GEOMETRY FOR 3D COMPUTER VISION

Václav Hlaváč

Czech Technical University, Faculty of Electrical Engineering
Department of Cybernetics, Center for Machine Perception
121 35 Praha 2, Karlovo nám. 13, Czech Republic

hlavac@fel.cvut.cz, http://cmp.felk.cvut.cz

LECTURE PLAN

1. Brief introduction to my group – Center for Machine Perception.
2. Mathematical model of a single perspective camera.
3. Epipolar constraint.
4. Correspondence problem.
5. Results: state-of-the-art stereo, uncalibrated 3D reconstruction, VR model.

- Research group, head Prof. Václav Hlaváč, established 1986 as computer vision lab, under the name CMP since 1996.

- $12\frac{1}{2}$ staff ($1\frac{1}{2}$ Prof., 1 Assoc. Prof., 3 PhD, 7 MSc);
  out of it 2 mathematicians, 2 physicists, 8 engineers)
  $+$ 8 full time PhD students.

- Interests: computer vision, pattern recognition, mathematical models for treating uncertainty.

- Links to industry mainly via a spin-off company Neovision Prague (10 people).

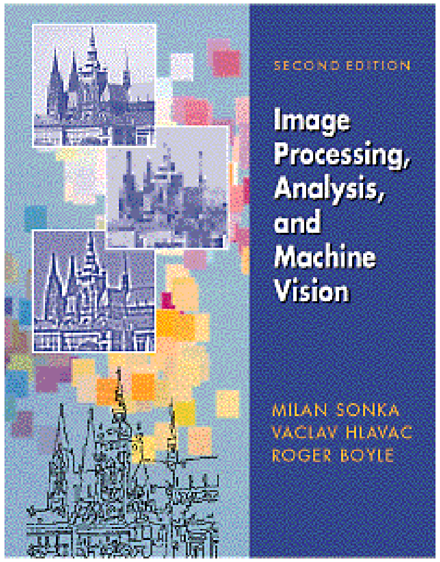
  E.g. Samsung, Boeing, Texas Instruments, Robert Bosch, Kyocera, Hitachi.

MAIN RUNNING PROJECTS

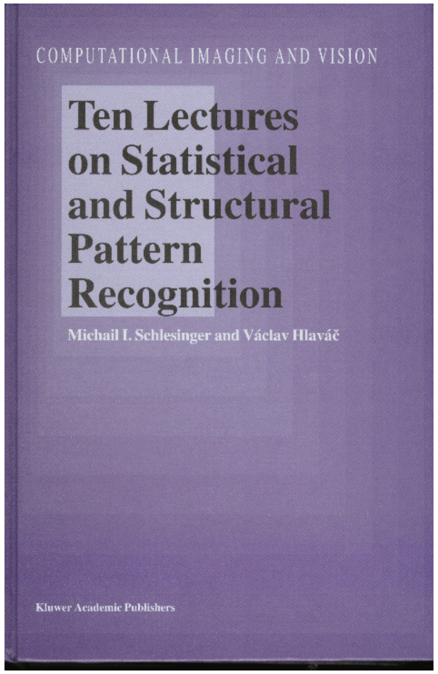
- ActIPret (R&D, 2001-2003, IST-2001-32184) Interpreting and Understanding Activities of Expert Operators for Teaching and Education (V. Hlaváč, J. Matas).
- ISAAC (Trial, 2002, IST-2001-33266) Inspecting Sewerage Systems And Image Analysis by Computer (V. Hlaváč).
- Reconstruction of 3D scene from multiple uncalibrated views (V. Hlaváč).
- Computational stereo (R. Šára).
- Omni-directional vision. (T. Pajdla).
- Authentication based on face recognition (J. Matas).
- Pattern recognition theory (V. Hlaváč).

V. Hlaváč, books

Šonka M., Hlaváč V., Boyle R.B.: *Image Analysis, Processing and Machine Vision*, 2nd edition, PWS Boston, USA, 1999 (China edition 2002), 800 p, USD 105.

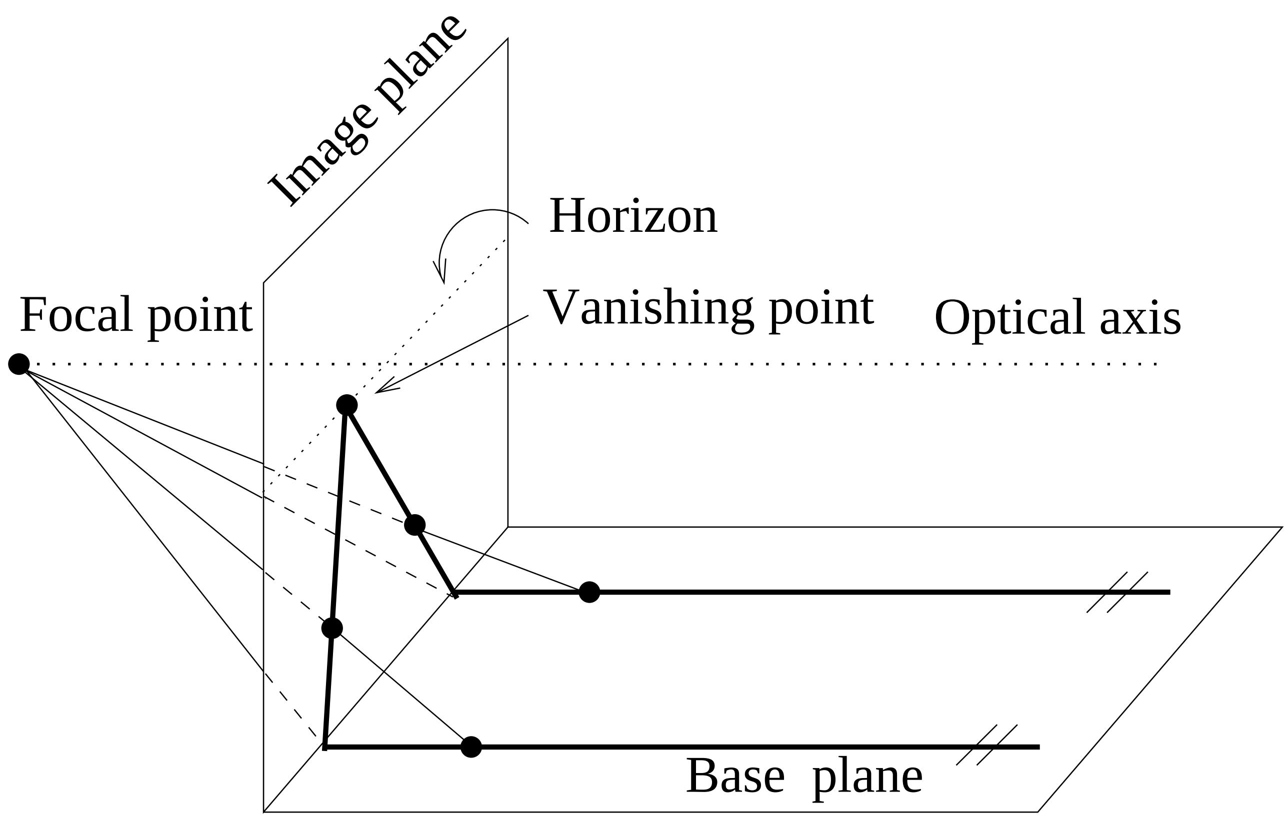


Schlesinger M.I., Hlaváč V.: *Ten Lectures on Statistical and Structural Pattern Recognition* Kluwer, Dordrecht, May 2002, EUR 165.



BASICS OF PROJECTIVE GEOMETRY

- Pinhole model - the simplest geometrical model of human eye, photographic and TV camera.
- Perspective projection, also central projection.
- Parallel lines in the world do not remain parallel in the image (e.g., view along the straight section of a railroad).



PROJECTIVE SPACE

Consider $(n+1)$ dimensional vector space without its origin, $\mathcal{R}^{n+1} - \{(0,\ldots,0)\}$.

Define an equivalence relation

$$\begin{array}{rrcl} & [x_1,\ldots,x_{n+1}]^T & \equiv & [x'_1,\ldots,x'_{n+1}]^T \\ \text{iff } \exists\, \alpha \neq 0 : & [x_1,\ldots,x_{n+1}]^T & = & \alpha \; [x'_1,\ldots,x'_{n+1}]^T \end{array}$$

Projective space $\mathcal{P}^n$ is the quotient space of this equivalence relation.

