

STRATÉGIES POUR ADAPTER UNE CHAÎNE DE RECONSTRUCTION 3D AU MILIEU SOUS-MARIN : DES IDÉES À LA PRATIQUE

Laurent Beaudoin¹, Loïca Avanthey¹

¹: Équipe SEAL (*Sense, Explore, Analyse and Learn*), ÉPITA Recherche, 14-16 rue Voltaire, 94270 Le Kremlin-Bicêtre, France

Résumé

Pour de nombreuses applications (environnementales, archéologiques, industrielles ou de défense et sécurité par exemple), disposer d'un modèle 3D très haute résolution de zones locales peut être important. Dans cet article, nous nous intéressons à l'obtention de nuages de points 3D denses par reconstruction. Mais les conditions d'acquisition propres au milieu sous-marin (milieu dynamique et hostile, visibilité réduite, localisation et communication sans fil complexes, etc.) rendent la tâche ardue. La plupart des travaux présents dans la littérature s'appuient sur l'expérience issue du monde aérien. Les contributions que nous présentons dans cet article portent sur l'appariement de couples d'images simultanées ou successives : détection automatique des zones d'information fiable en modifiant le détecteur de Harris, génération de germes d'appariement fiabilisée par un filtrage statistique sur le flot local, densification de l'appariement par propagation de l'information autour des germes avec exclusion automatique des zones sans information. Les résultats obtenus sur une base de données diversifiée issue de plusieurs campagnes terrain et la qualité des apports sont ensuite discutés.

Mots clés : Reconstruction 3D sous-marine, Détecteur de Harris modifié, génération de germes d'appariement fiables, densification par propagation.

Abstract

For many applications (environmental, archaeological, industrial or defense and security for example), having a very high resolution 3D model of local areas can be important. In this article, we are interested in obtaining dense 3D point clouds by reconstruction. But the acquisition conditions specific to the underwater environment (dynamic and hostile environment, reduced visibility, complex location and wireless communication, etc.) make the task difficult. Most of the work in the literature is based on experience from the air environment. The contributions we present in this article are related to the matching of simultaneous or successive image pairs : automatic detection of areas with reliable information by modifying the Harris detector, generation of matching germs made reliable by a statistical filtering on the local flow, densification of the matching by propagation of the information around the germs with automatic exclusion of areas without information. The results obtained on a diversified database resulting from several field campaigns and the quality of the contributions are then discussed.

Keywords : Underwater 3D reconstruction, modified Harris detector, generation of reliable pairing germs, densification by propagation.

1. Introduction

La cartographie tridimensionnelle sous-marine et sub-aquatique centimétrique présente un intérêt tout particulier pour la communauté scientifique travaillant sur des thématiques environnementales locales ou archéologiques, mais aussi pour le monde industriel et les applications en défense et sécurité.

Cependant, les conditions d'opération et d'acquisition particulièrement dures en milieu sous-marin rendent complexes tout le processus d'obtention de cette cartographie spécifique notamment, à partir d'images dans le visible.

L'information 3D s'extrait de l'écart de position entre des points appariés (identiques) dans des images observant la scène d'intérêt sous des angles différents. Or, non seulement les propriétés environnementales du milieu aquatique au moment de l'acquisition modifient les modèles de projection habituellement utilisés (Treibitz

et al., 2012), mais elles viennent également mettre en difficulté les algorithmes d'appariement classiques (ceux qui fonctionnent en aérien).

Les trois principaux handicaps pour ces algorithmes sont (1) la qualité dégradée et très variable des images (faible contraste, faible luminosité, perte des couleurs, éclairage non uniforme, flou, bruit, etc.), (2) l'absence d'information (texture) sur les portions de l'image où les objets sont trop éloignés du capteur (absorption) et (3) la non-rigidité de la scène observée (dynamisme des milieux naturels) qui est particulièrement prononcée dans ce milieu d'étude.

1.1. Les stratégies d'appariement dense sous-marin dans la littérature

Les stratégies d'appariement dense (tous les points) s'appuient sur un premier appariement épars des points les plus remarquables des images observant la scène d'intérêt. En effet, d'un point de vue algorithmique, plus le

nombre de pixels à apparier est grand, plus le problème devient complexe, car chaque objet représenté par un pixel sur une image peut potentiellement correspondre à n'importe quel pixel sur une autre image, si toutefois il s'y trouve. L'utilisation d'algorithmes brute-force (test systématique de toutes les combinaisons possibles) pour effectuer un tel appariement excède de loin les capacités de calculs des unités standards. Afin de rendre ce problème solvable, il est nécessaire de réduire l'espace des candidats avant d'effectuer les mesures de similarité.

La détection et la sélection des points remarquables reposent principalement au sein des travaux sous-marins sur l'emploi de trois détecteurs de points caractéristiques : SIFT (Lowe, 1999), SURF (Bay et al., 2006) et Harris (Harris et Stephens, 1988).

Le détecteur SIFT présente la particularité de fournir de nombreux points, répartis sur l'ensemble de l'image. Son descripteur lui permet, de plus, d'être robuste aux transformations (rotations, translations, homographies...). Pour ces raisons, il est très populaire au sein de la communauté aérienne et nous le retrouvons *de facto* très présent dans les chaînes sous-marines (Brandou et al., 2007; Schmidt et Rzhanov, 2012; O'Byrne et al., 2014). Cependant, Skarlatos et al. (2012) ont montré que ce détecteur était particulièrement sensible au bruit de chatonnement (*speckle*) et à l'effet de vignettage (assombrissement sur les bords) ce qui a pour conséquence d'abaisser ses performances sur des images sous-marines où ces phénomènes sont aggravés.

Partant des mêmes fondements que SIFT, le détecteur SURF favorise la vitesse d'exécution au nombre de points extraits. Néanmoins, bien que cet aspect soit très attractif pour les travaux où le temps de calcul est critique, il reste peu exploité pour le moment dans les chaînes sous-marines (Beall et al., 2010).

Quant au détecteur de Harris — que nous retrouvons souvent sous la version Shi-Tomasi (Shi et Tomasi, 1994) —, il s'attache aux zones texturées et en marque les ruptures. Très économe en temps et en mémoire par rapport aux deux précédents, nous le retrouvons au cœur des algorithmes denses fondés sur le critère spatio-temporel (suivi de cible). Il est cependant très sensible aux changements d'échelle et son descripteur de base (corrélation) est sensible aux transformations géométriques, c'est pourquoi il est couramment utilisé à travers des versions modifiées pour compenser ces défauts (Pizarro et al., 2004; Kunz et Singh, 2010).

L'appariement éparé réalisé à partir de ces points remarquables sélectionnés est ensuite utilisé en combinaison avec un critère pour réduire l'espace des candidats lors de l'étape d'appariement dense. Les principaux critères dédiés à cet usage sont les suivants : géométrique, spatio-temporel et voisinage spatial (Avanthey et al., 2016).

L'ensemble des chaînes de reconstruction 3D dense sous-marines que nous avons recensées s'appuient plutôt sur le critère géométrique (Brandou et al., 2007; Sedlazeck et al., 2009; Kunz et Singh, 2010; Prabhakar et Kumar, 2012; Schmidt et Rzhanov, 2012).

Certaines équipes ont aussi testé l'utilisation sous l'eau de logiciels de reconstruction développés spécifiquement pour le milieu aérien. Parmi ces dernières, Balletti et al. (2015), Diamanti et Vlachaki (2015) et Van Damme (2015) ont employé le logiciel propriétaire PhotoScan (Agisoft) qui s'appuie sur le critère géométrique. Toujours sur le même critère, le logiciel libre (*open source*) MicMac (Pierrot-Deseilligny, 2014) a été utilisé par Drap (2012). Enfin Drap (2012), Skarlatos et al. (2012) et O'Byrne et al. (2014) ont essayé le logiciel libre PMVS (Furukawa et Ponce, 2007) qui utilise une combinaison des critères géométrique et de voisinage spatial.

