

# Segmentation basée mouvement 3D pour la détection d'objets indépendants

J. Delhumeau

L. Morin

S. Pateux

IRISA (Institut de Recherche en Informatique et Systèmes Automatiques)

Campus universitaire de Beaulieu, 35042 Rennes cedex FRANCE

{jdelhume lmorin spateux}@irisa.fr

## Résumé

Nous proposons une méthode pour l'extraction de modèles 3D à partir de séquences vidéos monoculaires, sans connaissance à-priori sur la scène. Les méthodes utilisées dans ce domaine font généralement l'hypothèse d'une scène rigide. Notre méthode réduit cette contrainte à une scène contenant plusieurs objets rigides indépendants. Les différentes matrices fondamentales sont calculées à l'aide d'une méthode éparsée. Une segmentation dense du champ de mouvement en est déduite, ce qui permettra une reconstruction indépendante de chaque objet.

## Mots clefs

analyse de séquences vidéo, segmentation d'objets, estimation de mouvement 2D, mouvement 3D rigide, matrice fondamentale.

## 1 Introduction

L'extraction de modèles 3D d'objets à partir de scènes réelles filmées à l'aide d'une simple caméra est un thème ouvrant de nombreuses perspectives d'applications. Ainsi, à partir de ces modèles, il est possible de coder à très bas débit une vidéo, d'augmenter le réalisme via la possibilité d'une navigation contrainte dans la scène, de proposer un rendu en relief, ou bien encore d'insérer des objets synthétiques ou naturels issus d'autres sources [1].

La littérature propose de nombreuses solutions pour la segmentation selon le mouvement 3D ou la reconstruction 3D à partir de vidéos. Certaines utilisent des dispositifs à plusieurs caméras pour segmenter [2], ou reconstruire à partir d'une simple segmentation [3]. Notre but est à la fois de segmenter et de reconstruire à partir de séquences vidéos à une seule caméra. Il y a alors une ambiguïté : Une déformation ou un mouvement de l'objet peuvent aboutir au même mouvement apparent. On peut faire l'hypothèse que la scène se déforme [4], mais dans ce cas un modèle géométrique ou structural de l'objet observé est nécessaire. L'hypothèse duale, selon laquelle l'objet est rigide et a un mouvement 3D ne nécessite pas de modèle, et est plus fréquemment faite [1, 5, 6]. Dans le problème étudié, on fera l'hypothèse de plusieurs objets indépendants, mais dont le mouvement est rigide. Dans ce cadre, les travaux de G.

Adiv [7], ou de J. Weber et J. Malik [8], sont basés sur des méthodes de fusion de régions à partir de champs de mouvement denses. Ceux de E. Steinback [9] proposent une solution itérative dans laquelle sont combinés l'estimation du champ de mouvement dense et l'estimation des paramètres de mouvement 3D. D'autres approches sont basées sur un champ de mouvement éparsé [10, 11], mais n'aboutissent pas à une segmentation dense. Notre méthode peut se décomposer en deux étapes : le calcul des paramètres de mouvement 3D, ainsi que la séparation des différents objets, basés sur un champ de mouvement éparsé ; la segmentation finale qui s'effectue sur un champ dense (de manière à pouvoir effectuer une reconstruction des objets).

Après avoir rapidement présenté les concepts géométriques de base sur lesquels est basé notre travail, nous présenterons notre méthode d'identification des matrices fondamentales, puis de segmentation dense.

## 2 Bases théoriques

Nous présenterons rapidement la géométrie épipolaire, qui sert de base à la reconstruction 3D, et à l'identification des objets indépendants. Elle permet de formaliser les systèmes projectifs à deux caméras, en énonçant les relations liant les deux projetés d'un même point 3D. Intuitivement, la contrainte épipolaire est la suivante : pour chaque point de l'une des deux images, la droite sur laquelle se trouve son correspondant sur l'autre image est connue (cf. Figure 1).

