2018-17119 박수현 컴퓨터비전 HW3 theory writeup

Question1: Camera model (30 points)

Consider the camera model below.

$$\underset{\text{image}}{\begin{pmatrix} u \\ v \\ 1 \end{pmatrix}} = \mathbf{K} \left(\mathbf{R} \mid \mathbf{t} \right) \underset{\text{world}}{\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}} \quad (1)$$

camera -> image, world -> camera

$$\mathbf{K} = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}, \mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}, \quad (2)$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

The coordinate $[u, v, 1]^T$ and $[X, Y, Z, 1]^T$ represent homogeneous coordinates in the image and world coordinate system, respectively. $\mathbf{K}$ is the camera intrinsic matrix, $\mathbf{R}$ is the rotation matrix, and $\mathbf{t}$ is the translation vector. Note that the rotation matrix has some unique properties, that each row vectors and column vectors are unit vectors, and $\mathbf{R}^{-1} = \mathbf{R}^T$. (You can use these properties without any proof.)

(a) What is the image coordinates of the world coordinate system's origin $O(0, 0, 0)$? [5 points]

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} = \begin{bmatrix} ft_1 \\ ft_2 \\ t_3 \end{bmatrix}$$

Converting from homogeneous coordinates, $u = \frac{ft_1}{t_3}$, $v = \frac{ft_2}{t_3}$

$\therefore \left(f\frac{t_1}{t_3}, f\frac{t_2}{t_3} \right)$

(b) What is the coordinate of the camera in the world coordinate system? [5 points]

The camera coordinate frame is related to the world coordinate frame by a rotation (to align axes) and a translation (to align origins).

Let coordinates of a point in the camera frame $X_c = (X_c, Y_c, Z_c)$

coordinates of a point in the world frame $X_w = (X_w, Y_w, Z_w)$

coordinates of the camera center in the world frame $C = (C_1, C_2, C_3)$

$\Rightarrow X_c = R(X - C)$, and in matrix representation, $X_c = [R \mid -RC] \begin{bmatrix} X_w \\ 1 \end{bmatrix}$

($\because$ rotation & translation)

As the extrinsic camera parameters in the model can be represented as a 4x4 3D transformation matrix $[R \mid t]$, $t = -RC$

$t = -RC \iff -t = RC \iff (R^{-1})(-t) = C \iff R^T(-t) = C$

$$C = (R^T)(-t) = \begin{bmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{bmatrix} \begin{bmatrix} -t_1 \\ -t_2 \\ -t_3 \end{bmatrix} = \begin{bmatrix} -r_{11}t_1 - r_{21}t_2 - r_{31}t_3 \\ -r_{12}t_1 - r_{22}t_2 - r_{32}t_3 \\ -r_{13}t_1 - r_{23}t_2 - r_{33}t_3 \end{bmatrix}$$

$$\therefore \begin{bmatrix} -r_{11}t_1 - r_{21}t_2 - r_{31}t_3 \\ -r_{12}t_1 - r_{22}t_2 - r_{32}t_3 \\ -r_{13}t_1 - r_{23}t_2 - r_{33}t_3 \end{bmatrix}$$ (3x1 matrix)

(c) In the camera coordinate system, camera is looking at $[0, 0, 1]^T$ (*i.e.* $+z$). Convert this direction into the world coordinate system with a unit vector. [8 points]

Followed by Question 1 (b), $X_c = RX_w + T \iff X_w = R^T(X_c - t)$

Given $X_c = (0, 0, 1)$,

$$R^T(X_c - t) = \begin{bmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{bmatrix} \begin{bmatrix} -t_1 \\ -t_2 \\ 1-t_3 \end{bmatrix} = \begin{bmatrix} -r_{11}t_1 - r_{21}t_2 + r_{31}(1-t_3) \\ -r_{12}t_1 - r_{22}t_2 + r_{32}(1-t_3) \\ -r_{13}t_1 - r_{23}t_2 + r_{33}(1-t_3) \end{bmatrix} \Leftarrow V \text{ (a } 3\times 1 \text{ matrix)}$$

To find a unit vector with the same vector as V, we divide v by its magnitude.

