

Maillage 3D de scènes complexes à partir de séquences non calibrées*

Lionel Oisel, Luce Morin, Claude Labit

IRISA / INRIA Rennes
Campus de Beaulieu
35042 Rennes Cedex, France
fax: (+33) 99.84.71.71,

e-mail: loisel@irisa.fr, lmorin@irisa.fr, labit@irisa.fr

1 Introduction

Les travaux que nous présentons se placent dans le cadre d'un contrat CTI-CNET dont l'objectif est l'obtention d'informations 3D à partir d'une séquence d'images non calibrées ne contenant pas d'objets en mouvement. Ces hypothèses recouvrent un large spectre de séquences possibles allant de la séquence stéréoscopique à des séquences monoculaires quelconques mais dans lesquelles les objets en mouvement auront été segmentés et supprimés par un algorithme de détection de mouvement [OB95] (cf fig 1). A partir d'informations sur le mouvement apparent dans les images, la scène peut être reconstruite par triangulation, soit de manière projective (sans calibration euclidienne), soit de manière 3D euclidienne (après une étape d'auto-étalonnage). Cette représentation permet alors le rendu en temps réel pour des applications telles que la navigation interactive.

Des travaux se rapportant à ce sujet ont déjà été réalisés. Deux approches sont généralement proposées utilisant des informations 3D différentes :

- approche dense : ici un champ de disparité dense est calculé. Ainsi, connaissant l'ensemble des points se correspondant dans deux images (ou plus), une triangulation en chaque point est effectuée [LP97][BM97]. Cette méthode présente deux inconvénients majeurs. Le premier réside dans la difficulté à obtenir une information dense et cohérente (problème des zones d'occultation, des zones uniformes ...). Le deuxième problème réside dans le coût de calcul associé à la triangulation rendant le rendu difficilement calculable en temps réel.

*Cette recherche est menée avec le soutien et dans le cadre de l'action CTI-CNET n° 96ME06 réunissant l'INRIA et L'INT intitulée: Modélisation et représentation hiérarchique de scènes 3D pour des services multimédia. La coordination scientifique de cette action CNET est assurée par M P. Leray du CNET/DIH, Laboratoires du CCETT.



FIG. 1 – *Images originales et segmentées*

- approche éparses : l'information 3D est ici décrite par un ensemble limité de points en correspondance. Ces points qui caractérisent les sommets des différents plans 3D de la scène sont généralement extraits de façon semi-automatique [BR97]. Le modèle 3D de la scène est alors un modèle par facettes permettant des manipulations très rapides de l'information. Ces approches sont particulièrement pertinentes pour la reconstruction de bâtiments où le nombre de sommets à extraire est limité. Par contre dans le cas de scènes complexes, le coût de l'intervention humaine devient prohibitif.

En raison de la contrainte de temps réel que nous nous sommes fixée, notre travail s'inscrit dans le deuxième groupe de méthodes. Afin de pouvoir traiter des scènes complexes, nous cherchons ici à automatiser la phase de segmentation en facettes planes. L'approche utilisée consiste à trouver des régions caractérisées par un modèle de mouvement à partir de couples d'images, chaque région estimée correspondant à une facette plane de l'espace.

Cet article va donc dans un premier temps présenter le modèle de mouvement retenu ainsi qu'une brève discussion sur les méthodes de résolutions envisagées. Le deuxième point portera alors sur la méthode retenue qui sera suivie de divers résultats.

2 Le modèle de mouvement planaire

Afin de manipuler de manière rapide un modèle 3D, nous cherchons à segmenter en facettes planes deux ou plusieurs images. Il existe alors une contrainte géométrique liant un point d'une facette de l'image droite à son correspondant dans l'image gauche. Cette contrainte s'exprime par un modèle mathématique (homographie) à huit paramètres. Segmenter en facettes planes équivaut donc à segmenter au sens d'un mouvement homographique. Tout point d'une région vérifie alors le même modèle de mouvement. Le déplacement apparent en chaque point d'une région s'écrit alors :

$$\begin{cases} du' = \frac{-h_{31}u^2 - h_{32}uv + h_{11}u + h_{12}v + h_{13} - h_{33}u}{h_{31}u + h_{32}v + h_{33}} \\ dv' = \frac{-h_{32}v^2 - h_{31}uv + h_{21}u + h_{22}v + h_{23} - h_{33}v}{h_{31}u + h_{32}v + h_{33}} \end{cases} \quad (1)$$

Une première technique consiste à segmenter et à estimer les paramètres homographiques de chaque région simultanément [OB95]. Les modèles généralement

utilisés sont affines (6 paramètres et linéaires) ou quadratiques (8 paramètres). Ces modèles ne sont valides que sous certaines conditions [Fra91] d'orientation et de distance du plan 3D au plan caméra. Pour un modèle homographique, le critère à minimiser (DFD) est non linéaire en fonction des paramètres du modèle. Ceci rend l'utilisation de cette première technique d'un coup prohibitif.

