

Examen Vision 3D

Tous documents autorisés - Tablette, téléphone et objets communicants interdits

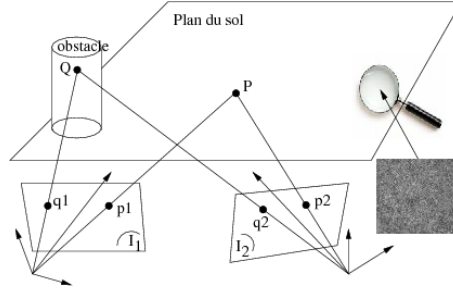


Figure 1: Principe de la détection d'obstacles par vision stéréo.

On souhaite mettre en oeuvre un algorithme de détection d'obstacles pour un véhicule mobile se déplaçant sur un sol plat à partir d'un système binoculaire embarqué. Cette méthode repose classiquement sur la mise en correspondance inter-image de primitives. **Toutes les questions, exceptées 6, 7 et 8 sont indépendantes.**

1. Caractériser et différencier **en une phrase** les techniques dites de stéréo-corrélation dense, stéréo-corrélation épars, stéréo-segment ? On considère que le plan du sol est texturé comme illustré sur la figure 1. Dans ce contexte et cadre applicatif, on privilégie une stratégie de stéréo-corrélation. Justifier.
2. L'application requiert un champ de vue induisant une focale $\in [7, 10]$ mm qui donne lieu à distorsion dans les images. On privilégie le modèle de distorsion ci après. Justifier le choix de ce modèle et préciser pour quels pixels image la distorsion est minimale/maximale.

$$\begin{cases} u_d &= u_0 + \alpha_u \cdot U \cdot (1 + k_1 \cdot R^2) \\ v_d &= v_0 + \alpha_v \cdot V \cdot (1 + k_1 \cdot R^2) \end{cases} \quad (1)$$

où (u_d, v_d) et (u, v) sont respectivement les coordonnées distordues et non distordues en posant : $U = \frac{u-u_0}{\alpha_u}$, $V = \frac{v-v_0}{\alpha_v}$, $R^2 = U^2 + V^2$.

3. On souhaite s'affranchir de la distorsion et donc générer une image "sans" distorsion grâce au modèle (1). La démarche est de construire incrémentalement cette image non distordue par association des coordonnées non distordues (u, v) et distordues (u_d, v_d) issues de l'image originale/distordue. Proposer un algorithme associé en précisant comment affecter les niveaux de gris dans l'image "non" distordue.
4. La stéréocorrélation dense repose sur le calcul de score de similarité qui peut être prohibitif en ressource CPU. Proposer quelques heuristiques pour réduire ces temps de calcul.
5. On supposera ici que le sol est une surface plane. Les coordonnées des points $\mathbf{P}_1 = (X_1, Y_1, Z_1)'$ exprimé dans le repère caméra 1 (noté R_{c_1}) et appartenant au plan du sol de normale $\mathbf{n} = (n_x, n_y, n_z)'$ vérifient une équation de la forme suivante. Ce plan est noté Π^c par la suite.

$$n_x \cdot X_1 + n_y \cdot Y_1 + n_z \cdot Z_1 = 1 \quad (2)$$

Sachant que l'on utilise une mire plane, comment étalonner en pratique le plan Π^c du sol ? On s'appuyera sur quelques développements mathématiques pour justifier vos propos.

6. On note $R = \{r_{ij}\}_{1 \leq i \leq 3, 1 \leq j \leq 3}$ la matrice de rotation résultant de $R_\gamma \cdot R_\beta \cdot R_\alpha$, et $\mathbf{t} = (t_x, t_y, t_z)'$ la translation exprimant le passage relatif entre les deux repères caméras. Soit I_1 (resp. I_2) l'image prise par la caméra 1 (resp. la caméra 2). Démontrer **par le calcul** que

le correspondant $\mathbf{p}_2 = (x_2, y_2, f_2)'$ de I_2 (et exprimé dans R_{c_2}) du pixel $\mathbf{p}_1 = (x_1, y_1, f_1)'$ dans I_1 et exprimé dans R_{c_1} vérifient la relation :

$$\begin{cases} x_2 &= f_2 \cdot \frac{r_{11} \cdot x_1 \cdot Z_1 + r_{12} \cdot y_1 \cdot Z_1 + r_{13} \cdot Z_1 \cdot f_1 + t_x \cdot f_1}{r_{31} \cdot x_1 \cdot Z_1 + r_{32} \cdot y_1 \cdot Z_1 + r_{33} \cdot Z_1 \cdot f_1 + t_z \cdot f_1} \\ y_2 &= f_2 \cdot \frac{r_{21} \cdot x_1 \cdot Z_1 + r_{22} \cdot y_1 \cdot Z_1 + r_{23} \cdot Z_1 \cdot f_1 + t_y \cdot f_1}{r_{31} \cdot x_1 \cdot Z_1 + r_{32} \cdot y_1 \cdot Z_1 + r_{33} \cdot Z_1 \cdot f_1 + t_z \cdot f_1} \end{cases} \quad (3)$$