Points in the projective space are expressed in homogeneous co-ordinates (called also projective coordinates) $\tilde{\mathbf{x}} = [x'_1,\ldots,x'_n,1]^T$.

RELATION BETWEEN EUCLIDEAN AND PROJECTIVE SPACES

Consider Euclidean space $\mathcal{R}^n$.

The one-to-one mapping from the $\mathcal{R}^n$ into $\mathcal{P}^n$

$$[x_1, \ldots, x_n]^T \rightarrow [x_1, \ldots, x_n, 1]^T$$

Projective points $[x_1, \ldots, x_n, 0]^T$ do not have an Euclidean counterpart and represent points at infinity in a particular direction.

Consider $[x_1, \ldots, x_n, 0]^T$ as a limiting case of $[x_1, \ldots, x_n, \alpha]^T$ that is projectively equivalent to $[x_1/\alpha, \ldots, x_n/\alpha, 1]^T$, and assume that $\alpha \rightarrow 0$.

This corresponds to a point in $\mathcal{R}^n$ going to infinity in the direction of the radius vector $[x_1/\alpha, \ldots, x_n/\alpha] \in \mathcal{R}^n$.

PROJECTIVE TRANSFORMATION (also CO-LINEATION)

Co-lineation is any mapping $\mathcal{P}^n \rightarrow \mathcal{P}^n$.

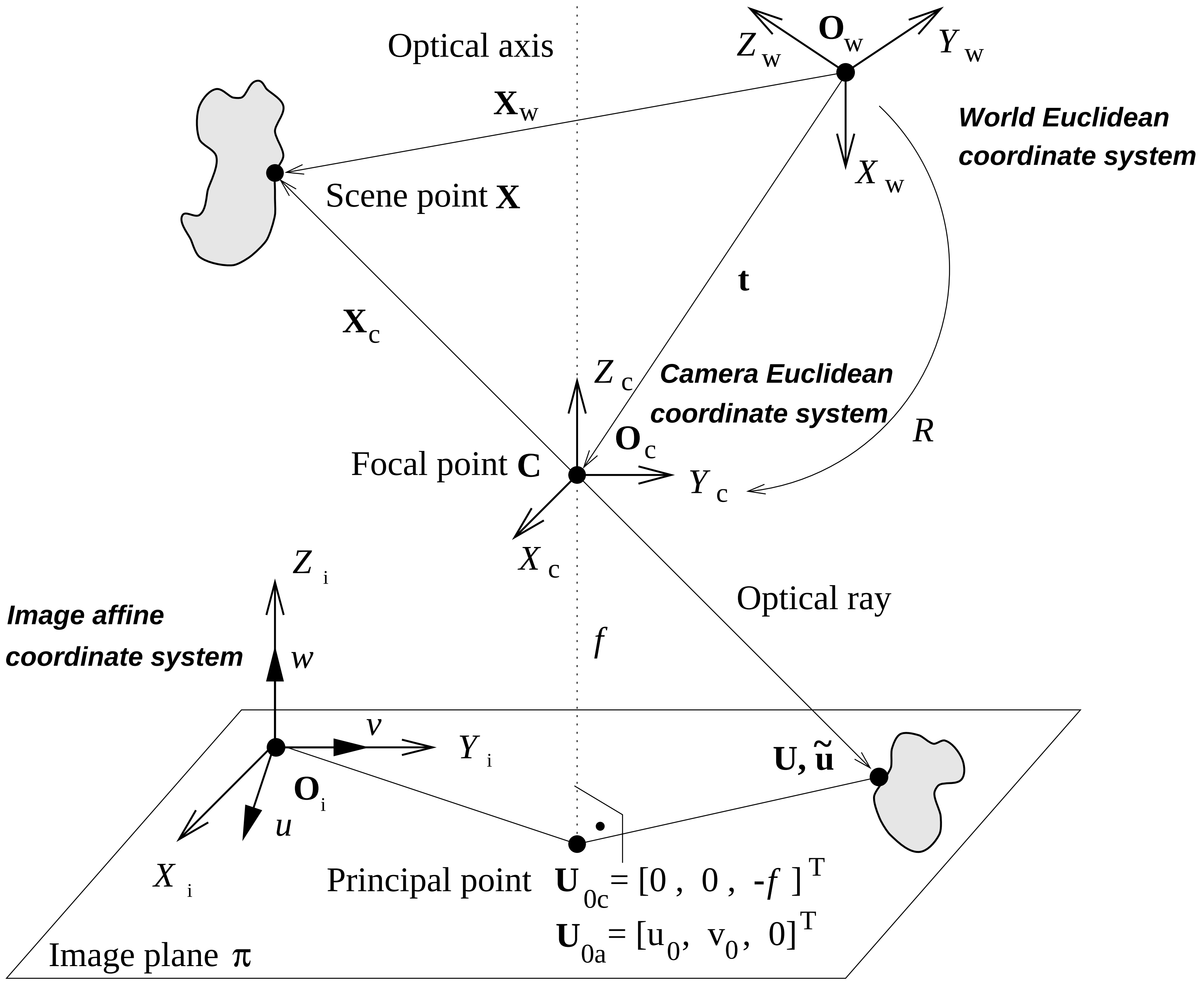
Defined by a regular $(n+1) \times (n+1)$ matrix A, $\tilde{\mathbf{y}} = A\ \tilde{\mathbf{x}}$.

Matrix A is defined up to a scale factor.

Co-lineations map hyperplanes to hyperplanes.

A special case is the mapping of lines to lines that is often used in computer vision.

SINGLE PERSPECTIVE CAMERA, pinhole model



CAMERA: $\mathcal{P}^3 \rightarrow \mathcal{P}^2$

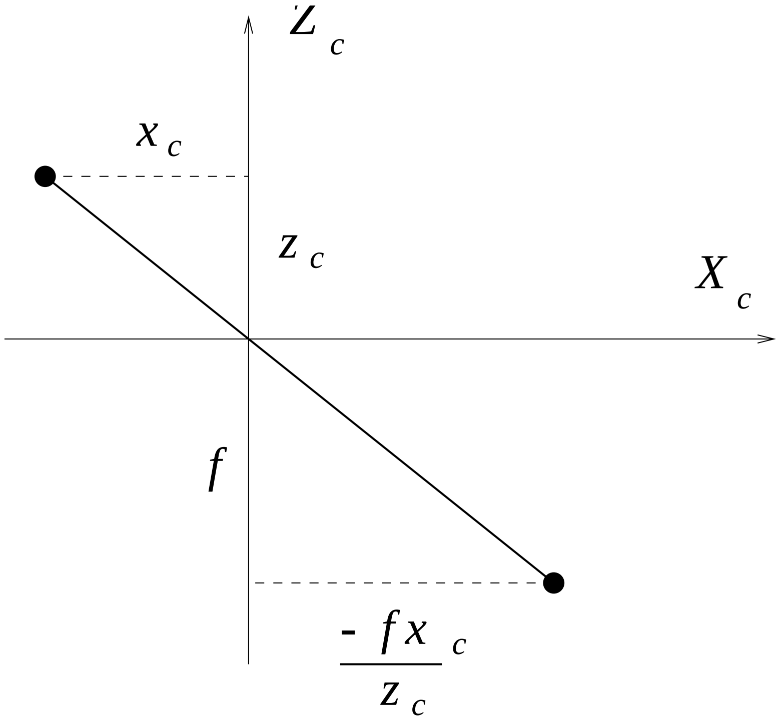
A scene point $\mathbf{X}_w$ in the world Euclidean co-ordinate system is a 3×1 vector.

The same point $\mathbf{X}_c$ in the camera Euclidean co-ordinate system is transformed by translation $\mathbf{t}$ (vector) and rotation R (orthogonal matrix).

$$\mathbf{X}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R\,(\mathbf{X}_w - \mathbf{t})$$

CAMERA: $\mathcal{P}^3 \rightarrow \mathcal{P}^2$ (2)

The point $\mathbf{X}_c$ is projected to the image plane π as point $\mathbf{U}_c$.



$$\mathbf{U}_c = \left[\; \frac{-f x_c}{z_c}, \quad \frac{-f y_c}{z_c}, \quad -f \; \right]^T , \qquad \mathbf{U}_{0a} = [u_0, v_0, 0]^T .$$

CAMERA: $\mathcal{P}^3 \rightarrow \mathcal{P}^2$ (3)

Projected point in the 2D image plane π in homogeneous co-ordinates

$$\tilde{\mathbf{u}} = \begin{bmatrix} U \\ V \\ W \end{bmatrix} = \begin{bmatrix} a & b & -u_0 \\ 0 & c & -v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{-f x_c}{z_c} \\ \frac{-f y_c}{z_c} \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} -fa & -fb & -u_0 \\ 0 & -fc & -v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{x_c}{z_c} \\ \frac{y_c}{z_c} \\ 1 \end{bmatrix}$$

2D Euclidean counterpart is $\mathbf{u} = [u, v]^T = [\frac{U}{W}, \frac{V}{W}]^T$.

