Reconnaissance d’objets et vision artificielle 2011

Instance-level recognition I. - Camera geometry and image alignment

Josef Sivic

http://www.di.ens.fr/~josef

INRIA, WILLOW, ENS/INRIA/CNRS UMR 8548

Laboratoire d’Informatique, Ecole Normale Supérieure, Paris

With slides from: S. Lazebnik, J. Ponce, and A. Zisserman http://www.di.ens.fr/willow/teaching/recvis11/

Lecture	Date	Topic
1	Sep 27	Introduction (J. Ponce, I. Laptev, J. Sivic); Instance-level recognition I. - Camera geometry (J. Sivic / I. Laptev);
2	Oct 4	Instance-level recognition II. - Local invariant features (C. Schmid) *Assignment 1 out.*
3	Oct 11	Instance-level recognition III. - Correspondence, efficient visual search (J. Sivic) Assignment 2 out. Final projects out.
4	Oct 18	Very large scale image indexing. Bag-of-feature models for category-level recognition (C. Schmid) ***Assignment 1 due.***
5	Oct 25	Sparse coding and dictionary learning for image analysis (J. Ponce) Category-level localization I. *(J. Sivic)* **Assignment 2 due.** *Assignment 3 out.*
6	Nov 1	***(Holidays, no lecture.)***
7	Nov 8	Neural networks; Optimization methods (N. Le Roux) ***Assignment 3 due.***
8	Nov 15	Category-level localization II. - Efficient fitting of pictorial structures using distance transform (J. Sivic / I. Laptev) Human pose estimation (I. Laptev)
9	Nov 22	Motion and human actions (I. Laptev)
10	Nov 29	Face detection and recognition, segmentation (C. Schmid)
11	Dec 6	Scenes and objects (I. Laptev, J. Sivic)
12	Dec 9 Dec 12	Final project presentations and evaluation (I. Laptev, J. Sivic) *Note unusual time: Friday Dec 9 (14:00-17:00) / Monday Dec 12 (10:00-13:00)*

Object recognition and computer vision 2011

Class webpage:

http://www.di.ens.fr/willow/teaching/recvis11/

Grading:

- 3 programming assignments (60%)
  - Panorama stitching
  - Image classification
  - Basic face detector

- Final project (40%)

More independent work, resulting in the report and a class presentation.

Matlab tutorial

Friday 30/09/2011 at 10:30-12:00.

The tutorial will be at 23 avenue d'Italie - Salle Rose.

Come if you have no/limited experience with Matlab.

Research

Both WILLOW (J. Ponce, I. Laptev, J. Sivic) and LEAR (C. Schmid) groups are active in computer vision and visual recognition research.

http://www.di.ens.fr/willow/

http://lear.inrialpes.fr/

with close links to SIERRA – machine learning (F. Bach)

http://www.di.ens.fr/sierra/

There will be master internships available. Talk to us if you are interested.

Outline

Part I - Camera geometry – image formation

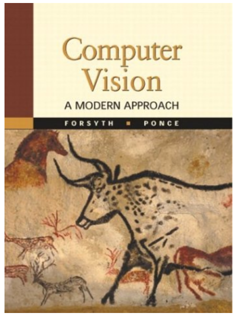
- Perspective projection
- Affine projection
- Projection of planes

Part II - Image matching and recognition with local features

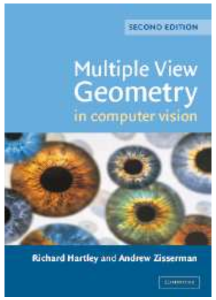
- Correspondence
- Semi-local and global geometric relations
- Robust estimation – RANSAC and Hough Transform

Reading: Part I. Camera geometry

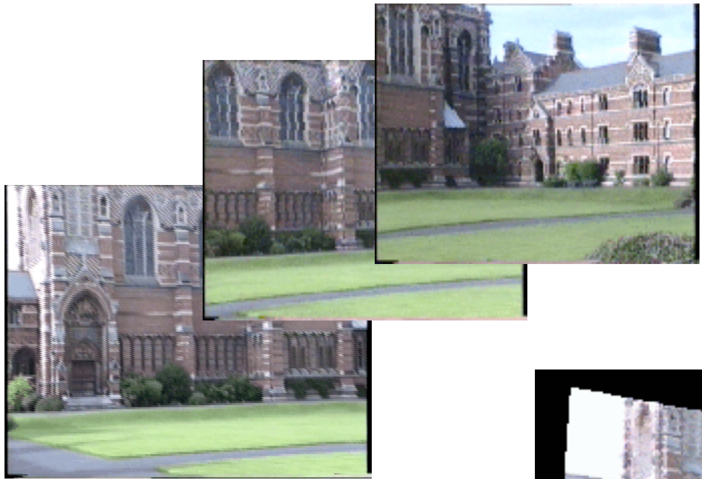
Forsyth&Ponce – Chapters 1 and 2

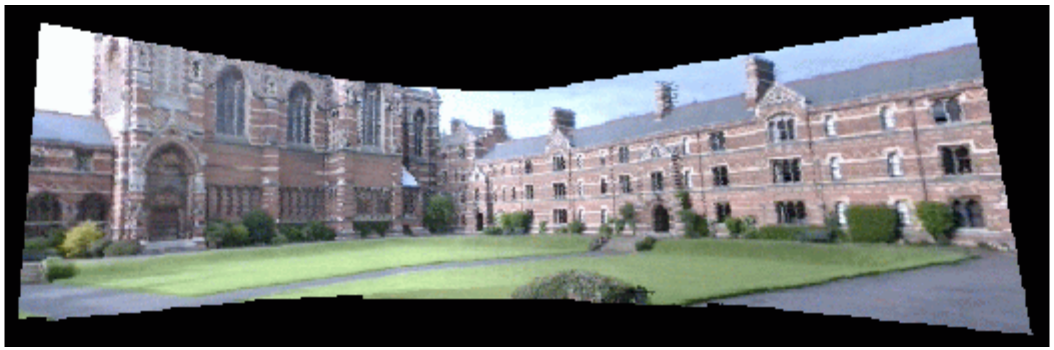


Hartley&Zisserman – Chapter 6: “Camera models”



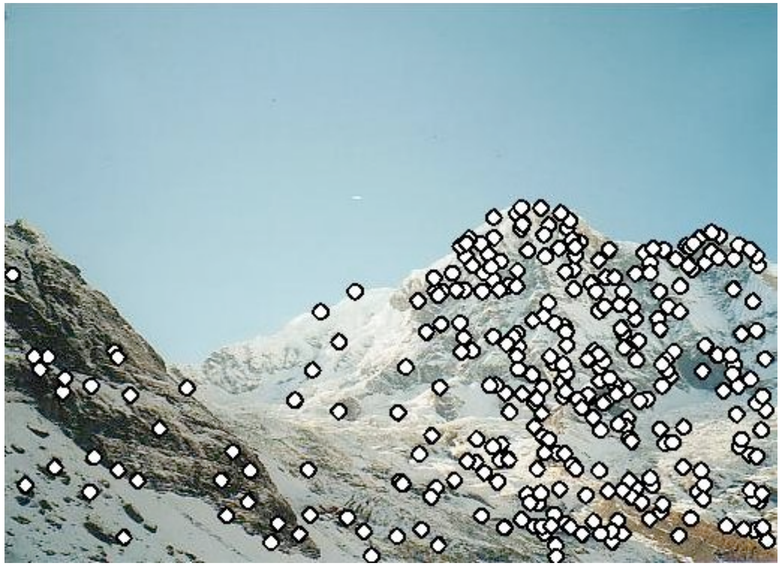
Motivation: Stitching panoramas

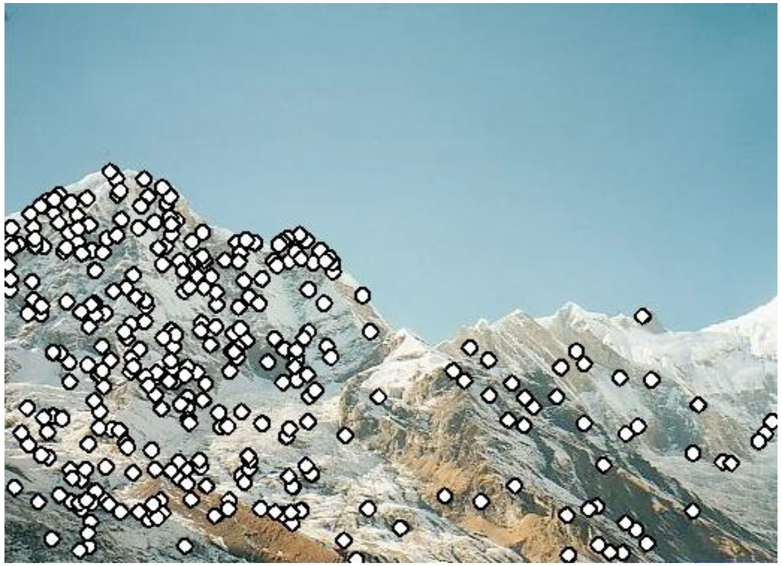




Feature-based alignment outline

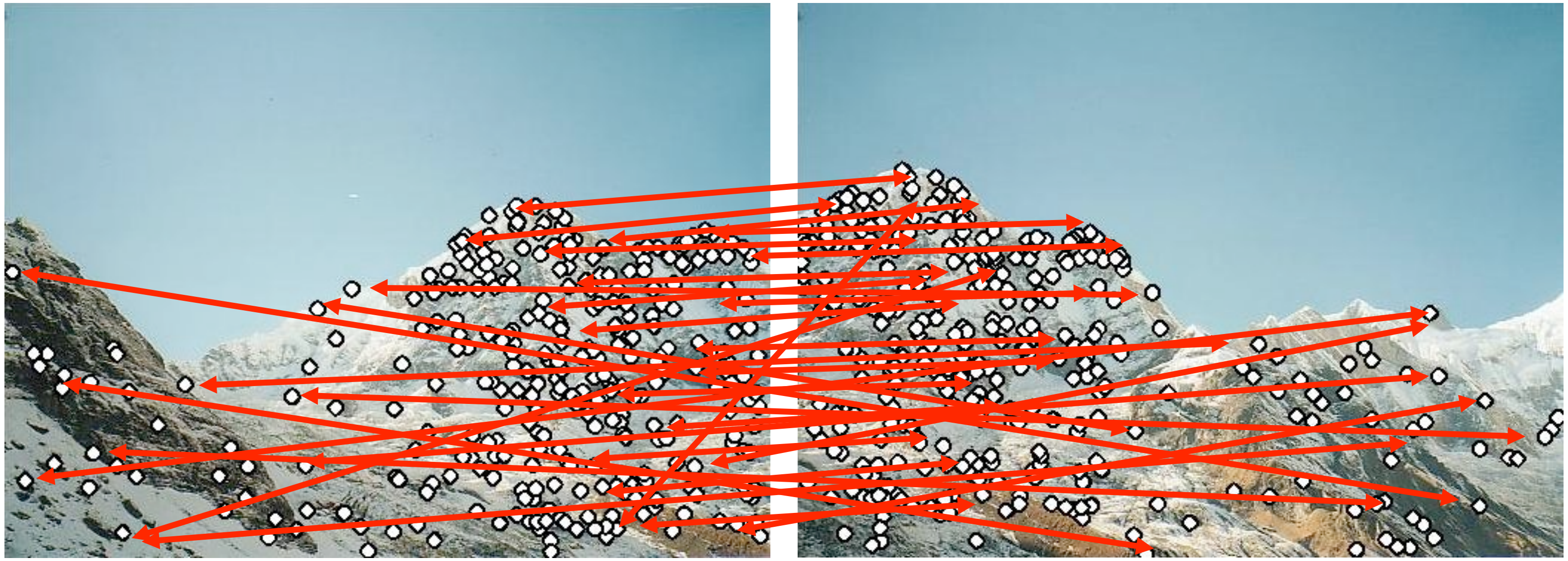
Feature-based alignment outline





Extract features

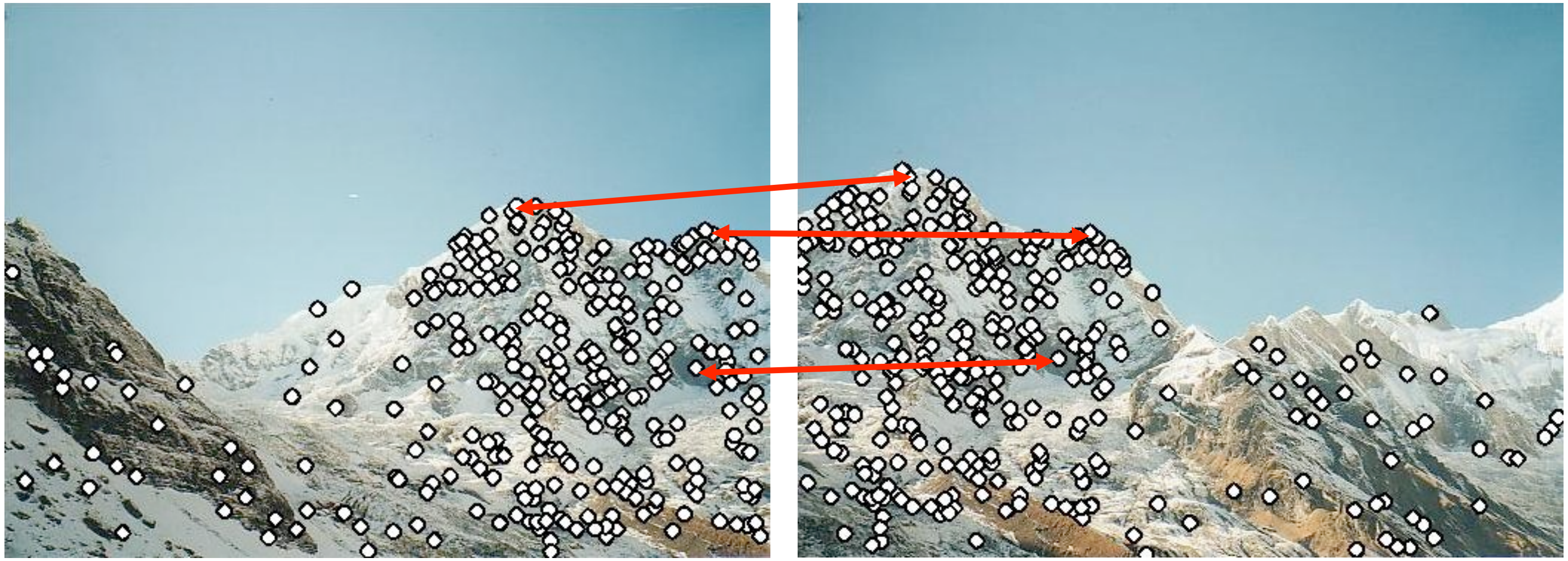
Feature-based alignment outline



Extract features

Compute *putative matches*

Feature-based alignment outline



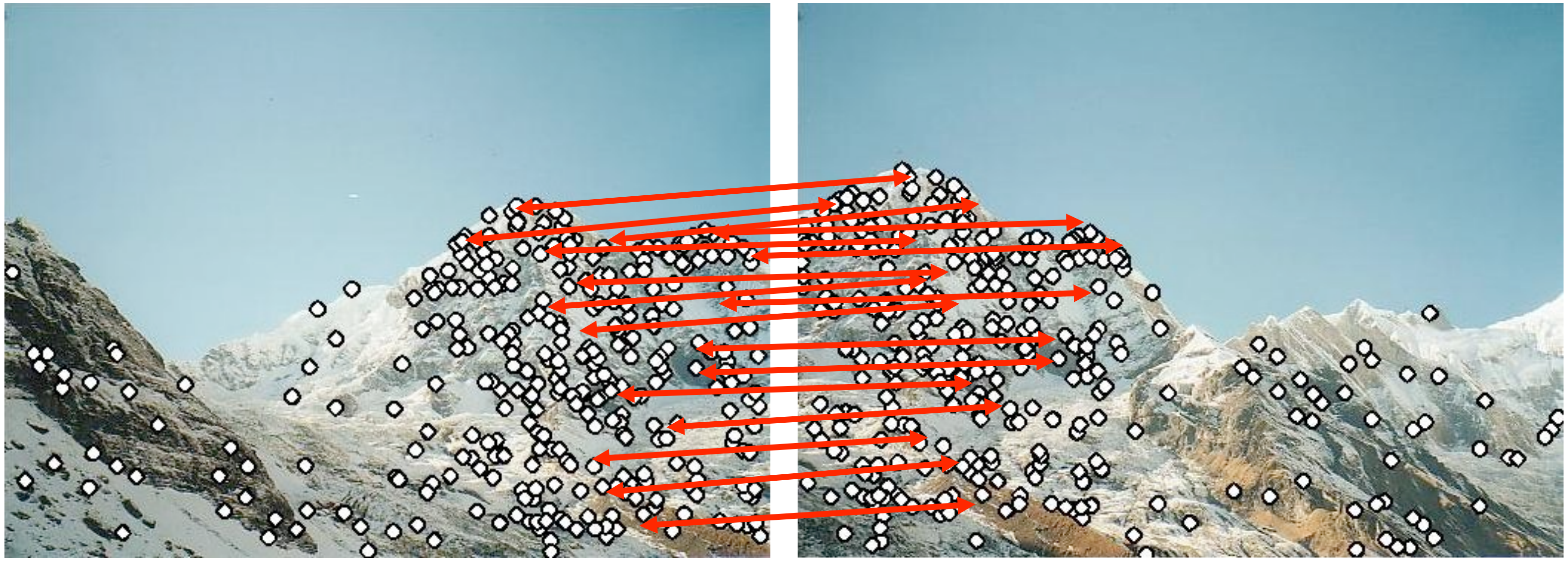
Extract features

Compute *putative matches*

Loop:

