



Approche photométrique pour une triangulation précise

El Mustapha Mouaddib, Eric Marchand, François Chaumette

► **To cite this version:**

El Mustapha Mouaddib, Eric Marchand, François Chaumette. Approche photométrique pour une triangulation précise. RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle), Jan 2012, Lyon, France. pp.978-2-9539515-2-3, 2012. <hal-00656556>

HAL Id: hal-00656556

<https://hal.archives-ouvertes.fr/hal-00656556>

Submitted on 17 Jan 2012

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.

Approche photométrique pour une triangulation précise

El Mustapha Mouaddib^{1,3}, Eric Marchand^{2,3}, François Chaumette³

¹ Laboratoire MIS (Modélisation, Information et Systèmes),
Université de Picardie Jules Verne
MIS-UPJV 33, Rue Saint Leu - 80033 Amiens, France

² Université de Rennes 1 - IRISA

³ INRIA Rennes - Bretagne Atlantique
Campus de Beaulieu, 35042 Rennes, France

mouaddib@u-picardie.fr

Résumé

La triangulation est l'opération qui consiste à estimer la position d'un point 3D à partir de points mis en correspondance dans les images. Elle a souvent été abordée uniquement d'un point de vue géométrique, c'est à dire, seules les positions des points ainsi que la géométrie des vues sont utilisées. Lorsqu'il n'y a pas de bruit, le problème de la triangulation est trivial et se résout en inversant les équations linéaires qui lient la position 3D aux positions des points dans les images. Lorsqu'on est en présence de bruit, ce qui est toujours le cas avec des images réelles, une estimation précise et robuste devient nécessaire. C'est ainsi que des méthodes optimales, sous l'hypothèse d'un bruit gaussien, ont été développées. L'inconvénient majeur de ces méthodes est qu'elles ne tiennent pas compte de l'information photométrique contenue dans les pixels. Nous montrerons dans cet article en quoi ceci peut être gênant. Nous proposerons une nouvelle approche qui est basée sur l'ajustement photométrique. Les résultats sur des images de synthèse et sur des images réelles montreront que cette prise en compte de l'information photométrique améliore la précision de reconstruction 3D.

Mots Clef

Triangulation, ajustement géométrique, ajustement photométrique, reconstruction 3D.

Abstract

Triangulation is the process of estimating the 3D position of a matched points from images. It has often been addressed from a geometric point of view, that is, where only positions of the points and the geometry of the views are used. When there is no noise, this triangulation issue is trivial and is solved by inverting the equations that link the view lines to the 3D point. In the presence of noise, which is always the case with real images, an accurate and robust estimation is needed. Under Gaussian noise assumption,

optimal methods have been developed. The major drawback of these methods is that they do not use the photometric information contained in the pixel neighborhood. We propose a new approach that is based on Photometric Adjustment. The results on synthetic images and real images show that taking into account the photometric information improves the accuracy of 3D reconstruction.

Keywords

Triangulation, geometric, photometric, 3D reconstruction.

1 Problématique et motivations

La reconstruction 3D à partir d'une paire d'images repose souvent sur le paradigme suivant :

- extraction de points d'intérêt dans les deux images,
- mise en correspondance de ces points,
- estimation des lignes de vues,
- triangulation pour estimer le point 3D à reconstruire.

Notons \mathbf{u}_1 et \mathbf{u}_2 deux points image appariés issus des deux premières étapes. Ils sont la projection, sur les plans images I_1 et I_2 , d'un point \mathbf{M} de \mathbf{R}^3 . Ces projections sont modélisées à l'aide de deux matrices de projections \mathbf{P}_1 et \mathbf{P}_2 supposées connues.

$$\begin{cases} \mathbf{u}_i = (u_i, v_i, 1)^T \\ \mathbf{M} = (X, Y, Z, 1)^T \\ w \mathbf{u}_i = \mathbf{K}_i \mathbf{H}_i \mathbf{M} = \mathbf{P}_i \mathbf{M} \end{cases} \quad (1)$$

où :

\mathbf{K} est la matrice des paramètres intrinsèques.

\mathbf{H} est la transformation rigide entre le repère de \mathbf{M} et le repère de la caméra.

La triangulation est l'opération qui consiste à déduire les coordonnées du point \mathbf{M} à partir de \mathbf{u}_1 , \mathbf{u}_2 , \mathbf{P}_1 et \mathbf{P}_2 . C'est un problème fondamental et bien connu en vision par ordinateur [8].

Dans un système parfait, la triangulation revient à calculer

l'intersection des deux lignes de vue issues de \mathbf{u}_1 et \mathbf{u}_2 , qui coïncide avec \mathbf{M} . Dans le cas réaliste, les deux lignes de vues peuvent ne pas s'intersecter (comme on peut le voir sur la figure 1) ou ne pas coïncider avec le point réel \mathbf{M} . Ceci peut être dû à :

1. des erreurs de localisation des points
2. des erreurs d'estimation de la matrice de projection (paramètres d'étalonnage et transformation rigide entre les vues)

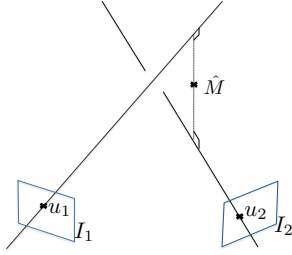


FIGURE 1 – Illustration de l'approximation de la triangulation par le point milieu

