la séquence « Pied de pont » qui a erreur moyenne élevée en portée. En fait, la séquence « Épave » est acquise en mode basse fréquence contrairement aux séquences « Pied de pont » et « Piscine » qui sont acquises en mode haute fréquence. Le mode d'acquisition des images influence la précision de l'extraction des points puisque la résolution des images en azimut et en portée est différente. En azimut, pour le même secteur angulaire de l'ensemble des faisceaux (29°), l'ouverture d'un faisceau en basse fréquence est de 0,4 degrés alors qu'en haute fréquence elle est de 0,3 degrés. Quant à la résolution en portée, elle dépend des caractéristiques d'acquisition telles que la portée maximale et le retard de l'enregistrement des images. Les résultats obtenus mettent en évidence que l'extraction de points des séquences acquises en mode haute fréquence présente une meilleure localisation en azimut par rapport à celle acquise en mode basse fréquence (Épave).

En conclusion, si l'on ne s'intéresse qu'aux facteurs relatifs au traitement des images pour évaluer l'approche d'extraction, alors les résultats proposés dans le Tableau 4 mettent en évidence que l'erreur en azimut est plus importante que celle en portée.

3.4.5 Discussion des résultats

L'évaluation des trois critères de qualité : la répétabilité, la bonne détection et la bonne localisation ont permis de conclure que les résultats de l'extraction de points saillants des trois types de séquences sont satisfaisants. En effet, les points saillants détectés à partir des « coins » des contours extraits non seulement caractérisent la forme des objets de la scène mais sont aussi détectables sur plusieurs images successives. Les

observations supplémentaires suivantes ont été notées à propos des résultats obtenus au niveau de chaque étape :

3.4.5.1 Observations concernant la détection de contours

Les contours extraits par le filtre de Canny représentent les objets de la scène. Mais en plus de ces contours, de nombreux faux contours dus au bruit présents dans les images ont été détectés. Ceci constitue un inconvénient pour l'extraction de points saillants dans les prochaines étapes. La mise en place de l'étape d'analyse multi-échelle par une pyramide gaussienne était une solution performante en ce qui a trait à l'élimination de ces faux contours.

3.4.5.2 Observations concernant l'analyse multi-échelle

L'analyse multi-échelle a contribué fortement à l'élimination de faux contours dans toutes les séquences utilisées. Dans les images où les contours ont été détectés sans passer par une analyse multi-échelle, la mise en correspondance de ces contours n'était pas suffisante pour éliminer tous les faux contours même en augmentant le seuil de sélection des contours appariés. L'ajout d'un filtre gaussien pyramidal selon des tailles d'échelle croissante a permis de conserver les contours pertinents qui proviennent réellement d'une structure dans l'image. Ces traitements s'adaptent aux séquences d'images utilisées en s'ajustant automatiquement en fonction des images analysées. En effet, pour les images acquises en milieu marin, les itérations combinant un filtrage gaussien et une mise en correspondance ont donné le meilleur résultat au bout de la 8^{ème} itération. À ce niveau le nombre de contours devient approximativement stable pour quelques itérations avant de commencer à diminuer rapidement de nouveau. Par exemple, pour la séquence « Pied de pont » et la séquence « Épave » c'est respectivement à la 8^{ème} itération et à la 6^{ème} itération que le nombre des contours devient stable. Sachant que la séquence « Pied de pont » et celle « Épave » sont différentes dans leurs contenus, ceci explique que le nombre d'itérations soit différent pour ces deux séquences. Quant aux images acquises en piscine le nombre d'itérations est encore plus petit. Il est logiquement plus faible dans un bassin intérieur où le milieu de propagation des ondes acoustiques est moins bruité. Il est égal à 3 itérations. Dès que la condition de la stabilité du nombre de contours est atteinte, les itérations de l'analyse multi-échelle s'arrêtent indiquant ainsi la sélection des contours pertinents pour chaque séquence.

L'approche ainsi développée est une approche qui s'adapte en fonction des caractéristiques de chaque séquence utilisée. En effet, elle identifie automatiquement l'échelle d'analyse **E** appropriée pour chaque séquence afin de permettre à l'étape suivante d'assurer une détection robuste des points saillants des contours.

Il faut noter que le filtrage des contours par leur mise en correspondance n'est efficace que lorsqu'il s'agit d'images très proches dans la séquence, en particulier des images successives. En fait, dans le présent travail

le déplacement entre deux images successives varie de 1 à 3 pixels ce qui justifie la dimension du voisinage choisi (matrice 3*3) dans l'algorithme de mise en correspondance des contours dans une paire d'images (une image de référence et une 2^{ème} image). Afin d'adapter l'algorithme à tout type de déplacement entre les images utilisées, la solution d'un voisinage variable en fonction du déplacement de la caméra devrait être utilisée. Pour ce faire, une étape d'estimation du déplacement de la caméra s'avère nécessaire. En traitant la séquence par paire d'images successives ou par paire d'images pour lesquelles le déplacement de la caméra est faible (de l'ordre de quelques pixels), il est aisé et plus rapide de retrouver les mêmes contours pertinents qui caractérisent les objets de la scène. De ces contours pertinents, il est possible d'extraire des points saillants pertinents et fiables pour une reconstruction 3D. Bien que cette dernière implique probablement l'utilisation d'images acquises selon des points de vue suffisamment différents, elle nécessite avant tout des points saillants pertinents et robustes. L'approche proposée est ainsi fondamentale pour garantir la qualité et la pertinence des points saillants.

3.4.5.3 Observations concernant la détection de points saillants

Selon les tests d'évaluation, les points détectés caractérisent bien la forme des objets en marquant les coins et les points d'intersection des contours. Plus de 77 % des points saillants de référence ont été détectés par l'algorithme avec une moyenne d'erreur de localisation dans l'image de l'ordre de 2 à 6,2 cm en portée et aux alentours de 0,3 à 06 degré en azimut. Ces points se caractérisent aussi par un taux de répétabilité intéressant. Ils se retrouvent avec une moyenne de répétabilité de l'ordre de 61 % dans la séquence « Pied de pont », de 65 % dans la séquence « Épave » et de 73 % dans la séquence « Piscine » sur une série de 11 images. Même si d'après ces résultats, tous les points saillants de référence ne sont pas totalement détectés et répétables le long d'une séquence d'images, ceci n'empêche pas leur exploitation dans la reconstruction tridimensionnelle de la scène.

3.5 La mise en correspondance des points saillants pour une reconstruction 3D

Ayant obtenu les points saillants de différentes images, on peut aborder l'étape de mise en correspondance. Elle permet d'identifier les points homologues à partir des différents ensembles de points extraits d'images multi-vues. Contrairement aux contours mis en correspondance dans la section précédente, ces points saillants appartiennent à des images non successives d'une même séquence avec des vues différentes de la scène ou à des images de séquences différentes acquises de différents angles de vue.

L'étape de la mise en correspondance de points est nécessaire dans la démarche de reconstruction 3D afin de s'appuyer sur des ensembles de points correspondants dont on est sûr de leur exactitude et de leur robustesse dans la représentation du même élément de la scène dans des images différentes. En fait de

simples tests d'appariement réalisés dans le présent travail (présentés dans les paragraphes qui suivent) ont montré la détection de fausses paires de points correspondants d'où la nécessité de la mise en place d'approches d'appariement plus pertinentes. Il existe plusieurs approches possibles dont celles qui s'appuient sur l'information locale des images (ex. intensité, texture, gradient, etc.), celles basées sur la comparaison des caractéristiques géométriques telles que la position, la forme, l'orientation et la taille des objets, celles qui combinent ces deux informations (intensité et géométrie), celles opérant sur l'information globale des images (ex. programmation dynamique (Wang *et al*, 2011), etc. Le choix d'un type d'appariement est tributaire de la nature de l'application, la forme des objets et en particulier la nature des images.

L'étape de mise en correspondance peut être effectuée en s'appuyant sur des approches impliquant l'exploitation de connaissances *a priori*, en appliquant des approches classiques d'appariement géométrique et aussi par l'application d'approches simples qui se basent uniquement sur l'intensité des images pour l'identification de points correspondants. Afin d'évaluer la faisabilité de ces approches simples sur des points extraits des images acoustiques, nous avons réalisés des tests. Ceux-ci se basent sur deux étapes classiques d'appariement :

- 1) Extraction, dans chaque image, des points saillants et de leurs descripteurs associés représentant la variation de l'intensité dans leur voisinage. Le voisinage est défini par une fenêtre de taille limitée centrée sur le point saillant.
- 2) Mise en correspondance des points saillants à l'aide d'une mesure de (dis-)similarité entre les descripteurs.

Une première façon de procéder est de comparer des fenêtres carrées centrées sur chaque point, par corrélation. Cependant, la corrélation ne prend pas en compte les déformations liées aux changements de point de vue des images. De plus, la simple mesure de corrélation entre deux voisinages de points n'est pas invariante au changement d'intensité entre les images. Il existe des versions plus élaborées centrées et normalisées qui sont plus performantes. La corrélation centrée et normalisée C entre deux fenêtres A et B de même taille de n pixels et d'intensité moyenne \bar{A} et \bar{B} s'exprime de la façon suivante :

$$C = \frac{\sum_{i,j=1}^n (A_{i,j} - \bar{A})(B_{i,j} - \bar{B})}{\sqrt{\sum_{i,j=1}^n (A_{i,j} - \bar{A})^2 (B_{i,j} - \bar{B})^2}} \quad (10)$$

Avec $A_{i,j}$ l'intensité du pixel (i, j) dans la fenêtre A et $B_{i,j}$ l'intensité du pixel (i, j) dans la fenêtre B .

En s'appuyant sur cette caractéristique pour la mise en correspondance des points saillants, nous avons obtenu les résultats suivants :

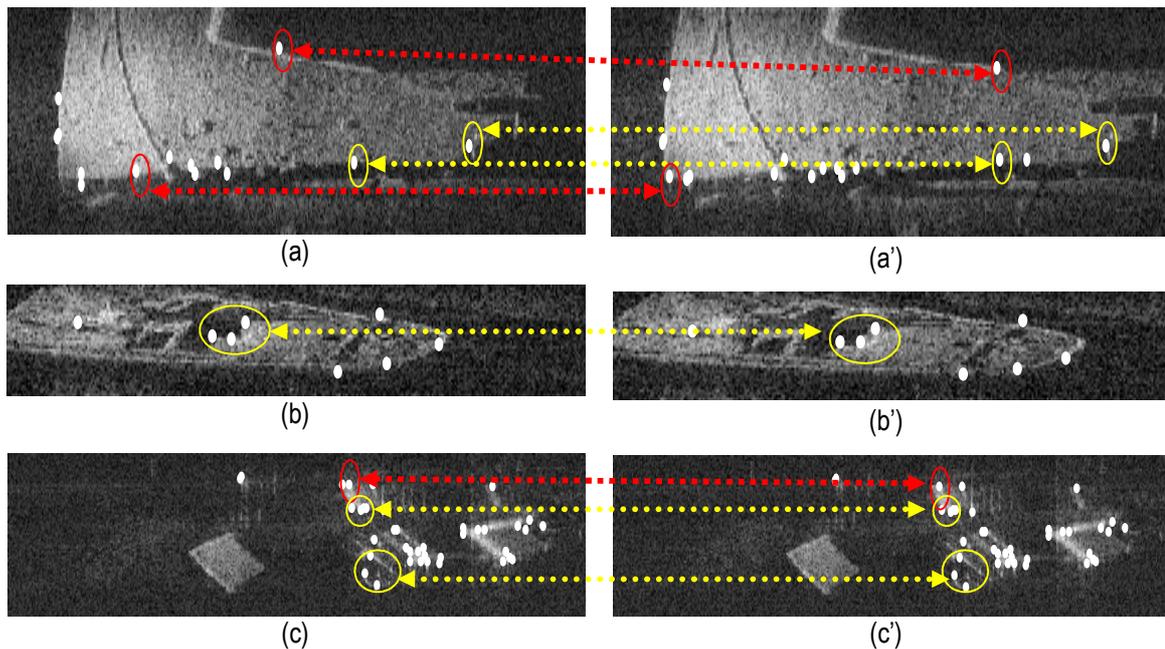


Figure 24 : Extraction de points saillants correspondants par corrélation normalisée : (a, a') exemple de points correspondants d'une paire d'images i et $i+1$ extraites de la séquence du « Pied de pont »; (b, b') exemple de points correspondants d'une paire d'images i et $i+1$ extraites de la séquence « Epave»; (c, c') exemple de points correspondants d'une paire d'images i et $i+1$ extraites de la séquence de la piscine.

Les résultats obtenus en appliquant la corrélation normalisée montrent une détection correcte de points représentant la même structure géométrique dans la scène (cf. Figure 24, points encadrés en jaune). Néanmoins, on remarque également parmi ces points, l'identification de certaines fausses paires de points correspondants alors qu'ils ne représentent pas la même structure dans la scène (cf. Figure 24 : points encadrés en rouge). Cette approche pourrait être améliorée en ajoutant plusieurs descripteurs dont notamment des descripteurs géométriques. Il s'agit de l'application d'un appariement géométrique en cherchant à identifier pour chaque point, ses caractéristiques géométriques telles que sa position par rapport aux autres points, son appartenance à un objet, un contour, etc. ou même l'exploitation des descripteurs de points disponibles tels que les descripteurs de SIFT ou de SURF.

Des travaux portant sur l'appariement géométrique ont proposé des approches performantes telles que le travail de Daniel (1998) portant sur la mise en correspondance des objets de différentes images sonar en s'appuyant sur la conservation de leurs positions relatives d'une image à une autre et sur l'association des objets voisins pour les identifier. En fait, le problème d'appariement est résolu par l'analyse des ombres et la disposition de trois objets voisins et à travers un raisonnement hypothétique sous la forme d'un arbre de décision (Daniel *et al*, 1998). Bouziani et Pouliot (2004) ont analysé des opérateurs d'appariement géométrique d'objets géographiques linéaires et ont montré que l'analyse de la position des points constituant

ces objets est un traitement de base des opérateurs d'appariement géométrique. Il s'agit, en effet, de définir les points saillants des lignes (ex. contours) comme vertex et nœuds et d'identifier les lignes correspondantes entre images en comparant les caractéristiques géométriques de leurs nœuds et vertex (ex. position, voisinage, etc.). Plusieurs autres approches ont été abordées dans la littérature offrant des solutions pertinentes pour exploiter les connaissances géométriques dont on dispose sur les points saillants extraits et effectuer l'étape de mise en correspondance.

L'étape de mise en correspondance ne constitue donc pas une problématique de recherche dans notre travail mais plutôt une étape de développement et de mise en œuvre d'approches déjà disponibles. Ainsi, afin de pouvoir investir le temps restant disponible et les efforts dans la conception et le développement de l'approche de reconstruction 3D multi-vues, les points saillants impliqués dans cette reconstruction ont été appariés manuellement. Ces points seront exploités pour l'ensemble des tests réalisés pour les expérimentations.

3.6 Synopsis du chapitre 3

Ce chapitre a présenté une revue de différentes primitives qui peuvent caractériser les objets d'une scène. En s'appuyant sur l'étude des caractéristiques des différentes primitives et sur les caractéristiques des images acoustiques, nous avons opté dans le cadre de ce travail, pour la sélection des primitives ponctuelles. Ces dernières sont plus adaptées au contexte de la reconstruction 3D multi-vues, en vue d'obtenir des points d'ancrage 3D dans la scène.

Afin d'extraire ces primitives ponctuelles, nous avons présenté une approche qui s'appuie sur la détection de contours pertinents qui caractérisent uniquement la structure géométrique de la scène. Pour ce faire, une approche itérative qui permet de supprimer les faux contours issus du bruit des images a été mise en œuvre. Elle se base sur une analyse multi-échelle des images à travers une pyramide gaussienne de filtrage où à chaque itération l'image est filtrée, les contours sont détectés par le filtre de Canny puis mis en correspondance et comptés. C'est l'étape de mise en correspondance des contours qui joue un rôle primordiale dans la sélection des contours pertinents. En tenant compte du déplacement de la caméra entre deux images successives, la mise en correspondance des contours est réalisée en s'appuyant sur la recherche de contours correspondants selon leur localisation dans l'image de référence et en considérant le déplacement de la caméra pour les retrouver dans la deuxième image. Ces étapes (filtrage, détection de contours et mise en correspondance) prennent fin **automatiquement** lorsque le nombre de contours devient stable au travers de l'analyse multi-échelle indiquant ainsi la détection des contours pertinents de l'image. De ces contours sont extraits les points saillants qui sont les points qui se caractérisent par une forte courbure. Selon les critères d'évaluation de l'approche (répétabilité, bonne détection et localisation), les résultats de détection de points sont satisfaisants.

L'approche d'extraction de primitives est proposée afin d'extraire des points saillants exploitables dans une approche de reconstruction 3D multi-vues. Cette approche se caractérise en particulier, par une auto-adaptation à la nature des séquences d'images utilisées (ex. qualité d'images, contenu de la scène) et extrait des points représentatifs de la géométrie de la scène dont la robustesse est indispensable pour une reconstruction 3D fiable.

En revanche, cette approche repose sur l'exploitation d'images consécutives présentant de faibles variations géométriques dues à la différence de points de vue. Dans le cadre d'une reconstruction 3D de scène, il semble clair que les points de vue pour remonter aux informations 3D manquantes, devront probablement être plus importants afin de compenser les incertitudes mentionnées précédemment (résolution de l'acquisition, erreurs de détection, ...). Ainsi, l'étape décrite dans ce chapitre permet l'extraction fondamentale de points saillants robustes qu'il faudra ensuite « tracker » (poursuivre) entre les images ou entre les séquences d'images pour disposer d'appariement impliquant des différences de points de vue plus importants. Ce tracking intra ou inter-séquences soulève des problématiques qui dépassent le cadre de cette thèse; c'est pourquoi les reconstructions ultérieures se baseront sur des points saillants robustes extraits par l'approche décrite mais dont les appariements inter-séquences (pour bénéficier de points de vue différents) seront réalisés manuellement. En effet, l'objectif de cette thèse est tout d'abord de vérifier que l'on dispose de suffisamment d'informations avec simplement les images acquises pour déjà permettre de remonter aux informations 3D manquantes.

Ainsi, dans le prochain chapitre qui est consacré à l'étape de la reconstruction 3D, nous avons donc réalisé un appariement manuel pour exploiter des points correspondants issus de séquences d'images différentes. Ce chapitre présentera l'approche de reconstruction 3D proposée, les expérimentations et les résultats obtenus.

Chapitre 4. Reconstruction 3D de scènes sous-marines à partir de séquences d'images acoustiques

Dans le chapitre précédent, nous avons présenté l'approche développée pour extraire des primitives ponctuelles depuis des images acquises par des caméras acoustiques. Ces primitives sont exploitées dans le présent chapitre afin de reconstruire la scène observée en 3D. Nous proposons ainsi deux approches de reconstruction 3D: une approche curviligne et une approche volumique. Ces deux approches utilisent le même algorithme d'optimisation et exploitent les mêmes primitives mais au travers de deux représentations différentes, afin de calculer les déplacements inter-images de la caméra pour ensuite remonter aux coordonnées 3D des points de la scène. Plusieurs expérimentations ont été réalisées pour évaluer et comparer ces approches.

La première section de ce chapitre est consacrée à la présentation de l'algorithme d'optimisation retenu. Cette section est suivie par la description des approches curviligne et volumique ainsi que les expérimentations réalisées pour les évaluer.

4.1 Optimisation & algorithmes évolutionnaires

La reconstruction 3D multi-vues de scènes sous-marines à partir de séquences d'images acoustiques est clairement un problème difficile et le recours aux algorithmes évolutionnaires résulte d'un choix argumenté par rapport à la méthodologie de résolution proposée.

4.1.1 Pourquoi se tourner vers les algorithmes évolutionnaires ?

Lors de l'acquisition de la séquence d'images, la caméra se déplace en enregistrant des images de la même scène, à partir de différents points de vue. Les points saillants qui caractérisent la géométrie de la scène ont été extraits de ces images. Ils représentent les points de forte courbure des contours des objets et vont être exploités lors de l'étape de la reconstruction de la scène observée.

L'étape de reconstruction 3D consiste alors à déterminer les paramètres du déplacement de la caméra qui 'expliquent' le déplacement des points saillants sur les images acquises, pour pouvoir ensuite estimer les coordonnées 3D (dans la scène) de chacun de ces points saillants. Par ailleurs, le calcul de ces inconnues (déplacement de la caméra et coordonnées 3D de chaque point saillant) implique l'utilisation du modèle de la caméra acoustique présenté dans l'équation (1) (Chapitre 1, paragraphe 2.2.2), qui formalise d'une part la projection des points de la scène dans les images acquises c'est-à-dire à quel pixel de l'image (en coordonnées polaires) contribue chaque point de la scène (en coordonnées cartésiennes) et d'autre part la

projection des points images (en coordonnées polaires) dans la scène (en coordonnées cartésiennes) c'est-à-dire de quelle région de la scène est issu chaque point image.

Ainsi, afin de calculer les coordonnées 3D des points extraits des images, dans la scène, on applique le modèle inverse de la caméra qui projette les points extraits des images $P(r, \theta)$ dans la scène $P(x, y, z)$. Ce modèle fournit pour chaque point saillant un système de trois équations, chaque équation représentant la projection du point en ses trois coordonnées respectivement en X, en Y et en Z. Ce système de trois équations admet 7 inconnues (i.e. les trois paramètres associés à la rotation de la caméra, les trois paramètres associés à la translation de la caméra et l'élévation du point considéré). Il est donc caractérisé par un grand nombre de degrés de liberté. Celui-ci croît au fur et à mesure qu'augmente le nombre d'images exploitées pour la reconstruction ainsi que le nombre de points pour caractériser la géométrie de la scène. En effet, par exemple, pour chaque point extrait d'une nouvelle image, s'ajoutent trois nouvelles équations qui caractérisent chacune de ses coordonnées (X, Y et Z) et qui incluent les 6 paramètres du mouvement (rotation et translation) qu'a effectué la caméra pour acquérir cette nouvelle image d'où l'augmentation du nombre de degré de liberté.

Par ailleurs, la complexité de ce système s'accroît aussi lorsqu'on cherche à prendre en considération la résolution de la caméra. En effet, lors de l'acquisition des images acoustiques, l'intensité de chaque pixel d'une image résulte de l'enregistrement des ondes réfléchies par la surface d'une région dans la scène, région comprise dans une plage en azimut ($\Delta\theta$) et une plage en portée (Δr). Par conséquent, la géométrie d'un pixel correspond en réalité à un polygone et non pas, de manière simplifiée, à un point (θ, r) . Si l'on considère que la géométrie d'un pixel est un simple point, alors cela induit que l'ensemble des points de la scène qui ont potentiellement contribué à la valeur d'intensité de ce pixel, se trouve sur un arc de cercle couvrant une plage en élévation, d'où la dénomination d'approche curviligne. En revanche, si chaque pixel correspond à un polygone, l'ensemble des points de la scène qui ont potentiellement contribué à la valeur de ce pixel, devient un ruban 3D possédant des dimensions en azimut, portée et élévation d'où la dénomination d'approche volumique. Ainsi, en fonction de l'approche, le système d'équations globalement représente soit l'intersection d'arcs de cercle, soit l'intersection de rubans 3D ce qui induit alors une complexité de représentation accrue.

Le recours aux algorithmes d'optimisation pour résoudre notre problématique de reconstruction 3D est, alors, une nécessité due au fait qu'il n'existe pas de solutions analytiques directes permettant de résoudre un système d'équations complexe avec un nombre élevé d'inconnues. En revanche, comme dans de nombreux problèmes complexes, nous disposons d'une fonction objective mesurant la qualité d'une solution potentielle. La minimisation de cette fonction objective jusqu'à l'optimum permet de déterminer les meilleures solutions au

problème. De ce fait, on fait face à un problème typique d'optimisation qui requiert le choix d'un algorithme d'optimisation robuste. En effet, vu la nature du modèle de la caméra et le contexte de la reconstruction 3D qui exige la recherche de solutions pertinentes, notre problème est hautement non-linéaire et possède un espace de recherche des solutions dont le nombre de dimensions est grand.

L'optimisation est actuellement un domaine de recherche en pleine évolution grâce au développement d'outils de modélisation, d'algorithmes et de moyens de calcul de plus en plus rapides et performants (Ejday, 2011). Les approches proposées dans la littérature sont nombreuses. Parmi ces approches un algorithme appartenant à la catégorie des algorithmes d'optimisation évolutionnaires a été appliqué dans le cadre du présent travail. Ce choix se justifie par la logique utilisée par de tels algorithmes dans le calcul de la solution optimale pour résoudre un problème donné. En effet, les algorithmes évolutionnaires sont une abstraction de la théorie de l'évolution. Ils traduisent le processus de l'évolution et de la sélection naturelle sur ordinateur, afin de chercher dans un laps de temps raisonnable une solution optimale et globale adaptée à un problème donné parmi un grand nombre de solutions possibles. Cette combinaison entre optimisation du temps de calcul et qualité de la solution calculée est la raison principale de notre choix de cette catégorie d'algorithmes d'optimisation. Plus de précisions et de détails sur le principe et les caractéristiques de ces algorithmes sont fournis dans ce chapitre.

4.1.2 Que sont les algorithmes évolutionnaires?

Les algorithmes évolutionnaires sont des méthodes stochastiques. Ils font partie des méthodes métaheuristiques¹⁴ qui consistent en une stratégie de choix pouvant piloter une ou plusieurs heuristiques¹⁵ de manière abstraite. Les méthodes métaheuristiques sont ainsi des méthodes itératives, à population de solutions puisqu'elles manipulent et maintiennent un ensemble de solutions au fil des itérations (Craciun, 2011).

Étant inspirés de la théorie de l'évolution, les algorithmes évolutionnaires mettent en œuvre les trois grands principes darwiniens à savoir l'hérédité, la variation et la sélection dans un objectif de survie de l'individu le mieux adapté (Ejday, 2011) au fil des générations. Ces trois principes se traduisent en trois étapes qui se répètent sur un certain nombre d'itérations ou bien jusqu'à l'obtention de solutions satisfaisantes (voire les meilleures) pour le problème en question. Ainsi, pour une itération (ou génération), ces trois étapes sont :

¹⁴ Méthodes métaheuristiques : des algorithmes d'optimisation stochastiques itératifs visant à résoudre des problèmes d'optimisation complexes.

¹⁵ Une heuristique : Une hypothèse visant à donner la bonne direction pour trouver une solution possible en faisant appel à des connaissances « expertes » sur le problème.

- 1) la sélection d'individus (parmi la population des solutions courantes) généralement avec un biais en faveur des meilleurs individus c'est-à-dire ceux représentant les meilleures solutions courantes du problème;
- 2) la variation de ces individus en faisant appel à des opérateurs de croisement et de mutation des individus;
- 3) l'évaluation des individus résultants.

La Figure 25 illustre les mécanismes d'évolution d'une population d'individus à travers les trois étapes principales d'un algorithme évolutionnaire, au fil des générations.

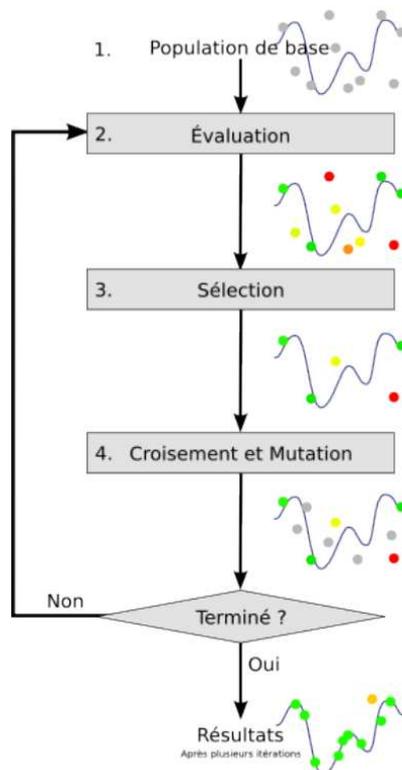


Figure 25 : Vue d'ensemble d'un algorithme évolutionnaire (Balmisse, 2002)

Les algorithmes évolutionnaires sont des algorithmes d'optimisation à caractère global, qui peuvent de plus, traiter des fonctions objectives non régulières ou des fonctions définies sur des espaces de recherche non standards. Cette souplesse permet à ces algorithmes de s'appliquer aux problèmes complexes où les algorithmes classiques d'optimisation, d'apprentissage ou de conception automatique sont incapables de produire des résultats satisfaisants. Ils sont essentiellement utilisés quand on fait face aux deux problématiques suivantes (Balmisse, 2002):

- 1) L'espace de recherche de solutions est très grand et plusieurs paramètres doivent être déterminés et optimisés simultanément.
- 2) Le problème ne peut être résolu avec une simple solution analytique.

Le caractère global est dû au fait que ces algorithmes déterminent l'optimum global de la fonction objective du problème en évitant les optima locaux (cf. Figure 26). Ce dernier point est le problème majeur des méthodes classiques et des méthodes heuristiques. De plus, ces méthodes évolutionnaires exploitent et explorent l'espace de recherche de solution d'une manière très efficace puisqu'elles font usage de l'expérience accumulée durant la recherche de l'optimum en déterminant les nouvelles solutions à chaque itération en fonction des solutions obtenues aux itérations précédentes, guidant ainsi le processus de recherche (Craciun, 2011).

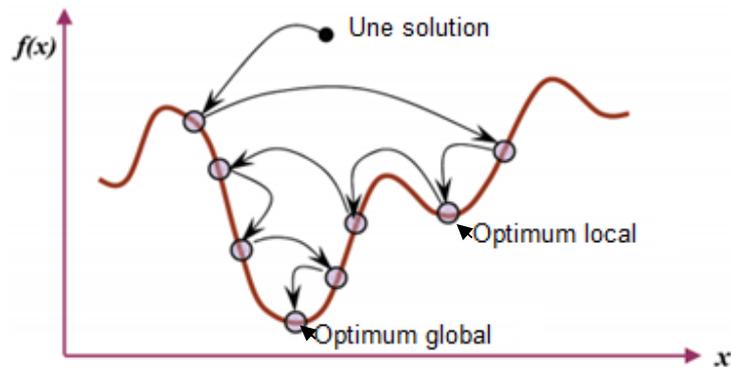


Figure 26 : Représentation de la recherche d'une solution pour une fonction objective $f(x)$ (Alkhatib, 2008). La figure illustre un parcours de recherche dans lequel un optimum local est rencontré mais l'exploration de l'espace des solutions continuant, un meilleur optimum est ensuite rencontré (en l'occurrence l'optimum global).

Une caractéristique importante des algorithmes évolutionnaires est qu'ils ne requièrent pas obligatoirement le calcul du gradient de la fonction objective pour élaborer de nouvelles solutions à la différence de l'algorithme de descente de gradient (Klein *et al*, 2009) et de l'algorithme Levenberg-Marquardt (Hanke, 1997). Cette propriété confère aux algorithmes évolutionnaires une flexibilité supplémentaire leur permettant d'être utilisés dans une grande variété de problèmes tels que les problèmes fortement non-linéaires ou les problèmes non-continus.

Grâce à leurs nombreux avantages, les algorithmes évolutionnaires sont appliqués avec beaucoup de succès dans le domaine industriel et l'ingénierie. De nombreux problèmes liés aux systèmes électriques sont résolus par les algorithmes évolutionnaires telles que les études de sûreté des systèmes électriques (Alkhatib, 2010),

la compression et la reconstruction des images (Moore et Babb, 2006), le recalage des images médicales (Jenkinson et Smith, 2001) etc.

De ces algorithmes évolutionnaires sont issues historiquement quatre familles d'optimisation stochastique : les algorithmes génétiques, la programmation génétique, les stratégies d'évolution et la programmation évolutionnaire (Gagné, 2005, Craciun, 2011).

Ce travail s'intéresse tout particulièrement aux stratégies d'évolution pour les raisons suivantes :

- 1) Ces méthodes ont été appliquées avec succès pour résoudre divers problèmes d'optimisation : multimodaux, non continus, contraints, multi-objectifs, etc. (Berro, 2001);
- 2) Elles sont flexibles vis-à-vis de l'intégration de contraintes et de la prise en compte de certains critères de recherche;
- 3) Comparativement à d'autres méthodes d'optimisation dont notamment les algorithmes génétiques, elles sont rapides et moins coûteuses en termes de temps de calcul (Le Riche *et al*, 2007).

4.1.3 Les stratégies d'évolution

Les approches de type stratégies d'évolution (SE) ont été proposées par Ingo Rechenberg en 1965 (Rechenberg, 1965) à l'Université Technique de Berlin, en Allemagne, puis développées en collaboration avec Hans-Paul Schwefel. Elles constituent un ensemble d'algorithmes évolutionnaires particulièrement adaptés à l'optimisation de fonctions avec des paramètres à valeur réelle (Gagné, 2005). En fait, les SE représentent chaque individu (ou solution au problème) par un vecteur de nombres réels, de dimension fixe n . Chaque composante du vecteur (nombre réel) correspond donc à une caractéristique (ou une dimension) de la solution potentielle représentée. L'ensemble des individus forme une population, dont le contenu change de génération en génération (au fil du temps). Ainsi, pour produire une nouvelle génération, les SE appliquent les trois opérateurs suivants à une population dite population de parents: la sélection, le croisement et la mutation. Ces opérations agissent d'une façon séquentielle selon l'ordre suivant :

- 1) Sélection : de la population initiale dite parent de taille $\mu \geq 1$, l'algorithme génère aléatoirement des individus formant une population d'enfants de taille λ . Pour engendrer une nouvelle génération de parents, les μ meilleurs individus sont choisis soit parmi les λ enfants (approche (μ, λ) où $\lambda > \mu$), soit parmi les μ parents et les λ enfants (approche $(\mu + \lambda)$).
- 2) Croisement : c'est une opération facultative pour les SE. L'objectif général de cet opérateur est de recombinaison des caractéristiques des individus (la nouvelle génération de parents) sélectionnés dans

l'opération précédente (la sélection) pour créer de nouveaux individus. Contrairement aux algorithmes génétiques, le croisement peut utiliser plus de deux individus parents choisis aléatoirement pour créer un seul individu enfant (Ejday, 2011). Lorsque l'opérateur croisement n'est pas utilisé dans l'algorithme d'optimisation, les individus qui seront exploités par l'opérateur suivant (la mutation) sont la copie des individus sélectionnés à l'opération de sélection.

- 3) Mutation : la particularité des algorithmes des stratégies d'évolution est que lors de l'opération de mutation, la génération des enfants se fait par une mutation gaussienne adaptative des parents. Elle consiste à additionner aux enfants (les individus issus du croisement) des valeurs générées aléatoirement selon une fonction de densité de probabilité paramétrée (Gagné, 2005). L'originalité de ces méthodes par rapport aux méthodes génétiques réside dans l'utilisation automatique du codage réel et de la mutation gaussienne. La mutation gaussienne sera détaillée dans le prochain paragraphe. Quant au codage, il correspond à l'approche utilisée pour la modélisation des individus. Ces individus doivent être « codés » pour que l'algorithme d'optimisation puisse appliquer les différentes étapes de traitement. Cette représentation codée peut consister en une variable ou un élément de la solution. La manière la plus utilisée de codage par algorithme génétique est le codage en vecteurs. Chaque individu est ainsi représenté par un vecteur. Ce vecteur peut être binaire où le codage s'appuie sur les deux éléments {0,1} pour coder les individus. Ce codage est simple et pratique pour la résolution d'un certain nombre de problèmes mais il reste limité et non performant pour d'autres problèmes d'optimisation numérique. Le codage réel utilisé par les SE, où chaque individu est représenté par une série de valeurs réelles quelconques est désormais le plus utilisé, notamment dans les domaines applicatifs pour l'optimisation de problèmes à variables réelles.

En conclusion, les algorithmes SE améliorent les algorithmes génétiques, notamment au niveau de l'opération de mutation. Plus de précisions sur la mutation gaussienne, ses paramètres et l'algorithme évolutionnaire SE-AMC sont fournies dans les paragraphes suivants.

4.1.3.1 La mutation gaussienne

La mutation gaussienne est l'opérateur de mutation principal des SE. Cet opérateur modifie la population en ajoutant du bruit normal centré. Pour ce faire, les algorithmes SE s'appuient sur la distribution gaussienne multivariée de la population $N(m, C)$, où m correspond à la moyenne ou à l'individu central (point central) de la population et C correspond à la matrice de covariance (matrice symétrique définie positive) qui caractérise la forme et l'orientation de la distribution de la population.

Plus simplement, pour un individu x , la mutation gaussienne appliquée à celui-ci s'écrit sous la forme suivante et traduit un déplacement de l'individu concerné dans l'espace de recherche selon une direction définie par la distribution gaussienne multivariée:

$$x \rightarrow x + \sigma N(0, C) \quad (11)$$

où σ est un paramètre positif qui correspond au pas de la mutation dans les directions principales de la distribution donnée par la matrice de covariance C . Ces deux grandeurs à savoir le pas de la mutation σ et la matrice de covariance C , qui définissent les paramètres de l'opérateur de mutation, évoluent en fonction de l'avancement de la recherche de solutions, au fil des générations.

Ajuster les paramètres d'un algorithme SE revient alors à rechercher les meilleures valeurs pour le pas et la matrice de covariance au cours de l'évolution des individus de la population (Hansen, 2011). Pour ce faire, les algorithmes SE utilisent un mécanisme d'adaptation du pas de la mutation ainsi que de la matrice de covariance dont le principe est abordé dans les paragraphes qui suivent.

4.1.3.2 Les paramètres de la mutation gaussienne

Le pas de la mutation et la matrice de covariance jouent un rôle très important dans le calcul des meilleures solutions. Hansen (2011) et Le Riche *et al* (2007) ont bien expliqué l'utilité du pas dans la sélection des meilleurs individus. En fait, au fil des générations, le pas, en progressant dans la direction définie par la matrice de covariance, décroît. La distance moyenne entre deux individus successifs est de ce fait proportionnelle au pas de mutation. Plus l'algorithme s'approche des meilleures solutions d'une génération à une autre, plus le pas devient petit.

Les algorithmes des SE utilisent la mutation auto-adaptative. Elle se caractérise par le fait que le pas et la matrice de covariance sont associés à chaque individu et sont aussi soumis à des mutations d'une itération à une autre (Back *et al*, 1995). En fonction des paramètres de mutation utilisés et de la complexité du modèle de la matrice de covariance, on distingue trois variantes de mutation auto-adaptative:

- 1) Mutation isotrope : elle n'utilise qu'un pas σ par individu et la matrice de covariance correspond à la matrice identité. La mutation consiste alors à modifier initialement la valeur du pas selon une loi log-normale et puis à muter les variables à l'aide de la nouvelle valeur du pas.
- 2) Mutation non-isotrope : elle attache à chaque individu un vecteur de pas $(\sigma_1, \dots, \sigma_n)$ (n est la dimension du vecteur) et la matrice de covariance est alors la matrice diagonale ayant σ_i^2 en $i^{\text{ème}}$ position sur la diagonale. Les différents individus sont mutés indépendamment en fonction du pas mais aussi en fonction de la direction du pas dans l'espace de recherche.

- 3) Mutation corrélée : elle attache à chaque individu une matrice de covariance complète. C'est une matrice symétrique définie positive. L'opération de mutation revient à muter les deux paramètres de la mutation puis à muter les individus du problème avec ces nouveaux paramètres.

La mutation auto-adaptative n'a pas eu un grand succès compte tenu de sa lenteur. Selon Le Riche *et al* (2007), dans le cas le plus simple de la mutation isotrope, si le pas initial n'est pas proportionnel à la distance entre la solution initiale et la solution optimale, un temps de calcul important est nécessaire pour qu'il le devienne et que l'algorithme donne des résultats satisfaisants. Cette limite a poussé les chercheurs, notamment Hansen et Ostermeier à développer l'approche SE-AMC ou Stratégie d'Évolution avec Adaptation de la Matrice de Covariance (« Covariance Matrix Adaptation » en anglais) qui introduit une méthode déterministe d'adaptation du pas (Baumgartner, 2008, Hansen, 2011).

L'algorithme SE-AMC connaît à l'heure actuel un succès croissant dans diverses applications industrielles comme la reconnaissance des formes, l'optimisation de la forme des lentilles, la conception automatisée de systèmes de lentilles (Gagné, 2005), la détermination de l'emplacement idéal des puits pour l'extraction des hydrocarbures (Bouzarkouna, 2012), etc. Il a été utilisé pour déterminer les paramètres de la reconstruction 3D dans le cadre du présent travail.

4.1.3.3 L'algorithme SE-AMC

L'algorithme SE-AMC ou Stratégie d'Évolution avec Adaptation de la Matrice de Covariance est l'algorithme le plus robuste, complet et actuel des stratégies d'évolution (Gagné, 2005, Hansen, 2011). Il s'appuie sur l'utilisation d'une matrice de covariance globale pour paramétrer les mutations gaussiennes et adapter cette matrice de covariance à partir des chemins cumulatifs de mutation ayant donné des résultats satisfaisants. La notion de chemin cumulatif provient du principe qui stipule que si deux mutations successives sont allées dans la même direction (définie par l'orientation de la matrice de covariance), il faut probablement augmenter le pas de la mutation afin d'aller plus vite vers les bonnes solutions. Une étude détaillée sur le développement et la mise en œuvre de l'algorithme SE-AMC est fournie dans la référence (Hansen et Ostermeier, 2001).

SE-AMC utilise un algorithme (μ, λ) : il génère λ enfants à partir d'un seul parent issu d'une moyenne pondérée des μ parents. L'étape de sélection s'applique ainsi uniquement sur les λ enfants (sachant que $\lambda > \mu$) (Baumgartner, 2008). Généralement, trois critères d'arrêt différents peuvent être utilisés dans les algorithmes d'optimisation afin de mettre un terme à la recherche de solutions :

- 1) Un critère sur le déplacement entre les solutions issues de générations successives. Par exemple, si x_k est la solution associée à la génération k et x_{k+1} est celle de la génération $k+1$, alors on

considère que l'algorithme a atteint un optimum si la mesure $\|x_{k+1} - x_k\|$ est inférieure à un seuil prédéterminé. Dans ce cas, la conformité au critère signifie que les solutions n'évoluent plus.

