

CMU Fall24 16820 Homework 1

Patrick Chen

Collaborators:

September 7, 2024

Q1.1 at page 2

Ans:

Given two camera projection matrices P_1 and P_2 corresponding to two cameras, we have the following equations (1) and equations (2), where x_1 represents the projected 2D point, on the image plane of camera1, of a 3D point, X_π , from another 3D plane π , and x_2 represents the projected 2D point of the same 3D point, X_π , on the image plane of camera2.

$$x_1 = P_1 X_\pi \quad (1)$$

$$x_2 = P_2 X_\pi \quad (2)$$

Both of equation (1) and (2) has the same 3D point X_π , it leads to the following equation (3):

$$P_1^{-1} x_1 = P_2^{-1} x_2 \quad (3)$$

Then we move P_1 to the right hand side by both multiply P_1 in equation (1) and (2), then we got equation (4).

$$x_1 = P_1 P_2^{-1} x_2 \quad (4)$$

It is the same form of the given equation (1) at page 2, which is the equation (5) here, where λ is a scale:

$$x_1 = \lambda H x_2 \quad (5)$$

So we get equation (6):

$$H = \lambda P_1 P_2^{-1} \quad (6)$$

Consequently, if we need to prove that H does exist, we need to prove that P_2^{-1} does exist. As X_π lies on a plane, it only has two degrees of freedom, and can be represented as a 2D point. So the projection matrix P_1 and P_2 can be reduced to containing only intrinsic matrix K_1 and K_2 , since the extrinsic matrix is no longer needed as the projection now is only 2D plane to 2D plane transformation and can be included in Homography, H. Since $P_1 = K_1$ and $P_2 = K_2$, and K_2 is invertible, so H does exist: $H = \lambda K_1 K_2^{-1}$.

Q1.2 at page 3

Ans:

1. h has 9 variables, but it only has 8 degrees of freedom. Specifically, h is the transformation of two 2D homogeneous coordinates, and if it is multiplied by a scale, λ , then the transformation equation still holds because of homogeneous coordinates. As a result, we have to add a constraint like setting the last element of h to be 1 or setting $\|h_2\| = 1$, and it decreases the degree for freedom by 1. So, h has 8 degrees of freedom.
2. Since h has 8 unknowns, and each point pair can contribute to 2 equations for solving h , we need 4 pairs of points to solve h .
3. Substitute the given equation of $\mathbf{x}_1^i \equiv H\mathbf{x}_2^i$ ($i \in \{1 \dots N\}$) with variables:

$$\begin{bmatrix} x_1^i \\ y_1^i \\ 1 \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} = \begin{bmatrix} x_2^i \\ y_2^i \\ 1 \end{bmatrix}$$

Then, we expand the equations to get:

$$\begin{aligned} x_1^i &= h_{11}x_2^i + h_{12}y_2^i + h_{13} \\ y_1^i &= h_{21}x_2^i + h_{22}y_2^i + h_{23} \\ 1 &= h_{31}x_2^i + h_{32}y_2^i + h_{33} \end{aligned}$$

By dividing the last element of homogeneous coordinates to get real plane coordinates, we get the following two equations:

$$\begin{aligned} x_1^i &= \frac{h_{11}x_2^i + h_{12}y_2^i + h_{13}}{h_{31}x_2^i + h_{32}y_2^i + h_{33}} \\ y_1^i &= \frac{h_{21}x_2^i + h_{22}y_2^i + h_{23}}{h_{31}x_2^i + h_{32}y_2^i + h_{33}} \end{aligned}$$

Though multiplying through by denominator and then rearrange:

$$\begin{aligned} h_{11}x_2^i + h_{12}y_2^i + h_{13} - h_{31}x_2^ix_1^i - h_{32}y_2^ix_1^i - h_{33}x_1^i &= 0 \\ h_{21}x_2^i + h_{22}y_2^i + h_{23} - h_{31}x_2^iy_1^i - h_{32}y_2^iy_1^i - h_{33}y_1^i &= 0 \end{aligned}$$

Then we rearrange it to fit into the matrix form $A_i h = 0$

$$\begin{bmatrix} x_2^i & y_2^i & 1 & 0 & 0 & 0 & -x_1^i x_2^i & -x_1^i y_2^i & -x_1^i \\ 0 & 0 & 0 & x_2^i & y_2^i & 1 & -y_1^i x_2^i & -y_1^i y_2^i & -y_1^i \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Then we derive A_i :

$$A_i = \begin{bmatrix} x_2^i & y_2^i & 1 & 0 & 0 & 0 & -x_1^i x_2^i & -x_1^i y_2^i & -x_1^i \\ 0 & 0 & 0 & x_2^i & y_2^i & 1 & -y_1^i x_2^i & -y_1^i y_2^i & -y_1^i \end{bmatrix}$$

4. (a) A trivial solution for h is all 0:

$$h = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- (b) A is not full rank. In homogeneous system, if h is not trivial, then A must not be invertible, or the only solution of h would be a zero vector (trivial). It means that A should be singular, in other words, A must not be full rank.
- (c) The impact of A being not full rank is that the smallest singular values would be zero or near zero, allowing to find a non-trivial solution h.

