Assignment_4 Solution

1. Camera Calibration 1

a.

- Forward projection: Given world point $\underline{P}_i$ and project matrix M, compute image point P_i
- Calibration: Given world point $\underline{P}_i$ and image point P_i, compute project matrix M
- Reconstruction: Given image point P_i and project matrix M, compute the world point $\underline{P}_i$
- Forward projection is the easiest
- Reconstruction is the hardest

b. The word coordinate of point and corresponding image coordinate of point

c.

1. Estimate projection matrix M
2. Find parameters: K^*, R^*, T^*

d. $(18/7, 2)$

e. $\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 & 0 & 0 & -100 & -200 & -300 & -100 \\ 0 & 0 & 0 & 0 & 1 & 2 & 3 & 1 & -200 & -400 & -600 & -200 \end{bmatrix}$

f.

- 6 points for non-planar calibration and 5 points for planar calibration
- Solution obtained by:
    1. Write down the AX from world point and image point
    2. Use SVD to decompose the matrix AX, and the last column of matrix V^T is the $\hat{M}$
    3. $M = \xi\hat{M}$, where $|\xi| = 1/|a_3|$ and a_3 is the transpose of last row of matrix $\hat{M}$ except last column element

g. The principal is that $M = K^*[R^*|T^*] = \xi\hat{M}$, where R^* is orthogonal matrix which means the row of vector in $R^*: r_1, r_2, r_3$ is orthogonal to each other and is unit vector. So we can use dot product and cross product to compute the parameters

h.

$$\text{MSE: } E(K^*, R^*, T^*) = \frac{1}{n}\sum_{i=1}^{n}\left((x_i - \frac{m_1^T P_i'}{m_3^T P_i'})^2 + (y_i - \frac{m_2^T P_i'}{m_3^T P_i'})^2\right)$$

i.

- The principle of planar camera calibration:
    1. Estimate the 2D Homography Matrix between calibration target and image.

1. The coordinate of z of world point is 0, so estimate the 2D Homography Matrix H.
    2. Write down the AX from world point and image point.
    3. Use SVD to decompose the matrix AX, and the last column of matrix V^T is the $\hat{H}$.
    4. Then we assume $H = \alpha\hat{H}$
2. Estimate intrinsic parameters from several views.
    1. According to $H = \alpha\hat{H}$, and the rotation vector r_1 and r_2 are orthogonal to each other, we get two dot product equations: ($S = {K^*}^{-T}{K^*}^{-1}$)
        - $r_1 \cdot r_1 = r_2 \cdot r_2 = 1 \Rightarrow \hat{h}_1^T S \hat{h}_1 = \hat{h}_2^T S \hat{h}_2$
        - $r_1 \cdot r_2 = 0 \Rightarrow \hat{h}_1^T S \hat{h}_2 = 0$
    2. We need at least 3 Homography Matrix, then we can build 6 equations to solve the K^* with 5 unknown parameters
3. Compute extrinsic parameters for any views. According to $H = \alpha\hat{H}$, we can get:
    - $r_1 = \alpha{K^*}^{-1}\hat{h}_1 \quad \text{and} \quad r_2 = \alpha{K^*}^{-1}\hat{h}_2$
    - $T^* = \alpha{K^*}^{-1}\hat{h}_3$

- The difference is that all points in same planar in planar calibration, while it is not in non-planar calibration

j.

- The difference is that there are two rotation vectors r_1 and r_2 in Homography matrix, while there are three rotation vectors r_1, r_2, and r_3 in projection matrix
- We assume the z coordinate is 0 in world point $\underline{P}_i$

---

2. Camera Calibration 2

a. $\begin{bmatrix} 3 & 4 & 5 & 1 & 0 & 0 & 0 & 0 & -3 & -4 & -5 & -1 \\ 0 & 0 & 0 & 0 & 3 & 4 & 5 & 1 & -6 & -8 & -10 & -2 \end{bmatrix}$

b. $M = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \end{bmatrix}$

- $u_0 = |\xi|^2 a_1 \cdot a_3 = \frac{a_1 \cdot a_3}{|a_3|^2} = \frac{3+8+15}{9+16+25} = \frac{26}{50} = 0.52$
- $v_0 = |\xi|^2 a_2 \cdot a_3 = \frac{a_2 \cdot a_3}{|a_3|^2} = \frac{6+12+20}{9+16+25} = \frac{38}{50} = 0.76$

c. 1.7337

d. $I + Q = \left[\begin{array}{c|c} R^*_{3\times3} & \mathbf{0}_{3\times1} \\ \hline \mathbf{0}_{1\times3} & 1 \end{array}\right]_{\text{3DH}} = \left[\begin{array}{ccc|c} 6 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \end{array}\right]_{\text{3DH}}$

- $R = (R^*)^T = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}_{\text{3D}}$
- $T = -RT^* = -(R^*)^T T^* = [-6 \quad -2 \quad -3]^T_{\text{3D}}$

e. $\begin{bmatrix} 3 & 4 & 1 & 0 & 0 & 0 & -3 & -4 & -1 \\ 0 & 0 & 0 & 3 & 4 & 1 & -6 & -8 & -2 \end{bmatrix}$

---

3. Multiple View Geometry 1

a.

- Sparse stereo matches specific points and dense stereo matches each pixel
- Sparse can be used for far (between frames) views, while dense should be used for close (between frames) views. Both cannot match uniform patch, invisible in one view, and ambiguity.

b.

- NCC computes the normalized cross correlation of two windows around the point. SSD computes the sum of square distance between two windows around the point. They are used to compute similarity to determine correspondence.
- If allowing the search space to be the entire image, we might get more mismatching points as we are more likely to get errors.
- The solution is to compute the epipole line.

c. $1000/3$

d. The point pair may not match correctly. Incorrect matching will lead to incorrect depth recovery.

e.

- $R = R_l^T R_r$
- $T = R_l^T (T_r - T_l)$

---

4. Multiple View Geometry 2

a. $20/3$

b. $\begin{bmatrix} 0 & -3 & 2 \\ 3 & 0 & -1 \\ -2 & 1 & 0 \end{bmatrix}$

c. 68

d. The left point is $P_i' = (x' = 1, y' = 2)$ and right point is $P_i = (x = 2, y = 3)$
$(P_r^T F P_l = P_i^T F P_i' = 0)$

$[xx' \quad xy' \quad x \quad yx' \quad yy' \quad y \quad x' \quad y' \quad 1] = [2 \quad 4 \quad 2 \quad 3 \quad 6 \quad 3 \quad 1 \quad 2 \quad 1]$