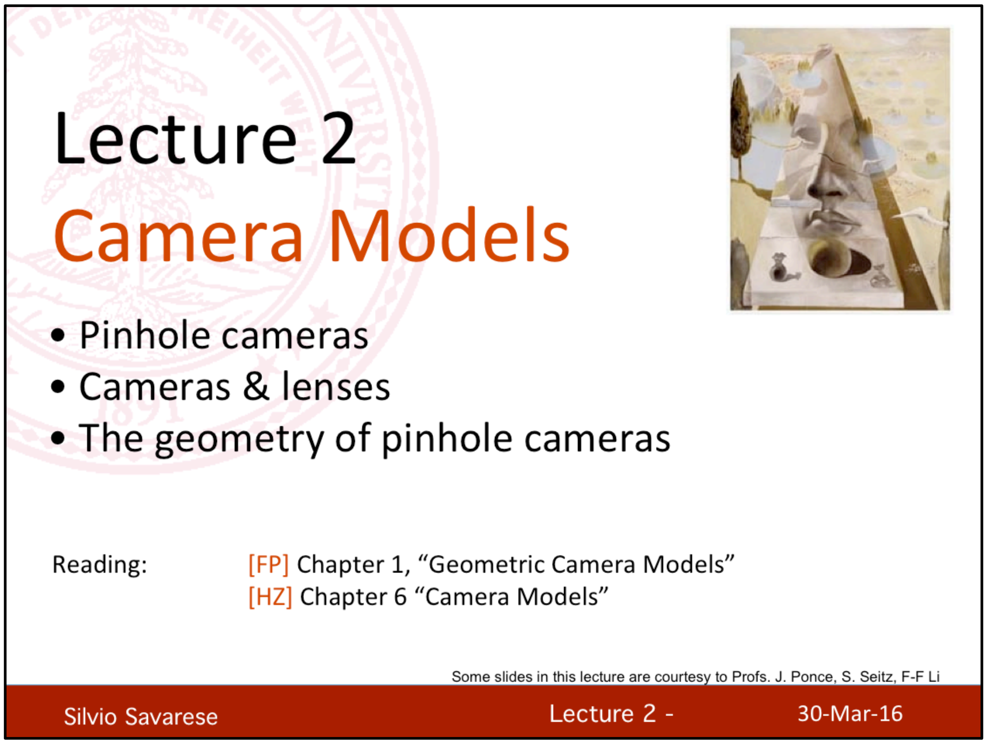
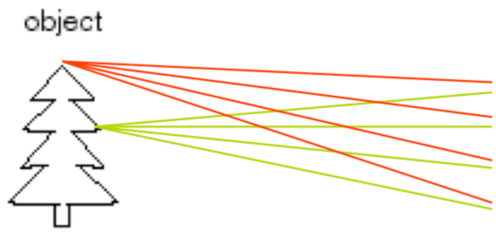
Lecture 2
Camera Models

Professor Silvio Savarese
Computational Vision and Geometry Lab



In this lecture we will talk about camera models and the basic properties of a camera. We will start from the most basic model called pinhole camera model. We will also talk about lenses— a key component for the design of a camera. We will discuss in details the geometrical properties of the pinhole camera model and conclude the lecture by illustrating examples of other types of camera models.

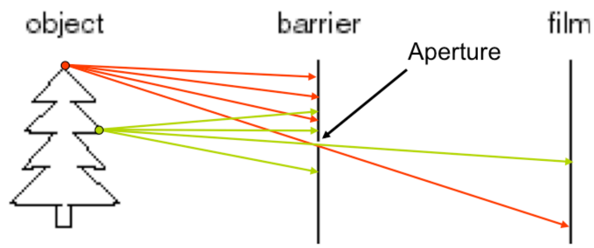
How do we see the world?



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

How would you design a camera?

Pinhole camera



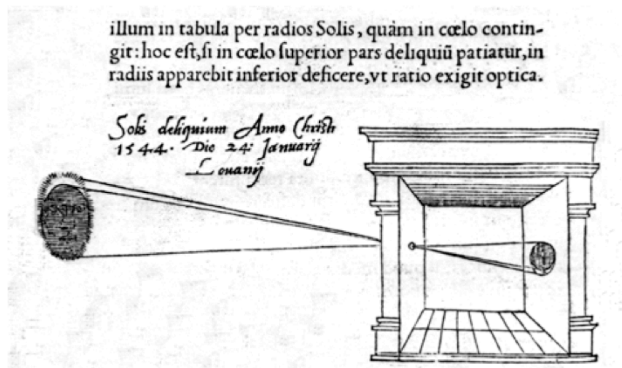
- Add a barrier to block off most of the rays
  - This reduces blurring
  - The opening known as the **aperture**

Let's design the simplest camera system – a system that can record an image of an object or scene in the 3D world. The simplest camera system can be designed by placing a barrier with a small aperture between the 3D object and a photographic film or sensor (that is, a material that is sensitive to light and can be used to record the light that impresses the film/sensor). As the slide shows, each point on the 3D object emits multiple rays of light outwards; different points and related rays of light are shown in different colors. Only one (or a few) of these rays of light passes through the aperture and hit the film. So an approximated 1-to-1 mapping can be established between the 3D object and the film. The result is that the film gets exposed by an "image" of the 3D object by means of this mapping. This simple camera model is called ***pinhole camera***.

Some history…

Milestones:

• Leonardo da Vinci (1452-1519):
first record of camera *obscura (1502)*

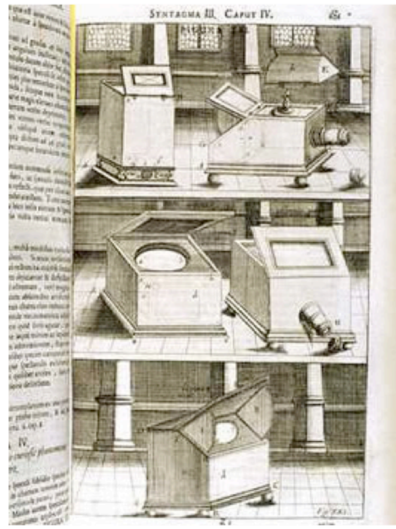


History of cameras:
The concept pinhole camera was first formalized by Leonardo da Vinci in 1502. He called it “Camera Obscura”.

Some history…

Milestones:
• Leonardo da Vinci (1452-1519): first record of camera *obscura*

• Johann Zahn (1685): first portable camera



Johann Zahn, more than a century later, introduces the first portable camera.

Some history…

Milestones:

• Leonardo da Vinci (1452-1519): first record of camera *obscura*

• Johann Zahn (1685): first portable camera

• Joseph Nicéphore Niépce (1822): first photo - birth of photography

Photography (Niépce, "La Table Servie," 1822)

We have to wait until 1822 for first photo which is credited to Joseph Niepce – this marks the birth of photography!

Some history…

Milestones:

• Leonardo da Vinci (1452-1519): first record of camera *obscura*

• Johann Zahn (1685): first portable camera

• Joseph Nicéphore Niépce (1822): first photo - birth of photography

•Daguerréotypes (1839)
• Photographic Film (Eastman, 1889)
• Cinema (Lumière Brothers, 1895)
• Color Photography (Lumière Brothers, 1908)

Photography (Niépce, "La Table Servie," 1822)

Other important events include:

- Daguerréotypes (1839)
- Photographic Film (Eastman, 1889)
- Cinema (Lumière Brothers, 1895)
- Color Photography (Lumière Brothers, 1908)

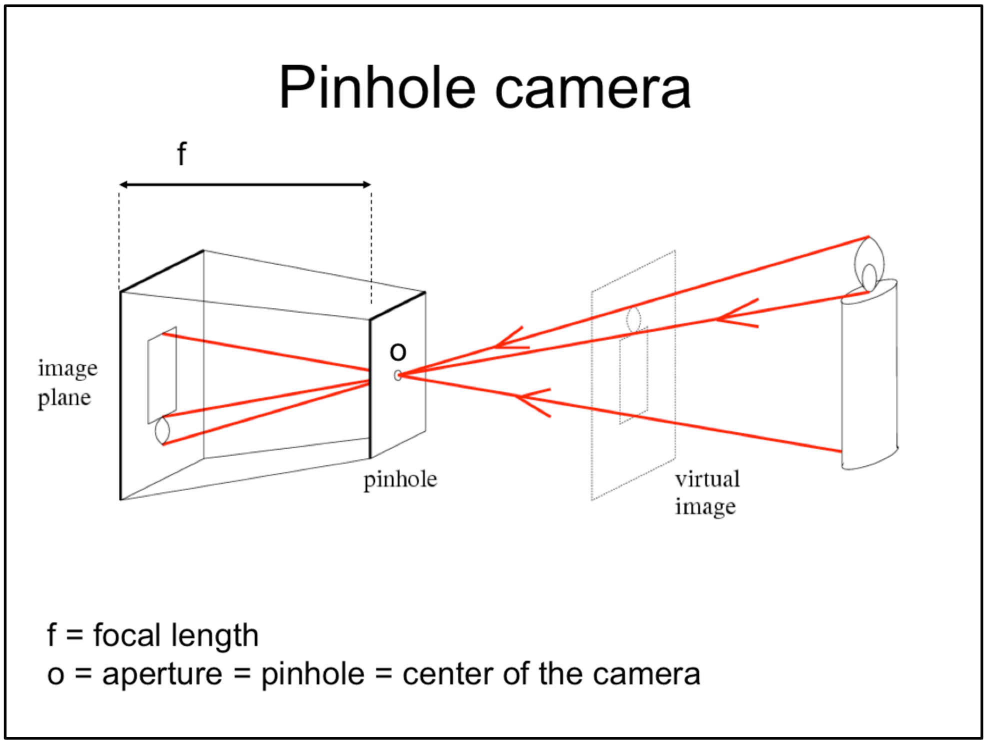
Let's also not forget…

Motzu
(468-376 BC)
Oldest existent book on geometry in China

Aristotle
(384-322 BC)
Also: Plato, Euclid

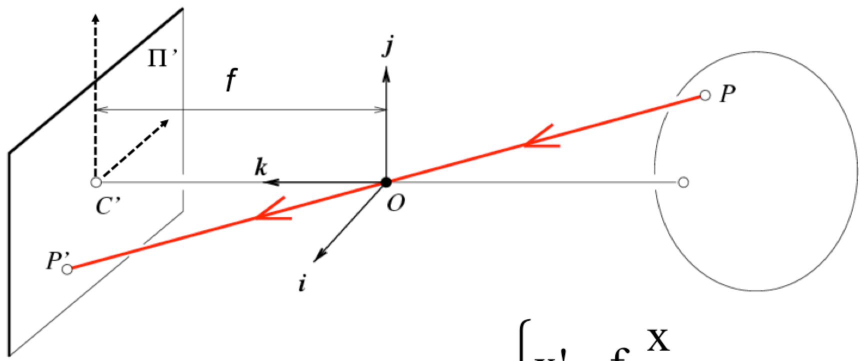
Al-Kindi (c. 801–873)
Ibn al-Haitham
(965-1040)

All this amazing progress wouldn't be possible without to the foundational work by philosophers such as Motzu, Aristotle and Al-kindi. The latter is credited to be one of the founders of modern optics.



A more formal construction of the pinhole camera model is shown here. The film (now on the left of the object) is commonly called ***image or retinal plane.*** The aperture is called ***pinhole*** O or ***center of the camera***. The distance between the image plane and O is the ***focal length f.***
Often, the retinal plane is placed between O and the 3D object at a distance f from O. In this case it is called ***virtual image*** or ***virtual retinal plane***. Note that the projection of the object in the image plane and the image of the object in virtual image plane are identical up to a scale (similarity) transformation.

Pinhole camera



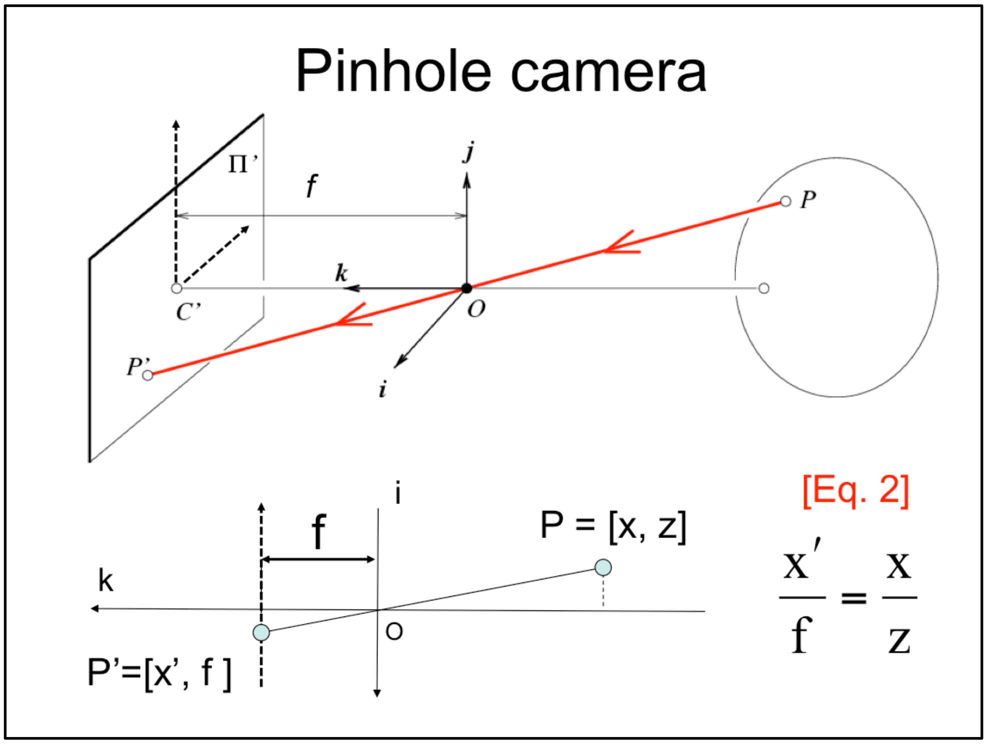
$$P = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \rightarrow P' = \begin{bmatrix} x' \\ y' \end{bmatrix} \qquad \begin{cases} x' = f\,\dfrac{x}{z} \\[2ex] y' = f\,\dfrac{y}{z} \end{cases} \quad \text{[Eq. 1]}$$

Derived using similar triangles

Let P be a point on the 3D object; P= $[x\ y\ z]^T$. P is mapped (projected) into a point P' = $[x'\ y']^T$ in the image plane Π' following the pinhole mapping through the aperture O. P' can be expressed as $[x'\ y']^T$. The projection of O into the image plane gives C'.

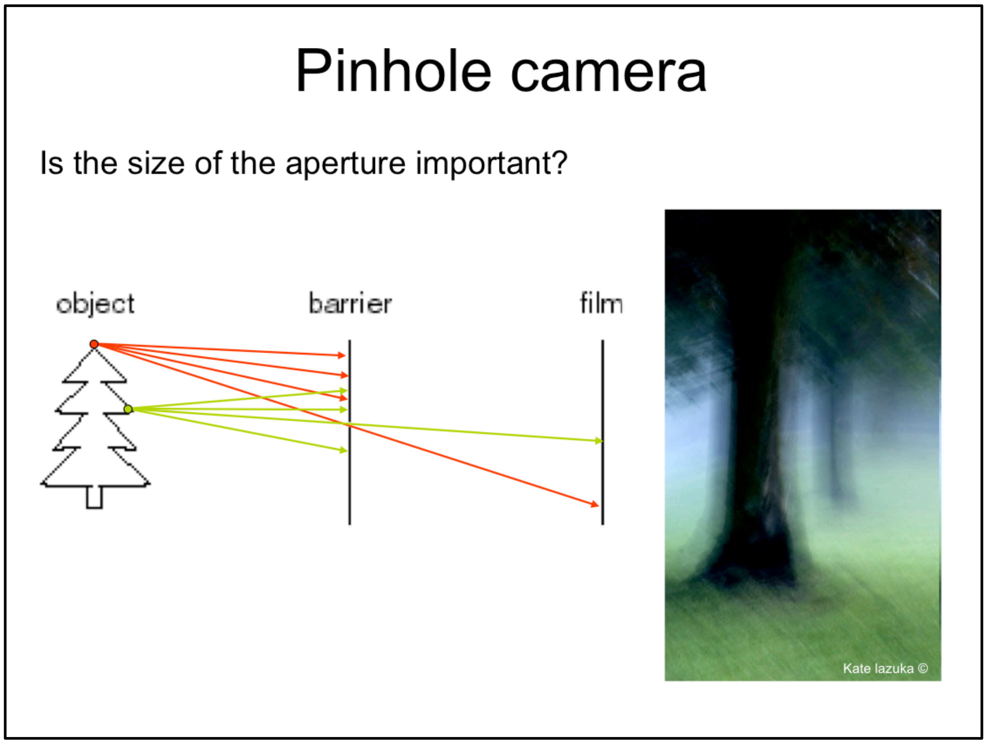
Here a coordinate system [i,j,k] is centered at O such that the axis k is perpendicular to the image plane and points toward it. This is the **camera reference system.** The line defined by C' and O is called the **optical axis** of the camera system.

The relationship between P and P' is shown in Eq. 1 and it is derived using similar triangles as shown next in a few more details.

