Lecture 1: The Pinhole Camera Model

1 Mathematical Model

The most commonly used model, which we will also use in the course, is the so called pinhole camera. The model is inspired by the simplest cameras. It has the shape of a box, light from an object enters though a small hole (the pinhole) in the front and produces an image on the back camera wall (see Figure 1).

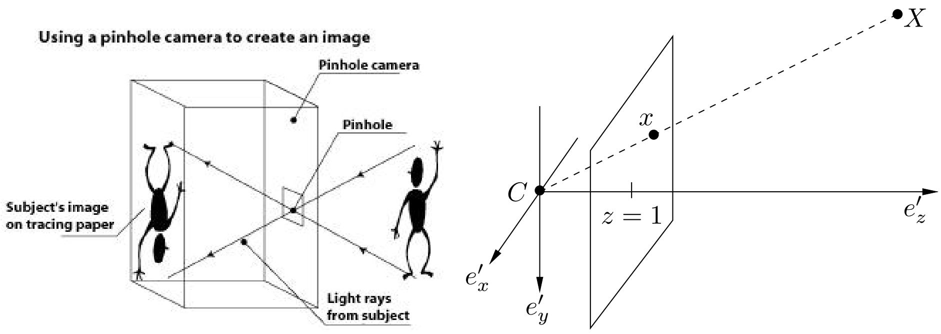

Figure 1: The Pinhole camera (left), and a mathematical model (right).

To create a mathematical model we first select a coordinate system $\{e'_x, e'_y, e'_z\}$. We will refer to this system as the **camera coordinate system**. The origin $C = (0,0,0)$ will represent the so called **camera center** (pinhole). To generate a projection $x = (x_1, x_2, 1)$ of a scene point $X = (X'_1, X'_2, X'_3)$ we form the line between X and C and intersect it with the plane $z = 1$. We will refer to this plane as the **image plane** and the line as the **viewing ray** associated with x or X. The plane $z = 1$ has the normal e_z and lies at the distance 1 from the camera center. We will refer to e_z as the **viewing direction**. Note that in contrast to a real pinhole camera we have placed the image plane in front of the camera center. This has the effect that the image will not appear upside down as in the real model.

Since $X - C$ is a direction vector of the viewing ray (see Figure 2) we can parametrize it by the expression

$$C + s(X - C) = sX, \quad s \in \mathbb{R}. \tag{1}$$

To find the intersection between this line and the image plane $z = 1$ we need to find and s such that the third coordinate sX'_3 of sX fulfills $sX'_3 = 1$. Therefore, assuming $X'_3 \neq 0$, we get $s = 1/X'_3$ and the projection

$$x = \begin{pmatrix} X'_1/X'_3 \\ X'_2/X'_3 \\ 1 \end{pmatrix} \tag{2}$$

Exercise 1. Compute the image of the cube with corners in $(\pm 1, \pm 1, 2)$ and $(\pm 1, \pm 1, 4)$, see Figure 3.

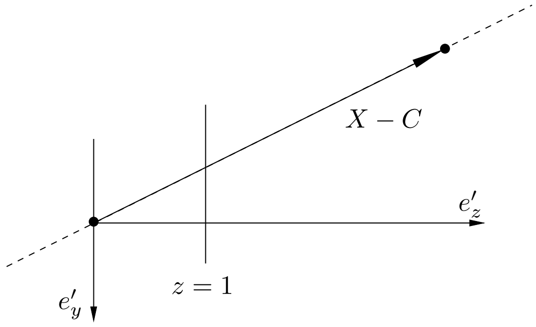

Figure 2: The model viewed from the side. (The vector e'_x points out of the figure.)

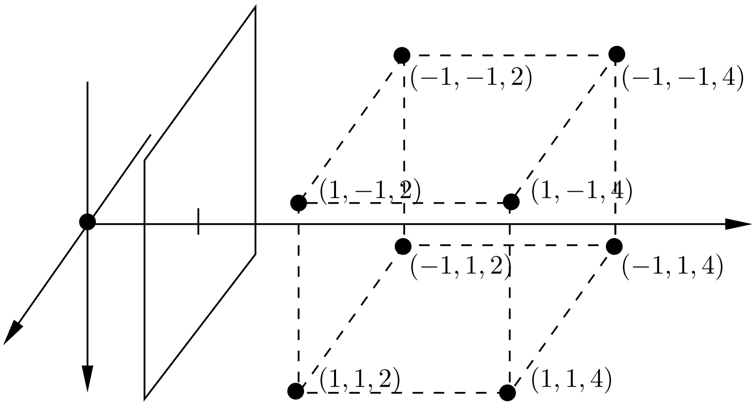

Figure 3: The Cube in the camera coordinate system.

2 Moving Cameras

In our applications we will frequently have cameras that have been capturing images from different viewpoints. Therefore we need to have a way of modeling camera movements. A camera can undergo translation and rotation. We will represent the translation with a vector $t \in \mathbb{R}^3$ and the rotation with a 3×3 matrix R. Since R is a rotation matrix it has to fulfill $R^T R = I$ and $\det(R) = 1$.

To encode camera movements we introduce a new reference coordinate system $\{e_x, e_y, e_z\}$, see Figure 4. We will refer to this coordinate system as the **global coordinate system**, since all the camera movements will be related to this system. Typically all the scene point coordinates are also specified in this coordinate system. Let's assume that a scene point X has coordinates (X'_1, X'_2, X'_3) in the camera coordinate system and (X_1, X_2, X_3) in the global coordinate system. Since the camera can be rotated and translated there is a rotation matrix R and translation vector t that relates the two coordinate systems via

$$\begin{pmatrix} X'_1 \\ X'_2 \\ X'_3 \end{pmatrix} = R \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} + t. \tag{3}$$

Exercise 2. What is the position of the camera (the coordinates of the camera center) in the global coordinate system? What is the viewing direction in the global coordinate system?