Q1.4.1 at page 4

Ans:

Given two cameras separated by a pure rotation, we have the following two equations:

$$x_1 = K_1IX \quad (7)$$

$$x_2 = K_2RX \quad (8)$$

Since X is the same data point in 3D space, then we can substitute X in equation (7) with X in equation (8) and result in the following equation:

$$x_1 = K_1R^{-1}K_2^{-1}x_2 \quad (9)$$

Since $x_1 = \lambda Hx_2$, then we can know that:

$$H = \frac{1}{\lambda}K_1R^{-1}K_2^{-1} \quad (10)$$

Since K_2 and R are invertible, $\det(K_2) \neq 0$ and $\det(R) \neq 0$, H does exist.

Q1.4.2 at page 4

Ans:

Recall from Q1.4.1, $H = K_1 R(\theta)^{-1} K_2^{-1}$, given that θ is the angle of rotation. Since $K_1 = K_2 = K$, we have:

$$H = K R(\theta)^{-1} K^{-1} \quad (11)$$

By multiplying two H together:

$$H^2 = (K R(\theta)^{-1} K^{-1})(K R(\theta)^{-1} K^{-1}) \quad (12)$$

By removing redundant identity matrix:

$$H^2 = K R(\theta)^{-1} R(\theta)^{-1} K^{-1} \quad (13)$$

According to the property of rotation matrix $R(\theta)^{-1} = R(\theta)^T = R(-\theta)$, and $R(-\theta)R(-\theta) = R(-2\theta)$, we have $R(\theta)^{-1}R(\theta)^{-1} = R(2\theta)^{-1}$, so we have:

$$H^2 = K R(2\theta)^{-1} K^{-1} \quad (14)$$

It demonstrates that H^2 corresponds to a rotation of 2θ .

Q1.4.3 at page 4

Ans:

It is because that either one of the following two conditions is needed to be held for the planar homography to exist:

1. The points in the 3D world lie on a 2D plane.
2. Only pure rotation between two views.

As a result, planar homography is not completely sufficient to map any scene image to another viewpoint.

Q1.4.4 at page 5

Ans:

The projection matrix P is a 3×4 matrix that can map a 3D point, X , to a 2D point, x :

$$x = PX \quad (15)$$

A line in 3D space can be represented as the following equation, where t is a parameter:

$$X(t) = X_1 + t(X_2 - X_1) \quad (16)$$

If we multiply the projection matrix, P , on both hand sides of equation (16), then we have:

$$PX(t) = PX_1 + tP(X_2 - X_1) \quad (17)$$

By moving P into parenthesis, then:

$$PX(t) = PX_1 + t(PX_2 - PX_1) \quad (18)$$

Then, we got the following equation:

$$x(t) = x_1 + t(x_2 - x_1) \quad (19)$$

Equation (19) is the equation of line on a 2D plane. As a result, the line can be preserved through perspective projection P .

Q2.1.1 at page 5

Ans:

FAST detector has the following differences when comparing with Harris corner detector:

1. FAST detector only compare intensities with neighbors in a circular range around a candidate point, while Harris corner detector uses much more expansive computations like first order derivative (gradient), summation of a window of gradients, and solving eigenvalue problems, leading to the result that FAST detector is much more faster than Harris corner detector.
2. FAST detector is more sensitive to noise, while Harris corner detector is more robust to noise because Harris corner detector gets more information from gradients in multiple directions of a window plus solving eigenvalue problems to get optimized candidates of corners.
3. FAST detector is faster but less accurate in the localizing features, while Harris corner detector is slower but more accurate in localizing feature points.

Q2.1.2 at page 5

Ans:

BRIEF descriptor has the following differences when comparing with filterbanks descriptor:

1. BRIEF descriptor chooses n pairs of points within a predefined window to form a n -bit binary descriptor by setting each bit 1 or 0 according to the comparison of each pair of selected points, while filterbanks convolve the image with a series of predefined filters to form an array of responses in each pixel as the descriptor.
2. BRIEF descriptor is faster than filterbanks descriptor as BRIEF descriptor only use sampling and comparison of intensity, while filterbanks descriptor applies convolution of different filters to produce different response as the descriptor.
3. As for the robustness, filterbanks descriptor is more robust to rotation and scale as it can apply different filters to mitigate the impact from scaling and rotating, while BRIEF descriptor only compares the intensity within a window and it is inherently not rotation-invariant and scale-invariant.

Q2.1.3 at page 5

Ans:

The Hamming distance is to calculate the number of different bits in two binary descriptors. For example, let x_1 be a d-dimensional binary descriptor: $x_1 = (x_1^0, x_1^1, \dots, x_1^{(d-1)})$, and x_2 is another d-dimensional binary descriptor: $x_2 = (x_2^0, x_2^1, \dots, x_2^{(d-1)})$, the Hamming distance is defined as the following:

$$d_H(\mathbf{x}_1, \mathbf{x}_2) = \sum_{i=1}^d \delta(x_1^i, x_2^i), \text{ where } \delta(x_1^i, x_2^i) = \begin{cases} 0 & \text{if } x_1^i = x_2^i, \\ 1 & \text{if } x_1^i \neq x_2^i. \end{cases} \quad (20)$$

Hamming distance has the following benefits over conventional Euclidean distance measure:

1. Hamming distance is faster than Euclidean distance because Hamming distance only needs to compute the number of different bits among two binary vectors, and it can be computed by efficient XOR and counter operations, while Euclidean distance needs to compute several complex operations, like vector subtraction, summation of square and square root operations, making Euclidean distance slow and expensive for computation.
2. In our setting, we use BREIF descriptor on our detected feature points, resulting in vectors in binary form, and it is definitely suitable to use Hamming distance to compare two binary feature descriptor vectors.

Q2.1.3 at page 5

Ans: