Introduction to Computer Vision

Lecture 9- 3D Vision I

Prof. He Wang

Logistics

- Assignment 2: released on 4/11 (this Friday evening), due on 4/26 11:59PM (Saturday)

- Some functions are required to be implemented without for loop.

- If 1 day (0 - 24 hours) past the deadline, 15% off
- If 2 day (24 - 48 hours) past the deadline, 30% off
- Zero credit if more than 2 days.

Logistics: Midterm

- Midterm (30% of the total score)
  - 4/30 Wednesday, 3:10 – 5:00, in class
  - **Location: to be determined**
  - First two sections, in total 110 min
  - The third class will be used for lecture

- One-page double-sided A4 cheat sheet (handwrite or print both OK)

- Scope: from Lecture 1 - Lecture 10

Logistics: Midterm

- Questions:
  - Multi-choice questions
  - True or False
  - Short answer questions
    - simply state why and how
    - some may require mathematical derivations
  - Simple calculation questions

From 2D to 3D

Some slides are borrowed and modified from Stanford CS 231A

2D Image Representations

$$H \times W \times 3$$

Beyond Single Frame and Single View

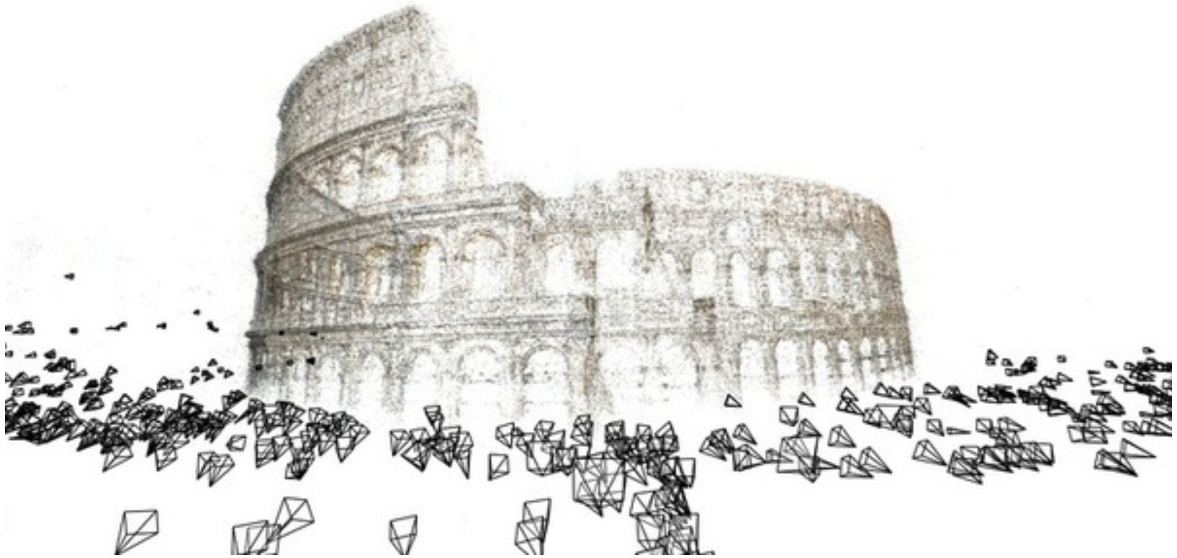
Stereo images

Multiview images

Panoramic images

We Live in a 3D World.

From partial observations to aggregate complete 3D scenes.



"Building Rome in a day." Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz and Richard Szeliski
International Conference on Computer Vision, 2009, Kyoto, Japan.

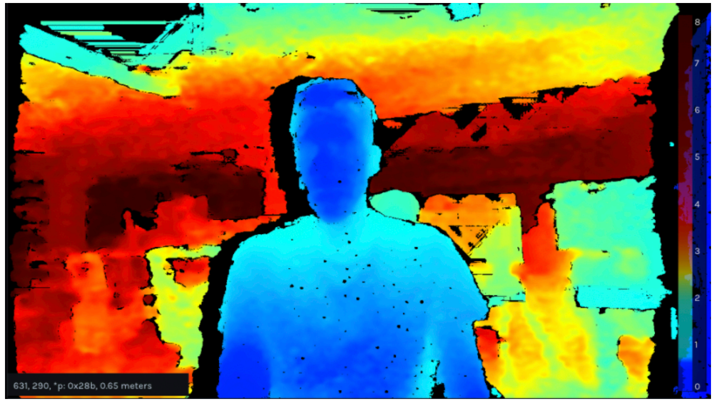
- Different types of sensors and visual data

RGB camera

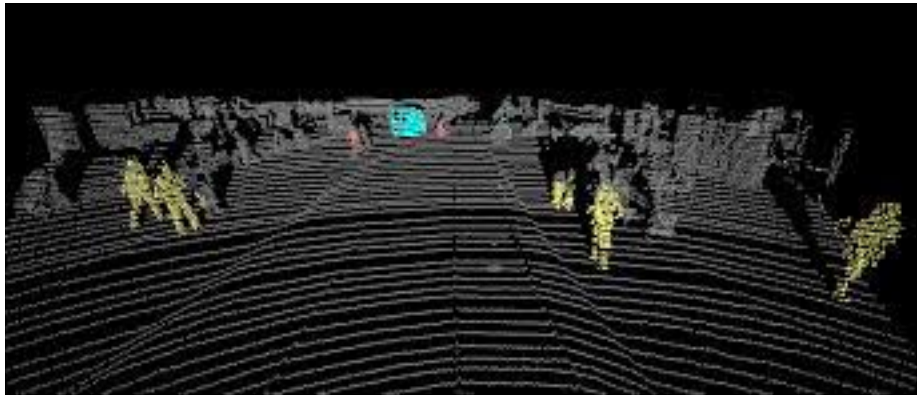
Depth camera

LiDAR

RGB image

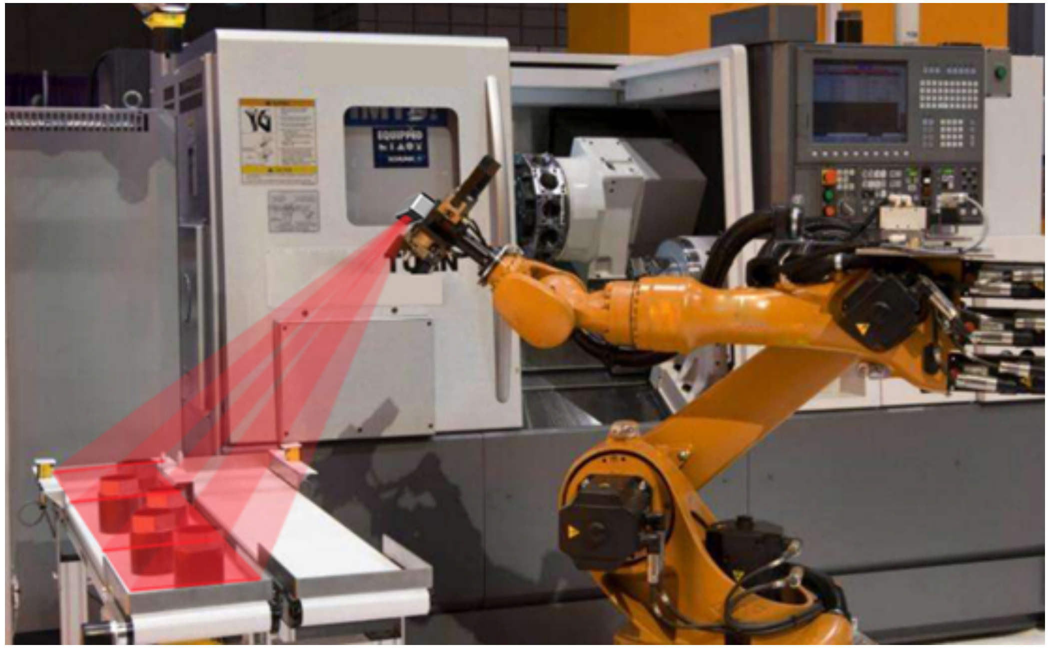


Depth image

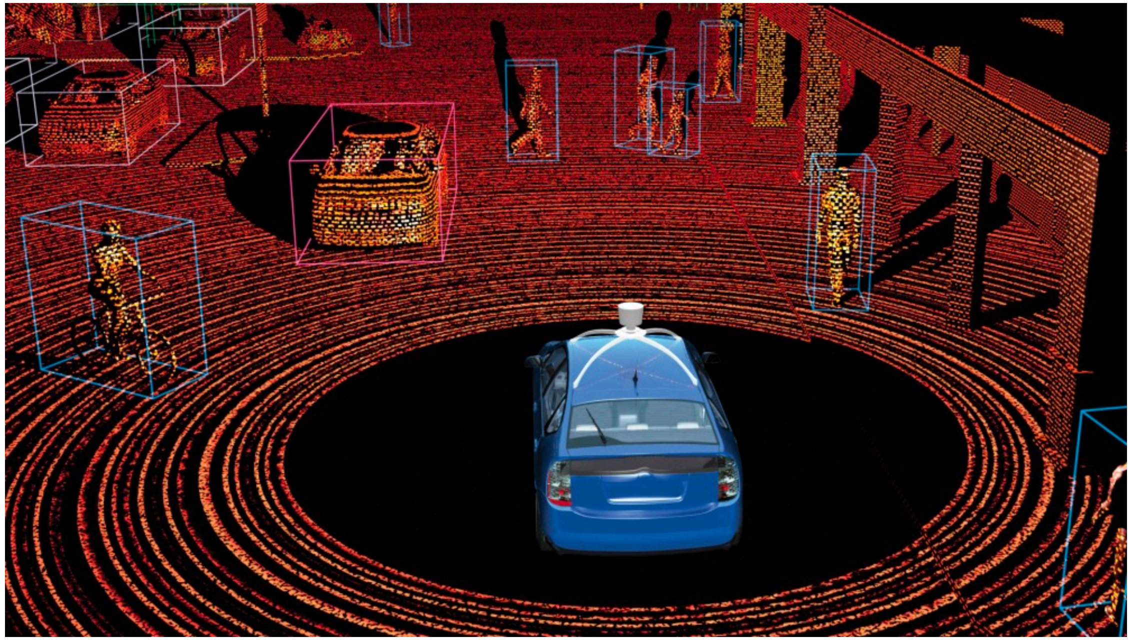


LiDAR point cloud

Robots Need 3D Vision!



- Industrial robots

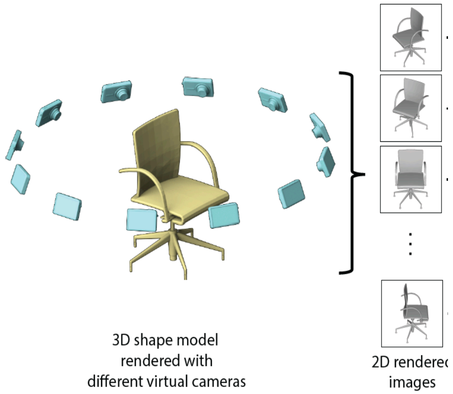


- Autonomous driving

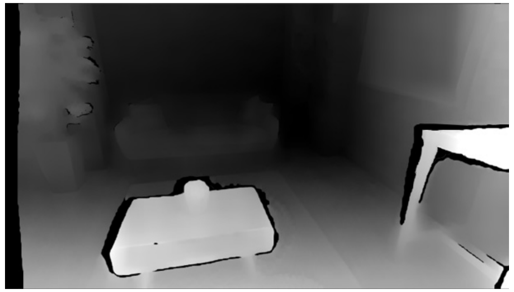
Both of them need accurate and robust 3D distance information!

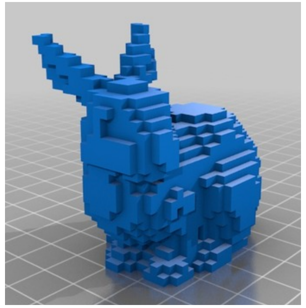
Various Representations of 3D Data

Regular form



Multi-view images

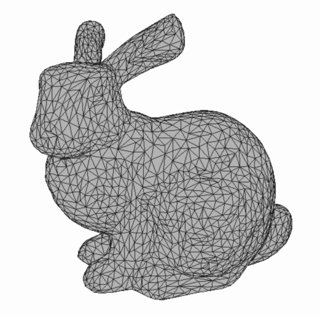




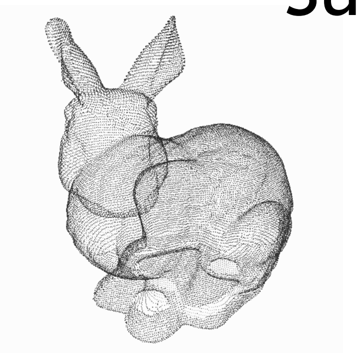
Depth

Volumetric

Irregular form



Surface Mesh



Point Cloud

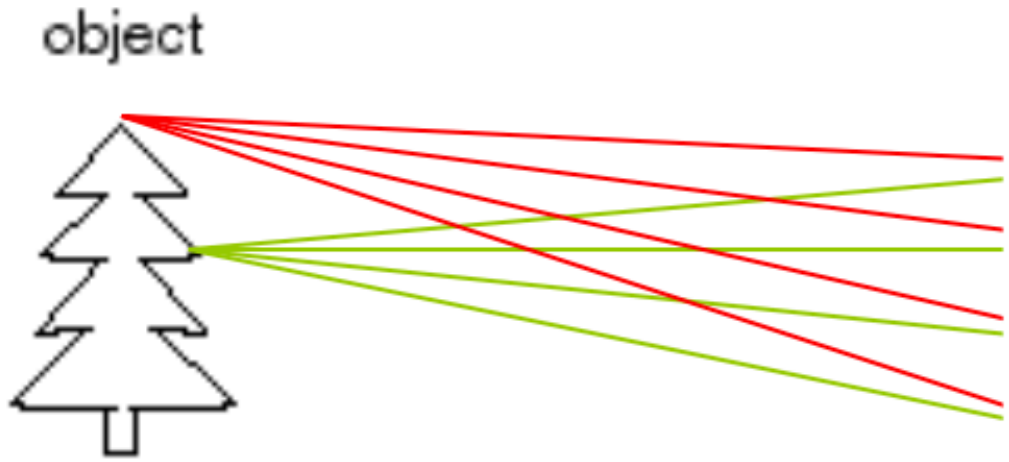
$$F(x) = 0$$

Implicit representation

Camera Model

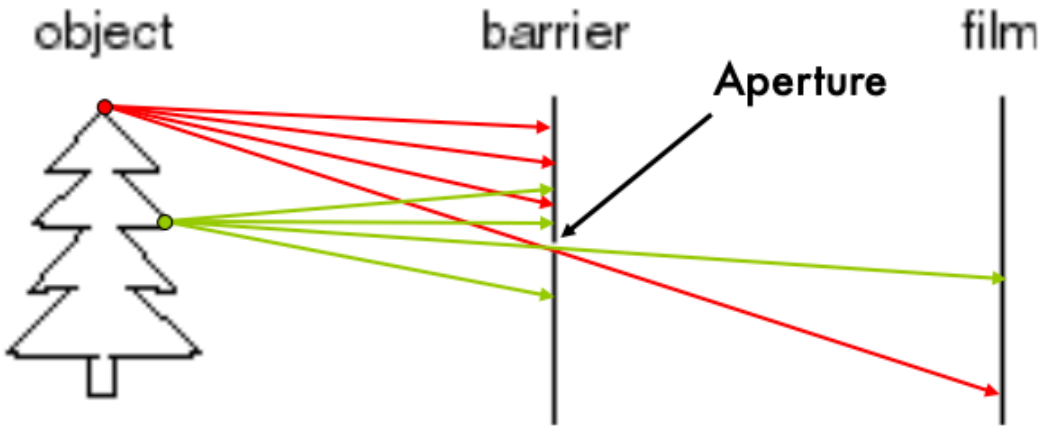
Some slides are borrowed and modified from Stanford CS 231A

How do We See a World?



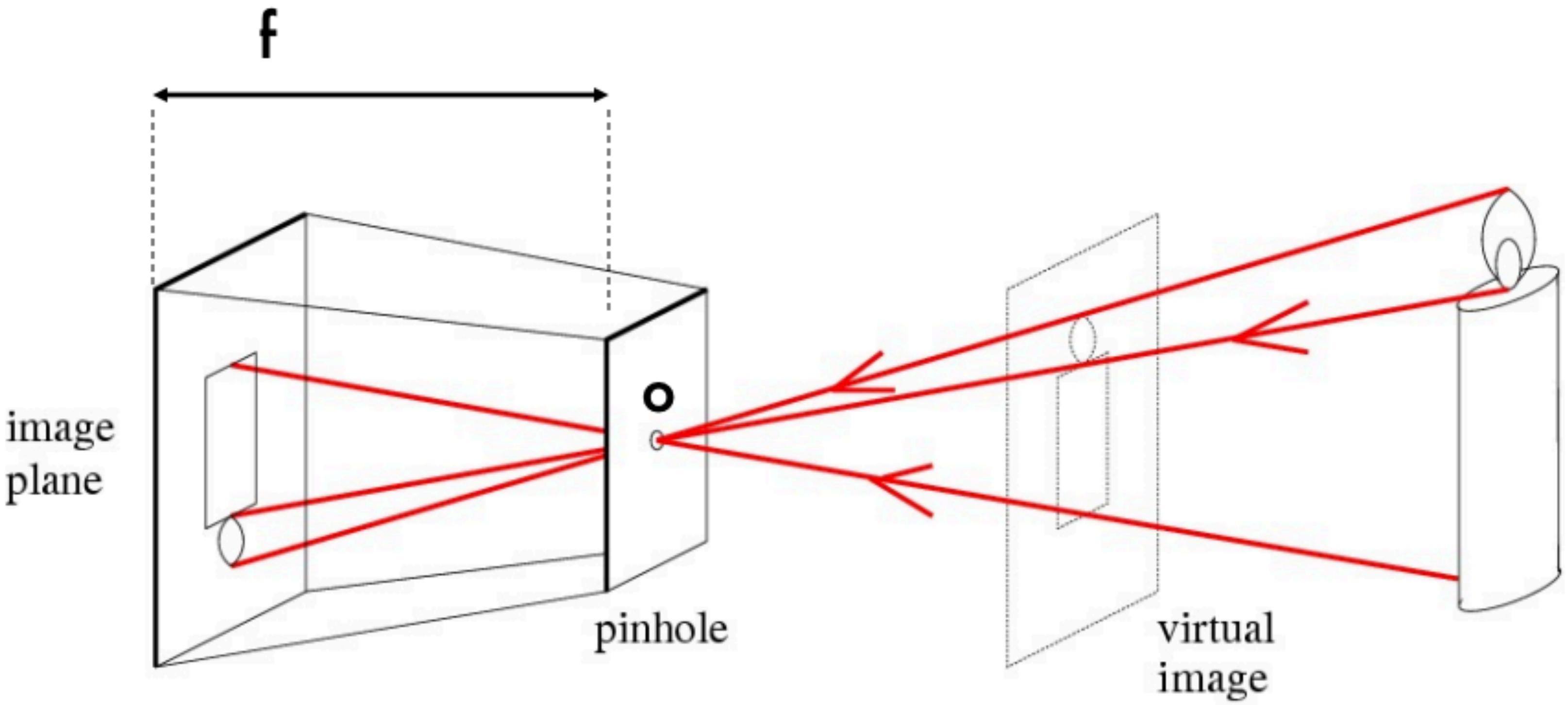
- Let's design a camera
  - Idea 1: put a piece of film in front of an object
  - Do we get a reasonable image?

Pinhole Camera



- Idea 2: Add a barrier to block off most of the rays
  - This reduces blurring
  - The opening known as the **aperture**

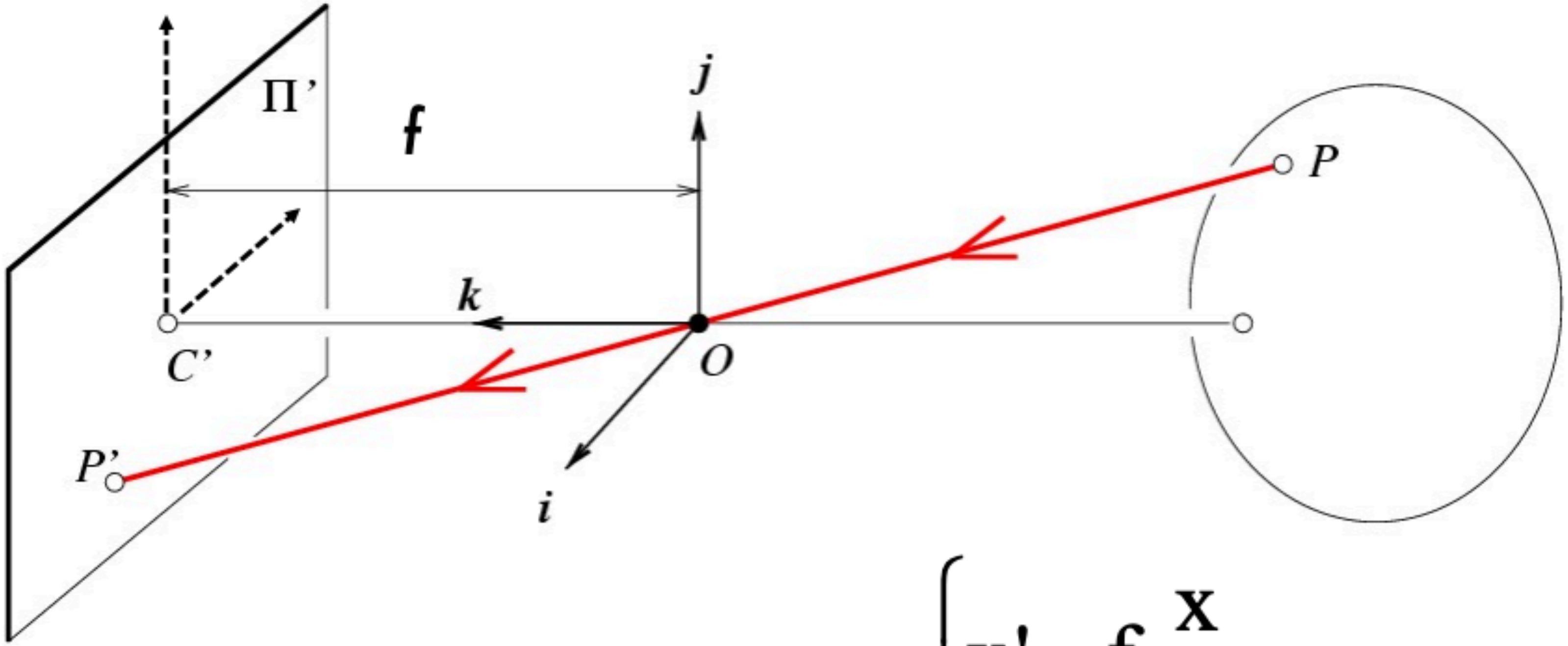
Pinhole Camera



f = focal length

o = aperture = pinhole = center of the camera

Pinhole Camera



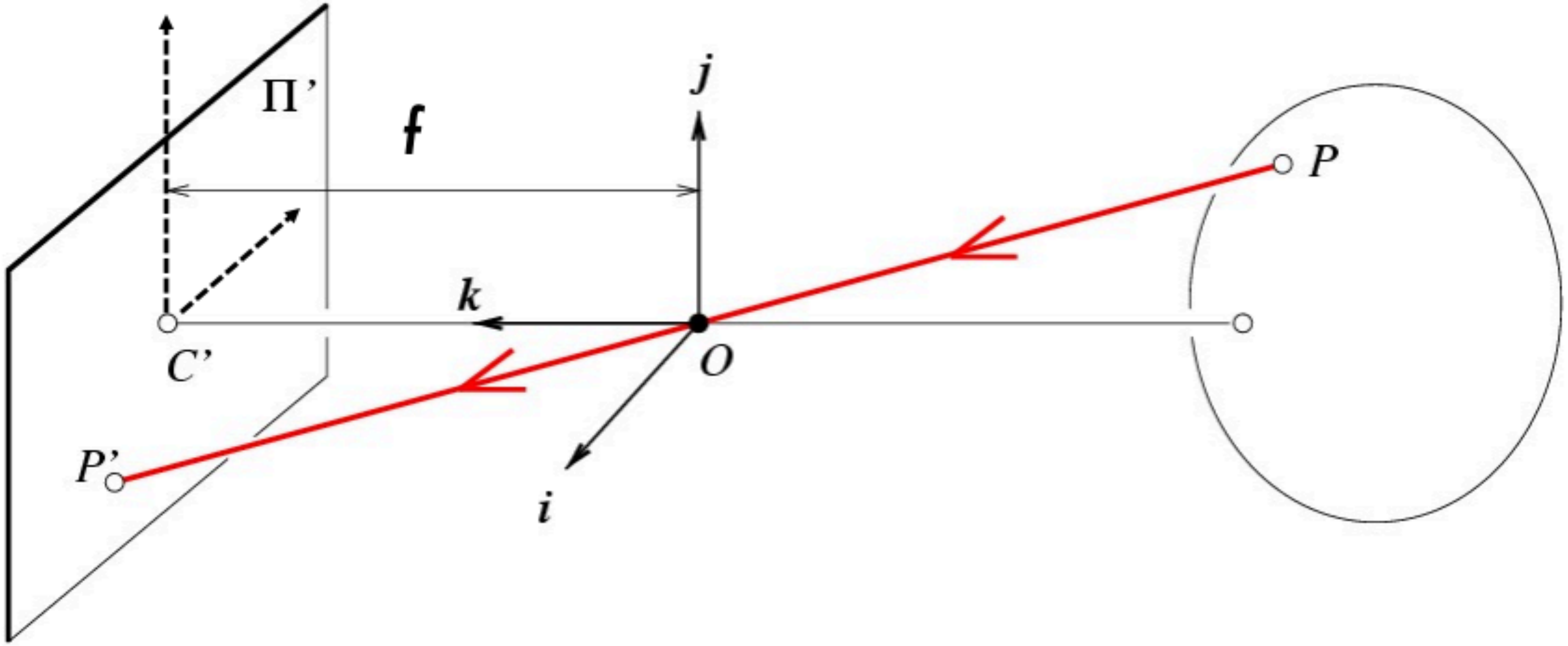
$$P = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \rightarrow P' = \begin{bmatrix} x' \\ y' \end{bmatrix}$$

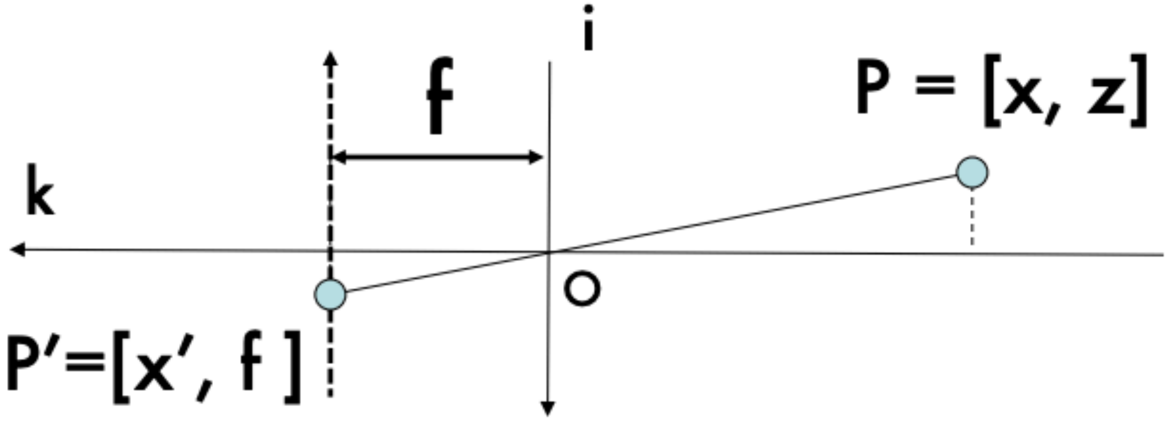
$$\begin{cases} x' = f\,\dfrac{x}{z} \\ y' = f\,\dfrac{y}{z} \end{cases}$$

[Eq. 1]

Derived using similar triangles

Pinhole Camera



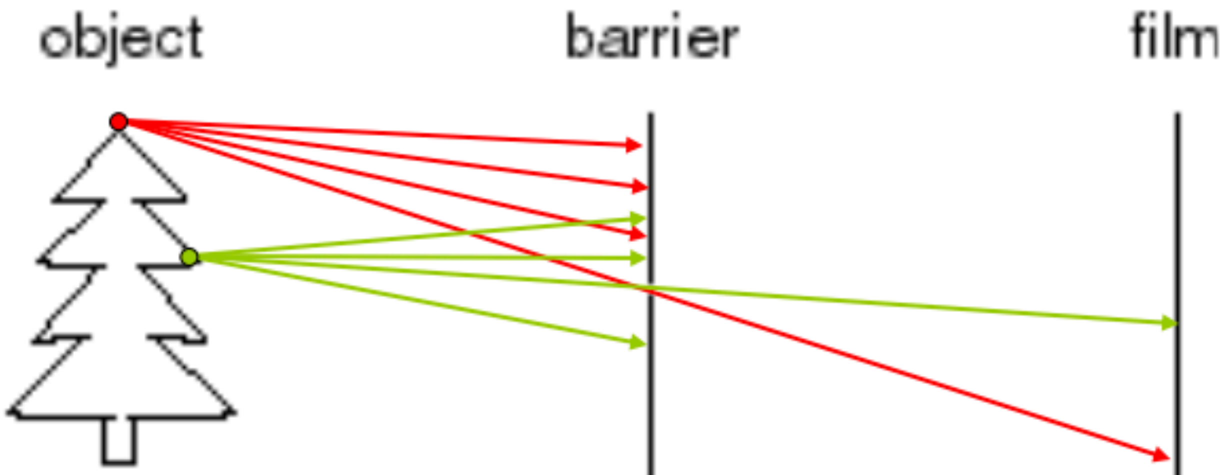


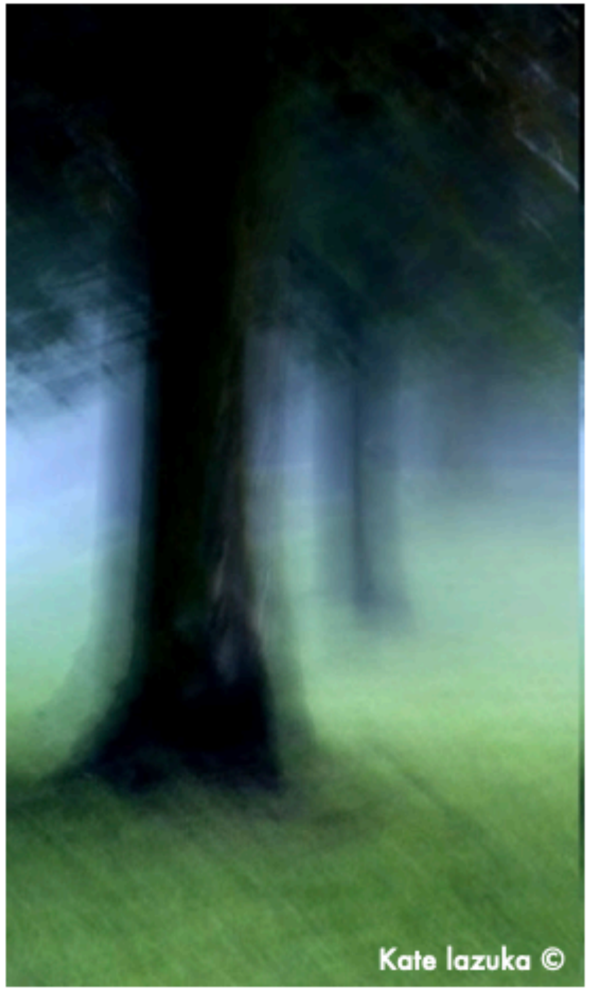
[Eq. 2]

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

Pinhole Camera

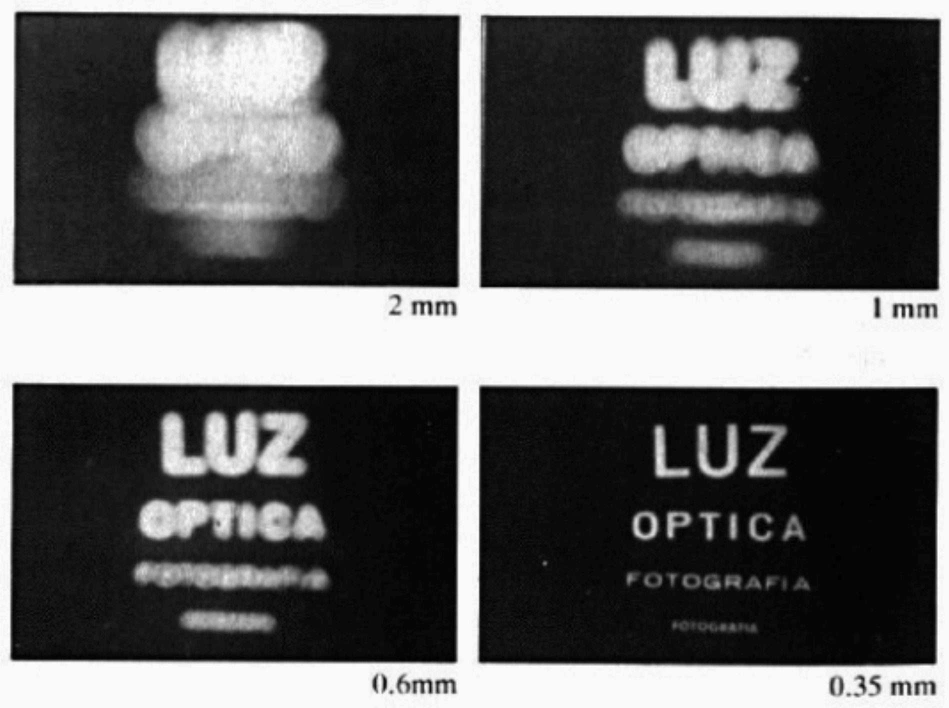
Is the size of the aperture important?