La méthode du point milieu (à mi distance entre les deux lignes de vue) est une méthode qui donne de bons résultats et qui est fréquemment utilisée pour estimer la position du point 3D. D'autres méthodes existent. On peut en trouver une très bonne synthèse et analyse dans [7]. Dans cet article, et dans [6] et [5], les auteurs ont montré que, si on suppose que le bruit qui accompagne les mesures des points image est gaussien, alors maximiser la vraisemblance entre les points détectés et les points corrigés revient à minimiser le critère suivant :

$$d(\mathbf{u}_1, \hat{\mathbf{u}}_1)^2 + d(\mathbf{u}_2, \hat{\mathbf{u}}_2)^2 \quad (2)$$

qui correspond à l'erreur de reprojection des points images, où les projections dans les deux images de l'estimé $\hat{\mathbf{M}}$ de \mathbf{M} sont $\hat{\mathbf{u}}_1$ et $\hat{\mathbf{u}}_2$ et s'obtiennent avec l'équation :

$$w \hat{\mathbf{u}}_i = \mathbf{K}_i \mathbf{H}_i \hat{\mathbf{M}} = \mathbf{P}_i \hat{\mathbf{M}} \quad (3)$$

L'idée consiste à déplacer les points pour que les lignes de vue soient coplanaires, en minimisant le critère ci-dessus tout en respectant la contrainte épipolaire :

$$\hat{\mathbf{u}}_2^T \mathbf{F} \hat{\mathbf{u}}_1 = 0 \quad (4)$$

où \mathbf{F} est la matrice fondamentale.

La résolution de ce problème permet de corriger les positions des images du point \mathbf{M} pour obtenir les nouvelles positions $\hat{\mathbf{u}}_1$ et $\hat{\mathbf{u}}_2$ qui satisfont la contrainte (4) tout en minimisant (2). On peut aussi reformuler cette contrainte à l'aide de la matrice essentielle.

Hartley et Sturm [7] ont apporté une solution basée sur la résolution d'un polynôme de degré six. Cette méthode est souvent appelée triangulation optimale ou méthode polynomiale.

Dans un même temps et sous les mêmes hypothèses sur le bruit des mesures, [10] ont également proposé une méthode basée sur les mêmes critères, mais la résolution est faite par linéarisation de la contrainte épipolaire puis par itérations successives. Un peu plus tard, dans [11], puis dans [9], un nouvel algorithme itératif de la méthode a été proposé.

Enfin, dans un article publié en 2010 [12], l'auteur revisite les contributions mentionnées ci-dessus et propose plusieurs variantes d'algorithmes itératifs pour résoudre ce problème. Il montre, et c'est là sa contribution importante, que la solution optimale peut être trouvée en deux itérations et propose *in fine* un algorithme rapide et très précis pour la triangulation.

Dans ce paradigme, seule l'information géométrique, i.e., la position des points images, est utilisée. Georjel et al. [2] ont proposé d'utiliser également l'information photométrique des points images et de leur voisinage pour l'estimation de la pose de la caméra et pour la reconstruction 3D. Ils ont ainsi élaboré une approche qui combine l'information géométrique et photométrique pour l'estimation de la pose de la caméra. Cette approche est qualifiée d'unifiée. L'ajout de l'information photométrique est destiné à assurer la cohérence et la consistance photométrique des points. Les fonctions de coût sont normalisées pour rendre quantitativement homogènes les deux termes.

Cependant, le choix des valeurs des coefficients de la pondération entre les deux fonctions de coût pose problème. En effet, il est difficile de savoir *a priori* s'il faut privilégier la contribution du géométrique au détriment du photométrique ou vice versa.

Nous proposons dans la section suivante une approche qui se base uniquement sur l'information photométrique. Elle est très proche de celle développée dans [1] en vue de la reconstruction 3D dense. Elle présente l'avantage de garantir la consistance ou cohérence photométrique et évite le problème de la pondération tout en tenant compte de la contrainte géométrique. Le principe de notre approche est le même que celui des méthodes LSM (Least Square Method) développées dans [3] et [4]. Cependant, notre travail vise essentiellement à réaliser la triangulation en utilisant des patches là où les méthodes LSM ont pour objectif de recalculer des surfaces. Le problème est légèrement différent parce que dans notre cas, les positions des points images appariés sont ajustées en même temps que le calcul du point 3D.

2 Approche proposée

L'objectif est de reconstruire la position (X, Y, Z) d'un point \mathbf{M} dans \mathbf{R}^3 de manière très précise et sûre. L'approche repose sur l'hypothèse que le motif photométrique (le patch) associé à un point \mathbf{M} de \mathbf{R}^3 se projette en deux motifs similaires et visibles dans les deux images I_1 et I_2 .

La figure 2 matérialise en couleur rouge les patches Γ_1 et Γ_2 autour des deux points images \mathbf{u}_1 et \mathbf{u}_2 et autour du point \mathbf{M} .

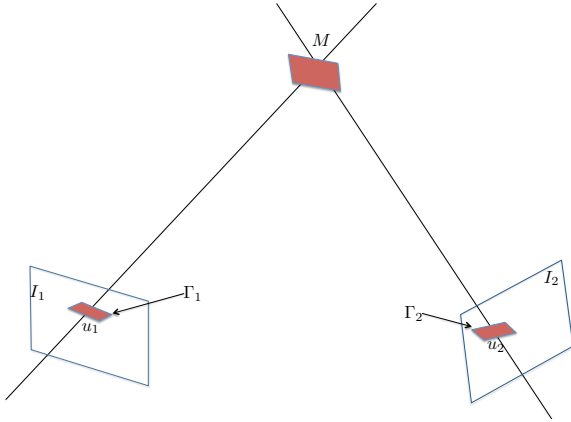


FIGURE 2 – Images du patch de \mathbf{M} .

