

# Examen Vision par Ordinateur

Documents autorisés

## 1 Localisation d'objets par un bras manipulateur équipé de stéréovision

L'objectif est ici de saisir un objet polyédrique connu à l'aide d'un bras manipulateur équipé d'une puis deux caméras montées sur l'effecteur, par une localisation 3D de l'objet à partir des images acquises par celle-ci (figure ??). L'étude sera menée pour une caméra puis étendue à une tête stéréoscopique étant supposé que un seul objet est sur le plan de travail. Nous supposons connu *a priori* : (i) le modèle CAO de l'objet polyédrique à saisir, (ii) une estimée initiale grossière, notée par les paramètres extrinsèques<sup>1</sup>  $(\alpha, \beta, \gamma, T_u, T_v, T_w)$  et la transformation homogène associée  $[\mathcal{M}_{oc1}]_{t=0}$ , entre le repère objet  $\mathcal{R}_o$  dans sa position initiale ( $t = 0$ ) et le repère caméra 1 noté  $\mathcal{R}_{c1}$  lors du processus de saisie. Le système mono- puis multi-caméra sera supposé étalonné *a priori*.

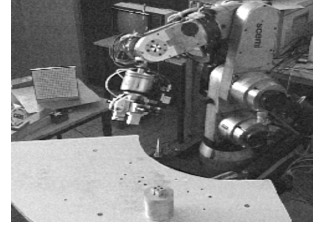


Figure 1: Bras robotisé.

1. Une boucle commande du bras-localisation de l'objet est mise en oeuvre pour saisir l'objet. Pour une image  $\mathcal{I}_{1,t}$  acquise à l'instant  $t$  par la caméra 1, la localisation repose sur des appariements 2D/3D entre segments de droite  $l_i$  extraits de  $\mathcal{I}_{1,t}$  et segments 3D du modèle CAO de l'objet projetés dans  $\mathcal{I}_{1,t}$ . Expliciter la relation mathématique entre un point 3D noté  $P_i^o = (X_i, Y_i, Z_i)'$ ,  $i = 1, 2$  extrémités du segment 3D noté  $L_i$  dans  $\mathcal{R}_0$  et leurs coordonnées image  $p_i = (u_i, v_i)'$ ,  $i = 1, 2$  en fonction de  $[\mathcal{M}_{oc1}]_t$  et  $[\mathcal{M}_{int}]$ .
2. Proposer des critères pour appairier des segments image  $l_i$  de  $\mathcal{I}_{1,t}$  avec les segments 3D  $L_i$  (du modèle CAO de l'objet) ainsi projetés dans  $\mathcal{I}_{1,t}$ . Doit-on privilégier une stratégie d'appariement priorisant ou non certaines primitives images ? Quelle hypothèse permet d'initialiser le processus d'appariement sur la première image acquise de la scène ( $t=0$ ) ?
3. La localisation 3D de l'objet dans  $\mathcal{R}_{c1}$  repose alors sur une résolution numérique et itérative à partir des appariements  $\{l_i, L_i\}_{i>3}$ . Cette résolution repose sur le critère  $F_i = N_i \cdot P_i^{c1}$  avec  $P_i^{c1} = [\mathcal{M}_{oc1}]_k \cdot P_i^o$  qui requiert à chaque itération  $k$  le calcul d'une jacobienne (matrice notée  $A$  sur le support de cours) et donc les dérivées de  $F_i$  par rapport aux paramètres  $(\alpha, \beta, \gamma, T_x, T_y, T_z)$ . On se focalise ici sur les dérivées  $\frac{\partial F_i}{\partial \alpha}$ ,  $\frac{\partial F_i}{\partial \beta}$ ,  $\frac{\partial F_i}{\partial \gamma}$ . L'estimation de ces paramètres à l'itération  $k$  vise à calculer un incrément<sup>2</sup> par rapport aux paramètres estimés à  $k-1$  de sorte que on peut approximer  $\alpha \sim 0, \beta \sim 0, \gamma \sim 0$ . Montrer avec ces approximations que :

$$\frac{\partial F_i}{\partial \alpha} = N_i \cdot \begin{pmatrix} 0 \\ -Z_i \\ Y_i \end{pmatrix}, \quad \frac{\partial F_i}{\partial \beta} = N_i \cdot \begin{pmatrix} Z_i \\ 0 \\ -X_i \end{pmatrix}, \quad \frac{\partial F_i}{\partial \gamma} = N_i \cdot \begin{pmatrix} -Y_i \\ X_i \\ 0 \end{pmatrix}$$

On s'appuiera sur les rappels et conventions suivants :

$$R_\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{pmatrix}, \quad R_\beta = \begin{pmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{pmatrix}, \quad R_\gamma = \begin{pmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

4. En déduire les dérivées  $\frac{\partial F_i}{\partial T_x}$ ,  $\frac{\partial F_i}{\partial T_y}$ ,  $\frac{\partial F_i}{\partial T_z}$  sur le même principe.
5. Le processus d'étalonnage non linéaire évoqué précédemment est transposable à un processus de localisation 3D d'objets à une variante près. Laquelle ? Justifier. Les critères à optimiser entre les deux méthodes sont alors différents. Expliquer cette différence.

