CSCI 677: Advanced Computer Vision - Fall 2021

Instructor: Prof. Nevatia

Assignment 1

Due on September 16, 2021 before 17:00 PDT

Submit a PDF file on course page; you may scan the hand-written solution and need not format the mathematical expressions.

Problem 1. (4 points) Consider a line l in the image, given by parameters (a, b, c), in the image coordinate system. We know that the corresponding 3-D line casting this image lies in a plane. Derive the equation of this plane (in the camera coordinate system). You may assume that the intrinsic calibration matrix, $\boldsymbol{K}$, is given.

Solution. The line l in the image is given by the parameters (a, b, c), thus the line in the 2-D image can be written as

$$ax + by + cz = 0$$

In the world coordinate system, we assume our plane π is

$$\alpha X + \beta Y + \gamma Z + \phi = 0$$

where $[X, Y, Z, 1]^T$ is the point in the world coordinate. This plane has to go through the origin, which is $(0, 0, 0)$ in the camera coordinate system, thus we get $\phi = 0$. The function of the plane π is

$$\alpha X + \beta Y + \gamma Z = \begin{bmatrix}\alpha & \beta & \gamma\end{bmatrix} \cdot \begin{bmatrix}X \\ Y \\ Z\end{bmatrix} = 0 \tag{1}$$

Since the intrinsic calibration matrix, $\boldsymbol{K}$, is given, and we have $\boldsymbol{R}$ as $\boldsymbol{I}_{3\times3}$ and $\boldsymbol{T}$, we derive the projection matrix $\boldsymbol{M}$ as

$$\boldsymbol{M} = \boldsymbol{K}[\boldsymbol{Id}, \boldsymbol{0}]$$

Thus the projection of the point $[X, Y, Z, 1]^T$ on the plane π be written as

$$\begin{bmatrix}x \\ y \\ 1\end{bmatrix} = \frac{1}{Z}\boldsymbol{M}\begin{bmatrix}X \\ Y \\ Z \\ 1\end{bmatrix} = \frac{1}{Z}\boldsymbol{K}[\boldsymbol{Id}, \boldsymbol{0}]\begin{bmatrix}X \\ Y \\ Z \\ 1\end{bmatrix} = \frac{1}{Z}\boldsymbol{K}([\boldsymbol{Id}; \boldsymbol{0}]\begin{bmatrix}X \\ Y \\ Z \\ 1\end{bmatrix}) = \frac{1}{Z}\boldsymbol{K}\begin{bmatrix}X \\ Y \\ Z\end{bmatrix}$$

Since on the image plane, the function of our line is $ax + by + c = 0$. Thus we have

$$\begin{bmatrix}a & b & c\end{bmatrix} \cdot \begin{bmatrix}x \\ y \\ 1\end{bmatrix} = \frac{1}{Z}\begin{bmatrix}a & b & c\end{bmatrix} \cdot \boldsymbol{K}\begin{bmatrix}X \\ Y \\ Z\end{bmatrix} = 0 \tag{2}$$

Since the points $[X, Y, Z, 1]^T$ are the points on the plane π in the world coordinate and we have Eq. 1 and Eq. 2, if we want to have Eq. 1 and Eq. 2 represent the same plane in the world coordinate, we can derive that $[\alpha, \beta, \gamma] = \lambda[a, b, c]\boldsymbol{K}$. For any $\lambda \neq 0$, the plane is the same. Thus we assume $\lambda = 1$ and the function of plane π can be written as

$$\begin{bmatrix}a & b & c\end{bmatrix} \cdot \boldsymbol{K}\begin{bmatrix}X \\ Y \\ Z\end{bmatrix} = 0$$

Grading details:

1. 1 point is given if the equation of the plane is shown.

2. 1 point is given if the equation of the line is shown.

3. 1 point is given if the equation of projection matrix is shown.

4. 1 point is given if the points in the projection matrix are used correctly.

