

Mise en correspondance semi-dense de points de vue éloignés et non calibrés

B. ALBOUY¹, S. TREUILLET², Y. LUCAS³

¹LAIC, IUT de Clermont-Ferrand, 8 rue Jean-Baptiste Fabre – BP 219, 43006 LE PUY EN VELAY Cedex

²LASMEA, Université de Clermont-Ferrand II, 24 avenue des Landais, 63177 AUBIERE Cedex

³LVR, IUT de Bourges, 63 avenue De Lattre de Tassigny, 18020 BOURGES Cedex

benjamin.albouy@iut.u-clermont1.fr

Résumé – Cet article propose une méthode générale de mise en correspondance semi-dense d'une paire d'images couleur prises de points de vue éloignés et non calibrés. Après une initialisation de la géométrie épipolaire et des appariements par le descripteur d'images SIFT, la contrainte épipolaire est récursivement resserrée en vue d'affiner la mise en correspondance. En fin du processus itératif, une étape de densification par corrélation affine permet d'obtenir entre 1733 et 10717 appariements entre deux images de résolution comprise entre 800x600 et 1024x768. Le processus complet s'exécute en environ 2 minutes sur un ordinateur de type Pentium IV à 3GHz, sans optimisation particulière.

Abstract – This article proposes a general semi-dense matching method between two uncalibrated wide-baseline colour images. After initialising the epipolar geometry and matches thank to SIFT image descriptor, the epipolar constraint is recursively tightened in order to refine the matches. At the end of this iterative process, an affine correlation based densification step allows to obtain between 1733 and 10717 matches between two images of resolution included between 800x600 and 1024x768. The whole process runs in around 2 minutes on a 3GHz cadenced Pentium IV computer, without any code optimisation.

1. Introduction

La qualité d'un modèle 3D reconstruit à partir d'images repose essentiellement sur la performance de la mise en correspondance entre les images et la position des prises de vues. Le passage à des points de vue éloignés améliore la triangulation mais change profondément le contenu géométrique et photométrique des images, rendant très délicate la mise en correspondance. La qualité de la reconstruction est alors directement liée à la mise en correspondance des images. Cette difficulté est encore renforcée lorsque la paire stéréoscopique n'a pas été préalablement calibrée, et que l'on ne dispose pas de la contrainte épipolaire. Il devient alors très difficile d'obtenir une densité convenable d'appariements robustes.

Cependant, dans certaines applications, il n'est pas possible de disposer d'une séquence d'images étoffée. Les méthodes de mise en correspondance qualifiées de denses, qui consistent à appairer chaque pixel, ont été développées pour des systèmes stéréoscopiques calibrés [1]. Dans le cas de vues non calibrées, la géométrie épipolaire est généralement retrouvée à partir de correspondances éparées. Ces dernières s'appuient sur des descripteurs locaux qui s'affranchissent des déformations géométriques et colorimétriques entre les deux images [2][3][4][5]. La méthode SIFT [6] a été reconnue comme l'une des plus robustes aux changements de points de vue par une étude comparative très complète [7].

Bien que ces dernières techniques soient suffisantes pour retrouver la géométrie épipolaire, il est nécessaire d'obtenir une mise en correspondance plus dense pour la visualisation du modèle ou des évaluations précises de volume. Pour obtenir un appariement plus dense, Lhuillier propose une technique de propagation à partir d'un ensemble d'appariements initiaux [8]. Le principe est similaire à un algorithme de croissance de région basé sur le score de corrélation, soumis à une contrainte de continuité. Ce type de méthode a été introduit par Otto pour les images de terrain [9].

Dans cet article, nous proposons une méthode générale de mise en correspondance semi-dense d'une paire d'images couleur prises de points de vue éloignés et non calibrés. Le nombre d'appariements robustes est d'abord augmenté grâce à un processus itératif au cours duquel la géométrie épipolaire est affinée progressivement. Les correspondances sont ensuite multipliées par l'application de transformations locales affines dans les images. La première partie de l'article présente le processus itératif, puis la technique d'appariement semi-dense par corrélation affine est décrite dans la seconde partie. Les résultats de cet algorithme sont ensuite donnés dans la troisième section avant de conclure.

