# Image Acquisition

## Geometric Projections

Technical University of Applied Sciences Würzburg-Schweinfurt

**2D and 3D Geometry**

- Euclidean Space
- Homogeneous Coordinates
- Line, Plane, Curve

**Geometric Projections**

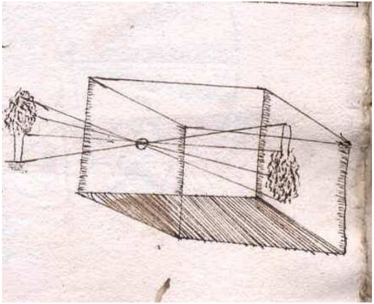
- Central projection
- Intrinsics & Extrinsics
- Projections of Lines & Planes
- Camera Calibration I

**Optics: The Lens**

- Characteristic Values
- Thin Lense
- Imaging Errors
- Camera Calibration II

# Geometric Projections

## Motivation



*Camera obscura*: Pen and ink drawing in the margin of a lecture manuscript on the *Principia Optices* from the 17th century.

www.wikipedia.org

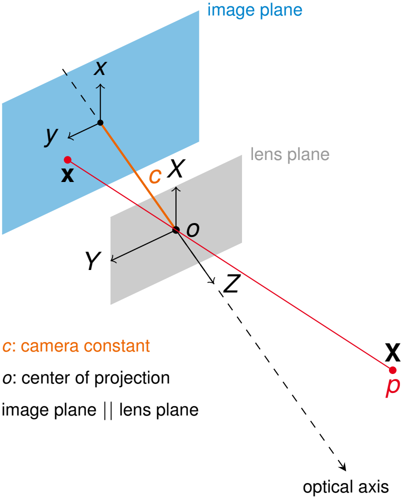
Image processing attempts to interpret images, i.e. to draw conclusions about the scene from the image. This inverse problem can be solved only if one has an understanding of how the image is formed and what information of the scene is preserved. From a purely geometric point of view, capturing an image is a projection of the scene onto the camera's sensor. In this process, 3D world points are mapped onto 2D image points. This mapping is reproduced by a suitable camera model.

# Geometric Projections

## Central projection - Pinhole model

The simplest camera model is the ideal pinhole camera model. All connecting lines (light rays) between points $p$ in space with camera coordinates $\mathbf{X} = [X, Y, Z]^\top$ and the image plane with the image coordinates $\mathbf{x} = [x, y]^\top$ pass through a point in the lens plane, the projection center $o$ (the pinhole). From the ray theorem we get the following relation between camera and image coordinates of the point:

$$x = -c\frac{X}{Z}\,, \quad y = -c\frac{Y}{Z}\ .$$




$c$: camera constant

$o$: center of projection

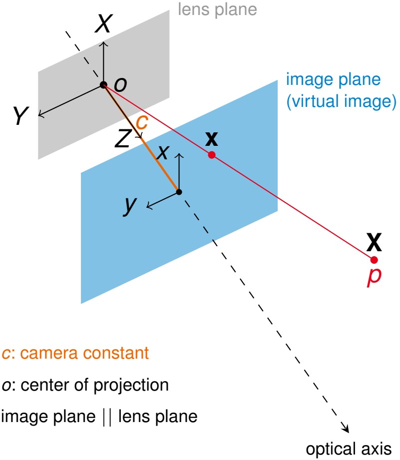
image plane || lens plane

# Geometric Projections

## Central projection - Virtual image

Since $-c/Z$ is negative, every object is imaged upside down on the image plane. This can be avoided by either setting the image coordinates negative, i.e. turning the image around, or placing the image plane in front of the projection center. If the image plane is placed in front of the projection center $o$, the following results for the image $\mathbf{x} = [x, y]^\top$ of the point $p$

$$x = c\frac{X}{Z}, \quad y = c\frac{Y}{Z} .$$




$c$: camera constant

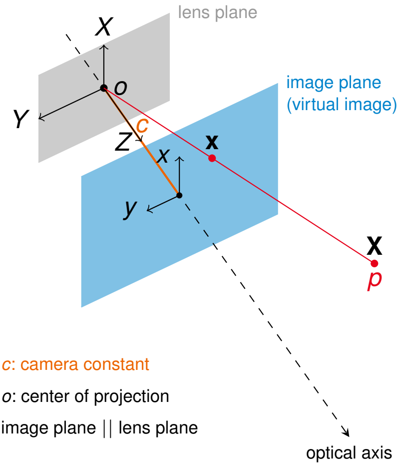
$o$: center of projection

image plane || lens plane

# Geometric Projections

## Central projection - Pinhole camera

Due to the central projection, the image coordinates no longer contain absolute information about the camera coordinates. They result from the multiplication of $X$ and $Y$ with the scaling factor $c/Z$. Thus the distance of an object, as well as the true size over $x$ and $y$ is no longer determinable.
In addition to the depth information, the information about the angles in the scene is also lost, since the scene is imaged differently depending on the depending on the angle of view of the camera.




$c$: camera constant

$o$: center of projection

image plane || lens plane

# Geometric Projections

## Central projection - Loss of information

During projection

- of a 3D point the information about the distance to the projection center is lost,
- of a 3D straight line the information about the distance to the projection center and the orientation in the 2D subspace is lost. (which is spanned by the straight line and the projection center) is lost,
- of a 3D plane the information about the distance to the projection center and the orientation in the 3D space is lost.

... construct a sketch ...

# Geometric Projections

## Central projection - Perspective effects

Technical University
of Applied Sciences
Würzburg-Schweinfurt

The following conclusions can be drawn from the projection equations:

- Objects of the same size appear smaller the further away they are.
- As the distance increases, an object is resolved less and less accurately at constant image resolution. If the object size and the distance are changed by the same factor $\alpha$, the projected size remains the same:

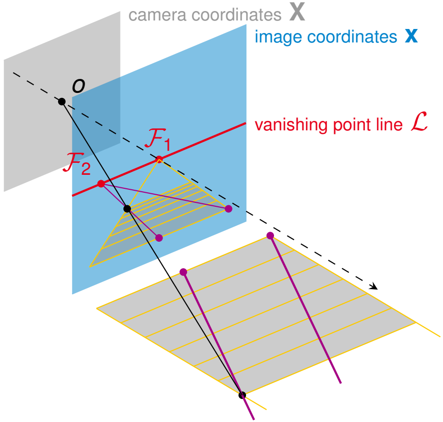
$$x = c\frac{X}{Z} = c\frac{\alpha X}{\alpha Z} \ .$$

- The change of the projected size $\partial x$, $\partial y$ due to a change of depth $\partial Z$ is inversely proportional to the square of the absolute depth $\propto \frac{1}{Z^2}$:

$$\partial x = \frac{cX}{Z^2}\partial Z \ , \ \partial y = \frac{cY}{Z^2}\partial Z \ .$$

## Central projection - Geometric properties

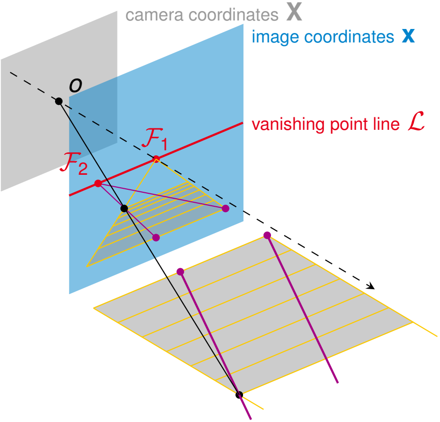
- Parallel lines converge in a fluctuation point $\mathcal{F}$. Different orientations result in different vanishing points. All vanishing points of lines in a plane lie on a vanishing point line $\mathcal{L}$.
- The image horizon $\mathcal{H}$ results as the intersection line between the image plane and a plane perpendicular to the image plane and contains the center of projection.



