Chapter 3: Perspective Projection

Konrad Koniarski

2016/07/16

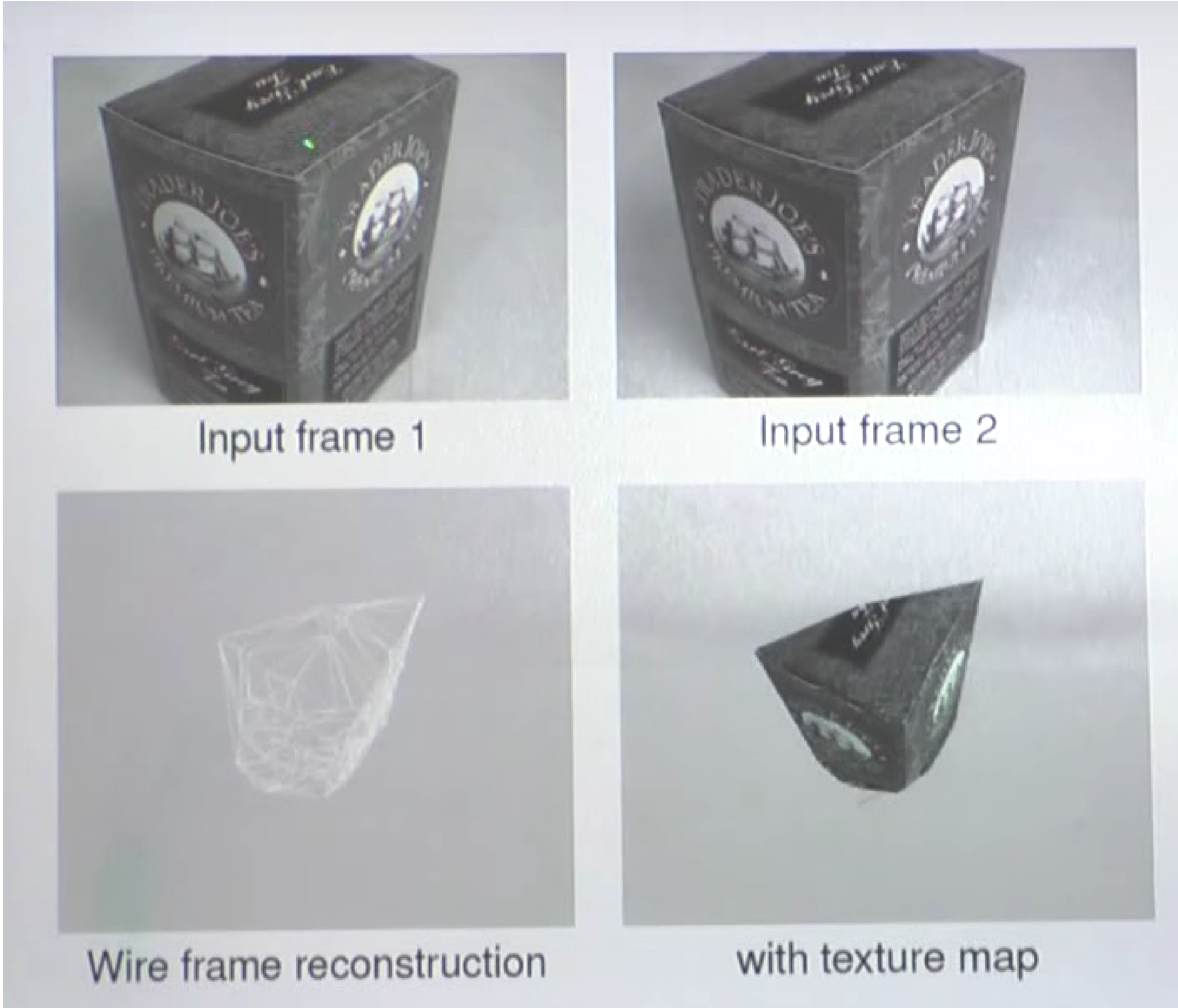
1 Historic Remarks

1.1 Some Historic Remarks

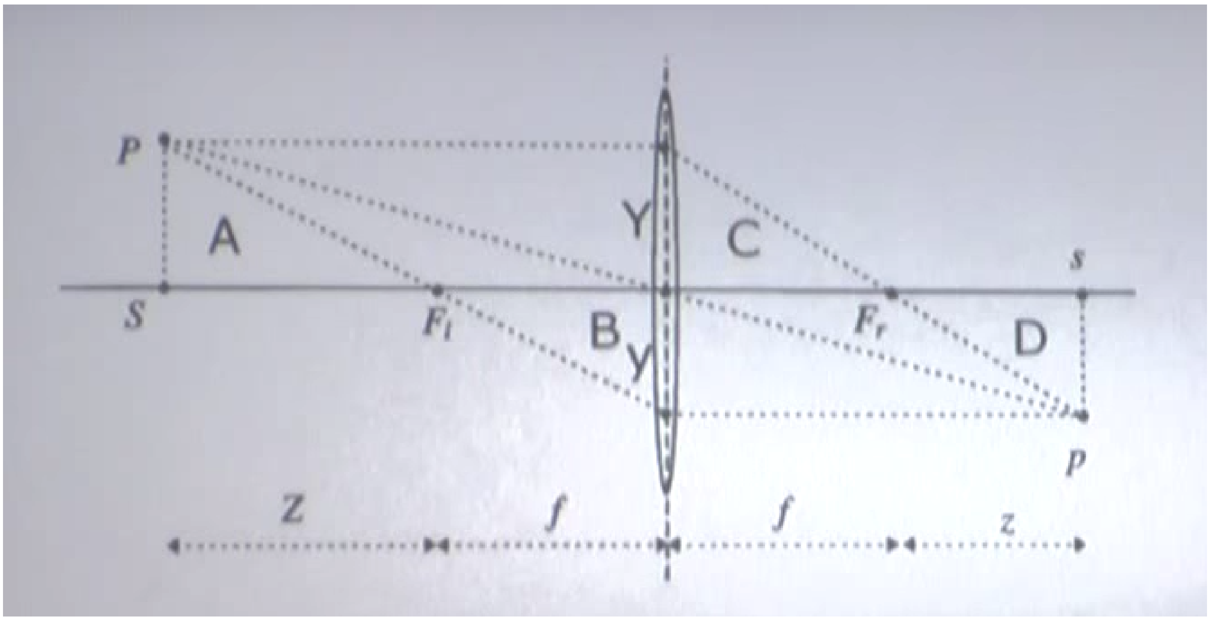
The study of the image formation process has a long history. The earliest formulations of the geometry of image formation can be traced back to Euclid (4th century B.C.). Examples of a partially correct perspective projection are visible in frescoes and mosiaics of Pompeii (1 B.C.). The skills seen to have been lost with the fall of the Roman empire. Correct perspective projection emerged again around 1000 years later in early Renaissance art. Among the proponents of perspective projection are the Renaissance artists Brunellesch, Donatello and Alberti. The first treatise on the projection process, "Della Pittura" was published by Leon Battista Alberti). Apart from the geometry of image formation, the study of the interaction of light with matter was propagated by artists liek Leonardo da Vini in the 1500s and by Renaissance painters such as Caravaggio and Raphael.