On s'appuiera sur la relation de projection perspective reliant un point image $\{p_i\}_{i=1,2}$ au point 3D correspondant $\{P_i\}_{i=1,2}$, soit :

$$x_i = f_i \cdot \frac{X_i}{Z_i}, \quad y_i = f_i \cdot \frac{Y_i}{Z_i}$$

7. Montrer que l'équation (2) peut s'écrire :

$$\frac{1}{Z_1} = \frac{1}{f_1} \cdot n_x \cdot x_1 + \frac{1}{f_1} \cdot n_y \cdot y_1 + n_z$$

Déduire en posant $\mathbf{r}_i = (r_{i1} \ r_{i2} \ r_{i3})_{1 \leq i \leq 3}$ que :

$$\begin{cases} x_2 &= f_2 \cdot \frac{\mathbf{r}_1 \cdot \mathbf{p}_1 + t_x \cdot (\mathbf{n}' \cdot \mathbf{p}_1)}{\mathbf{r}_3 \cdot \mathbf{p}_1 + t_z \cdot (\mathbf{n}' \cdot \mathbf{p}_1)} \\ y_2 &= f_2 \cdot \frac{\mathbf{r}_2 \cdot \mathbf{p}_1 + t_y \cdot (\mathbf{n}' \cdot \mathbf{p}_1)}{\mathbf{r}_3 \cdot \mathbf{p}_1 + t_z \cdot (\mathbf{n}' \cdot \mathbf{p}_1)} \end{cases} \quad (4)$$

8. Montrer que la relation précédente peut s'écrire comme suit :

$$s. \begin{pmatrix} x_2 \\ y_2 \\ f_2 \end{pmatrix} = f_2 \cdot (\mathbf{R} + \mathbf{t} \cdot \mathbf{n}'). \begin{pmatrix} x_1 \\ y_1 \\ f_1 \end{pmatrix} \quad (5)$$

où $(\mathbf{R} + \mathbf{t} \cdot \mathbf{n}')$ est une matrice 3×3 notée homographie. Comment exploiter au final cette relation pour distinguer deux points 3D appartenant au sol ou à un obstacle, notamment Q et P sur la figure 1 ?

9. Notre système repose classiquement sur deux caméras. Quel serait l'apport d'une troisième caméra dans le processus d'appariement ?

CORRECTION

1. stéréo-corrélation éparses et denses sont relatifs au nombre de pixels à apparier. Stéréo-segment est relatif à l'appariement de segments. Pour la stéréo-corrélation, il faut une texture aléatoire pour permettre les appariements... ce qui est le cas ici.
2. Focale $\geq 7 \text{ mm}$ donc $r_1 t_0$. La distorsion est minimale pour (u_0, v_0) , qui n'est pas le centre de l'image, et maximale pour les coins image
3. Le principe est de parcourir l'image non distordue (vierge au départ). Pour chaque pixel de coordonnées non distordues (u, v) , caculer ses coordonnées distordues (u_d, v_d) à partir de l'équation (1). Appliquer le niveau $I(u_d, v_d)$ ou le niveau $I(\cdot)$ du plus proche voisin de (u_d, v_d) au pixel (u, v) de l'image non distordue. Le principe est illustré figure 2; les coordonnées non distordues (resp. distordues) sont ici notées (x, y) (resp. (x', y')).

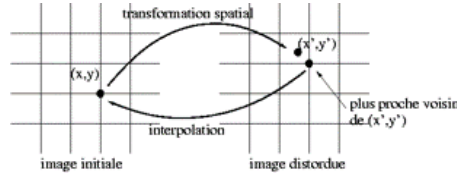


Figure 2: Correction de distorsion : principe.

4. On pourra évoquer notamment : (1) exploiter des images rectifiées et limiter la plage de disparité sur la droite épipolaire, (2) calculer la variance *a priori* de chaque pixel pour éviter la corrélation sur des zones homogènes, et (3) réduire la résolution des images avant traitement.
5. Il faut prendre une vue où la mire est placée sur le sol. Puis on cacule le plan du sol à partir des paramètres extrinsèques $(\alpha^s, \beta^s, \gamma^s, t_x^s, t_y^s, t_z^s)$ obtenues pour cette vue. Pour une position donnée de la mire, l'équation du plan dans l'espace sera défini par sa normale $N^c = (n_x, n_y, n_z)'$ et un point 3D quelconque P_1 dans le repère caméra 1. En notant $n_x^c \cdot X_1 + n_y^c \cdot Y_1 + n_z^c \cdot Z_1 + d = 0$ l'équation du plan Π , on a :