# Geometric Projections

## Central projection - Geometric properties

- Points become points.
- Lines become lines, except when they pass through the projection center, then they become points.
- planes become the whole image or half planes, except when they include the projection center, then they become lines.
- Polygons with *n* edges become polygons with *n* edges, unless the vertices of the polygon are in a plane containing the center of projection.

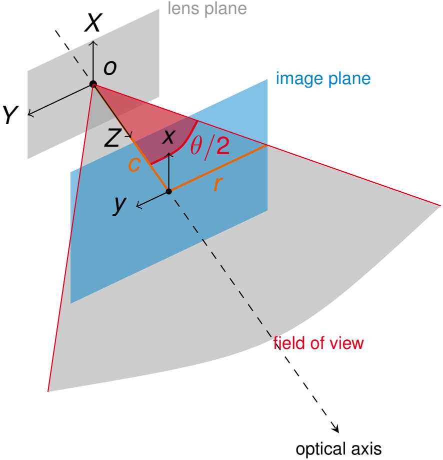


## Central projection - Field of view

In practice, the image plane is spatially limited, which means that not every point in space can be projected onto the image plane. This results in a restricted field of view *field of view* (FOV), which is defined by the angle $\theta$. This angle results from the camera constant *c* and the maximum extension *r* of the image plane at

$$\theta = 2\arctan\left(\frac{r}{c}\right) \; .$$

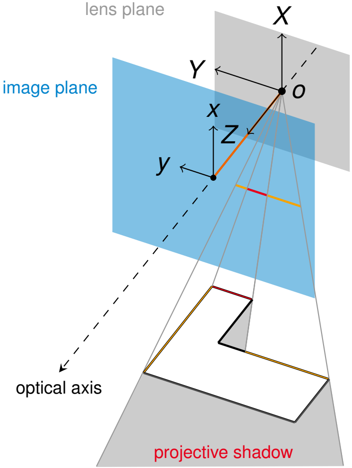
For a flat image plane, this angle is always less than $\theta < 180^\circ$.



# Geometric Projections

## Central projection - Occlusion

All points in space that lie on a ray are mapped to a point in the image plane. As a result, the entire area of space that lies behind an opaque object cannot be seen. Depending on the position and the viewing direction of the camera and the shape of the surface of an object, different areas of this object are imaged on the image plane. However, some areas of the object always remain hidden because they lie in the projection shadow.



# Geometric Projections

## Central projection - Approximations

If all points of a scene lie approximately on a plane parallel to the image plane and the distances of the scene points to this plane are small compared to the distance of the plane $Z_e$ from the projection center $o$, then it can be assumed that the distance of the scene points $Z$ is constant and corresponds to the distance of the plane $Z_e$. This gives the projection equations of the weak perspective projection:

$$x \approx c\frac{X}{Z_e} \propto X \,, \quad y \approx c\frac{Y}{Z_e} \propto Y \ .$$

If the points parallel to the optical axis are projected orthogonally onto the image plane, then one speaks of an orthographic projection and obtains the simplest of all projection equations:

$$x = X \,, \quad y = Y \ .$$

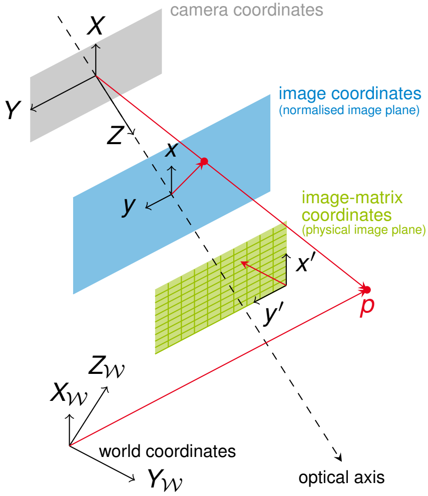
# Geometric Projections

## Mapping process - Mathematical description

The mathematical model of the entire mapping process of a point in space onto the surface of an image matrix can be described with three successive transformations:

- A transformation between camera and world coordinates.
- A central projection from 3D space to 2D image coordinates.
- A transformation between image-matrix and image coordinates.

From now on, there is no distinction between homogeneous and non-homogeneous coordinates when it is clear from the context.



# Geometric Projections

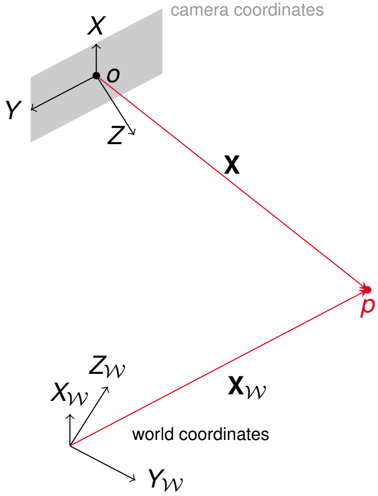
## Mapping process - Mathematical description

Affine transformation from world to camera coordinates:

$$\mathbf{X} = \mathbf{R}\mathbf{X}_{\mathcal{W}} + \mathbf{T} \quad \in \mathbb{R}^3 .$$

Linear transformation from homogeneous world to camera coordinates:

$$\underbrace{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} X_{\mathcal{W}} \\ Y_{\mathcal{W}} \\ Z_{\mathcal{W}} \\ 1 \end{bmatrix}}_{\mathbf{X}_{\mathcal{W}}} \quad \in \mathbb{R}^4 .$$



# Geometric Projections

## Mapping process - Mathematical description

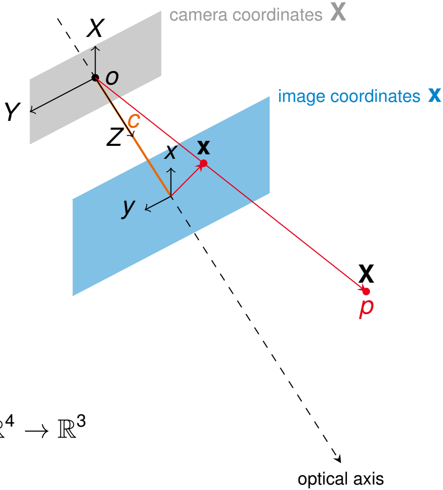
Nonlinear central projection from 3D spatial to 2D image coordinates:

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \frac{c}{Z} \begin{bmatrix} X \\ Y \end{bmatrix} .$$

Equals homogeneous transformation:

$$Z \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$Z \quad \mathbf{x} = \begin{bmatrix} c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \mathbf{X} \quad : \mathbb{R}^4 \rightarrow \mathbb{R}^3$$



# Geometric Projections

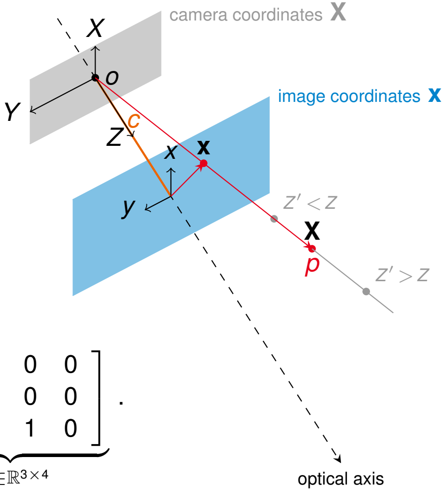
## Mapping process - Mathematical description

The depth $Z$ can be described by any positive scaling factor $\lambda \in \mathbb{R}_+$:

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}.$$

The following decomposition makes sense:

$$\begin{bmatrix} c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \underbrace{\begin{bmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}_c \in \mathbb{R}^{3\times 3}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\mathbf{\Pi}_0 \in \mathbb{R}^{3\times 4}}.$$



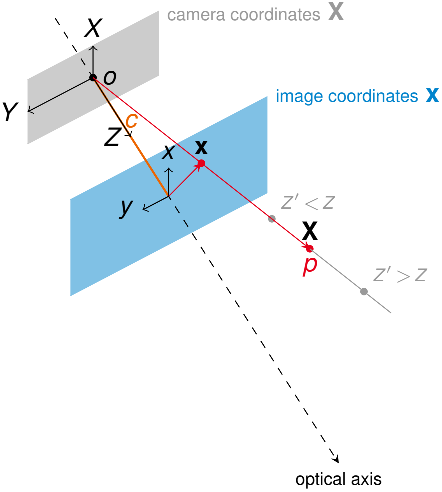
# Geometric Projections

## Mapping process - Mathematical description

From previous equation follows that to each object point **X** belongs an image point. To each image point **x** belongs however an infinite number of object points. Therefore, from an image a spatial object cannot be reconstructed without additional knowledge about the distance.

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \frac{c}{Z} \begin{bmatrix} X \\ Y \end{bmatrix},$$

$$\mathbf{X} = \begin{bmatrix} X \\ Y \end{bmatrix} = \frac{Z}{c} \begin{bmatrix} x \\ y \end{bmatrix}.$$



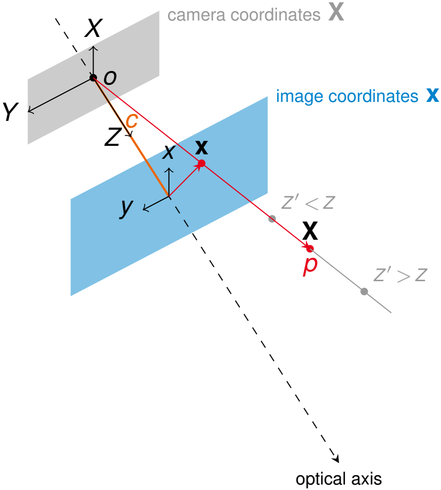
# Geometric Projections

## Mapping process - Mathematical description

One obtains a compact representation of the central projection as a transformation from homogeneous 3D space to 2D image coordinates:

$$\lambda \mathbf{x} = \mathbf{K}_c \mathbf{\Pi}_0 \mathbf{X} \quad ,$$

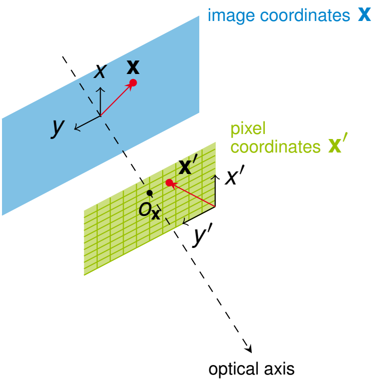
using the canonical projection matrix $\mathbf{\Pi}_0$. This model is valid only if the lens plane is parallel to the image plane, the origin of the camera coordinate system lies in the projection center, and the Z coordinate lies on the optical axis.



## Mapping process - Mathematical description

A transformation from image to image matrix coordinates is necessary for two reasons:

- The origin of the coordinate system of the image matrix lies in the upper left corner of the matrix and mostly does not coincide with the principal point $o_{\mathbf{x}} = \left[o_x, o_y\right]^\top$.
- The unit of the coordinate axes is typically given in millimeters, but the coordinate axes do not have to be perpendicular to each other and the shape of a pixel does not have to be square.



# Geometric Projections

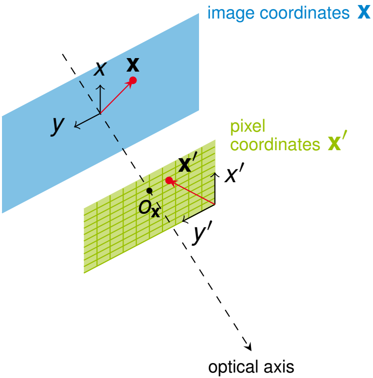
## Mapping process - Mathematical description

Three transformations can be used to map these properties:

- A scaling with the factors $s_{\mathbf{x}} = \left[s_x, s_y\right]^\top$.
- A shift around the coordinates of the principal point $o_{\mathbf{x}} = \left[o_x, o_y\right]^\top$.
- A shear by the skew factor $s_\theta$.

The result is the homogeneous representation:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} s_x & s_\theta & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}_s \in \mathbb{R}^{3\times 3}} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix},$$



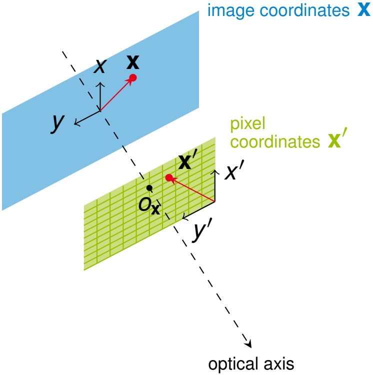
# Geometric Projections

## Mapping process - Mathematical description

respectively

$$\mathbf{x}' = \mathbf{K}_s \mathbf{x} \, .$$

In many practical applications the shear $s_\theta = 0$ is neglected, because the angle $\theta$ between the image matrix axes is almost equal to $90^\circ$. If the pixel shape is square, then the scaling factors $s_x = s_y$ are identical. The pixel size is equal to $1/s_x \, [mm] \times 1/s_y \, [mm]$ respectively there are $s_x \times s_y$ pixels per square millimeter.