2 Mathematical Representation

2.1 Mathematics of Perspective Projection

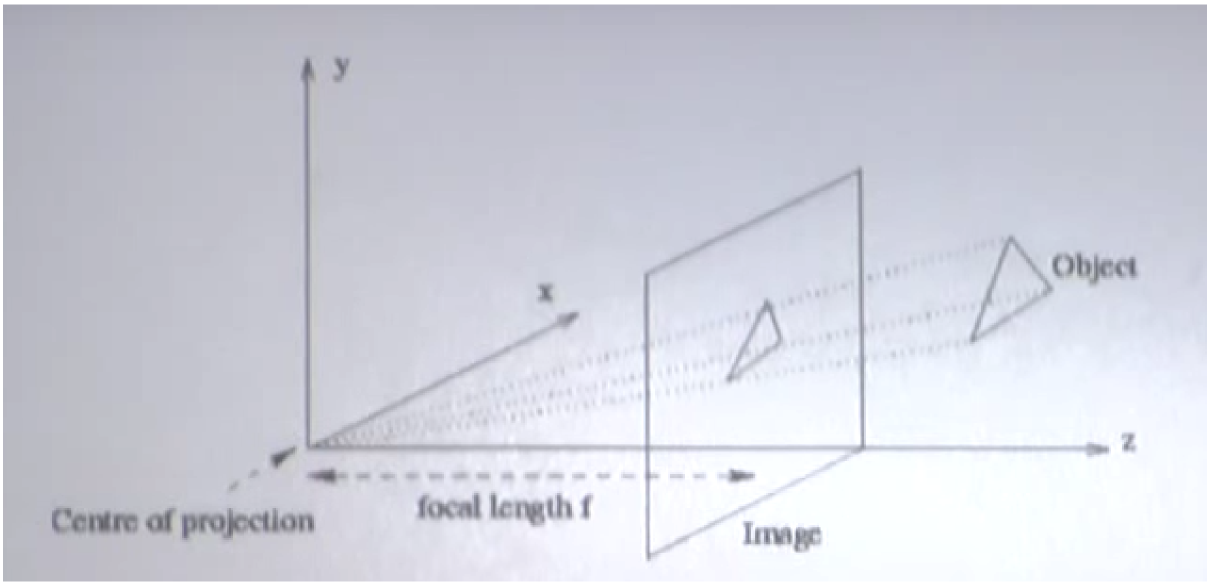


The perspective tranformation π from a point with coordinates $X = (X, Y, Z) \in \mathbb{R}^3$ relative to the reference frame centered at the optical center and with x-axis being the optical axis (of the lens) is obtained by comparing similar triangles A and B:



$$\frac{Y}{Z} = -\frac{y}{f} \Leftrightarrow y = -f\frac{Y}{Z} \tag{1}$$

2.2 Mathematic of Perspective Projection



To simplify equations, one flips the sign of x- and y-axes, which amounts to considering the image plane to be in front of the center of projection (rather the behind id). The perspective transformation π is therefore given by

$$\pi : \mathbb{R}^2 \to \mathbb{R}^2; X \to x = \pi(X) = \begin{pmatrix} f\frac{X}{Z} \\ f\frac{Y}{Z} \end{pmatrix} \tag{2}$$

2.3 An Ideal Perspective Camera 1

In homogeneous coordinates, the perspective transformation is given by:

$$Zx = Z \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = K_f \Pi_0 X \tag{3}$$

when we have introduced the two matrices

$$K_t \equiv \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \Pi_0 \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{4}$$

The matrix Π_0 in referred to as the **standard projection matrix**. Assuming Z to be a constant $\lambda > 0$, we obtain:

$$\lambda x = K_f \Pi_0 X \tag{5}$$

2.4 An Ideal Perspective Camera

From the previous lectures, we know that due to the **rigid motion of the camera**, the X **in camera coordinates** is given as a function of the point in **world coordinates** x_0 by:

$$X = RX_0 + T \tag{6}$$

or in homogeneous coordinates $X = (X, Y, Z, 1)^T$:

$$X = gX_0 = \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix} x_0 \tag{7}$$

In total, the transformation from world coordinates to image coordinates is therefore given by

$$\lambda x = K_t \Pi_0 g X_0 \tag{8}$$

In the focal length f is known, it can be normalized to 1 (by changing the units of the image coordinates), such that:

$$\boxed{\lambda x = \Pi_0 X = \Pi_0 g X_0} \tag{9}$$

3 Intristic Parameters

3.1 Intrinsic Camera Parameters

If the camera is not centered at the optical center, we have an additional translation o_x, o_y and if pixel coordinates do not have unit scale, we need to introduce an additional scaling in x- and y-direction by s_x and s_y. If the pixels are not rectangular, we have a **skew factor** s_θ.

The pixel coordinates $(x', y', 1)$ as a function of homogeneous camera coordiantes X are the given by:

$$\lambda \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} s_x & s_\theta & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{pmatrix}}_{\equiv K_s} \underbrace{\begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\equiv K_f} \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\equiv \Pi_0} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{10}$$

After the perspective projection Π_0 (with focal length 1), we have an additional transformation which depends on the (intrinsic) camera parameters. **This can be expressed by intristic parameters matrix** $K = K_s K_f$.

3.2 The Intrinsic Parameter Matrix

All intrinsic camera parameters therefore enter the **intrinsic parameter** matrix

$$K \equiv K_s K_f \equiv \begin{pmatrix} fs_x & fs_\theta & 0_x \\ 0 & fs_y & o_y \\ 0 & 0 & 1 \end{pmatrix} \tag{11}$$

As a function of the world coordinates X_0, we therefore have:

$$\boxed{\lambda x' = K\Pi_0 X = K\Pi_0 g X_0 \equiv \Pi X_0} \tag{12}$$

The 3×4 matrix $\Pi \equiv K\Pi_0 g = (KR, KT)$ is called a **general projection matrix**. Although the above equation looks like a linear one, we still have the scale parameter λ. Dividing by λ gives:

$$x' = \frac{\pi_1^T X_0}{\pi_3^T X_0}, y' = \frac{\pi_2^T X_0}{\pi_3^T X_0}, z' = 1 \tag{13}$$

where π_1^T, π_2^T, $\pi_3^T \in \mathbb{R}^4$ are the three rows of the projection matrix Π.