- 2) Un critère sur la progression de la fonction objective. Par exemple, si $f(x_k)$ représente le résultat de la fonction objective pour la solution x_k associée à la génération k et si $f(x_{k+1})$ représente le résultat de la fonction objective pour la solution x_{k+1} associée à la génération k+1, alors on considère que l'algorithme a atteint un optimum si la mesure $\|f(x_{k+1}) - f(x_k)\|$ qui représente la distance entre $f(x_{k+1})$ et $f(x_k)$ atteint une valeur très faible voire nulle et garde cette même valeur sur plusieurs générations. Cela signifie alors que la fonction objective a atteint un minimum.
- 3) Un critère sur le nombre d'itérations ou le temps de calcul.

La particularité de l'approche SE-AMC est qu'elle combine ces trois critères d'arrêt. Ainsi, elle contrôle la recherche par le suivi de l'évolution de la mutation des solutions à travers les générations et le calcul de la distance entre deux générations consécutives de solutions. Le nombre d'itérations intervient également comme critère d'arrêt de la recherche de solutions. La détermination du nombre d'itérations au bout duquel l'algorithme converge vers les meilleures solutions dépend de la complexité du problème et du nombre de solutions à déterminer. Il doit être suffisamment grand pour permettre une convergence complète de l'algorithme (Hansen, 2011, Gagné, 2005). Plus de détails sur l'algorithme sont décrits dans le tutoriel « Covariance Matrix Adaptation » (Hansen, 2011).

Selon Le Riche *et al* (2007), Gagné (2005) et Hansen (2011), SE-AMC est sans contestation possible, la meilleure méthode évolutionnaire pour l'optimisation paramétrée. L'étude comparative menée par Suganthan *et al* (2005) sur les performances de différentes approches stochastiques telles que SE-AMC, Differential Evolution (DE) ou encore Particule Swarm Optimization (PSO) a montré que SE-AMC donne les meilleurs résultats sur l'ensemble des expérimentations comparatives réalisées. Ces performances ont été mesurées lors de la résolution de problèmes complexes et réels, selon plusieurs critères dont le temps de calcul nécessaire, l'évolution des solutions calculées et la capacité à trouver l'optimum global de la fonction objective.

Compte tenu de l'efficacité de l'algorithme SE-AMC pour la résolution de problèmes complexes d'optimisation, celui-ci a été retenu dans le cadre de cette étude, pour l'étape de reconstruction 3D. Ainsi, les sections suivantes vont décrire les deux approches de reconstruction 3D proposées, ces deux approches faisant appel à SE-AMC puisque les deux modélisations sous-jacentes (une curviligne et une volumique) formalisent le problème de reconstruction via l'optimisation de fonctions objectives spécifiques.

4.2 Reconstruction 3D selon une approche curviligne

Comme introduite précédemment, cette première approche est qualifiée de curviligne car elle associe à tout pixel d'une image, un arc de cercle dans la scène. Cet arc de cercle correspond à l'ensemble des points de la scène qui peuvent potentiellement contribuer à l'intensité du pixel du fait de la géométrie d'acquisition et du modèle descriptif associé au pixel, c'est-à-dire un azimut et une distance connus par rapport à la caméra.

L'objectif de la reconstruction est donc de retrouver l'élévation (inconnue) du point de la scène qui contribue à deux pixels appariés sur deux images différentes, en recherchant l'intersection des deux arcs de cercle ainsi formés. Ainsi, en plus de l'élévation, le mouvement de la caméra entre les deux images est aussi à déterminer puisqu'il intervient dans les équations des arcs de cercle.

Le principe de cette approche est similaire à celui de l'approche « Structure From Motion » (SFM) (Moulon *et al*, 2012). Cette dernière est largement utilisée en imagerie optique pour estimer automatiquement le mouvement de la caméra et les coordonnées 3D des points extraits à partir d'une séquence d'images. À la différence des images optiques où tout pixel provient de la projection centrale d'un point de la scène à travers un rayon passant par le centre de la caméra et se projetant en un point du plan image (i.e. modèle sténopé associé à la caméra optique), en imagerie acoustique, tout pixel correspond à la contribution de tous les points de la scène se trouvant à une distance donnée de la caméra, selon un azimut donné, c'est-à-dire se trouvant sur un arc de cercle (i.e. modèle de projection associé à la caméra acoustique, cf. Chapitre 2, paragraphe 2.2.2). L'appariement de deux points saillants dans deux images acoustiques différentes traduit l'occurrence d'un point 3D dans la scène. De ce fait, chaque point 3D calculé à partir de deux points saillants appariés se trouve à l'intersection des arcs de cercle centrés sur les positions de la caméra lors de l'acquisition des différentes images. En appliquant le modèle de la caméra acoustique, la fonction objective définie pour l'algorithme d'optimisation vise à déterminer le déplacement de la caméra qui traduit la projection des points saillants d'une image 1 vers une image 2 pour pouvoir, par la suite, calculer les coordonnées 3D de ces points. D'une manière itérative, la fonction objective va, donc, minimiser la distance entre 1) l'arc de cercle issu de la projection d'un point saillant extrait d'une 1^{ère} image dans le repère de la 2^{ème} image et 2) l'arc de cercle issu de la projection du point correspondant extrait de la 2^{ème} image. Idéalement, cette distance doit être égale à zéro si le déplacement inter-images est celui qui a eu lieu lors de l'acquisition. Le point d'intersection de ces arcs de cercle dans la scène définit alors le point 3D recherché. Toutefois à cause du bruit présent dans les images, des imperfections du modèle de la caméra et des traitements appliqués, notamment lors de l'extraction des points saillants, une telle intersection ne peut être obtenue. Le déplacement de caméra recherché est donc celui qui minimise pour ce point saillant, la distance entre les arcs associés à ce point saillant. Afin de considérer l'ensemble des points saillants extraits d'une image, le but global de la fonction

objective est de déterminer le déplacement de la caméra qui minimise la somme des distances minimales entre les arcs de cercle calculées sur l'ensemble des points saillants utilisé.

4.2.1 Reconstruction à partir de deux images

Dans ce paragraphe, l'approche curviligne utilise un ensemble de points appariés extraits à partir de deux images acquises selon deux points de vue différents. Les points saillants issus de chaque image ont été détectés en appliquant l'approche présentée au chapitre 3. Comme indiqué à la fin du chapitre précédent, afin de tester la faisabilité d'une reconstruction 3D de la scène observée, nous n'avons pas mis en œuvre une approche d'appariement automatique entre images issues de différentes séquences. Ainsi, l'appariement des points saillants extraits et issus de différentes séquences a été effectué manuellement en associant les points saillants qui caractérisent un même point dans la scène (cf. Chapitre 3, paragraphe 3.5). Ayant ainsi à disposition des points saillants homologues, l'algorithme SE-AMC a été utilisé afin de calculer le déplacement de la caméra entre les deux acquisitions puis l'élévation de ces points.

4.2.1.1 Les paramètres de l'espace de recherche

Les paramètres recherchés lors de la mise en œuvre de l'approche SE-AMC sont :

1. les trois paramètres de la rotation $\mathbf{R} : (\alpha, \beta, \gamma)$. Ce sont les angles qui caractérisent, respectivement , le mouvement en rotation de la caméra par rapport à l'axe des X, des Y et des Z, entre les deux images.
2. les trois paramètres de la translation $\mathbf{T} : (T_X, T_Y, T_Z)$. Ce sont les translations de la caméra selon l'axe des X, des Y et des Z effectuées lors de son déplacement entre les points de vue d'acquisition des images.

L'espace de recherche de SE-AMC est donc constitué de ces 6 paramètres. Dans cette configuration à deux images, quel que soit le nombre de points extraits des images, la dimension de cet espace de recherche reste inchangée.

4.2.1.2 La fonction objective

Soit M_k , un point de la scène, observé sur deux images. Chaque image disposant de son repère d'acquisition propre, centré sur la position de la caméra à l'instant d'acquisition, les coordonnées dans la scène, de ce point M_k , calculées à partir de l'image 1 (resp. image 2) dans le repère \mathbf{R}_1 (resp. \mathbf{R}_2) sont $P_{1,k}$ (resp. $P_{2,k}$). Selon le modèle de la caméra acoustique, les coordonnées de $P_{1,k}$ et de $P_{2,k}$ peuvent également s'écrire de deux manières, à savoir en coordonnées polaires ou en coordonnées cartésiennes, dans leurs repères d'acquisition respectifs:

$$P_{1,k} = \begin{pmatrix} [r_{1,k} \sin \theta_{1,k}] \cos \varphi_{1,k} \\ [r_{1,k} \cos \theta_{1,k}] \cos \varphi_{1,k} \\ [-r_{1,k}] \sin \varphi_{1,k} \end{pmatrix} = \begin{pmatrix} a_{1,k} \cos \varphi_{1,k} \\ b_{1,k} \cos \varphi_{1,k} \\ c_{1,k} \sin \varphi_{1,k} \end{pmatrix} = \begin{pmatrix} x_{P_{1,k}} \\ y_{P_{1,k}} \\ z_{P_{1,k}} \end{pmatrix} \quad (12)$$

$$P_{2,k} = \begin{pmatrix} [r_{2,k} \sin \theta_{2,k}] \cos \varphi_{2,k} \\ [r_{2,k} \cos \theta_{2,k}] \cos \varphi_{2,k} \\ [-r_{2,k}] \sin \varphi_{2,k} \end{pmatrix} = \begin{pmatrix} a_{2,k} \cos \varphi_{2,k} \\ b_{2,k} \cos \varphi_{2,k} \\ c_{2,k} \sin \varphi_{2,k} \end{pmatrix} = \begin{pmatrix} x_{P_{2,k}} \\ y_{P_{2,k}} \\ z_{P_{2,k}} \end{pmatrix} \quad (13)$$

avec :

- $r_{1,k}$ (resp. $r_{2,k}$): distance entre la caméra et le point $P_{1,k}$ (resp. $P_{2,k}$). Ces valeurs sont enregistrées sur les images 1 et 2.
- $\theta_{1,k}$ (resp. $\theta_{2,k}$): angle en azimut par rapport à l'axe de pointage de la caméra et la droite passant par la caméra centrée en \mathbf{R}_1 (resp. \mathbf{R}_2) et le point $P_{1,k}$ (resp. $P_{2,k}$). Ces valeurs sont enregistrées sur les images 1 et 2.
- $\varphi_{1,k}$ (resp. $\varphi_{2,k}$): angle en élévation du point $P_{1,k}$ (resp. $P_{2,k}$). Ces deux valeurs ne sont pas enregistrées au travers des images et sont, donc, inconnues.

L'ensemble des lieux possibles dans la scène pour le point M_k correspond donc à l'intersection de deux arcs de cercle $C_{1,k}$ et $C_{2,k}$ (définis par $P_{1,k}$ et $P_{2,k}$), c'est-à-dire à un point d'intersection. Ces arcs de cercle s'étendent sur une ouverture égale à l'ouverture verticale de la caméra acoustique. L'élévation du point d'intersection sera, donc, recherchée dans un intervalle qui s'étend sur 10° .

Sachant que la transformation entre les repères \mathbf{R}_1 et \mathbf{R}_2 est modélisée par la combinaison d'une rotation $R_{1 \rightarrow 2} = (\alpha, \beta, \gamma)$ et d'une translation $T_{1 \rightarrow 2} = (T_X, T_Y, T_Z)$, il est alors possible de calculer $P'_{1,k}$, les coordonnées de M_k dans le repère \mathbf{R}_2 à partir des coordonnées $P_{1,k}$ de M_k dans le repère \mathbf{R}_1 .

$$P'_{1,k} = R_{1 \rightarrow 2} \cdot P_{1,k} + T_{1 \rightarrow 2} \quad (14)$$

avec :

$$R_{1 \rightarrow 2} = \begin{pmatrix} R_{1,1} & R_{1,2} & R_{1,3} \\ R_{2,1} & R_{2,2} & R_{2,3} \\ R_{3,1} & R_{3,2} & R_{3,3} \end{pmatrix} \quad (15)$$

$$= \begin{pmatrix} \cos \alpha \cos \gamma - \sin \alpha \sin \beta \sin \gamma & -\sin \alpha \cos \beta & \cos \alpha \sin \gamma - \sin \alpha \sin \beta \cos \gamma \\ \sin \alpha \cos \gamma + \cos \alpha \sin \beta \sin \gamma & \cos \alpha \cos \beta & \sin \alpha \sin \gamma - \cos \alpha \sin \beta \cos \gamma \\ -\cos \alpha \cos \gamma & \sin \beta & \cos \alpha \cos \gamma \end{pmatrix} \quad (16)$$

Comme $P_{1,k}$ et $P_{2,k}$ correspondent au même point M_k de la scène et $(R_{1 \rightarrow 2}, T_{1 \rightarrow 2})$ correspond au mouvement de la caméra entre l'image 1 et l'image 2, alors $P_{2,k}$ devrait être égal à $P'_{1,k}$ et la relation (17) exprimant une telle égalité, vérifiée.

$$\begin{cases} a_{2,k} \cos \varphi_{2,k} &= a_{1,k} \cos \varphi_{1,k} R_{1,1} + b_{1,k} \cos \varphi_{1,k} R_{1,2} + c_{1,k} \sin \varphi_{1,k} R_{1,3} + T_X \\ b_{2,k} \cos \varphi_{2,k} &= a_{1,k} \cos \varphi_{1,k} R_{2,1} + b_{1,k} \cos \varphi_{1,k} R_{2,2} + c_{1,k} \sin \varphi_{1,k} R_{2,3} + T_Y \\ c_{2,k} \sin \varphi_{2,k} &= a_{1,k} \cos \varphi_{1,k} R_{3,1} + b_{1,k} \cos \varphi_{1,k} R_{3,2} + c_{1,k} \sin \varphi_{1,k} R_{3,3} + T_Z \end{cases} \quad (17)$$

Le système de trois équations auquel on aboutit pour le point M_k , comprend plus d'inconnues à déterminer (i.e. $T_X, T_Y, T_Z, \alpha, \beta, \gamma, \varphi_{1,k}$ et $\varphi_{2,k}$) que le nombre d'équations du système. Afin de déterminer ces inconnues et calculer les coordonnées 3D des points saillants, l'approche proposée s'appuie, initialement, sur le calcul de 6 paramètres de déplacement de la caméra entre les images. Ces paramètres qui traduisent le déplacement de la caméra lors de l'acquisition des images permettront par la suite de calculer les élévations des points et de déduire les coordonnées 3D des points. L'espace de recherche est donc réduit à 6 paramètres de déplacement ($T_X, T_Y, T_Z, \alpha, \beta, \gamma$). Le calcul des angles des élévations des points ($\varphi_{1,k}$ et $\varphi_{2,k}$) se fait au fur et à mesure que l'algorithme d'optimisation propose des solutions à ces 6 paramètres de déplacement. Les angles des élévations sont déterminés à partir d'un intervalle de valeurs égal à l'ouverture verticale de la caméra.

Afin d'être en mesure de résoudre ce système de 3 équations à 6 inconnues et déterminer la transformation entre l'image 1 et l'image 2, un seul point M_k ne peut donc suffire. C'est pourquoi un ensemble de primitives ponctuelles $\{P_{1,k}\}$ et $\{P_{2,k}\}$ extraites des images puis appariées, est utilisé. En effet, toutes les paires de points issues d'une même paire d'images partagent la même transformation géométrique. Chaque paire de points introduit donc de nouvelles équations avec les mêmes paramètres de déplacement inconnus.

Par ailleurs, dans un cas idéal, il devrait y avoir intersection entre les arcs de cercle issus d'une paire de points. Néanmoins, il semble illusoire de chercher de telles intersections à la vue des différentes sources d'imperfection mises en évidence, jusqu'à présent. Par conséquent, l'idée est de rechercher la transformation (de paramètres $T_X, T_Y, T_Z, \alpha, \beta$ et γ) qui minimise la distance entre les arcs de cercle $C_{1,k}$ et $C_{2,k}$. En effet, connaissant cette transformation, les deux arcs de cercle $C_{1,k}$ et $C_{2,k}$ sont complètement définis. Calculer la distance entre ces arcs de cercle revient à trouver les deux points $\widehat{P'_{1,k}}$ de $C'_{1,k}$ et $\widehat{P_{2,k}}$ de $C_{2,k}$, les plus proches. $C'_{1,k}$ est l'expression de l'arc de cercle $C_{1,k}$ dans le repère \mathbf{R}_2 . On peut alors déduire l'élévation de M_k soit à partir de $\varphi_{1,k}$ provenant de $\widehat{P'_{1,k}}$, soit à partir de $\varphi_{2,k}$ provenant de $\widehat{P_{2,k}}$. Ceci explique pourquoi l'espace de recherche se limite aux seuls paramètres de la transformation.

L'idée est alors de chercher les valeurs $R_{1 \rightarrow 2}^*$, $T_{1 \rightarrow 2}^*$ de $R_{1 \rightarrow 2}$ et $T_{1 \rightarrow 2}$ qui minimisent la fonction objective définie comme étant la somme des distances entre les paires d'arcs de cercle $(C'_{1,k}, C_{2,k})$, chaque paire k d'arcs provenant de la paire k de points $(P_{1,k}, P_{2,k})$.

Le problème de reconstruction se résume donc au problème d'optimisation exprimé dans l'équation 18.

$$R_{1 \rightarrow 2}^*, T_{1 \rightarrow 2}^* = \operatorname{argmin}_{R_{1 \rightarrow 2}, T_{1 \rightarrow 2}} \sum_k \operatorname{dist}(C'_{1,k}, C_{2,k}) \quad (18)$$

où $R_{1 \rightarrow 2}^*$, $T_{1 \rightarrow 2}^*$ sont donc les paramètres de rotation et de translation calculés par SE-AMC.

Comme mentionné, toute estimation de la fonction objective pour une transformation $R_{1 \rightarrow 2}$, $T_{1 \rightarrow 2}$, nécessite en étape intermédiaire, le calcul pour chaque paire k de points, du couple de points $(\widehat{P}_{1,k}, \widehat{P}_{2,k})$ et de ce fait permet de remonter à l'élévation du point M_k correspondant aux élévations $\varphi_{1,k}$ et $\varphi_{2,k}$. Il faut noter que toutes les valeurs d'élévation ne sont pas acceptables, En tenant compte de l'angle d'inclinaison de la caméra, ces élévations se trouvent au sein de l'ouverture verticale de la caméra qui ne couvre que 10° .

Ainsi, l'estimation de la fonction objective qui aboutit à $R_{1 \rightarrow 2}^*$, $T_{1 \rightarrow 2}^*$, donne directement une estimation de l'élévation de chacun des points M_k .

4.2.1.3 Remarque sur l'implémentation de l'algorithme de reconstruction curviligne à partir de deux images

Cette remarque concerne la méthode utilisée pour rechercher la distance minimale entre deux arcs de cercle dans un repère tridimensionnel. En effet, il n'existe pas de formule mathématique explicite et simple pour exprimer cette distance. Ainsi, un algorithme itératif par dichotomie a été utilisé pour trouver quels sont les points $\widehat{P}_{1,k}$ de l'arc $C'_{1,k}$ et $\widehat{P}_{2,k}$ de $C_{2,k}$ qui produisent la distance minimale entre les deux arcs.

Cet algorithme doit donc calculer la distance entre tous les points formant l'arc $C'_{1,k}$ et tous les points formant l'arc $C_{2,k}$. Pour ce faire, les deux arcs de cercle, représentés par les équations 19 et 20, sont échantillonnés uniformément en élévation.

$$C_{2,k} = \begin{cases} a_{2,k} \cos \varphi_{2,k} \\ b_{2,k} \cos \varphi_{2,k} \\ c_{2,k} \sin \varphi_{2,k} \end{cases} \quad (19)$$

$$C'_{1,k} = \begin{cases} a_{1,k} \cos \varphi_{1,k} R_{1,1} + b_{1,k} \cos \varphi_{1,k} R_{1,2} + c_{1,k} \sin \varphi_{1,k} R_{1,3} + T_X \\ a_{1,k} \cos \varphi_{1,k} R_{2,1} + b_{1,k} \cos \varphi_{1,k} R_{2,2} + c_{1,k} \sin \varphi_{1,k} R_{2,3} + T_Y \\ a_{1,k} \cos \varphi_{1,k} R_{3,1} + b_{1,k} \cos \varphi_{1,k} R_{3,2} + c_{1,k} \sin \varphi_{1,k} R_{3,3} + T_Z \end{cases} \quad (20)$$

Toutes les distances sont alors calculées et le couple $(\widehat{P}_{1,k}, \widehat{P}_{2,k})$ produisant la distance minimale est retenu. De nouvelles plages d'élévation autour de ces points sur chaque arc sont définies pour être échantillonnées de manière plus fine et l'opération de calcul des distances est relancée. Ces opérations sont répétées jusqu'à aboutir à la précision souhaitée définie par la valeur de la fonction objective minimale.

Cette approche classique par dichotomie permet de réduire drastiquement le nombre d'estimation de distances pour une résolution donnée.

4.2.2 Reconstruction à partir de 'n' images

La scène sous-marine pouvant être observée plusieurs fois sous différents angles, un même point M_k de la scène peut apparaître sur plusieurs images. La reconstruction d'une scène à partir de 'n' images avec $n > 2$, reprend et étend les mécanismes précédents, en exploitant l'occurrence conjointe de ces points sur plusieurs images. L'idée sous-jacente est de déterminer dans quelle mesure, l'augmentation du nombre d'images appariées va impacter la qualité de la reconstruction par rapport à une reconstruction à partir de deux images seulement.

Utiliser 'n' images de la même scène implique que certains points M_k de la scène vont apparaître sur les 'n' images : $P_{1,k}$ sur l'image 1, $P_{2,k}$ sur l'image 2, ..., $P_{j,k}$ sur l'image j, ..., $P_{n,k}$ sur l'image n. Ceci veut donc dire que la recherche de l'intersection d'arcs de cercle ne va pas impliquer 2 arcs mais n arcs. Les coordonnées du point M_k dans la scène vont, par conséquent, correspondre au point d'intersection entre les différents arcs de cercle $C_{j,k}$ issus de chaque point $P_{j,k}$. Pour ce faire, en utilisant les paramètres de la transformation géométrique proposés, la fonction objective va chercher la distance minimale entre un point reconstruit $P_{j,k}$ d'un arc de cercle $C_{i,k}$ et tous les points situés sur l'ouverture verticale des autres arcs de cercle. Le calcul de cette distance minimale se fait par paire d'arcs. Une fois les distances minimales entre toutes les paires d'arc calculées, le rôle de la fonction objective est de permettre la minimisation de la somme de ces distances jusqu'à l'obtention d'une valeur minimale. En fonction de cette valeur, deux cas se présentent. Si cette valeur minimale est égale à zéro : tous les arcs de cercle se croisent en un seul point. Si elle est différente de zéro : les arcs de cercle ne se croisent pas. Dans ce cas, l'idée est de minimiser la distance entre tous les points des arcs de cercle qui sont à l'origine des distances minimales entre arcs, afin d'estimer le point de la scène qui minimise cette distance inter-arcs et donc qui correspond à une intersection conjointe de ces arcs. On retrouve la formulation de cette fonction objective dans la Figure 27 qui explicite le pseudo-code de l'algorithme SE-AMC chargé de retrouver la meilleure série de transformations géométriques entre les points de vue inter-images.

Entrées : Fonction objective f
Sorties : Meilleures solutions de transformations trouvées

1) Génération de la population initiale ($R_{1 \rightarrow 2}^0, T_{1 \rightarrow 2}^0, R_{1 \rightarrow 3}^0, T_{1 \rightarrow 3}^0, \dots, R_{1 \rightarrow n}^0, T_{1 \rightarrow n}^0$): SE-AMC utilise un algorithme (μ, λ) : il génère λ enfants à partir d'un seul parent issu d'une moyenne pondérée des μ parents.

2) Évaluation de chaque individu u de la population $R_{1 \rightarrow 2}^0, T_{1 \rightarrow 2}^0, R_{1 \rightarrow 3}^0, T_{1 \rightarrow 3}^0, \dots, R_{1 \rightarrow n}^0, T_{1 \rightarrow n}^0$ selon la fonction objective f

Pour chaque $u = R_{1 \rightarrow 2}^0, T_{1 \rightarrow 2}^0, R_{1 \rightarrow 3}^0, T_{1 \rightarrow 3}^0, \dots, R_{1 \rightarrow n}^0, T_{1 \rightarrow n}^0$

Recherche des distances minimales par dichotomie des arcs de cercle en élévation
Ces distances sont calculées entre paires d'arcs et c'est leur somme qui est minimisée par f

$$f(u) = \sum_k^N \text{dist} (C_{1,k}, C_{2,k}^{(1)}, \dots, C_{n,k}^{(1)})$$

avec N , le nombre de points appariés.
 n , le nombre d'images exploitées.
 $C_{i,k}^{(j)}$, l'arc de cercle correspondant au point k , issu de l'image i , exprimé dans le repère j .

Si la somme des distances minimales entre toutes les paires d'arcs de cercles = 0

$$\text{dist} (C_{1,k}, C_{2,k}^{(1)}, \dots, C_{n,k}^{(1)}) = 0$$

Alors les arcs s'intersectent

$$C_{1,k} \cap C_{2,k}^{(1)} \cap \dots \cap C_{n,k}^{(1)} \neq \emptyset$$

Fin pour

Sinon l'algorithme continue sa recherche de solutions optimales de transformation en suivant les étapes ci-dessous :

3) Sélection d'une partie de la population : $R_{1 \rightarrow 2}^*, T_{1 \rightarrow 2}^*, R_{1 \rightarrow 3}^*, T_{1 \rightarrow 3}^*, \dots, R_{1 \rightarrow n}^*, T_{1 \rightarrow n}^*$ les μ meilleurs individus sont choisis parmi les λ enfants qui correspondent aux paramètres des transformations géométriques inter-images. Les meilleurs individus/solutions sont ceux qui ont engendré la plus faible valeur de f

4) Mutation de $R_{1 \rightarrow 2}^*, T_{1 \rightarrow 2}^*, R_{1 \rightarrow 3}^*, T_{1 \rightarrow 3}^*, \dots, R_{1 \rightarrow n}^*, T_{1 \rightarrow n}^*$. Elle consiste à additionner aux individus sélectionnés dans l'étape 3 des valeurs générées aléatoirement selon une fonction de densité de probabilité paramétrée;

5) Remplacement de la population ($R_{1 \rightarrow 2}^*, T_{1 \rightarrow 2}^*, R_{1 \rightarrow 3}^*, T_{1 \rightarrow 3}^*, \dots, R_{1 \rightarrow n}^*, T_{1 \rightarrow n}^*$) issue de la génération 1 par la nouvelle génération de population issue de l'étape 4;

6) Retour à l'étape 2 : Évaluation

Les étapes 3, 4, et 6 se répètent jusqu'à ce que f ne s'améliore plus : $f = \text{minimum}$

Figure 27 : Pseudo-code associé à l'approche SE-AMC dans une reconstruction 3D multi-images. $R_{1 \rightarrow 2}^0, T_{1 \rightarrow 2}^0, R_{1 \rightarrow 3}^0, T_{1 \rightarrow 3}^0, \dots, R_{1 \rightarrow n}^0, T_{1 \rightarrow n}^0$: solutions initiales de la rotation et la translation entre les images 1 et 2, 1 et 3, ..., 1 et n ; $R_{1 \rightarrow 2}^*, T_{1 \rightarrow 2}^*, R_{1 \rightarrow 3}^*, T_{1 \rightarrow 3}^*, \dots, R_{1 \rightarrow n}^*, T_{1 \rightarrow n}^*$: solutions optimales de la rotation et la translation entre les images 1 et 2, 1 et 3, ..., 1 et n à l'itération t ; f fonction objective.

4.3 Reconstruction 3D selon une approche volumique

À la différence de l'approche curviligne, l'approche volumique considère que le lieu géométrique dans la scène associé à un pixel dans l'image acoustique ne se situe pas le long d'un arc de cercle mais au sein d'un volume. En effet, chaque pixel de coordonnées image correspondant à une distance en portée « r » et un azimut « θ » collecte toutes les rétrodiffusions de la scène en provenance d'une plage $[r, r+\Delta r]$ en distance et $[\theta, \theta+\Delta\theta]$ en azimut. $\Delta\theta$ et Δr correspondent aux résolutions en azimut et en distance de la caméra acoustique. L'intensité enregistrée dans le pixel rassemble donc toutes les réflexions des objets se trouvant dans le volume délimité par 4 arcs de cercle formant ainsi un ruban, pour toutes les valeurs d'élévation correspondant à l'ouverture verticale du capteur.

L'exploitation des paramètres de résolution ($\Delta\theta$ et Δr) est un atout important de l'approche volumique. Elle permet en effet d'intégrer au processus de reconstruction 3D des connaissances *a priori* ou non, caractérisant des sources d'imprécision potentielles (résolution de la caméra, précision des algorithmes de détection, ...). Les reconstructions ainsi obtenues pourront donc supporter des attributs quantitatifs de qualité : par exemple, le résultat de la reconstruction d'un point M_k pourra être donné sur un intervalle d'élévation plutôt que par un point unique.

4.3.1 Reconstruction à partir de deux images

Comme il a été mentionné dans le paragraphe précédent, compte tenu de la résolution de la caméra acoustique, chaque pixel de l'image acquise couvre une plage Δr en distance et $\Delta\theta$ en azimut dans la scène. Par conséquent, chaque pixel de coordonnées image (r, θ) reçoit les contributions de tous les points de la scène dont les coordonnées en distance et en azimut, dans le repère de la caméra, appartiennent au polygone \mathcal{P} défini par les sommets suivants $\{(r, \theta), (r + \Delta r, \theta), (r + \Delta r, \theta + \Delta\theta), (r, \theta + \Delta\theta)\}$. Chaque polygone \mathcal{P} induit donc un volume \mathbf{V} dans la scène (cf. Figure 28), volume délimité en distance et en azimut par les quatre arcs de cercle définis à partir des sommets de ce polygone selon le modèle de la caméra et par la plage $[\varphi, \varphi + \Delta\varphi]$ en élévation. Cette plage $\Delta\varphi$ en élévation correspond à l'ouverture verticale de la caméra et caractérise donc l'étendue du volume \mathbf{V} en élévation.

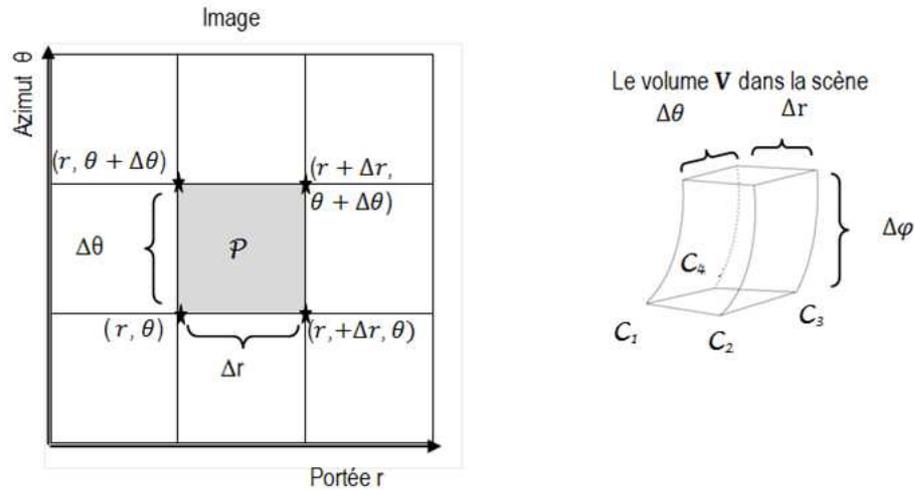


Figure 28 : Un pixel de l'image correspond en réalité à un polygone \mathcal{P} (en portée et en azimut) qui induit un volume V dans la scène. C_1 , C_2 , C_3 et C_4 sont les quatre arcs de cercle issus des quatre coins de \mathcal{P} $\{(r, \theta), (r, \theta + \Delta\theta), (r + \Delta r, \theta + \Delta\theta), (r + \Delta r, \theta)\}$ et qui s'étendent en élévation sur l'intervalle $[\varphi, \varphi + \Delta\varphi]$. $\Delta\theta$ et Δr sont les résolutions en azimut et en distance, $\Delta\varphi$ est l'ouverture en élévation du volume V .

Comme dans le cas de l'approche curviligne, un ensemble de pixels homologues est exploité pour déterminer l'élévation des points saillants correspondant de la scène. Ainsi, un point M_k de la scène, observé sur deux images, a pour coordonnées $(r_{1,k}, \theta_{1,k})$ dans l'image 1 et $(r_{2,k}, \theta_{2,k})$ dans l'image 2. Dans l'image 1 (resp. image 2), ce point est donc associé au polygone $\mathcal{P}_{1,k}$ de sommets $\{(r_{1,k}, \theta_{1,k}), (r_{1,k} + \Delta r, \theta_{1,k}), (r_{1,k} + \Delta r, \theta_{1,k} + \Delta\theta), (r_{1,k}, \theta_{1,k} + \Delta\theta)\}$ (resp. $\mathcal{P}_{2,k}$ de sommets $\{(r_{2,k}, \theta_{2,k}), (r_{2,k} + \Delta r, \theta_{2,k}), (r_{2,k} + \Delta r, \theta_{2,k} + \Delta\theta), (r_{2,k}, \theta_{2,k} + \Delta\theta)\}$), définissant alors dans la scène le ruban $\mathbf{V}_{1,k}$ (resp. $\mathbf{V}_{2,k}$). Le point M_k se trouve en conséquence dans la scène au sein du volume $\mathbf{V}_{1,k} \cap \mathbf{V}_{2,k}$, intersection des rubans $\mathbf{V}_{1,k}$ et $\mathbf{V}_{2,k}$.

Le principe de la méthode de reconstruction 3D proposée consiste donc à exploiter la stratégie d'optimisation SE-AMC afin de retrouver à partir des pixels appariés, la transformation géométrique (rotation $R_{1 \rightarrow 2}$ et translation $T_{1 \rightarrow 2}$) entre les points de vue ayant servi à l'acquisition des deux images, telle que les rubans $\mathbf{V}_{1,k}$ et $\mathbf{V}_{2,k}$ aient une intersection non vide.

Pour schématiser, alors que l'approche curviligne repose sur l'estimation d'intersections d'arcs de cercle, l'approche volumique étudie des intersections de rubans volumiques. L'intersection de rubans volumiques étant des volumes, l'élévation associée à chaque point M_k reconstruit devient de ce fait, une plage d'élévations. En effet, dès l'instant qu'il existe un volume d'intersection entre les deux rubans, tous les points de ce volume sont des candidats valides pour M_k . Ceci permet d'estimer la précision de calcul du point M_k puisque on peut lui attribuer, par exemple, la valeur moyenne des valeurs du volume d'intersection, auquel il

appartient, tout en précisant son intervalle de variation autour de cette moyenne. Ces informations (la moyenne et l'intervalle de variation des coordonnées) peuvent être utilisées par la suite comme des paramètres quantitatifs de la qualité de la reconstruction de chaque point de la scène.

4.3.1.1 Les paramètres de l'espace de recherche

Les paramètres recherchés lors de la mise en œuvre de l'approche SE-AMC sont :

1. les trois paramètres de la rotation $\mathbf{R} : (\alpha, \beta, \gamma)$.
2. les trois paramètres de la translation $\mathbf{T} : (T_x, T_y, T_z)$.

Ce sont aussi 6 paramètres qui constituent l'espace de recherche de SE-AMC dans l'approche volumique.

4.3.1.2 La fonction objective

Dans l'approche volumique proposée, la fonction objective vise à minimiser la distance euclidienne entre deux rubans volumiques. On retrouve le même principe qu'avec l'approche curviligne, sauf que celle-ci visait la minimisation de la distance entre deux arcs de cercle.

Sachant que la transformation entre l'image 1 et l'image 2 est modélisée par la combinaison d'une rotation $R_{1 \rightarrow 2} = (\alpha, \beta, \gamma)$ et d'une translation $T_{1 \rightarrow 2} = (T_x, T_y, T_z)$, pour chaque polygone $\mathcal{P}_{1,k}$ issu d'un point saillant k détecté et apparié, il est alors possible de calculer $\mathbf{V}'_{1,k}$ coordonnées du volume $\mathbf{V}_{1,k}$ dans le repère \mathbf{R}_2 selon l'équation (21).

$$\mathbf{V}'_{1,k} = R_{1 \rightarrow 2} \cdot \mathbf{V}_{1,k} + T_{1 \rightarrow 2} \quad (21)$$

La fonction objective va donc consister à mesurer la distance qui existe entre les deux rubans volumiques issus d'une paire de polygones $(\mathcal{P}_{1,k}, \mathcal{P}_{2,k})$ pour une transformation $(R_{1 \rightarrow 2}, T_{1 \rightarrow 2})$ inter-acquisitions donnée. Cette distance est construite par la somme de la distance en élévation entre les rubans, et de la distance horizontale entre les polygones. Tant que cette distance est strictement positive, il n'existe aucune intersection entre les deux rubans et de ce fait, aucun point de la scène ne peut contribuer aux deux polygones.

Dans la scène, le polygone $\mathcal{P}'_{1,k}$ définit un volume $\mathbf{V}'_{1,k}$ (un ruban dans la scène) délimité par les quatre arcs de cercle à partir de ses sommets et par la plage $[\varphi_{1,k}, \varphi_{1,k} + \Delta\varphi]$ en élévation. Le volume d'intersection entre $\mathbf{V}'_{1,k}$ et $\mathbf{V}_{2,k}$ ($\mathbf{V}_{2,k}$ est formé par les quatre arcs de cercle issu des sommets du polygone correspondant $\mathcal{P}_{2,k}$ et par la plage d'élévation $[\varphi_{2,k}, \varphi_{2,k} + \Delta\varphi]$) est le volume \mathbf{V} . Afin de trouver les coordonnées et la plage d'élévation de ce volume \mathbf{V} dans la scène, il suffit de trouver l'intersection entre les volumes $\mathbf{V}'_{1,k}$ et $\mathbf{V}_{2,k}$. Pour ce faire, la fonction objective calcule la distance entre les deux rubans volumiques $\mathbf{V}'_{1,k}$ et $\mathbf{V}_{2,k}$. Cette

distance est la somme de la distance euclidienne entre les polygones et la distance en élévation entre les polygones les plus proches. La somme nulle de ces deux distances engendre une intersection conjointe entre les rubans sur le plan horizontale et en élévation. Cette intersection correspond à un volume dont la dimension sur le plan horizontale est celle des polygones, et en élévation l'écart en élévation entre le polygone ayant la plus faible élévation (sommet inférieur du volume) et le polygone ayant la plus grande élévation (sommet supérieur du volume). L'objectif de l'algorithme d'optimisation SE-AMC étant de retrouver la transformation géométrique $R_{1 \rightarrow 2}^*, T_{1 \rightarrow 2}^*$ qui minimise cette distance (idéalement 0), comme indiqué dans l'équation (22).

De manière similaire à l'approche curviligne, la fonction objective exploite l'ensemble des appariements inter-images et somme donc les distances entre les rubans volumiques pour tous les polygones appariés.

$$R_{1 \rightarrow 2}^*, T_{1 \rightarrow 2}^* = \operatorname{argmin}_{R_{1 \rightarrow 2}, T_{1 \rightarrow 2}} \sum_k \operatorname{dist}(\mathbf{V}'_{1,k}, \mathbf{V}_{2,k}) \quad (22)$$

$R_{1 \rightarrow 2}^*, T_{1 \rightarrow 2}^*$ sont donc les paramètres de rotation et translation calculés par l'algorithme SE-AMC qui minimisent globalement la distance entre les paires de rubans volumiques. Pour chaque paire de rubans, une distance (*dist*) nulle indique que l'algorithme d'optimisation a trouvé une rotation et translation qui permettent une intersection entre ces rubans. La plage d'élévation pour chaque point M_k reconstruit, découle donc automatiquement de la plage en élévation du volume d'intersection de la paire de rubans associée.

4.3.1.3 Remarque sur l'implémentation de l'algorithme de reconstruction volumique à partir de deux images

Afin de calculer les paramètres de déplacement inter-images l'algorithme recherche itérativement la distance minimale entre les deux rubans volumiques $\mathbf{V}_{1,k}$ et $\mathbf{V}_{2,k}$ issus de l'image 1 et l'image 2 dans le repère de la scène. Pour ce faire, l'algorithme utilise la dichotomie pour discrétiser uniformément en élévation les rubans en un ensemble de polygones et déterminer par la suite les polygones $\widehat{\mathbf{P}}'_{1,k}$ du volume $\mathbf{V}'_{1,k}$ et $\widehat{\mathbf{P}}'_{2,k}$ de $\mathbf{V}_{2,k}$ qui produisent la distance minimale entre les deux rubans.

Cet algorithme parcourt ainsi tout l'ensemble de polygones qui forment le volume $\mathbf{V}'_{1,k}$ et calcule donc la distance entre ces polygones et tous ceux qui forment le volume $\mathbf{V}_{2,k}$. Afin de déterminer cette distance, l'algorithme calcule toutes les distances euclidiennes, exprimées par l'équation (23), entre tous les sommets de deux polygones et détermine la distance minimale. Cette dernière correspond, ainsi, à la distance entre ces deux polygones.

$$\operatorname{dist} = \sqrt{(x - x')^2 + (y - y')^2} \quad (23)$$

En plus de la distance entre les paires de polygones, la fonction objective tient compte de la distance en élévation des rubans. Cette distance correspond à l'écart en élévation entre les polygones les plus proches.

La somme de ces deux distances est, alors, calculée. Les rubans qui se croisent (en élévation et en polygone de résolution $(\Delta r, \Delta \theta)$) en des zones volumiques communes sont donc les rubans qui ont engendré une distance égale à zéro. Ces zones volumiques regroupent les couples de polygones qui ont produit la distance minimale.

Comme dans le cas de l'approche curviligne afin d'augmenter la précision de calcul, l'algorithme continue sa recherche itérative en définissant de nouvelles plages d'élévation autour de ces polygones sur chaque ruban volumique. Ces derniers sont alors re-échantillonnés de manière plus fine et l'algorithme relance sa recherche de couples de polygones produisant les distances minimales entre les rubans.