$$\|V\| = \sqrt{(-r_{11}t_1 - r_{21}t_2 + r_{31}(1-t_3))^2 + (-r_{12}t_1 - r_{22}t_2 + r_{31}(1-t_3))^2 + (-r_{13}t_1 - r_{23}t_2 + r_{33}(1-t_3))^2}$$

$$\therefore \frac{V}{\|V\|} = \frac{1}{\sqrt{(-r_{11}t_1 - r_{21}t_2 + r_{31}(1-t_3))^2 + (-r_{12}t_1 - r_{22}t_2 + r_{31}(1-t_3))^2 + (-r_{13}t_1 - r_{23}t_2 + r_{33}(1-t_3))^2}} \begin{bmatrix} -r_{11}t_1 - r_{21}t_2 + r_{31}(1-t_3) \\ -r_{12}t_1 - r_{22}t_2 + r_{32}(1-t_3) \\ -r_{13}t_1 - r_{23}t_2 + r_{33}(1-t_3) \end{bmatrix}$$

(d) The straight line in the world coordinate system can be expressed as $[d_1, d_2, d_3]^T t + [x_1, x_2, x_3]^T$ ($t \in \mathbf{R}$). Assume that this line is not parallel to the xy-plane and does not path through the origin in the camera coordinate system. What is the expression of this line in the image? [12 points]

$$\text{Let line } \ell = \underbrace{\begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}}_{\text{direction}} t + \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\text{point}}, \quad t \in \mathbb{R}$$

$$K(R|t)\ell = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \left(\begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ 1 \end{bmatrix} t + \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} fr_{11} & fr_{12} & fr_{13} & ft_1 \\ fr_{21} & fr_{22} & fr_{23} & ft_2 \\ fr_{31} & fr_{32} & fr_{33} & ft_3 \end{bmatrix} \left(\begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ 1 \end{bmatrix} t + \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} \right)$$

$$= \div \left(\begin{bmatrix} fr_{11}d_1 + fr_{12}d_2 + fr_{13}d_3 + ft_1 \\ fr_{21}d_1 + fr_{22}d_2 + fr_{23}d_3 + ft_2 \\ fr_{31}d_1 + fr_{32}d_2 + fr_{33}d_3 + ft_3 \end{bmatrix} t + \begin{bmatrix} fr_{11}x_1 + fr_{12}x_2 + fr_{13}x_3 + ft_1 \\ fr_{21}x_1 + fr_{22}x_2 + fr_{23}x_3 + ft_2 \\ fr_{31}x_1 + fr_{32}x_2 + fr_{33}x_3 + ft_3 \end{bmatrix} \right) \div$$

$$\therefore \frac{1}{fr_{31}d_1 + fr_{32}d_2 + fr_{33}d_3 + ft_3} \begin{bmatrix} fr_{11}d_1 + fr_{12}d_2 + fr_{13}d_3 + ft_1 \\ fr_{21}d_1 + fr_{22}d_2 + fr_{23}d_3 + ft_2 \end{bmatrix} t$$

$$+ \frac{1}{fr_{31}x_1 + fr_{32}x_2 + fr_{33}x_3 + ft_3} \begin{bmatrix} fr_{11}x_1 + fr_{12}x_2 + fr_{13}x_3 + ft_1 \\ fr_{21}x_1 + fr_{22}x_2 + fr_{23}x_3 + ft_2 \end{bmatrix}, \quad \text{for } t \in \mathbb{R}$$

Question2: Camera Calibration (10 points)

A camera is a mapping between the 3D world $\mathbf{X_i} = \begin{bmatrix} X_i & Y_i & Z_i & 1 \end{bmatrix}^T$ and a 2D image $\mathbf{x_i} = \begin{bmatrix} u_i & v_i & 1 \end{bmatrix}^T$ (both are homogeneous coordinates). A camera projection matric $\mathbf{P} = [m_{ij}]$ maps $\mathbf{x_i} = \mathbf{P}\mathbf{X_i}$. Given $n \geq 6$ pairs of correspoinding points, we have an over-determined system of equations, $\mathbf{Ap} = \mathbf{0}$, where $\mathbf{p}$ is the vectorized form of $\mathbf{P}$. Consider the SVD of $\mathbf{A}$ as $\mathbf{A} = \mathbf{U\Sigma V^T}$. Show that the solution $\mathbf{p} = \min_p \|Ap\|, \|p\| = 1$ is the last column of $\mathbf{V}$ corresponding to smallest ~~eigenvalue~~ singular value of $\mathbf{A}$.

