

Estimation des positions d'objets 3D à partir d'une séquence d'images monoculaire

Christophe BOUCHER, Jean-Charles NOYER, Mohammed BENJELLOUN

Laboratoire d'Analyse des Systèmes du Littoral, EA 2600
Maison de la Recherche Blaise Pascal, Université du Littoral Côte d'Opale
195 rue M. Luther-King, BP 649, 62228 Calais Cedex, France

Christophe.Boucher@lasl-gw.univ-littoral.fr, Jean-Charles.Noyer@lasl-gw.univ-littoral.fr,
Mohammed.Benjelloun@lasl-gw.univ-littoral.fr

Résumé – Cet article présente une méthode d'estimation des positions d'objets 3D à partir de primitives 2D extraites d'une séquence d'images monoculaire. L'estimation de la structure est ici rendue possible par l'utilisation de la modélisation de Plücker, dont l'avantage est de présenter une formulation invariante du mouvement en 2D et 3D. Cette méthode est bien adaptée à la prise en compte de tout objet que l'on peut décrire par une approximation polygonale.

Abstract – This paper presents an estimation method of 3D objects positions using 2D primitives of a monocular images sequence. The structure estimation is made possible by the use of the Plücker representation which leads to two similar equations of the 2D-3D dynamic. This method is well matched to objects which can be described by a polygonal approximation.

1 Introduction

L'un des points clefs en vision par ordinateur, que ce soit en vision polynoculaire ou monoculaire, dynamique ou statique, est la reconstruction géométrique d'une scène dans un environnement 3D à partir d'indices visuels 2D. Plusieurs tentatives ont été effectuées dans ce but selon les primitives et le contexte choisis [1].

Dans cet article, nous présentons une méthode géométrique d'estimation de la structure 3D à partir du mouvement. L'approche proposée présente le cas où l'observateur et l'objet sont mobiles. Le modèle d'état non-linéaire d'estimation des positions 2D établi en représentation de Plücker se généralise au cas de droites 3D. Notre contribution se situe à deux niveaux :

- la poursuite des positions 2D : le premier problème vient de l'étiquetage des segments (obtenus par les méthodes usuelles de segmentation [4]) dans la séquence d'images. Pour cela, on prédit à l'aide d'un filtre de Kalman les positions successives des points et des droites supports des segments suivis. Cette prédiction sert à définir l'espace de recherche dans le plan image pour effectuer la mesure. La mesure ainsi choisie, par minimisation d'un critère de distance [5] entre la mesure prédite et les mesures courantes, permet de mettre à jour le filtre qui effectue la mise en correspondance des segments.
- l'estimation des positions 3D : la position 3D des points et segments de droites est obtenue à partir de leur suivi dans le plan image et de la connaissance du mouvement de la caméra et de l'objet. Les relations en représentation de Plücker liant une droite 3D et sa projection perspective dans le plan rétinien [6] permettent alors d'estimer la position spatiale 3D

des points et segments de droite qui est obtenue sans hypothèses très restrictives sur les mouvements 3D.

2 Modèle 2D-3D

La primitive géométrique qui définit notre modèle est la droite. Son déplacement entre deux images successives est décrit par un mouvement 3D rigide. L'exploitation des propriétés de la représentation de Plücker permet d'aboutir à deux formulations du suivi 2D et de la reconstruction 3D de la même forme [3], dont l'expression générale est donnée par :

$$\begin{cases} \vec{U}_{k+1} = A_k \vec{U}_k \\ \vec{L}_{k+1} = A_k \vec{L}_k + \vec{D}_k \wedge A_k \vec{U}_k \\ \vec{\Theta}_{k+1} = A_k \vec{\Theta}_k + \vec{D}_k \end{cases} \quad (1)$$

avec :

- (\vec{U}_k, \vec{L}_k) et (\vec{u}_k, \vec{l}_k) les coordonnées de Plücker 3D et 2D de la droite à l'instant k ;
- $\vec{\Theta}_k$ et $\vec{\theta}_k$ les coordonnées cartésiennes 3D et 2D d'un point du segment à l'instant k ;
- $\begin{cases} A_k = R_{ex} R_c^{-1} R_l R_{ex}^{-1} \\ \vec{D}_k = R_{ex} R_c^{-1} (\vec{T}_l - \vec{T}_c) + (I - A_k) \vec{T}_{ex} \end{cases}$
 (R_c, \vec{T}_c) , (R_l, \vec{T}_l) les mouvements 3D de la caméra et de l'objet et (R_{ex}, \vec{T}_{ex}) les paramètres extrinsèques de la caméra.

