



HAL
open science

Reconstruction 3D d'un polyèdre convexe à partir de ses silhouettes

Baptiste Brument, Lilian Calvet, Jean Mélou, Jean-Denis Durou

► **To cite this version:**

Baptiste Brument, Lilian Calvet, Jean Mélou, Jean-Denis Durou. Reconstruction 3D d'un polyèdre convexe à partir de ses silhouettes. 18èmes journées francophones des jeunes chercheurs en vision par ordinateur (ORASIS 2021), Centre National de la Recherche Scientifique [CNRS]; Equipe REVA, IRIT : Institut de Recherche en Informatique de Toulouse, Sep 2021, Saint Ferréol, France. hal-03339659

HAL Id: hal-03339659

<https://hal.science/hal-03339659v1>

Submitted on 9 Sep 2021

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.

Reconstruction 3D d'un polyèdre convexe à partir de ses silhouettes

Baptiste BRUMENT¹

Lilian CALVET¹

Jean MÉLOU¹

Jean-Denis DUROU¹

¹ IRIT, UMR CNRS 5505, Université de Toulouse

baptiste.brument@etu.toulouse-inp.fr

Résumé

La reconstruction 3D d'un objet plongé dans un milieu réfringent tel que le verre nécessite de connaître non seulement la géométrie de la surface de séparation entre le milieu réfringent et l'air, appelée dioptre, mais également les paramètres internes et les poses de la caméra. Dans ce travail, nous nous plaçons dans un environnement suffisamment contrôlé pour pouvoir acquérir ces connaissances. Le dioptre, assimilé à un polyèdre convexe, est placé sur une table tournante. Les images sont acquises par une caméra statique de paramètres internes connus. La géométrie du dioptre est estimée à partir de ses silhouettes.

Mots Clef

Reconstruction 3D, polyèdre convexe, silhouettes.

Abstract

The 3D-reconstruction of an object immersed in a refractive medium such as glass requires knowledge not only of the geometry of the interface between the refractive medium and the surrounding air, but also of the internal parameters and the camera poses. In this work, we place ourselves in a sufficiently controlled environment to acquire this knowledge. The interface, assimilated to a convex polyhedron, is placed on a turning table. The images are acquired by a fixed camera with known intrinsic parameters. The geometry of the interface is estimated from its silhouettes.

Keywords

3D-reconstruction, convex polyhedron, silhouettes.

1 Introduction

La reconstruction 3D d'insectes préhistoriques présente un grand intérêt pour les entomologistes. En effet, certaines espèces ne nous sont parvenues qu'au travers de spécimens conservés dans l'ambre, une matière dans laquelle ils ont été piégés dans des ères géologiques anciennes. Les spécimens conservés dans l'ambre baltique (cf. figure 1) ont entre 40 et 50 millions d'années et sont proches d'espèces actuelles. Des spécimens plus anciens piégés dans l'ambre birmane, qui ont plus de 90 millions d'années et sont des témoins d'espèces totalement disparues, constituent une source d'information inestimable aux yeux des spécialistes

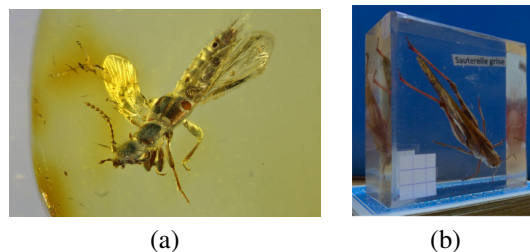


FIGURE 1 – (a) Insecte préhistorique piégé dans l'ambre (≈ 40 millions d'années). Image aimablement fournie par le Muséum d'Histoire Naturelle de Copenhague (Danemark). (b) Insecte plongé dans un bloc de verre. La forme du dioptre est un polyèdre convexe.

de l'évolution. La modélisation 3D précise du spécimen prisonnier de l'ambre permettrait de faciliter son étude, mais il est impossible d'extraire l'insecte de sa gangue sous peine de le faire tomber en poussière. L'utilisation d'un scanner médical est envisageable pour obtenir un modèle 3D de l'insecte, mais cela est long et coûteux. Il semble plus raisonnable d'opter pour le scannage par photogrammétrie, dont la mise en œuvre ne requiert qu'un appareil photographique et des ressources numériques.

Cependant, l'utilisation directe des techniques classiques du *shape-from-shading* [5], du *structure-from-motion* [9, 20], du SLAM [17] ou de la stéréoscopie multi-vues [7, 11] est impossible à cause de la réfraction. Certaines de ces techniques ont été adaptées à la prise en compte de ce phénomène [1, 2, 3, 4, 15], en particulier la stéréoscopie multi-vues [1, 3, 15], qui permet d'obtenir une reconstruction 3D dense. La mise en œuvre de ces approches nécessite cependant un certain nombre de connaissances a priori. D'une part, la stéréoscopie multi-vues requiert la connaissance des poses de la caméra. D'autre part, si le milieu réfringent dévie la trajectoire des rayons lumineux qui le traversent, la forme du dioptre, c'est-à-dire de la surface de séparation entre ce milieu et l'air, influe fortement sur l'apparence de l'objet à reconstruire. Fort heureusement, il n'est pas interdit de tailler le bloc d'ambre pour lui donner une forme particulière, tant que cela ne risque pas d'endommager l'insecte. Nous supposons donc que des coupes planes ont conféré au dioptre la forme d'un polyèdre convexe.

Nous nous plaçons dans le cadre d’un environnement d’acquisition semi-contrôlé : le dioptré est positionné sur une table tournante et photographié par une caméra statique de paramètres internes connus. Dans ce travail, nous nous attachons à l’estimation des sommets du polyèdre et des poses de la caméra à partir d’une collection de vues, en nous inspirant de la technique du *shape-from-silhouettes* [10] (cf. figure 2). Les différentes étapes de la solution proposée sont les suivantes :

1. extraction de la silhouette du dioptré dans chacune des vues ;
2. détection des arêtes du polygone que constitue cette silhouette ;
3. extraction des sommets de ce polygone, qui sont des images des sommets du polyèdre ;
4. mise en correspondance robuste des sommets des différents polygones ;
5. reconstruction 3D initiale des sommets du polyèdre et des poses de la caméra ;
6. ajustement de faisceaux.

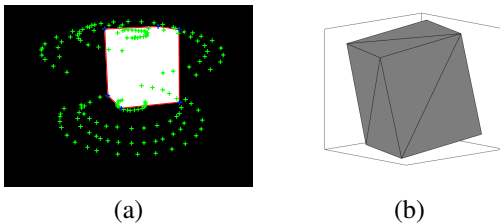


FIGURE 2 – Grâce au suivi des sommets du polyèdre (a), notre algorithme peut reconstruire ce dernier en 3D (b).