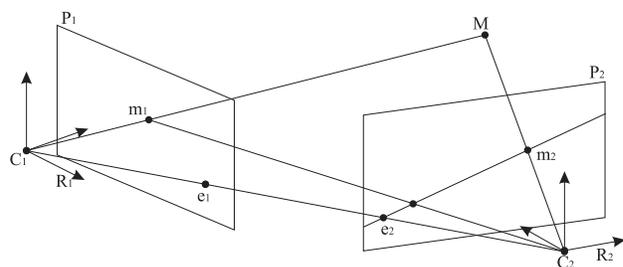


Figure 1 – Le modèle épipolaire

Dans la suite, on notera  $m$  le projeté de  $M$  sur le plan image,  $R_C$  le repère caméra et  $R_0$  le repère scène.

## 2.1 Projection perspective

Elle s'exprime de manière linéaire en utilisant les coordonnées homogènes ( $M_{R_C} = (X, Y, Z, \Omega)^T$  et  $m = (x, y, \omega)^T$ ):

$$m = \begin{pmatrix} 0 \\ \mathbf{A} & 0 \\ 0 & 0 \end{pmatrix} M_{R_C}$$

où la matrice  $\mathbf{A}$  regroupe les paramètres intrinsèques de la caméra.

Par ailleurs en coordonnées homogènes, le changement de repère euclidien entre  $R_0$  et  $R_C$  s'écrit ainsi :

$$M_{R_C} = \begin{pmatrix} \mathbf{R} & t \\ 0 & 0 & 0 & 1 \end{pmatrix} M_{R_0}$$

avec  $\mathbf{R}$  matrice unitaire de rotation et  $t$  vecteur de translation.  $\mathbf{R}$  et  $t$  sont appelés paramètres extrinsèques de la caméra.

On a donc, en regroupant l'ensemble :

$$m = \mathbf{P} M_{R_0}, \text{ avec } \mathbf{P} = \mathbf{A}(\mathbf{R}|t)$$

## 2.2 Matrice fondamentale

Si le repère de la première caméra est choisit comme repère scène, on peut écrire (avec  $p_1$  et  $p_2$  correspondant respectivement à  $m_1$  et  $m_2$  en coordonnées pixel) :

$$\begin{cases} m_1 = \mathbf{P}_1 M = \mathbf{A}_1(\mathbf{I}|0)M \\ m_2 = \mathbf{P}_2 M = \mathbf{A}_2(\mathbf{R}|t)M \end{cases}$$

On peut montrer qu'il existe une matrice  $\mathbf{F}$  telle que :

$$m_2^T \mathbf{F} m_1 = 0$$

Cette matrice est appelée matrice fondamentale et dépend de  $\mathbf{A}$  et du déplacement (défini par  $\mathbf{R}$  et  $t$ ) entre les deux caméras. Elle est de rang 2 et possède sept degrés de liberté.

## 2.3 Utilisation

Les matrices fondamentales permettent de caractériser la géométrie épipolaire, soit le mouvement 3D de la caméra par rapport à un objet. Elles peuvent se calculer à partir d'au moins 8 points mis en correspondance. Pour nous, le but de ce calcul est double :

- retrouver la géométrie épipolaire de façon robuste (ce qui permettra de reconstruire les objets)
- identifier les différents objets indépendants en identifiant les différentes matrices.

En effet, le mouvement relatif de la caméra par rapport à deux objets indépendants est différent, et est donc représenté par deux matrices fondamentales, qui imposent des contraintes différentes sur le mouvement des points. Chaque point peut alors être attribué à l'objet dont il satisfait la contrainte épipolaire.

## 3 Calcul du champ de mouvement

Le calcul du champ de mouvement est une étape essentielle pour le calcul du mouvement 3D. En effet, de la précision des mises en correspondances de points va dépendre à la fois la qualité des estimations des matrices fondamentales et la précision de la segmentation. Nous avons opté pour l'utilisation d'un estimateur par maillage 2D déformable [12], qui consiste en une minimisation d'EQM par un maillage triangulaire dont les sommets sont mobiles. Le maillage est uniforme sur l'image  $I_t$ , l'estimation porte alors sur le mouvement de chaque sommet vers l'image  $I_{t+1}$ . En chaque point, le mouvement est donné par :  $v(s) = \sum_j \alpha_j(s)v(j)$  où  $\alpha_j(s)$  représente les coordonnées barycentriques de  $s$  associées au sommet  $j$ . L'erreur quadratique est alors minimisée en ajustant les  $v(j)$ , soit le mouvement des sommets. Cette minimisation est effectuée via une méthode hiérarchique multirésolution. La méthode par maillage assure une continuité du mouvement, et permet une bonne estimation du mouvement dans les zones uniformes, contrairement à une méthode par flot optique par exemple. Ce champ de mouvement dense sera utilisé pour la segmentation. En revanche, pour l'estimation des matrices fondamentales, il est préférable de n'utiliser que les sommets du maillage, car le mouvement est localement affine. Ces sommets seront décimés en utilisant un détecteur de Harris et Stephens [13] pour ne pas utiliser les points situés sur des zones uniformes. Les points que l'on conserve seront appelés *points d'intérêt*.