3 Poursuite des segments dans le plan image

La mise en correspondance des segments d'une image à l'autre constitue une des principales étapes du processus

de reconstruction. Elle est réalisée par un filtre de Kalman en 2D qui prédit la position du segment dans l'image suivante et permet de définir une région de confiance dans laquelle doit se trouver la mesure correspondante [2].

3.1 Estimation des positions du segment

Le mouvement des segments dans le plan image est décrit par un modèle affine bi-dimensionnel :

$$\theta_{k+1} = a_k \theta_k + d_k + w_k \quad (2)$$

où θ_k sont les coordonnées 2D d'un point du segment et w_k est un bruit blanc gaussien de dynamique, de moyenne nulle et de variance Q_k .

Les équations d'état du filtre de Kalman s'écrivent (pour les extrémités d'un segment) :

1. Dynamique :

$$\begin{aligned} \theta'_{k+1} &= \begin{pmatrix} \theta_{k+1}^1 \\ \theta_{k+1}^2 \end{pmatrix} = \begin{pmatrix} x_{k+1}^1 \\ y_{k+1}^1 \\ x_{k+1}^2 \\ y_{k+1}^2 \end{pmatrix} \\ &= \begin{pmatrix} a_k & 0_2 \\ 0_2 & a_k \end{pmatrix} \begin{pmatrix} x_k^1 \\ y_k^1 \\ x_k^2 \\ y_k^2 \end{pmatrix} \\ &\quad + \begin{pmatrix} d_k^x \\ d_k^y \\ d_k^x \\ d_k^y \end{pmatrix} + \begin{pmatrix} w_k^{1,1} \\ w_k^{1,2} \\ w_k^{2,1} \\ w_k^{2,2} \end{pmatrix} \end{aligned}$$

où 0_2 est une matrice nulle de dimension 2, et $w_k^{i,j}$ ($i, j = \{1, 2\}$) sont des bruits blancs gaussiens de moyenne nulle et $E(w_k^{i,j} w_k^{i',j'}) = \sigma_Q^2 \delta_{ii'} \delta_{jj'}$.

2. Mesure :

$$Y_{k+1} = \theta'_{k+1} + V_{k+1} = \begin{pmatrix} x_{k+1}^1 \\ y_{k+1}^1 \\ x_{k+1}^2 \\ y_{k+1}^2 \end{pmatrix} + \begin{pmatrix} v_{k+1}^{1,1} \\ v_{k+1}^{1,2} \\ v_{k+1}^{2,1} \\ v_{k+1}^{2,2} \end{pmatrix}$$

où v_{k+1} est un bruit blanc gaussien de moyenne nulle et de matrice de covariance $R_{k+1} = \sigma_R^2 I_4$ (I_4 : matrice identité de dimension 4).

3.2 Mise en correspondance

L'étape précédente permet de prédire la position du segment dans l'image à l'instant $k+1$. On détermine maintenant le segment correspondant dans l'image segmentée par minimisation de la distance de Mahalanobis [5] :

$$\begin{aligned} d(\hat{\theta}_{k+1|k}, Y_{k+1}^i) &= \\ (\hat{\theta}_{k+1|k} - Y_{k+1}^i)^T (\tilde{P}_{k+1|k} + R_{k+1}) (\hat{\theta}_{k+1|k} - Y_{k+1}^i) & \end{aligned} \quad (3)$$

où $\hat{\theta}_{k+1|k}$ est l'état prédit du système à $k+1$, $\tilde{P}_{k+1|k}$ la covariance de l'erreur de prédiction du filtre et Y_{k+1}^i ($i = 1 \dots n$) l'ensemble des droites issues de la segmentation.

Le segment i mis en correspondance est celui pour lequel la distance se trouve en deçà d'un seuil déterminé.