Dans le paragraphe 2, nous passons en revue les principaux travaux de vision 3D liés aux différentes étapes de notre approche. Après avoir introduit les notations utiles dans le paragraphe 3, nous présentons dans le paragraphe 4 notre algorithme de mise en correspondance robuste des images des sommets du polyèdre. La reconstruction 3D des sommets du polyèdre et des poses de la caméra est décrite dans le paragraphe 5. Enfin, une évaluation de l’ensemble de la chaîne de traitement proposée, à la fois quantitative (sur un jeu de données de synthèse) et qualitative (sur des données réelles), est fournie dans le paragraphe 6.

2 Travaux antérieurs

Reconstruction 3D d’un objet disposé sur une table tournante – Dans [6], il est montré que l’estimation de la matrice fondamentale nécessite de connaître sept paires de points dans le cas général, et seulement cinq paires dans le cas d’une caméra calibrée. La reconstruction 3D d’un polyèdre convexe à partir de ses sommets semble donc a priori impossible pour un tétraèdre, qui comporte un nombre minimal de sommets. Nous nous inspirons de la méthode proposée dans [12, 13], qui montre comment retrouver la géométrie multi-vues à partir de la trajectoire de deux points,

lorsque la scène est placée sur une table tournante. Nous adoptons cette approche et proposons de reconstruire l’ensemble des sommets du polyèdre à partir de correspondances inter-vues.

Par ailleurs, il est montré dans [22] comment retrouver la géométrie multi-vues à partir de tangentes épipolaires qui sont mieux adaptées à la reconstruction 3D d’objets lisses.

Estimation des poses de la caméra en présence d’un objet réfringent

– L’estimation des poses en présence d’un objet réfringent a été étudiée, en particulier, dans les travaux de Chari et Sturm, qui ont montré dans [4] comment étendre la géométrie épipolaire au cas où la caméra est séparée de la scène 3D par un dioptré plan. La prise en compte explicite de la réfraction dans l’estimation de la pose par *structure-from-motion* (SfM) a également été étudiée récemment par Chadebecq et al. [2]. Ces contributions sont validées sur des images sous-marines.

L’une des principales difficultés de ce problème vient de l’appariement des points, les méthodes de suivi ou de mise en correspondance comme SIFT [16] ou ORB [19] n’étant pas utilisables en l’état. En effet, les hypothèses d’invariance photométrique sur lesquelles reposent les calculs des descripteurs ne s’étendent pas à l’utilisation d’objets semi-transparents, notamment à cause des reflets.

Mise en correspondance des images des sommets

– Toutefois, l’hypothèse d’un mouvement circulaire du polyèdre, relativement à la caméra, nous permet de résoudre le problème de la mise en correspondance à partir d’un raisonnement purement géométrique. Les trajectoires des sommets du polyèdre sont des cercles 3D parallèles dont les centres sont situés sur l’axe de rotation de la table. Les trajectoires des images des sommets constituent donc des ellipses. Ces propriétés sont exploitées pour concevoir une méthode de mise en correspondance robuste des images des sommets, qui se ramène à la résolution d’un problème de partitionnement d’un ensemble de points, dont chaque partie correspond aux images d’un sommet du polyèdre.

Shape-from-silhouettes – Cette méthode classique de reconstruction 3D multi-vues permet de retrouver une enveloppe convexe d’un objet à partir des silhouettes de ses projections dans les différentes vues. Sa mise en œuvre dans le cas d’un objet disposé sur une table tournante a été proposée dans [10]. Cependant, cette enveloppe convexe peut être assez éloignée de l’objet à reconstruire, dans le cas où l’objet est lui-même un polyèdre convexe.

Contributions – Notre première contribution est un paradigme de reconstruction 3D semi-automatique d’un polyèdre convexe, qui comprend les détails nécessaires à sa mise en œuvre : détection et mise en correspondance des images des sommets du polyèdre dans la collection de vues (cf. paragraphe 4); reconstruction 3D des sommets du polyèdre et des poses de la caméra à partir des correspondances inter-vues (cf. paragraphe 5). Notre deuxième contribution est un algorithme de mise en correspondance des images des sommets.

3 Notations

Nous considérons une scène constituée d'un polyèdre convexe comportant N sommets, disposé sur une table tournante, dont une caméra statique de paramètres internes connus prend J photographies. Dans la $j^{\text{ème}}$ image, $j \in \{1, \dots, J\}$, l'image \mathbf{x}_n^j du sommet \mathbf{X}_n du polyèdre, $n \in \{1, \dots, N\}$, a pour expression :

$$\mathbf{x}_n^j \sim P^j \mathbf{X}_n \quad (1)$$

où la matrice de projection perspective P^j liée à la $j^{\text{ème}}$ vue s'écrit :

$$P^j = K [R_{\text{réf}} R_{\theta^j} | \mathbf{t}_{\text{réf}}] \quad (2)$$

Dans cette expression, K désigne la matrice de calibrage, et $R_{\text{réf}}$ et $\mathbf{t}_{\text{réf}}$ désignent respectivement l'orientation du repère de la table tournante et la position de son origine, exprimés dans le repère de la caméra, de telle sorte que les deux premières colonnes de $R_{\text{réf}}$ définissent le plan de support de la table, et que sa troisième colonne soit un vecteur normal à la table dirigé vers le haut. Enfin, la matrice R_{θ^j} décrit la rotation de la table d'un angle θ^j autour de son axe :

$$R_{\theta^j} = \begin{bmatrix} \cos \theta^j & -\sin \theta^j & 0 \\ \sin \theta^j & \cos \theta^j & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

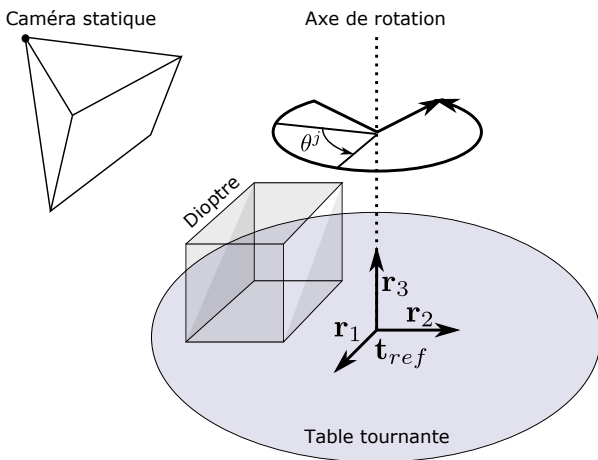


FIGURE 3 – Dispositif d'acquisition des vues du polyèdre. Le polyèdre est disposé sur une table tournante. Les vues sont acquises par une caméra statique. L'origine $\mathbf{t}_{\text{réf}}$ est l'intersection du plan de la table avec son axe de rotation. Les vecteurs $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ sont les colonnes de la matrice de rotation $R_{\text{réf}}$ qui décrit l'orientation de la table tournante, relativement au repère de la caméra.

