**FACULTY OF MATHEMATICS,**
**PHYSICS AND INFORMATICS**
Comenius University
Bratislava

3D Vision

# Lecture 2: Pinhole Camera Model
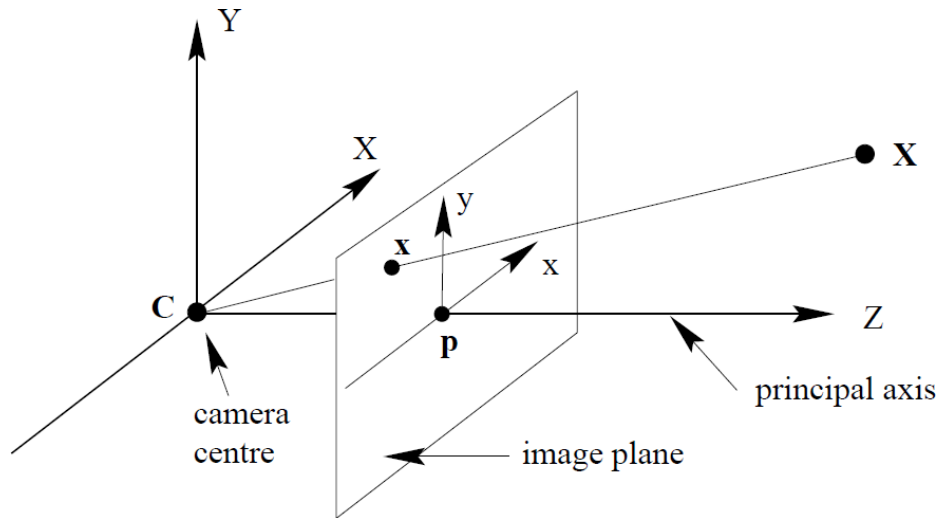
Ing. Viktor Kocur, PhD.

5.3.2024

# Contents

- Pinhole Camera Model
- Distortions
- Effect on Lines and Planes
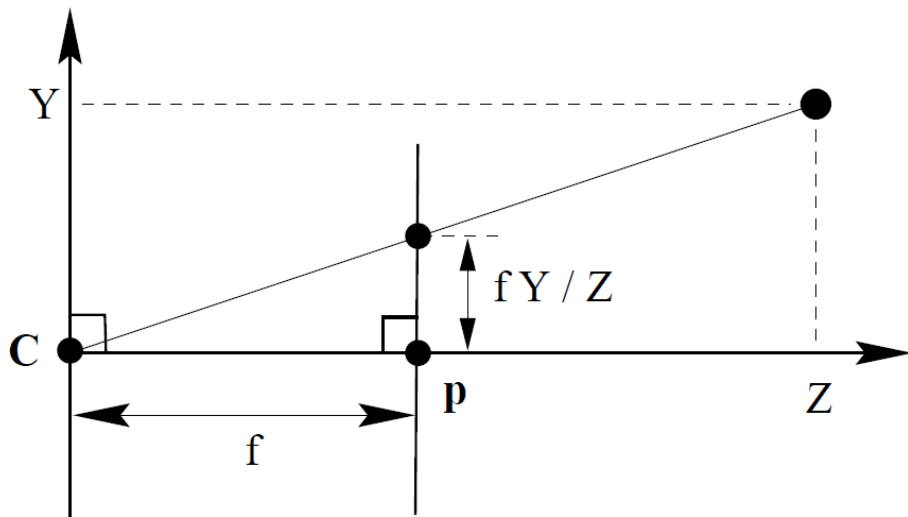
A point $\boldsymbol{X} = (X, Y, Z)^T$ in world coordinates is projected onto the imaging plane to obtain coordinates $\boldsymbol{x} = (x, y)^T$ using the following equation:

$$\boldsymbol{x} = \pi(\boldsymbol{X}) = \begin{pmatrix} f\frac{X}{Z} \\ f\frac{Y}{Z} \end{pmatrix}, \tag{1}$$

where $f$ is the focal length.

# Homogeneous Coordinates

Homogeneous coordinates can be used to represent points and vectors. In 2D, a point with homogeneous coordinates $(x, y, w)$ represents the point $(x/w, y/w)$ in Cartesian coordinates.