Pinhole Camera

Shrinking aperture size

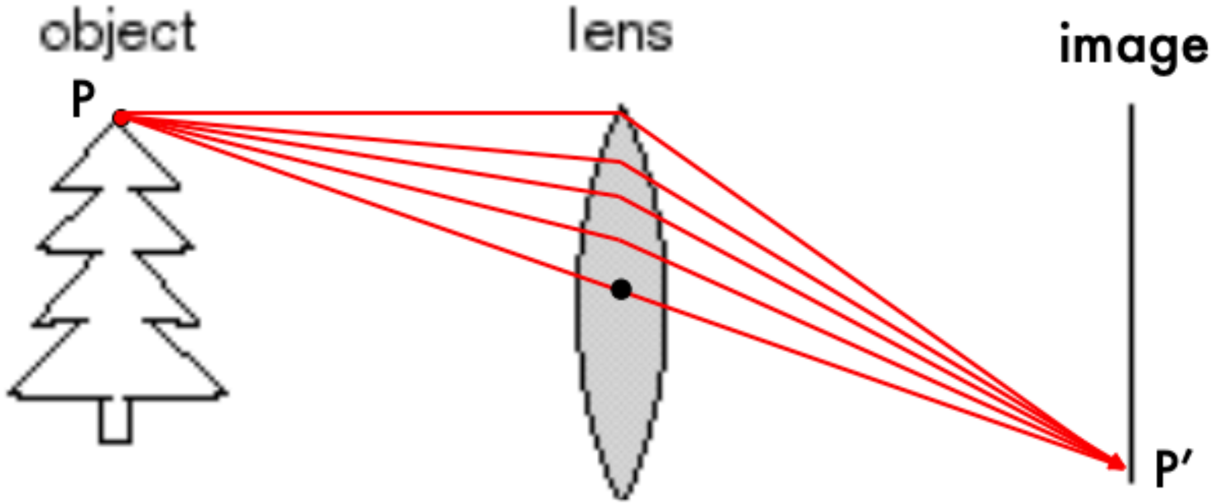


-What happens if the aperture is too small?

-Less light passes through

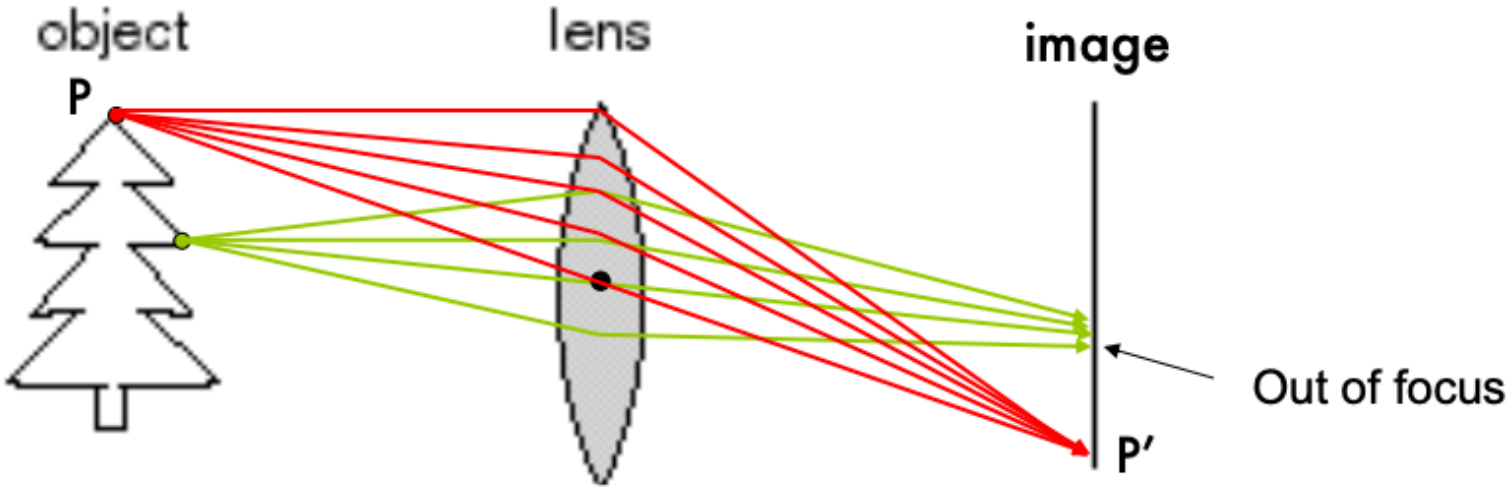
Adding lenses!

Camera and Lense



- A lens focuses light onto the film

Camera and Lense

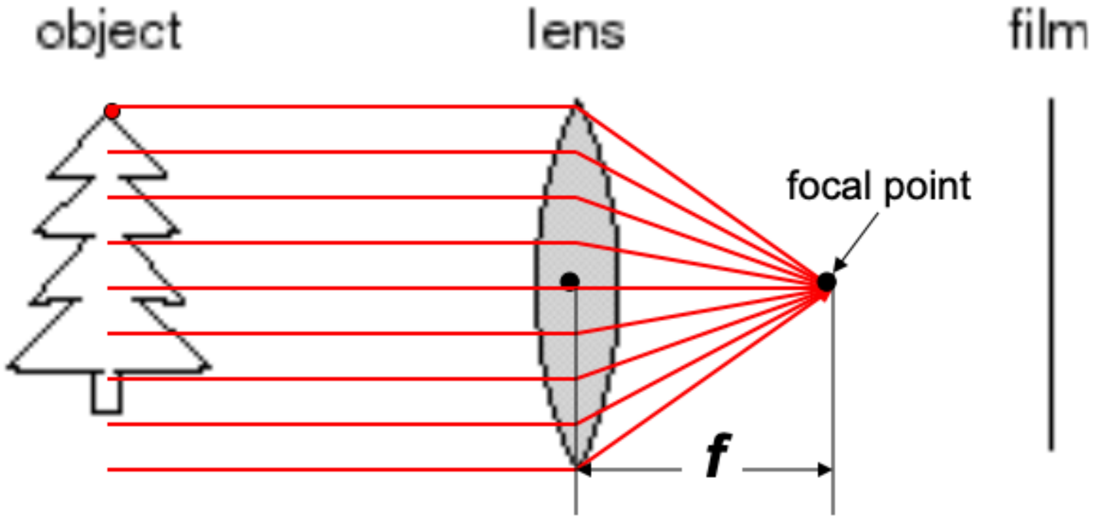


- A lens focuses light onto the film
  - There is a specific distance at which objects are "in focus"
  - Related to the concept of depth of field

Camera and Lense

- A lens focuses light onto the film
  - There is a specific distance at which objects are "in focus"
  - Related to the concept of depth of field

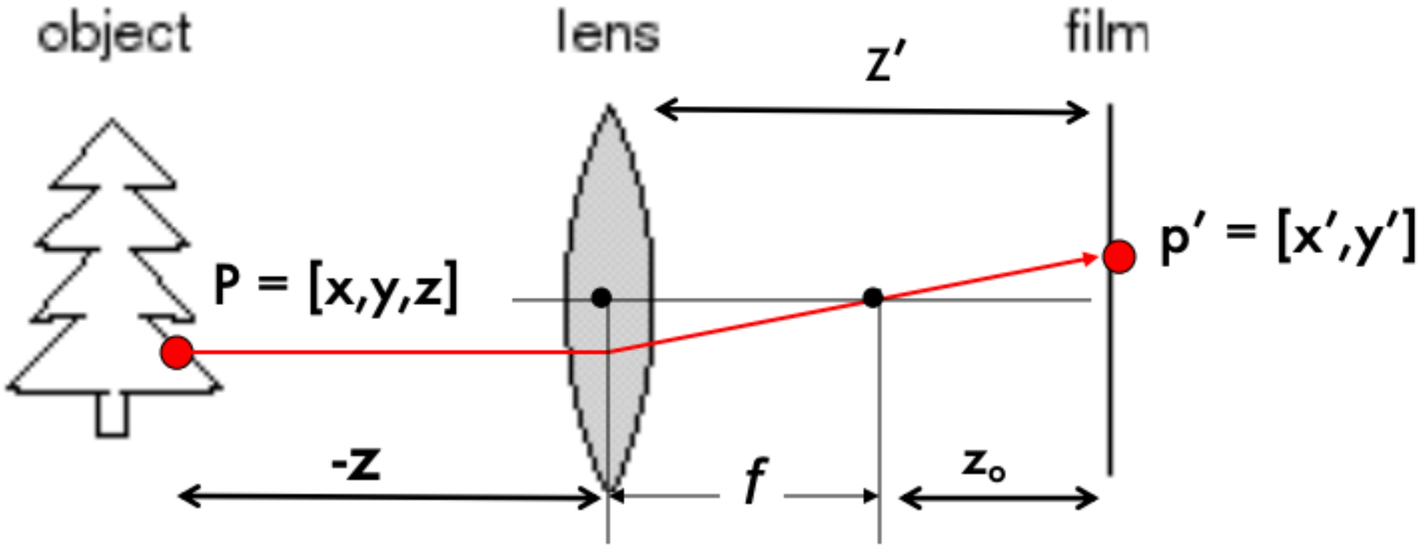
Camera and Lense



– **A lens focuses light onto the film**

- All rays parallel to the optical (or principal) axis converge to one point (the *focal point*) on a plane located at the *focal length f* from the center of the lens.
- Rays passing through the center are not deviated

Paraxial Refraction Model



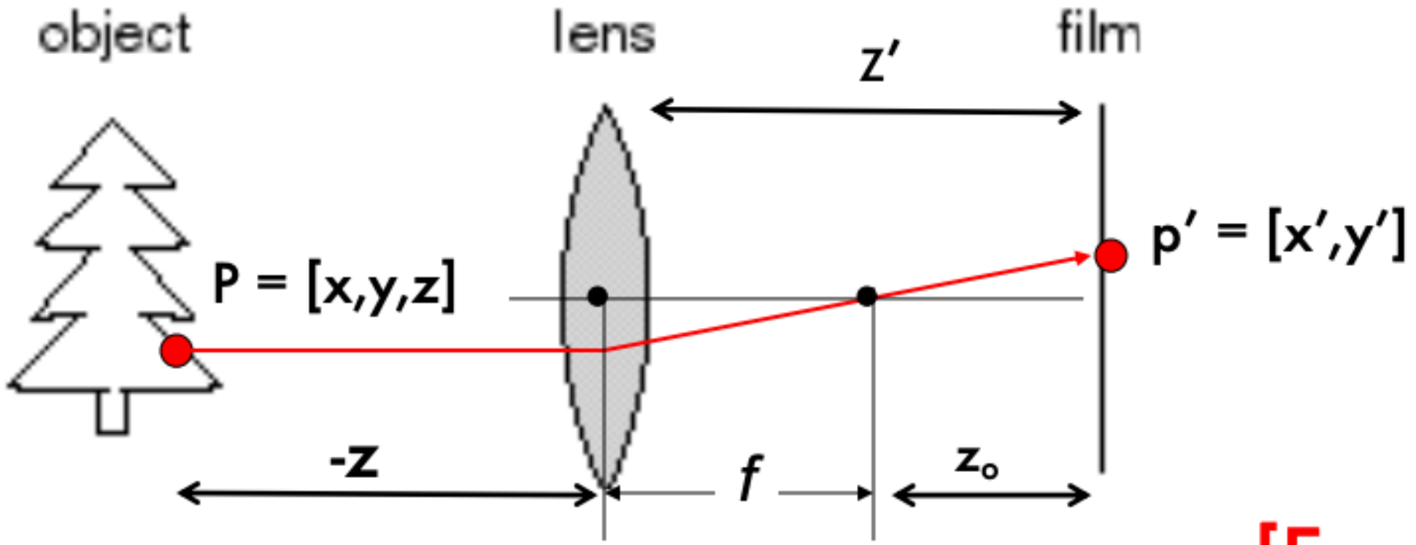
From Snell's law:

[Eq. 3]

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

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

[Eq. 1]



[Eq. 4]

$$\boxed{z' = f + z_o}$$

From Snell's law:

[Eq. 3]

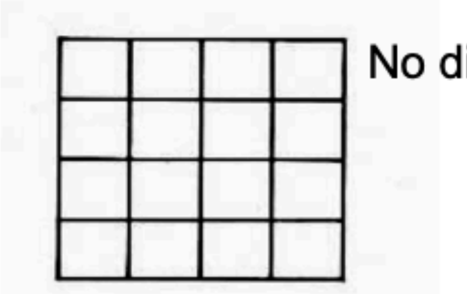
$$\begin{cases} x' = z'\dfrac{x}{z} \\ y' = z'\dfrac{y}{z} \end{cases}$$

$$f = \frac{R}{2(n-1)}$$

[FP] sec 1.1, page 8

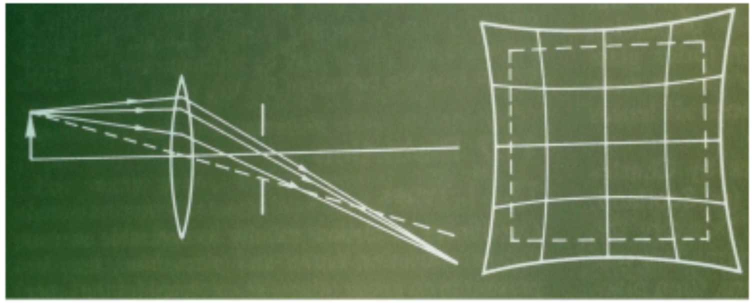
Issues with Lenses: Radial Distortion

– Deviations are most noticeable for rays that pass through the edge of the lens



No distortion

Pin cushion



Barrel (fisheye lens)

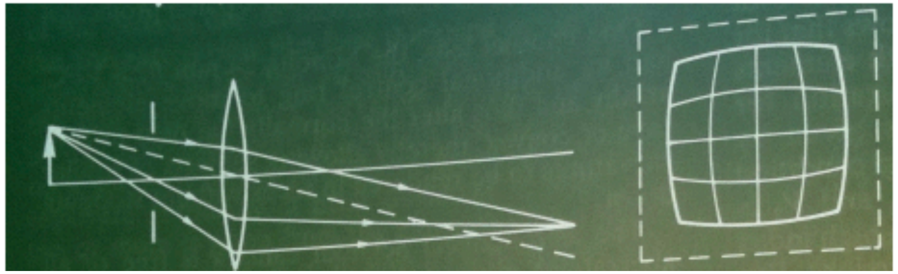


Image magnification decreases with distance from the optical axis

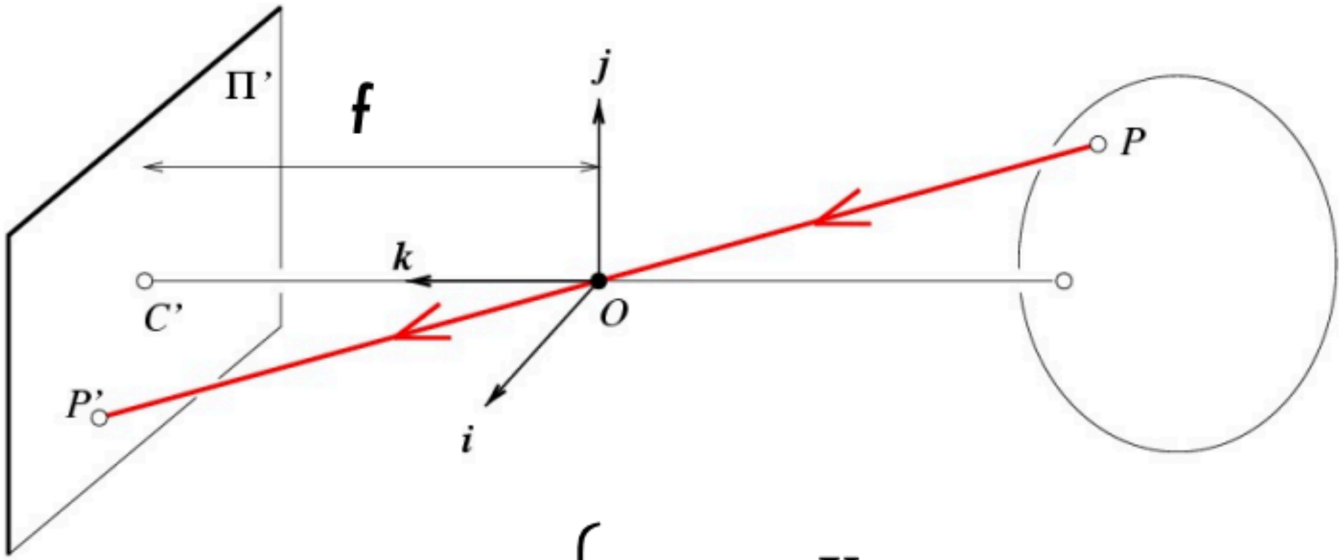
The Geometry of Pinhole Camera

- Intrinsics
  - The intrinsic properties of the camera

- Extrinsics
  - The pose of the camera (in the world reference frame)

Camera Model:Intrinsics

Pinhole Camera



$$P = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \rightarrow P' = \begin{bmatrix} x' \\ y' \end{bmatrix}$$

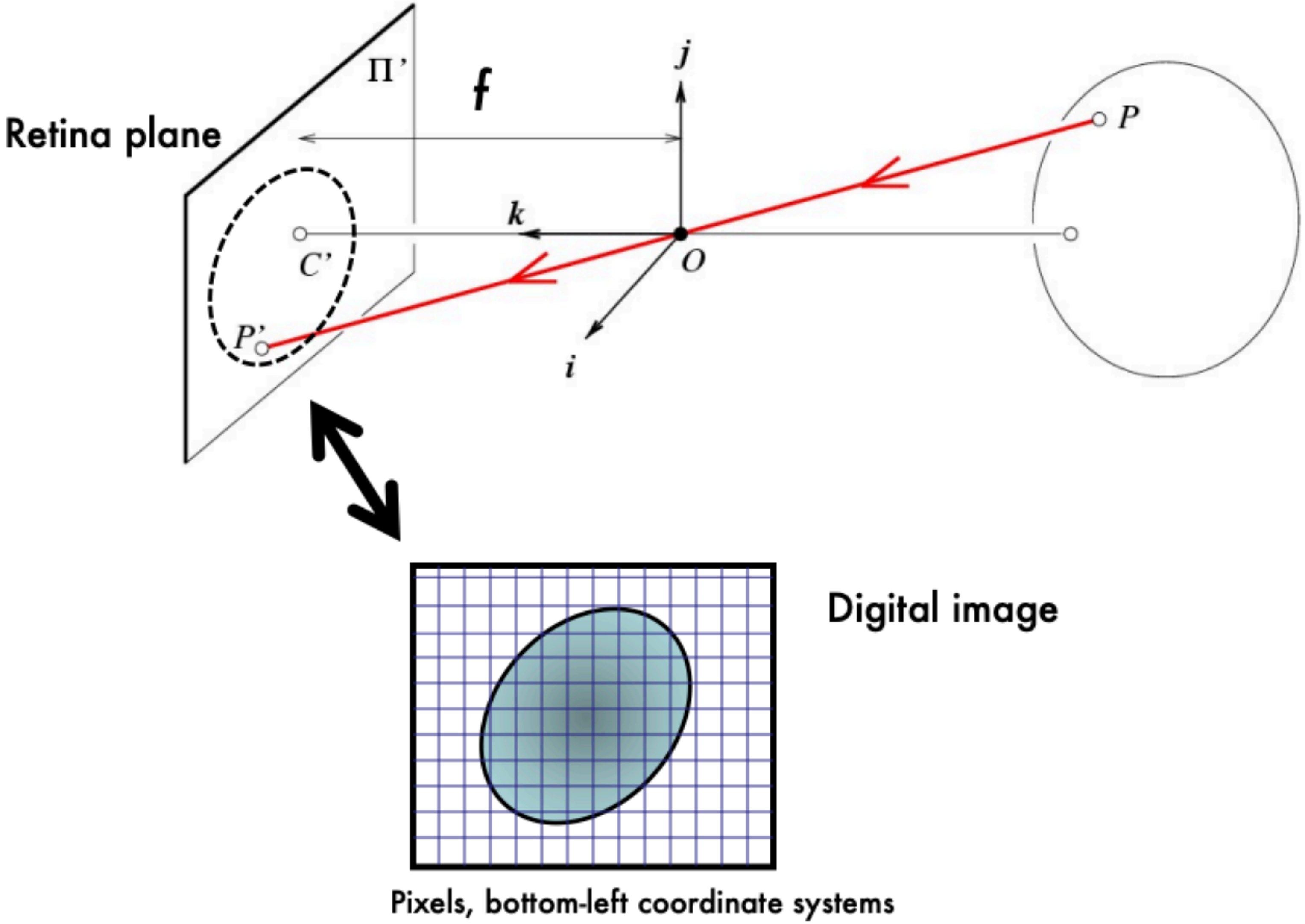
$$\begin{cases} x' = f\,\dfrac{x}{z} \\[2ex] y' = f\,\dfrac{y}{z} \end{cases}$$

$$\Re^3 \xrightarrow{E} \Re^2$$

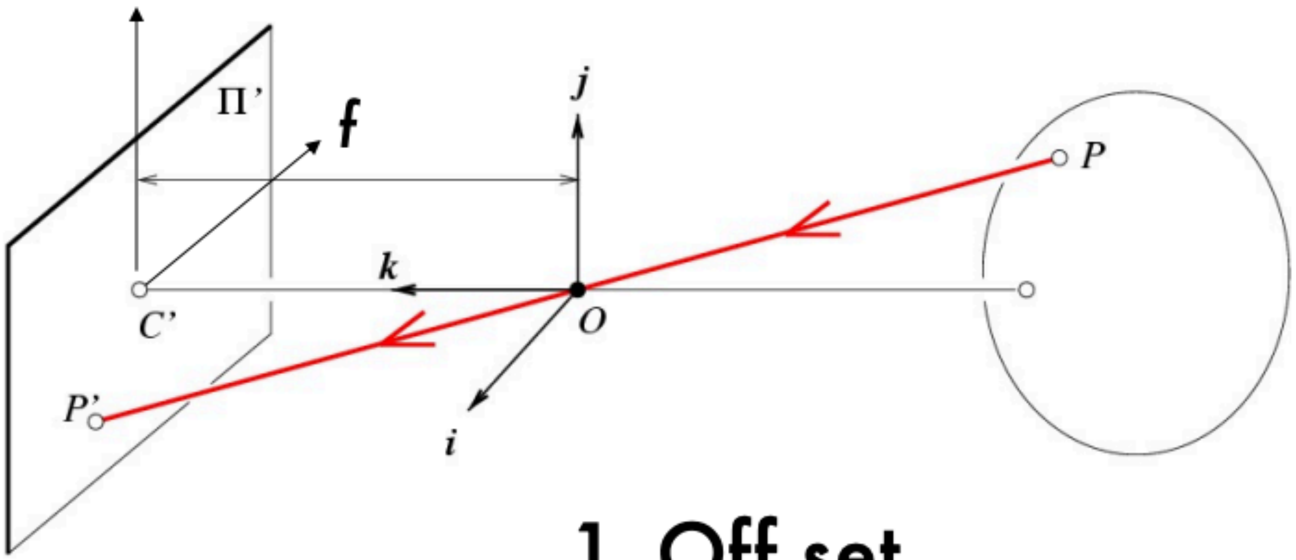
[Eq. 1]

f = focal length
o = center of the camera

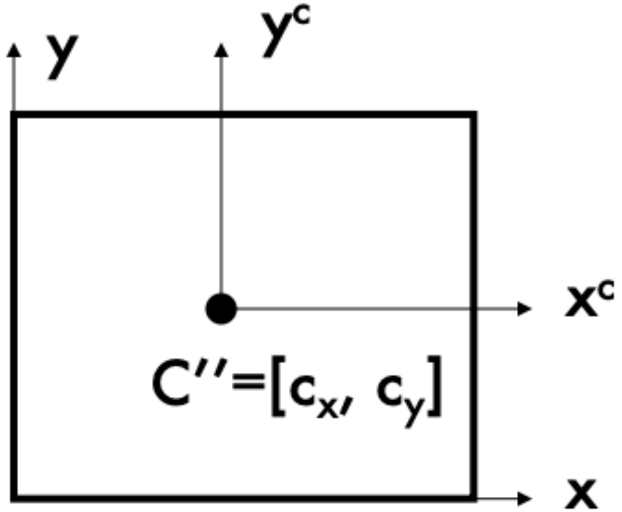
From Retina Plane to Images



Coordinate Systems



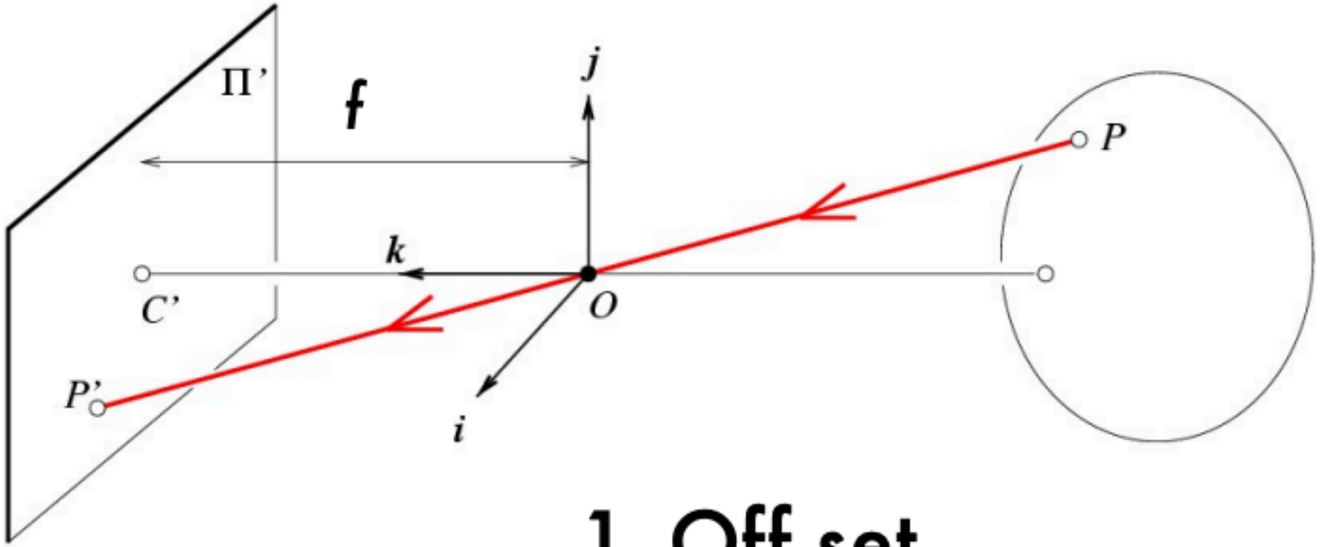
1. Off set

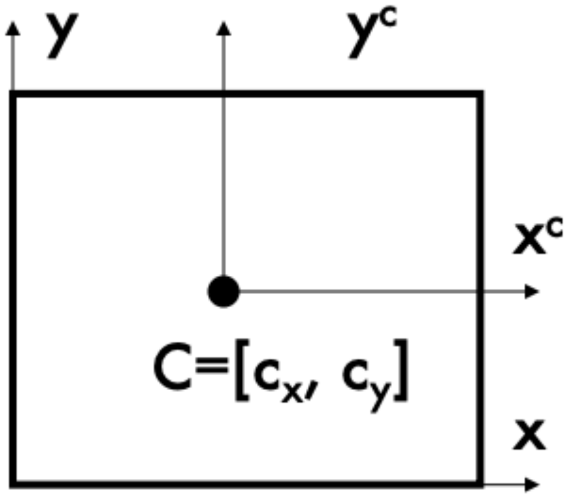


$$(x, y, z) \rightarrow (f\frac{x}{z} + c_x, f\frac{y}{z} + c_y)$$

[Eq. 5]

Converting to Pixels





1. Off set
2. From metric to pixels

$$(x, y, z) \rightarrow (\underbrace{\boxed{f\,k}}_{\alpha}\frac{x}{z} + c_x, \underbrace{\boxed{f\,l}}_{\beta}\frac{y}{z} + c_y)$$

[Eq. 6]

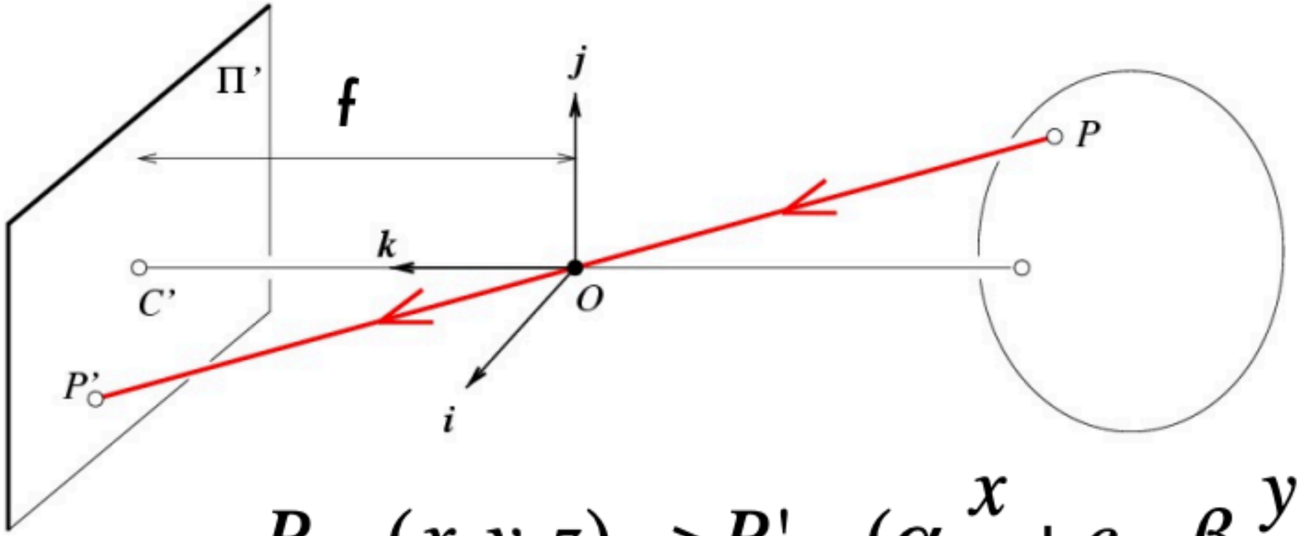
Units: k,l : pixel/m

f : m

Non-square pixels

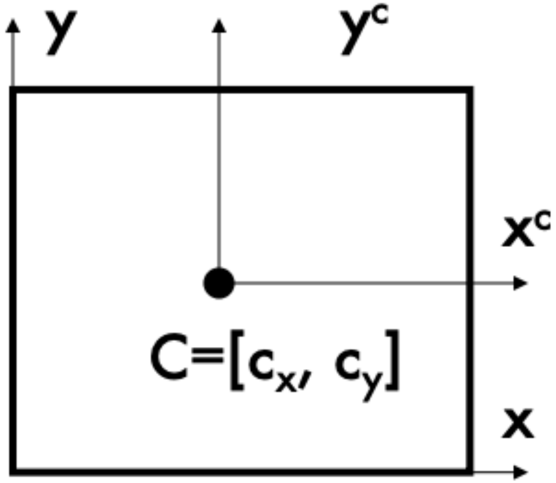
α, β : pixel

Projective Transformation



$$P = (x, y, z) \rightarrow P' = (\alpha \frac{x}{z} + c_x, \beta \frac{y}{z} + c_y)$$

[Eq. 7]



- Is this a linear transformation?
  No – division by z is nonlinear
- Can we express it in a matrix form?

