# Basics of Epipolar Geometry

Hellmuth Stachel

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

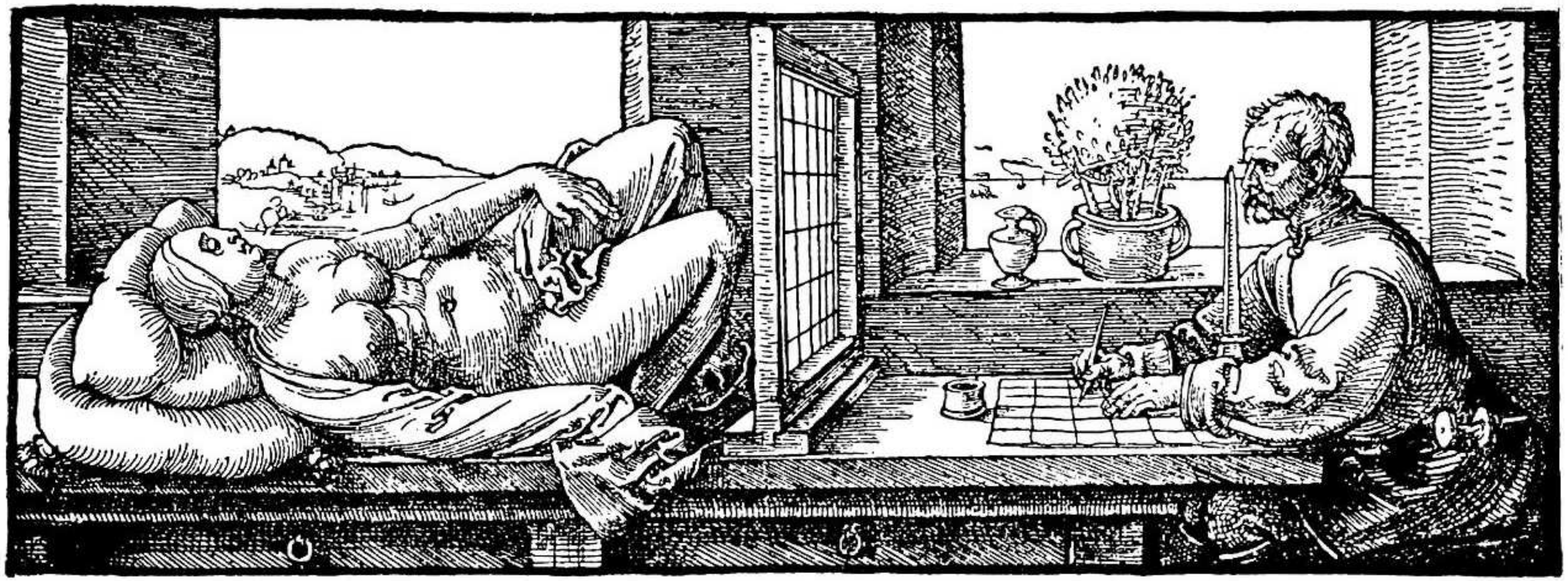
stachel@dmg.tuwien.ac.at — http://www.geometrie.tuwien.ac.at/stachel

Seminar at the Institute of Mathematics and Physics
Faculty of Mechanical Engineering, May 26, 2014, STU Bratislava

# Table of contents

# 1. Linear images of the 3-space

The **central projection** produces a linear image = "central perspective"

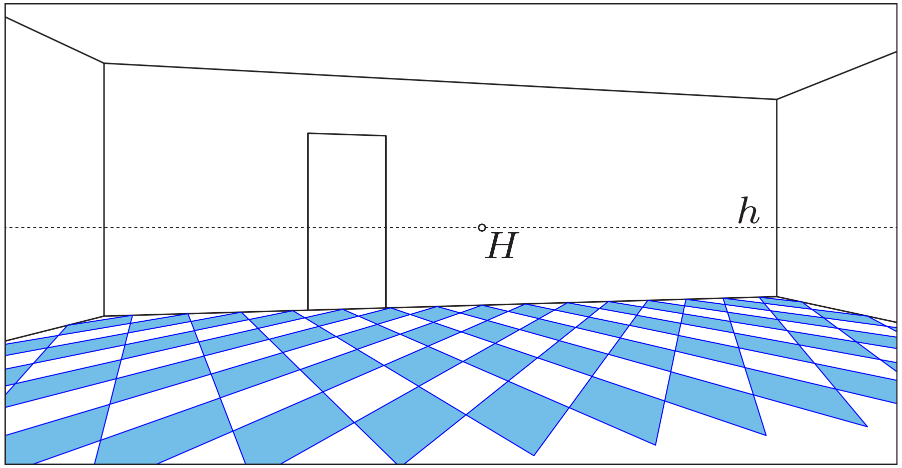


. . . according to A. DÜRER's woodcarving (1512-1525)
(in 'Underweysung der Messung mit dem Zirckel und Richtscheyt')

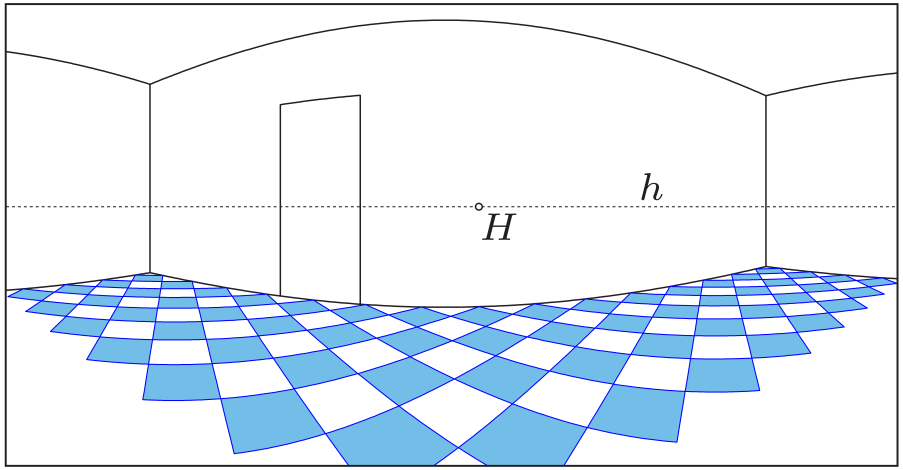
# 1. Linear images of the 3-space


A **panorama** is a nonlinear image — but can be generated from a perspective by a planar transformation according to the unfolding of a right cylinder.

# 1. Linear images of the 3-space



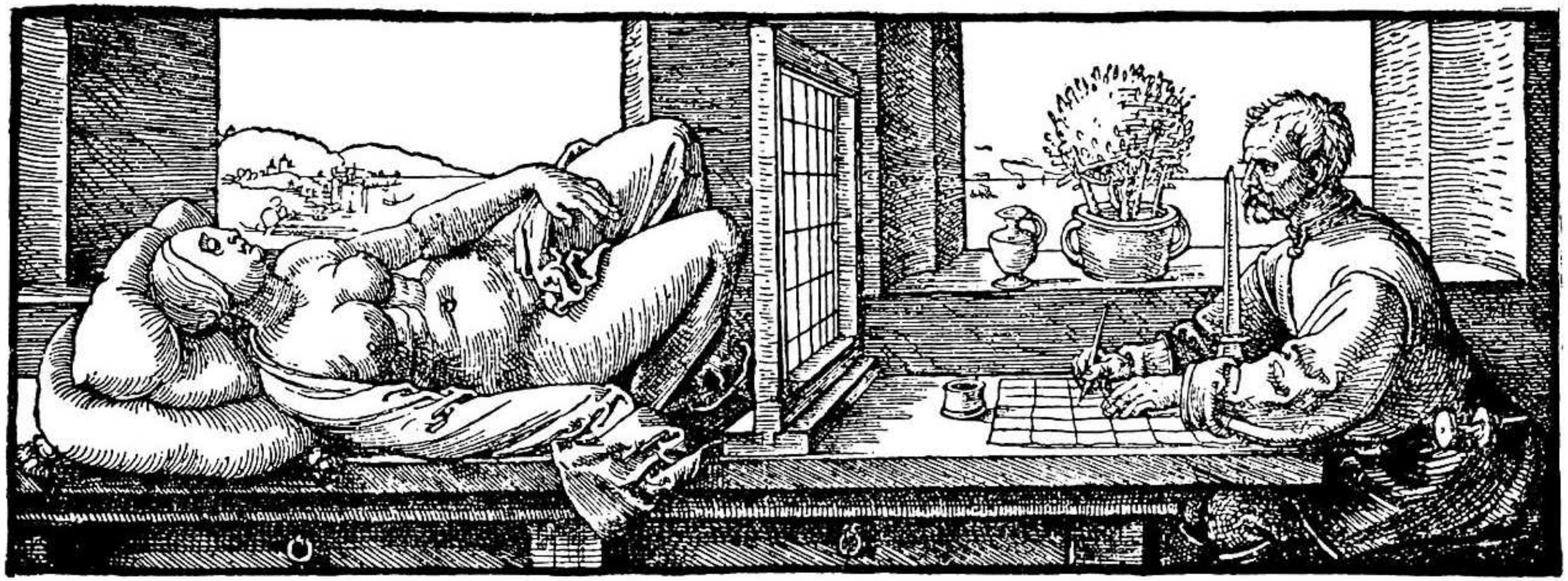

$\rightarrow$




$$(x, y) \mapsto (x^p, y^p) \quad \text{with} \quad x^p = r \arctan \frac{x}{d}, \qquad y^p = \frac{ry}{\sqrt{d^2 + x^2}}$$

# Central projection

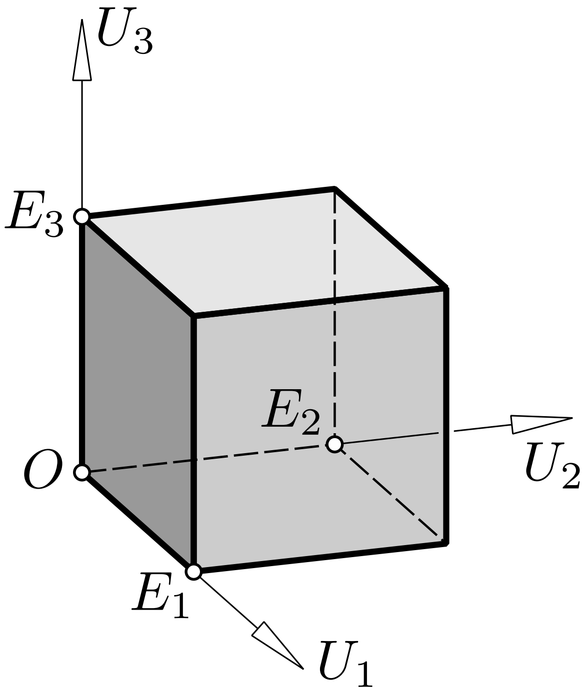
The central projection (according to A. Dürer)



can be generalized by a central axonometry.

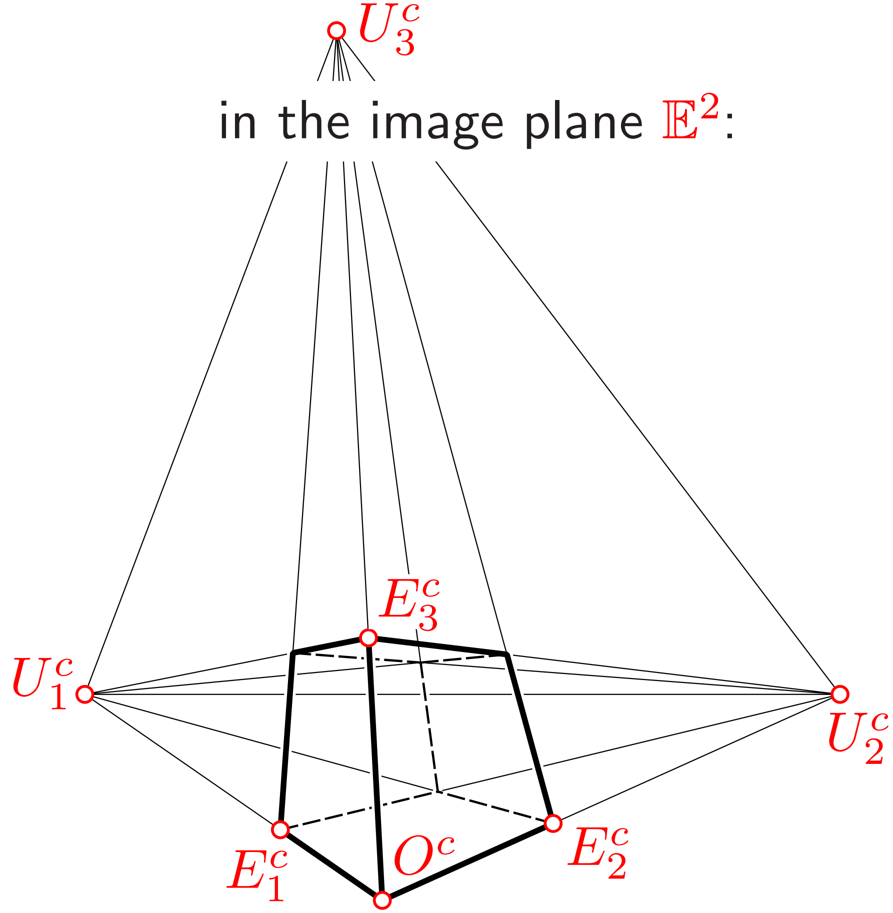
# Central axonometric principle

in space $\mathbb{E}^3$:




cartesian basis $O; E_1, E_2, E_3$
and points at infinity $U_1, U_2, U_3$

in the image plane $\mathbb{E}^2$:



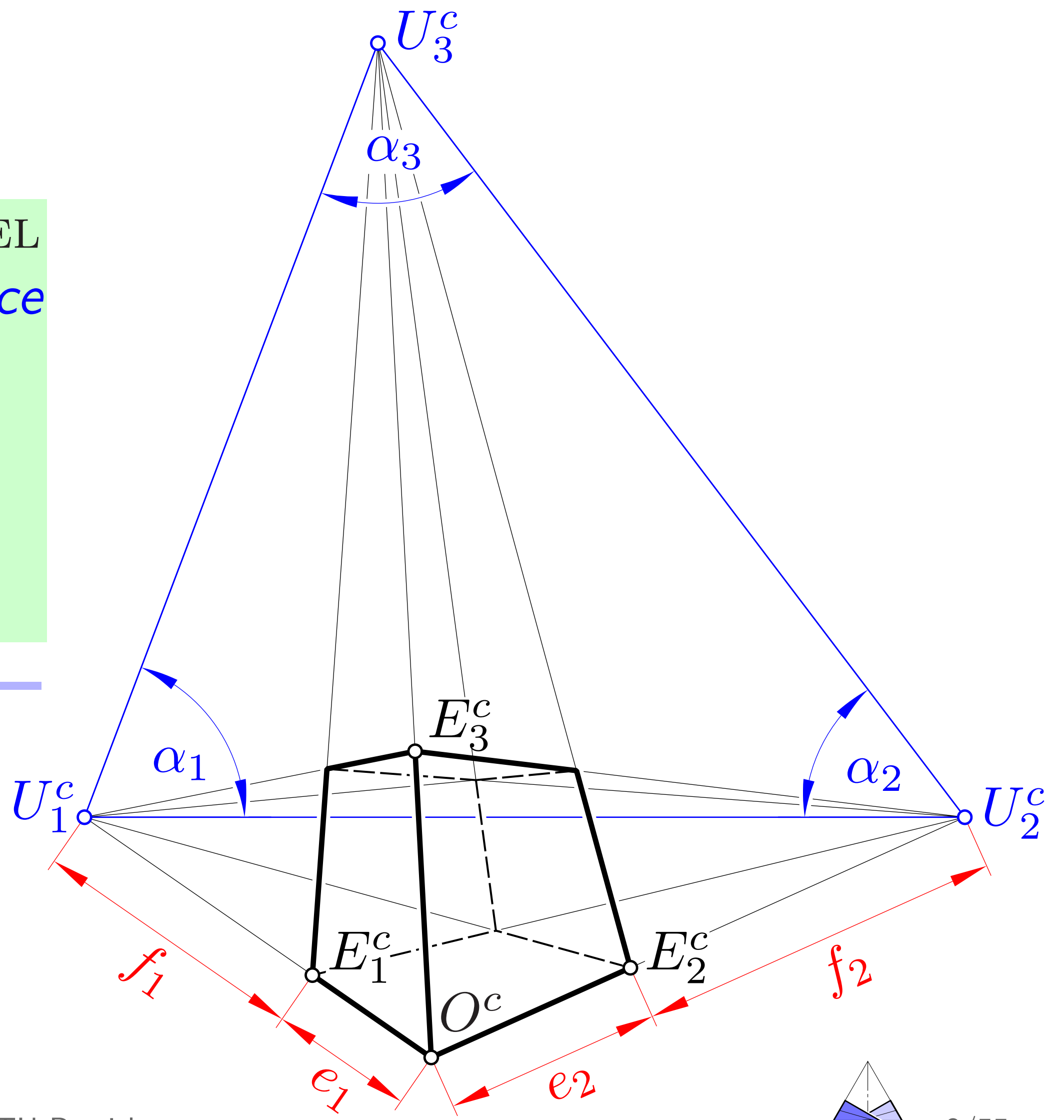

GIVEN: central axonometric reference system
$O^c; E_1^c, E_2^c, E_3^c; U_1^c, U_2^c, U_3^c$

# Central axonometric principle

**Theorem** (Szabó, H.S., Vogel 1994): *A central axonometric reference system defines a central projection*

$$\Longleftrightarrow \quad \left(\frac{e_1}{f_1}\right)^2 : \left(\frac{e_2}{f_2}\right)^2 : \left(\frac{e_3}{f_3}\right)^2 = \\ = \tan\alpha_1 : \tan\alpha_2 : \tan\alpha_3 \,.$$

**Theorem:** *Each central axonometry is affine to a central projection.*



# Photo versus linear image

Photo (= central perspective) or photo of a photo (= linear image) ?