3.3 The Intrinsic Parameter Matrix

The entries of the intrinsic parameter matrix

$$K = \begin{pmatrix} fs_x & fs_\theta & o_x \\ 0 & fs_y & o_y \\ 0 & 0 & 1 \end{pmatrix} \tag{14}$$

can be interpreted as follows:

o_x	x-coordinate of principal point in pixels
o_y	y-coordinate of principal point in pixels
$fs_x = \alpha_x$	size of unit length in horizontal pixels
$fs_y = \alpha_y$	size of unit length in vertical pixels
α_x/α_y	aspect ratio σ
fs_θ	skew of the pixel, ofter close to zero.

4 Spherical Projection

4.1 Spherical Perspective Projection

The parspective pinhole camera introduced above considers a planar imaging surface. Instead, one can consider a spherical projection sufrace given by the unit sphere $\mathbb{S}^2 \equiv \{x \in \mathbb{R}^3; |x| = 1\}$. The **spherical projection** π_s of a 3D point X is given by:

$$\pi_s : \mathbb{R}^3 \to \mathbb{S}^2; X \to x = \frac{X}{|X|} \tag{15}$$

The pixel coordinates x' as a function of the world coordinates X_0 are:

$$\lambda x' = K\Pi_0 g X_0 \tag{16}$$

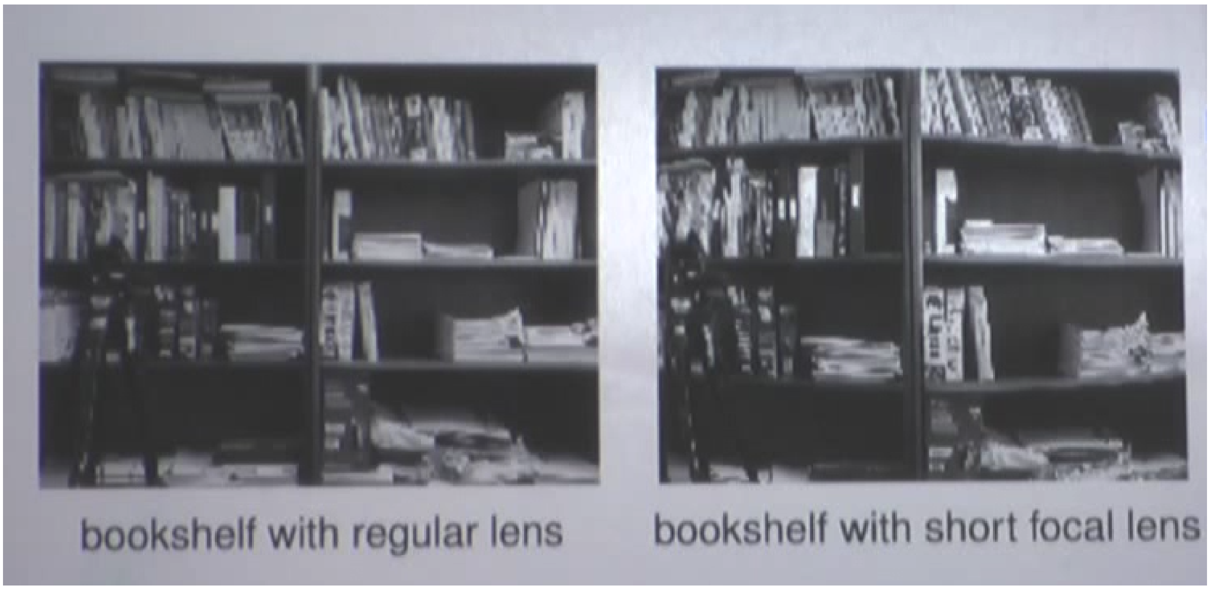
except that the scalar factor is now $\lambda = |X| = \sqrt{X^2 + Y^2 + Z^2}$. One often writes $x \sim y$ for homogeneous vector x and y if they are equal up to a scalar factor. Then we can write:

$$x' \sim \Pi X_0 = K\Pi_0 g X_0 \tag{17}$$

This property holds for any imaging surface, as long as they ray between X and the origin intersects the imagine surface.