4.3.2 Reconstruction à partir de 'n' images

Le principe de la reconstruction à partir de 'n' images consiste à exploiter plusieurs polygones homologues ($\mathcal{P}_{1,k}, \mathcal{P}_{2,k}, \mathcal{P}_{3,k}, \mathcal{P}_{4,k}, \dots$) afin de déduire l'élévation du point M_k correspondant dans la scène. Dans le cas précédent, la reconstruction s'appuie sur le calcul des élévations à partir de l'intersection entre deux rubans issus de la projection de deux polygones dans la scène. Or, dans le contexte de la reconstruction à partir de 'n' images, la recherche des élévations est effectuée à partir de l'intersection de 'n' rubans. Ainsi les paramètres de rotation et de translation sont déterminés de sorte que la fonction objective donne une distance minimale en utilisant des primitives extraites à partir de plus de deux images. L'exploitation de 'n' images ajoute ainsi plus de contraintes dans la recherche de solutions optimales en imposant à l'algorithme pour chaque point saillant, la recherche d'un volume d'intersection commun entre tous les rubans. Il semble que ceci pourrait fournir des solutions plus précises que celles issues d'une reconstruction à partir de deux images; en effet, le croisement de 'n' rubans devrait logiquement aboutir à une intersection commune plus petite que celle obtenue simplement à partir de 2 rubans.

Les expérimentations qui seront abordées dans les prochaines sections vont permettre d'établir une comparaison entre les solutions issues d'une reconstruction à partir de deux images et une reconstruction multi-images. L'influence de l'augmentation du nombre d'images (mais surtout de la diversité des points de vue d'acquisition) sur les solutions calculées sera également abordée.

Le pseudo-code qui intègre la fonction objective de la reconstruction volumique 3D multi-images dans l'algorithme d'optimisation SE-AMC est proposé ci-dessous (cf. Figure 29). Les algorithmes détaillant le calcul effectif de la fonction objectif sont présentés dans l'Annexe B.

Entrées : Fonction objective f
Sorties : Meilleures solution de transformation trouvées

1) Génération de la population initiale ($R_{1 \rightarrow 2}^0, T_{1 \rightarrow 2}^0, R_{1 \rightarrow 3}^0, T_{1 \rightarrow 3}^0, \dots, R_{1 \rightarrow n}^0, T_{1 \rightarrow n}^0$): SE-AMC utilise un algorithme (μ, λ) : il génère λ enfants à partir d'un seul parent issu d'une moyenne pondérée des μ parents.

2) Évaluation de chaque individu u de la population $R_{1 \rightarrow 2}^0, T_{1 \rightarrow 2}^0, R_{1 \rightarrow 3}^0, T_{1 \rightarrow 3}^0, \dots, R_{1 \rightarrow n}^0, T_{1 \rightarrow n}^0$ selon la fonction objective f

Pour chaque $u = R_{1 \rightarrow 2}^0, T_{1 \rightarrow 2}^0, R_{1 \rightarrow 3}^0, T_{1 \rightarrow 3}^0, \dots, R_{1 \rightarrow n}^0, T_{1 \rightarrow n}^0$
Recherche des distances minimales par dichotomie des volumes en élévation

$$f(u) = \sum_k^N \text{dist} (V_{1,k}, V_{2,k}^{(1)}, \dots, V_{n,k}^{(1)})$$

avec N , le nombre de points appariés.
 n , le nombre d'images exploitées.
 $V_{i,k}^{(j)}$, le volume correspondant au point k , issu de l'image i , exprimé dans le repère j

Si la distance conjointe entre tous les volumes est égale à zéro

$$\text{dist} (V_{1,k}, V_{2,k}^{(1)}, \dots, V_{n,k}^{(1)}) = 0$$

> Alors les volumes s'intersectent $V_{1,k} \cap V_{2,k}^{(1)} \cap \dots \cap V_{n,k}^{(1)} \neq \emptyset$

Fin pour

Sinon l'algorithme continue la recherche de solutions optimales de transformation en suivant les étapes ci-dessous :

3) Sélection d'une partie de la population : $R_{1 \rightarrow 2}^*, T_{1 \rightarrow 2}^*, R_{1 \rightarrow 3}^*, T_{1 \rightarrow 3}^*, \dots, R_{1 \rightarrow n}^*, T_{1 \rightarrow n}^*$ les μ meilleurs individus sont choisis parmi les λ enfants qui correspondent aux paramètres des transformations géométriques inter-images. Les meilleurs individus/solutions sont ceux qui ont engendré la plus faible valeur de f

4) Mutation de $R_{1 \rightarrow 2}^*, T_{1 \rightarrow 2}^*, R_{1 \rightarrow 3}^*, T_{1 \rightarrow 3}^*, \dots, R_{1 \rightarrow n}^*, T_{1 \rightarrow n}^*$. Elle consiste à additionner aux individus sélectionnés dans l'étape 3 des valeurs générées aléatoirement selon une fonction de densité de probabilité paramétrée;

5) Remplacement de la population ($R_{1 \rightarrow 2}^*, T_{1 \rightarrow 2}^*, R_{1 \rightarrow 3}^*, T_{1 \rightarrow 3}^*, \dots, R_{1 \rightarrow n}^*, T_{1 \rightarrow n}^*$) issue de la génération 1 par la nouvelle génération de population issue de l'étape 4 :

6) Retour à l'étape 2 : Évaluation

Les étapes 3, 4, 5 et 6 se répètent jusqu'à ce que $f = 0$ et on détermine, par conséquent, le volume d'intersection et les solutions optimales de transformation.

Figure 29 : Pseudo-code associé à l'approche SE-AMC dans une reconstruction volumique 3D multi-images. $R_{1 \rightarrow 2}^0, T_{1 \rightarrow 2}^0, R_{1 \rightarrow 3}^0, T_{1 \rightarrow 3}^0, \dots, R_{1 \rightarrow n}^0, T_{1 \rightarrow n}^0$: solutions initiales de la rotation et la translation entre les images 1 et 2, 1 et 3, ..., 1 et n ; $R_{1 \rightarrow 2}^*, T_{1 \rightarrow 2}^*, R_{1 \rightarrow 3}^*, T_{1 \rightarrow 3}^*, \dots, R_{1 \rightarrow n}^*, T_{1 \rightarrow n}^*$: solutions optimales de la rotation et la translation entre les images 1 et 2, 1 et 3, ..., 1 et n à l'itération t ; f fonction objective.

4.3.3 Principales différences entre l'approche curviligne et l'approche volumique

Les reconstructions 3D curviligne et volumique sont deux approches différentes. La première approche s'appuie sur un modèle simplifié de la caméra où un point saillant dans l'image correspond à un point dans la scène. L'approche volumique propose d'introduire certaines caractéristiques du modèle d'acquisition effectif de la caméra où un point dans l'image correspond à un volume dans la scène. L'application de ces deux approches aboutit à deux résultats différents : les coordonnées 3D d'un point dans la scène (approche curviligne) et un intervalle d'élévation sur lequel s'étend le volume (approche volumique). Les expérimentations qui vont être présentées dans les prochaines sections vont mettre en œuvre ces deux approches et analyser leurs résultats. La comparaison de ces résultats permettra de mettre en évidence quantitativement les avantages et les inconvénients de chaque approche.

4.4 Contexte d'expérimentation

4.4.1 Jeux de données exploités

4.4.1.1 Images utilisées

Les séquences utilisées dans les présentes expérimentations sont celles acquises dans la piscine de l'Université Laval. Dans chacune de ces séquences, les images sont enregistrées selon un point de vue d'acquisition similaire ce qui a engendré une faible différence dans les vues de la scène d'une image à une autre. Des expérimentations avec des images consécutives acquises au sein de la même séquence ont montré qu'il n'est pas possible d'obtenir une reconstruction 3D fiable de la scène. Une telle tentative de reconstruction 3D par approche volumique est illustrée dans la Figure 30. Il s'agit d'une reconstruction avec deux images successives de la séquence 17 en utilisant les 20 points saillants représentés dans la Figure 31. Dans cet exemple de reconstruction, des intersections entre les rubans issus des deux images sont obtenues. Ces intersections illustrent les solutions possibles que chaque point saillant peut avoir comme hauteur. Néanmoins, elles s'étendent sur presque toute l'ouverture verticale des rubans ne permettant pas ainsi de réduire le nombre de solutions possibles ni d'en sélectionner des hauteurs précises pour chaque point saillant.

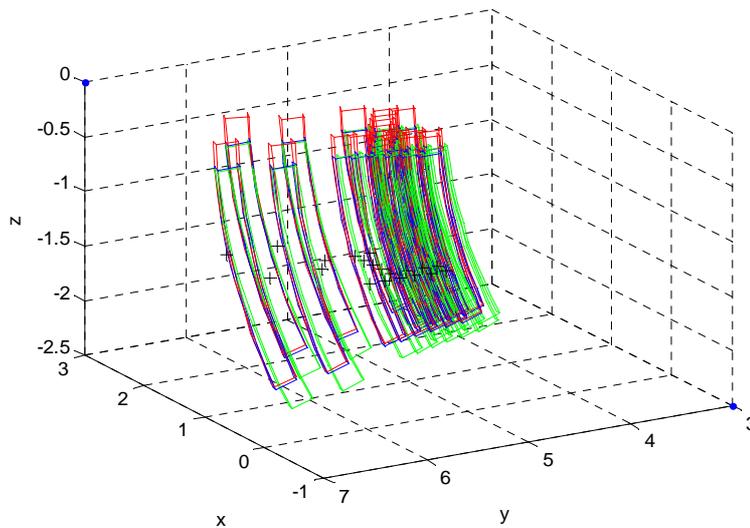
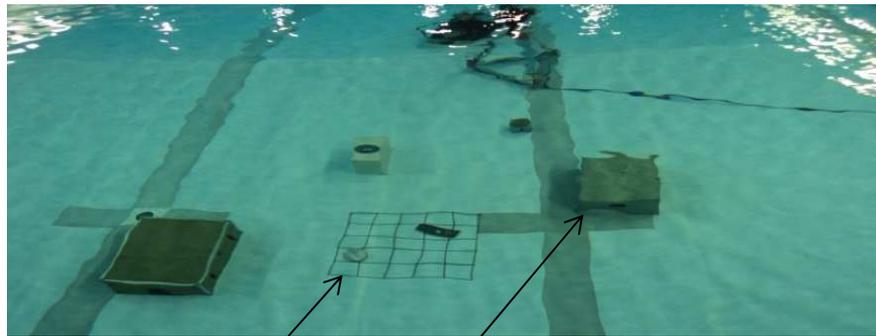
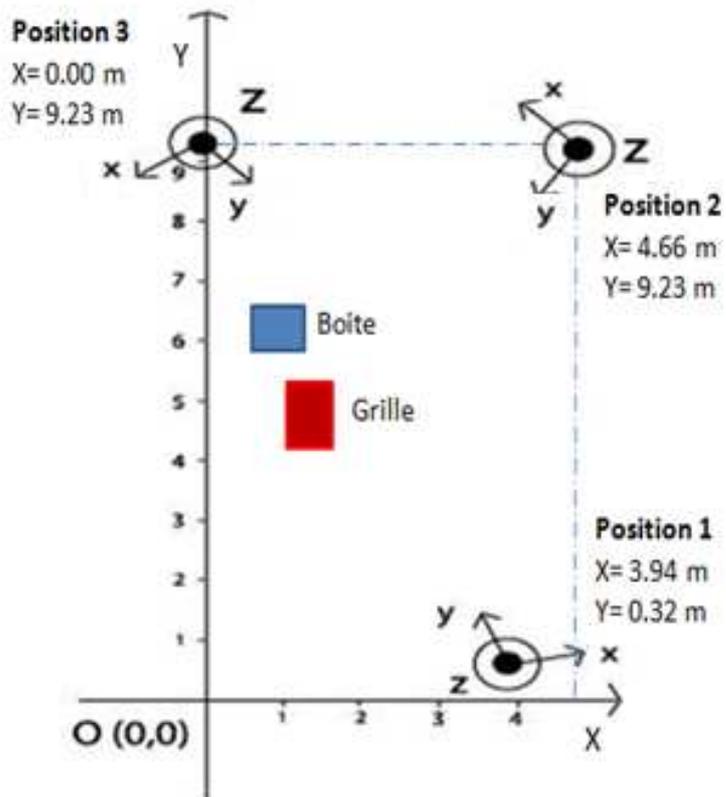


Figure 30 : Reconstruction 3D (exprimées en mètre) calculées par l'approche volumique. Les points en noir correspondent aux coordonnées réelles des points saillants. Les rubans en rouge, en vert et en cyan sont issus des points saillants, respectivement, des trois images successives (24, 25 et 26) extraite de la séquence 17.

Pour réaliser une reconstruction 3D multi-vues il faut, ainsi, utiliser des images ayant des points de vue différents. Pour ce faire, nous nous sommes appuyés sur la connaissance de la réalité terrain, c'est-à-dire la connaissance de la position des objets dans la piscine et de la position de la caméra par rapport aux objets à chaque acquisition d'une séquence (cf. Figure 31) pour sélectionner des séquences offrant des points de vue différents. De ces séquences nous avons extrait des images qui présentent la scène avec des angles de vue différents. Celles-ci sont présentées en Figure 32. Quant aux points saillants, les mêmes points ont été identifiés et par la suite extraits manuellement de ces images.



Grille Boite
(a)



(b)

Figure 31 : (a) La scène enregistrée dans la piscine de l'Université Laval. (b) Les positions de la caméra par rapport à la scène lors de l'acquisition des séquences d'images : La séquence 17 (image 1) est acquise à partir de la position 1; Les séquences 21 (image 2) et 24 (image 3) sont acquises à partir de la même position 2; La séquence 8 (image 4) est acquise à partir de la position 3.

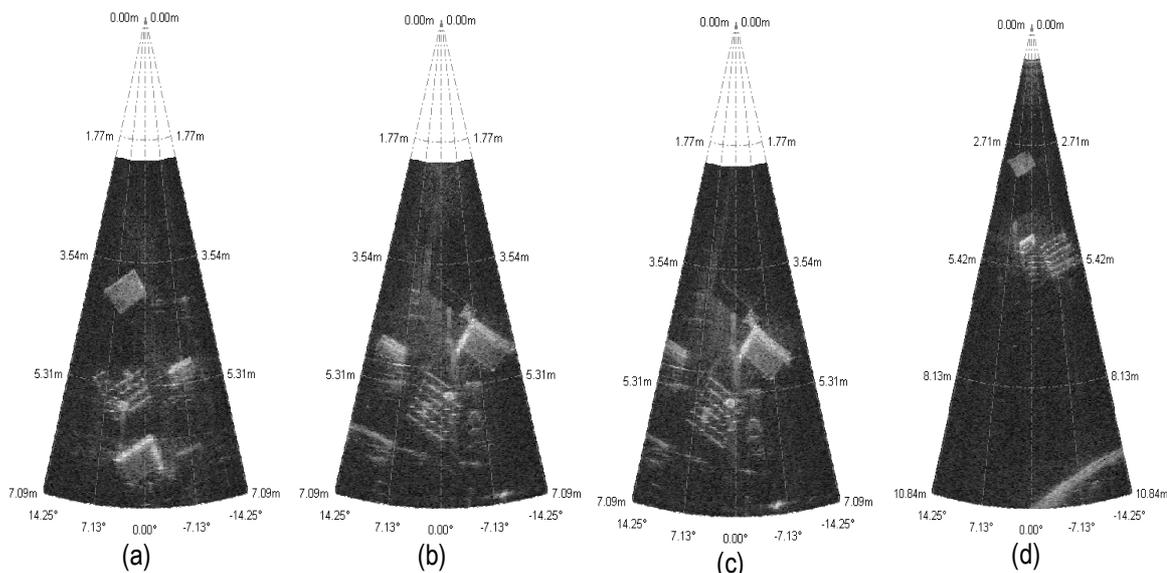


Figure 32 : Images acquises selon différents angles de vue dans le bassin de l'Université Laval et exploitées lors des expérimentations afin d'estimer les performances des approches de reconstruction : (a) Image 1 (séquence 17, position 1); (b) Image 2 (séquence 21, position 2); (c) Image 3 (séquence 24, position 2); (d) image 4 (séquence 8, position 3).

Les coordonnées des différentes positions de la caméra dans le repère de la piscine sont affichées dans le tableau suivant :

Tableau 6. Coordonnées des positions de la caméra dans le repère de la piscine

	Roll (°)	Pitch (°)	Yaw (°)	X (m)	Y (m)	Z (m)
Position 1	0,12	-12,78	54,67	3,94	0,32	1,61
Position 2	0,10	-16,86	77,41	4,66	9,23	1,61
Position 3	4,60	-13,74	87,69	0,00	9,23	1,61

4.4.1.2 Points saillants extraits des images

A partir des objets qui figurent dans la scène, nous avons sélectionné manuellement un ensemble de points saillants. Ces points sont illustrés sur les images de la Figure 33 puis dans le repère de la piscine proposé à la Figure 34. Ils sont classés en deux groupes : le premier groupe est constitué des quatre points de la boîte qui se situent à une hauteur de 0,30 m par rapport au fond de la piscine; le deuxième groupe est constitué des 16 points de la grille qui sont situés sur le fond de la piscine.

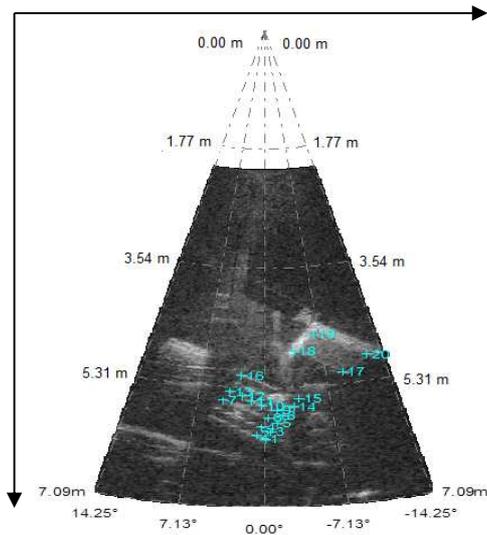
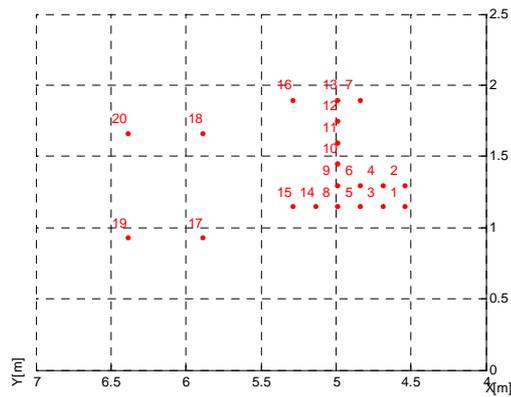
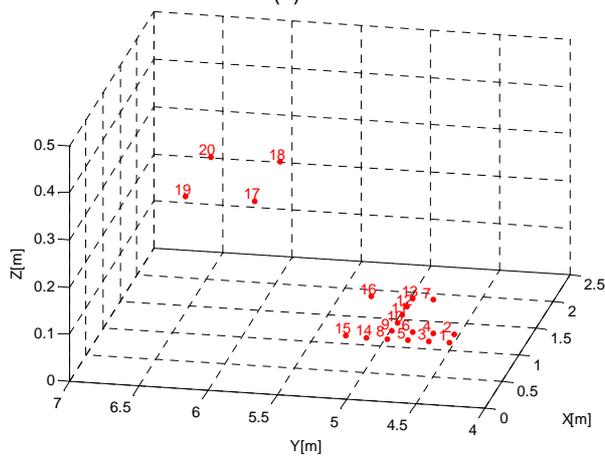


Figure 33 : Position des 20 points saillants extraits de l'image 1 (séquence 17) : 16 points associés à la grille et 4 points associés à la boîte (à droite de l'image).



(a)



(b)

Figure 34 : Position des points de la boîte et de la grille en (X, Y) (a) et en (X, Y, Z) (b) dans le repère de la piscine.

La connaissance des coordonnées réelles des points de la grille et de la boîte dans la piscine nous a permis de confronter les résultats de reconstruction obtenus avec la réalité de la scène, et ainsi d'évaluer les expérimentations de la reconstruction 3D de la scène pour valider les algorithmes développés.

4.4.2 Mise en œuvre de l'algorithme SE-AMC

L'algorithme d'optimisation SE-AMC va intervenir aussi bien dans l'approche curviligne que dans l'approche volumique de reconstruction 3D. Il va donc effectuer les recherches de solutions itérativement, en s'appuyant sur l'évolution de la fonction objective par rapport aux solutions déjà envisagées et estimées. C'est donc bien la formulation de la fonction objective qui diffère entre l'approche curviligne et volumique en intégrant des modélisations géométriques différentes pour la reconstruction 3D des points de la scène à partir des points saillants extraits des images.

Les paramètres de rotation et de translation qui engendrent les plus faibles valeurs des fonctions objective (c'est-à-dire les plus faibles distances entre arcs ou rubans volumiques) sont finalement les solutions retenues à l'issue du processus de recherche. L'évolution de ces solutions tout au long des itérations ainsi que l'évolution de la valeur fournie par la fonction objective peuvent être visualisées sur des graphiques.

4.5 Application de l'approche curviligne

4.5.1 Tests de reconstruction 3D multi-vues

Les tests avec les images réelles acquises en bassin visent plusieurs objectifs. Ils cherchent en premier lieu, à expérimenter les algorithmes développés et valider la méthodologie proposée. Ils visent également à observer le comportement de la convergence de la recherche vers les solutions optimales et finalement, ils cherchent à évaluer la précision des coordonnées 3D déterminées à partir des solutions calculées par l'algorithme d'optimisation. À cet effet, l'erreur entre les coordonnées 3D issues de l'approche SE-AMC et celles réelles, mesurées en bassin sera calculée. La distance euclidienne sera utilisée comme mesure de cette erreur. Les tests ciblent également la détermination des paramètres qui influencent le résultat de l'optimisation.

4.5.1.1 Reconstruction 3D à partir de deux images

La première série de tests a été menée à partir des images 1 (séquence 17) et 2 (séquence 21) (cf. paragraphe 4.4.1.1, Figure 32). L'application de l'approche curviligne de reconstruction 3D permet d'obtenir les paramètres de la rotation (α, β, γ) et de la translation (T_x, T_y, T_z). Ces paramètres sont :

- (α, T_x): L'angle de la rotation et la translation selon l'axe des X dans le repère de l'image 1. C'est au niveau de cet angle que se localise l'ouverture verticale des faisceaux d'insonification de la caméra

acoustique (10°). La rotation de la caméra, autour de l'axe des X, caractérise l'inclinaison de la caméra par rapport à la scène.

- (β, T_Y) : L'angle de la rotation et la translation selon l'axe des Y. Ce dernier est situé au niveau de la ligne centrale de l'image et caractérise la direction de pointage de la caméra.
- (γ, T_Z) : L'angle de la rotation et la translation selon l'axe des Z.

L'origine du repère composé des axes (X, Y et Z) est le centre de la caméra. Les valeurs des paramètres du déplacement de la caméra sont obtenues suite à plusieurs itérations de la fonction objective (cf. Figure 35). Ils sont fournis dans le Tableau 7. À partir de ces paramètres, les coordonnées 3D des points saillants sont calculées. Elles sont illustrées dans la Figure 37 et comparées aux coordonnées réelles dans le Tableau 8 et le Tableau 9.

Afin d'évaluer l'erreur de calcul du déplacement de la caméra, nous disposons des mesures réelles des angles d'inclinaison de la caméra entre les images et les coordonnées de ses positions dans le repère de la piscine. Ces données ont été, en effet, été enregistrés avant chaque nouvelle acquisition de séquences (à des fins de validation des transformations estimées).

- **Évolution de la fonction objective au cours de la recherche**

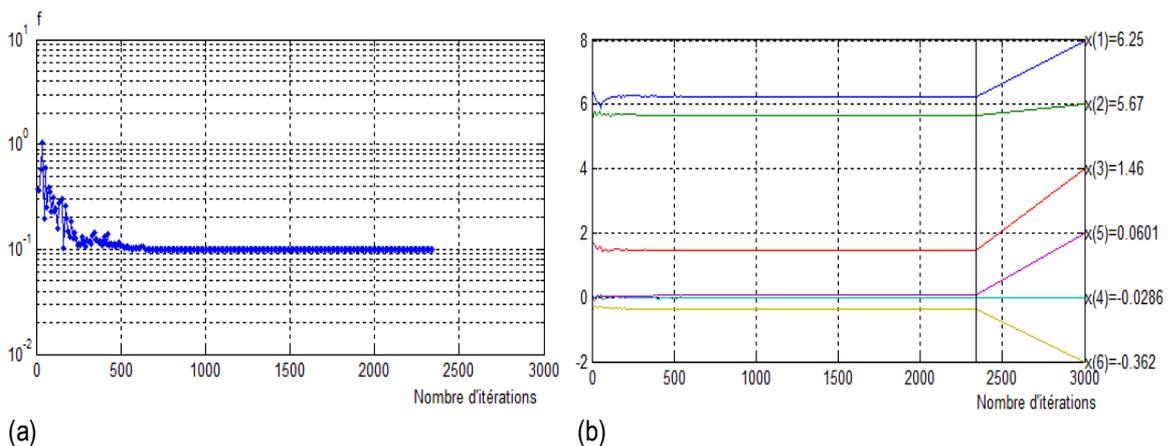


Figure 35 : Évolution de la fonction objective f (a) et des solutions optimales $[x(1), x(2), \dots, x(6)]$ du déplacement de la caméra entre l'image 1 et l'image 2 (b) en fonction du nombre d'itérations

À partir de la 1000^{ème} itération, l'algorithme d'optimisation a convergé vers une solution de rotation et de translation pour laquelle la fonction objective atteint une valeur égale à 0,10 m. La valeur de la fonction objective indique qu'il n'y a pas eu d'intersection entre tous les arcs de cercle issus de tous les points saillants

puisque'elle est différente de zéro. Même si cette valeur est faible pour l'ensemble de 20 points, elle peut avoir un impact sur la précision du calcul du déplacement et par conséquent sur les coordonnées 3D calculées.

D'ailleurs, en utilisant la transformation idéale (déplacement effectif, cf. Tableau 7) entre les images 1 et 2, la valeur de la fonction objective obtenue est différente de zéro, elle est égale à 0,12 m. Ainsi, même avec la transformation idéale, les arcs de cercle ne s'intersectent pas. Les valeurs des fonctions objectives obtenues, respectivement, avec la transformation estimée par optimisation et avec la transformation idéale sont proches (différence de 2 cm). L'algorithme d'optimisation a convergé, par conséquent, vers un minimum global.

- **Déplacement de la caméra**

Tableau 7. Paramètres de la rotation et de la translation de la caméra entre les images 1 et 2: α , β , γ sont les angles de la rotation selon les axes X, Y et Z; T_x , T_y , T_z sont les paramètres de la translation selon les axes X, Y et Z. Ces paramètres sont calculés dans le repère R1 de l'image 1.

	Rotation (°)			Translation (m)		
	A	B	γ	T_x	T_y	T_z
Déplacement effectif	0,02	4,08	-22,74	0,72	3,62	0,00
Déplacement estimé	-1,63	3,44	-20,74	6,25	5,66	0,00

Les déplacements (effectif et estimé) entre les deux images, affichés dans le Tableau 7, sont calculés dans le repère de l'image 1 (**R1**). L'expression de ces déplacements dans le même repère a pour objectif de comparer et évaluer l'erreur entre les paramètres réels et les paramètres calculés.

L'évaluation de la transformation inter-image calculée, nous a amené à conclure que le déplacement calculé par l'algorithme d'optimisation, ne coïncide pas avec le déplacement effectif, tout particulièrement au niveau des paramètres de la translation.

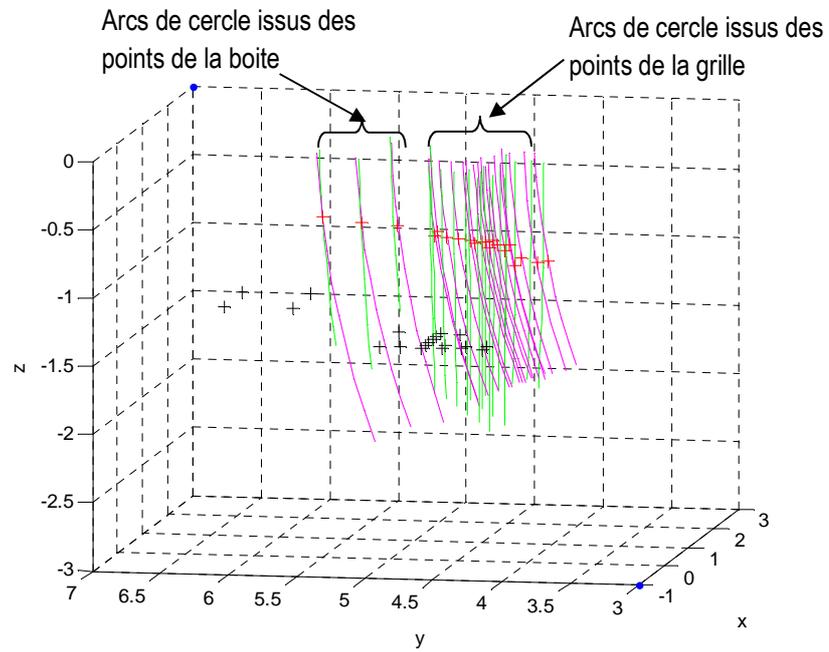
En translation, seul le paramètre en Z correspond à la réalité. En fait, connaissant les conditions expérimentales d'acquisition des séquences d'images, nous avons intégré une connaissance *a priori* dans l'algorithme de reconstruction. Cette connaissance correspond à l'absence d'un déplacement vertical lors du déplacement inter-acquisition de la caméra. En réalité, lors de l'acquisition des séquences dans la piscine, la caméra a été fixée sur un trépied, limitant les déplacements selon l'axe des Z aux très faibles variations de la topographie de la piscine. L'intégration de cette information, dans l'algorithme de reconstruction, a deux objectifs. Le premier est de limiter la recherche de solutions de translation aux paramètres selon lesquels la caméra a effectué réellement un déplacement inter-acquisition. Ces paramètres sont, donc, la translation selon les axes X et Y. Le deuxième est de permettre l'exploitation de cette connaissance (translation nulle selon l'axe des Z) afin d'estimer un déplacement qui corresponde le plus à la configuration réelle d'acquisition.

En fait, plus le mouvement estimé est proche du mouvement effectif, meilleure devrait être l'estimation des coordonnées 3D.

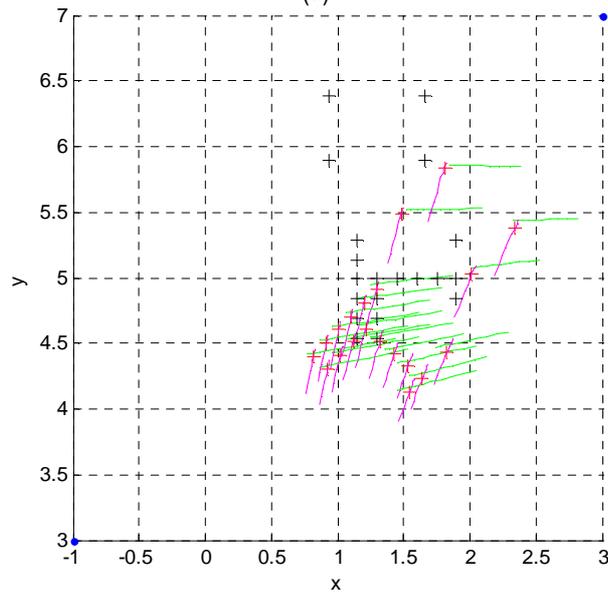
Néanmoins, la comparaison du mouvement estimé obtenu avec le mouvement effectif a révélé que les paramètres de translation en X et Y sont très éloignés de leurs valeurs réelles. Ils affichent une erreur au niveau de chaque axe égale à [5,53, 2,04, 0,00] m. Ces erreurs sont très élevées vis-à-vis de la translation effective et, aussi, de la configuration de la scène (périmètre de la scène, disposition des objets et des caméras dans la piscine).

En ce qui concerne les paramètres de la rotation calculés, ils sont plus précis que les paramètres de la translation calculés. Ils présentent une erreur de l'ordre de [1,65°, 0,64°, -2,00°], respectivement, selon les axes X, Y et Z. Toutefois, avec les erreurs élevées de la translation calculée, on ne peut qualifier le déplacement de la caméra, estimé par l'algorithme d'optimisation, que d'insatisfaisant. Il est différent de la géométrie réelle d'acquisition des images et peut engendrer des imperfections dans le calcul de la reconstruction. La Figure 37, la Figure 35, le Tableau 8 et le Tableau 9 montrent, graphiquement et quantitativement, l'erreur de la reconstruction des points de la grille et des points de la boîte.

- Reconstruction 3D des points saillants



(a)



(b)

Figure 36 : Coordonnées 3D calculées par l'approche curviligne (exprimées en mètre). Les points en noir correspondent aux coordonnées réelles des points saillants et ceux en rouge aux coordonnées calculées. Les arcs de cercle en magenta et en vert sont issus, respectivement, des images 1 et 2. (a) Localisation 3D des points saillants dans le repère de la piscine; (b) Localisation des points saillants dans le plan (X, Y).

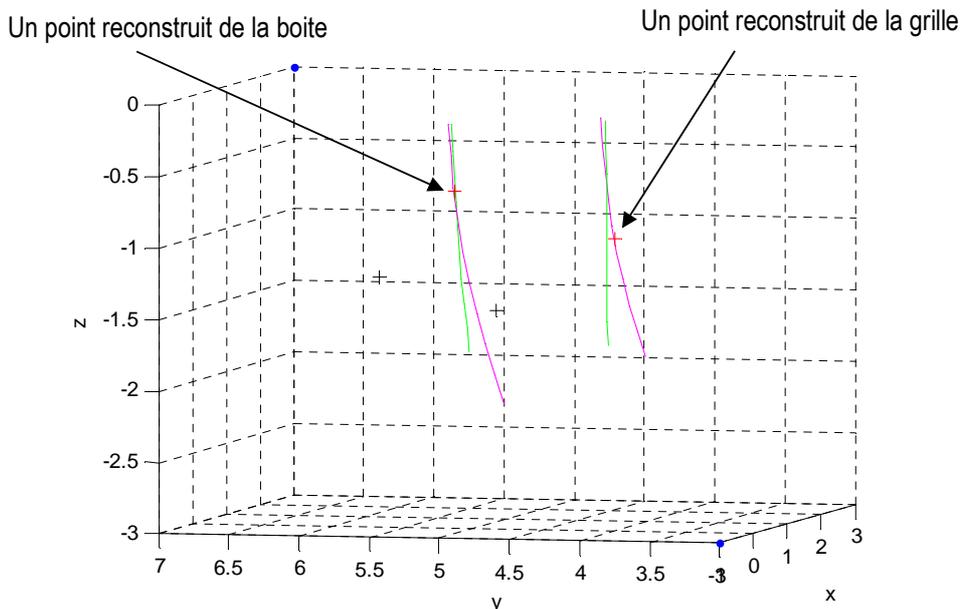


Figure 37 : Un exemple de reconstruction d'un point de la grille et d'un point de la boîte (coordonnées 3D exprimées en mètre). Les arcs de cercle en magenta et en vert sont issus, respectivement, des images 1 et 2. Les points en noir correspondent aux coordonnées réelles des points saillants et ceux en rouge aux coordonnées calculées.

Tableau 8. Les erreurs minimales, maximales et moyennes des coordonnées 3D calculées pour les points de la grille par rapport à leurs coordonnées 3D réelles.

Grille	Erreur minimale (m)	Erreur maximale (m)	Erreur moyenne (m)
X	0,04	0,37	0,19
Y	0,14	0,85	0,42
Z	0,52	0,84	0,73

Tableau 9. Les erreurs minimales, maximales et moyennes des coordonnées 3D calculées pour les points de la boîte par rapport à leurs coordonnées 3D réelles.

Boîte	Erreur minimale (m)	Erreur maximale (m)	Erreur moyenne (m)
X	0,35	0,87	0,61
Y	0,39	1,01	0,70
Z	0,42	0,56	0,49

La représentation des coordonnées dans le plan (X, Y) dans la Figure 37 (b) montre que la disposition des points, respectivement de la boîte et de la grille, correspond à la forme réelle des objets auxquels ils sont associés. Toutefois, cette représentation dans le plan (X, Y) met en évidence un écart entre les coordonnées réelles et celles calculées par l'algorithme. Cet écart est dû à la mauvaise estimation du mouvement de la caméra entre les deux acquisitions d'images. Ceci est confirmé par les résultats affichés dans le Tableau 7 et discutés dans le paragraphe précédent.

En ce qui concerne la hauteur des points, la hauteur moyenne de la boîte estimée à partir des quatre coordonnées Z calculées est égale à -0,81 m. Elle ne correspond pas à la hauteur réelle (-1,31 m, par rapport à la caméra) et présente une erreur moyenne égale à 0,49 m (cf. Tableau 9). Même la hauteur moyenne des points de la grille présente une erreur moyenne élevée de 0,73 m vis-à-vis de la hauteur réelle de ces points. Elle est estimée à -0,87 m alors qu'elle est égale à -1,61 m (par rapport à la caméra) dans la scène.

À partir de ces données, nous avons calculé l'écart entre les deux hauteurs moyennes des points de la grille et des points de la boîte. Celui-ci est égal à 0,06 m. Cet écart est très faible par rapport à la valeur réelle (0,30 m) et ne permet pas de différencier l'ensemble des points de la grille de celui de la boîte.

Les erreurs minimales, maximales et moyennes des coordonnées des points selon les trois axes (X, Y et Z) sont élevées (cf. Tableau 7 et Tableau 8). D'ailleurs, ces erreurs sont non seulement observées au niveau de la représentation 3D des points reconstruits mais aussi au niveau des arcs de cercle qui ne se croisent pas (cf. Figure 37). Les points 3D représentés dans la Figure 37 sont, en effet, les points ayant produit la plus faible distance entre ces arcs de cercle.

- **Principales observations tirées de la reconstruction curviligne 3D à partir de deux images**

Les résultats issus de la reconstruction 3D à partir de deux images ne sont pas satisfaisants. Des erreurs importantes sont observées dans le déplacement de la caméra et, par conséquent, au niveau de chaque coordonnée calculée. À cause de ces erreurs, notamment en déplacement, la géométrie d'acquisition reconstruite ne correspond pas à la géométrie réelle et la configuration des points 3D dans la scène n'est pas correcte. D'ailleurs, on ne distingue pas la différence de hauteur entre les points de la grille et les points de la boîte. En fait, les solutions de transformation, calculées par l'algorithme d'optimisation, ne permettent pas l'obtention de points d'intersection entre tous les arcs de cercle. De plus, en réalisant plusieurs tests, on remarque qu'il est possible, pour les mêmes images, d'avoir la même valeur pour la fonction objective mais des solutions optimales différentes. Ceci nous amène à déduire que la configuration des arcs dans la scène ne permet pas leur intersection en des points 3D communs. Ils sont disposés d'une manière presque verticale dans la scène. Le modèle géométrique, utilisé dans ce test de reconstruction 3D à partir de deux images, ne permet pas d'aboutir à une reconstruction 3D précise de la scène.

Les résultats de la reconstruction à partir de deux images, présentés dans cette section, sont jugés insatisfaisants. Dans ce contexte, nous avons tenté d'évaluer si une amélioration ne pourrait être obtenue en exploitant trois images, l'introduction d'une vue supplémentaire devant contraindre davantage la recherche de solution en ce qui concerne le mouvement de la caméra et les coordonnées 3D des points saillants. Le paragraphe suivant est consacré à l'étude de l'impact de l'augmentation du nombre d'images sur la qualité des résultats. Ces expérimentations ainsi que les résultats obtenus y sont présentés.

4.5.1.2 Reconstruction 3D de partir de trois images

Les résultats de la reconstruction à partir de deux images n'étant pas satisfaisants, nous essayons à travers l'augmentation du nombre d'images d'améliorer la qualité de la reconstruction. En fait, en augmentant le nombre d'images, on ajoute plus de contraintes dans la recherche de solutions optimales par l'algorithme d'optimisation. Par exemple, pour un ensemble de trois images, la fonction objective va minimiser, pour un point saillant donné, la distance non pas entre deux arcs de cercle mais plutôt entre trois arcs de cercle. Le point 3D qui engendre la plus faible distance entre ces trois arcs de cercle est, ainsi, plus précis en coordonnées que celui déterminé à partir de la minimisation de la distance entre deux arcs de cercle.

Dans la présente section, la troisième image utilisée est l'image 3 de la séquence 24. Cette image est acquise par la caméra de la même position que l'image 2 mais avec un mouvement de rotation, plus particulièrement selon l'axe des Z (cf. Figure 32). Les expérimentations réalisées dans cette section visent à étudier l'effet de l'ajout de cette image sur la reconstruction 3D et à déterminer si, avec seulement un mouvement de rotation entre les angles d'acquisition des images 2 et 3, l'estimation du mouvement va s'améliorer.

L'approche multi-images décrite dans le paragraphe (4.2.2) est ainsi appliquée. Les résultats concernant le déplacement de la caméra entre les images sont présentés dans le Tableau 10. Quant aux coordonnées 3D des points, elles sont représentées dans la Figure 39. Les erreurs minimales, maximales et moyennes des coordonnées 3D calculées par rapport aux coordonnées réelles des points de la grille et de la boîte sont rassemblées, respectivement, dans le Tableau 11 et le Tableau 12. Ces résultats sont issus de la recherche sur plusieurs itérations des solutions optimales par l'algorithme d'optimisation.

Étant donné que l'algorithme d'optimisation recherche le déplacement de la caméra entre trois acquisitions, le nombre de paramètres intervenant dans ce problème d'optimisation est passé de 6 à 12 (i.e. 6 paramètres par transformation entre les images 1 et 2 et entre les images 1 et 3). L'évolution de la fonction objective et des solutions optimales, au fur et à mesure des itérations, est illustrée dans la Figure 38.

