Geometry of a single view (a single camera case)

Václav Hlaváč

Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics
Center for Machine Perception
`http://cmp.felk.cvut.cz/~hlavac, hlavac@fel.cvut.cz`

Courtesy: T. Pajdla

Outline of the talk:

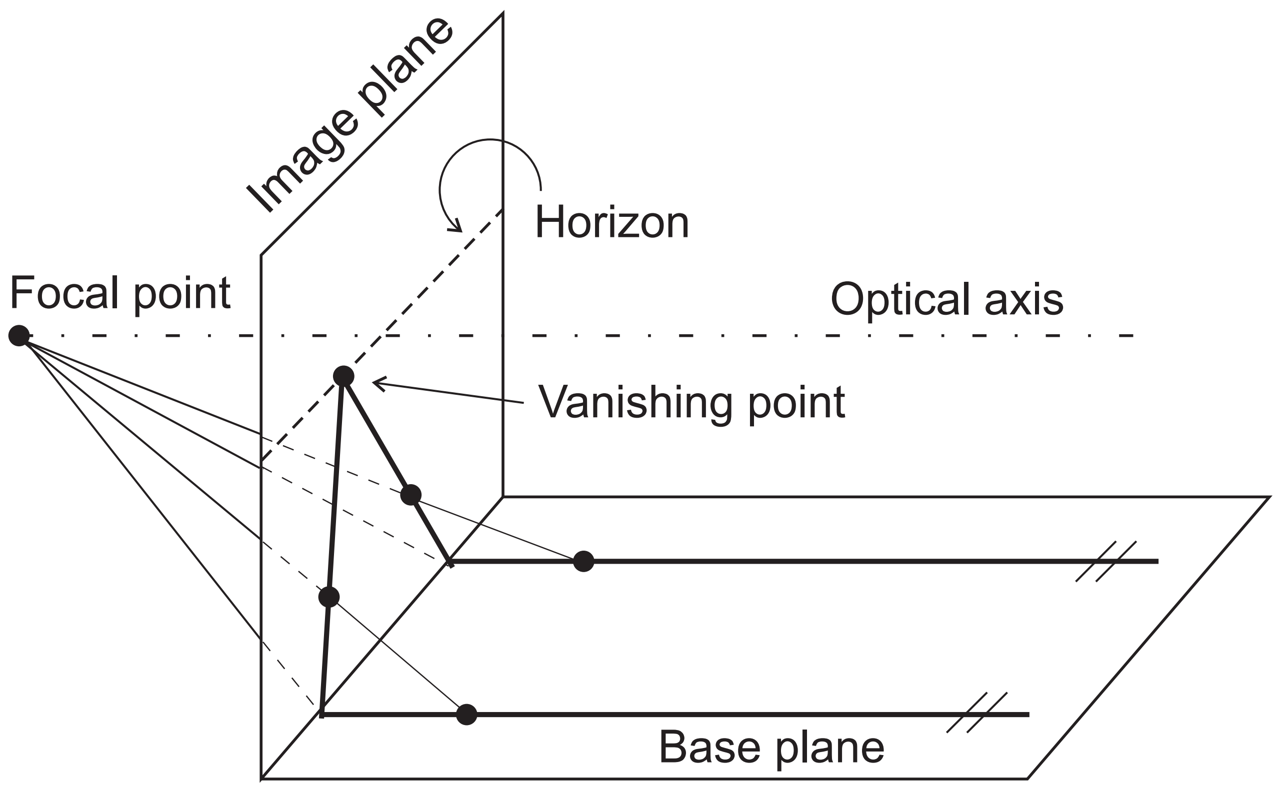
- Projectivity
- Projective space
- Homography
- Projective camera
- Camera calibration
- Radial distortion

Perspective transformation, motivation

Parallel lines do not look like parallel lines under perspective projection.

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).



Multiple view geometry

- 3D points in the scene (and, more generally, lines and other simple geometric objects),
- their camera projections, and
- relations among multiple camera projections of a 3D scene.

Projective space

- Consider $(d+1)$-dimensional vector space without its origin, $\mathbb{R}^{d+1} - \{(0,\ldots,0)\}$.
- Define an equivalence relation

$$\begin{array}{rrcl} & [x_1,\ldots,x_{d+1}]^\top & \equiv & [x'_1,\ldots,x'_{d+1}]^\top \\ \text{iff } \exists\, \alpha \neq 0 : & [x_1,\ldots,x_{d+1}]^\top & = & \alpha \; [x'_1,\ldots,x'_{d+1}]^\top \end{array}$$

- Projective space $\mathbb{P}^d$ is the quotient space of this equivalence relation.
- Points in the projective space are expressed in homogeneous co-ordinates (called also projective coordinates) $\mathbf{x} = [x'_1,\ldots,x'_d,1]^\top$.

Relation between Euclidean and projective spaces

- Consider Euclidean space $\mathbb{R}^d$.
- Non-homogeneous coordinates represent a point in $\mathbb{R}^d$ occupying the plane with equation $x_{d+1} = 1$ in $\mathbb{R}^{d+1}$.
- The one-to-one mapping from the $\mathbb{R}^d$ into $\mathbb{P}^d$

$$[x_1, \ldots, x_d]^\top \to [x_1, \ldots, x_d, 1]^\top$$

- Projective points $[x_1, \ldots, x_d, 0]^\top$ do not have an Euclidean counterpart and represent points at infinity in a particular direction.
- Consider $[x_1, \ldots, x_n, 0]^\top$ as a limiting case of $[x_1, \ldots, x_n, \alpha]^\top$ that is projectively equivalent to $[x_1/\alpha, \ldots, x_n/\alpha, 1]^\top$, and assume that $\alpha \to 0$.
- This corresponds to a point in $\mathbb{R}^d$ going to infinity in the direction of the radius vector $[x_1/\alpha, \ldots, x_d/\alpha] \in \mathbb{R}^d$.

Homogeneous coordinates of hyperplanes in $\mathbb{P}^d$

- A hyperplane in $\mathbb{P}^d$ is represented by the $(d+1)$-vector $\mathbf{a} = [a_1, \ldots, a_{d+1}]^\top$ such that all points $\mathbf{x}$ lying on the hyperplane satisfy $\mathbf{a}^\top \mathbf{x} = 0$ (where $\mathbf{a}^\top \mathbf{x}$ denotes the scalar product).
- Considering the points in the form $\mathbf{x} = [x_1, \ldots, x_d, 1]^\top$ yields the familiar formula $a_1 x_1 + \cdots + a_d x_d + a_{d+1} = 0$.
- The hyperplane defined by d distinct points represented by vectors $\mathbf{x}_1, \ldots, \mathbf{x}_d$ lying on it is represented by a vector $\mathbf{a}$ orthogonal to vectors $\mathbf{x}_1, \ldots, \mathbf{x}_d$. This vector $\mathbf{a}$ can be computed, e.g., by SVD.
- Symmetrically, the point of intersection of d distinct hyperplanes $\mathbf{a}_1, \ldots, \mathbf{a}_d$ is the vector $\mathbf{x}$ orthogonal to them.

Two useful hyperplanes in computer vision

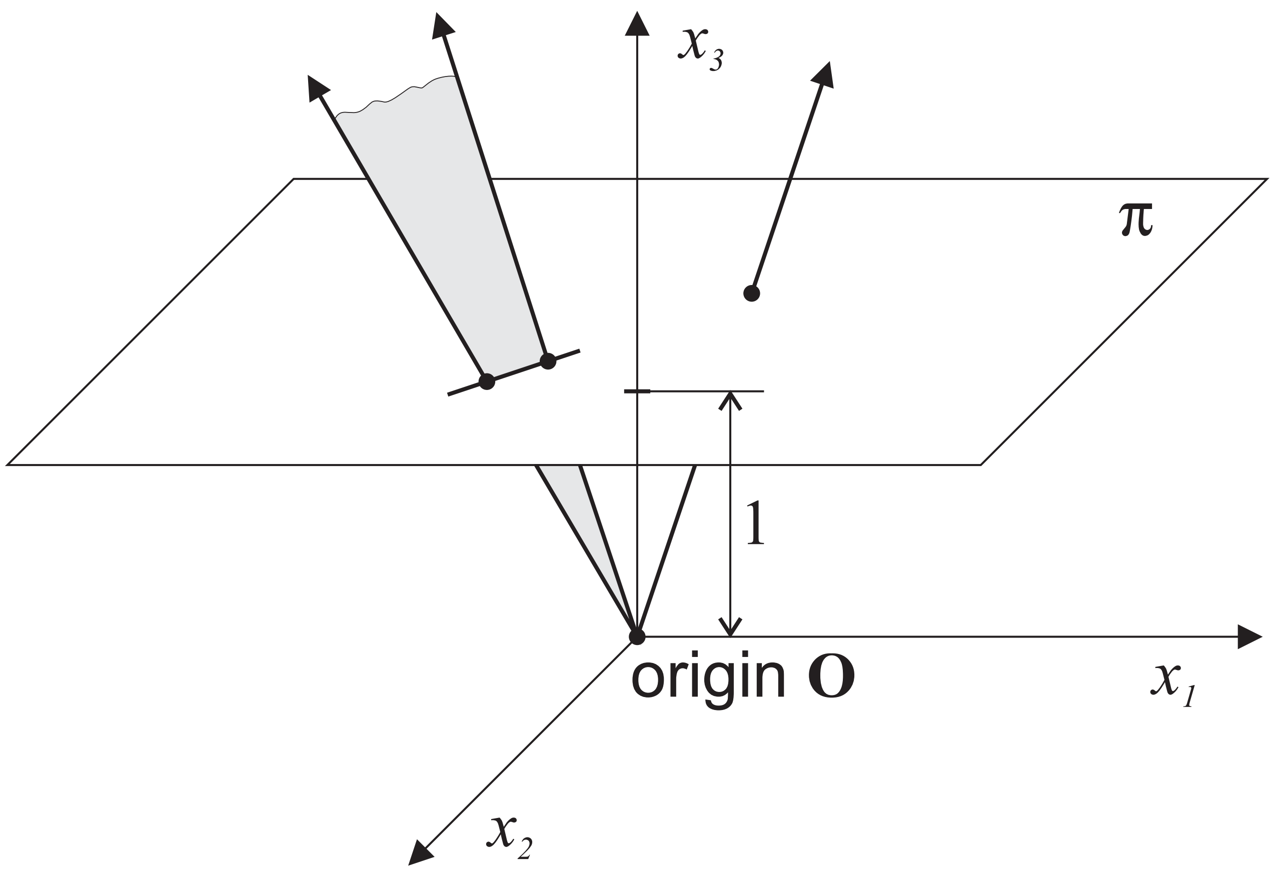
The projective plane $\mathbb{P}^2$.

- We will denote points in $\mathbb{P}^2$ by $\mathbf{u} = [u, v, w]^\top$, lines (hyperplanes) in $\mathbb{P}^2$ by $\mathbf{l}$.
- Line through 2 points: $\mathbf{l} = \mathbf{x} \times \mathbf{y}$.
- Point as intersection of 2 lines: $\mathbf{x} = \mathbf{l} \times \mathbf{m}$.

The projective 3-space $\mathbb{P}^3$.

- We will denote points in $\mathbb{P}^3$ by $\mathbf{X} = [X, Y, Z, W]^\top$.
- In $\mathbb{P}^3$, hyperplanes become planes and one more entity occurs that has no counterpart in the projective plane: a 3D line. The elegant homogeneous representation by 4-vectors, available for points and planes in $\mathbb{P}^3$, does not exist for lines. A 3D line can be represented either by a pair of points lying on it but this representation is not unique, or by a (Grassmann-)Plücker matrix.

Projective space $\mathbb{P}^2$, illustration



Points and lines in $\mathbb{P}^2$ are represented by rays and planes, respectively, which pass through the origin in the Euclidean space $\mathbb{R}^3$.

Homography

- Also projective transformation or co-lineation.
- Co-lineation is any mapping $\mathbb{P}^d \to \mathbb{P}^d$ linear in the embedding space $\mathbb{R}^{d+1}$.
- Co-lineation is defined up to unknown scale as $\mathbf{u}' \simeq H\,\mathbf{u}$, where H is a $(d+1) \times (d+1)$ matrix.
- The transformation maps any triplet of collinear points to a triplet of collinear points (hence one of its names—collineation).
- If H is regular then distinct points are mapped to distinct points.
- In $\mathbb{P}^2$, homography is the most general transformation which maps lines to lines.

Example of an image mapped by a 2D homography

Different form of homography for hyperplanes and points

- It can be derived from the fact that if the original point $\mathbf{u}$ and a hyperplane $\mathbf{a}$ are incident, $\mathbf{a}^\top \mathbf{u} = 0$.
- They have to remain incident after the transformation too, $\mathbf{a}'^\top \mathbf{u}' = 0$.
- Using equation $\mathbf{u}' \simeq H\,\mathbf{u}$, we obtain that $\mathbf{a}' \simeq H^{-\top}\mathbf{a}$, where $H^{-\top}$ denotes the transposed inverse of H.