# Definition of a linear image

There is a unique collinear transformation

$$\kappa\colon\ \mathbb{E}^3 \to \mathbb{E}^2 \ \text{ mit } \ O \mapsto O^c,\ E_i \mapsto E_i^c,\ U_i \mapsto U_i^c, \quad i = 1, 2, 3.$$

---

Any two-dimensional image of $\mathbb{E}^3$ under a collinear transformation is called *linear*.

$$\Longrightarrow \quad \begin{cases} \text{collinear points have collinear or coincident images} \\ \text{cross-ratios of any four collinear points are preserved.} \end{cases}$$

# Definition of a linear image

There is a unique collinear transformation

$$\kappa\colon\ \mathbb{E}^3 \to \mathbb{E}^2 \ \text{ mit }\ O \mapsto O^c,\ E_i \mapsto E_i^c,\ U_i \mapsto U_i^c, \quad i = 1, 2, 3.$$

---

Any two-dimensional image of $\mathbb{E}^3$ under a collinear transformation is called *linear*.

$$\Longrightarrow \quad \left\{ \begin{array}{l} \text{collinear points have collinear or coincident images} \\ \text{cross-ratios of any four collinear points are preserved.} \end{array} \right.$$

# Central projection in coordinates

**Notation:**

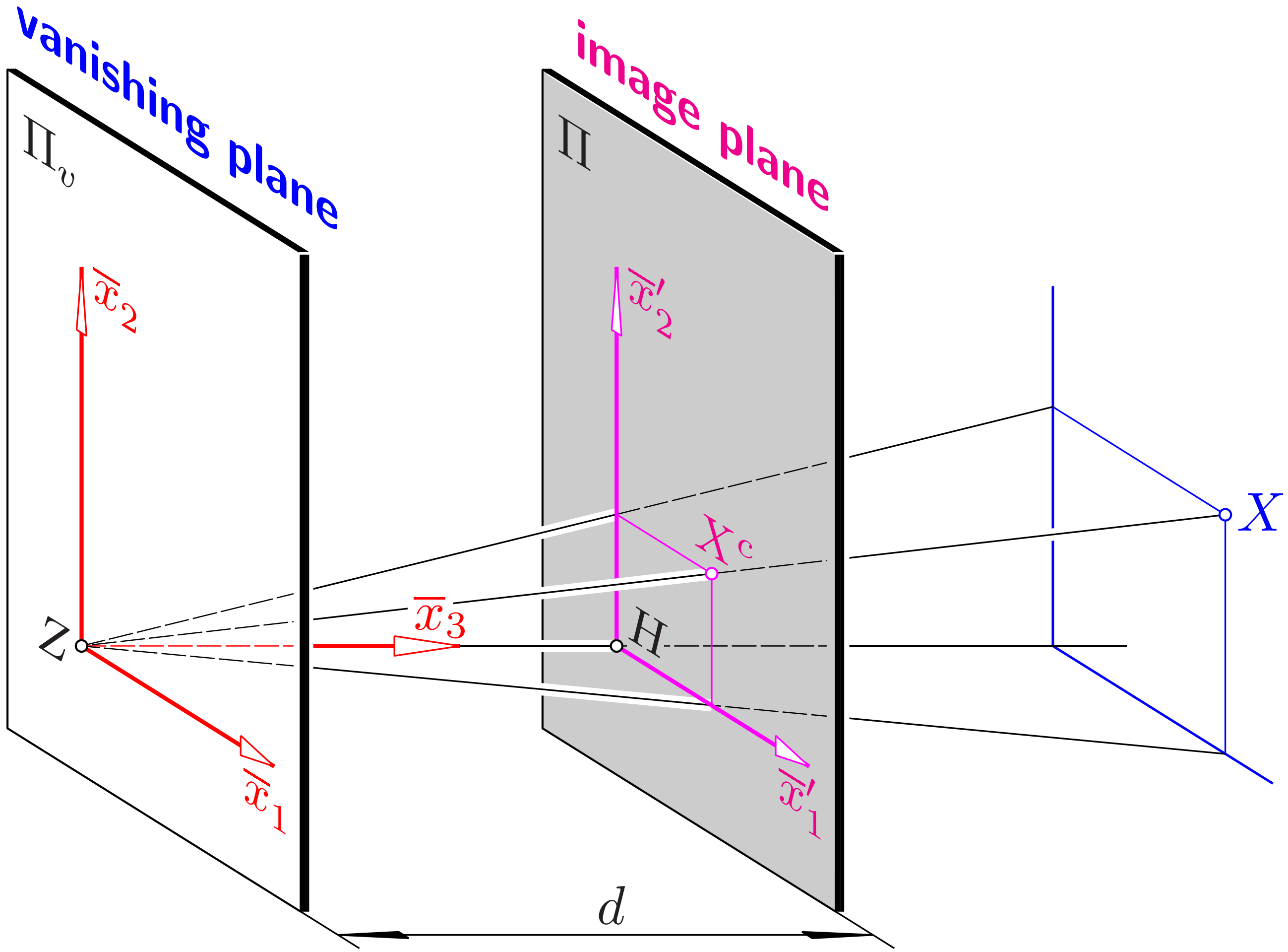
$Z$ . . . center

$H$ . . . principal point

$d$ . . . focal length

$\overline{x}_1, \overline{x}_2, \overline{x}_3$ . . . camera frame

$\overline{x}'_1, \overline{x}'_2$ . . . image coordinate frame



# Central projection in coordinates

$$\begin{pmatrix} \overline{x}'_1 \\ \overline{x}'_2 \end{pmatrix} = \frac{d}{\overline{x}_3} \begin{pmatrix} \overline{x}_1 \\ \overline{x}_2 \end{pmatrix}, \text{ or homogeneous } \begin{pmatrix} \overline{\xi}'_0 \\ \overline{\xi}'_1 \\ \overline{\xi}'_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & d & 0 & 0 \\ 0 & 0 & d & 0 \end{pmatrix} \begin{pmatrix} \overline{\xi}_0 \\ \vdots \\ \overline{\xi}_3 \end{pmatrix}.$$

Transformation from the camera frame $(\overline{x}_1, \overline{x}_2, \overline{x}_3)$ into arbitrary world coordinates $(x_1,\, x_2,\, x_3)$ and translation from the particular image frame $(\overline{x}'_1, \overline{x}'_2)$ into arbitrary $(x'_1, x'_2)$ gives in homogeneous form

$$\begin{pmatrix} \xi'_0 \\ \xi'_1 \\ \xi'_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ h'_1 & d\,f_1 & 0 \\ h'_2 & 0 & d\,f_2 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \left(\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline o_1 & & & \\ \vdots & & \mathbf{R} & \\ o_3 & & & \end{array}\right)}_{\text{matrix } \mathbf{A}} \begin{pmatrix} \xi_0 \\ \vdots \\ \xi_3 \end{pmatrix}.$$

# Central projection in coordinates

$$\begin{pmatrix} \overline{x}'_1 \\ \overline{x}'_2 \end{pmatrix} = \frac{d}{\overline{x}_3} \begin{pmatrix} \overline{x}_1 \\ \overline{x}_2 \end{pmatrix}, \text{ or homogeneous } \begin{pmatrix} \overline{\xi}'_0 \\ \overline{\xi}'_1 \\ \overline{\xi}'_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & d & 0 & 0 \\ 0 & 0 & d & 0 \end{pmatrix} \begin{pmatrix} \overline{\xi}_0 \\ \vdots \\ \overline{\xi}_3 \end{pmatrix}.$$

Transformation from the camera frame $(\overline{x}_1, \overline{x}_2, \overline{x}_3)$ into arbitrary world coordinates $(x_1,\ x_2,\ x_3)$ and translation from the particular image frame $(\overline{x}'_1, \overline{x}'_2)$ into arbitrary $(x'_1, x'_2)$ gives in homogeneous form

$$\begin{pmatrix} \xi'_0 \\ \xi'_1 \\ \xi'_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ h'_1 & d\,f_1 & 0 \\ h'_2 & 0 & d\,f_2 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \left( \begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline o_1 & & & \\ \vdots & & \mathbf{R} & \\ o_3 & & & \end{array} \right)}_{\text{matrix } \mathbf{A}} \begin{pmatrix} \xi_0 \\ \vdots \\ \xi_3 \end{pmatrix}.$$

# Central projection in coordinates

Left hand matrix: $(h'_1, h'_2)$ are image coordinates of the principal point $H$, $(f_1, f_2)$ are possible scaling factors, and $d$ is the focal length.

These parameters are called the intrinsic calibration parameters.

Right hand matrix: $\mathbf{R}$ is an orthogonal matrix.

The position of the camera frame with respect to the world coordinates defines the extrinsic calibration parameters.

Photos with known interior calibration parameters are called calibrated images, others (like central axonometries) are uncalibrated.

# Central projection in coordinates

Left hand matrix: $(h'_1, h'_2)$ are image coordinates of the principal point $H$, $(f_1, f_2)$ are possible scaling factors, and $d$ is the focal length.

These parameters are called the intrinsic calibration parameters.

---

Right hand matrix: $\mathbf{R}$ is an orthogonal matrix.

The position of the camera frame with respect to the world coordinates defines the extrinsic calibration parameters.

---

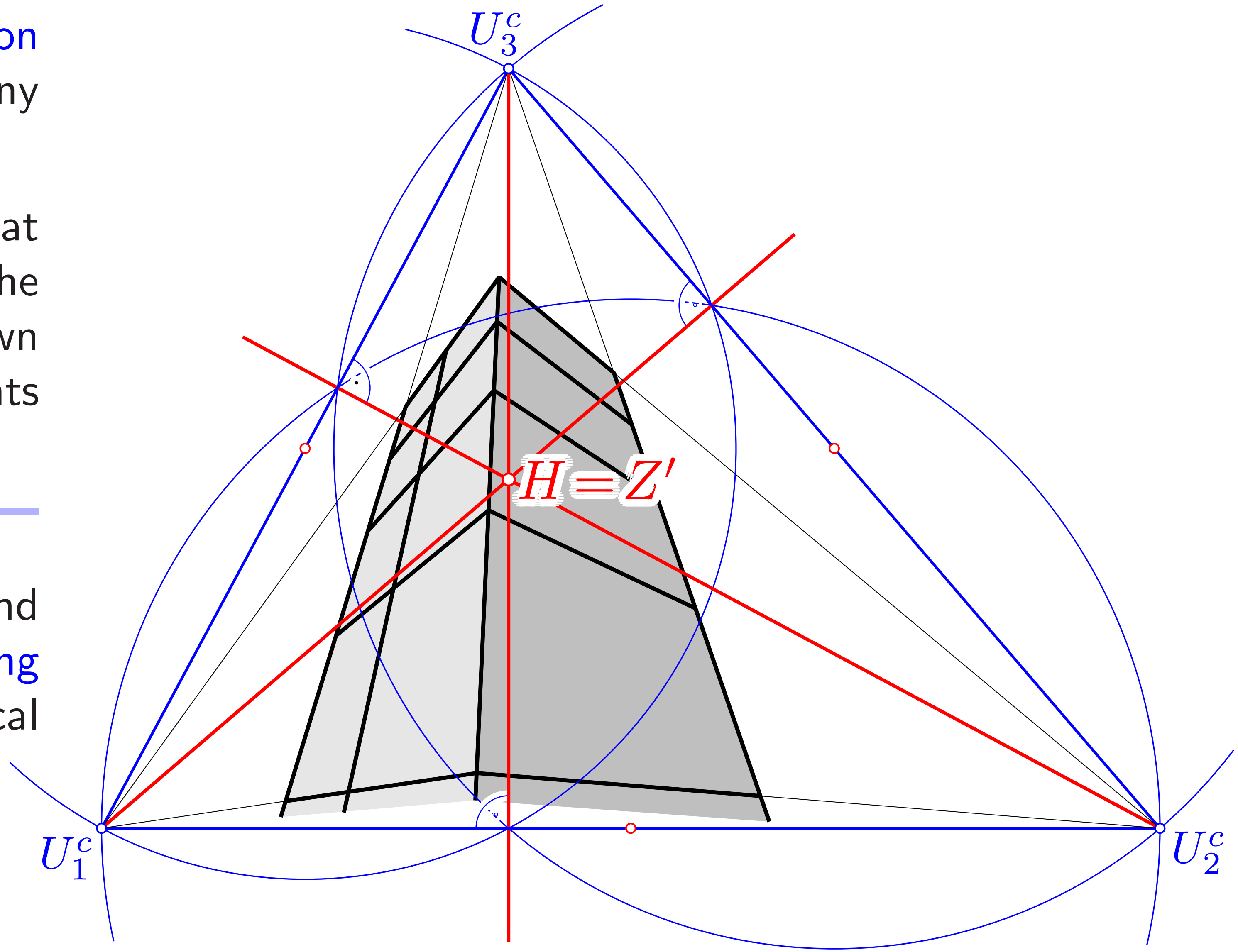
Photos with known interior calibration parameters are called calibrated images, others (like central axonometries) are uncalibrated.

How to check the interior calibration parameters of any camera ?