Dans le cas d'images réelles, les erreurs d'extraction des points images appariés provoquent des décalages dans la superposition des patches. L'idée principale de l'algorithme est que si on ajuste les patches alors on va améliorer la précision de l'estimation du point 3D \mathbf{M} . C'est ce que nous appelons la triangulation photométrique. Elle fonctionne en faisant varier les coordonnées (X, Y, Z) du point \mathbf{M} dans l'espace de manière itérative (voir l'algorithme dans la section 2.1), afin de minimiser les écarts entre les patches liés aux points appariés.

Comme l'ont montré [1], il est possible d'estimer également la normale du patch 3D. Le travail que nous proposons se limitant à la question de la triangulation, nous n'estimons que la position 3D des points.

Contrairement aux méthodes géométriques qui ne privilégient aucun des deux points images, l'approche que nous proposons considère que le point extrait dans I_1 est le point à reconstruire et par conséquent son patch sera utilisé comme patch de référence. On appellera ce point image \mathbf{q} : le point d'ancrage.

2.1 Principe

Pour minimiser l'écart entre les patches, nous proposons d'utiliser la SSD (Sum of Squared Difference) comme fonction de coût (eq. (5)) et (eq. (6)) :

$$\hat{\mathbf{M}} = \arg \min_{\mathbf{M}} \epsilon(\mathbf{M}) \quad (5)$$

avec :

$$\epsilon(\mathbf{M}) = \sum_{\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2} (I_1(\mathbf{r}) - I_1(\mathbf{v}_1))^2 + (I_1(\mathbf{r}) - I_2(\mathbf{v}_2))^2 \quad (6)$$

où :

- $\mathbf{M} \in \mathbf{R}^3$: coordonnées du point 3D à ajuster.
- $\mathbf{r} \in \Lambda$: coordonnées des pixels du patch Λ ,
- $\mathbf{v}_1 \in \Gamma_1$: coordonnées des pixels du patch Γ_1 ,

- $\mathbf{v}_2 \in \Gamma_2$: coordonnées des pixels du patch Γ_2 ,

Ces patches sont des régions carrées de pixels, centrés respectivement sur \mathbf{q} , \mathbf{u}_1 et \mathbf{u}_2 . Dans l'équation (6), le point \mathbf{q} est connu et ne varie pas (point d'ancrage) et par conséquent, le patch Λ est également connu et constant (Fig. 3). Les patches Γ_1 et Γ_2 correspondent aux patches issus, respectivement, de la projection \mathbf{M} sur I_1 et I_2 . Ils sont représentés en vert sur la figure 3.

Même si \mathbf{q} sert de référence, il est primordial d'inclure \mathbf{v}_1 et Γ_1 dans la minimisation pour garantir (autant que possible) que \mathbf{M} appartienne à la ligne de vue associée à \mathbf{q} . Ceci constitue une contrainte plus forte que la contrainte épipolaire.

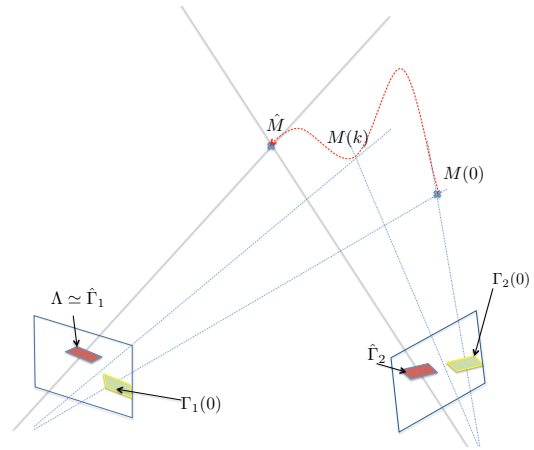


FIGURE 3 – Illustration du principe de l'ajustement. Les patches rouges sont les patches théoriques. La ligne rouge indique un exemple de positions successives prises par le point 3D lors du processus itératif.

Pour résoudre l'équation (5), plusieurs algorithmes sont possibles (voir [13] pour une discussion approfondie). En ce qui nous concerne, nous avons utilisé l'algorithme de Levenberg-Marquardt pour cette minimisation et l'équation dérivée de l'hypothèse d'illumination constante :

$$I(\mathbf{v}, s) \simeq I(\mathbf{v} + \Delta \mathbf{v}, s + \Delta s)$$

$$I(\mathbf{v}, s) \simeq I(\mathbf{v}, s) + \frac{\delta I}{\delta \mathbf{v}} \Delta \mathbf{v} + \frac{\delta I}{\delta s} \Delta s \quad (7)$$

$$\frac{\delta I}{\delta \mathbf{v}} \Delta \mathbf{v} \simeq -\frac{\delta I}{\delta s} \Delta s$$

$$\frac{\delta I}{\delta u} \Delta u + \frac{\delta I}{\delta v} \Delta v \simeq -\frac{\delta I}{\delta s} \Delta s$$

et ensuite :

$$\frac{\delta I}{\delta u} \frac{\delta u}{\delta \mathbf{M}} \Delta \mathbf{M} + \frac{\delta I}{\delta v} \frac{\delta v}{\delta \mathbf{M}} \Delta \mathbf{M} \simeq -\frac{\delta I}{\delta s} \Delta s \quad (8)$$

Les différents points images du patch permettent de construire un système surdéterminé d'équations à l'aide de (8) et d'estimer ainsi $\Delta \mathbf{M}$.