# Geometric Projections

## Mapping process - Mathematical description

If one summarizes all three transformations

- $\mathbf{X} = \mathbf{G}\mathbf{X}_{\mathcal{W}}$ ,
- $\lambda\mathbf{x} = \mathbf{K}_c\mathbf{\Pi}_0\mathbf{X}$ ,
- $\mathbf{x}' = \mathbf{K}_s\mathbf{x}$ ,

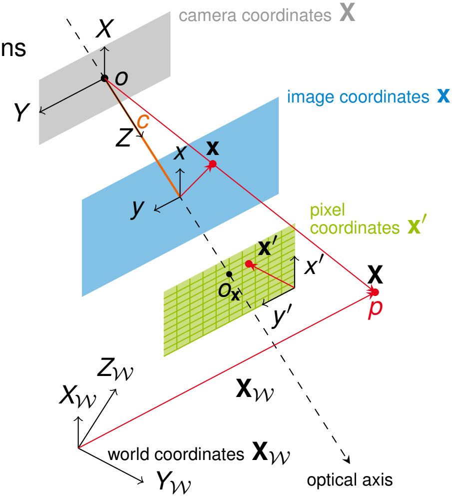
to one transformation, you get

$$\lambda\mathbf{x}' = \mathbf{K}_s\mathbf{K}_c\mathbf{\Pi}_0\mathbf{G}\mathbf{X}_{\mathcal{W}} .$$

With the abbreviated designations

- Calibration matrix: $\mathbf{K} = \mathbf{K}_s\mathbf{K}_c$ and
- Projection matrix: $\mathbf{\Pi} = \mathbf{K}\mathbf{\Pi}_0\mathbf{G}$

you reach



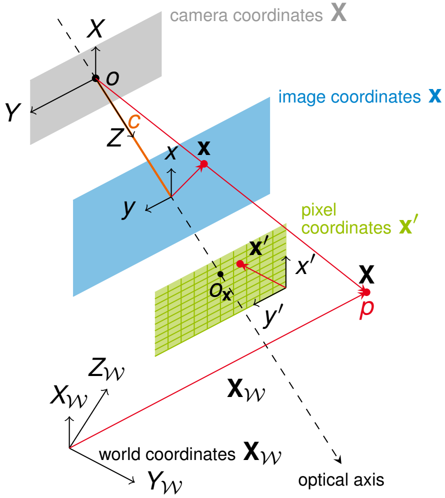
thws Technical University of Applied Sciences Würzburg-Schweinfurt

## Mapping process - Summary

the final transformation

$$\lambda \mathbf{x}' = \mathbf{\Pi} \mathbf{X}_{\mathcal{W}} = \mathbf{K} \mathbf{\Pi}_0 \mathbf{G} \mathbf{X}_{\mathcal{W}} \,,$$

where all the parameters of the calibration matrix are called intrinsic parameters resp. the intrinsics of the camera. The extrinsic parameters resp. the extrinsics of a camera are the parameters of the rigid body motion $\mathbf{G} = g(\mathbf{R}, \mathbf{T})$, i.e. the parameters that capture the pose of the camera with respect to a world coordinate system.



# Geometric Projections

## Mapping process - Intrinsics & extrinsics

Technical University of Applied Sciences Würzburg-Schweinfurt

Intrinsics:

$$\mathbf{K} = \begin{bmatrix} cs_x & cs_\theta & o_x \\ 0 & cs_y & o_y \\ 0 & 0 & 1 \end{bmatrix} .$$

- $c$ : camera constant
- $s_{\mathbf{x}}$ : scale factors $s_{\mathbf{x}} = \left[s_x, s_y\right]^\top$
- $s_\theta$ : skew factor
- $o_{\mathbf{x}}$ : principal point $o_{\mathbf{x}} = \left[o_x, o_y\right]^\top$

Extrinsics:

$$\mathbf{G} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix} .$$

- **R** : Three parameters of the rotation matrix e.g. $\Psi$, $\Theta$, $\Phi$ or $\boldsymbol{\omega}$
- **T** : three parameters of the translation vector

The internal orientation also includes the parameters that characterize the optical distortion. Depending on which model is chosen to describe the distortions, a different number of distortion parameters results. Since distortions are caused by the imaging through a real lens system, they are treated in the next section - *The lens*.

# Geometric Projections

## Mapping process - Projection matrix

The single parameters of the projection matrix

$$\boldsymbol{\Pi} = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} \\ \pi_{31} & \pi_{32} & \pi_{33} & \pi_{34} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\pi}_1^\top \\ \boldsymbol{\pi}_2^\top \\ \boldsymbol{\pi}_3^\top \end{bmatrix}$$

are composed of the individual parameters of the intrinsics and extrinsics as follows:

$$\boldsymbol{\Pi} = \begin{bmatrix} cs_x R_{11} + cs_\theta R_{21} + o_x R_{31} & cs_x R_{12} + s_\theta R_{22} + o_x R_{32} & cs_x R_{13} + cs_\theta R_{23} + o_x R_{33} & cs_x T_1 + cs_\theta T_2 + o_x T_3 \\ cs_y R_{21} + o_y R_{31} & cs_y R_{22} + o_y R_{32} & cs_y R_{23} + o_y R_{33} & cs_y T_2 + o_y T_3 \\ R_{31} & R_{32} & R_{33} & T_3 \end{bmatrix},$$

where

$$\mathbf{K} = \begin{bmatrix} cs_x & cs_\theta & o_x \\ 0 & cs_y & o_y \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{G} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0} & 1 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}, \mathbf{T} = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix}.$$

## Projection matrix - Decomposition

Technical University of Applied Sciences Würzburg-Schweinfurt

The projection matrix

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} \\ \pi_{31} & \pi_{32} & \pi_{33} & \pi_{34} \end{bmatrix} = \mathbf{K} \left[ \ \mathbf{R} | \mathbf{T} \ \right] = \left[ \ \mathbf{KR} | \mathbf{KT} \ \right]$$

can be decomposed with a RQ-decomposition (not QR!) into the matrices **K**, **R** and **T**, since **K** is an upper triangular matrix and **R** is an orthogonal matrix.

$$\text{RQ-Decomposition:} \quad \mathbf{KR} = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{bmatrix} \quad \Rightarrow \quad \{\mathbf{K}, \mathbf{R}\} \, ,$$

$$\Rightarrow \quad \mathbf{T} = \mathbf{K}^{-1} \begin{bmatrix} \pi_{14} \\ \pi_{24} \\ \pi_{34} \end{bmatrix} .$$

## Projection matrix - Interpretation of components

The columns $\mathbf{p}_i$ with $i = 1, \dots, 4$ resp. rows $\boldsymbol{\pi}_j^\top$ with $j = 1, \dots 3$ of the projection matrix have certain geometric meanings.