- *Hypothesize* transformation *T* (small group of putative matches that are related by *T*)

Feature-based alignment outline



Extract features

Compute *putative matches*

Loop:

- *Hypothesize* transformation *T* (small group of putative matches that are related by *T*)
- *Verify* transformation (search for other matches consistent with *T*)

Feature-based alignment outline

Extract features

Compute *putative matches*

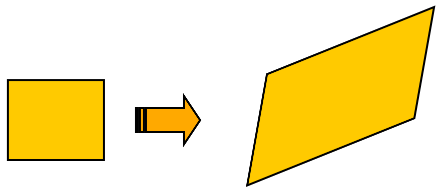
Loop:

- *Hypothesize* transformation *T* (small group of putative matches that are related by *T*)
- *Verify* transformation (search for other matches consistent with *T*)

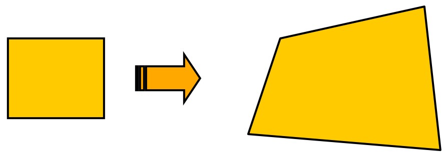
2D transformation models

Similarity
(translation,
scale, rotation)

Affine

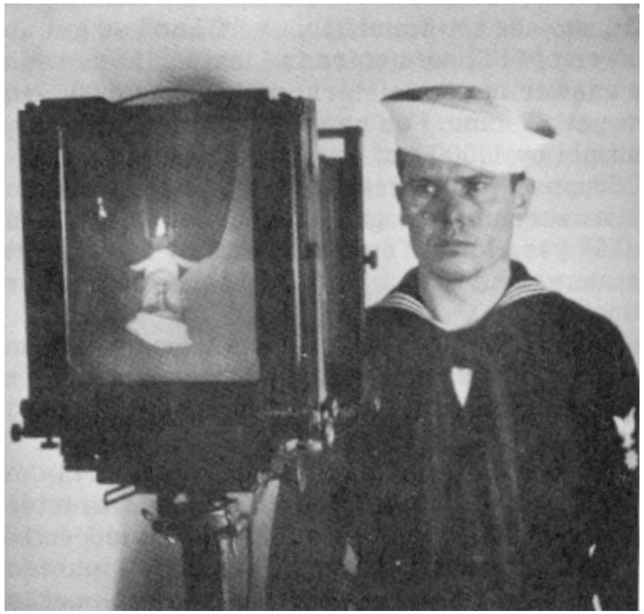


Projective
(homography)

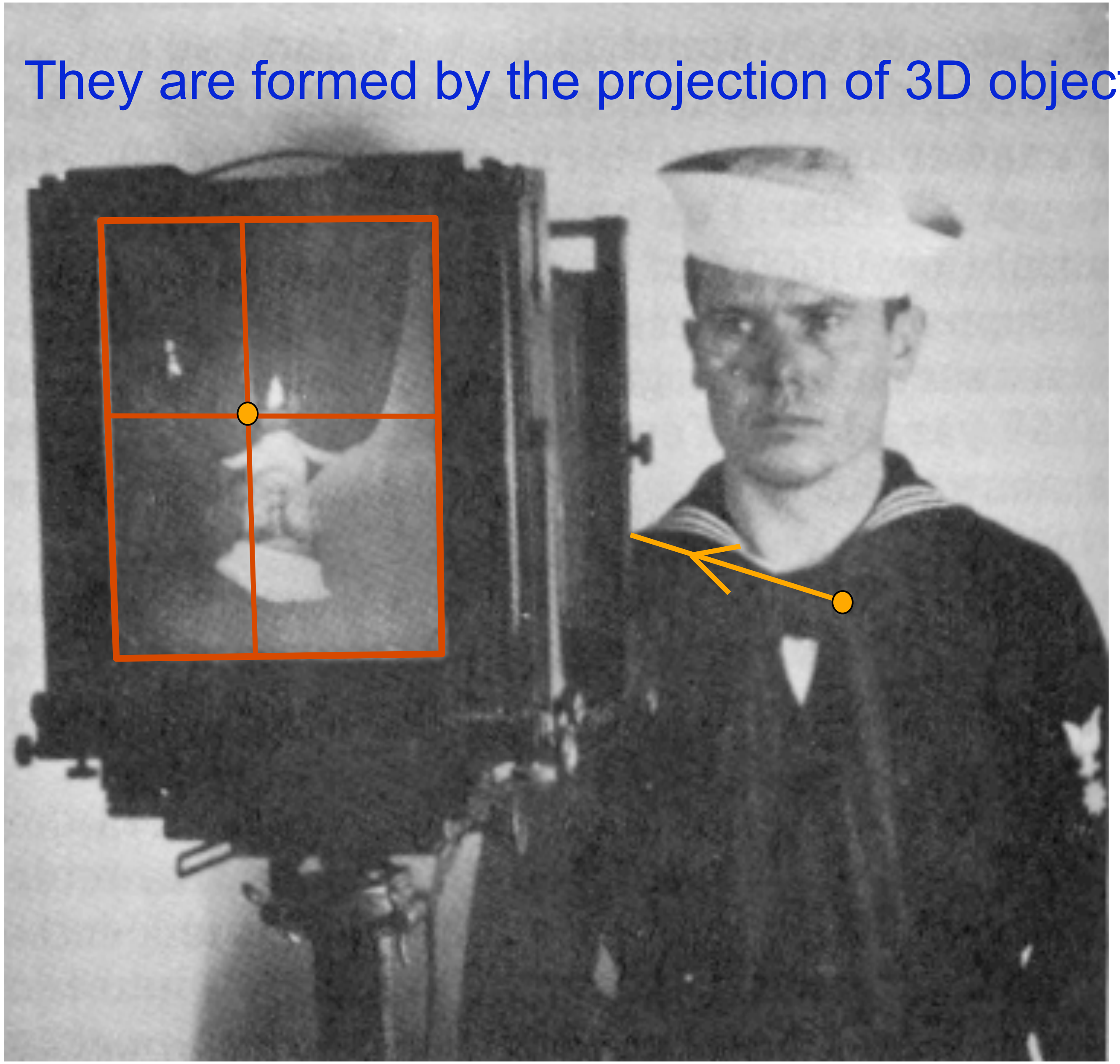


Why these transformations ???

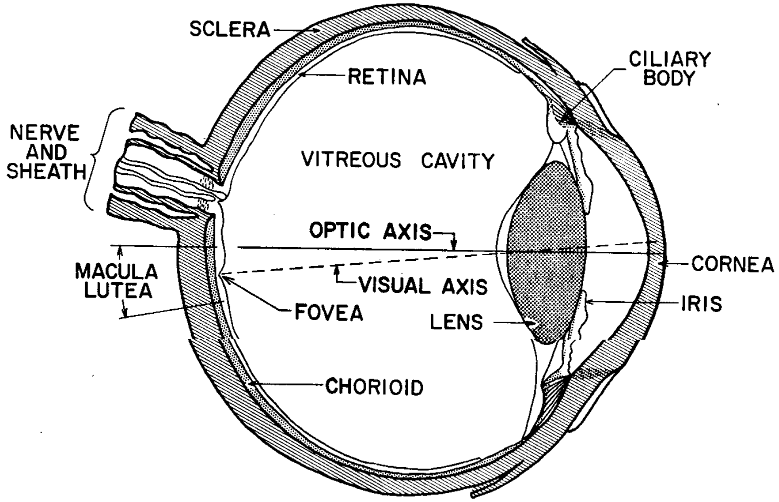
Camera geometry



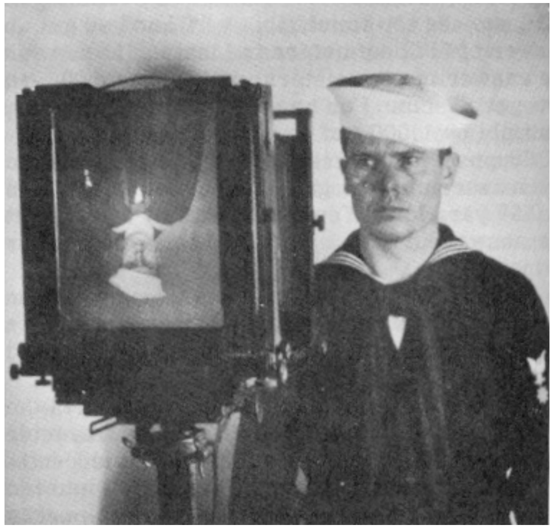
They are formed by the projection of 3D objects.



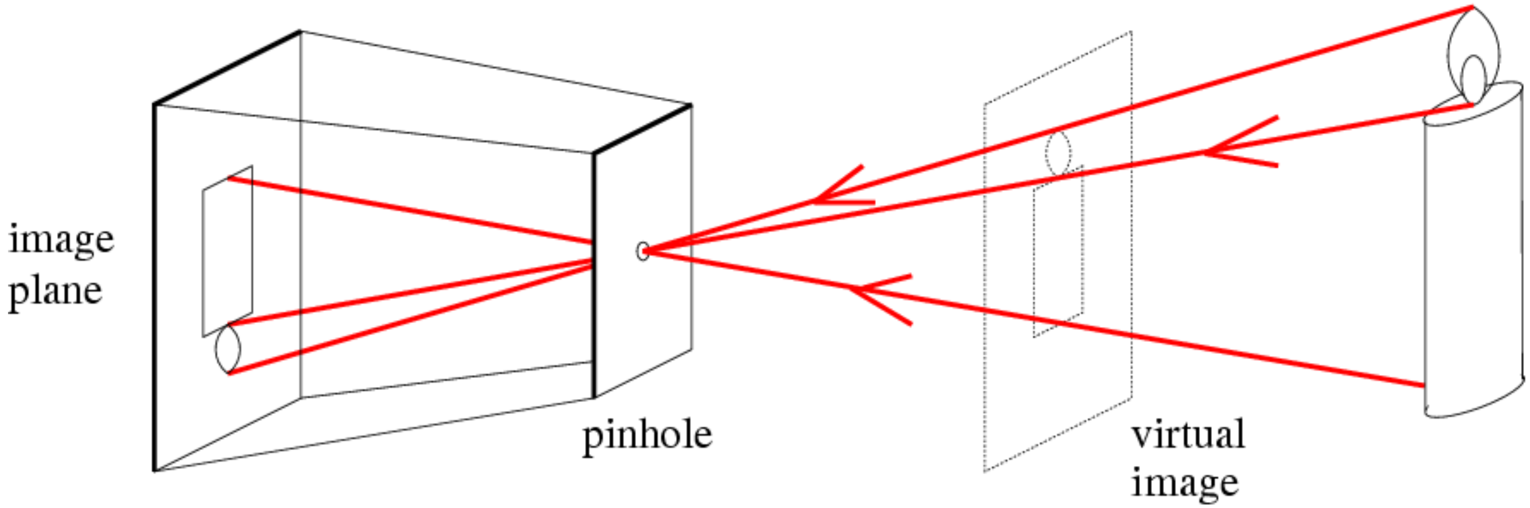
Images are two-dimensional patterns of brightness values.



Animal eye: a looonnng time ago.



Photographic camera: Niepce, 1816.



Pinhole perspective projection: Brunelleschi, XVth Century.
Camera obscura: XVIth Century.