## 4 Identification des différentes matrices fondamentales

Nous ne faisons pas d'hypothèses sur la profondeur de la scène (telle que l'hypothèse de perspective faible), qui permettent de réduire le nombre de paramètres à estimer, mais imposent de fortes contraintes sur la géométrie de la scène, qui ne sont pas nécessairement réalisées, en particulier dans le cas où la caméra se déplace à l'intérieur de la scène. Pour le calcul des matrices fondamentales une méthode existante sera utilisée : le calcul aux moindres carrés médians. Après l'avoir présentée, nous expliquerons son intérêt dans ce cas précis.

A titre d'illustration, nous présenterons nos résultats sur une séquence résultant de la composition de la séquence *flower garden* (déplacement latéral) et de la séquence *rue* (déplacement vers l'avant).

### 4.1 Méthode des moindres carrés médians

L'utilisation de la méthode des moindres carrés médians (LMedS) pour le calcul de matrices fondamentales a été proposé dans [6]. Dans le cadre de l'analyse de scènes rigides, les moindres carrés médians permettent d'effectuer un calcul aux moindres carrés (qui permet d'avoir une bonne estimation malgré l'erreur supposée gaussienne d'estimation de mouvement) tout en ignorant les points dont le mouvement est totalement erroné (*faux match*), qui

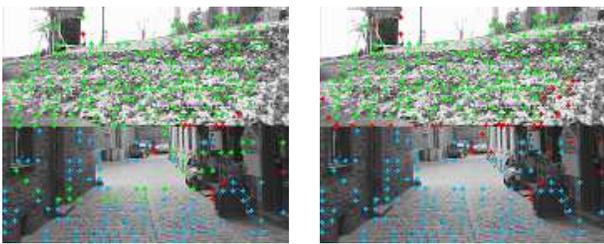
biaiseraient le calcul. Pour cela, la matrice qui minimise la médiane des erreurs (ne tenant donc pas compte de la moitié des points) est tout d’abord calculée. Les points situés au delà de 2,5 fois l’écart type attendu pour une distribution gaussienne (*outliers*) sont ensuite exclus, puis un calcul aux moindres carrés est effectué sur les points restants, pour cela, la méthode de la parallaxe virtuelle [14] sera utilisée.

## 4.2 Dans notre cadre

Cette méthode est utilisée pour séparer les différentes matrices fondamentales : si plusieurs objets sont présents dans la scène et qu’une estimation de matrice fondamentale est effectuée sur l’ensemble des *points d’intérêt*, la matrice correspondant à l’objet le plus important en terme de nombre de *points d’intérêt* est obtenue (on fait l’hypothèse qu’il regroupe au moins la moitié de ceux-ci), tous les *points d’intérêt* appartenant à des objets plus petits sont alors partie des *outliers* et sont exclus du calcul, de même que les *faux match*.

En itérant l’estimation sur les *outliers*, l’objet suivant par ordre de taille est obtenu, et ainsi de suite. Ne restent au final que les *faux match*. Dans l’exemple (Figure 2a), les points verts correspondent aux points associés au premier objet, les points bleus au deuxième objet, et les points rouges aux *outliers* finaux. Si le nombre d’objets est connu, le nombre d’itérations est fixé. Sinon, la condition d’arrêt peut être un nombre de points restant insuffisant pour calculer une matrice fondamentale de façon robuste. Dans la suite, cette méthode sera appelée *moindres carrés médians récursifs*.

