

# Estimation des positions d'objets 3D par analyse monoculaire en représentation de Plücker

---

## Structure estimation of 3D objects by monocular analysis using Plücker representation

par C BOUCHER, J.-C. NOYER, M. BENJELLOUN

Laboratoire d'Analyse des Systèmes du Littoral (EA 2600), Université du Littoral Côte d'Opale  
50, rue Ferdinand Buisson, B.P. 699, 62228 Calais Cedex, France

### *résumé et mots clés*

Cet article présente une méthode géométrique d'estimation de la structure d'objets 3D à partir de la connaissance de leur mouvement et d'une séquence d'images monoculaire, en tenant compte de l'ensemble de la chaîne de perception. L'estimation des positions 3D est ici rendue possible par l'exploitation des propriétés de la représentation de Plücker, dont l'avantage est de présenter une formulation géométrique invariante de la droite 2D et 3D. À noter que cette méthode est bien adaptée à la prise en compte de tout objet que l'on peut décrire par une approximation polygonale.

Estimation de la structure à partir du mouvement, analyse du mouvement 2D, représentation de Plücker

### *abstract and key words*

This article deals with a geometric method that estimates the 3D structure of an object from the knowledge of their motion and a monocular image sequence, taking into account the whole acquisition scheme. The estimation of the 3D positions is done via the properties of the Plücker modelization, whose benefit is to ensure an invariant geometric formulation of the 2D and 3D straight line. Besides, this method is well suited to deal with objects which can be described by a polygonal approximation.

Structure estimation from motion, 2D motion analysis, Plücker representation

## 1. introduction

Dans le domaine de la vision par ordinateur, un des points importants demeure la reconstruction géométrique d'une scène dans un environnement 3D à partir d'indices visuels 2D, comme il est souvent nécessaire, par exemple, pour les applications de la robotique mobile, l'imagerie médicale, la télésurveillance où l'on cherche à localiser un objet dans une scène 3D. L'étude de ce problème a donné lieu à de nombreux travaux qui utilisent différents contextes ou primitives [AS00, TK95, Sha95, HN94, LH88]. De nombreuses méthodes ont été développées pour estimer la structure 3D d'objets à partir de plusieurs vues. Elles peuvent être regroupées essentiellement en trois approches : les méthodes basées sur le suivi de primitives, celles basées sur des approches différentielles et finalement celles qui exploitent les corrélations spatio-temporelles. On peut notamment estimer la structure 3D à partir de l'intensité des pixels de l'image ou de la texture des objets [BG98, Gar93, KC89] des contours extraits [MDMC91, BAS<sup>+</sup>87] de connaissances géométriques sur l'objet 3D [Csu96, ER87, HS87] de l'analyse stéréoscopique de la scène [FA97, Fau93, Aya89, DA89], du mouvement de la caméra et/ou des objets de la scène [AM99, BZM97, DZB92, ER87, LHP80], etc.

La méthode proposée se place dans le cadre des travaux d'estimation de la structure à partir de la connaissance du mouvement (« structure from motion ») [FIH99, OG99, BZM97, DZB92]. Il s'agit d'une méthode géométrique de reconstruction 3D basée sur la représentation de Plücker [Bou00, BNB99b], dans le cas où l'observateur et l'objet sont mobiles. Par nature, la représentation de Plücker repose sur des primitives de type point et segment qui sont exploitées dans cet article. Dans ce genre d'approche, le mouvement peut être par ailleurs estimé par des méthodes telles que [Zha95, JJ84].

L'étude porte principalement sur deux points :

- le suivi des primitives 2D, où la difficulté réside dans l'étiquetage des segments de l'objet (obtenus par segmentation [Der87], chaînage des contours et approximation polygonale [HM95, Gir87]) 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 [DF92] entre la mesure prédite et les mesures courantes, permet de mettre à jour le filtre qui effectue la mise en correspondance des segments de même que le calcul des caractéristiques du mouvement 2D relatif de l'objet.
- l'estimation des positions 3D des points et des segments de droites, qui est obtenue à partir du suivi 2D dans la séquence d'images, de leur projection dans le plan rétinien, de la connaissance du mouvement de la caméra et de l'objet, et des paramètres de calibrage. L'exploitation des relations en représenta-

tion de Plücker [Sdi93, Xie89, ER87] liant une droite 3D à sa projection perspective dans le plan rétinien permet alors d'estimer la position spatiale 3D des points et segments de droite [Bou00, BNB99b, BSC95].

Pour valider la méthode de reconstruction proposée, des expérimentations sur données synthétiques et réelles sont également présentées.

## 2. suivi des segments 2D

Une des principales étapes du processus d'estimation de la structure à partir du mouvement concerne la mise en correspondance des segments d'une image à l'autre.

Dans notre cas, elle est réalisée par un filtre de Kalman qui prédit la position du segment dans l'image suivante et définit une région de confiance dans laquelle doit se trouver la mesure correspondante. On a choisi de décrire le mouvement des segments dans le plan image par un modèle affine bidimensionnel. L'intérêt du mouvement affine [Sha95] réside dans sa capacité à décrire une vaste classe de déplacements, même non-rigides dans certaines proportions.

L'équation générale du mouvement affine  $(a_k, d_k)$  s'écrit :

$$p_{k+1} = a_k p_k + d_k \quad (1)$$

où  $p_k (u_k, v_k)^T$  est le vecteur des coordonnées « pixel » d'un point 2D à l'instant  $k$ ,  $a_k$  est une matrice carrée dimension deux et  $d_k$  un vecteur de dimension deux. Ces paramètres caractérisent le mouvement affine 2D.

On choisit de modéliser le problème du suivi de primitives segments (décrits par leurs extrémités) sous forme d'équations d'état (dynamique et observation) :

$$\begin{cases} X_{k+1} = A'_k X_k + D'_k + W_k & (a) \\ Y_{k+1} = X_{k+1} + V_{k+1} & (b) \end{cases} \quad (2)$$

où :

- $X_k (p_k^1, p_k^2)^T$  est le vecteur d'état composé des coordonnées 2D de chaque extrémité  $p_k^i$  du segment, à l'instant  $k$  ;
- $(A'_k, D'_k)$  est le mouvement affine du segment perçu dans l'image avec  $A'_k = \begin{pmatrix} a_k & \mathbf{0}_2 \\ \mathbf{0}_2 & a_k \end{pmatrix}$  ( $\mathbf{0}_2$  est la matrice carrée nulle de dimension 2) et  $D'_k (d_k, d_k)^T$  ;
- $Y_{k+1}$  constitue le vecteur de mesures issues de l'étape de segmentation [Der87] des images, à l'instant  $k + 1$ . Il est donc composé des coordonnées pixel des extrémités des segments ;

–  $W_k (w_k^{1,u}, w_k^{1,v}, w_k^{2,u}, w_k^{2,v})^T$  et  $V_k (v_k^{1,u}, v_k^{1,v}, v_k^{2,u}, v_k^{2,v})^T$  sont des bruits blancs gaussiens additifs de moyennes nulles et de matrices de covariance respectives  $Q_k = \sigma_Q^2 I_4$  et  $R_k = \sigma_R^2 I_4$  ( $I_n$  représente la matrice identité de dimension  $n$ ).

**Remarque :** La présence du terme de bruit de modèle  $W_k$  permet notamment de tenir compte des dérives éventuelles du modèle de dynamique des primitives. Le vecteur  $V_k$  modélise les erreurs de calibrage de la caméra et de segmentation (qui dépendent elles même des bruits inhérents de capteur, notamment).

## 2.1. prédiction des positions 2D

La solution de ce problème d'estimation (eq. 2) repose sur un filtre de Kalman [BSL93] qui calcule, à chaque instant, les positions du segment. La structure générale de ce filtre se décompose en deux étapes : une étape de prédiction de l'état du système (à partir de l'équation de dynamique 2a) et une étape de correction de cette estimation (à partir de l'équation d'observation 2b). Les équations de prédiction du filtre s'écrivent (état prédit et covariance de l'erreur de prédiction) :

$$\begin{cases} \hat{X}_{k+1/k} = A'_k \hat{X}_{k/k} + D'_k \\ \tilde{P}_{k+1/k} = A'_k \tilde{P}_{k/k} A_k'^T + Q_k \end{cases} \quad (3)$$

où  $\hat{X}_{k+1/k}$  représente l'estimation de l'état à l'instant  $k+1$  à partir des mesures jusqu'à l'instant  $k$  ;  $\tilde{P}_{k/k}$  est la covariance de l'erreur d'estimation du filtre à l'instant  $k$ .

## 2.2. suivi des primitives dans la séquence d'images

La difficulté réside maintenant dans le suivi de chaque segment le long de la séquence d'images afin d'initialiser l'équation de mesure du filtre (eq. 2b). En effet, le problème principal de la description d'une scène 3D par un ensemble de segments est qu'elle engendre un nombre important de mesures potentielles et qu'une seule doit être choisie. Il existe plusieurs solutions à ce problème de suivi multi-cibles, telles que les méthodes d'association probabiliste de données (PDAF), les techniques de suivi multi-hypothèses (MHT) [BSL95], la minimisation d'une fonction de similitude, etc.

La solution retenue ici est celle de la minimisation de la distance de Mahalanobis [DF90] entre deux vecteurs de paramètres (les coordonnées 2D des extrémités, par exemple). Ce critère est largement utilisé dans les applications de vision par ordinateur. La méthode consiste à trouver le correspondant le plus plausible

parmi l'ensemble des segments de la zone étudiée et détecter éventuellement l'absence de correspondant dans l'image courante. On calcule donc cette distance pour toutes les primitives de l'image : le correspondant le plus probable est celui qui minimise la fonction, soit celui pour lequel la distance est en dessous d'un seuil défini.