$\frac{\delta u}{\delta \mathbf{M}}$ est évaluée à partir de la relation (1). Les matrices \mathbf{P}_i sont supposées connues (caméras calibrées).

$\frac{\delta I}{\delta u}$ et $\frac{\delta I}{\delta v}$ sont évaluées à l'aide de convolutions avec des dérivées de Gaussiennes.

$\frac{\delta I}{\delta s}$ est approximée par une différence entre les intensités des patches. Ce qui donne $I_1(\mathbf{r}) - I_1(\mathbf{v}_1)$ pour le premier terme de (6), et $I_1(\mathbf{r}) - I_2(\mathbf{v}_2)$ pour le second terme, si on pose que $I_1(\mathbf{r}) = I_1(\mathbf{r}, s)$, $I_1(\mathbf{v}_1) = I_1(\mathbf{v}_1, s_1)$ et $I_1(\mathbf{v}_2) = I_2(\mathbf{v}_2, s_2)$.

La solution $\tilde{\mathbf{M}}$ est approchée par itérations successives $\mathbf{M}(k) = \mathbf{M}(k-1) + \Delta\mathbf{M}$ en partant d'une position initiale $\mathbf{M}(0)$ du point 3D. L'objectif du travail présenté dans cet article étant la triangulation photométrique, $M(0)$ est calculé soit par triangulation à partir des coordonnées des points images appariés, soit directement avec la valeur du point 3D fourni par l'algorithme de triangulation géométrique. À l'itération suivante, l'algorithme estime une correction à appliquer pour obtenir $\mathbf{M}(1)$. La nouvelle position de ce point est projetée sur I_1 en $\mathbf{u}_1(1)$ et sur I_2 en $\mathbf{u}_2(1)$. Ces points servent à extraire les patches $\Gamma_1(1)$ et $\Gamma_2(1)$ suivants et ainsi de suite jusqu'à la convergence vers une solution.

Le principe exposé ci-dessus est immédiatement généralisable à plusieurs images. En effet, il suffit de rajouter les termes concernant les autres images dans la fonction de coût.

2.2 Interpolation

Comme pour le calcul de la disparité, il est indispensable de procéder à l'interpolation des patches pour améliorer la précision de l'ajustement photométrique [15], [14].

La projection du point $\mathbf{M}(k)$ dans I_1 et I_2 fournit des coordonnées réelles pour $\mathbf{u}_1(k)$ et $\mathbf{u}_2(k)$. La caméra a une résolution finie, ce qui a comme conséquence :

$$I_i(\mathbf{u}(k)) = I_i([\mathbf{u}(k) - \Delta\mathbf{u}, \mathbf{u}(k) + \Delta\mathbf{u}]) \quad (9)$$

$\Delta\mathbf{u}$ étant la demi-largeur et la demi-longueur d'un pixel.

Lorsqu'on réalise une reconstruction 3D à partir de $\mathbf{u}_1(k)$ et $\mathbf{u}_2(k)$, il y a donc fatalement une incertitude sur la position de $\mathbf{M}(k)$ parce que les pixels sont connus à $\Delta\mathbf{u}$ près. Dans ce cas, l'ajustement photométrique n'améliore pas les résultats de la reconstruction 3D si les patches sont ajustés au pixel près.

On peut envisager au moins deux solutions. La première consiste à augmenter la résolution de l'image en réalisant un ré-échantillonnage. Cette approche, outre le problème du temps de calcul parce que relativement important, ne fait que repousser le problème du *pixellocking*. En effet, le raisonnement développé ci-dessus pour la limite de la précision de l'ajustement photométrique, peut être à nouveau appliqué avec une résolution plus importante.

La seconde solution consiste à effectuer une interpolation locale des images. Nous avons retenu une approche basée sur l'interpolation bicubique. À chaque projection de $\mathbf{M}(k)$ sur les images en $\mathbf{u}_1(k)$ et $\mathbf{u}_2(k)$, les intensités des deux patches $\Gamma_1(k)$ et $\Gamma_2(k)$ sont interpolés. La SSD est

ensuite calculée sur ces deux patches et sur Λ qui aura été interpolé auparavant.

3 Résultats expérimentaux

L'objectif est d'évaluer la précision de reconstruction des points 3D et le comportement de l'algorithme en fonction de la présence ou non de l'interpolation, du bruit géométrique, et du bruit photométrique. Nous avons d'abord procédé à une évaluation sur des images de synthèse.

Expérience 1 : Validation sur des images de synthèse.

Pour cela, un couple d'images a été généré (Fig. 4). Les images se composent de 200 points répartis dans un volume de $400 \times 400 \times 40 \text{ cm}^3$ situé à 300cm de la caméra :

- 40 points avec $Z = 300\text{cm}$
- 120 points avec $Z = 280\text{cm}$
- 40 points avec $Z = 320\text{cm}$

La résolution de l'image est de 1000×1000 pixels. Les paramètres intrinsèques de la caméra perspective sont : $u_0 = v_0 = 500$, $a_u = a_v = 714$.

Le mouvement de la caméra entre les deux images est composé des translations et rotations suivantes : $t_x = -10\text{cm}$; $t_y = -5\text{cm}$; $t_z = -5\text{cm}$; $Rot_x = 3^\circ$; $Rot_y = 2^\circ$; $Rot_z = 1^\circ$. Il a été choisi volontairement faible pour montrer l'apport de la méthode dans le cas critique de faible base line. Les coins des objets sont extraits avec un algorithme de Harris avec une précision subpixelique.

On a procédé à la comparaison entre la reconstruction des points 3D avec la méthode de triangulation géométrique [12] décrite dans la Section 1 et notre méthode. Ces méthodes seront nommées *GEOM* et *PHOTO* dans la suite de cet article. Le tableau (Table 1) montre un résultat qui confirme l'amélioration (attendue) apportée par l'interpolation dans le cas sans bruit.

TABLE 1 – Comparaison sans bruit et apport de l'interpolation.

	Moy. (cm)	std (cm)
GEOM	0.63	0.32
PHOTO sans interp.	1.98	1.83
PHOTO avec interp.	0.41	0.18

Nous avons ensuite effectué des expérimentations pour évaluer et comparer la méthode photométrique et géométrique.

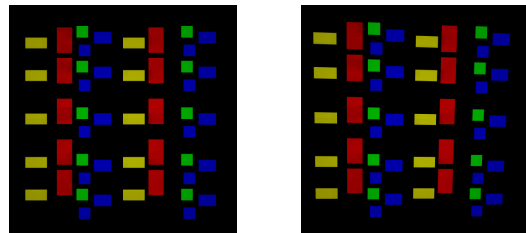


FIGURE 4 – Paire d'images utilisée pour l'évaluation de la méthode proposée

trique en présence de bruit. Deux types de bruit ont été considérés : un bruit, noté w_g , sur les coordonnées des points images utilisées pour la triangulation et un bruit, noté w_p , sur l'image (les intensités de l'image) avant extraction des points. La figure 5 montre le résultat de la reconstruction des points 3D avec : $w_g = 0$ et $w_p = 5$. Pour une meilleure lisibilité des résultats, les points reconstruits ont été projetés sur les plans XZ et YZ . Les points reconstruits à l'aide de la méthode photométrique (en rouge) sont moins dispersés que ceux reconstruits par la reconstruction géométrique (en bleu). Les points théoriques répartis (en vert) avec $Z = -200, Z = 0, Z = 200$.

Ces résultats ont été confortés par une évaluation quantitative. Nous avons fait varier les deux paramètres de bruit w_g et w_p pour étudier l'erreur 3D i.e. l'écart entre le point 3D estimé et le point 3D théorique (Fig. 6). Sur cette figure, chaque graphe représente l'évolution de l'erreur de reconstruction 3D en fonction du bruit photométrique w_p pour un bruit géométrique w_g donné. Ces résultats ont été obtenus avec un voisinage de taille égale à 5 pixels. Chaque erreur a été calculée en moyennant 100 mesures.

On remarque que la reconstruction 3D faite par *PHOTO* est dans tous les cas meilleure par rapport à *GEOM* et ceci de manière très prononcée.

Expérience 2 : Evaluation sur des images réelles. Nous présentons ici une comparaison avec les images réelles et

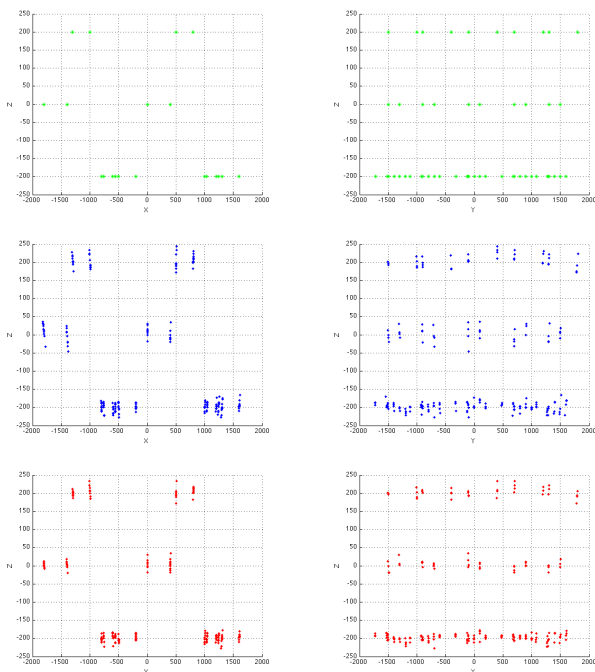


FIGURE 5 – Reconstruction en présence de bruit. En vert, la vérité terrain (points de synthèse). En bleu, la reconstruction obtenue par *GEOM*. En rouge, la reconstruction obtenue par *PHOTO*.

les données fournies par le VGG¹. La séquence se compose de cinq images et on peut voir un extrait de quatre images dans la figure 7.

Ces données contiennent également les points images mis en correspondance, les matrices de projection et les points 3D reconstruits. Ces points 3D vont servir pour la comparaison et seront considérés comme étant la vérité terrain. La séquence contient 1331 points.

Chaque point 3D fait l'objet d'une mise en correspondance entre au moins deux images parmi les cinq. La méthode est utilisable avec autant d'images qu'on le souhaite. Dans l'exemple de la séquence de cinq images, très peu de points ont été mis en correspondance sur plus de deux images. Pour des raisons d'homogénéisation des résultats, nous avons volontairement limité à deux le nombre de correspondances (deux images parmi les cinq de la séquence) pour chaque point 3D.