L’hypothèse faite ici selon laquelle le plus gros objet regroupe au moins la moitié des points d’intérêt est imposée par la méthode aux moindres carrés médians qui n’est valable que si l’on a moins de 50% d’*outliers*. Une méthode de type Ransac permettrait éventuellement de supprimer cette limitation. Un autre problème se pose pour les objets trop petits, pour lesquels le nombre de points d’intérêt pour effectuer une estimation robuste de la matrice fondamentale est insuffisant. Il pourrait alors être intéressant de raffiner localement l’estimation de mouvement pour obtenir le nombre de points suffisant pour calculer une matrice fondamentale.



(a)

(b)

Figure 2 – Points utilisés pour le calcul des matrices, (a) en utilisant la méthode des moindres carrés médians récursifs (§4.2), (b) en utilisant l’approche alternée (§5.2).

## 5 Segmentation

Nous allons maintenant utiliser le champ de mouvement dense obtenu par l’estimateur par maillage pour effectuer une segmentation, soit décider pour chaque point de l’objet auquel il appartient. La distance du correspondant estimé par le flot optique à la droite épipolaire sur laquelle il devrait se trouver s’il appartenait à un objet (appelé résidu épipolaire) fournit une mesure pour décider de l’appartenance d’un point à cet objet.

Pour chaque objet, une carte dense des résidus épipolaires est calculée. En théorie, pour tous les points appartenant à l’objet, le résidu épipolaire est nul. Malheureusement, un résidu nul (ou faible, suivant l’incertitude des calculs précédents) n’est pas une condition suffisante pour assurer qu’un point appartient à l’objet.

### 5.1 Approche markovienne

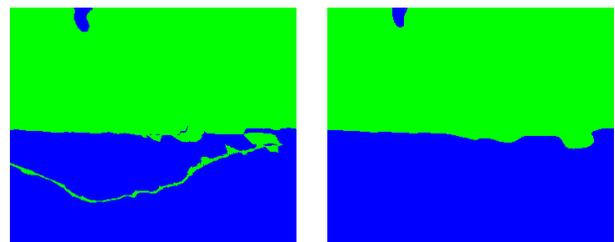
Étant donné que l’on a accès aux différentes matrices épipolaires, la solution la plus simple consiste à attribuer chaque point à l’objet pour lequel son résidu épipolaire est le plus faible. Cependant, dans certaines zones ambiguës, les résidus relatifs à plusieurs géométries sont faibles, et un point peut être attribué à un objet dont il ne fait pas effectivement partie, mais dont le mouvement est compatible (cf. Figure 3a). Ces zones sont généralement de petite dimension, et une régularisation spatiale est alors suffisante pour les faire disparaître.

Dans ce but, une modélisation probabiliste a été adoptée, avec un terme de régularisation de type markovien. Il s’agit de minimiser la fonction d’énergie suivante :

$$\sum_{i,k} \left[ (p(i,k) - p_0(i,k))^2 + \lambda \sum_{j \in V_i} (p(i,k) - p(j,k))^2 \right]$$

où  $p(i,k)$  est la probabilité du point  $i$  d’appartenir à l’objet  $k$ , et où  $p_0(i,k) = \text{normalisée}(\exp(-\text{résidu}_k(i)^2))$  avec *normalisée* fonction de normalisation qui ramène en chaque point la somme des probabilités à 1. La minimisation se fait à l’aide d’une technique itérative (apparentée à une relaxation markovienne, sous contrainte).

Un masque *régularisé* (cf. Figure 3b), où n’apparaissent plus les petites zones mal assignées évoquées précédemment, est ainsi obtenu.



(a)

(b)

Figure 3 – Cartes de segmentation, (a) par affectation directe, (b) en utilisant l’approche markovienne.

## 5.2 Approche alternée

Elle consiste à utiliser la méthode des *moindres carrés médians récurrents* comme initialisation d'un algorithme de classification de type réallocation-recentrage :

1. Calcul des matrices fondamentales et segmentation markovienne.
2. Partition de la liste des points d'intérêt en sous listes correspondant au masque de chaque objet.
3. Recalcul de chaque matrice fondamentale.
4. Calcul de la segmentation associée aux nouvelles matrices.
5. Retour éventuel à l'étape 2 (il y a rarement une amélioration significative après la deuxième itération)

Cette méthode permet d'éviter que des points mal situés ne viennent biaiser le calcul de la matrice épipolaire.