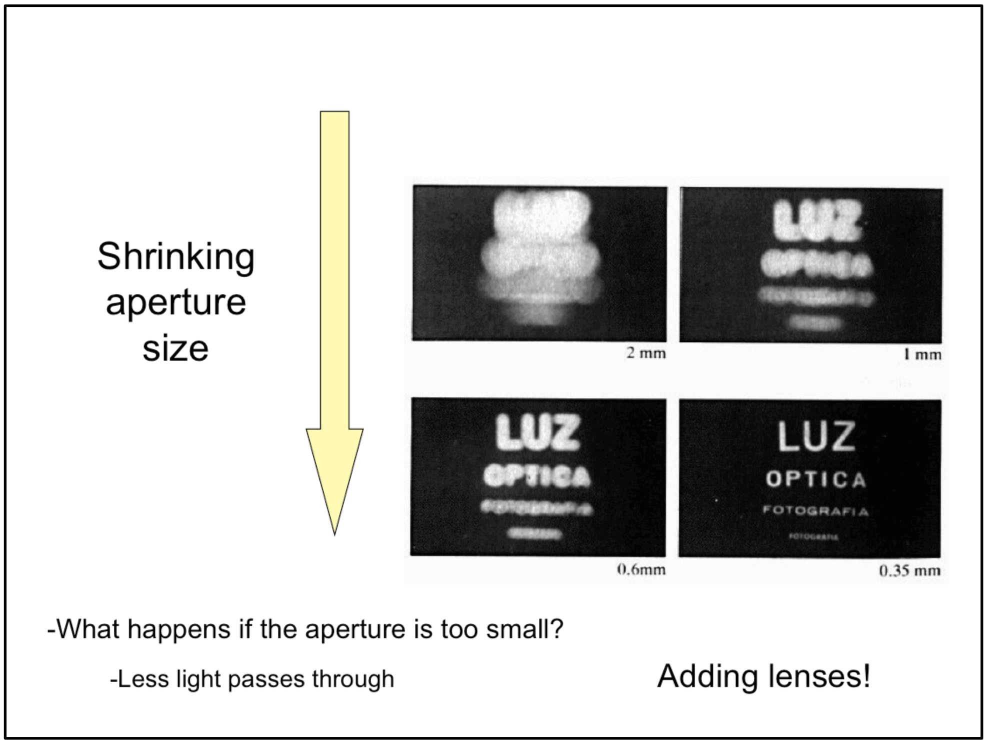


$$\frac{x'}{f} = \frac{x}{z}$$

We derive the relationship between P and P' for the simpler case where P, O and P' belongs to the same plane defined by i and k (which is perpendicular to j). The bottom figure shows such a plane. The more general case is a simple extension of this and we leave it as an exercise.
As the figure shows, the triangle defined by P'=[x',f], O and [0,f] is similar to the triangle defined by P, O and [0,z]; from this we can easily derive Eq.2



This slide illustrates the importance of the size of the aperture:
The larger is the size of the aperture, the greater is the number of rays of lights that pass through the aperture. Thus, each point on the film may corresponds to multiple points on the object; As a result, the image of the object becomes blurred.

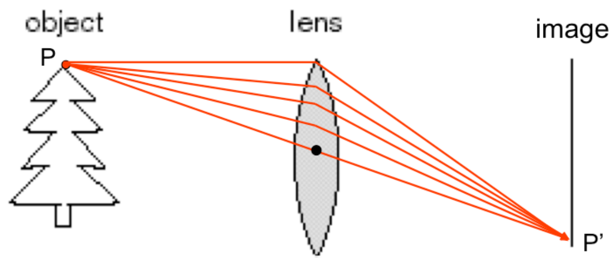


On the other hand, the smaller is the size of the aperture, the fewer are the rays of lights that pass through the aperture, which results into having a crispier but darker image of the object.

The figure illustrates this property; for a 2mm aperture the letters looks blurred; as the aperture goes from 2mm to 0.35mm, the letters look more focused but less bright.

Is there a way to overcome this problem?
Yes, by adding lenses to replace the aperture!

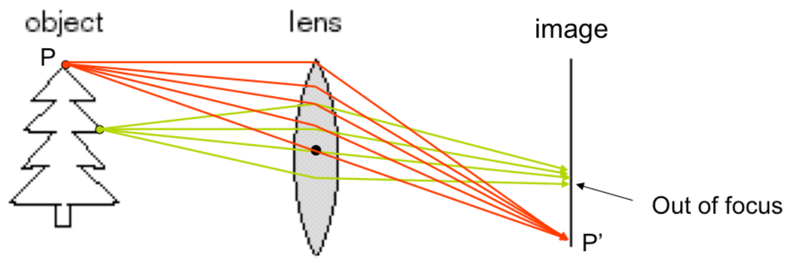
Cameras & Lenses



- A lens focuses light onto the film

This gives rise to a more complex camera model. If a lens is properly placed and has proper size, it satisfies the following property: all rays of light that are emitted by a point P are refracted by the lens such that they converge to a single point P' in the image plane.

Cameras & Lenses



- 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

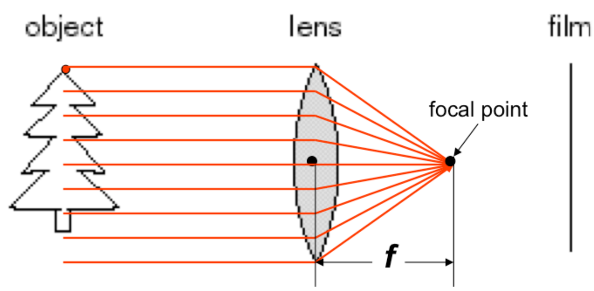
This is true for all the points on the object that are equidistant from the image plane. This no longer true for points that are closer or further to the image plane than P is (e.g. the green point in the figure). The corresponding projection into the image is now blurred or out of focus.

Cameras & Lenses

- 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

This is related to the concept of depth of field: There is a specific distance at which objects are "in focus"

Cameras & Lenses

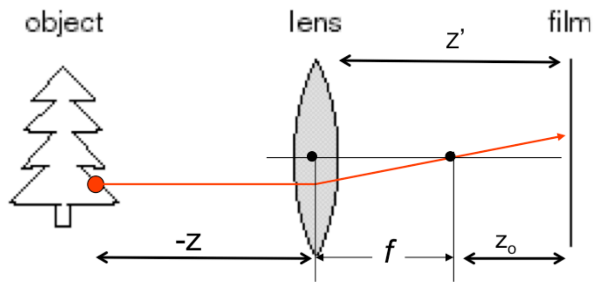


- A lens focuses light onto the film
  - All rays parallel to the optical (or principal) axis converge to one point (the *focal point*) on a plane located at the *focal length f* from the center of the lens.
  - Rays passing through the center are not deviated

Lenses have other interesting properties:

- A lens focuses light onto the film such that all rays parallel to the optical (or principal) axis converge to one point called **focal point**. The distance between the focal point and center of the lens is the *focal length f.*
- Rays passing through the center of the lens (i.e., the aperture in the pinhole model) are not deviated

Paraxial refraction model



From Snell's law:

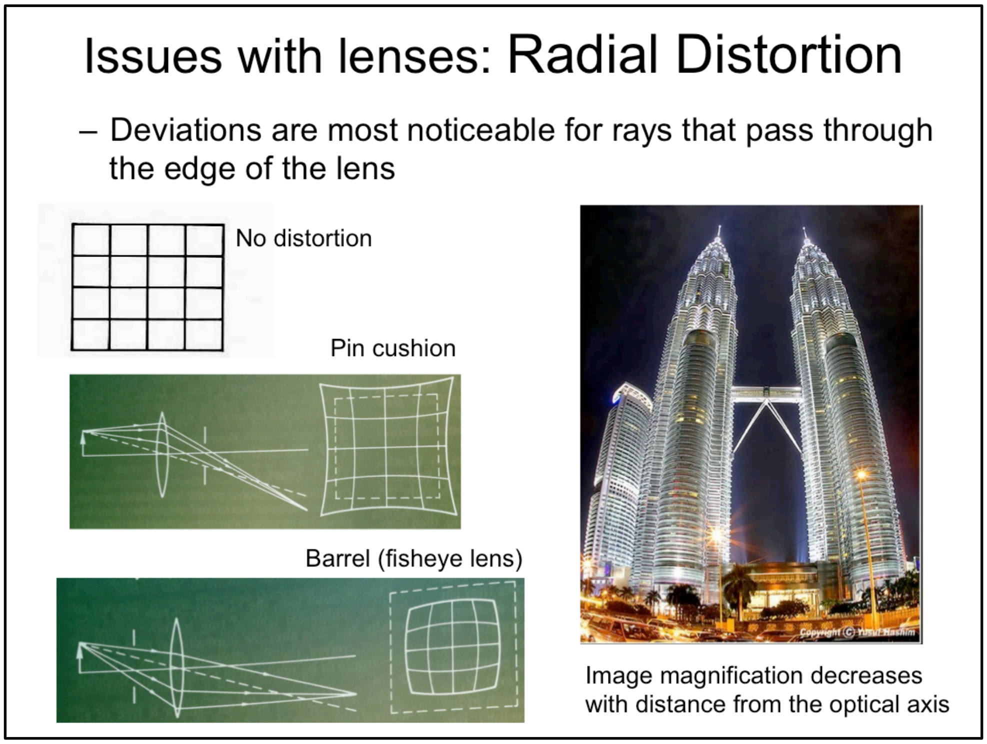
[Eq. 3]

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

[Eq. 4]

$$\boxed{\begin{aligned} z' &= f + z_o \\ f &= \frac{R}{2(n-1)} \end{aligned}}$$

Camera models based on lenses give rise to a very similar construction that allows to relate a point P in 3D and its corresponding "image" in the image plane –see Eq.3. This is essentially identical to Eq. 1. The difference is in the definition of z'. Instead of having z'=f as in Eq 1, z' is now given by Eq.4 which establishes a more complex dependency on f (following the thin lens assumption). Here R is the radius of curvature of the spherical lens, n is index of refraction and z_0 is the distance between the focal point and the film. Eq. 4 can be derived using the Snell's law ($n_1 \sin\alpha_1 = n_2 \sin\alpha_2$) under the assumption that the angle between the direction of the light rays and the lens surface normal is small (in this case $\sin\alpha = \alpha$). This model is called the *paraxial refraction model.* A complete derivation goes beyond the scope of this lecture. More details can be found in [FP] sec 1.1, page 8.



