

Mise en correspondance de points 3D obtenus avec une grande “base-line”

N. Soontranon^{1,2}

M. Pierrot-Deseilligny¹

E. Mouaddib²

¹ Laboratoire MATIS, IGN

73 Avenue de Paris – 94160 Saint Mandé, France

² Laboratoire MIS, UPJV

33 Rue Saint Leu – 80039 Amiens, France

narut.soontranon@ign.fr – marc.pierrot-deseilligny@ensg.eu – mouaddib@u-picardi.fr

Résumé

Nous présentons dans cet article deux approches pour obtenir les points homologues entre des images prises avec une base longue d’une scène fixe. Récemment, de nombreux dispositifs d’acquisition peuvent fournir non seulement l’image mais aussi l’information 3D associée (par exemple LiDAR, caméra stéréo, multi-stéréoscopie). Nous montrons que l’information 3D peut être utilisée pour calculer des images qui sont invariantes pour les transformations géométriques (affine et perspective) entre les images stéréo. Les images transformées sont ensuite appariées avec SIFT. Dans la première approche, pour une scène composée de surfaces planes, l’image est transformée en une orthoimage rectifiée pour chaque plan. Avec la deuxième approche, pour une scène de surfaces en relief, l’image est transformée par une cartographie conforme (conformal mapping). Les résultats de comparaison montrent que les points homologues obtenus avec nos approches sont nombreux et bien répartis dans les images.

Mots Clef

Point homologue, reconstruction 3D, stéréo, SIFT, orthoimage, cartographie conforme.

Abstract

This paper presents two approaches to obtain the homologous points between wide baseline images, for a fixed scene. Recently, many acquisition devices can provide not only the image but also associated 3D information (e.g. LiDAR, stereo camera, multi-stereoscopic). We show that the 3D information can be used to compute the images, which are invariant to the geometric transformations (affine and perspective) between the stereo image. The transformed images are then matched by SIFT. First approach, for a planar scene, the image is transformed to the orthorectification for each planar region. Second approach, for a smooth scene, the image is transformed by using a conformal mapping. The results, which are compared between SIFT and our approaches, show that the homologous points are increased and distributed on the images.

Keywords

Homologous point, 3D reconstruction, stereo, SIFT, ortho rectified, conformal mapping.

1 Introduction

La production de modèles tridimensionnels (3D) de scènes fixes est utilisée dans de nombreux domaines : études architecturales, urbanisme, archéologie, pour atteindre différents objectifs tel que le tourisme, la modélisation 3D de la ville et l’archivage numérique du patrimoine. Elle est reliée à la vision par ordinateur et à la photogrammétrie.

L’état de l’art actuel de la modélisation en 3D de manière précise et robuste des scènes complexes à partir de simples photos se réduit au cas de scènes bien texturées et couvertes avec suffisamment d’images. Le schéma de reconstruction est généralement le suivant :

- Calcul de points homologues entre images par des méthodes invariantes aux principales transformations géométriques et radiométriques « Scale-Invariant Feature Transform (SIFT) [1], Speeded-Up Robust Features (SURF) [2], Maximally Stable Extremal Regions (MSER) [3] ».
- Calibration et orientation des images [12] à partir des points homologues précédents.
- Utilisation des orientations pour la mise en correspondance de manière suffisamment dense.

Les méthodes de détection des points homologues dépendent fortement de la base stéréoscopique utilisée. À titre d’exemple, le produit Photofly d’Autodesk recommande pour la modélisation d’un objet de prendre une photo tous les 10-15 degrés. Le calcul des points homologues est une première étape obligatoire dans toutes les approches de reconstruction de modèles 3D à partir de plusieurs des images [10, 11, 12, 18]. L’algorithme SIFT [1] est le plus connu et le plus utilisé en pratique. SIFT est très efficace, mais il est limité quand la géométrie varie trop fortement entre les deux images. La figure 1 présente un exemple utilisant SIFT qui n’a pas trouvé de points homologues sur l’objet “Borne F” en raison d’une forte variation de la géométrie entre les deux images.