5 Radial Distortion

5.1 Radial Distortion



bookshelf with regular lens bookshelf with shor focal lens

5.2 Radial Distortion

The intrinsic parameters in the matrix K model linear distortions in the transformation to pixel coordinates. In practice, however, one can also encounter significant **distorions along the radial axis**, in particular if a wide field of view is used or if one uses cheaper cameras such as webcam. A simple effective model for such distorions is

$$x = s_d(1 + a_1 r^2 + a_2 r^4), y = y_d(1 + a_1 r^2 + a_x r^4) \tag{18}$$

where $x_d \equiv (x_d, y_d)$ is the distorted point, $r_2 = x_d^2 + y_d^2$. If a calibration rig is avaliable, the distorion parameters a_1 and a_2 can be estimated. Alternatively, one can estimate a distorion model directly from images. A more general model (Devernay and Faugeras 1995) is

$$x = c + f(r)(x_d - c), \text{ with } f(r) = 1 + a_1 r + a_2 r^2 + a_3 r^3 + a_4 r^4 \tag{19}$$

here, $r = |x - c|$ is the distance to an arbitraty center of distortion c and the distorion to an arbitrary center of distortion c and the **distorion correction factor f(r)** is an arbitraty 4-th order expression. Parameters are computed **from distortions of strainght lines** of **simultaneously with the 3D reconstruction** (Zhang '96, Stein' 97, Fitzgibbon 0'1).

6 Preimage and Coimage

6.1 Preimage of Points and Lines

The perspective transformation introduced above allows to define images for arbitrary geometric entities by simply transforming all points of the entity. However, due to the unknows scale factor, each point is mapped not to a single point x, but to an **equivalence class of points** $y \sim x$. It is therefore usefull to **study how lines are transformed**. A line L in 3D is characterized by a base point $X_0 = (X_0, Y_0, Z_0, 1) \in \mathbb{R}^4$ and a vector $V = (V_1, V_2, V_3, 0) \in \mathbb{R}^4$:

$$X = X_0 + \mu V, \mu \in \mathbb{R} \tag{20}$$

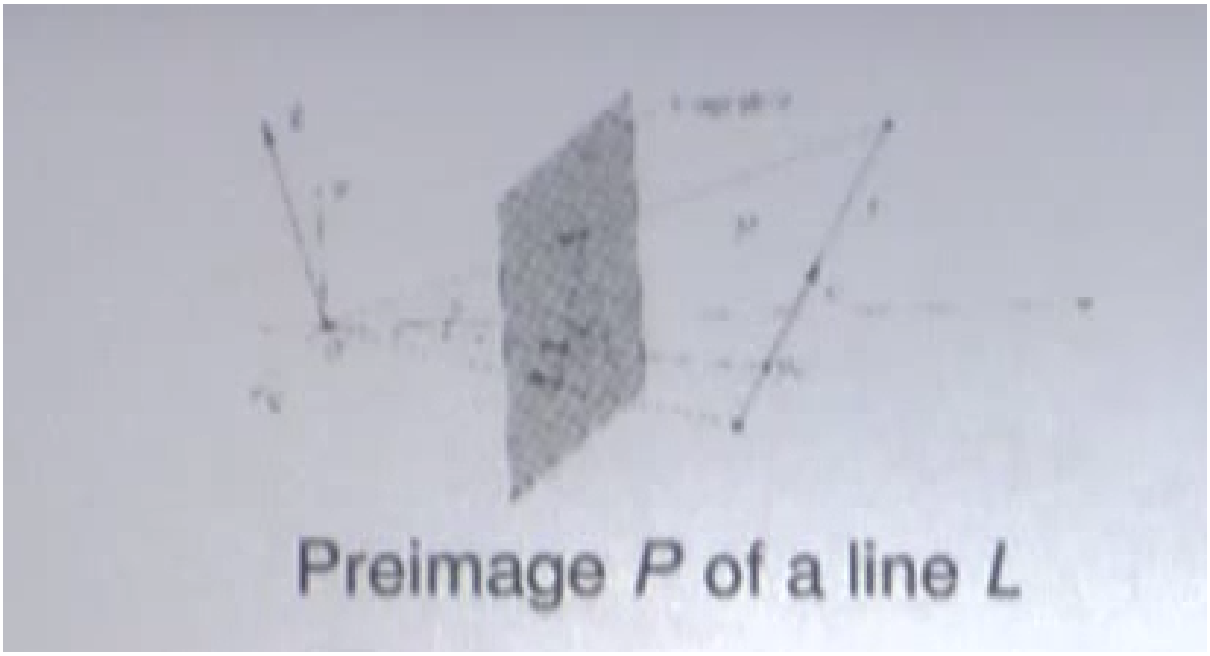
The image of the line L is given by

$$x \sim \Pi_0 X = \Pi_0(X_0 + \mu V) = \Pi_0 X_0 + \mu \Pi_0 V \tag{21}$$

All points x treated as vectors from the origin o span a 2D subspace P. The intersection of this plane P with the image plane give the image of the line. P is called the preimage of the line.