Homogeneous Coordinate System

E→H

$$(x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

homogeneous image coordinates

$$(x, y, z) \Rightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

homogeneous scene coordinates

- Converting back *from* homogeneous coordinates

H→E

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w)$$

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow (x/w, y/w, z/w)$$

Projective Transformation in H

$$P_h' = \begin{bmatrix} \alpha x + c_x z \\ \beta y + c_y z \\ z \end{bmatrix} = \begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

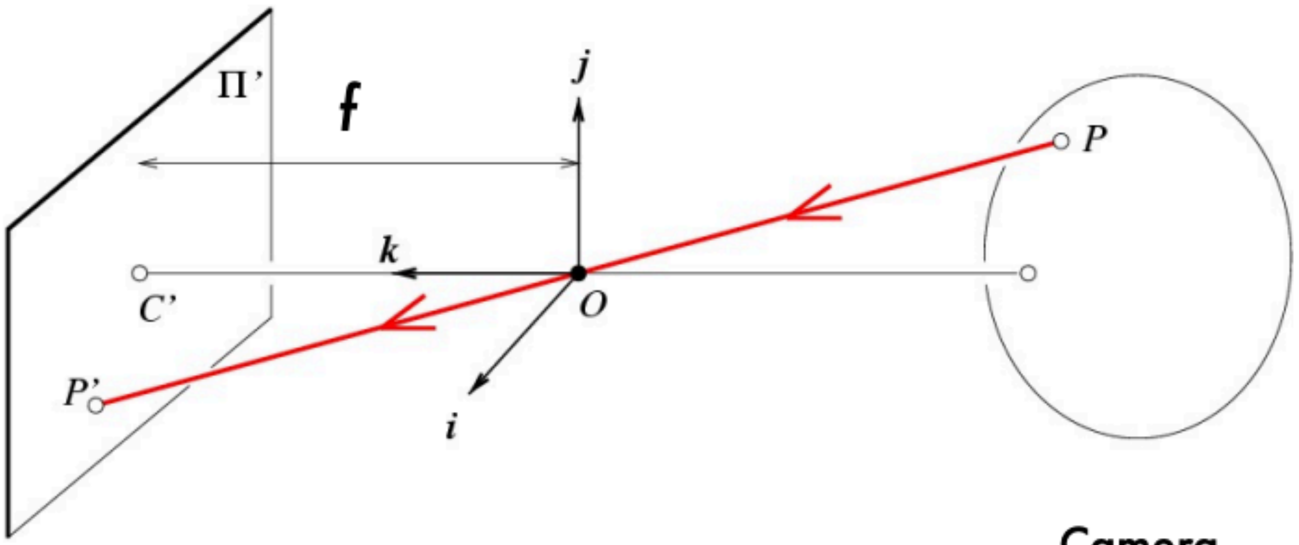
P_h

[Eq.8]

$$M = \begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\overbrace{P_h'}^{\text{Homogenous}} \rightarrow \overbrace{P' = (\alpha \frac{x}{z} + c_x, \beta \frac{y}{z} + c_y)}^{\text{Euclidian}}$$

The Camera Matrix



Camera matrix K

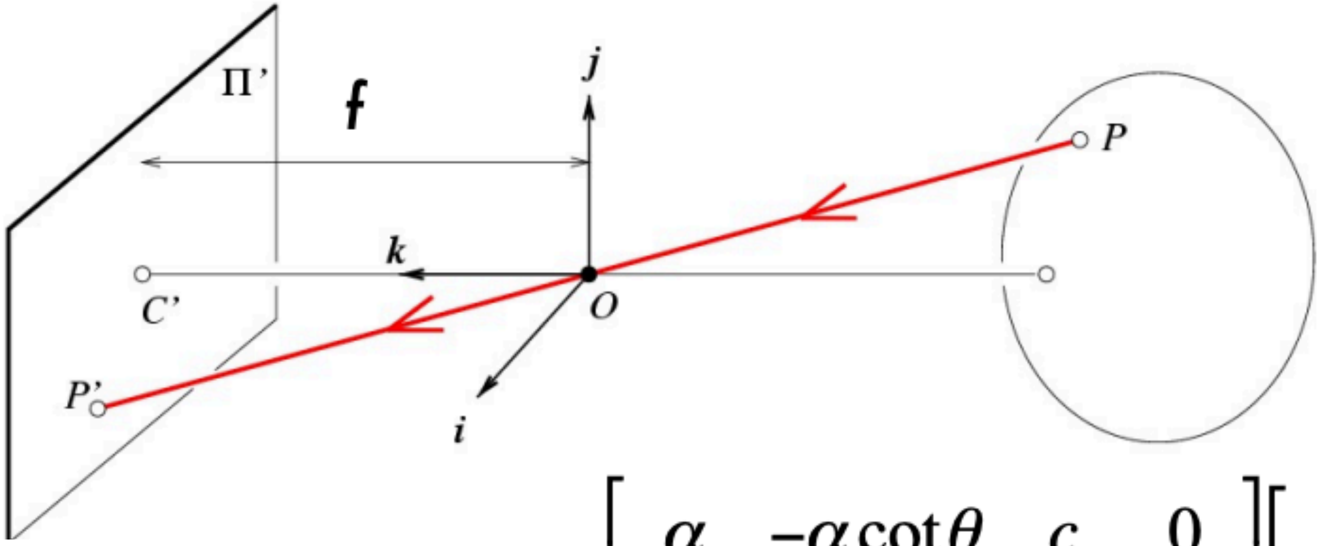
[Eq.9]

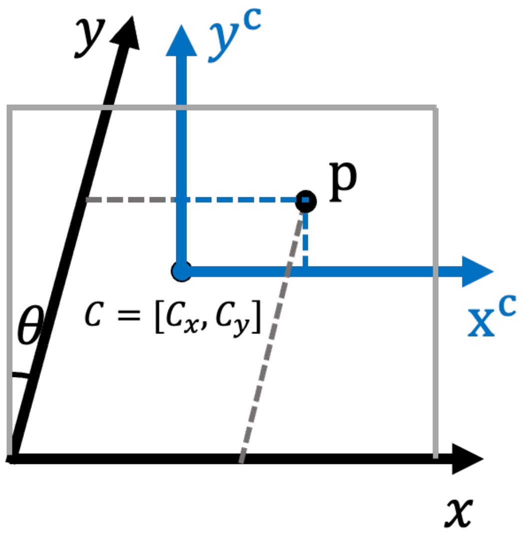
$$P' = M\,P$$

$$= K\begin{bmatrix} I & 0 \end{bmatrix} P$$

$$P' = \begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Camera Skewness

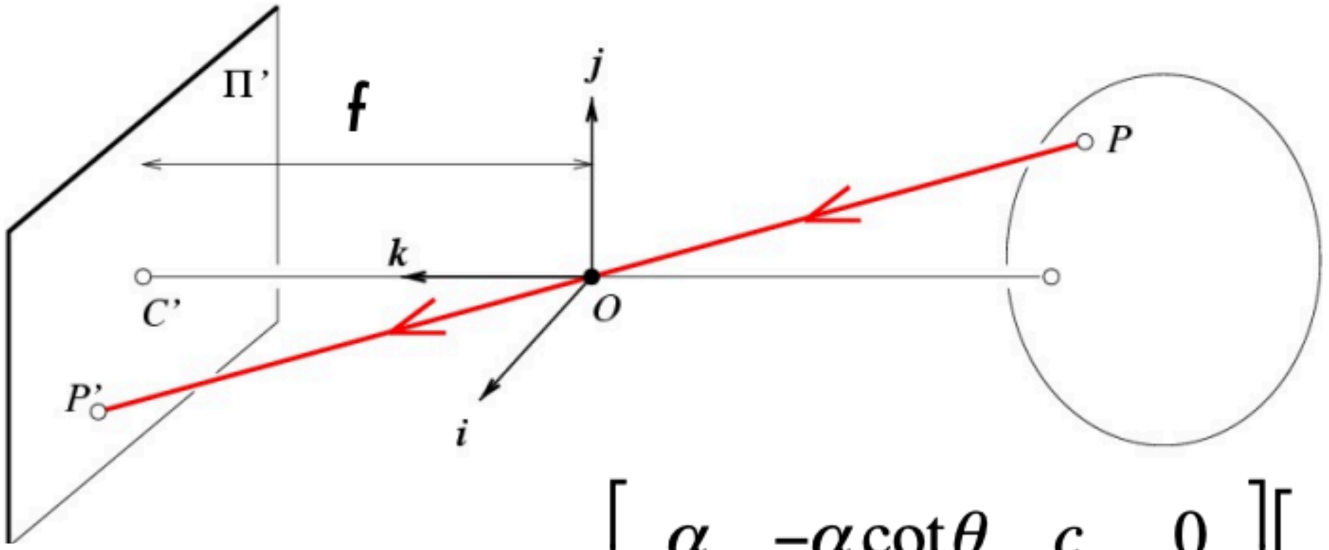


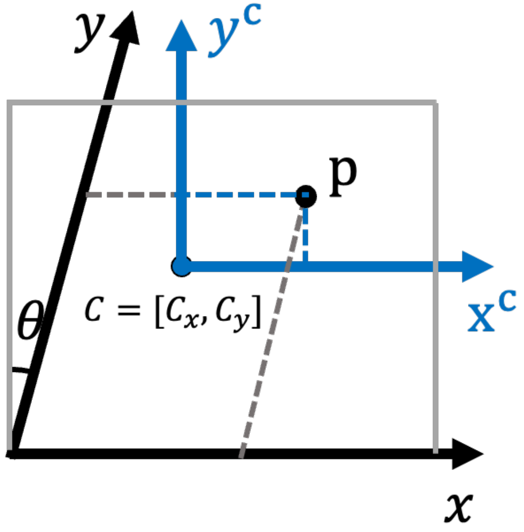


$$P' = \begin{bmatrix} \alpha & -\alpha\cot\theta & c_x & 0 \\ 0 & \frac{\beta}{\sin\theta} & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

sec. 1.2.2 [FP] or sec. 6.2.4 [HZ]

Degree of Freedom of K





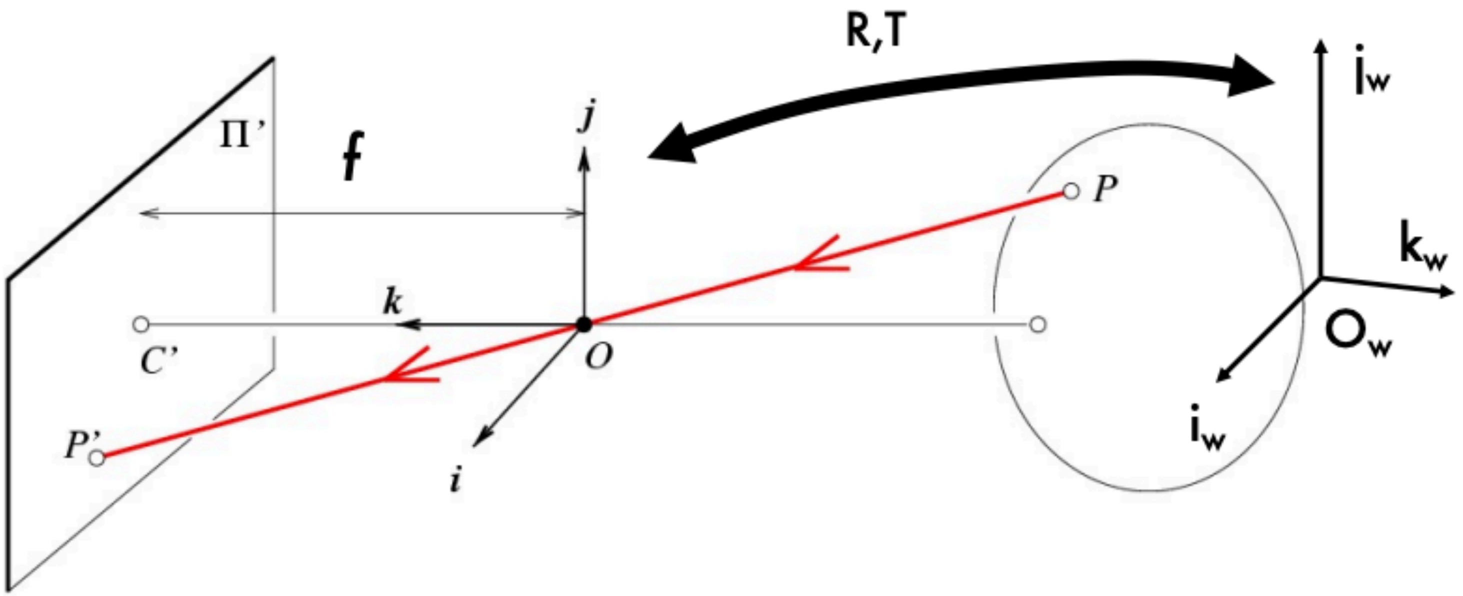
$$P' = \begin{bmatrix} \alpha & -\alpha\cot\theta & c_x & 0 \\ 0 & \dfrac{\beta}{\sin\theta} & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

How many degrees of freedom does K have?

5 degrees of freedom!

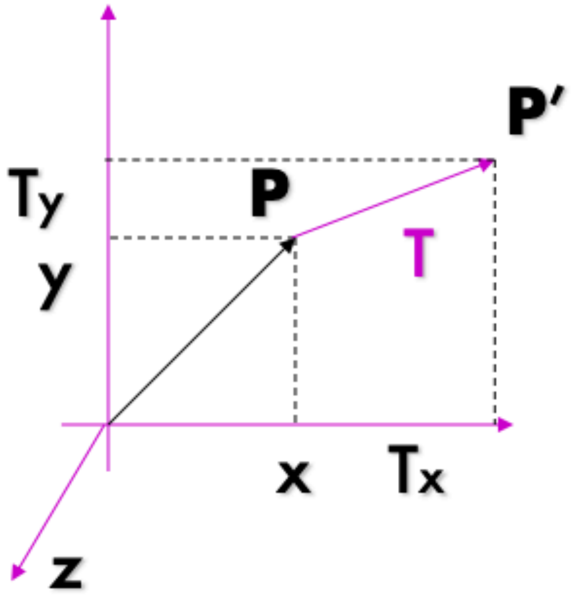
Camera Model:Extrinsics

World Reference Frame



- The mapping so far is defined within the camera reference system
- What if an object is represented in the world reference system?
- Need to introduce an additional mapping from world ref system to camera ref system

3D Translation of Points



$$T = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$$

$$P' \rightarrow \begin{bmatrix} \mathbf{I} & T \\ 0 & 1 \end{bmatrix}_{4\times4} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

A translation vector in 3D has 3 degrees of freedom

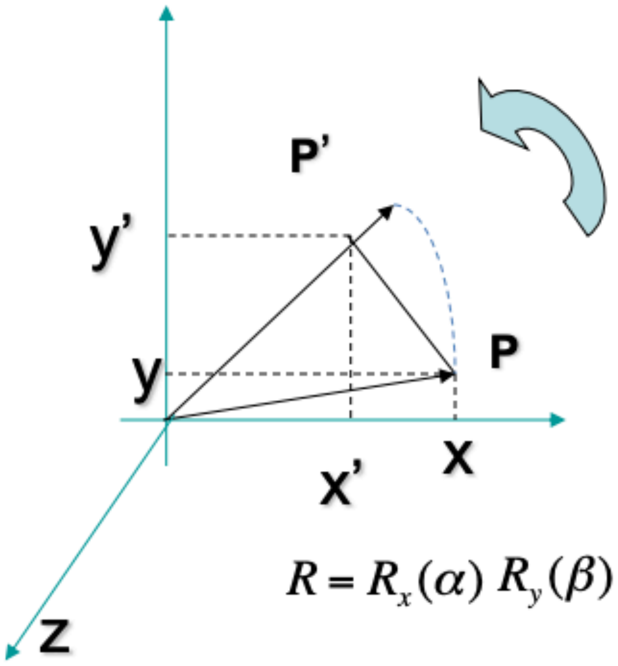
3D Rotation of Points

Rotation around the coordinate axes, **counter-clockwise**:

$$R_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}$$

$$R_y(\beta) = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}$$

$$R_z(\gamma) = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



$$R = R_x(\alpha)\, R_y(\beta)\, R_z(\gamma)$$

$$P' \to \begin{bmatrix} R & 0 \\ 0 & 1 \end{bmatrix}_{4\times 4} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

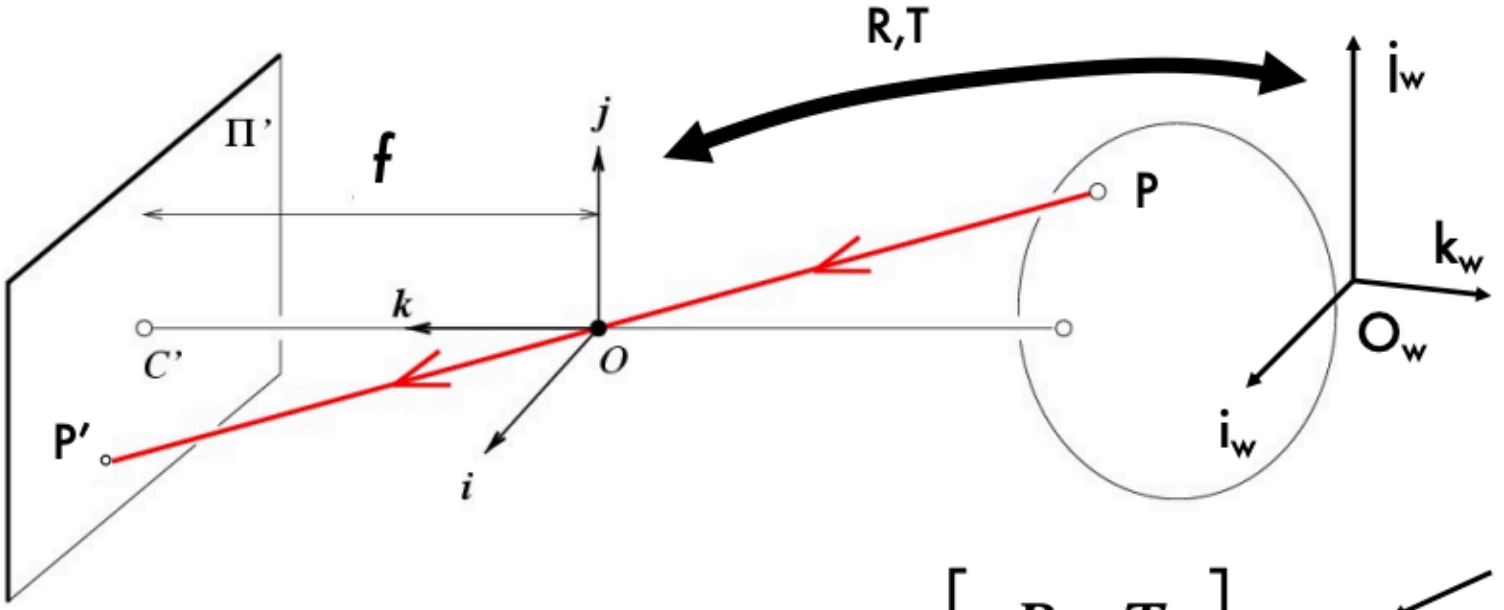
A rotation matrix in 3D has 3 degrees of freedom

3D Rotation and Translations

$$R = R_x(\alpha)\ R_y(\beta)\ R_z(\gamma) \qquad T = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$$

$$P' \rightarrow \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}_{4\times 4} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

World Reference System



In 4D homogeneous coordinates: $P = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}_{4\times 4} P_w$ $\qquad \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$

Internal parameters

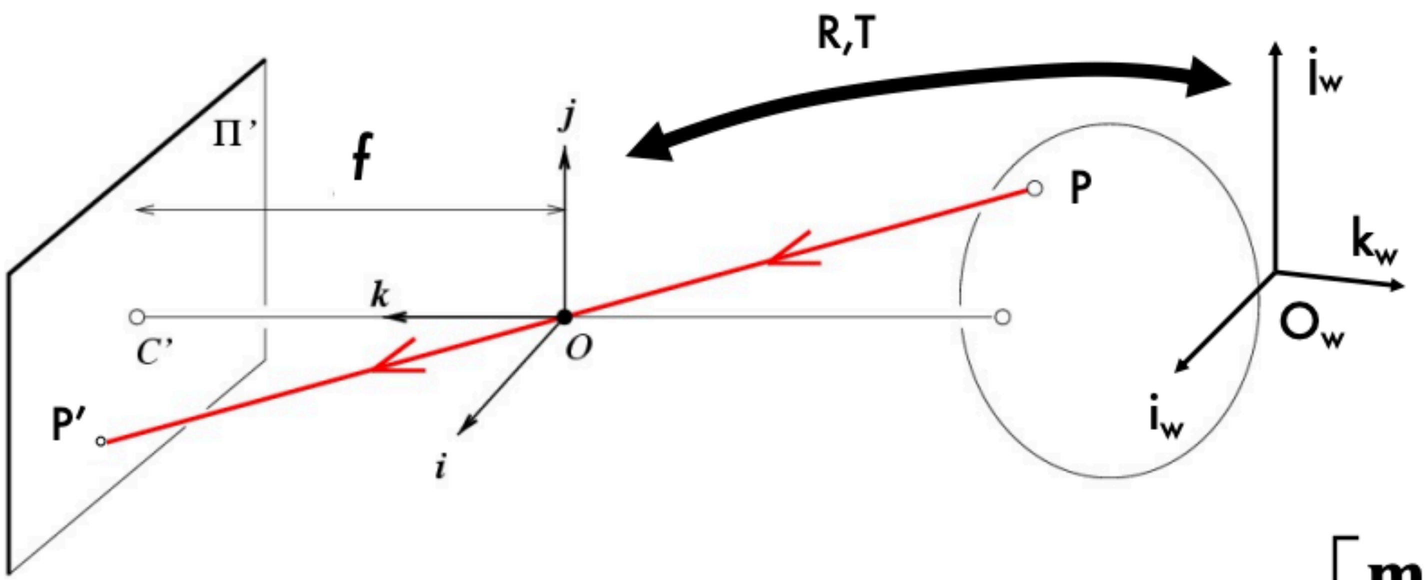
External parameters **(Camera pose)**

[Eq.9]

$$P' = K\begin{bmatrix} I & 0 \end{bmatrix} P = K\begin{bmatrix} I & 0 \end{bmatrix}\begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}_{4\times 4} P_w = \boxed{K}\begin{bmatrix} \boxed{R \quad T} \end{bmatrix} P_w$$

[Eq.11]

The Projective Transformation



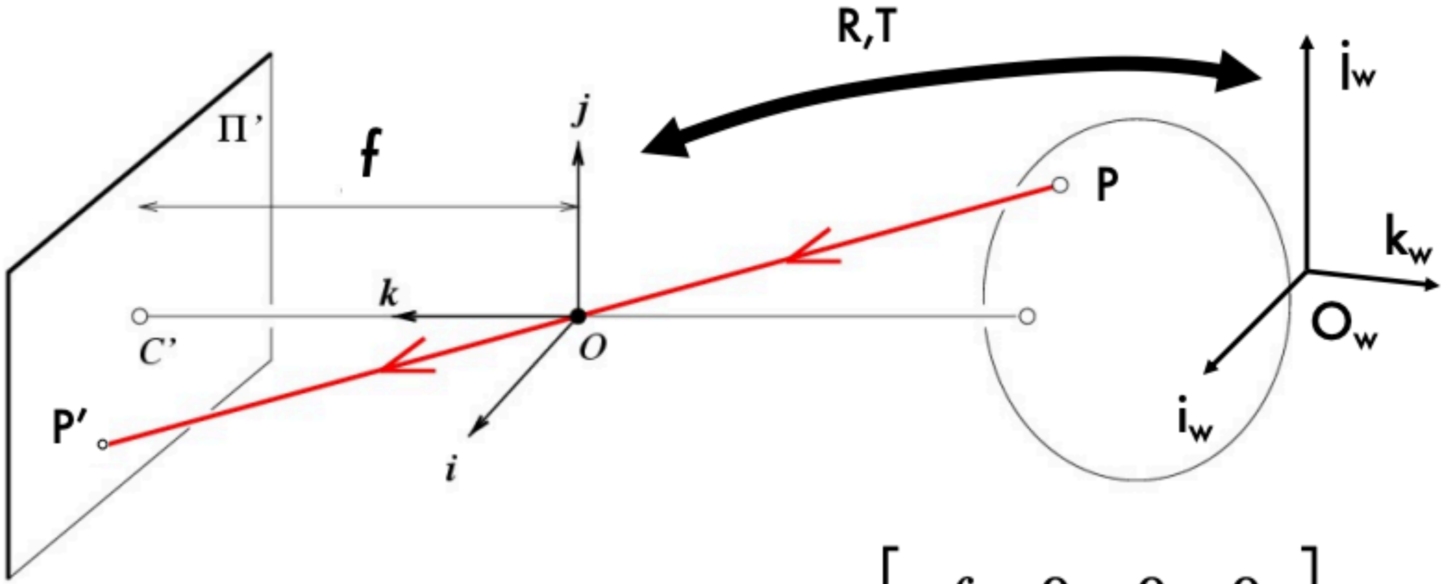
$$P'_{3\times1} = M\ P_w \quad = K_{3\times3}\begin{bmatrix} R & T \end{bmatrix}_{3\times4} P_{w\,4\times1} \qquad M = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{bmatrix} P_W = \begin{bmatrix} \mathbf{m}_1 P_W \\ \mathbf{m}_2 P_W \\ \mathbf{m}_3 P_W \end{bmatrix} \quad \mathrm{E} \rightarrow \left(\frac{\mathbf{m}_1 P_w}{\mathbf{m}_3 P_w}, \frac{\mathbf{m}_2 P_w}{\mathbf{m}_3 P_w}\right)$$ [Eq.12]

Properties of Projective Transformations

- Points project to points
- Lines project to lines
- Distant objects look smaller