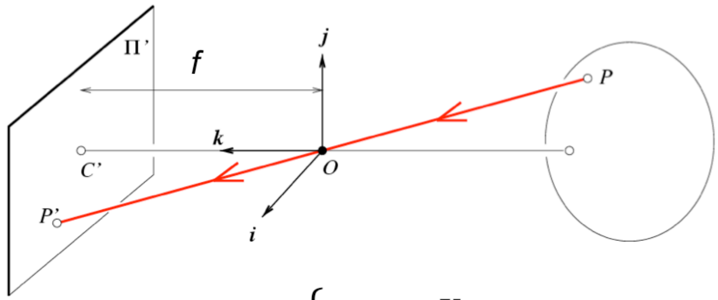
Because the paraxial refraction model is only an approximation (for instance, α can be in general large and $\sin \alpha$ cannot approximated by α), this creates a number of aberrations. The most common one is referred to as “radial distortion”. The radial distortion is caused by the fact that the focal length of a lens is not uniform but different portions of the lens have different focal lengths. The effect of a radial distortion is that the image magnification decreases or increases as function of the distance separating the optical axis from the point of interest p’. If the magnification increases we talk about pin **cushion distortion**; if the magnification decreases we talk about **barrel distortion**. The latter typically occurs when one uses fisheye lens.

Lecture 2

Camera Models

- Pinhole cameras
- Cameras & lenses
- The geometry of pinhole cameras
  - Intrinsic
  - Extrinsic

Pinhole camera

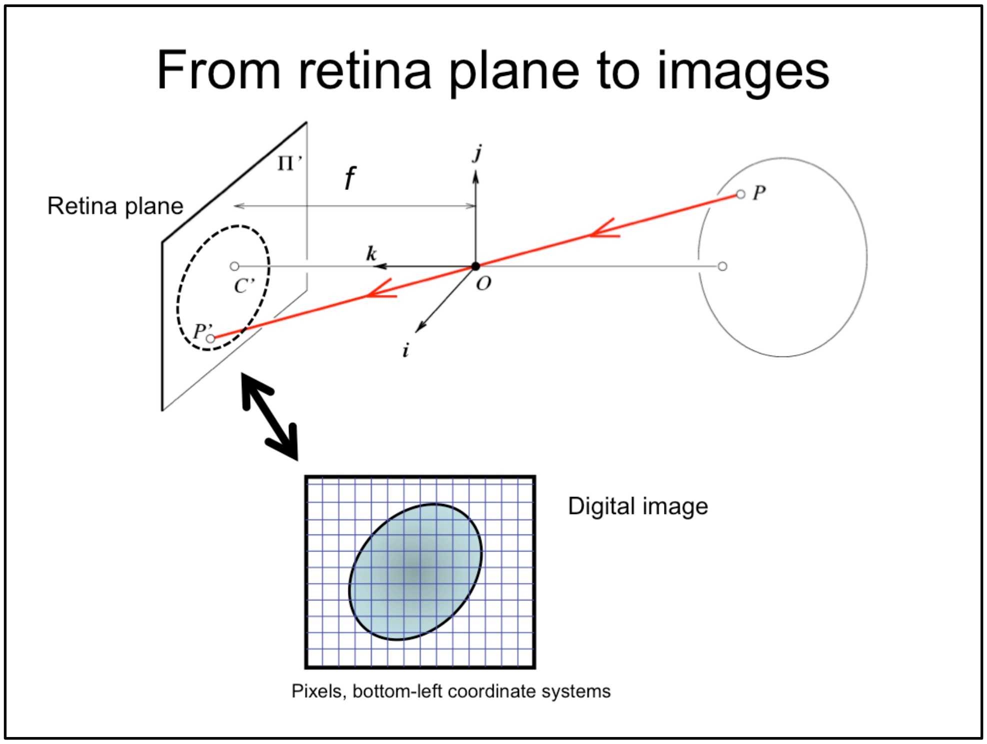


$$P = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \rightarrow P' = \begin{bmatrix} x' \\ y' \end{bmatrix} \qquad \begin{cases} x' = f\dfrac{x}{z} \\[2ex] y' = f\dfrac{y}{z} \end{cases} \qquad \Re^3 \xrightarrow{E} \Re^2$$

[Eq. 1]

f = focal length
o = center of the camera

In the following, we will discuss in details the geometry of the pinhole camera model. All the results derived using this camera model also hold for the paraxial refraction model. As we discussed earlier, in the pinhole camera model, a point P in 3D (in the camera reference system) is mapped (projected) into a point P' in the image plane Π'. This is a R^3 to R^2 mapping as indicated by Eq.1. P' can be expressed as $[x'\ y']^T$. The camera reference system [i,j,k] is centered at O, and the projection of O into the image plane gives C'. This mapping or transformation is also referred to *projective* transformation.



The projection of 3D points into the image plane does not directly correspond to what we see in actual digital images (i.e., what we see in a display or monitor): i) points in the digital display are, in general, in a different reference system than those in the image/retinal plane; ii) Digital images are divided into pixels whereas points in the image/retina plane are measured in centimeters; iii) The physical sensors can introduce non-linearity (e.g. distortion) to the mapping. Next we introduce a number additional transformations that allow to map any 3D point from the 3D world directly into a point in pixel coordinates in the digital image.

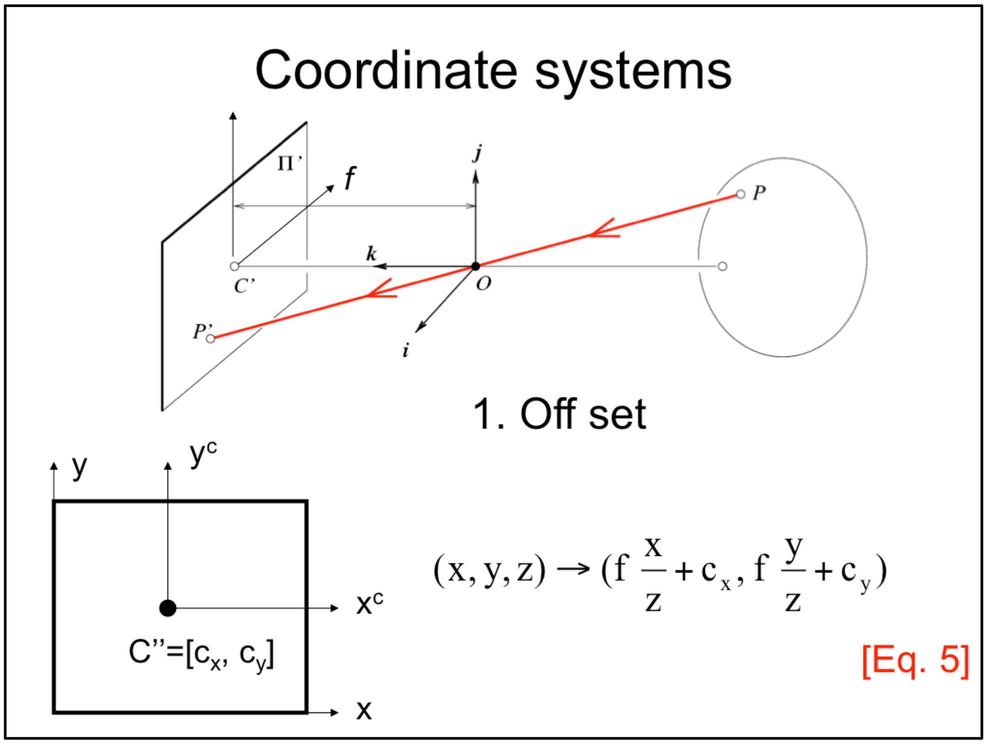
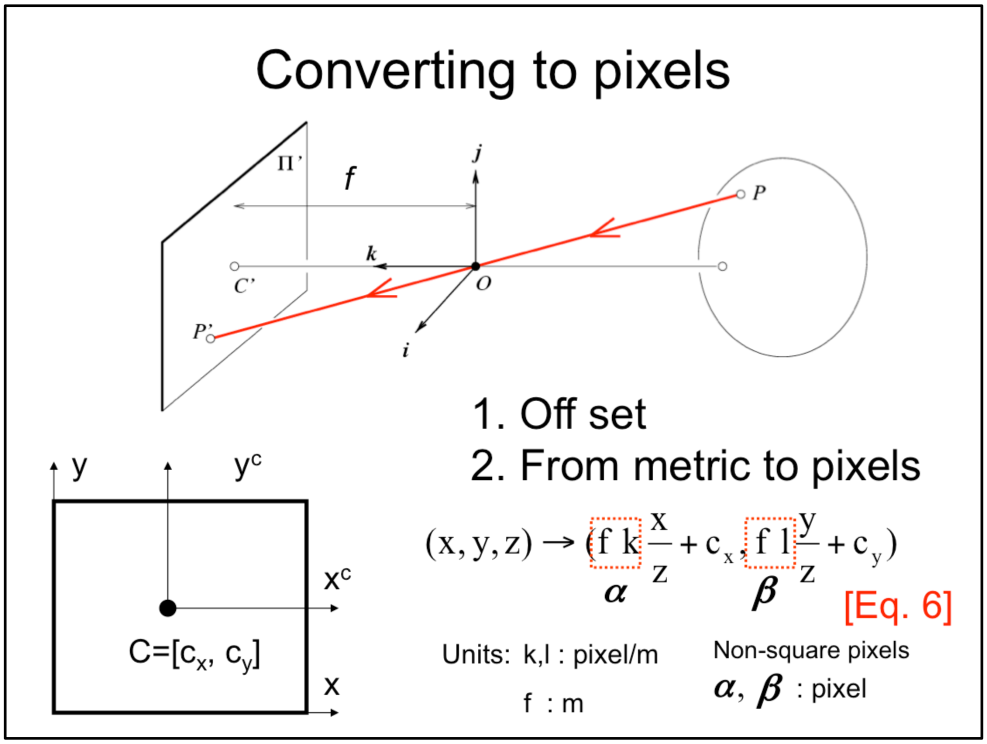
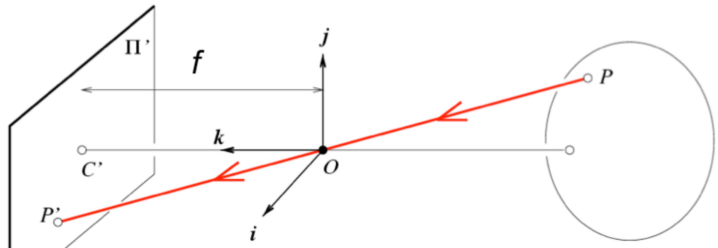