- Points with $w \neq 0$ represent finite points in the Euclidean plane. In this case, the coordinates $(x, y, w)$ are proportional to the Cartesian coordinates $(x/w, y/w)$ of the corresponding point in 2D space.

- Points with $w = 0$ represent points at infinity. In this case, the coordinates $(x, y, 0)$ represent the direction of the line that passes through the point at infinity.

- The triple $(0, 0, 0)$ is not valid in 2D homogeneous coordinates.

This also works in 3D coordinates where $(X, Y, Z, W)$ in homogeneous coordinates represents the point $(X/W, Y/W, Z/W)$ in Cartesian coordinates.

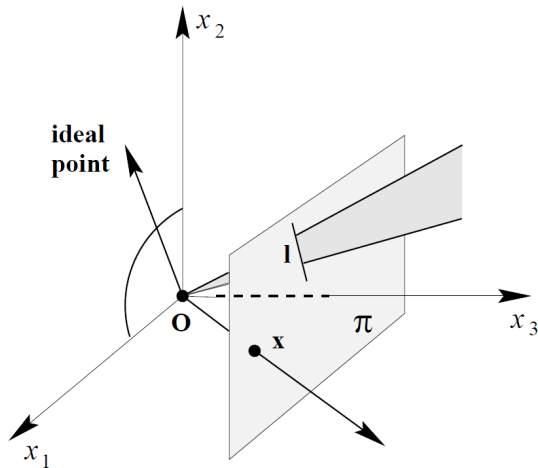Note that the homogeneous coordinates for points are not unique since $\boldsymbol{x}$ and $\alpha\boldsymbol{x}$ represent the same point if $\alpha \neq 0$. This creates an equivalence class defined by relation $\sim$:

$$\boldsymbol{p} \sim \boldsymbol{q} \iff (\exists\alpha \neq 0)(\alpha\boldsymbol{p} = \boldsymbol{q}). \tag{2}$$

Note that this sort of equivalence will also hold for most of the matrices we will work with as for $\boldsymbol{y} = A\boldsymbol{x}$ $\alpha\boldsymbol{y} = \alpha(A\boldsymbol{x}) = (\alpha A)\boldsymbol{x}$.

We can represent a 2D line by a vector $\boldsymbol{h} = (a, b, c)^T$ such that it represents the line given by the following equation:

$$ax + by + c = 0. \tag{3}$$

If we have a point in 2D expressed in homogeneous coordinates in a vector $\boldsymbol{x}$ and a line $\boldsymbol{h}$. Then the point lies on the line iff $\langle \boldsymbol{x}, \boldsymbol{h} \rangle = \boldsymbol{x}^T \boldsymbol{h} = 0$.

We can also find the intercept of two lines $\boldsymbol{h}$ and $\boldsymbol{l}$ as a point $\boldsymbol{y} = \boldsymbol{h} \times \boldsymbol{l} = [\boldsymbol{h}]_\times \boldsymbol{l}$. We can also find a line $\boldsymbol{k}$ connecting two points $\boldsymbol{p}$ and $\boldsymbol{q}$ in the same fashion as $\boldsymbol{k} = \boldsymbol{p} \times \boldsymbol{q} = [\boldsymbol{p}]_\times \boldsymbol{q}$. Note that same as with homogeneous coordinates, multiplying by non-zero scalar still gives the same line.

Similarly to 2D we can represent a plane in 3D space by a vector $\boldsymbol{\pi} = (a, b, c, d)^T$ such that it represents the plane given by the following equation:

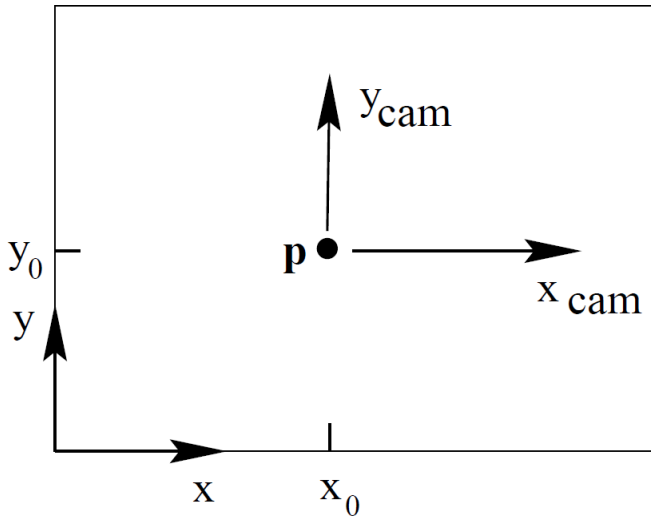$$aX + bY + cZ + d = 0. \tag{4}$$

Again if we multiply by a non-zero scalar we get the same plane. A 3D point $\boldsymbol{X}$ expressed in homogeneous coordinates lies on the plane if $\boldsymbol{\pi}^T \boldsymbol{X} = 0$. The vector $\boldsymbol{h} = (a, b, c)^T$ is in the direction of the plane normal and we can get the normal by calculating $\boldsymbol{n} = \frac{\boldsymbol{h}}{|\boldsymbol{h}|}$.

Using homogeneous coordinates we can express the simple pinhole camera model using the following equation:

$$\boldsymbol{x} \sim w \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = P\boldsymbol{X} = K_f \Pi_0 \boldsymbol{X} = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \qquad (5)$$

Image adopted from: Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003

12 / 29

To shift the coordinates we modify the equation to obtain:

$$P\boldsymbol{X} = K\Pi_0\boldsymbol{X} = \begin{pmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \tag{6}$$

where $(x_0, y_0)$ is the **principal point** (hlavný bod).

So far we have consider a coordinate system where the camera center is in the origin, the optical axis is the z axis and the remaining two axes are aligned with the image coordinates. However, we will often need to consider camera which has its own rotation $R$ and translation $\boldsymbol{t}$ w.r.t the world coordinate system. In such case we may rewrite the equation into the form:

$$P\boldsymbol{X} = K[R|\boldsymbol{t}]\boldsymbol{X}. \tag{7}$$

We then call the matrix $K$ the **intrinsic matrix** (*matica vnútorných parametrov*) or also the calibration matrix or the camera matrix and $[R|\boldsymbol{t}]$ the **extrinsic matrix** (*matica vonkajších parametrov*).

The camera center **C** is the right null-vector of $P$ ($P\mathbf{C} = \mathbf{0}$). This means that it is the only point which does not get projected to valid homogeneous coordinates.

Given $R$ and **t** the camera center can be calculated as:

$$\mathbf{C} = -R^T\mathbf{t}. \tag{8}$$

Note that $R$ and **t** denote the rotation of the world coordinate system w.r.t. the camera which is not the same as the position of the camera in the original coordinate frame. Therefore, usually $\mathbf{t} \neq \mathbf{C}$.

In a more general case $K$ can be slightly more complicated:

$$K = \begin{pmatrix} f_x & s & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad (9)$$

where $s$ is skew of the lens and $f_x$ and $f_y$ are different focal lengths for their corresponding directions. However, for most modern cameras we can assume that $f_x = f_y$, $s = 0$ and $(x_0, y_0)$ lies in the center of the image.

The pinhole camera model may not account for some distortions of the image arising from the geometry of the camera lens. Most common such distortion is the radial distortion. As a result of this distortion straight lines from the imaged scene are curved in the image. The rate at which the lines are distorted is approximately symmetric around the principal point. Thus it is possible to model this type of distortion as a function of the distance of a pixel $(x, y)$ from the principal point $(x_0, y_0)$:

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2}. \tag{10}$$

To correct for the distortion we can calculate the undistorted position of the pixels $(\hat{x}, \hat{y})$:
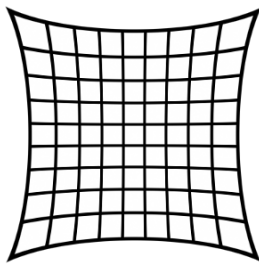
$$\hat{x} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \tag{11}$$

$$\hat{y} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \tag{12}$$

where $k_1, k_2, k_3$ are the parameters of the model. In some cases it is sufficient to assume that $k_3$ or even $k_2$ is zero. The mappings (11) and (12) can be used to transform the captured image to remove the effects of the distortion. Radial distortion is usually most pronounced in cheaper cameras or cameras with special lens.
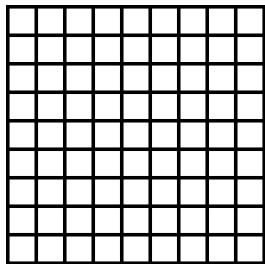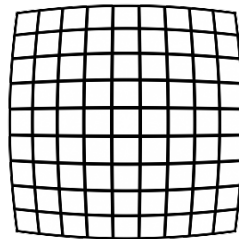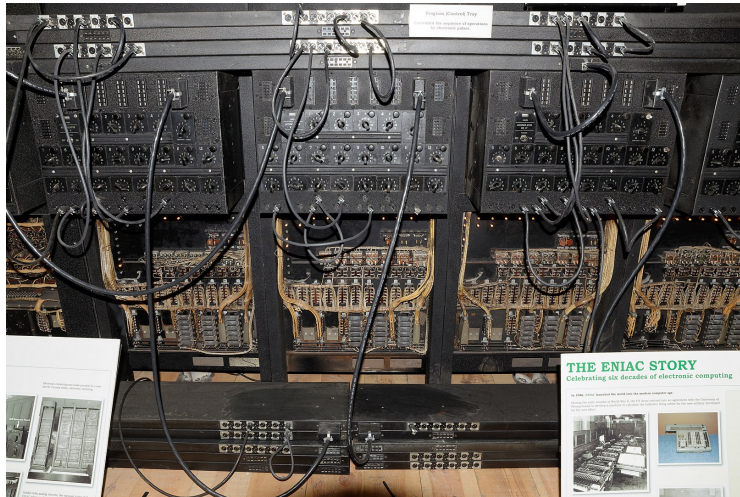
a

b

c
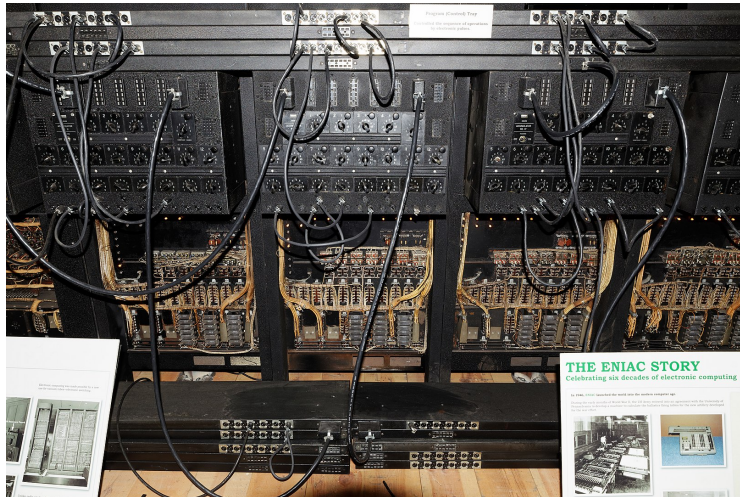
Image adopted from: Jud McCranie. *Wikimedia Commons*.

Image adopted from: Jud McCranie. *Wikimedia Commons*.

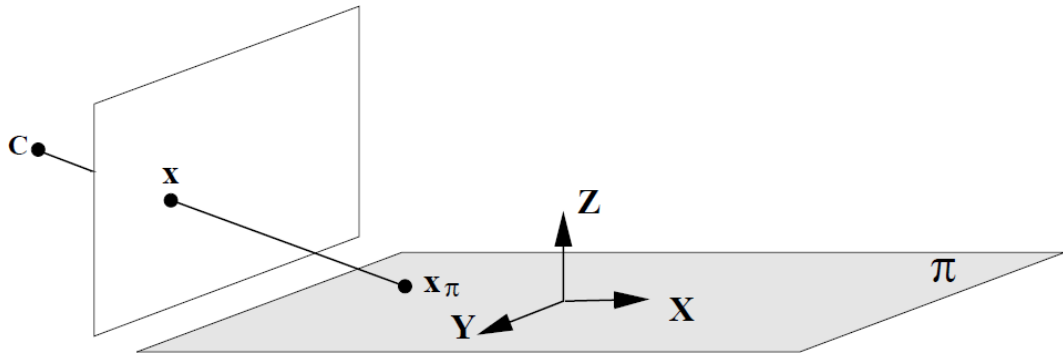We have discussed how points in the world are projected. Now we will consider how planes get projected onto an image. Suppose we have a plane in 3D space defined by $Z = 0$. All points on the plane then have coordinates $(X, Y, 0, 1)$. This leads to images in the form:

$$\boldsymbol{x} \sim P\boldsymbol{X} = K[R|\boldsymbol{t}] \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix}. \tag{13}$$