Nous noterons vis-à-vis du critère géométrique que Drap et al. (2015) soulignent l'importance de modifier le modèle utilisé par ce critère, comme pour la reprojection (Treibitz et al., 2012), afin de tenir compte du phénomène de réfraction à travers les hublots plans en environnement aquatique.

1.2. Approches de la littérature sur les problématiques spécifiques au milieu

La qualité de l'appariement dense réalisé — et donc de la reconstruction 3D résultante — dépend fortement de la qualité de l'appariement éparé, c'est-à-dire du nombre de points remarquables appariés correctement. Dans le milieu sous-marin, on relève deux problèmes majeurs : la faible quantité de points remarquables détectés et les erreurs d'appariement de ces derniers.

En ce qui concerne la quantité de points remarquables détectés, elle est généralement faible dans le milieu sous-marin et est directement reliée à la qualité des prises de vues. Dans la littérature, ce problème est essentiellement abordé par le biais de prétraitements appliqués sur les images (Singh et al., 2007; Beall et al., 2010; Jenkin et al., 2010; Johnson-Roberson et al., 2010) : rehaussement du contraste, correction du vignettage et de l'éclairage, correction des couleurs, débruitage, renforcement de la netteté, etc. L'idée est de se rapprocher le plus possible de la qualité des images aériennes pour maximiser le nombre de points appariés.

Une autre manière d'y arriver est de travailler directement sur l'acquisition : avec des capteurs haut de gamme et une bonne maîtrise de l'éclairage artificiel, il est possible d'obtenir des images de très bonne qualité (Rives et al., 1978). Mais ces méthodes sont compliquées à mettre en œuvre sur des plateformes automatisées.

Certains travaux proposent de combiner l'usage de plusieurs détecteurs pour augmenter le nombre de points caractéristiques en jouant sur la différence des critères de sélection (Singh et al., 2007; Méline et al., 2012).

Enfin, d'autres travaux équipent le terrain avec des marqueurs pour fournir plus de points fiables (Skarlatos et Rova, 2010; Burns et al., 2015; Diamanti et Vlachaki, 2015), mais du fait qu'il soit essentiellement réalisé par des plongeurs, ce procédé est coûteux en temps et limite les champs d'application.

En ce qui concerne la fiabilité de l'appariement éparé, les mouvements qui ont lieu au sein du milieu comptent

parmi les sources majeures d'erreurs de mise en correspondance. En effet, si les objets appariés se sont déplacés entre les acquisitions des deux points de vue stéréoscopiques, leur géométrie locale n'est plus conforme à la géométrie globale et il sera alors impossible de retrouver correctement leur position.

Les motifs répétitifs ou ressemblants (algues, galets) sont aussi source d'erreurs en provoquant de faux appariements (différenciation difficile). Or renforcer l'exigence du critère de similitude comporte le risque d'écarter de bons appariements (par définition, les deux points de vue ne peuvent être parfaitement identiques).

La contre-mesure la plus courante dans la littérature consiste alors à identifier et retirer les couples erronés. Dans les travaux cités précédemment, ce filtrage s'effectue principalement par le biais du critère géométrique. Il a lieu généralement en même temps que l'estimation de la géométrie elle-même : un grand nombre d'échantillons de points appariés sont étudiés et celui qui permet de rassembler le plus de couples autour de son modèle géométrique est considéré comme le plus fiable, les couples divergents sont alors rejetés.

Quant aux portions de l'image qui contiennent très peu de texture réelle, elles sont appariées sans distinction par le critère géométrique, ce qui dégrade la qualité de la reconstruction. En général, ces portions sont retirées manuellement des nuages 3D créés. Sedlazeck et al. (2009) ont proposé un algorithme de segmentation permettant de distinguer automatiquement les objets étudiés des zones sans intérêt et de rejeter tous les points remarquables de celles-ci.

1.3. Notre approche

Nous nous intéressons dans nos travaux à la reconstruction 3D systématique à la demande de petites zones d'études sous-marines et sub-aquatiques peu profondes (< 100 m). Dans ce cadre, l'agilité et la facilité de déploiement sont des critères importants. Nos vecteurs d'acquisition (robots sous-marins) sont donc petits, légers et à bas coût. Il en va de même pour leurs charges utiles stéréoscopiques. Dans ces conditions, nous obtenons donc des images de qualité très variable, soumises de surcroît aux variations de l'éclairage naturel.

Pour garantir la qualité des reconstructions 3D denses réalisées à partir de couples stéréoscopiques (composés d'images simultanées ou successives) quelles que soient les conditions d'acquisition, nous proposons une approche originale de l'étape d'appariement, inspirée des algorithmes de croissance de régions utilisés pour les couples aériens à base très large. Dans notre cas, nous utilisons le critère de voisinage spatial pour propager l'information d'appariement à partir des zones les plus fiables (c'est-à-dire les plus texturées) de nos images jusqu'aux frontières des zones les moins fiables. De cette manière, nous évitons les étapes de pré-traitement (amélioration de la qualité des images) et de post-traitement (élimination des zones sans information).

Notre algorithme, dont le diagramme est proposé à la figure 1, se découpe en trois grandes étapes :

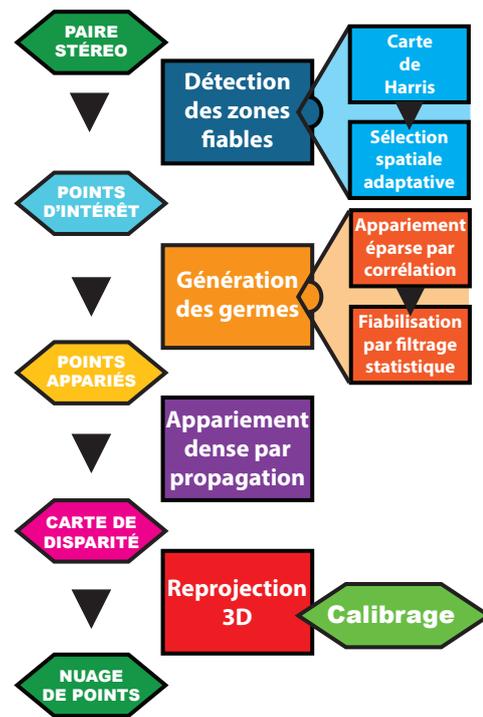


FIGURE 1: Diagramme représentant notre algorithme sous-marin de reconstruction 3D dense à partir d'une paire stéréoscopique [© L. Avanthey et al.].

1. La détection automatique des zones d'information fiable quelle que soit la qualité des images grâce à l'emploi et à une modification du détecteur de Harris.
2. La génération de germes d'appariement fiables grâce à un filtrage statistique sur le flot local.
3. La densification de l'appariement par propagation de l'information autour des germes avec une exclusion automatique des zones sans information.

Les trois parties suivantes décrivent chacune des étapes de notre algorithme. La dernière partie de l'article présente les résultats obtenus et discute de la qualité des contributions proposées. Le lecteur intéressé par les informations plus spécifiques au vecteur léger d'acquisition sous-marin, à sa charge utile et à sa mise en œuvre pourra consulter (Avanthey et al. (2013) ; Avanthey et al. (2016)).