On peut voir sur la figure 8, une visualisation de la reconstruction 3D obtenue par *PHOTO* et les points 3D de la vérité terrain.

La figure 9 présente les écarts entre les points 3D reconstruits en utilisant notre méthode et la vérité terrain, en fonction des points.

Pour comprendre les raisons de ces écarts, nous avons analysé plus finement les points pour lesquels cet écart est supérieur à 0.3 (ce seuil permet d'avoir un nombre de points qui ne rend pas trop fastidieux la vérification manuelle). Le tableau 2 présente les numéros de ces points et les valeurs de ces écarts. Sur les 16 points relevés :

1. L'algorithme *PHOTO* n'a pas amélioré la reconstruction pour un point. Ceci est probablement dû à l'insuffisance de la méthode d'interpolation.

1. Visual Geometry Group - <http://www.robots.ox.ac.uk/vgg/data/dataview.html>

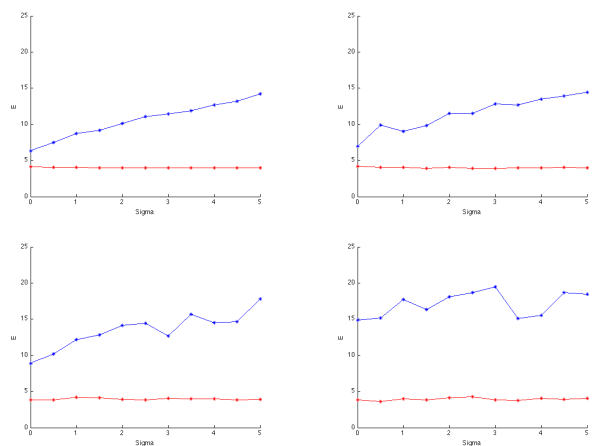


FIGURE 6 – Evolution de l'erreur en fonction du bruit w_p dans l'image (courbe en rouge pour *PHOTO* et courbe en bleu pour *GEOM*). Les graphes correspondent successivement de haut en bas et de gauche vers la droite à $w_g = 0, 0.2, 0.4, 0.6$

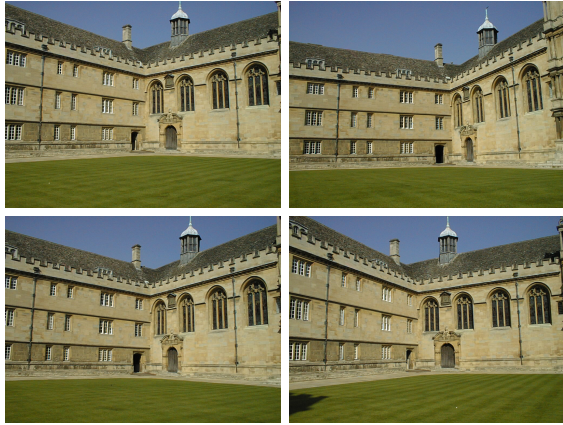


FIGURE 7 – Exemples d’images de la séquence utilisée pour la reconstruction

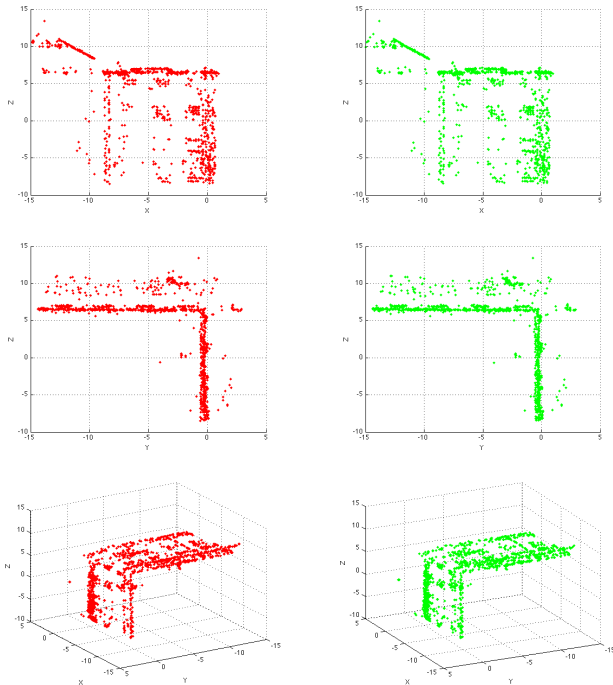


FIGURE 8 – Reconstructions 3D des points estimés par PHOTO et les points vérité terrain. Vues projetées sur les plans XZ et YZ et vues perspectives.

