Feuille d'exercices #2 — Vision par ordinateur - Solutions

IFT 3345

Question 1. *Backprojection*

Un robot mobile est équipé d'une caméra RGB-D. Un pixel est détecté dans l'image et sa profondeur est connue. Votre tâche est de calculer la position 3D de ce point dans le **repère mondial**.

Supposons que le repère mondial soit "Est-Nord-Haut" - c'est-à-dire x vers la droite, y vers le haut et z sortant du sol. Supposons que le repère du robot soit également "Est-Nord-Haut" - avec x sortant de l'avant du robot, y vers la gauche et l'axe z pointant vers le ciel. Le repère de la caméra a x vers la droite, y vers le bas et z sortant de la caméra.

Les paramètres intrinsèques de la caméra sont donnés par :

$$K = \begin{bmatrix} 500 & 0 & 250 \\ 0 & 500 & 250 \\ 0 & 0 & 1 \end{bmatrix}$$

La caméra est montée à une position de $(0.2, 0.0, 0.2)$ *dans le repère du robot* avec une orientation alignée avec la direction du robot et parallèle au plan du sol.

La position du robot dans le repère mondial est $\mathbf{x} = (1.5, 1.5)$ (sur le plan du sol) avec un cap (yaw) de $\psi = 90^o$.

Une caractéristique est observée au pixel $(u, v) = (400, 200)$ (depuis le coin supérieur gauche). La profondeur associée est $d = 2.5m$. Pour trouver la position de la caractéristique dans le repère mondial, nous devons procéder comme suit :

(a) Convertir le pixel en un point 3D dans le **repère de la caméra**.

(b) Transformer le point dans le **repère du robot**.

(c) Transformer le point dans le **repère mondial**.

Solution 1.

(a) Pixel → Repère de la caméra (Backprojection) :

Pixel sous forme homogène :

$$\tilde{p} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 400 \\ 200 \\ 1 \end{bmatrix}$$

[Remarque : nous utiliserons la convention selon laquelle les coordonnées homogènes sont notées avec un ˜ au-dessus]

Tout d'abord, nous pouvons convertir le pixel en coordonnées normalisées de l'image :

$$\tilde{x}_C = K^{-1}\tilde{p}$$

Intrinsèques inverses :

$$K^{-1} = \begin{bmatrix} 1/500 & 0 & -1/2 \\ 0 & 1/500 & -1/2 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\tilde{x}_C = K^{-1}\tilde{p} = \begin{bmatrix} 0.3 \\ -0.1 \\ 1 \end{bmatrix}$$

Ce vecteur homogène définit le rayon passant par le centre de la caméra et le pixel $(400, 200)$. Pour le restreindre à un point individuel en 3D, nous avons besoin d'une information supplémentaire. Dans ce cas, nous savons que la profondeur est 2.5m (ce qui correspond exactement au facteur d'échelle par lequel il faut multiplier la coordonnée z) pour obtenir une coordonnée 3D à 2.5m de la caméra.

Multiplication par la profondeur :

$$X_C = \lambda \tilde{x}_C = \begin{bmatrix} 0.75 \\ -0.25 \\ 2.5 \end{bmatrix}$$

Ceci est exactement le point dans le repère de la caméra (x vers la droite, y vers le bas et z sortant de la caméra par le centre de projection).

Pour les calculs futurs, nous pouvons le convertir en coordonnées homogènes :

$$\tilde{X}_C = \begin{bmatrix} 0.75 \\ -0.25 \\ 2.5 \\ 1 \end{bmatrix}$$

(b) Caméra $\rightarrow$ Repère du robot

La première étape consiste à écrire la matrice de transformation $SE(3)$ qui transforme les points du repère de la caméra vers le repère du robot. Pour ce faire, nous devons simplement convertir la pose de la caméra dans le repère du robot en un élément de $SE(3)$.

Un détail très important ici est que les conventions des axes dans le repère de la caméra sont différentes de celles du repère du robot. Comme décrit ci-dessus, dans le repère de la caméra, x est vers la droite, y vers le bas et z sort de la caméra. Dans le repère du robot, x sort de l'avant du robot, y vers la gauche et z vers le haut.