Exercise

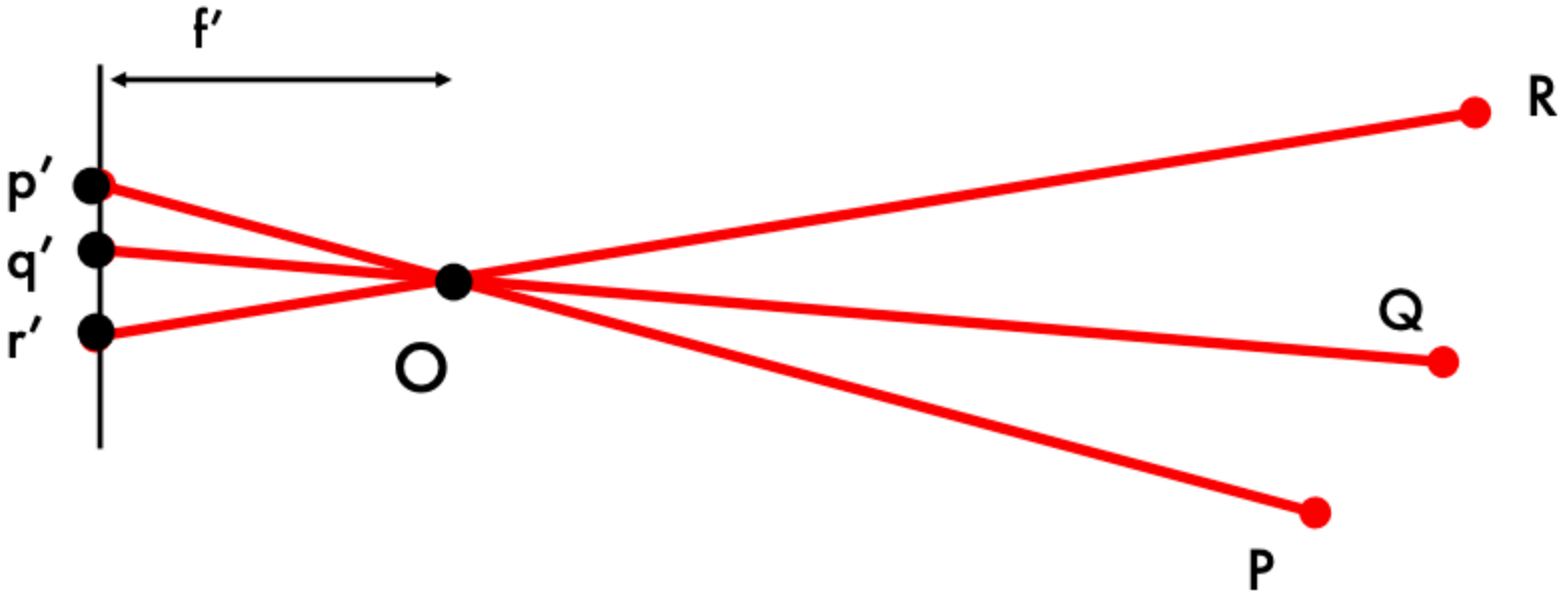


$$M = K\begin{bmatrix} R & T \end{bmatrix} = K\begin{bmatrix} I & 0 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$P_w = \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

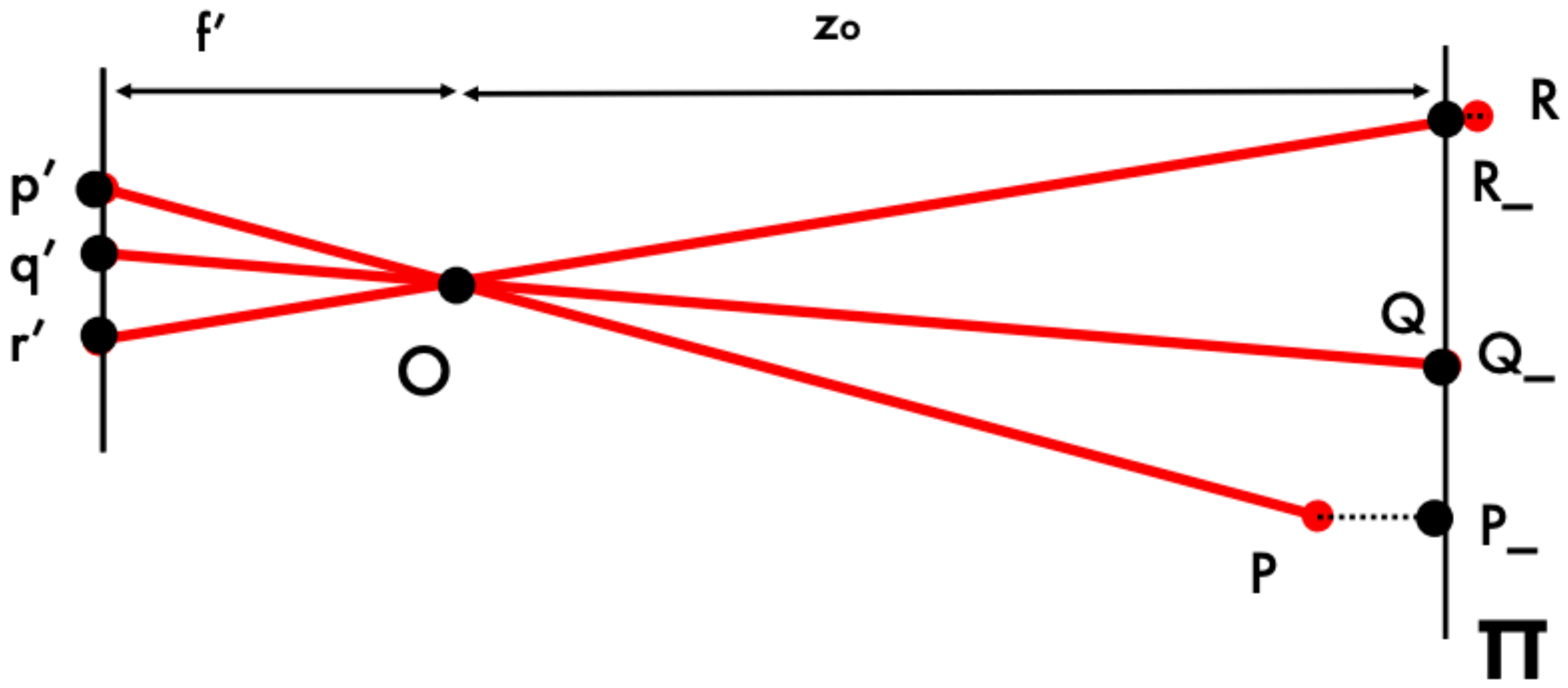
$$\rightarrow P'_E = \left(\frac{\mathbf{m}_1 P_w}{\mathbf{m}_3 P_w}, \frac{\mathbf{m}_2 P_w}{\mathbf{m}_3 P_w}\right) = \left(f\frac{x_w}{z_w}, f\frac{y_w}{z_w}\right)$$

Projective Camera

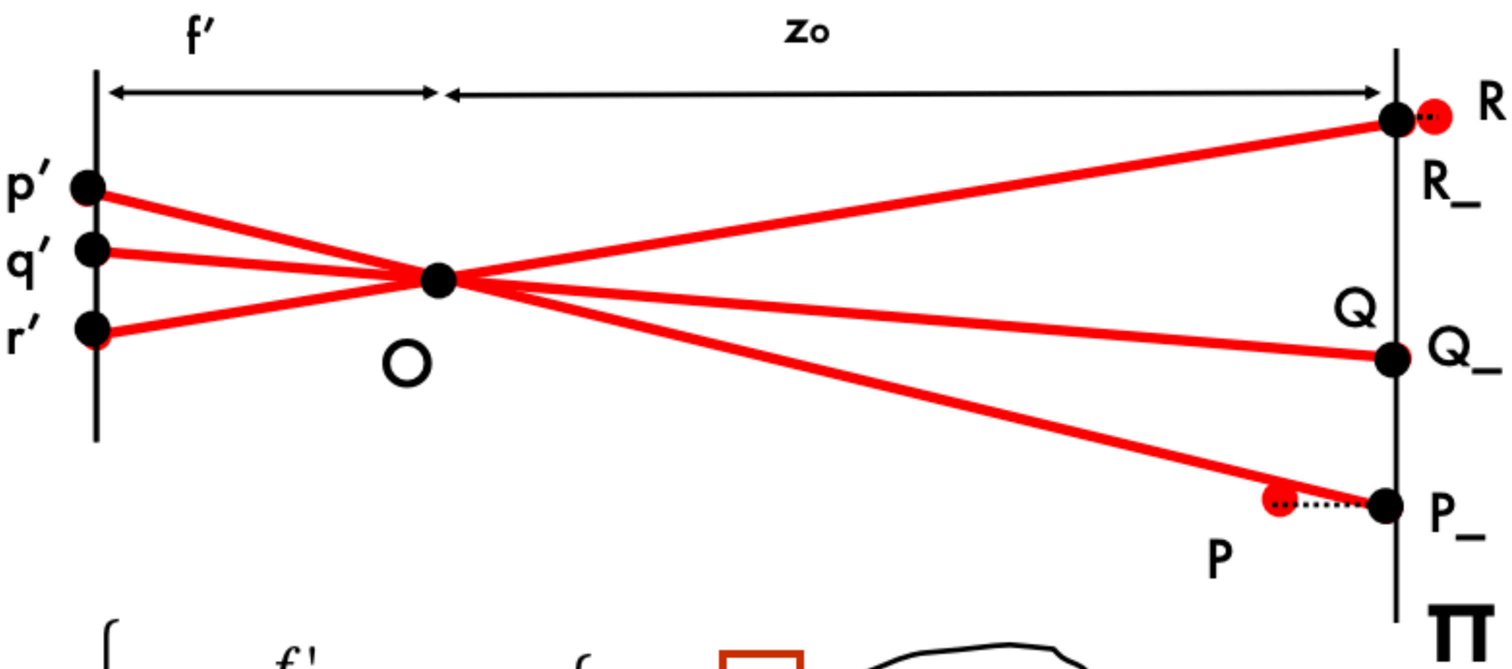


Weak Projective Camera

When the relative scene depth is small compared to its distance from the camera



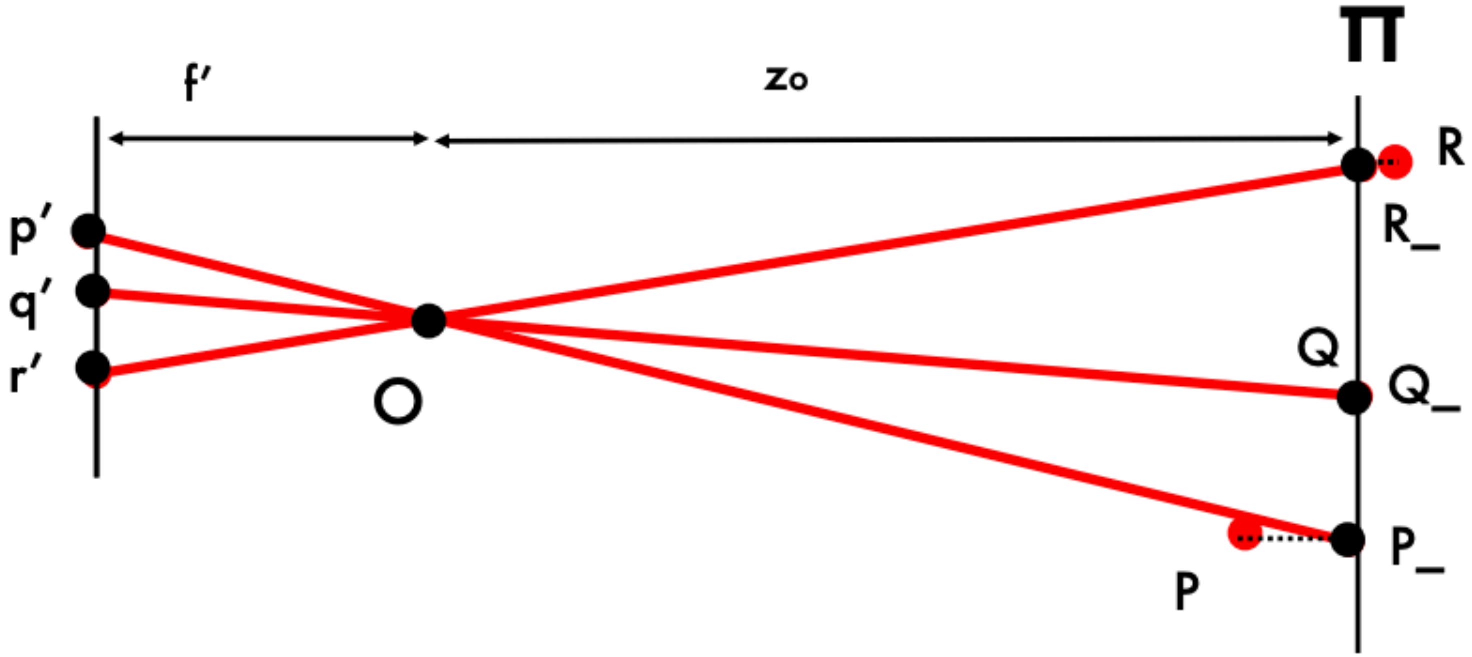
Weak Projective Camera



$$\begin{cases} x' = \dfrac{f'}{z} x \\ y' = \dfrac{f'}{z} y \end{cases} \rightarrow \begin{cases} x' = \boxed{\dfrac{f'}{z_0}} x \\ y' = \dfrac{f'}{z_0} y \end{cases}$$

Magnification m

Weak Projective Camera



Projective (perspective) Weak perspective

$$M = K\begin{bmatrix} R & T \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{v} & 1 \end{bmatrix} \rightarrow M = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ 0 & 1 \end{bmatrix}$$

Perspective vs. Weak Perspective

$$P' = M\,P_w = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{bmatrix} P_W = \begin{bmatrix} \mathbf{m}_1 P_W \\ \mathbf{m}_2 P_W \\ \mathbf{m}_3 P_W \end{bmatrix} \qquad M = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{v} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{bmatrix}$$

$$\xrightarrow{E} \left(\frac{\mathbf{m}_1 P_w}{\mathbf{m}_3 P_w}, \frac{\mathbf{m}_2 P_w}{\mathbf{m}_3 P_w}\right)$$

Perspective

$$P' = M\,P_w = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{bmatrix} P_W = \begin{bmatrix} \mathbf{m}_1 P_W \\ \mathbf{m}_2 P_W \\ 1 \end{bmatrix} \qquad M = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{bmatrix} = \begin{bmatrix} & \mathbf{m}_1 & & \\ & \mathbf{m}_2 & & \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

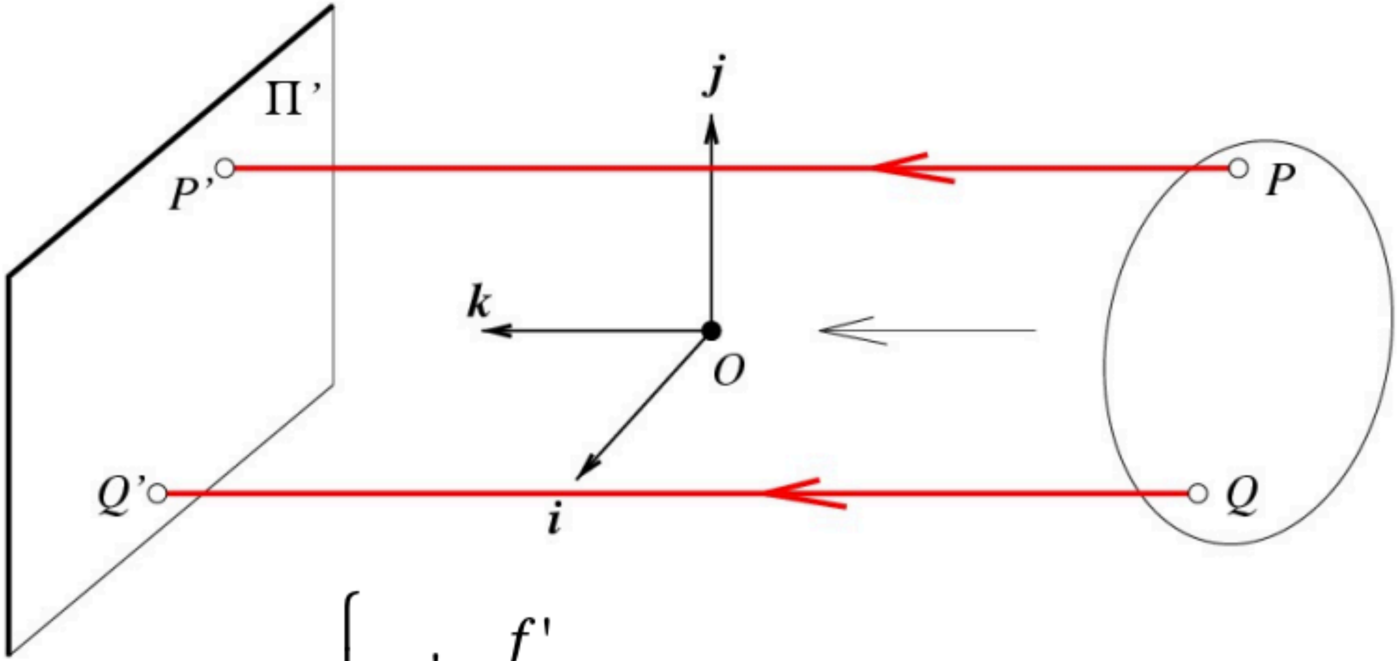
$$\xrightarrow{E} (\mathbf{m}_1 P_w, \mathbf{m}_2 P_w)$$

magnification

Weak perspective

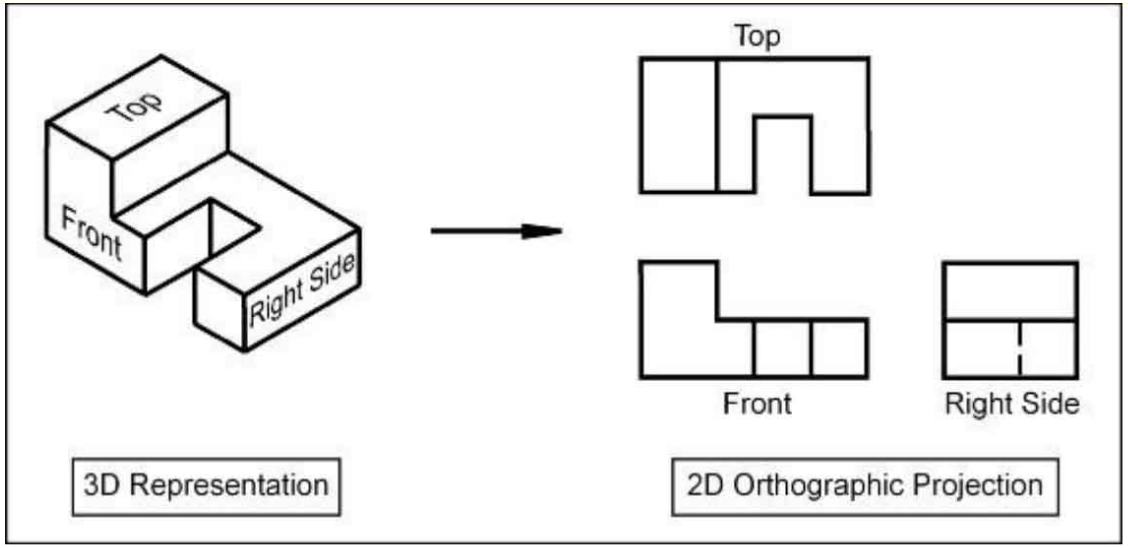
Orthographic (Affine) Projection

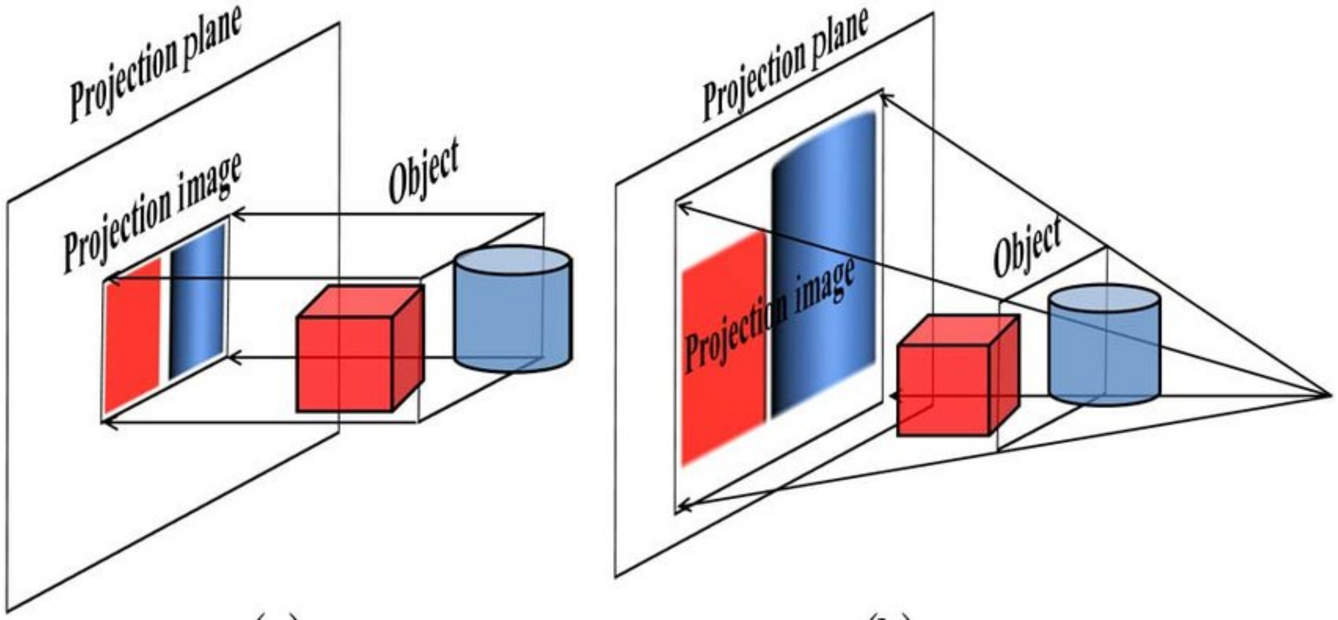
Distance from center of projection to image plane is infinite



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

Orthographic Projection vs. Perspective Projection



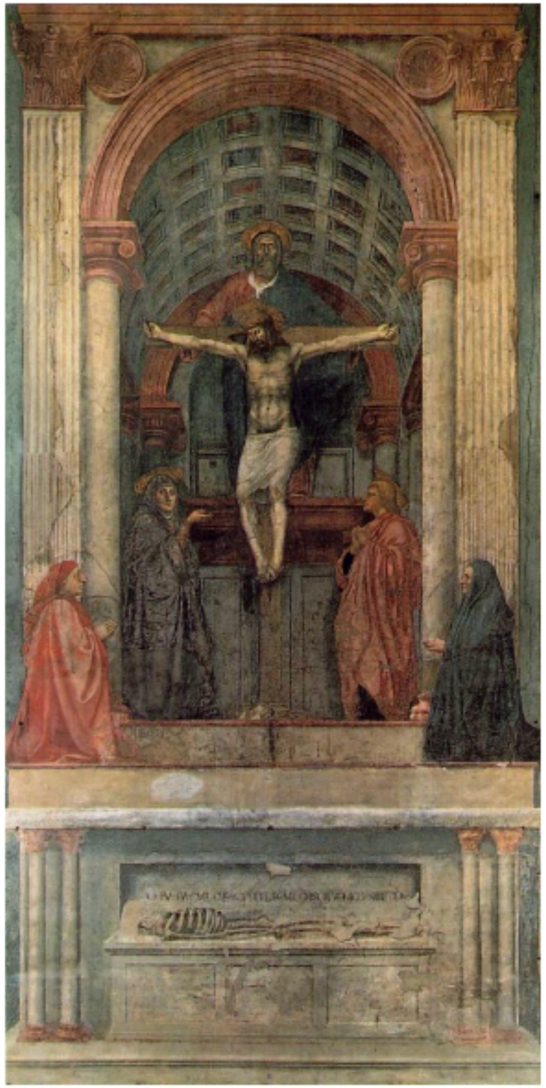


(a) (b)

Pros and Cons of the Camera Models

- Weak perspective results in much simpler math.
  - Accurate when object is small and distant.
  - Most useful for recognition.
- Pinhole perspective is much more accurate for modeling the 3D-to-2D mapping.
  - Used in structure from motion or SLAM.

One-Point Perspective



Masaccio, *Trinity*, Santa Maria Novella, Florence, 1425-28

il Canaletto The Piazzetta, Venice,

Weak Perspective Projection

The Kangxi Emperor's Southern Inspection Tour (1691-1698) by Wang Hui

Camera Calibration

Some slides are borrowed from Stanford CS231A.

Why Camera Calibration?

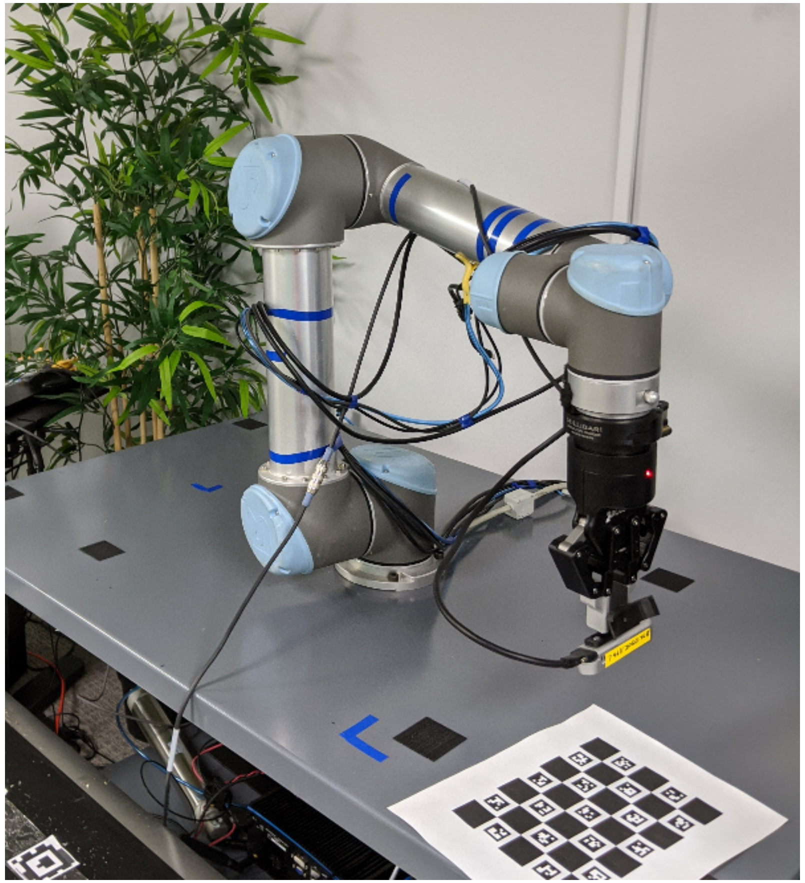
- Imagine how we picks up an object

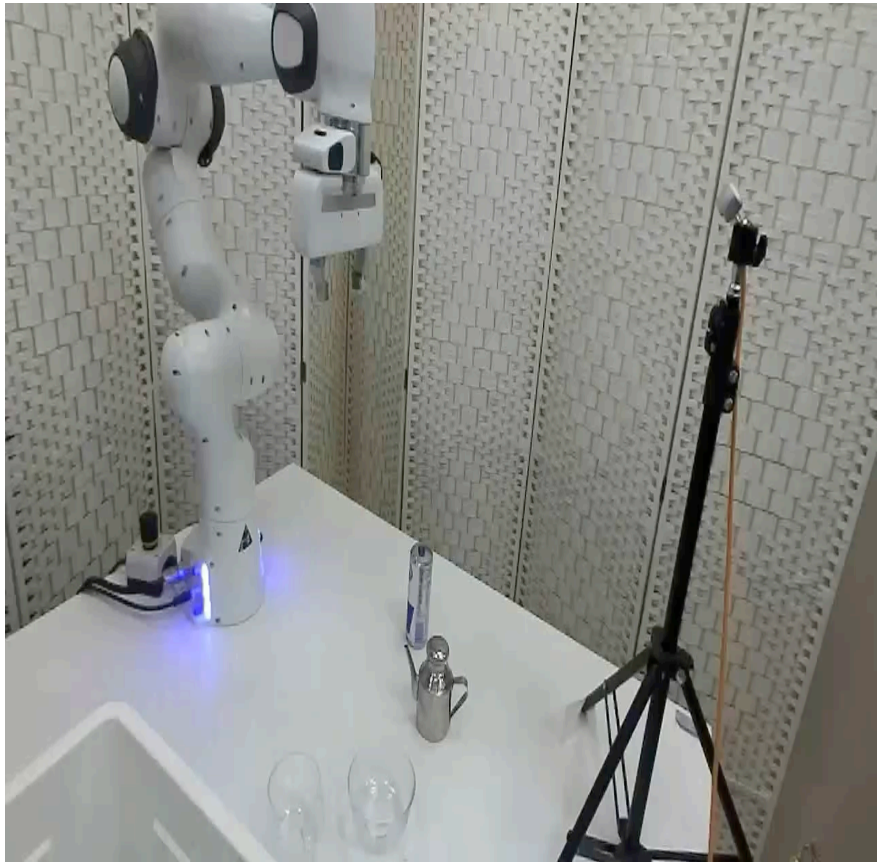
  - Our eyes capture images of the object.
  - Our brain processes these images, finds the object, and tells our arms and hands where to go and how to pick up the object .
  - To connect the space from our eye and the space of our body, we need camera calibration.

https://blog.zivid.com/importance-of-3d-hand-eye-calibration

Camera Calibration in Robot-Camera System

- Hand-eye calibration: transfer the end-effector target pose from camera space to robot space.