Equation (13) can be simplified to:

$$\boldsymbol{x} \sim K[\boldsymbol{r}_{:,1}, \boldsymbol{r}_{:,2}, \boldsymbol{t}] \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} = H \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}, \tag{14}$$

where $\boldsymbol{r}_{:,i}$ is the i-th column of the rotation matrix. We can observe that $H = K[\boldsymbol{r}_{:,1}, \boldsymbol{r}_{:,2}, \boldsymbol{t}]$ is a $3 \times 3$ matrix. If $[\boldsymbol{r}_{:,1}, \boldsymbol{r}_{:,2}, \boldsymbol{t}]$ is regular then also $H$ is also regular. This is also known as a **homography**. $H$ can be singular if the camera center lies on the plane. In that case $H$ is not a homography.

Homographies preserve straight lines. A line $\boldsymbol{h}$ is transformed to $\boldsymbol{h}' \sim H^{-T}\boldsymbol{h}$.
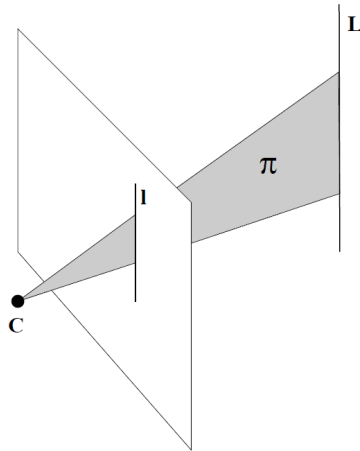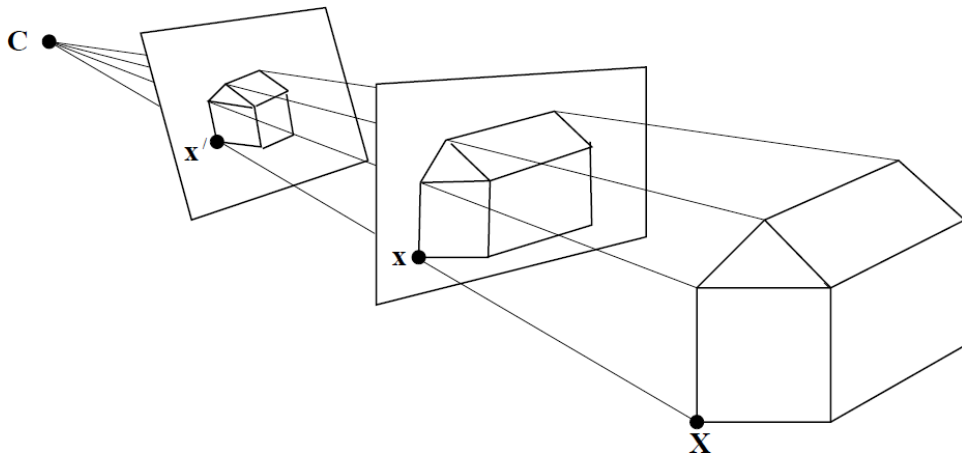
Image adopted from: Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003

Unless the camera center lies on a line, the line in the world is projected to a line in the image. Points imaged to a line $\boldsymbol{h}$ have to lie on a plane $\pi$ given by

$$\boldsymbol{\pi} = P^T \boldsymbol{h}. \tag{15}$$

Suppose our coordinate system is such that we have $\boldsymbol{t} = \boldsymbol{0}$ for two different cameras. $P = K[R|\boldsymbol{0}]$ and $P' = K'[R'|\boldsymbol{0}]$. We can then show a relationship between real point $\boldsymbol{X}$ and its images in $\boldsymbol{x}$ and $\boldsymbol{x}'$ in the two cameras:

$$\boldsymbol{x}' \sim P'\boldsymbol{X} \sim (K'R')(KR)^{-1}P\boldsymbol{X} \sim (K'R')(KR)^{-1}\boldsymbol{x} = H\boldsymbol{x}. \tag{16}$$

Note that the resulting matrix is a homography. This result can be useful for creating synthetic views with different camera paramaters from the same camera center. Note that this also means that it is not possible to recover the scene structure from two cameras with the same center.