$$\begin{pmatrix} n_x^c \\ n_y^c \\ n_z^c \end{pmatrix} = R_{\gamma^s} \cdot R_{\beta^s} \cdot R_{\alpha^s} \cdot \begin{pmatrix} n_x^m \\ n_y^m \\ n_z^m \end{pmatrix}$$

où $(n_x^m, n_y^m, n_z^m)'$ est la normale au plan de la mire dans la repère mire. De plus, l'origine du repère mire *i.e.* $P^m = (0, 0, 0)'$ a pour coordonnées $P_1 = (t_x^s, t_y^s, t_z^s)'$ dans le repère caméra 1, noté R_{c_1} . Au total, on a :

$$d = -(n_x \cdot t_x^s + n_y \cdot t_y^s + n_z \cdot t_z^s)$$

Donc tous les paramètres du plan Π sont ainsi déterminés. On normalise pour obtenir l'équation (2).

6. Avec les notations, on exprime le passage entre R_{c_1} et R_{c_2} par :

$$\begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \\ 1 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{pmatrix}$$

On développe cette expression matricielle. On rappelle que :

$$x_1 = f_1 \cdot \frac{X_1}{Z_1}, \quad y_1 = f_1 \cdot \frac{Y_1}{Z_1}, \quad x_2 = f_2 \cdot \frac{X_2}{Z_2}, \quad y_2 = f_2 \cdot \frac{Y_2}{Z_2} \quad (6)$$

A partir de (6), on remplace X_2, Y_2, X_1, Y_1 dans l'expression ci-dessus. On déduit alors la relation (3).

7. En remplaçant avec (6) dans la relation (2), on montre que :

$$Z_1(n_x.x_1.\frac{1}{f_1} + n_y.y_1.\frac{1}{f_1} + n_z) = 1 \quad (7)$$

L'équation (3) devient :

$$\begin{cases} x_2 &= f_2 \cdot \frac{\mathbf{r}_1 \cdot \mathbf{p}_1 + t_x \cdot f_1 \cdot (1/Z_1)}{\mathbf{r}_3 \cdot \mathbf{p}_1 + t_z \cdot f_1 \cdot (1/Z_1)} \\ y_2 &= f_2 \cdot \frac{\mathbf{r}_2 \cdot \mathbf{p}_1 + t_y \cdot f_1 \cdot (1/Z_1)}{\mathbf{r}_3 \cdot \mathbf{p}_1 + t_z \cdot f_1 \cdot (1/Z_1)} \end{cases} \quad (8)$$

Puis, on remplace $\frac{1}{Z_1}$, à partir de (7), dans (8) pour démontrer la relation (4).

8. On développe :

$$\begin{cases} x_2 \cdot [\mathbf{r}_3 \cdot \mathbf{p}_1 + t_z \cdot (n'_z \cdot \mathbf{p}_1)] &= f_2 \cdot [\mathbf{r}_1 \cdot \mathbf{p}_1 + t_x \cdot (n'_x \cdot \mathbf{p}_1)] \\ y_2 \cdot [\mathbf{r}_3 \cdot \mathbf{p}_1 + t_z \cdot (n'_z \cdot \mathbf{p}_1)] &= f_2 \cdot [\mathbf{r}_2 \cdot \mathbf{p}_1 + t_y \cdot (n'_y \cdot \mathbf{p}_1)] \end{cases}$$

On pose alors : $s = \mathbf{r}_3 \cdot \mathbf{p}_1 + t_z \cdot (n'_z \cdot \mathbf{p}_1)$.

La relation (5) se déduit alors aisément sachant que :

$$\mathbf{t} \cdot \mathbf{n}' = \begin{pmatrix} t_x \cdot n_x & t_x \cdot n_y & t_x \cdot n_z \\ t_y \cdot n_x & t_y \cdot n_y & t_y \cdot n_z \\ t_z \cdot n_x & t_z \cdot n_y & t_z \cdot n_z \end{pmatrix}$$

Deux points correspondant aux projections d'un point P du sol vérifient donc la relation *i.e.* $s \cdot \mathbf{p}_2 = (\mathbf{R} + \mathbf{t} \cdot \mathbf{n}') \cdot \mathbf{p}_1$ alors que deux pixels ne correspondant pas à un point du sol ne la vérifient pas *i.e.* $s \cdot \mathbf{q}_2 \neq (\mathbf{R} + \mathbf{t} \cdot \mathbf{n}') \cdot \mathbf{q}_1$.

9. Une troisième caméra permet de : (i) vérifier, (ii) lever les ambiguïtés, (iii) gérer les occultations.