Image coordinates have their origin C' at the image center where the k axis of the camera reference system intersects the image plane. On the other hands, typically, digital images have their origin at the lower-left corner of the image. Thus, 2D points in the image plane and 2D points in a digital image are equal up to an offset translation vector [c_x, c_y], where C''= [c_x, c_y] is the location of C' in the digital image.
Thus, in order to accommodate this change of coordinate metric systems, the mapping is now described by [Eq.5]



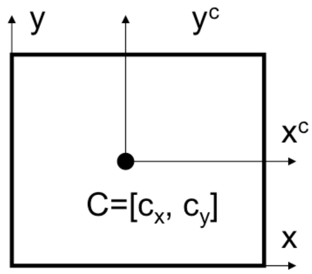
As second effect we need to take into account for is the fact that points in a digital image are expressed in pixels whereas points in the image/retina plane are expressed in (e.g.,) centimeters. In order to accommodate this further change of coordinate systems, the mapping is now described by [Eq.6], where *k* and *l* are parameters expressed in pixel × m^{-1}. Notice that k is different from l because the aspect ratio of the unit element of the sensor (CCD or CMOS) is not necessarily 1 or the pixel aspect ratio is modified by the imaging system.

Is this projective transformation linear?



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

[Eq. 7]



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

Now I'd like to ask a question: is the projective mapping (or transformation) described in Eq. 7 a linear one? What's the definition of linear mapping? Please refer to the CA review session on linear algebra for details or refer to your favorite text book on linear algebra for the definition of linear transformations. The transformation in Eq. 7 is not linear since the x and y coordinates are divided by z the coordinate – a non linear operation.

It is easy to show that a linear transformation (mapping) can always be expressed as a product of a matrix and the input signal $[x\ y\ z]^T$. Is there a way we can express Eq.7 as product of a matrix and the vector $[x\ y\ z]^T$? In general, we can't, because, as we said before, the transformation is not linear. But….

Homogeneous coordinates

E→H

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

homogeneous image coordinates | homogeneous scene coordinates

- Converting back *from* homogeneous coordinates

H→E

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

One way to get around this problem is to change coordinate system and go from the Euclidean reference system (our original coordinate system) to the so called *Homogenous* coordinate system. The process of going back and forth from Euclidean and Homogenous coordinate system is non-linear, but the good news is that once we are in the homogenous coordinate system a projective transformation like the one expressed by Eq. 7 becomes linear. We will discuss this in a few more details next. Please refer to HZ section 2.2 for more details and examples. The change of reference system from Euclidean to Homogenous goes as follows:

E→H: a point X in R^n (in the Euclidean coordinate system) is represented as R^{n+1} vector (in the Homogenous coordinate system) by adding a final coordinate of 1. See slide for examples.

H→E: a generic point X in R^{n+1} (in the Homogenous coordinate system) is represented as a R^n vector (in the Euclidean coordinate system) by suppressing the last coordinate and dividing the first n coordinates of X by the $n+1^{th}$ coordinate. See slide for examples.

The benefit of working with homogenous coordinates will become more evident throughout the upcoming lectures. For now let's go back to Eq. 7 and propose an alternative formulation in the homogenous coordinate systems.

Projective transformation in the homogenous coordinate system

$$P_h' = \begin{bmatrix} \alpha x + c_x z \\ \beta y + c_y z \\ z \end{bmatrix} = \begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad [\text{Eq.8}]$$

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

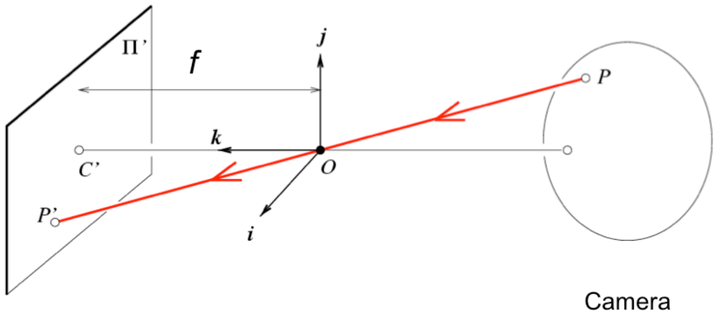
$$\overbrace{P_h'}^{\text{Homogenous}} \rightarrow \overbrace{P' = (\alpha \frac{x}{z} + c_x, \beta \frac{y}{z} + c_y)}^{\text{Euclidian}}$$

A point $P=[x\ y\ z]^T$ in the 3D Euclidean reference system is expressed as $P_h = [x\ y\ z\ 1]^T$ in Homogenous reference system. Let's multiply the matrix M (which contains an arrangement of the camera parameters) by P_h: the result is the vector P'_h as expressed by Eq. 8. Now let's convert P'_h back to the Euclidean reference system. The result is exactly the point P' in Eq. 7. As this process illustrates:

- The projective transformation can be expressed as a linear transformation in the Homogenous reference system
- This transformation is expressed as a multiplication of the matrix M and the vector P_h.
- The conversion from Homogenous to Euclidean returns the actual point P' (that's where the non-linearity takes place).

The key advantage of this representation is that we have obtained a much more compact expression for capturing the projective transformation. As we'll see next, as we consider additional linear transformations to characterize the camera model, these can be nicely and compactly added as matrix multiplications to the expression in Eq. 8.

The Camera Matrix



[Eq.9]

$$P' = M\,P$$

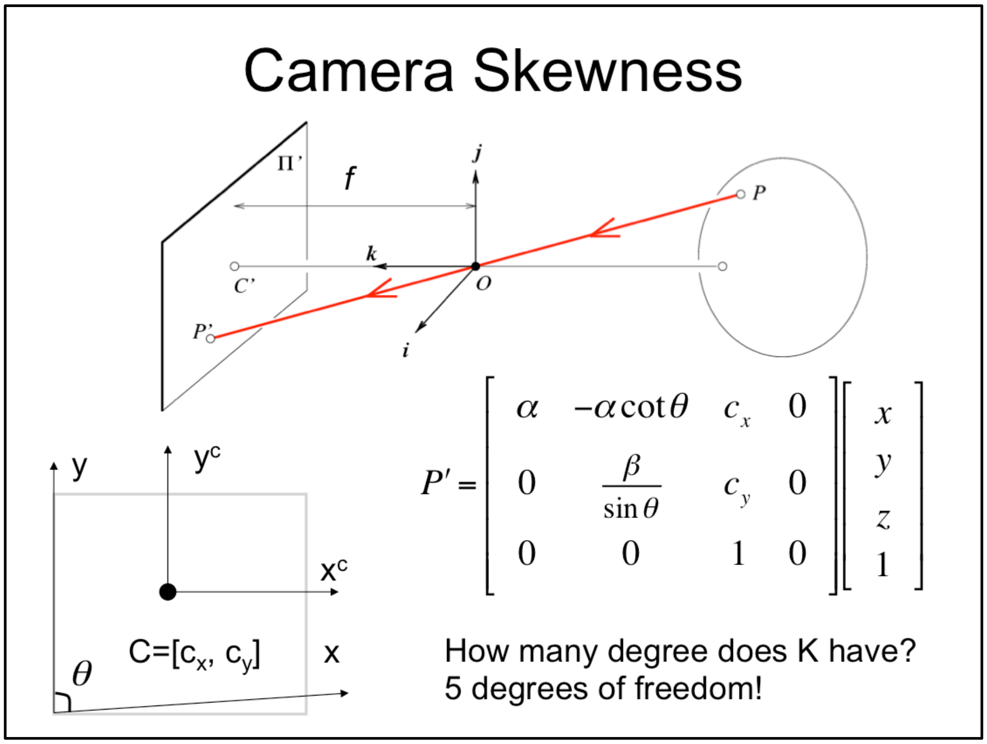
$$= K\begin{bmatrix} I & 0 \end{bmatrix} P$$

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

Camera matrix K

From this point on, let's assume that we work on homogenous coordinates and that a generic point P in 3D is expressed in such coordinate system. To keep the notation light, we drop the index h (we'll add the index back when the discussion becomes ambiguous).