Projective Camera

$$P' = M\,P_w \quad = \underbrace{K}_{\text{Internal parameters}}\underbrace{\begin{bmatrix} R & T \end{bmatrix}}_{\text{External parameters}} P_w$$

$$\boxed{\mathcal{M} = \begin{pmatrix} \alpha \boldsymbol{r}_1^T - \alpha \cot\theta \boldsymbol{r}_2^T + u_0 \boldsymbol{r}_3^T & \alpha t_x - \alpha \cot\theta t_y + u_0 t_z \\ \dfrac{\beta}{\sin\theta}\boldsymbol{r}_2^T + v_0 \boldsymbol{r}_3^T & \dfrac{\beta}{\sin\theta} t_y + v_0 t_z \\ \boldsymbol{r}_3^T & t_z \end{pmatrix}_{3\times 4}}$$

$$K = \begin{bmatrix} \alpha & -\alpha\cot\theta & u_o \\ 0 & \dfrac{\beta}{\sin\theta} & v_o \\ 0 & 0 & 1 \end{bmatrix} \qquad R = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{bmatrix} \qquad T = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

Goal of Calibration

- Estimate camera intrinsics and extrinsic from one or multiple images

$$P' = M\,P_w \quad = K\begin{bmatrix} R & T \end{bmatrix} P_w$$

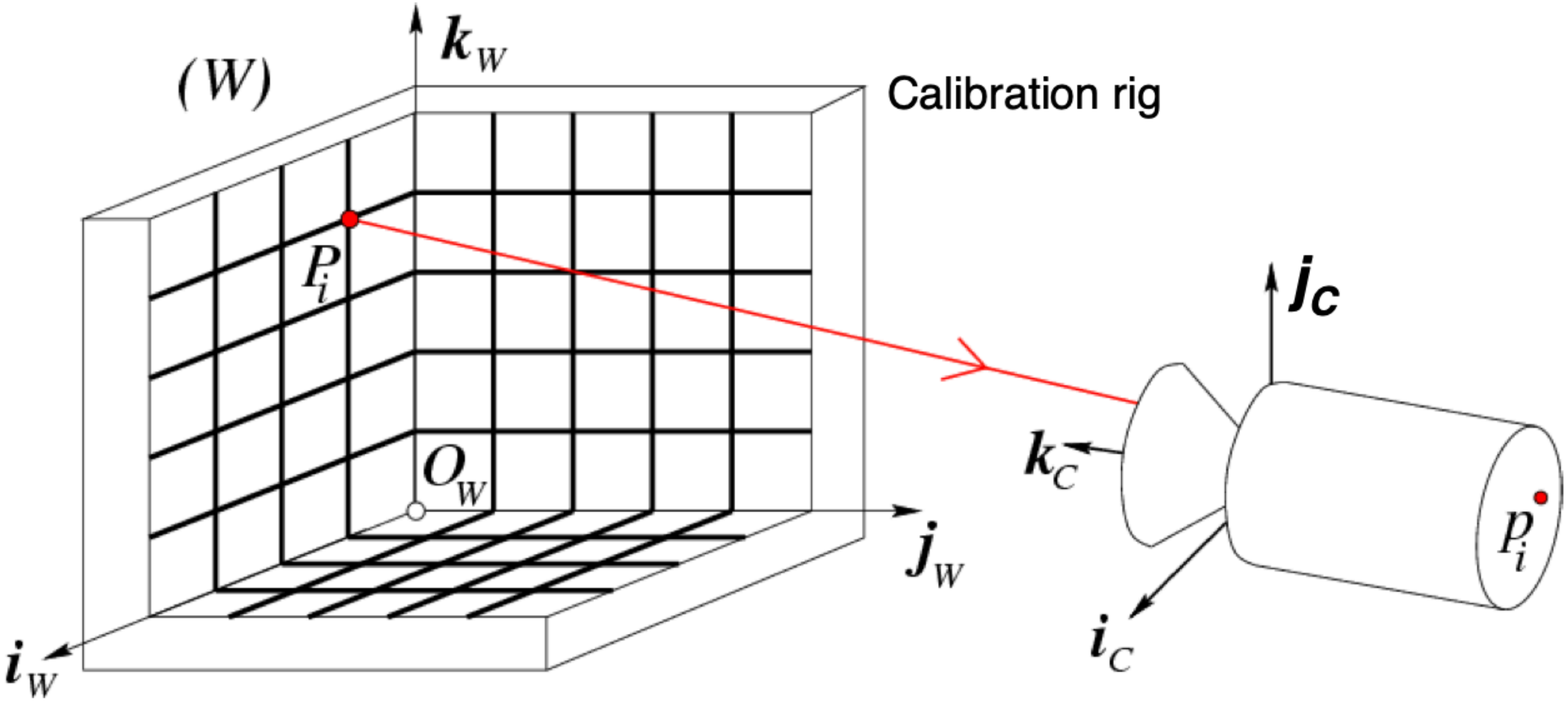
$$\mathcal{M} = \begin{pmatrix} \alpha \boldsymbol{r}_1^T - \alpha \cot\theta \boldsymbol{r}_2^T + u_0 \boldsymbol{r}_3^T & \alpha t_x - \alpha \cot\theta t_y + u_0 t_z \\ \frac{\beta}{\sin\theta} \boldsymbol{r}_2^T + v_0 \boldsymbol{r}_3^T & \frac{\beta}{\sin\theta} t_y + v_0 t_z \\ \boldsymbol{r}_3^T & t_z \end{pmatrix}_{3\times 4}$$

$$K = \begin{bmatrix} \alpha & -\alpha\cot\theta & u_o \\ 0 & \frac{\beta}{\sin\theta} & v_o \\ 0 & 0 & 1 \end{bmatrix} \qquad R = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{bmatrix} \qquad T = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

Change notation:
$P = P_w$

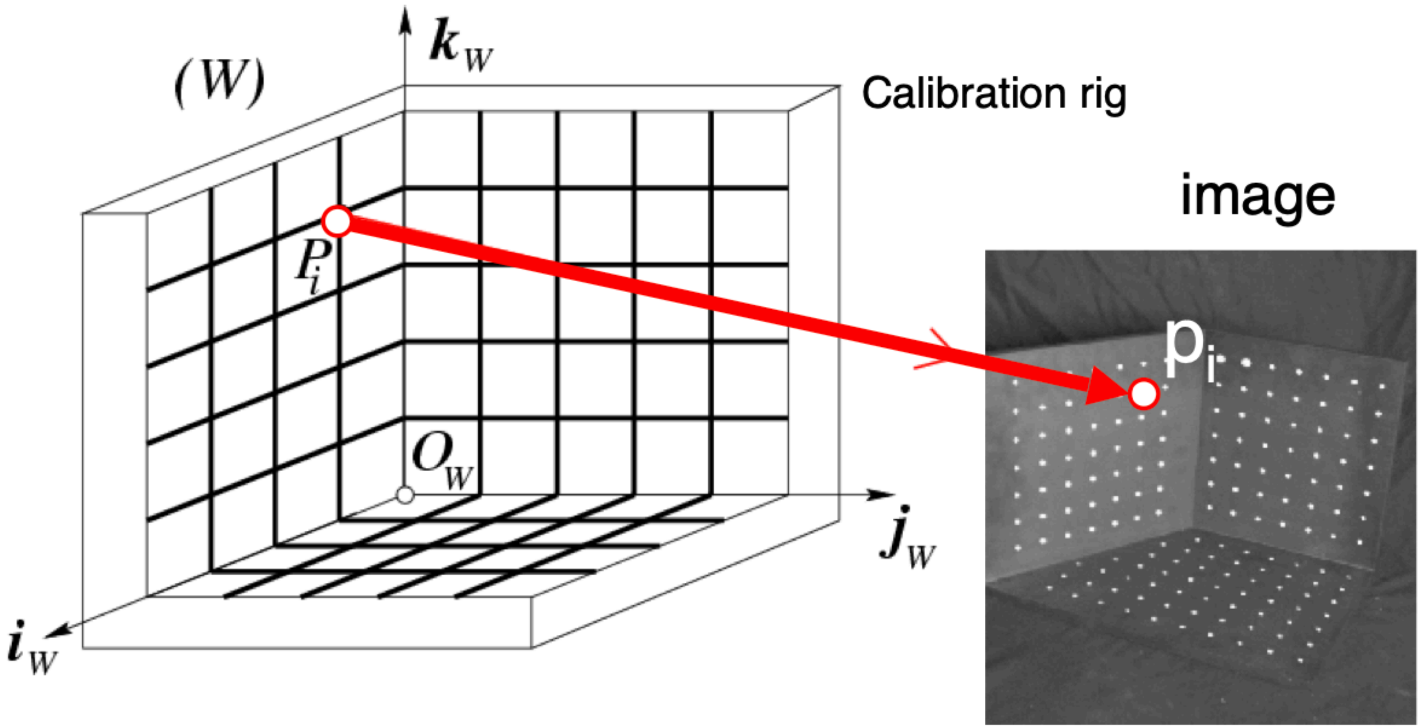
$p = P'$

Calibration Problem



- $P_1 \ldots P_n$ with known positions in $[O_w, i_w, j_w, k_w]$

Calibration Problem



- $P_1 \ldots P_n$ with **known** positions in $[O_w, i_w, j_w, k_w]$
- $p_1, \ldots p_n$ **known** positions in the image

Goal: compute intrinsic and extrinsic parameters

Assuming known correspondence between P_n and p_n

Calibration Problem

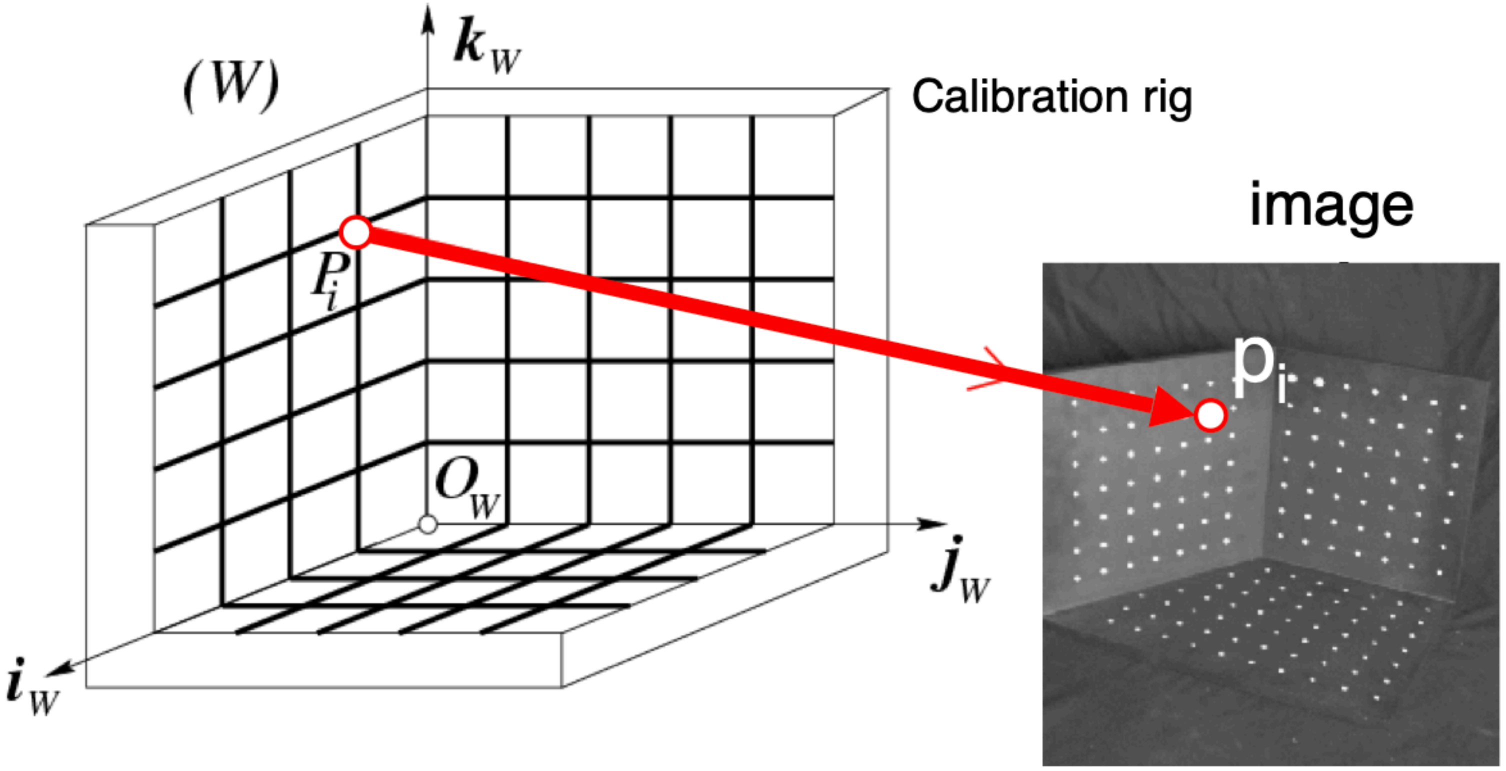
- The degree of freedom of M: 5 + 3 + 3 = 11

- We need 11 equations
- Thus, 6 correspondence would suffice

$$p = K[R\ T]P$$

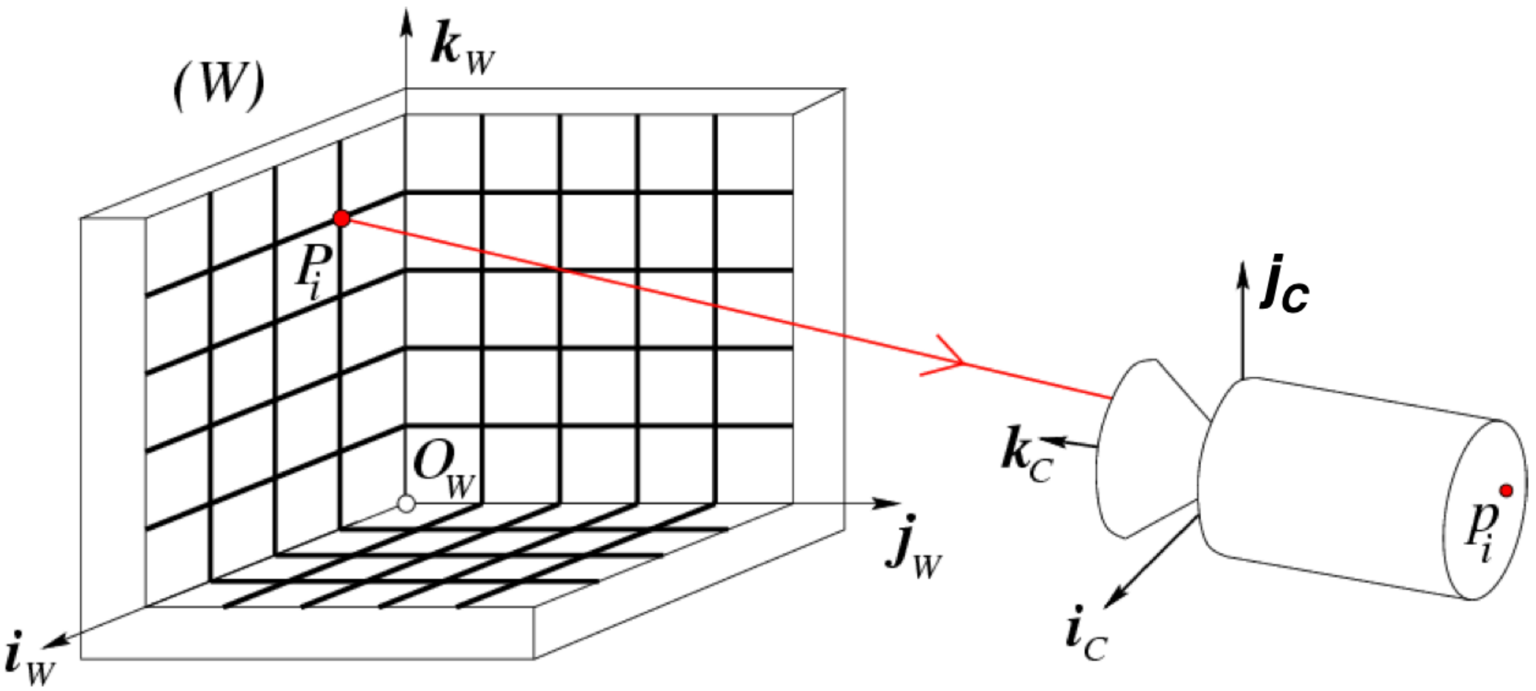
$$K = \begin{bmatrix} \alpha & -\alpha\cot\theta & u_o \\ 0 & \dfrac{\beta}{\sin\theta} & v_o \\ 0 & 0 & 1 \end{bmatrix} \qquad R = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{bmatrix} \qquad T = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

Calibration Problem



In practice, using more than 6 correspondences enables more robust results

Calibration Problem



$$P_i \rightarrow M\,P_i \rightarrow p_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} \dfrac{\mathbf{m}_1\,P_i}{\mathbf{m}_3\,P_i} \\ \dfrac{\mathbf{m}_2\,P_i}{\mathbf{m}_3\,P_i} \end{bmatrix} \qquad M = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{bmatrix}$$

in pixels

[Eq. 1]

Calibration Problem

[Eq. 1] $$\begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} \dfrac{\mathbf{m}_1\ P_i}{\mathbf{m}_3\ P_i} \\ \dfrac{\mathbf{m}_2\ P_i}{\mathbf{m}_3\ P_i} \end{bmatrix}$$

$$u_i = \frac{\mathbf{m}_1\ P_i}{\mathbf{m}_3\ P_i} \rightarrow u_i(\mathbf{m}_3\ P_i) = \mathbf{m}_1\ P_i \rightarrow u_i(\mathbf{m}_3\ P_i) - \mathbf{m}_1\ P_i = 0$$

$$v_i = \frac{\mathbf{m}_2\ P_i}{\mathbf{m}_3\ P_i} \rightarrow v_i(\mathbf{m}_3\ P_i) = \mathbf{m}_2\ P_i \rightarrow v_i(\mathbf{m}_3\ P_i) - \mathbf{m}_2\ P_i = 0$$

[Eqs. 2]

Calibration Problem

$$\begin{cases} u_1(\mathbf{m}_3 P_1) - \mathbf{m}_1 P_1 = 0 \\ v_1(\mathbf{m}_3 P_1) - \mathbf{m}_2 P_1 = 0 \\ \vdots \\ u_i(\mathbf{m}_3 P_i) - \mathbf{m}_1 P_i = 0 \\ v_i(\mathbf{m}_3 P_i) - \mathbf{m}_2 P_i = 0 \\ \vdots \\ u_n(\mathbf{m}_3 P_n) - \mathbf{m}_1 P_n = 0 \\ v_n(\mathbf{m}_3 P_n) - \mathbf{m}_2 P_n = 0 \end{cases}$$

[Eqs. 3]

Calibration Problem

$$\begin{cases} -u_1(\mathbf{m}_3\ P_1) + \mathbf{m}_1\ P_1 = 0 \\ -v_1(\mathbf{m}_3\ P_1) + \mathbf{m}_2\ P_1 = 0 \\ \vdots \\ -u_n(\mathbf{m}_3\ P_n) + \mathbf{m}_1\ P_n = 0 \\ -v_n(\mathbf{m}_3\ P_n) + \mathbf{m}_2\ P_n = 0 \end{cases} \longrightarrow \boxed{\mathbf{P}\,\mathbf{m} = 0}$$

known

unknown

[Eq. 4]

Homogenous linear system

$$\mathbf{P} \stackrel{\text{def}}{=} \begin{pmatrix} \mathrm{P}_1^T & \mathbf{0}^T & -u_1\mathrm{P}_1^T \\ \mathbf{0}^T & \mathrm{P}_1^T & -v_1\mathrm{P}_1^T \\ & \vdots & \\ \mathrm{P}_n^T & \mathbf{0}^T & -u_n\mathrm{P}_n^T \\ \mathbf{0}^T & \mathrm{P}_n^T & -v_n\mathrm{P}_n^T \end{pmatrix}_{2n \times 12}$$

1x4

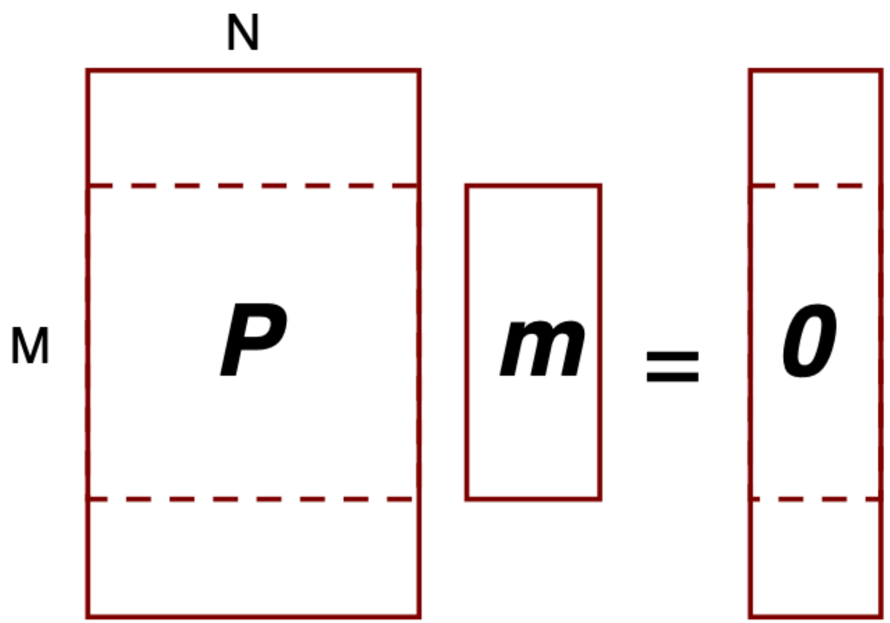
$$\boldsymbol{m} \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{m}_1^{\mathrm{T}} \\ \mathbf{m}_2^{\mathrm{T}} \\ \mathbf{m}_3^{\mathrm{T}} \end{pmatrix}_{12\times 1}$$

4x1

Calibration Problem

Homogeneous M x N Linear Systems

M=number of equations = 2n
N=number of unknown = 11



Rectangular system (M>N)

- 0 is always a solution

Calibration Problem

- How do we solve this homogenous linear system?

$$Pm = 0$$

Calibration Problem

- How do we solve this homogenous linear system?

$$Pm = 0$$

- Add a constraint to **m** to avoid trivial solution: $|m|^2 = 1$
- Then we can solve the following minimization problem using SVD:

Minimize $|\mathbf{P}\ \mathbf{m}|^2$

under the constraint $|\mathbf{m}|^2 = 1$

Calibration Problem

$$\boxed{\mathbf{P}}\ \mathbf{m} = 0$$

$$\downarrow \text{SVD decomposition of P}$$

$$\boxed{U_{2n\times 12}\ D_{12\times 12}\ V^{T}_{12\times 12}}$$

Last column of V gives $\boldsymbol{m}$

Why? See pag 592 of HZ*

$$\boldsymbol{m} \xrightarrow{\text{Convert } 1\times 12 \text{ into } 3\times 4} \hat{M}$$

$$\boldsymbol{m} \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{m}_1^T \\ \mathbf{m}_2^T \\ \mathbf{m}_3^T \end{pmatrix}$$

*: R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, 2003.

Calibration Problem

- Since $||\hat{M}||_F = 1$, we need to find a scale ρ to unnormalize it:

$$M = \rho\hat{M} = \left(\begin{array}{c|c} \alpha \boldsymbol{r}_1^T - \alpha\cot\theta \boldsymbol{r}_2^T + c_x \boldsymbol{r}_3^T & \alpha t_x - \alpha\cot\theta t_y + c_x t_z \\ \frac{\beta}{\sin\theta}\boldsymbol{r}_2^T + c_y \boldsymbol{r}_3^T & \frac{\beta}{\sin\theta} t_y + c_y t_z \\ \boldsymbol{r}_3^T & t_z \end{array}\right) = K[R \quad T]$$

$$\underbrace{\qquad}_{A} \qquad \underbrace{\qquad}_{\mathbf{b}}$$

$$K = \begin{bmatrix} \alpha & -\alpha\cot\theta & c_x \\ 0 & \frac{\beta}{\sin\theta} & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

Note that we can also represent $\hat{M} = \left[\hat{A}_{3\times3}\ \hat{b}_{1\times3}\right]$, which satisfies $A = \rho\hat{A}, b = \rho\hat{b}$.

Calibration Problem

$$M = \rho\hat{M} = \begin{pmatrix} \alpha \boldsymbol{r}_1^T - \alpha\cot\theta \boldsymbol{r}_2^T + c_x \boldsymbol{r}_3^T & \alpha t_x - \alpha\cot\theta t_y + c_x t_z \\ \dfrac{\beta}{\sin\theta}\boldsymbol{r}_2^T + c_y \boldsymbol{r}_3^T & \dfrac{\beta}{\sin\theta} t_y + c_y t_z \\ \boldsymbol{r}_3^T & t_z \end{pmatrix} = K\begin{bmatrix} R & T \end{bmatrix}$$

$$\underbrace{}_{A} \qquad \underbrace{}_{\mathbf{b}}$$

$$K = \begin{bmatrix} \alpha & -\alpha\cot\theta & c_x \\ 0 & \frac{\beta}{\sin\theta} & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

Box 1

$$\hat{A} = \begin{bmatrix} \hat{\mathbf{a}}_1^T \\ \hat{\mathbf{a}}_2^T \\ \hat{\mathbf{a}}_3^T \end{bmatrix} \qquad \hat{b} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{bmatrix}$$

Estimated values from $\hat{M}$

Intrinsic

$$\rho = \frac{\pm 1}{|\hat{\mathbf{a}}_3|} \qquad c_x = \rho^2(\hat{\mathbf{a}}_1 \cdot \hat{\mathbf{a}}_3)$$

$$c_y = \rho^2(\hat{\mathbf{a}}_2 \cdot \hat{\mathbf{a}}_3)$$

$$\cos\theta = \frac{(\hat{a}_2 \times \hat{a}_3) \cdot (\hat{a}_3 \times \hat{a}_1)}{|\hat{a}_2 \times \hat{a}_3| \cdot |\hat{a}_3 \times \hat{a}_1|}$$

Theorem (Faugeras, 1993)

Let $\mathcal{M} = (\mathcal{A} \quad \boldsymbol{b})$ be a 3×4 matrix and let $\boldsymbol{a}_i^T$ $(i = 1, 2, 3)$ denote the rows of the matrix $\mathcal{A}$ formed by the three leftmost columns of $\mathcal{M}$.

- A necessary and sufficient condition for $\mathcal{M}$ to be a perspective projection matrix is that $\text{Det}(\mathcal{A}) \neq 0$.

- A necessary and sufficient condition for $\mathcal{M}$ to be a zero-skew perspective projection matrix is that $\text{Det}(\mathcal{A}) \neq 0$ and

$$(\boldsymbol{a}_1 \times \boldsymbol{a}_3) \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3) = 0.$$

- A necessary and sufficient condition for $\mathcal{M}$ to be a perspective projection matrix with zero skew and unit aspect-ratio is that $\text{Det}(\mathcal{A}) \neq 0$ and

$$\begin{cases} (\boldsymbol{a}_1 \times \boldsymbol{a}_3) \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3) = 0, \\ (\boldsymbol{a}_1 \times \boldsymbol{a}_3) \cdot (\boldsymbol{a}_1 \times \boldsymbol{a}_3) = (\boldsymbol{a}_2 \times \boldsymbol{a}_3) \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3). \end{cases}$$

Calibration Problem

$$M = \rho\hat{M} = \left(\underbrace{\begin{matrix}\alpha \boldsymbol{r}_1^T - \alpha\cot\theta \boldsymbol{r}_2^T + c_x \boldsymbol{r}_3^T \\ \dfrac{\beta}{\sin\theta}\boldsymbol{r}_2^T + c_y \boldsymbol{r}_3^T \\ \boldsymbol{r}_3^T\end{matrix}}_{A} \quad \underbrace{\begin{matrix}\alpha t_x - \alpha\cot\theta t_y + c_x t_z \\ \dfrac{\beta}{\sin\theta} t_y + c_y t_z \\ t_z\end{matrix}}_{\mathbf{b}}\right) = K\begin{bmatrix}R & T\end{bmatrix}$$

$$K = \begin{bmatrix}\alpha & -\alpha\cot\theta & c_x \\ 0 & \frac{\beta}{\sin\theta} & c_y \\ 0 & 0 & 1\end{bmatrix}$$

Box 1

$$\hat{A} = \begin{bmatrix}\hat{\mathbf{a}}_1^T \\ \hat{\mathbf{a}}_2^T \\ \hat{\mathbf{a}}_3^T\end{bmatrix} \qquad \hat{b} = \begin{bmatrix}\hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3\end{bmatrix}$$

Estimated values from $\hat{M}$

Intrinsic

$$\alpha = \rho^2 \left|\hat{\mathbf{a}}_1 \times \hat{\mathbf{a}}_3\right| \sin\theta$$

$$\beta = \rho^2 \left|\hat{\mathbf{a}}_2 \times \hat{\mathbf{a}}_3\right| \sin\theta$$

Calibration Problem

$$M = \rho \hat{M} = \begin{pmatrix} \alpha \boldsymbol{r}_1^T - \alpha \cot\theta \boldsymbol{r}_2^T + c_x \boldsymbol{r}_3^T & \alpha t_x - \alpha \cot\theta t_y + c_x t_z \\ \dfrac{\beta}{\sin\theta} \boldsymbol{r}_2^T + c_y \boldsymbol{r}_3^T & \dfrac{\beta}{\sin\theta} t_y + c_y t_z \\ \boldsymbol{r}_3^T & t_z \end{pmatrix} = \mathrm{K}[\mathrm{R} \quad \mathrm{T}]$$

$$\underbrace{\qquad}_{A} \qquad \underbrace{\qquad}_{\mathbf{b}}$$

$$\mathrm{K} = \begin{bmatrix} \alpha & -\alpha \cot\theta & c_x \\ 0 & \dfrac{\beta}{\sin\theta} & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

Box 1

$$\hat{\mathrm{A}} = \begin{bmatrix} \hat{\mathbf{a}}_1^{\mathrm{T}} \\ \hat{\mathbf{a}}_2^{\mathrm{T}} \\ \hat{\mathbf{a}}_3^{\mathrm{T}} \end{bmatrix} \qquad \hat{\mathrm{b}} = \begin{bmatrix} \hat{\mathrm{b}}_1 \\ \hat{\mathrm{b}}_2 \\ \hat{\mathrm{b}}_3 \end{bmatrix}$$