□

Problem 2. (11 points) Suppose that we have a right-handed camera coordinate system (X_c, Y_c, Z_c) associated with its origin at the lens center (or the pinhole), as in the examples discussed in class. Suppose that the imaging plane is at a distance of 50 millimeters from the lens center, the imaging surface (a planar patch) is 1200×1200 pixels, each pixel is .04 millimeters in each dimension, and that the principal ray intersects the imaging surface in the center. Let the image (or retinal) coordinate system have its origin at the upper-left corner of the imaging sensor, the x-axis along the top-row, and the y-axis points downward at an acute angle of 88 degrees to the x-axis). Assume that the x-axis in the image plane is parallel to the x-axis in the normalized image plane.

- For these conditions, derive the intrinsic matrix K, which helps map a point, specified in the normalized image coordinate frame to the image coordinates $(x, y, 1)^T$ expressed in pixel units (ignore the issue of rounding off pixel coordinates to integers).
- Now suppose that the camera is placed in a world coordinate system (O_w, X_w, Y_w, Z_w) such that O_c is at location $(4, -3, 2)$ in the world coordinate system (all distances expressed in meters); X_c is parallel to X_w and then the camera is rotated by 15 degrees about the X_c axis in a clockwise direction (visualize as a person taking a picture with camera pointing down slightly). Compute the final projection matrix, M.
- Consider a set of parallel lines in the horizontal plane (i.e. the $X_w - Z_w$ plane). Find the vanishing point, in the image coordinates, of this set of lines in terms of the direction of the lines for the above configuration.

Solution. We present the solution of the problem as follows.

- Focal length $f = 50$mm. As the square pixel is 0.04 mm/pixel, $k = l = \frac{1}{0.04}$pixel/mm. Thus, we have

$$\alpha = kf = 1250 \text{ pixel}, \tag{3}$$
$$\beta = lf = 1250 \text{ pixel}. \tag{4}$$

  The flipping (retinal origin at upper left corner, the x-axis along the top-row and the y-axis points downward) and skewing are represented in the $\theta = -88°$.

  The optical center is in the center of the square sensor, but its coordinates (x_0, y_0) are in the non-standard axes. Thus

$$x_0 = 600 - 600 \cot|\theta| = 579.048, \tag{5}$$
$$y_0 = \frac{600}{sin|\theta|} = 600.366 \tag{6}$$

  Thus, we have

$$K = \begin{bmatrix} \alpha & -\alpha\cot\theta & x_0 \\ 0 & \frac{\beta}{\sin\theta} & y_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1250 & 43.651 & 579.048 \\ 0 & -1250.762 & 600.366 \\ 0 & 0 & 1 \end{bmatrix} \tag{7}$$

  Common Mistakes:

  - **"Skewed offset issue"**: 1 point deducted. Some case take 0.5 point since the student is aware of this issue but didn't get it completely correct. Please refer to Fig. 1 for references.
  - **Use $\theta = 88°$**: 1 point deducted.

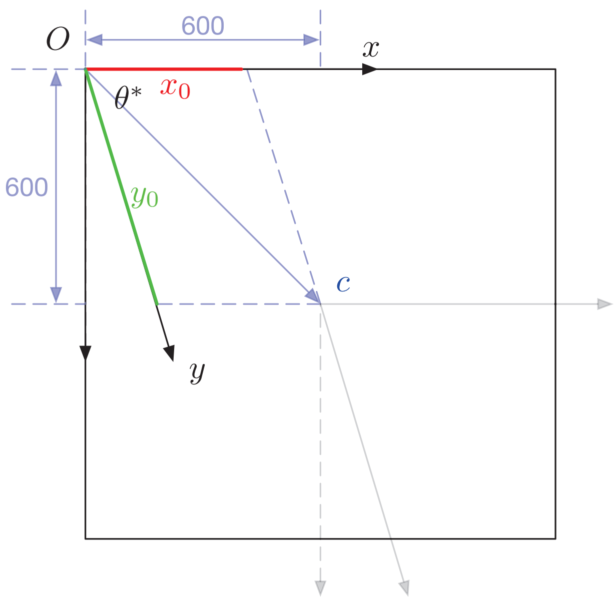

Figure 1: The offset coordinates (x_0, y_0) corresponds to the skewed retina coordinate system.