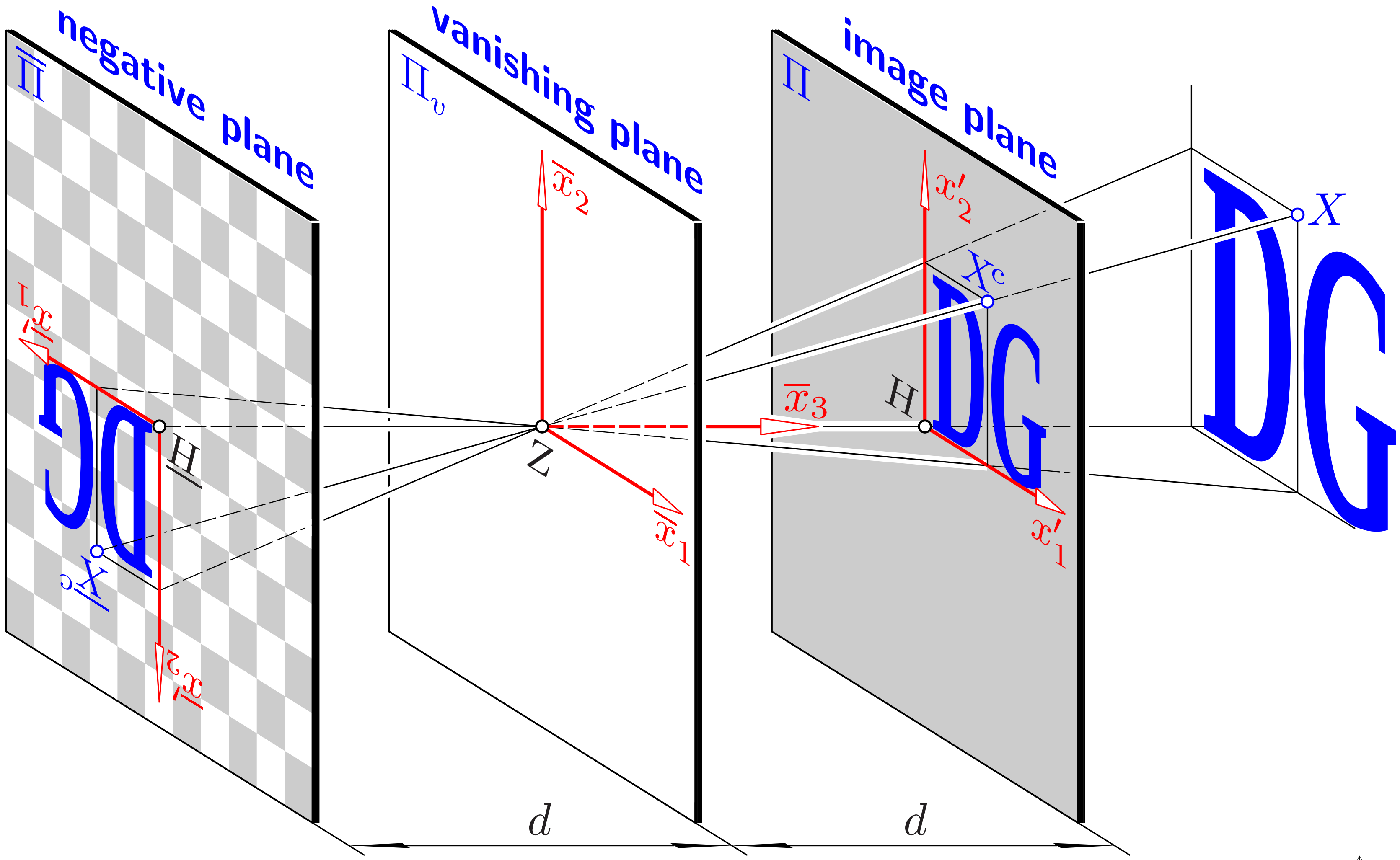
The center $Z$ lies at the intersection of the three spheres drawn over the segments $U_1^c U_2^c, \ldots$

---

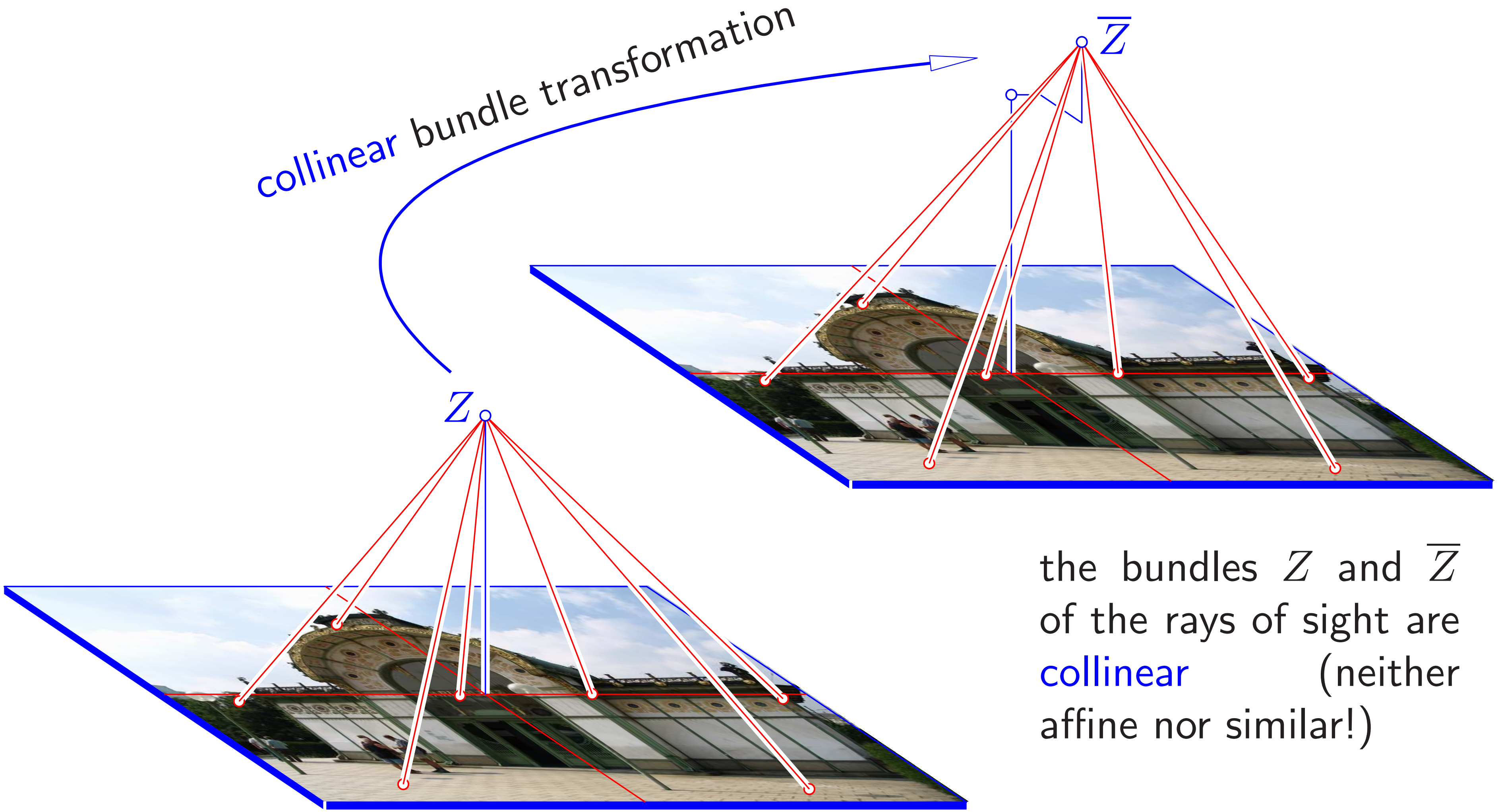
Note: Zooming and (automatic) focussing changes the focal distance $d$ !



# Positive and negative central pespective



# unknown interior calibration parameters




the bundles $Z$ and $\overline{Z}$ of the rays of sight are collinear (neither affine nor similar!)

# 2. Geometry of two images

GIVEN: Two linear images or two photographs showing the same object.

WANTED: Dimensions of the depicted 3D-object.

Historical 'Stadtbahn' station Karlsplatz in Vienna (Otto Wagner, 1897)

# 2. Geometry of two images

The geometry of two images is a classical subject of **Descriptive Geometry**. Its results have become standard (Finsterwalder, Kruppa, Krames, Wunderlich, Hohenberg, Tschupik, Brauner, Havlicek, H.S., ...).

**Photogrammetry** (Remote sensing) deals with the practical usage of these results.

---

**Why now ?** Advantages of digital images:

- less distorsion, because no paper prints are needed,
- exact boundary is available, and
- precise coordinate measurements are possible using standard software.

# 2. Geometry of two images

The geometry of two images is a classical subject of **Descriptive Geometry**. Its results have become standard (FINSTERWALDER, KRUPPA, KRAMES, WUNDERLICH, HOHENBERG, TSCHUPIK, BRAUNER, HAVLICEK, H.S., . . . ).

**Photogrammetry** (Remote sensing) deals with the practical usage of these results.

---

**Why now ?** Advantages of digital images:

- less distortion, because no paper prints are needed,
- exact boundary is available, and
- precise coordinate measurements are possible using standard software.

# Computer Vision

**Why now ?**

The geometry of two images is important for Computer Vision, a topic with the main goal *to endow a computer with a sense of vision*.

---

Basic problems:

- Which information can be extracted from digital images ?
- How to preprocess and represent this information ?

Sensor-guided robots, automatic vehicle control, 'Big Brother', . . .

# Computer Vision

**Why now ?**

The geometry of two images is important for Computer Vision, a topic with the main goal *to endow a computer with a sense of vision*.

---

Basic problems:

- Which information can be extracted from digital images ?
- How to preprocess and represent this information ?

Sensor-guided robots, automatic vehicle control, ‘Big Brother’, . . .

# Computer Vision

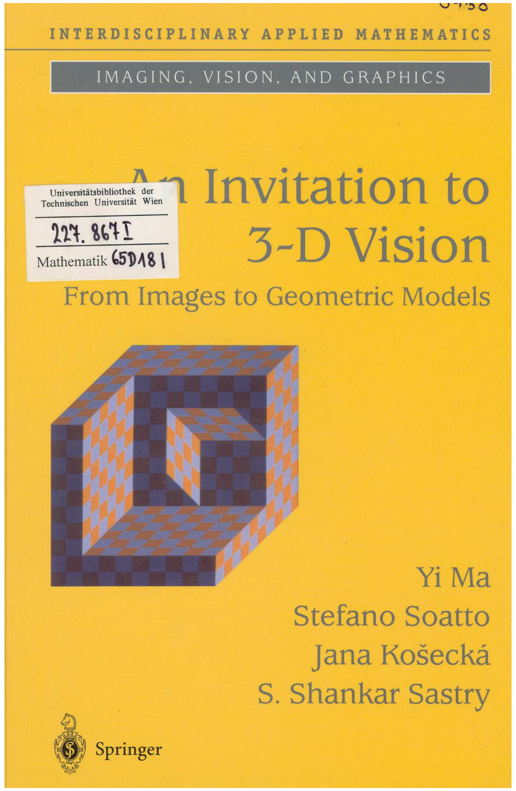
Recent textbooks:

Yi Ma, St. Soatto, J. Košecká, S.S. Sastry: *An Invitation to 3-D Vision.* Springer-Verlag, New York 2004

R. Hartley, A. Zisserman: *Multiple View Geometry in Computer Vision.* Cambridge University Press 2000

---

Fortunately the authors in the cited book refer to some of these standard results (Krames, Kruppa, Wunderlich)



# Computer Vision

Recent textbooks:

Yi MA, St. SOATTO, J. KOŠECKÁ, S.S. SASTRY: *An Invitation to 3-D Vision.* Springer-Verlag, New York 2004

R. HARTLEY, A. ZISSERMAN: *Multiple View Geometry in Computer Vision.* Cambridge University Press 2000

Fortunately the authors in the cited book refer to some of these standard results (Krames, Kruppa, Wunderlich)

10 Chapter 1. Introduction

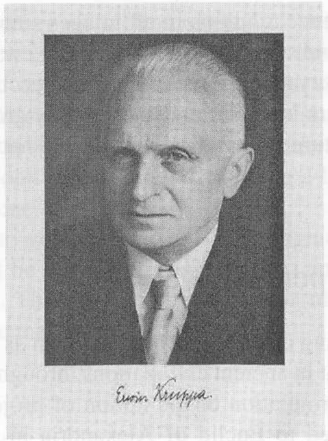



Figure 1.8. Photograph of Erwin Kruppa (courtesy of A. Kruppa).

The first work that is directly related to multiple-view geometry is believed to be a 1913 paper by the German mathematician Kruppa (see Figure 1.8).[7] He proved that two views of five points are sufficient to determine both the relative transformation between the views and the 3-D location of the points up to finitely many solutions.[8] Kruppa's proof was done in the traditional projective geometry setting [Kruppa, 1913].

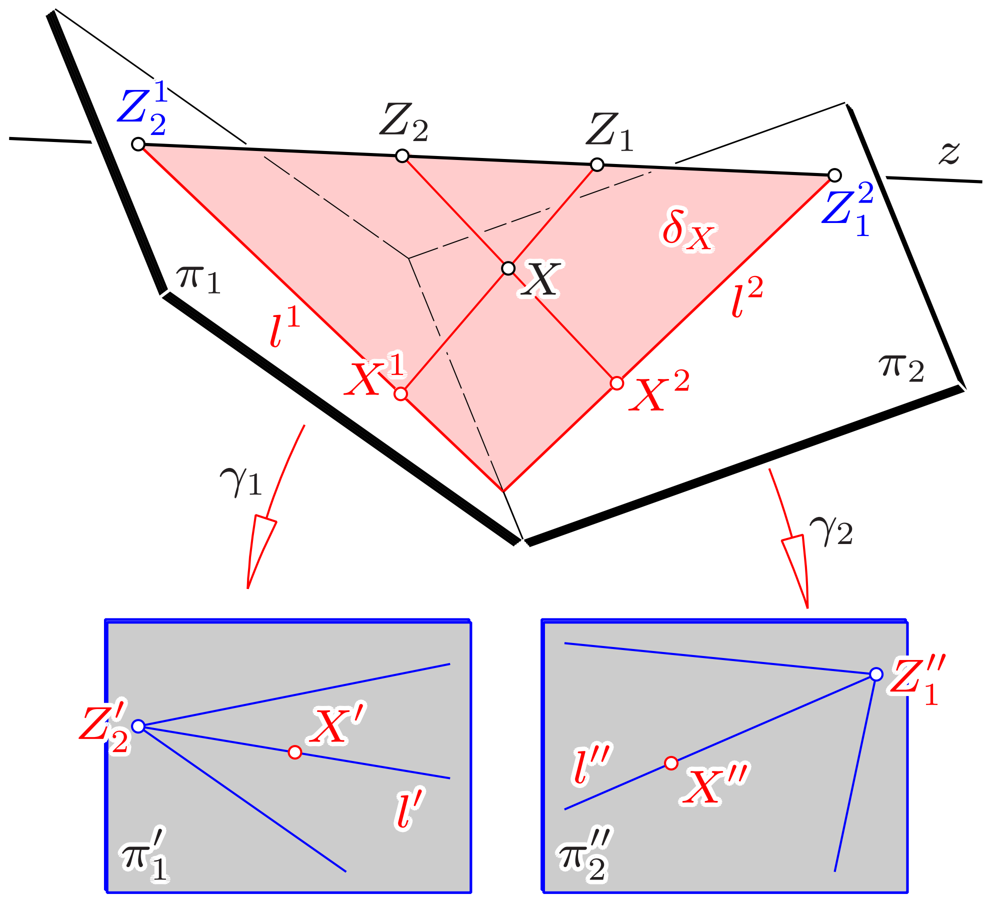
# Geometry of two images (epipolar geometry)

*viewing situation*
(two different centers $Z_1 \neq Z_2$)

two central projections together
with two collinear transformations

result in two images

$Z_1'' =$ 2nd image of the 1st center
$Z_2' =$ 1st image of the 2nd center



# Geometry of two images (epipolar geometry)

**Notations:**

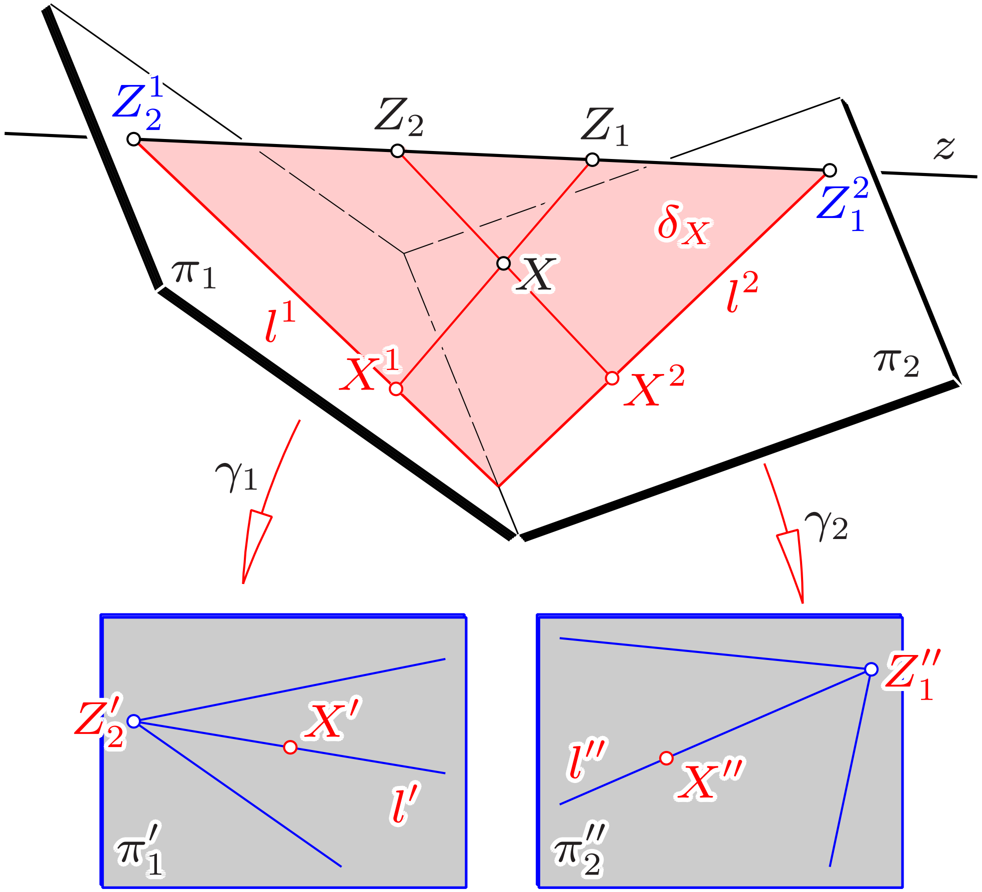
line $z = Z_1 Z_2$ . . . baseline,

$Z_2', Z_1''$ . . . epipoles
(German: Kernpunkte),

$\delta_X$ . . . epipolar plane (it is twice projecting),

$l', l''$ . . . pair of epipolar lines
(German: Kernstrahlen, Ordner)

$(X', X'')$ . . . corresponding views.