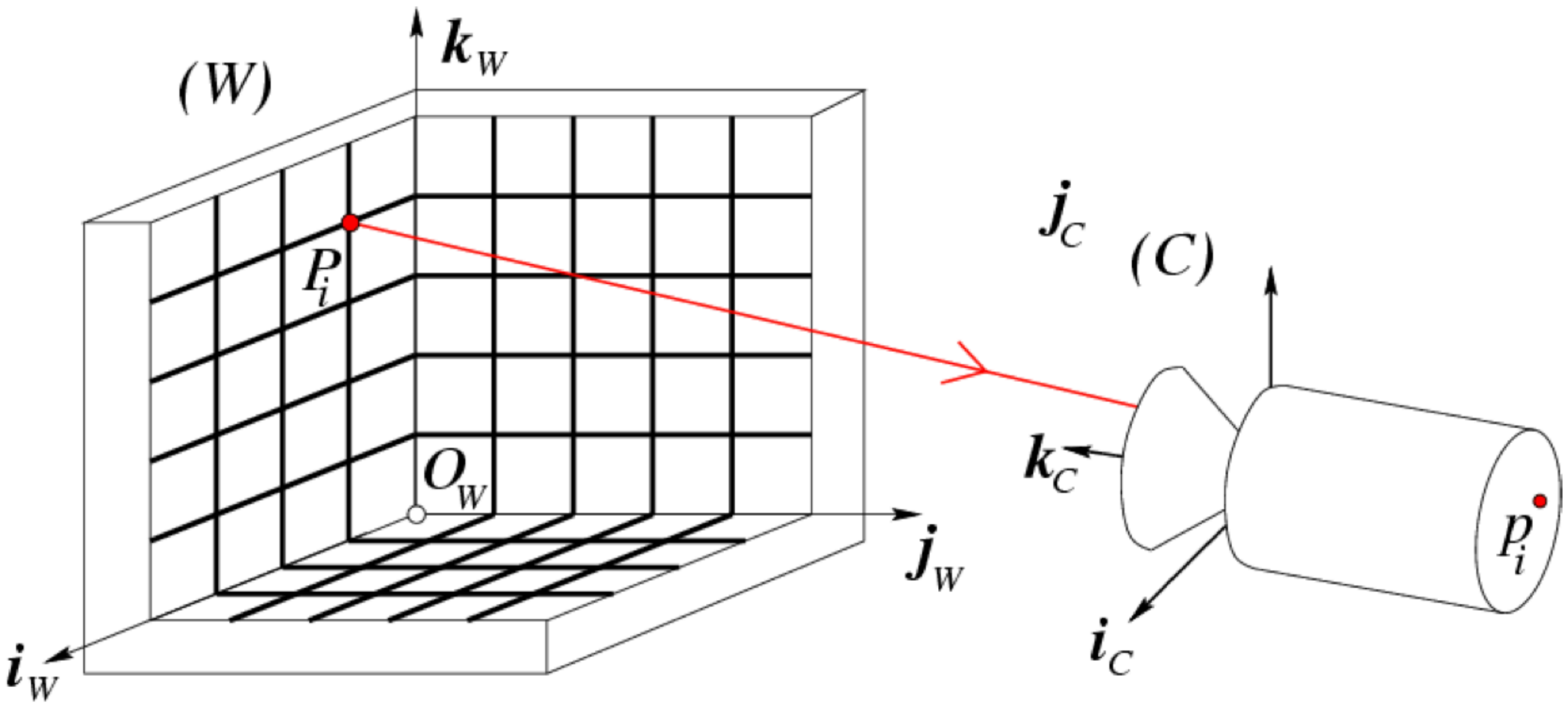
Estimated values from $\hat{M}$

Extrinsic

$$\mathbf{r}_1 = \frac{(\hat{\mathbf{a}}_2 \times \hat{\mathbf{a}}_3)}{|\hat{\mathbf{a}}_2 \times \hat{\mathbf{a}}_3|} \qquad \mathbf{r}_3 = \frac{\pm \hat{\mathbf{a}}_3}{|\hat{\mathbf{a}}_3|}$$

$$\mathbf{r}_2 = \mathbf{r}_3 \times \mathbf{r}_1 \qquad \mathrm{T} = \rho \, \mathrm{K}^{-1} \hat{\mathbf{b}}$$

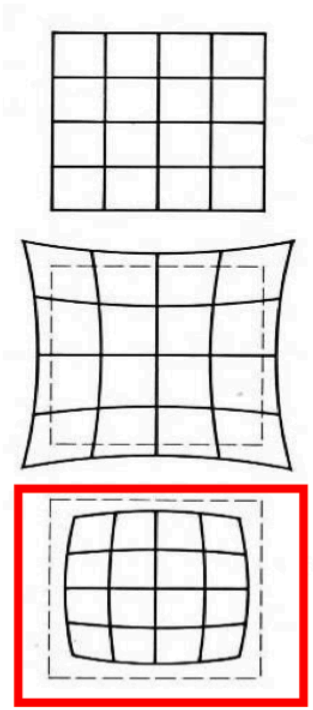
Calibration Problem: Degeneration Case



- P_i's cannot lie on the same plane!
  - Points cannot lie on the intersection curve of two quadric surfaces

Camera Calibration with Radial Distortion

- Image magnification (in)decreases with distance from the optical axis
- Caused by imperfect lenses
- Deviations are most noticeable for rays that pass through the edge of the lens



No distortion

Pin cushion

Barrel

General Camera Calibration

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} \dfrac{\mathbf{q}_1 \mathrm{P}_\mathrm{i}}{\mathbf{q}_3 \mathrm{P}_\mathrm{i}} \\ \dfrac{\mathbf{q}_2 \mathrm{P}_\mathrm{i}}{\mathbf{q}_3 \mathrm{P}_\mathrm{i}} \end{bmatrix} \longrightarrow X = f(Q) \quad \text{[Eq .8]}$$

i=1…n

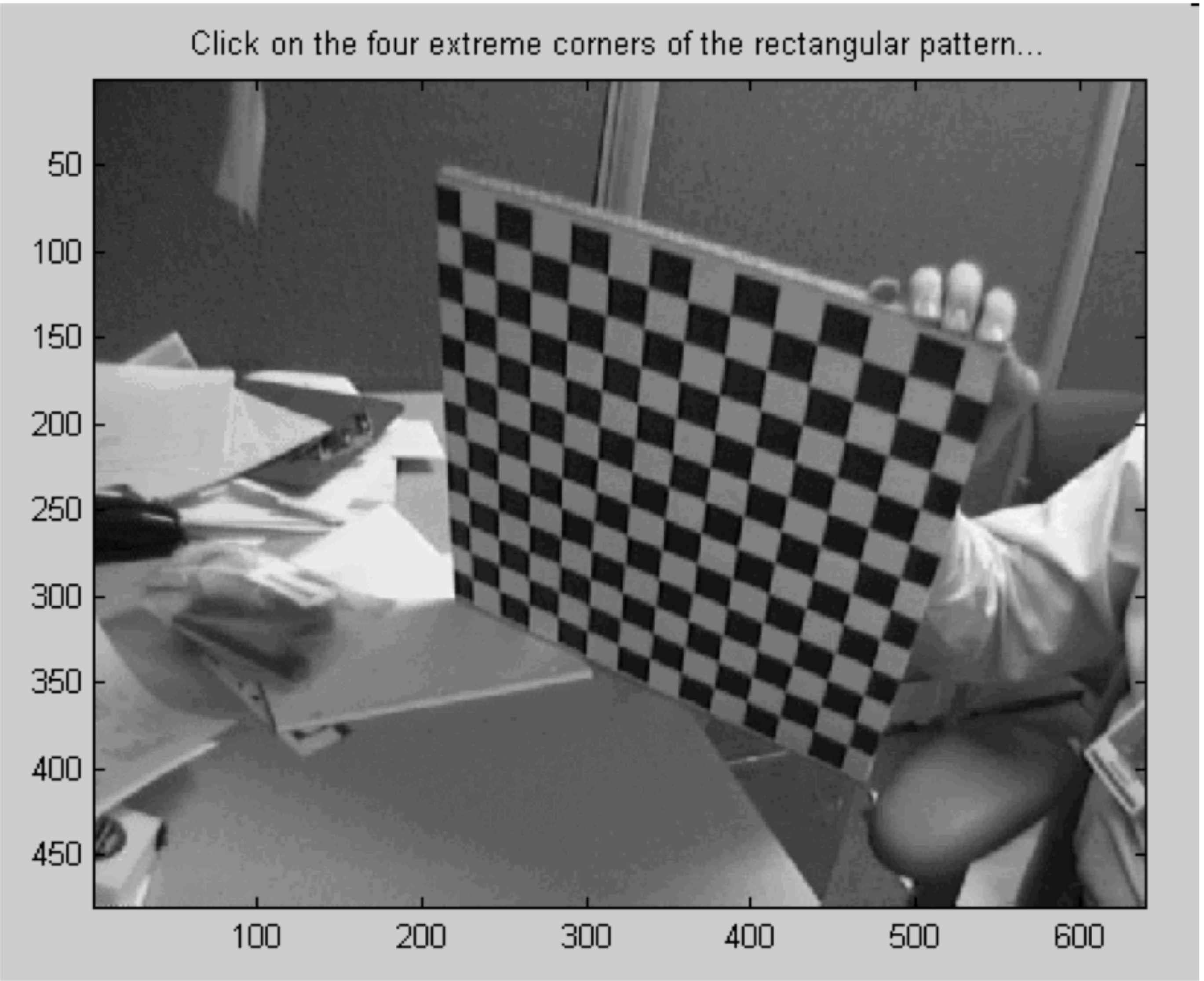
measurements

parameters

f( ) is the nonlinear mapping

- Reference:
  - Chapter 1 in D. A. Forsyth and J. Ponce. Computer Vision: A Modern Approach (2nd Edition). Prentice Hall, 2011.
  - Chapter 7 in R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, 2003.

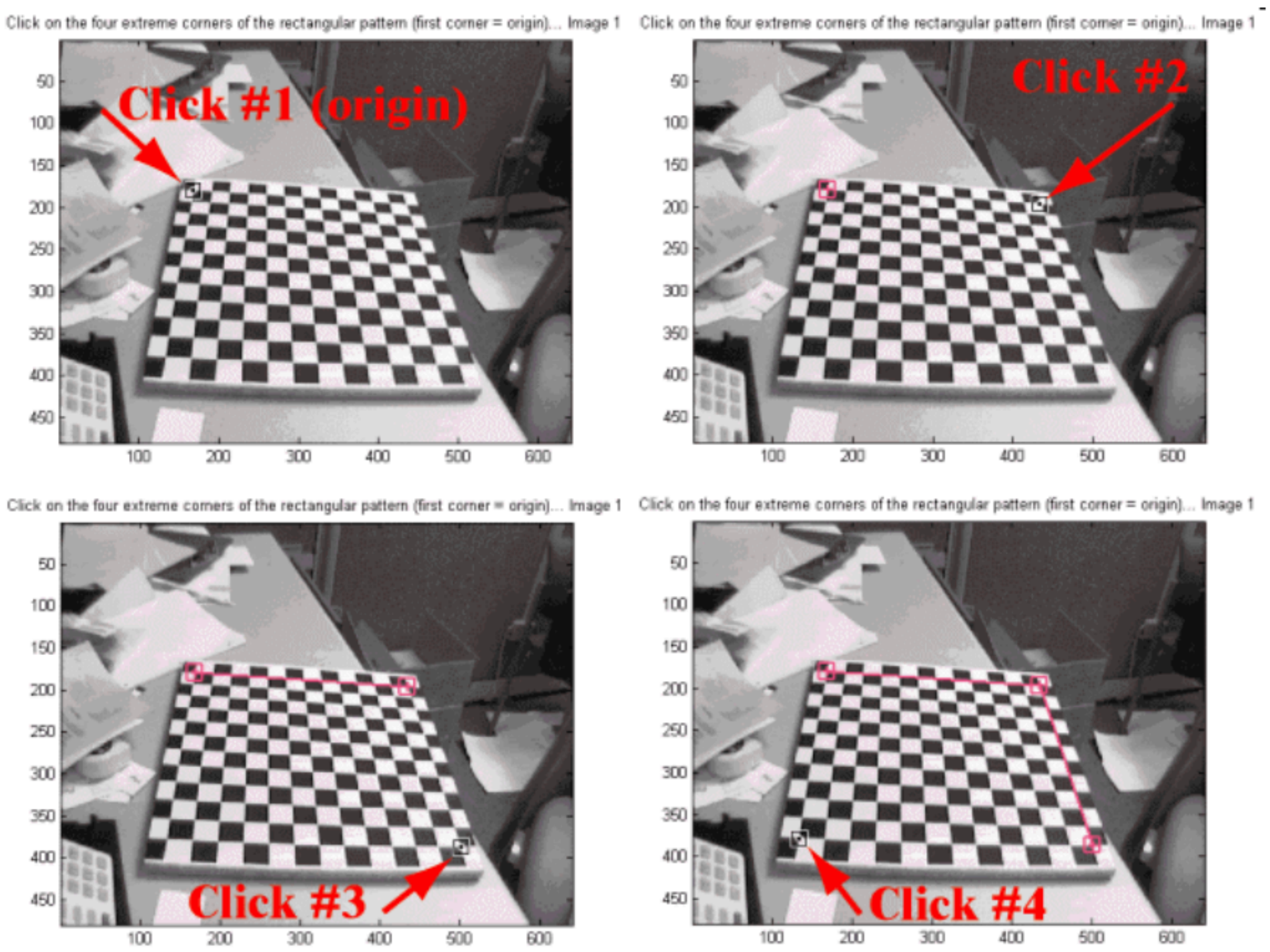
Calibration Procedure



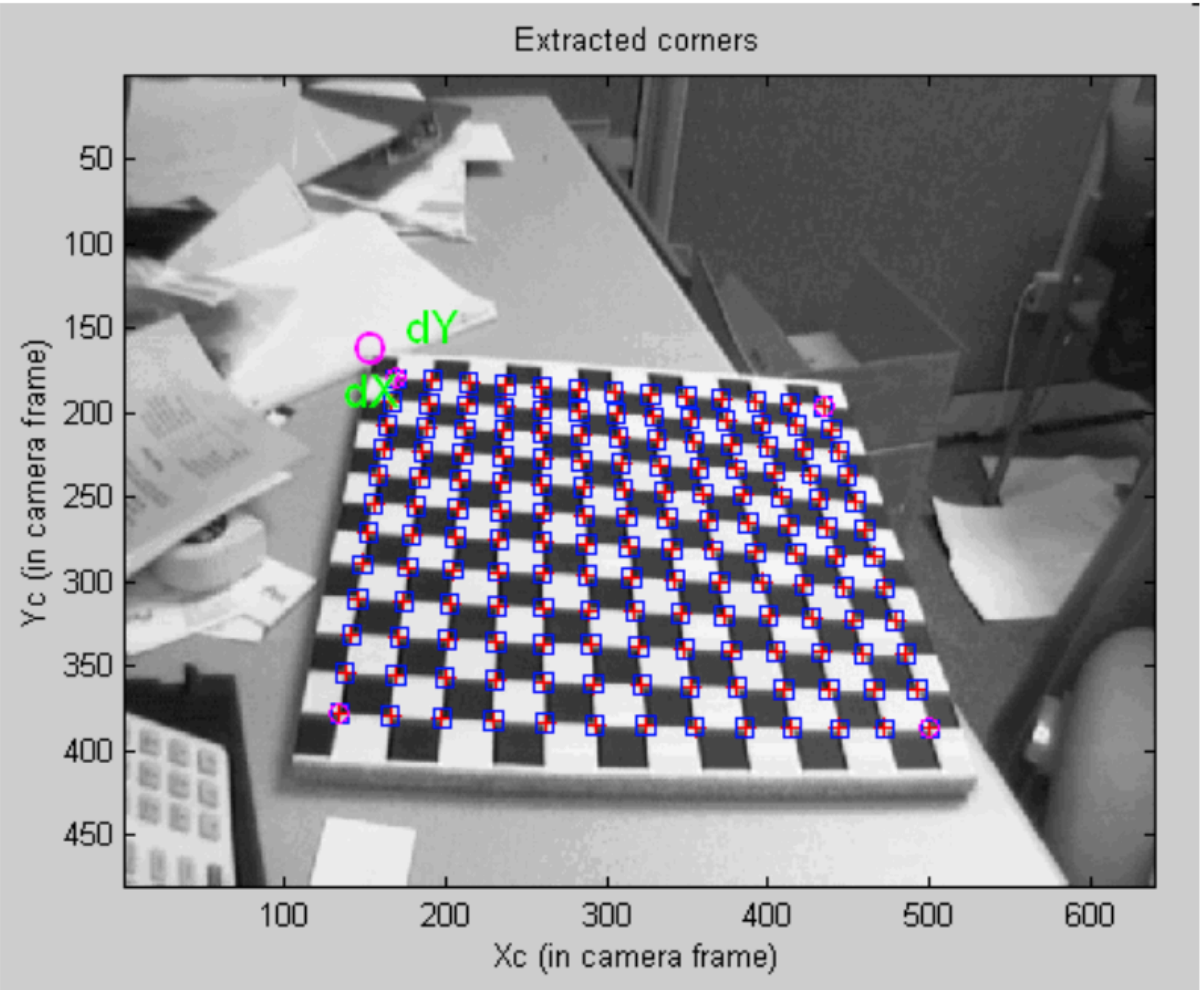
Camera Calibration Toolbox for Matlab
J. Bouguet – [1998-2000]

http://robots.stanford.edu/cs223b04/JeanYvesCalib/htmls/example.html

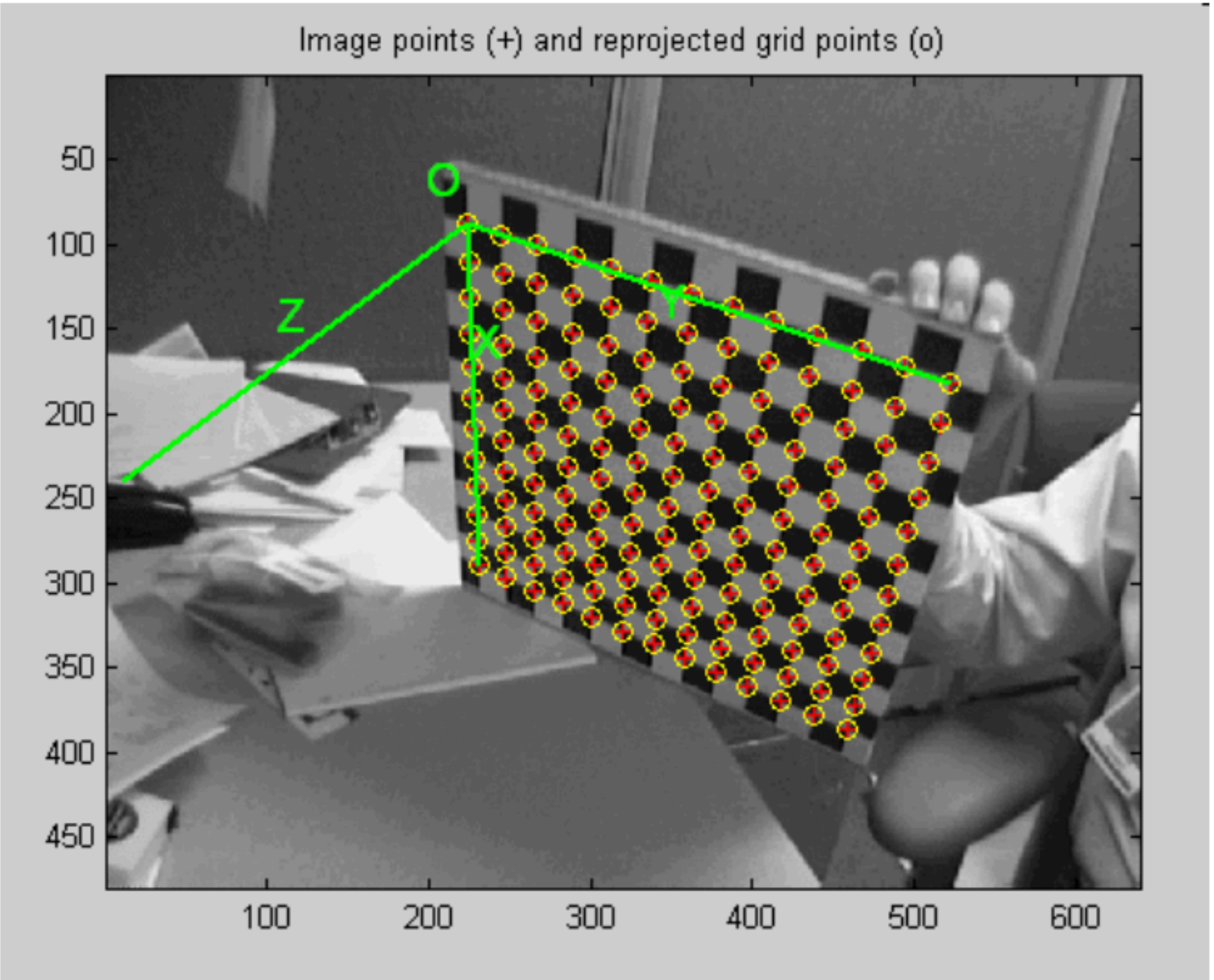
Calibration Procedure



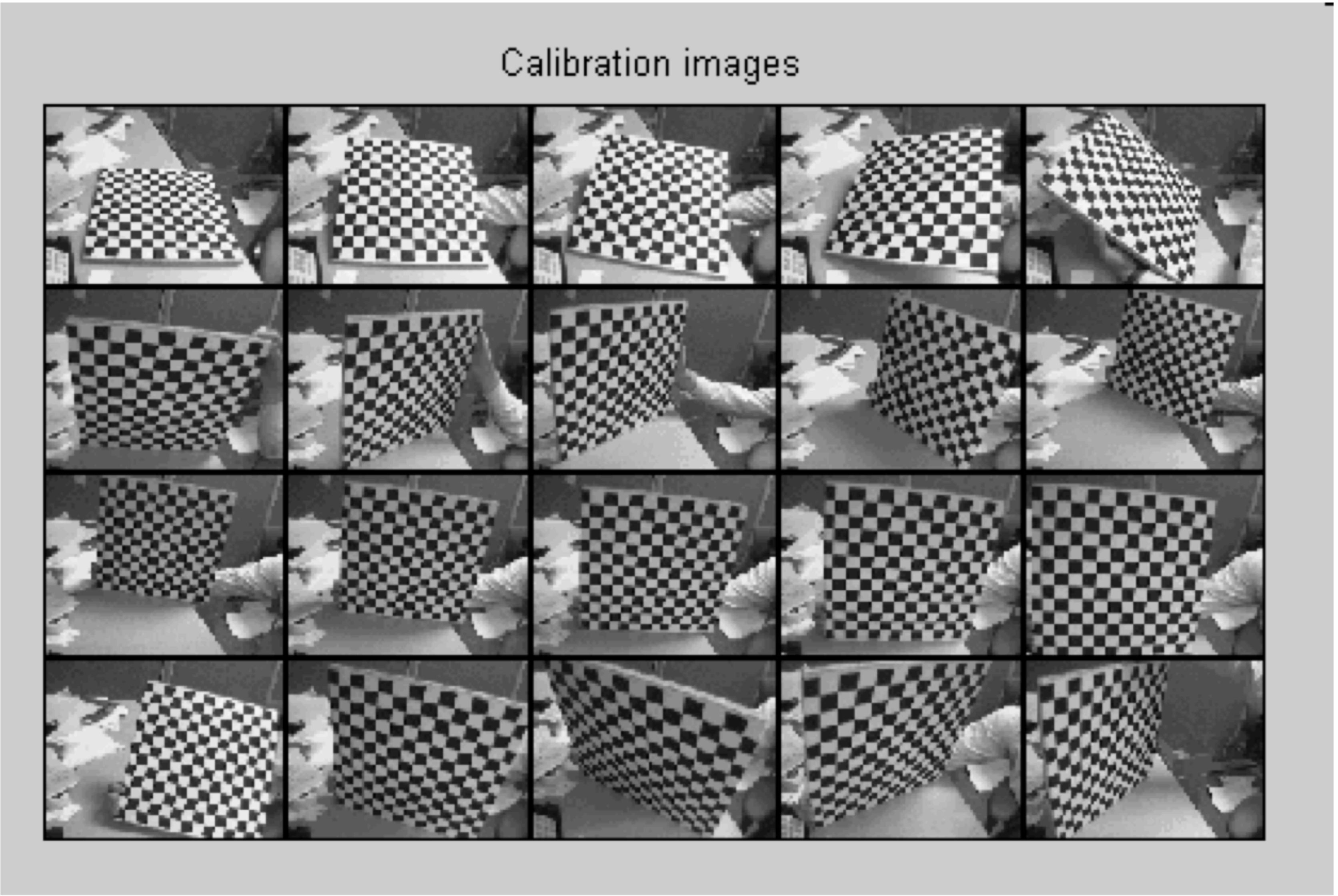
Calibration Procedure



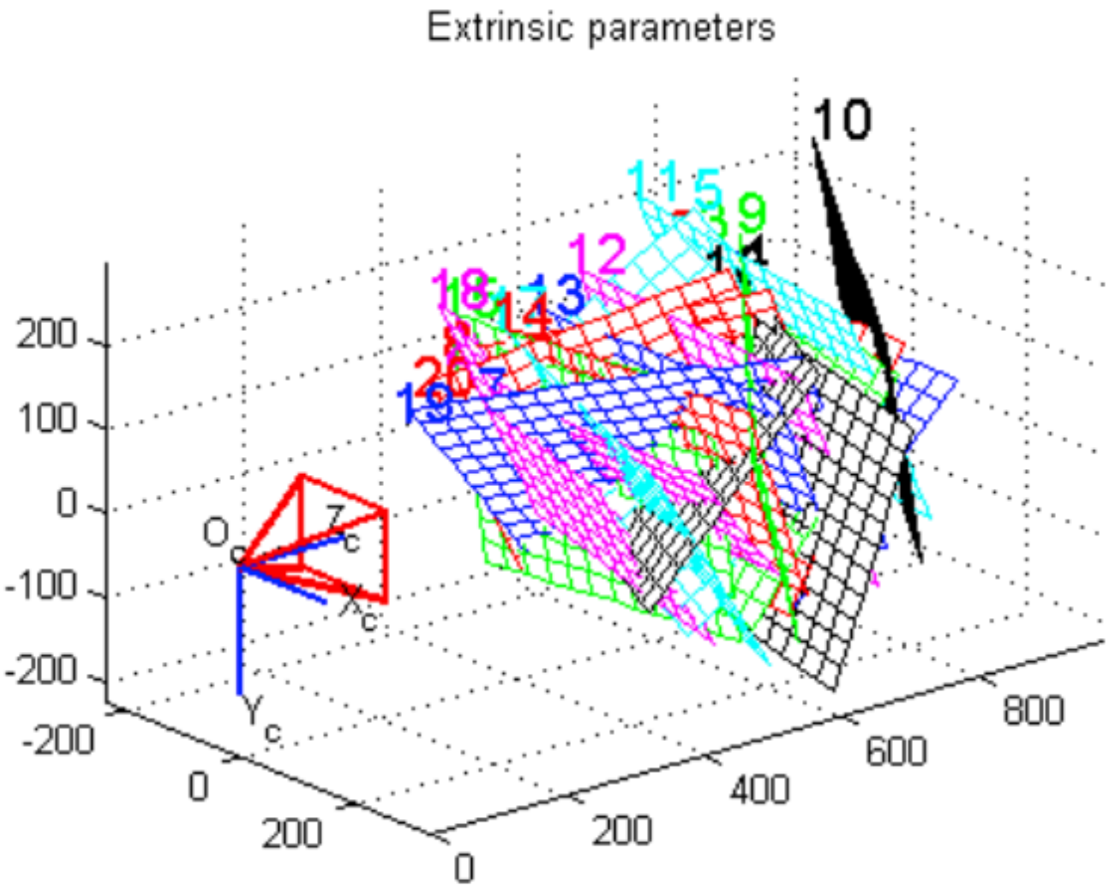
Calibration Procedure



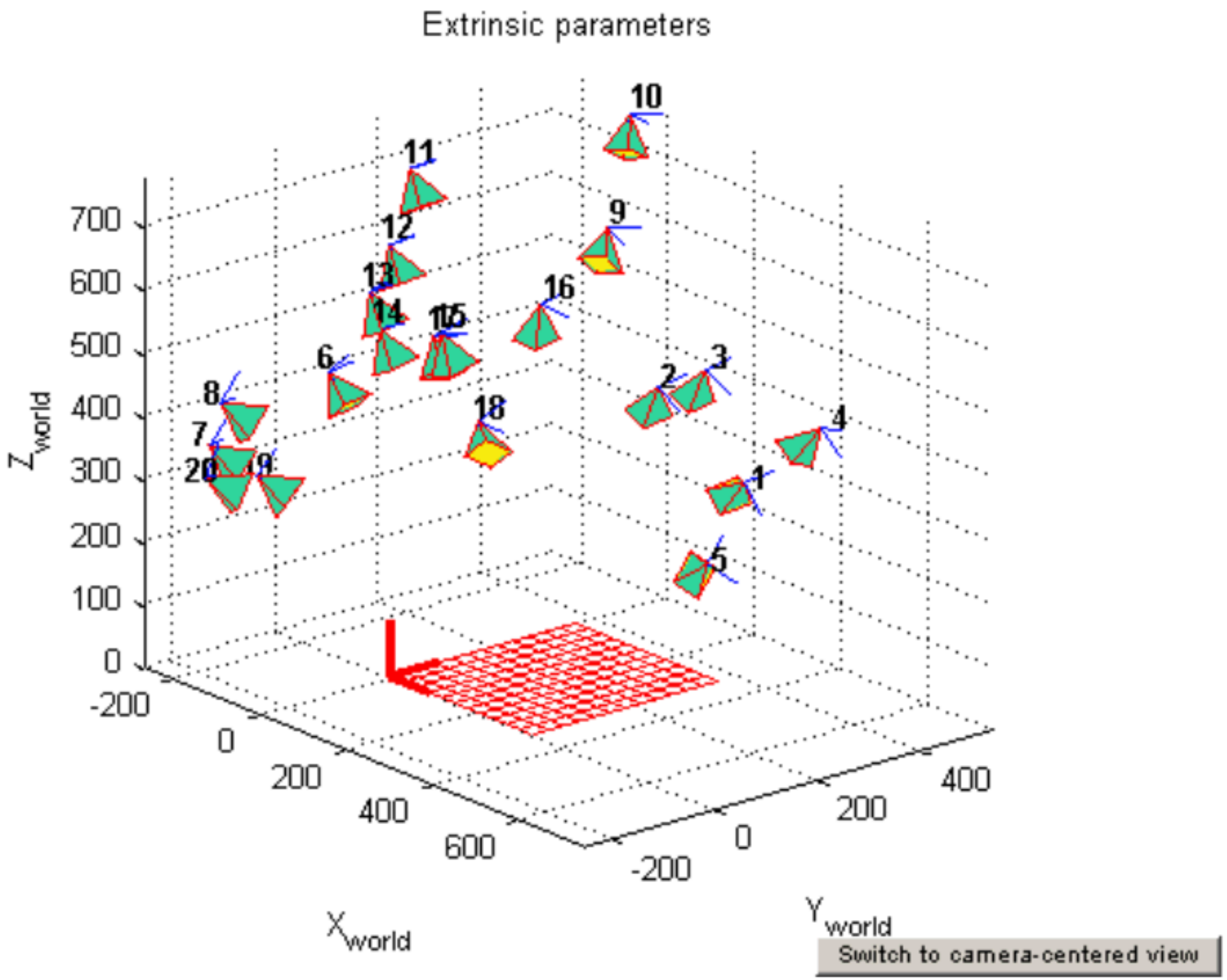
Calibrating Intrinsics + Multiple Extrinsics



