hw9

November 5, 2025

Hoang Anh Nguyen

1 Exercise 1

1.1 Step 1 – Expand $MX = X'$

For a correspondence $(x, y) \mapsto (x', y')$,

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x' \\ y' \end{bmatrix}$$

This expands to two linear equations:

$$x\, m_{11} + y\, m_{12} + 0\, m_{21} + 0\, m_{22} = x'$$

$$0\, m_{11} + 0\, m_{12} + x\, m_{21} + y\, m_{22} = y'$$

1.2 Step 2 – Construct matrices Q and b

Using the coordinates of points A, B, C, D and their corresponding transformed points A', B', C', D', we obtain 8 equations (2 per point):

$$Qm = b, \quad \text{where } m = \begin{bmatrix} m_{11} \\ m_{12} \\ m_{21} \\ m_{22} \end{bmatrix}$$

Given:

$$\begin{aligned} A &= (1, 1) \to A' = (-0.9,\ 0.8) \\ B &= (1.5, 0.5) \to B' = (-0.1,\ 1.3) \\ C &= (2, 1) \to C' = (-0.4,\ 1.9) \\ D &= (2.5, 2) \to D' = (-1.25,\ 2.55) \end{aligned}$$

The stacked system is:

$$Q = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1.5 & 0.5 & 0 & 0 \\ 0 & 0 & 1.5 & 0.5 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 2.5 & 2 & 0 & 0 \\ 0 & 0 & 2.5 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} -0.9 \\ 0.8 \\ -0.1 \\ 1.3 \\ -0.4 \\ 1.9 \\ -1.25 \\ 2.55 \end{bmatrix}$$

1.3 Step 3

[6]:
```
import numpy as np

Q = np.array([
    [1, 1, 0, 0],
    [0, 0, 1, 1],
    [1.5, 0.5, 0, 0],
    [0, 0, 1.5, 0.5],
    [2, 1, 0, 0],
    [0, 0, 2, 1],
    [2.5, 2, 0, 0],
    [0, 0, 2.5, 2]
])

b = np.array([-0.9, 0.8, -0.1, 1.3, -0.4, 1.9, -1.25, 2.55])

m, _, _, _ = np.linalg.lstsq(Q, b, rcond=None)

M = m.reshape(2, 2)
print("Transformation matrix M =\n", M)
```

```
Transformation matrix M =
 [[ 0.332  -1.0808]
 [ 0.876   0.1256]]
```

2 Exercise 2

- **Line 35 – Matrix K:**
  `K` is the camera intrinsic matrix a 3×3 matrix containing the focal lengths (fx, fy) and the principal point (cx, cy).
  It defines the internal camera parameters and maps 3D camera coordinates to 2D image coordinates:

$$x = K[R\ t]X$$

- **Line 37 – Image keypoints and descriptors:**
  The algorithm extracts SIFT (Scale-Invariant Feature Transform) keypoints and descriptors,

which are invariant to scale and rotation.
These allow reliable matching between images taken from different viewpoints.

- **Line 39 – Lowe's threshold:**
  A ratio threshold of 0.6 is used in Lowe's ratio test to filter out ambiguous or incorrect matches, keeping only the most distinctive feature correspondences.

- **Line 44 – Variables F and `inlier_mask`:**
  `F` is the fundamental matrix, a 3×3 rank-2 matrix that encodes the epipolar geometry between two uncalibrated images.
  The `inlier_mask` identifies which feature matches satisfy the epipolar constraint.

$$x'^T F x = 0$$

- **Line 46 – Variable E:**
  `E` is the essential matrix computed as $E = K^T F K$.
  It relates normalized image coordinates from two calibrated cameras and encodes their relative rotation (R) and translation (t)

$$x'^T E x = 0$$

- **Line 47 – Objective:**
  This step recovers the relative pose — rotation `R` and translation `t` — between the two camera views from the essential matrix using SVD decomposition.

- **Lines 49–56 – Objective and output X:**
  These lines perform triangulation to compute the 3D coordinates of scene points from their corresponding 2D image points in both views.
  The result `X` represents the reconstructed 3D point cloud.

$$X = \text{triangulate}(x, x', P_1, P_2)$$