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} \\ \pi_{31} & \pi_{32} & \pi_{33} & \pi_{34} \end{bmatrix} = \overbrace{\begin{bmatrix} \boldsymbol{\pi}_1^\top \\ \boldsymbol{\pi}_2^\top \\ \boldsymbol{\pi}_3^\top \end{bmatrix}}^{\text{Zeilen}} = \overbrace{\begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 \end{bmatrix}}^{\text{Spalten}} .$$

The columns $\mathbf{p}_1$, $\mathbf{p}_2$ and $\mathbf{p}_3$ are the images of the vanishing points of the directions of the world coordinate axes (points in the $\infty$ along the axes) and $\mathbf{p}_4$ is the image of the world coordinate origin:

$$\begin{aligned} \mathbf{p}_1 &= \mathbf{\Pi}(1,0,0,0)^\top , \qquad \mathbf{p}_2 = \mathbf{\Pi}(0,1,0,0)^\top , \\ \mathbf{p}_3 &= \mathbf{\Pi}(0,0,1,0)^\top , \qquad \mathbf{p}_4 = \mathbf{\Pi}(0,0,0,1)^\top . \end{aligned}$$

## Geometric Projections

## Projection matrix - Interpretation of components

The columns $\mathbf{p}_i$ with $i = 1, \ldots, 4$ resp. rows $\boldsymbol{\pi}_j^\top$ with $j = 1, \ldots 3$ of the projection matrix have certain geometric meanings.

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} \\ \pi_{31} & \pi_{32} & \pi_{33} & \pi_{34} \end{bmatrix} = \overbrace{\begin{bmatrix} \boldsymbol{\pi}_1^\top \\ \boldsymbol{\pi}_2^\top \\ \boldsymbol{\pi}_3^\top \end{bmatrix}}^{\text{Zeilen}} = \overbrace{\begin{bmatrix} \mathbf{p_1} & \mathbf{p_2} & \mathbf{p_3} & \mathbf{p_4} \end{bmatrix}}^{\text{Spalten}} .$$

The lines $\boldsymbol{\pi}_1^\top$ and $\boldsymbol{\pi}_2^\top$ describe the planes which arise as back projections of the image coordinate axes. The third line $\boldsymbol{\pi}_3^\top$ corresponds to the objective plane.

$$\begin{aligned} \text{Back projection x-axis:} \quad \boldsymbol{\pi}_1^\top \overline{\mathbf{X}} &= 0 , \\ \text{Back projection y-axis:} \quad \boldsymbol{\pi}_2^\top \overline{\mathbf{X}} &= 0 , \\ \text{lens plane:} \quad \boldsymbol{\pi}_3^\top \overline{\mathbf{X}} &= 0 . \end{aligned}$$

## Projection matrix - Center of projection

The world coordinates of the center of projection $\mathbf{o}_{\mathcal{W}}$ are given by the following projective mapping

$$\mathbf{0} = \mathbf{\Pi}\overline{\mathbf{o}}_{\mathcal{W}} = \mathbf{KR}\mathbf{o}_{\mathcal{W}} + \mathbf{KT}\,.$$

($\overline{\mathbf{o}}_{\mathcal{W}}$ = Null space of projection matrix, $\mathbf{T}$ = Camera coordinates of the origin of the world coordinate system)

That equals the intersection point $\overline{\mathbf{X}} = \overline{\mathbf{o}}_{\mathcal{W}}$ of the following three planes $\pi_1^\top \overline{\mathbf{X}} = 0$, $\pi_2^\top \overline{\mathbf{X}} = 0$ and $\pi_3^\top \overline{\mathbf{X}} = 0$.

$$\Rightarrow \mathbf{o}_{\mathcal{W}} = -\left[\ \mathbf{KR}\ \right]^{-1} \mathbf{KT}\,.$$

Thus, the projection matrix can be alternatively represented as follows:

$$\mathbf{\Pi} = \left[\ \mathbf{KR} \middle| -\mathbf{KR}\mathbf{o}_{\mathcal{W}}\ \right] = \mathbf{K}\left[\ \mathbf{R} \middle| -\mathbf{R}\mathbf{o}_{\mathcal{W}}\ \right]\,.$$

# Geometric Projections

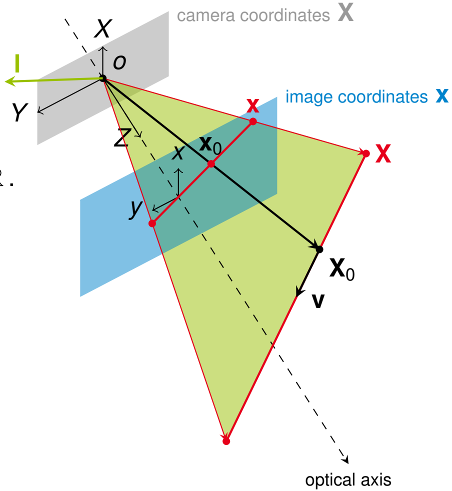
## Central projection of a line

Representation of a straight line in space in camera coordinates:

$$\mathbf{X} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix} + \mu \begin{bmatrix} v_x \\ v_y \\ v_z \\ 0 \end{bmatrix}, \ \mu \in \mathbb{R} .$$

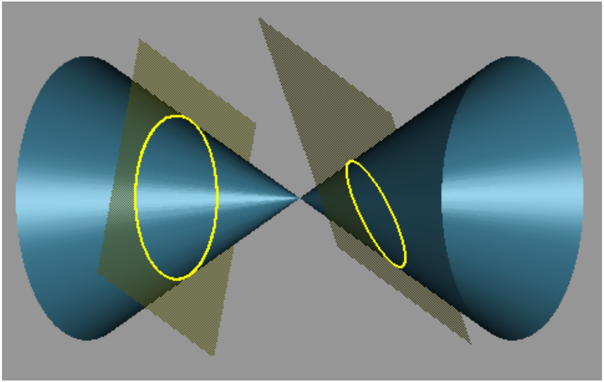
The 2D straight line on the image plane can be represented as a normal vector $\mathbf{l} = [a, b, c]^\top \in \mathbb{R}^3$ of the plane, which is spanned by the 3D straight line and the projection center:

$$\mathbf{l}^\top \mathbf{x} = \mathbf{l}^\top \mathbf{\Pi}_0 \mathbf{X} = 0 .$$



# Geometric Projections

## Central projection of ellipses



# Geometric Projections

## Central projection of a plane - Homography

If we include the homogeneous plane equation

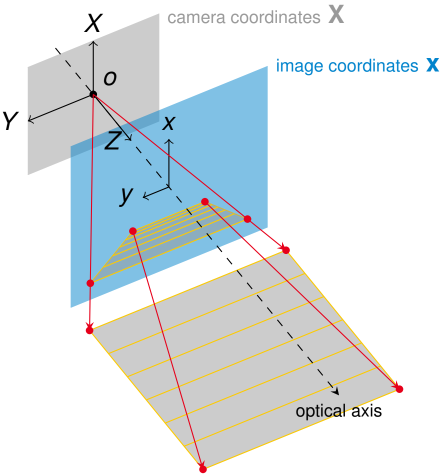
$$\mathbf{e}^\top \mathbf{X}_{\mathcal{W}} = 0\,, \text{ with } \mathbf{e} = [e_1, e_2, e_3, e_4]\,,$$

solved for $Z$

$$Z_{\mathcal{W}} = -\frac{e_1}{e_3} X_{\mathcal{W}} - \frac{e_2}{e_3} Y_{\mathcal{W}} - \frac{e_4}{e_3}$$

into the general projection equation

$$\lambda \mathbf{x}' = \mathbf{\Pi} \mathbf{X}_{\mathcal{W}} = \begin{bmatrix} | & | & | & | \\ \pi_1 & \pi_2 & \pi_3 & \pi_4 \\ | & | & | & | \end{bmatrix} \mathbf{X}_{\mathcal{W}}\,,$$



# Geometric Projections

## Central projection of a plane - Homography

you get the dependency

$$\lambda \mathbf{x}' = \begin{bmatrix} | & | & | \\ (\pi_1 - \frac{e_1}{e_3}\pi_3) & (\pi_2 - \frac{e_2}{e_3}\pi_3) & (\pi_4 - \frac{e_4}{e_3}\pi_3) \\ | & | & | \end{bmatrix} \begin{bmatrix} X_{\mathcal{W}} \\ Y_{\mathcal{W}} \\ 1 \end{bmatrix},$$

which is a homography **H**, also called collinear mapping

$$\lambda \mathbf{x}' = \lambda \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} | & | & | \\ \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} X_{\mathcal{W}} \\ Y_{\mathcal{W}} \\ 1 \end{bmatrix} = \mathbf{H} \begin{bmatrix} X_{\mathcal{W}} \\ Y_{\mathcal{W}} \\ 1 \end{bmatrix},$$

which maps, for example, rectangles to general quadrilaterals
and straight lines back to straight lines.

# Geometric Projections

## Central projection of a plane - Homography

If you put the world coordinate system in the plane, e.g.

$$\mathbf{e} = [0, 0, 1, 0] \ , \quad \rightarrow \quad Z_{\mathcal{W}} = 0 \, ,$$

then the homography is composed of three column vectors of the projection matrix, e.g.

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3] = [\pi_1, \pi_2, \pi_4] \ .$$

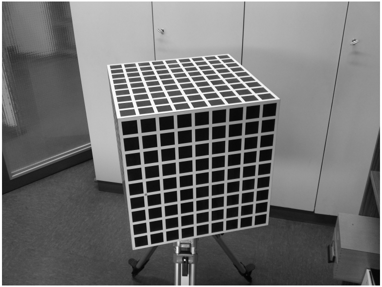
An affine homography imposes further restrictions on the mapping $h_{31} = h_{32} = 0$ and $h_{33} = 1$. This also preserves parallelism and rectangles are mapped to parallelograms.

$$\mathbf{H}_{\text{affine}} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ 0 & 0 & 1 \end{bmatrix} .$$

# Geometric Projections

## Camera Calibration - Calibration rig

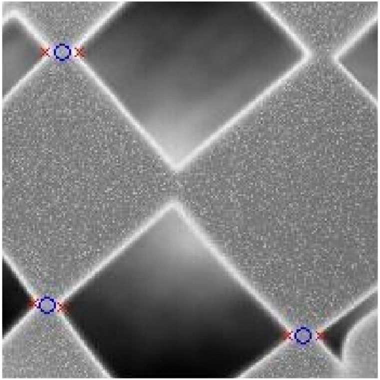
If a 3D calibration body with $N$ accurately measured 3D points $\mathbf{X}_i$, $i = 1, \ldots, N$, exists, then the projection matrix $\mathbf{\Pi}$ can be calculated from a single image acquisition of the calibration body. For this purpose, the 2D coordinates $\mathbf{x}_i$ of the projected 3D points of the calibration body must be determined as accurately as possible.

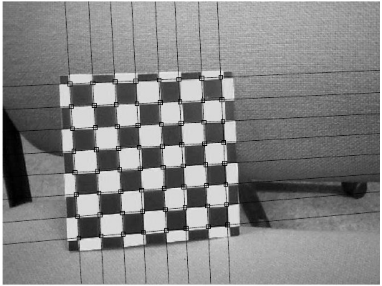


($\mathbf{x}_i$ now corresponds to the coordinates of the image matrix $\mathbf{x}'_i$ and the 3D points $\mathbf{X}_i$ are measured with respect to an arbitrarily chosen coordinate system of the calibration body.)

# Geometric Projections

## Camera Calibration - Main problem



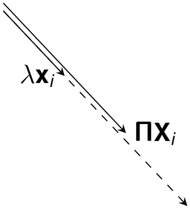


# Geometric Projections

## Camera Calibration - Linear approach: DLT

For each 2D-3D pair of points $\{\mathbf{x}_i, \mathbf{X}_i\}$ the transformation rule holds

$$\lambda \mathbf{x}_i = \mathbf{\Pi} \mathbf{X}_i = \begin{bmatrix} \boldsymbol{\pi}_1^\top \\ \boldsymbol{\pi}_2^\top \\ \boldsymbol{\pi}_3^\top \end{bmatrix} \mathbf{X}_i \,.$$




The cross product of both equation sides with $\mathbf{x}_i$ results in three equations which are linear in the 12 parameters $\boldsymbol{\pi} = [\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3]^\top$ of the projection matrix

$$\lambda \widehat{\mathbf{x}}_i \mathbf{x}_i = \mathbf{0} = \widehat{\mathbf{x}}_i \mathbf{\Pi} \mathbf{X}_i = \underbrace{(\widehat{\mathbf{x}}_i \otimes \mathbf{X}_i^\top)}_{\mathbf{M}_i} \boldsymbol{\pi} \,.$$

$$\widehat{\mathbf{x}}_i = \begin{bmatrix} 0 & -1 & y_i \\ 1 & 0 & -x_i \\ -y_i & x_i & 0 \end{bmatrix}$$

Technical University of Applied Sciences Würzburg-Schweinfurt

## Camera Calibration - Linear approach: DLT

$$\mathbf{M}_i \pi = (\widehat{\mathbf{x}}_i \otimes \mathbf{X}_i^\top)\pi \quad = \quad \mathbf{0}\,,$$

$$(\widehat{\mathbf{x}}_i \otimes \mathbf{X}_i^\top)\pi = \begin{bmatrix} \mathbf{0}^\top & -\mathbf{X}_i^\top & y_i\mathbf{X}_i^\top \\ \mathbf{X}_i^\top & \mathbf{0}^\top & -x_i\mathbf{X}_i^\top \\ -y_i\mathbf{X}_i^\top & x_i\mathbf{X}_i^\top & \mathbf{0}^\top \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} \quad = \quad \mathbf{0}\,.$$

Only two of the three equations are linearly independent. Therefore each pair of points $\{\mathbf{x}_i, \mathbf{X}_i\}$ provide two equations:

$$\begin{bmatrix} \mathbf{0}^\top & -\mathbf{X}_i^\top & y_i\mathbf{X}_i^\top \\ \mathbf{X}_i^\top & \mathbf{0}^\top & -x_i\mathbf{X}_i^\top \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \mathbf{0}\,.$$

This transformation is also called Direct Linear Transformation (DLT).