<sup>1</sup> $\alpha, \beta, \gamma$  sont les angles d'Euler selon les axes x, y, z du repère.

<sup>2</sup>On parle alors de localisation par recalages successifs/itératifs.

6. On rajoute la seconde caméra en configuration stéréo. Avec cet ajout, quel est le gain attendu sur la localisation ? On souhaite alors étalonner le système à deux caméras. On étalonne alors indépendamment la seconde caméra. Décrire alors l'étape supplémentaire à réaliser pour étalonner le système complet. En quoi cette étape est-elle indispensable ? Soit  $[\mathcal{M}_{c_1 c_2}] = \{m_{ij}\}_{1 \leq i, j \leq 4}$  la matrice de passage entre caméras 1 et 2 estimée. Comment en déduire de façon précise la distance entre les centres optiques des deux caméras ? Cette étape a-t-elle une influence sur le processus de localisation ?
7. Une seconde caméra, cette fois-ci déportée dans l'environnement, serait-elle pertinente pour faciliter la saisie de l'objet ?

## CORRECTION

### 2 Localisation d'objets par un bras manipulateur équipé de stéréovision

1. La transformation met en jeu paramètres intrinsèques  $(u_0, v_0, \alpha_u, \alpha_v)$  et extrinsèques *via* la transformation homogène  $[\mathcal{M}_{oc1}]_t$ , soit :

$$\begin{pmatrix} s.u_i \\ s.v_i \\ s \end{pmatrix} = [\mathcal{M}_{int}].[\mathcal{M}_{oc1}]_t \cdot \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix}$$

2. Les critères sont : (1) similarité d'orientation entre arête projetée et segment image ( $D_1$ ), (2) distance entre arête projetée et segment image ( $D_2$ ), (3) recouvrement entre arête projetée et segment image ( $D_3$ ). Le score global d'appariement est alors une somme pondérée *e.g.* :  $score = (0,45 \times D_1) + (0,45 \times D_2) + (0,1 \times D_3)$  car on accorde une confiance relative au critère  $D_3$ . Oui, la stratégie priorise certains appariements *e.g.* les **segments image longs** plutôt que les courts car plus fiables. Pour initialiser ( $t=0$ ), on suppose que la position objet est connue **approximativement** dans le repère monde *i.e.* robot.

3. Pour rappel :  $F_i = N'_i.[R_\gamma R_\beta R_\alpha.P_i + T_{xyz}]$ . On déduit :

$$\frac{\partial F_i}{\partial \alpha} = N'_i.R_\gamma.R_\beta.\frac{\partial R_\alpha}{\partial \alpha}.P_i = N'_i \cdot \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = N'_i \cdot \begin{pmatrix} 0 \\ -Z_i \\ Y_i \end{pmatrix}$$

De même :

$$\frac{\partial F_i}{\partial \beta} = N'_i.R_\gamma.\frac{\partial R_\beta}{\partial \beta}.R_\alpha.P_i = N'_i \cdot \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = N'_i \cdot \begin{pmatrix} Z_i \\ 0 \\ -X_i \end{pmatrix}$$

De même :

$$\frac{\partial F_i}{\partial \gamma} = N'_i.\frac{\partial R_\gamma}{\partial \gamma}.R_\beta.R_\alpha.P_i = N'_i \cdot \begin{pmatrix} -Y_i \\ X_i \\ 0 \end{pmatrix}$$

4. On déduit :

$$\frac{\partial F_i}{\partial T_x} = N'_i.\frac{\partial T_{xyz}}{\partial T_x} = N'_i \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \frac{\partial F_i}{\partial T_y} = N'_i \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \frac{\partial F_i}{\partial T_z} = N'_i \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

5. Il suffit de fixer les paramètres intrinsèques dans le processus d'étalonnage. Critères 3D pour la méthode de localisation versus 2D pour la méthode d'étalonnage adaptée.
6. La localisation 3D sera plus précise, notamment le calcul de la profondeur. L'étape supplémentaire est la détermination du positionnement relatif caméra 1/caméra 2. Pour ce faire, on utilise une paire d'image caméra 1/ caméra 2 correspondant à une position commune de la mire. Cette étape est nécessaire pour trianguler et donc apparier. On utilise les éléments  $m_{ij}$  de la matrice  $M_{12}$  de passage correspondant à la translation. La distance entre les deux centres optiques des caméras est donnée par :

$$d = \sqrt{m_{14}^2 + m_{24}^2 + m_{34}^2}$$

Cette étape influence la précision de localisation au final.

7. Oui car à proximité de l'objet, la vision embarquée est inopérante (objet hors du champ de vue)...