L'étape précédente (eq. 3) 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 entre l'état prédit et chaque mesure potentielle :

$$d(\hat{X}_{k+1/k}, Y_{k+1}^i) = (\hat{X}_{k+1/k} - Y_{k+1}^i)^T (\tilde{P}_{k+1/k} + R_{k+1})^{-1} (\hat{X}_{k+1/k} - Y_{k+1}^i) \quad (4)$$

où  $(Y_{k+1}^i)_{i=1..n}$  représente l'ensemble des primitives issues de la segmentation.

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

## 2.3. estimation des positions 2D

Les segments mis en correspondance sont utilisés pour mettre à jour l'estimation réalisée par le filtre. Les équations du filtre de Kalman s'écrivent alors :

$$\begin{cases} K_{k+1} = \tilde{P}_{k+1/k} (\tilde{P}_{k+1/k} + R_{k+1})^{-1} \\ \tilde{P}_{k+1/k+1} = \tilde{P}_{k+1/k} - K_{k+1} \tilde{P}_{k+1/k} \\ \hat{X}_{k+1/k+1} = \hat{X}_{k+1/k} + K_{k+1} (Y_{k+1}^* - \hat{X}_{k+1/k}) \end{cases}$$

où  $Y_{k+1}^*$  est la mesure (ici le segment) retenue dans l'étape de mise en correspondance.

Si aucun segment n'est retenu, le filtre fonctionne alors en prédiction pure.

## 2.4. estimation du modèle de dynamique

Pour éviter toute divergence du filtre consécutive à une erreur de modélisation du mouvement, on calcule de manière adaptative les paramètres du mouvement affine. Ce calcul est effectué périodiquement lorsqu'aucun segment n'a été mis en correspondance. En effet, cette absence de détection peut avoir plusieurs origines. La première provient de l'hypothèse de stationnarité du modèle de dynamique qui ne peut être garantie à long terme. Cette dérive de modèle implique une erreur dans la prédiction de la position du segment, qui interdit toute mise en correspondance.

La deuxième raison est issue du processus de génération des mesures. Les « mesures potentielles » sont extraites de l'étape de segmentation de l'image étudiée, qui peut donner lieu (selon les conditions opératoires) à une absence de segment à mettre en correspondance (phénomène d'occultation, mauvaise segmentation, etc.).

En résumé, on choisit de corriger les paramètres du mouvement lorsque l'on est en absence de mesure potentielle dans l'image courante (le seuil sur la distance de Mahalanobis est dépassé pour l'ensemble des segments candidats). On utilise pour cela les segments précédemment étiquetés pour recalculer par moindres carrés les matrices de mouvement affine  $(a_k, d_k)$ . On prédit ensuite la position du segment dans l'image courante à partir des nouvelles caractéristiques du mouvement. S'il n'y a toujours pas de segments à mettre en correspondance, le filtre fonctionne simplement en prédiction.

### 3. reconstruction 3D

La méthode de reconstruction 3D retenue repose sur la représentation de Plücker [Sdi93, DRLR89, Xie89] d'une droite. Elle présente l'avantage d'être invariante en 2D et 3D en terme d'opérations géométriques comme le produit scalaire ou le produit vectoriel. Un certain nombre de relations déduites de cette considération permettent d'aboutir à une reconstruction 3D de la primitive segment.

#### 3.1. représentation de Plücker

Soient  $\vec{P}_0$  et  $\vec{P}$  deux points de l'espace tridimensionnel appartenant à une même droite [Xie89]. Si l'on représente la direction de cette droite par le vecteur unitaire  $\vec{U}$ , l'équation liant ces deux points s'écrit :

$$\vec{P} = \vec{P}_0 + \alpha \vec{U} \tag{6}$$

où  $\alpha$  est un coefficient réel. Celle-ci peut se réécrire sous forme d'un produit vectoriel :

$$(\vec{P} - \vec{P}_0) \wedge \vec{U} = \vec{0} \tag{7}$$

La représentation d'une droite par ses coordonnées de Plücker est définie par le couple  $(\vec{U}, \vec{L})$  tel que :

$$\vec{L} = \vec{P} \wedge \vec{U} \tag{8}$$

de sorte que le produit scalaire est nul :

$$\vec{L} \cdot \vec{U} = 0 \tag{9}$$

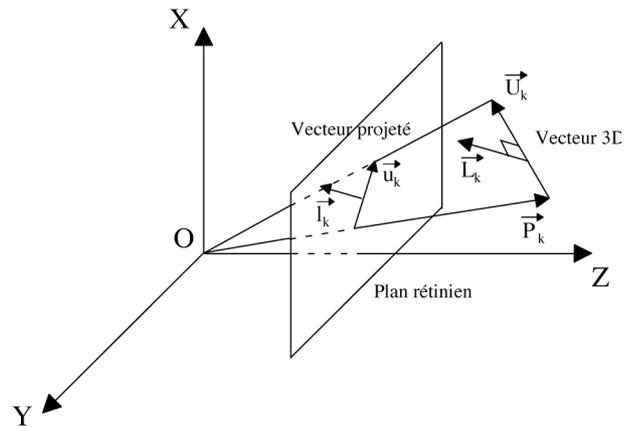


Figure 1. – Représentation d'une droite 2D et 3D.

Un segment du plan rétinien est la projection perspective d'un segment du monde 3D. Par rapport au même repère, le formalisme de représentation est rigoureusement identique pour une droite 2D ou une droite 3D (figure 1). Cette propriété est exploitée pour l'estimation de la structure 3D de segments de droites dans une séquence d'images. L'équation d'une droite 2D liant ses coordonnées de Plücker  $(\vec{u}, \vec{l})$  s'écrit (éq. 8) :

$$\vec{l} = \vec{p} \wedge \vec{u} \tag{10}$$

De plus, il existe en coordonnées de Plücker des relations intrinsèques liant les coordonnées d'une droite 3D et sa projection perspective [BSC95], valables à chaque instant  $k$  :

$$\vec{l}_k \cdot \vec{U}_k = 0 \tag{11}$$

$$\vec{l}_k \wedge \vec{L}_k = \vec{0} \tag{12}$$

#### 3.2. équations d'évolution de la droite 3D

Par construction, les primitives géométriques mises en jeu dans le formalisme de Plücker sont le point et la droite. Le but est ici de modéliser l'évolution temporelle des coordonnées de Plücker d'une droite 3D dans le cas d'un mouvement rigide par rapport à un repère fixe lié à la scène.

Soient une droite 3D et sa projection dans le plan rétinien aux instants  $k$  et  $k + 1$ . On désigne par :

- $(\vec{u}_k, \vec{l}_k)$  et  $(\vec{U}_k, \vec{L}_k)$  les coordonnées de Plücker 2D et 3D du segment à l'instant  $k$  ;
- $\vec{\theta}_k$  et  $\vec{\Theta}_k$  les coordonnées cartésiennes 2D et 3D d'un point du segment à l'instant  $k$  ;
- $(R_c, \vec{T}_c)$  le mouvement rigide de la caméra ;
- $(R_l, \vec{T}_l)$  le mouvement rigide de l'objet.

On peut montrer [Bou00, BNB99b, BSC95] que l'évolution 3D de l'objet peut s'écrire :

$$\begin{cases} \vec{U}_{k+1} = A_k \vec{U}_k \\ \vec{L}_{k+1} = A_k \vec{L}_k + \vec{D}_k \wedge A_k \vec{U}_k \\ \vec{\Theta}_{k+1} = A_k \vec{\Theta}_k + \vec{D}_k \end{cases} \quad (13)$$

avec :

$$- A_k = R_{ex} R_c^{-1} R_l R_{ex}^{-1} ;$$

$$- D_k = R_{ex} R_c^{-1} (\vec{T}_l - \vec{T}_c) + (I - A_k) \vec{T}_{ex}$$

Le couple  $(R_{ex}, \vec{T}_{ex})$  correspond aux paramètres extrinsèques de la caméra, c'est-à-dire les informations permettant de localiser le capteur dans un repère absolu 3D.

### 3.3. détermination de la structure 3D

On propose ici une méthode de reconstruction 3D à partir des primitives 2D, qui s'appuie sur le formalisme proposé par [BSC95]. L'équation d'évolution (éq. 13) va permettre d'obtenir les relations géométriques autorisant l'estimation de la structure 3D d'un objet à partir de la connaissance d'indices visuels 2D dans le plan image. L'ambiguïté du passage du plan image au monde 3D est levée par la connaissance des mouvements de la caméra et de l'objet (« structure from motion »).

On détaille par la suite les principales étapes du processus de reconstruction 3D.

#### 3.3.1. calcul du vecteur directeur

En combinant la relation 11 à  $k+1$  et la première équation du modèle d'évolution 13, on obtient l'expression suivante :

$$(A_k^T l_{k+1}^{\vec{}}) \cdot \vec{U}_k = 0 \quad (14)$$

Posons  $\vec{K} = A_k^T l_{k+1}^{\vec{}}$ . La solution vectorielle donnant le vecteur directeur  $\vec{U}_k$  est obtenue en utilisant les équations 11 et 14, soit :

$$\vec{U}_k = \frac{\vec{K} \wedge \vec{l}_k}{\|\vec{K} \wedge \vec{l}_k\|} \quad (15)$$

#### 3.3.2. calcul de la normale

Afin de déterminer le vecteur normal  $\vec{L}_k$ , on doit résoudre la dernière équation du modèle de dynamique 13. Soit  $\vec{\theta}_k (\theta_k^x, \theta_k^y, 1)^T$ , la projection perspective du vecteur 3D

$\vec{\Theta}_k (\Theta_k^X, \Theta_k^Y, \Theta_k^Z)^T$ . La loi de la projection centrale aux instants  $k$  et  $k+1$ , et la troisième équation du modèle d'évolution donnent les relations suivantes :

$$\vec{\Theta}_k = \alpha \vec{\theta}_k \quad (16)$$

avec  $\alpha = \Theta_k^Z$ , d'où :

$$\vec{\theta}_{k+1} \wedge \vec{\Theta}_{k+1} = \vec{0} \quad (17)$$

On peut donc trouver un vecteur  $\sigma_{k+1}^{\vec{}}$  tel que :

$$\sigma_{k+1}^{\vec{}} \cdot \vec{\Theta}_{k+1} = 0 \quad (18)$$

avec  $\sigma_{k+1}^{\vec{}} (1 - \theta_{k+1}^y, \theta_{k+1}^x - 1, \theta_{k+1}^y - \theta_{k+1}^x)^T$ .