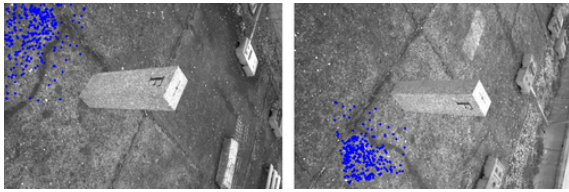


FIGURE 1 – Les points homologues obtenus par SIFT sont en bleu. Nous remarquons que les points ne peuvent pas être trouvés sur l’objet “Borne F” (gauche :vue 1, droite :vue 2).

Aujourd’hui, il existe de nombreux dispositifs d’acquisition qui sont utilisés pour obtenir une image avec sa profondeur associée. La figure 2 montre les dispositifs d’acquisition qui sont disponibles actuellement ; LiDAR (Light Detection And Ranging), caméra stéréo, multi-stéréoscopie. Dans cet article, nous nous concentrons sur les caméras ; caméra stéréo ou multi-stéréoscopie parce que ces appareils sont disponibles pour le grand public. Les autres avantages de la caméra sont qu’elle est moins cher et plus pratique que le LiDAR.



FIGURE 2 – Les systèmes d’acquisition “image avec 3D” : LiDAR, caméra stéréo, multi-stéréoscopie et autres systèmes.

Notre article est organisé comme suit : Section 2, nous rappelons l’état de l’art qui peut-être classé en trois catégories en s’appuyant sur les types de sources utilisées : image-image, 3D-3D, approche hybride. Dans la section 3, nous présentons notre approche dans le cadre d’une approche hybride qui permet d’obtenir les points homologues qui sont invariants et robustes à toutes les transformations géométriques. Un processus est proposé pour traiter les surfaces planes et un autre pour les surfaces en relief. Les ré-

sultats sont montrés dans la section 4. Enfin, la section 5 conclut notre article avec les perspectives de ces travaux.

2 Travaux antérieurs

En s’appuyant sur une approche classique de reconstruction, la première étape importante en reconstruction 3D est de trouver les points correspondants dans les images prises de différents points de vue. Dans une scène fixe, nous distinguons trois catégories d’approches pour l’obtention de points homologues en fonction du type de données utilisées.

2.1 Recalage image-image

Dans cette catégorie, deux images sont acquises de différents points de vue. C’est la radiométrie qui peut être utilisée pour obtenir l’invariance géométrique.

- Détecter des points d’intérêt (ou points clés) : généralement, on extrait les points saillants (ou coins) [15, 16] ou les régions de Laplacien maximum [1, 2, 3].
- Déterminer les descripteurs : typiquement, on calcule le descripteur à partir des pixels du voisinage de ces points.
- Apparier les descripteurs : L’algorithme optimise l’appariement des descripteurs pour une grande base de descripteurs [19].

Les algorithmes image-image comme SIFT permettent de trouver des points homologues qui sont invariants par translation, rotation et homothétie dans le plan 2D, mais en théorie, ils ne sont pas invariants par transformation affine ni par effet de perspective. Ils ne garantissent donc pas d’apparier les projections d’un même objet 3D selon différents angles de vue. Dans le cas d’images prises avec une base longue, on peut se référer à ces travaux [4, 5]. Les résultats sont améliorés, mais les temps de calcul sont longs.

2.2 Recalage 3D-3D

La production d’objets en 3D par LiDAR a suscité beaucoup d’attention durant la dernière décennie en raison d’équipements devenus disponibles au grand public. Pour recalcr entre eux deux nuages de points, l’algorithme classique, appelé ICP (Iterative Closest Point) [6] fonctionne bien si l’algorithme est initialisé avec une solution proche de la solution recherchée, mais il échoue lorsqu’il y a une transformation géométrique trop importante entre les deux nuages. Dans ce cas, il faut initialiser les positions relatives avec des stratégies de type RANSAC (RANdom Sample Consensus) [14] par exemple, mais la recherche est encore active dans ce domaine.