# Epipolar constraint

**Theorem** (synthetic version): For any two linear images of a scene, there is a projectivity between two line pencils

$$Z_2'(\delta_X') \barwedge Z_1''(\delta_X'')$$

such that the points $X', X''$ are corresponding $\iff$ they are located on (corresponding $=$) epipolar lines.

**Theorem** (analytic version): Using homogeneous coordinates for both images, there is a bilinear form $\beta$ of rank 2 such that two points $X' = \mathbf{x}'\mathbb{R} = (\xi_0' : \xi_1' : \xi_2')$ and $X'' = \mathbf{x}''\mathbb{R} = (\xi_0'' : \xi_1'' : \xi_2'')$ are corresponding

$$\iff \beta(\mathbf{x}', \mathbf{x}'') = \sum_{i,j=0}^{2} b_{ij}\, \xi_i'\, \xi_j'' = (\xi_0'\ \xi_1'\ \xi_2') \cdot (b_{ij}) \cdot \begin{pmatrix} \xi_0'' \\ \xi_1'' \\ \xi_2'' \end{pmatrix} = \mathbf{x}'^{T} \cdot \mathbf{B} \cdot \mathbf{x}'' = 0\,.$$

# Epipolar constraint

**Theorem** (synthetic version): For any two linear images of a scene, there is a projectivity between two line pencils

$$Z_2'(\delta_X') \barwedge Z_1''(\delta_X'')$$

such that the points $X', X''$ are corresponding $\iff$ they are located on (corresponding $=$) epipolar lines.

**Theorem** (analytic version): Using homogeneous coordinates for both images, there is a bilinear form $\beta$ of rank 2 such that two points $X' = \mathbf{x}'\mathbb{R} = (\xi_0' : \xi_1' : \xi_2')$ and $X'' = \mathbf{x}''\mathbb{R} = (\xi_0'' : \xi_1'' : \xi_2'')$ are corresponding

$$\iff \beta(\mathbf{x}', \mathbf{x}'') = \sum_{i,j=0}^{2} b_{ij}\, \xi_i'\, \xi_j'' = (\xi_0'\ \xi_1'\ \xi_2') \cdot (b_{ij}) \cdot \begin{pmatrix} \xi_0'' \\ \xi_1'' \\ \xi_2'' \end{pmatrix} = \mathbf{x}'^{\,T} \cdot \mathbf{B} \cdot \mathbf{x}'' = 0\,.$$

# Epipolar constraint

*Proof* (analytic version): Using homogeneous line coordinates, the projectivity between the line pencils can be expressed as

$$\beta\colon\ (\mathbf{u}_1'\lambda_1 + \mathbf{u}_2'\lambda_2)\mathbb{R} \ \mapsto\ (\mathbf{u}_1''\lambda_1 + \mathbf{u}_2''\lambda_2)\mathbb{R} \ \text{ for all } \ (\lambda_1, \lambda_2) \in \mathbb{R}^2 \setminus \{(0,0)\}.$$

$\mathbf{x}'$ and $\mathbf{x}''$ are corresponding $\iff$ there is a nontrivial pair $(\lambda_1, \lambda_2)$ such that

$$\begin{array}{rcl} (\mathbf{u}_1'\lambda_1 + \mathbf{u}_2'\lambda_2)\cdot \mathbf{x}' & = & 0 \\ (\mathbf{u}_1''\lambda_1 + \mathbf{u}_2''\lambda_2)\cdot \mathbf{x}'' & = & 0\,. \end{array}$$

These two linear homogeneous equations in the unknowns $(\lambda_1, \lambda_2)$ have a nontrivial solution $\iff$ the determinant vanishes, i.e.,

$$\beta(\mathbf{x}', \mathbf{x}'') := (\mathbf{u}_1'\cdot\mathbf{x}')(\mathbf{u}_2''\cdot\mathbf{x}'') - (\mathbf{u}_2'\cdot\mathbf{x}')(\mathbf{u}_1''\cdot\mathbf{x}'') = \textstyle\sum_{i,j=0}^{2} b_{ij}\, \xi_i'\, \xi_j'' = 0.$$

There are singular points of this correspondance: $Z_2'$ corresponds to all $X''$, and vice versa all points $X'$ correspond to $Z_1''$ $\implies$ $\mathrm{rk}(b_{ij}) = 2\,.$ $\square$

# Epipolar constraint

$$\beta(\mathbf{x}', \mathbf{x}'') = \sum_{i,j=0}^{2} b_{ij}\, \xi_i'\, \xi_j'' = (\xi_0' \;\, \xi_1' \;\, \xi_2') \cdot (b_{ij}) \cdot \begin{pmatrix} \xi_0'' \\ \xi_1'' \\ \xi_2'' \end{pmatrix} = \mathbf{x}'^{\,T} \cdot \mathbf{B} \cdot \mathbf{x}'' = 0\,.$$

The epipoles solve systems of homogeneous linear equations with the coefficient matrices $\mathbf{B}$ and $\mathbf{B}^\top$:

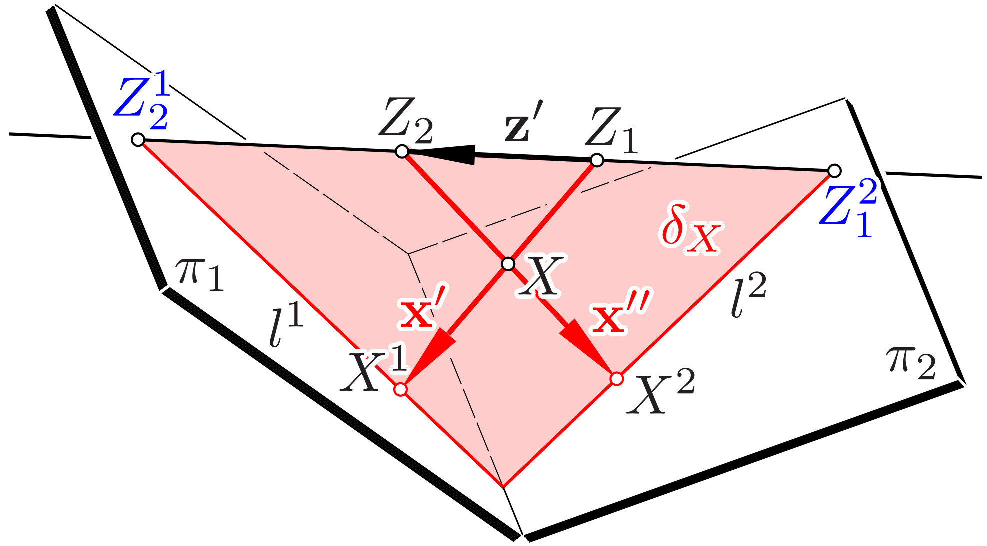
$Z_2' = (\xi_0' : \xi_1' : \xi_2')$ corresponds to all $X''$ $\iff$ $\sum_{i=0}^{2} b_{ij}\, \xi_i' = 0$ for $j = 0, 1, 2\,.$

$Z_1'' = (\xi_0'' : \xi_1'' : \xi_2'')$ corresponds to all $X'$ $\iff$ $\sum_{j=0}^{2} b_{ij}\, \xi_j'' = 0$ for $i = 0, 1, 2.$

# Epipolar constraint in the calibrated case

**Theorem:** *In the calibrated case the* *essential matrix* $\mathbf{B} = (b_{ij})$ *is the product of a* *skew symmetric* *matrix and an* *orthogonal* *one, i.e.,*

$$\mathbf{B} = \mathbf{S} \cdot \mathbf{R}\,.$$




*Proof:* We use both camera frames and the homogeneous coordinates

$$\mathbf{x}' = \overrightarrow{Z_1 X'}, \quad \mathbf{x}'' = \overrightarrow{Z_2 X''}.$$

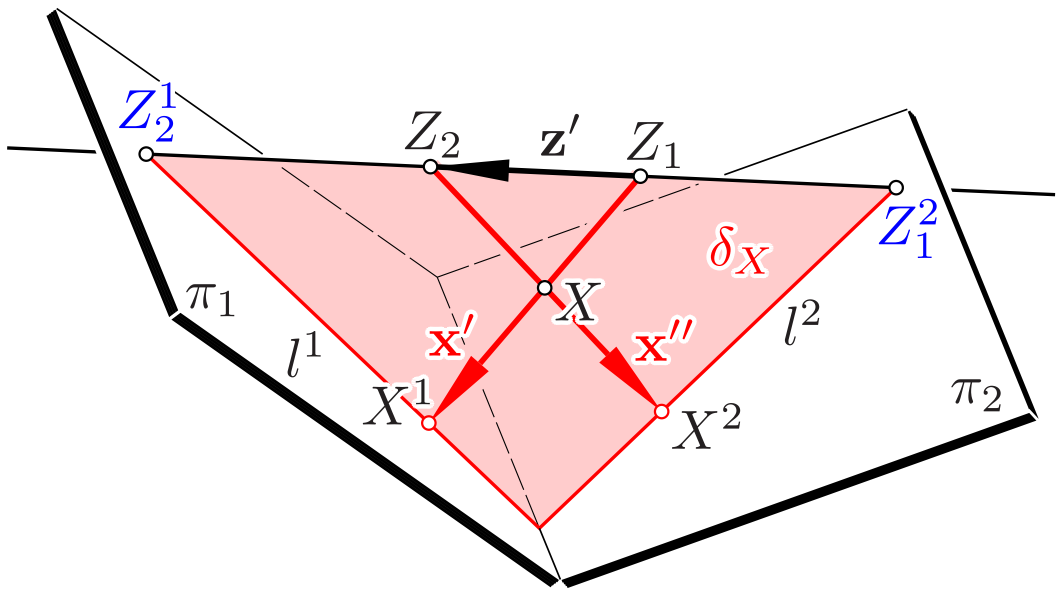
# Epipolar constraint in the calibrated case

For transforming the coordinates from the second camera frame into the first one, there is an orthogonal matrix $\mathbf{R}$ such that

$$\mathbf{x}_1'' = \mathbf{z}' + \mathbf{R}\cdot\mathbf{x}'' \text{ with } \mathbf{R}^\top = \mathbf{R}^{-1} \text{ and } \mathbf{z}' = (z_1',\, z_2',\, z_3')^\top = \overrightarrow{Z_1Z_2}.$$

The points $X^1$, $X^2$, $Z_1$, $Z_2$ are coplanar $\iff$ the triple product of the vectors $\mathbf{x}'$, $\mathbf{z}'$ and $\mathbf{x}_1'' = \overline{Z_1X^2}$ vanishes, i.e.,

$$\det(\mathbf{x}', \mathbf{z}', \mathbf{x}_1'') = \mathbf{x}'\cdot(\mathbf{z}'\times\mathbf{x}_1'') = 0\,.$$



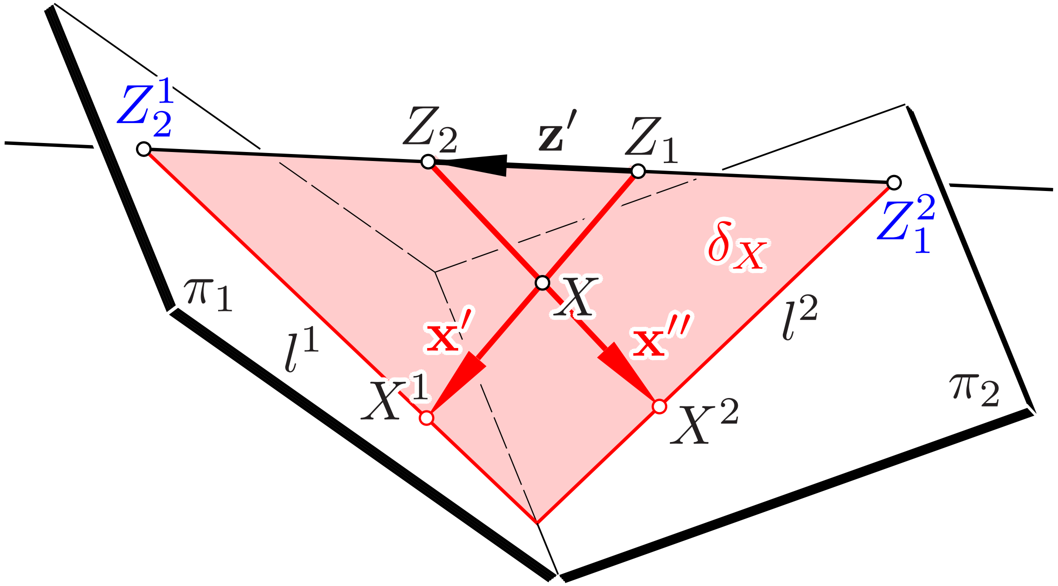
# Epipolar constraint in the calibrated case

For transforming the coordinates from the second camera frame into the first one, there is an orthogonal matrix $\mathbf{R}$ such that

$$\mathbf{x}_1'' = \mathbf{z}' + \mathbf{R}\cdot\mathbf{x}'' \text{ with } \mathbf{R}^\top = \mathbf{R}^{-1} \text{ and } \mathbf{z}' = (z_1', z_2', z_3')^\top = \overrightarrow{Z_1 Z_2}.$$

The points $X^1$, $X^2$, $Z_1$, $Z_2$ are coplanar $\iff$ the triple product of the vectors $\mathbf{x}'$, $\mathbf{z}'$ and $\mathbf{x}_1'' = \overline{Z_1 X^2}$ vanishes, i.e.,

$$\det(\mathbf{x}', \mathbf{z}', \mathbf{x}_1'') = \mathbf{x}'\cdot(\mathbf{z}'\times\mathbf{x}_1'') = 0\,.$$



# Epipolar constraint in the calibrated case

We replace the vector product $(\mathbf{z}' \times \mathbf{x}_1'')$ by

$$\mathbf{z}' \times (\mathbf{z}' + \mathbf{R} \cdot \mathbf{x}'') = \mathbf{z}' \times (\mathbf{R} \cdot \mathbf{x}'') = \mathbf{S} \cdot \mathbf{R} \cdot \mathbf{x}'' \quad \text{with} \quad \mathbf{S} = \begin{pmatrix} 0 & -z_3' & z_2' \\ z_3' & 0 & -z_1' \\ -z_2' & z_1' & 0 \end{pmatrix}.$$

Matrix $\mathbf{S}$ is skew symmetric and $\mathbf{R}$ is orthogonal.