- Évolution de la fonction objective au cours de la recherche

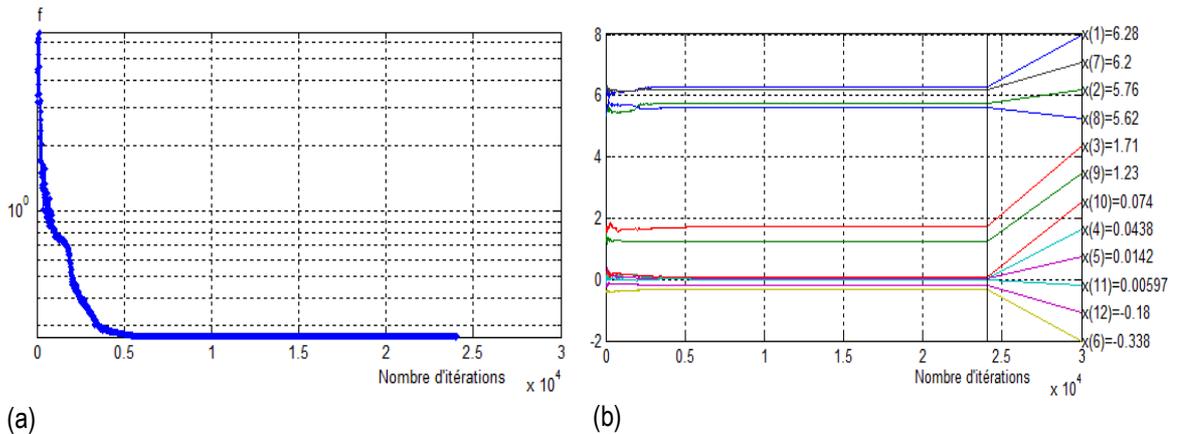


Figure 38 : Évolution de la fonction objective f (a) et des solutions optimales $[x(1), x(2), \dots, x(12)]$ du déplacement de la caméra entre les images 1 et 2 et les images 1 et 3 (b) en fonction du nombre d'itérations

La recherche des solutions optimales s'est déroulée en 24037 itérations. A l'issue de ces itérations, la fonction objective a atteint une valeur égale à 0,26 m. Par comparaison avec la reconstruction à partir de deux images (où la fonction objective est égale à 0,12 m), la valeur de la fonction objective a augmenté. Ceci est lié à l'augmentation du nombre d'images et, par conséquent, à l'augmentation du nombre de paires de points. D'ailleurs, avec une fonction objective égale à 0,26 m, l'erreur entre chaque paire de points est de l'ordre de 0,6 cm.

La valeur de la fonction objective, ci-haut citée, est un indicateur de la « non-intersection » des arcs de cercles puisqu'elle est différente du zéro. Ceci engendrera donc des erreurs au niveau des coordonnées 3D calculées. De plus, elle est proche de la valeur de la fonction objective obtenue par les transformations inter-image idéales (0,27 m). La distance inter-arc calculée est presque la même que la distance inter-arc réelle. L'algorithme d'optimisation a, par conséquent, convergé vers un optimum global.

- Déplacement de la caméra

Tableau 10. Mouvement calculé de la caméra entre les images, respectivement, (1 et 2) et (1 et 3). α, β, γ sont les angles de la rotation selon les axes X, Y et Z. T_x, T_y, T_z sont les paramètres de la translation selon les axes X, Y et Z. Ces paramètres sont calculés dans le repère R1 de l'image 1.

		Rotation (°)			Translation (m)		
		α	β	γ	T_x	T_y	T_z
Images (1-2)	Déplacement effectif	0,02	4,08	-22,74	0,72	3,62	0,00
	Déplacement estimé	2,51	0,81	-19,35	6,27	5,76	0,00
Images (1-3)	Déplacement effectif	0,31	3,76	-26,88	0,72	3,62	0,00
	Déplacement estimé	4,24	0,34	-10,32	6,20	5,61	0,00

Les déplacements affichés dans le Tableau 10 sont calculés par rapport au même repère **R1** (le repère de l'image 1). Ces déplacements correspondent aux mouvements (effectif et estimé) effectués par la caméra entre les trois images, respectivement, (1 et 2) et (1 et 3).

Par comparaison entre le déplacement estimé dans le présent test et le déplacement estimé dans la section précédente (reconstruction à partir de deux images), on remarque que le changement apporté par l'ajout de la troisième image est faible en translation et élevé en rotation. En effet, dans le Tableau 10, les valeurs montrent une modification au niveau de la rotation estimée entre l'image 1 et l'image 2 par rapport à la reconstruction précédente, en particulier au niveau de l'axe des X (α) et l'axe des Y (β). Cette modification est estimée à $(+4,14^\circ)$ selon l'axe des X et à $(-3,21^\circ)$ selon l'axe des Y. Par contre, au niveau de la translation, une faible modification (d'ordre centimétrique) est observée entre les valeurs calculées à partir de deux images et celles calculées à partir de trois images. Cette variation est, respectivement, de l'ordre de 0,02 m et 0,10 m.

Quant au déplacement entre les images 1 et 3, il affiche une translation proche de celle déterminée entre les images 1 et 2 avec une différence de l'ordre de 0,07 m en translation selon X et 0,15 m en translation selon Y. En fait, l'image 2 et l'image 3 ont été enregistrées, quasiment, à la même position avec un mouvement de rotation en Z par rapport à l'image 2. Ce mouvement est bien observé dans les résultats, la rotation calculée entre l'image 1 et l'image 3 par l'algorithme, est concentrée au niveau de l'axe des Z $(+9,03^\circ)$ par rapport à la rotation en Z, entre l'image 1 et l'image 2). Ceci coïncide avec le mouvement effectué par la caméra dans la scène. Néanmoins, entre les valeurs de chaque paramètre du déplacement calculées par l'algorithme d'optimisation et celles réelles, il y a une grande marge d'erreur.

En conclusion aux observations précédentes, le déplacement estimé par l'algorithme d'optimisation qu'il soit entre les images (1-2) ou les images (1-3) ne correspond pas au déplacement effectif. La marge d'erreur entre ces deux types de déplacement est élevée :

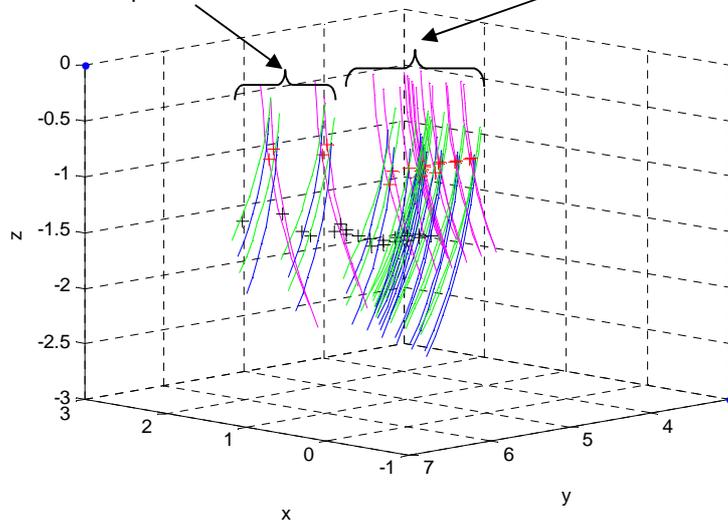
- Entre les images (1-2), en rotation, l'erreur est de l'ordre de $(-2,49, 3,27, -3,39)^\circ$, respectivement, selon les axes des X, des Y et des Z et, en translation, elle est de l'ordre de $(-5,55, 2,14, 0,00)$ m, respectivement, selon les axes des X, des Y et des Z.
- Entre les images (1-3), en rotation, l'erreur est de l'ordre de $(-3,93, 3,42, -16,56)^\circ$, respectivement, selon les axes des X, des Y et des Z et, en translation, elle est de l'ordre de $(-5,48, -1,99, 0,00)$ m, respectivement, selon les axes des X, des Y et des Z.

D'ailleurs, ceci est prévisible puisque la fonction objective a déterminé les paramètres du déplacement avec une erreur égale à 0,26 m. La géométrie d'acquisition des images dans la scène établie à partir de ces résultats est, par conséquent, incohérente avec la géométrie d'acquisition réelle.

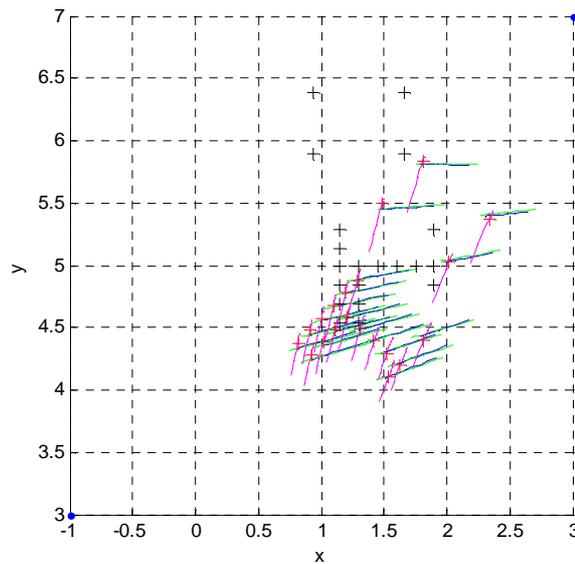
- **Reconstruction 3D des points saillants**

Arcs de cercle issus des points de la boîte

Arcs de cercle issus des points de la grille



(a)



(b)

Figure 39 : Coordonnées 3D (exprimées en mètre) calculées par l'approche curviligne. Les points en noir correspondent aux coordonnées réelles des points saillants et ceux en magenta aux coordonnées calculées. Les arcs de cercle en magenta, en vert et en bleu sont issus, respectivement, des images 1, 2 et 3. (a) Localisation 3D des points saillants dans le repère de la piscine; (b) Localisation des points saillants dans le plan (X, Y).

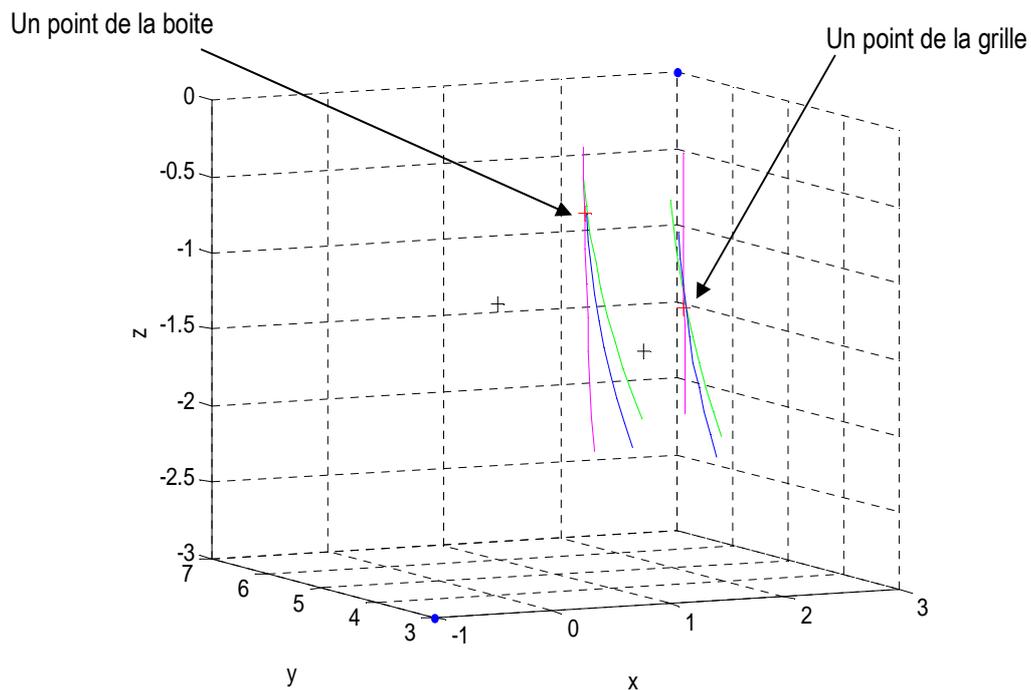


Figure 40 : Un exemple de reconstruction d'un point de la grille et d'un point de la boîte (coordonnées exprimées en mètre). Les points en noir correspondent aux coordonnées réelles des points saillants et ceux en rouge aux coordonnées calculées. Les arcs de cercle en magenta, vert et bleu sont issus des points saillants, respectivement, des images 1, 2 et 3.

Tableau 11. Les erreurs minimales, maximales et moyennes des coordonnées 3D calculées des points de la grille par rapport à leurs coordonnées 3D réelles.

Grille	Erreur minimale (m)	Erreur maximale (m)	Erreur moyenne (m)
X	0,00	0,43	0,23
Y	0,27	1,19	0,62
Z	0,02	0,51	0,15

Tableau 12. Les erreurs minimales, maximales et moyennes des coordonnées 3D calculées des points de la boîte par rapport à leurs coordonnées 3D réelles.

Boîte	Erreur minimale (m)	Erreur maximale (m)	Erreur moyenne (m)
X	0,37	0,87	0,62
Y	0,38	1,00	0,68
Z	0,53	0,68	0,61

Dans la Figure 39, Figure 40, le Tableau 11 et le Tableau 12, on remarque que la reconstruction à partir de trois images a, d'une part, amélioré les résultats au niveau de la précision en hauteur de la grille et, d'autre part, a affecté la précision de la reconstruction de la boîte. En fait, les erreurs des coordonnées 3D calculées pour les points de la boîte restent élevées par rapport aux coordonnées réelles (cf. Tableau 12). D'ailleurs, en

observant la disposition des arcs de cercle dans la scène et la position des points reconstruits, on remarque que tous les points de la boîte sont situés à l'extrémité des arcs contrairement à la majorité des points de la grille situés au milieu des arcs. Ceci signifie que les arcs de cercle des points de la boîte ne se croisent pas, ils sont disposés presque verticalement les uns par rapport aux autres. La plus faible distance entre ces arcs des points de la boîte, déterminée par la fonction objective, est au niveau de leurs extrémités.

En évaluant l'erreur moyenne des deux ensembles de points, on trouve qu'elle est égale à 0,15 m pour la grille et 0,61 m pour la boîte. L'erreur moyenne de la grille a, ainsi, diminué de 0,23 m par rapport à l'erreur moyenne calculée lors de la reconstruction à partir de deux images, contrairement aux points de la boîte qui affichent une augmentation de l'erreur moyenne de 0,12 m.

L'écart déterminé entre les hauteurs moyennes des points de la grille, estimée à 1,62 m, et des points de la boîte, estimée à 0,69 m, est égal à 0,93 m. Il présente, ainsi, une erreur élevée de l'ordre de 0,63 m par rapport à l'écart réel entre les deux ensembles de points. Toutefois, même si cet écart de hauteur ne coïncide pas avec la valeur réelle, il permet de différencier en termes de hauteur les deux groupes de points, contrairement aux résultats de la reconstruction à partir de deux images.

Finalement, malgré cette faible amélioration au niveau des hauteurs calculées, observée particulièrement au niveau des points de la grille, ces dernières ne correspondent pas aux hauteurs réelles des points. L'erreur moyenne des trois coordonnées de tous les points demeure élevée et ne permet pas de conclure que l'ajout de cette troisième image, à cause de son acquisition d'un angle de vue légèrement différent de celui de la deuxième image, a fait converger la recherche de l'algorithme vers les solutions optimales. Elle n'a pas eu un apport significatif sur la qualité de la reconstruction 3D.

- **Principales observations tirées de la reconstruction curviligne 3D à partir de trois images**

La reconstruction à partir de ces trois images spécifiques a permis d'obtenir une amélioration au niveau des hauteurs des points de la grille. En revanche, elle n'a pas permis d'apporter une meilleure précision dans le calcul des coordonnées des points de la boîte, par rapport à la reconstruction à partir de deux images. Ces résultats sont liés :

1. À la faible différence entre les angles de vue de la scène dans l'image 3 par rapport à l'image 2. Par conséquent, elle n'a pas eu beaucoup d'apport dans l'amélioration de la reconstruction de la scène, notamment au niveau des points de la boîte.
2. À la position des points dans les images. Les points de la grille sont situés au début de la portée, alors que les points de la boîte sont à la fin de la portée. Cette disposition des points par rapport à la

portée impacte les résultats parce que la résolution entre début et fin de la portée est différente. Cette dernière est plus élevée en fin de portée qu'au début. D'ailleurs, ceci affecte la reconstruction de la boîte parce que plus la résolution est grossière, plus la distance entre les arcs de cercle est grande empêchant leur intersection.

L'erreur observée dans la reconstruction de la scène est principalement due à la mauvaise estimation des solutions optimales par l'algorithme d'optimisation. Ces solutions correspondent aux paramètres du déplacement estimé qui sont utilisés pour le calcul des coordonnées 3D des points saillants. Même si le déplacement estimé affiche un comportement cohérent avec le déplacement effectif de la caméra (faible translation et une rotation dominante en Z entre les images (2-3)), l'erreur au niveau de leur estimation a affecté la qualité de la reconstruction de la scène. Cette mauvaise estimation des solutions optimales n'est pas liée à l'algorithme d'optimisation mais plutôt à plusieurs autres facteurs, dont la résolution des images (paragraphe précédent), la configuration des arcs dans la scène, le modèle géométrique utilisé, etc.

Étant donné que l'image 3 n'a pas apporté beaucoup d'amélioration dans les résultats de la reconstruction, notamment à cause du faible changement de point de vue avec l'image 2, on a ajouté, dans le test qui suit, une quatrième image acquise d'un angle de vue différent de celui des images 2 et 3. L'objectif du prochain test est non seulement d'étudier l'impact de l'image 4 sur les résultats de la reconstruction 3D mais, en outre, d'évaluer son apport dans l'estimation du mouvement entre les images 1 et 2 et les images 1 et 3. En fait, même s'il y a un mouvement relativement faible entre les images 2 et 3, l'algorithme d'optimisation ne l'a pas estimé correctement. Les expérimentations et les résultats sont présentés dans le paragraphe suivant.

4.5.1.3 Reconstruction 3D à partir de quatre images

Dans le test précédent, la reconstruction à partir de trois images a légèrement amélioré la reconstruction 3D par rapport à la reconstruction à partir de deux images. Cette amélioration est observée uniquement dans les coordonnées 3D des points de la grille. L'analyse de l'ensemble des résultats (le déplacement estimé, les coordonnées 3D, la configuration des arcs dans la scène) nous a amené à conclure qu'à cause de la similarité entre les vues des images 2 et 3, la troisième image n'a pas eu beaucoup d'apport dans la recherche de solutions optimales. Les déplacements estimés ne correspondent pas aux déplacements effectifs ni entre l'image 1 et 2 ni entre l'image 1 et 3. À travers l'ajout d'une 4^{ème} image on cherche à améliorer l'estimation de ces déplacements. La quatrième image (l'image 4 extraite de la séquence 8, cf. Figure 32) est acquise selon un angle de vue différent des autres images. À cause de l'augmentation du nombre d'images, le nombre de paramètres recherchés par l'algorithme d'optimisation a augmenté. Il y en a maintenant 18 qui correspondent aux déplacements entre les 3 paires d'images utilisées, à savoir (1 et 2), (1 et 3), (1 et 4).

L'évolution de la fonction objective ainsi que les valeurs des solutions optimales, au fur et à mesure des itérations, sont illustrées dans la Figure 41. Les déplacements calculés par l'algorithme sont affichés dans le Tableau 13. La représentation de la reconstruction obtenue ainsi que l'erreur minimale, maximale et moyenne de la hauteur des points saillants sont affichés, respectivement, dans la Figure 42, le Tableau 16 et le Tableau 17.

- **Évolution de la fonction objective au cours de la recherche**

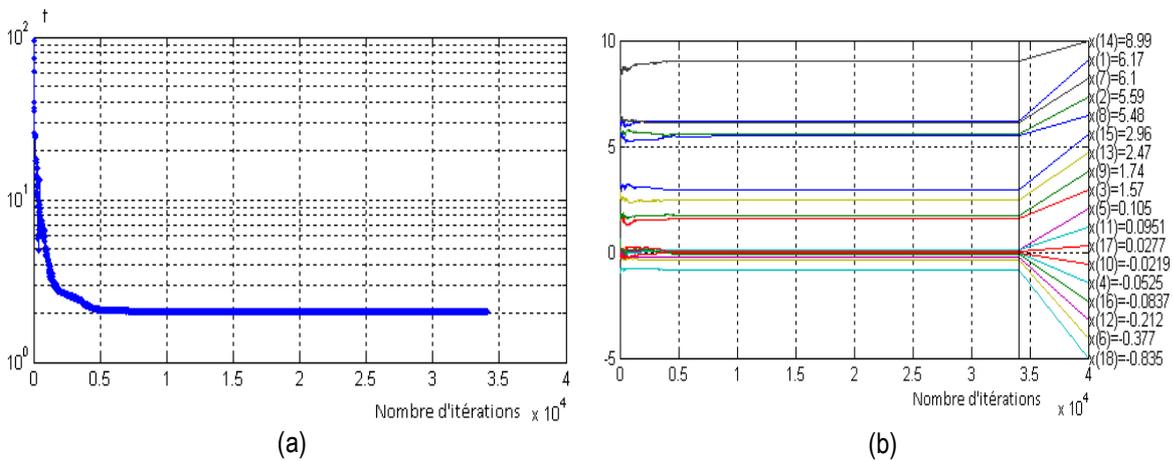


Figure 41 : Évolution de la fonction objective f (a) et des solutions optimales $[x(1), x(2), \dots, x(18)]$ du déplacement de la caméra entre les images 1 et 2, les images 1 et 3 et les images 1 et 4 (b) en fonction du nombre d'itérations

Selon la Figure 41, au bout de la 34042^{ème} l'algorithme d'optimisation a convergé vers une solution de rotation et de translation pour laquelle la fonction objective atteint une valeur égale à 2,06 m. Cette valeur d'erreur est élevée par rapport à la valeur de la fonction objective obtenue avec les transformations inter-images idéales (0,57 m) et par rapport aux reconstructions précédentes. Avec une telle valeur, on a une erreur de l'ordre de 0,03 m pour chaque paire de points (une reconstruction avec 4 images implique la minimisation d'une distance entre 60 paires de points). Elle peut affecter, par conséquent, la qualité des coordonnées 3D calculées. De plus, une telle valeur montre que, comme dans les reconstructions précédentes, l'algorithme n'a pas abouti à une intersection entre tous les arcs de cercle des points saillants.

- Déplacement de la caméra

Tableau 13. Mouvement calculé de la caméra entre les images, respectivement, (1 et 2), (1 et 3) et (1 et 4). α , β , γ sont les angles de la rotation selon les axes des X, Y et Z. T_x , T_y , T_z sont les paramètres de la translation selon les axes des X, Y et Z. Ces paramètres sont calculés dans le repère R1 de l'image 1.

		Rotation (°)			Translation (m)		
		A	β	γ	T_x	T_y	T_z
Images (1-2)	Déplacement effectif	0,02	4,08	-22,74	0,72	3,62	0,00
	Déplacement estimé	-3,00	6,01	-21,59	6,17	5,59	0,00
Images (1-3)	Déplacement effectif	0,31	3,76	-26,88	0,72	3,62	0,00
	Déplacement estimé	-1,25	5,44	-12,15	6,09	5,48	0,00
Images (1-4)	Déplacement effectif	-4,48	0,96	-33,02	-3,94	3,62	0,00
	Déplacement estimé	-4,79	1,58	-47,85	2,47	8,99	0,00

Le Tableau 13 affiche le mouvement de la caméra calculé entre les quatre images. Tous les déplacements (effectifs et estimés) sont calculés dans le repère **R1** de l'image 1.

Les déplacements obtenus ont été modifiés par rapport à ceux issus de la reconstruction à partir de 3 images. Ces modifications concernent les paramètres de la rotation et les paramètres de la translation. Une amélioration est, ainsi, observée dans les déplacements entre les images 1 et 2 et les images 1 et 3. L'erreur en rotation s'est réduite de même que l'erreur en translation. Par conséquent, malgré l'augmentation de la valeur de la fonction objective, l'image 4 a eu un apport conséquent sur l'estimation des paramètres de déplacement par l'algorithme d'optimisation. Ceci est lié, notamment, à la différence dans l'angle de vue de l'image 4 par rapport aux images 1, 2 et 3. Elle a impacté la recherche de solutions de déplacement en faisant converger l'algorithme vers des solutions optimales meilleures

Ces résultats restent, quand même, insatisfaisants vu qu'ils sont différents des déplacements effectifs de la caméra entre les différentes positions d'acquisition des séquences. Ils présentent des erreurs de rotation et des erreurs de translation. Ces erreurs sont présentées, respectivement, dans le Tableau 14 et le Tableau 15.

Tableau 14. L'erreur en translation de la caméra entre les images (1 et 2), (1 et 3) et (1 et 4) selon les trois axes (X, Y et Z).

	Erreur en translation (m)		
	T_x	T_y	T_z
Images (1 et 2)	5,45	1,97	0,00
Images (1 et 3)	5,37	1,86	0,00
Images (1 et 4)	1,47	5,37	0,00

Tableau 15. L'erreur en rotation de la caméra entre les images (1 et 2), (1 et 3) et (1 et 4) selon les trois axes (X, Y et Z).

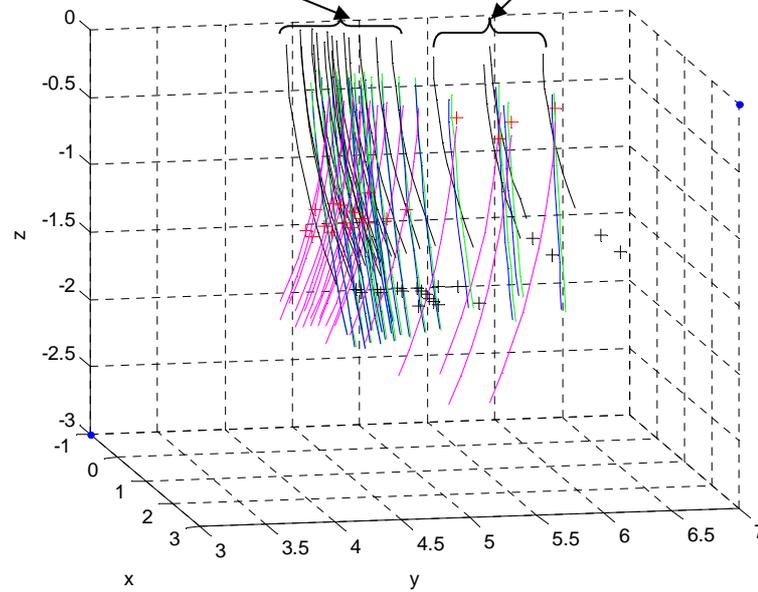
	Erreur en rotation (°)		
	α	β	γ
Images (1 et 2)	3,02	1,93	1,15
Images (1 et 3)	1,56	1,68	14,73
Images (1 et 4)	0,31	0,62	14,83

Même si l'erreur de la rotation estimée en Z est élevée par rapport à sa valeur réelle, la rotation entre l'image 1 et l'image 4 a un comportement cohérent avec la géométrie d'acquisition réelle. En fait, afin d'acquérir une séquence d'image d'un angle de vue différent de celui des images 2 et 3, la caméra a fait un mouvement dominant en rotation selon l'axe des Z. Ceci apparaît dans le mouvement estimé par l'algorithme d'optimisation.

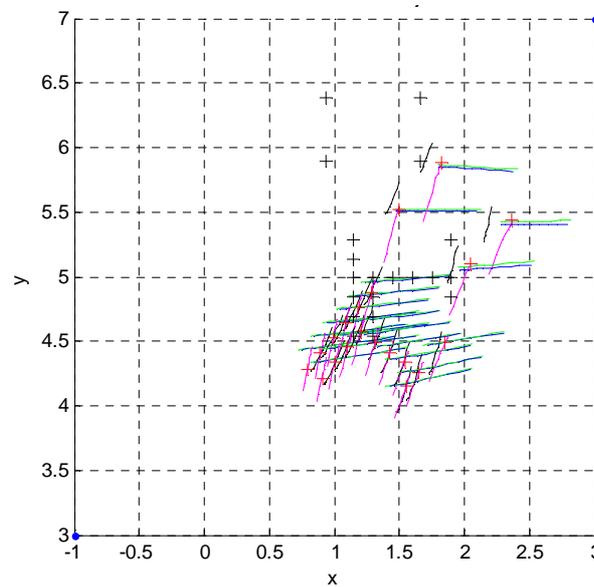
- Reconstruction 3D des points saillants

Arcs de cercle issus des points de la grille

Arcs de cercle issus des points de la boîte



(a)



(b)

Figure 42 : Coordonnées 3D (exprimées en mètre) calculées par l'approche curviligne. Les points en noir correspondent aux coordonnées réelles des points saillants et ceux en rouge aux coordonnées calculées : (a) Localisation 3D des points saillants; (b) Localisation des points saillants dans le plan (X, Y). Les arcs de cercle en magenta, vert, bleu et noir sont issus des points saillants, respectivement, des images 1, 2, 3 et 4

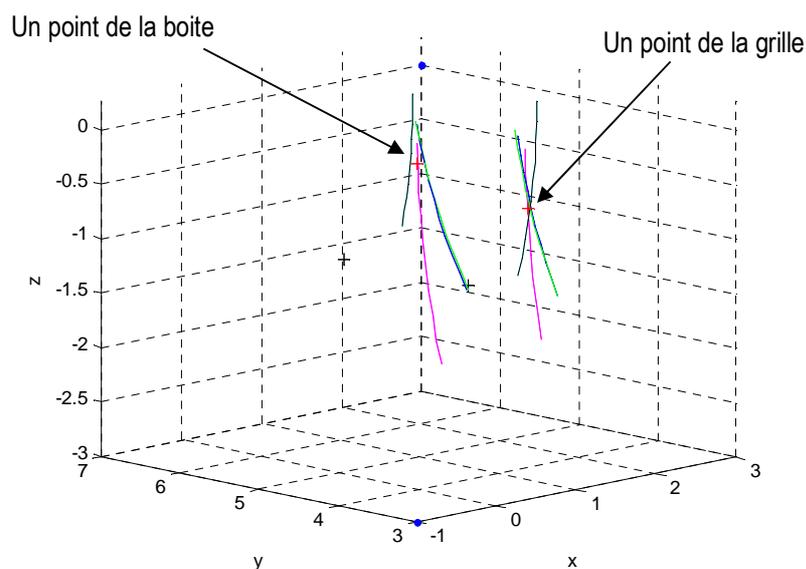


Figure 43 : Un exemple de reconstruction d'un point de la grille et d'un point de la boîte (coordonnées exprimées en mètre). Les points en noir correspondent aux coordonnées réelles des points saillants et ceux en rouge aux coordonnées calculées. Les arcs de cercle en magenta, vert, bleu et noir sont issus des points saillants, respectivement, des images 1, 2, 3 et 4

Tableau 16. Les erreurs minimales, maximales et moyennes des coordonnées 3D calculées pour les points de la grille par rapport à leurs coordonnées réelles.

Grille	Erreur minimale (m)	Erreur maximale (m)	Erreur moyenne (m)
X	0,03	0,39	0,19
Y	0,25	0,80	0,46
Z	0,34	0,84	0,55

Tableau 17, Les erreurs minimales, maximales et moyennes des coordonnées 3D calculées pour les points de la boîte par rapport à leurs coordonnées réelles.

Boîte	Erreur minimale (m)	Erreur maximale (m)	Erreur moyenne (m)
X	0,38	0,89	0,63
Y	0,37	0,95	0,46
Z	0,84	1,12	1,04

L'analyse du Tableau 16 et du Tableau 17 montre que les erreurs moyennes des coordonnées 3D calculées des points restent très élevées. Les erreurs moyennes, notamment en hauteur sont supérieures à celles calculées lors de la reconstruction à partir de trois images. L'amélioration n'est observée qu'au niveau des coordonnées en X et en Y des points où une diminution de quelques centimètres est enregistrée. Ces erreurs sont prévisibles, puisque l'erreur globale déterminée par la fonction objective est relativement élevée (estimée à 2,06 m) affectant ainsi la précision du calcul des coordonnées 3D. Quant à l'écart en hauteur moyenne entre les points de la grille (1,05 m) et les points de la boîte (0,26 m), il est égal à 0,79 m. Cette valeur est largement supérieure à la valeur réelle. D'après ces résultats, l'ajout d'une quatrième image n'induit pas une amélioration

de la qualité de la reconstruction. Au contraire, on observe une augmentation de l'erreur au niveau des coordonnées des points de la boîte et au niveau de la différence de hauteur entre les deux ensembles de points (grille et boîte).

- **Principales observations tirées de la reconstruction curviligne 3D à partir de quatre images**

La reconstruction 3D multi-images, par l'approche curviligne en utilisant quatre images, a donné des résultats insatisfaisants. En effet, en reconstruction de points, la marge d'erreur de calcul des points a augmenté, en particulier au niveau des points de la boîte. Même si l'erreur moyenne en hauteur des points de la grille est élevée, elle reste plus faible que celle des points de la boîte. Pour la même raison citée dans le test précédent, cette différence au niveau de la précision de calcul entre les deux ensembles de points est liée à la géométrie de formation des images, plus le point est loin du centre de la caméra plus sa résolution est faible.

La configuration des points dans la scène est correcte, on retrouve la disposition réelle des points les uns par rapport aux autres. Toutefois, on remarque un décalage entre ces points reconstruits et la disposition des points réels. Ceci est observé, aussi, dans tous les tests de reconstructions précédents. Ce décalage est dû à la mauvaise estimation des paramètres du déplacement même s'ils présentent une cohérence avec le mouvement effectué par la caméra lors de l'acquisition (mouvement de rotation, faible translation, etc.)

4.5.2 Évaluation de la performance de l'approche curviligne

Suite à l'évaluation des tests de reconstruction 3D, deux conclusions importantes ont été tirées :

- 1) *La performance de l'approche curviligne est affectée par la résolution des images et le modèle géométrique utilisé.*

Sur les différentes solutions de reconstruction 3D déterminées en fonction du nombre d'images avec l'approche curviligne, les erreurs de calcul des coordonnées des points en X et en Y, illustrées sur les figures de représentation des reconstructions 3D, sont faiblement variables, tout particulièrement en X. Elles sont plus considérables pour les points de la boîte que pour les points de la grille. En ce qui concerne la hauteur des points (coordonnée en Z), les résultats sont insatisfaisants pour l'ensemble des points puisqu'ils présentent une marge d'erreur élevée et ne s'améliorent pas en fonction du nombre d'images. La comparaison entre les erreurs des deux ensembles de points (boîte et grille) nous a amené à déduire que la position des points dans l'image impacte la précision de calcul des coordonnées. D'ailleurs, la résolution des images en azimut au niveau des points de la grille est de l'ordre de 0,02 m alors qu'elle est de l'ordre de 0,06 m au niveau des points de la boîte. De plus, on retrouve ce même ordre de grandeur d'erreur entre les points de la boîte et ceux de la grille (cf. Tableau 11 et Tableau 12).

Il n'est pas possible d'intégrer la résolution des images dans l'approche curviligne. En fait, le modèle géométrique utilisé pour la projection des points dans la scène est un modèle simplifié. Il associe, à tout pixel d'une image, un arc de cercle dans la scène. Cet arc de cercle correspond à l'ensemble des points de la scène qui peuvent contribuer à l'intensité du pixel. Il est défini par un azimut, une distance connue par rapport à la caméra (portée) et une ouverture verticale égale à 10° .

2) *L'approche curviligne ne permet pas d'estimer l'imprécision des coordonnées 3D calculées*

L'inconvénient majeur de l'approche curviligne est la difficulté de l'évaluation des imprécisions des coordonnées 3D calculées. Elle ne fournit qu'une solution unique en (X, Y, Z) pour les points reconstruits. Cette solution correspond aux coordonnées des points qui ont produit la plus faible distance entre les différents arcs de cercle issus de la projection des points saillants des images dans la scène. En fait, les tests de reconstruction ont montré qu'il est difficile d'aboutir à une intersection entre les arcs de cercle et de déduire en conséquence les coordonnées 3D des points d'intersection avec précision. Cet inconvénient est une conséquence de l'utilisation du modèle géométrique simplifié dans l'approche curviligne.

Dans la section suivante, on présente les tests de la reconstruction 3D par approche volumique où le modèle géométrique idéal de la caméra a été utilisé.

4.6 Application de l'approche volumique

Comme dans le cas de l'approche curviligne, l'approche volumique a été testée sur les séquences d'images réelles acquises en bassin. Afin de comparer les résultats de l'approche volumique et ceux de l'approche curviligne, les mêmes tests ont été réalisés à savoir que le même ensemble de points saillants et les mêmes images ont été utilisés. L'approche volumique vise à déterminer un intervalle de solutions possibles en termes de valeur d'élévation, alors que l'approche curviligne vise à déterminer une valeur unique d'élévation pour chaque point. Dans ce contexte, sur les figures qui vont être proposées, dans les paragraphes suivants, pour illustrer la position 3D des points saillants, l'ensemble de coordonnées en Z possibles correspondant à l'intervalle de solutions d'élévation pour chaque point fourni par l'approche volumique sera présenté.

Les expérimentations présentées dans cette section sont réalisées avec des paramètres de résolution égaux à [$\Delta\theta = 0,9^\circ$ et $\Delta r = 0,20$ m]. Nous cherchons avec ces paramètres à déduire les zones d'intersection entre des rubans volumiques ayant des dimensions assez larges. En fait, la dimension de ces paramètres prend en compte la résolution du capteur et inclut aussi l'imperfection issue des traitements appliqués pour l'extraction des points saillants. Sachant que la résolution en azimut de la caméra DIDSON est égale à $0,3^\circ$, en spécifiant une résolution en azimut égale à $0,9^\circ$ dans les expérimentations, on tient compte d'une erreur d'un pixel de chaque côté du pixel du point saillant, dans la direction de l'azimut. Quant à la résolution en portée de la

caméra DIDSON, elle dépend des caractéristiques d'acquisition telles que la portée maximale et la portée minimale c'est-à-dire la distance par rapport à laquelle commence l'enregistrement de l'image. Le choix d'une résolution en portée égale à 0,2 m, pour les images utilisées dans ces expérimentations, permet d'intégrer une erreur en distance qui s'étend sur 8 pixels de chaque côté du pixel du point saillant.

4.6.1 Reconstruction 3D à partir de deux images

Les images utilisées dans ce paragraphe sont l'image 1 de la séquence 17 et l'image 2 de la séquence 21. Le calcul des solutions optimales des transformations entre les images est réalisé sur plusieurs itérations. La condition d'arrêt de ces itérations est l'atteinte d'une fonction objective égale à 0,00 m. Dès que la fonction objective atteint cette valeur, l'algorithme termine la recherche des solutions optimales et calcule par la suite les coordonnées 3D des points saillants (cf. Figure 44).

Le mouvement de la caméra entre les deux acquisitions estimé par l'approche volumique est présenté dans le Tableau 18 et comparé à celui obtenu par l'approche curviligne dans le paragraphe qui suit le tableau. Les intervalles de solutions des coordonnées en Z, calculées pour chaque point saillant, sont définis par une valeur minimale et une autre maximale. Ils sont affichés dans le Tableau 19 (points de la grille) et le Tableau 20 (points de la boîte). Ces intervalles sont illustrés dans la Figure 46.

- **Évolution de la fonction objective au cours de la recherche**

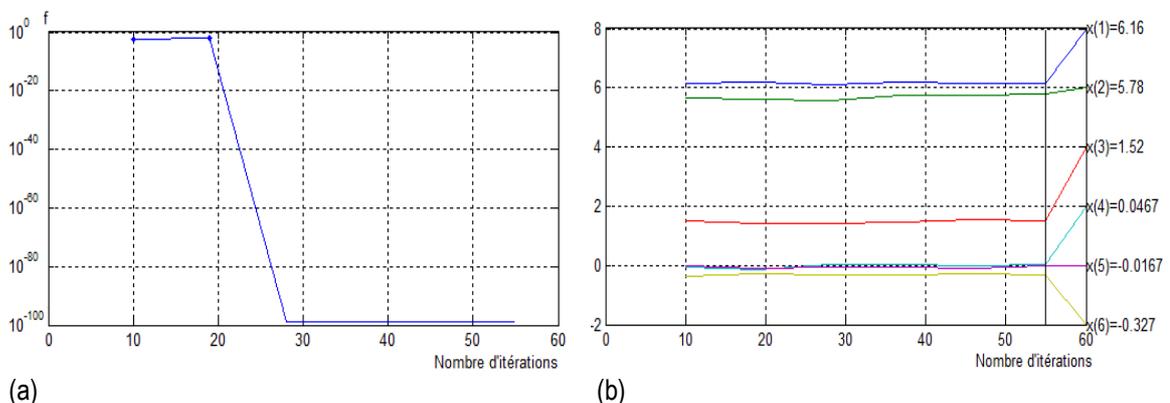


Figure 44 : Évolution de la fonction objective f (a) et des solutions optimales $[x(1), x(2), \dots, x(6)]$ du déplacement de la caméra entre les images 1 et 2 en fonction du nombre d'itérations.

L'estimation des paramètres de la rotation et de la translation par l'approche volumique est réalisée en un faible nombre d'itérations comparativement à l'approche curviligne. D'ailleurs, le nombre d'itérations n'a pas dépassé les 56 itérations, au bout desquelles la fonction objective a atteint une valeur égale à 0,00 m (la même valeur de la fonction objective obtenue à partir de la transformation idéale). Cette valeur indique que les

rubans volumiques des points saillants se croisent et que les solutions optimales ont été calculées avec une erreur nulle.

- **Déplacement de la caméra**

Tableau 18. Mouvement de la caméra entre les images 1 et 2 par l'approche volumique (a) et par approche curviligne (a'), calculé dans le repère R1 de l'image 1. α , β , γ sont les angles de la rotation selon, respectivement, l'axe des X, des Y et des Z. T_x , T_y , T_z sont les paramètres de la translation selon, respectivement, l'axe des X, des Y et des Z.

	Rotation (°)			Translation (m)		
	α	β	γ	T_x	T_y	T_z
Déplacement effectif	0,02	4,08	-22,74	0,72	3,62	0,00
Déplacement estimé (a)	2,67	0,95	-18,71	6,16	5,77	0,00
Déplacement estimé (a')	-1,63	3,44	-20,74	6,25	5,66	0,00

Le Tableau 18 affiche les déplacements (effectif et estimé) effectué par la caméra entre les images (1 et 2). Ces déplacements sont calculés dans le repère **R1** de l'image 1.