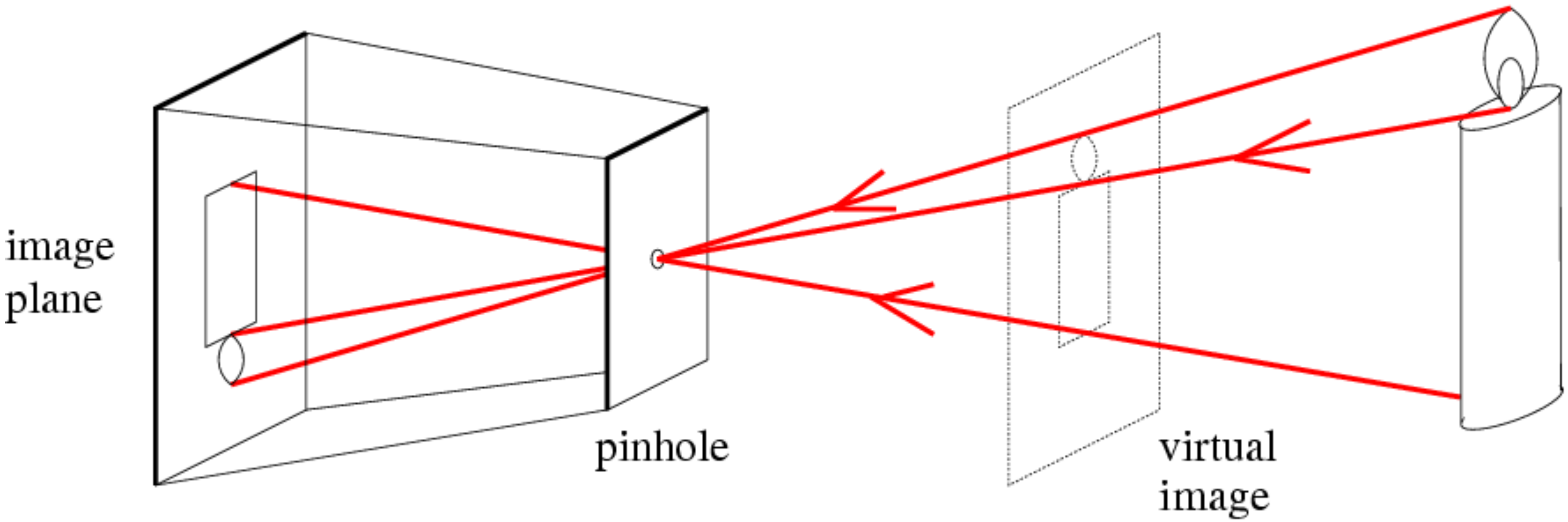
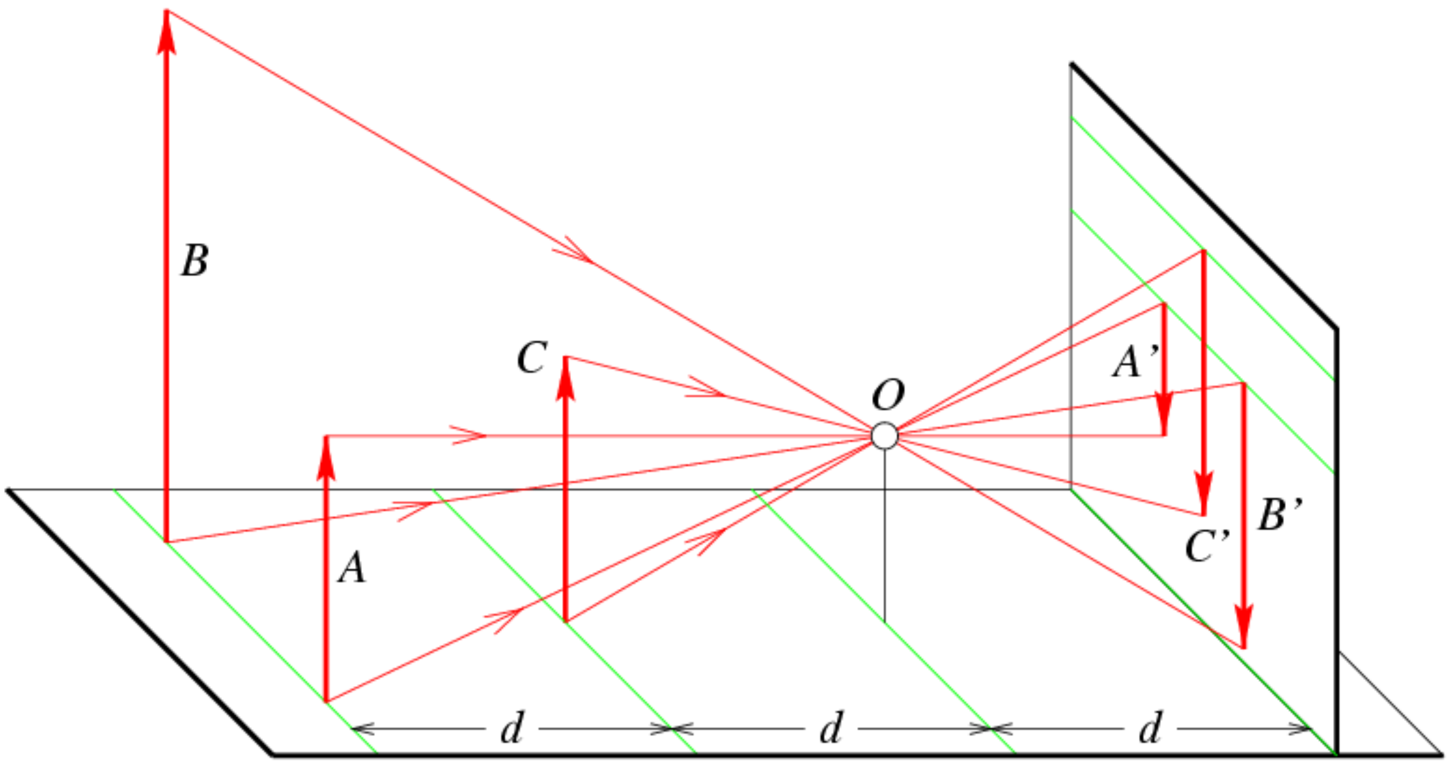
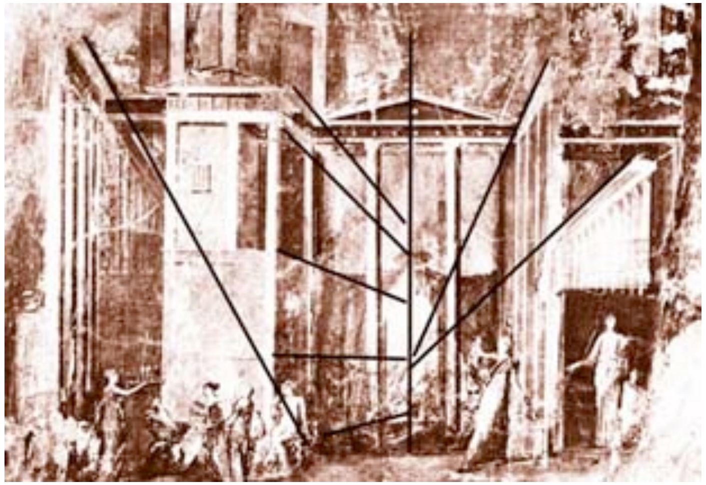


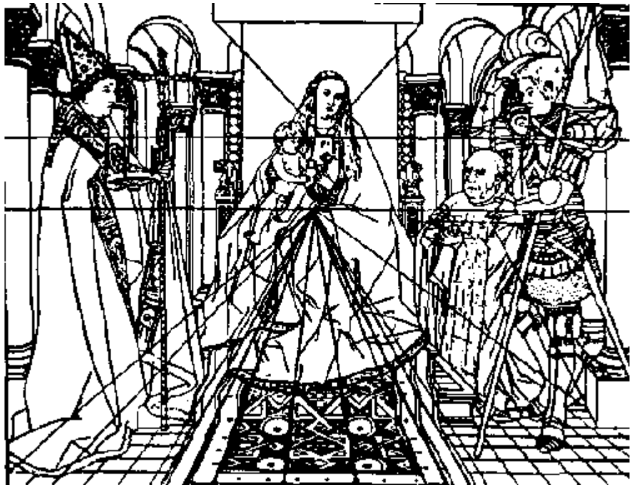
image plane
pinhole
virtual image



B
C
A
O
A'
C'
B'
d
d
d

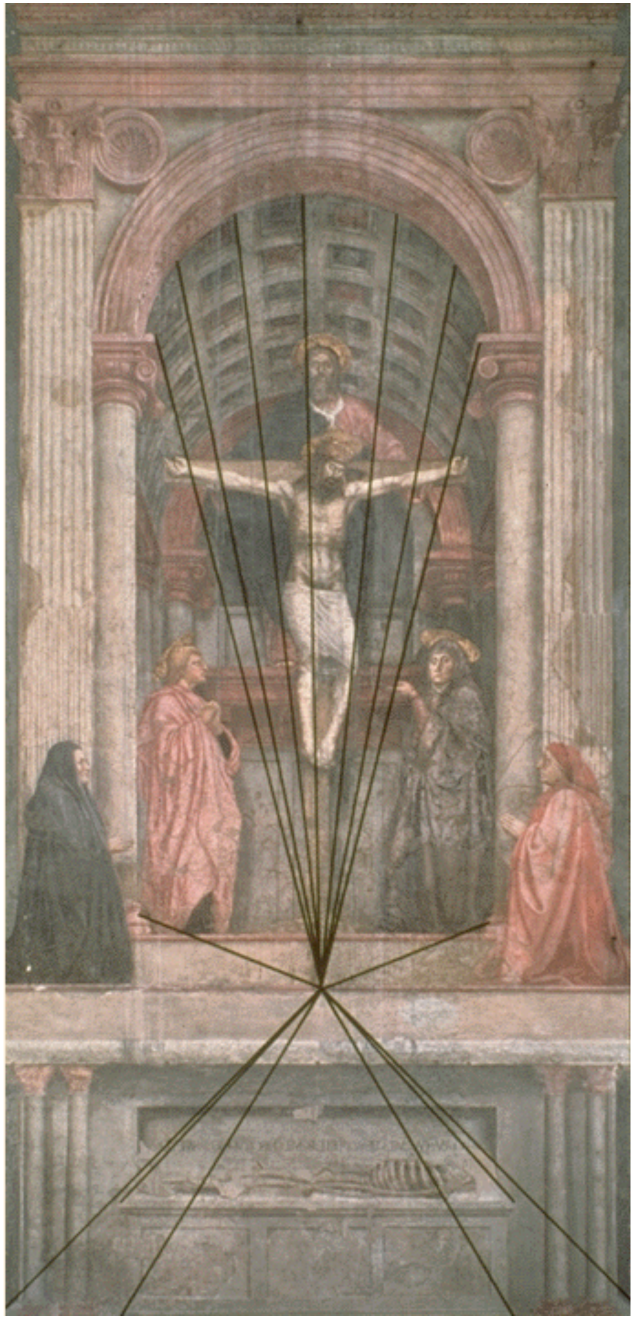


Pompei painting, 2000 years ago.



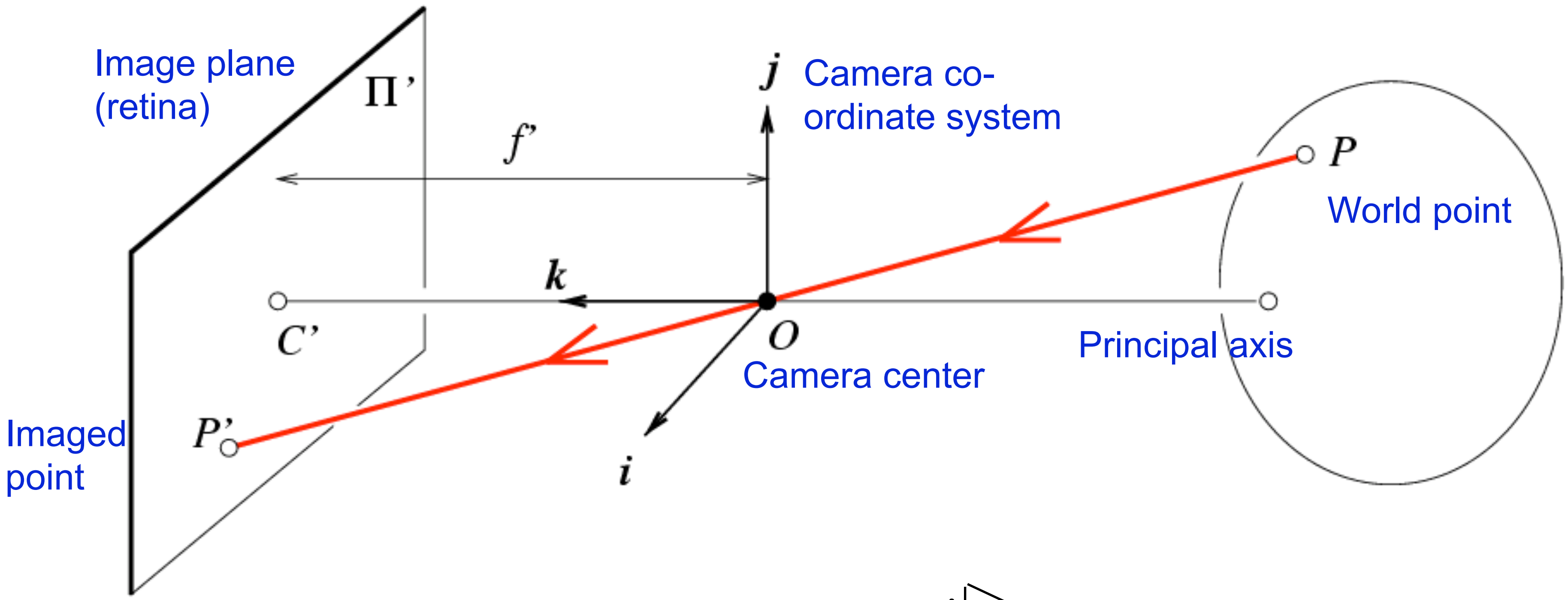
Van Eyk, XIVth Century

Brunelleschi, 1415



Massaccio’s Trinity, 1425

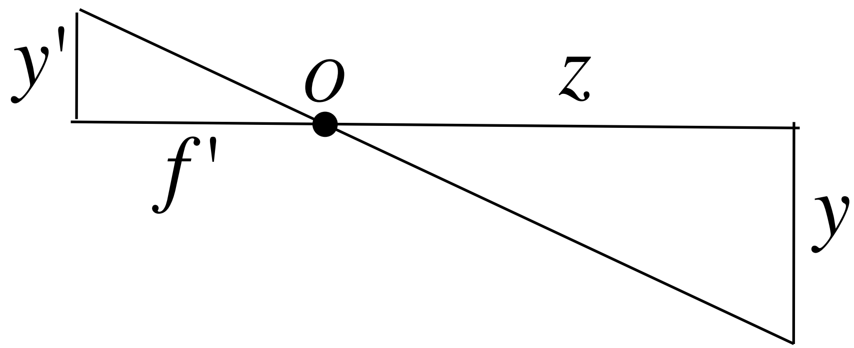
Pinhole Perspective Equation



$$P = (x, y, z)^T$$

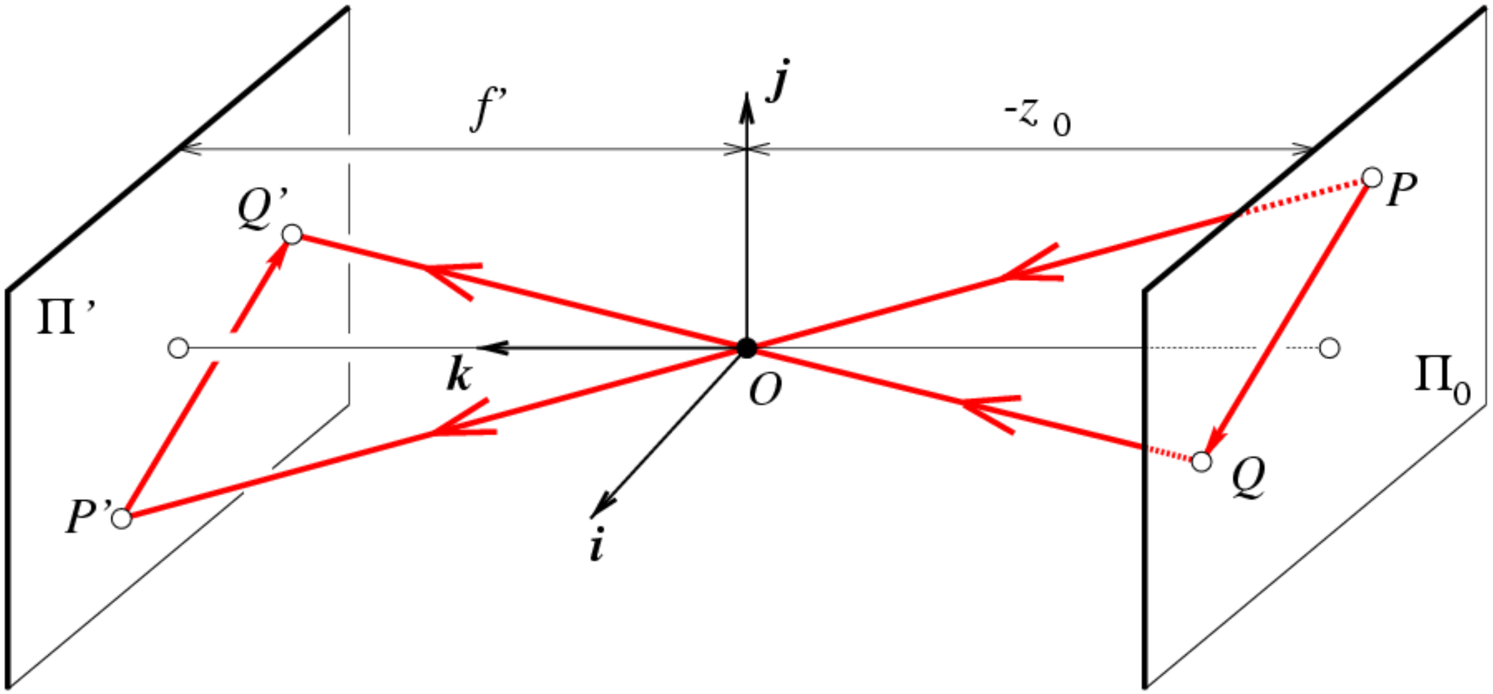
$$P' = (x', y')^T$$

$$\boxed{\begin{cases} x' = f'\dfrac{x}{z} \\ y' = f'\dfrac{y}{z} \end{cases}}$$



NOTE: *z* is always negative..