2. Un point correspond à un appariement vérité terrain aberrant comme on peut le voir sur la première ligne de la figure 10 (confusion due aux fenêtres) i.e. la position 3D vérité terrain n’est pas correcte.
3. Trois points correspondent à des situations où il est très difficile de décider par insuffisance de texture.
4. Onze points sont des points pour lesquels l’appariement est correct mais peu précis. En effet, la figure 10 montre un cas de l’amélioration apportée par la méthode PHOTO. Sur la partie basse de cette figure,

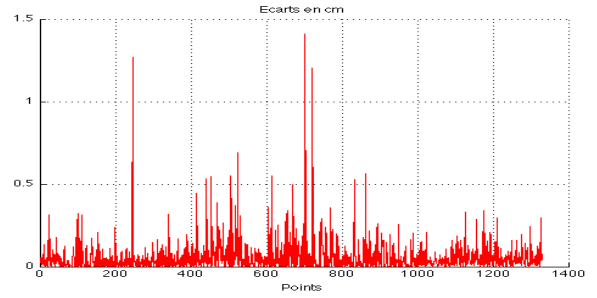


FIGURE 9 – Courbes des écarts des points reconstruits par PHOTO pour $T = 5$.

l’image de gauche correspond à l’image I_1 et celle de droite à I_2 . Sur toutes les images, les croix vertes indiquent les positions vérité terrain et celles estimées par PHOTO sont rouges. On peut constater sur cette figure que ces dernières correspondent à une mise en correspondance plus précise. On remarque également que le point u_1 a subi une correction plus faible que le point u_2 , ce qui est normal puisque le point u_1 sert de référence.

TABLE 2 – Points avec écart $>0.3cm$. O pour PHOTO a amélioré la précision. N pour le contraire. ? pour difficile de trancher.

Num.	Ecart (cm)	Vérification
87	0.32	N
152	0.33	?
246	0.42	Vérité terrain aberrante
341	0.34	O
500	0.35	O
524	0.74	O
535	0.31	O
606	0.47	O
615	0.40	O
670	0.45	?
672	0.37	O
674	0.37	O
833	0.38	O
857	0.38	O
863	0.44	O
876	0.32	?

La figure 9 montre les écarts obtenus avec un voisinage de taille $T = 5$ et la figure 11 ceux obtenus avec $T = 10$. L’augmentation de la taille du voisinage améliore la précision. En outre, elle améliore la convexité de la surface de la fonction de coût, mais pénalise le temps de calcul. Enfin, l’évaluation de la moyenne et de l’écart type de l’erreur photométrique (sur une échelle allant de 0 à 255 niveau de gris), confirme l’avantage de la méthode photométrique comme le montre la table 3.

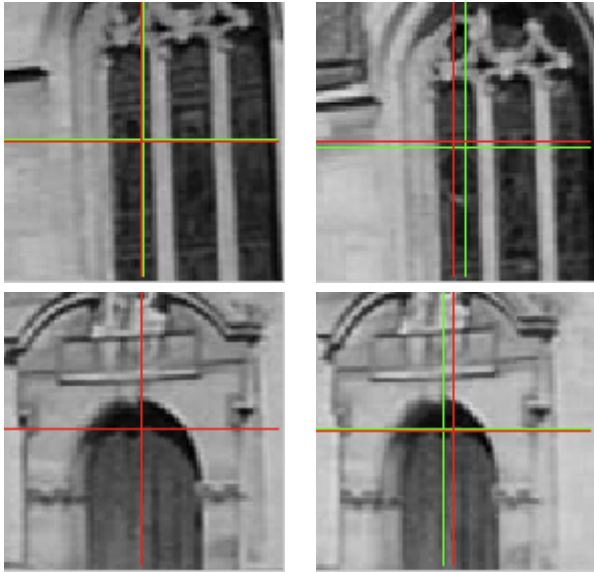


FIGURE 10 – Ajustement de positions pixels. En rouge celles estimées par *PHOTO* et en vert celles de la vérité terrain. Voisinage de taille 5. Les deux sous figures du haut montrent le point 246 qui est aberrant. Les deux sous figures du bas montrent l’apport de *PHOTO* pour le point 615.

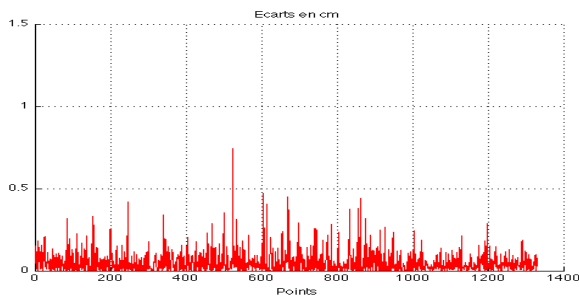


FIGURE 11 – Courbes des écarts des points reconstruits par *PHOTO* pour $T = 10$.

TABLE 3 – Résidus photométriques.

	Moy. (ndg)	Ecart type (ndg)
GEOM	9.1	5.8
PHOTO	7.0	4.5

Expérience 3 : Robustesse aux positions initiales.

L’expérience suivante est destinée à évaluer la robustesse de notre approche à l’initialisation des positions des points images dans I_2 . Pour cela, les positions \mathbf{u}_2 ont été perturbées avec un bruit gaussien de variance 5. La figure 12 montre les résultats de reconstruction 3D obtenus. On peut remarquer, contrairement à *GEOM*, que l’approche proposée corrige les positions pixels et permet d’obtenir une bonne reconstruction 3D. Il faut cependant souligner que les corrections sont limitées par la taille du voisinage utilisé. Il va de soi que la distance de la position initiale de

\mathbf{u}_2 ne peut être supérieure à $T/2$. La table 4 montre que l’erreur résiduelle photométrique est plus importante dans le cas de *GEOM* alors que pour *PHOTO*, elle reste très proche de celle obtenue avec des positions initiales très favorables.

TABLE 4 – Résidus photométriques (positions initiales bruitées).

	Moy. (ndg)	Ecart type (ndg)
GEOM	15.6	10.3
PHOTO	7.6	4.9

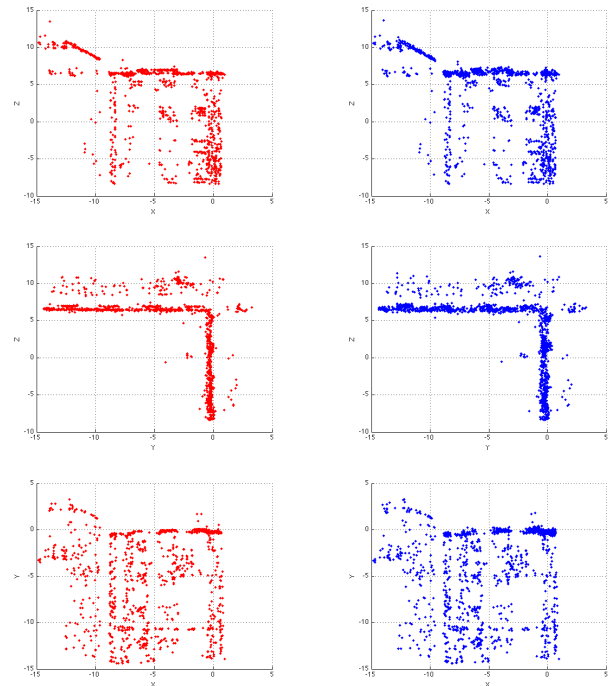


FIGURE 12 – Reconstructions 3D des points estimés par *PHOTO* et les points reconstruits par *GEOM* directement à partir des points bruités.

Expérience 4 : Symétrie et choix du point d’ancrage.

Il s’agit ici de quantifier la précision de reconstruction en permutant le patch de référence. X_1 est l’ensemble des points 3D reconstruits en utilisant I_1 comme image des patches de référence. X_2 est l’équivalent en utilisant I_2 comme base des patches de référence. L’influence de l’image de référence est faible : $Moy|X_1 - X_2| = 0,0036cm$. On peut considérer que la méthode est symétrique. Cependant, cette méthode privilégie le point de référence dans le sens où elle considère que c’est ce point qui est à reconstruire.

4 Conclusion

Les algorithmes de triangulation basés exclusivement sur les positions de points mis en correspondance peuvent se

révéler insuffisamment précis. Ils ont cependant l'avantage d'être très rapides, même dans leur version itérative. Nous avons proposé un algorithme de triangulation précise basé sur l'utilisation de l'information photométrique qui entoure les pixels à trianguler. Même si cet algorithme a un coût plus important en temps de calcul, il améliore la précision et la robustesse au bruit, que ce soit pour du bruit géométrique et/ou photométrique. Les résultats obtenus sur les images de synthèse confirment cette amélioration. Nous avons également montré l'apport de cette approche sur des images réelles. Il faut aussi souligner que ces bons résultats sont obtenus avec seulement deux correspondances par point, alors que les résultats utilisés pour la comparaison (vérité terrain) ont été obtenus par ajustement de faisceaux sur l'ensemble de la séquence. Enfin, la précision et la robustesse pourraient être encore améliorées en ayant recours à d'autres méthodes d'interpolation.

Références

- [1] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8), 2010.
- [2] P. Georgel, S. Benhimane, and N. Navab. A unified approach combining photometric and geometric information for pose estimation. In *19th British Machine Vision Conference (BMVC)*, Sep 2008.
- [3] A. Gruen and D. Akca. Least squares 3d surface matching. In *IAPRS*, 34(5/W16), 2004.
- [4] A. W. Gruen. Adaptive least squares correlation : A powerful image matching technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography*, 14 :175–187, 1985.
- [5] R. Hartley and P. Sturm. Triangulation. In *ARPA Image Understanding Workshop, Monterey, California, USA*, pages 957–966, Nov 1994.
- [6] R. Hartley and P. Sturm. Triangulation. In *6th International Conference on Computer Analysis of Images and Patterns, Prague, Czech Republic*, pages 190–197, Sep 1995.
- [7] R. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2) :146–157, 1997.
- [8] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN : 0521623049, 2000.
- [9] K. Kanatani, Y. Sugaya, and H. Niitsuma. Triangulation from two views revisited : Hartley-sturm vs. optimal correction. In *BMVC'08, Proc. 9th British Machine Vision Conference*, pages 173–182, 2008.
- [10] Y. Kanazawa and K. Kanatani. Reliability of 3-d reconstruction by stereo vision. *IEICE Trans. Inf. and Syst. E78-D(10)*, pages 1301–1306, 1995.
- [11] K. Kenichi, S. Yasuyuki, and N. Hirotsuka. Triangulation from two views : Hartley-sturm vs. optimal correction. In *14th Symp. Sensing via Image Information, June 2008, Yokohama, Japan*, 2008.
- [12] P. Lindstrom. Triangulation made easy. In *The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010*, pages 1554–1561, 2010.
- [13] E. Malis. Improving vision-based control using efficient second-order minimization techniques. In *Proceedings of the 2004 IEEE International Conference on Robotics and Automation, ICRA 2004, April 26 - May 1, 2004, New Orleans, LA, USA*, pages 1843–1848, 2004.
- [14] M. Shimizu and M. Okutomi. Sub-pixel estimation error cancellation on area-based matching. *Int. J. Comput. Vision*, 63 :207–224, July 2005.
- [15] R. Szeliski and D. Scharstein. Sampling the disparity space image. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26 :419–425, 2004.