Par comparaison avec le déplacement estimé par l'approche curviligne, une faible différence dans les valeurs des paramètres de la translation est observée. En fait, selon l'approche curviligne les paramètres de la translation sont égaux à [$T_x = 6,25$, $T_y = 5,66$ et $T_z = 0,00$] m. Ils ont, ainsi, une différence d'ordre centimétrique avec les paramètres de la translation du présent test. Par contre, un décalage entre les paramètres de rotation estimés par les deux approches est remarquable (en X = +4,03°, en Y = -2,49° et en Z = 2,03°). Ce changement dans les résultats induira une différence dans la reconstruction 3D de la scène à partir des deux approches. Il est issu de la différence entre les caractéristiques des deux approches (notamment le modèle géométrique utilisé). En fait, la résolution intégrée dans l'approche volumique a conduit l'algorithme d'optimisation à la détermination de solutions de déplacement qui ont permis l'intersection entre les rubans volumique. Par contre, dans l'approche curviligne, les solutions de déplacement calculées n'ont pas conduit à l'intersection entre les arcs de cercle. Néanmoins, les solutions du déplacement estimé affichent, quand même, des erreurs élevée en rotation (-2,65, 3,13, -4,03)° et en translation (5,44, 2,15, 0,00) m. Ces erreurs affecteront la précision de calcul des coordonnées 3D de la scène, notamment l'intervalle des hauteurs des points reconstruits. Dans le paragraphe suivant, on va présenter et analyser les intervalles des hauteurs calculés à partir de la transformation obtenue par l'approche volumique.

- Reconstruction 3D des points saillants

Tableau 19. Les intervalles des coordonnées 3D des points de la grille

Points de la grille								
ID	1	2	3	4	5	6	7	8
Intervalle des Z	[-1,44,	[-1,45,	[-1,45,	[-1,44,	[-1,45,	[-1,44,	[-1,48,	[-1,44,
m [Zmin, Zmax]	-0,84]	-0,91]	-0,82]	-0,88]	-0,80]	-0,86]	-1,03]	-0,77]
Intervalle des X	[0,83,	[0,94,	[0,93,	[1,04,	[1,03,	[1,14,	[1,56,	[1,12,
m [Xmin, Xmax]	0,86]	0,97]	0,96]	1,07]	1,06]	1,17]	1,61]	1,16]
Intervalle des Y	[4,32,	[4,23,	[4,43,	[4,34,	[4,54,	[4,45,	[4,09,	[4,64,
m [Ymin, Ymax]	4,47]	4,37]	4,58]	4,48]	4,69]	4,59]	4,20]	4,79]
ID	9	10	11	12	13	14	15	16
Intervalle des Z	[-1,45,	[-1,44,	[-1,44,	[-1,46,	[-1,47,	[-1,45,	[-1,44,	[-1,45,
m [Zmin, Zmax]	-0,84]	-0,90]	-0,96]	-1,00]	-1,03]	-0,75]	-0,75]	-1,04]
Intervalle des X	[1,23,	[1,34,	[1,45,	[1,55,	[1,66,	[1,22,	[1,32,	[1,86,
m [Xmin, Xmax]	1,27]	1,38]	1,49]	1,60]	1,71]	1,26]	1,36]	1,90]
Intervalle des Y	[4,55,	[4,47,	[4,38,	[4,29,	[4,20,	[4,75,	[4,86,	[4,42,
m [Ymin, Ymax]	4,69]	4,60]	4,50]	4,40]	4,31]	4,90]	5,00]	4,51]

Tableau 20. Les intervalles des coordonnées 3D des points de la boîte

Points de la boîte				
ID	1	2	3	4
Intervalle des Z m [Zmin, Zmax]	[-1,51, -0,66]	[-1,41, -0,62]	[-1,51, -0,60]	[-1,40, -0,55]
Intervalle des X m [Xmin, Xmax]	[1,50, 1,55]	[2,05, 2,10]	[1,84, 1,88]	[2,38, 2,44]
Intervalle des Y m [Ymin, Ymax]	[5,43, 5,58]	[5,00, 5,14]	[5,78, 5,93]	[5,36, 5,49]

D'après les résultats affichés dans le Tableau 19 et le Tableau 20 et la Figure 45 (affichée ci-dessous), l'approche volumique a réussi à effectuer la reconstruction 3D de la scène. Elle a pu, d'une part, déterminer des zones d'intersection entre les rubans issus des images et, d'autre part, déduire des intervalles de solutions incluant la hauteur réelle de certains points.

Selon le Tableau 19 et le Tableau 20, les intervalles des coordonnées en Z affichent des solutions de hauteurs différentes entre les points de la grille et les points de la boîte. Ils présentent les hauteurs issues des élévations des points qui ont induit une erreur de reconstruction nulle donc une intersection entre les rubans volumiques. Ce résultat est prévisible puisque ces deux ensembles de points (les points de la boîte et les points de la grille) ont des caractéristiques différentes. Ils ont des hauteurs différentes dans la scène et aussi des positions différentes dans les images. La Figure 46 présente une illustration graphique de ces intervalles.

Quant aux intervalles des coordonnées en X et Y, présentés dans le Tableau 19 et le Tableau 20, ils n'incluent pas les coordonnées réelles des points. Par comparaison aux coordonnées réelles en X et en Y, l'erreur de calcul s'avère plus importante pour les points de la boîte que pour les points de la grille. Ceci est dû, probablement, aux différentes positions des points de la boîte et de la grille en portée et en azimut dans l'image.

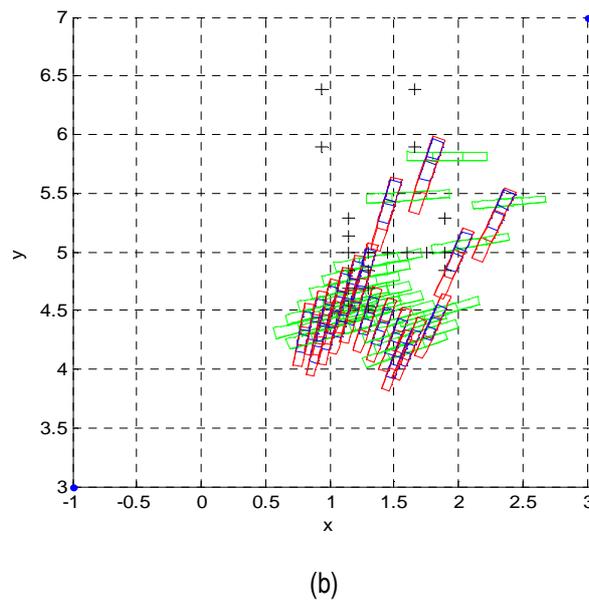
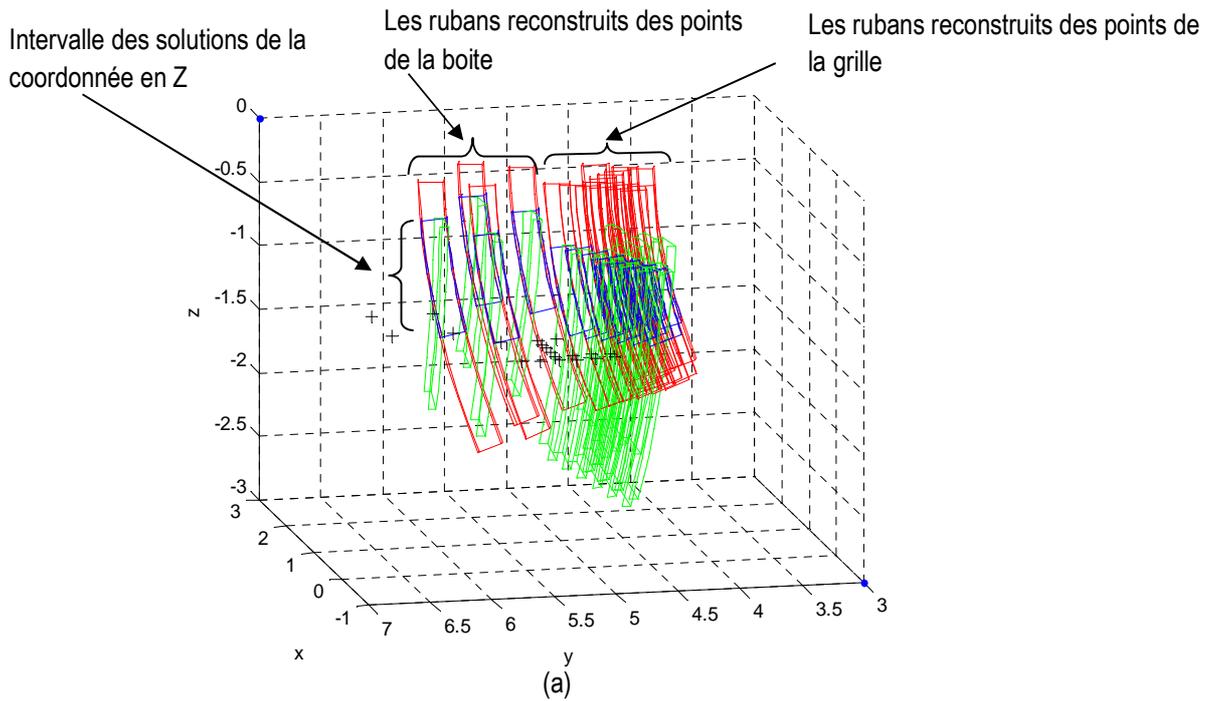


Figure 45 : Coordonnées 3D (exprimées en mètre) calculées par l'approche volumique. Les points en noir correspondent aux coordonnées réelles. Les rubans en bleu sont les intervalles des hauteurs possibles des points saillants. Les rubans en rouge et en vert sont issus des points saillants, respectivement, des images 1 et 2: (a) Localisation des hauteurs possibles des points saillants; (b) Localisation des rubans et des points réels dans le plan (X, Y).

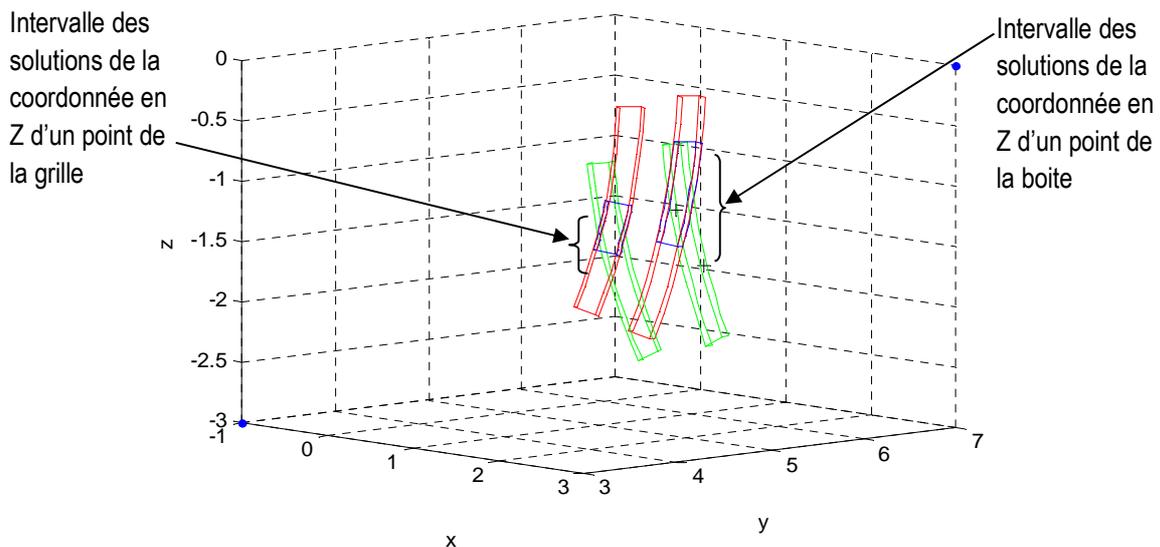


Figure 46 : Exemple de reconstruction d'un point de la grille et d'un point de la boîte (coordonnées 3D exprimées en mètre) par l'approche volumique. Les points en noir correspondent aux coordonnées réelles. Les rubans en bleu sont les intervalles des hauteurs possibles des points saillants. Les rubans en rouge et en vert sont issus des points saillants, respectivement, des images 1 et 2: (a) Localisation des hauteurs possibles des points saillants; (b) Localisation des rubans et des points réels dans le plan (X, Y).

L'intersection des rubans volumiques des deux images (rubans en rouge et rubans en vert) est illustrée par les zones en bleu (cf. Figure 45 et Figure 46). Cette intersection est le résultat de l'intégration de la résolution dans l'approche volumique et, donc, l'utilisation du modèle idéal de la caméra. À cause de cette résolution, les rubans se présentent sous forme d'un volume qui s'étend sur toute l'ouverture verticale des images (10°), sur un azimut de $0,9^\circ$ et une portée de 0,20 m. De plus, avec une telle résolution, la configuration des rubans dans la scène permet de déterminer les solutions en hauteur possibles des points saillants. Ces hauteurs correspondent aux élévations qui ont engendré des distances minimales en élévation entre les rubans et en distance entre les paires de polygones composant les rubans volumiques. Toutes ces hauteurs sont ainsi des candidates déterminées par l'approche volumique et qui peuvent correspondre à la hauteur réelle de chaque point.

La connaissance *a priori* des coordonnées réelles des points nous a permis de déduire les points dont la hauteur réelle figure dans ces intervalles. En fait, tous les intervalles des points de la boîte contiennent la hauteur réelle des points (-1,31 m). Ceci n'est pas le cas pour les intervalles des points de la grille où on ne retrouve pas les hauteurs réelles parmi les solutions possibles. Ces résultats sont liés d'une part à la dimension des intervalles, les intervalles des points de la boîte sont plus larges et présentent plus de solutions en hauteur. D'autre part, les zones d'intersection sont déterminées à partir du déplacement estimé par l'algorithme d'optimisation. Or, le déplacement, entre l'image 1 et l'image 2, est estimé avec une marge d'erreur par rapport au déplacement effectif.

Par comparaison avec la reconstruction par l'approche curviligne, l'approche volumique présente un meilleur résultat. En fait, dans l'approche curviligne, aucun point de la boîte ne présente une hauteur qui correspond à la réalité.

- **Principales observations tirées de la reconstruction volumique 3D à partir de deux images**

Contrairement à l'approche curviligne où il s'est avéré difficile d'aboutir à une intersection entre les arcs de cercle, l'intégration de la résolution dans l'approche volumique a contribué à l'extraction de zones d'intersection entre les rubans volumiques. Ces zones fournissent un ensemble de hauteurs possibles pour chaque point saillant. Ces hauteurs appartiennent à des intervalles de solutions assez larges en dimension. Bien que certaines hauteurs réelles se soient retrouvées dans ces intervalles (points de la boîte), il serait plus avantageux d'avoir des intervalles de solutions plus réduits en dimension afin de réduire le nombre de solutions possibles pour chaque point. Ainsi, dans le paragraphe suivant on va présenter des expérimentations de reconstruction 3D à partir de trois images pour étudier l'effet de l'augmentation du nombre d'images sur la dimension de ces intervalles de solutions et sur l'augmentation du nombre d'intervalles qui incluent la hauteur réelle des points.

4.6.2 Reconstruction 3D à partir de trois images

Comme décrit dans le paragraphe précédent, la reconstruction à partir de deux images par l'approche volumique a permis l'extraction de zones d'intersection entre les rubans volumiques des points. Néanmoins, ces zones se caractérisent par une dimension assez large, donc, un nombre élevé de solution de hauteurs. De plus, ce ne sont pas toutes les zones d'intersection qui incluent les hauteurs réelles des points. Alors, afin d'améliorer la qualité des résultats issus de la reconstruction à partir de deux images, nous avons testé l'approche volumique sur les trois images 1, 2 et 3, respectivement, des séquences 17, 21 et 24.

Le mouvement de la caméra est calculé entre les trois images (cf. Tableau 21). Comme dans le paragraphe précédent, l'approche volumique permet d'estimer l'intervalle de solutions de hauteurs pour chaque point saillant. Ces intervalles sont limités par une valeur minimale Z_{min} et une valeur maximale Z_{max} . Ils sont déterminés sur plusieurs itérations (cf. Figure 47) et affichent une différence par rapport à ceux déterminés dans le paragraphe précédent (cf. Tableau 22 et Figure 23). Ils sont illustrés dans la Figure 48.

- Évolution de la fonction objective au cours de la recherche

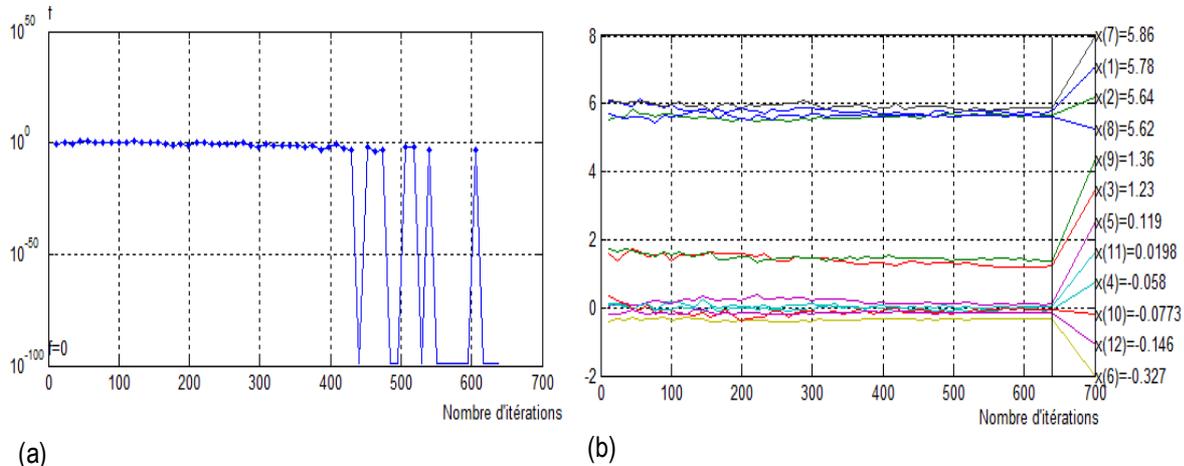


Figure 47 : Évolution de la fonction objective f (a) et des solutions optimales $[x(1), x(2), \dots, x(12)]$ du déplacement de la caméra entre les images 1 et 2 et les images 1 et 3 en fonction du nombre d'itérations.

Les solutions optimales relatives à la rotation et à la translation ont été estimées tout au long des 640 itérations. La fonction objective a atteint une valeur minimale égale à 0,00 m indiquant que l'algorithme a trouvé un optimum global. La valeur nulle de la fonction objective informe aussi qu'une intersection entre tous les rubans volumiques des points saillants a été déterminée.

- Déplacement de la caméra

Tableau 21. Mouvement de la caméra entre les images 1 et 2 et les images 1 et 3 par approche volumique (a et b) et par approche curviligne (a' et b'), calculé dans le repère R1 de l'image 1. α, β, γ sont les angles de la rotation selon, respectivement, l'axe des X, des Y et des Z. T_x, T_y, T_z sont les paramètres de la translation selon, respectivement, l'axe des X, des Y et des Z.

		Rotation (°)			Translation (m)		
		α	β	γ	T_x	T_y	T_z
Images (1-2)	Déplacement effectif	0,02	4,08	-22,74	0,72	3,62	0,00
	Déplacement estimé (a)	-3,32	6,82	-18,74	5,78	5,64	0,00
	Déplacement estimé (a')	2,51	0,81	-19,35	6,27	5,76	0,00
Images (1-3)	Déplacement effectif	0,31	3,76	-26,88	0,72	3,62	0,00
	Déplacement estimé (b)	-4,43	1,13	-8,34	5,86	5,62	0,00
	Déplacement estimé (b')	4,24	0,34	-10,32	6,20	5,61	0,00

Le Tableau 21 présente les déplacements (effectifs et estimés) entre les trois images 1, 2 et 3 dans le repère de l'image 1 (R1).

Le déplacement calculé entre l'image 1 et l'image 2 affiche des paramètres de rotation en Z et de translation proches de ceux calculés par reconstruction 3D à partir de deux images. Toutefois, une différence est observée au niveau de la rotation selon l'axe des Y et l'axe des X de l'ordre de $[+6,00^\circ$ et $-7,77^\circ]$. C'est l'ajout de la troisième image qui a engendré ce changement dans le calcul du déplacement entre les images 1 et 2. Quant au déplacement entre l'image 1 et l'image 3, il affiche une faible variation en translation par rapport au déplacement entre l'image 1 et l'image 2. Ceci correspond à la réalité des acquisitions puisque les images 2 et 3 sont extraites de deux séquences acquises de, quasiment, la même position (position 2). La rotation entre l'image 1 et l'image 3 est cohérente avec la géométrie d'acquisition des images, même si sa valeur ne correspond pas à la valeur réelle.

En comparant les valeurs des paramètres estimés et effectifs, on déduit que le déplacement a été calculé avec une erreur élevée. Donc, même si l'algorithme d'optimisation a estimé les solutions optimales des déplacements avec une fonction objective nulle, ceci ne signifie pas que ces solutions sont exactes et correspondent à la réalité d'acquisition. En fait, des sources d'imperfection peuvent contribuer à ce résultat. Ces imperfections sont liées aux erreurs dans l'extraction des points des images et aussi à la non-calibration du modèle géométrique idéal utilisé.

- **Reconstruction 3D des points saillants**

Tableau 22. Les intervalles des coordonnées 3D des points de la grille

Points de la grille								
ID	1	2	3	4	5	6	7	8
Intervalle des Z m	[-1,42,	[-1,43,	[-1,42,	[-1,43,	[-1,43,	[-1,43,	[-1,49,	[-1,43,
[Zmin, Zmax]	-1,17]	-1,17]	-1,16]	-1,16]	-1,15]	-1,15]	-1,17]	-1,14]
Intervalle des X m	[0,84,	[0,94,	[0,93,	[1,04,	[1,03,	[1,14,	[1,56,	[1,12,
[Xmin, Xmax]	0,85]	0,96]	0,95]	1,06]	1,05]	1,16]	1,60]	1,14]
Intervalle des Y m	[4,33,	[4,24,	[4,44,	[4,34,	[4,54,	[4,45,	[4,08,	[4,65,
[Ymin, Ymax]	4,40]	4,31]	4,51]	4,42]	4,62]	4,53]	4,17]	4,72]
ID	9	10	11	12	13	14	15	16
Intervalle des Z m	[-1,43,	[-1,44,	[-1,46,	[-1,47,	[-1,49,	[-1,43,	[-1,43,	[-1,48,
[Zmin, Zmax]	-1,15]	-1,14]	-1,15]	-1,16]	-1,17]	-1,13]	-1,13]	-1,15]
Intervalle des X m	[1,23,	[1,34,	[1,45,	[1,55,	[1,66,	[1,22,	[1,32,	[1,86,
[Xmin, Xmax]	1,25]	1,36]	1,47]	1,58]	1,69]	1,24]	1,34]	1,89]
Intervalle des Y m	[4,56,	[4,47,	[4,37,	[4,28,	[4,19,	[4,76,	[4,86,	4,41,
[Ymin, Ymax]	4,63]	4,55]	4,46]	4,37]	4,28]	4,83]	4,94]	4,49]

Tableau 23. Les intervalles des coordonnées 3D des points de la boîte

ID	Points de la boîte			
	1	2	3	4
Intervalle des Z m [Zmin, Zmax]	[-1,31, -1,15]	[-1,35, -1,00]	[-1,32, -1,28]	[-1,32, -1,01]
Intervalle des X m [Xmin, Xmax]	[1,52, 1,53]	[2,05, 2,08]	[1,85, 1,85]	[2,39, 2,41]
Intervalle des Y m [Ymin, Ymax]	[5,47, 5,51]	[5,02, 5,09]	[5,82, 5,83]	[5,37, 5,43]

L'objectif de l'augmentation du nombre d'image dans le présent test a été atteint. En fait, la troisième image ajoutée à la reconstruction a modifié la dimension des intervalles des hauteurs des points en la réduisant et en augmentant par conséquent la précision des solutions calculées (cf. Tableau 22 et Tableau 23). Le Tableau 24 et le Tableau 25 illustrent la valeur de la réduction de la dimension de ces intervalles entre la reconstruction à partir de deux images et la reconstruction à partir de trois images. La réduction de ces intervalles est considérable pour les points de la boîte (cf. Tableau 25). Par contre, elle est moins élevée pour les points de la grille (cf. Tableau 24). La précision du calcul de la coordonnée en Z des points de la boîte est ainsi meilleure que celle des points de la grille. Une illustration de ces intervalles est présentée dans la Figure 48. Quant aux coordonnées en X et en Y, l'augmentation du nombre d'images n'a pas influencé leur calcul. En fait, les intervalles des coordonnées en X et Y sont presque les mêmes que ceux calculés à partir de deux images. La modification est de l'ordre de 1 cm. Ainsi, l'écart entre les coordonnées réelles et calculées en X et en Y a été conservé.

La résolution choisie pour la reconstruction de la scène est plus adéquate pour la reconstruction des points de la boîte que pour la reconstruction des points de la grille. D'ailleurs, malgré la réduction de leur dimension, les quatre intervalles des points de la boîte incluent la hauteur réelle des points.

Tableau 24. Différence entre les dimensions des intervalles de hauteurs des points de la grille reconstruits à partir de deux images (intervalles 1) et trois images (intervalles 2)

ID	Points de la grille							
	1	2	3	4	5	6	7	8
Différence m (dimension intervalle1 - dimension intervalle 2)	0,35	0,28	0,36	0,28	0,37	0,30	0,12	0,38
ID	9	10	11	12	13	14	15	16
Différence m (dimension intervalle 1 - dimension intervalle 2)	0,31	0,24	0,17	0,14	0,12	0,39	0,39	0,08

Tableau 25. Différence entre les dimensions des intervalles de hauteur des points de la boîte reconstruits à partir de deux images (intervalles 1) et trois images (intervalles 2)

ID	Points de la boîte			
	1	2	3	4
Différence m (dimension intervalle 1 - dimension intervalle 2)	0,68	0,44	0,86	0,54

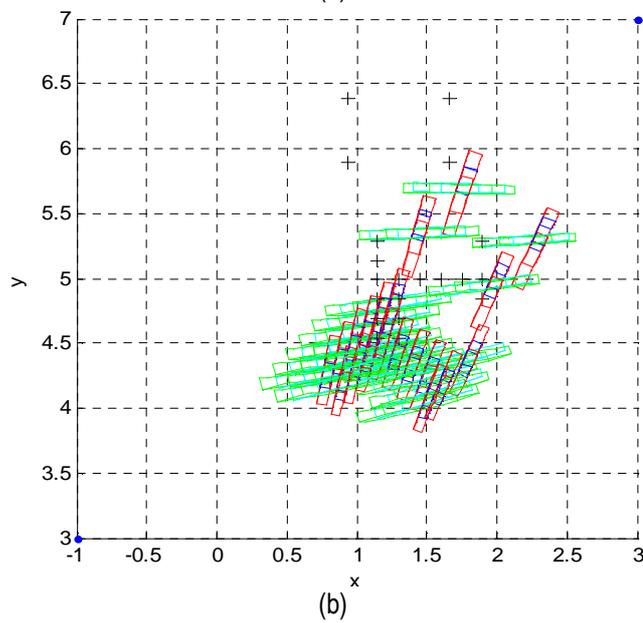
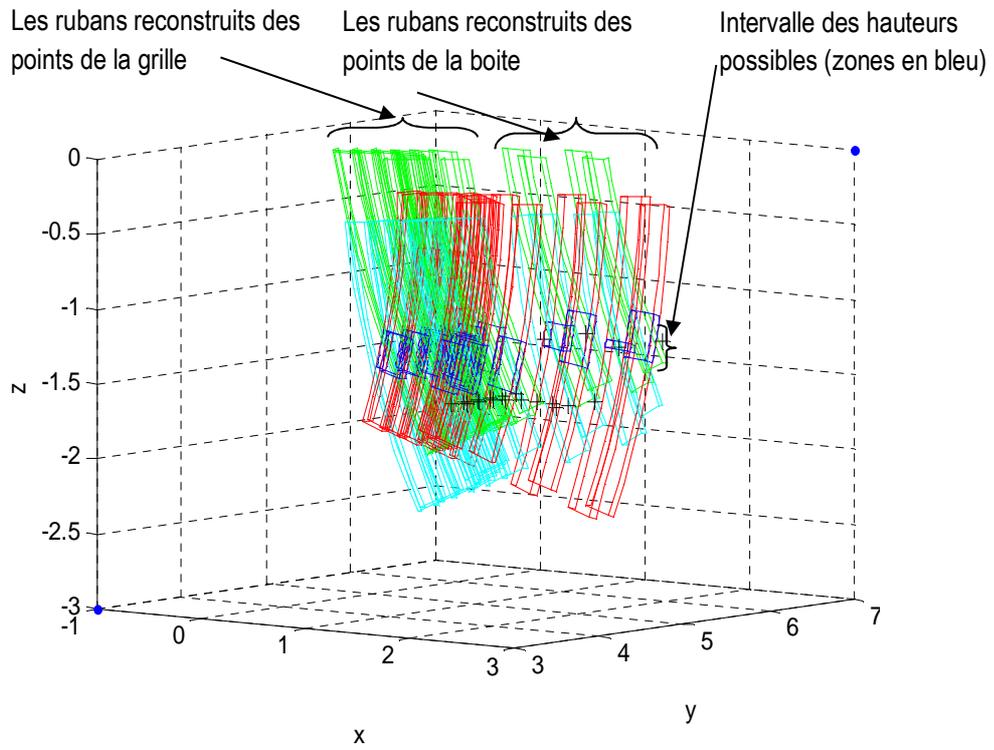


Figure 48 : Coordonnées 3D (exprimées en mètre) calculées par l'approche volumique. Les points en noir correspondent aux coordonnées réelles. Les rubans en bleu sont les intervalles des hauteurs possibles des points saillants. Les rubans en rouge, en vert et en cyan sont issus des points saillants, respectivement, des images 1, 2 et 3: (a) Localisation des hauteurs possibles des points saillants; (b) Localisation des rubans et des points réels dans le plan (X, Y).

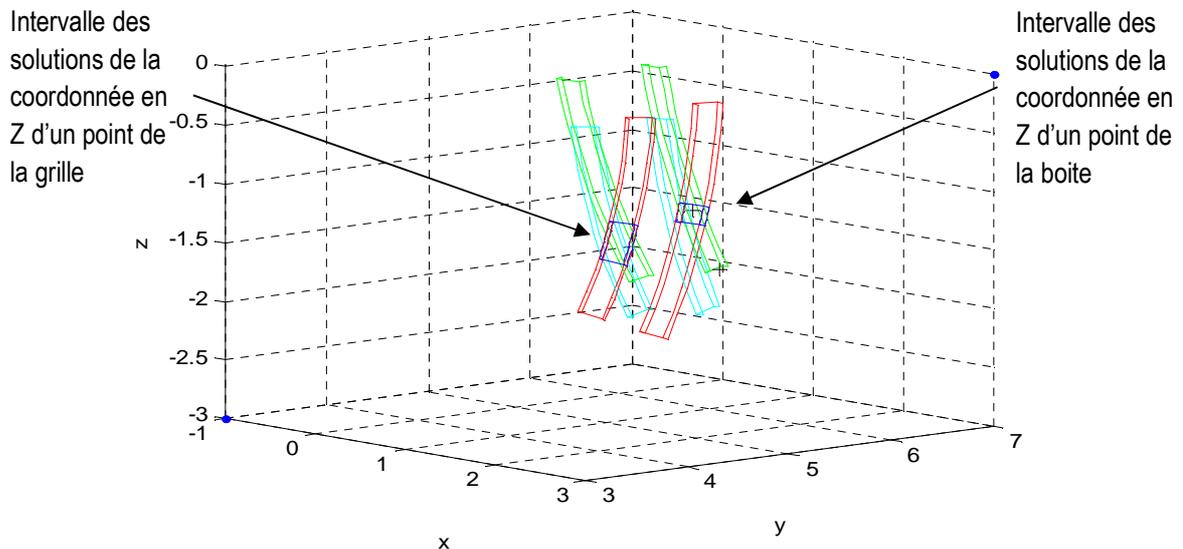


Figure 49 : Exemple de reconstruction d'un point de la grille et d'un point de la boîte par l'approche volumique (coordonnées 3D exprimées en mètre). Les points en noir correspondent aux coordonnées réelles. Les rubans en bleu sont les intervalles des hauteurs possibles des points saillants. Les rubans en rouge, en vert et en cyan sont issus des points saillants, respectivement, des images 1, 2 et 3: (a) Localisation des hauteurs possibles des points saillants; (b) Localisation des rubans et des points réels dans le plan (X, Y).

Comme discuté dans le paragraphe précédent, les intervalles de hauteurs des points de la grille et des points de la boîte ont des dimensions et des solutions différentes de celles obtenues dans la reconstruction à partir de deux images. En fait, la reconstruction à partir de trois images a induit une réduction de la dimension des intervalles des hauteurs pour chaque point reconstruit et, aussi, une augmentation de l'écart en hauteur entre les deux ensembles de points. Cet écart est égal à 0,12 m. La recherche de zones d'intersection, pour un point saillant donné, entre trois rubans a changé les solutions en élévation qui ont induit une intersection entre deux rubans dans la reconstruction précédente.

Afin de bien distinguer la réduction de la dimension des intervalles et l'augmentation de l'écart en hauteur, nous représentons, dans la Figure 50, un exemple de deux points (un point de la grille et un point de la boîte) reconstruits à partir de deux images et à partir de trois images.

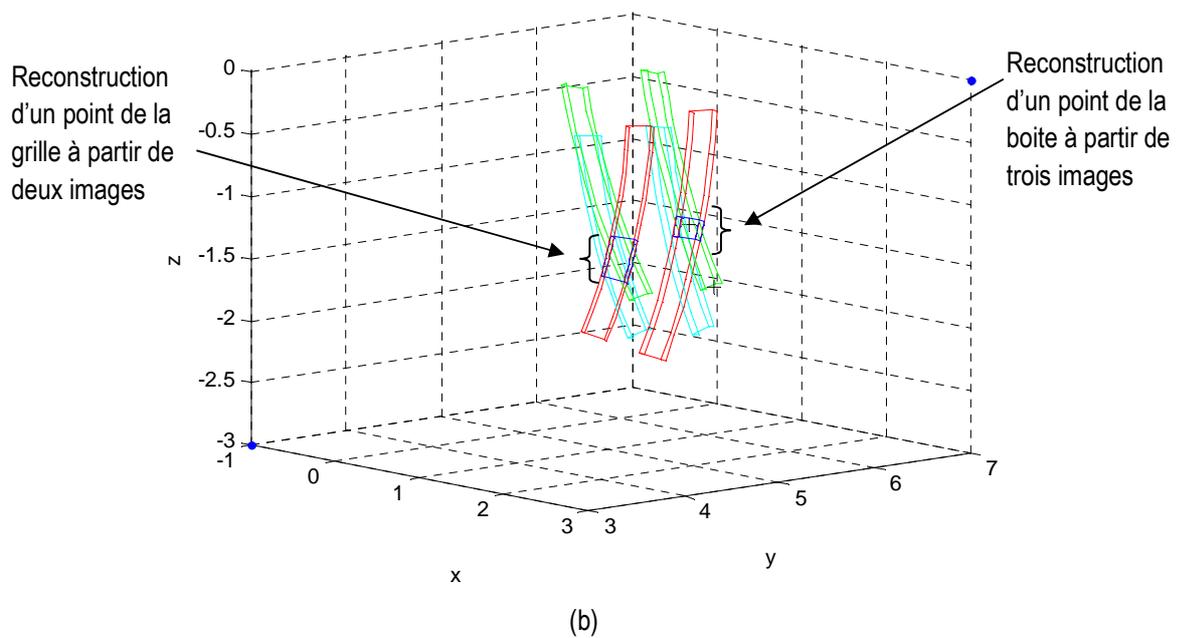
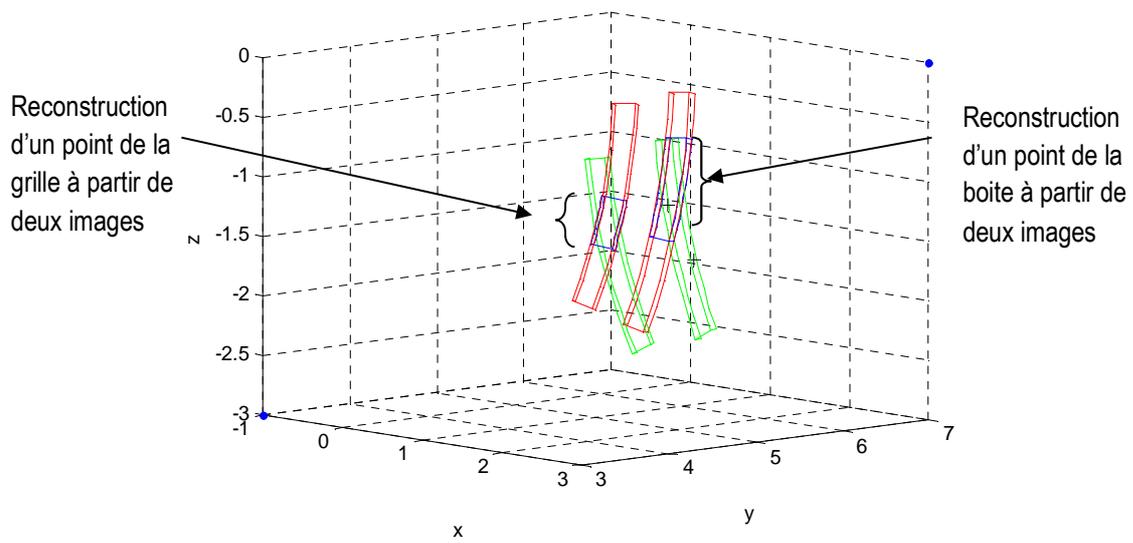


Figure 50 : Coordonnées 3D (exprimées en mètre) d'un point de la grille et d'un point de la boîte calculées par l'approche volumique à partir de deux images et de trois images. Les points en noir correspondent aux coordonnées réelles. Les rubans en bleu sont les intervalles des hauteurs possibles des points saillants. Les rubans en rouge, en vert et en cyan sont issus des points saillants, respectivement, des images 1, 2 et 3: (a) Reconstruction à partir de deux images; (b) Reconstruction à partir de trois images

- **Principales observations tirées de la reconstruction volumique 3D à partir de trois images**

La reconstruction à partir de trois images, par l'approche volumique, a apporté plus de précision en réduisant les intervalles de solutions de hauteurs possibles pour chaque point. En comparant ces résultats avec ceux issus de l'approche curviligne, on déduit que l'approche volumique apporte plus d'information sur la hauteur réelle des points. Dans l'approche curviligne, chaque point est représenté par une hauteur unique qui ne correspond pas à sa hauteur réelle. Alors que dans l'approche volumique, non seulement on a un ensemble de solutions possibles pour chaque point mais la hauteur réelle de quelques points, notamment les points de la boîte figure aussi dans ces solutions. C'est l'utilisation du modèle géométrique idéal de la caméra qui a permis à l'approche volumique de produire des résultats plus satisfaisants que ceux de l'approche curviligne.

Étant donné que toutes les hauteurs réelles ne figurent pas dans les intervalles de solution calculés (notamment la hauteur des points de la grille), nous réalisons, dans la prochaine section, un test de reconstruction à partir de quatre images. On cherche, ainsi à évaluer l'effet de l'ajout d'une quatrième image sur la précision des intervalles de hauteur calculés. Les résultats obtenus sont abordés dans le paragraphe qui suit.

4.6.3 Reconstruction 3D à partir de quatre images

Les résultats présentés et discutés dans le test précédent, ont montré que l'augmentation du nombre d'images a amélioré la reconstruction de la scène. Cette amélioration concerne l'augmentation de la précision des solutions de hauteurs. Toutefois, on a remarqué que contrairement aux points de la boîte, les hauteurs réelles des points de la grille ne figurent pas parmi les solutions calculées. Dans ce contexte, il apparaît pertinent d'évaluer s'il y a un gain effectif à ajouter encore plus d'images afin d'améliorer le calcul des hauteurs des points de la grille et des paramètres du mouvement entre les images. Des tests impliquant quatre images ont donc été réalisés.

Ainsi, nous avons ajouté l'image 4 aux images utilisées dans le paragraphe précédent. Le mouvement calculé entre les images (1 - 2, 1 - 3 et 1 - 4) est présenté dans le Tableau 26. L'évolution de la fonction objective et des solutions optimales est présentée dans la Figure 51. Quant aux intervalles des solutions des coordonnées en Z, ils sont présentés dans le Tableau 27 et le Tableau 28.

- Évolution de la fonction objective au cours de la recherche

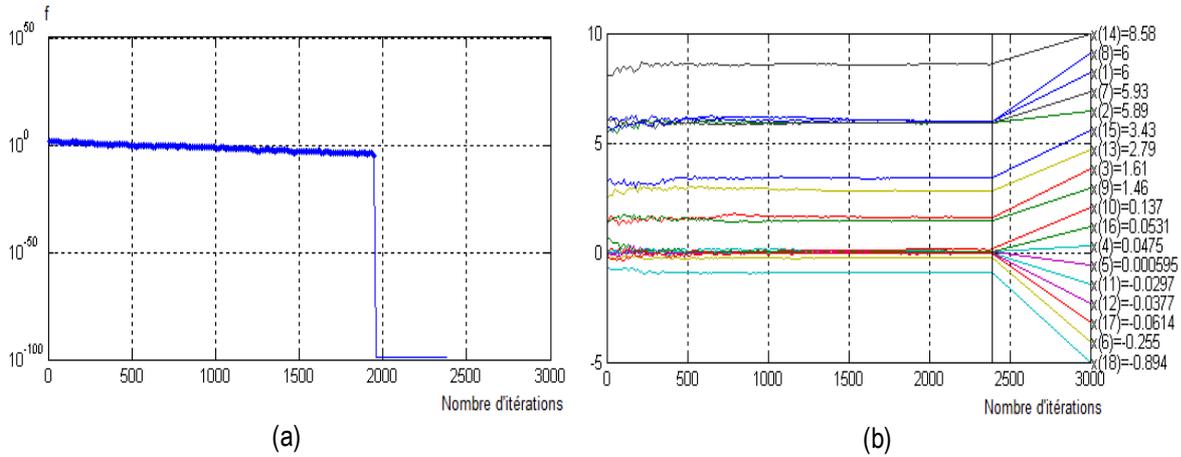


Figure 51 : Évolution de la fonction objective f (a) et des solutions optimales $[x(1), x(2), \dots, x(18)]$ du déplacement de la caméra entre les images 1 et 2, les images 1 et 3 et les images 1 et 4 en fonction du nombre d'itérations.

L'estimation des solutions optimales des paramètres de la rotation et de la translation est effectuée sur un nombre d'itérations encore plus élevé que celui de la reconstruction à partir de trois images. L'augmentation du nombre d'itérations est due à l'augmentation du nombre d'images utilisées, ce qui implique un temps de calcul plus élevé pour la recherche des solutions optimales qui permettent l'intersection entre tous les rubans des points saillants. La fonction objective a atteint la valeur minimale zéro au bout de 2390 itérations (cf. Figure 51).