3.3 Estimation du modèle de dynamique

Les paramètres du modèle affine sont calculés adaptativement de manière à éviter toute divergence du filtre suite à une erreur de modélisation. La périodicité du calcul est fixée par le seuil introduit dans la mise en correspondance. Dans un premier temps, le dépassement du seuil peut être engendré par une mauvaise prédiction de la position due à une erreur de modèle. On utilise donc les segments précédemment étiquetés pour recalculer par moindres carrés les matrices du mouvement affine (a_k, d_k). On traite à nouveau les données pour décider si il y a absence de segment à mettre en correspondance (phénomène d'occultation, mauvaise segmentation, etc.) dans l'image courante. Dans ce cas, le filtre fonctionne seulement en prédiction et n'est pas corrigé par les mesures à l'instant courant.

4 Détermination de la structure 3D

On pose ici le problème d'estimation de la structure 3D à partir de la connaissance d'indices visuels 2D dans le plan image. Cette opération n'est possible que par l'adjonction d'une contrainte supplémentaire levant l'ambiguïté du passage du plan image au monde 3D. La connaissance des mouvements de la caméra et de l'objet s'avère suffisante. En effet, outre le modèle d'état précédent, il existe en coordonnées de Plücker des relations intrinsèques liant les coordonnées d'une droite 3D et sa projection perspective [6] :

$$\vec{l}_k \cdot \vec{U}_k = 0 \quad \text{et} \quad \vec{l}_{k+1} \cdot \vec{U}_{k+1} = 0 \quad (4)$$

$$\vec{l}_k \wedge \vec{L}_k = \vec{0} \quad \text{et} \quad \vec{l}_{k+1} \wedge \vec{L}_{k+1} = \vec{0} \quad (5)$$

À partir des considérations précédentes, des relations (4), (5) et du système d'état, nous avons abouti à une reconstruction 3D exacte des primitives 2D rétiniennes :

$$\vec{U}_k = \frac{\vec{K} \wedge \vec{l}_k}{\|\vec{K} \wedge \vec{l}_k\|} \quad (6)$$

$$\vec{\Theta}_k = \frac{-\vec{\sigma}_{k+1} \cdot \vec{D}_k}{\vec{\sigma}_{k+1} A_k \vec{\theta}_k} \cdot \vec{\theta}_k \quad (7)$$

$$\vec{L}_k = \frac{-\vec{J}_{k+1} \cdot [\vec{D}_k \wedge A_k \vec{U}_k]}{\vec{J}_{k+1} A_k \vec{l}_k} \vec{l}_k \quad (8)$$

avec :

$$\begin{aligned} -\vec{K} &= A_k^T \cdot \vec{l}_{k+1}; \\ -\vec{\sigma}_{k+1}^T &= (1 - \theta_2, \theta_1 - 1, \theta_2 - \theta_1)_{k+1}; \\ -\vec{J}_{k+1}^T &= (l_3 - l_2, l_1 - l_3, l_2 - l_1)_{k+1}. \end{aligned}$$

5 Résultats de reconstruction

La méthode proposée est ici appliquée à des images en niveaux de gris synthétiques d'objets en mouvement de rotation/translation selon les trois axes, pour lesquelles l'observateur décrit le même type de mouvement.

Nous présentons, à titre d'exemple, les résultats de simulation de la méthode sur une séquence de 36 images synthétiques. La figure 1 en présente deux échantillons aux instants 8 et 23. La caméra est décrite par un modèle sténopé : $\alpha_u = \alpha_v = 512$ et $u_0 = v_0 = 256$.

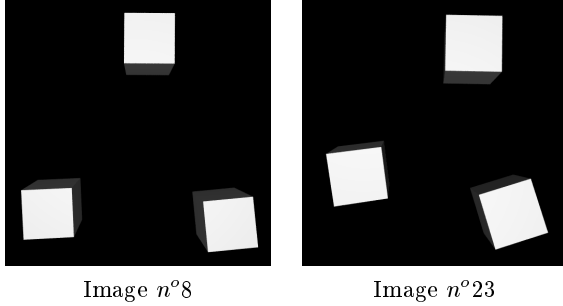


FIG. 1 – Images de la séquence de synthèse

Les mouvements 3D (rotation/translation) de l'objet et de la caméra sont résumés dans le tableau ci-dessous :

TAB. 1 – Mouvement 3D de l'objet et de la caméra

	R_x	R_y	R_z	T_x	T_y	T_z
Caméra	0.2°	0.1°	0.15°	-0.05	-0.1	-0.15
Cube n°2	0.1°	0.1°	0.1°	0.1	-0.15	0.05

La figure 2 présente les résultats de mise en correspondance du segment 0, par exemple, aux instants 8 et 23. Pour ce même segment, nous avons tracé sur la figure 3 l'évolution de la distance de Mahalanobis en fonction du temps (le seuil de décision est fixé à 30). Les instants où celle-ci chute à zéro correspondent à une absence de mesure pertinente : le filtre fonctionne alors uniquement en prédiction.