2. Identification des zones fiables par un détecteur de Harris modifié

Notre première contribution porte sur l'identification automatique robuste des zones les plus texturées en s'adaptant à la qualité très variable des images sous-marines acquises par nos capteurs.

Le détecteur de Harris, dont la spécificité est de distinguer les fortes variations de texture, nous semble le plus approprié pour sélectionner les pixels candidats.

Dans son fonctionnement classique, la première étape du détecteur de Harris consiste à fournir une mesure d'intérêt pour chacun des pixels de l'image créant ainsi une carte de Harris. La seconde étape s'occupe

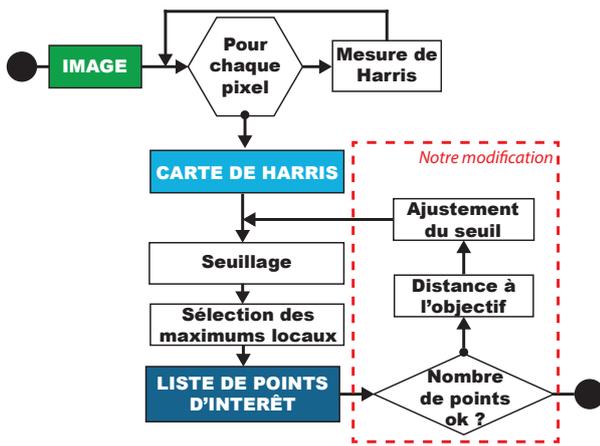


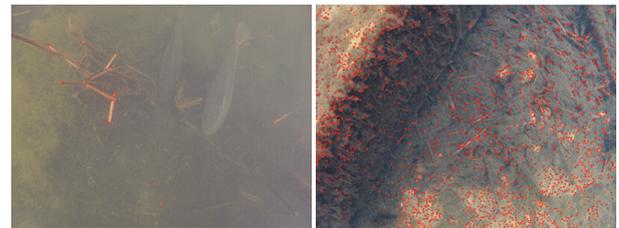
FIGURE 2: Diagramme de notre version modifiée de l'algorithme du détecteur de Harris [© L. Avanthey et al.].

de la sélection des points caractéristiques. Un seuil arbitraire, fixe et universel, permet tout d'abord de séparer les pixels d'intérêt (ceux qui présentent une forte variabilité de texture, soit une valeur élevée sur la carte de Harris) des autres pixels. Enfin, parmi eux, seuls les maximums locaux sont conservés pour former la liste finale de points de Harris.

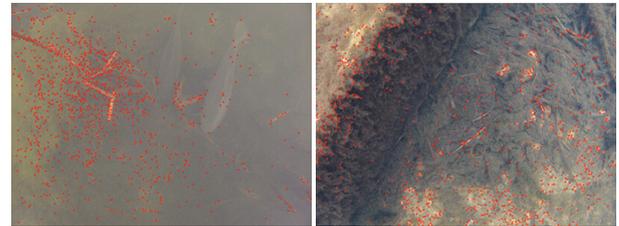
C'est cette seconde étape qui détient la clef du contrôle du nombre de points extraits. Un nombre insuffisant de points caractéristiques pose un problème sur la qualité de la reconstruction. Plus la qualité de l'image est dégradée, plus le seuil doit être tolérant. Mais en fixant par défaut un seuil très bas, on se retrouve cette fois-ci avec trop de points caractéristiques, ce qui augmente la complexité algorithmique de la suite de la chaîne. Nous proposons donc de rendre ce seuillage adaptatif pour contrôler de manière optimale la quantité de points caractéristiques extraits.

Lors de la première itération de l'algorithme, nous fixons un seuil initial d'après un pourcentage de la plus grande valeur de la carte de Harris (généralement à 75 %) afin de prendre en compte les mesures réelles. Un premier ensemble de points d'intérêt spatialement répartis est ensuite extrait. Si le nombre de points obtenus n'est pas inclus dans l'intervalle voulu, un nouveau seuil est calculé en conséquence. L'algorithme sélectionne un nouvel ensemble de points à partir de la carte de Harris. Les deux étapes de la phase d'extraction des points caractéristiques qui permettent de passer de la carte de Harris à la liste des points d'intérêt (seuillage et sélection des maximums locaux) sont alors répétées jusqu'à ce que le nombre final de points de Harris obtenus soit conforme à la demande.

Pour converger le plus rapidement possible vers l'objectif, nous ajustons la nouvelle valeur du seuil à partir de la distance entre le nombre de points obtenus et le milieu de l'intervalle cible : si nous sommes loin de l'objectif, une forte modification permettra de s'en rapprocher efficacement, sinon une petite modification évitera une oscillation autour de la cible. Ce type d'algorithme en boucle fermée est inspiré des correcteurs proportionnels issus



Seuillage arbitraire fixe (détecteur de Harris classique)



Seuillage adaptatif (notre contribution)

FIGURE 3: Détecteur de Harris et qualité variable des images : seuillage arbitraire de la version originale (seuil fixé à 189 \Rightarrow 20 points obtenus sur la première image et 2078 sur la seconde) et seuillage adaptatif de notre version modifiée (intervalle cible est fixé à [1000 – 1100] \Rightarrow 1023 et 1004 points obtenus pour des seuils finaux de 95 et 213) [© L. Avanthey et al.].

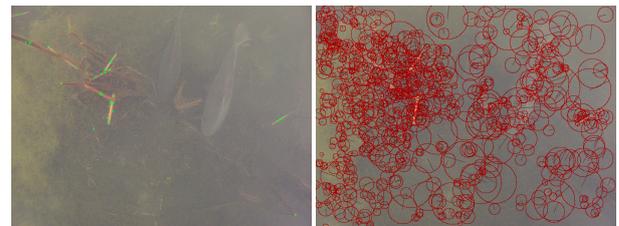


FIGURE 4: À titre de comparaison, les résultats du détecteur SIFT (à gauche, 60 points obtenus) et du détecteur SURF (à droite, 905 points obtenus) sur la première image de la figure précédente (3).

de la communauté du contrôle et de l'automatique (Minorsky, 1922).

Toujours pour minimiser le nombre d'itérations nécessaire à la convergence du seuil, nous exploitons le fait que les images prises à proximité (ce qui est le cas des images stéréoscopiques simultanées ou successives) auront une bonne probabilité d'avoir une qualité similaire. Nous utilisons donc le seuil précédemment calculé comme valeur initiale pour le traitement de l'autre image.

Enfin, pour exclure automatiquement les images qui ne présentent pas ou trop peu d'information (absorption quasi totale ne laissant visible que la masse d'eau ou absence presque totale de luminosité), nous fixons une limite à la capacité d'adaptation : si plus de 10 % des pixels de l'image sont sélectionnés sans qu'il n'y ait convergence sur l'intervalle cible, alors l'image doit être rejetée.

À la fin de cette étape, nous obtenons un échantillon de points représentatifs des zones les plus texturées pour chaque image des couples stéréoscopiques. Le diagramme de notre version modifiée de l'algorithme de Harris adaptée aux images sous-marines est donné

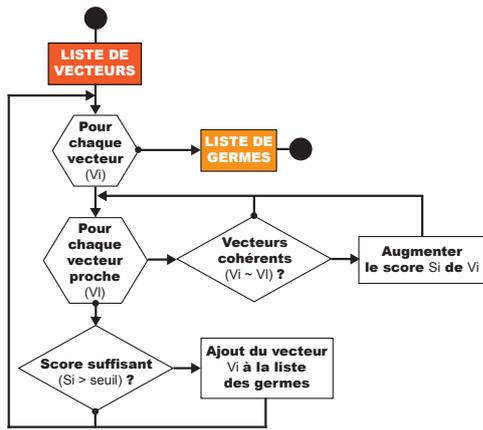


FIGURE 5: Diagramme de notre algorithme de filtrage statistique sur le flot local de vecteurs d'appariement [© L. Avanthey et al.].