2. Augmentation itérative du nombre des correspondances

Le processus itératif de mise en correspondance est initialisé par un ensemble éparé d'appariements robustes, puis la géométrie épipolaire est itérativement renforcée par l'ajout de nouvelles correspondances respectant des contraintes épipolaire, de disparité et de voisinage.

2.1 Initialisation robuste du processus

L'algorithme de mise en correspondance s'initialise en trouvant un nombre limité d'appariements grâce au descripteur SIFT [6] associé à une stratégie de type « winner takes all » croisée. Seul les appariements les moins ambigus sont conservés, c'est-à-dire ceux dont la distance euclidienne entre leurs descripteurs est inférieure à 0,6. Ces appariements initiaux sont utilisés pour calculer une première estimation de la matrice fondamentale régissant la géométrie épipolaire. Ce calcul est effectué grâce à la méthode robuste Least Median of Square (LMedS) permettant de rejeter les outliers de l'ensemble des correspondances, suivie par une minimisation aux moindres carrés orthogonaux sur l'ensemble des inliers [10]. Par ailleurs, pour alimenter le processus itératif, un ensemble de 2000 points d'intérêt est automatiquement détecté dans les images en utilisant le détecteur de coins de Harris et Stephens [11] adapté aux images couleurs par Montesinos et al [2]. Ces deux initialisations réalisées, le processus itératif peut être exécuté.

2.2 Renforcement itératif de la contrainte épipolaire

De nouveaux appariements sont générés en appliquant la contrainte épipolaire déterminée par la matrice fondamentale estimée à l'étape d'initialisation ou à l'itération précédente. Après rectification des images, la contrainte épipolaire est réduite au cas canonique où les points correspondants se trouvent sur les mêmes lignes de pixels dans les images.

L'application de cette contrainte permet de réduire la recherche des correspondants dans une bande rectangulaire autour des droites épipolaires. La largeur de ce rectangle centré sur le point d'intérêt de la première image correspond à une limite de disparité alors que sa hauteur permet de prendre en considération les imperfections de l'estimation de la géométrie épipolaire. Un score de corrélation ZSAD (Zero-mean Sum of Absolute Distance) est calculé sur une fenêtre circulaire de 31 pixels de diamètre pour tous les appariements possibles dans la zone rectangulaire de recherche, puis le meilleur appariement est conservé en appliquant la stratégie « winner takes all » symétrique. Certaines erreurs d'appariement peuvent subsister si les scores de corrélations sont proches. Pour rejeter ces erreurs, nous avons choisi d'appliquer une contrainte de voisinage, qui consiste à vérifier la cohérence des appariements sur les images non rectifiées. Pour chaque correspondance, un vecteur d'appariement moyen est calculé à partir des correspondances situés à moins de

50 pixels, puis le vecteur d'appariement est comparé à ce vecteur moyen. Une tolérance de 25% sur la norme et de $\pi/10$ sur l'argument est acceptée.

En général, l'estimation de la matrice fondamentale et la sélection des inliers ne sont réalisées qu'une seule fois. L'originalité de notre approche est de renforcer progressivement la contrainte épipolaire au fil des itérations. La hauteur habituellement constante du rectangle de recherche épipolaire est alors remplacée par une hauteur linéairement décroissante en fonction de l'indice d'itération. Cette réduction a le double intérêt d'augmenter le nombre de correspondances et d'affiner la géométrie épipolaire. Dans les premières itérations, la géométrie épipolaire estimée à partir de peu d'appariements risque de présenter de fortes erreurs, ainsi une grande zone de recherche est nécessaire. Dans les itérations suivantes, la géométrie épipolaire est de plus en plus robuste et la réduction de la zone de recherche permet de lever les ambiguïtés des scores d'appariements sur un ensemble de candidats plus réduit.