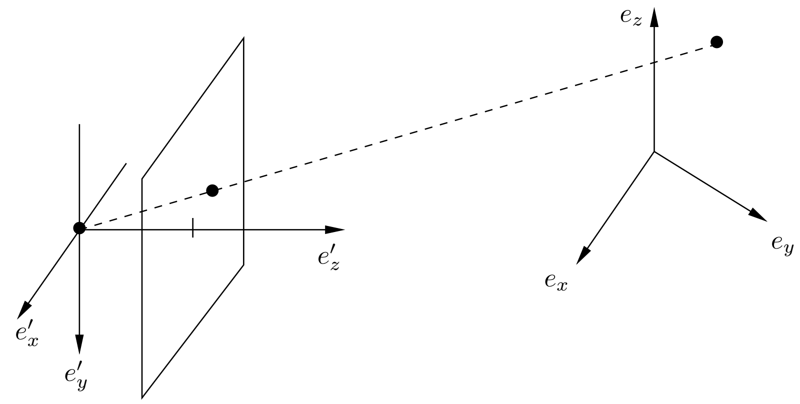

Figure 4: New global coordinate system.

If we add an extra 1 to the scene point X we can write (3) in matrix form

$$X'_3 \begin{pmatrix} X'_1/X'_3 \\ X'_2/X'_3 \\ 1 \end{pmatrix} = \begin{pmatrix} X'_1 \\ X'_2 \\ X'_3 \end{pmatrix} = [R \quad t] \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{pmatrix}. \tag{4}$$

Here $[R \quad t]$ is the 3×4 matrix where the first 3×3 block is R and the last column is t. As we saw previously, the projection of X is given by (2). Therefore we conclude from (4) that the projection x of a scene point X (with coordinates given in the global coordinate system) is obtained by computing the vector $v = [R \quad t] \begin{bmatrix} X \\ 1 \end{bmatrix}$ and dividing the elements of v by the third coordinate of v. From here on we will always assume that scene point coordinates are given in the global coordinate system, if nothing else is stated.

Exercise 3. Compute the projection of $X = (0, 0, 1)$ in the cameras

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & \sqrt{2} & 0 & 0 \\ 1 & 0 & 1 & \sqrt{2} \end{pmatrix}. \tag{5}$$

3 Depth of a Point

We say that a scene point is in front of the camera (or has positive depth) if its third coordinate is positive in the camera coordinate system. According to (4) the third coordinate is given by

$$X'_3 = [R_3 \quad t_3] \begin{bmatrix} X \\ 1 \end{bmatrix}, \tag{6}$$

where R_3 is the third row R and t_3 is the third coordinate of t. To determine whether a point is in front of the camera we therefore compute $v = [R \quad t] \begin{bmatrix} X \\ 1 \end{bmatrix}$ and check if the third coordinate of v is positive.

4 The Inner Parameters

In the pinhole camera model the image plane is embedded in $\mathbb{R}^3$. That is, image projections are given in the length unit of $\mathbb{R}^3$ (e.g. meters). Furthermore, the center of the image will be located in $(0, 0, 1)$, and will therefore have image coordinate $(0, 0)$. For real cameras we typically obtain images where the coordinates are measured in pixels with $(0, 0)$ in the upper left corner. To be able to do geometrically meaningful computations we need to transform pixel coordinates into the length unit of $\mathbb{R}^3$. We do this by adding a mapping from the image plane embedded in $\mathbb{R}^3$ to the real image, see Figure 5.

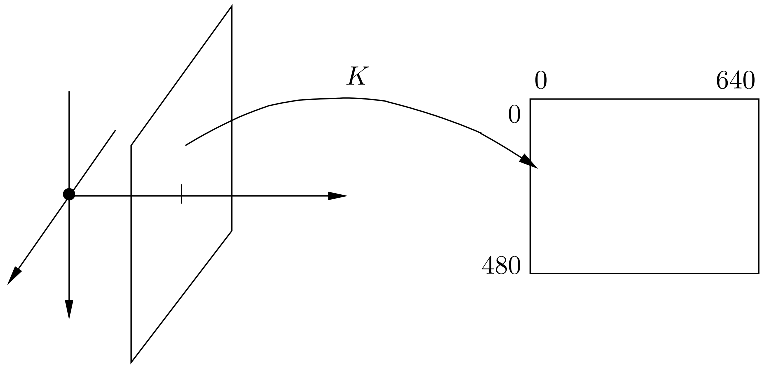

Figure 5: The mapping K from the image plane to the real image.

The mapping is represented by an invertible triangular 3×3 matrix K. This matrix contains what is usually referred to as the **inner parameters** of the camera, that is, focal length, principal point etc. (We will discuss this more in Lecture 3.) The projection (reference coordinate system to real image) is now given by

$$\lambda \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}}_{=\mathbf{x}} = \underbrace{K \begin{bmatrix} R & t \end{bmatrix}}_{=P} \underbrace{\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{pmatrix}}_{=\mathbf{X}}, \tag{7}$$

or in matrix form

$$\lambda \mathbf{x} = P\mathbf{X}. \tag{8}$$

Equation (8) is usually called the **camera equations** and the matrix P is called the **camera matrix**.

Exercise 4. Suppose that P is a 3×4 matrix that can be factorized into $P = K \begin{bmatrix} R & t \end{bmatrix}$ where K is upper triangular with positive diagonal elements and R is a rotation. How can we find the camera center and viewing direction from P without computing the factorization?