à la figure 2. La figure 3 illustre les résultats obtenus sur deux images prises sur un même lieu à quelques minutes d'écart en comparaison à l'algorithme classique. À titre de comparaison, la figure 4 montre les résultats des détecteurs SIFT et SURF. Sur cet exemple, le nombre de points détectés par SIFT est insuffisant. SURF, quant à lui, trouve une bonne quantité de points, mais nous verrons dans la partie suivante que la majorité de ces derniers ne sont pas assez pertinents, pour former des appariements fiables.

3. Génération des germes fiables pour la propagation de l'appariement

La deuxième étape de notre algorithme porte sur la création d'un ensemble de germes fiables pour initialiser la propagation depuis les zones les plus texturées : il s'agit donc d'apparier nos points caractéristiques de Harris au sein d'un couple stéréoscopique.

Pour cet appariement éparé, le voisinage du point est utilisé comme descripteur et la similitude de deux voisinages est mesurée par corrélation. Ce choix est possible car les couples stéréoscopiques sur lesquels nous travaillons sont simultanés ou successifs : les conditions d'éclairage et d'orientation des deux images sont donc similaires.

Si le score de ressemblance entre deux voisinages est trop faible, l'appariement du couple potentiel est rejeté. Si la mesure de corrélation propose plusieurs candidats, seule la combinaison offrant le meilleur score est retenue, car un point ne peut être apparié qu'une seule et unique fois.

Notre mesure de corrélation calcule par défaut la somme des carrés des différences (SSD — *Sum of Squared Differences*) sur une fenêtre de 13×13 qui offre en moyenne le meilleur compromis. Dans le cas où le nombre de couples appariés est trop faible, l'algorithme change pour un critère plus stable, mais plus complexe : la corrélation croisée normalisée centrée (ZNCC — *Zero-mean Normalized Cross-Correlation*).

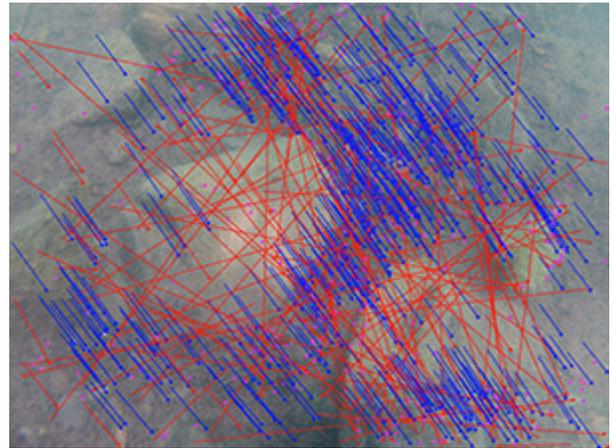


FIGURE 6: Sur les 969 points d'intérêt détectés sur chaque image du couple, 660 appariements initiaux ont été formés par corrélation (68 % des points), 143 sont rejetés par notre algorithme de filtrage (soit 22 % des appariements initiaux, marqués en rouge) et 517 germes sont déclarés fiables (marqués en bleu) [© L. Avanthey et al.].

Parmi les appariements réalisés à cette étape, certains sont erronés. Pour les éliminer, nous proposons un algorithme de filtrage basé sur un critère statistique appliqué au flot local des vecteurs d'appariement (figure 5). Rapide et efficace, cet algorithme nous évite de recourir au critère géométrique. L'étalonnage sous-marin des caméras n'étant pas exempt d'erreurs, nous préférons éviter de les cumuler et n'employer le modèle géométrique que pour la reprojection.

Un vecteur d'appariement représente spatialement la relation entre deux points appariés (disparité) en replaçant leurs coordonnées dans le même plan image. Notre algorithme de filtrage extrait pour chaque vecteur d'appariement v_i son flot local de vecteurs, c'est-à-dire le sous-ensemble de vecteurs situés à proximité de v_i .

Il calcule ensuite combien parmi eux sont cohérents avec v_i (norme et direction). S'ils sont peu nombreux, le vecteur d'appariement v_i , considéré alors comme erroné, est retiré de la liste des candidats. À l'inverse, si le vote est satisfaisant, v_i est ajouté à la liste des germes.

Cette manière de procéder élimine systématiquement les vecteurs isolés, même s'ils sont valides. Mais ce comportement est souhaitable, car les zones qu'ils représentent ne sont probablement pas suffisamment texturées pour permettre une propagation fiable.

La figure 6 nous montre un exemple du résultat de notre algorithme de formation de germes après filtrage de l'appariement éparé sur un couple d'images sub-aquatiques. Nous remarquons que plus de 75 % des vecteurs candidats sont valides. Tous les appariements erronés ont été correctement filtrés et quelques bons appariements isolés ont été exclus (moins de 2 % des appariements rejetés pour ce couple).

La figure 7 compare le résultat de notre algorithme de génération de germes sur le même couple avec celui obtenu par l'implémentation d'un algorithme fondé sur le détecteur-descripteur SIFT. Pour aider à la comparaison,

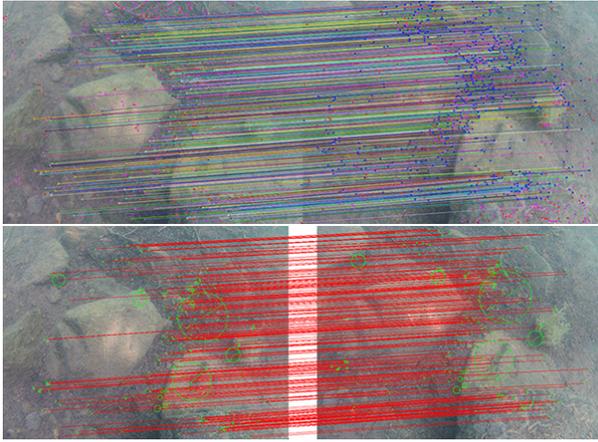


FIGURE 7: Résultat de l'étape de génération de germes par notre algorithme (en haut) et par une implémentation de l'algorithme SIFT (en bas) [© L. Avanthey et al.]. Notre algorithme apparie avec succès 517 points sur les 969 détectés (soit 53 % des points) alors que SIFT apparie seulement 260 points avec succès sur les 954 détectés (soit 27 % des points).

nous avons adapté l'intervalle cible de notre algorithme pour correspondre au nombre de points d'intérêt trouvés par SIFT. Sur cet exemple, notre algorithme apparie 53 % des points d'intérêt détectés contrairement à SIFT qui en apparie à peine 30 %, soit presque 50 % en moins.

Lorsque les scènes observées sont dynamiques (présence de la faune ou de la flore par exemple), nous synchronisons finement les acquisitions de manière à nous abstraire des mouvements. Tous les objets apparaissent alors statiques : nous pouvons les reconstruire sans inconsistances géométriques. Le lecteur intéressé pourra trouver plus de détails à ce propos dans (Avanthey et al. (2016)).

4. Densification de l'appariement par propagation autour des germes

Une fois les germes créés au cœur des zones les plus texturées, la phase suivante de notre algorithme consiste à propager l'information d'appariement dans leur voisinage afin de produire une carte dense de disparité.

Comme vu précédemment, un germe désigne un couple fiable de pixels appariés : G_g sur la vue gauche et G_d sur la vue droite. Chaque germe est étudié indépendamment les uns des autres et sert de point de départ local à la croissance de la région qu'il représente.