2.3 Hybride

Les approches hybrides utilisent les deux types d’information : image et 3D. Notre approche appartient à cette catégorie, elle se fonde sur les approches hybrides qui utilisent des images associées à une carte de profondeur. L’objectif de ces travaux est d’utiliser la 3D pour la correction géométrique. Les méthodes proposées [8, 9] ont extrait des plans à partir de l’information de profondeur pour obtenir

des images qui sont invariantes par transformations géométriques. On prend comme point de départ les travaux de [17] qui a utilisé la cartographie conforme pour transformer l'image initiale en une image conforme pour des applications de reconnaissance faciale.

3 Approche proposée

On cherche à utiliser l'information de profondeur pour rendre l'étape d'appariement plus robuste à ces transformations induites par la géométrie 3D. En entrée, on dispose de deux images avec leurs profondeurs associées (voir la figure 3). Ces deux cartes ne sont pas dans le même repère, on ne peut donc pas comparer directement l'information 3D issue des deux images. Dans notre approche, on utilise la 3D pour extraire des images texturées : des images ortho-rectifiées pour les surfaces planes, ou des images conformes pour les surfaces en relief. On obtient donc des images où le relief a été "aplani" et où l'algorithme SIFT ne sera pas altéré par les déformations affines ou perspectives.

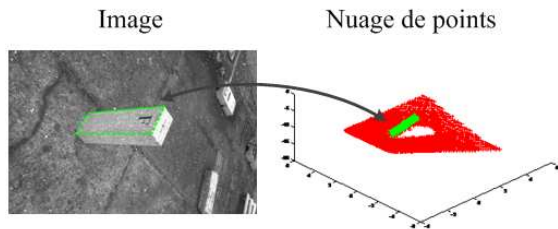


FIGURE 3 – Données en entrée de l'algorithme pour chaque point de vue : l'image (gauche) avec son nuage de points associé (droite).

La méthode s'applique localement à des régions de l'image. Dans chaque région, on extrait ce qu'on appelle une «texture de région». Dans chacune de ces «textures de région», les points homologues SIFT sont calculés. Ils doivent ensuite être associés à des pixels dans les images d'origine pour trouver la transformation entre les deux images. Pour cela, il faut calculer la transformation inverse de celle qui permet de passer de l'image à la texture de région.

On peut expliquer l'idée générale comme suit :

- Calculer des images qui sont localement ortho rectifiées. Mémoriser également la transformation entre l'image initiale et l'image ortho rectifiée.
- Obtenir les points homologues entre les ortho rectifiées.
- Importer les points homologues dans les repères des images initiales en utilisant la transformation inverse.

3.1 Cas des surfaces planes

Notre première contribution est une méthode appliquée pour les scènes de surfaces planes (voir la figure 4). Nous avons implémenté la méthode pour trouver les points homologues dans ce cas. Dans la figure 5, en utilisant des



FIGURE 4 – Exemple de scène de surfaces planes.

images avec leurs profondeurs, l'algorithme détaillé est le suivant :

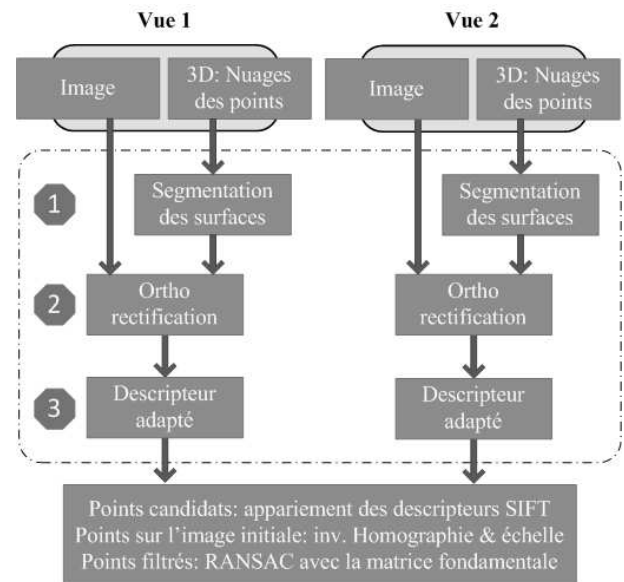


FIGURE 5 – Méthode pour une scène de surfaces planes.