4 Mise en correspondance

La première étape de notre approche consiste à extraire de chaque vue la silhouette du polyèdre, qui est représentée sous la forme d'un masque binaire. Dans ce travail, l'extraction des silhouettes est réalisée manuellement,

puis les images des sommets du polyèdre situées sur les contours des silhouettes sont extraites automatiquement. Pour ce faire, nous appliquons aux masques l'algorithme de détection de segments de droites LSD [18]. Pour chaque vue, nous calculons les intersections de toutes les paires de droites portées par les segments détectés. Toute intersection située au voisinage de deux extrémités de segments est potentiellement l'image d'un sommet du polyèdre. Il est notable, cependant, que l'algorithme LSD tend à morceler les arêtes du polygone, notamment à cause du bruit d'acquisition, d'artefacts de compression ou du crénelage induit par l'échantillonnage spatial. D'autre part, l'estimation de l'image d'un sommet est très peu précise lorsque l'intersection est estimée entre deux arêtes quasi-parallèles. Pour ces raisons, les intersections entre deux segments quasi-parallèles sont rejetées. Ce mécanisme de rejet n'est pas critique, en pratique, si le nombre de vues dont nous disposons est suffisant pour effectuer la reconstruction 3D de l'ensemble des sommets du polyèdre.

Dans ce qui suit, nous notons V l'ensemble des points candidats qui ont été conservés à l'issue du calcul des intersections décrit ci-dessus. Cet ensemble constitue la donnée d'entrée du problème de mise en correspondance inter-vues, qui est exprimé sous la forme d'un problème de *partitionnement* robuste de V en sous-ensembles de points situés sur des ellipses qui sont les images de cercles 3D parallèles. Le nombre de sous-ensembles recherchés, à savoir le nombre de sommets du polyèdre, est supposé inconnu. Dans la suite de ce paragraphe, un *point* fait référence à un point de l'ensemble V .

4.1 Recherche exhaustive

Cinq points sont nécessaires pour estimer les paramètres d'une ellipse. Le problème de partitionnement peut donc être résolu par la recherche exhaustive des sous-ensembles de cinq points tels qu'un grand nombre de points soient situés à proximité de l'ellipse passant par les cinq points sélectionnés. En pratique, la combinatoire de cette recherche est trop élevée pour pouvoir être appliquée. Elle peut cependant être réduite en ajoutant deux hypothèses. Nous supposons dorénavant que l'axe de rotation de la table tournante se projette verticalement dans l'image, et que l'abscisse du centre des ellipses est connue. L'estimation d'une ellipse se ramène alors au calcul de trois paramètres, et donc à la sélection de trois points.

Nous constituons les $\mathcal{C}_3^{|V|}$ triplets de points pouvant être formés à partir de l'ensemble V . Les triplets contenant au moins deux points de la même image sont retirés de cet ensemble, car un sommet ne peut avoir qu'une image par vue. Les ellipses sont estimées pour l'ensemble des triplets restants. Cet ensemble d'ellipses est noté E .

Dans le cas d'un polyèdre de géométrie quelconque et d'un scénario sans bruit de mesure, si le nombre de vues est supérieur au nombre de sommets du polyèdre, c'est-à-dire si $J > N$, les ellipses de E passant par le plus grand nombre de points correspondent aux trajectoires des images des

sommets du polyèdre, et les sous-ensembles de points ainsi formés fournissent les correspondances inter-vues.

Nous proposons donc de trier les ellipses estimées selon le nombre de points situés dans leur voisinage. Les ellipses obtenant les scores les plus élevés correspondent aux sous-ensembles de points dont la probabilité de représenter les images d'un même sommet est la plus élevée. Le score est calculé comme suit.

Score – Un point est *situé à proximité d'une ellipse* A s'il est à l'intérieur de l'enveloppe elliptique $\mathcal{E}(A_{+\delta}, A_{-\delta})$, où $A_{+\delta}$ (resp. $A_{-\delta}$) désigne l'ellipse ayant le même centre et la même orientation que A , et dont les demi-axes sont égaux à ceux de A augmentés (resp. diminués) de δ . Il s'agit d'une approximation du seuillage de la distance euclidienne d'un point à une ellipse inférieure proposée dans [14]. En effet, le calcul de la distance euclidienne d'un point à une ellipse nécessite la résolution d'une équation quartique coûteuse en temps de calcul. L'appartenance d'un point de coordonnées cartésiennes augmentées \mathbf{x} à $\mathcal{E}(A_{+\delta}, A_{-\delta})$ est déterminée par le signe de $(\mathbf{x}^\top A_{+\delta} \mathbf{x})(\mathbf{x}^\top A_{-\delta} \mathbf{x})$. Si ce produit est négatif, alors le point se situe à l'intérieur de l'enveloppe, sinon il se situe à l'extérieur. Le score d'une ellipse A correspond au nombre de points de signe négatif. Si dans l'ensemble des points situés à proximité de A , plusieurs points apparaissent dans une même image, alors ces points sont comptabilisés une seule fois, car un sommet ne peut apparaître qu'une fois dans une vue.

Singularité A – Lorsqu'un sommet est situé sur l'axe de rotation de la table, sa trajectoire se réduit à un point. Une telle configuration pose problème pour le calcul du score à cause de la tolérance $\delta > 0$. Les enveloppes elliptiques contenant ce point obtiennent en général des scores très élevés, alors qu'elles associent les images du sommet à celles d'autres sommets. Nous contourignons ce problème grâce à un partitionnement préalable des points de V . Les groupes constitués d'au moins quatre points très proches, c'est-à-dire situés à une distance deux à deux inférieure à 0,012 sont exclus du calcul, après conditionnement des points dans le pavé $[-1, 1]^2$.

4.2 Élimination des erreurs

En pratique, les ellipses obtenant les scores les plus élevés correspondent dans certains cas aux images de plusieurs sommets. Ceci peut être dû à l'utilisation d'une tolérance δ non nulle dans le calcul du score, à la présence de bruit de mesure ou au croisement dans l'image des trajectoires de différents sommets. Un exemple d'ellipses associées aux scores les plus élevés correspondant à des regroupements aberrants est illustré sur la figure 4(a). La méthode d'estimation de paramètres que nous utilisons pour la reconstruction 3D des sommets du polyèdre et des poses de la caméra n'est pas robuste à de telles correspondances aberrantes. L'utilisation directe des correspondances fournies par les ellipses de scores les plus élevés est donc impossible. Néanmoins, un résultat pratique important pour la ré-