CALIBRATION MATRIX K

$$
\begin{aligned}
z_c\, \tilde{\mathbf{u}} &= z_c \begin{bmatrix} -fa & -fb & -u_0 \\ 0 & -fc & -v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{x_c}{z_c} \\ \frac{y_c}{z_c} \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} -fa & -fb & -u_0 \\ 0 & -fc & -v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \\
&= \begin{bmatrix} -fa & -fb & -u_0 \\ 0 & -fc & -v_0 \\ 0 & 0 & 1 \end{bmatrix} R\,(\mathbf{X}_w - \mathbf{t}) = K R\,(\mathbf{X}_w - \mathbf{t})
\end{aligned}
$$

Calibration parameters:
intrinsic (matrix K) vs. extrinsic (vector $\mathbf{t}$, matrix R).

PROJECTION MATRIX M

$$
\begin{aligned}
\tilde{\mathbf{u}} = \begin{bmatrix} U \\ V \\ W \end{bmatrix} &= \frac{1}{z_c} K R \left(\mathbf{X}_w - \mathbf{t} \right) \\
&= [K R \,|\, - K \; R \, \mathbf{t}] \begin{bmatrix} \mathbf{X}_w \\ 1 \end{bmatrix} \\
&= M \begin{bmatrix} \mathbf{X}_w \\ 1 \end{bmatrix} \\
&= M \tilde{\mathbf{X}}_w
\end{aligned}
$$

SINGLE CAMERA CALIBRATION, overview

Intrinsic parameters only - seeking matrix K.

Intrinsic + extrinsic parameters - seeking matrix M.

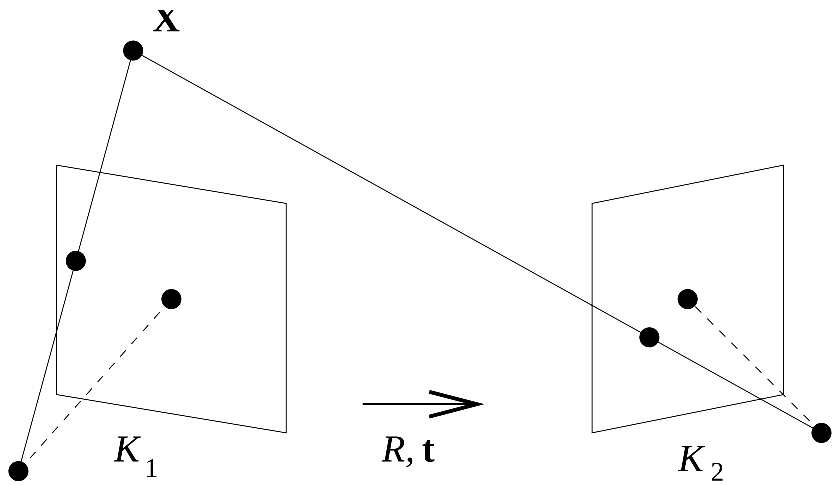
1. Known scene: A set of n non-degenerate (not co-planar) points in the 3D world (e.g., a calibration object), and the corresponding 2D image points are known.

   Each correspondence between a 3D scene and 2D image point provides one equation

$$\alpha_j \tilde{\mathbf{u}}_j = M \begin{bmatrix} \mathbf{X}_j \\ 1 \end{bmatrix} .$$

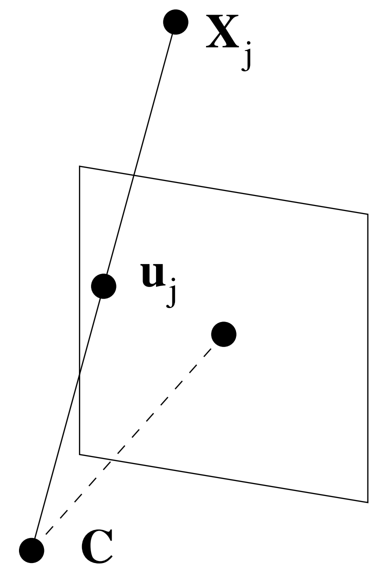
2. Unknown scene: More views are needed to calibrate the camera. The intrinsic camera parameters will not change for different views, and the correspondence between image points in different views must be established.

CALIBRATION FROM UNKNOWN SCENE (cont.)



1. Known camera motion: Three cases according to the known motion constraint:
   (a) *Both rotation and translation*, general case.
   (b) *Pure rotation*
   (c) *Pure translation*, a linear solution proposed by [Pajdla, Hlaváč 1995].
2. Unknown camera motion: The most general case, sometimes called *camera self-calibration*. At least three views are needed and the solution is nonlinear. Numerically hard.

CAMERA CALIBRATION FROM A KNOWN SCENE (1)



Typically a two stage process.

1. Estimate the projection matrix M is estimated from the co-ordinates of points with known scene positions.
2. The extrinsic and intrinsic parameters are estimated from M.

Note: The second step is not always needed – the case of stereo vision is an example.

CAMERA CALIBRATION FROM A KNOWN SCENE (2)

Each correspondence between scene point $\mathbf{X} = [x, y, z]^T$ and 2D image point $[u, v]^T$ gives one equation

$$\begin{bmatrix} \alpha u \\ \alpha v \\ \alpha \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \alpha u \\ \alpha v \\ \alpha \end{bmatrix} = \begin{bmatrix} m_{11}x & + & m_{12}y & + & m_{13}z & + & m_{14} \\ m_{21}x & + & m_{22}y & + & m_{23}z & + & m_{24} \\ m_{31}x & + & m_{32}y & + & m_{33}z & + & m_{34} \end{bmatrix}$$

CAMERA CALIBRATION FROM A KNOWN SCENE (3)

$$u(m_{31}x + m_{32}y + m_{33}z + m_{34}) = m_{11}x + m_{12}y + m_{13}z + m_{14}$$
$$v(m_{31}x + m_{32}y + m_{33}z + m_{34}) = m_{21}x + m_{22}y + m_{23}z + m_{24}$$

Two linear equations, each in 12 unknowns $m_{11}, \ldots, m_{34}$, for each known corresponding scene and image point (actually only 11 unknowns due to unknown scaling). 6 corresponding points needed, at least.

If n such points are available, we can write it as a $2n \times 12$ matrix.

$$\begin{bmatrix} x & y & z & 1 & 0 & 0 & 0 & 0 & -ux & -uy & -uz & -u \\ 0 & 0 & 0 & 0 & x & y & z & 1 & -vx & -vy & -vz & -v \\ & & & & & & \vdots & & & & & \end{bmatrix} \begin{bmatrix} m_{11} \\ m_{12} \\ \vdots \\ m_{34} \end{bmatrix} = 0$$

Overconstraint linear system. Robust least squares. Result $= M$.

SVD—Singular Value Decomposition

SVD is a linear algebra technique for solving linear equations in the least square sense. SVD works for singular matrices or matrices numerically close to singular. Contained, e.g., in MATLAB.

Any $m \times n$ matrix A, $m \geq n$ can be factorized as $A = UDV^T$.

U has orthonormal columns, D is non-negative diagonal, and V^T has orthonormal rows.

SVD locates the closest possible solution in a least square sense.

Sometimes need for the 'closest' singular matrix to the original matrix A – this decreases the rank from n to $n-1$. Replace the smallest diagonal element of D by zero. This new matrix is the closest to the original one with respect to the Frobenius norm (which is calculated as a sum of the squared values of all matrix elements).