Two simple homographies useful in computer vision

1. A projection of a planar scene by one pinhole camera are related by a 2D homography. This can be used to rectify images of planar scenes (e.g., building facades) to frontoparallel view.
2. Two images of a 3D scene (planar or non-planar) by two pinhole cameras sharing a single center of projection is a 2D homography. This can be used for stitching panoramic images from photographs

Homography vs. non-homography (1)

- Let us illustrate how the non-homogeneous 2D point $[u, v]^\top$ (e.g., a point in an image) is actually mapped to the non-homogeneous image point $[u', v']^\top$ by H using $\mathbf{u}' \simeq H\,\mathbf{u}$.
- With the components and the scale written explicitly, the equation reads

$$\alpha \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} .$$

Homography vs. non-homography (2)

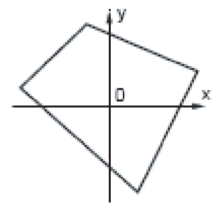
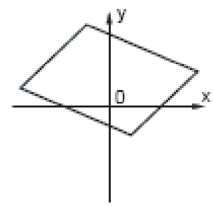
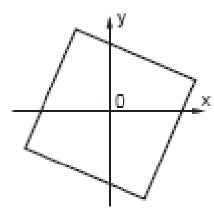
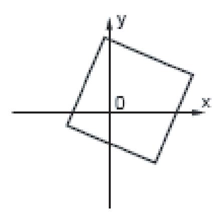
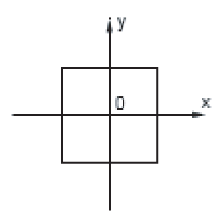
Writing 1 in the third coordinate of $\mathbf{u}'$, we tacitly assume that $\mathbf{u}'$ is not a point at infinity, that is, $\alpha \neq 0$. To compute $[u', v']^\top$, we need to eliminate the scale α. This yields the expression

$$u' = \frac{h_{11}u + h_{12}v + h_{13}}{h_{31}u + h_{32}v + h_{33}}\,, \qquad v' = \frac{h_{21}u + h_{22}v + h_{23}}{h_{31}u + h_{32}v + h_{33}}\,,$$

familiar to people who do not use homogeneous coordinates.

Note that compared to this, the expression $\mathbf{u}' \simeq H\,\mathbf{u}$ is simpler, linear, and can handle the case when $\mathbf{u}'$ is a point at infinity. These are the practical advantages of homogeneous coordinates.

Subgroups of homographies

Name	Constraints on H	2D example	Invariants
projective	$\det H \neq 0$		collinearity tangency cross ratio
affine	$H = \begin{bmatrix} A & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}$ $\det A \neq 0$		projective invariants + parallelism + length ratio on parallels + area ration + linear combinations of vectors centroid
similarity	$H = \begin{bmatrix} sR & -R\mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}$ $R^\top R = I$ $\det R = 1$ $s > 0$		affine invariants + angles + ratio of lengths
metric (Euclidean, isometric)	$H = \begin{bmatrix} R & -R\mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}$ $R^\top R = I$ $\det R = 1$		similarity invariants + length + area (volume)
identity	$H = I$		trivial case everything is invariant

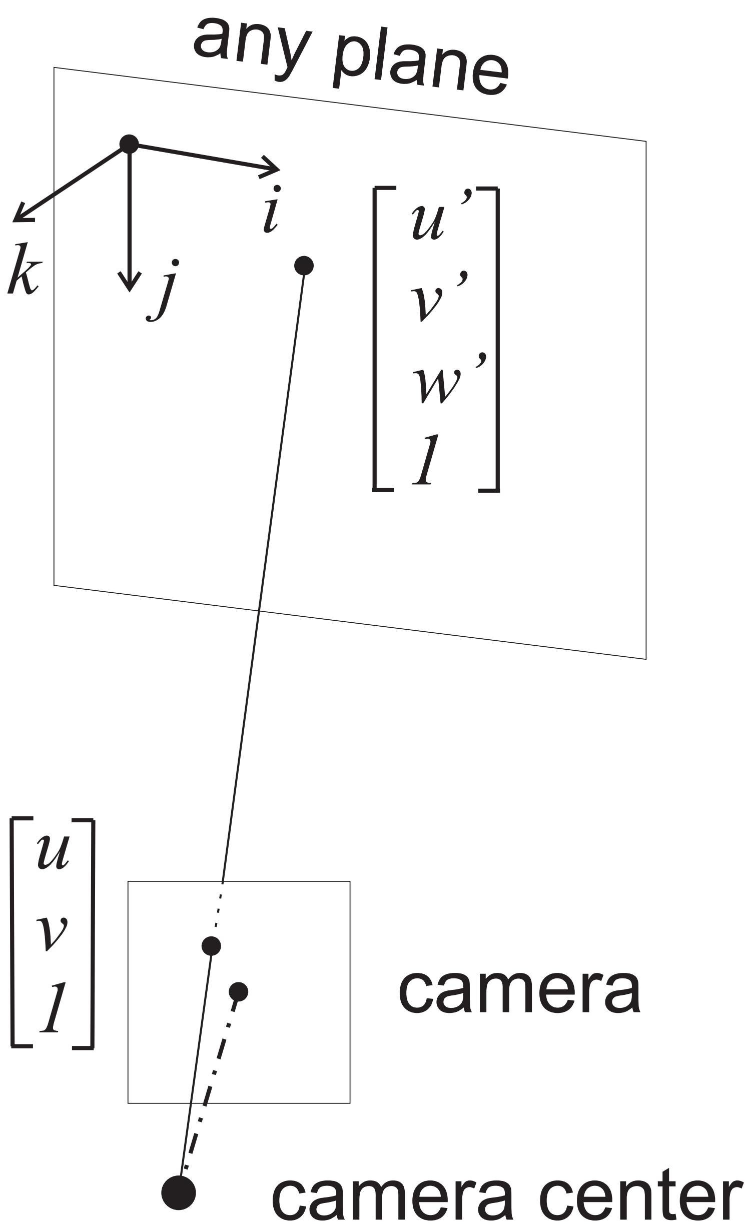
Decomposition of homographies

Any homography can be uniquely decomposed as $H = H_P\, H_A\, H_S$, where

$$H_P = \begin{bmatrix} I & \mathbf{0} \\ \mathbf{a}^\top & b \end{bmatrix}, \quad H_A = \begin{bmatrix} K & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix}, \quad H_S = \begin{bmatrix} R & -R\mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix},$$

- Matrix K is upper triangular.
- Matrices of the form of H_S represent Euclidean transformations.
- Matrices $H_A H_S$ represent affine transformations; thus matrices H_A represent the 'purely affine' subgroup of affine transformations, i.e., what is left of the affine group after removing from it (more exactly, factorizing it by) the Euclidean group.
- Matrices $H_P H_A H_S$ represent the whole group of projective transformations; thus matrices H_P represent the 'purely projective' subgroup of the projective transformation.

Homography

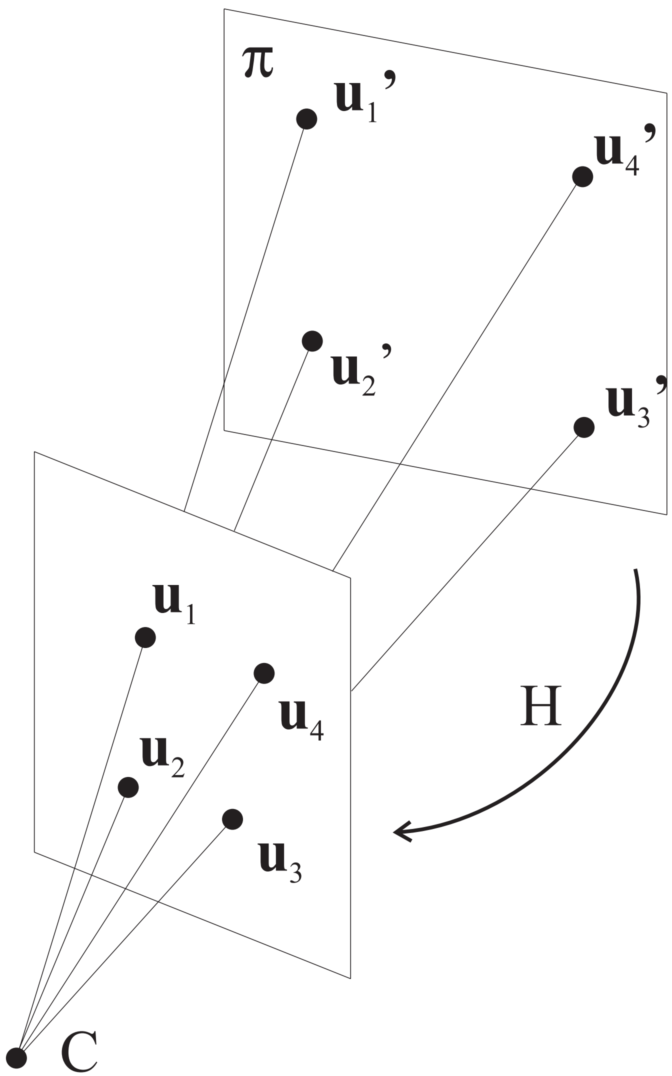


Homography maps a plane to a plane.

$$\alpha \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = H \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

H – $[3 \times 3]$ homography matrix

Example: Distance measurement in a plane

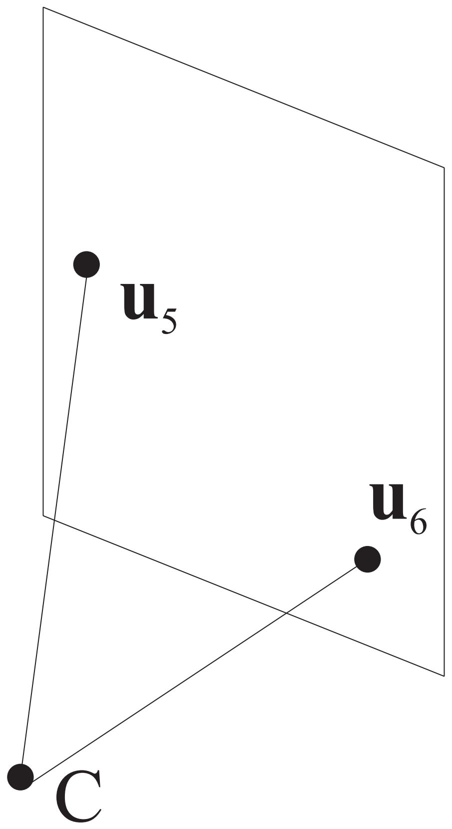


- We know coordinates of four points $\mathbf{u}_i'$, $i = 1, \ldots, 4$ in a plane in which we intend to measure distances.
- We observe images of these four points in image plane, $\mathbf{u}_i$, $i = 1, \ldots, 4$ and get their coordinates.

$$\alpha\, \mathbf{u}_i' = \alpha \begin{bmatrix} u_i' \\ v_i' \\ 1 \end{bmatrix} = H\, \mathbf{u}_i = H \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix}$$

- Courtesy T. Pajdla.

Example (2) Distance between points 5 and 6 ?



- We observe $\mathbf{u}_5$, $\mathbf{u}_6$.
- Calculate $\mathbf{u}'_5$, $\mathbf{u}'_6$.
- Calculate $d = ||\mathbf{u}'_5 - \mathbf{u}'_6||$.

Example (3) Calculation of $\mathbf{x}_5$, $\mathbf{x}_6$

Our plan

$$\begin{aligned} \alpha\, \mathbf{u}'_i &= H\, \mathbf{u}_i \qquad \text{det}(H) \neq 0 \\ \alpha\, H^{-1}\, \mathbf{u}'_i &= \mathbf{u}_i \\ \alpha \neq 0\,, \qquad H^{-1}\, \mathbf{u}'_i &= \alpha\, \mathbf{u}_i \qquad \text{Linear in } \alpha,\, H\,. \end{aligned}$$

Elimination of α_i

$$\begin{bmatrix} \alpha_i\, u'_i \\ \alpha_i\, v'_i \\ \alpha_i \end{bmatrix} = H\, \mathbf{u}_i = \begin{bmatrix} \mathbf{h}_1^\top \\ \mathbf{h}_2^\top \\ \mathbf{h}_3^\top \end{bmatrix} \mathbf{u}_i = \begin{bmatrix} \mathbf{h}_1^\top\, \mathbf{u}_i \\ \mathbf{h}_2^\top\, \mathbf{u}_i \\ \mathbf{h}_3^\top\, \mathbf{u}_i \end{bmatrix}$$

$$\alpha_i = \mathbf{h}_3^\top\, \mathbf{u}_i$$

Example (4) Calculation of $\mathbf{x}_5$, $\mathbf{x}_6$, cont.

$$\begin{aligned} u_i' \, \mathbf{h}_3^\top \, \mathbf{u}_i &= \mathbf{h}_1^\top \, \mathbf{u}_i \\ v_i' \, \mathbf{h}_3^\top \, \mathbf{u}_i &= \mathbf{h}_2^\top \, \mathbf{u}_i \end{aligned}$$

$$\begin{aligned} u_i' \, \mathbf{h}_3^\top \, \mathbf{u}_i - \mathbf{h}_1^\top \, \mathbf{u}_i &= 0 \\ v_i' \, \mathbf{h}_3^\top \, \mathbf{u}_i - \mathbf{h}_2^\top \, \mathbf{u}_i &= 0 \end{aligned}$$

$$\begin{aligned} u_i' \, (h_{31} u_i + h_{32} v_i + h_{33}) - (h_{11} u_i + h_{12} v_i + h_{13}) &= 0 \\ v_i' \, (h_{31} u_i + h_{32} v_i + h_{33}) - (h_{21} u_i + h_{22} v_i + h_{23}) &= 0 \end{aligned}$$

- We obtained two homogeneous linear equations for each point.
- Homography matrix H contains 9 unknowns. However, one of them remains unresolved due to unknown scale.
- Thus 8 unknowns remain $\Rightarrow$ 4 points are needed to calculate them at least.

Example (5)

Calculation of u_5, u_6, system of equations

$$
\begin{array}{ccccccccccc}
-u_1h_{11} & -y_1h_{12} & -h_{13} & & & & +u'_1u_1h_{31} & +u'_1v_1h_{32} & +u'_1h_{33} & =0 \\
 & & & -x_1h_{21} & -v_1h_{22} & -h_{23} & +v'_1u_1h_{31} & +v'_1v_1h_{32} & +v'_1h_{33} & =0 \\
 & & & & & \vdots & & & & \\
-u_4h_{11} & -v_4h_{12} & -h_{13} & & & & +u'_4x_ih_{31} & +u'_4v_4h_{32} & +u'_4h_{33} & =0 \\
 & & & -u_4h_{21} & -v_4h_{22} & -h_{23} & +v'_4u_4h_{31} & +v'_4v_4h_{32} & +v'_4h_{33} & =0
\end{array}
$$

$$A\,\mathbb{H} = 0$$

- Matrix A $[8 \times 9]$ contains measured data.
- Vector $\mathbb{H}$ $[9 \times 1]$ contains unknowns in the homography matrix.

Example (6) Solution of underconstrained homogenous system of linear equations

- Linear system $A\,\mathbb{H} = 0$, has 9 unknowns and only 8 equations.
- There is always trivial solution $\mathbb{H} = 0$ because $A \cdot 0 = 0$.
- We are interested in $\mathbb{H} \neq 0$. Thus $\mathbb{H}$ has to have rank < 9.
- Why?

$$
\begin{aligned}
[\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_9][h_1, h_2, \ldots, h_9]^\top &= 0 \\
\mathbf{a}_1 h_1 + \mathbf{a}_2 h_2 + \ldots + \mathbf{a}_9 h_9 &= 0
\end{aligned}
$$

This is linear combination of vectors $\mathbf{a}_i$. If it is zero $\Rightarrow$ it is linearly dependent.

- We look for matrix A of rank 8.

Example (7) Zero space

- $A\,\mathbb{H} = 0$, i.e., A maps to zero $\Rightarrow$ $\mathbb{H}$ is the right zero space.
- The zero space can be found by SVD.
- If noise is present and more points N available, $N > 4$ then
  - $A\ [2N \times 9]$, $\mathbb{H}\ [9 \times 1]$.
  - rank$(A) = 8 \Rightarrow \exists$ infinite 1D space satisfying the equation. We choose solution with Euclidean norm $||\mathbb{H}|| =$0.
  - Real data with noise provide the full rank $\tilde{A} \in \mathbb{R}^{[2n\times 9]}$ … rank$(\tilde{A}) =$ 9.

- Task formulation: We seek A, $A\mathbb{H} = 0$, rank$(A) =$ 8 with minimal $||\tilde{A} - A||_F$ (Frobenius norm. i.e., $\sum_i \sum_j a_{ij}^2$).

SVD—Singular Value Decomposition

- SVD is a linear algebra technique for solving linear equations in the least square sense. SVD works for general matrices (including singular matrices or matrices numerically close to singular). SVD is contained, e.g., in MATLAB.
- Any $m \times n$ matrix A, $m \geq n$ can be factorized as $A = UDV^\top$.
- U has orthonormal columns, D is non-negative diagonal, and $V^\top$ has orthonormal rows.
- SVD locates the closest possible solution in a least square sense.
- Often 'closest' singular matrix to the original matrix A is needed. This decreases the rank from n to $n - 1$. How? Replace the smallest diagonal element of D by zero. This new matrix is the closest approximation to A with respect to the Frobenius norm (which is calculated as a sum of the squared values of all matrix elements).

EXAMPLE (8) SVD applied to $\tilde{A}$

$$\mathsf{SVD}(\tilde{A}) = U\tilde{D}V^T = \underset{\mathbb{R}^{1\times n}}{[u_1, \ldots, u_m]} \underset{\mathbb{R}^{n\times n}}{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}} \underset{\mathbb{R}^{n\times 1}}{\begin{bmatrix} v_1^\top \\ \vdots \\ v_n^\top \end{bmatrix}}$$

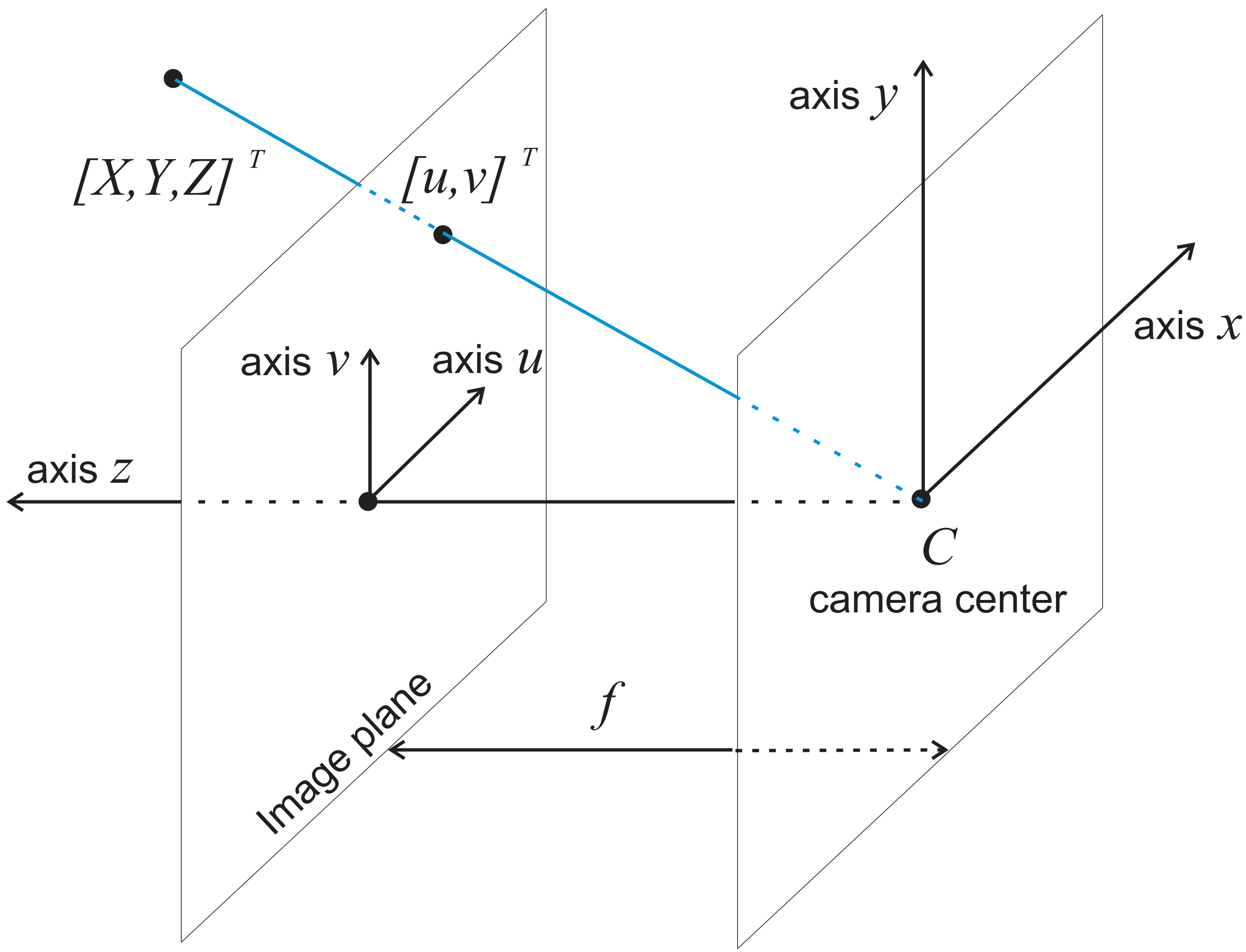
- We zero the smallest singular value σ_n in matrix $\tilde{D}$. (Note: the eigenvalue is a special case of the singular value for a square matrix).
- Observation: $||\tilde{A} - A||_F = \sigma_n^2$ which is minimal.

Projective camera, notation

- Image points will be denoted by lower-case letters
  - In Euclidean (non-homogeneous) coordinates $\mathbf{u} = [u, v]^\top$ or
  - In homogeneous coordinates $\mathbf{u} = [u, v, w]^\top$.
- 3D scene points will be denoted by upper-case letters
  - In Euclidean coordinates $\mathbf{X} = [X, Y, Z]^\top$
  - In homogeneous coordinates $\mathbf{X} = [X, Y, Z, W]^\top$

Subscripts will be used to distinguish different coordinate systems if necessary.

Perspective (pinhole) camera in a canonical configuration



f – focal length.

$$u = f \frac{X}{Z}, \qquad v = f \frac{Y}{Z}.$$

Linear in f, X, Y.

Nonlinear in Z !!

Points, lines in the image

POINTS

- An image point $[u, v]^\top$ represents a spatial direction $\mathbf{u} = [u_1, u_2, u_3]^\top$.
- $u = f\,\frac{u_1}{u_3}\,, \quad v = f\,\frac{u_2}{u_3}$.
- $\alpha \neq 0\,, \quad \alpha\,\mathbf{x} \sim \mathbf{u}$.
- Ideal point $u_3 = 0$.

LINES

- An image line $[u, v, f]^\top$ represents a spatial plane $\mathbf{n} = [n_1, n_2, n_3]^\top$.
- The equation of a plane: $n_1\,u + n_2\,v + n_3\,f = 0$.
- $\alpha \neq 0\,, \quad \alpha\,\mathbf{n} \sim \mathbf{n}$.
- Ideal line $n_1 = n_2 = 0$.

Camera projection matrix

In homogeneous coordinates. $$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} .$$

$\mathbf{u} = M\,\mathbf{X}$. Projection matrix $M = [Q, \mathbf{q}] = \begin{bmatrix} \mathbf{q}_1^\top & q_{14} \\ \mathbf{q}_2^\top & q_{24} \\ \mathbf{q}_3^\top & q_{34} \end{bmatrix}$.

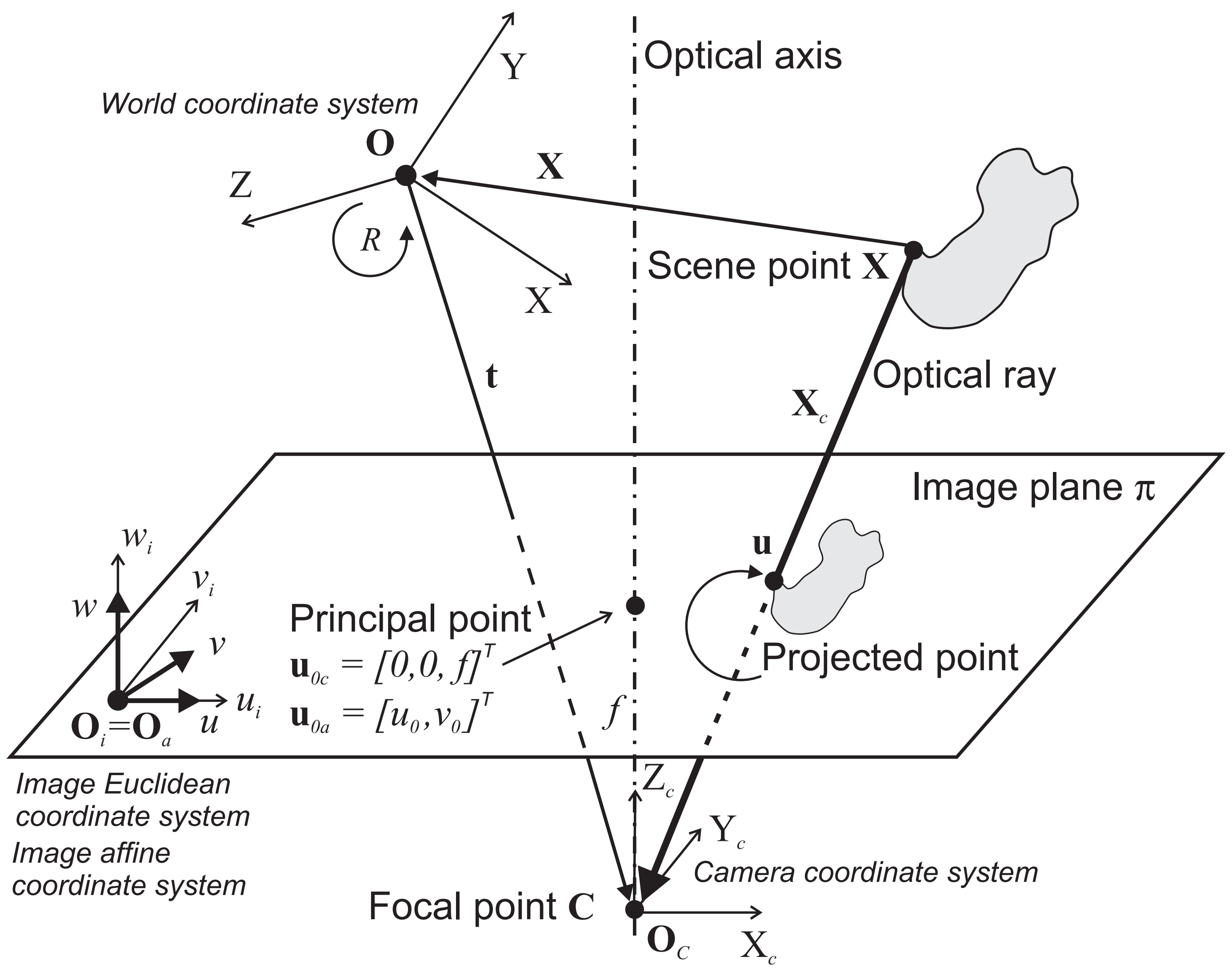
Optical center $C = -Q^{-1}\,\mathbf{q}$.

Optical axis q_3.

Optical ray (or direction) $\mathbf{d} = Q^{-1}\mathbf{u}$.

Optical plane $\mathbf{p} = Q^{-1}\mathbf{n}$.

Single perspective camera, a pinhole model



- World Euclidean coordinate system.
- Camera Euclidean coordinate system (subscript $_c$).
- Image Euclidean coordinate system (subscript $_i$).
- Image affine coordinate system (subscript $_a$).

The camera performs a linear transformation from $\mathbb{P}^3$ to $\mathbb{P}^2$. Optical ray reflected from a scene point $\mathbf{X}$ or originating from a light source hits the image plane at the projected point $\mathbf{u}$.

Factorization of the projective transformation

The projective transformation in the general case can be factorized into three simpler transformations which correspond to three transitions between above mentioned four different coordinate systems.

World $\rightarrow$ camera centered coordinate system .

Projection of the 3D scene point expressed in the camera centered coordinate system to the point in the image plane in the image coordinate system.

Affine mapping in the image plane from the image Euclidean coordinate system $\rightarrow$ the image affine coordinate system.

World to camera centered coordinates

The transformation constitutes transition from the (arbitrary) world coordinate system $(\mathbf{O}; X, Y, Z)$ to the camera centered coordinate system $(\mathbf{O}_c;$ $X_c, Y_c, Z_c)$.

$$\mathbf{X}_c = R\,(\mathbf{X} - \mathbf{t})\ .$$

---

6 degrees of freedom, 3 rotations, 3 translations.

Parameters R and $\mathbf{t}$ are called extrinsic camera calibration parameters.

... in homogeneous coordinates

We already know from that this can be done by a subgroup of homographies H_S

$$\mathbf{X}_c = \begin{bmatrix} R & -R\,\mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \mathbf{X}\,.$$

Projection

- The $\mathbb{R}^3 \longrightarrow \mathbb{R}^2$ projection in non-homogeneous coordinates gives two equations non-linear in Z_c

$$u_i = \frac{X_c\, f}{Z_c}\,, \qquad v_i = \frac{Y_c\, f}{Z_c}\,,$$

where f is the focal length.

- Embedding in the projective space. Projection $\mathbb{P}^3 \longrightarrow \mathbb{P}^2$ writes as

$$\mathbf{u}_i \simeq \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{X}_c\,.$$

Camera with normalized image plane also camera in canonical configuration

Special case: a camera with the focal length $f = 1$.

$$\mathbf{u}_i \simeq \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{X}_c \,.$$

Affine mapping in the image plane

- It is advantageous to gather all parameters intrinsic to a camera (the focal length f is one of them) into a 3×3 matrix K, called the intrinsic calibration matrix.
- Matrix K is upper triangular and expresses the mapping $\mathbb{P}^2 \to \mathbb{P}^2$ which is a special case of the affine transformation.

$$\mathbf{u} \simeq K\,\mathbf{u}_i = \begin{bmatrix} f & s & -u_0 \\ 0 & g & -v_0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{u}_i \,.$$

Intrinsic calibration matrix parameters

- Parameter f (focal length) gives the scaling along the u axis.
- Parameter g gives scaling along the v axis.

  Often, both values are equal to the focal length, $f = g$.
- Parameter s (shear) gives the degree of shear of the coordinate axes in the image plane. It is assumed that the v axis of the image affine coordinate system is co-incident with the v_i axis of the image Euclidean coordinate system. The value s shows how far the u axis is slanted in the direction of axis v. The shear s is introduced in practice to cope with distortions caused by, e.g., placing a photosensitive chip off-perpendicular to the optical axis during camera assembly.

Projection in its full generality

It is a product of the three factors derived above

$$\mathbf{u} \simeq K \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & -R\,\mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \mathbf{X}\,.$$

The product of the second and the third factor exhibits a useful internal structure;

$$\mathbf{u} \simeq K \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & -R\,\mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \mathbf{X} = K \left[R \mid -R\,\mathbf{t} \right] \mathbf{X} = M\,\mathbf{X}\,.$$

Projection matrix

In homogeneous coordinates, the perspective projection can be expressed linearly using a single 3×4 matrix M, **projection matrix**(or camera matrix). The leftmost 3×3 submatrix of M describes a rotation and the rightmost column gives the translation.

The delimiter $|$ denotes that the matrix is composed of two submatrices. M contains all intrinsic and extrinsic parameters because

$$M = K \left[R \mid -R\,\mathbf{t} \right] . \tag{1}$$

These parameters can be obtained by decomposing M to K, R, and $\mathbf{t}$—this decomposition is unique. Denoting $M = [A \mid \mathbf{b}]$, we have $A = KR$ and $\mathbf{b} = -A\mathbf{t}$. Clearly, $\mathbf{t} = -A^{-1}\mathbf{b}$. Decomposing $A = KR$ where K is upper triangular and R is rotation can be done by RQ-decomposition, similar to the better known QR-decomposition.

Single camera calibration, an overview

Intrinsic parameters only - seeking matrix K.

Intrinsic + extrinsic parameters - seeking matrix M.

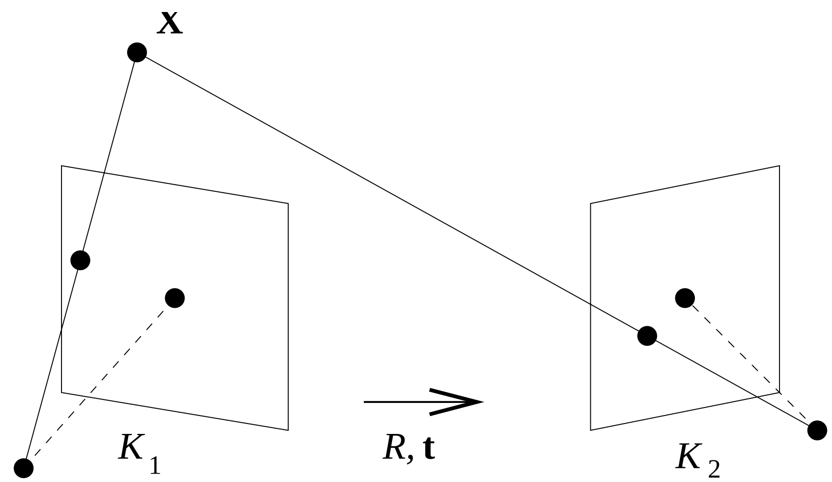
1. Known scene: A set of n non-degenerate (not co-planar) points in the 3D world (e.g., a calibration object), and the corresponding 2D image points are known.