À la fin du processus itératif, un dernier filtrage des appariements robustes est effectué après reconstruction des points 3D [10]. Cette reconstruction métrique utilise une méthode d'autocalibrage [12] applicable dans le cas où les images sont prises avec la même focale. Les points 3D qui violent l'hypothèse d'une surface lisse (pics positifs ou négatifs) sont considérés comme issus de mauvaises correspondances et supprimés de l'ensemble d'appariements.

3. Multiplication du nombre d'appariements par corrélation affine

Le processus itératif précédent offre un ensemble de correspondances robustes mais dont la densité est encore insuffisante pour un bon rendu visuel ou un calcul précis de volume (typiquement entre 200 et 1000 appariements). Pour booster le nombre de correspondances, nous proposons d'appliquer des déformations locales des images afin de rendre plus pertinent le calcul des scores de corrélation. Les déformations sont guidées par les correspondances disponibles en fin du processus itératif.

Une triangulation de Delaunay est effectuée sur les points appariés sur la première image, puis reportée sur la seconde image. Ainsi, nous disposons d'un ensemble de triangles homologues entre les deux images. En supposant que la portion d'objet inclus dans un triangle est plane, alors il existe une homographie H reliant strictement les deux triangles images. Si l'hypothèse du plan n'est pas strictement respectée, alors les erreurs provoquées par H sont liées aux défauts de planéité. Ces erreurs restent d'autant plus faibles que les triangles sont petits. Le calcul de H pourrait être basé sur la reconstruction 3D obtenue précédemment mais pour rendre la méthode plus générale, au cas où le modèle 3D n'est pas disponible, nous avons choisi d'approcher cette homographie par la transformation affine la plus proche. L'estimation de H par une transformation affine ne cause que peu d'erreur,

surtout si l'inclinaison du plan formé par le triangle dans l'espace est faible par rapport au plan image.

Après transformations des triangles homologues pour les placer dans un même cadre géométrique, des points d'intérêts sont détectés dans chacun des triangles par la méthode d'Harris puis appariés en appliquant l'algorithme de relaxation proposé par Zhang [13]. Le score de corrélation est pondéré par une gaussienne d'écart-type 5 pixels pour privilégier les points appartenant au plan moyen du triangle.



FIG. 1 : Images extraites de la base de test.

4. Résultats

Le processus itératif de mise en correspondance a été testé sur un lot de 24 paires d'images couleurs de résolutions comprises entre 800x600 et 1024x768 (FIG. 1). Cette base de données incorpore des images de la base de test de Mikolajczyk et al [7]. Ces images présentent diverses dégradations comme du flou, de forts changements de point de vue, des zooms et rotations. Par ailleurs, des images propres à notre application, présentant peu de texture, ont été utilisées. La performance de la méthode est mesurée par l'évolution de trois indices au cours des itérations : l'augmentation du nombre de correspondances (FIG. 2), le pourcentage d'appariements rejetés par l'estimation robuste de la matrice fondamentale (FIG. 3) et la distance moyenne des points à leurs droites épipolaires (FIG. 4). Les figures représentent les moyennes de ces indices sur les 24 paires d'images. Les barres de tolérance correspondent à plus ou moins l'écart-type des mesures sur les 24 paires.

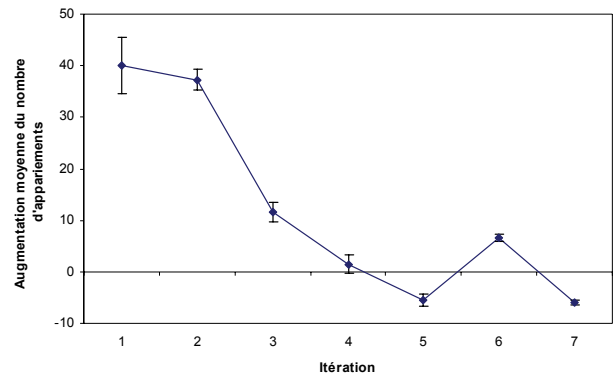


FIG. 2 : Augmentation moyenne du nombre de correspondances au cours des itérations

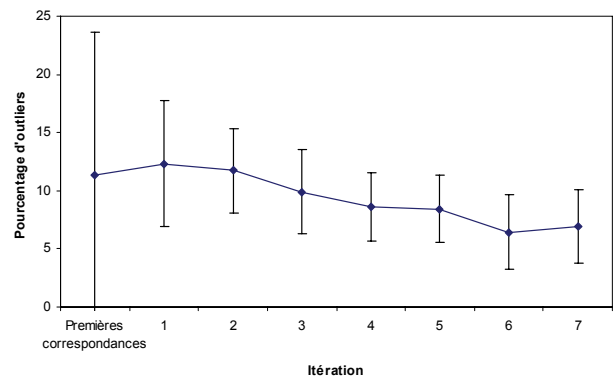


FIG. 3 : Pourcentage moyen d'outliers rejetés au cours des itérations

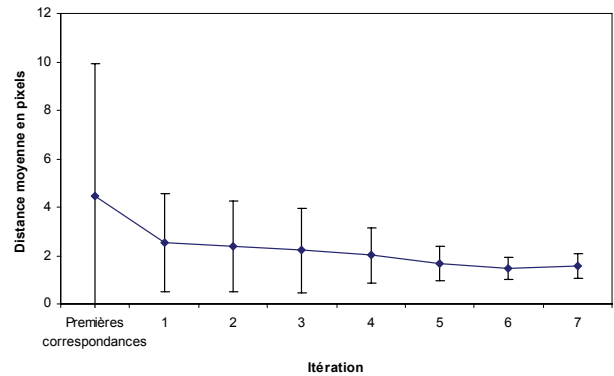


FIG. 4 : Distance moyenne des points aux droites épipolaires au cours des itérations

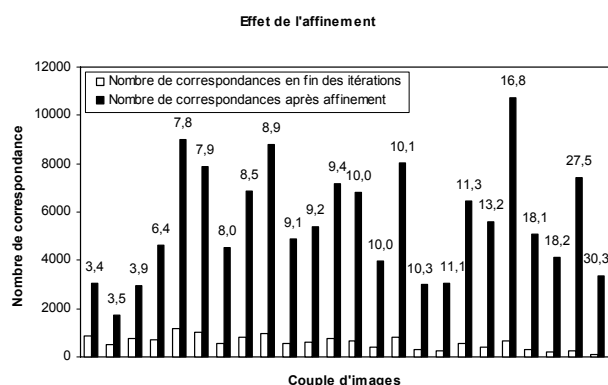


FIG. 5 : Effet de l'affinement sur le nombre de correspondances.

Ces résultats montrent que le nombre de correspondances correctes augmente progressivement au fil de l'affinement de la géométrie épipolaire. Le nombre des correspondances possibles est limité par les 2000 points initialement fournis à l'algorithme itératif. La distance moyenne aux droites épipolaires est bornée par la hauteur du rectangle de la contrainte épipolaire. Dans ces tests, cette valeur est initialisée à 35 puis linéairement réduite au cours des itérations jusqu'à atteindre 5 pixels. La figure 3 prouve que la qualité des appariements est améliorée durant le processus, par le rejet d'appariements de moins en moins nombreux proportionnellement aux correspondances conservées. La figure 5 synthétise le facteur multiplicatif entre le nombre de correspondances à la fin des itérations et après l'étape de raffinement. Ce facteur varie entre 3,4 et 30,3 dans nos tests, fournissant un nombre suffisant d'appariements (entre 1733 et 10717) pour une reconstruction 3D correcte. Enfin, cette méthode s'exécute en un temps raisonnable d'environ 2 minutes sur un PC standard (Pentium IV à 3GHz). Cet algorithme étant facilement parallélisable, le temps de calcul pourrait être divisé par deux sur les machines récentes en tirant profit des nouveaux processeurs à double cœur.