Appelons F_g la fenêtre carrée de 5×5 centrée sur G_g qui définit son voisinage et F_d la fenêtre correspondante centrée sur G_d . Chaque pixel P_{F_g} de F_g non déjà apparié est évalué individuellement pour chercher son potentiel correspondant dans une fenêtre 3×3 centrée sur le pixel équivalent P_{F_d} dans F_d .

Cela donne neuf candidats dont la similarité avec P_{F_g} est évaluée par une mesure de corrélation. Si le meilleur score obtenu est suffisamment élevé, le candidat correspondant est apparié avec P_{F_g} et l'information colorimétrique est sauvegardée.

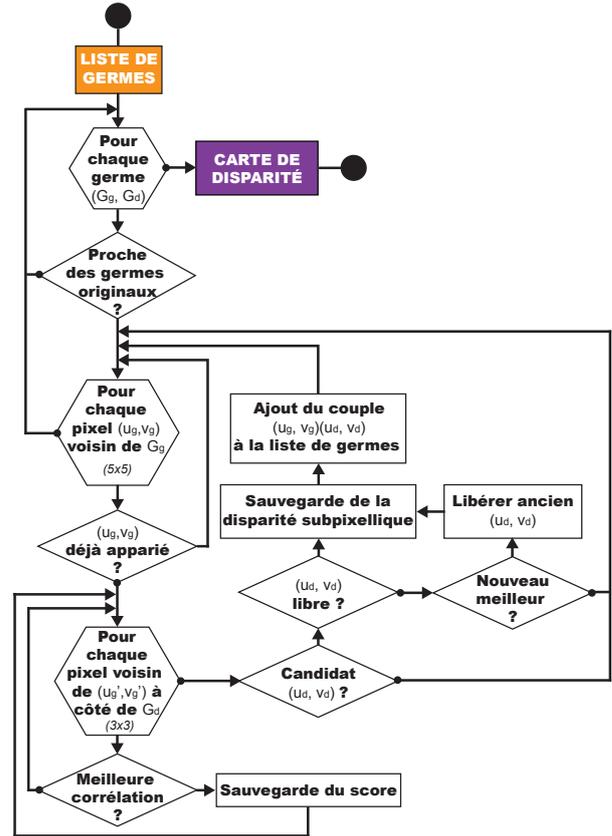


FIGURE 8: Diagramme de notre algorithme d'appariement dense par propagation [© L. Avanthey et al.].

Leur disparité est calculée puis affinée au niveau subpixelique pour améliorer la granularité du système. Cette interpolation est réalisée par la recherche de la position exacte de l'extremum du paraboloïde de corrélation et sauvegardée pour compléter la carte de disparité.

Un nouvel appariement peut s'avérer meilleur qu'un ancien déjà enregistré, car les fenêtres de recherche se chevauchent (les pixels étudiés dans le voisinage du germe sont côte à côte). Dans ce cas, le nouveau P_{F_g} est sauvegardé et l'ancien pixel P_{F_g} pourra être évalué à nouveau s'il se retrouve dans une nouvelle fenêtre d'étude de F_g .

Chaque pixel apparié dans le voisinage d'un germe se transforme à son tour en un nouveau germe pour propager l'appariement dans son propre voisinage. Tant qu'il y a des germes à étudier, les régions vont continuer à croître de cette manière jusqu'à la collision avec d'autres régions ou jusqu'à la limite d'expansion autorisée dans un rayon autour des germes originaux.

Les parties de l'image sans informations réelles sont de cette manière automatiquement exclues de la reconstruction. Une situation optimale est obtenue avec un nombre suffisant de germes pour permettre de fixer un petit rayon d'expansion et ne pas trop déborder des zones d'intérêt.

Le diagramme de notre algorithme de densification de l'appariement est donné à la figure 8. Le lecteur trouvera sur la figure 9 deux exemples présentant les cartes denses de disparité obtenues par propagation et démon-

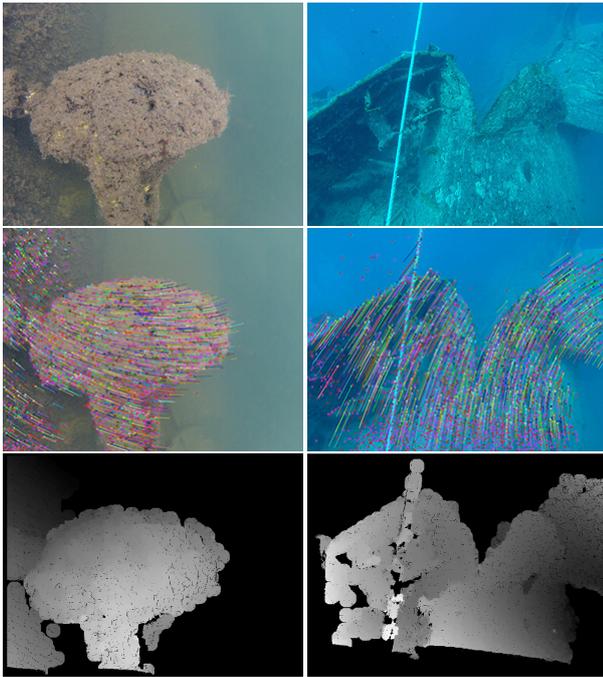


FIGURE 9: Densification de l'appariement par propagation et exclusion automatique des zones sans information réelle [© L. Avanthey et al.]. En haut : une image originale des couples stéréoscopiques. Au centre : génération des germes (flots de vecteurs) sur les zones les plus texturées. En bas : densification de l'appariement par propagation jusqu'à la limite des zones sans intérêt (cartes de disparité).

trant l'exclusion automatique des zones sans information réelle.

5. Résultats

Pour évaluer les différentes parties de notre algorithme et produire des statistiques générales, nous avons constitué une base de données à partir d'un échantillon représentatif d'images sous-marines. Elle contient une centaine de couples stéréoscopiques, composés d'images synchronisées, simultanées, ou successives et pris sur une dizaine de lieux différents (lacs, fleuves, Méditerranée, Mer Rouge).

Les conditions environnementales représentées par ces images sont très variées, que ce soit sur le type d'eau (salinité, turbidité, courant, etc.), les conditions météorologiques (nuages, soleil, vent, etc.) et les horaires des prises de vue (moment de la journée et saison). Aucun pré-traitement n'est appliqué à ces images.

Pour pouvoir effectuer des calculs statistiques systématiques sur l'ensemble de la base en des temps de calcul raisonnables, les images ont été sous-échantillonnées. Les GSD résultants vont de 0,25 mm à 10 mm, ce qui permet de distinguer des objets de quelques millimètres dans le meilleur des cas à une dizaine de centimètres sur les images les moins résolues.