Le vecteur  $\vec{\Theta}_k$  est donc déterminé à l'aide de l'équation suivante :

$$\vec{\Theta}_k = \frac{-\sigma_{k+1}^{\vec{}} \cdot \vec{D}_k \vec{\theta}_k}{\sigma_{k+1}^{\vec{}} A_k \vec{\theta}_k} \quad (19)$$

À ce stade, les vecteurs 2D  $\vec{u}_k$ ,  $\vec{l}_k$  et  $\vec{\theta}_k$  ainsi que les vecteurs 3D  $\vec{U}_k$  et  $\vec{\Theta}_k$  sont connus, tout comme le mouvement 3D  $(A_k, \vec{D}_k)$ . La relation 12 donne, à  $k+1$  :

$$l_{k+1}^{\vec{}} \wedge L_{k+1}^{\vec{}} = \vec{0} \quad (21)$$

Ce que l'on peut exprimer sous la forme suivante :

$$\lambda_{k+1}^{\vec{}} \cdot L_{k+1}^{\vec{}} = 0 \quad (22)$$

avec  $\lambda_{k+1}^{\vec{}} (l_{k+1}^z - l_{k+1}^y, l_{k+1}^x - l_{k+1}^z, l_{k+1}^y - l_{k+1}^x)^T$ .

En utilisant la relation 12, on peut écrire la relation suivante :

$$\vec{L}_k = c \vec{l}_k \quad (22)$$

où  $c$  est une constante réelle. À partir de la deuxième équation d'évolution 13 et des relations 21 et 22, on obtient l'égalité suivante :

$$\lambda_{k+1}^{\vec{}} A_k c \vec{l}_k + \lambda_{k+1}^{\vec{}} \cdot [\vec{D}_k \wedge A_k \vec{U}_k] = 0 \quad (23)$$

soit:

$$\vec{L}_k = \frac{-\lambda_{k+1}^{\vec{}} \cdot [\vec{D}_k \wedge A_k \vec{U}_k]}{\lambda_{k+1}^{\vec{}} A_k \vec{l}_k} \vec{l}_k \quad (24)$$

## 4. passage aux données rétinien

La combinaison des relations 15, 19 et 24 permet alors de reconstruire la structure 3D de l'objet. On remarque toutefois que cette étape passe par la connaissance des vecteurs  $\vec{u}_k$  et  $\vec{l}_k$ , ainsi que d'un point  $\vec{\theta}_k$  qui décrivent, en coordonnées de Plücker, le segment dans le plan rétinien. On détaille donc par la suite, la détermination des données rétinien à partir des données numériques issues de l'étape de suivi.

### 4.1. formulation du problème

La poursuite d'un segment dans la séquences d'images donne à chaque instant  $k$  la position de ses extrémités  $p_k^1 (u_k^1, v_k^1)^T$  et  $p_k^2 (u_k^2, v_k^2)^T$ , exprimée dans le plan image numérique. En utilisant les paramètres intrinsèques de la caméra, il est possible de calculer leurs positions correspondantes  $p_{r_k}^i (x_k^i, y_k^i, z_k^i)^T$ ,  $i=1,2$  dans le plan image rétinien, à l'aide des relations suivantes :

$$\begin{cases} x_k^i = \frac{u_k^i - u_0}{\alpha_u} \\ y_k^i = \frac{v_k^i - v_0}{\alpha_v} \\ z_k^i = 1 \end{cases} \quad (25)$$

où la focale est  $f = 1$  et  $(\alpha_u, \alpha_v, u_0, v_0)$  sont les paramètres intrinsèques identifiés de la caméra.

L'utilisation des relations 25 donne une estimation des coordonnées rétinien des extrémités du segment suivi. Les coordonnées de Plücker établies dans le plan rétinien sont définies comme suit. En ce qui concerne le point du segment, on peut prendre par exemple :

$$\vec{\theta}_k = p_{r_k}^2 \quad (26)$$

Le calcul du vecteur directeur est donné par la relation suivante :

$$\vec{u}_k = \frac{p_{r_k}^2 - p_{r_k}^1}{\|p_{r_k}^2 - p_{r_k}^1\|} \quad (27)$$

Quant à la normale, elle est obtenue à l'aide du produit vectoriel suivant :

$$\vec{l}_k = \vec{\theta}_k \wedge \vec{u}_k \quad (28)$$

L'idée est donc d'utiliser l'estimation des coordonnées des points rétinien  $p_{r_k}^1$  et  $p_{r_k}^2$  afin de calculer les quantités précédentes. L'expérience prouve qu'une détermination directe de ces quantités conduit à une reconstruction 3D du segment imprécise, du fait de l'erreur sur les coordonnées rétinien et, a fortiori, sur les paramètres intrinsèques de la caméra. Une étape de lissage des données rétinien s'avère donc indispensable.

### 4.2. étape de lissage

On utilise pour cela un filtre de Kalman qui s'appuie sur le modèle d'évolution du point  $\theta_k$  dans le plan rétinien et utilise comme mesures les points  $p_{r_k}^i$  obtenus précédemment :

$$\begin{cases} X_{k+1}^i = \Phi_k X_k^i + V_k \\ Y_{k+1}^i = X_{k+1}^i + W_k \end{cases} \quad (29)$$

où :

–  $X_k^i (p_{r_k}^i \ 1)^T$  est le vecteur d'état à l'instant  $k$  ;

–  $\Phi_k = \begin{pmatrix} A_{r_k} & D_{r_k} \\ \mathbf{0}_3^T & 1 \end{pmatrix}$  est la matrice d'évolution de l'état  $k$  à

l'état  $k+1$ , avec  $(A_{r_k}, D_{r_k})$  le mouvement perçu dans le plan rétinien ;

–  $Y_k^i$  est le vecteur de mesures à l'instant  $k$  ;

–  $V_k$  et  $W_k$  sont des vecteurs de bruits blancs gaussiens additifs de moyennes nulles et de matrices de covariance respectives  $Q_k = \sigma_Q^2 I_4$  et  $R_k = \sigma_R^2 I_4$ .

Le premier problème est d'obtenir une estimation correcte du mouvement « rétinien »  $(A_{r_k}, D_{r_k})$ .

Rappelons que les mouvements 3D de l'objet  $(R_l, \vec{T}_l)$  et de la caméra  $(R_c, \vec{T}_c)$  sont connus. L'évolution du point 3D correspondant (éq. 13) est donnée par :

$$P_{k+1}^i = A_k P_k^i + D_k \quad (30)$$

La projection perspective du point  $P_k^i (X_k^i, Y_k^i, Z_k^i)^T$  donne la relation suivante, à l'instant  $k$  :

$$p_{r_k}^i = \frac{P_k^i}{Z_k^i}$$

En exploitant cette relation à l'instant  $k+1$ , on peut réécrire l'équation 30 :

$$Z_{k+1}^i p_{r_{k+1}}^i = A_k (Z_k^i p_{r_k}^i) + D_k \quad (31)$$

soit :

$$p_{r_{k+1}}^i = \frac{Z_k^i}{Z_{k+1}^i} A_k p_{r_k}^i + \frac{1}{Z_{k+1}^i} D_k \quad (32)$$

Ainsi, on identifie le mouvement perçu dans le plan rétinien :

$$- A_{r_k} = \frac{Z_k^i}{Z_{k+1}^i} A_k ;$$

$$- \vec{D}_{r_k} = \frac{1}{Z_{k+1}^i} \vec{D}_k ;$$

Les rapports  $\frac{Z_k^i}{Z_{k+1}^i}$  et  $\frac{1}{Z_{k+1}^i}$  sont estimés par moindres carrés à partir de l'équation d'évolution des points dans le plan rétinien (éq. 32). Enfin, il est souhaitable d'avoir une bonne initialisation du filtre pour converger rapidement : on prendra la première mesure à notre disposition. On utilise également de faibles variances de bruits de modèle pour obtenir un lissage efficace des données.

## 5. résumé de la méthode

La méthode de reconstruction proposée peut donc se décomposer principalement en trois parties :

- une première étape de suivi des primitives 2D qui est réalisée par un filtre de Kalman dont les équations (éqs. 3, 4 et 5) permettent de réaliser un étiquetage des segments dans la séquence d'images.
- une étape d'estimation des coordonnées de Plücker du segment dans le plan rétinien à partir des positions 2D exprimées dans le plan image numérique (éqs. 25) ; afin de régulariser les résultats obtenus, on procède ensuite à une étape de lissage des coordonnées « rétiniennes » via la modélisation d'état 29.

- l'étape de reconstruction utilise les coordonnées lissées  $(\vec{u}_k, \vec{l}_k, \theta_k)$  pour estimer la structure 3D à chaque instant à partir des relations 15, 19 et 24.

Les trois étapes sont réalisées à chaque instant. Pour un traitement dynamique, on itère les étapes 1, 2, et 3.

## 6. résultats

La méthode proposée est appliquée à deux séquences d'images distinctes :

- des images de synthèse en niveaux de gris d'objets en mouvement de rotation et de translation selon les trois axes, pour lesquelles l'observateur décrit le même type de mouvement ;
- des images réelles en niveaux de gris d'un objet en mouvement de translation selon un axe, pour lesquelles l'observateur est fixe.

Les primitives géométriques utilisées sont ici le point et le segment de droite. Elles sont obtenues par chaînage et approximation polygonale des contours dans les mêmes conditions opératoires.

### 6.1. données simulées

Nous présentons les résultats de simulation de la méthode sur une séquence de 36 images synthétiques [PTh97] d'intensité  $(512 \times 512)$  pixels) composée de trois objets (C0, C1, C2) en mouvement de rotation/translation. Les figures 2 montrent des échantillons de la séquence aux instants 4, 8, 17, 21, 26 et 34.