- 1) Segmentation des surfaces planes : à partir des nuages de points, chaque surface plane est segmentée. Cette étape peut être remplacée par l'algorithme standard, par exemple la croissance de région montre quelques résultats dans la figure 6.
- 2) Calcul de l'ortho photo : chaque région de surface plane est convertie en ortho photo pour obtenir les invariances affine et perspective. La figure 7 présente une ortho-photo de la région segmentée. Cette ortho-photo est obtenue par une matrice translation et une matrice rotation. La figure 8 présente les points homologues qui sont trouvés entre les ortho-photos par la méthode SIFT.

- 3) Adaptation du descripteur SIFT (comme le montre la figure 9). En plus d'utiliser le descripteur SIFT classique, nous avons ajouté deux informations importantes - la taille de l'ortho photo et la matrice d'homographie. (utile pour la transformation inverse pour revenir dans l'image initiale)

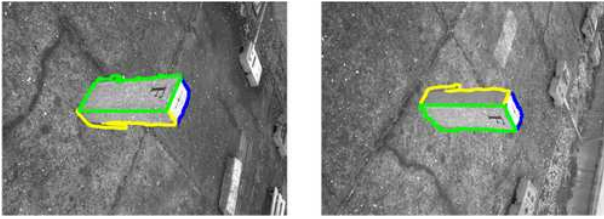


FIGURE 6 – Segmentation des surfaces planes avec les nuages de points (gauche : vue 1, droite : vue 2).

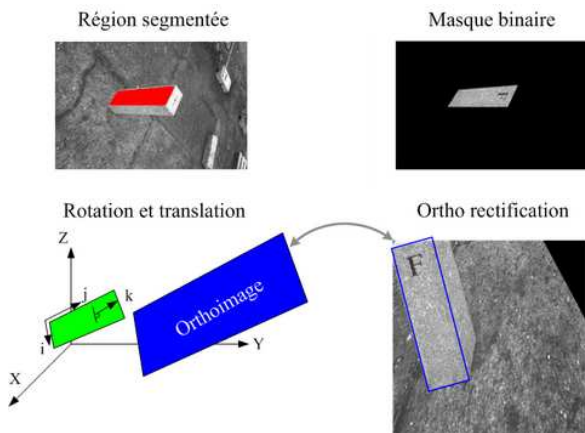


FIGURE 7 – L'ortho-photo de la région segmentée (vue 1).

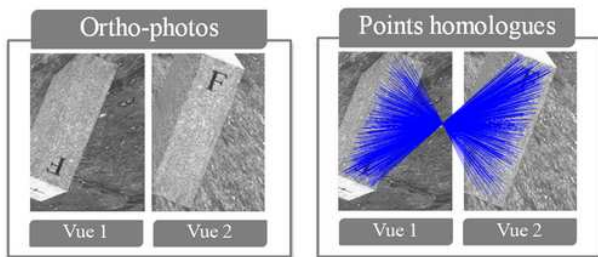


FIGURE 8 – Nous avons utilisé la méthode SIFT sur les ortho-photos. À gauche : les orthoimages rectifiées. À droite : les points homologues trouvés par cette approche.

3.2 Cas des surfaces en relief

Dans cette section, nous parlons des scènes plus réalistes (par exemple les sculptures). Le processus utilisé pour les

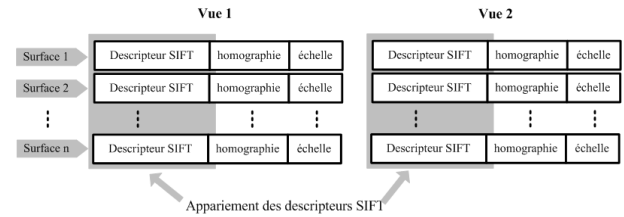


FIGURE 9 – Adaptation des descripteurs SIFT : l'information en plus est l'homographie et l'échelle.

surfaces en relief est présenté dans la figure 10 en utilisant comme données en entrée de l'algorithme, un couple d'images avec leurs profondeurs associées (figure 11).