Concernant la robustesse de notre algorithme à la qualité des images acquises, les tests conduits sur les 200 images unitaires de notre base confirment que la

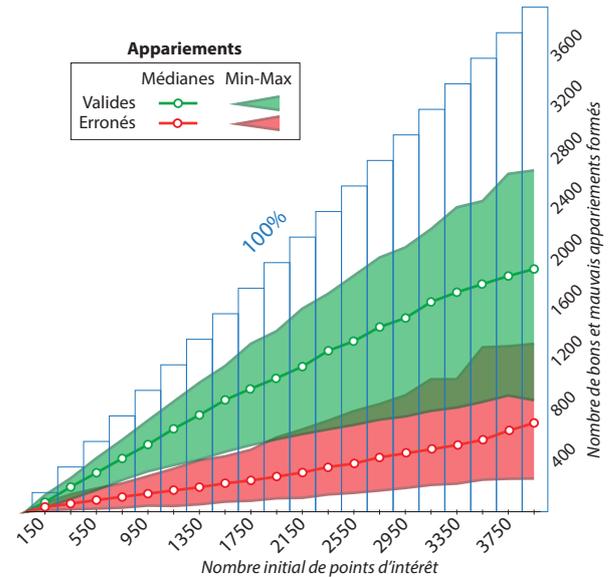


FIGURE 10: Nombre et qualité des appariements obtenus sur les couples de notre base de données [© L. Avanthey et al.].

modification de l'algorithme de Harris est nécessaire. En effet, sur des images acquises dans des conditions similaires, le seuil final sur lequel converge notre algorithme pour atteindre un nombre de points contenu dans l'intervalle cible (quelqu'il soit) présente une variation non négligeable de l'ordre de 20 %.

Les images de notre base acquises dans des conditions lumineuses très faibles (prises de vue durant la période autour du coucher du soleil) nous ont permis de tester les limites de notre algorithme. Ces images présentent un histogramme très étroit : toute leur dynamique est comprise sur 15 % seulement de la plage [0 – 255]. Sur ces dernières, le seuil adaptatif converge sur une valeur en moyenne 80 % plus faible que celles obtenues avec les cas généraux et sont toujours fiables pour permettre la reconstruction.

Notre version adaptative du détecteur de Harris présente une complexité linéaire, il appartient donc à la même classe algorithmique que la version originale. La différence en temps de calcul (nombre de cycles processeurs) entre la version originale et notre version est inférieure à 5 %. Sur les images de notre base de données, la convergence s'obtient en moins de cinq itérations.

Pour chaque couple d'images de notre base, nous avons calculé le nombre de germes générés par notre algorithme en fonction du nombre initial de points d'intérêt voulu ainsi que le nombre de vecteurs candidats rejetés (appariements erronés).

Les statistiques ont été produites pour vingt intervalles cibles de points d'intérêt de même largeur entre [100 – 200] et [3900 – 4000]. Ces résultats sont présentés à la figure 10 et au tableau 1.

Nous pouvons remarquer que le ratio bons appariements sur mauvais appariements est stable sur l'ensemble des intervalles (variation maximale de $\pm 4\%$) même lorsque le nombre de points d'intérêt initial est très bas

SURF			
Intervalle	Médiane	Min	Max
593 (médiane) [107 - 2256]	46 % (279)	24 % (37)	83 % (1200)

SIFT			
Intervalle	Médiane	Min	Max
950 (médiane) [231 - 1521]	35 % (307)	6 % (54)	72 % (610)

Nombre de germes

Notre algorithme			
Intervalle	Médiane	Min	Max
Tous	53 %	28 %	73 %
[500 - 600]	55 % (305)	31 % (170)	76 % (419)
[900 - 1000]	55 % (526)	33 % (317)	75 % (714)

TABLE 1: Pourcentage de génération de germes par notre algorithme, par SURF et SIFT sur notre base de données. Les intervalles cibles donnés à notre algorithme ont été choisis pour être cohérents avec les nombres médians de points d'intérêt détectés par SIFT et par SURF afin de permettre la comparaison.

ou très élevé.

De manière générale (médiane), notre algorithme permet d'extraire 53 % de germes à partir de nos points de Harris (73 % dans les meilleurs cas, 28 % dans les pires).

De plus, seulement 23 % des vecteurs candidats sont rejetés (7 % dans les meilleurs cas, 46 % pour les moins bons) et aucun faux positif n'a été trouvé. Ces résultats nous permettent de considérer que les deux premières étapes de notre algorithme sont robustes à la qualité variable des images, quel que soit le nombre de points d'intérêt demandé.

À titre de comparaison, nous avons testé sur notre base de données les algorithmes d'appariements épars basés sur les détecteurs-descripteurs SIFT et SURF implémentés dans la bibliothèque OpenCV. Les résultats obtenus sont présentés dans la table 1.

De manière générale, SURF présente une meilleure qualité d'appariement que SIFT sur ces images (46 % de germes extraits contre 35 %). Mais en revanche, il est beaucoup plus irrégulier pour la détection des points d'intérêt et donc sur le nombre de germes qu'il fournit.

De plus, son taux d'appariement médian baisse en dessous de 40 % lorsque le nombre initial de points d'intérêt augmente sensiblement (> 1000 points) et son filtrage laisse passer quelques mauvais appariements (environ 1 % des points détectés).

En comparaison, le détecteur SIFT présente un comportement plus stable et produit beaucoup moins de faux positifs (environ 0,2 % des points détectés). Cependant, le ratio de bons appariements sur le nombre initial de points d'intérêt est faible et diminue d'autant que le nombre initial de points d'intérêt augmente.

La capacité de notre algorithme à fournir un nombre de points d'intérêt conforme à une consigne le distingue des autres algorithmes d'appariement épars. Afin de pou-

GSD (mm)	Vérité terrain (mm)	Mesure 3D (mm)	Erreur (mm)
10	875	862,7	-12,3
10	298	307,2	9,2
10	49	42,3	-6,7
8	1273	1248,6	-24,4
3	732	734,2	2,2
3	218	216,7	-1,3
0.8	433	434,1	1,1
0.8	427	427,8	0,8
0.8	350	349,1	-0,9
0.6	220	219,5	-0,5
0.6	170	170,8	0,8
0.6	113	112,9	-0,1

TABLE 2: Échantillon de mesures réalisées *in situ* sur des rochers avec un mètre ruban accompagnées de leurs mesures correspondantes sur les nuages de points extraits à une taille d'échantillon au sol donnée.

voir comparer les résultats, nous avons utilisé ceux produits par notre algorithme sur les intervalles de points contenant les nombres médians de points d'intérêt détectés par SIFT et par SURF.

Dans les deux cas, nous obtenons avec notre algorithme un pourcentage de germes extraits plus élevé : 55 % contre 46 % pour SURF ([900 - 1000]) et 55 % contre 35 % pour SIFT ([500 - 600]). Nous pouvons donc conclure qu'en plus de présenter une très bonne stabilité, notre algorithme fournit en moyenne de meilleurs résultats dans le contexte sous-marin.

Concernant la densité des nuages de points créés, nos résultats montrent que notre algorithme permet l'appariement dense de 60 % à 85 % des pixels des images d'une paire stéréoscopique présentant un recouvrement entre 70 % et 90 %. La densité mesurée sur les nuages de points est consistante avec la taille de l'échantillon au sol des acquisitions.

Concernant la justesse des nuages de points créés, qui est liée à l'échelle d'acquisition, nous l'évaluons vis-à-vis de la taille d'échantillon au sol. Les mesures réalisées *in situ* sur des objets caractéristiques de l'environnement comme des pierres ou des rochers sont utilisées pour vérifier la justesse obtenue au sein d'un nuage.

La comparaison de ces mesures physiques réalisées *in situ* avec celles estimées sur les nuages de points produits (voir un échantillon sur la table 2) permet d'estimer une erreur inférieure à deux fois la taille de l'échantillon au sol ($j < \pm 2 \times \text{GSD}$).

Nous noterons que les mesures terrain sont difficiles à réaliser avec une grande exactitude sous l'eau (Holt, 2003; Bowens, 2011). En effet, la résolution des outils de mesure utilisés (mètre ruban) est d'environ ± 1 mm.

De surcroît, plus la distance mesurée est grande, plus sa justesse sera faible à cause des problèmes liés à la tension du ruban ou à la difficulté à rester dans le plan. De manière générale, les mesures ainsi effectuées présentent une erreur de quelques centimètres.