Nous pouvons écrire les correspondances comme suit :

$$\begin{aligned} x_r &= z_c \\ y_r &= -x_c \\ z_r &= -y_c \end{aligned}$$

Nous devons donc construire une matrice de rotation $R_{C \rightarrow R}$ qui effectue cette rotation :

$$R_{C \rightarrow R} = \begin{bmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}$$

Ce qui réalise bien la correspondance souhaitée :

$$\begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} z_c \\ -x_c \\ -y_c \end{bmatrix}$$

Nous pouvons donc construire la matrice de transformation complète en $SE(3)$ en utilisant cette rotation et la translation spécifiée ci-dessus :

$$T_{C \rightarrow R} = \begin{bmatrix} 0 & 0 & 1 & t_x \\ -1 & 0 & 0 & t_y \\ 0 & -1 & 0 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0.2 \\ -1 & 0 & 0 & 0.0 \\ 0 & -1 & 0 & 0.2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Nous pouvons simplement convertir notre point dans le repère de la caméra en un point dans le repère du robot en appliquant la transformation en coordonnées homogènes :

$$\tilde{X}_R = T_{C\to R}\tilde{X}_C$$

Ce qui donne (en coordonnées homogènes) :

$$\tilde{X}_R = \begin{bmatrix} 2.7 \\ -0.75 \\ 0.45 \\ 1 \end{bmatrix}$$

Ce qui correspond au point en coordonnées inhomogènes dans le repère du robot :

$$X_R = \begin{bmatrix} 2.7 \\ -0.75 \\ 0.45 \end{bmatrix}$$

(c) Robot $\rightarrow$ Repère mondial

Pour convertir ce point dans le repère mondial, nous avons besoin de la transformation (en $SE(3)$) correspondant à la pose du robot dans le repère mondial. C'est une légère modification des transformations précédentes en $SE(2)$, avec roulis (roll) et tangage (pitch) nuls et aucune translation sur l'axe z puisque le robot se déplace sur un plan :

$$T_{R\to W} = \begin{bmatrix} \cos\psi & -\sin\psi & 0 & t_x \\ \sin\psi & \cos\psi & 0 & t_y \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 & 1.5 \\ 1 & 0 & 0 & 1.5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Nous pouvons donc calculer le point dans le repère mondial :

$$\tilde{X}_W = T_{R\to W}\tilde{X}_R = \begin{bmatrix} 2.25 \\ 4.2 \\ 0.45 \\ 1 \end{bmatrix}$$

Le point final en coordonnées inhomogènes dans le repère mondial est donc :

$$X_W = \begin{bmatrix} 2.25 \\ 4.2 \\ 0.45 \end{bmatrix}$$

Question 2. *Points idéaux (points à l'infini)*

(a) Considérons le point homogène

$$\tilde{\mathbf{x}} = \begin{bmatrix} 3 \\ -2 \\ 0 \end{bmatrix}.$$

Cela représente un point à l'infini (un point idéal) puisque la valeur du troisième élément est 0. Quel est l'ensemble des lignes qui se rencontrent toutes en ce point particulier à l'infini ?

(b) Deux lignes dans l'image sont

$$\ell_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \qquad \ell_2 = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}.$$

Calculez leur intersection en utilisant le produit vectoriel. L'intersection est-elle finie ou à l'infini ?

Solution 2.

(a) La direction est obtenue à partir des deux premières coordonnées du point à l'infini :

$$\text{direction} = (3, -2).$$

L'ensemble des lignes pour lesquelles ceci est la normale est l'ensemble des lignes parallèles à la direction $(3, -2)$, ou en d'autres termes toutes les droites de pente

$$m = \frac{-2}{3}.$$

(b) Intersection de deux lignes :

$$\tilde{\mathbf{x}} = \ell_1 \times \ell_2$$

$$\tilde{\mathbf{x}} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & -1 & 0 \\ 2 & -2 & 1 \end{vmatrix} = \begin{bmatrix} (-1)(1) - 0(-2) \\ 0(2) - 1(1) \\ 1(-2) - (-1)(2) \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

Puisque $w = 0$, l'intersection se trouve **à l'infini**. Par conséquent, les lignes sont parallèles dans l'image.