Hence, the coplanarity of $\mathbf{x}'$, $\mathbf{x}''$ and $\mathbf{z}'$ is equivalent to

$$0 = \mathbf{x}' \cdot (\mathbf{z}' \times \mathbf{x}_1'') = \mathbf{x}'^T \cdot \underbrace{\mathbf{S} \cdot \mathbf{R}}_{\mathbf{B}} \cdot \mathbf{x}'', \quad \text{hence} \quad \mathbf{B} = \mathbf{S} \cdot \mathbf{R}\,. \qquad \square$$

The decomposition of the fundamental matrix $\mathbf{B}$ into these two factors defines the relative position of the second camera frame against the first one !

# Epipolar constraint in the calibrated case

We replace the vector product $(\mathbf{z}' \times \mathbf{x}_1'')$ by

$$\mathbf{z}' \times (\mathbf{z}' + \mathbf{R}\cdot\mathbf{x}'') = \mathbf{z}' \times (\mathbf{R}\cdot\mathbf{x}'') = \mathbf{S}\cdot\mathbf{R}\cdot\mathbf{x}'' \quad \text{with} \quad \mathbf{S} = \begin{pmatrix} 0 & -z_3' & z_2' \\ z_3' & 0 & -z_1' \\ -z_2' & z_1' & 0 \end{pmatrix}.$$

Matrix $\mathbf{S}$ is skew symmetric and $\mathbf{R}$ is orthogonal.

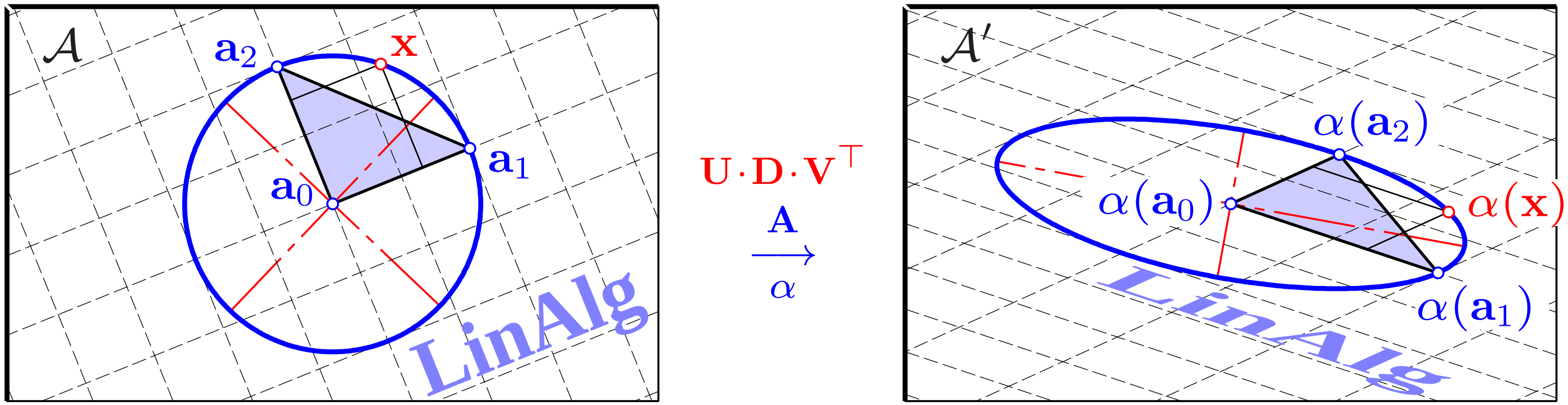
Hence, the coplanarity of $\mathbf{x}'$, $\mathbf{x}''$ and $\mathbf{z}'$ is equivalent to

$$0 = \mathbf{x}' \cdot (\mathbf{z}' \times \mathbf{x}_1'') = \mathbf{x}'^T \cdot \underbrace{\mathbf{S} \cdot \mathbf{R}}_{\mathbf{B}} \cdot \mathbf{x}'', \quad \text{hence} \quad \mathbf{B} = \mathbf{S} \cdot \mathbf{R}\,. \qquad \square$$

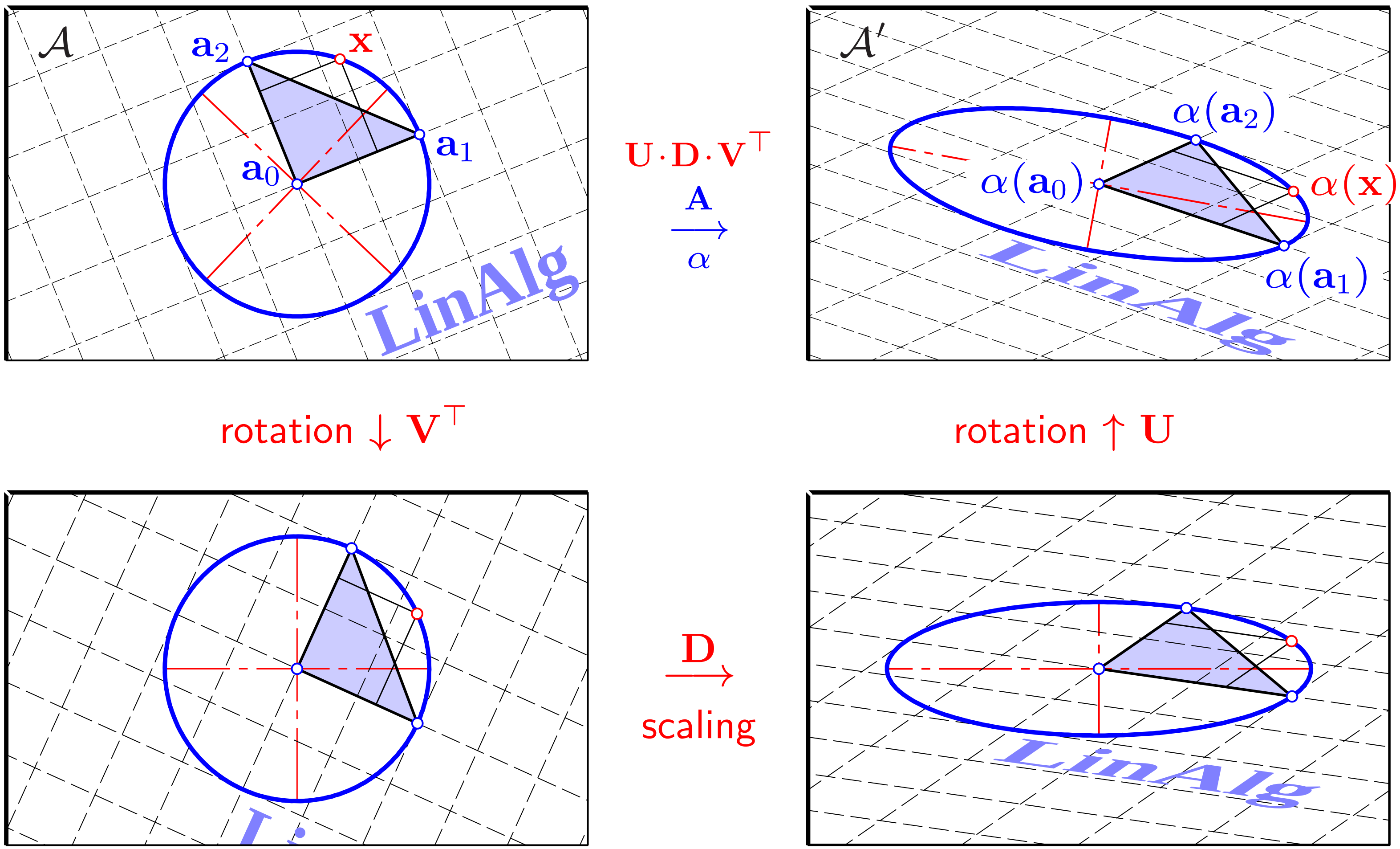
---

The decomposition of the fundamental matrix $\mathbf{B}$ into these two factors defines the relative position of the second camera frame against the first one !

# Singular value decomposition (SVD)



# Singular value decomposition (SVD)



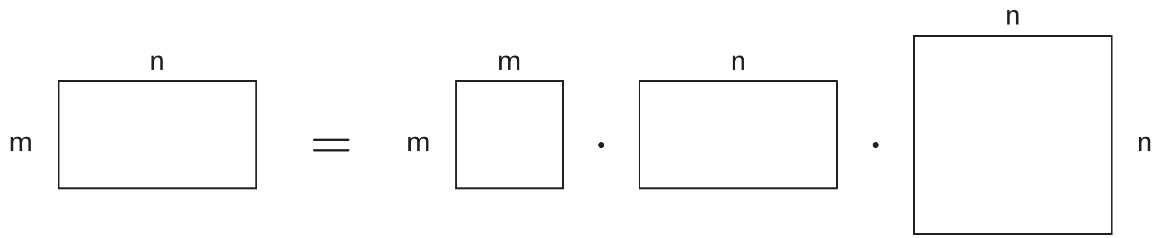
# Singular value decomposition (SVD)

**Theorem:** [Singular value decomposition]
*Any matrix* $\mathbf{A} \in M(m,n;\mathbb{R})$ *can be decomposed into a product*

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{V}^\top \text{ \textit{with orthogonal} } \mathbf{U}, \mathbf{V} \text{ \textit{and} } \mathbf{D} = \operatorname{diag}(\sigma_1, \ldots, \sigma_p)$$

*with* $\mathbf{D} \in M(m,n;\mathbb{R})$, $\sigma_i \geq 0$, *and* $p = \min\{m,n\}$.
The positive entries in the main diagonal of $\mathbf{D}$ are called singular values of $\mathbf{A}$.

The squares of the singular values are the non-zero eigenvalues of $\mathbf{A}^\top \cdot \mathbf{A}$.




$$\mathbf{A} \quad = \quad \mathbf{U}^\top \quad \cdot \quad \mathbf{D} \quad \cdot \quad \mathbf{V}$$

# Singular value decomposition (SVD)

**Theorem:** [Singular value decomposition]
*Any matrix* $\mathbf{A} \in M(m, n; \mathbb{R})$ *can be decomposed into a product*

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{V}^\top \text{ with orthogonal } \mathbf{U}, \mathbf{V} \text{ and } \mathbf{D} = \operatorname{diag}(\sigma_1, \ldots, \sigma_p)$$

*with* $\mathbf{D} \in M(m, n; \mathbb{R})$, $\sigma_i \geq 0$, *and* $p = \min\{m, n\}$.
The positive entries in the main diagonal of $\mathbf{D}$ are called singular values of $\mathbf{A}$.

The singular values of $\mathbf{A}$ can be seen as non-vanishing principal distortion factors of the affine transformation represented by $\mathbf{A}$, i.e., the semiaxes of the affine image of the unit sphere.

E.g., the singular values of an orthogonal projection are $(1, 1)$ as the unit sphere is mapped onto a unit disk.

# Singular values of the essential matrix

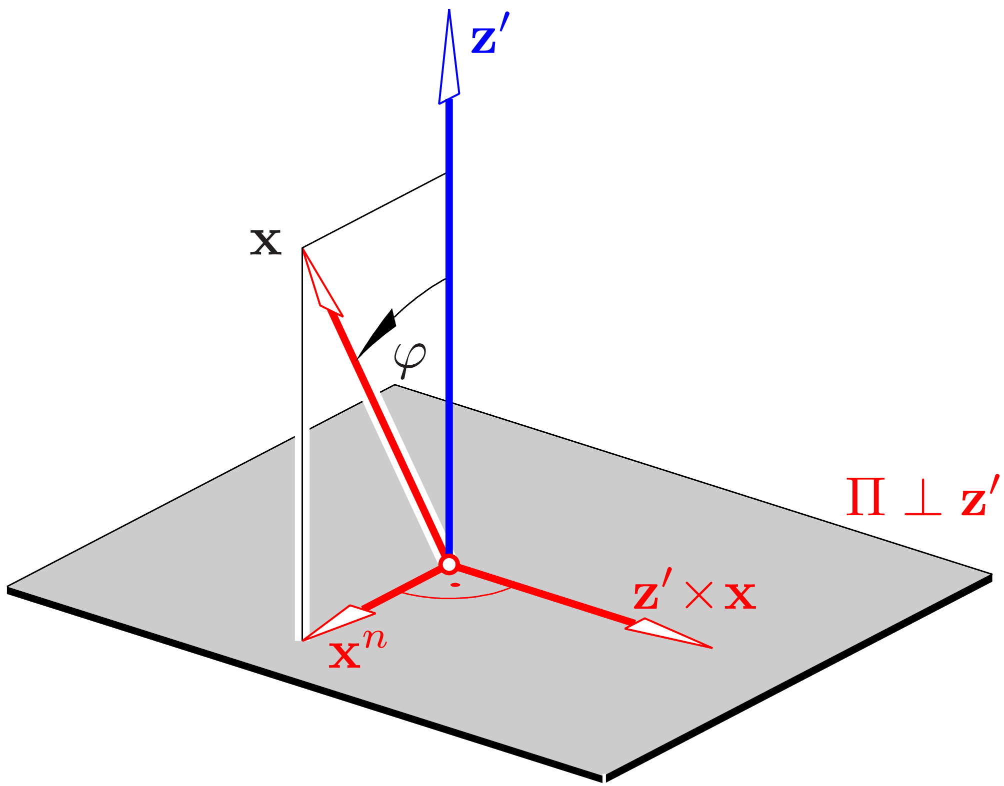
**Theorem:** In the calibrated case the essential matrix $\mathbf{B}$ has two equal singular values $\sigma := \sigma_1 = \sigma_2$.

*Proof:* We have $\mathbf{B} = \mathbf{S} \cdot \mathbf{R}$ with orthogonal $\mathbf{R}$. The vector

$$\mathbf{S} \cdot \mathbf{x} = \mathbf{z}' \times \mathbf{x}$$

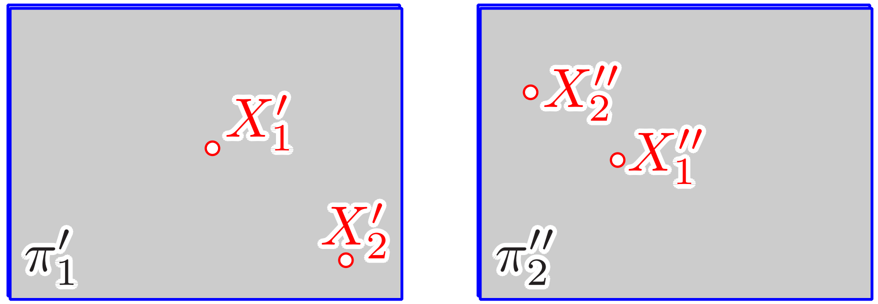
is orthogonal zu the orthogonal view $\mathbf{x}^n$, where

$$\begin{aligned} \|\mathbf{z}' \times \mathbf{x}\| &= |\sin\varphi| \, \|\mathbf{x}\| \, \|\mathbf{z}'\| = \\ &= \|\mathbf{x}^n\| \, \|\mathbf{z}'\| = \sigma \, \|\mathbf{x}^n\|. \end{aligned}$$



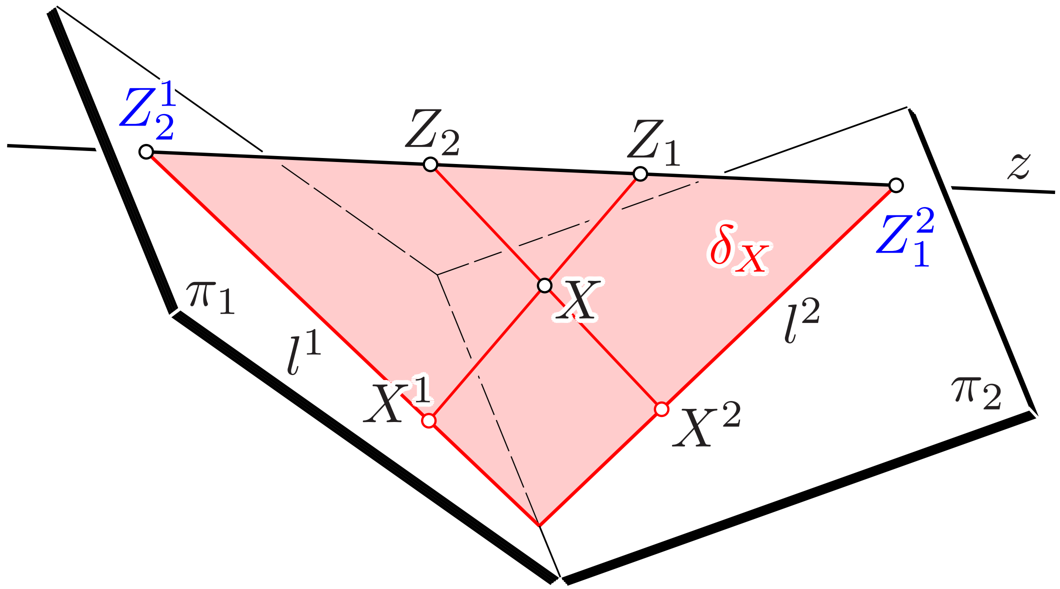
# What means 'reconstruction'

GIVEN: Two either calibrated or uncalibrated images.




