TP1 – Estimation de la matrice essentielle

Préambule

Le calcul de la structure à partir du mouvement (en anglais SfM, pour *structure-from-motion*) consiste à estimer les *changements de pose* d'un appareil photographique à partir de plusieurs vues d'un même objet, indéformable et opaque, prises sous différents angles, ce qui nécessite un mouvement (*motion*) de l'appareil photographique. Un autre résultat fort appréciable de cette technique est qu'elle fournit la *structure* de l'objet, sous la forme d'un nuage de points 3D.

Mise en correspondance

Dans le cas où l'on ne dispose que de deux images I_1 (image gauche) et I_2 (image droite), une étape préliminaire du SfM consiste à mettre en correspondance ces images. Un pixel $\mathbf{p}_1$ de I_1 et un pixel $\mathbf{p}_2$ de I_2 sont dits *homologues* s'ils correspondent à un même point 3D situé sur la surface de l'objet. Cette étape de *mise en correspondance* doit être menée avec soin si l'on souhaite obtenir une reconstruction 3D précise. Un certain nombre d'algorithmes, que vous étudierez dans l'UE VRAA (*Vision, Réalité Augmentée et Applications*), effectuent automatiquement la détection et l'appariement de *points d'intérêt*.

Le script `exercice_0` permet de constituer environ 10000 paires de points d'intérêt en quelques secondes (pour des raisons de lisibilité, seule une partie de ces paires est affichée). Commencez par lancer ce script.

Matrice fondamentale

Deux pixels homologues $\mathbf{p}_1 = [u_1, v_1]^\top \in I_1$ et $\mathbf{p}_2 = [u_2, v_2]^\top \in I_2$ (coordonnées exprimées dans les repères pixels) sont censés correspondre à un même point 3D noté $\mathbf{Q}$. Sachant qu'une photographie se forme par projection perspective sur le plan image, $\mathbf{p}_1$ et $\mathbf{p}_2$ se situent forcément sur le plan passant par $\mathbf{Q}$ et par les deux centres de projection $\mathbf{C}_1$ et $\mathbf{C}_2$. Cette contrainte sur les pixels homologues, dite *contrainte épipolaire*, peut être reformulée comme suit (voir cours) :

$$\widetilde{\mathbf{p}}_2^\top \, \mathbf{F} \, \widetilde{\mathbf{p}}_1 = 0 \tag{1}$$

où :

- les vecteurs $\widetilde{\mathbf{p}}_1 = [u_1, v_1, 1]^\top$ et $\widetilde{\mathbf{p}}_2 = [u_2, v_2, 1]^\top$ contiennent les *coordonnées homogènes* de $\mathbf{p}_1$ et $\mathbf{p}_2$;
- la matrice $\mathbf{F}$, de taille 3×3, est appelée *matrice fondamentale*.

Étant donné un pixel $\mathbf{p}_2 = [u_2, v_2]^\top$ de I_2, l'équation (1) peut être réécrite :

$$a_2 \, u_1 + b_2 \, v_1 + c_2 = 0 \tag{2}$$

où $[a_2, b_2, c_2] = \widetilde{\mathbf{p}}_2^\top \, \mathbf{F}$. L'équation (2) est l'équation cartésienne d'une droite dans I_1, dite *droite épipolaire*. Inversement, étant donné un pixel $\mathbf{p}_1 = [u_1, v_1]^\top$ de I_1, l'équation (1) peut être réécrite :

$$a_1 \, u_2 + b_1 \, v_2 + c_1 = 0 \tag{3}$$

où $[a_1, b_1, c_1] = \widetilde{\mathbf{p}}_1^\top \, \mathbf{F}^\top$. L'équation (3) est l'équation cartésienne d'une droite épipolaire dans I_2. La matrice fondamentale $\mathbf{F}$ caractérise donc entièrement la *géométrie épipolaire* de la paire d'images $\{I_1, I_2\}$.