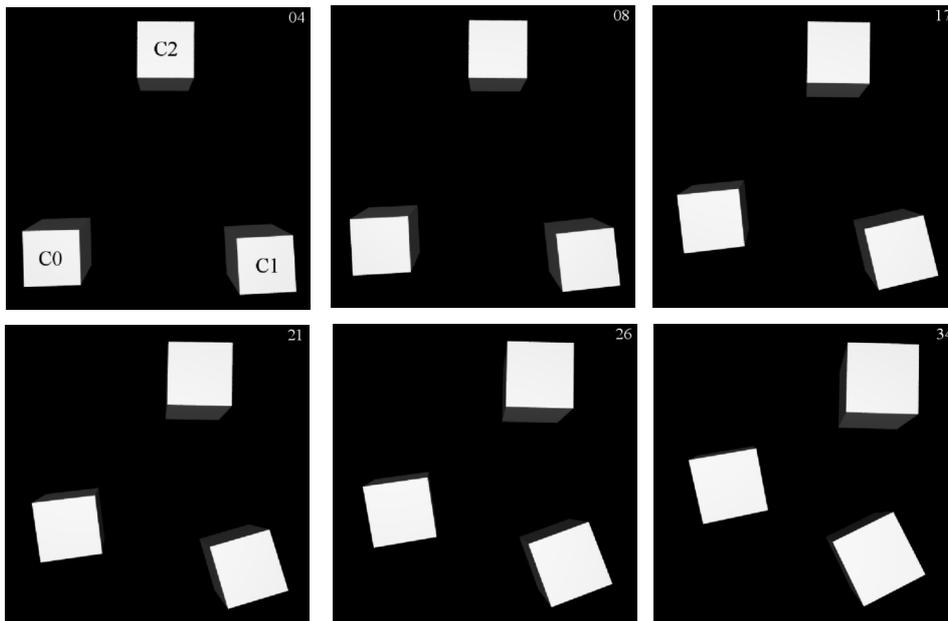


Figure 2. – Images de la séquence de synthèse.

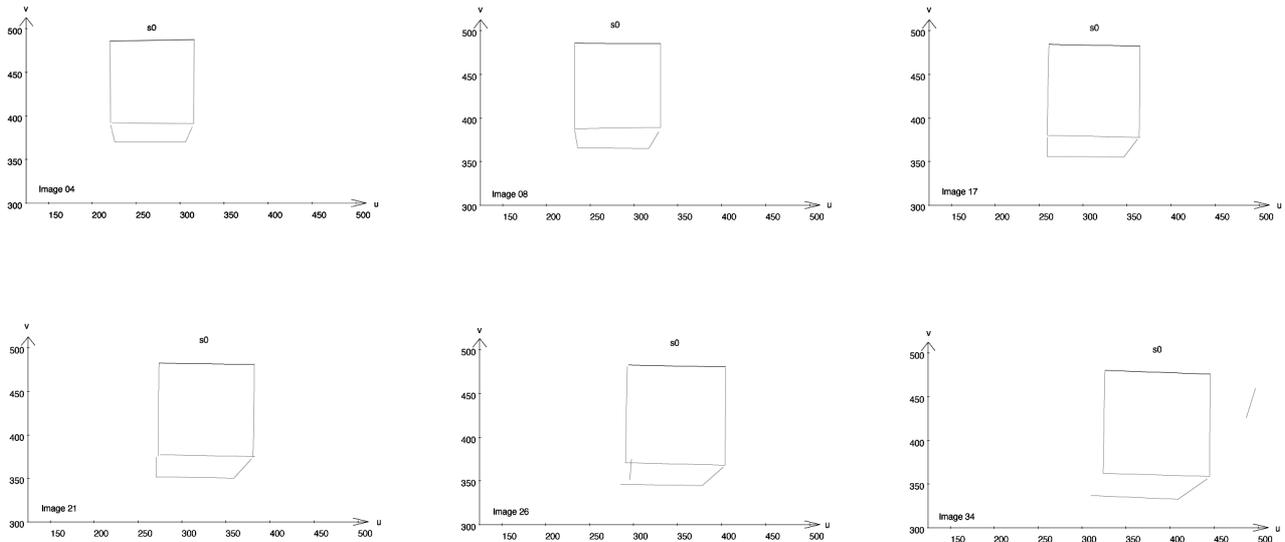


Figure 3. – Suivi des segments dans les images.

La caméra est décrite par un modèle sténopé dont les paramètres intrinsèques sont :

- $f = 1$  ;
- $\alpha_u = \alpha_v = 512$  ;
- $u_0 = v_0 = 256$ .

Les mouvements 3D (rotation et translation) inter-image de l'objet étudié (C2) et de la caméra sont résumés dans le tableau 1.

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

	$R_x$	$R_y$	$R_z$	$T_x$	$T_y$	$T_z$
Caméra	$0.2^\circ$	$0.1^\circ$	$0.15^\circ$	$-0.05^\circ$	$-0.1^\circ$	$0.15^\circ$
C2	$0.1^\circ$	$0.1^\circ$	$0.1^\circ$	$0.1^\circ$	$0.15^\circ$	$0.05^\circ$

### 6.1.1. suivi des segments de droite

On donne aux figures 3 les résultats de mise en correspondance temporelle et spatiale, aux mêmes instants, de l'ensemble des primitives 2D visibles. Pour le segment  $s_0$  en particulier, est tracée l'évolution de la distance de Mahalanobis en fonction du temps (figure 4) : le seuil de décision a été fixé à 30. Par convention, l'absence de mesure retenue est figurée par une distance nulle : le filtre fonctionne alors en prédiction pure.

### 6.1.2 reconstruction des objets 3D

Les figures 5 montrent l'évolution des coordonnées de Plücker 2D du segment  $s_0$  obtenues par la méthode proposée. On a

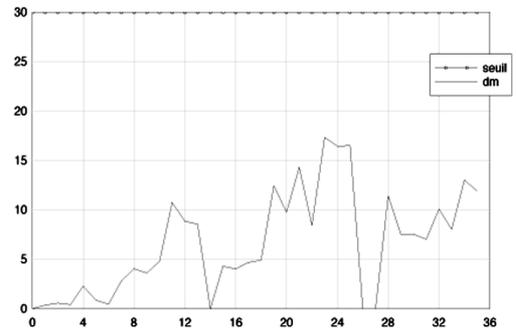


Figure 4. – Évolution de la distance de Mahalanobis (segment  $s_0$ ).

représenté sur ces courbes les coordonnées réelles, lissées et brutes. On remarque rapidement que l'erreur entre l'estimation brute d'une coordonnée (obtenue par inversion de la matrice des paramètres intrinsèques) et sa valeur exacte (par projection perspective des coordonnées 3D réelles des extrémités) fluctue assez fortement. Le lissage des coordonnées des extrémités conduit à une légère erreur sur les coordonnées 2D ( $\vec{u}, \vec{l}, \vec{\theta}$ ) du segment tout en supprimant les effets indésirables sur les coordonnées brutes.

Les courbes des figures 6 montrent les erreurs à la reconstruction des paramètres ( $\vec{U}, \vec{L}, \vec{\Theta}$ ) sur l'ensemble de la séquence, pour le segment  $s_0$ . On n'a tracé ici que les faces que l'on peut reconstruire, car on constate sur les figures 3 de suivi des primitives 2D que l'on « perd » un des segments par absence de mesures à mettre en correspondance.

On remarque donc sur les figures 6 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

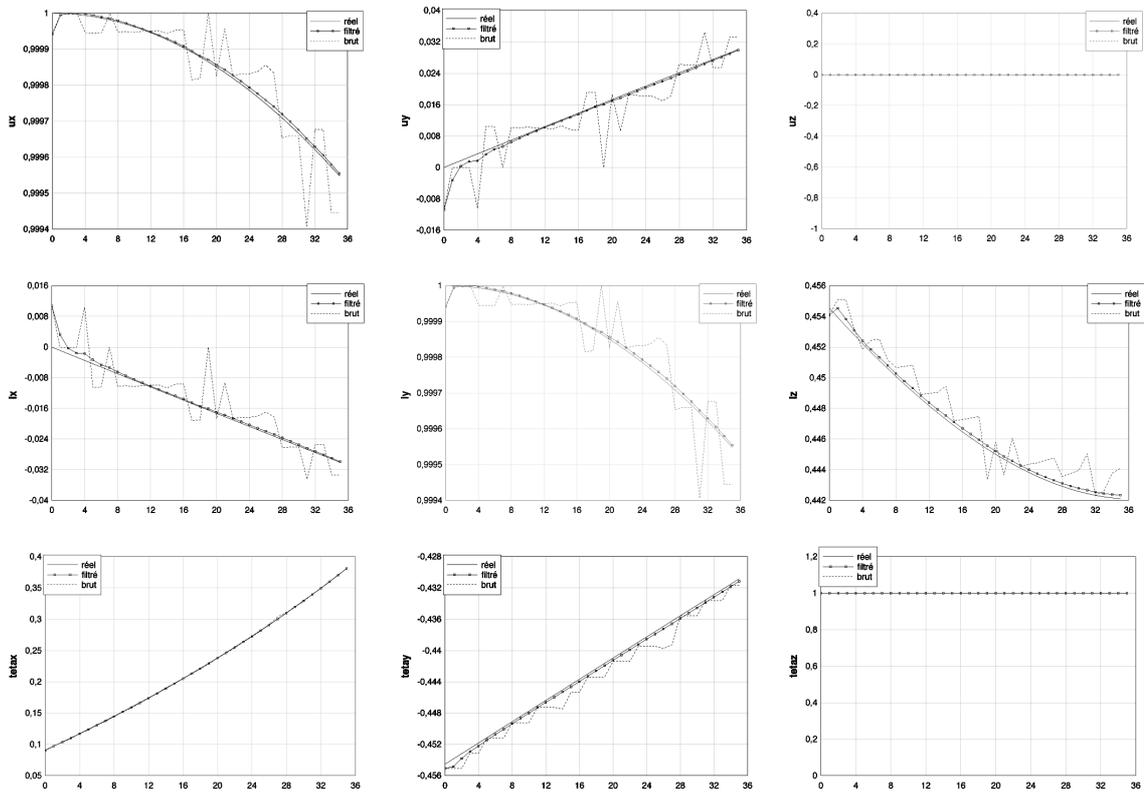


Figure 5. – Coordonnées de Plücker 2D du segment  $s_0$  :  $\vec{u}$ ,  $\vec{l}$  et  $\vec{\theta}$ .