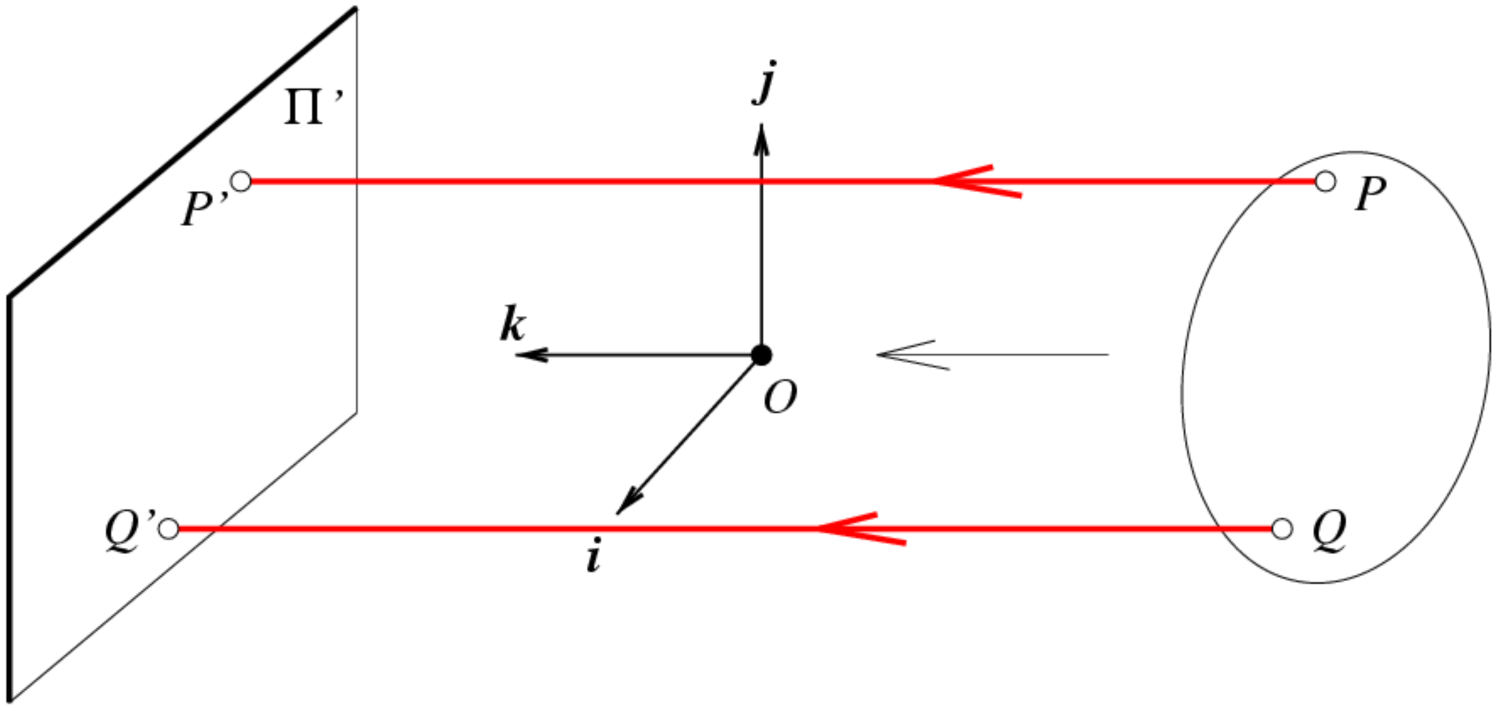
Affine projection models: Weak perspective projection



$$\boxed{\begin{cases} x' = -mx \\ y' = -my \end{cases}}$$ where $m = -\dfrac{f'}{z_0}$ is the magnification.

When the scene relief is small compared its distance from the Camera, *m* can be taken constant: weak perspective projection.

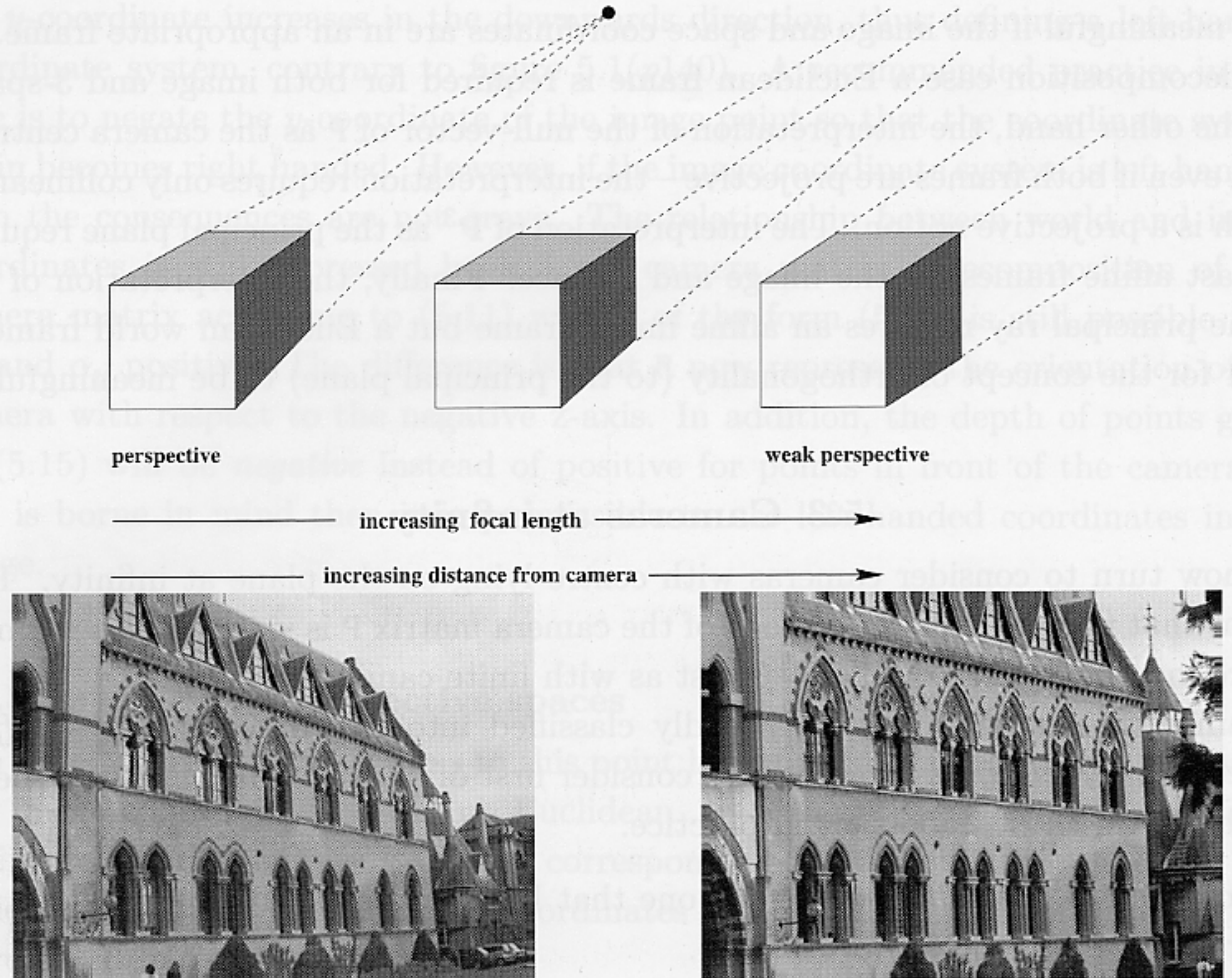
Affine projection models: Orthographic projection



$$\boxed{\begin{cases} x' = x \\ y' = y \end{cases}}$$

When the camera is at a (roughly constant) distance from the scene, take *m*=1.

Strong perspective:
Angles are not preserved
The projections of parallel lines intersect at one point

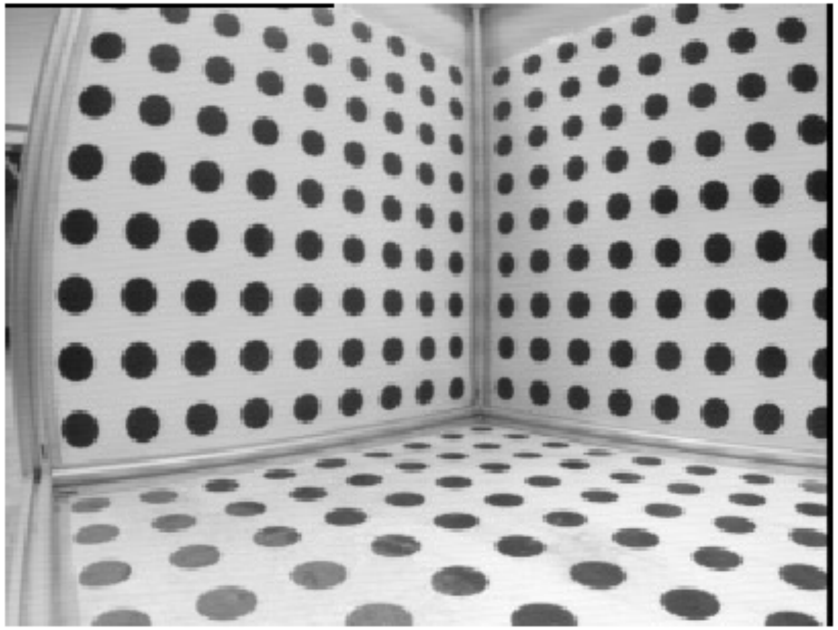


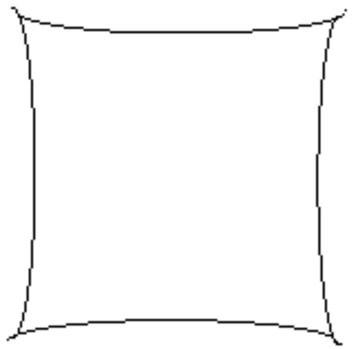
From Zisserman & Hartley

Strong perspective:
Angles are not preserved
The projections of parallel lines intersect at one point

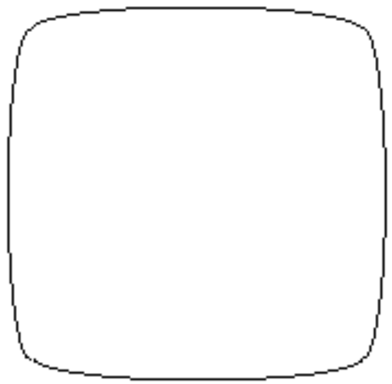
Weak perspective:
Angles are better preserved
The projections of parallel lines are (almost) parallel

Beyond pinhole camera model: Geometric Distortion

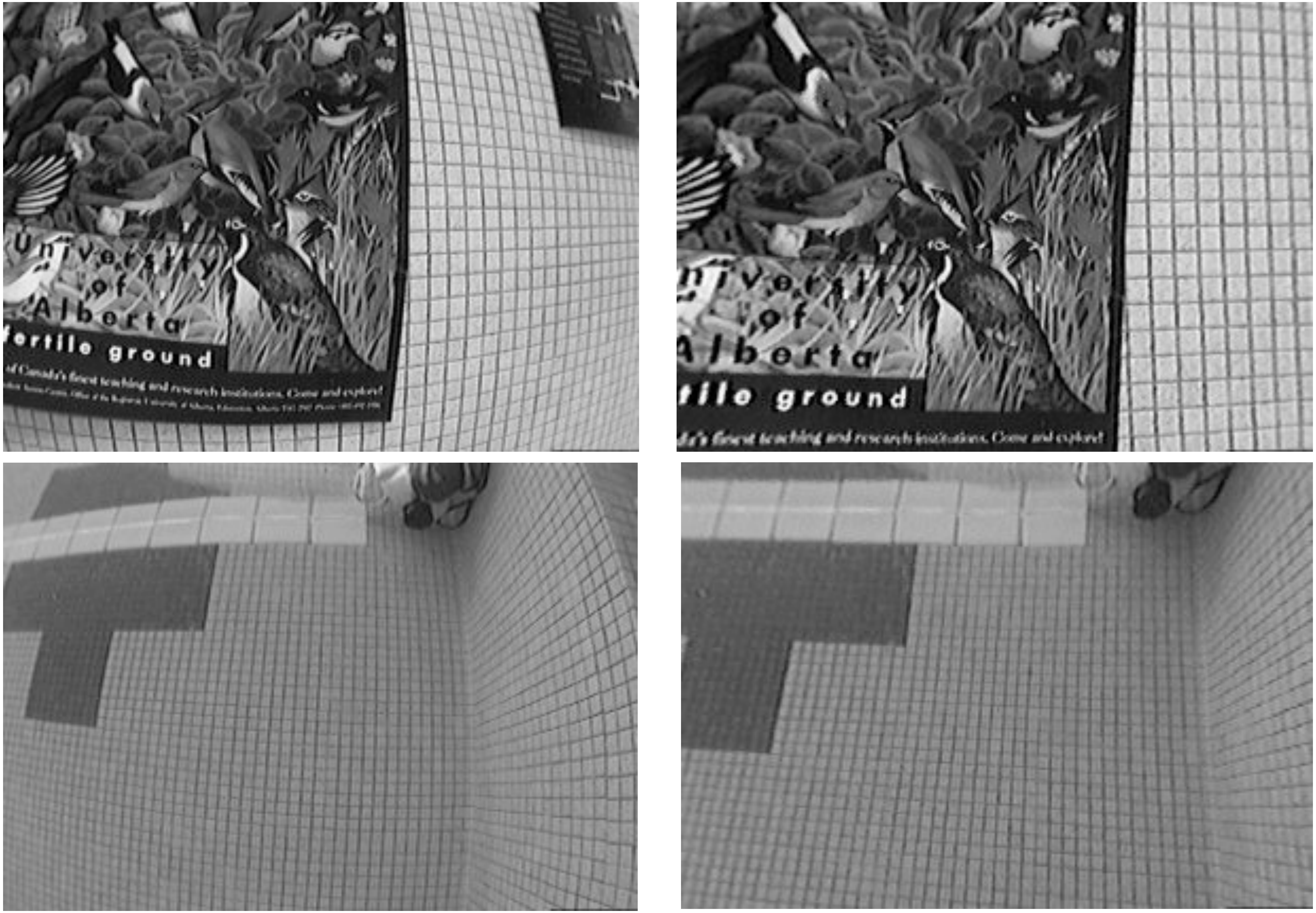




pincushion

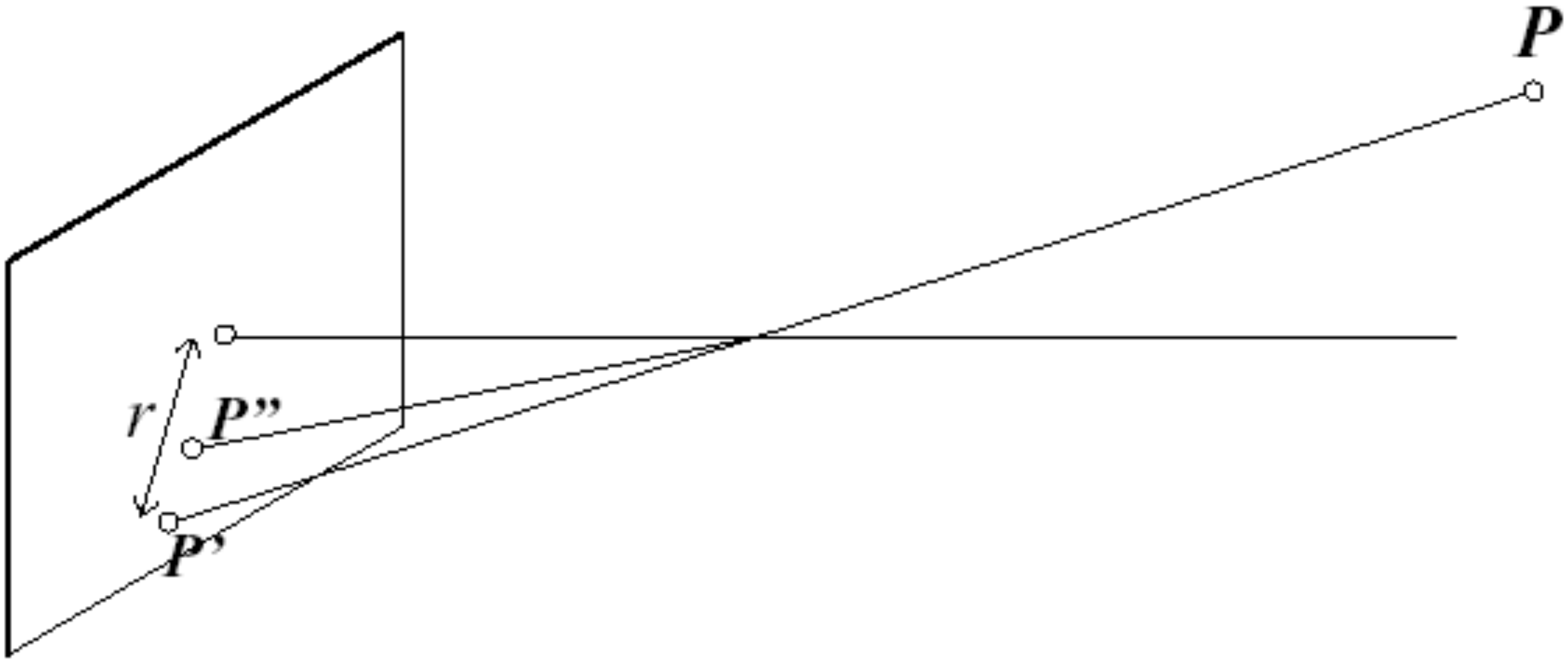


barrel



Rectification

Radial Distortion Model



Ideal:

$$x' = f\frac{x}{z}$$

$$y' = f\frac{y}{z}$$

Distorted:

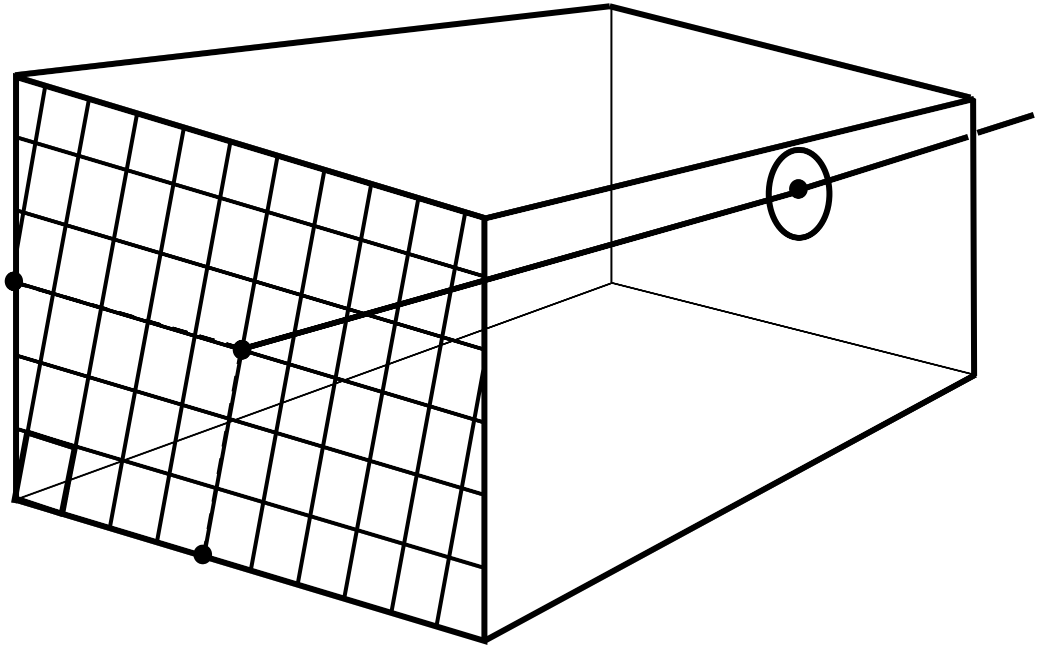
$$x'' = \frac{1}{\lambda}x'$$

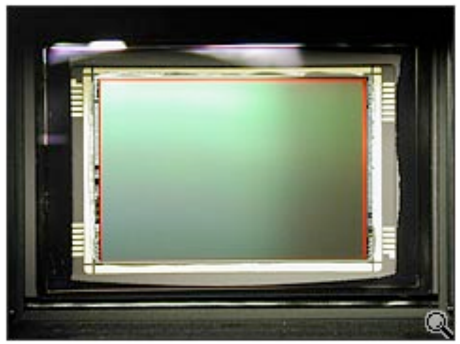
$$y'' = \frac{1}{\lambda}y'$$

$$\lambda = 1 + k_1 r^2 + k_2 r^4 + \cdots$$

Perspective Projection	$x' = f\frac{x}{z}$ $y' = f\frac{y}{z}$	x,y: World coordinates x',y': Image coordinates f: pinhole-to-retina distance
Weak-Perspective Projection (Affine)	$x' \approx -mx$ $y' \approx -my$ $m = -\frac{f}{\bar{z}}$	x,y: World coordinates x',y': Image coordinates m: magnification
Orthographic Projection (Affine)	$x' \approx x$ $y' \approx y$	x,y: World coordinates x',y': Image coordinates
Common distortion model	$x'' = \frac{1}{\lambda}x'$ $y'' = \frac{1}{\lambda}y'$ $\lambda = 1 + k_1 r^2 + k_2 r^4 + \cdots$	x',y': Ideal image coordinates x'',y'': Actual image coordinates

Cameras and their parameters





Images from M. Pollefeys

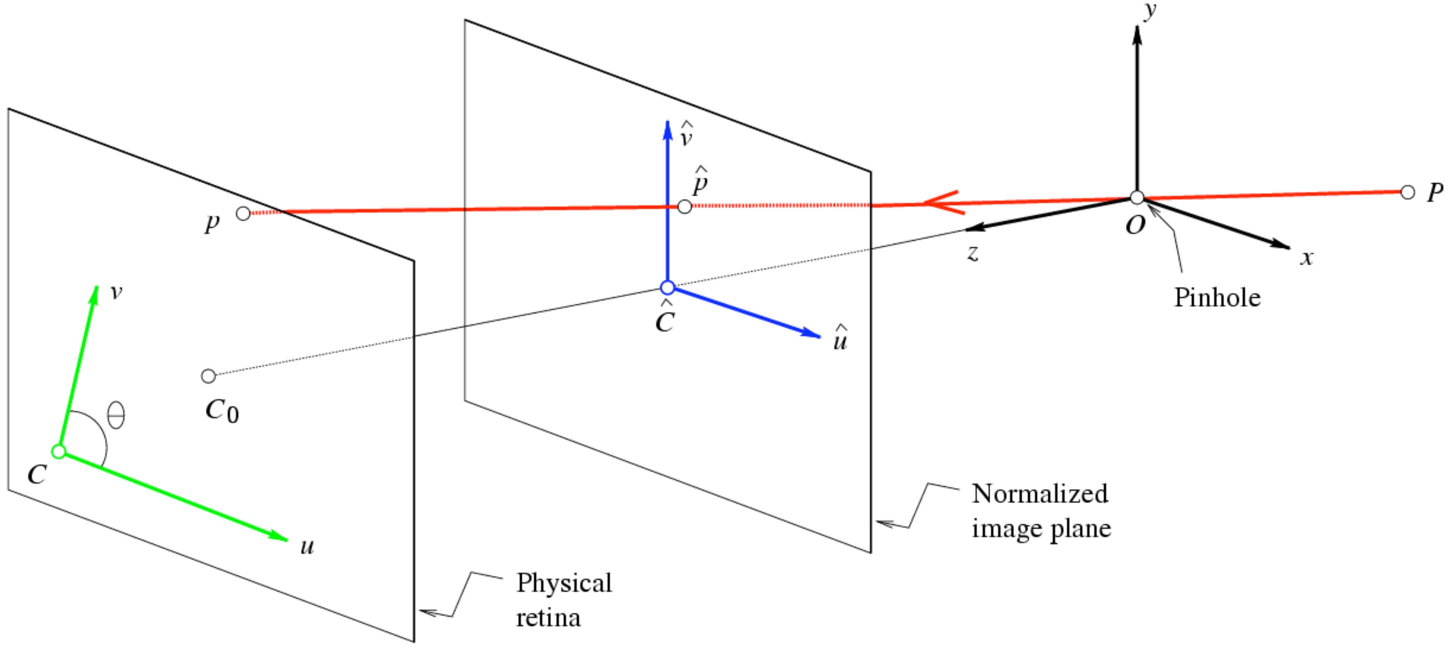
The Intrinsic Parameters of a Camera

Units:

k,l : pixel/m

f : m

α,β : pixel



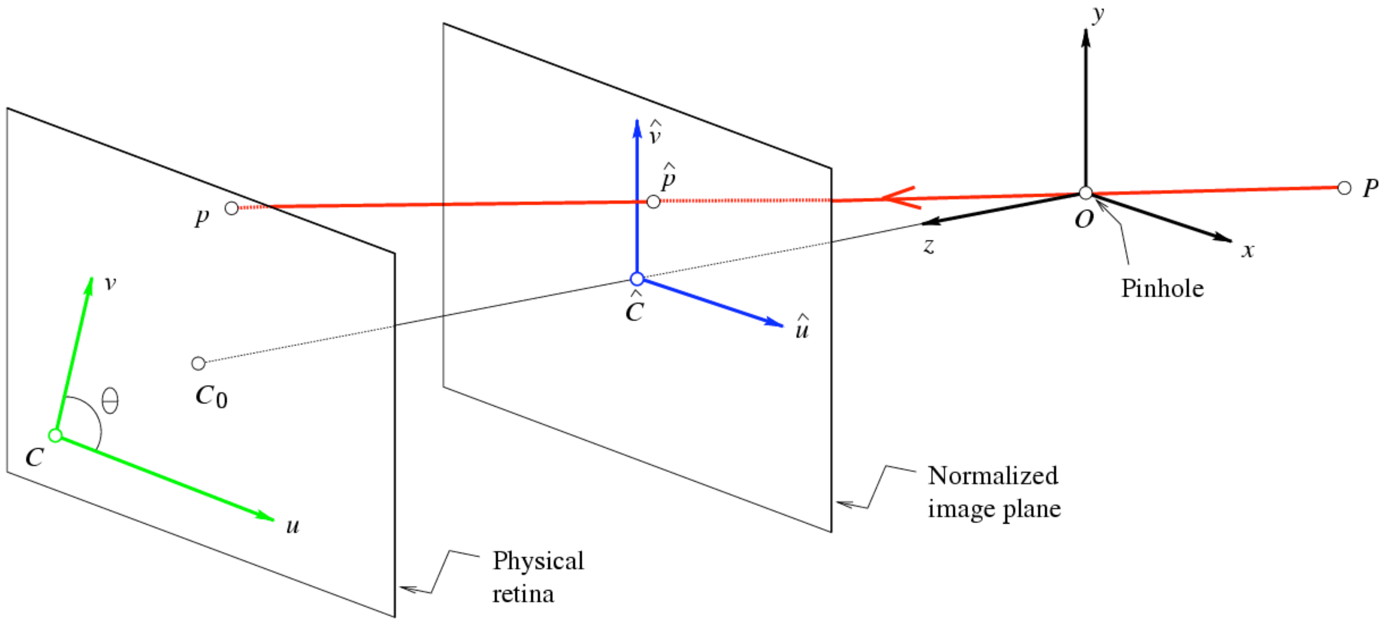
$$\begin{cases} \hat{u} = \dfrac{x}{z} \\ \hat{v} = \dfrac{y}{z} \end{cases} \Longleftrightarrow \hat{\boldsymbol{p}} = \frac{1}{z} \begin{pmatrix} \mathrm{Id} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{P} \\ 1 \end{pmatrix}$$

Normalized Image Coordinates

Physical Image Coordinates

$$\begin{cases} u = kf\dfrac{x}{z} \\ v = lf\dfrac{y}{z} \end{cases} \rightarrow \begin{cases} u = \alpha\dfrac{x}{z} + u_0 \\ v = \beta\dfrac{y}{z} + v_0 \end{cases} \rightarrow \begin{cases} u = \alpha\dfrac{x}{z} - \alpha\cot\theta\dfrac{y}{z} + u_0 \\ v = \dfrac{\beta}{\sin\theta}\dfrac{y}{z} + v_0 \end{cases}$$

The Intrinsic Parameters of a Camera



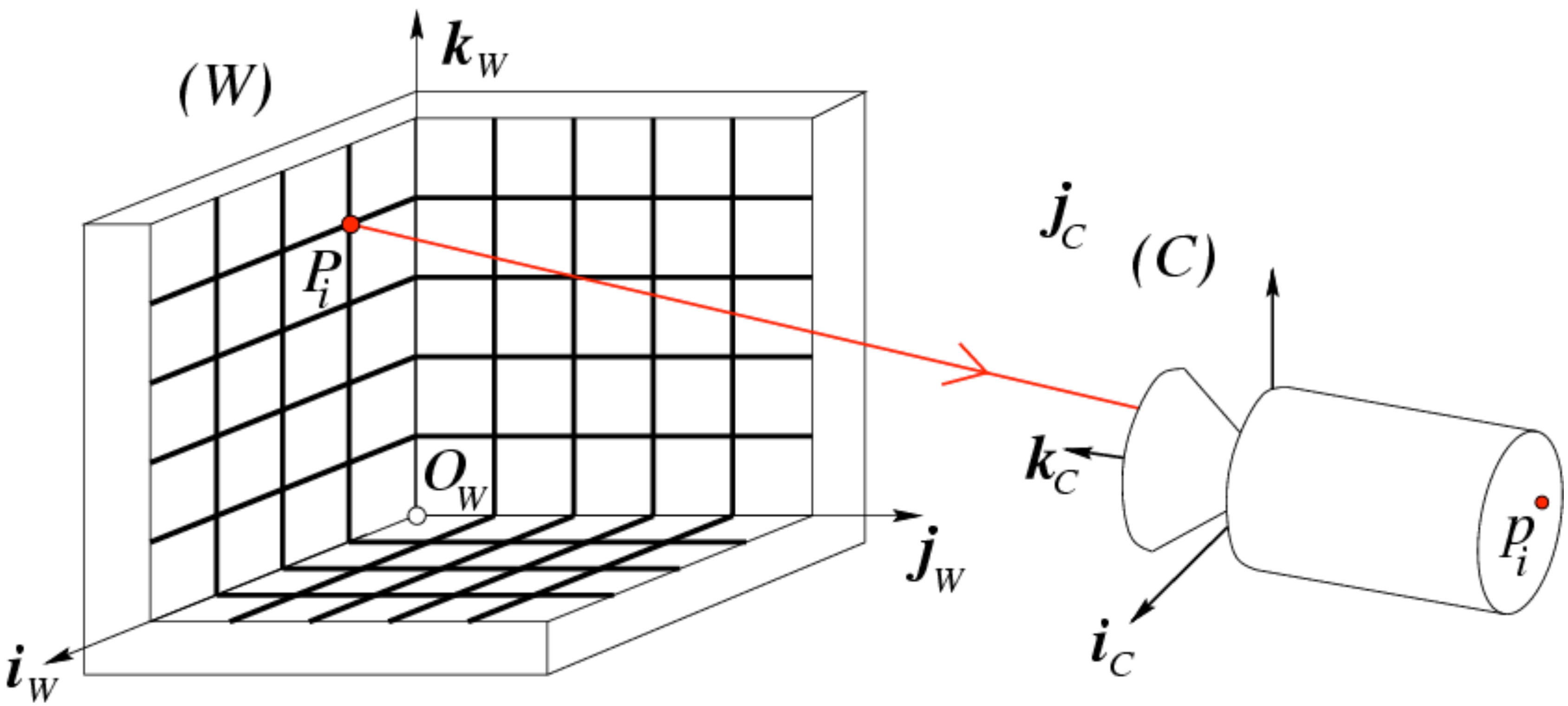
Calibration Matrix

$$\boldsymbol{p} = \mathcal{K}\hat{\boldsymbol{p}}, \quad \text{where} \quad \boldsymbol{p} = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \quad \text{and} \quad \mathcal{K} \stackrel{\text{def}}{=} \begin{pmatrix} \alpha & -\alpha\cot\theta & u_0 \\ 0 & \dfrac{\beta}{\sin\theta} & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