- Déplacement de la caméra

Tableau 26. Mouvement de la caméra entre les images 1 et 2, les images 1 et 3 et les images 1 et 4 par approche volumique (a, b et c) et par approche curviligne (a', b' et c'), calculé dans le repère R1 de l'image 1. α, β, γ sont les angles de la rotation selon, respectivement, l'axe des X, des Y et des Z. T_x, T_y, T_z sont les paramètres de la translation selon, respectivement, l'axe des X, des Y et des Z.

		Rotation (°)			Translation (m)		
		α	β	γ	T_x	T_y	T_z
Images (1-2)	Déplacement effectif	0,02	4,08	-22,74	0,72	3,62	0,00
	Déplacement estimé (a)	2,71	0,03	-14,60	5,99	5,88	0,00
	Déplacement estimé (a')	-3,00	6,01	-21,59	6,17	5,59	0,00
Images (1-3)	Déplacement effectif	0,31	3,76	-26,88	0,72	3,62	0,00
	Déplacement estimé (b)	7,82	-1,70	-2,16	5,93	5,99	0,00
	Déplacement estimé (b')	-1,25	5,44	-12,15	6,09	5,48	0,00
Images (1-4)	Déplacement effectif	-4,48	0,96	-33,02	-3,94	3,62	0,00
	Déplacement estimé (c)	3,04	3,51	-51,23	2,78	8,58	0,00
	Déplacement estimé (c')	-4,79	1,58	-47,85	2,47	8,99	0,00

Le Tableau 26 présente les déplacements (effectifs et estimés) effectués par la caméra entre les quatre images. Ces déplacements sont calculés dans le repère **R1** de l'image 1.

Le mouvement de la caméra calculé par l'approche volumique en utilisant quatre images est différent du mouvement calculé dans les paragraphes précédents, notamment en rotation. Par contre, les mêmes conclusions tirées de l'analyse du mouvement calculé par reconstruction à partir de trois images sont vérifiées dans le présent mouvement:

1) les paramètres de translation calculés selon les trois axes entre les images (1 - 2) et (1 - 3) sont très proches (la différence est d'ordre centimétrique), ceci est cohérent avec la géométrie d'acquisition des images puisque les deux images sont extraites de deux séquences acquises, quasiment, de la même position;

2) l'écart en rotation selon l'axe des Z est égal à $12,44^\circ$. Il est proche de celui déterminé par reconstruction à partir de trois images entre les images (1 - 2) et (1 - 3). Ceci confirme la cohérence des résultats avec la géométrie d'acquisition. Entre ces deux images, il existe un mouvement de rotation en Z dominant.

En ce qui concerne la transformation entre l'image 1 et l'image 4, elle affiche une rotation élevée selon l'axe des Z. En fait, l'image 4 est acquise d'une position située à l'opposé de la position d'acquisition de l'image 1, la valeur élevée de la rotation selon l'axe des Z par rapport aux valeurs de la rotation selon l'axe des X et des Y correspond à la réalité de la configuration d'acquisition des images.

Même si le mouvement estimé entre les trois paires d'images correspond à la configuration d'acquisition, les erreurs relatives à ces paramètres, notamment en translation, peuvent induire des imprécisions au niveau du calcul de l'élévation des points saillants.

- Reconstruction 3D des points saillants

Tableau 27. Les intervalles des coordonnées 3D des points de la grille

Points de la grille								
ID	1	2	3	4	5	6	7	8
Intervalle des Z m	[-1,74,	[-1,71,	[-1,78,	[-1,76,	[-1,83,	[-1,81,	[-1,73,	[-1,86,
[Zmin, Zmax]	-1,68]	-1,62]	-1,67]	-1,60]	-1,65]	-1,56]	-1,72]	-1,63]
Intervalle des X m	[0,81,	[0,92,	[0,91,	[1,01,	[1,00,	[1,10,	[1,53,	[1,09,
[Xmin, Xmax]	0,82]	0,93]	0,91]	1,03]	1,01]	1,13]	1,53]	1,11]
Intervalle des Y m	[4,22,	[4,14,	[4,31,	[4,23,	[4,40,	[4,32,	[3,93,	[4,50,
[Ymin, Ymax]	4,24]	4,17]	4,35]	4,29]	4,47]	4,41]	4,00]	4,59]
ID	9	10	11	12	13	14	15	16
Intervalle des Z m	[-1,84,	[-1,82,	[-1,80,	[-1,78,	[-1,77,	[-1,91,	[-1,87,	[-1,93,
[Zmin, Zmax]	-1,53]	-1,57]	-1,62]	-1,68]	-1,74]	-1,59]	-1,53]	-1,76]
Intervalle des X m	[1,20,	[1,30,	[1,41,	[1,51,	[1,62,	[1,18,	[1,28,	[1,79,
[Xmin, Xmax]	1,23]	1,33]	1,43]	1,53]	1,63]	1,21]	1,31]	1,82]
Intervalle des Y m	[4,42,	[4,34,	[4,26,	[4,18,	[4,10,	[4,60,	[4,72,	[4,26,
[Ymin, Ymax]	4,53]	4,43]	4,32]	4,21]	4,11]	4,71]	4,83]	4,32]

Tableau 28. Les intervalles des coordonnées 3D des points de la boîte

Points de la boîte				
ID	1	2	3	4
Intervalle des Z m				
[Zmin, Zmax]	[-1,55, -1,42]	[-1,73, -1,44]	[-1,43, -1,35]	[-1,62, -1,50]
Intervalle des X m				
[Xmin, Xmax]	[1,51, 1,51]	[2,01, 2,04]	[1,84, 1,85]	[2,36, 2,37]
Intervalle des Y m				
[Ymin, Ymax]	[5,42, 5,45]	[4,92, 4,99]	[5,80, 5,82]	[5,31, 5,34]

Le Tableau 27 et le Tableau 28 présentent les intervalles de hauteur pour les points de la grille et les points de la boîte obtenus à partir de la reconstruction à partir de quatre images.

Les résultats de la reconstruction à partir de quatre images permettent de tirer les cinq conclusions suivantes :

1) Plusieurs points de la grille présentent un intervalle de solutions de coordonnées en Z correct c'est-à-dire que la hauteur réelle figure parmi les solutions proposées dans les intervalles. La quatrième image ajoutée a contribué à faire converger la recherche de l'algorithme d'optimisation vers des solutions plus adéquates avec la reconstruction des points de la grille. Elle présente la scène sous une vue très différente de celle de l'image 1. D'ailleurs, l'orientation des rubans des points de l'image 4 (les rubans en magenta, cf.

Figure 52 et Figure 53) est différente de l'orientation des rubans issus des images 1, 2 et 3. La recherche des zones d'intersection entre les ensembles de quatre rubans a, par conséquent, donné des solutions différentes.

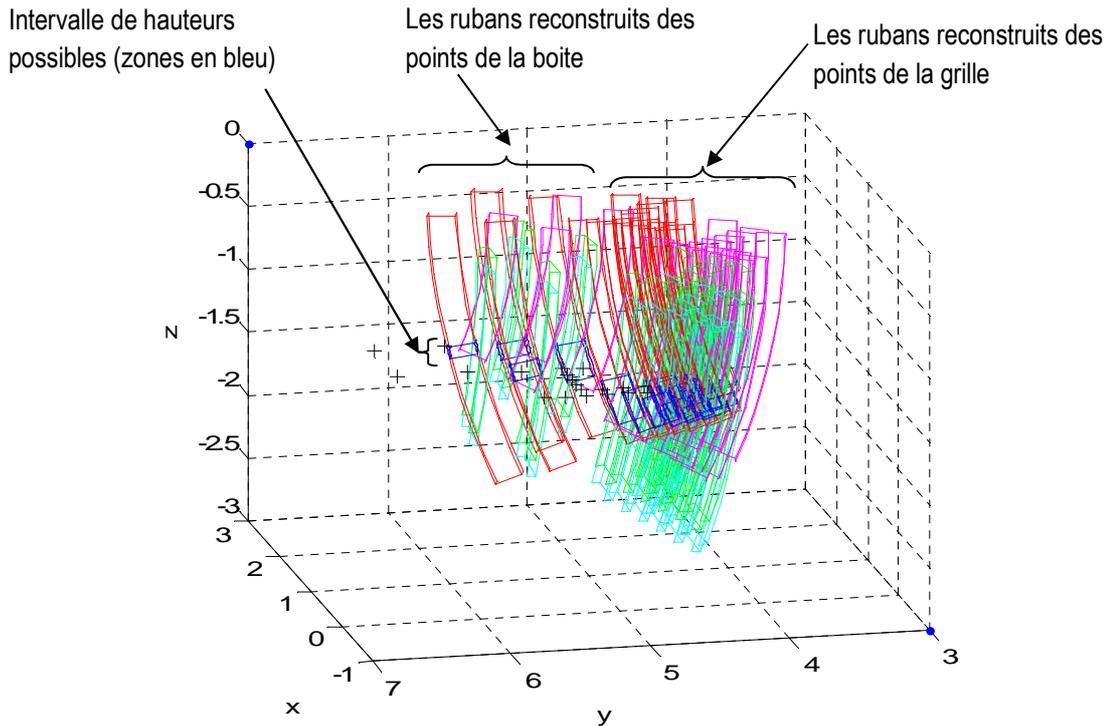
2) Les points de la grille dont la hauteur n'appartient pas à l'ensemble de solutions possibles ont des limites supérieures très proches de la hauteur réelle. Par exemple, l'intervalle de hauteurs du point 8 a une limite supérieure égale à -1,63 m, elle présente une faible différence de 2 cm par rapport à la hauteur réelle du point (-1,61 m). Une telle erreur (2 cm) n'affecte pas la reconstruction de la scène. En fait, tous les intervalles de hauteurs des points de la grille sont situés à un même niveau de hauteur dans la scène, mais avec des dimensions faiblement variable (cf. Figure 52). Par rapport à l'ouverture verticale de chaque ruban, la position de ces intervalles change. Par exemple, pour les rubans rouges, issus de l'image 1, ces intervalles se situent à l'extrémité inférieure du ruban. Par contre, pour les rubans en verts, ils se trouvent presque au milieu de l'ouverture verticale (cf. Figure 52). La dimension et la position en hauteur des intervalles de solution dépendent de la configuration des rubans dans la scène. Ajouter une image, acquise d'un point de vue différent, impacte les résultats en changeant la position des zones d'intersection entre les rubans.

3) Les intervalles présentent une dimension plus réduite que ceux issus de la reconstruction avec trois images, aussi bien pour les points de la grille que pour les points de la boîte. La Figure 52 illustre ces intervalles de solutions pour chaque point. Pour les mêmes raisons expliquées dans les deux premiers points, les zones d'intersection entre quatre rubans ont des dimensions plus réduites que ceux entre trois rubans, surtout quand il s'agit de rubans de différentes orientations dans la scène.

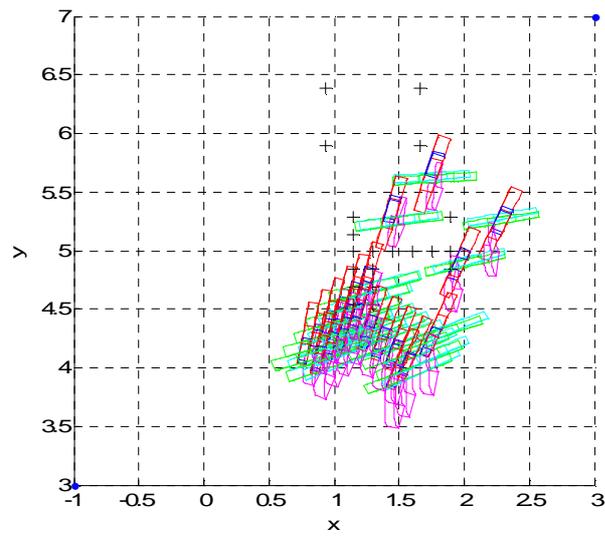
4) Contrairement aux reconstructions précédentes, les points de la boîte ont des intervalles de coordonnées en Z qui n'incluent plus leur hauteur réelle. La reconstruction à partir de quatre images a ainsi amélioré la reconstruction volumique des points de la grille et a tendance à éloigner la reconstruction des points de la boîte des solutions de hauteurs obtenues dans la reconstruction précédente. La quatrième image a influencé la recherche des solutions d'élévation qui induisent une intersection entre les quatre rubans. Les zones d'intersection entre les rubans des points de la boîte ont, par conséquent, changé de position par rapport à la reconstruction précédente. Toutefois, comme pour les points de la grille, les intervalles de hauteur des points de la boîte se situent tous à un même niveau d'élévation dans la scène. Sur la Figure 52, on peut différencier la présence de deux groupes d'intervalles avec deux hauteurs moyennes différentes.

5) L'écart moyen en hauteur entre les points de la grille et les points de la boîte est égal à 0,23 m. Cet écart a augmenté par rapport à celui calculé lors de la reconstruction à partir de trois images (0,12 m). Il est plus proche de l'écart réel entre les deux ensembles de points (0,30 m) (cf. Figure 53). La reconstruction à partir de quatre images a, ainsi, amélioré la reconstruction 3D de la scène en introduisant un écart en hauteur entre les points situés à des élévations différentes.

6) Les intervalles des coordonnées en X et en Y ont une dimension plus réduite que celle obtenue en reconstruction à partir de deux et de trois images. L'augmentation du nombre d'images a ainsi augmenté la précision de ces coordonnées. Toutefois, un écart entre les coordonnées réelles et les coordonnées calculées est toujours observé.



(a)



(b)

Figure 52 : Coordonnées 3D (exprimées en mètre) calculées par l'approche volumique. Les points en noir correspondent aux coordonnées réelles. Les rubans en bleu sont les intervalles des hauteurs possibles des points saillants. Les rubans en rouge, en vert, en cyan et en magenta sont issus des points saillants, respectivement, des images 1, 2, 3 et 4: (a) Localisation des hauteurs possibles des points saillants; (b) Localisation des rubans et des points réels dans le plan (X, Y).

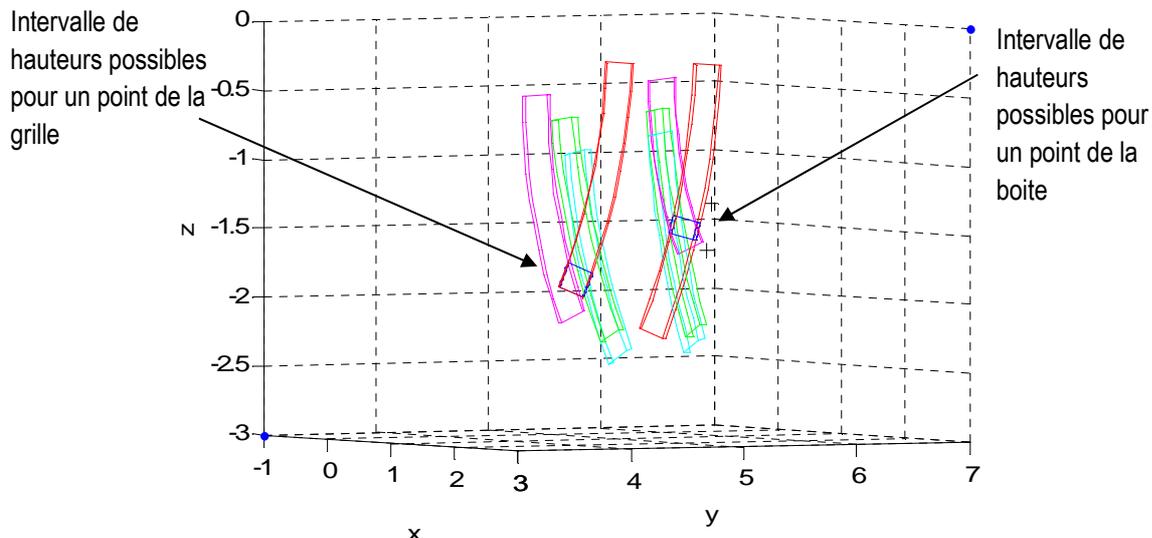


Figure 53 : Exemple de reconstruction d'un point de la grille et d'un point de la boîte par l'approche volumique (coordonnées 3D exprimées en mètre). Les points en noir correspondent aux coordonnées réelles. Les rubans en bleu sont les intervalles des hauteurs possibles des points saillants. Les rubans en rouge, en vert, en cyan et en magenta sont issus des points saillants, respectivement, des images 1, 2, 3 et 4: (a) Localisation des hauteurs possibles des points saillants; (b) Localisation des rubans et des points réels dans le plan (X, Y); Exemple de reconstruction d'un point de la grille et un point de la boîte

Afin d'illustrer la réduction de la dimension des intervalles de hauteurs des points, nous avons calculé la différence entre les intervalles de hauteurs calculés par reconstruction à partir de trois images et ceux calculés par reconstruction à partir de quatre images. Le Tableau 29 et le Tableau 30 présentent ces résultats. L'évolution des intervalles de hauteur d'un point de la grille et d'un point de la boîte au travers des trois types de reconstruction multi-vues est illustrée dans la Figure 54 :

Tableau 29. Différence entre les dimensions des intervalles de hauteur des points de la grille reconstruits à partir de trois images (intervalles 2) et quatre images (intervalles 3).

Points de la grille									
ID	1	2	3	4	5	6	7	8	
Différence m (dimension intervalle2 - dimension intervalle 3)	0,19	0,16	0,14	0,10	0,09	0,02	0,30	0,05	
ID	9	10	11	12	13	14	15	16	
Différence m (dimension intervalle2 - dimension intervalle 3)	0,02	0,04	0,13	0,21	0,29	0,02	0,04	0,15	

Tableau 30. Différence entre les dimensions des intervalles de hauteur des points de la grille reconstruits à partir de trois images (intervalles 2) et quatre images (intervalles 3).

ID	Points de la boîte			
	1	2	3	4
Différence m (dimension intervalle2 - dimension intervalle 3)	0,03	0,05	0,04	0,18

La réduction des intervalles des points de la grille est plus importante que celle des intervalles des points de la boîte. Ces derniers ont presque gardé la même dimension obtenue lors de la reconstruction à partir de trois images.

Dans la Figure 54 (ci-dessous), qui illustre un exemple de reconstruction d'un point de la grille et d'un point de la boîte au travers des trois types de reconstruction, montre que les dimensions des intervalles de hauteur des points se sont réduites au fur et à mesure de l'augmentation du nombre d'images. En fait, arrivant à une reconstruction à partir de quatre images (cf. Figure 52), certains points se trouvent représentés par des intervalles ayant des dimensions très petites (centimétriques). Par exemple, les points 7, 13 et 1 de la grille ont des intervalles ayant des dimensions égales, respectivement, à 1 cm, 3 cm et 6 cm et les points 3 et 1 de la boîte ont des intervalles ayant des dimensions égales à, respectivement, 8 cm et 13 cm.

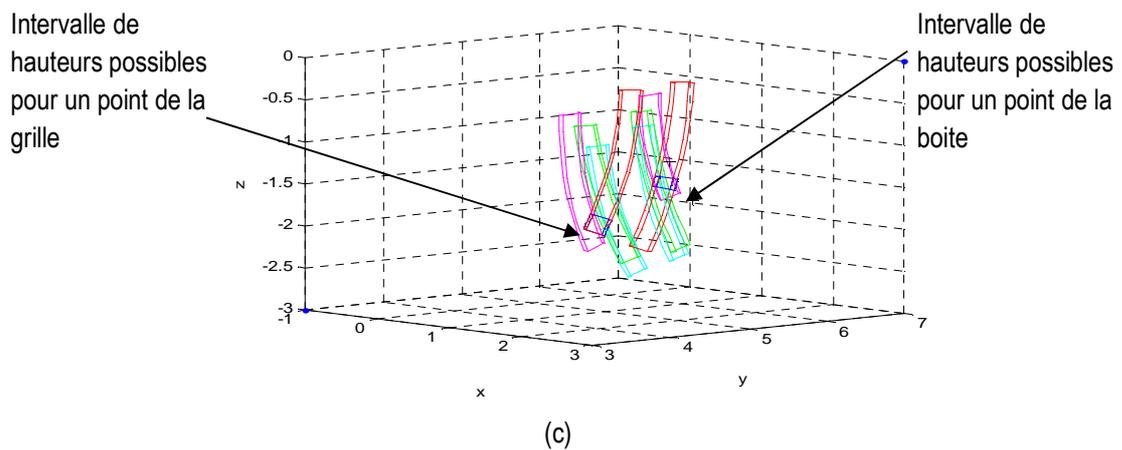
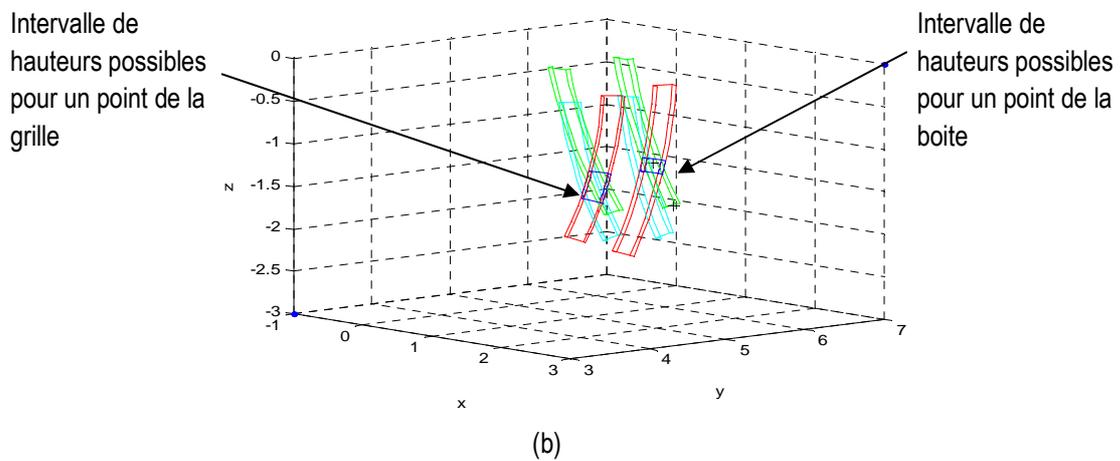
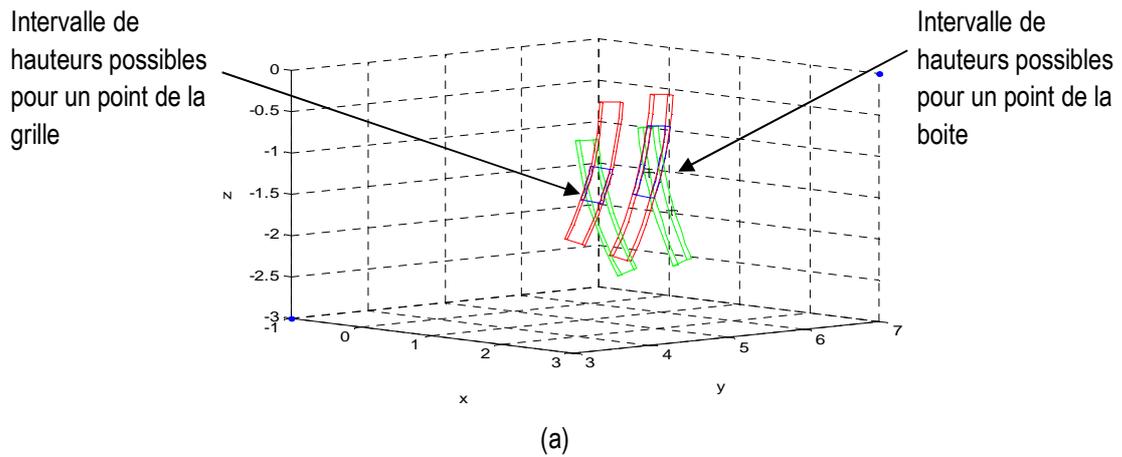


Figure 54 : Coordonnées 3D (exprimées en mètre) d'un point de la grille et d'un point de la boîte calculées par l'approche volumique à partir de deux images (a), de trois images (b) et de quatre images (c). Les points en noir correspondent aux coordonnées réelles. Les rubans en bleu sont les intervalles des hauteurs possibles des points saillants. Les rubans en rouge, en vert, en cyan et en magenta sont issus des points saillants, respectivement, des images 1, 2, 3 et 4.

- **Principales observations tirées de la reconstruction volumique 3D à partir de quatre images**

L'ajout d'une quatrième image à la reconstruction volumique a amélioré les résultats. Cette amélioration est observée au niveau de quatre principaux points :

- 1) La réduction de la dimension des intervalles de hauteur pour chaque point.
- 2) Tous les intervalles des points de la grille se trouvent à un même niveau de hauteur dans la scène.
- 3) Tous les intervalles des points de la boîte se trouvent, aussi, à un même niveau de hauteur dans la scène.
- 4) L'augmentation des intervalles qui incluent les hauteurs réelles des points de la grille.

En fait, en augmentant le nombre d'images au travers des différents tests, les intervalles de la coordonnée en Z des points de la grille ont convergé vers des solutions proches de la hauteur réelle. Par contre, ce n'est pas le cas des points de la boîte puisque la construction à partir de quatre images a modifié les solutions des intervalles, en donnant des solutions qui n'incluent plus les hauteurs réelles. Ceci est lié à la configuration des rubans dans la scène. La disposition des rubans issus de l'image 4, par rapport aux autres rubans, a induit le changement de position des zones d'intersection par rapport aux reconstructions précédentes. Néanmoins, même si les intervalles de points de la boîte n'incluent plus leur hauteur réelle, l'écart en hauteur entre les points de la boîte et les points de la grille s'est amélioré. Il correspond plus à la réalité.

De plus, puisque les points de la grille et les points de la boîte se situent à des positions différentes dans les images, donc, à des résolutions différentes, ceci nous permet de déduire que la résolution des images a, aussi, un impact sur la précision des solutions calculées. A partir de là, nous avons réalisé des tests de reconstruction en utilisant des résolutions variées. Les résultats sont présentés dans le paragraphe qui suit.

4.6.4 Impact de la résolution des images sur la précision de la reconstruction 3D selon l'approche volumique

Dans les paragraphes précédents, les tests ont montré que l'augmentation du nombre d'images a changé la dimension de l'intervalle de solutions associées à la coordonnée en Z. De plus, on a constaté que ces intervalles diffèrent dans leur dimension et dans la précision de leurs solutions en fonction de la position des points dans l'image. Ceci est lié, en partie, à la résolution des images acoustiques. Alors, afin de déduire l'influence de la résolution des images sur la précision des résultats, nous avons testé la reconstruction des points avec des résolutions différentes selon l'azimut et la portée. La recherche de solutions en élévation est réalisée avec la même ouverture verticale de la caméra (10°).

Dans les tests précédents, la résolution utilisée est égale à $0,9^\circ$ en azimut et 0,20 m en portée. Cette résolution tient compte de la résolution du capteur et des incertitudes dues à divers facteurs (ex. incertitudes des traitements d'extraction de points saillants). Les expérimentations, présentées dans la section précédente, ont montré qu'une telle résolution a permis d'extraire des zones d'intersection entre les rubans volumiques des points saillants. Dans les présents tests, nous avons choisi deux autres résolutions dont les dimensions sont plus réduites. La première résolution est égale à $0,9^\circ$ en azimut et 0,10 m en portée. Nous avons, ainsi, pris en compte une erreur d'un pixel de chaque côté du pixel du point saillant, en azimut, et d'une erreur de 4 pixels de chaque côté du pixel du point saillant, en portée. À travers la réduction de la dimension de la résolution, on essaye d'estimer la marge d'incertitude possible autour de chaque point saillant. Ce choix de $0,9^\circ$ en azimut et 0,10 m en portée est effectué aussi dans l'objectif de déduire l'impact de la diminution de la résolution, notamment en portée, sur la dimension des zones d'intersection entre les rubans. Quant à la deuxième résolution, nous avons choisi une résolution en azimut égale à celle du capteur ($0,3^\circ$) et une résolution en portée inférieure à la celle du capteur (0,01 m). Dans ce choix de résolution on ne tient pas compte des incertitudes possibles dans les données utilisées. La Figure 55, la Figure 56 et la Figure 57 présentent les résultats de la reconstruction obtenus, respectivement, à partir de deux, trois et quatre images.

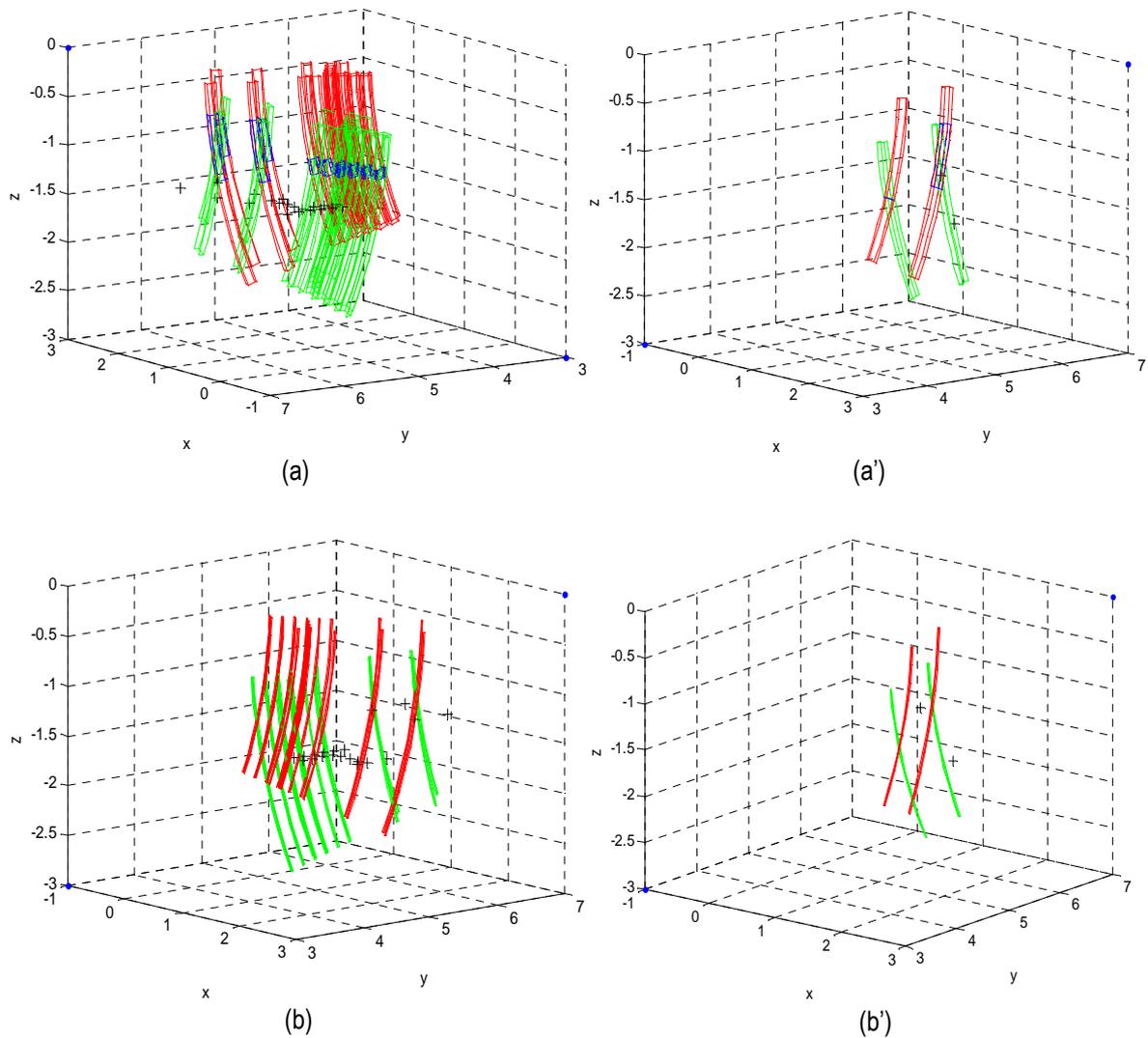


Figure 55 : Coordonnées 3D (exprimées en mètre) calculées par l'approche volumique à partir de deux images avec deux résolutions différentes. Les points en noir correspondent aux coordonnées réelles. Les rubans en bleu sont les intervalles des hauteurs possibles des points saillants. Les rubans en rouge sont issus des points saillants de l'image 1 et en vert de l'image 2. (a, a') Localisation des hauteurs possibles de l'ensemble des points saillants et un exemple d'un point de la grille et d'un point de la boîte à une résolution égale à [azimut = 0,9°, portée = 0,10 m], (b, b') Localisation des hauteurs possibles de l'ensemble des points saillants et un exemple d'un point de la grille et d'un point de la boîte à une résolution égale à [azimut = 0,3°, portée = 0,01 m].

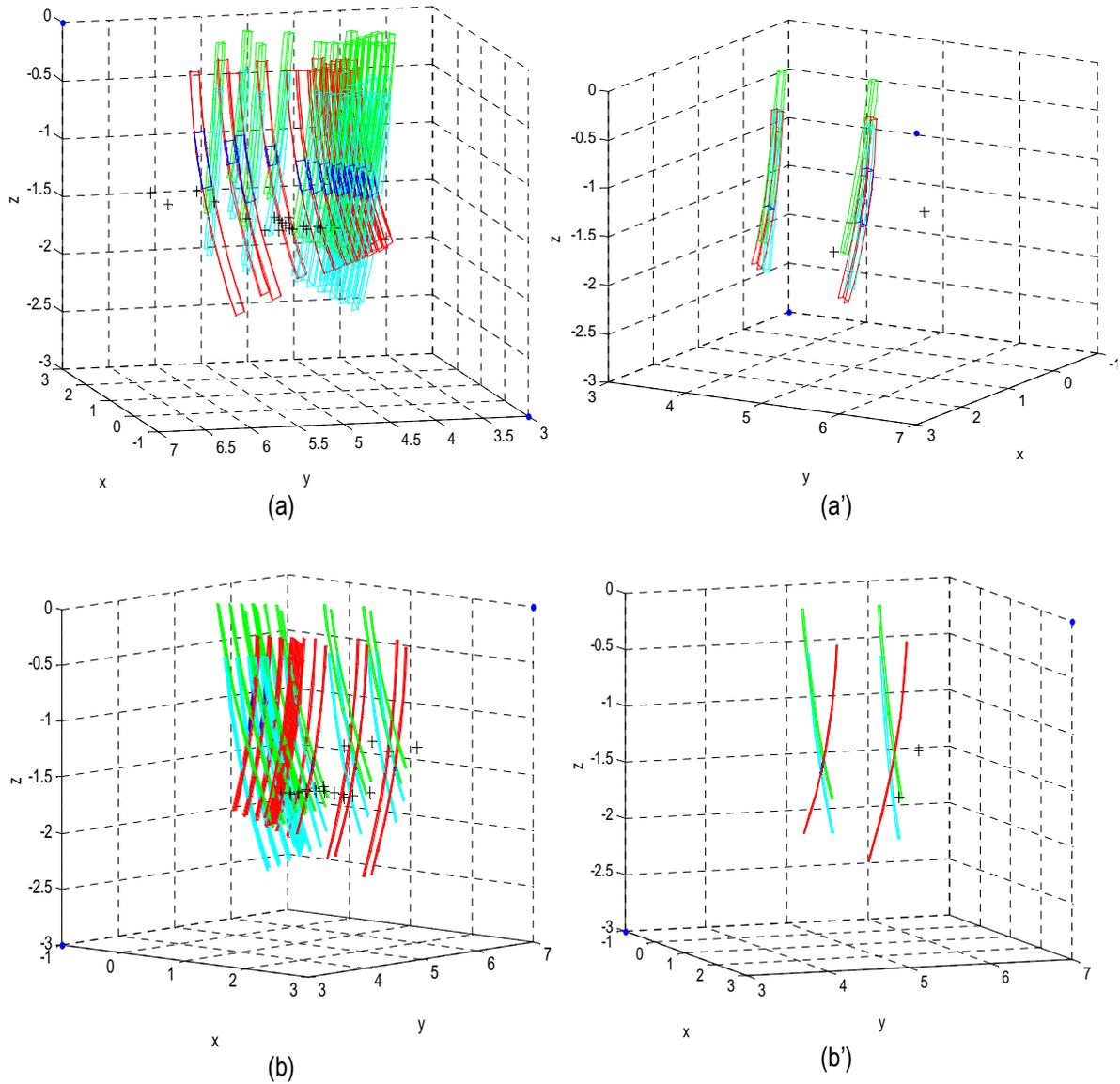


Figure 56 : Coordonnées 3D (exprimées en mètre) calculées par l'approche volumique à partir de trois images avec deux résolutions différentes. Les points en noir correspondent aux coordonnées réelles. Les rubans en bleu sont les intervalles des hauteurs possibles des points saillants. Les rubans en rouge sont issus des points saillants de l'image 1, en vert de l'image 2 et en cyan de l'image 3. (a, a') Localisation des hauteurs possibles de l'ensemble des points saillants et un exemple d'un point de la grille et un point de la boîte à une résolution égale à [azimut = $0,9^\circ$, portée = 0,10 m], (b, b') Localisation des hauteurs possibles de l'ensemble des points saillants et un exemple d'un point de la grille et d'un point de la boîte à une résolution égale à [azimut = $0,3^\circ$, portée = 0,01 m].

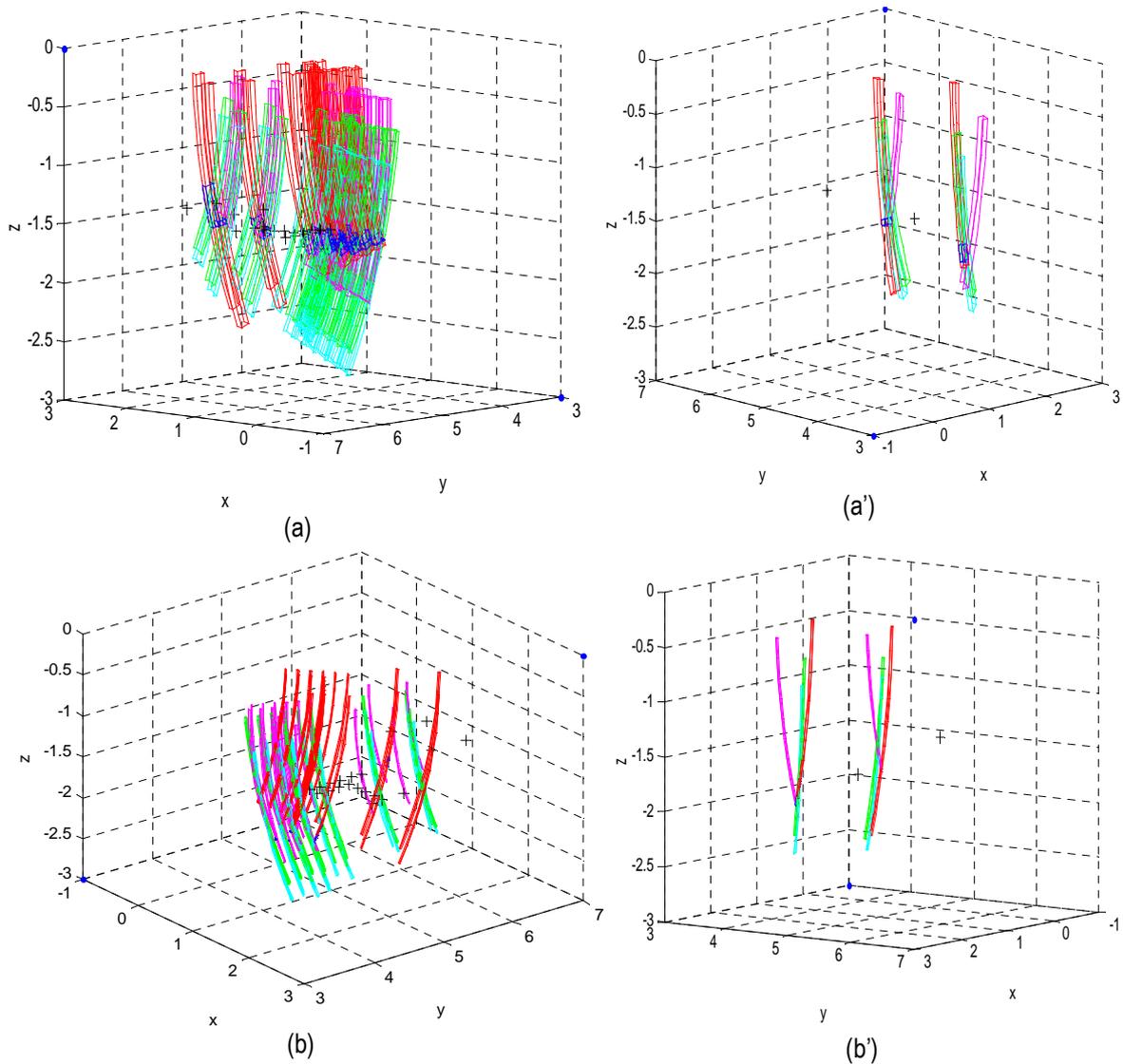


Figure 57 : Coordonnées 3D (exprimées en mètre) calculées par l'approche volumique à partir de quatre images avec deux résolutions différentes. Les points en noir correspondent aux coordonnées réelles. Les rubans en bleu sont les intervalles des hauteurs possibles des points saillants. Les rubans en rouge sont issus des points saillants de l'image 1, en vert de l'image 2, en cyan de l'image 3 et en magenta de l'image 4. (a, a') Localisation des hauteurs possibles de l'ensemble des points saillants et un exemple d'un point de la grille et d'un point de la boîte à une résolution égale à [azimut = $0,9^\circ$, portée = 0,10 m], (b, b') Localisation des hauteurs possibles de l'ensemble des points saillants et un exemple d'un point de la grille et d'un point de la boîte à une résolution égale à [azimut = $0, 3^\circ$, portée = 0,01 m].