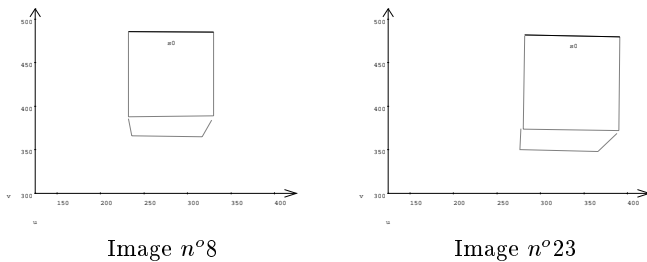


FIG. 2 – Poursuite du segment 0 (en trait plein) dans le plan numérique : images 8 et 23

Les figures 4, 5 et 6 présentent les erreurs à la reconstruction des paramètres (\vec{U} , \vec{L} , $\vec{\Theta}$) sur l'ensemble de la séquence, pour le segment 0. On remarque notamment que l'erreur de reconstruction 3D tend rapidement vers zéro (en une dizaine d'itérations environ). La vitesse de convergence du filtre est ici fortement liée à l'étape d'extraction des primitives 2D et à l'estimation des paramètres de calibrage [7].

Suivant le même principe, nous présentons à la figure 7 une reconstruction de l'ensemble des segments suivis, formant la face avant et inférieure du cube choisi, aux instants 8 et 23.

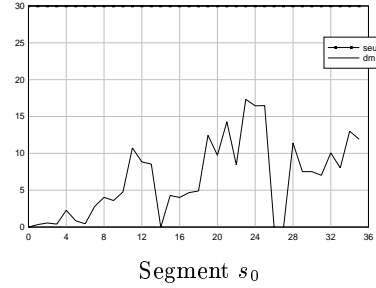


FIG. 3 – Evolution de la distance de Mahalanobis du segment 0

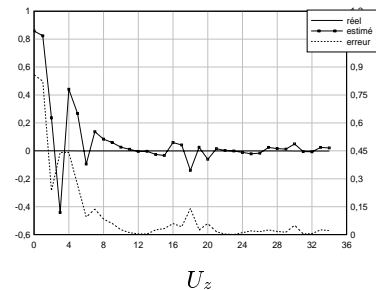
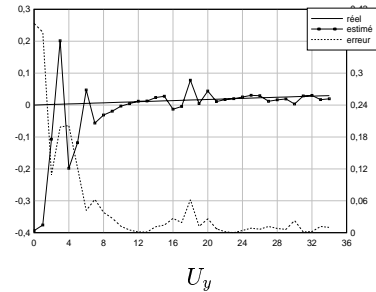
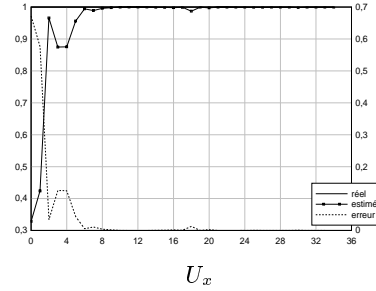


FIG. 4 – Reconstruction 3D du segment 0 : \vec{U} (Erreur en pointillé)

6 Conclusion

Cet article présente une méthode de suivi 2D et d'estimation des positions d'objets 3D dans le cas où l'observateur et l'objet sont mobiles. L'originalité réside dans l'utilisation des relations en représentation de Plücker liant une droite 3D et sa projection perspective dans le plan rétinien pour une estimation de la structure 3D. Cette méthode a montré son efficacité en reconstruction ainsi qu'une robustesse à de forts déplacements grâce au calcul adaptatif du modèle dynamique.