Matrice essentielle

Avec les notations de la figure 1, il a été vu en cours que le point 3D noté $\mathbf{w}$, défini par $\mathbf{Q} = Z\,\mathbf{w}$, vérifiait l'égalité suivante, où $\mathbf{K}$ est la *matrice de calibrage* :

$$\widetilde{\mathbf{p}} = \mathbf{K}\,\mathbf{w} \tag{4}$$

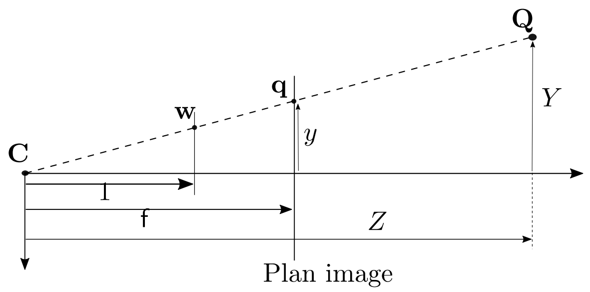

FIGURE 1 – Projection centrale, de centre $\mathbf{C}$, d'un point 3D $\mathbf{Q}$ en un point image $\mathbf{q}$. Le point 3D $\mathbf{w}$ est également obtenu par projection centrale sur un plan parallèle au plan image, ce plan étant distant de 1 du point $\mathbf{C}$.

Des équations (1) et (4), nous tirons l'égalité :

$$\mathbf{w}_2^\top\,\mathbf{K}^\top \mathbf{F}\,\mathbf{K}\,\mathbf{w}_1 = 0 \tag{5}$$

qui nous permet de réécrire la contrainte épipolaire (1) sous la forme suivante :

$$\mathbf{w}_2^\top\,\mathbf{E}\,\mathbf{w}_1 = 0 \tag{6}$$

qui fait apparaître la *matrice essentielle* $\mathbf{E}$, définie par :

$$\mathbf{E} = \mathbf{K}^\top\,\mathbf{F}\,\mathbf{K} \tag{7}$$

Estimation de la matrice essentielle

Grâce à l'appariement, il est possible d'estimer la matrice fondamentale et/ou la matrice essentielle. Pour des raisons de stabilité numérique, il est préférable d'estimer la matrice essentielle $\mathbf{E}$, mais cela nécessite que la caméra soit calibrée. De fait, si la matrice $\mathbf{K}$ est connue, l'équation (4) nous permet de calculer les coordonnées des points 3D $\mathbf{w}_1$ et $\mathbf{w}_2$ correspondant à deux pixels homologues $\mathbf{p}_1$ et $\mathbf{p}_2$:

$$\mathbf{w}_1 = \mathbf{K}^{-1}\,\widetilde{\mathbf{p}}_1 \qquad \text{et} \qquad \mathbf{w}_2 = \mathbf{K}^{-1}\,\widetilde{\mathbf{p}}_2 \tag{8}$$

Si nous notons les coordonnées de ces deux points 3D comme suit :

$$\mathbf{w}_1 = [r_1, s_1, 1]^\top \qquad \text{et} \qquad \mathbf{w}_2 = [r_2, s_2, 1]^\top \tag{9}$$

et si nous notons $\{e_{ij}\}_{i,j\in\{1,2,3\}}$ les éléments (inconnus) de la matrice essentielle $\mathbf{E}$, l'équation (6) nous procure l'équation linéaire homogène suivante :

$$r_2\,(r_1\,e_{11} + s_1\,e_{12} + e_{13}) + s_2\,(r_1\,e_{21} + s_1\,e_{22} + e_{23}) + r_1\,e_{31} + s_1\,e_{32} + e_{33} = 0 \tag{10}$$

En utilisant m paires de pixels homologues, les équations (10) forment un système linéaire homogène $\mathbf{A}\,\mathbf{e} = \mathbf{0}$, où $\mathbf{e} \in \mathbb{R}^9$ est obtenu par vectorisation de la matrice $\mathbf{E}$, et où $\mathbf{A}$ est une matrice de taille $m \times 9$, dont chaque ligne est égale au *produit de Kronecker* de $\mathbf{w}_2^\top$ par $\mathbf{w}_1^\top$ (fonction `kron` de Matlab). La matrice $\mathbf{E}$ étant définie seulement à un facteur près, il suffit de $m = 8$ paires de pixels homologues pour l'estimer.

Exercice 1 : estimation de la matrice essentielle

Le script `exercice_1` tire aléatoirement $m = 8$ paires de pixels homologues à l'aide de la fonction `randperm` de Matlab, parmi le très grand nombre de paires de pixels homologues fournies par le script `exercice_0`.