À l'issu de ces différents tests, les constatations suivantes ont été faites :

- Dans les différentes reconstructions obtenues, c'est la résolution [0,9° en azimut, 0,10 m en portée] qui a permis d'extraire les zones d'intersection entre les différents rubans. Ces zones sont de dimensions inférieures à celles obtenues avec la résolution [0,9° en azimut et 0,20 m en portée]. Elles ont permis l'obtention d'un nombre réduit de solutions de hauteur pour les points saillants. Ces dernières sont plus précises.
- Comme dans les tests de la section précédente, l'augmentation du nombre d'images a influencé la dimension des intervalles de hauteurs. Plus le nombre d'images utilisées pour la reconstruction est grand, plus la dimension de ces intervalles se réduit. Logiquement, la résolution [0,3° en azimut, 0,01 m en portée] n'a permis pour aucun test de reconstruction d'extraire des zones d'intersection entre les rubans, puisque la dimension des rubans en distance est inférieure à la résolution du capteur.

Finalement, à l'exception des intervalles de hauteur de quelques points de la boîte calculés à partir de deux et trois images, tous les autres intervalles n'incluent pas les hauteurs réelles des points. Ceci indique qu'il y a des imperfections dans le calcul des transformations géométriques inter-images ce qui a induit des erreurs de calcul des coordonnées 3D des points. Ces imperfections peuvent être issues de plusieurs sources dont notamment les incertitudes liées à l'étape d'extraction de points saillants des images et, aussi, aux imperfections, probablement, issues du modèle idéal de la caméra utilisé. En fait, nous n'avons pas réalisé une étape de calibration de la caméra acoustique pour vérifier l'exactitude de la projection des points de la scène en points image selon ce modèle idéal utilisé. Cette étape pourrait révéler l'existence d'une marge d'erreur entre le modèle effectif de la caméra et le modèle idéal.

4.7 Synopsis du chapitre 4

Dans ce chapitre nous avons présenté deux approches de reconstruction tridimensionnelle à partir d'images acoustiques. La première approche est une approche curviligne. Elle exploite un modèle géométrique simplifié qui projette les points saillants en arcs de cercle dans la scène afin de calculer le mouvement de la caméra et déduire les coordonnées 3D. La deuxième approche s'appuie sur une démarche volumique impliquant le modèle géométrique idéal de la caméra, la résolution spatiale des images acoustiques et la manipulation de polygones au lieu de points lors du calcul des coordonnées 3D. Les résultats issus de ces deux approches ont été présentés et discutés. Afin de valider les approches développées, nous avons comparé les coordonnées 3D calculées par chacune des deux approches avec les coordonnées réelles des points des objets de la scène. Les tests ont été réalisés à partir des séquences acquises en bassin fermé (la piscine de l'Université

Laval) pour lesquelles les coordonnées réelles de la scène et des connaissances supplémentaires sur le milieu et la stratégie d'acquisition sont disponibles.

L'approche volumique s'est avérée plus performante et a donné des résultats plus satisfaisants que l'approche curviligne. Cette performance est observée au niveau de trois points : l'intégration de la résolution du capteur dans l'approche, la proposition d'un intervalle de solutions possibles et non pas une solution unique tel que c'est le cas pour l'approche curviligne et finalement une convergence vers des solutions pertinentes et proches de la réalité de la coordonnées en Z. Les coordonnées en X et Y, comme dans l'approche curviligne affichent un écart par rapport coordonnées réelles. Cet écart est illustré dans les figures de reconstruction 3D dans le plan (X, Y). Même si, l'augmentation du nombre d'images a induit une réduction de la dimension de l'intervalle de solutions de ces deux coordonnées, l'erreur reste élevée (l'erreur moyenne de la coordonnée en X des points de la grille, dans la reconstruction à partir de quatre images, est 18 cm et celle des points de la boîte est 65 cm).

Le prochain chapitre aborde la conclusion générale de la thèse et présente les contributions et les perspectives du présent travail.

Chapitre 5. Conclusion générale

5.1 Synthèse de la recherche

L'objectif de la thèse était de proposer une méthodologie permettant la reconstruction 3D de l'environnement marin à partir d'images issues de caméras acoustiques afin d'offrir un moyen d'observation du milieu marin en trois dimensions. En effet, comme premières conséquences des changements climatiques, des signes de vulnérabilité ont été observés dans le milieu marin, notamment au niveau des fonds (érosion, dérives sédimentaires, etc.) et des structures marines (ex. vieillissement des quais, des digues, fissures des pipelines, etc.). Le besoin d'un outil de détection et de suivi de ces signes, pour la prévention de conséquences plus graves, s'est ainsi imposé.

Le chapitre 1 a mis en évidence qu'un tel outil nécessite une représentation tridimensionnelle du milieu marin pour une meilleure observation et interprétation des changements. Pour ce faire, nous avons proposé en premier lieu d'exploiter les caméras acoustiques comme moyen d'observation. En effet, en plus d'être maniables, ces caméras acquièrent des séquences d'images de haute résolution, avec une cadence élevée. En deuxième lieu, nous avons proposé d'exploiter les données issues de ces caméras pour reconstruire le milieu marin selon ses trois dimensions.

Afin de répondre à l'objectif principal de la thèse, nous avons identifié trois sous-objectifs :

- Proposition d'un cadre conceptuel permettant la mise en œuvre des différentes étapes d'une méthodologie de reconstruction tridimensionnelle complète d'une scène observée par des caméras acoustiques
- Proposition d'une approche d'extraction d'informations pertinentes caractérisant la géométrie de la scène observée, à partir de séquences d'images acoustiques
- Proposition d'une approche d'estimation de la géométrie 3D de la scène conforme à la réalité

5.1.1 Un cadre conceptuel de reconstruction 3D de scènes

Le premier sous-objectif de cette thèse visait donc à mettre en place un cadre conceptuel permettant la mise en œuvre des différentes étapes d'une méthodologie de reconstruction tridimensionnelle complète d'une scène observée par des caméras acoustiques. Pour ce faire, nous avons effectué une revue de littérature interdisciplinaire portant sur les caméras acoustiques, la reconstruction 3D en vision numérique et en imagerie acoustique. Celle-ci a mis en évidence l'absence d'approches et de traitements basés uniquement sur les données issues des caméras acoustiques à des fins de reconstruction 3D. Ainsi pour l'élaboration du cadre

conceptuel, nous nous sommes inspirés de la démarche et des approches de reconstruction 3D proposées en vision numérique. Ces approches ont été adaptées, d'une part, aux caractéristiques intrinsèques de ces capteurs (modèle géométrique de formation des images, importance du bruit, ...) et, d'autre part, au contexte particulier d'acquisition de ces séquences d'images (observations multiples d'une même scène sous-marine). Quant à la démarche suivie pour la définition des différents composants du cadre conceptuel, elle correspond aux différentes étapes d'une démarche classique d'une reconstruction 3D en vision numérique.

Ainsi, le cadre conceptuel proposé présente une architecture séquentielle qui permet d'aboutir à la reconstruction tridimensionnelle de la scène sous-marine observée, à partir d'un ensemble de séquences d'images et de connaissances *a priori*. Cette architecture a été implémentée sous la forme d'une chaîne de traitements qui se décompose en trois principaux éléments dédiés respectivement à la détection et l'extraction de primitives, la mise en correspondance de ces primitives et enfin à la reconstruction 3D de la scène à partir des primitives appariées. Ainsi, le premier élément correspond au premier composant du cadre conceptuel « Détection et extraction de primitives » où des prétraitements sont identifiés et proposés afin d'extraire un ensemble de primitives robustes représentant les points saillants de la scène. Ces points sont par la suite exploités dans le deuxième élément « La mise en correspondance des primitives » pour définir des groupes de points saillants correspondants extraits à partir de plusieurs images multi-vues. Enfin, l'élément qui concerne « La reconstruction 3D » regroupe les approches qui permettent de déterminer les coordonnées 3D de ces points saillants et par conséquent de reconstruire la scène.

Tout en validant le cadre conceptuel proposé, grâce à une première implémentation effective, les algorithmes développés au cours de cette thèse, implémentent chacun des éléments du cadre conceptuel par des approches innovantes en imagerie sonar, approches qui s'inspirent du domaine de la vision numérique tout en exploitant les spécificités des caméras acoustiques.

5.1.2 Une approche d'extraction d'informations pertinentes caractérisant la géométrie de la scène observée

Au sein du cadre conceptuel envisagé, le deuxième sous-objectif de cette thèse visait à proposer des approches permettant l'extraction d'informations pertinentes pour la reconstruction 3D de scènes sous-marines à partir des images acoustiques. En étudiant les caractéristiques des différentes primitives proposées dans la littérature (contours, régions et points), il s'est avéré qu'en vue d'un processus de reconstruction 3D, les primitives ponctuelles étaient les plus adaptées pour caractériser sans ambiguïtés, certaines propriétés géométriques de la scène observée à savoir les points anguleux (ou saillants) de cette scène, sur une succession d'images acoustiques. Toutefois, des expérimentations ont montré que l'extraction de ces primitives à partir d'images acoustiques n'était pas réalisable par une simple application des détecteurs de

points les plus utilisés dans la littérature (ex. SIFT, Harris, etc.). Ainsi, suite à plusieurs recherches et expérimentations, nous avons proposé une approche qui tient compte des caractéristiques des images acoustiques et qui permet effectivement d'extraire des points représentatifs de la géométrie de la scène.

L'approche proposée exploite tout particulièrement la redondance d'observation de la scène au sein d'une séquence d'images acoustiques, en s'appuyant sur trois principales étapes qui sont l'extraction de contours pertinents, leur mise en correspondance et enfin l'extraction des points saillants à partir de ces contours pertinents appariés. Lors de la détection des contours pertinents, une approche itérative est mise en œuvre. Elle permet de supprimer les faux contours provoqués par le bruit présent dans les images, en se basant sur une analyse multi-échelle de ces images qui exploite une pyramide gaussienne de filtrage. Ainsi, à chaque itération, l'image est filtrée, les contours sont détectés par le filtre de Canny puis mis en correspondance et comptés. L'étape de mise en correspondance s'appuie sur une connaissance très grossière du déplacement de la caméra entre deux images successives pour retrouver les contours correspondants d'une image sur l'autre. Ces deux étapes (extraction de contours pertinents et mise en correspondance) sont répétées pour chaque échelle considérée. L'échelle d'arrêt est déterminée automatiquement quand le nombre de contours devient stable au travers de l'analyse multi-échelle indiquant ainsi la détection des contours pertinents de l'image, c'est-à-dire ceux qui reflètent effectivement la géométrie de la scène observée. De ces contours sont, donc, extraits les points saillants. Ces points de contour ayant une forte courbure traduisent en effet, des caractéristiques géométriques anguleuses de la scène.

Afin de valider la robustesse de ces points extraits en vue d'une reconstruction 3D et valider par conséquent, la pertinence de l'approche proposée, plusieurs tests d'évaluation ont été réalisés. Chaque test a évalué un critère de qualité précis. Le premier critère est la répétabilité qui quantifie l'accumulation de détection d'un même point saillant dans une suite d'images et donc sous différents points de vue. Le second est la bonne détection qui correspond à la détection de points qui caractérisent uniquement la géométrie de la scène. Finalement, le troisième est la bonne localisation qui permet d'évaluer la précision de la localisation des points saillants extraits. Ces critères ont été évalués sur trois séquences (séquence « Pied de pont », séquence « Épave » et séquence « Piscine »). Les résultats obtenus révèlent que les points détectés sont répétables avec un taux supérieur à 60 % sur une séquence de 11 images pour l'ensemble des trois séquences. L'approche d'extraction de points saillants proposée se caractérise aussi par un bon taux de détection puisque l'algorithme détecte sur les séquences d'images exploitées, globalement 80 % des points de référence qui sont les points des images jugés être des points caractéristiques de la géométrie de la scène. En termes de bonne localisation, les évaluations menées ont montré que l'erreur maximale en portée était de l'ordre de 2 pixels pour la séquence « Pied de pont » et 3 pixels pour les deux autres séquences ce qui correspond à la moitié de la valeur de l'erreur calculée en azimut. Les erreurs en termes de localisation des points saillants

sont ainsi concentrées au niveau de l'azimut. Ceci est lié à la géométrie de formation des images qui engendre plus de précision en portée qu'en azimut.

D'après ces tests d'évaluation, l'approche proposée se caractérise donc par une bonne détection de points saillants. Ces points sont représentatifs de la géométrie de la scène et présentent un taux de répétabilité intéressant qui permet d'envisager leur utilisation dans une reconstruction 3D multi-vues. Par ailleurs, les tests relatifs à leur bonne localisation, ont permis de quantifier les marges d'erreur associées à leur localisation. La connaissance de ces incertitudes est importante car elle permet de les intégrer comme connaissance *a priori* dans l'étape de reconstruction 3D de la scène. L'approche d'extraction de primitives proposée répond ainsi au second objectif de la thèse et permet de fournir des primitives pertinentes pour une reconstruction 3D.

5.1.3 Une approche d'estimation de la géométrie 3D de la scène

Le troisième sous-objectif de la thèse visait à proposer une approche permettant de calculer à partir d'une séquence d'images acoustiques, l'élévation de la scène observée, pour aboutir à une scène tridimensionnelle reconstruite conforme à la réalité. Nos travaux de recherche nous ont amené à proposer deux approches de reconstruction tridimensionnelle : une approche curviligne et une approche volumique. Le choix de ces deux approches a été fait avec pour objectif de tester deux versions du modèle géométrique de formation des images: une version simplifiée (approche curviligne) et une version "idéale" du modèle (approche volumique) qui permet également l'intégration des incertitudes de localisation des points saillants considérés.

5.1.3.1 Différences et points-clés des deux approches proposées

Dans la première approche (curviligne), le lieu géométrique dans la scène associé à un pixel dans l'image acoustique, se situe le long d'un arc de cercle. Cet arc de cercle est défini par la portée « r », l'azimut « θ » et un angle « φ » situé dans un intervalle d'élévations qui correspond à l'ouverture verticale de la caméra. En revanche, l'approche volumique considère que le lieu géométrique dans la scène, d'un pixel de l'image, ne se situe pas le long d'un arc de cercle mais au sein d'un volume. Ce volume est délimité en portée et en azimut, par les quatre sommets d'un polygone \mathcal{P} correspondant aux quatre « coins » du pixel considéré. Le modèle géométrique retenu en termes de formation des images, est donc la première différence entre ces deux approches. Par ailleurs, l'approche volumique considère chaque pixel comme un polygone dont les dimensions correspondent aux résolutions en portée et en azimut des images. Il est alors possible de modifier ces dimensions pour, d'une part intégrer les erreurs de localisation des points saillants (connaissance *a priori*), et d'autre part estimer les incertitudes résultantes de positionnement des points 3D reconstruits grâce au volume final représentant le lieu géométrique possible de chaque point reconstruit. Seule l'approche volumique permet de manipuler de telles quantités qui permettent d'associer à une scène reconstruite une

certaine quantification de sa précision de reconstruction. Cette méta-connaissance est clairement un atout lors de l'alimentation des systèmes, tels que les SIG 3D, qui vont exploiter les reconstructions 3D ainsi produites.

Même si les modèles géométriques de représentation des images diffèrent entre les deux approches, celles-ci reposent ensuite sur une mécanique commune et générique d'estimation du déplacement de la caméra entre les images. En effet, à partir du déplacement estimé, les coordonnées 3D (approche curviligne) ou volumes 3D (approche volumique) associés aux points saillants peuvent être estimés. Du fait de la complexité de l'espace de recherche d'un tel déplacement, l'algorithme d'optimisation qui a été retenu est l'algorithme SE-AMC (ou Stratégie d'Évolution avec Adaptation de la Matrice de Covariance) qui est issu de la famille des algorithmes évolutionnaires. L'une des particularités du SE-AMC est d'une part l'utilisation d'une matrice de covariance globale pour paramétrer les mutations gaussiennes, et d'autre part l'adaptation dynamique de cette matrice de covariance à partir des chemins cumulatifs de mutation ayant donné des solutions satisfaisantes. La force de SE-AMC dans ce contexte, est ainsi de permettre une recherche globale dans l'espace des solutions possibles (afin d'éviter des solutions sous-optimales) tout en restant efficace dans la convergence vers la meilleure solution courante.

5.1.3.2 Reconstruction 3D par primitives curvilignes

Plusieurs tests portant sur la reconstruction à partir de deux images et la reconstruction multi-images ont été réalisés. Dans les deux approches les mêmes points saillants extraits de différentes images et l'algorithme d'optimisation SE-AMC ont été exploités. À l'issue de chaque test les coordonnées 3D des points saillants ont été calculées et comparées avec les coordonnées réelles. Lors de l'application de l'approche curviligne, le résultat obtenu correspond au calcul des points d'intersection conjointe entre tous les arcs de cercle issus de la projection des points saillants selon le modèle géométrique. Ces points d'intersection définissent les points 3D recherchés. À défaut de l'obtention d'une intersection conjointe entre les arcs de cercle, ce sont les points qui produisent la distance la plus faible qui sont sélectionnés par l'algorithme de reconstruction.

Dans tous les tests de reconstruction 3D réalisés et présentés dans la section de l'approche curviligne, aucune fonction objective n'a atteint la valeur minimale zéro. Ceci indique que les arcs de cercle issus des points saillants ne se croisent pas. Cette « non intersection » est liée à la configuration des arcs de cercle dans la scène. Leur disposition les uns par rapport aux autres ne permet pas leur intersection. Par contre, ils se trouvent tous proches d'un même point 3D dans la scène. D'ailleurs, c'est la distance entre ce point 3D et tous les arcs de cercle qui est minimisée par la fonction objective. Une fois la fonction objective a atteint la valeur minimale, l'algorithme d'optimisation présente en sortie les solutions optimales du déplacement de la caméra. Dans les différents tests, ces déplacements ont été calculés avec une marge d'erreur élevée par rapport au déplacement effectif. D'ailleurs, l'impact de cette marge d'erreur est observé dans la reconstruction 3D des objets d'une part, dans l'erreur de calcul des coordonnées 3D où la comparaison des coordonnées 3D

avec les coordonnées réelles des points saillants associés aux objets présents dans la scène a révélé la présence d'une marge d'erreur élevée selon les trois axes (X, Y et Z) et, d'autre part, dans le décalage uniforme en position entre tous les points reconstruits et tous les points réels. En effet, les tests de reconstruction (avec 2 images, 3 images et 4 images) conduisent, toujours, à une distribution spatiale identique des points. Cette disposition est conforme à la disposition réelle des points dans la scène. Elle caractérise, par conséquent, fidèlement la géométrie des objets.

Il est à noter, aussi, que malgré l'erreur affectant les déplacements estimés on a observé une cohérence dans les résultats avec la géométrie d'acquisition. Ceci est observé lors de la reconstruction à partir de trois et quatre images au niveau de deux points :

- 1) dans les translations estimées entre l'image 2 et l'image 3. Ces images, acquises quasiment de la même position, ont des translations estimées très proches par rapport à l'image 1.
- 2) dans l'estimation de la rotation effectuée entre l'image 1 et l'image 4. Cette dernière est acquise d'une position située à l'opposé de la position d'acquisition de l'image 1, alors, la rotation calculée selon l'axe des Z par rapport aux valeurs de la rotation selon l'axe des X et des Y est plus élevée. Elle correspond à la réalité de la configuration d'acquisition des images.

Finalement, d'après les résultats obtenus et discutés dans le chapitre 4, la reconstruction par primitives curvilignes n'a pas abouti à une bonne reconstruction de la scène. Certes, l'erreur de la reconstruction peut provenir, en partie, d'une imprécision dans les mesures lors de l'acquisition ou d'une imperfection issue des traitements appliqués aux images; mais en grande partie, l'erreur est due à l'utilisation d'un modèle géométrique simplifié. Ceci est confirmé par les résultats de l'approche volumique où le modèle géométrique idéal est utilisé. Ces résultats sont synthétisés dans le paragraphe suivant.

5.1.3.3 Reconstruction 3D par primitives volumiques

L'approche volumique permet d'estimer les déplacements de la caméra qui permettent à tous les rubans correspondant à un même point saillant d'avoir une intersection conjointe en un volume commun, et cela pour tous les points saillants extraits des différentes images. Les résultats obtenus sont, donc, de nature différente de ceux issus de l'approche curviligne. En effet, cette approche fournit une plage en distance, en azimut et en élévation pour chaque point reconstruit. Cette plage en distance permet de déduire l'ensemble de coordonnées 3D possibles pour un point saillant.

Les expérimentations conduites sur les données acquises en bassin ont montré que les coordonnées réelles des points saillants se retrouvaient (ex. les points de la boîte) dans les plages estimées par l'algorithme de reconstruction proposé. Il a aussi été montré que plus le nombre d'images utilisées augmente, plus les plages

estimées se réduisent autour des coordonnées réelles des points de la scène (ex. les points de la grille). En effet, pour chaque point saillant, chaque image supplémentaire fournit un ruban de plus avec une orientation différente (du fait du point de vue différent de la caméra).

En utilisant des primitives volumiques, la fonction objective minimise non seulement la distance, en azimuth et en portée (tel que pour l'approche curviligne), mais de plus la distance en élévation entre tous les rubans. La minimisation de la distance en élévation permet d'extraire les zones d'intersection commune en élévation entre tous les rubans volumiques. Dans tous les tests de reconstruction 3D (avec 2 images, 3 images et 4 images), la fonction objective a atteint la valeur minimale zéro indiquant l'obtention d'une intersection conjointe. Ceci met en évidence deux points :

- 1) La configuration des primitives volumiques dans la scène permet leur intersection. En fait, ceci est lié à leur résolution en portée et en azimuth. Dans les expérimentations réalisées dans cette thèse, ce sont les résolutions suivantes qui ont permis d'aboutir à des intersections entre les rubans et d'estimer les coordonnées 3D des points saillants : ($\Delta r = 0,1 \text{ m}$ & $\Delta \theta = 0,9^\circ$ et $\Delta r = 0,2 \text{ m}$ & $\Delta \theta = 0,9^\circ$).
- 2) L'utilisation du modèle géométrique idéal a favorisé l'intégration de la résolution et, par conséquent, le calcul des volumes d'intersection.

Malgré la différence entre le principe d'utilisation de primitives respectivement volumiques et curvilignes, certains résultats sont semblables. Ces derniers concernent, premièrement, la disposition des volumes reconstruits dans la scène. Ils respectent aussi la géométrie réelle des objets. De plus, en augmentant le nombre d'images, l'écart en hauteur entre les points de la boîte et les points de la grille s'approche de l'écart réel. La deuxième ressemblance, dans les résultats, est observée au niveau des déplacements estimés. Dans certains tests (reconstruction à partir de deux images), les déplacements estimés par approche curviligne et par approche volumique sont proches. Toutefois, malgré ces points de ressemblance dans les résultats, les coordonnées 3D estimées par l'approche volumique sont meilleures. Comme évoqué, ci-haut, on retrouve les hauteurs réelles des points saillants dans les intervalles de hauteurs calculés.

En conclusion, au travers des résultats obtenus à partir des deux approches de reconstruction 3D proposées (l'approche curviligne et l'approche volumique), le troisième objectif spécifique de la thèse « Reconstruction 3D » est atteint. La comparaison des résultats de ces deux approches a permis de déduire que l'approche volumique répond le mieux aux caractéristiques des images des caméras acoustiques (notamment en utilisant le modèle géométrique idéal de formation des images) et qui fournit une reconstruction proche de la réalité en termes de précision des hauteurs calculées.

5.2 Principales observations tirées des approches développées

De la synthèse des résultats, présentée dans le paragraphe précédent, diverses observations ont été tirées. Elles concernent les trois approches développées dans la thèse; l'approche d'extraction de primitives et les deux approches de reconstruction 3D. Ces observations sont présentées dans les paragraphes qui suivent.

L'approche d'extraction de primitives nous a permis d'extraire des points saillants représentatifs de la scène et répétables sur une série d'images consécutives. C'est une approche qu'on peut qualifier de spécifique aux caractéristiques des images des caméras acoustiques. Elle tient compte de la qualité bruitée des images au travers de ses diverses étapes de traitement et elle exploite le déplacement de la caméra entre deux images successives. De plus, c'est une approche automatique. Lorsque ses différentes étapes de traitement (filtrage, détection de contours et mise en correspondance) extraient les contours pertinents des images, elles prennent fin automatiquement, pour passer à l'étape finale « d'extraction des points saillants pertinents ». Selon les critères d'évaluation de l'approche (répétabilité, bonne détection et localisation), les résultats de la détection de points sont satisfaisants. Ces critères d'évaluation ont aussi révélé quelques limites. Ces limites accompagnées de suggestions pour les contourner sont présentées dans les paragraphes suivants :

- En évaluant la répétabilité, nous avons mis en évidence que les points sont moins répétables dès que les images sont décalées dans la séquence. Ceci est un inconvénient pour une reconstruction tridimensionnelle qui vise l'utilisation d'images décalées dans une séquence. Par contre, en exploitant la haute cadence des images et la faible variation des vues entre images successives, il est possible de pallier ce problème par une approche de suivi des points identiques le long d'une séquence.
- Des fausses alarmes ont été détectées lors de l'évaluation de la bonne détection de l'approche. Néanmoins, on a remarqué que ces fausses alarmes caractérisent particulièrement la présence des ombres des objets dans la scène. Ainsi, ajouter une approche de segmentation qui élimine les zones d'ombre et extrait les objets d'intérêt avant de procéder à la détection de points permettrait de réduire ces fausses alarmes.
- Des points ont été extraits avec une marge d'erreur de l'ordre de quelques centimètres en portée et plus d'un degré en azimuth. Ces erreurs peuvent influencer la précision de la reconstruction 3D mais la quantification qui a été effectuée de ces erreurs, est aussi une connaissance importante que les méthodes de reconstruction peuvent exploiter.

La prise en compte de ces limitations et des solutions proposées dans les travaux futurs permettra d'améliorer l'approche d'extraction de points saillants.

Quant aux approches de reconstruction 3D proposées, plusieurs observations ont été faites à partir des résultats obtenus. Elles sont synthétisées dans les paragraphes suivants.

L'approche curviligne ne fournit qu'une coordonnée 3D pour chaque point reconstruit. Dans la majorité des résultats obtenus, en utilisant deux images ou « n » images, les coordonnées 3D calculées par cette approche, affichent une marge d'erreur importante par rapport aux coordonnées 3D réelles des points. Par ailleurs, aucune indication d'incertitude n'est fournie avec ces coordonnées. Cette approche ne convient pas, par conséquent, à la reconstruction 3D d'une scène à partir d'images acoustiques. Et il n'est pas possible d'envisager une amélioration de ces résultats de reconstruction, étant donné qu'elle s'appuie sur l'utilisation d'un modèle géométrique simplifié. Ce modèle projette les points saillants en arc de cercle dont la configuration dans la scène ne permet pas leur intersection. Les points 3D reconstruits afficheront, donc, toujours une marge d'erreur par rapport au point 3D réel.

L'approche volumique fournit des résultats beaucoup plus intéressants. En effet, chaque point reconstruit consiste en un volume de points possibles, volume qui se réduit et converge vers le point réel au fur et à mesure de l'intégration d'images supplémentaires offrant des points de vue différents. L'approche volumique s'avère, ainsi, adéquate pour une reconstruction 3D multi-vues d'une scène à partir des images acoustiques. Elle permet non seulement de prendre en compte les caractéristiques de la formation géométrique des images au travers de l'utilisation du modèle géométrique idéal, mais aussi d'estimer la précision des coordonnées 3D. De plus, au travers de la variation de la résolution, il est également possible avec l'approche volumique d'intégrer les erreurs de localisation de points saillants.

Ainsi avec l'ensemble de ces observations, on peut affirmer que les trois sous-objectifs de la thèse ont été atteints. Nous avons ainsi proposé un cadre conceptuel, une approche d'extraction de primitives et deux approches de reconstruction 3D de la scène. L'atteinte de ces trois sous-objectifs, nous permet de conclure que l'objectif principal de la thèse a été atteint. La méthodologie proposée, développée et expérimentée permet de reconstruire en 3D une scène observée par une séquence d'images acoustiques. Cette méthodologie fournit une reconstruction qui est :

- Complète : tous les points saillants sélectionnés pour la caractérisation de la scène ont été reconstruits.
- Fidèle à la réalité : elle reconstruit fidèlement la géométrie de la scène. C'est-à-dire que la disposition des points reconstruits dans la scène, les uns par rapport aux autres, est conforme à la disposition des points réels.

- Robuste et fiable : elle est robuste puisqu'elle permet de retrouver les coordonnées 3D des points dans la scène tout en évaluant leur précision à travers la proposition d'un ensemble de solutions d'élévation possibles. Elle est ainsi fiable et efficace pour la reconstruction de l'environnement marin vu qu'elle intègre la notion d'incertitude dans le calcul des coordonnées 3D.

Les caractéristiques de cette méthodologie comptent parmi les contributions de cette thèse. Celles-ci sont détaillées dans le paragraphe qui suit.

5.3 Les contributions de notre travail

Les apports de notre travail concernent à la fois la proposition et le développement d'une méthodologie de reconstruction tridimensionnelle de scènes sous-marines à partir de séquences d'images acoustiques. En plus de cette contribution principale, ce travail de recherche a contribué de manière tangible à plusieurs réalisations et avancées décrites dans les points suivants :

- 1) Une étude approfondie des caractéristiques des caméras acoustiques et une étude comparative avec les caractéristiques des caméras optiques, largement utilisées pour explorer le milieu marin, afin de mettre en évidence la prééminence des caméras acoustiques comme outil d'observation du milieu marin.
- 2) Un cadre conceptuel spécifiant les fondements méthodologiques requis pour une reconstruction tridimensionnelle complète.
- 3) Une nouvelle approche d'extraction de primitives à partir des images acoustiques. Elle a été accompagnée de divers tests et expérimentations sur des séquences d'images acoustiques acquises en milieu contrôlé et en milieu naturel. C'est une approche robuste et fiable puisque : 1) elle surmonte le niveau de bruit des images acoustiques, 2) elle exploite la cadence des images qui est spécifique au type de caméra utilisée, 3) elle détermine automatiquement l'échelle de l'analyse multi-échelle en fonction de la séquence utilisée (ex. le contenu et le mode d'acquisition) et 4) elle répond aux exigences de la reconstruction 3D en détectant des points saillants pertinents, répétitifs et représentatifs de la géométrie de la scène.
- 4) Deux approches de reconstruction 3D développées et testées sur des images de caméras acoustiques. Ces deux approches (curviligne et volumique) sont novatrices et spécifiques aux caméras acoustiques puisqu'elles exploitent les caractéristiques de ces caméras dans leurs différentes étapes dont, notamment, le modèle de formation des images.

La présente thèse est riche en information touchant divers domaines, notamment les caméras acoustiques, la vision numérique et l'imagerie sonar. Les contributions de cette recherche ont été présentées, lors de 7 occasions, à la communauté scientifique internationale dans le cadre de conférences impliquant une évaluation des soumissions par les pairs (ex. MTS/IEEE Ocean's, 2010, 2011, 2012; Colloque FFGG¹⁶ 2010, Septième édition des ateliers de travail sur le traitement et l'analyse de l'information Taima 2011; Conférence Hydrographique du Canada CHC 2010, Colloque CIDCO¹⁷ 2013). Un article dédié à la présentation du cadre conceptuel a été publié dans la revue *Traitement du Signal* (Lavoisier).

5.4 Perspectives

Tout au long de notre travail de recherche, nous avons essayé d'explorer plusieurs approches que ce soit en termes d'extraction de primitives ou de reconstruction tridimensionnelle. Nous avons mis en œuvre des approches adaptées au contexte des caméras acoustiques et en adéquation avec l'objectif final d'une reconstruction 3D robuste et fiable. Ces approches ont fourni un résultat positif quant à la faisabilité d'une telle reconstruction avec des résultats prometteurs qui répondent aux objectifs du travail, notamment au niveau des points saillants extraits des images et au niveau des coordonnées 3D calculées. Néanmoins chacune des approches développées pourrait être plus explorées dans des projets subséquents pour des résultats encore plus performants en termes de précision.

5.4.1 Réflexions sur le processus de reconstruction

L'algorithme d'optimisation mis en place a pour objectif de calculer un jeu de déplacements de caméras inter-images qui soit cohérent avec la configuration spatiale des points saillants extraits sur ces différentes images. Ce jeu de déplacements permet d'estimer ensuite une plage ou un volume possible pour chaque point reconstruit. Du fait de la résolution du capteur et des erreurs de localisation des points saillants, il n'y a probablement pas qu'un seul jeu de déplacements qui explique la configuration spatiale des points saillants entre les images. D'autres jeux probablement très proches, peuvent également convenir, amenant d'autres volumes certainement très voisins, de reconstruction pour chaque point. Un axe de recherche consisterait à déterminer non pas un jeu de déplacements mais tous les jeux de déplacements afin de récupérer exhaustivement tous les volumes de reconstruction possibles pour chaque point. Une première étape serait de déterminer quelle serait la plus-value d'un tel effort en termes de précision de reconstruction effective par rapport aux coûts de recherche supplémentaires.

¹⁶ Colloque facultaire de la Faculté de Foresterie, de Géographie et de Géomatique

¹⁷ Centre Interdisciplinaire de Développement en Cartographie des Océans

5.4.2 Optimisation de la méthodologie de reconstruction

Dans la solution de reconstruction tridimensionnelle actuelle, plusieurs points peuvent être améliorés, notamment à travers la réalisation de plusieurs tests impliquant différents type de séquences. En fait, il serait pertinent de tester l'approche de reconstruction sur d'autres séquences d'images issues de différentes caméras acoustiques et acquises dans des contextes différents, notamment dans le milieu marin. Ces tests permettront d'adapter la méthodologie de reconstruction aux différents types de caméras et de généraliser son application. Avant de réaliser ces tests, la méthodologie actuelle pourrait être optimisée, notamment en améliorant la stratégie d'acquisition, en intégrant des connaissances *a priori* et en mettant en œuvre une approche pleinement automatique. Ces modifications sont présentées dans les paragraphes suivants.

5.4.2.1 Amélioration de la stratégie d'acquisition des données

Dans un contexte opérationnel, il conviendrait de proposer des scénarii d'acquisition dédiés aux différentes scènes & applications envisagées. Ces scénarii indiqueraient le nombre de trajectoires de la caméra et leur configuration spatiale afin de respecter un nombre minimum de vues avec des différences de points de vue suffisantes. Ces scénarii ont clairement pour objectif de maximiser les différences de points de vue de manière à augmenter la précision des points reconstruits, tout en permettant également l'exploitation d'images issues de plusieurs séquences. En effet, le tracking des points saillants est envisageable assez simplement au sein d'une séquence d'images puisque le flux d'informations est continu. Repérer les mêmes points saillants sur une autre séquence acquise selon une trajectoire différente devient très problématique à moins de connaître cette trajectoire. Définir et spécifier une liste de scénarii préétablis en fonction de la complexité géométrique de la scène à reconstruire (paroi d'un quai, pipeline, tête de puits, ...) et des contraintes spatiales d'acquisition, seraient par conséquent très intéressants dans un contexte opérationnel.

5.4.2.2 Intégration de connaissance *a priori* pour une meilleure reconstruction

De par sa conception modulaire, l'approche proposée semble très prometteuse pour la reconstruction de scène à partir de différents types de caméras acoustiques. Dans un contexte opérationnel, certaines connaissances géométriques sur les scènes à reconstruire sont disponibles (installations portuaires, surveillance de fonds connus). Il est alors totalement envisageable d'introduire certaines contraintes supplémentaires telles que les coordonnées réelles de quelques points d'appui dans la scène pour permettre à l'algorithme d'optimisation de converger plus rapidement vers des solutions plus précises. Par ailleurs, après introduction de ces connaissances *a priori*, l'analyse d'une convergence difficile ou d'une valeur finale de la fonction d'optimisation élevée, pourrait être un indicateur pertinent de changements au sein de la scène observée, dans un contexte de surveillance de l'environnement ou d'infrastructures portuaires (*a priori* connues) par exemple.

5.4.2.3 Mise en œuvre d'une approche de reconstruction pleinement automatique

Envisager la mise en œuvre d'une approche pleinement automatique de reconstruction 3D est important pour offrir un outil d'observation performant du milieu marin. En fait, dans le présent travail l'étape d'appariement des points n'a pas été abordée. Il serait donc intéressant d'envisager l'intégration de cette étape afin de compléter la démarche méthodologique proposée. Étant donné que les points sont répétables sur une série d'images successives, une approche d'appariement basée sur le suivi des points dans une séquence d'images est possible. Ceci est avantageux dans le contexte d'une reconstruction tridimensionnelle à partir d'une séquence d'images. En offrant une approche complète et automatique, il serait plus facile de mettre en œuvre une solution facilement manipulable par différents utilisateurs et qui fournit une reconstruction 3D en temps réel à partir de séquences d'images acoustiques.

5.4.3 De l'exportation des scènes reconstruites

Ayant des coordonnées 3D précises, la modélisation 3D de la scène constitue une prochaine étape du présent travail de thèse. Il existe trois types de modélisation 3D : filaire, volumique et surfacique. Ces modélisations permettent de visualiser la reconstruction 3D en modélisant l'information 3D calculée et de proposer ainsi une représentation « réelle » de la scène. Les données issues de la modélisation 3D pourront être par la suite structurées de manière à les intégrer dans un SIG 3D. Les SIG 3D sont des outils intéressants pour regrouper au sein d'un système unique toutes les données historiques et récentes sur l'évolution du milieu sous-marin. De plus les SIG servent d'interface d'accès aux données. Celles-ci peuvent être consultées et modifiées par le biais d'une mise à jour du modèle de reconstruction 3D. Ils fournissent ainsi un moyen de sauvegarde de données et d'observations spatio-temporelles intéressant dans divers applications marines (ex. observation de l'évolution du milieu, élaboration de cartes 3D, etc.).

5.4.4 Reconstruction 3D multi-sources

Cette thèse propose un cadre conceptuel de reconstruction 3D conjointement à une première implémentation fonctionnelle. L'avantage de cette implémentation est qu'elle permet une reconstruction de la scène uniquement avec des informations ponctuelles extraites des images acquises. Néanmoins il est logique d'envisager l'intégration d'informations supplémentaires comme solution pertinente pour aboutir à une reconstruction précise et complète de la scène. Ces informations supplémentaires peuvent être issues de différentes sources telles que des primitives extraites de nature différente, ou bien directement de capteurs différents.

Dans cette première famille, des primitives surfaciques pourraient venir compléter spatialement la reconstruction ponctuelle obtenue en indiquant les orientations des surfaces de la scène entre les points reconstruits. Pour ce faire, des algorithmes de type *Shape From Shading*. permettent 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. De telles primitives sont effectivement complémentaires car elles souffrent généralement d'un manque de positionnement absolu, positionnement absolu qui fait la force des primitives ponctuelles. L'information 3D ponctuelle servirait ainsi de point d'appui pour l'intégration de l'information surfacique et l'obtention d'un modèle tridimensionnelle continu.

Dans un contexte plus large, les informations de relief issues d'autres capteurs tels que les sonars interférométriques ou les sondeurs bathymétriques, peuvent apporter aussi une source d'informations complémentaire, en proposant des modèles de départ de la scène à reconstruire. Dans le cas où la résolution de ces modèles géométriques est plus grossière que celle qu'il est possible d'obtenir avec des caméras acoustiques, le processus de reconstruction vient affiner le modèle existant. Dans le cas où la résolution de ces modèles est équivalente, la maniabilité des caméras acoustiques devient un atout majeur dans la mise à jour de ces modèles (monitoring) et la détection d'éventuels changements.

Bibliographie

- Alharthy A., and J. Bethel, «Detailed building reconstruction from airborne laser data using a moving surface method», In Proceeding of the 20ème ISPRS congress, Istanbul, Turkey, International Society for Photogrammetry and Remote Sensing, vol 35, Partie B, 2004.
- Alharthy A., and J. Bethel, « Heuristic filtering and 3d feature extraction from LIDAR data», In Proceedings of the ISPRS Commission III symposium, Graz, Austria Institute for Computer Graphics and Vision Graz University of Thechnology, vol 34, Partie A, 2002.
- Alkama S, D. Berkani, « Segmentation Vectorielle des Images Multi-spectrales MSG2 par une Approche Markovienne», Studia Informatica Universalis, vol 8, pp. 89-110, 2010.
- Alkhatib, H., «Etude de la stabilité au petites perturbations dans les grands reseaux électriques : optimisation de la regulation par une méthode métaheuristique», Thèse, Université de Paul Cezanne d'Aix Marseille, 2010.
- Asada A., «Seamless Mosaicking of acoustic video images, visual surveillance technology for reinforcing an underwater security sonar system». Sea Technology, vol. 49, p. 46, December 2008.
- Brahim N., S. Daniel, D. Guériot and B. Solaiman, «Vers une méthodologie de reconstruction 3D de scènes sous-marines par caméras acoustiques : Problématiques et études préliminaires», Traitement du signal, Lavoisier, pp. 521-551, 2012.
- Brahim N., S. Daniel, D. Guériot, and B. Solaiman, (a), «3D reconstruction of underwater scenes using DIDSON acoustic sonar image sequences through evolutionary algorithms», Conference Oceans 2011, MTS/IEEE, Santander, Spain, pp. 1-6, 2011.
- Brahim N., S. Daniel, D. Guériot, and B. Solaiman, (b) «Combinaison de l'information curviligne et ponctuelle pour une caractérisation robuste de la forme des objets dans des séquences d'images acquises par des caméras acoustiques », Conférence Traitement et Analyse de l'Information Méthodes et Applications TALMA'2011, Hammamet, Tunisie, 2011.
- Brahim N., D. Guériot, S. Daniel and B. Solaiman, «3D Reconstruction of Underwater Scenes Using Image Sequences from Acoustic Camera», Conference Oceans 2010, MTS/IEEE, Sydney, Australie, 2010.
- Brahim N., «Évaluation du potentiel des capteurs acoustiques sous-marins dans l'inspection des infrastructures portuaires», Mémoire de maitrise, Université Laval, 107 p. 2009.
- Brahim N., S. Daniel and D. Guériot (e), «Potential of underwater sonar systems for port infrastructure inspection", OCEANS 2008, 15-18 Septembre 2008, Québec, Canada, 2008.
- Back T., M. Schutz and S. Khuri, «A comparative study of a penalty function, a repair heuristic, and stochastic operators with the set-covering problem», 9 p., 1995.
- Balmisse G., "Les algorithmes génétiques", 5p., 2002.
- Bauer, J., N. Sünderhauf and P. Protzel, «Comparing Several Implementations of Two Recently Published Feature Detectors», Proceeding of the International Conference on Intelligent and Autonomous Systems, IAV, Toulouse, France, pp. 1-6, 2007.
- Baumgartner B., «Particle Swarm SE-AMC», Université de Munich, 70 p., 2008.
- Bay H., T. Tuytelaars and L. V. Gool, «SURF: Speeded Up Robust Features», Proceedings of the ninth European Conference on Computer Vision, 14 p., 2006.
- Beaudet P., «Context dependent interpolation». Image Science Mathematics Symposium, pp. 234-236, 1976.