The Perspective Projection Equation

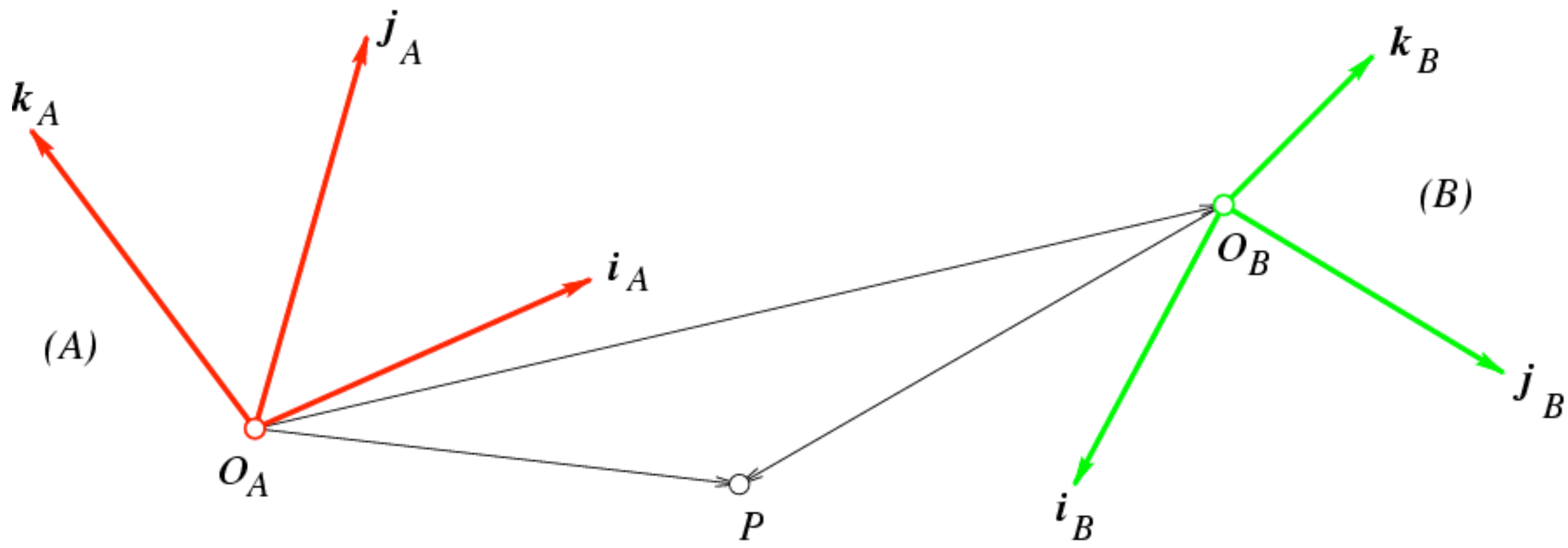
$$\boldsymbol{p} = \frac{1}{z}\mathcal{M}\boldsymbol{P}, \quad \text{where} \quad \mathcal{M} \stackrel{\text{def}}{=} \begin{pmatrix} \mathcal{K} & \boldsymbol{0} \end{pmatrix}$$

Notation



Euclidean Geometry

Recall:
Coordinate Changes and Rigid Transformations



$$\begin{bmatrix} {}^B P \\ 1 \end{bmatrix} = \begin{bmatrix} {}^B_A R & {}^B O_A \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} {}^A P \\ 1 \end{bmatrix} = \begin{bmatrix} {}^B_A R\,{}^A P + {}^B O_A \\ 1 \end{bmatrix}$$

$${}^B P = {}^B_A R\,{}^A P + {}^B O_A$$

The Extrinsic Parameters of a Camera

- When the camera frame (C) is different from the world frame (W),

$$\begin{pmatrix} {}^C P \\ 1 \end{pmatrix} = \begin{pmatrix} {}^C_W \mathcal{R} & {}^C O_W \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} {}^W P \\ 1 \end{pmatrix}.$$

- Thus,

$$\boxed{\boldsymbol{p} = \frac{1}{z} \mathcal{M} \boldsymbol{P},} \quad \text{where} \quad \begin{cases} \mathcal{M} = \mathcal{K} \begin{pmatrix} \mathcal{R} & \boldsymbol{t} \end{pmatrix}, \\ \mathcal{R} = {}^C_W \mathcal{R}, \\ \boldsymbol{t} = {}^C O_W, \\ \boldsymbol{P} = \begin{pmatrix} {}^W P \\ 1 \end{pmatrix}. \end{cases}$$

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

$${}^W P = \begin{pmatrix} {}^W x \\ {}^W y \\ {}^W z \end{pmatrix}$$

- Note: z is *not* independent of $\mathcal{M}$ and $\boldsymbol{P}$:

$$\mathcal{M} = \begin{pmatrix} \boldsymbol{m}_1^T \\ \boldsymbol{m}_2^T \\ \boldsymbol{m}_3^T \end{pmatrix} \Longrightarrow z = \boldsymbol{m}_3 \cdot \boldsymbol{P}, \quad \text{or} \quad \begin{cases} u = \dfrac{\boldsymbol{m}_1 \cdot \boldsymbol{P}}{\boldsymbol{m}_3 \cdot \boldsymbol{P}}, \\ v = \dfrac{\boldsymbol{m}_2 \cdot \boldsymbol{P}}{\boldsymbol{m}_3 \cdot \boldsymbol{P}}. \end{cases}$$

Explicit Form of the Projection Matrix

$$\mathcal{M} = \begin{pmatrix} \alpha \boldsymbol{r}_1^T - \alpha \cot\theta \boldsymbol{r}_2^T + u_0 \boldsymbol{r}_3^T & \alpha t_x - \alpha \cot\theta t_y + u_0 t_z \\ \dfrac{\beta}{\sin\theta}\boldsymbol{r}_2^T + v_0 \boldsymbol{r}_3^T & \dfrac{\beta}{\sin\theta} t_y + v_0 t_z \\ \boldsymbol{r}_3^T & t_z \end{pmatrix}$$

Note: If $\mathcal{M} = (\mathcal{A} \quad \boldsymbol{b})$ then $|\boldsymbol{a}_3| = 1$.

Replacing $\mathcal{M}$ by $\lambda\mathcal{M}$ in

$$\begin{cases} u = \dfrac{\boldsymbol{m}_1 \cdot \boldsymbol{P}}{\boldsymbol{m}_3 \cdot \boldsymbol{P}} \\ v = \dfrac{\boldsymbol{m}_2 \cdot \boldsymbol{P}}{\boldsymbol{m}_3 \cdot \boldsymbol{P}} \end{cases}$$

does not change u and v.

M is only defined up to scale in this setting!!

Weak perspective (affine) camera

$$z_r = m_3^T P = \text{const.}$$

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{z_r}\begin{bmatrix} m_1^T \\ m_2^T \\ m_3^T \end{bmatrix} P = \begin{pmatrix} m_1^T P / m_3^T P \\ m_2^T P / m_3^T P \\ m_3^T P / m_3^T P \end{pmatrix}$$

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{z_r}\begin{bmatrix} m_1^T \\ m_2^T \end{bmatrix} P$$

$$\begin{pmatrix} u \\ v \end{pmatrix} = [\underbrace{\begin{matrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{matrix}}_{A_{2\times 3}} \quad \underbrace{\begin{matrix} b_1 \\ b_2 \end{matrix}}_{b_{2\times 1}}] \begin{pmatrix} {}^W x \\ {}^W y \\ {}^W z \\ 1 \end{pmatrix} = A\,{}^W P + b$$

Geometric Interpretation

Projection equation:

$$u = \frac{m_1^T P}{m_3^T P} = \frac{a_1^T \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + b_1}{a_3^T \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + b_3}$$

Observations:

$$a_1^T \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + b_1 = 0$$

is the equation of a plane of normal direction a_1

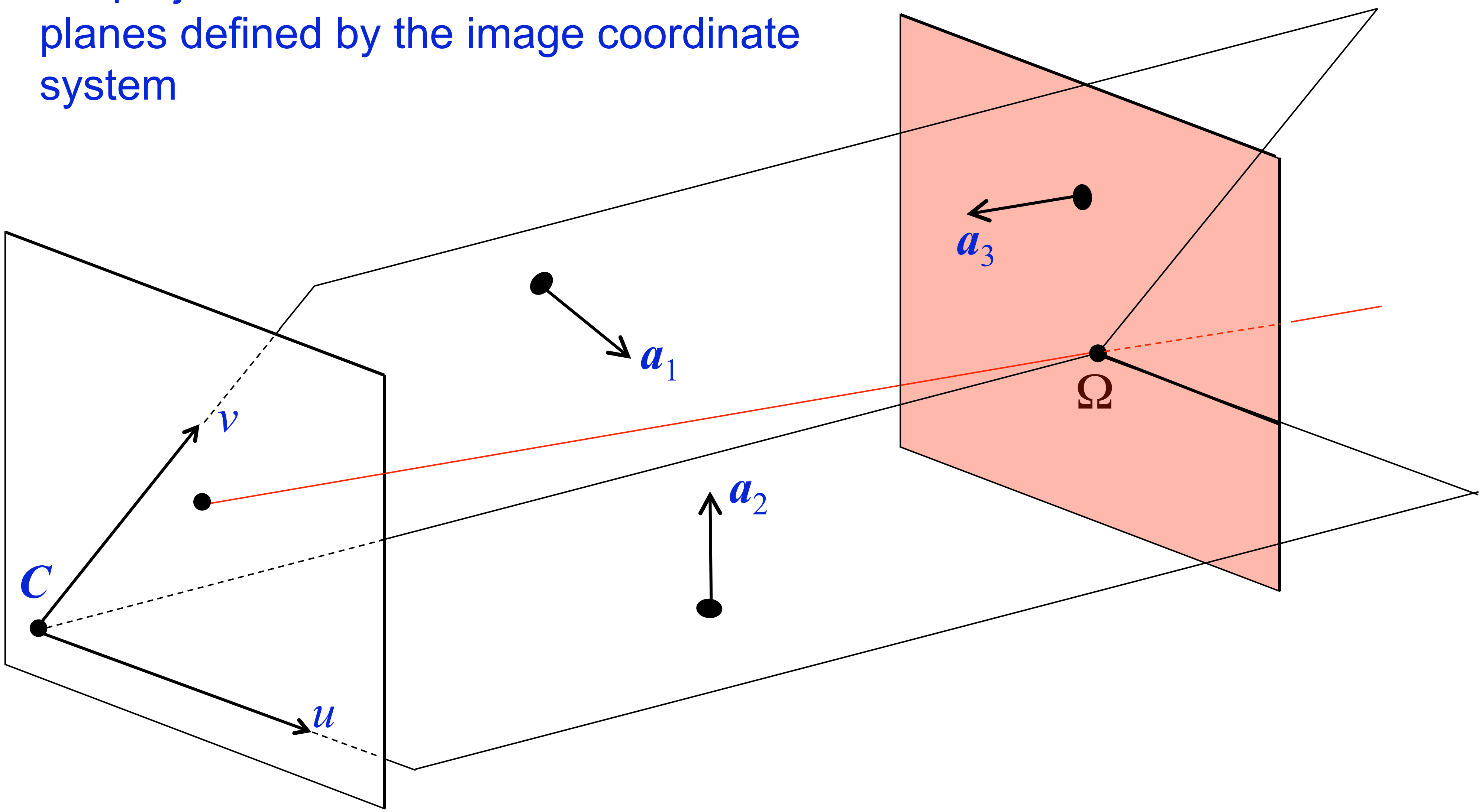
- From the projection equation, it is also the plane defined by: $u = 0$

- Similarly:
  - (a_2, b_2) describes the plane defined by: $v = 0$
  - (a_3, b_3) describes the plane defined by:

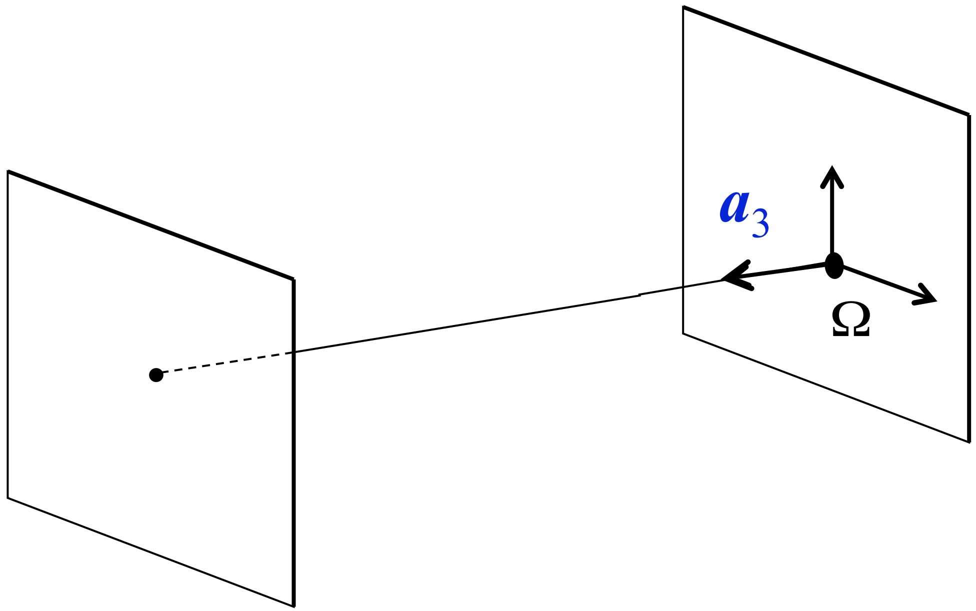
    $u = \infty \quad v = \infty$

  → That is the plane parallel to image plane passing through the pinhole ($z = 0$) – principal plane

Geometric Interpretation: The rows of the projection matrix describe the three planes defined by the image coordinate system

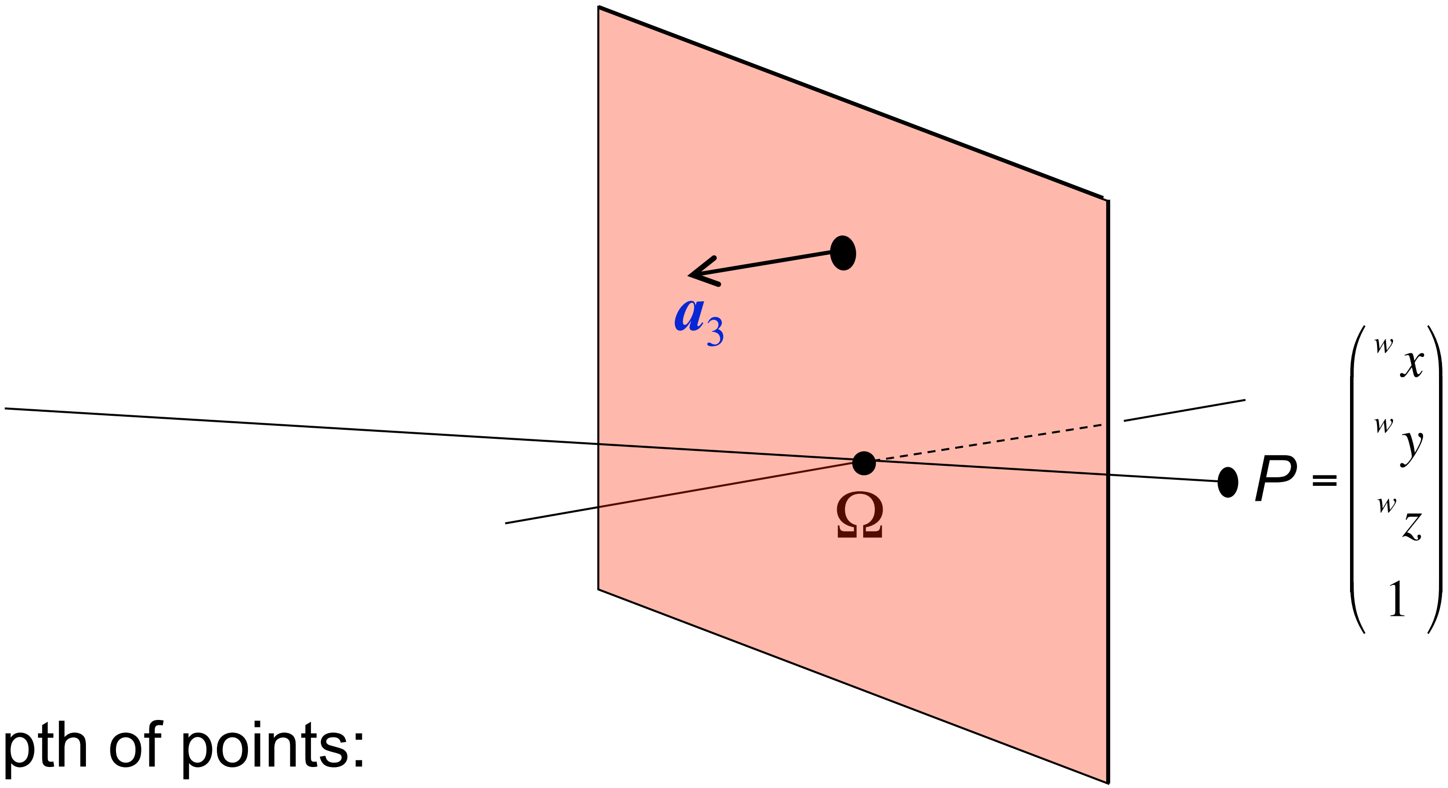


Other useful geometric properties



Principal axis of the camera:
The ray passing through the camera centre Ω with direction vector $\mathbf{a_3}$