- 1) Filtre passe-bas : compte tenu de l'information de profondeur, à la première étape, nous appliquons le filtre passe-bas pour obtenir un objet de surface lisse (la profondeur à partir d'une image stéréo est toujours un peu bruitée).
- 2) Sous-échantillonnage : une chose importante est de réduire le temps de calcul. Ainsi, l'algorithme de sous-échantillonnage (figure 12) est utilisé pour réduire la complexité du nuage de points.
- 3) Cartographie conforme (conformal mapping) [7] : pour obtenir l'image invariante par transformation affine, nous utilisons la méthode de cartographie conforme (voir la figure 13). Dans cet article, nous utilisons l'algorithme [13].

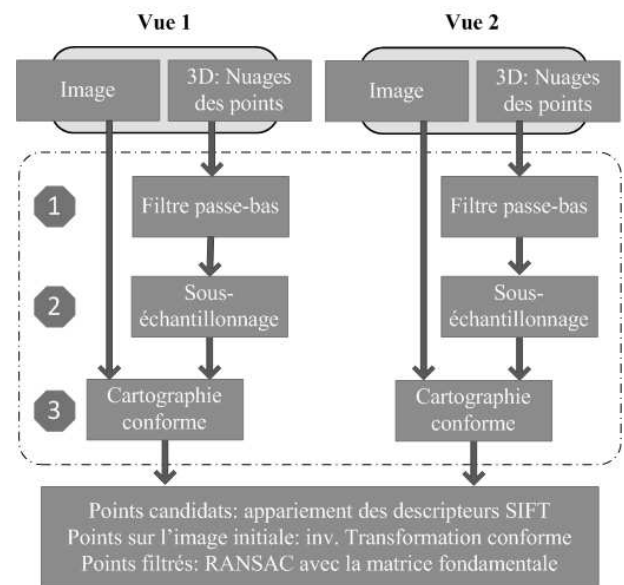


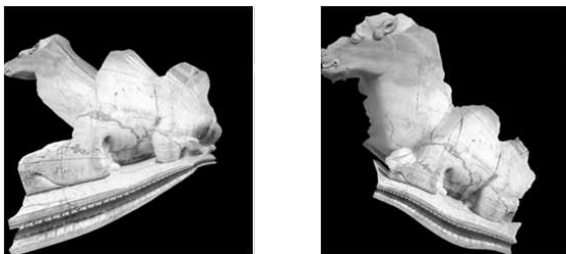
FIGURE 10 – Méthode pour une scène des surfaces en relief.



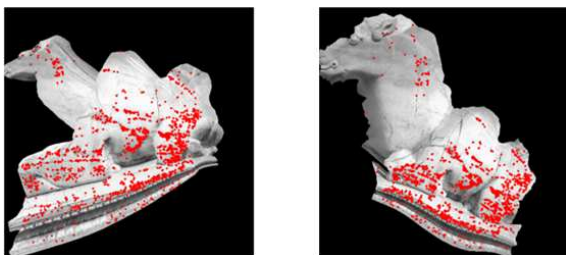
FIGURE 11 – Données en entrée de l’algorithme : un couple d’ images avec leurs profondeurs associées (gauche : vue 1, droite : vue2).



FIGURE 12 – À Gauche : le nuage de points initial (583.958 points). À droite : le nuage de points sous-échantillonné (3.478 points).



a) Les images conformes



b) Les points homologues par SIFT sont en rouge.

FIGURE 13 – Les images sont transformées en coordonnées conformes (gauche : vue 1, droite : vue 2).

4 Résultats

Nous comparons les résultats obtenus en appliquant directement SIFT sur les images initiales (le plus utilisé en pra-

tique) avec les résultats obtenus l’aide de nos approches.