A **preimage of a point or a line** in the image plane is the largest set of 3D points that give raise to an image equal to the given point or line.

6.2 Preimage and Coimage



Preimages can be defined for curves or other more complicated geometric structures. In the case of points and lines, however, the preimage is a subspace of $\mathbb{R}^3$. This subspace can also be represented by its orthogonal complements, i.e. the normal vector in the case of a plane. This complement is called the coimage. The **coimage of a point or a line** is the subspace $\mathbb{R}^3$ that is the (unique) ortogonal complement of its preimage. Image, preimage and coimage are **equivalent** because they uniquely determine oneanother:

$$\begin{array}{ll} \text{image} = \text{preimage} \cap \text{image plane}, & \text{preimage} = \text{span(image)} \\ \text{preimage} = \text{coimage}^{\perp}, & \text{coimage} = \text{preimage}^{\perp} \end{array}$$

6.3 Preimage and Coimage of Points and Lines

In the case of the line L, the preimage is a 2D subspace, characterized by the 1D coimage given by the span of its normal vector $\ell \in \mathbb{R}^3$. All points of the preimage, and hance all points x of the image of L are orthogonal to ℓ:

$$\ell^T x = 0 \tag{22}$$

The space of all vector orthogonal to ℓ is spanned by the row vector of $\hat{\ell}$, thus we have:

$$P = span(\hat{\ell}) \tag{23}$$

In the case that x is the image of a point p, the preimage is a line and the coimage is the plane orthogonal to x, i.e. it is spanned by the rows of the matrix $\hat{x}$. In summary we have the following table:

	Image	Preimage	Coimage
Point	span(x) $\cap$ image plane	span(x) $\subset \mathbb{R}^3$	span($\hat{x}$) $\subset \mathbb{R}^3$
Line	span($\hat{\ell}$) $\cap$ image plane	span($\hat{\ell}$) $\subset \mathbb{R}^3$	span(ℓ) $\subset \mathbb{R}^3$

6.4 Summary

In this part of the lecture, we studied the **perspective projection** which takes us from the 3D (4D) camera coordinates to 2D camera image coordinates and pixel coordinates. In homogeneous coordinates and pixel coordinates. In homogeneous coordinates, we have the transformations:

$$\text{4D world coordinates} \overset{g \in SE(3)}{\rightarrow} \text{4D Camera coordinates}$$

$$\text{3D image coordinates} \overset{K_s}{\rightarrow} \text{3D pixel coordinates.}$$

In particular, we can summarize the **(intrinsic) camera parameters** in the matrix:

$$K = K_s K_f \tag{24}$$

The full transformation from world coordinates X_0 to pixel coordinates x' is given by:

$$\lambda x' = K \Pi_0 g X_0 \tag{25}$$

Moreover, for the images of points and lines we introduced the notions of **preimage** (maximal points set which is consistent with a given image) and **coimage** (it orthogonal complement). Both can be used equivalently to the image.

7 Projective Geometry

7.1 Projective Geometry

In order to formally write transformations by linear operations, we made extensive use of **homogeneous coordinates** to represent a 3D point as a 4D-vector **(X, Y ,Z, 1)** with the last coordinate fixed to 1. This normalization is not always necessary: One can represent 3D points by a general 4D vector

$$X = (XW, YW, ZW, W) \in \mathbb{R}^4 \tag{26}$$

remembering that merely the direction of this vector is of importance. **We therefore identify the point in homogeneous coordinates with the line connecting it with the origin**. This leads to the definition of projective coordinates. **An n-dimensional projective space** $\mathbb{P}^n$ is the set of all one-dimensional subspace (i.e. lines through the origin) of the vector space $\mathbb{R}^{n+1}$. A point $p \in \mathbb{P}^n$ can then be assigned homogeneous coordinates $X = (x_1, ..., x_{n+1})^T$, among which at least one x is nonzero. For any nonzero $\lambda \in \mathbb{R}$, the coordinates $Y = (\lambda x_1, ..., \lambda x_{n+1})^T$ represent the same point p.