WANTED: 'viewing situation', i.e., determine

- the relative position of the two camera frames, and
- for each pair $(X', X'')$ of images the location of the orginal space point $X$.

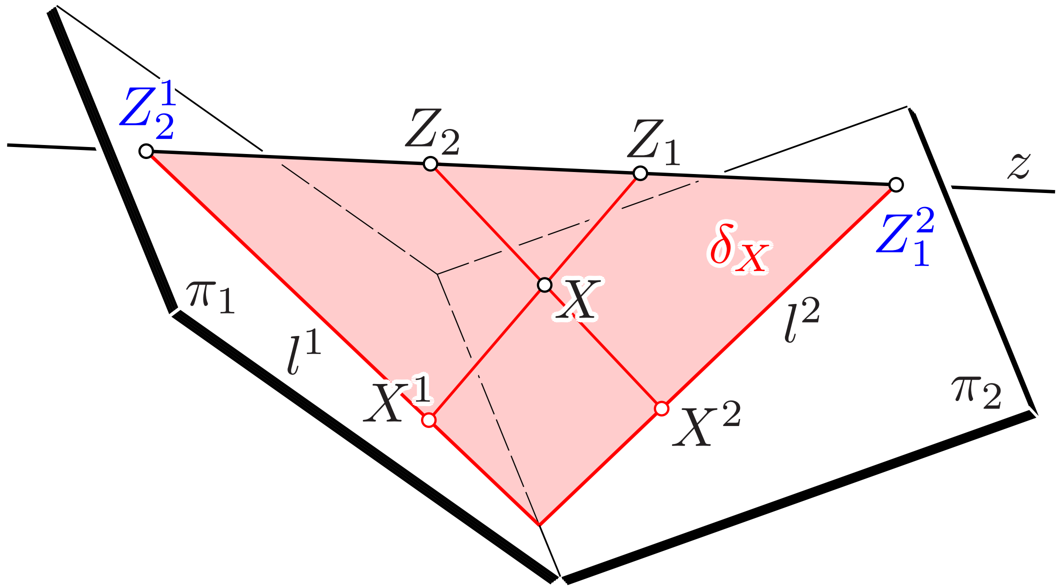


# First fundamental theorem

**Theorem 1:**
From two uncalibrated images with given projectivity between epipolar lines the depicted object can be reconstructed up to a collinear transformation.

*Sketch of the proof:*
The two images can be placed in space such that pairs of epipolar lines are intersecting. Then for arbitrary $Z_1, Z_2$ on the baseline $z = Z_1^2 Z_2^1$ there is a reconstructed 3D object.

Any other choice of the viewing situation gives a collinear transform of the 3D object. $\square$

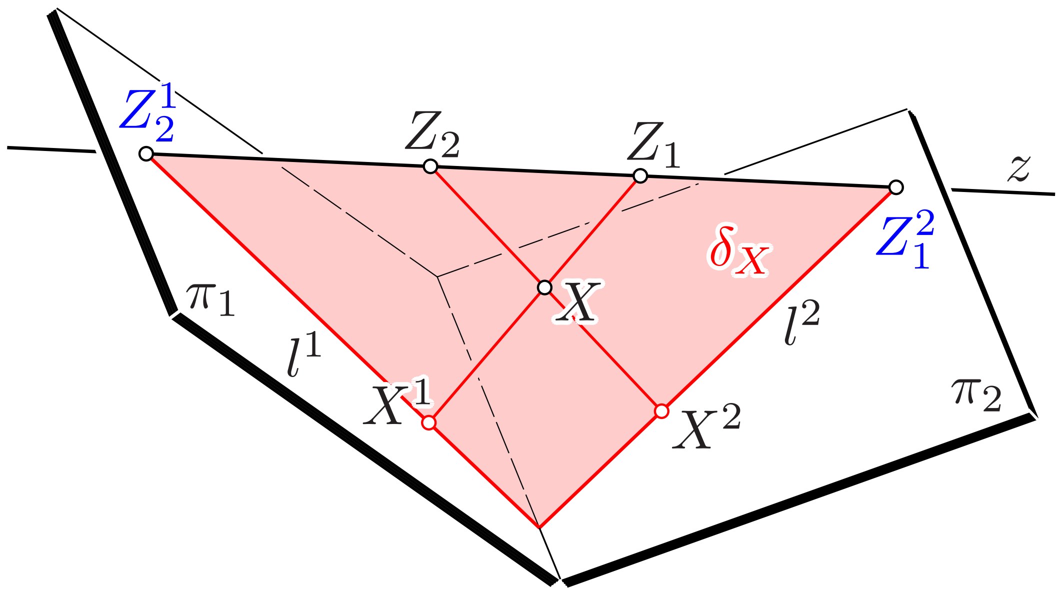


# Second fundamental theorem

**Theorem 2** (S. Finsterwalder, 1899):
From two calibrated images with given projectivity between epipolar lines the depicted object can be reconstructed up to a similarity.

*Sketch of the proof:*
Now in the two bundles of rays the pencils of epipolar planes $\delta_X$ are congruent, and they can be made coincident by a rigid motion. Then relative to the first bundle $Z_1$ for any $Z_2 \in z$ there is a reconstructed 3D object.

Any other choice of $Z_2$ gives a similar 3D object. □



# Aerial photographs

**Historical remarks:**

• High strategic importance of 'aerial photogrammetry', already from World War I on.

• Mechanical devices ('stereo comparators') were developed for the reconstruction from a pair of photos.

• Now, in the time of GPS, the exterior calibration parameters are always available with rather high precision available. Hence numerical methods are preferred.

Orthophoto, a geometrically corrected aerial photograph (cadastral boundaries)

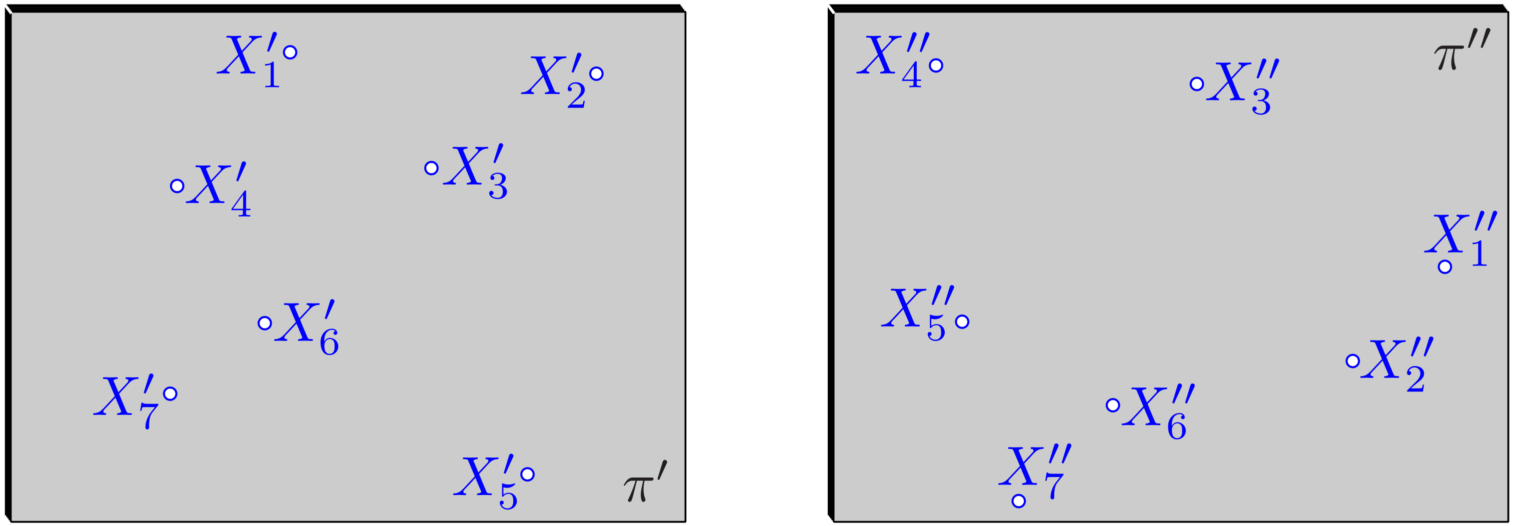
# 3. Numerical reconstruction from two images

**Problem of Projectivity:**

GIVEN: **7** pairs of corresponding points $(X_1', X_1''), \ldots, (X_7', X_7'')$.

WANTED: A pair of points $(S', S'')$ (= epipoles) such that there is a projectivity

$$S'([S'X_1'], \ldots, [S'X_7']) \barwedge S''([S'X_1''], \ldots, [S''X_7'']).$$



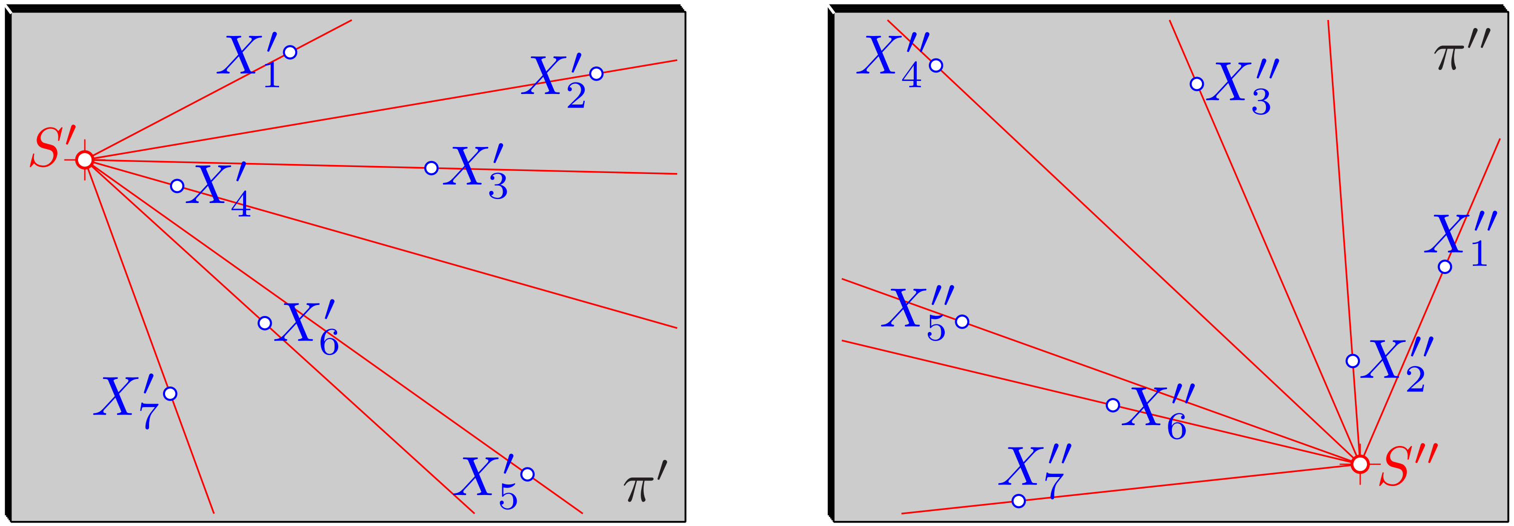
# Determination of epipoles — geometric meaning

**Problem of Projectivity:**

GIVEN: **7** pairs of corresponding points $(X_1', X_1''), \ldots, (X_7', X_7'')$.

WANTED: A pair of points $(S', S'')$ (= epipoles) such that there is a projectivity

$$S'([S'X_1'], \ldots, [S'X_7']) \;\barwedge\; S''([S'X_1''], \ldots, [S''X_7'']).$$



# Determination of epipoles — analytic solution

**Theorem:** If **7** pairs of corresponding points $(X_1', X_1''), \dots, (X_7', X_7'')$ ("control points") are given, the determination of the epipoles is a cubic problem.

*Proof:* 7 pairs of corresponding points give 7 linear homogeneous equations

$$\beta(\mathbf{x}_i', \mathbf{x}_i'') = \mathbf{x}_i^{\top} \cdot \mathbf{B} \cdot \mathbf{x}_i'' = 0, \quad i = 1, \dots, 7,$$

for the 9 entries in the $(3\times 3)$-matrix $\mathbf{B} = (b_{ij})$ — called essential matrix.

$\det(b_{ij}) = 0$ gives an additional cubic equation which fixes all $b_{ij}$ up to a common factor. $\square$

For noisy image points one should use more than 7 control points and then methods of least square approximation for obtaining the *'best fitting matrix'* $\mathbf{B}$:

# Determination of epipoles — analytic solution

**Theorem:** If **7** pairs of corresponding points $(X_1', X_1''), \ldots, (X_7', X_7'')$ ("control points") are given, the determination of the epipoles is a cubic problem.

*Proof:* 7 pairs of corresponding points give 7 linear homogeneous equations

$$\beta(\mathbf{x}_i', \mathbf{x}_i'') = \mathbf{x}_i'^{\top} \cdot \mathbf{B} \cdot \mathbf{x}_i'' = 0, \quad i = 1, \ldots, 7,$$

for the 9 entries in the $(3\times 3)$-matrix $\mathbf{B} = (b_{ij})$ — called essential matrix.

$\det(b_{ij}) = 0$ gives an additional cubic equation which fixes all $b_{ij}$ up to a common factor. $\square$

For noisy image points one should use more than 7 control points and then methods of least square approximation for obtaining the *'best fitting matrix'* $\mathbf{B}$*:*

# Determination of epipoles — analytic solution

**1)** Let $\mathbf{A}$ denote the coefficient matrix in the linear homogeneous system for the entries of $\mathbf{B}$. Then the 'least square fit' for this overdetermined system is an eigenvector for the smallest eigenvalue of the symmetric matrix $\mathbf{A}^\top \cdot \mathbf{A}$ (= smallest singular value of $\mathbf{A}$).

---

**2)** Since the essential matrix must have rank 2, we use the 'projection into the essential space'. This means, the singular value decomposition of $\mathbf{B}$ gives a representation

$$\mathbf{B} = \mathbf{U} \cdot \mathrm{diag}(\sigma_1,\, \sigma_2,\, \sigma_3) \cdot \mathbf{V}^\top \text{ with orthogonal } \mathbf{U}, \mathbf{V} \text{ and } \sigma_1 \geq \sigma_2 \geq \sigma_3\,.$$

Then in the uncalibrated case $\mathbf{B} = \mathbf{U} \cdot \mathrm{diag}(\sigma_1, \sigma_2, 0) \cdot \mathbf{V}^\top$ is optimal (with respect to the Frobenius norm) and in the calibrated case

$$\mathbf{B} = \mathbf{U} \cdot \mathrm{diag}(\sigma, \sigma, 0) \cdot \mathbf{V}^\top \quad \text{with} \quad \sigma = (\sigma_1 + \sigma_2)/2.$$

# Determination of epipoles — analytic solution

**1)** Let $\mathbf{A}$ denote the coefficient matrix in the linear homogeneous system for the entries of $\mathbf{B}$. Then the 'least square fit' for this overdetermined system is an eigenvector for the smallest eigenvalue of the symmetric matrix $\mathbf{A}^\top \cdot \mathbf{A}$ (= smallest singular value of $\mathbf{A}$).