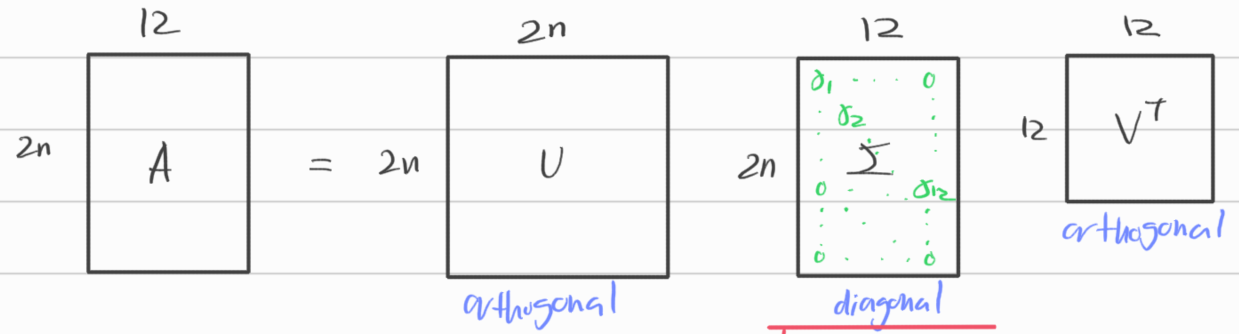
or smallest eigenvalue of A^TA

$$\overset{x_i}{\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}} = \overset{P}{\begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix}} \overset{X_i}{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}} \quad (3)$$

P : 12×1 matrix ("flatten-out" vectorized form of P)

A : $2n \times 12$ matrix ($2n \geq 12 \because n \geq 6$)

For background information, the SVD of A is a decomposition of the form



↳ diagonal elements are singular values $\sigma_1, \dots, \sigma_{12}$; others are all zeros.

Let the solution be $\hat{p}$, and I will prove that $\hat{p}$ is the last column of V corresponding to the smallest singular value of A.

$$\hat{p} = \min_p \|U\Sigma V^T p\|$$

$$= \min_p \|\Sigma V^T p\| \quad \leftarrow \|U\Sigma V^T p\| = \|\Sigma V^T p\|$$

$\because$ An orthogonal $m \times m$ matrix Q satisfies $Q^TQ = I_m$, i.e., $Q^T = Q^{-1}$ by definition.

Note that U and V^T are orthogonal matrices:

$$\|U\Sigma V^T p\| = \sqrt{U\Sigma V^T p \cdot U\Sigma V^T p} = \sqrt{(U\Sigma V^T p)^T U\Sigma V^T p}$$

$$= \sqrt{(\Sigma V^T p)^T \underbrace{U^T U}_{= I_{2n}} \Sigma V^T p} = \sqrt{(\Sigma V^T p)^T \Sigma V^T p} = \|\Sigma V^T p\|$$

$$= \min_p \|\Sigma y\| \quad \leftarrow \text{substitute } y = V^T p \ (12 \times 1 \text{ matrix})$$

In practice, the singular values in Σ are in descending order, i.e., $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{12}$.
(cf. numpy.linalg.svd outputs Σ as vectors with the singular values, within each vector sorted in descending order, as well)

Since $\hat{p} = \min_p \|\Sigma y\|$, in order for $\hat{p}$ to contain the smallest singular value σ_{12}, the best option for y is $y = [0, 0, \cdots, 0, 1]^T$.

Recalling that $y = V^T p$ from substitution,

$p = Vy$ ($\because$ V^T is orthogonal, so $V = (V^T)^{-1}$)

and $\|p\| = 1$ ($\because$ $\|p\| = \|Vy\| = \|y\| = 1$, V is orthogonal)

Therefore, the solution $\hat{p}$ where $\|p\| = 1$ is the last column of V, which corresponds to the smallest singular value.

Question3: Homography (10 points)

We can use homographies to create a panoramic image from multiple views of the same scene. This is possible when there is no camera translation between the views (*e.g.* only rotation about the camera center). In this case, corresponding points from two views of the same scene can be related by a homography:

$$\boldsymbol{p_1}^i \equiv \boldsymbol{H}\boldsymbol{p_2}^i \tag{4}$$

where $\boldsymbol{p_1}^i$ and $\boldsymbol{p_2}^i$ denote the homogeneous coordinates (*e.g.*, $\boldsymbol{p_1}^i \equiv (x, y, 1)^T$) of the 2D projection of the i-th point in images 1 and 2 respectively, and $\mathbf{H}$ is a 3×3 matrix representing the homography. Given N point correspondences and using (4), derive a set of 2N independent linear equations in the form $\mathbf{Ah} = \mathbf{b}$ where $\mathbf{h}$ is a 9×1 vector containing the unknown entries of $\mathbf{H}$.