5. Conclusion

Lorsque le niveau de précision requis du modèle 3D est important, la principale difficulté réside dans l'obtention d'un appariement suffisamment dense des images. Jusqu'ici la communauté vision s'est cantonnée aux faibles changements de points de vue appliqués aux têtes stéréoscopiques calibrées. Les progrès en auto-calibrage nous incitent à aborder les points de vue éloignés car ils offrent une meilleure triangulation pour peu que les difficultés d'appariement soient maîtrisées.

Deux améliorations majeures ont été introduites dans notre méthode d'appariement. Tout d'abord, le processus itératif utilise une contrainte épipolaire progressivement resserrée pour aider l'estimation de la matrice fondamentale et la sélection de nouveaux appariements. Ensuite, une transformation affine locale appliquée sur des triangles homologues permet d'obtenir une mise en

correspondance plus dense. L'augmentation du nombre de correspondances robustes permet d'obtenir un modèle 3D plus précis et stable. Bien que cette méthode ne fournisse pas une mise en correspondance aussi dense que celle proposée par [8], elle a l'avantage d'être plus simple et de fournir une mise en correspondance semi-dense sur des points de vue plus éloignés. Les améliorations futures porteront sur la gestion des occultations.

Références

- [1] D. Scharstein et R. Szeliski, *A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms*. IJCV, vol 47, p. 7-42, avril-juin 2002.
- [2] P. Montesinos, V. Gouet et R. Deriche, *Differential Invariants for Color Images*, ICPR, p. 838-840, 1998.
- [3] T. Tuytelaars et L. Van Gool, *Wide Baseline Stereo Matching based on Local, Affinely Invariant Regions*, Proceedings of the 11th British Machine Vision Conference, Bristol (UK), 2000.
- [4] F. Schaffalitzky et A. Zisserman, *Viewpoint Invariant Texture Matching and Wide Baseline Stereo*, ICCV, Vancouver (CA), p. 636-643, 2001.
- [5] J. Matas, O. Chum, M. Urban et T. Pajdla, *Robust Wide Baseline Stereo from Maximally Stable Extremal Regions*, Proceedings of the 13th British Machine Vision Conference, Cardiff (UK), p. 384-393, septembre 2002.
- [6] D. Lowe, *Distinctive Image Feature from Scale-Invariant Keypoints*, IJCV, vol. 60, n° 9, p.91-110, 2004.
- [7] K. Mikolajczyk et C. Schmid, *Performance Evaluation of Local Descriptors*, PAMI, vol. 27, n° 10, p. 1615-1630, 2005.
- [8] M. Lhuillier et L. Quan, *Match Propagation for Image-Based Modelling and Rendering*, PAMI, vol. 24, n° 8, p. 1140-1146, 2002.
- [9] G. Otto et T. Chau, *A Region Growing for Matching of Terrain Images*, Image & Vision Comp., vol. 7, n° 2, p. 83-94, 1989.
- [10] B. Albouy, S. Treuillet, Y. Lucas et D. Birov. *Fundamental matrix estimation revisited through a global 3D reconstruction framework*, ACIVS – Bruxelles, p. 185-192, 31 août – 3 sept. 2004.
- [11] C. Harris et M. Stephens, *A Combined Corner and Edge Detector*, 4th Alvey Vis. Conf., Manchester (UK), p. 147-151, 1988.
- [12] P. Sturm, *On Focal Length Calibration from Two Views*, CVPR, Kauai, Hawaï, vol. 2, p. 145-150, 8-14 déc. 2001.
- [13] Z. Zhang, R. Deriche, O. Faugeras et Q.-T. Luong, *A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry*, Rapport de Recherche INRIA, n° 2273, mai 1994.