The Austalian National University
School of Computing
Dylan Campbell

Semester 1, 2024
Tutorial 6

COMP/ENGN 4528/6528: Computer Vision

Question 1 **3D SFM and Image formation**

1. (a) Given two calibrated cameras, C_1 and C_2, C_1 has a focal length of 500 in x and 375 in y, (in pixel unit) the camera has resolution 512×512, and the camera centre projected to the image is at $(249, 249)$, with no skew. Suppose C_2 has the same image resolution and focal length as C_1, but the camera centre projected to the image is at $(251, 252)$. Write down the calibration matrix K_1 and K_2 for C_1 and C_2 respectively.

   (b) Suppose that a 3D world coordinate system ((X, Y, Z) coordinates as in Figure 1) is defined as aligned with the camera coordinate system of C_1. More specifically, the world origin is at the camera centre of C_1, the Z axis is aligned with the optical (principal) axis and the X and Y world coordinate systems are aligned parallel with the x and y axes of the image of C_1. Write down the matrices $K[R|t]$ which define the projection of a point in the world coordinate system to the image of C_1.

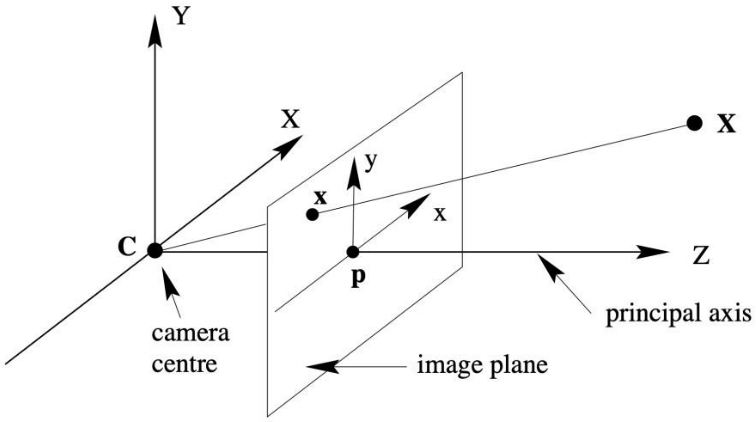

Figure 1: 3D coordinate systems

   (c) Suppose that the scene has a point, P_1, that in the world coordinate system defined above that lies at $(39, 35, 100)$. Note that the points in the world coordinate system are measured in centimetres. What location (to the nearest pixel) will that world point (P_1) map to in the image of C_1?

   (d) Suppose that with respect to the world coordinate system that is aligned with camera C_1, camera C_2 begins being aligned to C_1 and is then rotated by 45° about its vertical axis (Y axis) (as shown in Figure 2), and subsequently, the centre of C_2 is translated by 0.2 metre to the left of C_1 (along the X axis of C_1), then moved forward by 0.2 metre parallel to the optical axis of C_1. Write down the matrices $K[R|t]$, which define the projection of points in the world system (i.e, the same coordinate system of C_1) to the image of C_2.

   (e) Define the term "epipole".

   (f) For camera C_1, there is an epipole (or epipolar point) that relates to camera C_2. For the two-camera setup for predicting structure from motion, what is the position of the epipole in camera C_1 of camera C_2? (Hint: It is a point in the image coordinates of Camera C1).

Solution.

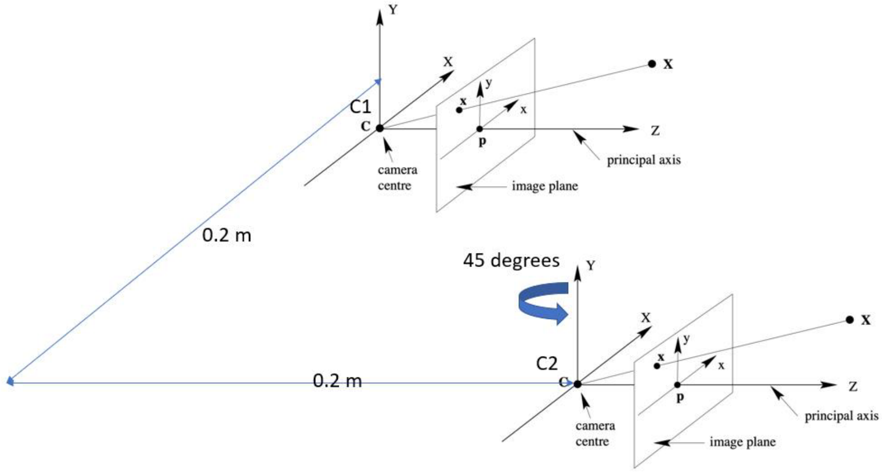

Figure 2: Visualisation of camera relative positions

(a)