- We aim at solving the extrinsic matrix, which consists of rotation matrix $\mathbf{R}$ and translation $\mathbf{T}$ that project any point from the world coordinate system $\{\mathcal{W}\}$ to the camera coordinate system $\{\mathcal{C}\}$: $\mathbf{P}^{\mathcal{C}} = \mathbf{R}\mathbf{P}^{\mathcal{W}} + \mathbf{T}$.

  Given that the view plane rotates clockwise along the x-axis by 15°, we denote the angle as $\phi = -15°$. (counterclockwise rotations have positive rotation angles.) Hence, we have

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.966 & -0.259 \\ 0 & 0.259 & 0.966 \end{bmatrix} \tag{8}$$

  We know that the origin of the camera coordinate system in $\{\mathcal{W}\}$ is $\mathbf{O}^{\mathcal{W}} = [4\ -3\,2]^T$. We also know that the coordinates of this origin in $\{\mathcal{C}\}$ is $\mathbf{O}^{\mathcal{C}} = [0\,0\,0]^T$. Given that $\mathbf{O}^{\mathcal{W}}, \mathbf{O}^{\mathcal{C}}$ satisfy the projection equation: $\mathbf{O}^{\mathcal{C}} = \mathbf{R}\mathbf{O}^{\mathcal{W}} + \mathbf{T}$, we can obtain the translation T as

$$\begin{aligned} \mathbf{T} &= -\mathbf{R}\mathbf{O}^{\mathcal{W}} && (9) \\ &= -\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.966 & -0.259 \\ 0 & 0.259 & 0.966 \end{bmatrix} \begin{bmatrix} 4 \\ -3 \\ 2 \end{bmatrix} && (10) \\ &= \begin{bmatrix} -4 \\ 3.416 \\ -1.155 \end{bmatrix}. && (11) \end{aligned}$$

  Then, we can derive the projection matrix $\mathbf{M}$ from $\mathbf{K}[\mathbf{R}\,\mathbf{T}]$:

$$\begin{aligned} \mathbf{M} &= \mathbf{K}[\mathbf{R}\,\mathbf{T}] && (12) \\ &= \begin{bmatrix} 1250 & 192.140 & 548.055 & -5519.689 \\ 0 & -1052.754 & 903.853 & -4965.968 \\ 0 & 0.259 & 0.966 & -1.155 \end{bmatrix} && (13) \end{aligned}$$

- The direction of all lines in the horizontal plane in $\{\mathcal{W}\}$ can be written as $[s_x\,0\,s_z]^T$, with arbitrary s_x and s_z that cannot be zero at the same time. Following the lecture, we can derive that all lines intersects Π_∞ at $[s_x\,0\,s_z\,0]^T$.

Therefore, we can have the corresponding vanishing point as

$$\tilde{V} = \mathbf{M} \begin{bmatrix} s_x \\ 0 \\ s_z \\ 0 \end{bmatrix} = \begin{bmatrix} 1250s_x + 548.055s_z \\ 903.853s_z \\ 0.966s_z \end{bmatrix}. \tag{14}$$

Then we can obtain the image coordinate as

$$\tilde{v} = \begin{bmatrix} 1294.00\frac{s_x}{s_z} + 567.344 \\ 935.666 \\ 1 \end{bmatrix} \tag{15}$$

Common Mistakes:

- **Rotation direction incorrect**: 1 point deducted. The sign of the rotation angle depends on how the Rotation matrix is written.
- **Translation incorrect**: 1 point deducted. The translation vector should be derived under the camera coordinate system.
- **Didn't unify the last element**: The last element of the homogeneous vector should be unified to 1. 0.5 point deducted.

□