---

**2)** Since the essential matrix must have rank 2, we use the 'projection into the essential space'. This means, the singular value decomposition of $\mathbf{B}$ gives a representation

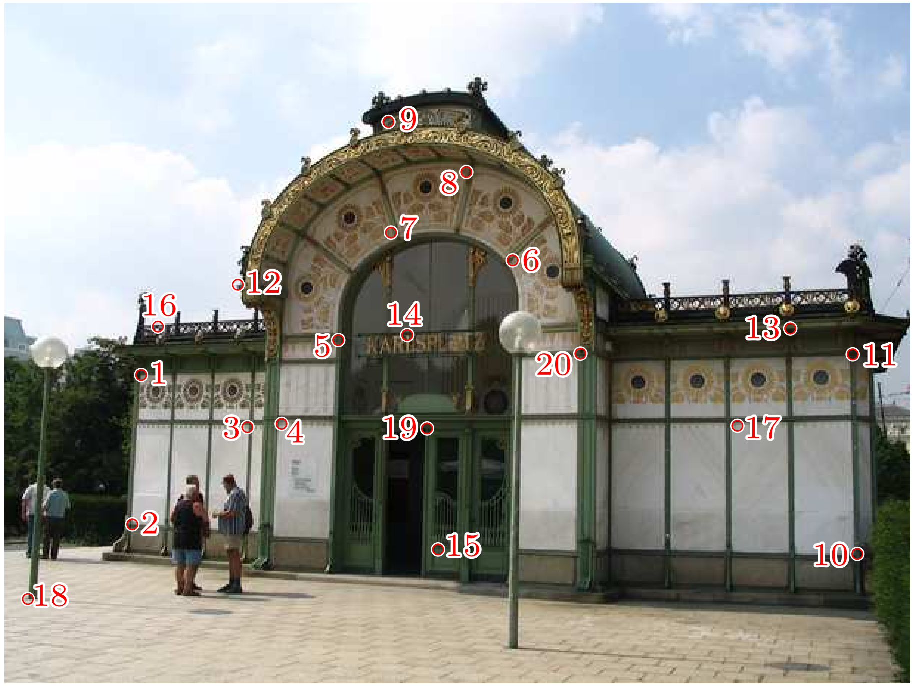
$$\mathbf{B} = \mathbf{U} \cdot \mathrm{diag}(\sigma_1,\, \sigma_2,\, \sigma_3) \cdot \mathbf{V}^\top \text{ with orthogonal } \mathbf{U}, \mathbf{V} \text{ and } \sigma_1 \geq \sigma_2 \geq \sigma_3\,.$$

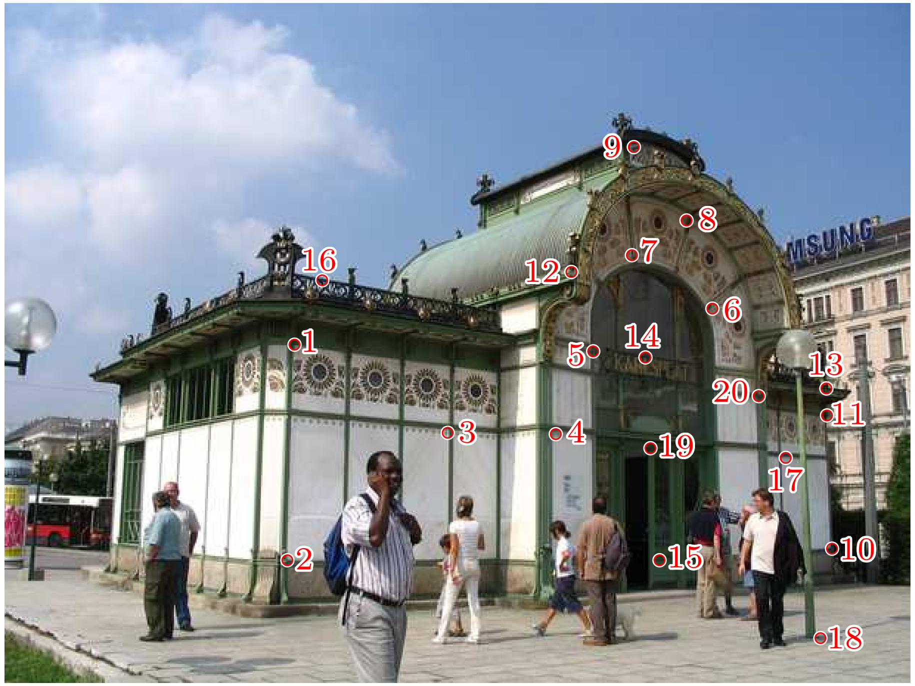
Then in the uncalibrated case $\mathbf{B} = \mathbf{U} \cdot \mathrm{diag}(\sigma_1, \sigma_2, 0) \cdot \mathbf{V}^\top$ is optimal (with respect to the Frobenius norm) and in the calibrated case

$$\mathbf{B} = \mathbf{U} \cdot \mathrm{diag}(\sigma, \sigma, 0) \cdot \mathbf{V}^\top \quad \text{with} \quad \sigma = (\sigma_1 + \sigma_2)/2.$$

# 3. Numerical reconstruction of two images

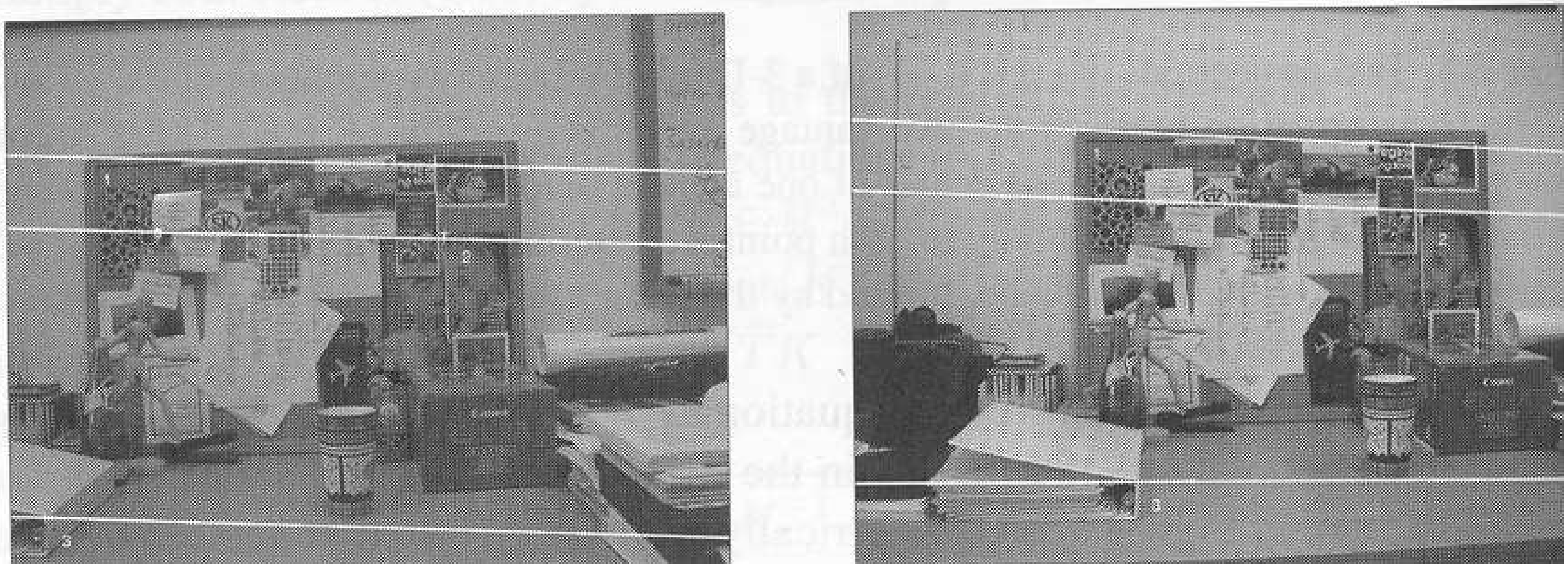
**Step 1:** Specify at least 7 reference points




. . . manually — or automatically by methods of pattern recognition

# Step 2: Compute the essential matrix



Two images with depicted epipolar lines [see J. Košecká et al.]

Remark: In the calibrated case $\geq 5$ pairs of corresponding points are needed, since in the decomposition $\mathbf{B} = \mathbf{S} \cdot \mathbf{R}$ each factor depends on 3 parameters only. Because of the homogeneity only 5 unknowns are essential.

# Step 2: Compute the essential matrix

**Step 2:** Compute the essential matrix $\mathbf{B}$ — including the pairs of epipolar lines

# Step 3: Factorize $\mathbf{B} = \mathbf{S}\cdot\mathbf{R}$

**Theorem:** There are exactly two ways of decomposing $\mathbf{B} = \mathbf{U}\cdot\mathbf{D}\cdot\mathbf{V}^\top$ with $\mathbf{D} = \mathrm{diag}(\sigma,\sigma,0)$ into a product $\mathbf{S}\cdot\mathbf{R}$ with skew-symmetric $\mathbf{S}$ and orthogonal $\mathbf{R}$:

$$\mathbf{S} = \pm\mathbf{U}\cdot\mathbf{R}_+\cdot\mathbf{D}\cdot\mathbf{U}^\top \quad \text{and} \quad \mathbf{R} = \pm\mathbf{U}\cdot\mathbf{R}_+^\top\cdot\mathbf{V}^\top \quad \text{with} \quad \mathbf{R}_+ = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

*Proof:*

a) It is sufficient to factorize $\mathbf{U}\cdot\mathbf{D} = \mathbf{S}\cdot\mathbf{R}'$ which implies $\mathbf{B} = \mathbf{S}\cdot(\mathbf{R}'\cdot\mathbf{V}^\top)$, i.e., $\mathbf{R} = \mathbf{R}'\cdot\mathbf{V}^\top$.

b) $\mathbf{D}$ represents the product of the orthogonal projection into the $x_1x_2$-plane and the scaling with factor $\sigma$. The rotation $\mathbf{U}$ transforms the $x_1x_2$-plane into the image plane of $\mathbf{U}\cdot\mathbf{D}$.

# Step 3: Factorize $\mathbf{B} = \mathbf{S}\cdot\mathbf{R}$

**Theorem:** There are exactly two ways of decomposing $\mathbf{B} = \mathbf{U}\cdot\mathbf{D}\cdot\mathbf{V}^\top$ with $\mathbf{D} = \mathrm{diag}(\sigma, \sigma, 0)$ into a product $\mathbf{S}\cdot\mathbf{R}$ with skew-symmetric $\mathbf{S}$ and orthogonal $\mathbf{R}$:

$$\mathbf{S} = \pm\mathbf{U}\cdot\mathbf{R}_+\cdot\mathbf{D}\cdot\mathbf{U}^\top \quad \text{and} \quad \mathbf{R} = \pm\mathbf{U}\cdot\mathbf{R}_+^\top\cdot\mathbf{V}^\top \quad \text{with} \quad \mathbf{R}_+ = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

*Proof:*

a) It is sufficient to factorize $\mathbf{U}\cdot\mathbf{D} = \mathbf{S}\cdot\mathbf{R}'$ which implies $\mathbf{B} = \mathbf{S}\cdot(\mathbf{R}'\cdot\mathbf{V}^\top)$, i.e., $\mathbf{R} = \mathbf{R}'\cdot\mathbf{V}^\top$.

b) $\mathbf{D}$ represents the product of the orthogonal projection into the $x_1x_2$-plane and the scaling with factor $\sigma$. The rotation $\mathbf{U}$ transforms the $x_1x_2$-plane into the image plane of $\mathbf{U}\cdot\mathbf{D}$.

# Step 3: Factorize $\mathbf{B} = \mathbf{S} \cdot \mathbf{R}$

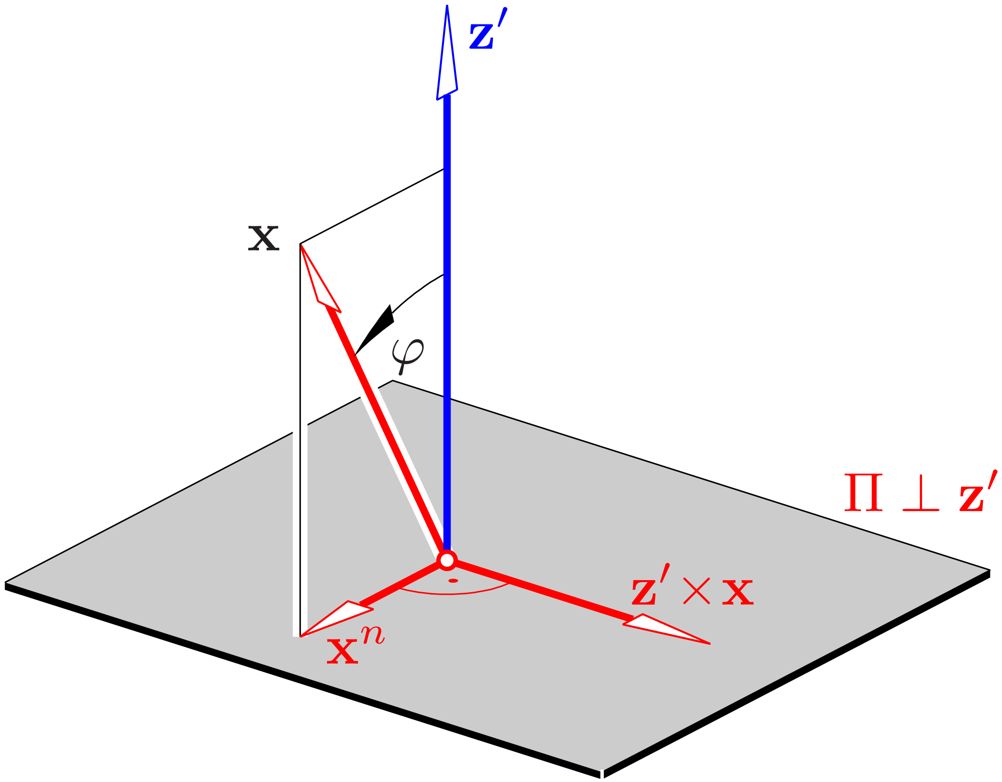
c) In the case

$$\mathbf{S} = \begin{pmatrix} 0 & -z_3' & z_2' \\ z_3' & 0 & -z_1' \\ -z_2' & z_1' & 0 \end{pmatrix}$$

we have

$$\mathbf{S}\,\mathbf{x} = \mathbf{z}' \times \mathbf{x} \quad \text{for} \quad \mathbf{z}' = \begin{pmatrix} z_1' \\ z_2' \\ z_3' \end{pmatrix}.$$

Hence, the skew symmetric matrix $\mathbf{S}$ represents the product of an orthogonal projection parallel to $\mathbf{z}'$, a $90°$-rotation about $\mathbf{z}'$ and a scaling with factor $\|\mathbf{z}'\|$.



# Step 3: Factorize $\mathbf{B} = \mathbf{S}\cdot\mathbf{R}$

d) Matrix $\mathbf{R}_+ = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ is orthogonal, $\mathbf{R}_+ \cdot \mathbf{D} = \begin{pmatrix} 0 & -\sigma & 0 \\ \sigma & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is skew-symmetric with $\mathbf{z}' = (0, 0, \sigma)$. We transform it by $\mathbf{U}$ to obtain the required position, i.e., $\mathbf{S} = \pm\mathbf{U}\cdot(\mathbf{R}_+\cdot\mathbf{D})\cdot\mathbf{U}^\top$.

$\mathbf{R}_+$ commutes with $\mathbf{D}$ $\implies$ $\mathbf{U}\cdot\mathbf{D} = \mathbf{U}\cdot\mathbf{D}\cdot\mathbf{R}_+\cdot\mathbf{U}^\top\cdot\mathbf{U}\cdot\mathbf{R}_+^\top =$

$$= \underbrace{\left[\pm\mathbf{U}\cdot\mathbf{R}_+\cdot\mathbf{D}\cdot\mathbf{U}^\top\right]}_{\mathbf{S}}\cdot\underbrace{\left[\pm\mathbf{U}\cdot\mathbf{R}_+^\top\right]}_{\mathbf{R}'}.$$

e) $\mathbf{B}$ represents an orthogonal axonometry; its column vectors are images of an orthonormal frame. We know from Descriptive Geometry that apart from translations there are not more than two different frames with given images. $\square$

# Step 3: Factorize $\mathbf{B} = \mathbf{S}\cdot\mathbf{R}$

d) Matrix $\mathbf{R}_+ = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ is orthogonal, $\mathbf{R}_+ \cdot \mathbf{D} = \begin{pmatrix} 0 & -\sigma & 0 \\ \sigma & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is skew-symmetric with $\mathbf{z}' = (0, 0, \sigma)$. We transform it by $\mathbf{U}$ to obtain the required position, i.e., $\mathbf{S} = \pm\mathbf{U}\cdot(\mathbf{R}_+\cdot\mathbf{D})\cdot\mathbf{U}^\top$.