(a) What are the expressions for $\mathbf{A}$ and $\mathbf{b}$? [3 points]

Let $p_1^i = (x_1, y_1, 1)^T$, $p_2^i = (x_2, y_2, 1)^T$

$p_1^i \equiv H p_2^i \iff \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \cong \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix}$, $\quad x_1 = \dfrac{h_{00}x_2 + h_{01}y_2 + h_{02}}{h_{20}x_2 + h_{21}y_2 + h_{22}}$, $\quad y_1 = \dfrac{h_{10}x_2 + h_{11}y_2 + h_{12}}{h_{20}x_2 + h_{21}y_2 + h_{22}}$

$\Rightarrow x_1(h_{20}x_2 + h_{21}y_2 + h_{22}) = h_{00}x_2 + h_{01}y_2 + h_{02}$

$\quad y_1(h_{20}x_2 + h_{21}y_2 + h_{22}) = h_{10}x_2 + h_{11}y_2 + h_{12}$

Using the linear equations, $p_1^i \equiv Hp_2^i$ can be expressed in the form $Ah = b$ where

$$\underbrace{\begin{bmatrix} x_2^1 & y_2^1 & 1 & 0 & 0 & 0 & -x_1^1x_2^1 & -x_1^1y_2^1 & -x_1^1 \\ 0 & 0 & 0 & x_2^1 & y_2^1 & 1 & -y_1^1x_2^1 & -y_1^1y_2^1 & -y_1^1 \\ & & & & \vdots & & & & \\ x_2^i & y_2^i & 1 & 0 & 0 & 0 & -x_1^ix_2^i & -x_1^iy_2^i & -x_1^i \\ 0 & 0 & 0 & x_2^i & y_2^i & 1 & -y_1^ix_2^i & -y_1^iy_2^i & -y_1^i \\ & & & & \vdots & & & & \\ x_2^N & y_2^N & 1 & 0 & 0 & 0 & -x_1^Nx_2^N & -x_1^Ny_2^N & -x_1^N \\ 0 & 0 & 0 & x_2^N & y_2^N & 1 & -y_1^Nx_2^N & -y_1^Ny_2^N & -y_1^N \end{bmatrix}}_{\substack{A \\ 2N\times 9}} \underbrace{\begin{bmatrix} h_{00} \\ h_{01} \\ h_{02} \\ h_{10} \\ h_{11} \\ h_{12} \\ h_{20} \\ h_{21} \\ h_{22} \end{bmatrix}}_{\substack{h \\ 9\times 1}} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}}_{\substack{b \\ 2N\times 1}}$$

(b) How many correspondences will be needed to solve for $\mathbf{h}$? [4 points]

4 or more $(N \geq 4)$

$\because$ While h has 9 unknown entries, we can multiply all h_{ij} $(i, j \in \{0,1,2\})$ by nonzero scaling factor k without changing the equations. Therefore, the degree of freedom is 8. Since each point has x and y coordinates.

with at least 4 correspondances, the problem would be defined as an overdetermined system of equations. As a result, we can apply the least squares technique.

(c) How can we get **H**? (Hint: Question 2) [3 points]

As soon as we find h, we can recover H by reshaping the matrix from 9×1 to 3×3.

h is a (non zero) solution to a least squares problem $\|Ah\|^2$.

Followed by Question 2, considering the SVD of A as $A = U \Sigma V^T$, the solution h for $\min_h \|Ah\|$ is the last column of V corresponding to the smallest singular value of A.

Note that minimizing $\|Ah\|^2$ and $\|Ah\|$ should both yield the same solution vector h.

In conclusion, we can get H through the following steps:

1. Construct a $2N \times 9$ matrix A using coordinates of p1 and p2, as presented in Question 3 (a).
2. Calculate the singular value decomposition of matrix A s.t. $A = U \Sigma V^T$
3. Retrieve the last column of V (= last row of V^T) as h.
4. Reshape h (9×1 matrix) into H (3×3 matrix).