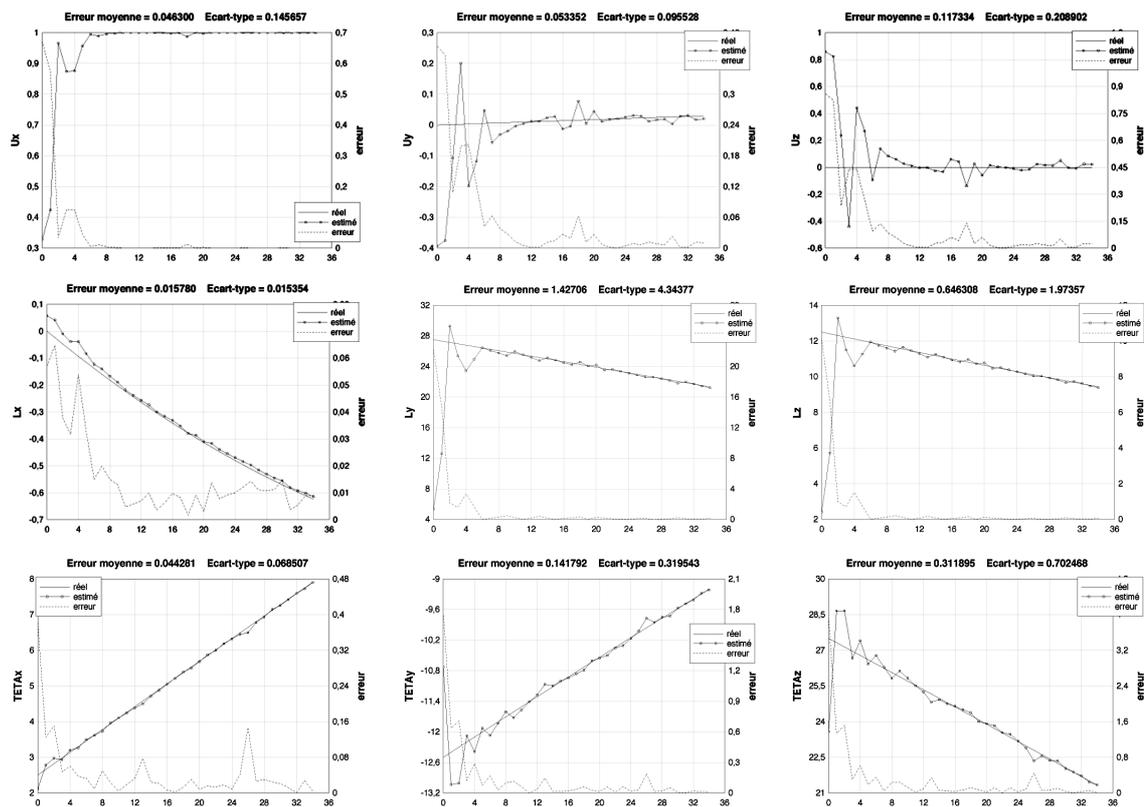


Figure 6. – Reconstruction 3D du segment  $s_0$  :  $\vec{U}$ ,  $\vec{L}$  et  $\Theta$ .

# Estimation des positions d'objets 3D par analyse monoculaire

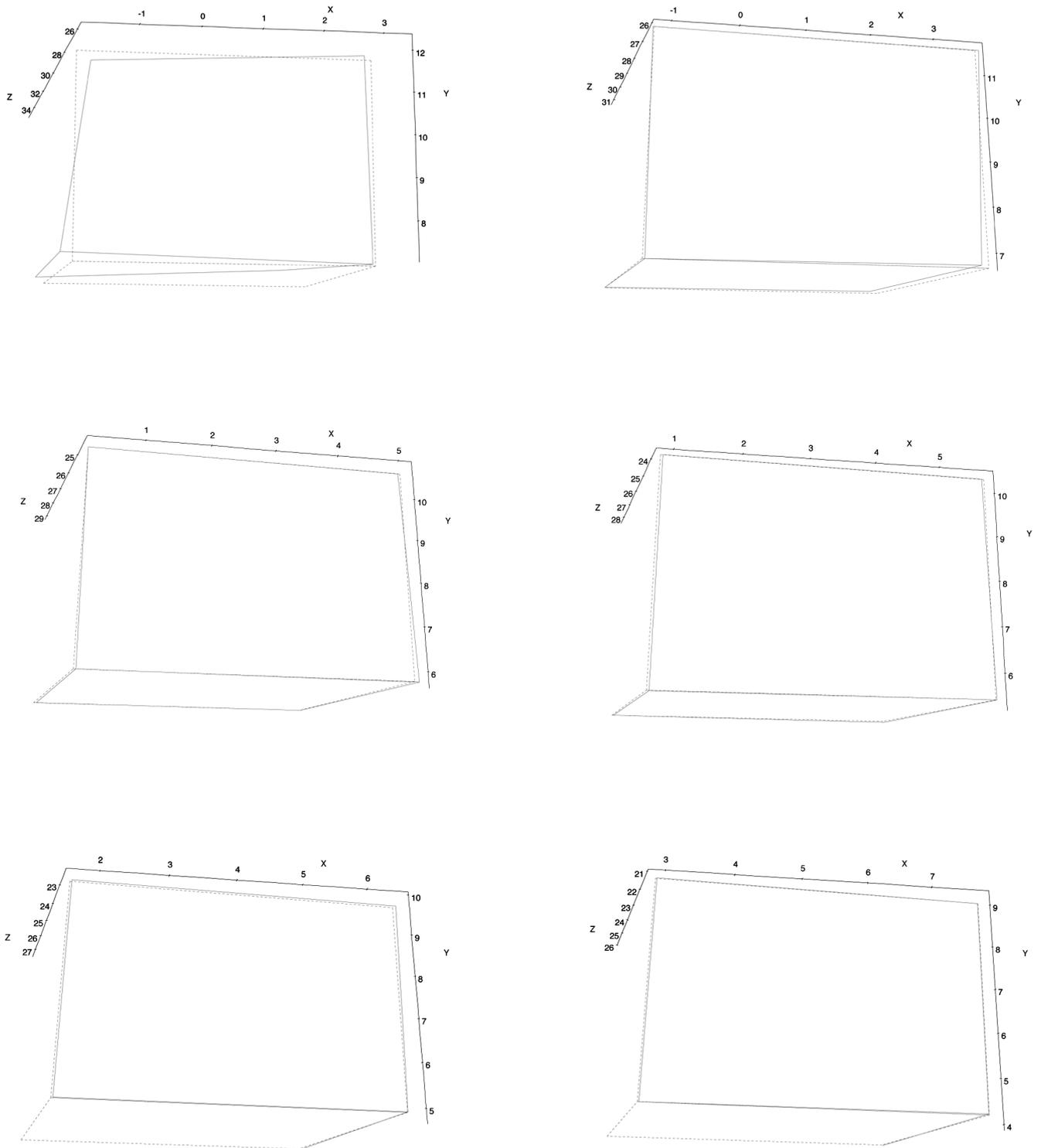


Figure 7. – Reconstruction 3D de l'objet C2 (objet réel en pointillé).

liée à l'étape d'extraction des primitives 2D et à l'estimation des paramètres de calibrage [Zel96].

Suivant le même principe, nous présentons une reconstruction de l'ensemble des segments suivis (obtenus par chaînage de leurs extrémités, éq. 13), formant la face avant et inférieure du cube choisi à différents instants (figure 7).

## 6.2. données réelles

Nous présentons (figure 8) les résultats de la méthode sur une séquence d'images réelles<sup>1</sup> (768 × 576 pixels) d'un objet ayant un mouvement de forte amplitude (il sort du champ de l'image en 16 itérations).

La caméra est décrite par un modèle sténopé dont les paramètres intrinsèques sont :

- $f = 0.025$  ;
- $\alpha_u = -2520.37, \alpha_v = -2421.54$  ;
- $u_0 = 399.98, v_0 = 264.41$ .

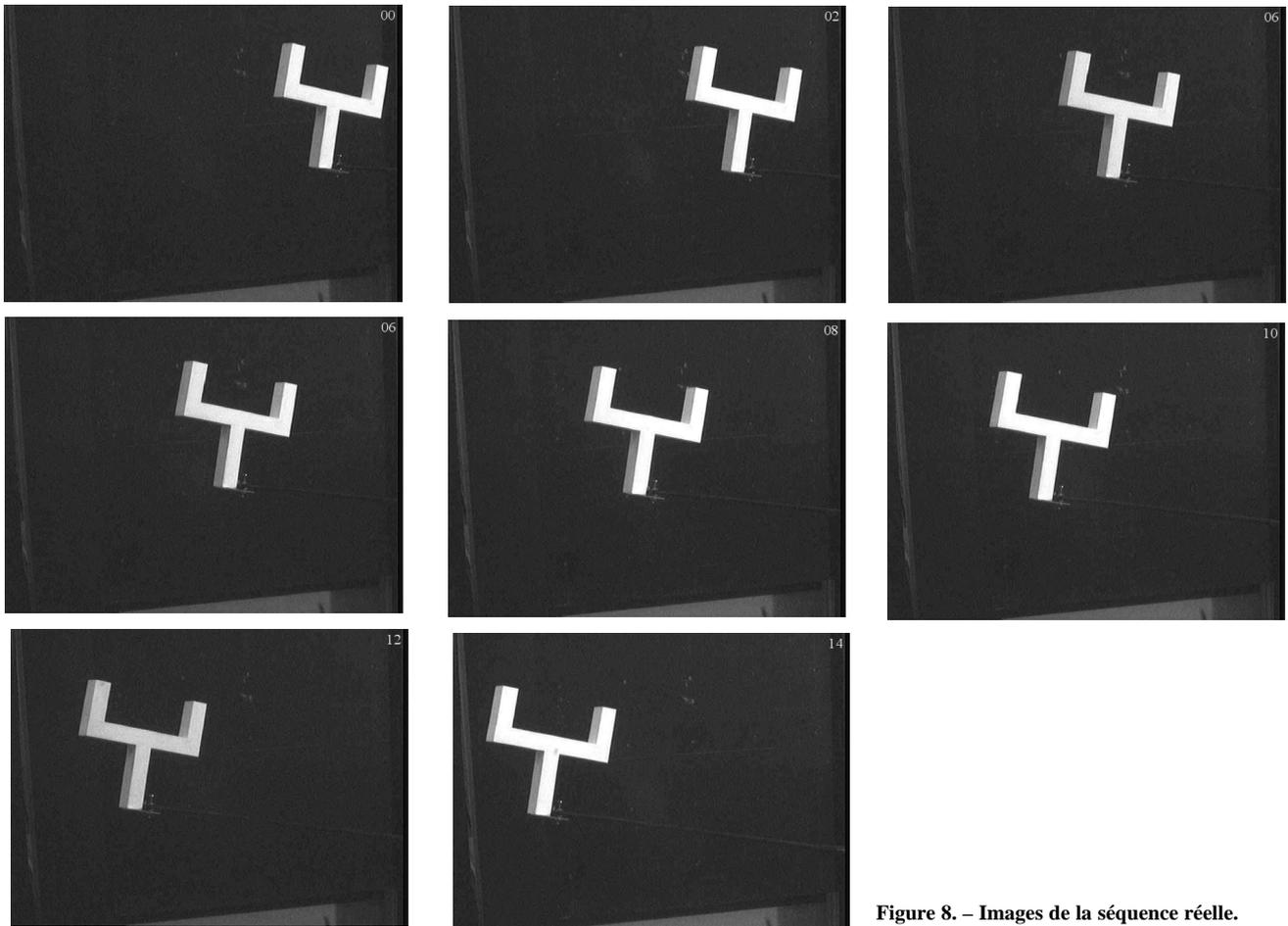


Figure 8. – Images de la séquence réelle.

Tableau 2. – Mouvement 3D de l'objet.

	$R_x$	$R_y$	$R_z$	$T_x$	$T_y$	$T_z$
Pince	0°	0°	0°	-0.02°	0	0

Le mouvement 3D (translation pure) inter-image de l'objet étudié est résumé dans le tableau 2.

### 6.2.1. poursuite des segments de droite

Les résultats de mise en correspondance temporelle et spatiale de l'ensemble des primitives 2D étudiées sont donnés à différents instants (figure 9). On a tracé l'évolution de la distance de Mahalanobis en fonction du temps, pour le segment  $s_0$  (figure 10) : le seuil de décision reste identique au cas des images de synthèse (égal à 30). Les instants où celle-ci chute à zéro correspondent à une absence de mesure pertinente: le filtre fonctionne alors uniquement en prédiction.

1. Cette séquence d'images a été réalisée et calibrée au laboratoire CEMIF par M. Shaheen

# Estimation des positions d'objets 3D par analyse monoculaire

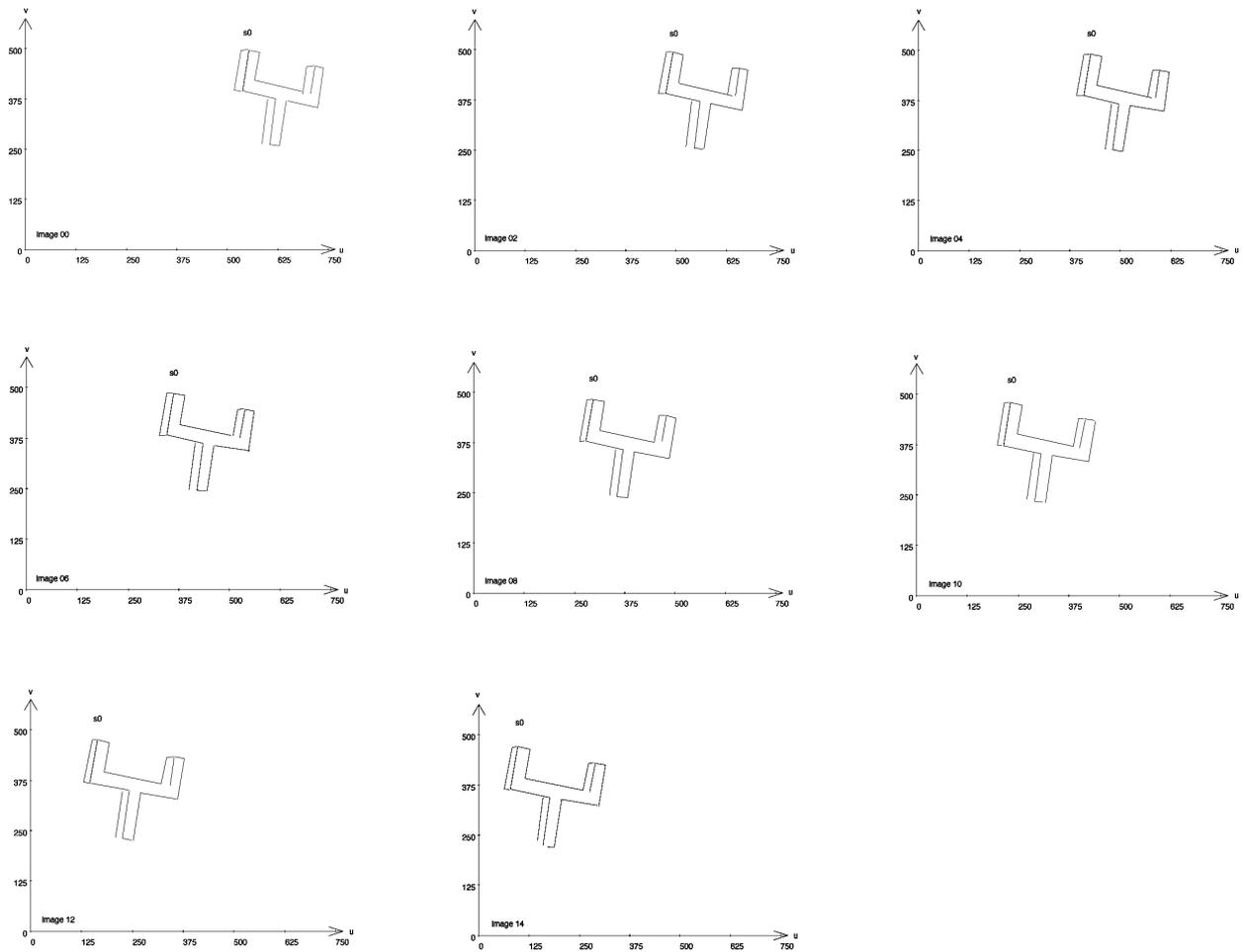


Figure 9. – Suivi des segments dans les images.

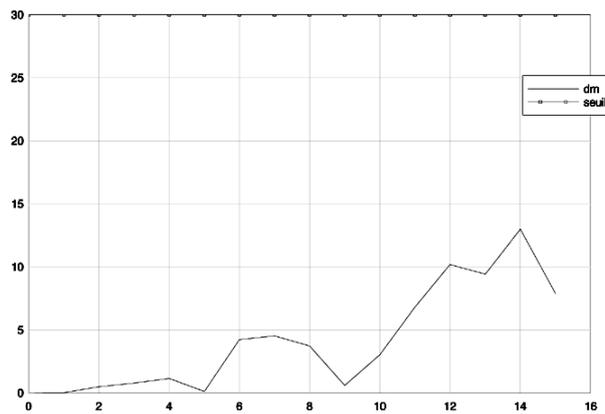


Figure 10. – Évolution de la distance de Mahalanobis (segment  $s_0$ ).

# Estimation des positions d'objets 3D par analyse monoculaire

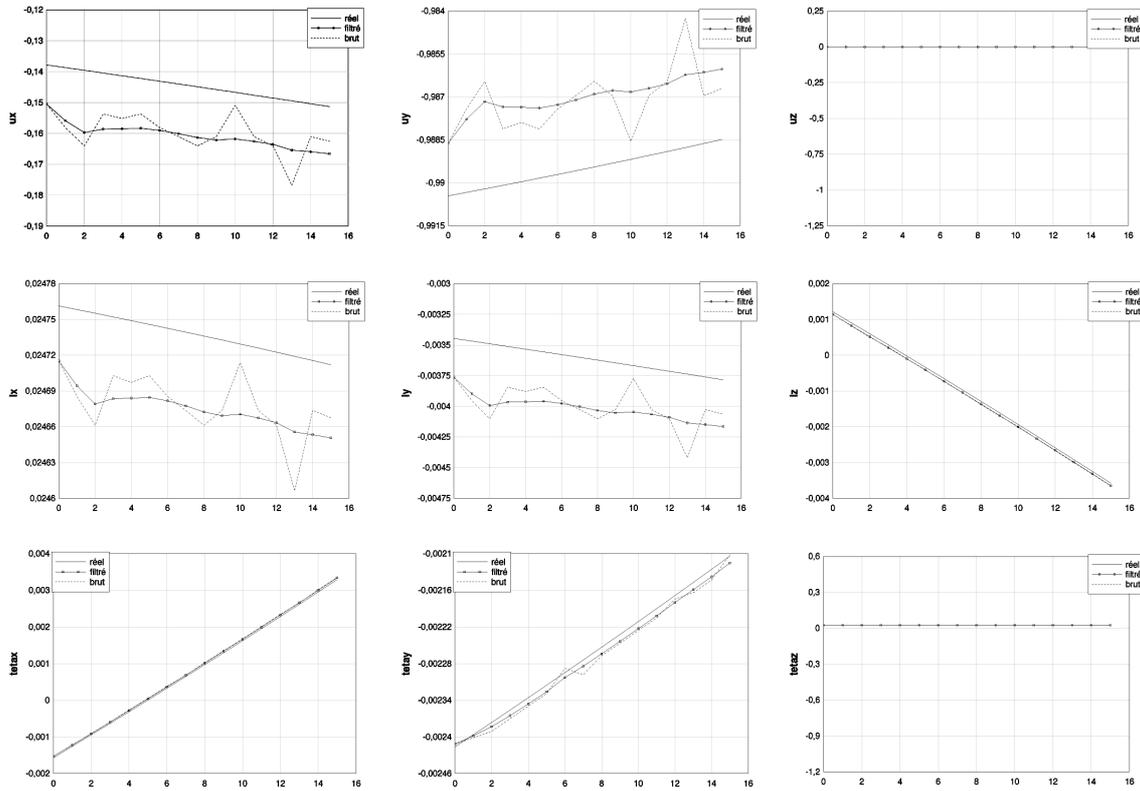


Figure 11. – Coordonnées de Plücker 2D du segment  $s_0$  :  $\vec{u}$ ,  $\vec{l}$  et  $\Theta$ .

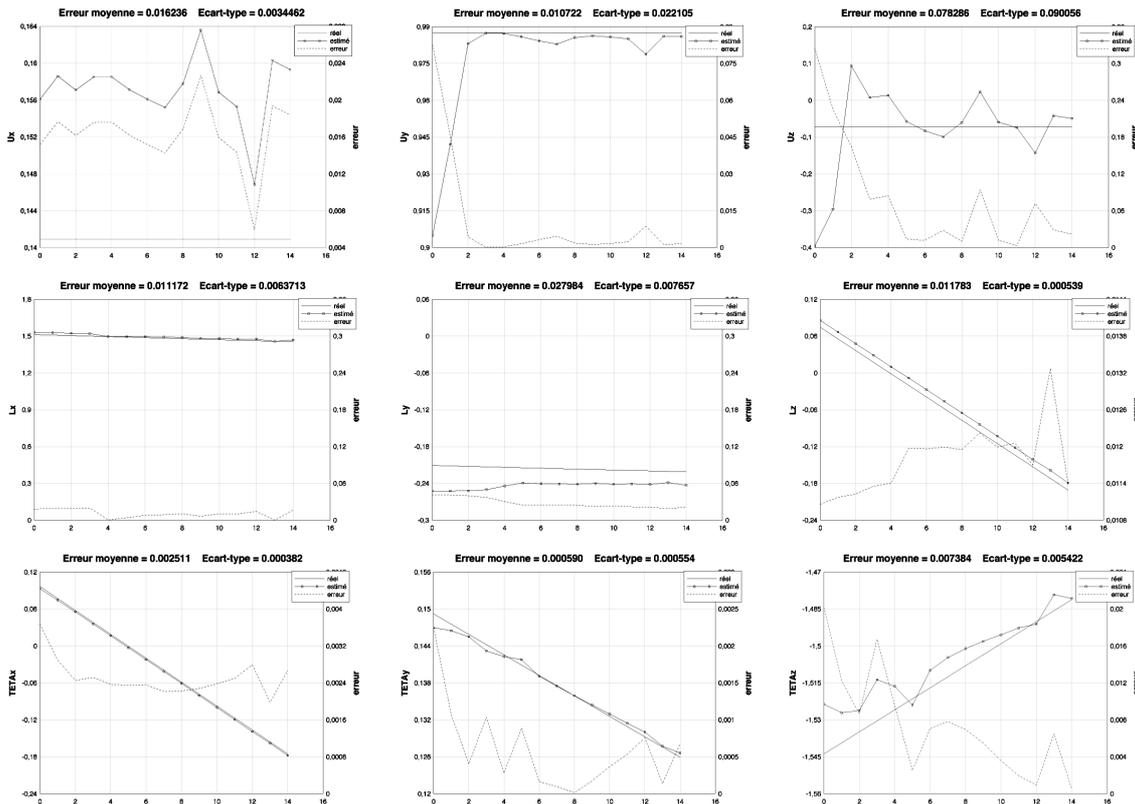
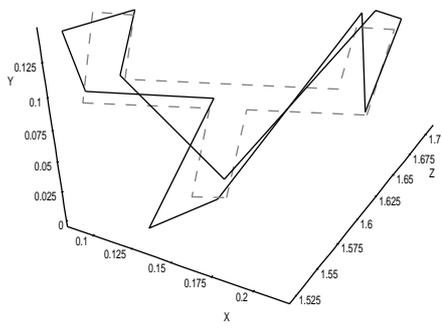
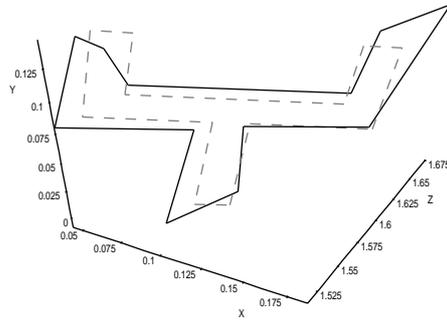


Figure 12. – Reconstruction 3D du segment  $s_0$  :  $\vec{U}$ ,  $\vec{L}$  et  $\Theta$ .

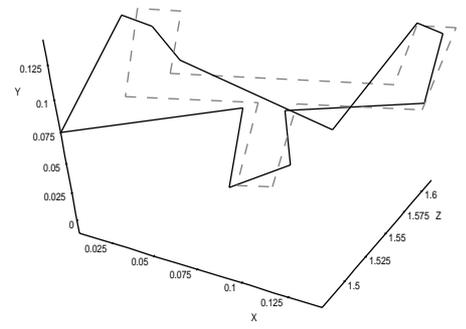
# Estimation des positions d'objets 3D par analyse monoculaire



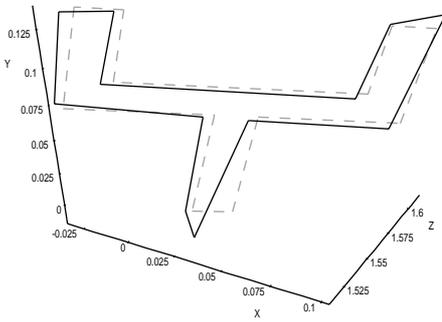
Scène 0



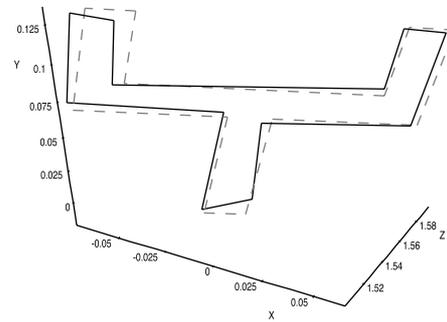
Scène 2



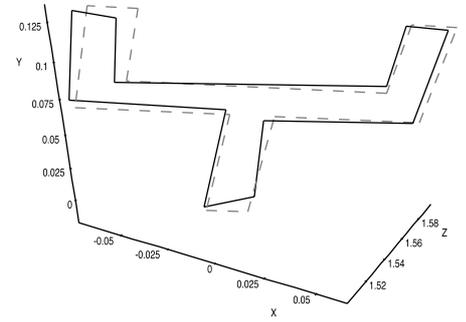
Scène 4



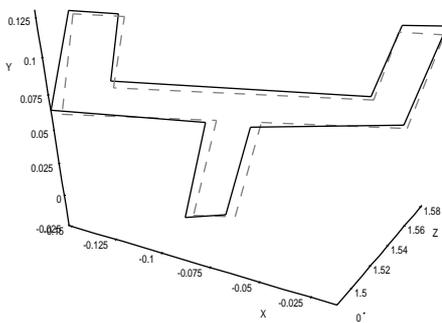
Scène 6



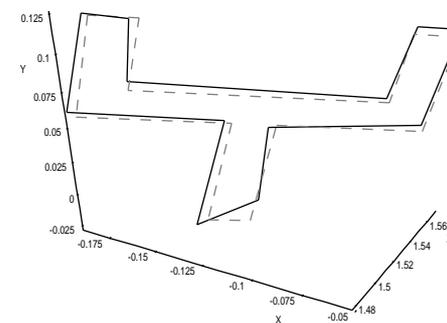
Scène 8



Scène 10



Scène 12



Scène 14

Figure 13. – Reconstruction 3D de la pince (pince réelle en pointillé).

### 6.2.2. reconstruction des objets

Comme précédemment, les courbes suivantes montrent l'évolution des coordonnées de Plücker 2D du segment  $s_0$  (figure 11). On a également représenté sur ces courbes les coordonnées réelles, lissées et brutes. Ajoutons que l'on est dans un cas d'étude réel et que l'erreur sur les paramètres intrinsèques de la caméra est plus importante que dans le cas précédent. Le filtrage des données brutes conduit à une erreur résiduelle mais supprime les discontinuités préjudiciables à l'étape de reconstruction.

Les courbes suivantes montrent les erreurs à la reconstruction des paramètres  $(\vec{U}, \vec{L}, \vec{\Theta})$  à différents instants, pour le segment  $s_0$  (figure 12). On remarque notamment que l'erreur de reconstruction 3D reste relativement faible au cours du temps et que la convergence de l'algorithme est rapide (quelques itérations).

Ce bon comportement est évidemment dû à la précision de l'étape de segmentation et d'identification des paramètres de calibrage.

Nous présentons également une reconstruction de l'ensemble des segments suivis qui décrivent l'objet étudié, pour la totalité de la séquence (figure 13). Ceci permet de constater les bonnes précisions en reconstruction malgré un fort déplacement inter-images.