Par exemple, sur la Figure 2a, des points sont associés à tort au premier objet (le jardin) alors qu'ils se situent sur le deuxième (la rue). Ces points disparaissent après l'étape 2 (cf. Figure 2b). Il s'avère que la segmentation n'est pas vraiment améliorée par le second calcul. Cependant, la matrice correspondant au jardin, calculée en utilisant des points mal assignés, était fautive. Ceci entraîne une reconstruction aberrante. Le recalcul des matrices fondamentales permet ici de corriger cette erreur.

Des problèmes de segmentation subsistent en raison des difficultés d'estimation de mouvement au voisinage du contour des objets, en particulier dans les zones de découverture/recouvrement, visibles dans des séquences moins triviales que celle présentée ici. Une nouvelle itération du processus complet, en utilisant la segmentation pour l'estimation de mouvement pourrait permettre de réduire ces problèmes.

## 6 Conclusion

Nous avons proposé une méthode pour segmenter une séquence vidéo suivant le mouvement 3D et identifier chaque mouvement, de manière à pouvoir reconstruire indépendamment chaque objet. Le champ de mouvement est d'abord estimé de façon robuste par un estimateur par maillages. Les différentes matrices fondamentales sont ensuite identifiées et calculées en utilisant un calcul moindres carrés médians de façon récursive sur une décimation des sommets du maillage. Puis l'image est segmentée d'après les résidus épipolaires après relaxation markovienne. Enfin les étapes de calcul de matrices fondamentales et de segmentation sont réitérées pour améliorer l'estimation des paramètres de mouvement, et obtenir une segmentation cohérente avec ces nouveaux paramètres.

Les résultats obtenus sont globalement intéressants, malgré les limitations que nous avons soulignées. Les objets sont bien localisés, et, s'ils possèdent suffisamment de points d'intérêt, bien reconstruits.

## Références

- [1] Franck Galpin. *Représentation 3D de séquences vidéo ; Schéma d'extraction automatique d'un flux de modèles 3D, applications à la compression et à la réalité virtuelle*. Thèse de doctorat, Université de Rennes 1, France, 1 2002.
- [2] D. Demirdjian et R. Horaud. A projective framework for scene segmentation in the presence of moving objects. Dans *CVPR99*, pages I: 2–8, 1999.
- [3] W. Matusik, C. Buehler, et L. McMillan. Polyhedral visual hulls for real-time rendering. Dans *12th Eurographics Workshops on Rendering*, pages 115–126, 2001.
- [4] L. Zhou et C. Kambhamettu. Hierarchical structure and nonrigid motion recovery from 2d monocular views. Dans *CVPR00*, pages II: 752–759, 2000.
- [5] L. Oisel. *Reconstruction 3D de scènes complexes à partir de séquences vidéo non calibrées : estimation et maillage d'un champ de disparité*. Thèse de doctorat, Université de Rennes 1, Novembre 1998.
- [6] Z.Y. Zhang, R. Deriche, O.D. Faugeras, et Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *AI*, 78(1-2):87–119, 1995.
- [7] G. Adiv. Determining 3-d motion and structure from optical flow generated by several moving objects. *PAMI*, 7(4):384–401, Juillet 1985.
- [8] J. Weber et J. Malik. Rigid-body segmentation and shape-description from dense optical-flow under weak perspective. *PAMI*, 19(2):139–143, Février 1997.
- [9] E. Steinbach, S. Chaudhuri, et B. Girod. Robust estimation of three-dimensional motion and structure of multiple objects from image sequences. Dans *Proceedings 3D Image Analysis and Synthesis*, pages 53–59, 1996.
- [10] G. Xu. A unified approach to image matching and segmentation in stereo, motion, and object recognition via recovery of epipolar geometry. *Videre*, 1(1):22–55, 1997.
- [11] T.Y. Tian et M. Shah. Recovering 3d motion of multiple objects using adaptive hough transform. *PAMI*, 19(10):1178–1183, October 1997.
- [12] G. Marquant. *Représentation par maillage adaptatif déformable pour la manipulation et la communication d'objets vidéo*. Thèse de doctorat, Université de Rennes 1, 2000.
- [13] C. Harris et M.J. Stephens. A combined corner and edge detector. Dans *Alvey88*, pages 147–152, 1988.
- [14] B.S. Boufama et R. Mohr. Epipole and fundamental matrix estimation using virtual parallax. Dans *ICCV95*, pages 1030–1036, 1995.