Nous avons donc opté pour une deuxième approche divisée en deux étapes distinctes. Dans un premier temps, le mouvement (disparité) est estimé en chaque point. Suit alors la phase d'estimation des différents modèles compatibles avec l'information dense.

3 Méthode proposée

En raison de sa complexité, l'estimation du modèle homographique est précédée d'une phase d'estimation du champ dense de mouvements. La technique de calcul d'un champ dense robuste et régularisé utilisée sera présentée dans un premier temps. Nous montrerons alors comment nous exploitons les différentes informations fournies par cette phase préliminaire afin d'obtenir une triangulation de l'image en facettes planes de l'espace.

3.1 Calcul d'un champ dense

Nous avons donc développé un algorithme dérivé de l'estimation du flot optique permettant d'obtenir un champ estimé robuste et régularisé [MP96]. Par cet aspect, il rejoint certaines techniques déjà proposées [RD95] [DS96]. L'originalité de notre contribution réside dans l'utilisation de la géométrie épipolaire couplée à un schéma multirésolution mêlant méthodes différentielles et discrètes qui semblent a priori incompatibles. Le champ dense vérifiant la géométrie épipolaire, il est géométriquement cohérent avec le modèle de projection perspective. L'utilisation d'un schéma multirésolution contraint permet d'assurer la convergence de l'algorithme pour un gain de temps important.

3.1.1 Principe général

L'hypothèse sur laquelle se base notre méthode est que les projections dans deux images d'un même point 3D ont la même intensité. Cela se traduit par la nullité de la DFD (Displaced Frame Difference) définie comme suit :

$$DFD(s, ds) = I_1(s) - I_2(s + ds) = 0, \quad (2)$$

où $I_i(s)$ représente l'intensité lumineuse dans la i^{e} image au point $s = (x, y)$ et $ds = (dx, dy)$ est le déplacement d'une image à l'autre d'un point matériel le long des axes x et y .

En utilisant l'hypothèse de rigidité de la scène, nous pouvons alors utiliser la relation épipolaire afin de reformuler cette DFD [LF95]. Cette relation qui peut s'exprimer sous forme matricielle par la matrice fondamentale notée F fournit une

droite de correspondants potentiels pour un point donné. Le déplacement se décompose alors sous la forme d'un vecteur normal \vec{N}_s et d'un vecteur tangent \vec{V}_s à la droite épipolaire associée à s (voir figure ??) . La DFD devient alors :

$$DFD(s, ds) = I_1(s) - I_2(s + \vec{N}_s + \lambda_s \vec{V}_s) = 0 \quad (3)$$

Le problème initial de recherche d'un champ de déplacement 2D est donc réduit à un problème 1D d'estimation d'abscisse λ_s le long des droites épipolaires.

3.1.2 Schéma multirésolution

La matrice fondamentale ne peut être correctement estimée que pour des déplacements importants entre deux prises de vues. À l'inverse, l'équation (3) est résolue en effectuant une linéarisation par développement limité par rapport à λ_s , ce qui suppose des déplacements faibles le long de la droite épipolaire. Afin de pouvoir coupler ces approches a priori incompatibles, un schéma multirésolution a été développé. À un niveau k de la pyramide, la disparité λ_s^k le long de la droite épipolaire est décomposée en une disparité λ_s^{k-1} , issue de la projection de l'estimation réalisée au niveau de plus basse résolution $k-1$ et d'un incrément $d\lambda_s^k$ à estimer. L'équation (3) s'écrit maintenant pour un niveau k donné :

$$I_1^k(s) - I_2^k(s + \vec{N}_s^k + \lambda_s^{k-1} \vec{V}_s^k + d\lambda_s^k \vec{V}_s^k) = 0 \quad (4)$$

où l'inconnue est à présent $d\lambda_s^k$.

Les pyramides d'images I_1^k et I_2^k sont construites à partir des images originales vers le niveau de résolution le plus faible par filtrage puis sous-échantillonnage. La matrice fondamentale est calculée en utilisant la méthode de la parallaxe virtuelle [BM95] à partir de points automatiquement extraits et appariés dans les deux images originales. Les matrices F^k sont alors calculées pour chaque niveau par changement de base. L'ensemble des F^k nous permet alors de calculer les pyramides de vecteur \vec{N}_s^k, \vec{V}_s^k .