Other useful geometric properties



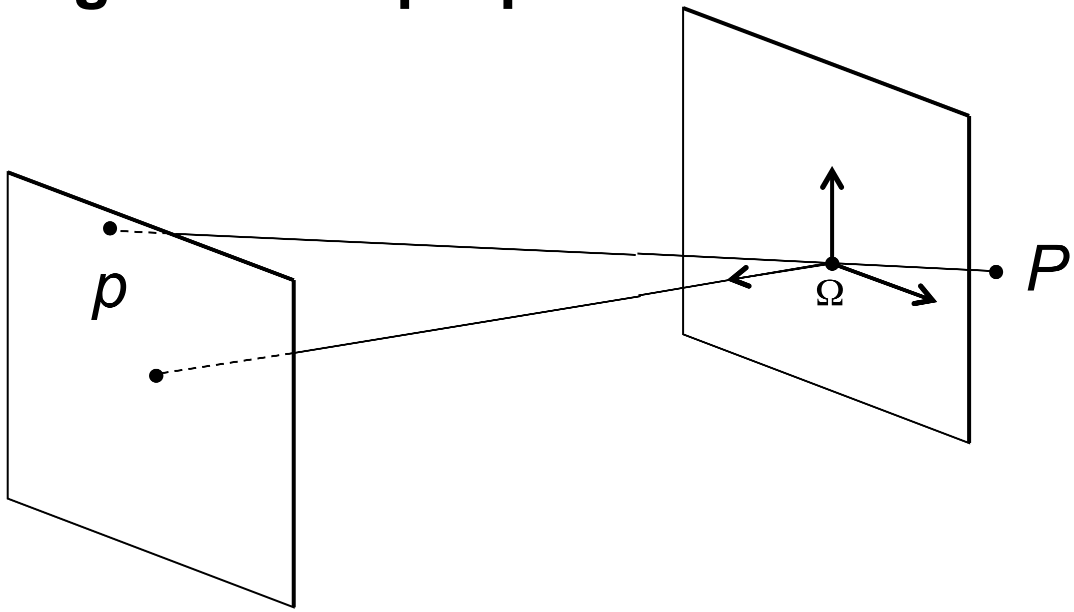
Depth of points:
How far a point lies from the principal plane of a camera?

$$d(P,\Pi_3) = \frac{a_3^T \begin{bmatrix} {}^w x \\ {}^w y \\ {}^w z \end{bmatrix} + b_3}{\|a_3\|} = \frac{m_3{}^T P}{\|a_3\|} = m_3{}^T P$$

If $||a_3||=1$

But for general camera matrices:
- need to be careful about the sign.
- need to normalize matrix to have $||a_3||=1$

Other useful geometric properties



Q: Can we compute the position of the camera center Ω?

A: $\boxed{\Omega = -A^{-1}b}$

Hint: Start from the projection equation. Show that the right null-space of camera matrix M is the camera center.

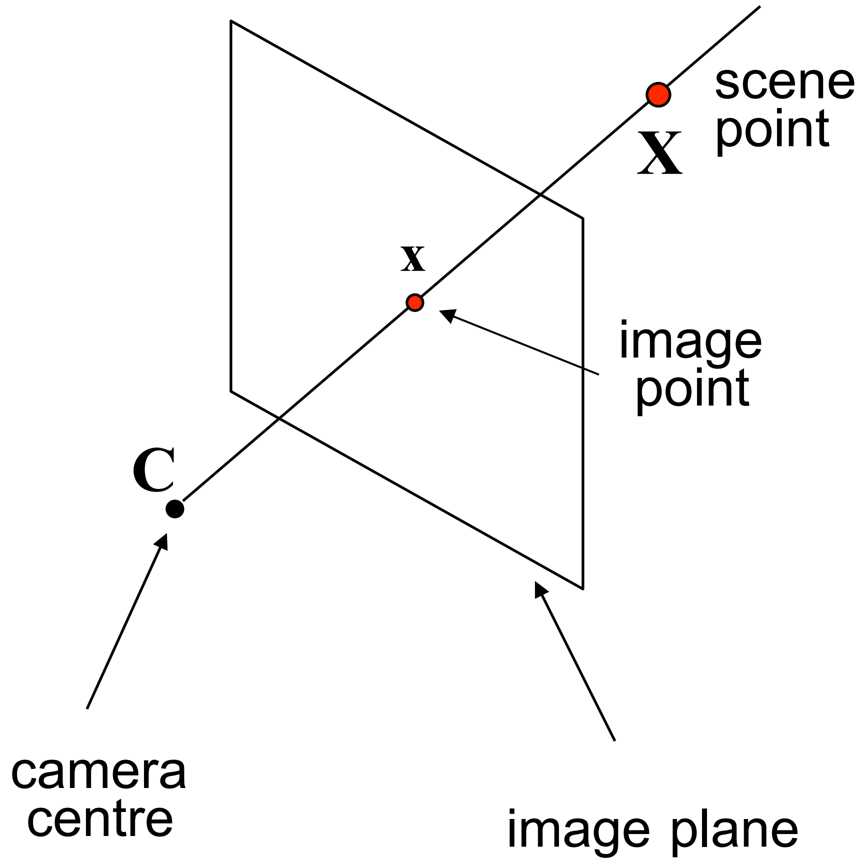
Q: Given an image point *p*, what is the direction of the corresponding ray in space?

A: $\boxed{w = A^{-1}p}$

Hint: Start from a projection equation and write all points along direction w, that project to point p.

Re-cap: imaging and camera geometry (with a slight change of notation)

- perspective projection
- camera centre, image point and scene point are collinear
- an image point back projects to a ray in 3-space
- depth of the scene point is unknown



Slide credit: A. Zisserman

Notation – A Projective Camera

The camera model for perspective projection is a linear map between homogeneous point coordinates

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \left[\quad \mathtt{P} \;\; (3 \times 4) \quad \right] \begin{pmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \\ 1 \end{pmatrix}$$

Image Point   Scene Point

$$\lambda \mathbf{x} = \mathtt{P} \quad \mathbf{X}$$

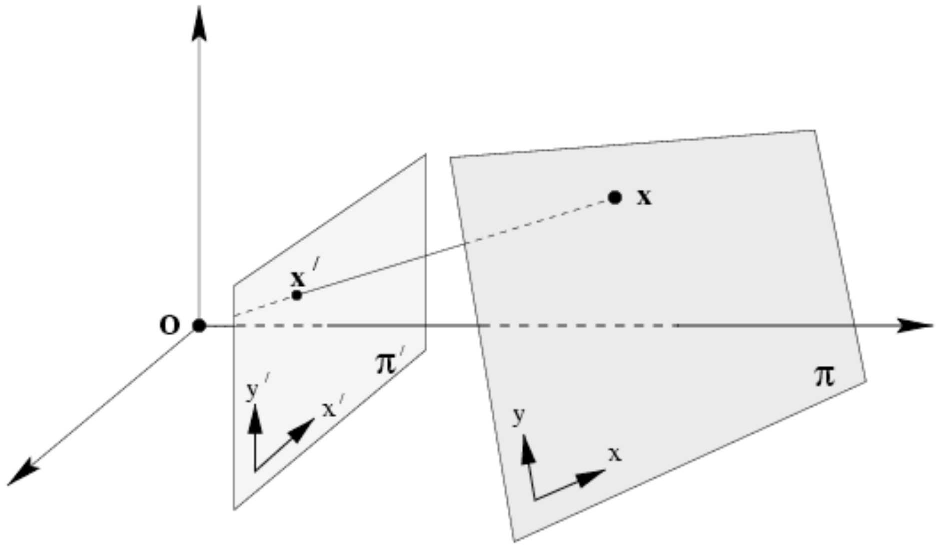
e.g. if $\mathtt{P} = [\mathtt{I}|0]$ then

$$x = \frac{\mathrm{X}}{\mathrm{Z}} \qquad y = \frac{\mathrm{Y}}{\mathrm{Z}}$$

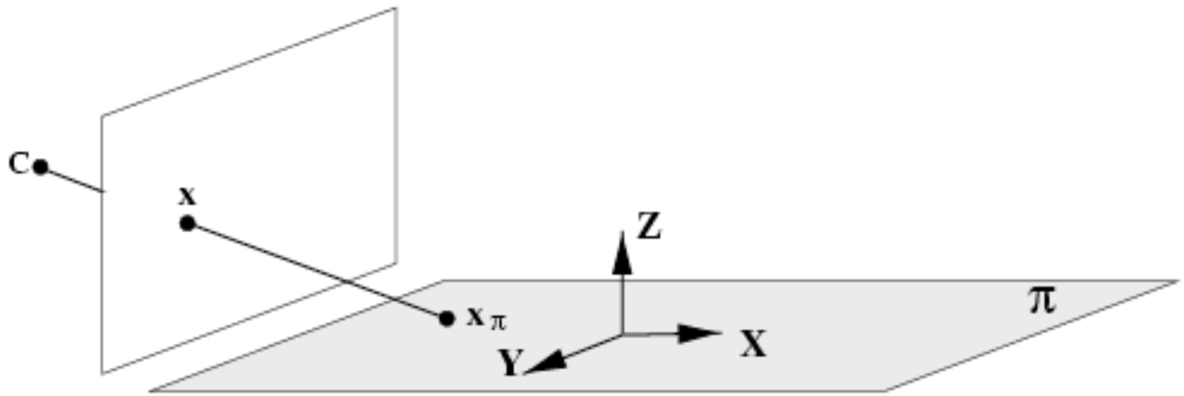
- $\mathtt{P}$ has 11 degrees of freedom (essential parameters).

Slide credit: A. Zisserman

How a “scene plane” projects into an image?



Plane projective transformations



Choose the world coordinate system such that the plane of the points has zero z coordinate. Then the 3×4 matrix $\mathtt{P}$ reduces to

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{pmatrix} \mathsf{X} \\ \mathsf{Y} \\ 0 \\ 1 \end{pmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{14} \\ p_{21} & p_{22} & p_{24} \\ p_{31} & p_{32} & p_{34} \end{bmatrix} \begin{pmatrix} \mathsf{X} \\ \mathsf{Y} \\ 1 \end{pmatrix}$$

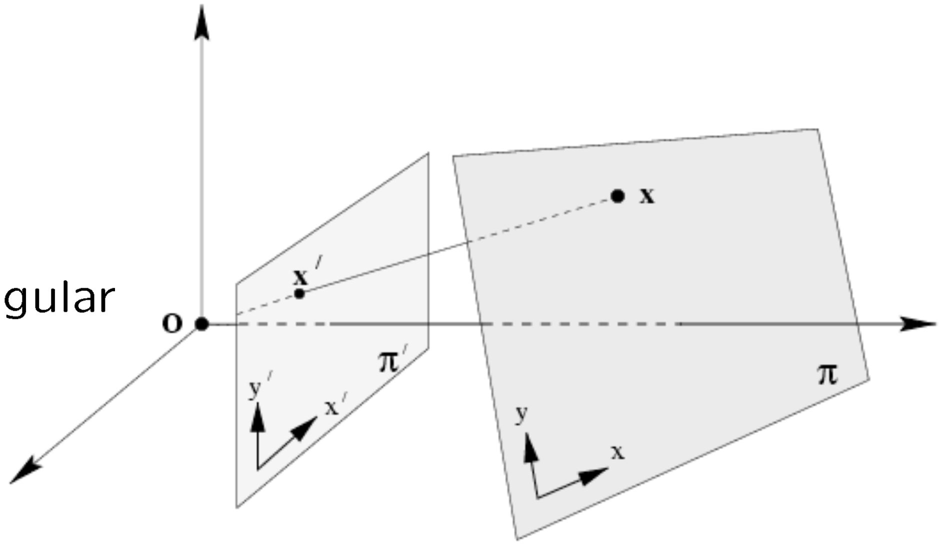
which is a 3×3 matrix representing a general plane to plane projective transformation.

Slide credit: A. Zisserman

Projective transformations continued

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

or $\mathbf{x}' = \mathtt{H}\mathbf{x}$, where $\mathtt{H}$ is a 3×3 non-singular homogeneous matrix.



- This is the most general transformation between the world and image plane under imaging by a perspective camera.
- It is often only the 3 x 3 form of the matrix that is important in establishing properties of this transformation.
- A projective transformation is also called a ``homography'' and a ``collineation''.
- H has 8 degrees of freedom.

Slide credit: A. Zisserman

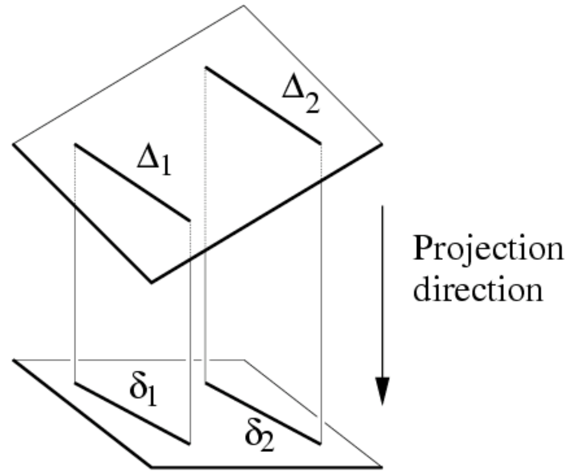
Planes under affine projection

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \end{bmatrix} \begin{pmatrix} x \\ y \\ 0 \\ 1 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = A_{2\times 2}P + b_{2\times 1}$$

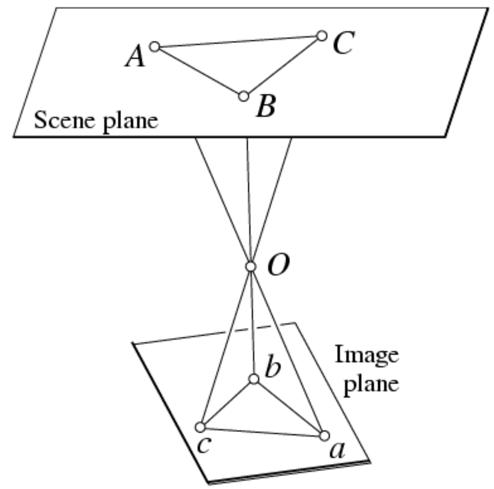
Points on a world plane map with a 2D affine geometric transformation (6 parameters)

Summary

• Affine projections induce affine transformations from planes onto their images.

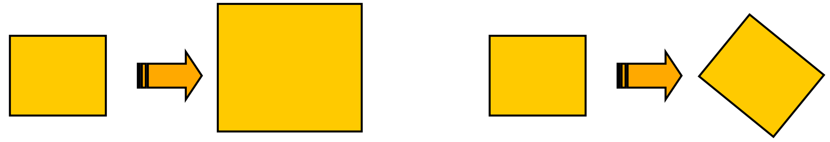


• Perspective projections induce projective transformations from planes onto their images.