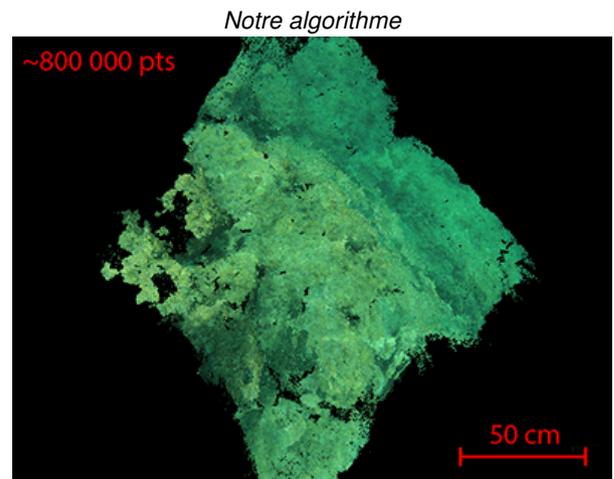
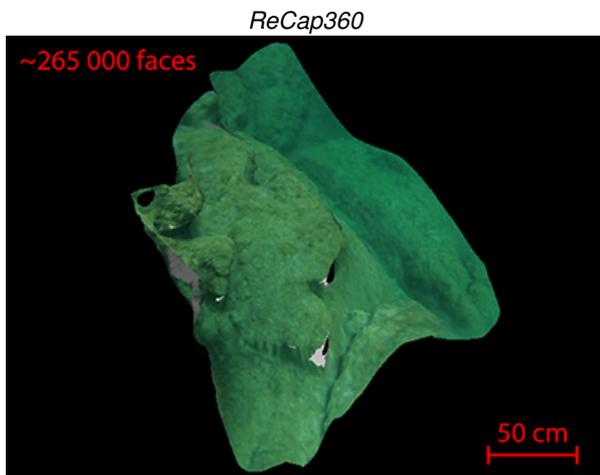
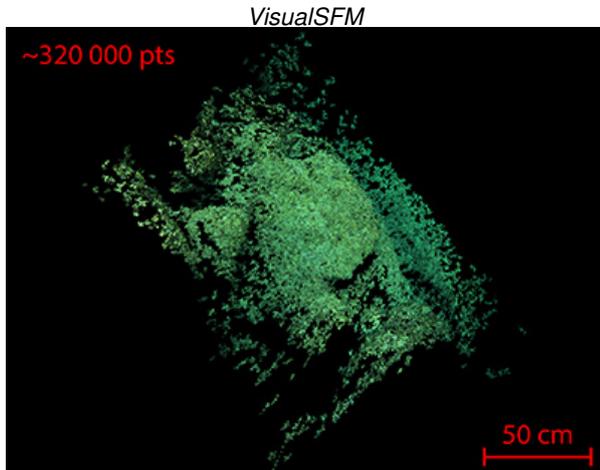
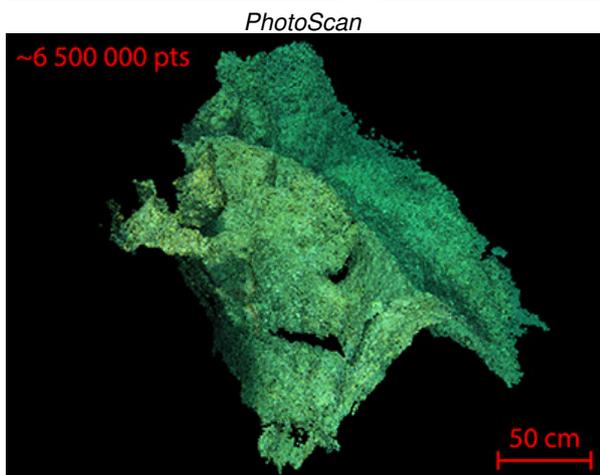
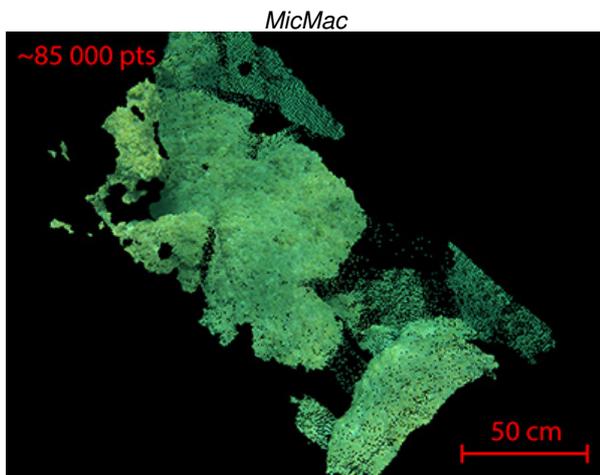


FIGURE 11: Résultats obtenus sur un jeu de 18 images séquentielles par notre algorithme [© L. Avanthey et al.] et avec les logiciels MicMac (Pierrot-Desseignly, 2014), PhotoScan, VisualSFM et ReCap360.

Pour terminer, nous avons comparé le nuage 3D dense produit par notre algorithme avec les résultats des logiciels accessibles de la littérature : MicMac (Pierrot-Desseignly, 2014) (*open source*), PhotoScan, VisualSFM (Wu, 2013) — qui implémente CMVS (Furukawa et Ponce, 2007) — et ReCap360.

Ces logiciels ne permettent pas de faire une reconstruction à partir d'une paire stéréoscopique seule, nous sommes donc partis d'une séquence de 18 images sous-marines acquises en Méditerranée.

MicMac et PhotoScan utilisent des algorithmes de reconstruction multi-images, Visual SFM est plutôt basé sur une reconstruction éparse par tracking et ReCap360 produit un modèle triangulé et texturé.

Quant aux nuages produits par notre algorithme, ils ne sont pas filtrés après la reprojection pour retirer les points aberrants, comme c'est souvent le cas des produits issus des logiciels cités. La figure 11 montre les différents résultats obtenus.

L'analyse qualitative de ces résultats, présentée dans le tableau 3, montre que notre nuage global présente une géométrie correcte malgré quelques erreurs de reprojection visibles à certains endroits (bavures sur le nuage). Notre nuage offre également une bonne densité de points sans que leur quantité ne soit excessive. La complétude est convenable (les zones d'occlusions résiduelles sont peu nombreuses) et la texture de la scène a été bien préservée.

6. Conclusion

Dans cet article, nous avons vu quelles étaient les difficultés spécifiques au monde sous-marin pour reconstruire l'information 3D à partir d'images dans le visible et plus particulièrement au niveau de la phase d'appariement dense.

Nous avons présenté les principales solutions utilisées dans la littérature. Puis nous avons proposé nos

	Nous	MicMac	PhotoScan	VisualSFM	ReCap360
Densité	***	***	***	*	**
Texture	***	***	***	**	*
Complétude	***	**	***	*	***
Redondance	**	*	***	**	*
Bruit	**	*	**	**	**

TABLE 3: Appréciations des résultats présentés à la figure 11.

contributions originales : un algorithme de propagation de l'information d'appariement depuis les zones les plus fiables jusqu'aux frontières des zones les moins fiables, robuste à la qualité variable des images acquises.

Nous avons détaillés les trois grandes étapes qui composent notre algorithme : détection automatique et robuste (pas de pré-traitement des images) des zones d'information fiable en utilisant et modifiant un détecteur de Harris, génération de germes d'appariement fiabilisée par un filtrage statistique sur le flot local, densification de l'appariement par propagation de l'information autour des germes avec exclusion automatique des zones sans information.