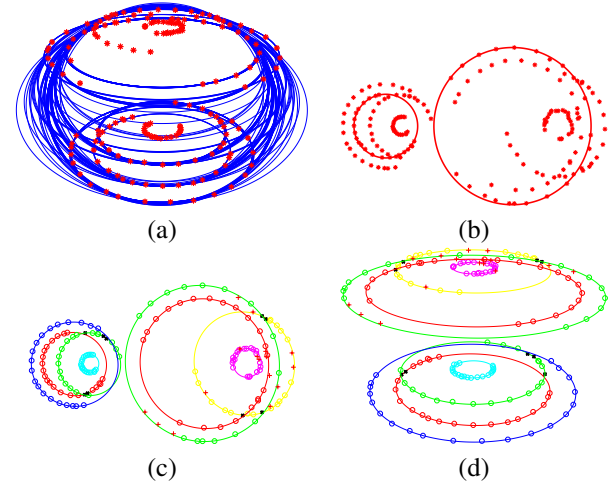


FIGURE 4 – Illustration de la mise en correspondance proposée. Les ellipses formées par les triplets de points sont triées selon le nombre de points (en rouge) situés dans leur voisinage. (a) Les K_{best} meilleures ellipses (en bleu). (b) Rectification euclidienne des points à partir des intersections d'une paire d'ellipses (en rouge) associées aux trajectoires de deux sommets. (c) Collecte des correspondances, résultat du partitionnement de points situés sur des cercles (chaque couleur est associée à un sommet). (d) Résultat de la mise en correspondance. Les croix noires indiquent les points pour lesquels la mise en correspondance est ambiguë, les points rouges les points non classés.

solution de notre problème est que les ellipses images des trajectoires des sommets font partie des ellipses ayant les scores les plus élevés. Ce résultat est crucial pour la mise en œuvre de la solution proposée dans ce qui suit.

À ce stade, le parallélisme des plans de support des trajectoires des sommets n'a pas encore été utilisé. Nous proposons d'exploiter cette propriété à partir de l'utilisation d'une paire de points à l'infini, complexes conjugués, inclus dans les trajectoires de tous les sommets du polyèdre. Ces points sont communément appelés *points cycliques* [9] du plan de support de la table tournante. Les intersections de coniques étant invariantes à toute transformation projective [9], l'ensemble des ellipses images des trajectoires des sommets intersectent les images d'une seule et même paire de points complexes conjugués notés IPC (pour *images des points cycliques*). Si les IPC sont identifiées, il devient alors possible de classer l'ensemble des ellipses collectées dans le paragraphe 4.1 en deux classes : celles qui appartiennent à la famille des trajectoires des images des sommets, lorsque celles-ci intersectent les IPC, et les autres.

Identification des IPC – En théorie, la seule famille d'ellipses de E de cardinal strictement supérieur à deux s'intersectant en une même paire de points complexes conjugués, à savoir les IPC, sont les images des trajectoires des sommets. Cette propriété nous permet de calculer les IPC du plan de support de la table comme la solution d'un problème de régression robuste des points d'intersec-

tion des ellipses de E . Notons que le hasard peut produire un sous-ensemble d'au moins trois ellipses de E s'intersectant en une même paire de points complexes conjugués autres que les IPC. Il est hautement improbable qu'un sous-ensemble d'ellipses s'intersectant en une même paire de points autres que les points cycliques ait un cardinal supérieur à celui d'une famille d'images de trajectoires de sommets. Seul le sous-ensemble d'ellipses de cardinal le plus élevé partageant des intersections communes est conservé. La solution du problème de régression robuste est calculée comme suit. Pour chaque paire d'ellipses $(A_1, A_2) \in E^2$, nous comptons le nombre d'ellipses de $E \setminus \{A_1, A_2\}$ passant par les intersections de (A_1, A_2) . Les intersections contenues par le plus grand nombre d'ellipses de E sont supposées être les IPC du plan de support de la table.

La solution proposée pour l'identification des IPC soulève cependant les deux sous-problèmes suivants.

Sous-problème 1 – Le premier sous-problème consiste à savoir comment compter le nombre d'ellipses de E passant par les intersections d'une paire d'ellipses. Deux ellipses images de cercles 3D parallèles s'intersectent en une paire de points réels et une paire de points complexes ou en deux paires de points complexes dont les IPC. En présence de bruit, les ellipses associées aux images des sommets ne contiennent pas les IPC du plan de support de la table. Plus généralement, ces ellipses ne s'intersectent pas en une même paire de points. Pour cette raison, le nombre d'ellipses passant par les IPC est nul, en pratique. L'espace projectif complexe $\mathcal{P}_2(\mathbb{C})$ n'est pas muni d'une métrique qui nous permettrait d'utiliser un seuil de tolérance sur la distance d'un point à une ellipse, pour déterminer si une ellipse contient ou non une paire donnée de points complexes conjugués. Nous proposons de contourner ce problème en considérant qu'une ellipse A *intersecte* les IPC si l'ellipse $H^T A H$, rectification de A par une certaine homographie :

$$H = [\mathbf{h}_1 \ \mathbf{h}_2 \ *] \quad (4)$$

où \mathbf{h}_1 et \mathbf{h}_2 sont les parties réelle et imaginaire de l'une des IPC, est un cercle. En pratique, la rectification est assimilée à un cercle si le ratio des demi-axes de $H^T A H$ est proche de 1. Pour une paire d'ellipses disjointes, deux homographies composées des parties réelles et imaginaires des deux paires de points sont appliquées à l'ensemble des ellipses. L'homographie présentant le plus grand nombre de bonnes rectifications est conservée [8]. Dans le cas d'une paire d'ellipses présentant des intersections réelles, seule les intersections complexes conjuguées sont testées.

Sous-problème 2 – Le deuxième sous-problème soulevé est un coût trop élevé pour le calcul des intersections des $|E|(|E| - 1)/2$ paires d'ellipses de E . Nous proposons d'utiliser un des résultats vus précédemment, à savoir que les ellipses de la famille des images des trajectoires des sommets sont parmi les ellipses possédant les meilleurs scores (cf. paragraphe 4.1). Nous proposons donc de réduire le nombre de calculs d'intersec-

tions à $K_{\text{best}}(K_{\text{best}} - 1)/2$, à savoir le calcul des intersections de toutes les paires d'ellipses pouvant être formées parmi les K_{best} meilleures ellipses. En pratique, la valeur de K_{best} n'est pas critique. Sa valeur fixée empiriquement doit être suffisamment grande pour garantir que les K_{best} meilleures ellipses incluent les ellipses images des trajectoires d'au moins deux sommets distincts. Nous choisissons $K_{\text{best}} = 200$, ce qui réduit le calcul des intersections à la résolution de 40000 systèmes de deux équations de degré 2 à deux inconnues. Un résultat du calcul des IPC est illustré sur la figure 4(b), où la paire d'ellipses maximisant le nombre d'ellipses rectifiées en cercles est affichée en rouge dans la représentation rectifiée.

Partitionnement – Au lieu de classer directement les ellipses intersectant les IPC comme appartenant à la famille des images des trajectoires des sommets et de collecter les correspondances associées, nous proposons d'appliquer une rectification euclidienne des points à partir de la connaissance des IPC et de partitionner les points rectifiés en sous-ensembles de points situés sur des cercles. Cette stratégie nous a permis d'améliorer les résultats. Ce constat peut très probablement être expliqué par une distribution des points dans la représentation rectifiée plus favorable à un partitionnement correct dans les cas de configurations ambiguës, comme par exemple pour des points distribués le long de portions d'ellipses très proches (voir l'exemple de la figure 4) ou de prises de vue très rasantes.

Les points sont rectifiés par l'homographie H^{-1} ayant pour expression (4). Les points rectifiés sont soumis à un partitionnement en sous-ensembles de points appartenant à des cercles. Les cercles sont triés par ordre décroissant du score décrit dans le paragraphe 4.1. Les correspondances inter-vues sont fournies par les sous-ensembles de points situés sur les cercles de scores les plus élevés. Chaque sous-ensemble associé à un cercle détecté est retiré en suivant un ordre décroissant du score, jusqu'à ce que le nombre de points restants soit inférieur à 5. Le résultat du partitionnement fournit les correspondances inter-vues.

Singularité B – Il se peut que le polyèdre soit positionné de telle sorte que les images des trajectoires de deux sommets soient situées sur une même ellipse. La méthode proposée ne permet pas de séparer les deux sous-ensembles de correspondances. En pratique, l'utilisateur peut éviter une telle configuration en excentrant le dioptré par rapport à l'axe de rotation de la table.

Nous disposons à présent des images des points cycliques $\mathbf{h}_1 \pm i\mathbf{h}_2$ et des correspondances inter-vues. Dans ce qui suit, nous montrons comment effectuer la reconstruction 3D des sommets du polyèdre et des poses de la caméra à partir de ces informations.

5 Reconstruction 3D

5.1 Formulation du problème

Hypothèses – Les hypothèses portant sur l’image de l’axe de rotation verticale et sur la connaissance a priori de l’abscisse de l’image du centre de rotation ne sont pas utilisées dans ce qui suit. En effet, à la différence du problème de mise en correspondance, l’utilisation de ces hypothèses pour la reconstruction 3D induirait des erreurs d’approximation de modèle trop importantes, relativement au niveau de précision de la solution attendue, pour être utilisées dans une application de stéréoscopie multi-vues.

Problème – Supposant connues les IPC $\mathbf{h}_1 \pm i\mathbf{h}_2$ et les correspondances inter-vues $\{\mathbf{x}_n^j\}$ calculées selon la méthode décrite dans le paragraphe 4, ainsi que la matrice de calibrage \mathbf{K} , le problème consiste à estimer les sommets et les poses du polyèdre dans le repère de la caméra. Plus précisément, il s’agit de calculer la matrice de rotation $\mathbf{R}_{\text{réf}}$, le vecteur de translation $\mathbf{t}_{\text{réf}}$, les coordonnées 3D des sommets $\{\mathbf{X}_n\}_{n=1..N}$ et les angles $\{\theta^j\}_{j=1..J}$.

5.2 Estimation de la pose de la caméra

Calcul de $\mathbf{R}_{\text{réf}}$ – Conformément à l’approche [21], nous cherchons une matrice orthogonale $\mathbf{R}_{\text{réf}}$ et un vecteur \mathbf{w} solutions du problème suivant :

$$\underset{\mathbf{R}_{\text{réf}}, \mathbf{w}, \lambda}{\operatorname{argmin}} \quad \left\| \lambda \mathbf{K}^{-1} \mathbf{H} - \mathbf{R}_{\text{réf}} \begin{bmatrix} 1 & 0 \\ 0 & 1 & \mathbf{w} \\ 0 & 0 \end{bmatrix} \right\|_F^2$$

s.c. $\mathbf{R}_{\text{réf}}^T \mathbf{R}_{\text{réf}} = \mathbf{I}_3$

Les deux premières colonnes de la solution en $\mathbf{R}_{\text{réf}}$ sont fournies par le produit $\mathbf{U}_{3 \times 2} \mathbf{V}_{2 \times 2}^T$, où les matrices $\mathbf{U}_{3 \times 2}$ et $\mathbf{V}_{2 \times 2}$ sont obtenues par la décomposition en valeurs singulières de $\mathbf{K}^{-1}[\mathbf{h}_1 \ \mathbf{h}_2]$:

$$\mathbf{K}^{-1}[\mathbf{h}_1 \ \mathbf{h}_2] = \mathbf{U}_{3 \times 2} \mathbf{\Sigma}_{2 \times 2} \mathbf{V}_{2 \times 2}^T \quad (5)$$

la troisième colonne de $\mathbf{R}_{\text{réf}}$ étant obtenue par le produit vectoriel des deux premières. La matrice \mathbf{H} est mise à jour par sa forme recomposée, à savoir :

$$\mathbf{H} \leftarrow \mathbf{K} \mathbf{R}_{\text{réf}} \begin{bmatrix} 1 & 0 \\ 0 & 1 & \mathbf{w} \\ 0 & 0 \end{bmatrix} \quad (6)$$

pour laquelle la valeur optimale de \mathbf{w} n’est pas requise. Nous encourageons le lecteur à se référer à [21] pour le calcul des valeurs optimales de \mathbf{w} et λ .

Calcul de $\mathbf{t}_{\text{réf}}$ – L’étape suivante consiste à calculer la matrice \mathbf{T} de translation garantissant que l’image du centre d’une trajectoire par $(\mathbf{T} \mathbf{H}^{-1})^{-1}$ est située à l’origine du repère attaché à la table tournante. Cette matrice s’écrit sous la forme suivante :

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & -x_c \\ 0 & 1 & -y_c \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

Dans (7), $[x_c, y_c]^T = f(\mathbf{H}^{-1} \mathbf{c}_n)$ désignent les coordonnées cartésiennes de l’image par \mathbf{H}^{-1} du centre \mathbf{c}_n , où la fonction f est telle que $f([u, v, w]^T) = [u/w, v/w]^T$ et \mathbf{c}_n est le vecteur de coordonnées homogènes de l’image du centre de la trajectoire, supposée circulaire, du sommet numéro n , obtenue à partir de la relation pôle-polaire $\mathbf{c}_n = \mathbf{A}_n^{-1} \mathbf{l}_\infty$, où \mathbf{l}_∞ désigne le vecteur de la ligne de fuite du plan de la table fournie par le produit vectoriel $\mathbf{l}_\infty = \mathbf{h}_1 \times \mathbf{h}_2$, et \mathbf{A}_n la matrice de l’ellipse image de la trajectoire du sommet numéro n . La matrice \mathbf{H} est mise à jour par $\mathbf{H} \leftarrow \mathbf{H} \mathbf{T}$. Le vecteur $\mathbf{t}_{\text{réf}}$ s’obtient alors ainsi :

$$\mathbf{t}_{\text{réf}} = \mathbf{K}^{-1} \mathbf{h}_3 \quad (8)$$

où \mathbf{h}_3 désigne la troisième colonne de \mathbf{H} .

5.3 Estimation des sommets et des poses du polyèdre

Calcul des angles $\{\theta^j\}$ – Nous proposons de calculer l’angle de rotation θ^j de la table entre une position de référence de la table, par exemple lors de l’acquisition de la première vue, et sa position dans la vue numéro j , comme la somme des angles $\theta^{k,k+1}$ correspondant à l’angle de rotation entre deux acquisitions successives, pour k tel que $1 \leq k$ et $k+1 \leq j$. La valeur de $\theta^{k,k+1}$ est obtenue en calculant la médiane des cosinus des angles estimés pour les différents sommets :

$$\theta^{k,k+1} = \operatorname{acos} \left(\operatorname{median}_{n \in \mathcal{D}^k} \left\{ \mathbf{d}_n^k \cdot \mathbf{d}_n^{k+1} \right\} \right)$$

où $\mathcal{D}^k \subset \{1, \dots, N\}$ désigne l’ensemble des indices des sommets ayant été détectés dans les images k et $k+1$, et où \mathbf{d}_n^k et \mathbf{d}_n^{k+1} désignent les vecteurs unitaires pointant vers les images par \mathbf{H}^{-1} des points correspondants \mathbf{x}_n^k et \mathbf{x}_n^{k+1} dans les deux vues consécutives k et $k+1$, soit :

$$\mathbf{d}_n^k = \frac{f(\mathbf{H}^{-1} \mathbf{x}_n^k) - f(\mathbf{H}^{-1} \mathbf{c}_n)}{\|f(\mathbf{H}^{-1} \mathbf{x}_n^k) - f(\mathbf{H}^{-1} \mathbf{c}_n)\|} \quad (9)$$

et idem pour \mathbf{d}_n^{k+1} . La table est censée tourner dans le sens trigonométrique durant l’acquisition. En outre, les valeurs de $\theta^{k,k+1}$ sont généralement comprises entre 0 et π . La valeur de θ^j s’écrit finalement :

$$\theta^j = \sum_{k=1}^{j-1} \theta^{k,k+1} \quad (10)$$

Calcul des positions des sommets – Nous proposons dans ce travail de reconstruire directement l’ensemble des sommets à partir des correspondances inter-vues et des matrices des ellipses $\{\mathbf{A}_n\}_{n=1..N}$, qui sont les images des trajectoires, calculées conformément à la méthode présentée dans le paragraphe 4.

Nous considérons dorénavant une représentation des sommets en coordonnées cylindriques. Les inconnues du problème, à savoir les rayons $\{\rho_n\}$ des sommets, leurs orientations $\{\phi_n\}$ et leurs cotes $\{z_n\}$, sont calculées comme suit.

En théorie, la cote z_n du sommet numéro n est égale à la cote du centre du cercle 3D que constitue sa trajectoire. Nous proposons de calculer la cote \hat{z}_n d'un point $\mathbf{t}_{\text{réf}} + \hat{z}_n \mathbf{r}_3$ situé sur l'axe de rotation de la table, ayant pour vecteur directeur la troisième colonne $\mathbf{r}_3 = [r_x, r_y, r_z]^\top$ de la matrice $\mathbf{R}_{\text{réf}}$, de manière à minimiser l'erreur de reprojec-tion égale à la distance euclidienne entre $f(\mathbf{t}_{\text{réf}} + \hat{z}_n \mathbf{r}_3)$ et l'image du centre de la trajectoire, supposée circulaire, du sommet $\mathbf{c}_n = \mathbf{A}^{-1} \mathbf{1}_\infty$. La valeur \hat{z}_n est donc solution du problème suivant :

$$\underset{z_n}{\operatorname{argmin}} \left(u_n - \frac{t_x + z_n r_x}{t_z + z_n r_z} \right)^2 + \left(v_n - \frac{t_y + z_n r_y}{t_z + z_n r_z} \right)^2 \quad (11)$$

Il est rappelé que $[u_n, v_n]^\top = f(\mathbf{K}^{-1} \mathbf{c}_n)$ désignent les coordonnées cartésiennes de l'image par \mathbf{K}^{-1} de \mathbf{c}_n , et que $\mathbf{t}_{\text{réf}} = [t_x, t_y, t_z]^\top$ désigne l'origine du repère de la table. La solution de (11) étant indépendante du repère de l'espace 3D euclidien choisi, nous changeons l'origine du repère de manière à annuler les valeurs de u_n et v_n . Dans ce nouveau repère, nous trouvons l'expression analytique suivante pour la solution du problème d'optimisation (11) :

$$\hat{z}_n = \frac{r_z(t_x^2 + t_y^2) - t_z(r_x t_x + r_y t_y)}{t_z(r_x^2 + r_y^2) - r_z(r_x t_x + r_y t_y)}$$

Les rayons des sommets, quant à eux, sont obtenus en calculant la médiane des normes des images par \mathbf{H}^{-1} des points correspondants, c'est-à-dire :

$$\rho_n = \operatorname{median}_{j \in \mathcal{M}_n} \{ \|f(\mathbf{H}^{-1} \mathbf{x}_n^j) - f(\mathbf{H}^{-1} \mathbf{c}_n)\| \}$$

où $\mathcal{M}_n \subset \{1, \dots, J\}$ contient les indices des images dans lesquelles le sommet numéro n a été détecté.

Enfin, dans un scénario sans bruit, l'angle polaire ϕ_n du sommet n dans la vue de référence (par exemple celle de numéro $j = 1$) est égal à la différence entre l'angle associé à l'image par \mathbf{H}^{-1} du point correspondant \mathbf{x}_n^j , et l'angle θ_j de rotation de la table calculé précédemment. Nous proposons donc une estimation de ϕ_n à partir de la moyenne de ces différences sur l'ensemble des vues dans lesquelles le sommet numéro n a été détecté, soit :

$$\phi_n = \operatorname{atan2}(a_n, b_n) \quad (12)$$

où :

$$\begin{bmatrix} a_n \\ b_n \end{bmatrix} = \frac{1}{|\mathcal{M}_n|} \sum_{j \in \mathcal{M}_n} \begin{bmatrix} \cos \theta_j & \sin \theta_j \\ -\sin \theta_j & \cos \theta_j \end{bmatrix} \mathbf{d}_n^j$$

5.4 Ajustement de faisceaux

L'ajustement de faisceaux consiste à optimiser simultanément les positions des sommets, les poses du polyèdre et la pose de la caméra. Il s'agit d'un problème aux moindres carrés non linéaire. Les paramètres à optimiser sont réunis dans le vecteur :

$$\nu = [\phi_x, \phi_y, \phi_z, \mathbf{t}_{\text{réf}}^\top, \mathbf{X}_1^\top, \dots, \mathbf{X}_N^\top, \theta^1, \dots, \theta^J]^\top \quad (13)$$

où ϕ_x, ϕ_y et ϕ_z désignent les angles d'Euler de la matrice de rotation $\mathbf{R}_{\text{réf}}$. La fonction de coût s'écrit :

$$\mathcal{C}(\nu) = \sum_{n=1..N} \sum_{j \in \mathcal{M}_n} d(\mathbf{x}_n^j, f(\mathbf{P}^j \mathbf{X}_n)) \quad (14)$$

où $d(\cdot, \cdot)$ désigne la distance euclidienne entre points 2D, $\{\mathbf{x}_n^j\}$ l'ensemble des correspondances extraites selon la méthode proposée dans le paragraphe 4, et \mathbf{P}^j la $j^{\text{ème}}$ matrice de projection, qui s'écrit sous la forme (2). L'optimisation de la fonction $\mathcal{C}(\nu)$ est effectuée par l'algorithme de Levenberg-Marquardt.

6 Résultats expérimentaux

6.1 Données de synthèse

Jeux de données – Dix jeux de données de synthèse ont été créés. Chaque vue comporte 1000×1000 pixels. Le point principal de la caméra est situé au centre de l'image. Les pixels sont carrés. La distance focale est fixée à 700 pixels. Pour chaque jeu de données, la caméra est située à une distance du centre de rotation de la table tournante tirée aléatoirement dans l'intervalle $[4, 5; 5, 5]$. La caméra pointe vers l'axe de rotation de la table. Son inclinaison est tirée aléatoirement entre -25 et -45 degrés. Au total, $J = 35$ angles θ^j sont répartis régulièrement entre 0 et 360 degrés, moyennant un bruit gaussien d'écart-type $\pi/60$, et triés par ordre croissant. Les sommets du polyèdre sont obtenus en calculant l'enveloppe convexe d'un ensemble de points 3D de cardinal arbitraire très supérieur à 8. Les coordonnées des points 3D sont tirées aléatoirement dans le pavé $[-1, 1]^3$. Des points tirés aléatoirement sont retirés de l'enveloppe convexe jusqu'à obtenir 8, 7, 6, 5 et 4 points, ce qui nous fournit cinq ensembles de sommets de polyèdres convexes. Les reprojctions de ces sommets sont générées. Un bruit blanc gaussien d'écart-type variant entre 0 et 1, 5 pixels, par pas de 0, 1 pixel, est ajouté aux abscisses et aux ordonnées des reprojctions obtenues. Il en résulte un total de 138000 jeux de données de synthèse.

Seuls les résultats de la reconstruction 3D sont évalués sur les données de synthèse, les correspondances étant supposées connues. La mise en correspondance sera donc seulement évaluée sur les données réelles.

Évaluation de la reconstruction 3D – Nous avons évalué la méthode de reconstruction 3D pour un écart-type du bruit gaussien allant de 0 à 1, 5 pixels, et pour un nombre de vues allant de 10 à 35. Les résultats sont caractérisés par la racine carrée de l'erreur quadratique moyenne :

$$\operatorname{REQM} = \sqrt{\frac{1}{N} \frac{1}{J} \sum_{n=1}^N \sum_{j=1}^J (\hat{\mathbf{X}}_n^j - \mathbf{X}_n^j)^2} \quad (15)$$

où $\hat{\mathbf{X}}_n^j$ et \mathbf{X}_n^j désignent, respectivement, l'estimation et la vérité terrain des coordonnées du sommet numéro n dans la vue j , exprimées dans le repère de la caméra, soit :

$$\hat{\mathbf{X}}_n^j = [\hat{\mathbf{R}}_{\text{réf}} \mathbf{R}_{\hat{\theta}^j} | \hat{\mathbf{t}}_{\text{réf}}] \hat{\mathbf{X}}_n \quad \text{et} \quad \mathbf{X}_n^j = [\mathbf{R}_{\text{réf}} \mathbf{R}_{\theta^j} | \mathbf{t}_{\text{réf}}] \mathbf{X}_n$$

Les résultats sont présentés sur la figure 6.

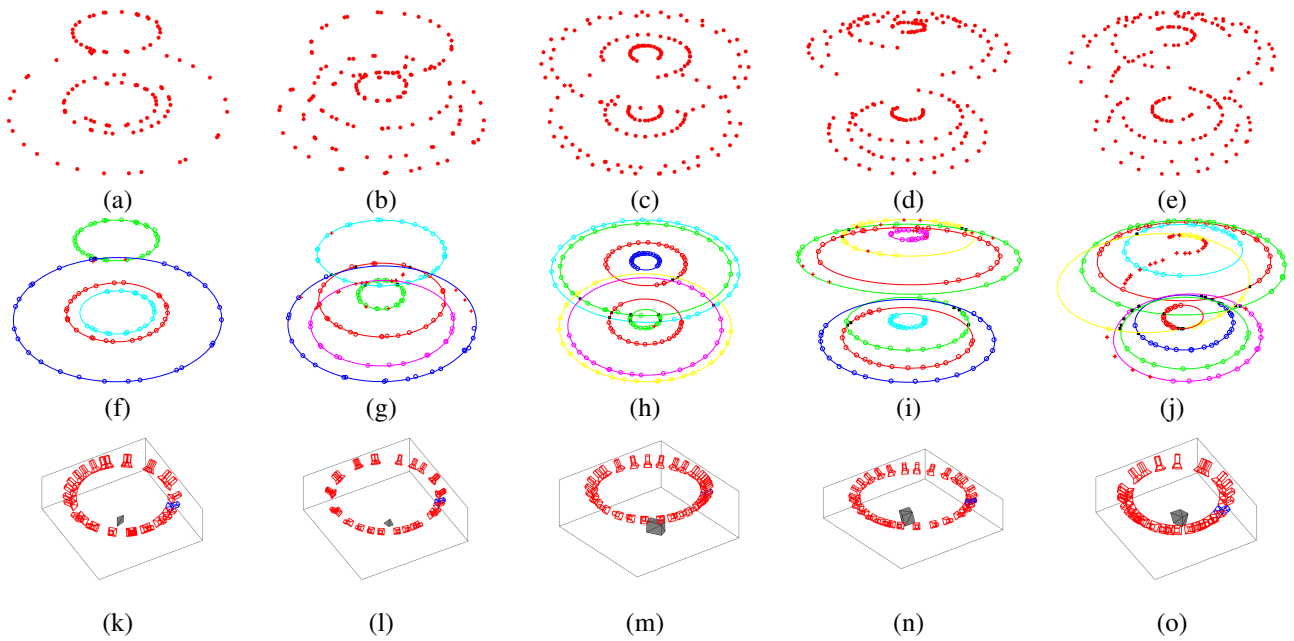


FIGURE 5 – (a-e) Points en entrée de la chaîne de traitement, pour les jeux de données *tétraèdre*, *pyramide carrée*, *pavé A*, *pavé B* et *polyèdre à 9 sommets*. (f-j) Correspondances obtenues. Une couleur est associée à chaque sommet. Les croix rouges désignent les points non classés, les croix noires les correspondances de classe ambiguë, qui sont exclues des données d’entrée de la reconstruction 3D. Le cas (j) constitue un cas d’échec de la méthode. (k-o) Reconstructions 3D obtenues.