Question 3. *Erreur de reprojection*

Supposons la même configuration (intrinsèques de la caméra, extrinsèques de la caméra, pose du robot) que dans la Question 1. Il y a une caractéristique connu sur le plan du sol ($z = 0$) à la position $x = 2$, $y = 4$ (dans le repère mondial). Il est détecté à l'emplacement du pixel :

$$x_{obs} = \begin{bmatrix} 370 \\ 300 \end{bmatrix}$$

Nous souhaitons calculer l'erreur de reprojection. Cela peut se faire selon les étapes suivantes :

(a) Transformer le point 3D du repère mondial vers le repère de la caméra.

(b) Projeter le point dans l'image.

(c) Calculer l'erreur de reprojection en pixels.

Solution 3.

(a) Repère mondial $\rightarrow$ Repère de la caméra

Nous pouvons commencer par convertir la position du repère mondial de la caractéristique dans le repère du robot.

$$T_{W\rightarrow R} = (T_{R\rightarrow W})^{-1} = \begin{bmatrix} 0 & 1 & 0 & -1.5 \\ -1 & 0 & 0 & 1.5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(Rappel : on obtient la transformation inverse en trouvant $\begin{bmatrix} R^T & -R^T t \\ 0^T & 1 \end{bmatrix}$)

Nous avons la coordonnée de la caractéristique dans le repère mondial en coordonnées homogènes :

$$\tilde{X}_W = \begin{bmatrix} 2 \\ 4 \\ 0 \\ 1 \end{bmatrix}$$

Nous pouvons donc calculer la position de la caractéristique dans le repère du robot :

$$\tilde{X}_R = T_{W\rightarrow R}\tilde{X}_W = \begin{bmatrix} 2.5 \\ -0.5 \\ 0 \\ 1 \end{bmatrix}$$

Nous répétons maintenant la procédure pour obtenir le point dans le repère de la caméra :

$$T_{R\rightarrow C} = (T_{C\rightarrow R})^{-1} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0.2 \\ 1 & 0 & 0 & -0.2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\tilde{X}_C = T_{R\rightarrow C}\tilde{X}_R = \begin{bmatrix} 0.5 \\ 0.2 \\ 2.3 \\ 1 \end{bmatrix}$$

Ainsi, notre point (inhomogène) dans le repère de la caméra est :

$$X_C = \begin{bmatrix} 0.5 \\ 0.2 \\ 2.3 \end{bmatrix}$$

(b) Projection du point dans l'image

Nous voulons trouver à quel pixel nous aurions dû observer la caractéristique. Il suffit de multiplier par la matrice de calibration intrinsèque, puis de normaliser.

$$\tilde{\mathbf{x}} = K\mathbf{X}_C = \begin{bmatrix} 500(0.5) + 250(2.3) \\ 500(0.2) + 250(2.3) \\ 2.3 \end{bmatrix} = \begin{bmatrix} 825 \\ 675 \\ 2.3 \end{bmatrix}$$

Nous obtenons les coordonnées du pixel en normalisant :

$$\mathbf{x}_{proj} = \begin{bmatrix} 825/2.3 \\ 675/2.3 \end{bmatrix} = \begin{bmatrix} 358.7 \\ 293.5 \end{bmatrix}$$

(c) Calcul de l'erreur de reprojection

$$\mathbf{e} = \mathbf{x}_{obs} - \mathbf{x}_{proj} = \begin{bmatrix} 370 \\ 300 \end{bmatrix} - \begin{bmatrix} 358.7 \\ 293.5 \end{bmatrix} = \begin{bmatrix} 11.3 \\ 6.5 \end{bmatrix}$$

$$\|\mathbf{e}\| \approx 13 \text{ pixels}$$

Question 4. *Intersection rayon-plan*

Une caméra calibrée montée sur un robot observe un point situé sur le plan du sol. Calculons les coordonnées 3D du point dans le repère mondial de deux manières différentes :

1. En intersectant le rayon avec le plan du sol
2. En calculant et en appliquant l'homographie inverse entre le plan du sol et le plan du capteur