SEPARATION OF EXTRINSIC PARAMETERS FROM M

Given: projection matrix M

Output: rotation matrix R and translation vector $\mathbf{t}$).

$$M = [KR \,|\, -KR\,\mathbf{t}] = [A\,|\mathbf{b}]$$

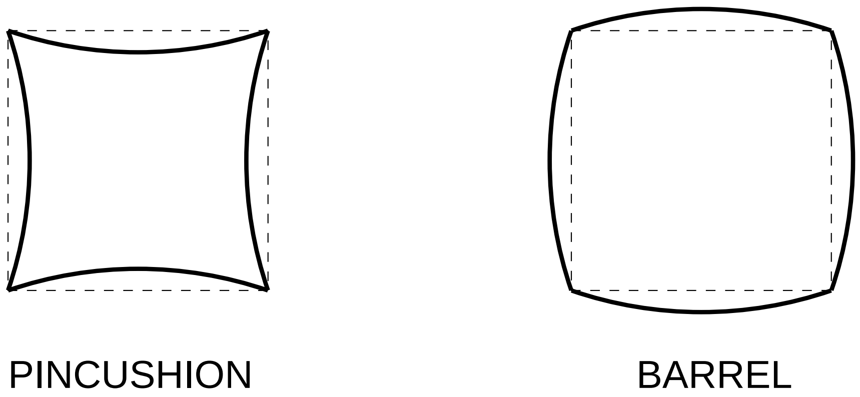
The 3×3 submatrix is denoted as A, and the rightmost column as $\mathbf{b}$.

Translation vector $\mathbf{t}$ is easy; $A = KR$, $\mathbf{t} = -A^{-1}\mathbf{b}$.

Rotation matrix R. Recall that the calibration matrix K is upper triangular and the rotation matrix is orthogonal.

The QR factorization method or SVD will decompose A into a product and hence recover K and R.

RADIAL DISTORTION AND DE-CENTERING



Often modelled as rotationally symmetric by polynomials.

u, v - correct image co-ordinates

$\tilde{u}$, $\tilde{v}$ - measured uncorrected image co-ordinates

$\hat{u}_0$, $\hat{v}_0$ - estimate of the position of the principal point

$$\tilde{u} = x - \hat{u}_0 \, , \quad \tilde{v} = y - \hat{v}_0$$

RADIAL DISTORTION (2)

$$u = \tilde{u} + \delta u \,, \quad v = \tilde{v} + \delta v$$

$$\begin{aligned} \delta u &= (\tilde{u} - u_p)(\kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6) \\ \delta v &= (\tilde{v} - v_p)(\kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6) \end{aligned}$$

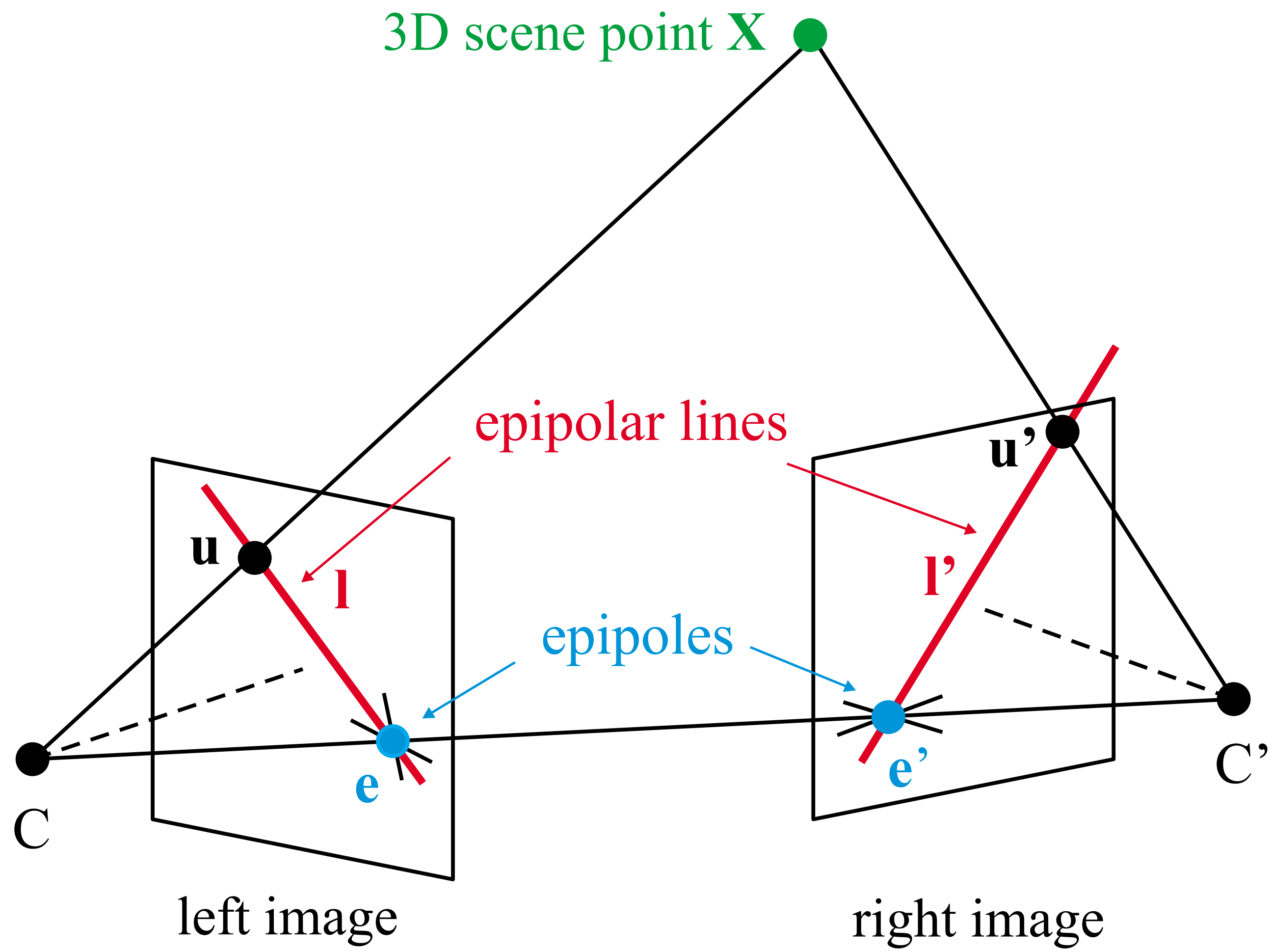
r^2 is the square of the radial distance from the center of the image.

$$r^2 = (\tilde{u} - u_p)^2 + (\tilde{u} - u_p)^2$$

u_p, v_p are corrections to $\hat{u}_0$, $\hat{v}_0$

$$\begin{aligned} u_0 &= \hat{u}_0 + u_p \\ v_0 &= \hat{v}_0 + v_p \end{aligned}$$

GEOMETRY OF 2 CAMERAS



Epipoles $\mathbf{e}$, $\mathbf{e}'$, epipolar lines $\mathbf{l}$, $\mathbf{l}'$.

$\mathbf{e}$, $\mathbf{e}'$, $\mathbf{l}$, $\mathbf{l}'$, C, C', $\mathbf{X}$ lie in a single plane.

Epipolar geometry. Seeking correspondences between two 1D signals.
Bilinear relation between $\mathbf{u}$, $\mathbf{u}'$.

FUNDAMENTAL MATRIX (1)

Left projection $\mathbf{u}$ and right projection $\mathbf{u}'$ of the scene point $\mathbf{X}$.

$$
\begin{aligned}
\mathbf{u} &\simeq [K|\mathbf{0}] \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} = K\,\mathbf{X}, \\
\mathbf{u}' &\simeq [K'R\,|-K'R\,\mathbf{t}] \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} \\
&= K'(R\,\mathbf{X} - R\,\mathbf{t}) = K'\mathbf{X}'
\end{aligned}
$$

Coplanarity of $\mathbf{X}$, $\mathbf{X}'$ and $\mathbf{t}$.

Distinguish co-ordinates of the left and right cameras by the subscript $_L$, $_R$.

Vector product $\times$.

FUNDAMENTAL MATRIX (2)

Coordinates rotation
$\mathbf{X}'_R = R\,\mathbf{X}'_L$, and hence $\mathbf{X}'_L = R^{-1}\mathbf{X}'_R$.

Coplanarity constraint $\mathbf{X}_L^T(\mathbf{t} \times \mathbf{X}'_L) = 0$.

Preparing for substitution
$\mathbf{X}_L = K^{-1}\mathbf{u}$, $\mathbf{X}'_R = (K')^{-1}\mathbf{u}'$, and $\mathbf{X}'_L = R^{-1}(K')^{-1}\mathbf{u}'$.

Epipolar constraint in vector form

$$(K^{-1}\mathbf{u})^T(\mathbf{t} \times R^{-1}\,(K')^{-1}\mathbf{u}') = 0\,.$$

Equation is homogeneous with respect to $\mathbf{t}$, so the scale is not determined.

Absolute scale cannot be recovered without 'yardstick'.

FUNDAMENTAL MATRIX (3)

Replacement of a vector product by a matrix multiplication.

The translation vector is $\mathbf{t} = [t_x, t_y, t_z]^T$, and a skew symmetric matrix $S(\mathbf{t})$ (i.e., $S^T = -S$) can be created from it if $\mathbf{t} \neq \mathbf{0}$.

$$S(\mathbf{t}) = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

Note that $\mathrm{rank}(S) = 2$ if and only if $\mathbf{t} \neq \mathbf{0}$.

FUNDAMENTAL MATRIX (4)

The vector product can be replaced by the multiplication of two matrices.

For any regular matrix A, we have

$$\mathbf{t} \times A \; = S(\mathbf{t})\, A\,.$$

Thus we can rewrite the epipolar constraint in a vector form

$$(K^{-1}\mathbf{u})^T\, (S(\mathbf{t})\, R^{-1}\, (K')^{-1}\mathbf{u}') = 0\,,$$

$$\mathbf{u}^T(K^{-1})^T S(\mathbf{t}) R^{-1}(K')^{-1}\mathbf{u}' = 0\,.$$

FUNDAMENTAL MATRIX (5)

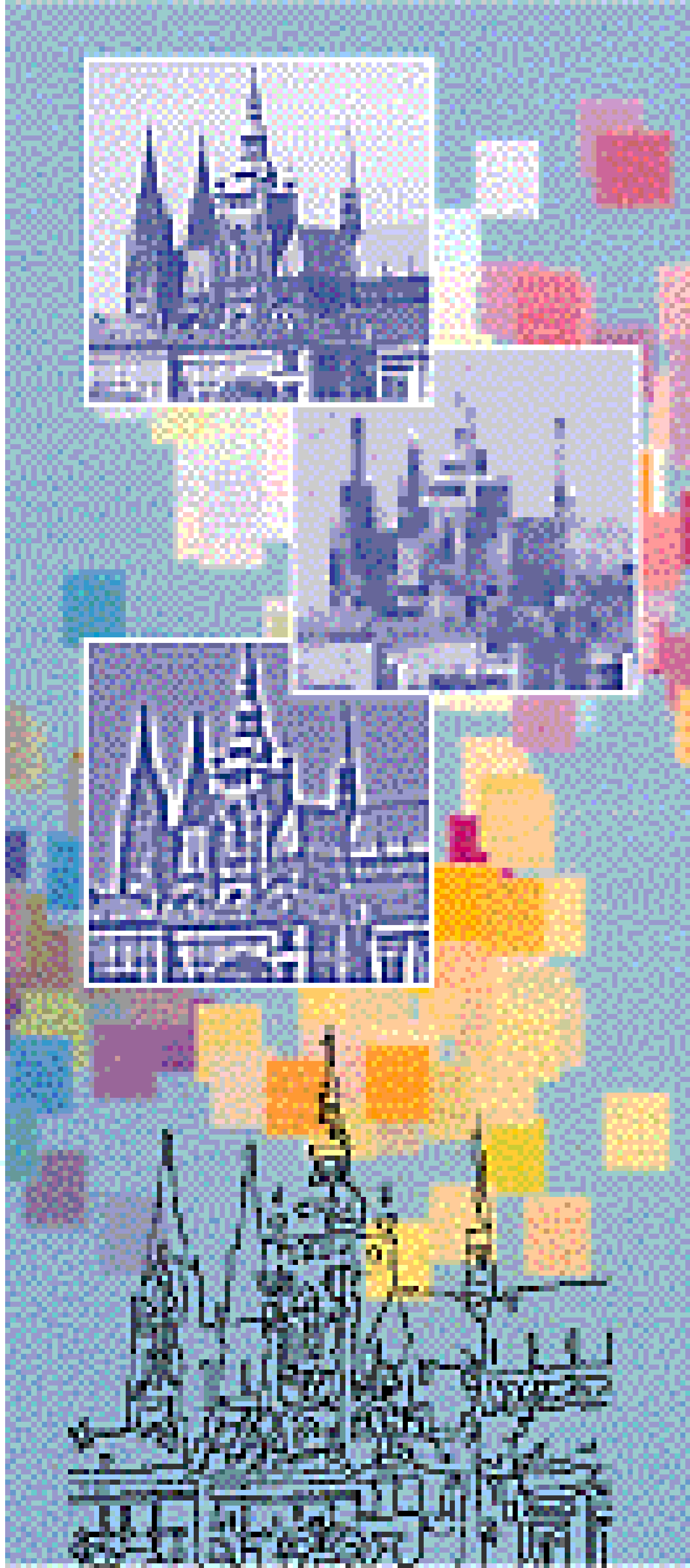
The middle part can be concentrated into a single matrix F called the fundamental matrix of two views,

$$F = (K^{-1})^T S(\mathbf{t}) R^{-1} (K')^{-1} \, .$$

With the substitution for F we finally get the bilinear relation (sometimes named after Longuet-Higgins) between any two views

$$\mathbf{u}^T \, F \, \mathbf{u}' = 0 \, .$$

It can be seen that the fundamental matrix F captures all information that can be recovered from a pair of images if the correspondence problem is solved.



SECOND EDITION

Image Processing, Analysis, and Machine Vision

MILAN SONKA
VACLAV HLAVAC
ROGER BOYLE

COMPUTATIONAL IMAGING AND VISION

Ten Lectures on Statistical and Structural Pattern Recognition

Michail I. Schlesinger and Václav Hlaváč

Kluwer Academic Publishers

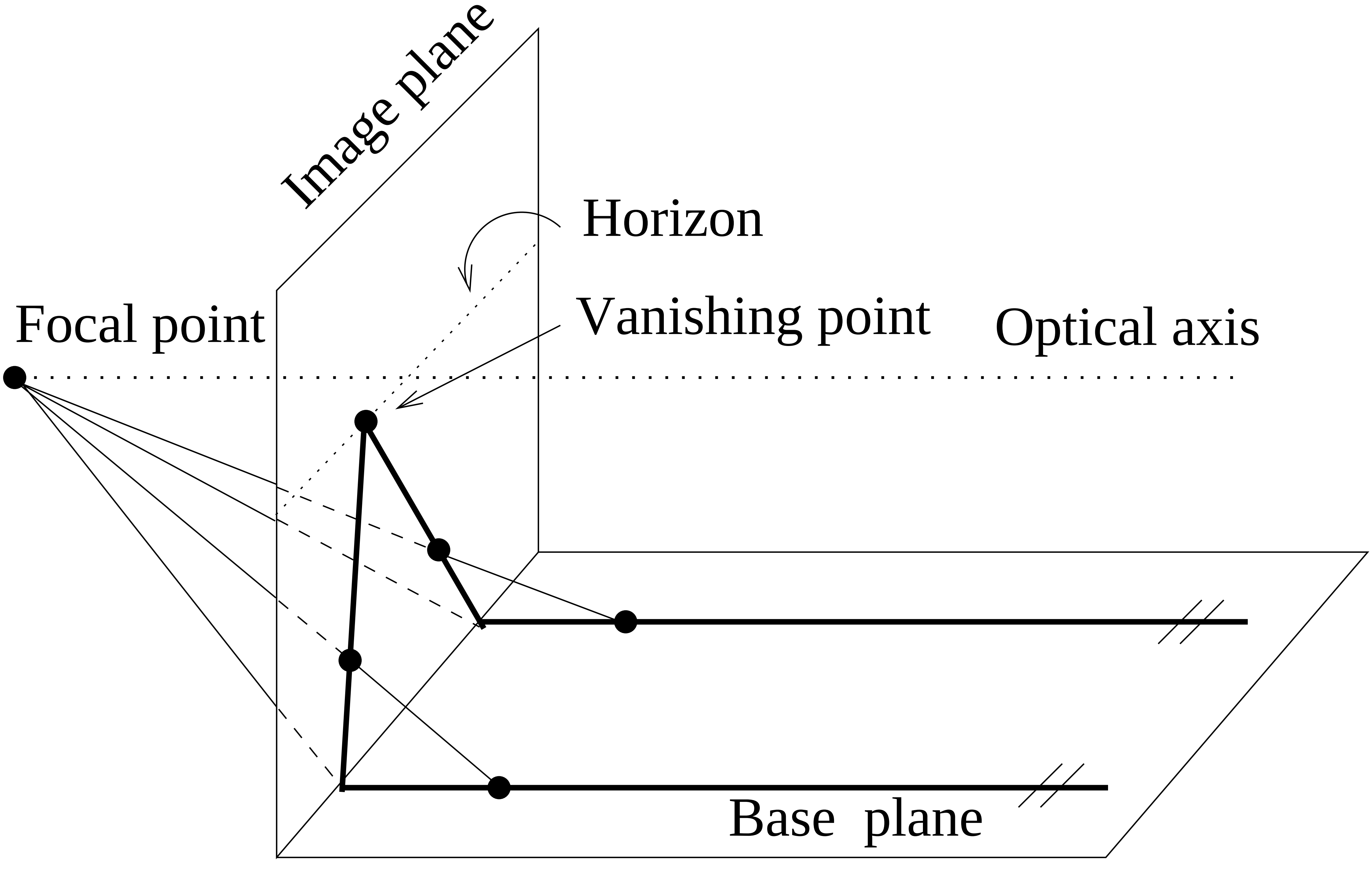
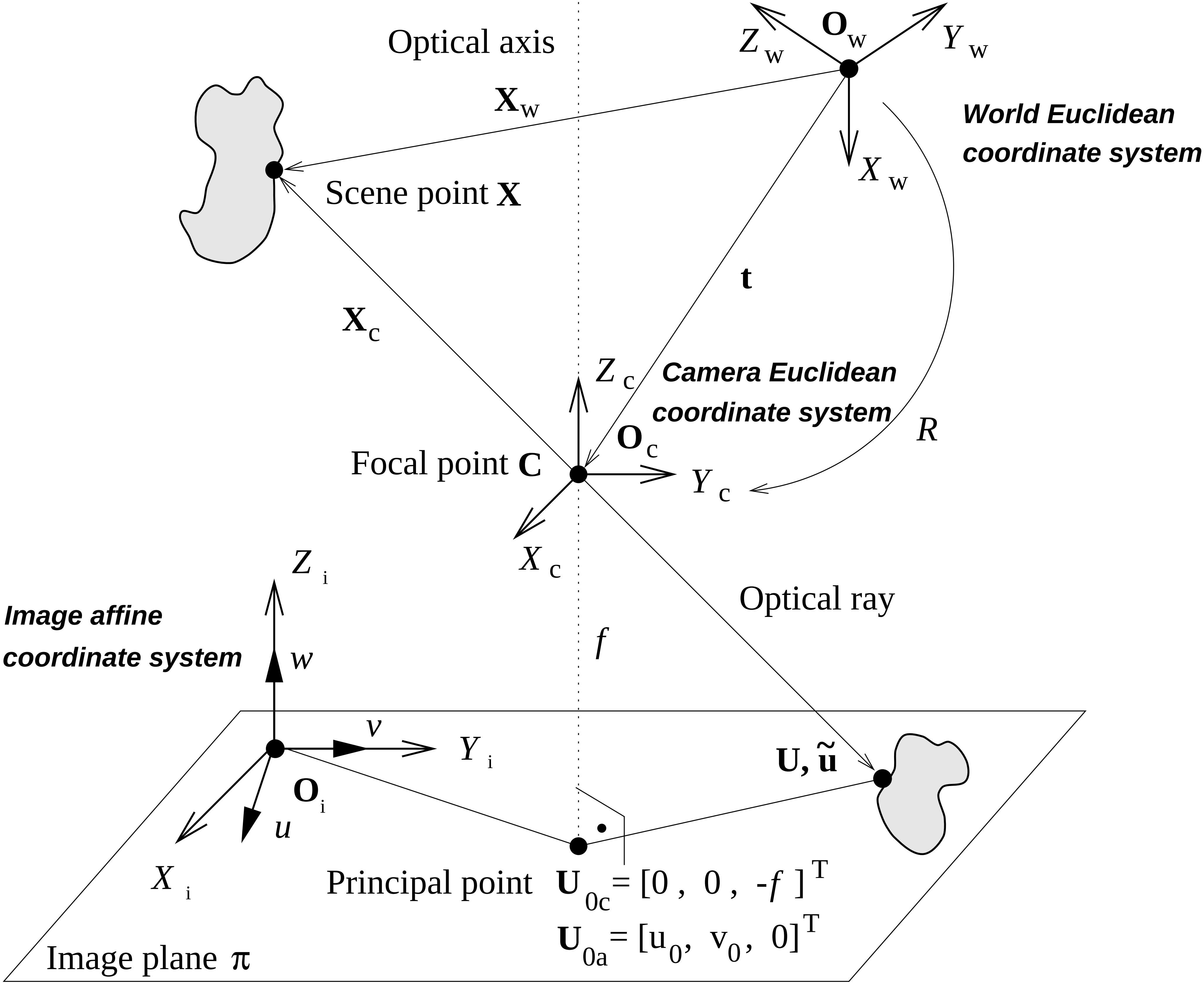
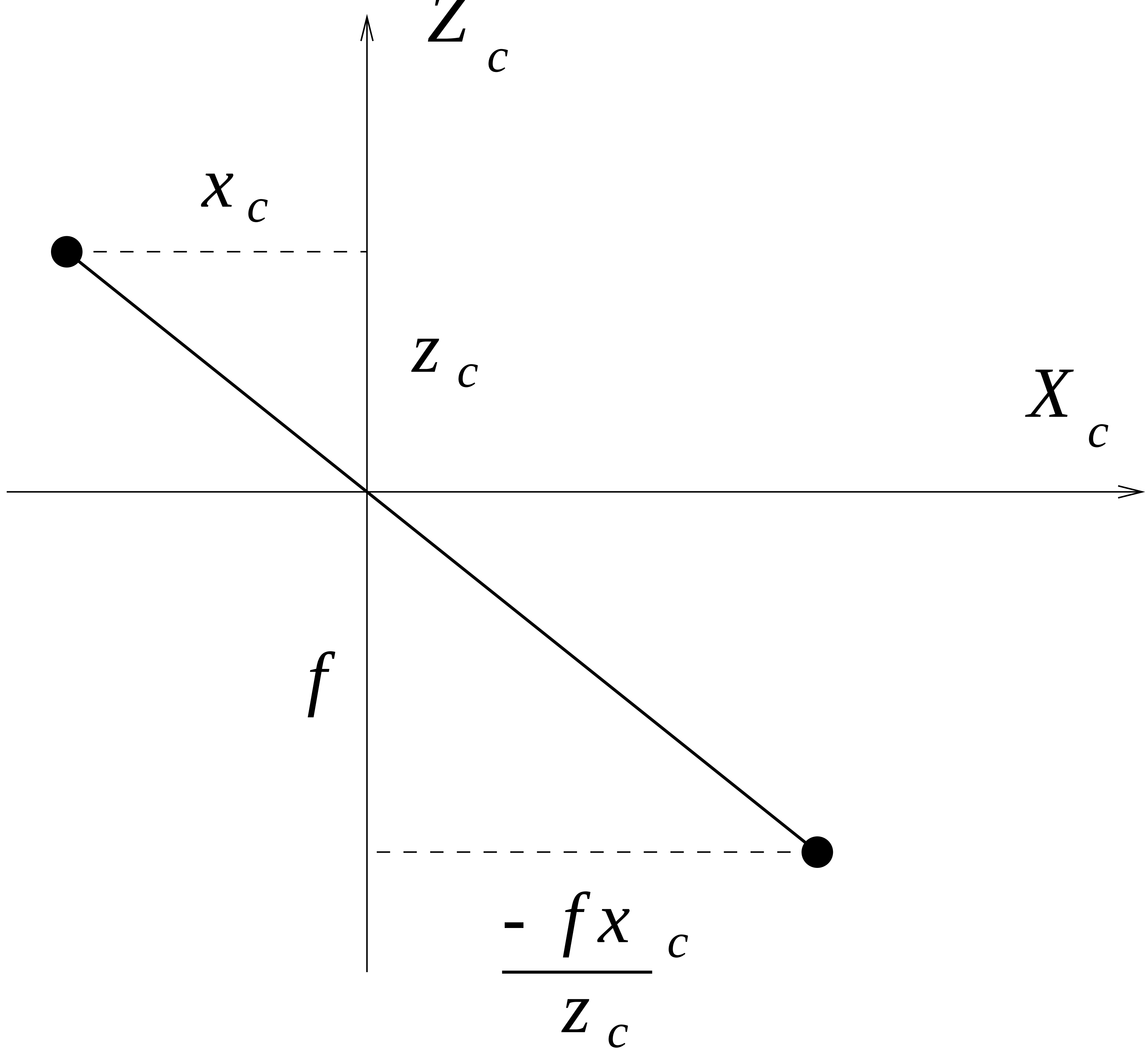


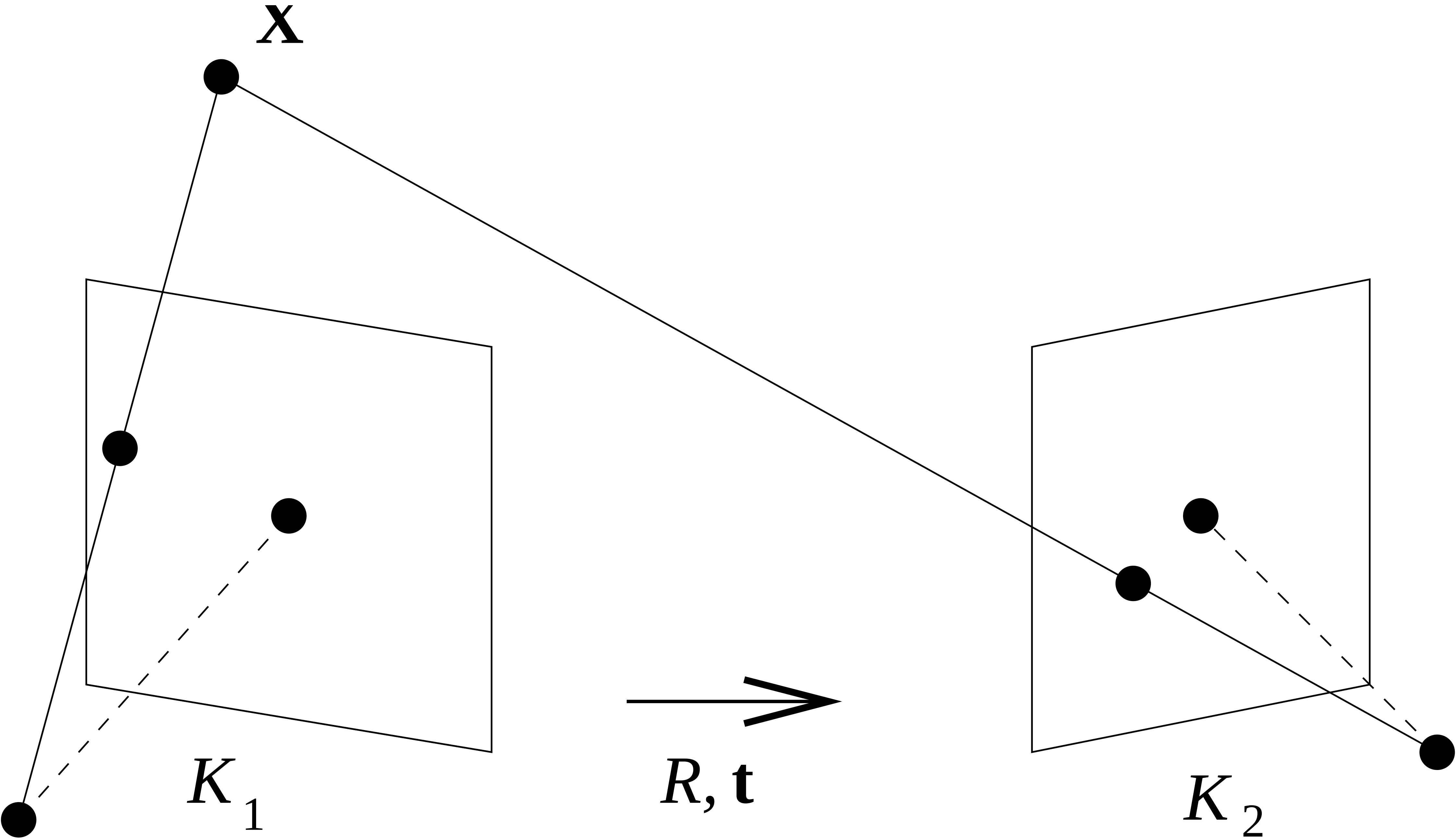
Image plane
Horizon
Vanishing point
Focal point
Optical axis
Base plane



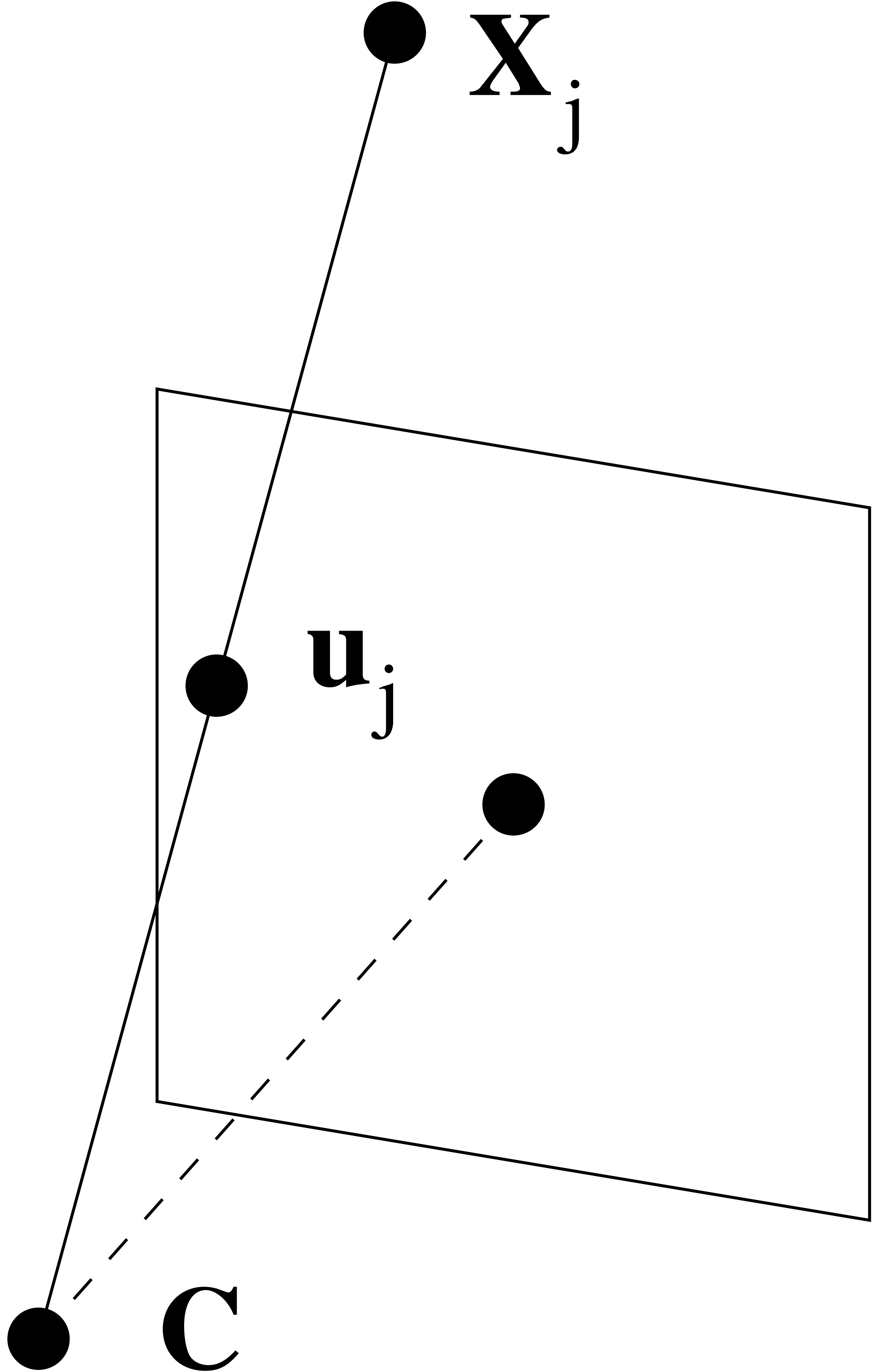
Optical axis
$\mathbf{O}_w$
Z_w
Y_w
X_w
$\mathbf{X}_w$
World Euclidean coordinate system
Scene point $\mathbf{X}$
$\mathbf{t}$
$\mathbf{X}_c$
Z_c
Camera Euclidean coordinate system
R
$\mathbf{O}_c$
Focal point $\mathbf{C}$
Y_c
X_c
Z_i
Optical ray
Image affine coordinate system
w
f
v
Y_i
$\mathbf{U}, \tilde{\mathbf{u}}$
$\mathbf{O}_i$
u
X_i
Principal point $\mathbf{U}_{0c} = [0,\ 0,\ -f\ ]^T$
$\mathbf{U}_{0a} = [u_0,\ v_0,\ 0]^T$
Image plane π



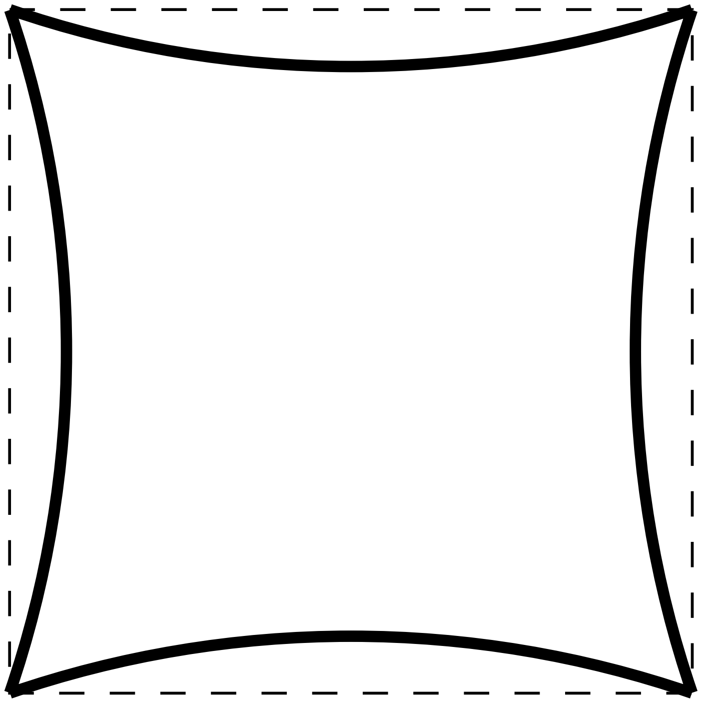
Z_c
x_c
z_c
X_c
f
$\frac{-f x_c}{z_c}$



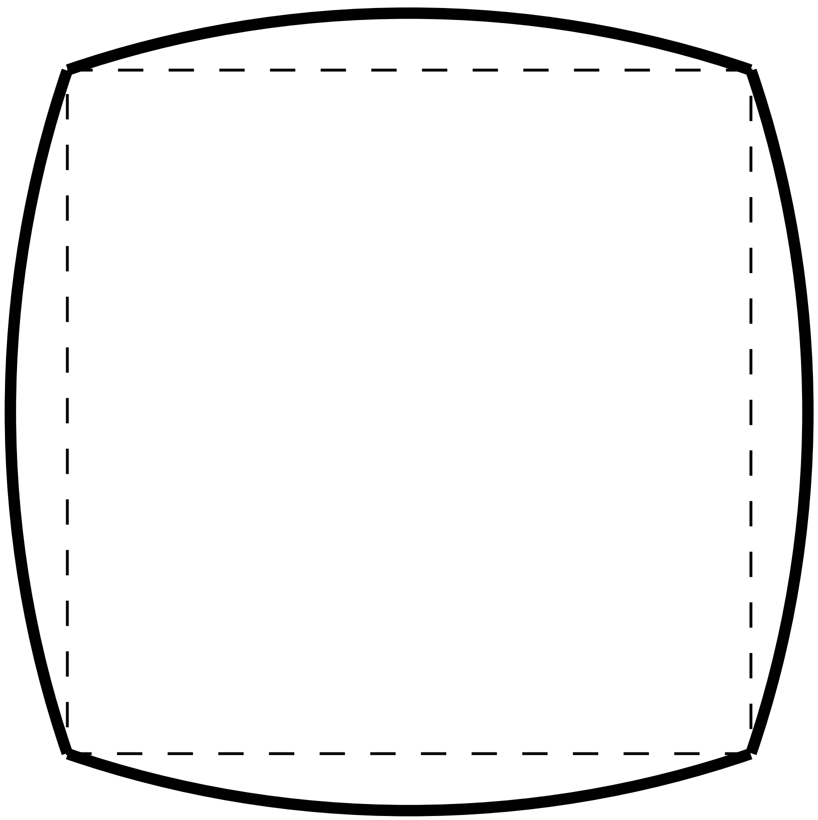
$\mathbf{X}$
K_1
$R, \mathbf{t}$
K_2



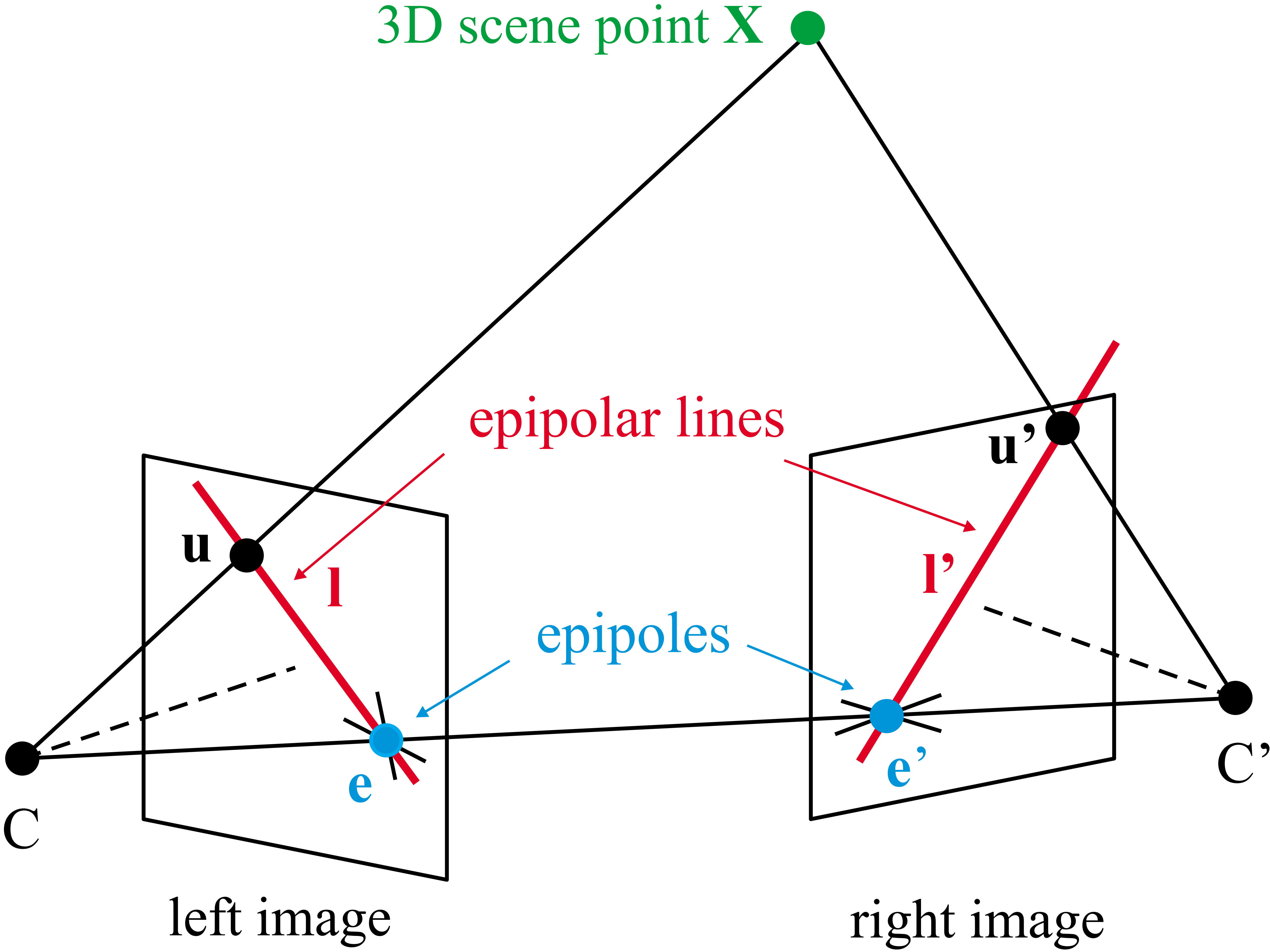
$\mathbf{X}_j$
$\mathbf{u}_j$
$\mathbf{C}$



PINCUSHION



BARREL


3D scene point X
epipolar lines
u
l
u'
l'
epipoles
e
e'
C
C'
left image
right image