By working In homogenous coordinates, we can write a very compact expression that regulates the projective transformation (see Eq. 9), where M is defined in the previous slide, I is a 3x3 identity matrix and and K is defined as the *camera matrix*. This matrix contains some of the critical parameters that are useful to characterize a camera model. Two parameters are missing: *skewness* and *distortion.* We will talk about skewness next and distortion during the next lecture.



We say that an image is skewed when the camera coordinate system is skewed; that is, the two image axes are not perpendicular (as they should) and the angle θ between the two axes is slightly larger or smaller than 90 degrees. While most of the cameras have zero-skew, some degree of skewness may occur because of sensor manufacturing errors or when the image is a picture of another picture (a photograph that has been "re-photographed"); In some other cases, pixels can potentially be skewed when the image is acquired by a frame grabber. In this particular case, the pixel grid may be skewed due to an inaccurate synchronization of the pixel-sampling process. Deriving an accurate model of the skewness goes beyond the scope of this lecture and we just report here the final expression of the camera matrix. For more details refer to sec. 1.2.2 [FP] or sec. 6.2.4 [HZ].

Given all the parameters we have introduced so far, how many degree of freedom does K have? 2 for focal length, 2 for off-set, 1 for skewness. So 5 in total.

Canonical Projective Transformation

[Eq.10]

$$P' = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\text{M}} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$P' = M\,P$$

$$\Re^4 \xrightarrow{\text{H}} \Re^3$$

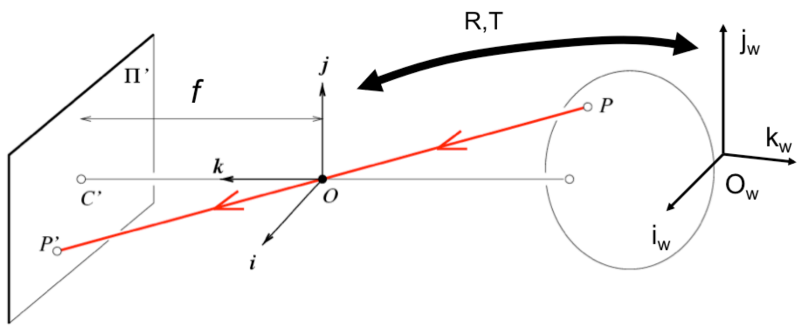
$$P_i' = \begin{bmatrix} \frac{x}{z} \\ \frac{y}{z} \end{bmatrix}$$

Before we move forward to the next topic, we introduce in Eq. 10 the *canonical form* of the projective transformation. This form captures the essence of a projective transformation and assumes that the focal length is 1, there is no offset and zero-skew.

Lecture 2
Camera Models

- Pinhole cameras
- Cameras & lenses
- The geometry of pinhole cameras
  - Intrinsic
  - Extrinsic
- Other camera models

World reference system



•The mapping so far is defined within the camera reference system

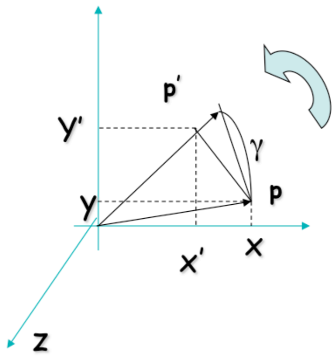
•What if an object is represented in the world reference system?

•Need to introduce an additional mapping from world ref system to camera ref system

So we far we have described a mapping between a point P in the 3D camera reference system to a point P' in the 2D image plane. But what if the information we have about points in 3D (or in general about any physical object) is available in the world coordinate system?
Then, we need to include to our mapping an additional transformation that allows to relate points from/to the world reference system to/from the camera reference system. This transformation is captured by a rotation matrix R and a translation vector T. Next we'll review the basics of rigid transformations.

3D Rotation of Points

Rotation around the coordinate axes, counter-clockwise:



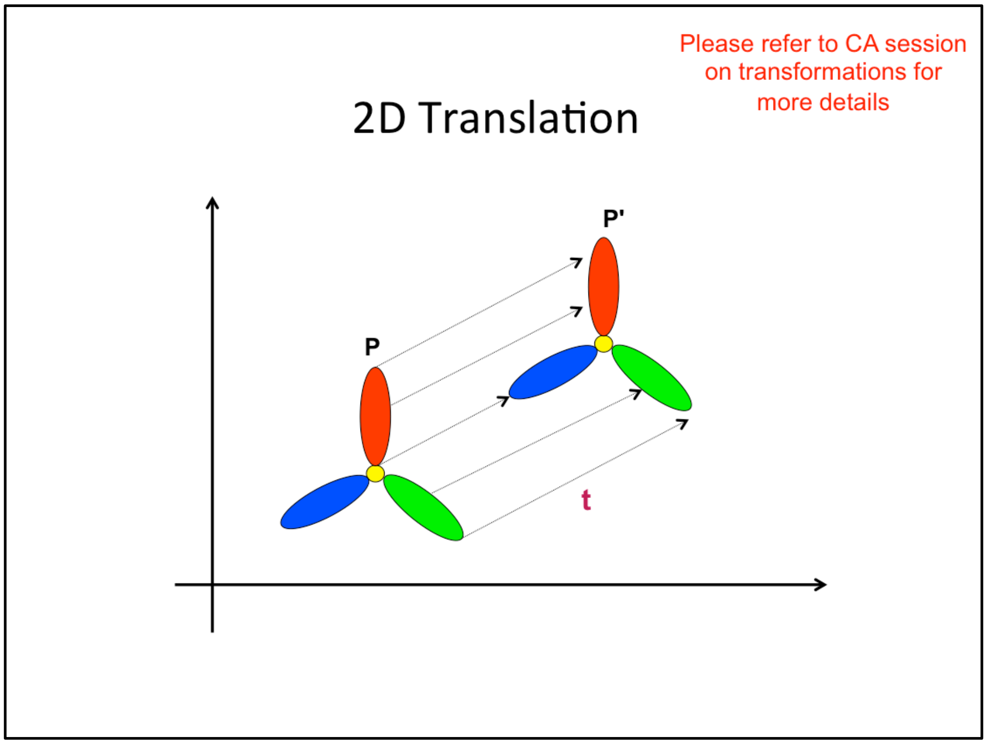
$$R_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}$$

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

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

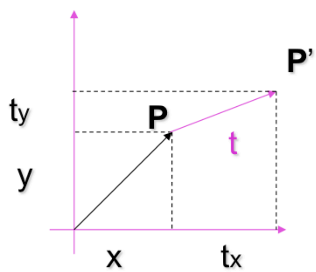
$$P' \rightarrow \begin{bmatrix} R & 0 \\ 0 & 1 \end{bmatrix}_{4\times4} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

A rotation matrix in 3D has 3 degree of freedom



Next I will briefly review the generic concepts of transformations in 2D and 3D. For details please refer to the CA session held on Friday Jan 9th.

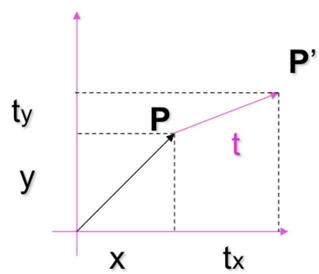
2D Translation Equation



$$\mathbf{P} = (x, y)$$
$$\mathbf{t} = (t_x, t_y)$$

$$\mathbf{P}' = \mathbf{P} + \mathbf{t} = (\mathrm{x} + \mathrm{t}_{\mathrm{x}}, \mathrm{y} + \mathrm{t}_{\mathrm{y}})$$

2D Translation using Homogeneous Coordinates

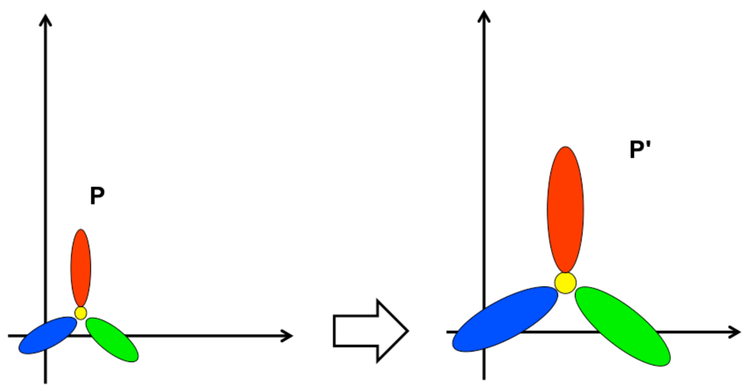


$$\mathbf{P} = (x, y) \rightarrow (x, y, 1)$$

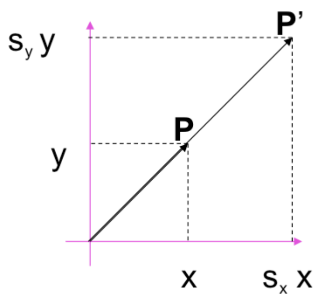
$$\mathbf{P}' \rightarrow \begin{bmatrix} x + t_x \\ y + t_y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{I} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{T} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Scaling



Scaling Equation

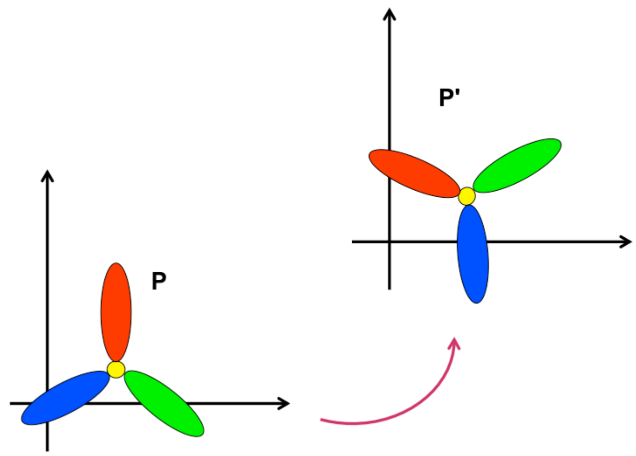


$$\mathbf{P} = (x, y) \rightarrow \mathbf{P}' = (s_x x, s_y y)$$

$$\mathbf{P} = (x, y) \rightarrow (x, y, 1)$$

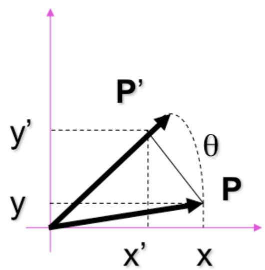
$$\mathbf{P}' \rightarrow \begin{bmatrix} s_x x \\ s_y y \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{S}} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{S}' & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{S} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Rotation



Rotation Equations

- Counter-clockwise rotation by an angle θ



$$x' = \cos\theta\, x - \sin\theta\, y$$

$$y' = \cos\theta\, y + \sin\theta\, x$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \qquad \mathbf{P'} = \mathbf{R}\ \mathbf{P}$$

How many degrees of freedom? 1

$$\mathbf{P'} \rightarrow \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Scale + Rotation + Translation

$$\mathbf{P}' \rightarrow \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \cos\theta & -\sin\theta & t_x \\ \sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \boxed{\begin{bmatrix} \mathbf{R}\,\mathbf{S} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

If $s_x = s_y$, this is a similarity transformation

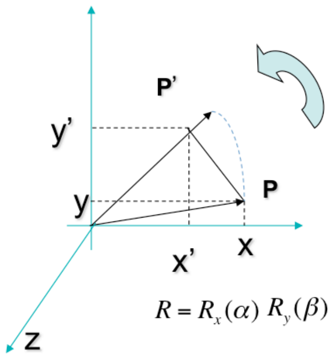
3D Rotation of Points

Rotation around the coordinate axes, counter-clockwise:

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

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

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

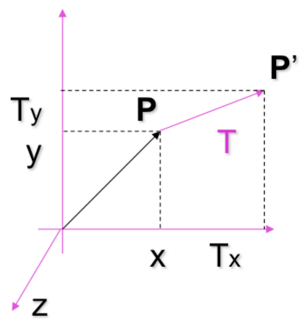


$$R = R_x(\alpha)\, R_y(\beta)\, R_z(\gamma)$$

$$P' \rightarrow \begin{bmatrix} R & 0 \\ 0 & 1 \end{bmatrix}_{4\times4} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

A rotation matrix in 3D has 3 degrees of freedom

3D Translation of Points



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

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

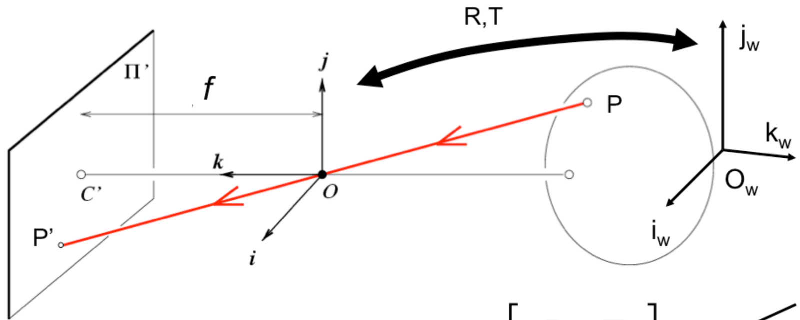
A translation vector in 3D has 3 degrees of freedom

3D Translation and Rotation

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

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

World reference system



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

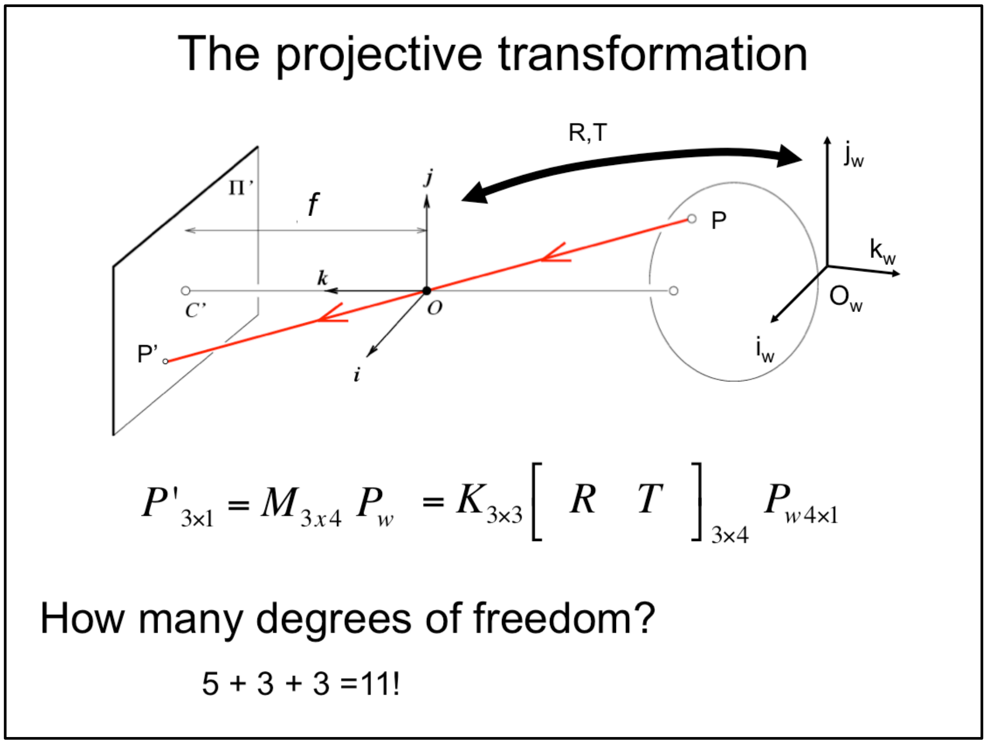
Internal parameters

External parameters

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

M [Eq.11]

The transformation that allows to relate points from/to the world reference system to/from the camera reference system is a rigid transformation and is regulated by a translation R and a rotation T in 3D. The benefit of working in a homogeneous coordinates is that these two transformations can be compactly described by the matrix [R T; O 1] as the slide illustrates. This is a 4x4 matrix and allows to map any point $P_w = [x_w\ y_w\ z_w\ 1]^T$ in the world reference system to a point $P = [x\ y\ z\ 1]^T$ in the camera reference system (in homogenous coordinates). The projective mapping from P_w to P' = [x' y' z'] is expressed by Eq. 11 and is compactly summarized as P'=M P. The projection matrix M = K [R T], where K are called internal or intrinsic parameter and [R T] are called the external or extrinsic parameters; the former are related to the intrinsic physical properties of the camera (e.g. focal length, etc…) where the latter are related to pose and location of the camera; Notice that these two set of parameters are nicely separated in this formulation.



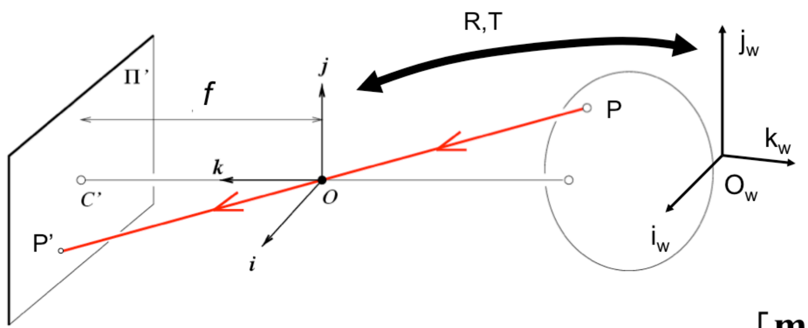
We have rewritten here Eq. 11 again to highlight the dimensionality of the matrices and vectors; For instance, M_{3x4} means that the matrix M has 3 rows and 4 columns.
How many degrees of freedom do M have?