4.1 Cas des surfaces planes

Les résultats sont présentés pour quatre ensembles d’images (borne F, wall, façade, graffiti). La figure 14 présente quelques exemples de résultats (façade, graffiti). En utilisant les images de la figure 4, la table 1 présente le nombre de points homologues obtenus par SIFT et notre approche. Pour toutes les images testées, notre approche «ortho rectifiée» donne les meilleurs résultats.

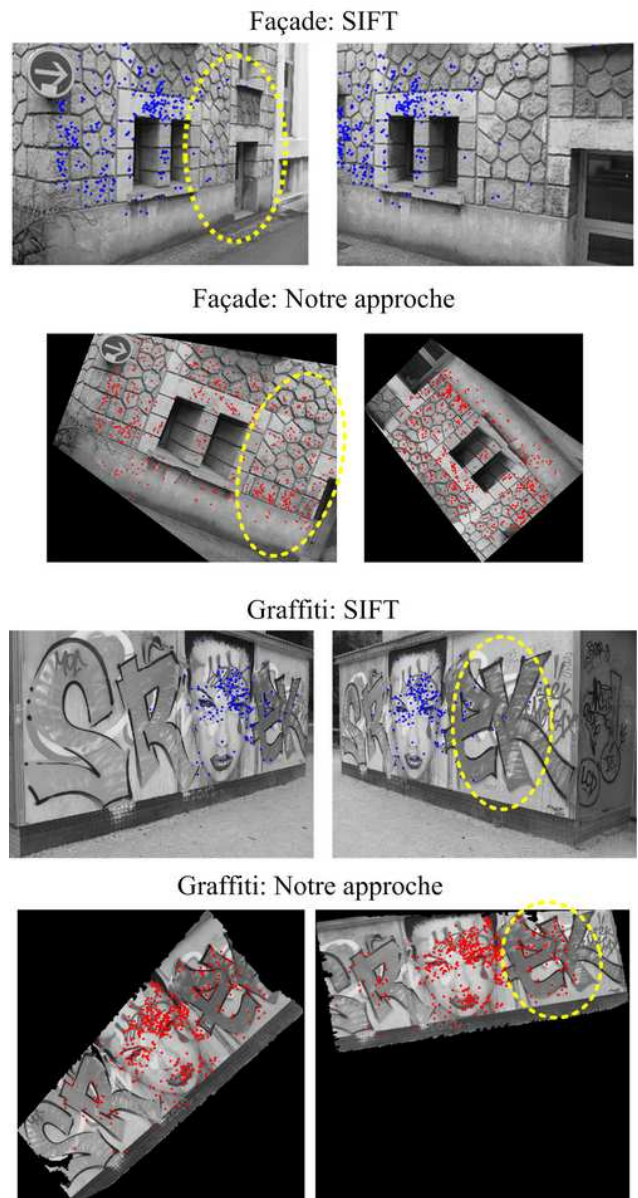


FIGURE 14 – Quelques résultats entre SIFT et notre approche “surface plane”. Nous avons remarqué que les points ne peuvent pas être trouvés par SIFT dans les zones de forte variation géométrique (voir les ellipses jaunes).

TABLE 1 – Comparaison quantitative des deux approches, SIFT et notre approche, pour quatre ensembles d’images. Les points d’intérêt pour chaque image et les points homologues qui sont trouvés entre images correspondantes sont présentés.

Borne F	SIFT	Ortho rectification
#points d’intérêt 1	8.088	5.670
#points d’intérêt 2	6.199	8.107
#points homologues	aucun	599
Wall	SIFT	Ortho rectification
#points d’intérêt 1	2.091	3.074
#points d’intérêt 2	1.267	1.583
#points homologues	43	67
Façade	SIFT	Ortho rectification
#points d’intérêt 1	25.629	12.190
#points d’intérêts 2	22.678	4.942
#points homologues	338	600
Graffiti	SIFT	Ortho rectification
#points d’intérêt 1	1.555	3.103
#points d’intérêt 2	1.974	3.314
#points homologues	190	543

4.2 Cas des surfaces en relief

Dans cette partie, nous regardons les résultats utilisant la cartographie conforme pour les scènes de surfaces en relief (chameau, ballon, bouddha). La figure 15 montre la comparaison des points homologues qui sont obtenus par SIFT (en bleu) et notre approche (en rouge). La répartition des points obtenus par notre approche est plus meilleur que pour SIFT, pour les trois ensembles d’images testées. Le nombre de points trouvés est montré dans la table 2.

TABLE 2 – Comparaison quantitative des deux approches, SIFT et notre approche, pour trois ensembles d’images.

Chameau	SIFT	Conforme
#points d’intérêt 1	67.866	25.712
#points d’intérêt 2	74.278	27.647
#points homologues	577	1.624
Ballon	SIFT	Conforme
#points d’intérêt 1	16.195	17.609
#points d’intérêt 2	16.427	17.183
#points homologues	694	2.141
Bouddha	SIFT	Conforme
#points d’intérêt 1	208.943	43.499
#points d’intérêt 2	261.045	83.186
#points homologues	48	615

5 Conclusion et perspectives

Nous avons présenté deux approches pour obtenir des points homologues entre des images prises avec une base longue en utilisant l’information 3D obtenue par une ca-

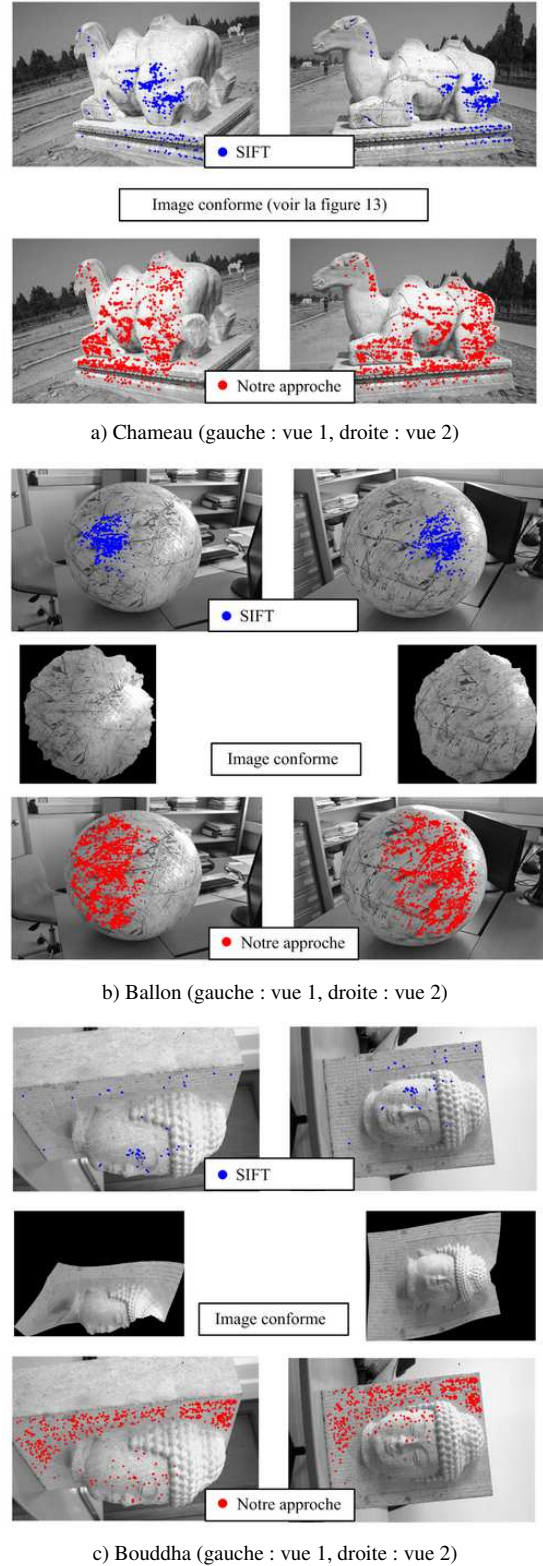


FIGURE 15 – Les résultats sont comparés entre SIFT et notre approche “surface en relief”. La répartition des points obtenus par notre approche (en rouge) est plus meilleur que pour SIFT (en bleu).

méra stéréo ou par multi-stéréoscopie. Dans la première approche, pour une scène composée de surfaces planes, l'image est transformée en une orthoimage rectifiée pour chaque plan. Avec la deuxième approche, pour une scène de surfaces en relief, l'image est transformée par une cartographie conforme. Les résultats de comparaison entre SIFT et nos approches montrent que les points homologues obtenus avec nos approches sont nombreux et bien répartis dans les images.