## Camera Calibration - Linear approach: DLT

The vector $\boldsymbol{\pi}$ has 12 entries $\pi_n$. Since the transformation rule is valid for arbitrary scalings, the vector $\boldsymbol{\pi}$ has only 11 degrees of freedom. To get an exact solution, exactly 11 linearly independent equations are necessary, i.e. 5.5 pairs of points. Since errors in the measurement of both the projected and the coordinates of the calibration body, more points should be used to determine $\boldsymbol{\pi}$. This leads to a solution of an overdetermined linear homogeneous equation system $\mathbf{M}\boldsymbol{\pi} = 0$, where $\mathbf{M}$ has dimension $2N \times 12$.

$$\underbrace{\begin{bmatrix} \mathbf{0}^\top & -\mathbf{X}_1^\top & y_1\mathbf{X}_1^\top \\ \mathbf{X}_1^\top & \mathbf{0}^\top & -x_1\mathbf{X}_1^\top \\ \vdots & \vdots & \vdots \\ \mathbf{0}^\top & -\mathbf{X}_i^\top & y_i\mathbf{X}_i^\top \\ \mathbf{X}_i^\top & \mathbf{0}^\top & -x_i\mathbf{X}_i^\top \\ \vdots & \vdots & \vdots \\ \mathbf{0}^\top & -\mathbf{X}_N^\top & y_N\mathbf{X}_N^\top \\ \mathbf{X}_N^\top & \mathbf{0}^\top & -x_N\mathbf{X}_N^\top \end{bmatrix}}_{\mathbf{M}} \underbrace{\begin{bmatrix} \boldsymbol{\pi}_1 \\ \boldsymbol{\pi}_2 \\ \boldsymbol{\pi}_3 \end{bmatrix}}_{\boldsymbol{\pi}} = \mathbf{0}\,.$$

$\boldsymbol{\pi}$ is null space of $\mathbf{M}$.

# Geometric Projections

## Camera Calibration - Solution: SVD

The homogeneous overdetermined system of equations can be solved by minimization of the quadratic error under a constraint on $\boldsymbol{\pi}$. This is the so-called

Algebraic Error

$$\hat{\boldsymbol{\pi}} = \text{argmin}_{\boldsymbol{\pi}} \|\mathbf{M}\boldsymbol{\pi}\|^2 , \quad \text{where} \quad \|\boldsymbol{\pi}\| = 1 .$$

The solution corresponds to the eigenvector $\hat{\boldsymbol{\pi}} = \mathbf{v}_{min}$ of the matrix $\mathbf{M}^\top\mathbf{M}$ of the smallest eigenvalue $\lambda_{min}$. This eigenvector results from a singular value decomposition

$$\mathbf{M} = \mathbf{USV}^\top , \quad \text{where} \quad \hat{\boldsymbol{\pi}} = \mathbf{v}_{12} = \mathbf{v}_{min} ,$$

with $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_{12}]$.

# Geometric Projections

## Camera Calibration - Refined solution

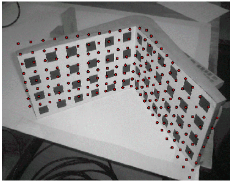
The solution via the algebraic error can be improved by a subsequent minimization of the squared Euclidean distance of measured $\mathbf{x}_i$ and projected $\hat{\mathbf{x}}_i = h(\mathbf{\Pi}\overline{\mathbf{X}}_i)$ 2D image coordinates.

Geometric Error (similar to reprojection error)

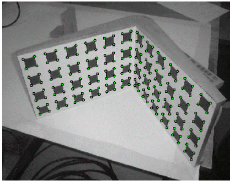
$$\hat{\mathbf{\Pi}} = \operatorname{argmin}_{\mathbf{\Pi}} \sum_{i=1}^{N} \left\| \mathbf{x}_i - h(\mathbf{\Pi}\overline{\mathbf{X}}_i) \right\|^2, \quad \text{mit} \quad h(\mathbf{X}) = \frac{[X, Y]^\top}{Z}.$$

The solution of this nonlinear minimization problem is obtained by iterative optimization methods, such as the Gauß-Newton method.

$\hat{\mathbf{\Pi}}_{initial}$



$\hat{\mathbf{\Pi}}_{converged}$



thws Technical University of Applied Sciences Würzburg-Schweinfurt

## Camera Calibration - Normalization

Also a normalization of the image $\mathbf{x}_i$ and space coordinates $\mathbf{X}_i$ of the calibration points can decisively improve the estimation result of the projection parameters $\hat{\pi}$. The following normalizations are useful:

$$\tilde{\mathbf{x}}_i = \mathbf{T}\mathbf{x}_i \,, \quad \text{with} \quad \mathbf{T} = \begin{bmatrix} s & 0 & -st_x \\ 0 & s & -st_y \\ 0 & 0 & 1 \end{bmatrix} ,$$

$$\text{where} \quad \mathbf{t} = \begin{bmatrix} t_x \\ t_y \\ 1 \end{bmatrix} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \quad \text{and} \quad \sqrt{2} = s\frac{1}{N}\sqrt{\sum_{i=1}^{N}(\mathbf{x}_i - \mathbf{t})^\top(\mathbf{x}_i - \mathbf{t})} \,.$$

## Camera Calibration - Normalization

If the distance $Z_i$ of the space coordinates $\mathbf{X}_i$ does not vary too much, then one uses a normalization $\tilde{\mathbf{X}}_i = \mathbf{U}\mathbf{X}_i$ analogous to the image coordinates, which places the coordinate origin in the geometric center of gravity and scales it so that the average distance of all coordinates to the origin is $\sqrt{3}$.

Since the projection matrix $\tilde{\mathbf{\Pi}}$ is calculated for the transformed coordinates, the actual projection matrix $\mathbf{\Pi}$ is obtained by the following inverse transformation:

$$\mathbf{\Pi} = \mathbf{T}^{-1}\tilde{\mathbf{\Pi}}\mathbf{U}\,.$$

## Camera Calibration - Number of points

In order to achieve a good estimation in spite of measurement errors, at least five times as many equations $2N$ as degrees of freedom $F$ should be chosen.

$$\text{Rule of Thumb}: \quad 2N \geq 5 \cdot F \rightarrow N \geq 2.5 \cdot 11 \approx 28\,.$$

Moreover, not every arbitrary point configuration in space leads to a unique solution of the transformation $\mathbf{\Pi}$.