- Beaudet P., «Rotationally invariant image operators». International Conference on Pattern Recognition, pp. 579-583, 1978.
- Belcher E. O., «Vision in Turbid Water», Proceedings of MTS/ADCI Underwater Intervention Conference, New Orleans, 5 p., 2007.
- Belcher E. O., J. R., Barone, D. G. Gallagher and R. E. Honaker, «Acoustic lens camera and underwater display combine to provide efficient and effective hull and berth inspections», MTS Proc. Of oceans, pp. 22-26, 2003.
- Ben Amor B., «Contributions à la reconstruction, la reconnaissance et l'authentification faciale 3D», Thèse de doctorat, Ecole Centrale de Lyon, Décembre 2006.
- Berro A., «Optimisation multiobjectif et stratégies d'évolution en environnement dynamique», Thèse de doctorat. Toulouse. France, 170 p., 2001.
- Beus H.L., and S.S.H. Tiu, «An Improved Corner Detection Algorithm Based on Chain Coded Plane Curves», Pattern Recognition, vol. 20, pp. 291-296, 1987.
- Boukerroui D., «Segmentation bayésienne d'images par une approche markovienne multirésolution. Application à l'imagerie échocardiographique et échographique 2D et 3D du sein», Thèse de Doctorat, INSA Lyon, 2000.
- Bouzarkouna Z., «Optimisation de placement des puits», Thèse, Université Paris-Sud, 132 p., 2012.
- Bouziani M. and J. Pouliot, «Inventaire et analyse des opérateurs d'appariement géométrique d'objets géographiques linéaires», Colloque National Géomatique de l'association canadienne des sciences géomatiques, Montréal, 15 p., 2004.
- Bray M., E. K. Meier, P. Müller, L. V. Gool and N. Schraudolph, «3D hand tracking by rapid stochastic gradient descent using a skinning model», 1st European conference on visual media production, pp. 59-68, 2004.
- Brown M., D. Burschka, D. Hager, «Advances in computational stereo», IEEE transaction on Pattern analysis and machine intelligence, vol. 25, pp. 993-1008, 2003.
- Burwen, D. L., S. J. Fleischman, and J. D. Miller, «Evaluation of a dual-frequency imaging sonar for estimating fish size in the Kenai River», Alaska Department of Fish and Game, Fishery Data Series No. 07-44, 2007.
- Busque M., «Reconnaissance d'objets à partir de points 3D obtenus par appariement stéréoscopique pour la localisation absolue d'un robot mobile», Mémoire Maitrise, Université Laval, 166 p., 2001.
- Caimi F. M., M. K. Donna Dalglish and J. Watson, «Underwater Imaging and Optics: Recent Advances», IEEE Oceanic Engineering Society Newsletter, pp. 21-29, April, 2010.
- Canny J., «A Computational Approach To Edge Detection», IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, p.679-714,1986.
- Carette V., «Amélioration de la représentation géométrique 2D et 3D des agrégations de poissons en support à l'étude de leur évolution spatio-temporelle», Maitrise, Département Sciences Géomatiques, Université Laval, 156 p., 2008.
- Chailloux, C., «Recalage d'images sonar par appariement de régions. Application à la génération d'une mosaïque». Thèse de doctorat, ENST Bretagne, Université de Rennes, 2007.
- Chuang K.S., H.L. Tzeng, S. Chen, J. Wu, T.J. Chen, «Fuzzy c-means clustering with spatial information for image segmentation», Computer Medcin Imaging Graph, vol 30, pp. 9-15, 2006.

- Collet C., P. Thourel, M., Mignotte, P. Perez and P. Bouthemy, «Segmentation markovienne hiérarchique multimodèle d'images sonar haute résolution», *Traitement du Signal*, vol 15, pp. 231-250, 1998.
- Com. Europ; Communiqué de presse, «Lancement du réseau d'excellence européen ESONET pour l'observation et la mesure des processus en fond de mer». Brest, le 21 mars 2007. <http://www.ifremer.fr/com/dossier-presse/Esonet>.
- Coustya J., L. Najmana, F. Diasa and J. Serraa, «Morphological filtering on graphs», *Computer Vision and Image Understanding*, vol 4, pp. 370-385, 2013.
- Craciun D. I., «Modélisation des équivalents dynamiques des réseaux électriques», Thèse, Université de Grenoble, 174 p., 2011.
- Cugier P. and P. Le Hir, «Development of a 3D Hydrodynamic Model for Coastal Ecosystem Modelling. Application to the Plume of the Seine River (France) », *Coastal and Shelf Science*, pp. 673-695, 2002.
- Cugier P., A. Menesguen and G. F. Guillaut, «Modélisation écologique tridimensionnelle (3D) de la baie de Seine orientale (Manche, France)», *Hydroécol*, pp 21-35, 2001.
- Daniel S., «Fusion multisource appliquée à la reconnaissance d'objets dans le milieu sous-marin», Thèse de doctorat, Université de Rennes I, 188 p., 1998.
- Daniel S., F. L. Léannec, C. Roux, B. Solaiman and E. P. Maillard, «Side-Scan Sonar Image Matching», *IEEE Journal of Oceanic Engineering*, vol 23, pp. 245-259, 1998.
- Demigny D., Pons, J., Boudouani, N. and Kessal, L., «Réalisation récursive temps réel de filtre RIF : filtre de Canny, filtre gaussien et ses dérivées», 18^{ème} Colloque sur le Traitement de Images et du Signa, 4p, 2001.
- Delaney, J., C. Barnes, P. Beauchamp, A. Chave, J. Madden, and M. McNutt, «Project NEPTUNE: An interactive, regional cabled ocean observatory in the northeast Pacific», *Proceeding IEEE, Oceans 2003*, San Diego, pp. 1231-1235, CA, 22-26, 2003.
- Deriche R. and Giraudon G., «Accurate corner detection: an analytical study», *Proceeding IEEE, ICCV*, pp.66-70, 1991.
- Douglas D.H. and T.K. Peucker, «Algorithms for the reduction of the number of points-required to represent a line or its caricature», *The Canadian Cartographer*, vol 10, pp. 112-122, 1973.
- Dura E, J. Bell and D. Lane, «Reconstruction of textured seafloors from side-scan sonar images», *IEE Proceeding, Radar, Sonar and Navigation*, vol 151, pp. 114-126, 2004.
- Ejday M., «Optimisation Multi-Objectifs à base de Métamodèle pour les Procédés de Mise en Forme», Thèse, L'Ecole nationale supérieure des mines de Paris, 136 p., 2011.
- Ferreira, M., V. Årskog, S. Jalali, and O.E. Gjør, (a), «Probability-Based Durability Analysis of Concrete Harbor Structures», *International conference of Concrete under sever conditions: Environment & loading, CONSEC'04*, Seoul, Korea, pp. 999-1006, 2004.
- Ferreira, M., S. Jalali and O.E. Gjør, «Probabilistic assessment of the durability performance of concrete structures», *Revista do DEC, UMinho*, 10 p., 2004.
- Follet H., «Caractérisation Biomécanique et Modélisation 3D par Imagerie X et IRM haute résolution de l'os spongieux humain : évaluation du risque fracturaire», Thèse de doctorat, Université de Lyon, 414 p. 2003.
- Gagné C., «Algorithmes évolutionnaires appliqués à la reconnaissance des formes et à la conception optique», Thèse, Université Laval, 212 p., 2005.

- GESMA, «Reconnaissance d'objets par l'analyse de séquences d'images acoustiques, Rapport : Poste 3 : Étude de mise en correspondance des images successives», Janvier 2009.
- GESMA, «Reconnaissance d'objets par l'analyse de séquences d'images acoustiques, Rapport : Poste 1 : Etude bibliographique & définition de la méthode envisagée», Mars 2008.
- Gonzalez C. and E. Woods, «Digital Image Processing» 3ed. Prentice-Hall, 954 p., 2008.
- Gonzalez, J. P. and U. Ozguner, «Lane detection using histogram-based segmentation and decision trees». Proceeding of the Intelligent Transportation Systems, IEEE, pp. 346-351, 2000.
- Grand-brochier, Christophe Tilmant, Michel Dhome, «Descripteur local d'image invariant aux transformations affines», Congrès des jeunes chercheurs en vision par ordinateur (ORASIS), Trégastel, France, 2009.
- Guo J., W.S. Cheng, C. Y. Ying, and T. Liu, «Image Registration for the Underwater Inspection Using the Maximum a Posteriori Technique» IEEE Journal of Ocean Engineering, vol 28, no 1, pp. 203-211, 2003.
- Hanke M., «A regularizing Levenberg - Marquardt scheme, with applications to inverse groundwater filtration problems», *Inverse Problems*, vol. 13, 1997.
- Hansen K., «The CMA Evolution Strategy: A Tutorial», 35 p., 2011.
- Hansen N. and A. Ostermeier, «Completely Derandomized Self-Adaptation in Evolution Strategies», *Evolutionary Computation*, vol. 9, pp. 159-195, 2001.
- Hartley R. and A. Zisserman, «Multiple View Geometry in Computer Vision Cambridge» University Press, 2004.
- Harris C. and M. Stephens. "A Combined Corner and Edge Detector», Proceedings of the 4th Alvey Vision Conference, pp.147-151, 1988.
- He X. and N. Yung, «Curvature Scale Space Corner Detector with Adaptive Threshold and Dynamic Region of Support», Proceeding 17th International Conference Pattern Recognition, pp. 791-794, 2004.
- Hu X. and L. Ye, «A fast and simple method of building detection from LiDAR data based on scan line analysis», The ISPRS Workshop on 3D Virtual City Modeling, Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol 2, pp. 7-13, 2013.
- Jenkinson M. and S. Smith, «A global optimisation method for robust affine registration of brain images», *Medical Image Analysis*, vol. 5, pp. 143-156, 2011.
- Johnson A.E. and M. Hebert, «Seafloor Map Generation for Autonomous Underwater Vehicle Navigation», *Autonomous Robots*, vol 3, pp. 145-168, 1996.
- Jun H. and A. Asada, «Acoustique counting method of upstream juvenile Ayu *Plecoglossus altivelis* by using DIDSON», *Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies*, pp. 459-462, 2007.
- Juneja M., P. S. Sandhu, «Performance Evaluation of Edge Detection Techniques for Images in Spatial Domain», *International Journal of Computer Theory and Engineering*, vol 1, pp. 614-621, 2009.
- Tarsha Kurdi F., «Extraction et reconstruction de bâtiments en 3D à partir de relevés lidar aéroportés», Thèse, Université Louis Pasteur, Strasbourg, 291p. 2008.
- Kant R. A., «Abs-Laplacian and Robert's cross operator offers high speed edge detection capabilities with comparable speed-quality tradeoffs», *International Journal of Computer Science & Engineering Technology (IJCSET)*, vol 4, pp. 977-986, 2013.

- Khatri R.P. and V. Sirivivatnanon, «Characteristic service life for concrete exposed to marine environments», *Cement Concrete Research*, vol. 34, pp. 745-752, 2004.
- Klein S., P.W. Josien Plum, M. Staring and A., Max «Viergever Adaptive Stochastic Gradient Descent Optimisation for Image Registration», *International Journal Computer Vision*, vol 81, pp. 227-239, 2009.
- Koussa C., «Implantation d'un système d'information géographique 3D sur Internet pour la gestion des modèles urbains», Thèses de doctorat, Université de Strasbourg, 230p. 2011.
- Lahamy H., «Outlining building using airborne Laser Scanner Data», Thèse, International Institute for Geo-Information Science and Earth Observation, 93p. 2008.
- Laligant O., Reynaud N. and Truchetet F., «Generalization of Shen-Castan and Canny-Deriche filters», *Intelligent Robots and Computer Vision XVII: Algorithms, Techniques, and Active Vision*, vol 54, 1998.
- Le Riche R., M. Schoenauer and M. Sebag, «Un état des lieux de l'optimisation évolutionnaire et de ses implications en sciences pour l'ingénieur», *Modélisation Numérique: défis et perspectives, Traité Mécanique et Ingénierie des Matériaux*, Edition, P. Breikopf and C. Knopf-Lenoir, pp. 187-259, 2007.
- Lecouvé M., «Conception et réalisations de filtres microondes à modes évanescents à l'aide d'un algorithme génétique : Egalisation d'amplitude par l'algorithme de Levenberg-Marquardt», Thèse de doctorat, Université Bordeaux 1, 160 p., 2000.
- Lee J., Y. Sun, and C. Chen, «Wavelet Transform for Corner Detection», *Proceeding. IEEE Conference Systems Eng.*, 1992.
- Lowe D. G., «Distinctive Image Features from Scale-Invariant Keypoints», *International Journal of Computer Vision*, vol 60, pp. 91-110, 2004.
- Lu, Y. and M. Oshima, «3-D reconstruction of the underwater acoustic images using the GPS positioning and the image matching technology», *Oceans'02 MTS/IEEE*, pp. 2273-2278, 2002.
- Lutton E., J. D. Fekete, «Visual analytics and experimental analysis of evolutionary algorithms», *Rapport de recherché, Institut National de Recherche en Informatique et en Automatique*, 22 p., 2011.
- Maître H., «Le traitement des images», Number ISBN 2-7462-0584-X. Hermès Science publique, 2003.
- Medioni G. and Y. Yasumoto, «Corner detection and curve representation using cubic splines», *Computer Vision, Graphics and Image Processing*, vol 39, pp. 279-290, 1987.
- Merad D., «Reconnaissance 2D/2D et 2D/3D d'objets à partir de leurs squelettes», Université d'Evry-Val d'Essonne, 210 p., 2004.
- Mingqiang Y., K. I. Kpalma, J. Ronsin, «A Survey of Shape Feature Extraction Techniques», *Pattern Recognition*, pp. 43-90, 2008.
- Mignotte M., «Segmentation d'images sonar par approche markovienne non supervisée et classification d'ombres portées par modèles statistiques». Thèse doctorat. Université de Bretagne Occidentale, 239 p. 1998.
- Mikolajczyk K., and C. Schmid, «Scale & Affine Invariant Interest Point Detectors», *International Journal of Computer Vision*, vol 60, pp. 63 – 86, 2004.
- Moore F. and B. Babb, «Revolutionary Image Compression and Reconstruction via Evolutionary Computation, Part 2: Multiresolution Analysis Transforms», *WSEAS Transactions on Signal Processing*, vol. 2, pp. 1209-1214, 2006.
- Moravec H., «Techniques towards automatic visual obstacle avoidance». In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, 584 p., 1977.

- Moulon P., P. Monasse and R. Marlet, «Adaptive Structure from Motion with a contrario model estimation», ACCV'12 Proceedings of the 11th Asian conference on Computer Vision, vol. Part IV, pp. 257-270, 2012.
- Mouragnon E., M. Lhuillier, M. Dhome, F. Dekeyser and P. Saydy, «3D reconstruction of complex structures with bundle adjustment: an incremental approach», Proceedings of the IEEE International Conference on Robotics and Automation, Orlando, USA, pp. 3055-3061, 2006
- Negahdaripour S., H. Sekkati, and H. Pirsiavash, «Opti-acoustic stereo imaging: on system calibration and 3-D target reconstruction», IEEE Transaction on Image Processing, vol 18, no 6, pp. 1203-1214, 2008.
- Negahdaripour S., H. Sekkati, «3-D Motion Estimation for Positioning from 2-D Acoustic Video Imagery», Pattern Recognition and Image Analysis, vol 4478, pp 80-88, 2007.
- Negahdaripour S., «Calibration of DIDSON forward-scan acoustic video camera», MTS/IEEE OCEANS '05, vol. 2, pp. 1287-1294, September 2005.
- Niemisto A., T. Korpelainen, R. Saleem, O. Yli-Harja, L. Aitchison, and L. Shmulevich, «A K-means segmentation method for finding 2-D object areas based on 3-D image stacks obtained by confocal microscopy», in Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'07), Lyon, France, 4p., 2007.
- Nillius, P. and J.-O. Eklundh, «Fast Block Matching with Normalized Cross-Correlation using Walsh Transforms», report, ISRN KTH/NA/P--02/11--SE, Sept. 2002.
- Nistér d., O. Naroditsky, J. R. Bergen, «Visual odometry for ground vehicle applications». J. Field Robot, vol 23, pp. 3-20, 2006.
- Nister D. «Preemptive RANSAC for Live Structure and Motion Estimation» Proceedings of the Ninth IEEE International Conference on Computer Vision, vol 2, p. 199, 2003.
- Oshin O., J. Illingworth and R. Bowden, «Spatio-Temporal Feature Recognition using Randomised Ferns» International Workshop on Machine Learning for Vision Based Motion Analysis, ECCV08, Marseille, France, 2008.
- Ozuysal M., P. Fua, V. Lepetit, «Fast keypoint recognition in ten lines of code», In Proceeding IEEE Conference on Computing Vision and Pattern Recognition, 8p. 2007.
- Parisot P., «Suivi d'objets dans des séquences d'images de scènes déformables: de l'importance des points d'intérêt et du maillage 2D», Thèse, Université de Toulouse, 207 p., 2009.
- Pham D. L., J. L. Prince, «An Adaptive Fuzzy C-Means Algorithm for Image Segmentation in the Presence of Intensity Inhomogeneities», Pattern Recognition Letters, vol 20, pp. 57-68, 1998.
- Pornon H., «3D et les SIG: Etat de l'art et perspectives», Géomatique Expert - N° 71, pp. 46-51, 2009.
- Pouliot, J., B. Lachance and B. Kirkwood, «L'importance de la modélisation géométrique 3D lors de l'élaboration d'un SIG 3D: Exemple du développement d'une structure topologique pour une application géologique», La revue internationale de géomatique: Information géographique tridimensionnelle, vol 16 no 1, pp. 29-49, 2006.
- Rabaud C., «Une nouvelle approche de mise en correspondance stéréoscopique dense par méthodes possibilistes», Thèse de doctorat, Université de Montpellier, 139 p, 2005.
- Ramos F., «Modélisation et Validation d'un Système d'Information Géographique 3D opérationnel». Thèse de doctorat, Université de Marne La Vallée, 175 p, 2003.
- Rapport SIVOM, SIVOM du Littoral des Maures, «Surveillance de l'herbier de Posidonie Observatoire Marin SIVOM du Littoral des Maures», Observatoire marin, 2008, <http://www.observatoire-marin.com>.

- Rechenberg I., «Cybernetic Solution Path of an Experimental Problem». Ministry of Aviation, Royal Aircraft Establishment (U.K.), 1965.
- Reed S., A. Cormack, K. Hamilton, T. Ruiz and D. Lane, «Automatic Ship Hull Inspection using Unmanned Underwater Vehicles (UUV's) », Proceedings from the 7th International Symposium on Technology and the Mine Problem. Monterey, USA. 2006.
- Rodriguez S. A., Frémont V. and Bonnifait P., «An Experiment of a 3D Real-Time Robust Visual Odometry for Intelligent Vehicles», IEEE International Conference on Intelligent Transportation Systems, ITSC '09, États-Unis d'Amérique, pp. 1-6, 2009
- Rominger C. and A. Martin, «Recalage et fusion d'images sonar multi-vues : utilisation du conflit», Atelier Fouille de données complexes - Complexité liée aux données multiples, Extraction et Gestion des Connaissances (EGC), Hammamet, Tunisie, 26 janvier 2010.
- Roussel D., «Reconstruction de Courbes et de Surfaces 3d en Stéréo-Acquisition». Thèse de doctorat de l'Université Paris XI, 1999.
- Rosenfeld A., and J. S. Weszka, «An improved method of angle detection on digital curves», IEEE Transactions on Computer Vision, Vol 24, pp. 940-941, 1975.
- Rosten E., R. Porter and T. Drummond, «Faster and better: A machine learning approach to corner detection», IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, pp. 105 –119, 2010.
- Rosten E. and T. Drummond, «Fusing points and lines for high performance tracking», IEEE International Conference on Computer Vision, pp. 1508-1511, 2005.
- Roux M, «Algorithmes de classification», éditions Masson, Paris, 81p. 2006.
- Schmid C., R. Mohr and C. Bauckhage, «Evaluation of Interest Point Detectors», International Journal of Computer Vision, vol. 37, pp. 151-172, 2000.
- Schomaker, L. and M. Bulacu, «Automatic writer identification using connected-component contours and edge-based features of uppercase Western script», IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, pp. 787 - 798, 2004.
- Sebari I. and D. C. HE, «Les approches de segmentation d'images par coopération régions-contours», Revue Télédétection, vol. 7, pp. 499-506, 2007.
- Smith M., J., Brady, «SUSAN—A New Approach to Low Level Image Processing», International Journal of Computer Vision, vol 23, pp. 45-78, 1997.
- Sosa J. C., R. Rodríguez, V. H. García Ortega, R. Hernández, «Real-time Optical-flow Computation for Motion Estimation under Varying Illumination Conditions», International Journal of Reconfigurable and Embedded Systems (IJRES) vol. 1, no. 1, pp. 25-36, 2012.
- Suganthan P. N., N. Hansen, J. J. Liang, K. Deb, Y.-P. Chen, A. Auger, and S. Tiwari, «Problem definitions and evaluation criteria for the cec 2005 special session on real-parameter optimization». Technical Report Kan-GAL Report 2005005, IIT Kanpur, India, 2005.
- Sun N., T. Shim and M. Cao, «3D reconstruction of seafloor from sonar images based on the multi-sensor method», IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, pp. 573-577, 2008.
- Tunncliffe, V., C. R. Barnes and R. Dewey, «Major advances in cabled ocean observatories (VENUS and NEPTUNE Canada) in coastal and deep sea settings», Proc. US/EU-Baltic International Symposium, IEEE/OES, 7p., 2008.
- Trucco A. and S. Curletto, «Extraction of 3-D Information from Sonar Image Sequences», IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics, vol. 33, no. 4, pp.687-699, 2003.

- Trucco E. and A. Verri, «Introductory 3D techniques for 3D computer vision», ISBN 0-13-261108-2, Prentice Hall, Upper Saddle River, USA, 343 p., 1998.
- Victor J. S. and J. Sentieiro, «The Role of Vision for Underwater Vehicles», IEEE International Symposium on Autonomous Underwater Vehicle Technology - AUVT94, Boston, USA, 1994.
- Vincent, O. R. and O. Folorunso, «A Descriptive Algorithm for Sobel Image Edge Detection». Proceeding of the Information Science and Information technology Education Joint Conference, Macon, Georgia USA, pp. 97-107, 2009.
- Walker, I.J. and J.V. Barrie, «Geomorphology and sea-level rise on one of Canada's most sensitive coasts: Northeast Graham Island, British Columbia», Journal of Coastal Research, pp.220-226, 2006.
- Walther GR, E. Post, P. Convey, A. Menzel, C. Parmesan, T.J. Beebee, JM. Fromentin, O. Hoegh-Guldberg and F. Bairlein, «Ecological responses to recent climate change», NATURE, pp. 389-95, 2002.
- Wang H., S. Tang, et X. Fan, «A new unsupervised approach to word segmentation. Computational Linguistics», vol 37, pp. 421-454, 2011.
- Wang X. «Laplacian Operator-Based Edge Detectors», IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 29, no 5, pp. 886-890, 2007.
- Weidner U., and W. Fostner, «Towards automatic building reconstruction from high resolution digital elevation models», ISPRS Journal of Photogrammetry and Remote Sensing, vol 50, pp. 30-49, 1995.
- Wenbin L., «Efficient Removal of Impulse Noise from Digital Images», IEEE Transactions on Consumer Electronics, vol. 52, no 2, pp 523-527, 2006.
- Wright D. J., and M. F. Goodchild, «Data from the deep: implications for the GIS community», International Journal of Geographical Information Science, pp. 523-528, 1997.
- Zerr, B. and B. Stage, «Three-dimensional reconstruction of underwater objects from a sequence of sonar images», Proc. Int. Conf. Image Processing, pp. 927-930, 1996.

Annexe A

Liste des publications et participations scientifiques

Article de revue avec comité de lecture

Brahim N., S. Daniel, D. Guériot and B. Solaiman, "Vers une méthodologie de reconstruction 3D de scènes sous-marines par caméras acoustiques : Problématiques et études préliminaires", *Traitement du signal*, Lavoisier, pp. 521 - 551, 2012.

Communications dans des conférences avec comité de lecture

Brahim N., S. Daniel, D. Guériot, and B. Solaiman, "3D reconstruction of underwater scenes using DIDSON acoustic sonar image sequences through evolutionary algorithms", *Conference Oceans'11, MTS/IEEE*, 06 - 09 June, Santander, Spain, 2011.

Brahim N., S. Daniel, D. Guériot, and B. Solaiman, "Combinaison de l'information curviligne et ponctuelle pour une caractérisation robuste de la forme des objets dans des séquences d'images acquises par des caméras acoustiques", *Conférence Traitement et Analyse de l'Information Méthodes et Applications TAIMA'2011*, 3 - 8 Octobre, Hammamet, Tunisie, 2011.

Brahim N., D. Guériot, S. Daniel and B. Solaiman, "3D Reconstruction of Underwater Scenes Using Image Sequences from Acoustic Camera", *Conference Oceans'10, MTS/IEEE*, 24 - 27 May, Sydney, Australie, 2010.

Brahim N., D. Guériot, S. Daniel and B. Solaiman, "Reconstruction tridimensionnelle de scènes sous-marines à partir d'images acoustiques", *Conférence Hydrographique Canadienne*, 21 - 23 June, Québec, Québec, 2010.

Annexe B

Dans les sections, ci-dessous, sont présentés les algorithmes de reconstruction 3D par approche volumique. Quant aux algorithmes de la reconstruction 3D par approche ponctuelle, ils ont le même principe que ceux de l'approche volumique sauf au niveau de la modélisation des points saillants (la résolution du capteur n'est intégrée dans le calcul) et au niveau de l'erreur calculée (seule l'erreur en portée et en azimut est calculée).

1) L'algorithme qui permet de calculer les solutions optimales du déplacement de la caméra par exploitation de l'algorithme d'optimisation SE-AMC.

```
% x_init : Vecteur des solutions initiales
%'sum_inter_combined' : la fonction objective
%Cmaes : Algorithme d'optimisation : Stratégie d'Evolution par Adaptation de la Matrice de Covariance
%data : Structure de données, elle contient le nombre d'images utilisées (data.nSets), le nombre de points
saillants (data.nPoints), les coordonnées des points en portée (data.r), en azimut (data.theta)
et l'inclinaison de la caméra (data.tilt).

% Précision du critère d'arrêt
opts.StopFitness= -0.001 ;

[ xmin, ... % Vecteur des solutions optimales du déplacement de la caméra
  f, ...   % Valeur de la fonction objective
] = cmaes ('sum_inter_combined', x_init , 0.1 , opts, data)

% Les solutions xmin vont être utilisées pour calculer les coordonnées des points saillants en appliquant le
modèle géométrique de la caméra.
```

2) La fonction objective

```
function [out_err] = sum_inter_combined (x , data)
% Retourne la distance cumulée entre les meilleures distances de reconstruction pour chaque point.
% [IN]
% x : Vecteur contenant les paramètres de la transformation à estimer
% data : Structure contenant les jeux de données à utiliser
% [OUT]
% out_err : Erreur totale estimée à partir de l'utilisation des solutions des paramètres de déplacement entre
les jeux de données

err_rt = zeros (1, data.nPoints) ; % erreur horizontale sur le plan (r, theta)
err_dz = zeros (1, data.nPoints) ; % erreur en élévation
% Calcul de la distance minimale entre rubans pour un point sur l'ensemble des jeux de données
okInter = zeros (data.nSets) ;
for kPoint = 1 : data.nPoints
    % En considérant une translation identique entre deux jeux de données
    % consécutifs, calcul des distances minimales entre les rubans pour le
    % point kPoint
    [e_rt , e_dz , okInterP] = find_common_combined (x, data, kPoint) ;
    % Conservation des erreurs
```

```

err_rt (kPoint) = e_rt ;
err_dz (kPoint) = e_dz ;
okInter = okInter + okInterP ;
end
% Erreur globale sur l'ensemble des points du jeu de données : sommation de toutes les distances
out_err = sum (err_rt(:)) + sum(err_dz(:));
end

```

3) Algorithme qui permet de calculer la distance minimale entre les rubans issus d'un même point saillant.

```

function [err_rt , err_dz , okInter] = find_common_combined (x , data , kPoint)
% Retourne la distance minimale entre les rubans issus d'un même point dans les différents jeux de données.
% Chaque point 2D (r, theta) apparié entre deux jeux induit la formation de deux rubans dont on veut calculer
l'intersection.
% Les deux valeurs retournées qualifient les intersections potentielles entre les rubans dans le plan (r, theta)
et en élévation (axe Z) pour un même point en considérant tous les jeux de données.
% [IN]
% x : Vecteur contenant la translation
% data : Jeux de données
% kPoint : Indice du point à considérer dans les différents jeux
% [OUT] :
% err_rt : Plus petite distance entre polygones dans le plan (r, theta)
% (0 si intersection)
% err_dz : Plus petite distance entre les rubans (en Z)
% (0 si intrestion)
% Si les err_rt et err_dz sont égales à 0, alors la transformation étudiée est compatible avec l'acquisition
traitée.

err_rt = 0.0 ;
err_dz = 0.0 ;
okInter = zeros (data.nSets , data.nSets) ;
l_low = zeros (data.nSets , data.nSets , 3) ;
l_low(:) = -inf ;
l_up = zeros (data.nSets , data.nSets , 3) ;
l_up(:) = inf ;

for kSet1 = 1 : data.nSets

    for kSet2 = kSet1+1 : data.nSets

        % Extraction de la transformation adéquate
        % xl est exprimée dans le repère du jeu 1
        xl = extract_transformation (x , kSet1 , kSet2) ;
        % Transformation jeu 1 vers kSet1
        xl1 = extract_transformation_one (x , kSet1 , 6) ;
        % Transformation exprimée dans le repère du jeu 1 du vecteur x
        % La transformation est à projeter dans le repère du jeu kSet1
        Rz = rotation (xl1(4:6)) ;
        Rzi = rotation(-xl1(4:6)) ;
        T = xl1(1:3) ;
    end
end

```

```

% Coordonnées de la caméra dans le repère de 1 (translation avec la rotation du repère 1)
xl(1:3) = (Rzi * xl(1:3))';
% Erreurs sur le point kPoint
[err_rtl , err_dzl , ll_low , ll_up] = compute_errors (xl , data , kPoint , kSet1 , kSet2) ;
% Les coordonnées de ll_low et ll_up doivent être remis dans le repère de jeu 1 pour être comparable
aux autres reconstructions
if (~isempty(ll_low) && ~isempty(ll_up))
    ll_low(kSet1,kSet2,:) = Rz * ll_low + T ;
    ll_up(kSet1,kSet2,:) = Rz * ll_up + T ;
end
% Cumul des erreurs en distance
err_rt = err_rt + err_rtl ;
err_dz = err_dz + err_dzl ;

if (err_rtl+err_dzl < 1e-16)
    okInter(kSet1,kSet2) = okInter(kSet1,kSet2) + 1 ;
end

end
end

% Déterminer l'intersection en Z entre tous les rubans
l = ll_low(:, :, 3) ;
min_z = max (l(:)) ;
l = ll_up(:, :, 3) ;
max_z = min (l(:)) ;

if (max_z < min_z)
    % Dans ce cas, il n'y a pas d'intersection entre les Delta Z
    err_dz = err_dz + (min_z - max_z) ;
end

end

```

4) Algorithme qui permet calculer l'erreur horizontale et verticale entre les polygones.

```

function [err_rt err_dz ll_low ll_up] = compute_errors (xl , data , kPoint , kSet1 , kSet2)
% Retourne l'erreur calculée entre les polygones en élévation et en (portée et azimut)
% Chaque point 2D (r , theta) apparié entre deux jeux induit la formation de deux rubans dont on veut calculer
l'intersection.
% Les deux valeurs retournées qualifient les intersections potentielles entre les rubans dans le plan (r , theta)
et en élévation (axe Z) pour un même point en considérant tous les jeux de données.
% [IN]
% xl : Vecteur contenant la translation
% data : Jeux de données
% kPoint : Indice du point à considérer dans les différents jeux
% [OUT] :
% err_rt : Plus petite distance entre polygones dans le plan (r , theta)
% (0 si intersection)
% err_dz : Plus petite distance entre les rubans (en Z)
% (0 si intrestion)

```

```

% l_low : si distance nulle, coordonnées de bas de l'intersection
% l_up : si distance nulle, coordonnées de haut de l'intersection
% Si les err_rt et err_dz sont égales à 0, alors la translation étudiée
% est compatible avec l'acquisition traitée (à la précision retenue).

l_low = [];
l_up = [];
% Extraction des matrices de la transformation entre le jeu kSet1 et kSet2
R = rotation(xl(4:6));
Rm = rotation(-xl(4:6));
T = [xl(1); xl(2); xl(3)];
% Distances entre polygones pour les paires de jeu de données
e_rt = [];
% Hauteurs min et max des rubans pour le point kPoint reconstruit à partir des différents jeux de données
z_min = [];
z_max = [];
z_best = [];

% Recherche de la meilleure intersection pour le point kPoint considéré entre les jeux de données kSet 1 et
kSet2 avec le déplacement (R, T)
% e : distance mini entre polygones sur le plan (r, theta)
% Si e == 0 alors
% [p2_min p2_max] : plage d'élévation (phi) avec intersection
% [h_min h_max] : plage d'élévation (en z) avec intersection
% Sinon
% p2_best : élévation ayant produit la distance la plus faible
% h_best : altitude ayant produit la distance la plus faible
[e1 p2_min p2_max h_min h_max p2_best h_best M_low M_up] = ...
search_elevation_range(data, kSet1, kSet2, kPoint, R, T, 1);

% Les coordonnées de M_low et M_up sont données dans le repère associé à kSet1
% Il faut tout replacer dans un même repère pour déterminer l'intersection en Z entre les rubans
if (isempty(M_low) == 0 && isempty(M_up) == 0)
% Conservation de la configuration d'intersection des rubans entre le jeu kSet1 et le jeu kSet2 (dans le plan
Z)
[z, kVertex] = max(M_up(3,:));
l1_up = M_up(:,kVertex);
[z, kVertex] = min(M_low(3,:));
l1_low = M_low(:,kVertex);
end
z_best = [z_best h_best];
% On fait la même recherche dans l'autre sens
[e2 p2_min p2_max h_min h_max p2_best h_best M_low M_up] = ...
search_elevation_range(data, kSet2, kSet1, kPoint, Rm, -T, 2);
% Les coordonnées de M_low et M_up sont données dans le repère associé à kSet2
% Il faut tout replacer dans un même repère pour déterminer l'intersection en Z entre les rubans
if (isempty(M_low) == 0 && isempty(M_up) == 0)
% Transformation des M21 vers le repère 1 pour affichage
M_low = R * M_low + repmat(T,1,4);
M_up = R * M_up + repmat(T,1,4);
% Conservation de la configuration d'intersection des rubans entre le jeu kSet1 et le jeu kSet2 (dans le plan

```

```

Z)
[z , kVertex] = max (M_up(3,:)) ;
l2_up = M_up (:,kVertex) ;
[z , kVertex] = min (M_low(3,:)) ;
l2_low = M_low (:,kVertex) ;
end
z_best = [z_best h_best] ;
% Calcul de l'erreur globale sur la distance entre polygones
%
err_rt = e1+e2 ;
% Calcul de l'erreur globale en altitude
% Y a-t-il intersection entre les polygones pour chaque paire de jeu ?
if (err_rt > 0)
% Il y a au moins une paire de jeux sans intersection entre les polygones. L'erreur en altitude est donc l'écart
entre les meilleures positions en Z des polygones les plus proches
err_dz = max (z_best(:))-min(z_best(:)) ;
else
% Y a-t-il intersection commune entre tous les rubans dans le plan Z ?
% oui si le max des altitudes inférieures (min_z) est inférieur au min des altitudes supérieures (max_z)
min_z = max(z_min(:)) ;
max_z = min(z_max(:)) ;
l = [l1_low l2_low] ;
[min_z , kVertex] = max (l(3,:)) ;
l_low = l(:,kVertex) ;
l = [l1_up l2_up] ;
[max_z , kVertex] = min (l(3,:)) ;
l_up = l(:,kVertex) ;
% Pas d'intersection commune en altitude entre tous les rubans.
% L'erreur résiduelle est donc la différence entre min_z et max_z.
% min_z correspond à un point moins profond que max_z
if (max_z < min_z)
% Dans ce cas, il n'y a pas d'intersection entre les Delta Z
err_dz = min_z - max_z ;
else
% Test où on a une erreur négative ce qui permet de privilégier les
% zones d'intersection maximale
%err_dz = -(max_z-min_z) ;
err_dz = 0.0 ;
end
end
if (isempty(err_dz))
err_dz
end
end

```

5) Algorithme qui permet calculer les angles d'élévation qui ont induit la distance minimale entre les polygones.

```
function [min_dist phi2_min phi2_max z_min z_max phi2_best z_best M_low M_up] = search_elevation_range
(data , kSet1 , kSet2 , k , R , T , sens)
% data: Données de la scène
% kSet1: Indice du premier jeu à apparier
% kSet2: Indice du second jeu à apparier
% k : Indice du point à analyser
% R et T : Les paramètres de la rotation et de la translation de la caméra
% min_dist : Distance minimale entre les polygones
% phi2_min : élévation où l'intersection commence
% phi2_max : élévation où l'intersection termine

% Approche volumique

% Définitions des rayons (R1) depuis l'image 1
R1.x = data.theta(kSet1,k) + [-data.delta_theta data.delta_theta data.delta_theta -data.delta_theta]/2;
R1.y = data.r(kSet1,k) + [-data.delta_r -data.delta_r data.delta_r data.delta_r]/2;
% Définitions des rayons (R2) depuis l'image 2
R2 = zeros (4 , 3) ;
vt2 = data.theta(kSet2,k) + [-data.delta_theta data.delta_theta data.delta_theta -data.delta_theta]/2 ;
vr2 = data.r(kSet2,k) + [-data.delta_r -data.delta_r data.delta_r data.delta_r]/2 ;

( % Approche ponctuelle
% Définitions des rayons (R1) depuis l'image 1
R1.x = data.theta(kSet1,k);
R1.y = data.r(kSet1,k);
% Définitions des rayons (R2) depuis l'image 2
R2 = zeros (4 , 3) ;
vt2 = data.theta(kSet2,k);
vr2 = data.r(kSet2,k);
)

%% Recherche par dichotomie
% Domaine de recherche sur phi d'une intersection éventuelle des deux volumes
% la dimension de la résolution en élévation est égale à l'ouverture verticale de la caméra (10°)
delta_phi = 10;
% Limite inférieure de l'intervalle de valeurs des élévations
binf = (data.tilt(kSet2)-delta_phi)*pi/180 ;
% Limite supérieure de l'intervalle de valeurs des élévations
bsup = (data.tilt(kSet2)+delta_phi)*pi/180;
% Prise en considération de l'inclinaison de la caméra
L(:,1)=(data.tilt(kSet2)-delta_phi);
% Prise en considération de l'inclinaison de la caméra
L(:,2)=(data.tilt(kSet2)+delta_phi);
binf_init = binf ;
bsup_init = bsup ;
% Distance minimale entre les deux régions possibles pour les points
% reconstruits dans le repère image 1
min_dist = inf ; % Distance minimale
```

```

phi2_best = binf ; % Meilleure solution d'élévation
z_best = 0 ; % Meilleure solution de hauteur
phi2_min = inf ; % valeur de l'élévation minimale
phi2_max = -inf ; % valeur de l'élévation maximale
z_min = inf ; % hauteur minimale
z_max = -inf ; % hauteur maximale
M_low = [] ;
M_up = [] ;
% Nombre d'échantillons par arc de cercle
n = 5 ;
% Recherche d'un minimum de distance par dichotomie
for iRun = 1:5
    % Découpage du ruban de l'image 2 en polygones
    tab_phi2 = linspace (binf, bsup, n) ;
    dst_phi2 = [] ;
    for phi2 = tab_phi2
        % M22 = (r2, theta2, phi2) -> M2/R2, conversion des coordonnées sphériques en coordonnées cartésiennes
        M22 = sph2cart (vr2, vt2, [phi2 phi2 phi2 phi2]) ;
        % Expression de M2/R2 dans R1 -> M2/R1
        if (sens == 1)
            M21 = R * M22 + repmat(T, 1, 4) ;
        else
            M21 = R * (M22 + repmat(T, 1, 4)) ;
        end
        % M21 : 3 x 4, retour vers les coordonnées sphériques
        R2(1,:) = cart2sph (M21(:,1))' ;
        R2(2,:) = cart2sph (M21(:,2))' ;
        R2(3,:) = cart2sph (M21(:,3))' ;
        R2(4,:) = cart2sph (M21(:,4))' ;
        R22.x = R2(:,2)' ;
        R22.y = R2(:,1)' ;
        % Détermination de l'intersection entre polygone R1 et R22
        dst = min_dist_between_two_polygons(R1 , R22 , 0) ;
        if (dst <= min_dist)
            min_dist = dst;
            phi2_best = phi2;
            z_best = mean(M21(3,:)) ;
            if (dst == 0.0)
                if (phi2_min >= phi2)
                    phi2_min = phi2 ;
                    M_low = M21 ;
                end
                if (phi2_max <= phi2)
                    phi2_max = phi2 ;
                    M_up = M21 ;
                end
                z_min = min ([z_min min(M21(3,:))]) ;
                z_max = max ([z_max max(M21(3,:))]) ;
            end
        end
        dst_phi2 = [dst_phi2 dst] ;
    end
end

```

```

end
% On continue la recherche si on n'a pas d'intersection
step_phi2 = (bsup-binf) / (n-1) ;
if (min_dist > 0.0)
    binf = phi2_best - step_phi2 ;
    bsup = phi2_best + step_phi2 ;
else
    % On encadre la région où on a trouvé une distance égale à 0
    binf = phi2_min - step_phi2 ;
    bsup = phi2_max + step_phi2 ;

    if (n == 100)
        break ;
    end
    n = 100 ;
end
% On ne peut pas dépasser les bornes initiales
if (binf < binf_init)
    binf = binf_init ;
end
if (bsup > bsup_init)
    bsup = bsup_init ;
end
end
end

```