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

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

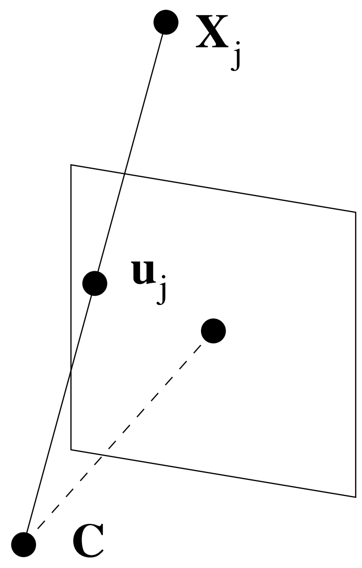
2. Unknown scene: More views are needed to calibrate the camera. The intrinsic camera parameters will not change for different views, and the correspondence between image points in different views must be established.

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 $\mathbf{X} = [x, y, z]^\top$ and 2D image point $[u, v]^\top$ 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 $= M$.

Separation of extrinsic parameters from M

Given: projection matrix M

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

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

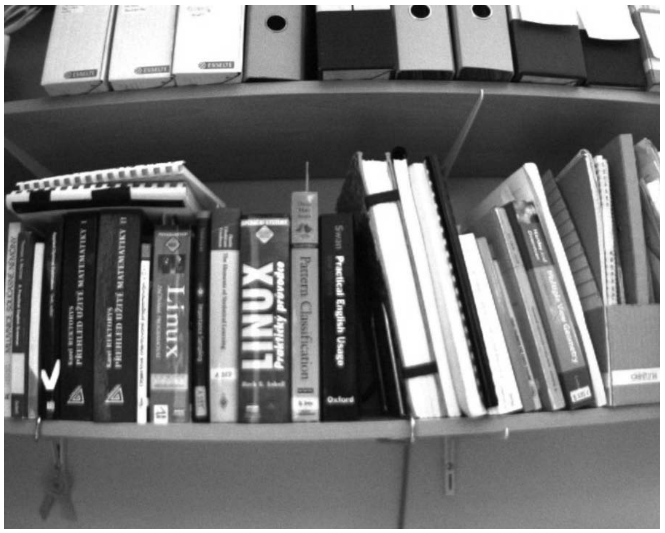
The 3×3 submatrix is denoted as A, and the rightmost column as $\mathbf{b}$.

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, a practical view



barrel

no distortion

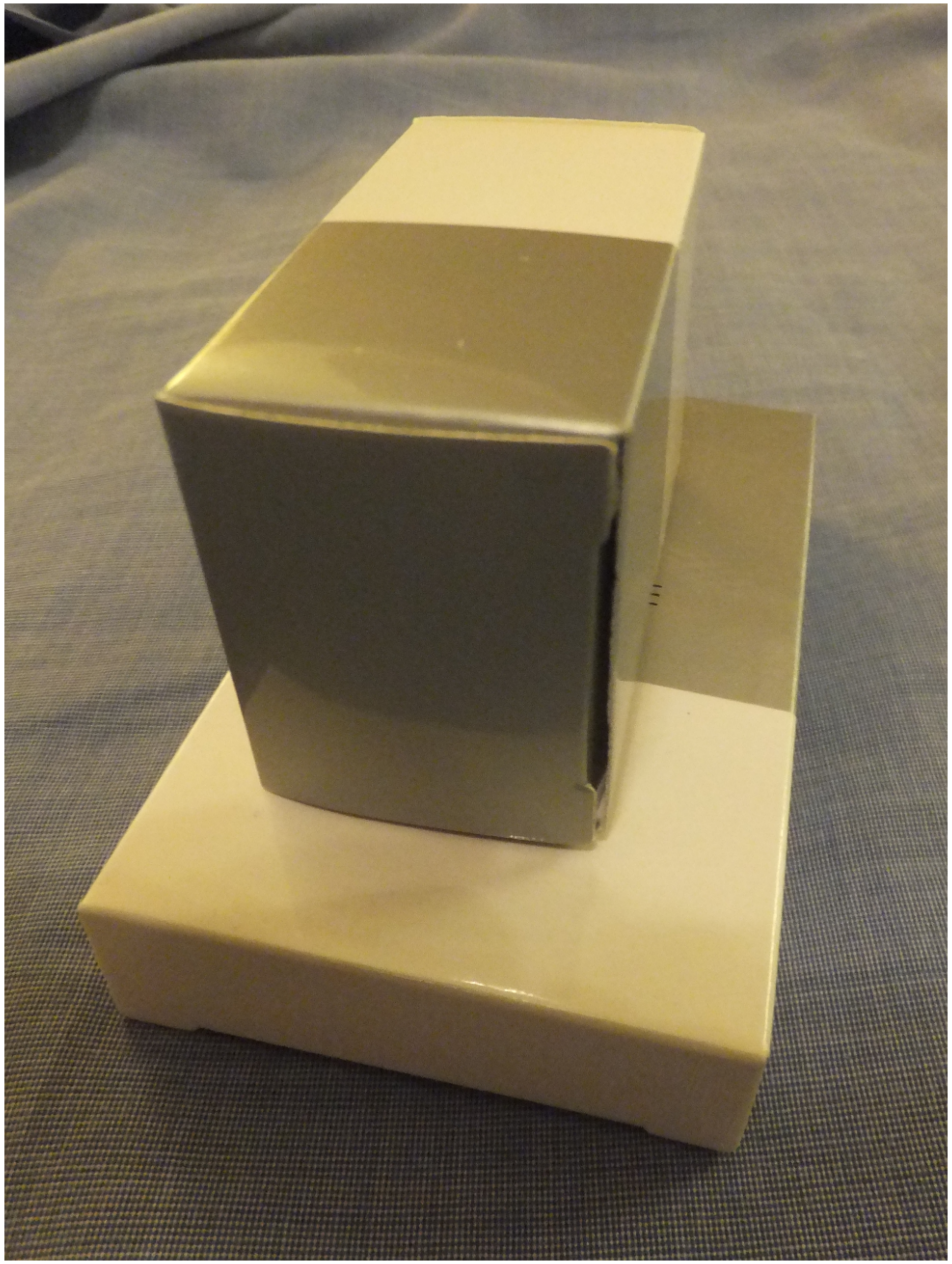
pincushion

Q: How to recognize that a significant radial distortion is present?

A: Straight lines are not mapped to straight lines any more.

Is the distortion radial or perspective? (1)

Is the distortion radial or perspective? (2)



Undoing radial distortion

- A dominant geometric distortion. It is more pronounced with wide-angle lenses.
- (x', y') are coordinates measured in the image (uncorrected); (x, y) are corrected coordinates; (x_0, y_0) are coordinates of the principal point; (Δ_x, Δ_y) are elements of the correction and r is a radius, $r = \sqrt{(x' - x_0)^2 + (y' - y_0)^2}$.
- The distortion is approximated by an even-order polynomial (why?), often only of the second order,

$$\begin{aligned}\Delta_x &= (x' - x_0)\left(\kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6\right) , \\ \Delta_y &= (y' - y_0)\left(\kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6\right) .\end{aligned}$$