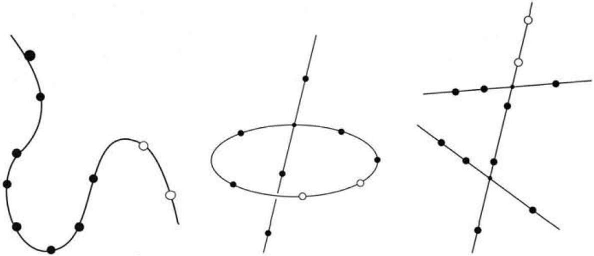
### Ambiguities

Ambiguities in the calculation of the transformation rule $\mathbf{\Pi}$ occur with certain point-camera relations. They depend only on the position of the projection center or the projection centers with respect to the points of the calibration body.

## Geometric Projections

### Camera Calibration - Ambiguities

Configurations of points and projection centers in space that result in identical images.

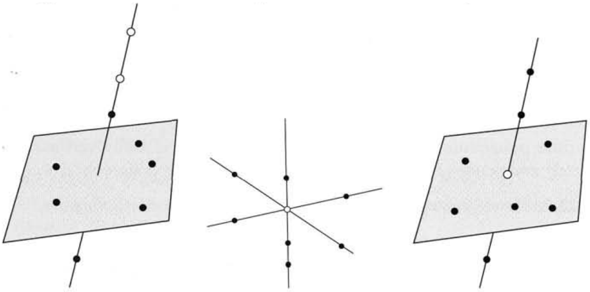


○: different projection centers
●: 3D point configurations

(Hartley & Zissermann p.533-539)

## Camera Calibration - Ambiguities

Configurations of points and projection centers in space that result in identical images.



○: different projection centers

●: 3D point configurations

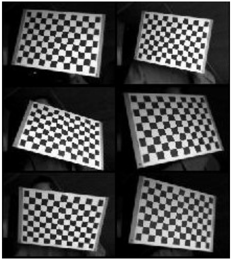
(Hartley & Zissermann p.533-539)

# Geometric Projections

## Camera Calibration - Planar rig

If a planar calibration rig is used, then a multi-image calibration method must be used. For this purpose, the world coordinate system is selected so that the X-Y plane corresponds to the plane calibration body:

$$\mathbf{X}_i = [X_i, Y_i, 0, 1]^\top \ .$$

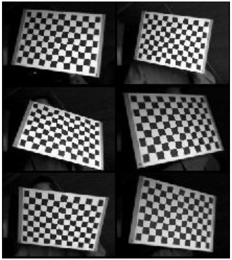


This gives a homography between the plane calibration body and the image plane:

$$\lambda \mathbf{x}_i = \mathbf{\Pi X}_i = \mathbf{K}\,[\mathbf{R}, \mathbf{T}]\,\mathbf{X}_i = \mathbf{K}\,[\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{T}]\,\mathbf{X}_i \quad \rightarrow \quad \lambda \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \underbrace{\mathbf{K}\,[\mathbf{r}_1, \mathbf{r}_2, \mathbf{T}]}_{\mathbf{H}} \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix} .$$

# Geometric Projections

## Camera Calibration - Planar rig



If at least four points on the plane calibration body are known, then this equation can be solved according to the same scheme, as for the single-body calibration method:

$$\widehat{\mathbf{x}}_i \mathbf{H} \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix} = 0\,, \quad \text{wobei} \quad \mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3]\,,$$

$$\rightarrow \quad \underbrace{\widehat{\mathbf{x}}_i \otimes [X_i, Y_i, 1]}_{\mathbf{M}_i} \mathbf{h} = 0\,, \quad \text{mit} \quad \mathbf{h} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \mathbf{h}_3 \end{bmatrix}.$$

The solution of the equation system is usually done via a singular value decomposition (analogous to slide 36-38).

# Geometric Projections

## Camera Calibration - Planar rig

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

If the homography $\mathbf{H}$ is known, then the following relation holds:

$$\mathbf{H} = \mathbf{K}[\mathbf{r}_1, \mathbf{r}_2, \mathbf{T}] \quad \text{also} \quad [\mathbf{h}_1, \mathbf{h}_2] \sim \mathbf{K}[\mathbf{r}_1, \mathbf{r}_2] \quad \text{bzw.} \quad \mathbf{K}^{-1}[\mathbf{h}_1, \mathbf{h}_2] \sim [\mathbf{r}_1, \mathbf{r}_2] \; .$$

Since $\mathbf{r}_1$ and $\mathbf{r}_2$ are orthogonal to each other and of equal amount $\|\mathbf{r}_1\| = \|\mathbf{r}_2\| = 1$, the following conditions for the calibration matrix $\mathbf{K}$ result:

1. $$\mathbf{r}_1^\top \mathbf{r}_2 = \left(\mathbf{K}^{-1}\mathbf{h}_1\right)^\top \mathbf{K}^{-1}\mathbf{h}_2 = \mathbf{h}_1^\top \underbrace{\mathbf{K}^{-\top}\mathbf{K}^{-1}}_{\mathbf{S}} \mathbf{h}_2 = 0 \, ,$$

2. $$\mathbf{r}_1^\top \mathbf{r}_1 = \mathbf{r}_2^\top \mathbf{r}_2 \quad \text{bzw.} \quad \mathbf{h}_1^\top \underbrace{\mathbf{K}^{-\top}\mathbf{K}^{-1}}_{\mathbf{S}} \mathbf{h}_1 = \mathbf{h}_2^\top \underbrace{\mathbf{K}^{-\top}\mathbf{K}^{-1}}_{\mathbf{S}} \mathbf{h}_2 \, .$$

### Camera Calibration - Planar rig

1. $\mathbf{h}_1^\top \mathbf{S} \mathbf{h}_2 = 0\,,$
2. $\mathbf{h}_1^\top \mathbf{S} \mathbf{h}_1 = \mathbf{h}_2^\top \mathbf{S} \mathbf{h}_2\,.$

That means, each image of the calibration body from a certain viewing angle provides two equations (1. and 2.) to calculate $\mathbf{S} = \mathbf{K}^{-\top}\mathbf{K}^{-1}$ with different homographies $\mathbf{H}_j$. Since the calibration matrix $\mathbf{K}$ has five parameters, one needs at least three different exposures of the calibration rig to uniquely determine $\mathbf{K}$. The matrix $\mathbf{S}$ can again be calculated by a singular value decomposition and the upper triangular matrix $\mathbf{K}$ is obtained e.g. via a Cholesky decomposition of the symmetric positive definite matrix $\mathbf{S}$.