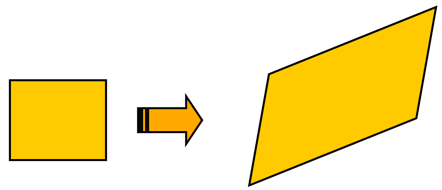


2D transformation models

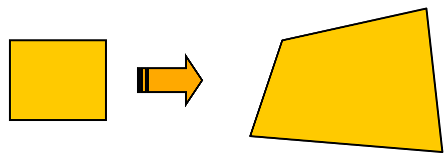
Similarity (translation, scale, rotation)



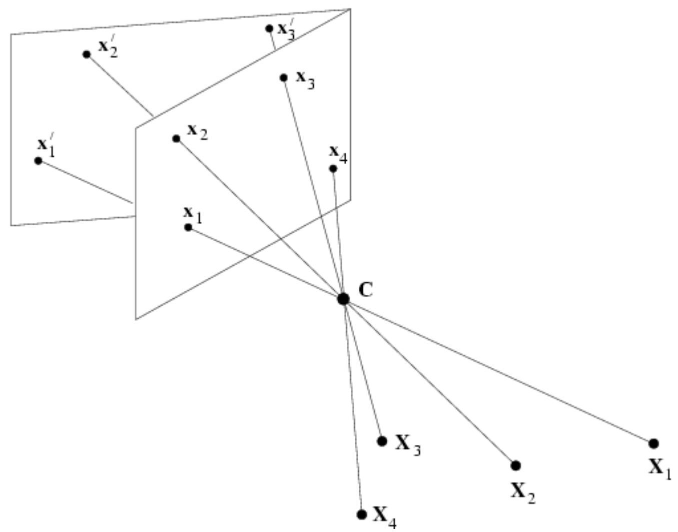
Affine

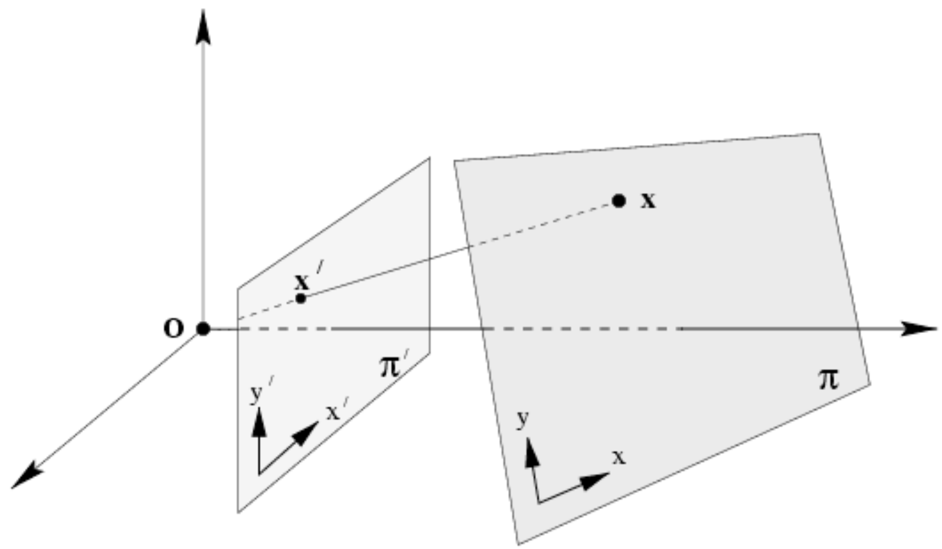


Projective (homography)

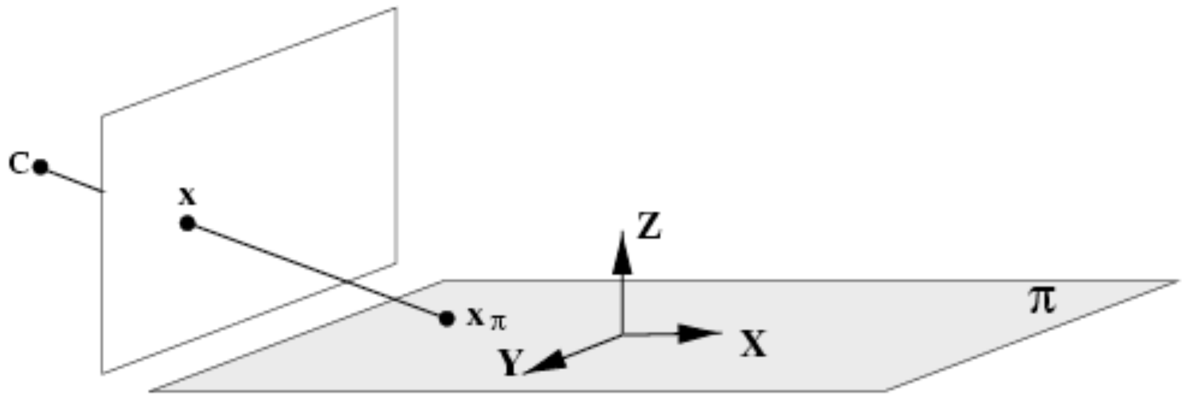


When is homography a valid transformation model?





Case I: Plane projective transformations



Choose the world coordinate system such that the plane of the points has zero z coordinate. Then the 3×4 matrix $\mathtt{P}$ reduces to

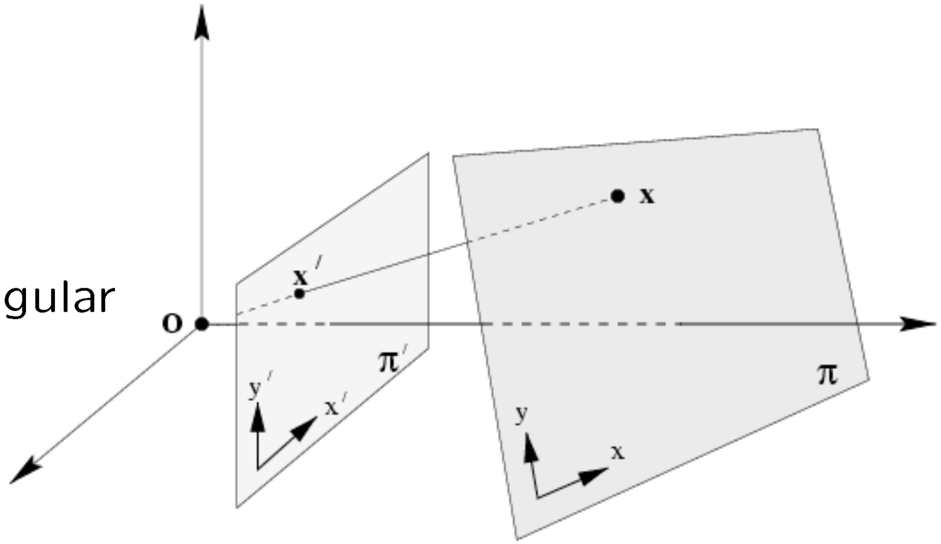
$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{pmatrix} \mathsf{X} \\ \mathsf{Y} \\ 0 \\ 1 \end{pmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{14} \\ p_{21} & p_{22} & p_{24} \\ p_{31} & p_{32} & p_{34} \end{bmatrix} \begin{pmatrix} \mathsf{X} \\ \mathsf{Y} \\ 1 \end{pmatrix}$$

which is a 3×3 matrix representing a general plane to plane projective transformation.

Case I: Projective transformations continued

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

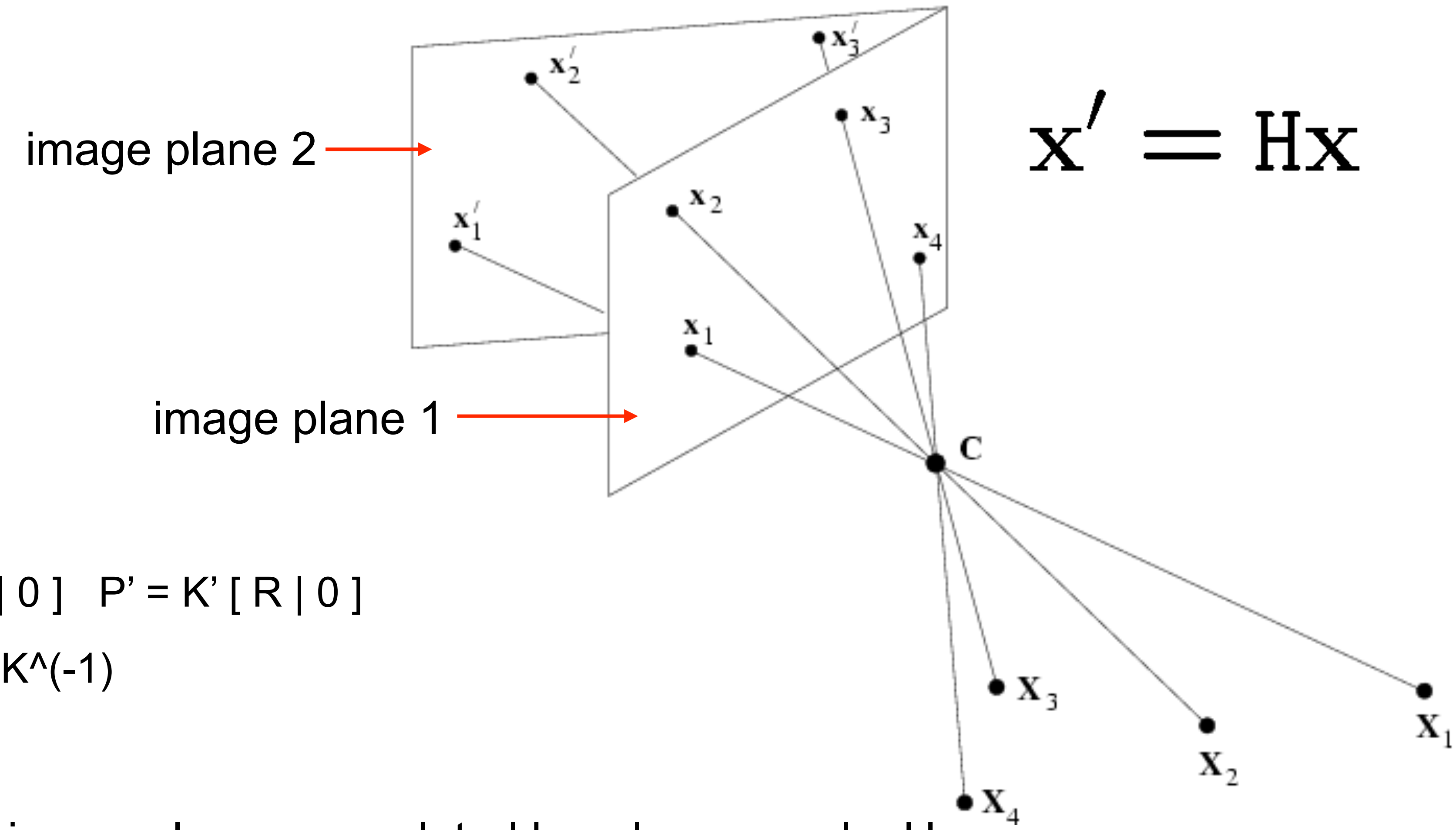
or $\mathbf{x}' = \mathtt{H}\mathbf{x}$, where $\mathtt{H}$ is a 3×3 non-singular homogeneous matrix.



- This is the most general transformation between the world and image plane under imaging by a perspective camera.
- It is often only the 3 x 3 form of the matrix that is important in establishing properties of this transformation.
- A projective transformation is also called a ``homography'' and a ``collineation''.
- H has 8 degrees of freedom.

Slide credit: A. Zisserman

Case II: Cameras rotating about their centre



$$\mathbf{x}' = \mathtt{H}\mathbf{x}$$

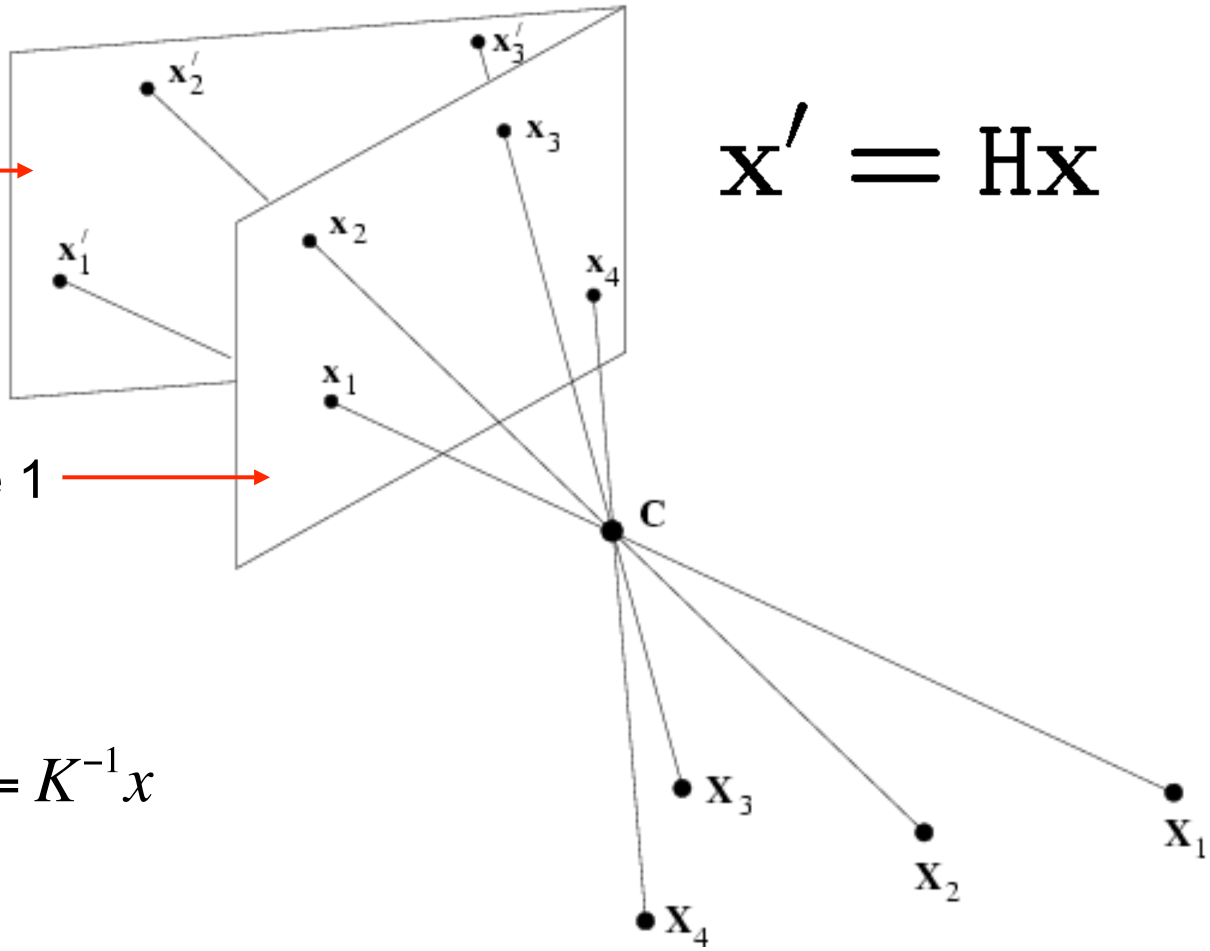
P = K [I | 0]   P' = K' [R | 0]

H = K' R K^(-1)

- The two image planes are related by a homography H
- H depends only on the relation between the image planes and camera centre, C, not on the 3D structure

Slide credit: A. Zisserman

Case II: Cameras rotating about their centre



$$\mathrm{x}' = \mathrm{H}\mathrm{x}$$

$$P = K[I \mid 0] \quad P' = K'[R \mid 0]$$

$$x = K[I \mid 0]\begin{pmatrix} X \\ 1 \end{pmatrix} = KX \Rightarrow X = K^{-1}x$$

$$x' = K'[R \mid 0]\begin{pmatrix} X \\ 1 \end{pmatrix} = K'RX$$

$$x' = \underbrace{K'RK^{-1}}_{H}x$$

Slide credit: A. Zisserman