## 7. conclusion

Cet article propose une méthode de reconstruction 3D à partir des informations de mouvement (caméra et objet). L'originalité réside dans l'exploitation des propriétés géométriques de la 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 sur données synthétiques et réelles, ainsi qu'une robustesse à de forts déplacements grâce, notamment, au calcul adaptatif du modèle dynamique. L'obtention des paramètres de mouvement 3D reste un point clé que l'on peut résoudre par des techniques telles que [Krü98, VLF94]. On peut également envisager l'immersion de cette méthode dans un système de perception multi-capteurs, avec l'ajout d'un capteur actif, comme une caméra laser de profondeur, qui permettrait d'en obtenir une estimation [BNB00, BNB99a].

*Manuscrit reçu le 14 novembre 2000*

### BIBLIOGRAPHIE

[AM99] P.M.Q. Aguiar and J.M.F. Moura. A fast algorithm for rigid structure from image sequences. In *International Conference on Image Processing*, page 27A02, 1999.

- [AS00] S. Avidan and A. Shashua. Trajectory triangulation : 3D reconstruction of moving points from a monocular image sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(4) : 348--357, April 2000.
- [Aya89] N. Ayache. *Vision Stéréoscopique et Perception Multisensorielle*. InterEditions, Paris, 1989.
- [BAS<sup>+</sup>87] C. Brown, J. Aloimonos, M. Swain, P. Choup, and A. Basu. Texture, contour, shape and motion. *Pattern Recognition Letters*, 5 : 151-169, 1987.
- [BG98] C. Ballester and M. Gonzalez. Affine invariant texture segmentation and shape from texture by variational methods. *Journal of Mathematical Imaging and Vision*, 9 (2) : 141-171, September 1998.
- [BNB99a] C. Boucher, J.-C. Noyer, and M. Benjelloun. 3D structure and motion estimation using range and intensity images. In *Proceedings of the 33<sup>rd</sup> IEEE Asilomar Conference on Signals, Systems and Computers*, pages 1431-1435, Pacific Grove, USA, October 1999.
- [BNB99b] C. Boucher, J.-C. Noyer, and M. Benjelloun. Estimation des positions d'objets 3D à partir d'une séquence d'images monoculaire. In *Actes du 17<sup>ème</sup> Colloque GRETSI*, pages 407-410, Vannes, FRANCE, September 1999.
- [BNB00] C. Boucher, J.-C. Noyer, and M. Benjelloun. 3D structure and motion recovery by fusing range and intensity image sequences. In *Proceedings of the 3rd International Conference on Information Fusion*, Paris, France, July 2000.
- [Bou00] C. Boucher. *Contribution à la fusion d'informations par filtrage non-linéaire : application à l'estimation de la structure du mouvement 3D dans un contexte multi-capteurs*. PhD thesis, Université du Littoral Côte d'Opale, Octobre 2000.
- [BSC95] 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.
- [BSL93] Y. Bar-Shalom and X.-R. Li. *Estimation and tracking: principles, techniques and software*. Artech House, 1993.
- [BSL95] Y. Bar-Shalom and X. R. Li. *Multitarget multisensor tracking*. YBS Publication, 1995.
- [BZM97] P.A. Beardsley, A.-Zisserman, and D.W. Murray. Sequential updating of projective and affine structure from motion. *International Journal of Computer Vision*, 23(3):235--259, 1997.
- [Csu96] G. Csurka. *Modélisation projective des objets tridimensionnels en vision par ordinateur*. PhD thesis, Université de Nice Sophia Antipolis, Avril 1996.
- [DA89] U.R. Dhond and J.K. Aggarwal. Structure from stereo: A review. *IEEE Transactions on systems, man and cybernetics*, 19(6):1489-1510, November 1989.
- [Der87] R. Deriche. Using Canny's criteria to derive an optimal edge detector recursively implemented. *The International Journal of Computer Vision*, April 1987.
- [DF90] R. Deriche and O. Faugeras. Tracking line segments. *Image and Vision Computing*, 8(4) : 261-270, 1990.
- [DRLR89] M. Dhome, M. Richetin, J. T. Lapeste, and G. Rives. Determination of the attitude of {3D} objects from a single perspective view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11 : 1265-1278, December 1989.
- [DZB92] S. Demey, A. Zisserman, and P.A. Beardsley. Affine and projective structure from motion. In *British Machine Vision Conference*, pages 49-58, 1992.
- [ER87] B. Espiau and P. Rives. Estimation récursive de primitives 3D au moyen d'une caméra mobile. Technical Report 0652, INRIA, 1987.
- [FA97] C. Fermuller and Y. Aloimonos. On the geometry of visual correspondence. *International Journal of Computer Vision*, 21(3) : 223-247, February 1997.
- [Fau93] O. Faugeras. *Three-Dimensional Computer Vision : A Geometric Viewpoint*. MIT Press, 1993.

- [FIH99] D. A. Forsyth, S. Ioffe, and J. Haddon. Bayesian structure from motion. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 660-665, 1999.
- [Gar93] J. Garding. Direct estimation of shape from texture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11) : 1202-1208, November 1993.
- [Gir87] G. Giraudon. Chaînage efficace de contour. Technical Report 605, INRIA, Février 1987.
- [HM95] R. Horaud and O. Monga. *Vision par ordinateur : outils fondamentaux*. Hermès, Paris, 1995.
- [HN94] T. S. Huang and A. N. Netravali. Motion and structure from feature correspondences : A review. *Proceedings of the IEEE*, 82(2) : 252-268, February 1994.
- [HS87] B. K. P. Horn and B. G. Schunck. New methods for matching 3D objects with single perspective views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3 : 401-412, May 1987.
- [JJ84] C. Jerian and R. C. Jain. Determining motion parameters for scenes with translation and rotation. *PAMI*, 6(4) : 523-530, July 1984.
- [KC89] K.I. Kanatani and T.C. Chou. Shape from texture: General principle. In *IEEE Computer Vision and Pattern Recognition*, pages 578-583, 1989.
- [Krü98] Stephan Krüger. *Motion Analysis and Estimation Using Multiresolution Affine Models*. PhD thesis, University of Bristol, July 1998.
- [LH88] Y. Liu and T. S. Huang. A linear algorithm for determining motion and structure from line correspondences. *Computer Vision, Graphics, and Image Processing*, 44(1) :35-57, 1988.
- [LHP80] H. C. Longuet-Higgins and K. Prazdney. The interpretation of moving retinal image. *Proc. Roy. Soc. London*, 208:385-397, July 1980.
- [MDMC91] O. Monga, R. Deriche, G. Malandain, and J.-P. Cocquerez. Recursive filtering and edge tracking : Two primary tools for 3D edge detection. *Image and Vision Computing*, 9 : 203-214, 1991.
- [OG99] J. Oliensis and Y. Genc. New algorithms for two-frame structure from motion. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 737-744, 1999.
- [PTh97] POV-Team and C. J. Cason (<http://www.povray.org>). POV-Ray(tm) rendering engine for Windows, 1991-1997. Version 3.02.
- [Sdi93] A. Sdiguï. *Contribution à l'analyse de scènes dynamiques par approche monoculaire : Application au suivi temporel 2D et à la reconstruction 3D de segments*. PhD thesis, Université d'Evry Val d'Essone, Octobre 1993.
- [Sha95] L. S. Shapiro. *Affine Analysis of Image Sequences*. Cambridge University Press, 1995.
- [TK95] C. J. Taylor and D. J. Kriegman. Structure and motion from line segments in multiple images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(11) : 1021-1032, November 1995.
- [VLF94] T. Viéville, Q.-T. Luong, and O. Faugeras. Motion of points and lines in the uncalibrated case. *International Journal of Computer Vision*, 1994.
- [Xie89] M. Xie. *Contribution à la vision dynamique: reconstruction d'objets 3D polyédriques par une Caméra mobile*. PhD thesis, Université de Rennes I, Juin 1989.
- [Zel96] C. Zeller. *Calibration projective affine et euclidienne en vision par ordinateur et application à la perception tridimensionnelle*. PhD thesis, Ecole Polytechnique, Février 1996.
- [Zha95] Z. Zhang. Estimating motion and structure from correspondences of line segments between two perspective images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(12) : 1129--1139, December 1995.

## LES AUTEURS

### Christophe BOUCHER



Christophe Boucher a obtenu son DEA d'Automatique et d'Informatique Industrielle à l'Université de Lille I en 1996. En octobre 2000, il a obtenu sa thèse de doctorat en Génie Informatique, Automatique et Traitement du Signal, préparée au Laboratoire d'Analyse des Systèmes du Littoral, Université du Littoral Côte d'Opale. Ses activités de recherche se situent en fusion d'informations et vision par ordinateur. Il enseigne également à l'IUT du Littoral dans le cadre de son contrat d'Attaché Temporaire d'Enseignement et de Recherche.

### Mohammed BENJELLOUN



Mohammed Benjelloun est docteur de 3<sup>ème</sup> cycle de l'Université de Lille I en Génie Électrique et habilité à diriger des recherches de l'Université d'Evry Val d'Essonne en sciences. Il est actuellement Professeur à l'Université du Littoral Côte d'Opale dans le Laboratoire d'Analyse des Systèmes du Littoral (LASL, UPRES EA 2600), dont il est directeur depuis 1997. Son domaine d'intérêt en recherche est la vision artificielle et la fusion d'informations.

### Jean-Charles NOYER



Jean-Charles Noyer a obtenu le titre de docteur de l'Université Paul Sabatier en 1996 en Automatique et Traitement du Signal, au sein du LAAS à Toulouse. Il est Maître de Conférences à l'Université du Littoral Côte d'Opale depuis 1997 et travaille au Laboratoire d'Analyse des Systèmes du Littoral à Calais. Son domaine de recherche concerne le filtrage non-linéaire et la fusion d'informations appliqués à la vision artificielle.