3.1.3 Méthode d'estimation régularisée

Nous nous plaçons maintenant à un niveau de résolution donné k . Pour des raisons de clarté l'indice k sera omis dans les équations qui suivent.

L'équation (4) est linéarisée vis à vis de l'inconnue $d\lambda_s$ autour de $s + \vec{N}_s + \lambda_s \vec{V}_s$. $d\lambda_s$ est considéré comme une réalisation d'un champ de Markov aléatoire. Le meilleur champ de disparités en accord avec le critère Bayésien du *M.A.P.* (Maximum A Posteriori) revient au problème de minimisation globale suivant :

$$\widehat{d\lambda} = \arg \min_{d\lambda \in \mathbb{R}} H(d\lambda) = \arg \min_{d\lambda \in \mathbb{R}} (H_1(d\lambda) + \alpha[H_2(d\lambda)]) \quad (5)$$

où α est une constante réelle ayant pour but d'équilibrer les deux termes énergétiques.

Le terme H_1 est le terme d'énergie lié aux observations. Il provient directement de la linéarisation de la DFD. Le terme H_2 est le terme de lissage qui vise à favoriser

des vecteurs déplacements similaires d_s et d_r pour toute paire $\langle s, r \rangle$ de positions voisines .

Cependant afin de pouvoir tenir compte des erreurs liées au non respect de la DFD (resp. pour autoriser des discontinuités dans le champ de disparité), un M-estimateur est associé à H_1 (resp H_2). Sous certaines conditions [BA92], H_1 et H_2 peuvent être exprimées comme des fonctions quadratiques en $d\lambda_s$. La contribution d'un point s à H_1 est pondérée par un facteur δ_s : plus sa contribution énergétique est forte, plus δ_s est faible. Il en va de même pour le terme de lissage mais avec une pondération par une quantité β_{sr} dépendant de la différence entre les vecteurs déplacements $\|d_s - d_r\|$.

À un niveau de résolution donné, un schéma itératif de Gauss-Seidel est mis en œuvre pour résoudre le problème de minimisation. La minimisation est effectuée alternativement sur le champ de disparités $d\lambda_s$ et sur le champ des poids δ_s et β_{sr} . En raison du caractère local de la minimisation, une bonne initialisation du champ de disparité est nécessaire. Celle-ci est réalisée par interpolation des vecteurs déplacement issus des paires de points d'intérêt extraits pour le calcul de la géométrie épipolaire.

3.2 Segmentation en facettes planes

La phase de segmentation comporte deux opérations alternées : la triangulation et le calcul de l'homographie associée a chaque triangle. Chaque triangle est alors redécoupé si le modèle homographique ne correspond pas au champ dense préalablement calculé.

3.2.1 Estimation des homographies

L'estimation de l'homographie est effectuée à partir des paires de points en correspondance sur chaque facette en utilisant une méthode proposée par Robert [RF95]. Cette méthode présente l'intérêt de prendre en compte la contrainte épipolaire (exprimée par la matrice $F(3 \times 3)$) pour calculer la matrice homographique à partir d'au moins trois paires de points en correspondance.

Soient $H(3 \times 3)$ la matrice homographique homogène, m_d^i et m_g^i les i^e points correspondants de l'image droite et gauche. Une condition nécessaire pour que H soit compatible avec la géométrie épipolaire est que la matrice symétrique homogène $F^T.H + H^T.F$ soit nulle. Cette relation nous donne 6 équations indépendantes. Chaque paire de point en correspondance nous fournit une équation scalaire supplémentaire :

$$[m_d^i, Fm_g^i, Hm_g^i] = 0$$

où $[a, b, c]$ définit le triple produit. L'ensemble de ces équations nous donne un système où les inconnues à déterminer sont les coefficients h_{ij} de la matrice H :

$$\begin{cases} F^T.H + H^T.F & = 0 \\ [m_d^i, Fm_g^i, Hm_g^i] & = 0 \\ \vdots & \end{cases} \quad (6)$$

Ce système peut alors être réécrit sous la forme $Ah = 0$. Le vecteur h contenant les inconnues est calculé par décomposition en valeurs singulières de A^tA .