$$\mathbf{K}_1 = \begin{bmatrix} 500 & 0 & 249 \\ 0 & 375 & 249 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{K}_2 = \begin{bmatrix} 500 & 0 & 251 \\ 0 & 375 & 252 \\ 0 & 0 & 1 \end{bmatrix}$$

(b) Since the world origin overlaps with the camera centre of C_1, there is no rotation and translation when projecting from the world coordinate to the camera coordinate. So, we can infer that $\mathbf{R}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\mathbf{t}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

$$\begin{aligned} \mathbf{P}_1 &= \mathbf{K}_1[\mathbf{R}_1|\mathbf{t}_1] \\ &= \begin{bmatrix} 500 & 0 & 249 \\ 0 & 375 & 249 \\ 0 & 0 & 1 \end{bmatrix} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \\ &= \begin{bmatrix} 500 & 0 & 249 & 0 \\ 0 & 375 & 249 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{aligned}$$

(c)

$$\begin{aligned} \begin{bmatrix} x \\ y \\ w \end{bmatrix} &= \begin{bmatrix} 500 & 0 & 249 & 0 \\ 0 & 375 & 249 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 39 \\ 35 \\ 100 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 44400 \\ 38025 \\ 100 \end{bmatrix} \end{aligned}$$

Then, we transform from the homogeneous coordinate to the ordinary coordinate and obtain the location $(444, 380.25) \approx (444, 380)$.

(d)

$$\mathbf{K}_2 = \begin{bmatrix} 500 & 0 & 251 \\ 0 & 375 & 252 \\ 0 & 0 & 1 \end{bmatrix}$$

Since we use centimetres in the previous question, we convert the unit of measurement from metres to centimetres. Also, given $\tilde{\mathbf{X}} = \mathbf{R}(\tilde{\mathbf{X}}_{world} - \mathbf{C})$, $\tilde{\mathbf{X}} = (20, 0, -20)$ and $\tilde{\mathbf{X}}_{world} = (0, 0, 0)$, we can get vector $\mathbf{C}_2$ as follows.

$$\mathbf{C}_2 = \begin{bmatrix} -20 \\ 0 \\ 20 \end{bmatrix}$$

We counterclockwise rotate the Y axis for $45°$ as shown in Figure 2. Since the rotation matrices provided in the lecture slides are clockwise, we should use $-45°$ in the calculation.

$$\mathbf{R}_2 = \mathbf{R}_y(-45°) = \begin{bmatrix} \cos(-45°) & 0 & \sin(-45°) \\ 0 & 1 & 0 \\ -\sin(-45°) & 0 & \cos(-45°) \end{bmatrix}$$

Then, we can compute the camera projection in two ways as follows.

i.

$$\begin{aligned}
\mathbf{P}_2 &= \mathbf{K}_2\mathbf{R}_2[\mathbf{I}| - \mathbf{C}_2] \\
&= \begin{bmatrix} 500 & 0 & 251 \\ 0 & 375 & 252 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(-45°) & 0 & \sin(-45°) \\ 0 & 1 & 0 \\ -\sin(-45°) & 0 & \cos(-45°) \end{bmatrix} \left[\begin{array}{ccc|c} 1 & 0 & 0 & -20 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 20 \end{array}\right] \\
&= \begin{bmatrix} 531.037 & 0 & -176.070 & 14142.136 \\ 178.191 & 375 & 178.191 & 0 \\ 0.707 & 0 & 0.707 & 0 \end{bmatrix}
\end{aligned}$$

ii.

$$\begin{aligned}
\mathbf{t}_2 &= -\mathbf{R}_2\mathbf{C}_2 \\
&= -\begin{bmatrix} \cos(-45°) & 0 & \sin(-45°) \\ 0 & 1 & 0 \\ -\sin(-45°) & 0 & \cos(-45°) \end{bmatrix} \begin{bmatrix} -20 \\ 0 \\ 20 \end{bmatrix} \\
&= \begin{bmatrix} 20\sqrt{2} \\ 0 \\ 0 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\mathbf{P}_2 &= \mathbf{K}_2[\mathbf{R}_2|\mathbf{t}_2] \\
&= \begin{bmatrix} 500 & 0 & 251 \\ 0 & 375 & 252 \\ 0 & 0 & 1 \end{bmatrix} \left[\begin{array}{ccc|c} \cos(-45°) & 0 & \sin(-45°) & 20\sqrt{2} \\ 0 & 1 & 0 & 0 \\ -\sin(-45°) & 0 & \cos(-45°) & 0 \end{array}\right] \\
&= \begin{bmatrix} 531.037 & 0 & -176.070 & 14142.136 \\ 178.191 & 375 & 178.191 & 0 \\ 0.707 & 0 & 0.707 & 0 \end{bmatrix}
\end{aligned}$$

(e) Given a setup consisting of two cameras, the epipole is defined as the projection of one camera centre at the image plane of the other one.

(f) Two ways to solve this question are as follows.

i. The camera center of camera C_2 is located at $(-20, 0, 20)$ in the world coordinate. We can calculate the epipole in camera C_1 of camera C_2 as follows.

$$
\begin{aligned}
w_1 \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} &= \mathbf{P}_1 \begin{bmatrix} -20 \\ 0 \\ 20 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} 500 & 0 & 249 & 0 \\ 0 & 375 & 249 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} -20 \\ 0 \\ 20 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} -5020 \\ 4980 \\ 20 \end{bmatrix} \\
&= 20 \begin{bmatrix} -251 \\ 249 \\ 1 \end{bmatrix}
\end{aligned}
$$

Thus, we can conclude $(u_1, v_1) = (-251, 249)$.

ii. Denote $\mathbf{X}_1$ as the coordinates under camera C_1, and $\mathbf{X}_2$ as the coordinates under camera C_2.

$$
\mathbf{X}_2 = \mathbf{R}_2 \mathbf{X}_1 + \mathbf{t}_2
$$

$$
\begin{aligned}
\mathbf{X}_1 &= \mathbf{R}_2^{-1}(\mathbf{X}_2 - \mathbf{t}_2) \\
\mathbf{X}_1 &= -\mathbf{R}_2^{-1}\mathbf{t}_2
\end{aligned}
$$

$$
\begin{aligned}
w_1 \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} &= \mathbf{K}_1(\mathbf{R}_1 \mathbf{X}_1 + \mathbf{t}_1) \\
&= \mathbf{K}_1(-\mathbf{R}_1 \mathbf{R}_2^{-1} \mathbf{t}_2 + \mathbf{t}_1) \\
&= \mathbf{K}_1(-\mathbf{R}_1 \mathbf{R}_2^{-1} (-\mathbf{R}_2 \mathbf{C}_2) + \mathbf{t}_1) \\
&= \mathbf{K}_1(\mathbf{R}_1 \mathbf{C}_2 + \mathbf{t}_1) \\
&= \begin{bmatrix} 500 & 0 & 249 \\ 0 & 375 & 249 \\ 0 & 0 & 1 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 20 \\ 0 \\ -20 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right) \\
&= \begin{bmatrix} 5020 \\ -4980 \\ -20 \end{bmatrix} \\
&= -20 \begin{bmatrix} -251 \\ 249 \\ 1 \end{bmatrix}
\end{aligned}
$$

So, we can conclude $(u_1, v_1) = (-251, 249)$.

Question 2 — Epipolar geometry

1. Describe an algorithm to recover a fundamental matrix between two cameras given a set of 50 putative matching points between the cameras (that may contain errors).

   Solution. We can use the RANSAC algorithm to handle outliers in this scenario.

(a) Repeat until $\tau(\#inliers, \#samples) > 95\%$ or too many iterations.

   i. Randomly select $N \geq 8$ matched points pairs.

   ii. Compute a fundamental matrix $\mathbf{F}$ based on chosen points pairs using the **Normalised DLT algorithm**.

   iii. Check how many of the other matches are less than a threshold distance based on the calculated fundamental matrix. If it is the largest number of inliers so far, set this to be the best match $\mathbf{F}$.

(b) Take the inlier set from the best match $\mathbf{F}$.

(c) Recompute $\mathbf{F}$ using the **Normalised DLT algorithm** using the whole inlier set.

Normalised DLT algorithm:

(a) Objective

   i. Given $N \geq 8$ points correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$, determine the 3×3 fundamental matrix $\mathbf{F}$ such that $\mathbf{x}'^\top \mathbf{F}\mathbf{x} = 0$

(b) Algorithm

   i. Normalise points: $\tilde{\mathbf{x}}_i = \mathbf{T}\mathbf{x}_i$, $\tilde{\mathbf{x}}'_i = \mathbf{T}'\mathbf{x}'_i$.

   ii. For each correspondence $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$ compute

$$\mathbf{A}_i = \begin{bmatrix} u_i u'_i & v_i u'_i & u'_i & u_i v'_i & v_i v'_i & v'_i & u_i & v_i & 1 \end{bmatrix}$$

   iii. Assemble N row vectors into a single $N \times 9$ matrix $\mathbf{A}$.

   iv. Compute the SVD of $\mathbf{A}$. The solution for $\mathbf{f}$ is the last column of $\mathbf{V}$.

   v. Rearrange (reshape) vector $\mathbf{f}$ into matrix $\tilde{\mathbf{F}}$.

   vi. Enforcing fundamental matrix to have rank = 2. Let $\tilde{\mathbf{F}} = \mathbf{U}\Sigma\mathbf{V}^\top$ (SVD of matrix $\tilde{\mathbf{F}}$), where $\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}$. Let $\Sigma' = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, then $\hat{\mathbf{F}} = \mathbf{U}\Sigma'\mathbf{V}^\top$.

   vii. Denormalise the recovered solution

$$\mathbf{F} = \mathbf{T}'^\top \hat{\mathbf{F}} \mathbf{T}$$

(c) Normalisation

   i. Suppose $\mathbf{X}$ is a matrix representing a set of 2D points, with shape $2 \times N$. It is consist of column vectors $\mathbf{x}_i = \begin{bmatrix} u_i & v_i \end{bmatrix}^\top$.

$$\mathbf{X} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \cdots & u_N \\ v_1 & v_2 & \cdots & v_N \end{bmatrix}$$

   ii. Compute the mean for each row of the matrix $\mathbf{X}$.

$$\mu_X = \begin{bmatrix} \mu_\mathbf{u} \\ \mu_\mathbf{v} \end{bmatrix} = \begin{bmatrix} \frac{1}{N}\sum_i u_i \\ \frac{1}{N}\sum_i v_i \end{bmatrix}$$

   iii. Subtract the mean of each dimension from the original coordinates. The new matrix is denoted by $\mathbf{X}'$.

$$\mathbf{X}' = \begin{bmatrix} u'_1 & u'_2 & \cdots & u'_N \\ v'_1 & v'_2 & \cdots & v'_N \end{bmatrix} = \begin{bmatrix} u_1 - \mu_\mathbf{u} & u_2 - \mu_\mathbf{u} & \cdots & u_N - \mu_\mathbf{u} \\ v_1 - \mu_\mathbf{v} & v_2 - \mu_\mathbf{v} & \cdots & v_N - \mu_\mathbf{v} \end{bmatrix}$$

   iv. Then, calculate the mean distance s using the formula below.

$$s = \frac{\sqrt{2}}{\frac{1}{N}\sum_i \sqrt{(u'_i)^2 + (v'_i)^2}}$$

v. The normalisation matrix $\mathbf{T}$ is calculated as follows. Before using this matrix, 2D coordinates should be transformed into the homogeneous coordinate.

$$\mathbf{T} = \begin{bmatrix} s & 0 & -s \times \mu_{\mathbf{u}} \\ 0 & s & -s \times \mu_{\mathbf{v}} \\ 0 & 0 & 1 \end{bmatrix}$$

Question 3 **Camera Calibration**

1. Complete the coding questions in `COMP4528_lab6_code.ipynb`.

   Solution. Please refer to the `COMP4528_lab6_code_sol.ipynb` file on Wattle.