- 5 from K (as we have seen earlier).
- 3 from R (a rotation matrix in 3D has 3 degree of freedom).
- 3 from the translation vector T.

In total 11!

Eq. 11 describes the projective transformation and M is a projective matrix.

The projective transformation



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

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

The mapping between P_w and P' is in homogenous coordinates. If we want to obtain P' in Euclidean coordinates (which corresponds to our actual measurements in the image), we need to change coordinates from homogenous to Euclidean by dropping the last coordinate of P' and dividing the first two coordinates by the last coordinate. This operation is illustrated in Eq. 12. Here $\mathbf{m}_1$, $\mathbf{m}_2$ and $\mathbf{m}_3$ are the first, second and third row vectors of M, respectively. Thus, each of the $\mathbf{m}_i$ is 1x4 vector. As Eq. 12 shows, the final outcome of the most generic formulation of a projective transformation in Euclidean coordinates is a rather complex (and non linear) intermix of the intrinsic and extrinsic parameters. However, by deriving this mapping in the homogenous coordinates, we managed to have a good understanding of the geometrical properties of the problem where intrinsic and extrinsic parameters were separated.

Theorem (Faugeras, 1993)

$$M = K\begin{bmatrix} R & T \end{bmatrix} = \begin{bmatrix} KR & KT \end{bmatrix} = \begin{bmatrix} A & b \end{bmatrix} \qquad A = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix}$$

[Eq.13]

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

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

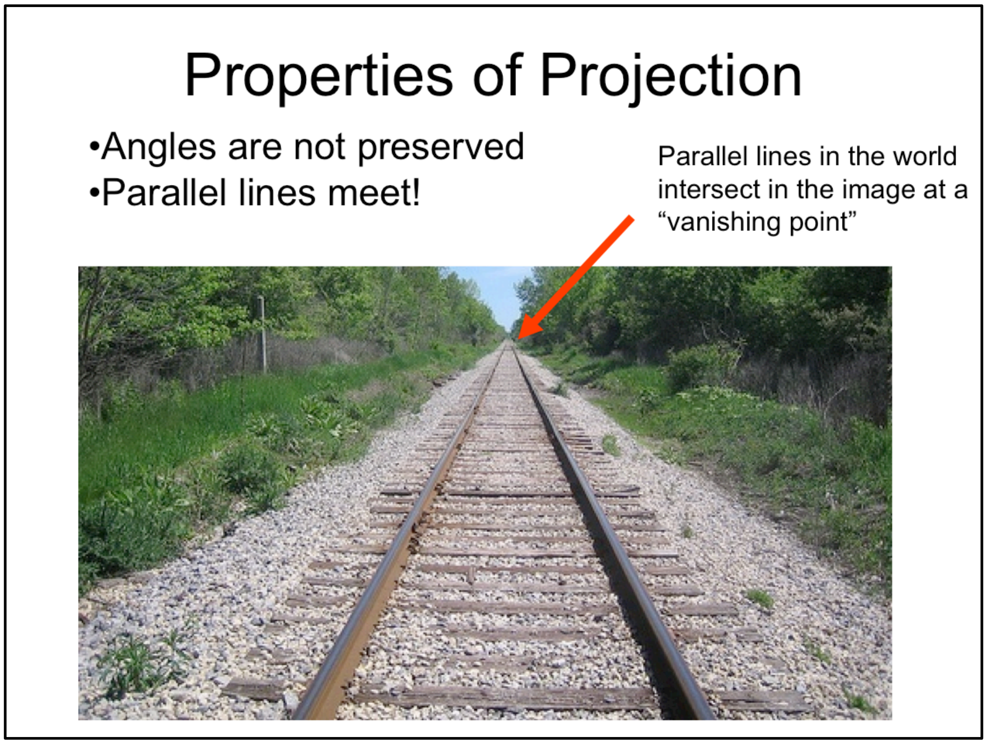
In 1993 Olivier Faugeras derived a powerful theorem that relates the properties of a generic projective transformation M with the properties of the matrix A_{3x3} as defined in Eq. 13. A is defined as K R and $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ are the first, second and third row vectors of A, respectively. These relationships are summarized in the 3 necessary and sufficient conditions illustrated in the slide.

Properties of projective transformations

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

Other more intuitive properties can be pointed out about projective transformations:

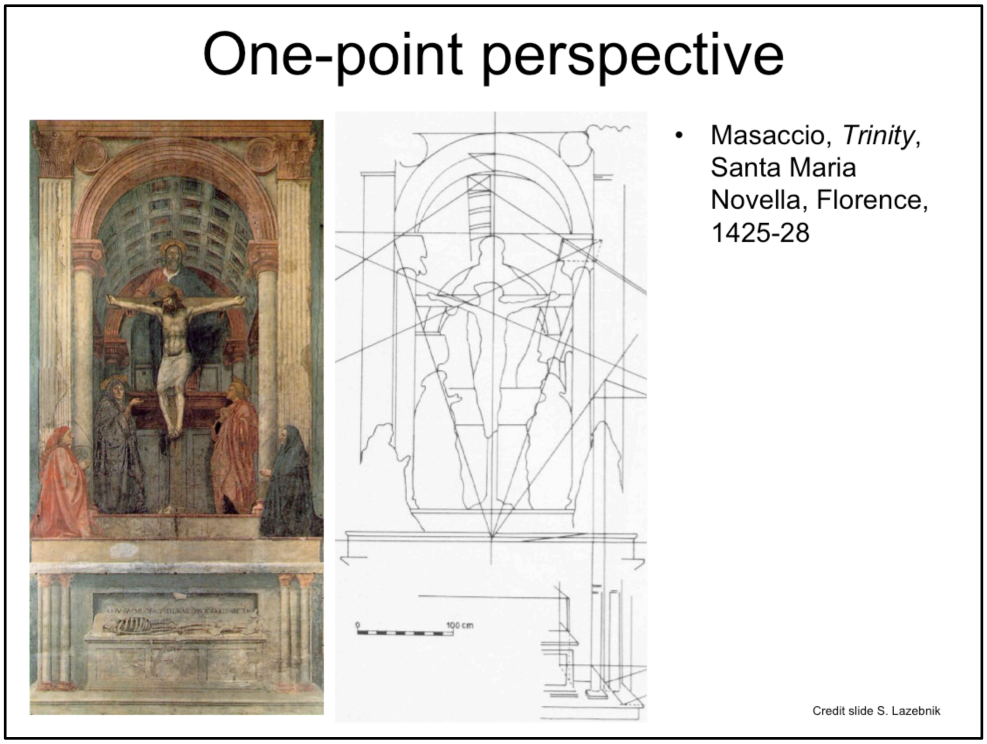
- A point is still a point after a projective transformation
- A line is still a line after a projective transformation
- Distant objects look smaller; Can you explain why?



- Two parallel lines in 3D are no longer necessarily parallel after a projective transformation; that is, in general parallel lines in 3D intersect at one point after a projective transformation. As we will discuss in more details in lecture 4, this point is called *vanishing point*. An example of vanishing point is illustrated above.

Horizon line (vanishing line)

- By considering multiple pairs of parallel lines on the same plane (the ground plane in this example), we obtain a set of vanishing points which have the property of being collinear. The line that passes through these vanishing points is called horizon.



The concept of central projective transformation (or one-point perspective) was first studied during the Renaissance and the slide shows a fine example by Masaccio.

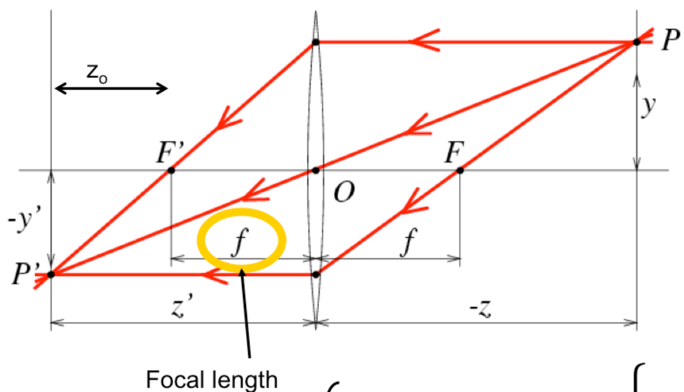
Next lecture

- How to calibrate a camera?

Supplemental material

Thin Lenses

[FP] sec 1.1, page 8.



$$z' = f + z_o$$

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

Snell's law:

$n_1 \sin \alpha_1 = n_2 \sin \alpha_2$ ⇨

Small angles:

$n_1 \alpha_1 \approx n_2 \alpha_2$

n_1 = n (lens)

n_1 = 1 (air)

⇨

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

n_i = index of refraction

R=radius of the lens

Horizon line (vanishing line)