3.2.2 Triangulation itérative

Afin d'initialiser la maillage, une triangulation des points d'intérêt extraits pour le calcul de la géométrie épipolaire est effectuée. Pour chaque triangle l'homographie est calculée sur l'ensemble des points du triangle. Cependant, pour ne pas prendre en compte des données non fiables, seules sont conservées les paires de points qui ne s'éloignent pas trop du modèle d'illumination constant (i.e. dont la disparité ne provient pas uniquement du terme de lissage). Le triangle est alors découpé si un critère C de distance entre le champ dense et le modèle calculé est inférieur à un seuil ϵ donné. Ce critère prend en compte les informations de discontinuités provenant de l'estimateur robuste associé au terme de lissage, de la distance euclidienne entre le correspondant trouvé par l'homographie et celui trouvé par le champ dense :

$$C(\sum_{\langle s,r \rangle} \beta_{sr}, \|Hm_g - (m_g + \vec{d}_g)\| + \|H^{-1}(m_g + \vec{d}_g) - m_g\|) < \epsilon$$

où $\|\cdot\|$ représente la norme.

4 Résultats

Nous présentons ici les résultats de notre algorithme sur une séquence d'images prise par un camescope grand public sans aucune connaissance sur la calibration. Après la phase de calcul de la géométrie épipolaire, il restait 121 paires de points en correspondance. La figure 2 présente les images originales ainsi que les résultats provenant du calcul de champ dense. Il est à noter qu'en raison de l'importance des valeurs de déplacement que nous traitons, les algorithmes classiques de calcul de champ dense (sans contrainte épipolaire) ne convergent pas. La carte de discontinuités montrent des zones de discontinuités importantes au niveau de la table et de l'armoire. La carte d'observation quant à elle met en évidence les zones d'occultations ou de fort gradient (où la DFD n'est pas vérifiée).

La figure 3 présente la triangulation initiale réalisée sur les points d'intérêt extraits. Pour chaque triangle, la matrice homographique a été calculée par la méthode décrite dans le paragraphe 3.2.1 en prenant les trois sommets et la matrice fondamentale. Chaque triangle a ensuite été reconstruit en appliquant la matrice en chaque point et en effectuant une interpolation bilinéaire. L'image c de la figure 3 montre la triangulation en sortie du schéma itératif. On peut constater que le raffinement est plus important dans les zones d'occultation (près de la table) que sur le mur du fond. L'image reconstruite à partir de cette triangulation montre bien le gain de qualité par rapport à la reconstruction précédente. La qualité de la reconstruction peut d'ailleurs se vérifier sur l'image d'erreur résultante.

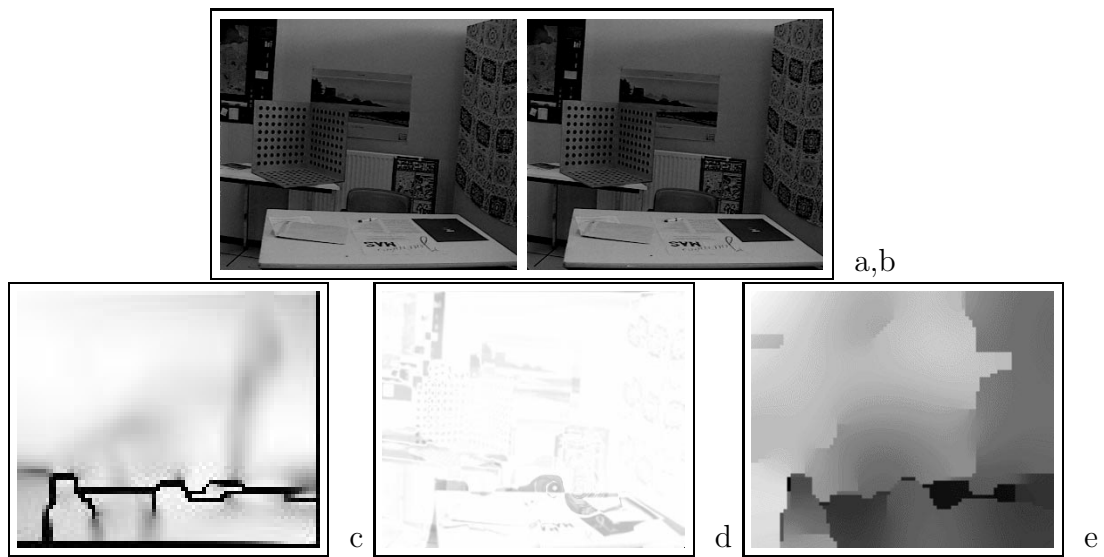


FIG. 2 – *a,b* : images originales droites et gauches - *c,d* : carte de discontinuités et d'observations (plus la zone est sombre, plus faible est la pondération de l'énergie) - *e* : carte de disparité