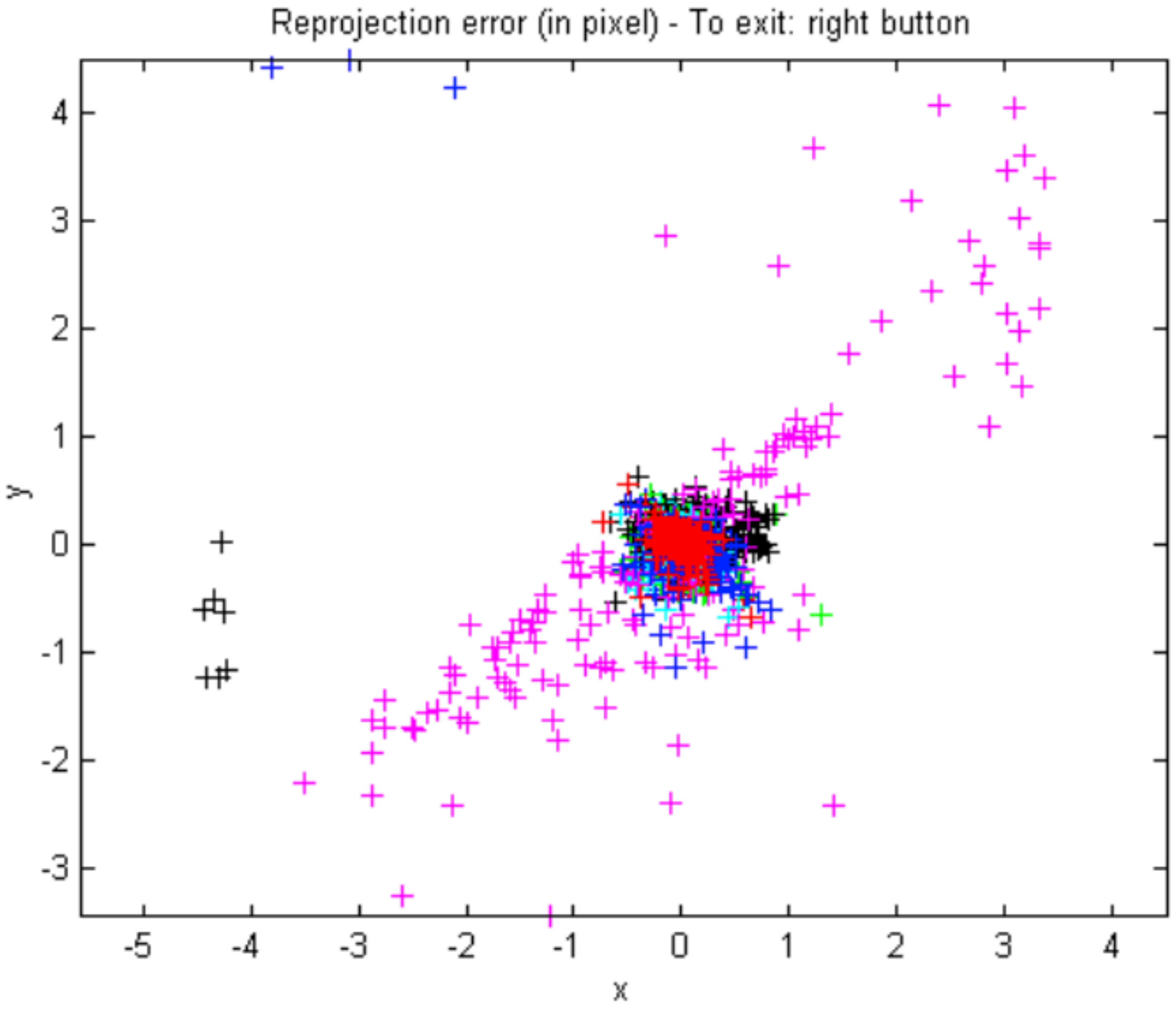
Calibrating Intrinsics + Multiple Extrinsics



Calibrating Intrinsics + Multiple Extrinsics



Visualization of Reproduction Errors

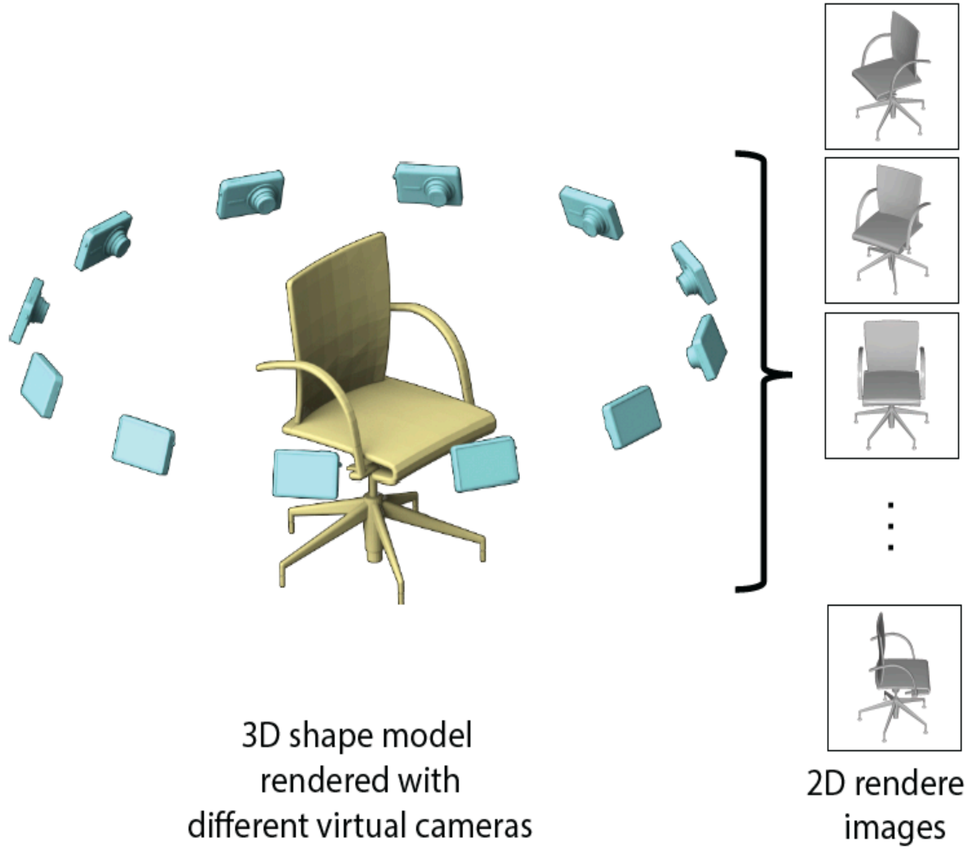


Depth Images

2D Image Representations

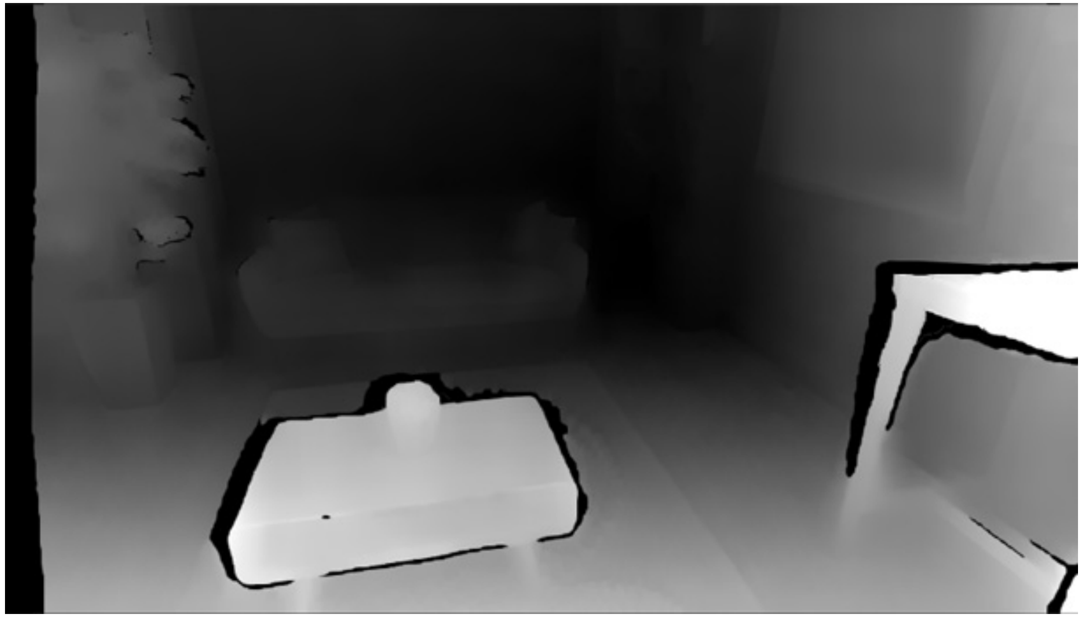
$$H \times W \times 3$$

Multi-View Images



- Multiple images from different viewpoints
- Contain 3D information
- Indirect, not a true 3D representation

Depth Image



- A single-channel image filled by depth values
- A 2.5D representation

True 3D representation should enable distance measurement between two points.

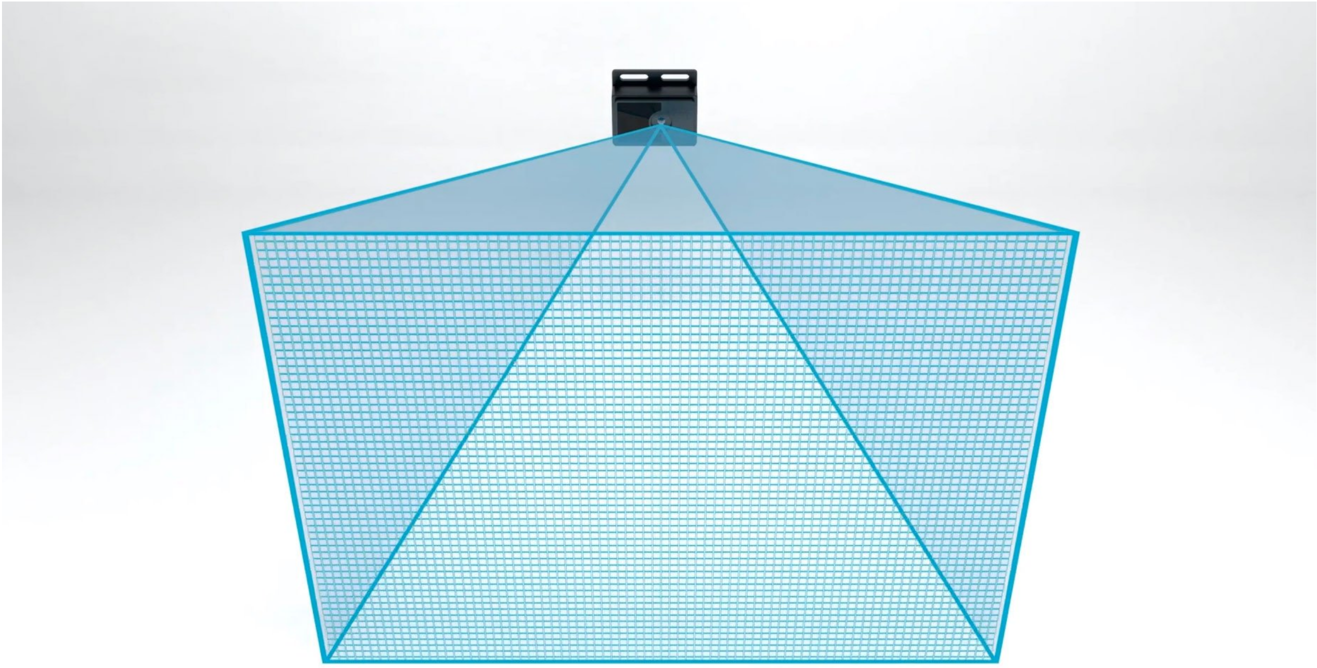
3D Data: from Sensors or Graphics

Real 3D data acquired by 3D sensing

Synthetic 3D data

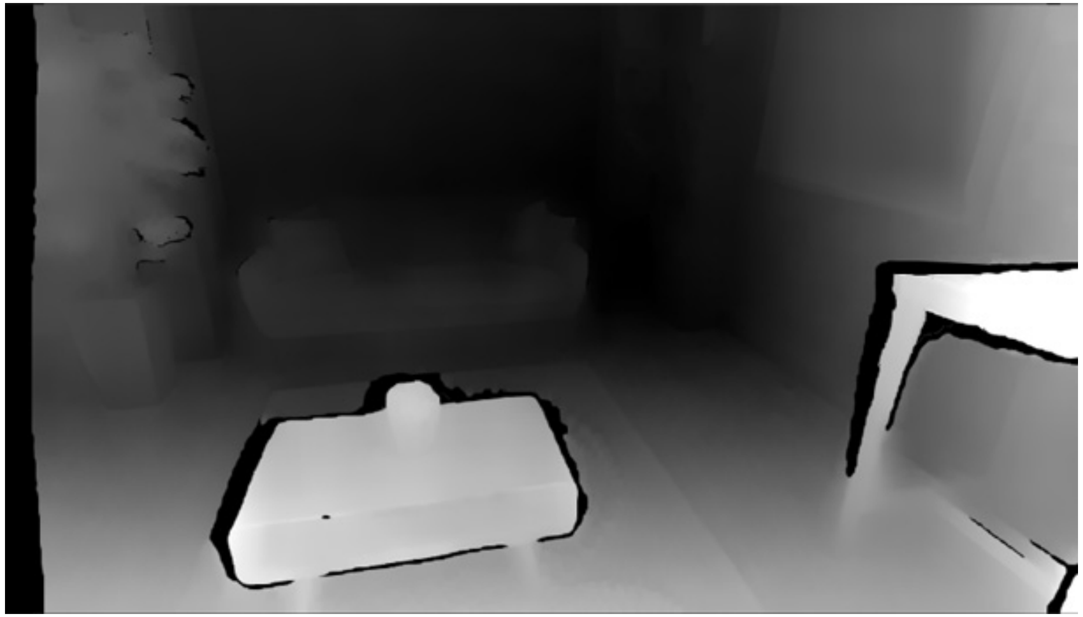
Depth Sensors

- Depth sensors are a form of 3D range finder
- Measure multi-point distance information across a wide Field-of-View (FoV)



https://www.terabee.com/depth-sensors-precision-personal-privacy

Depth Image



- A single-channel image filled by depth values
- A 2.5D representation

True 3D representation should enable distance measurement between two points.

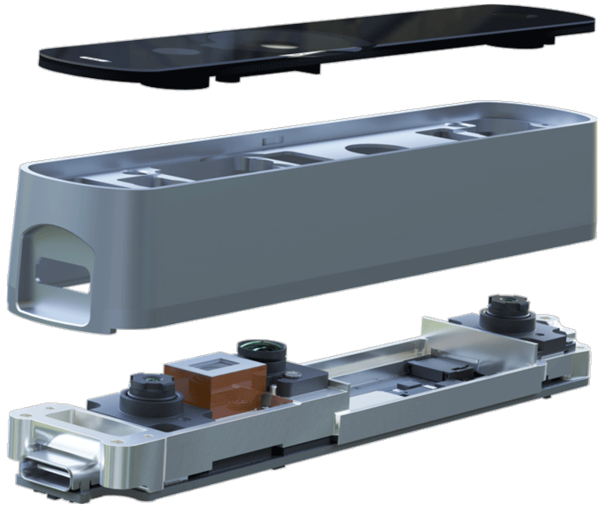
Stereo Sensors

- Mechanism: estimate correspondence, compute disparity and then turn it into depth.

Stereolabs Zed

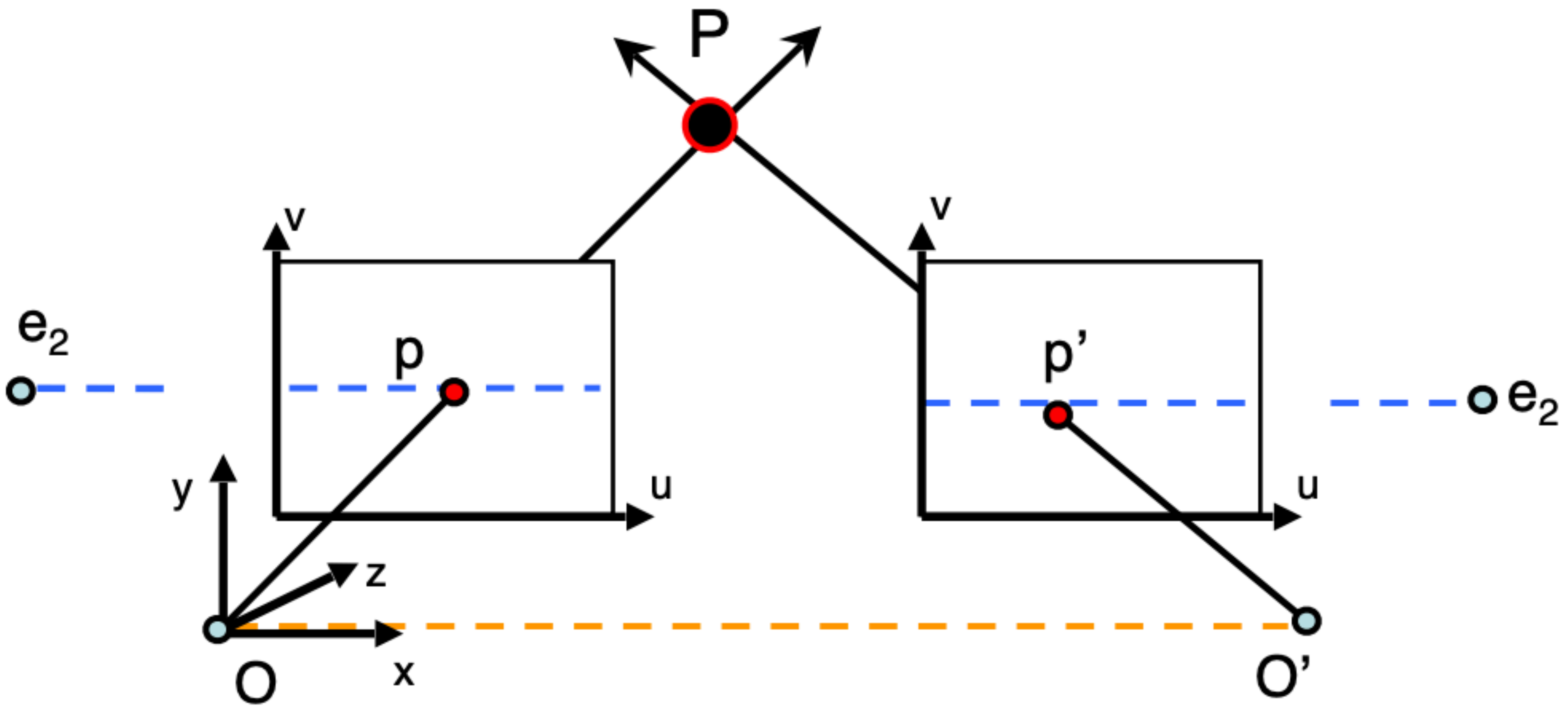
Intel RealSense

Ensenso

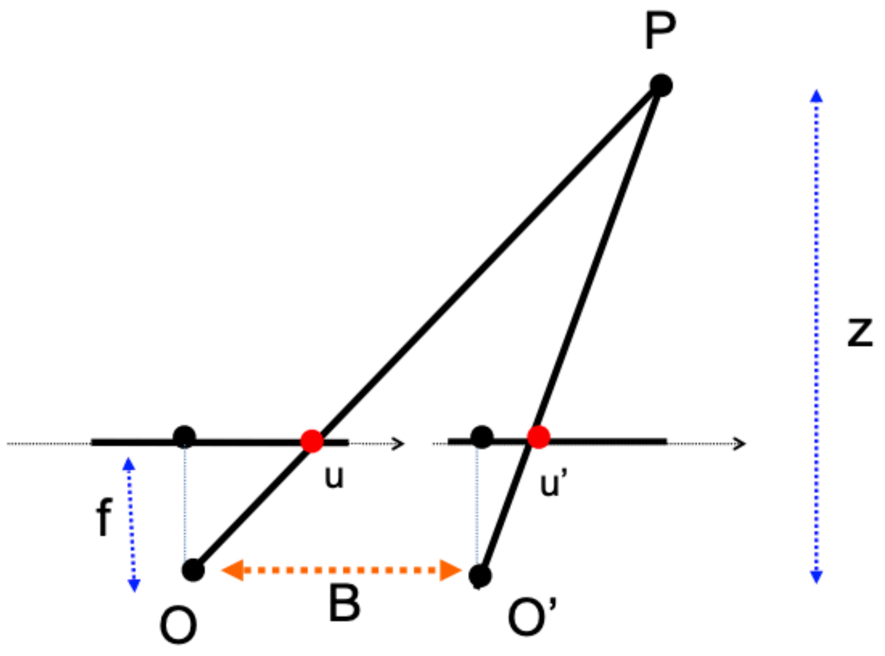


Occipital Structure Core

Point Triangulation



Computing Depth



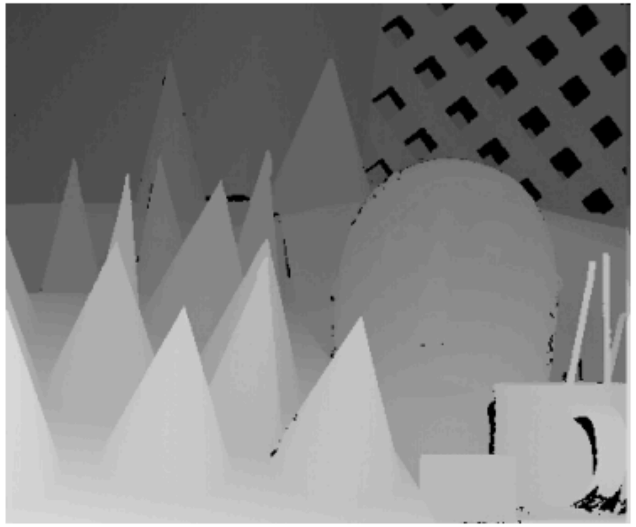
$$u - u' = \frac{B \cdot f}{z} = \text{disparity} \qquad \text{[Eq. 1]}$$

Note: Disparity is inversely proportional to depth

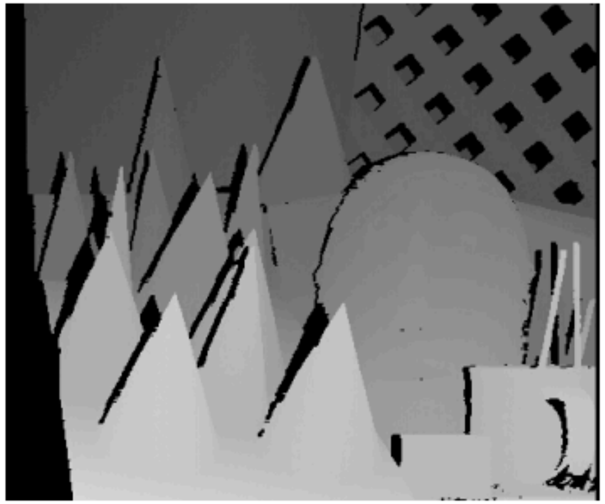
Disparity Maps

$$u - u' = \frac{B \cdot f}{z}$$

Stereo pair



Disparity map / depth map



Disparity map with occlusions

http://vision.middlebury.edu/stereo/

Advantages and Disadvantages of Stereo Sensors

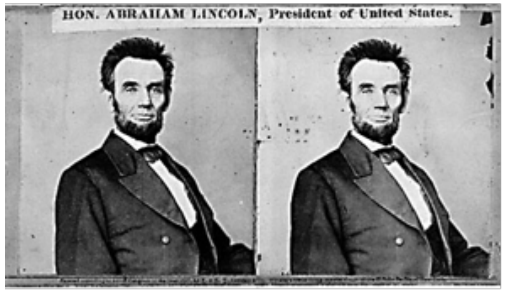
Advantages:

1. Robust to the illumination of direct sunlight
2. Low implementation cost

Disadvantage:

Finding correspondences along $Image_L$ and $Image_R$ is hard and erroneous

Failure of correspondence search



Textureless surfaces

Occlusions, repetition

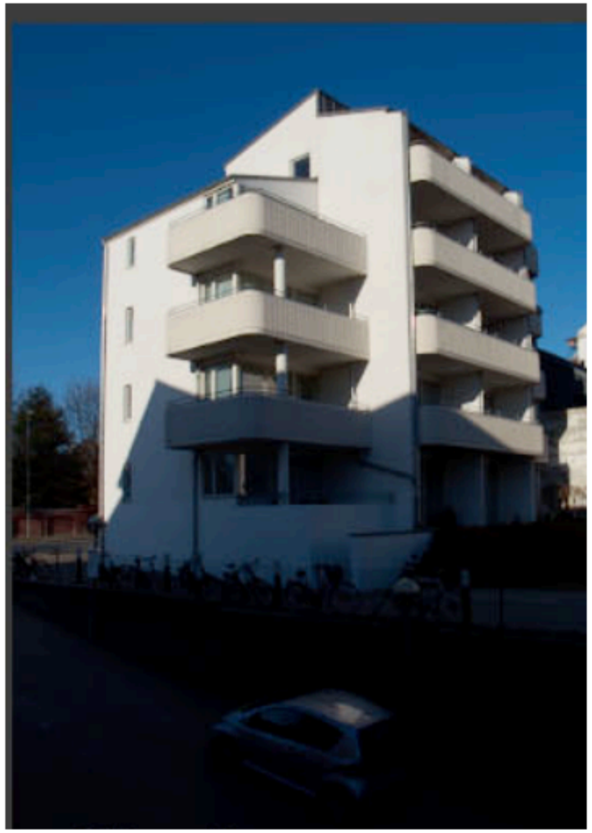
Non-Lambertian surfaces, specularities

Correspondence is Difficult

- Occlusions
- Fore shortening
- Baseline trade-off
- Homogeneous regions
- Repetitive patterns

Challenges

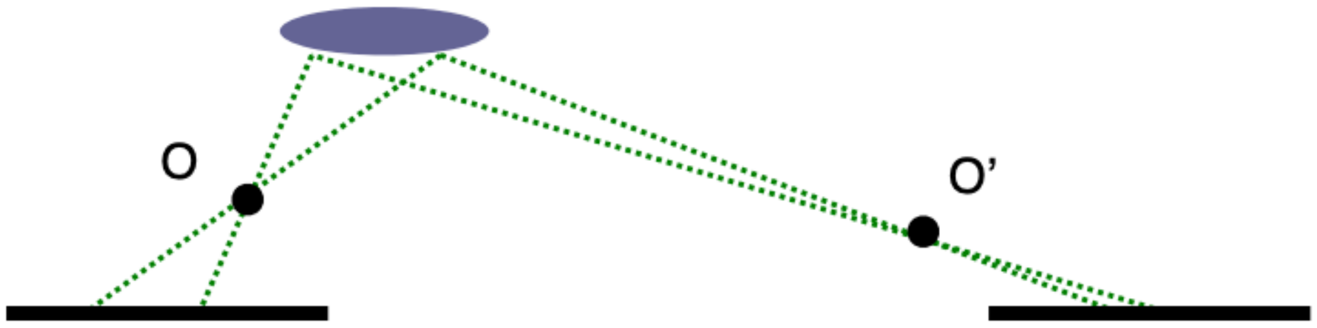
Changes of brightness/exposure



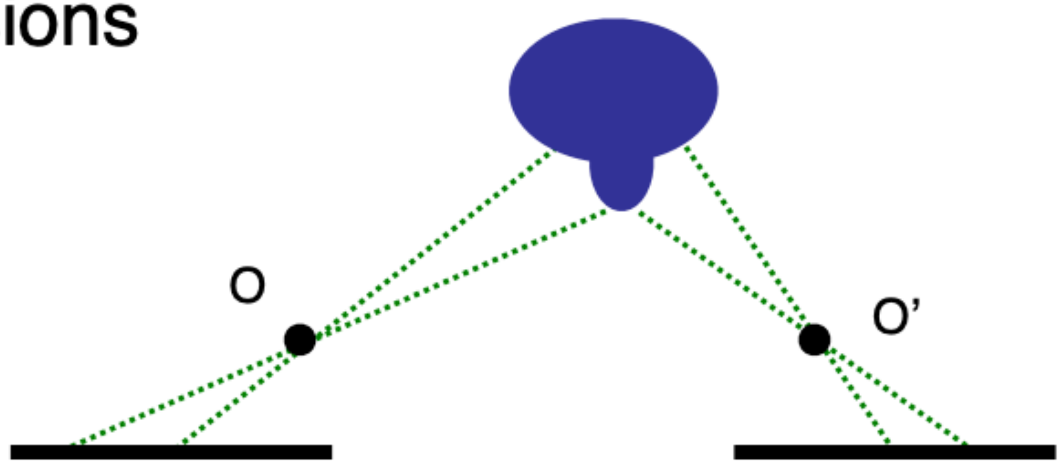
Changes in the mean and the variance of intensity values in corresponding windows!