Vous devriez obtenir la **même réponse** avec les deux méthodes.
Comme précédemment, supposons la matrice intrinsèque de la caméra :

$$K = \begin{bmatrix} 500 & 0 & 250 \\ 0 & 500 & 250 \\ 0 & 0 & 1 \end{bmatrix}$$

Le pixel qui détecte le point est :

$$p = \begin{bmatrix} 400 \\ 400 \end{bmatrix}$$

Cette fois, la caméra est située à 1 m au-dessus du plan du sol (et toujours orientée parallèlement au plan du sol). Supposons que le repère mondial soit placé sur le plan du sol directement sous la caméra. Nous allons calculer le point sur le plan du sol en utilisant les deux méthodes ci-dessus. Voici la répartition des tâches :

(a) Backprojecter le pixel pour obtenir le rayon visuel dans le repère de la caméra.
(b) Intersecter le rayon avec le plan du sol dans le repère de la caméra.
(c) Calculer la position du point du sol dans le repère mondial.
(d) Calculer l'homographie induite par le plan du sol vers l'image.
(e) Utiliser l'homographie inverse pour calculer la position du point du sol dans le repère mondial.

Solution 4.

(a) Backprojection du pixel vers le rayon de la caméra

$$K^{-1} = \begin{bmatrix} 1/500 & 0 & -1/2 \\ 0 & 1/500 & -1/2 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\tilde{p} = \begin{bmatrix} 400 \\ 400 \\ 1 \end{bmatrix}$$

$$\tilde{x}_C = K^{-1}\tilde{p} = \begin{bmatrix} 0.3 \\ 0.3 \\ 1 \end{bmatrix}$$

L'ensemble de tous les points 3D sur ce rayon est donné par

$$\lambda\tilde{x}_C = \begin{bmatrix} \lambda 0.3 \\ \lambda 0.3 \\ \lambda \end{bmatrix}$$

Nous devons simplement résoudre pour λ afin de trouver le point spécifique. On peut le faire en intersectant le rayon avec le plan du sol. Pour cela, nous définissons le plan du sol sous la forme :

$$n^T x_C + d = 0$$

où n est la normale au plan et d le décalage. Comme le rayon est exprimé dans le repère de la caméra, nous devons définir le plan dans le repère de la caméra. Rappel : dans le repère de la caméra, x vers la droite, y vers le bas et z sort du capteur. La normale au plan (par convention du plan du sol, la normale pointe vers le haut) sera :

$$n = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$$

c'est-à-dire pointant dans la direction négative de y (vers le haut). Le décalage sera $d = 1$, puisque le plan est à 1 mètre sous la caméra (y positif).

Ainsi, nous pouvons résoudre pour λ :

$$\lambda = -\frac{d}{n^T \tilde{x}_C} = 1/0.3$$

On obtient alors le point 3D associé :

$$X_c = \lambda \tilde{x}_C = \begin{bmatrix} 1 \\ 1 \\ 1/0.3 \end{bmatrix}$$

(b) Point dans le repère mondial

Comme précédemment, nous pouvons convertir le point dans le repère mondial en utilisant les coordonnées homogènes et la transformation entre le repère de la caméra et le repère mondial définie précédemment (avec une translation légèrement différente) :

$$\tilde{X}_W = T_{C\to W} \tilde{X}_C = \begin{bmatrix} 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1/0.3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/0.3 \\ -1 \\ 0 \\ 1 \end{bmatrix}$$

Ce qui correspond au point

$$X_W = \begin{bmatrix} 1/0.3 \\ -1 \\ 0 \end{bmatrix}$$

dans le repère mondial (coordonnées inhomogènes).

(c) Homographie induite par le plan

Rappelons la définition de la matrice de projection de la caméra :

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \simeq K \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} T_{W\to C} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Comme précédemment :

$$T_{W\to C} = (T_{C\to W})^{-1} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

En appliquant la contrainte $z = 0$ on obtient :

$$H = K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} = \begin{bmatrix} 500 & 0 & 250 \\ 0 & 500 & 250 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 250 & -500 & 0 \\ 250 & 0 & 500 \\ 1 & 0 & 0 \end{bmatrix}$$

Cette homographie mappe les points du plan du sol vers le plan du capteur (pixels). Ici, nous avons besoin de l'inverse :

$$H^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ -1/500 & 0 & 1/2 \\ 0 & 1/500 & -1/2 \end{bmatrix}$$

qui convertit les points de l'espace pixel vers le plan du sol.