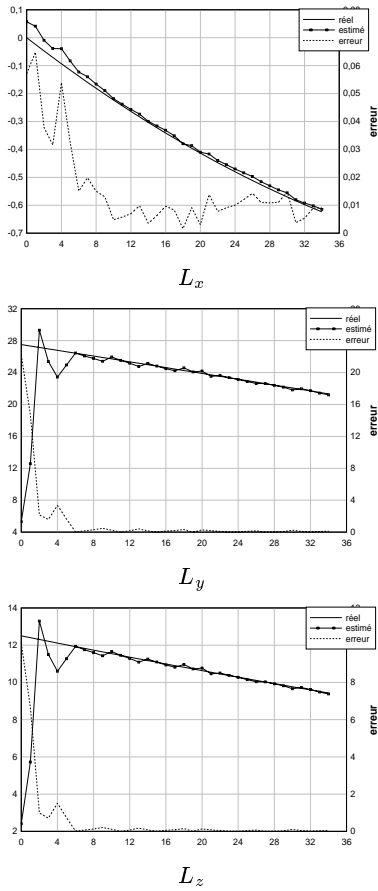


FIG. 5 – Reconstruction 3D du segment 0 : \vec{L} (Erreur en pointillé)

7 Remerciements

Les auteurs tiennent à remercier la région Nord/Pas-de-Calais pour son soutien financier par l'intermédiaire du GRAISyHM.

Références

- [1] Numéro spécial reconstruction. *Traitement Du Signal*, 13(4), 1996.
- [2] Y. Bar-Shalom and X.-R. Li. *Estimation and Tracking : Principles, Techniques and Software*. Artech House, 1993.
- [3] M. Benjelloun, A. Sdigui, and F. Chavand. Moving object tracking and reconstruction from 3D straight lines using a mobile camera. In *Motion Control*, pages 745–752. LBM, October 1995.
- [4] R. Deriche. Using Canny's criteria to derive an optimal edge detector recursively implemented. *The International Journal of Computer Vision*, April 1987.
- [5] R. Deriche and O. Faugeras. Tracking line segments. *Image and Vision Computing*, 8 :261–270, 1990.
- [6] A. Sdigui and M. Benjelloun. Moving scene reconstruction from 3D segments using a fixed camera. In *International Conference on Image Processing*, Austin, Texas, USA, November 1994. IEEE.

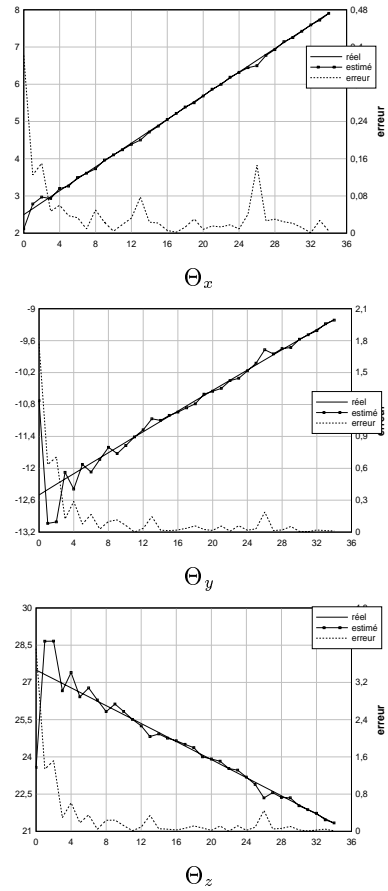


FIG. 6 – Reconstruction 3D du segment 0 : $\vec{\Theta}$ (Erreur en pointillé)

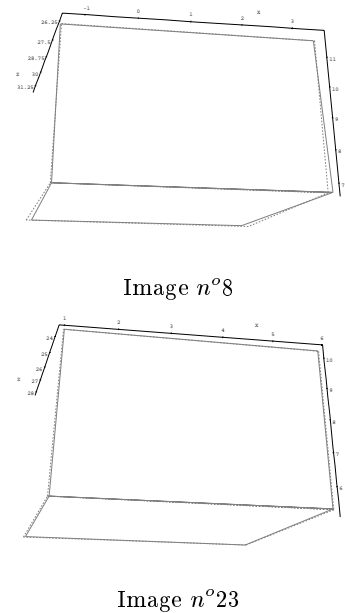


FIG. 7 – Reconstruction 3D du cube 2 (Cube réel en pointillé)

- [7] C. Zeller. *Calibration Projective, Affine et Euclidienne En Vision Par Ordinateur et Application À la Perception Tridimensionnelle*. PhD thesis, Université de Rennes I, Février 1996.