Ces travaux sont la première étape pour valider nos approches en théorie. Des travaux futurs sont prévus pour :

- La segmentation automatique à partir de l'information de profondeur entre les régions planes, les régions en relief ou les régions bruitées (par exemple les arbres) doit être développée.
- Des méthodes plus sophistiquées que la cartographie conforme, par exemple une contrainte d'isométrie non linéaire, peuvent être appliquées. Ces nouvelles approches seraient utilisables pour les scènes non lisses.

Références

- [1] D. G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, *IJCV*, pp.91–110, 2004.
- [2] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, SURF : Speeded Up Robust Features, *CVIU*, pp.346–359, 2008.
- [3] J. Matas, O. Chum, U. Martin, and T. Pajdla, Robust Wide Baseline Stereo from Maximally Stable Extremal Regions, *BMVC*, pp.384–393, 2002.
- [4] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, A Comparison of Affine Region Detectors, *IJCV*, 2005.
- [5] J. M. Morel and G. Yu, ASIFT : A New Framework for Fully Affine Invariant Image Comparison, *SIAM, on Imaging Sciences*, pp.438–469, 2009.
- [6] P. J. Besl and N. D. McKay, A Method for Registration of 3-D Shapes, *PAMI*, pp.239–256, 1992.
- [7] B. Lévy, S. Petitjean, N. Ray, and J. Maillot, Least Squares Conformal Maps for Automatic Texture Atlas Generation, *ACM Transactions on Graphics*, pp.362–371, 2002.
- [8] C. Wu, B. Clipp, X. Li, J.-M. Frahm, and M. Pollefeys, 3D Model Matching with Viewpoint-Invariant Patches (VIP), *IEEE CVPR*, 2008.
- [9] K. Köser and R. Koch, Perspectively Invariant Normal Features, *ICCV Workshop on 3D Representation for Recognition*, 2007.
- [10] M. Pollefeys, D. Nistér, J.-M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.-J. Kim, P. Merrell, C. Salmi, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewénus, R. Yang, G. Welch, and H. Towles, Detailed Real-Time Urban 3D Reconstruction From Video, *IJCV*, 2008.
- [11] N. Snavely, S. M. Seitz, and R. Szeliski, Photo Tourism : Exploring Image Collections in 3D, *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2006)*, 2006.
- [12] M. Pierrot-Deseilligny and I. Cléry, APERO, An Open Source Bundle Adjustment Software for Automatic Calibration and Orientation of A Set of Images, *ISPRS Commission V*, 2011. (<http://www.micmac.ign.fr>)
- [13] CGAL, Computational Geometry Algorithms Library : <http://www.cgal.org>
- [14] M. A. Fischler and R. C. Bolles, Random Sample Consensus : A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Comm. of the ACM*, pp.381–395, 1981.
- [15] W. Förstner and E. Gülchl, A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features, *ISPRS Conf. on Fast Processing of Photogrammetric Data*, pp.281–305, 1987.
- [16] C. Harris and M. Stephens, A Combined Corner and Edge Detector, in *4th Alvey Vision Conf.*, pp.147–151, 1988.
- [17] S. Wang, Y. Wang, M. Jin, X. Gu, and D. Samaras, Conformal Geometry and Its Applications on 3D Shape Matching, Recognition, and Stitching, *PAMI*, pp.1209–1220, 2007.
- [18] Y. Furukawa and J. Ponce, Accurate, Dense, and Robust Multi-view Stereopsis, *PAMI*, pp.1362–1376, August 2010.
- [19] ANN, Approximate Nearest Neighbor Searching : <http://www.cs.umd.edu/~mount/ANN/>