(d) <u>Calcul du point sur le plan du sol</u>

Nous pouvons maintenant calculer le point sur le plan du sol :

$$\tilde{X}_W = H^{-1}\tilde{p} = \begin{bmatrix} 1 \\ -0.3 \\ 0.3 \end{bmatrix}$$

En normalisant ce point, on obtient :

$$\tilde{X}_W = \begin{bmatrix} 1/0.3 \\ -1 \\ 1 \end{bmatrix}$$

ce qui correspond au point avec $X = 1/0.3$ et $Y = -1$ sur le plan du sol, c'est-à-dire dans le repère mondial :

$$X_W = \begin{bmatrix} 1/0.3 \\ -1 \\ 0 \end{bmatrix}$$

Ce qui correspond exactement à ce que nous avons calculé en (b).

Question 5. *Gradients d'image et détection de coins Harris*

Considérons le patch d'image en niveaux de gris 5×5 suivant, normalisé entre 0 et 1 :

$$I = \begin{bmatrix} 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.5 & 0.5 & 0.1 & 0.1 \\ 0.1 & 0.5 & 0.9 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix}$$

(a) En utilisant un « filtre à différences centrales », calculez I_x et I_y pour les pixels intérieurs 3×3. Le filtre à différences centrales approxime le gradient en x comme : $\frac{\partial f}{\partial x} = \frac{f(x+1,y)-f(x-1,y)}{2}$, implémenté par le filtre $1/2 \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$, et $\frac{\partial f}{\partial y} = \frac{f(x,y+1)-f(x,y-1)}{2}$, implémenté par le filtre $1/2 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$.

(b) Calculez la magnitude et l'orientation du gradient du pixel central.

(c) En utilisant une fenêtre uniforme 3×3 centrée sur le pixel central, calculez le tenseur de structure (matrice des moments du 2 ordre) pour ce pixel.

(d) Calculez la réponse Harris, R, du pixel central en supposant $\alpha = 0.04$. Classifiez le pixel central comme région plate, bord ou coin.

Solution 5.

(a) <u>Gradients avec le filtre à différences centrales</u>
Les filtres à différences centrales en x et y utilisés pour calculer les gradients d'image sont donnés par :

$$S_x = 1/2 \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}, \qquad S_y = 1/2 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Après application de S_x et S_y sur la région intérieure 3×3, on obtient les gradients suivants :

$$I_x = \begin{bmatrix} 0.2 & -0.2 & -0.2 \\ 0.4 & -0.2 & -0.4 \\ 0 & 0 & 0 \end{bmatrix}$$

$$I_y = \begin{bmatrix} 0.2 & 0.4 & 0 \\ -0.2 & -0.2 & 0 \\ -0.2 & -0.4 & 0 \end{bmatrix}$$

(b) <u>Magnitude et orientation du pixel central</u>
Au centre :

$$I_x = -0.2, \quad I_y = -0.2$$

Magnitude :

$$|\nabla I| \approx 0.28$$

Orientation :

$$\theta = \text{atan2}(-0.2, -0.2) = 225^\circ$$

(c) Tenseur de structure au pixel central
Calcul des sommes sur la fenêtre 3×3 :

$$\sum I_x^2 = 0.48, \qquad \sum I_y^2 = 0.48, \qquad \sum I_x I_y = -0.08$$

Ainsi :

$$M = \begin{bmatrix} 0.48 & -0.08 \\ -0.08 & 0.48 \end{bmatrix}$$

(d) Réponse Harris

$$\det(M) = 0.48^2 - (-0.08)^2 = 0.224$$

$$\text{trace}(M) = 0.48 + 0.48 = 0.96$$

$$R = 0.224 - 0.04(0.96)^2 = 0.187$$

Puisque $R > 0$ et relativement grand par rapport à l'énergie du gradient, le pixel central est classé comme un **coin**.