La qualité de ces apports a été éprouvée sur une base de données diversifiée issue de plusieurs campagnes terrain et les résultats obtenus montrent un très bon comportement de notre algorithme y compris quand on le compare avec ceux de la littérature.

Références

- Avanthey, L., Beaudoin, L., Gademer, A., Roux, M., 2016. Tools to Perform Local Dense 3D Reconstruction of Shallow Water Seabed. *Sensors* 16 (5), 712.
- Avanthey, L., Gademer, A., Beaudoin, L., Roux, M., 2013. First Steps for Operational Dense and High-Resolution Mapping of Shallow Water Using Dedicated Robots. Dans : *Ocean and Coastal Observation : Sensors and observing systems, numerical models and information Systems (OCOSS)*. Nice, France.
- Balletti, C., Beltrame, C., Costa, E., Guerra, F., Vernier, P., 2015. Underwater Photogrammetry and 3D Reconstruction of Marble Cargos Shipwreck. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS) XL-5/W5*.
- Bay, H., Tuytelaars, T., Van Gool, L., 2006. SURF : Speeded Up Robust Features. Dans : *European Conference on Computer Vision (ECCV)*. Graz, Autriche, pp. 404–417.
- Beall, C., Lawrence, B. J., Ila, V., Dellaert, F., 2010. 3D Reconstruction of Underwater Structures. Dans : *IEEE International Conference on Intelligent Robots and Systems (IROS)*. pp. 4418–4423.
- Bowens, A. (Ed.), 2011. *Underwater archaeology : the NAS guide to principles and practice*. John Wiley & Sons.
- Brandou, V., Allais, A.-G., Perrier, M., Malis, E., Rives, P., Sarrazin, J., Sarradin, P.-M., 2007. 3D Reconstruction of Natural Underwater Scenes Using the Stereovision System IRIS. Dans : *IEEE OCEANS Conference*. Aberdeen, Écosse, pp. 674–679.
- Burns, J. H. R., Delparte, D., Gates, R. D., Takabayashi, M., 2015. Integrating Structure-From-Motion Photogrammetry with Geospatial Software as a Novel Technique for Quantifying 3D Ecological Characteristics of Coral Reefs. *PeerJ*.
- Diamanti, E., Vlachaki, F., 2015. 3D Recording of Underwater Antiquities in the South Euboean Gulf. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1.
- Drap, P., 2012. Underwater Photogrammetry for Archaeology. Dans : *Special Applications of Photogrammetry*. InTech, pp. 111–136.
- Drap, P., Merad, D., Hijazi, B., Gaoua, L., Nawaf, M. M., Saccone, M., Chemisky, B., Seinturier, J., Sourisseau, J.-C., Gambin, T., Castro, F., 2015. Underwater Photogrammetry and Object Modeling : A Case Study of Xlendi Wreck in Malta. *Sensors* 15 (12), 30351–30384.
- Furukawa, Y., Ponce, J., 2007. Accurate, Dense, and Robust Multi-View Stereopsis. Dans : *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Minneapolis, États-Unis, pp. 1–8.
- Harris, C., Stephens, M., 1988. A Combined Corner and Edge Detector. Dans : *Alvey Vision Conference*. Vol. 15. Manchester, Royaume-Uni, pp. 147–152.
- Holt, P., 2003. An Assessment of Quality in Underwater Archaeological Surveys Using Tape Measurements. *International Journal of Nautical Archaeology (IJNA)* 32 (2), 246–251.
- Jenkin, M., Verzijlenberg, B., Hogue, A., 2010. Progress Towards Underwater 3D Scene Recovery. Dans : *ACM C* Conference on Computer Science and Software Engineering (C3S2E)*. Montréal, Canada, pp. 123–128.
- Johnson-Roberson, M., Pizarro, O., Williams, S. B., Mahon, I., 2010. Generation and Visualization of Large-Scale Three-Dimensional Reconstructions from Underwater Robotic Surveys. *Journal of Field Robotics (JFR)* 27 (1), 21–51.
- Kunz, C., Singh, H., 2010. Stereo Self-Calibration for Seafloor Mapping Using AUVs. Dans : *IEEE Autonomous Underwater Vehicles (AUV)*. pp. 1–7.
- Lowe, D. G., 1999. Object Recognition from Local Scale-Invariant Features. Dans : *IEEE International Conference on Computer Vision (ICCV)*. Vol. 2. Corfou, Grèce, pp. 1150–1157.
- Méline, A., Triboulet, J., Jouvencel, B., 2012. Comparative Study of Two 3D Reconstruction Methods for Underwater Archaeology. Dans : *IEEE International Conference on Intelligent Robots and Systems (IROS)*.
- Minorsky, N., 1922. Directional Stability of Automatically Steered Bodies. *Journal of the American Society for Naval Engineers (ASNE)* 42 (2), 280–309.
- O'Byrne, M., Pakrashi, V., Schoefs, F., Ghosh, B., 2014. A Comparison of Image Based 3D Recovery Methods for Underwater Inspections. Dans : *European Workshop on Structural Health Monitoring (EWSHM)*.
- Pierrot-Deseilligny, M., 2014. *Apero, Pastis and Other Beverages in a Nutshell*. Manuel, IGN, Saint-Mandé, France.
- Pizarro, O., Eustice, R., Singh, H., 2004. Large Area 3D Reconstructions from Underwater Surveys. Dans : *IEEE OCEANS Conference*. Vol. 2. pp. 678–687.
- Prabhakar, C. J., Kumar, P. U. P., 2012. 3D Surface Reconstruction of Underwater Objects. Dans : *National Conference on Advanced Computing and Communications (NCACC)*.
- Rives, C., Chenz, Pétron, C., 1978. *La prise de vue sous-marine*. Fillipacchi.
- Schmidt, V. E., Rzhanov, Y., 2012. Measurement of Micro-Bathymetry with a GOPRO Underwater Stereo Camera Pair. Dans : *IEEE OCEANS Conference*. Virginia Beach, États-Unis, pp. 1–6.
- Sedlazeck, A., Köser, K., Koch, R., 2009. 3D Reconstruction Based on Underwater Video from ROV Kiel 6000 Considering Underwater Imaging Conditions. Dans : *IEEE OCEANS Conference*. Brême, Allemagne, pp. 1–10.

- Shi, J., Tomasi, C., 1994. Good Features to Track. Dans : IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, États-Unis, pp. 593–600.
- Singh, H., Roman, C., Pizarro, O., Eustice, R., Can, A., 2007. Towards High-Resolution Imaging from Underwater Vehicles. *The International Journal of Robotics Research (IJRR)* 26 (1), 55–74.
- Skarlatos, D., Demestiha, S., Kiparissi, S., 2012. An 'open' method for 3D modelling and mapping in underwater archaeological sites. *International Journal of Heritage in the Digital Era* 1 (1), 1–24.
- Skarlatos, D., Rova, M., 2010. Photogrammetric Approaches for the Archaeological Mapping of the Mazotos Shipwreck. Dans : *International Conference on Science and Technology In Archaeology and Conservation (STIAC)*.
- Treibitz, T., Schechner, Y. Y., Kunz, C., Singh, H., 2012. Flat Refractive Geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 34 (1), 51–65.
- Van Damme, T., 2015. Computer Vision Photogrammetry for Underwater Archaeological Site Recording in a Low-Visibility Environment. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS)* 1 (-), 231–238.
- Wu, C., 2013. Towards Linear-Time Incremental Structure from Motion. Dans : *IEEE International Conference on 3D Vision (3DV)*. Seattle, États-Unis, pp. 127–134.