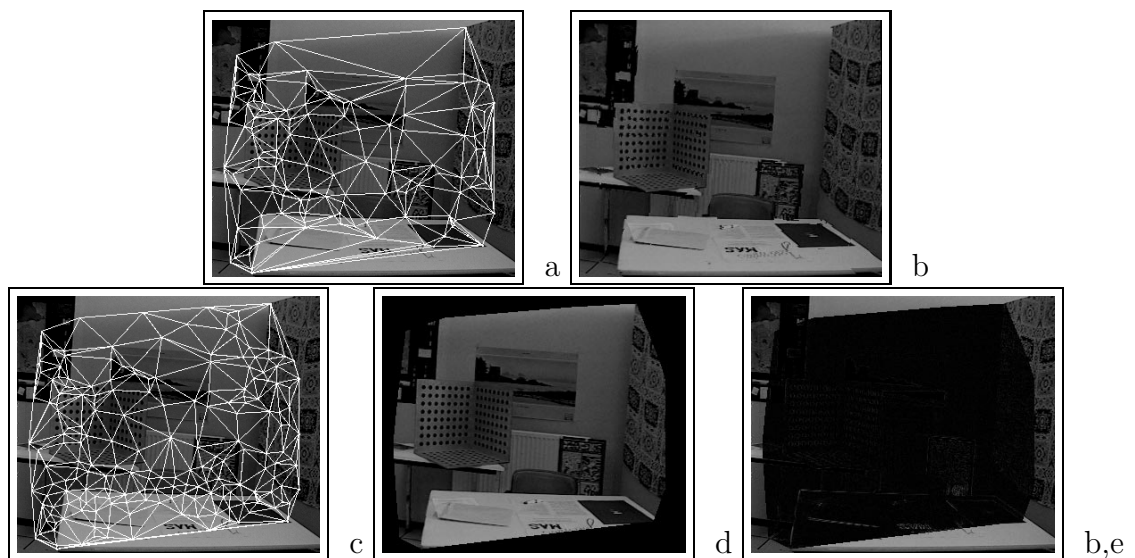


FIG. 3 – *a* : triangulation initiale - *b* : reconstruction à partir des sommets - *c* : triangulation finale - *d* : reconstruction finale - *e* : image d'erreur

5 Conclusion et Perspectives

Nous venons de présenter une méthode de segmentation en facettes planes pouvant s'appliquer sur des scènes complexes. Nous travaillons actuellement sur la phase de reconstruction. Cette reconstruction pourra être projective si l'on ne s'intéresse qu'au rendu 2D ou euclidienne si l'objectif est de remonter à une information 3D euclidienne de la scène (description au format VRML par exemple).

Références

- [BA92] M. Black and P. Anandan. Robust incremental optic flow. In *Proceedings of the Conf. on Computer Vision and Pattern Recognition*, 1992.
- [BM95] B. Boufama and R. Mohr. Epipole and fundamental matrix estimation using the virtual parallax property. In *Proceedings of the International Conf. on Computer Vision*, pages 1030–1036, Cambridge, Massachusetts, 1995.
- [BM97] J. Blanc and R. Mohr. Towards fast and realistic image synthesis from real views. In *Scandinavian Conference on Image Analysis*, Finland, 1997.
- [BR97] S. Bougnoux and L. Robert. Totalcalib: a fast and reliable system for off-line calibration of images sequences. In *Proceedings of the Conf. on Computer Vision and Pattern Recognition*, 1997.
- [DS96] J. Demarty and F. Schmitt. Reconstruction 3d dense à partir de séquences d'images. In *Journées ORASIS 96*, Clermont-Ferrand, 1996.
- [Fra91] E. François. *Interprétation qualitative du mouvement à partir d'une séquence d'images*. PhD thesis, IFSIC-IRISA, Rennes, France, 1991.
- [LF95] Quang-Tuan Luong and Olivier Faugeras. The fundamental matrix: theory, algorithms, and stability analysis. *International Journal on Computer Vision*, 17(1):43–76, January 1995.
- [LP97] G. Lemestre and D. Pelé. Analyse de scènes 3d pour des services de télécommunication. In *CORESA*, France, 1997.
- [MP96] E. Memin and P. Perez. Robust discontinuity-preserving model for estimating optical flow. In *Proceedings of 13th International Conf. on Pattern Recognition*, pages 920–924, 1996.
- [OB95] J.M Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(4):348–365, Décembre 1995.
- [RD95] L. Robert and R. Deriche. Dense depth map reconstruction using a multiscale regularization approach which preserves discontinuities. In *Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging*, pages 32–39, Septembre 1995.
- [RF95] L. Robert and O. Faugeras. Relative 3-D positioning and 3-D convex hull computation from a weakly calibrated stereo pair. *Image and Vision Computing*, 13(3):189–197, 1995.