Écrivez la fonction `estimation_E`, appelée par `exercice_1`, qui est censée résoudre le problème d'optimisation suivant (*moindres carrés linéaires*) :

$$\min_{\mathbf{e}\in\mathbb{R}^9} \|\mathbf{A}\,\mathbf{e}\|^2 \tag{11}$$

Comme ce problème admet la solution triviale $\mathbf{e} = \mathbf{0}$, nous imposons la contrainte $\|\mathbf{e}\| = 1$. Il a été vu en cours que la solution de ce problème d'*optimisation sous contrainte* est un des deux vecteurs propres de norme 1 de la matrice $\mathbf{A}^\top\mathbf{A}$, associés à sa plus petite valeur propre (la matrice $\mathbf{A}^\top\mathbf{A}$ étant symétrique réelle, semi-définie positive, le théorème spectral implique que toutes ses valeurs propres sont positives ou nulles).

La matrice $\mathbf{E}$ se déduit facilement de la solution en $\mathbf{e}$ du problème (11). Néanmoins, comme cela a été vu en cours, cette matrice $\mathbf{E}$ ne possède pas les caractéristiques d'une matrice essentielle : une de ses valeurs singulières devrait être nulle, et les deux autres valeurs singulières devraient être non nulles et égales (on les fixe généralement à 1, mais cela n'est pas indispensable, car une matrice essentielle est définie à un facteur près). Cela se traduit par le fait que, si les droites épipolaires qui se déduisent de la contrainte épipolaire passent effectivement par les $m = 8$ paires de pixels homologues utilisés pour la résolution du problème (11), elles ne forment pas deux faisceaux de droites, dès lors que la localisation des pixels homologues est imprécise ou erronée.

Modifiez la fonction `estimation_E`, de manière à forcer les valeurs singulières de la matrice $\mathbf{E}$ à être égales à 1, 1 et 0 (utilisez pour cela la fonction `svd` de Matlab). En lançant plusieurs exécutions du script `exercice_1`, vous constaterez alors que, si les $m = 8$ droites épipolaires de chaque image forment maintenant un faisceau, ces droites ne passent généralement plus par les pixels tirés aléatoirement. Il s'ensuit que, d'un tirage aléatoire à l'autre, la position des épipôles varie fortement. Il est donc nécessaire de rendre l'estimation de $\mathbf{E}$ robuste.

Exercice 2 : estimation robuste de la matrice essentielle

Si les $m = 8$ paires de pixels homologues $\{\mathbf{p}_1, \mathbf{p}_2\}$ utilisés pour l'estimation de $\mathbf{E}$ ont été correctement localisées, et si aucune de ces paires ne constitue un « faux appariement », alors les droites épipolaires D_1 et D_2 doivent passer par $\mathbf{p}_1$ et $\mathbf{p}_2$, respectivement. Un critère permettant de savoir si les $m = 8$ paires de pixels tirés aléatoirement se prêtent bien à l'estimation de $\mathbf{E}$ consiste donc à mesurer, pour chaque paire de pixels $\{\mathbf{p}_1, \mathbf{p}_2\}$, la distance euclidienne de $\mathbf{p}_2$ à la droite épipolaire D_2 et la distance euclidienne de $\mathbf{p}_1$ à la droite épipolaire D_1 (les équations cartésiennes de ces droites épipolaires se déduisent de $\mathbf{E}$).

Écrivez la fonction `estimation_E_robuste`, appelée par le script `exercice_2`, qui tire aléatoirement $m = 8$ paires de pixels homologues « en boucle », tant que la médiane (fonction `median` de Matlab) des carrés des distances aux droites épipolaires *de la totalité des points d'intérêt* fournis par le script `exercice_0` est supérieure à $S = 5.10^{-6}$. Attention : ces distances doivent être exprimées en mètres, et non en pixels. Il est rappelé que le carré de la distance d'un point $P = (x, y)$ à une droite D d'équation cartésienne $a\,x + b\,y + c = 0$ s'écrit :

$$d(P, D)^2 = \frac{(a\,x + b\,y + c)^2}{a^2 + b^2} \tag{12}$$

Remarques :

- Bien entendu, la fonction `estimation_E_robuste` doit appeler `estimation_E`.
- La médiane est préférée à la moyenne, car elle constitue un *estimateur robuste*.

Lancez plusieurs exécutions du script `exercice_2`, et observez que les positions des épipôles, qui se situent d'ailleurs en dehors des images, semblent maintenant stables.