$\mathbf{R}_+$ commutes with $\mathbf{D}$ $\implies$ $\mathbf{U}\cdot\mathbf{D} = \mathbf{U}\cdot\mathbf{D}\cdot\mathbf{R}_+\cdot\mathbf{U}^\top\cdot\mathbf{U}\cdot\mathbf{R}_+^\top =$

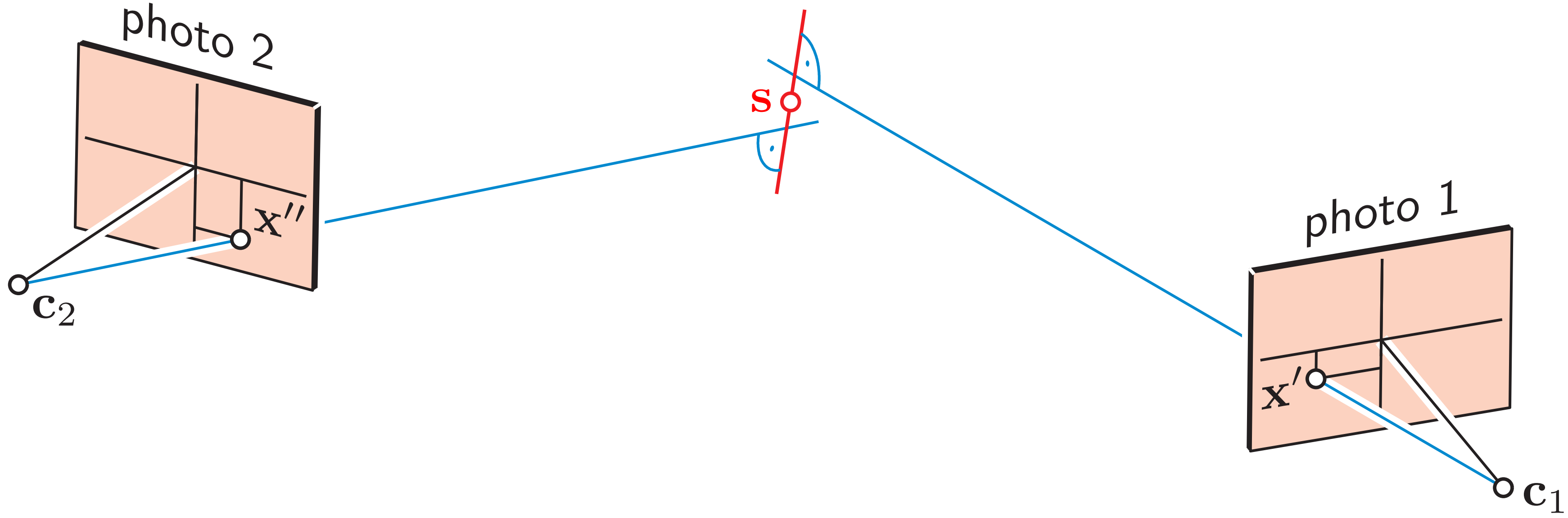
$$= \underbrace{\left[\pm\mathbf{U}\cdot\mathbf{R}_+\cdot\mathbf{D}\cdot\mathbf{U}^\top\right]}_{\mathbf{S}} \cdot \underbrace{\left[\pm\mathbf{U}\cdot\mathbf{R}_+^\top\right]}_{\mathbf{R}'}.$$

---

e) $\mathbf{B}$ represents an orthogonal axonometry; its column vectors are images of an orthonormal frame. We know from Descriptive Geometry that apart from translations there are not more than two different frames with given images. $\square$

# Step 4: Intersecting corresponding rays

In one of the frames compute the approximate point of intersection between corresponding rays.




For the center of the common perpendicular line segment the sum of squared distances is minimal.

# Summary of algorithm

1) Specify $n > 7$ pairs $(X_i', X_i'')$, $i = 1, \ldots, n$.

2) Set up linear system of equations for the essential matrix $\mathbf{B}$ and seek best fitting matrix (eigenvector of the smallest eigenvalue).

3) Compute the closest rank 2 matrix $\mathbf{B}$ with two equal singular values.

4) Factorize $\mathbf{B} = \mathbf{S} \cdot \mathbf{R}$ ; this reveals the relative position of the two camera frames.

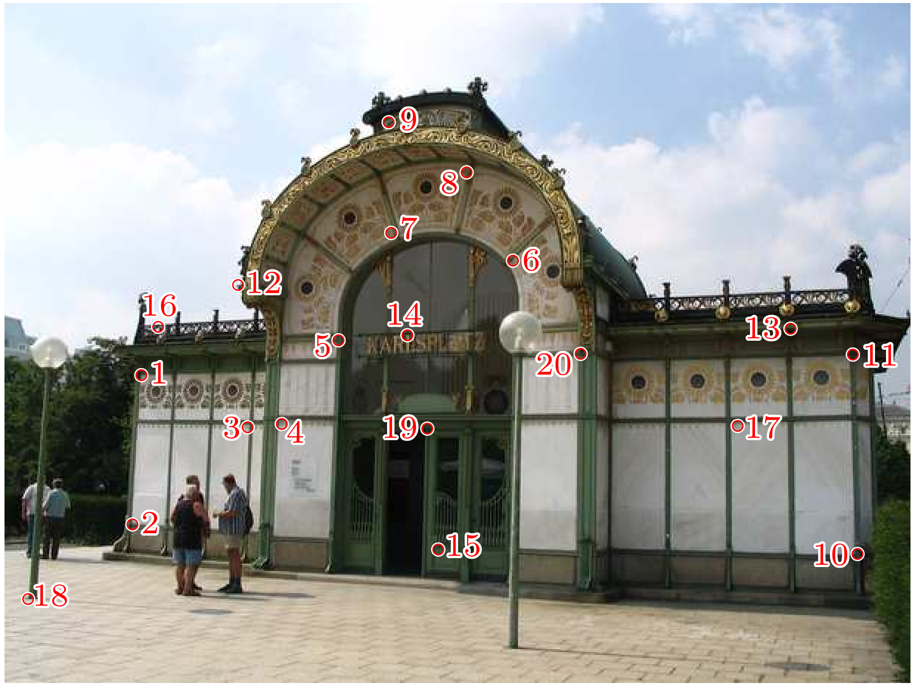
5) In one of the frames compute the approximate point of intersection between corresponding rays.

6) Transform the recovered coordinates into world coordinates.

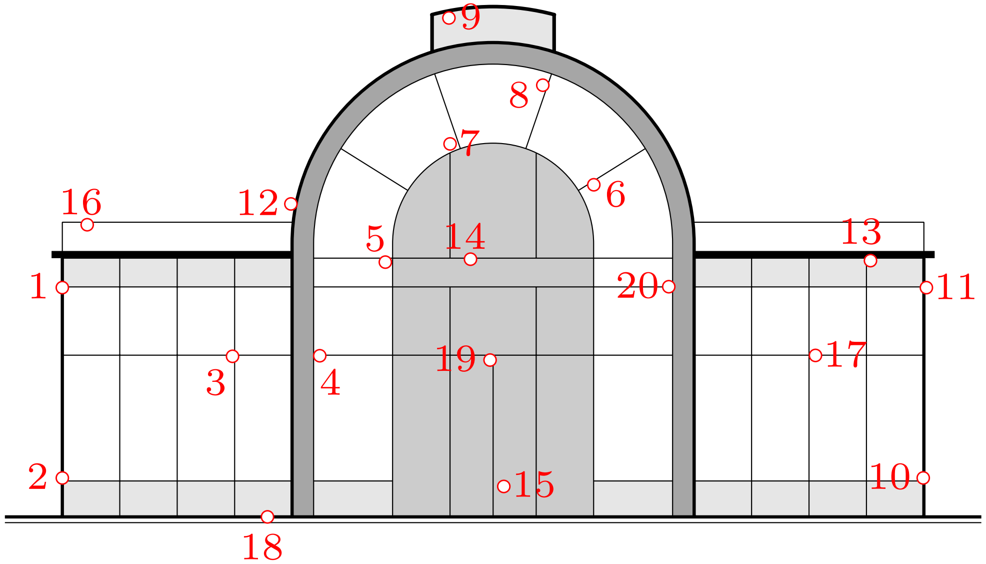
# Remaining problems

- Analysis of precision (e.g., no precise linear images due to effects of lenses)
- Effects of automated calibration (autofocus and zooming change the focal distance $d$),
- There are critical configurations (e.g., all passpoints within one plane) where the problem of projectivity has no unique solution despite of an arbitrarily big number of control points. How to figure out the correct one?
- How to find the correct decomposition of $\mathbf{B} = \mathbf{S} \cdot \mathbf{R}$ among the two possible one?

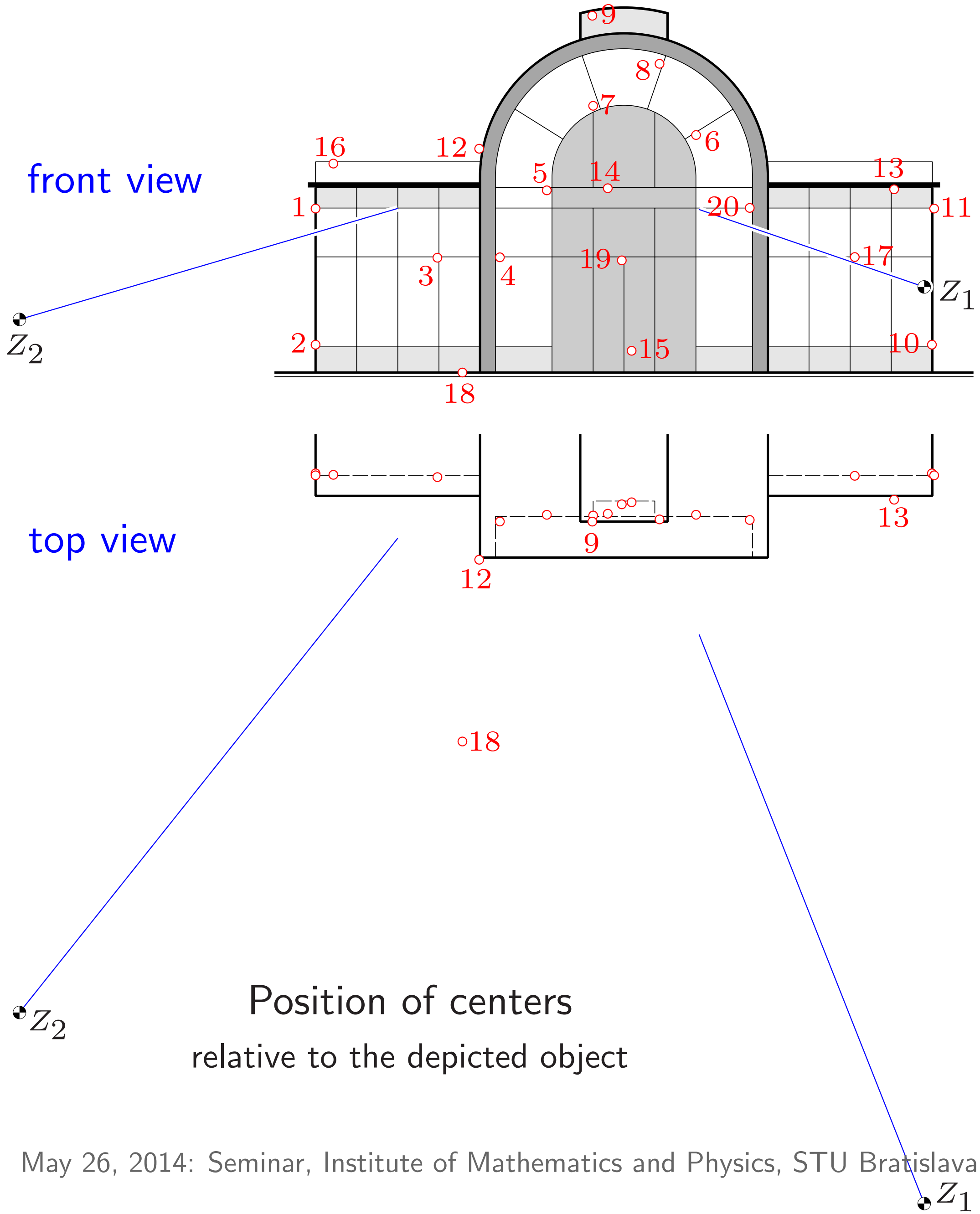
# The solution




original image




the reconstruction (M $\sim 1:100$)




## Position of centers

relative to the depicted object

Photo 1

Photo 2

# Literatur


- H. Brauner: *Lehrbuch der konstruktiven Geometrie*. Springer, Wien 1986.
- H. Brauner: *Lineare Abbildungen aus euklidischen Räumen*. Beitr. Algebra Geom. **21**, 5–26 (1986).
- A. Dür: *An Algebraic Equation for the Central Projection*. J. Geometry Graphics **7**, 137–143 (2003).
- O. Faugeras: *Three-Dimensional Computer Vision. A Geometric Viewpoint*. MIT Press, Cambridge, Mass., 1906 .
- O. Faugeras, Q.-T. Luong: *The Geometry of Multiple Images*. MIT Press, Cambridge, Mass., 2001.

- R. Harley, A. Zisserman: *Multiple View Geometry in Computer Vision*. Cambridge University Press 2000.
- V. Havel: *O rozkladu singulárních lineárních transformací*. Časopis Pést. Mat. **85**, 439–447 (1960).
- H. Havlicek: *On the Matrices of Central Linear Mappings*. Math. Bohem. **121**, 151–156 (1996).
- M. Hoffmann: *On the Theorems of Central Axonometry*. J. Geometry Graphics **2**, 151–155 (1997).
- M. Hoffmann, P. Yiu: *Moving Central Axonometric Reference Systems*. J. Geometry Graphics **9**, (2005) (in press).
- E. Kruppa: *Zur achsonometrischen Methode der darstellenden Geometrie*. Sitzungsber., Abt. II, österr. Akad. Wiss., Math.-Naturw. Kl. **119**, 487–506 (1910).

- YI MA, ST. SOATTO, J. KOŠECKÁ, S. SH. SASTRY: *An Invitation to 3-D Vision*. Springer-Verlag, New York 2004.
- H. STACHEL: *Mehrdimensionale Axonometrie*. Proceedings of the Congress of Geometry, Thessaloniki 1987, 159–168.
- H. STACHEL: *Parallel Projections in Multidimensional Space*. Proceedings of Compugraphics '91, Sesimbra (Portugal) 1991: Vol. I, 119–128.
- H. STACHEL: *Zur Kennzeichnung der Zentralprojektionen nach H. Havlicek*. Sitzungsber., Abt. II, österr. Akad. Wiss., Math.-Naturw. Kl. **204**, 33–46 (1995).
- H. STACHEL: *On Arne Dür's Equation Concerning Central Axonometries*. J. Geometry Graphics **8**, 215–224 (2004).
- E. STIEFEL: *Zum Satz von Pohlke*. Comment. Math. Helv. **10**, 208–225 (1938).

- E. Stiefel: *Lehrbuch der darstellenden Geometrie*. 3. Aufl., Basel, Stuttgart 1971.
- H. Stachel: *Descriptive Geometry Meets Computer Vision — The Geometry of Two Images*. J. Geometry Graphics **10**, 137–153 (2006).
- J. Szabó, H. Stachel, H. Vogel: *Ein Satz über die Zentralaxonometrie*. Sitzungsber., Abt. II, österr. Akad. Wiss., Math.-Naturw. Kl. **203**, 3–11 (1994).
- J. Tschupik, F. Hohenberg: *Die geometrische Grundlagen der Photogrammetrie*. In Jordan, Eggert, Kneissl (eds.): *Handbuch der Vermessungskunde III a/3.* 10. Aufl., Metzlersche Verlagsbuchhandlung, Stuttart 1972, 2235–2295.
- K. Vala: *A propos du théorèm de Pohlke*. Ann. Acad. scient. Fennicae **270**, 3–5 (1959).