Challenges

- Fore shortening effect

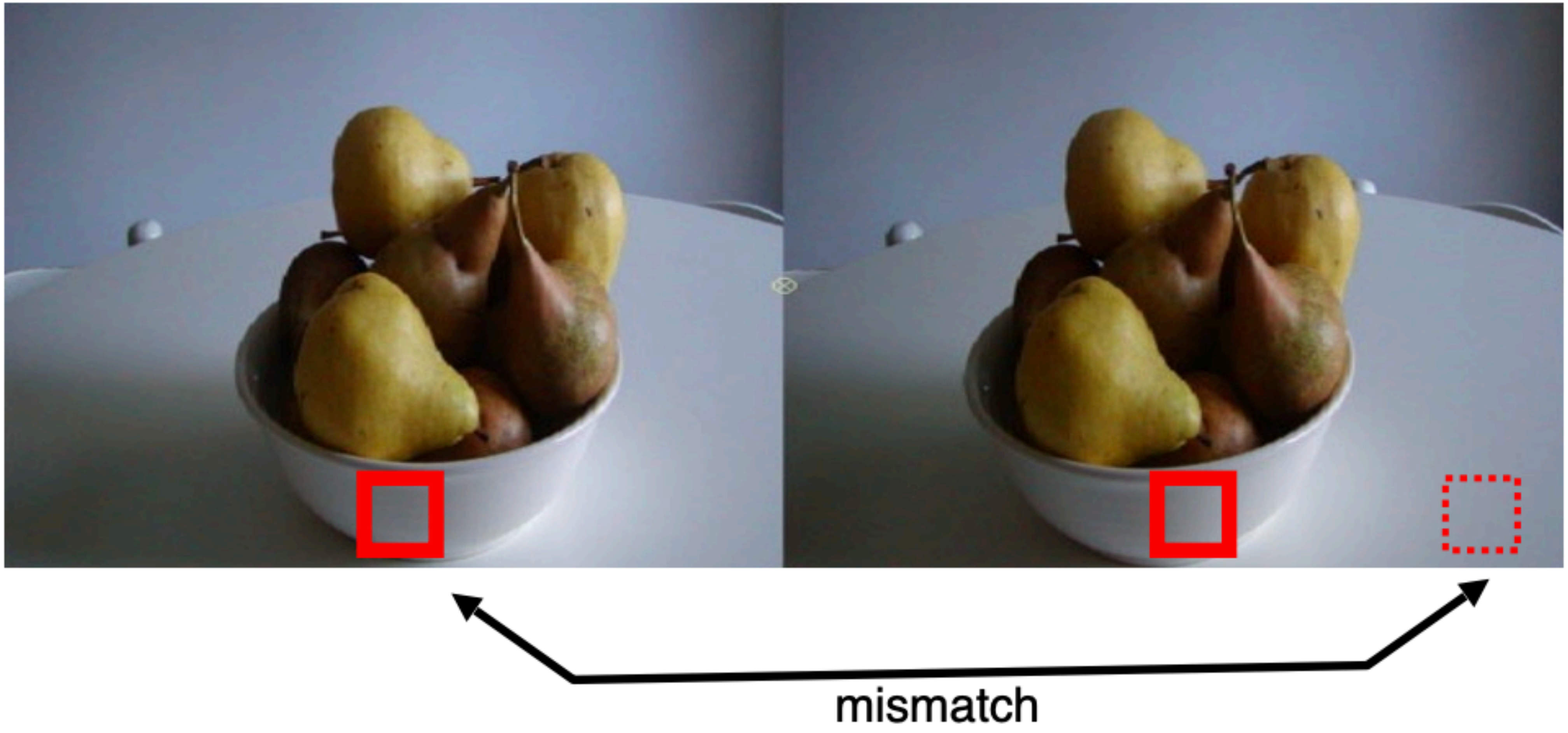


- Occlusions



Challenges

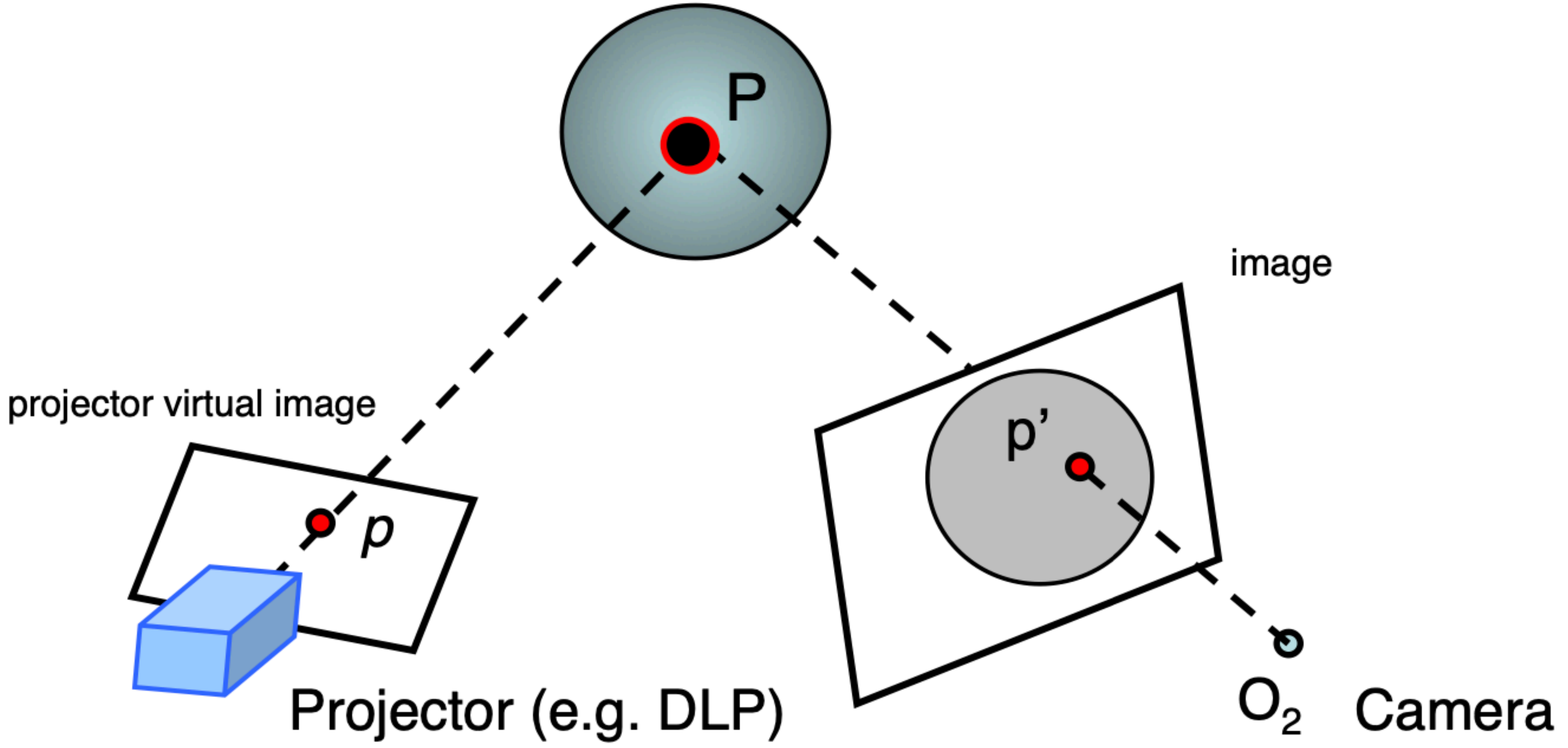
- Homogeneous regions



Challenges

- Repetitive patterns

Active Stereo



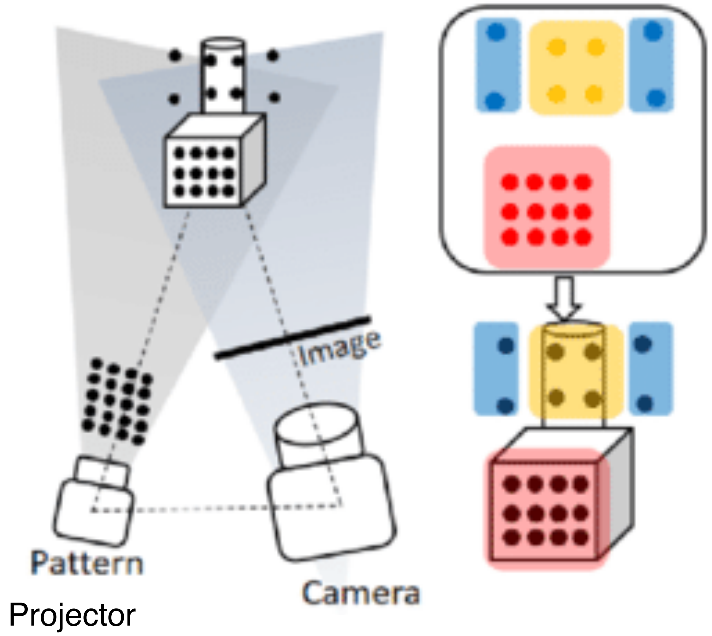
Replace one of the two cameras by a projector
- Single camera
- Projector geometry calibrated
- What's the advantage of having the projector? Correspondence problem solved!

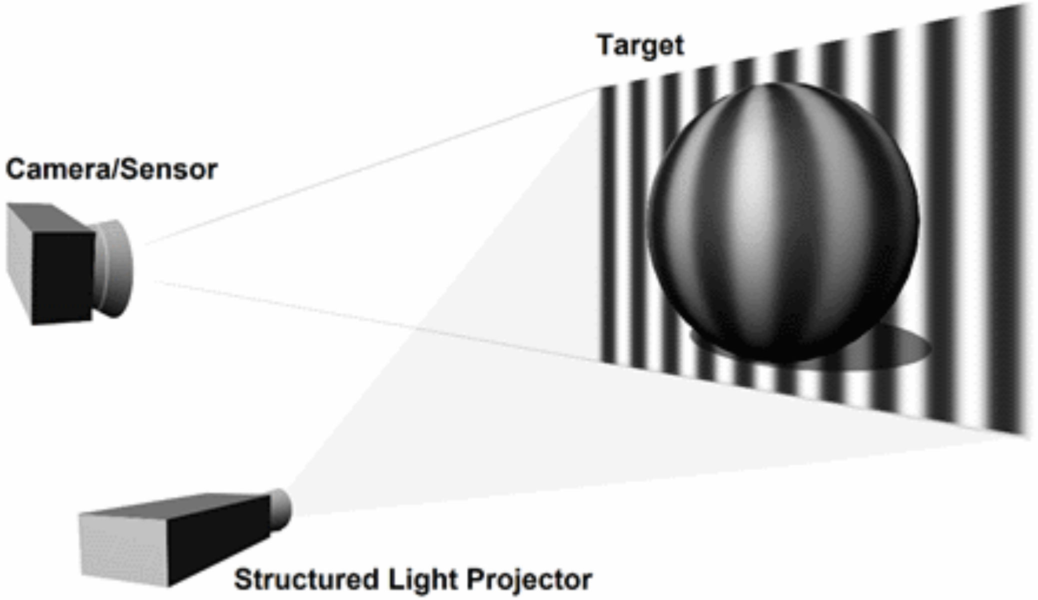
Structured Light

- Belongs to active stereoscopic approaches
- One camera replaced by an infrared projection unit
- Generates a pattern by projecting on the imaged surface

Advantage:
1. Simplify the correspondence problem

Drawback:
1. Near field
2. Indoor





Structure Light

RealSense D415

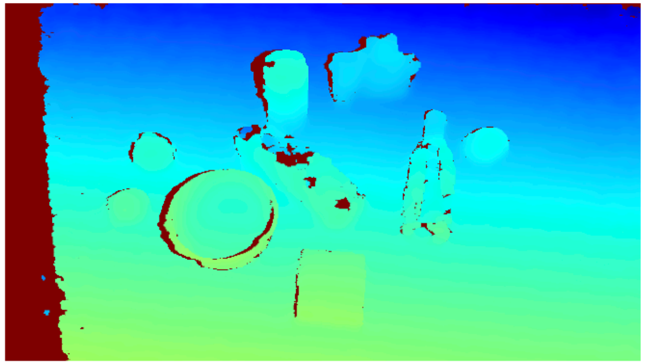
RealSense D435

RealSense D455

- RGB camera, infrared projector, left and right infrared cameras.
- Captures video data in 3D under any ambient light conditions.

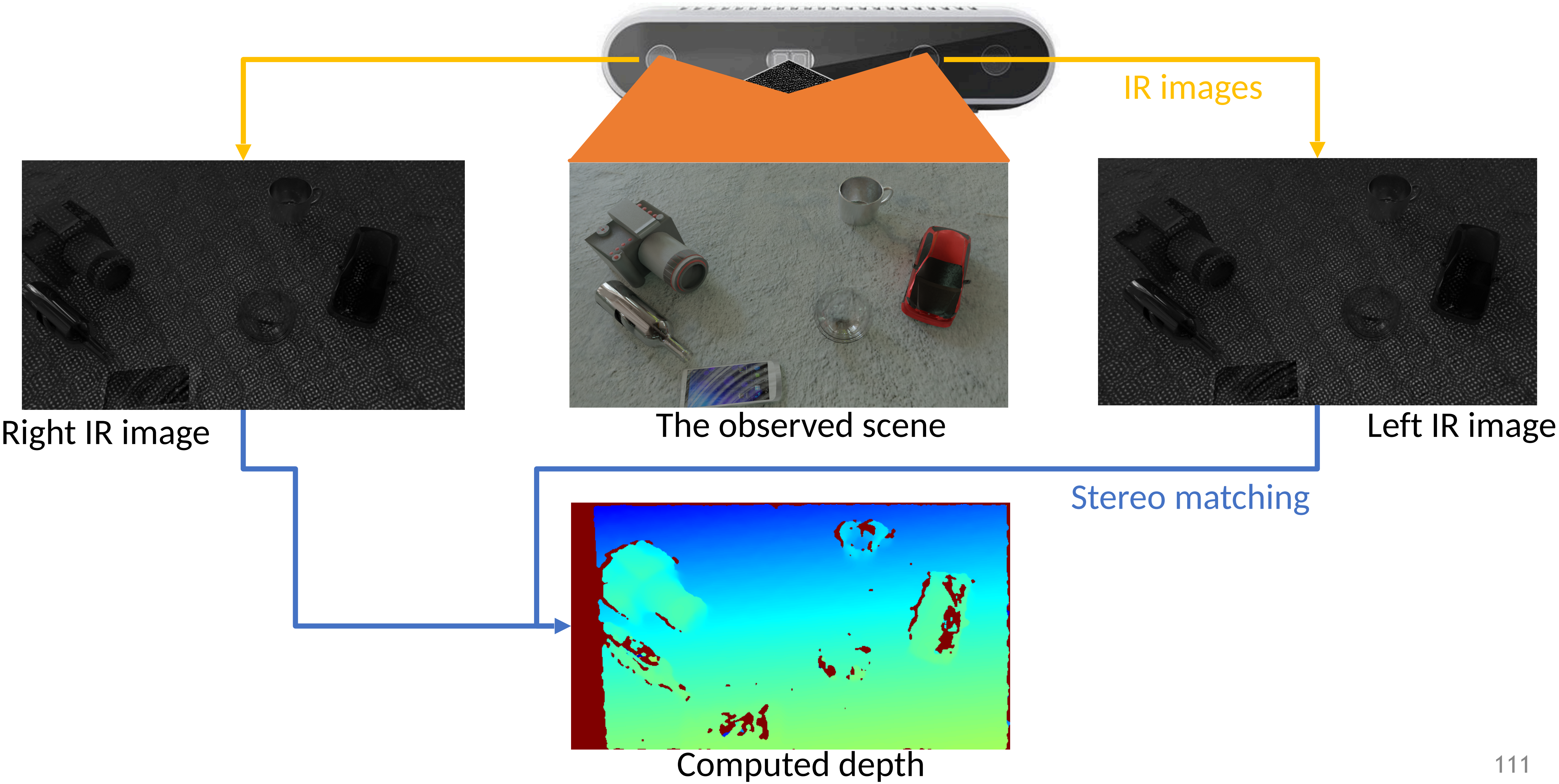
RGB image

Pattern of projected infrared points to generate a dense 3D image



Depth map

Structural Light



Time-of-Flight (ToF) Sensors

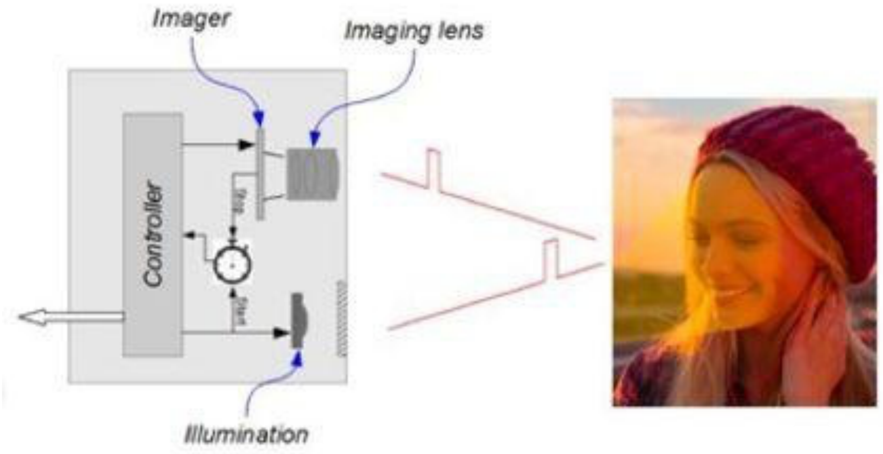
Microsoft Kinect v2 (2013)

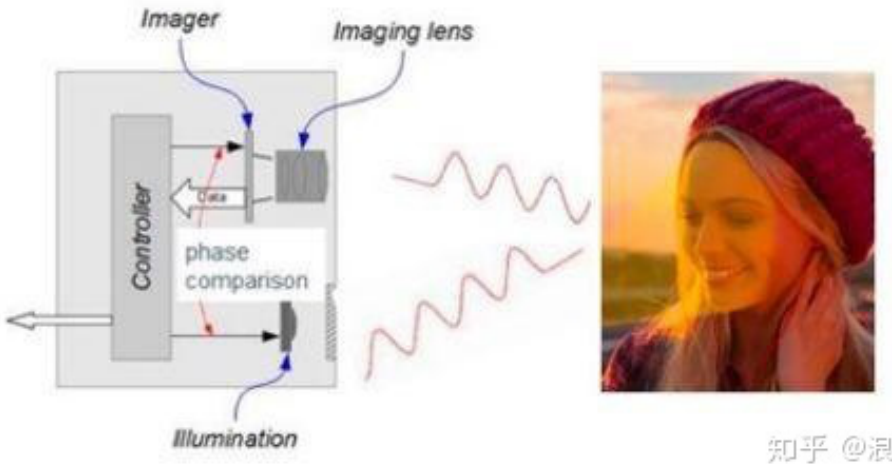
Microsoft Azure Kinect (2020)

iPad Pro 2019 LiDAR

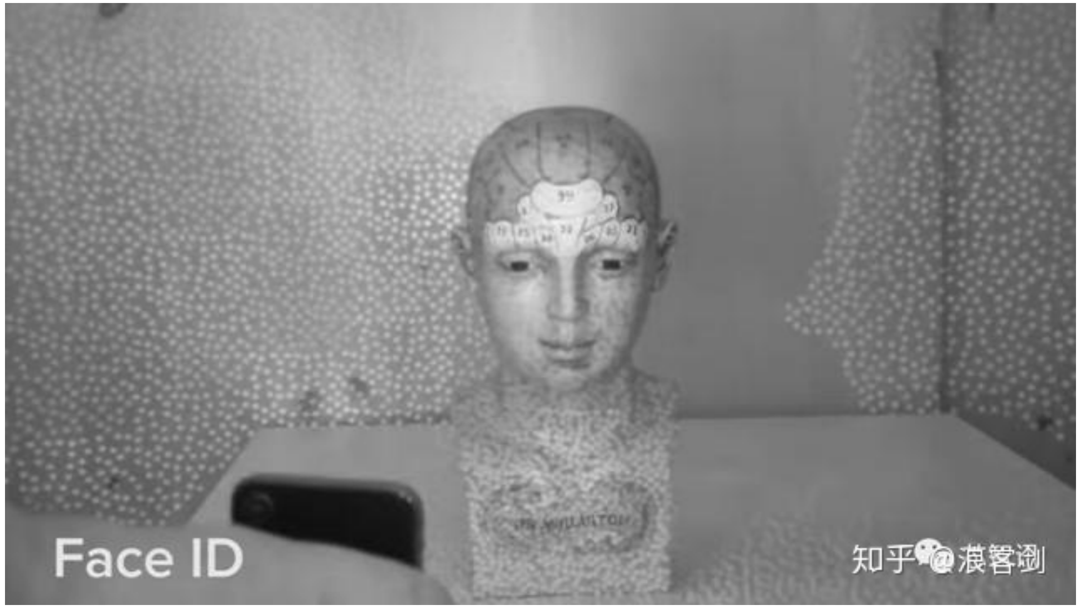
iToF vs. dToF

- dToF (the future)
  - Direct time-of-flight
  - Pulse wave
  - Long range
  - Theoretically higher precision but currently lower resolution
  - Expensive (needs SPAD)
- iToF (Classic 3D imaging)
  - Indirect time-of-flight
  - Sin wave and solve for phase shift
  - Lower range
  - Lower precision but higher resolution
  - Cheaper





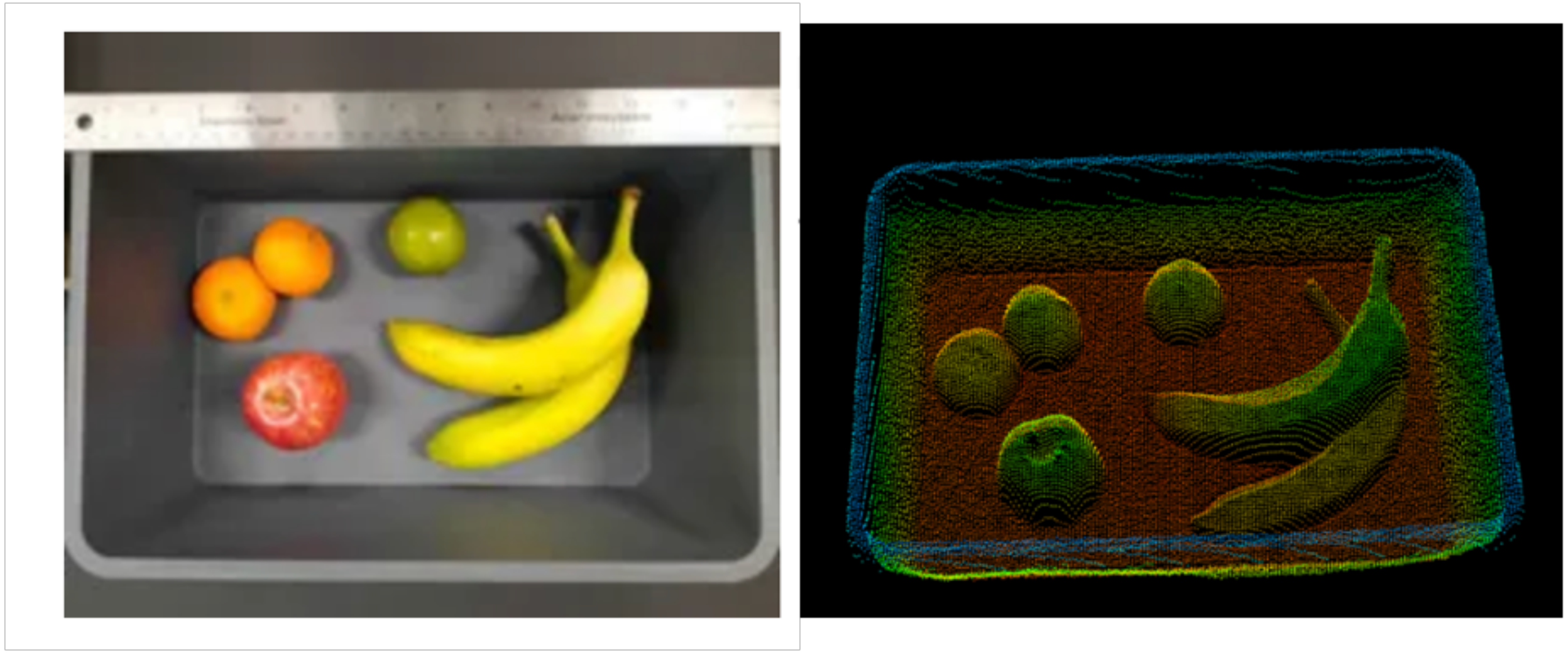
iPad Pro Front Structure Light vs. LiDAR



dToF in iPad Pro is not there yet.

Next Generation ToF

- Industrial level 3D sensor
- https://thinklucid.com/helios-time-of-flight-tof-camera/



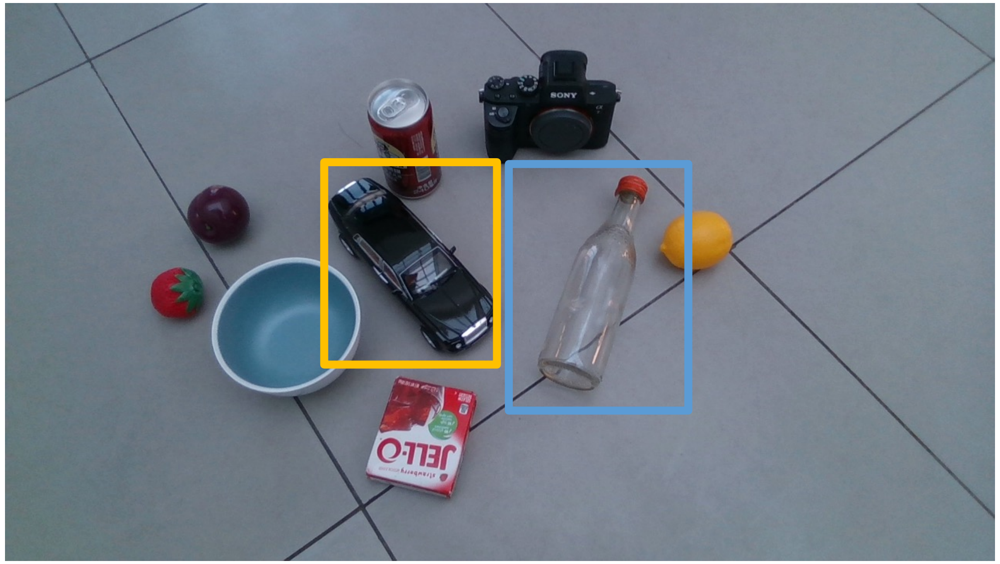
Helios2 sensor from Lucid Vision Labs

Summary of Different Depth Sensors

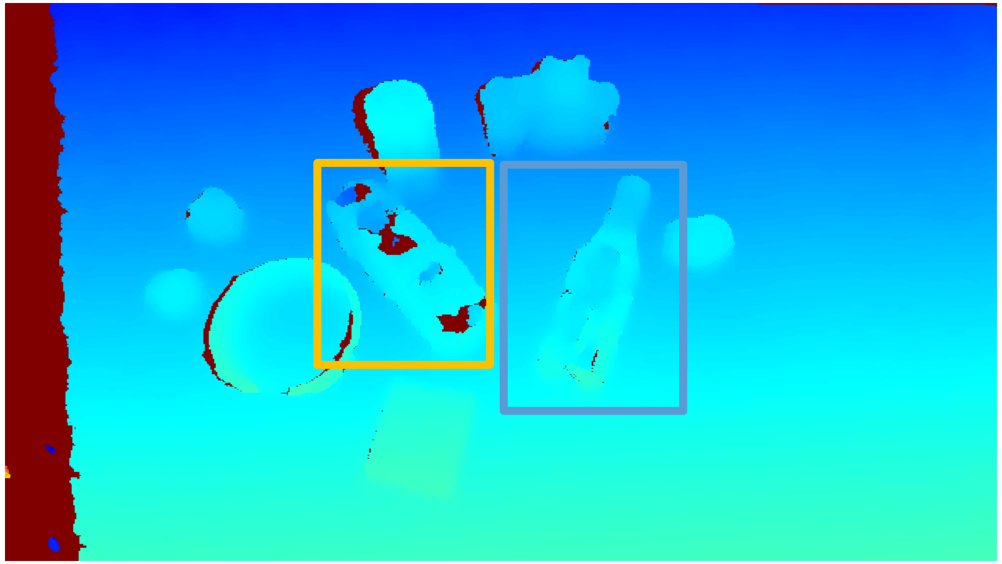
CONSIDERATIONS	STEREO VISION	STRUCTURED-LIGHT	TIME-OF-FLIGHT (TOF)
Software Complexity	High	Medium	Low
Material Cost	Low	High	Medium
Compactness	Low	High	Low
Response Time	Medium	Slow	Fast
Depth Accuracy	Low	High	Medium Quickly improving!
Low-Light Performance	Weak	Good	Good
Bright-Light Performance	Good	Weak	Good
Power Consumption	Low	Medium	Scalable
Range	Limited	Scalable	Scalable

Failure Cases in Depth Sensing

RGB

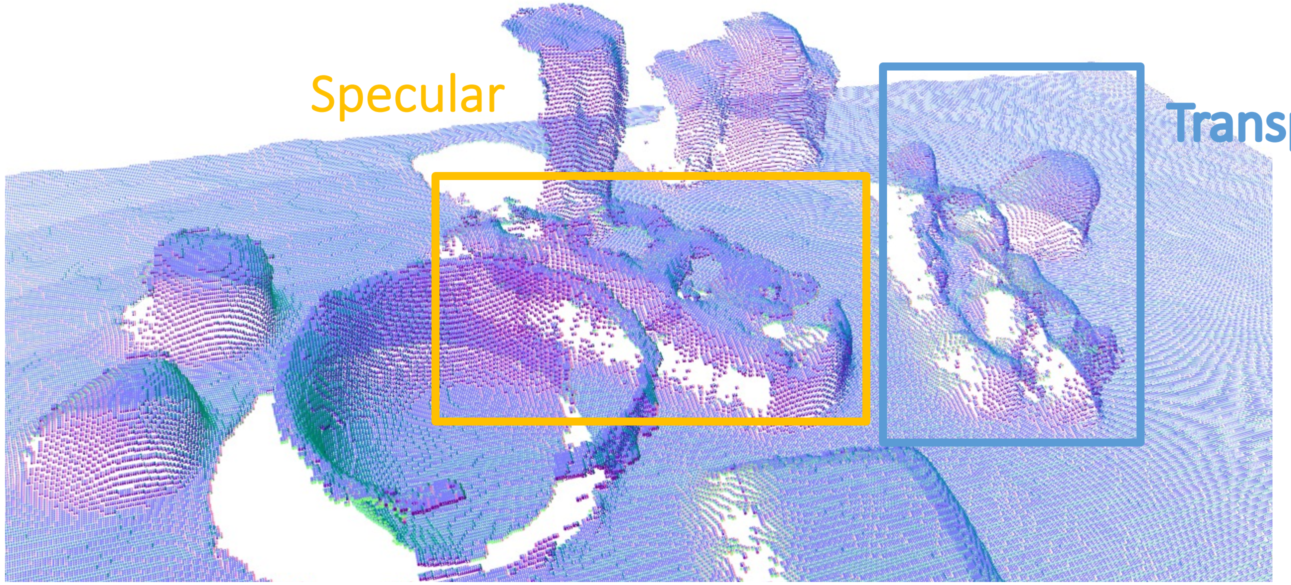


Depth



Point cloud

Specular



Transparent

Next week: Lecture 10,
3D Vision II