Exercice 3 : reconstruction 3D par triangulation

La reconstruction 3D par *triangulation* nécessite de connaître le mouvement de la caméra entre les deux prises de vue, c'est-à-dire la « transformation rigide » (*isométrie*) entre les repères $\mathcal{R}_1$ et $\mathcal{R}_2$. Or, il a été vu en cours que la matrice essentielle $\mathbf{E}$ peut s'exprimer en fonction de $\mathbf{t}_{1/2} = [t_1, t_2, t_3]^\top$ et de $\mathbf{R}_{1/2}$:

$$\mathbf{E} = \left[\mathbf{t}_{1/2}\right]_\wedge \mathbf{R}_{1/2} \tag{13}$$

où la matrice antisymétrique $\left[\mathbf{t}_{1/2}\right]_\wedge$ est définie par :

$$\left[\mathbf{t}_{1/2}\right]_\wedge = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix} \tag{14}$$

Même s'il a déjà été vu en cours comment calculer $\mathbf{t}_{1/2}$ et $\mathbf{R}_{1/2}$ à partir de la matrice $\mathbf{E}$, nous vous fournissons dans ce premier TP la solution en $\mathbf{R}_{1/2}$, dont le vecteur $\mathbf{t}_{1/2}$ se déduit facilement en utilisant l'égalité (13). Notez toutefois que le vecteur $\mathbf{t}_{1/2}$ est connu seulement au signe près, puisque tel est le cas de la matrice $\mathbf{E}$ (cette ambiguïté sera levée dans le TP2).

Les coordonnées d'un point 3D $\mathbf{Q}$ dans le repère $\mathcal{R}_1$ sont proportionnelles à celles de $\mathbf{w}_1$:

$$\mathbf{Q}_1 = Z_1 \, \mathbf{w}_1 \tag{15}$$

De la même façon, dans le repère $\mathcal{R}_2$:

$$\mathbf{Q}_2 = Z_2 \, \mathbf{w}_2 \tag{16}$$

Or, $\mathbf{Q}_1$ et $\mathbf{Q}_2$ sont liés par la formule de changement de repère :

$$\mathbf{Q}_2 = \mathbf{R}_{1/2} \, \mathbf{Q}_1 + \mathbf{t}_{1/2} \tag{17}$$

De (15), (16) et (17), nous tirons l'équation matricielle suivante :

$$Z_2 \, \mathbf{w}_2 = \mathbf{R}_{1/2} \, (Z_1 \, \mathbf{w}_1) + \mathbf{t}_{1/2} \tag{18}$$

qui constitue un système de trois équations linéaires à deux inconnues Z_1 et Z_2. À cause des imprécisions sur la localisation des points d'intérêt, ce système n'admet généralement aucune solution exacte. Cela correspond au fait que les deux rayons de projection issus des pixels homologues $\mathbf{p}_1$ et $\mathbf{p}_2$ ne se croisent jamais *exactement*. L'équation (18) doit donc être résolue en moindres carrés :

$$\min_{(Z_1, Z_2) \in \mathbb{R}^2} \| Z_2 \, \mathbf{w}_2 - \mathbf{R}_{1/2} \, (Z_1 \, \mathbf{w}_1) - \mathbf{t}_{1/2} \|^2 \tag{19}$$

La solution de ce problème nous donne la paire de points 3D les plus proches situés sur les deux rayons de projection. La méthode de triangulation dite *du point milieu* consiste à positionner le point $\mathbf{Q}$ recherché au centre du segment défini par ces deux points 3D. Les coordonnées de $\mathbf{Q}$ dans le repère $\mathcal{R}_2$ s'écrivent donc :

$$\mathbf{Q}_2 = \frac{1}{2} \left[Z_2 \, \mathbf{w}_2 + \mathbf{R}_{1/2} \, (Z_1 \, \mathbf{w}_1) + \mathbf{t}_{1/2} \right] \tag{20}$$

Écrivez les fonctions `estimation_t` et `reconstruction_3D`, appelées par le script `exercice_3`, qui est censé appliquer cette méthode pour calculer les coordonnées 3D des points $\mathbf{Q}$, exprimées dans le repère $\mathcal{R}_2$, correspondant à l'ensemble des paires de points d'intérêt fournies par le script `exercice_0`.