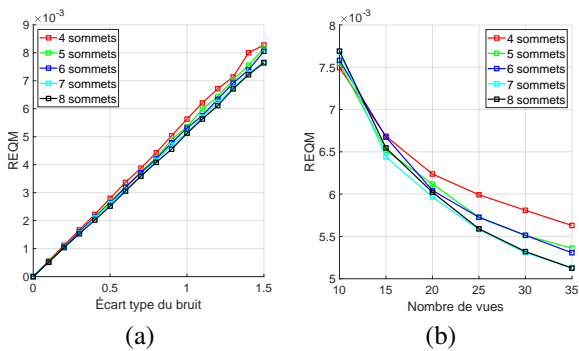


FIGURE 6 – (a) REQM en fonction de l’écart-type du bruit sur les mesures pour 35 vues. (b) REQM en fonction du nombre de vues pour un bruit d’écart-type égal à 1 pixel.

6.2 Données réelles

La figure 7 montre le dispositif d’acquisition de données réelles, et un exemple d’image réelle obtenue.

Les algorithmes de mise en correspondance et de reconstruction 3D ont été évalués sur cinq jeux de données réelles. Le premier jeu correspond aux vues d’un tétraèdre qui correspond au polyèdre comportant le nombre minimal de sommets, à savoir $N = 4$. Les autres jeux de données correspondent à une pyramide à base carrée ($N = 5$), à deux pavés droits ($N = 8$) et à un polyèdre à $N = 9$ sommets. Les résultats de la mise en correspondance et de la reconstruction 3D sont présentés sur la figure 5.

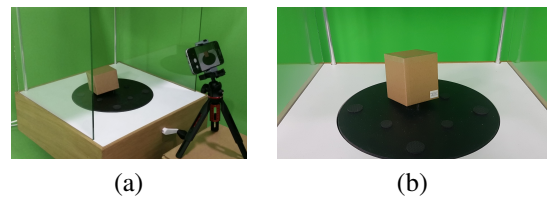


FIGURE 7 – (a) Dispositif d’acquisition de données réelles. (b) Exemple d’acquisition.

7 Conclusion et perspectives

Dans cet article, nous avons présenté un paradigme de reconstruction 3D des sommets et des poses d’un polyèdre convexe, et les détails nécessaires à sa mise en œuvre. Un algorithme d’extraction et de mise en correspondance des images des sommets d’un polyèdre disposé sur une table tournante est proposé. La méthode de reconstruction 3D, qui est évaluée sur 138000 jeux de données de synthèse, présente de bonnes performances. La mise en correspondance est évaluée sur cinq jeux de données réelles.

Nous envisageons d’améliorer la méthode en intégrant l’estimation des IPC à partir de l’ensemble des images des trajectoires proposée dans [8], au lieu d’utiliser seulement deux ellipses. La méthode de mise en correspondance est encore instable, notamment à cause de trois paramètres critiques, dont la valeur de la tolérance δ utilisée pour le calcul des scores. Enfin, l’algorithme de mise en correspondance doit faire l’objet d’une évaluation quantitative.

Références

- [1] M. Cassidy, J. Mérou, Y. Quéau, F. Lauze, and J.-D. Durou. Refractive Multi-view Stereo. In *Proceedings of 3DV*, 2020.
- [2] F. Chadebecq, F. Vasconcelos, R. Lacher, E. Maneas, A. Desjardins, S. Ourselin, T. Vercauteren, and D. Stoyanov. Refractive Two-View Reconstruction for Underwater 3D Vision. *International Journal of Computer Vision*, 128(5):1101–1117, 2020.
- [3] Y.-J. Chang and T. Chen. Multi-view 3D reconstruction for scenes under the refractive plane with known vertical direction. In *Proceedings of ICCV*, 2011.
- [4] V. Chari and P. F. Sturm. Multi-View Geometry of the Refractive Plane. In *Proceedings of BMVC*, 2009.
- [5] J.-D. Durou, M. Falcone, Y. Quéau, and S. Tozza. *Advances in Photometric 3D-Reconstruction*. Advances in Computer Vision and Pattern Recognition. Springer International Publishing, 2020.
- [6] A. W. Fitzgibbon, G. Cross, and A. Zisserman. Automatic 3D Model Construction for Turn-Table Sequences. In *European Workshop on 3D Structure from Multiple Images of Large-Scale Environments*, 1998.
- [7] Y. Furukawa and C. Hernández. Multi-view stereo: A tutorial. *Foundations and Trends in Computer Graphics and Vision*, 9(1-2):1–148, 2015.
- [8] P. Gurdjos, P. F. Sturm, and Y. Wu. Euclidean Structure from $N \geq 2$ Parallel Circles: Theory and Algorithms. In *Proceedings of ECCV*, 2006.
- [9] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Second edition, 2004.
- [10] C. E. Hernández and F. Schmitt. Silhouette and Stereo Fusion for 3D Object Modeling. *Computer Vision and Image Understanding*, 96(3):367–392, 2004.
- [11] M. Jancosek and T. Pajdla. Multi-view reconstruction preserving weakly-supported surfaces. In *Proceedings of CVPR*, 2011.
- [12] G. Jiang, H.-T. Tsui, L. Quan, and A. Zisserman. Single Axis Geometry by Fitting Conics. In *Proceedings of ECCV*, 2002.
- [13] G. Jiang, H.-T. Tsui, L. Quan, and A. Zisserman. Geometry of Single Axis Motions Using Conic Fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1343–1348, 2003.
- [14] K.-i. Kanatani and N. Ohta. Automatic Detection Of Circular Objects By Ellipse Growing. *International Journal of Image and Graphics*, 4(1):35–50, 2004.
- [15] L. Kang, L. Wu, and Y. Yang. Two-View Underwater Structure and Motion for Cameras under Flat Refractive Interfaces. In *Proceedings of ECCV*, 2012.
- [16] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [17] R. Mur-Artal and J. D. Tardós. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262, 2017.
- [18] V. Pătrăucean, P. Gurdjos, and R. Grompone von Gioi. Joint A Contrario Ellipse and Line Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):788–802, 2017.
- [19] E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski. ORB: An efficient alternative to SIFT or SURF. In *Proceedings of ICCV*, 2011.
- [20] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the World from Internet Photo Collections. *International Journal of Computer Vision*, 80(2):189–210, 2008.
- [21] P. F. Sturm. Algorithms for Plane-Based Pose Estimation. In *Proceedings of CVPR*, 2000.
- [22] H. Zhang and K. K. Wong. Self-calibration of turntable sequences from silhouettes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):5–14, 2009.