

# MEAM 620 - Sensing

Spring 2020

C.J. Taylor

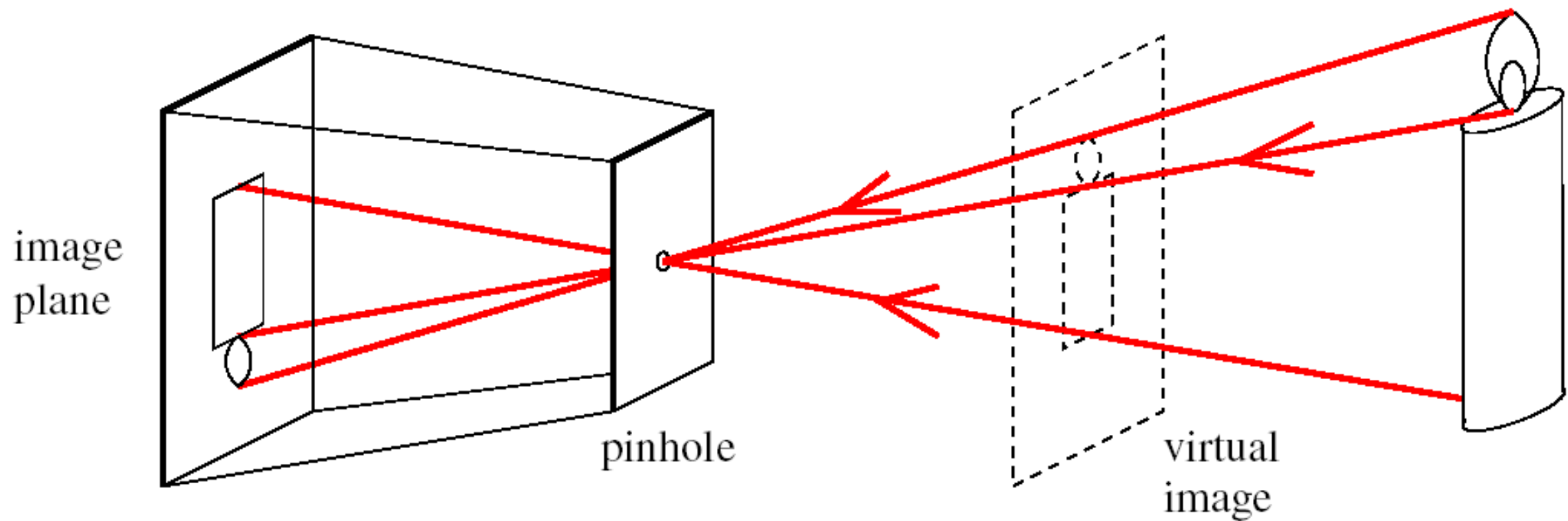
# How Cameras Work

# Cameras - a brief history

- From the latin *Camera Obscura* - Dark Chamber
- The cameras of the 11th century were literally dark rooms with a small opening on one side

# The Pinhole Camera

- Light enters a darkened chamber through a pinhole opening and forms an image on the further surface



# The invention of the lens

- The first major improvement occurred in the 16th century when Giovanni Battista della Porta added a lens to the design
- The lens improved the light gathering power of the device so brighter images were possible
- The problem with being ahead of your time...
  - ▼ Unfortunately Signor della Porta's contribution was not universally appreciated at first. He was called before the papal court on charges of sorcery

# Forming images

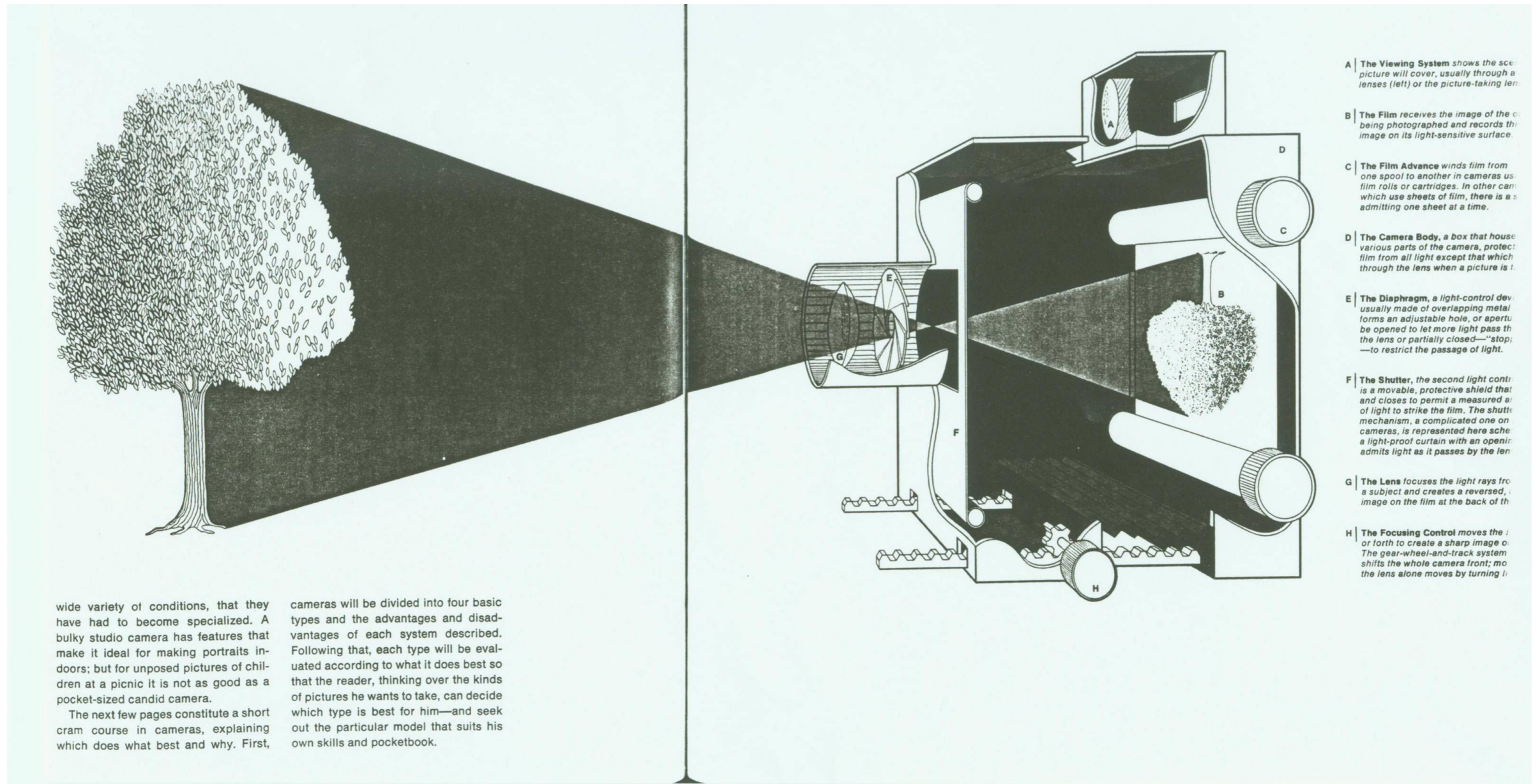
- The first cameras were used by painters who would literally sit in the box and trace the outline of the image on their canvases
- Film
  - ▼ The first film was developed by Niepce (1822) who took the worlds first photograph of a farmyard in central France – it required an exposure time of eight hours.
  - ▼ He joined forces with Louis Jacques Mande Daguerre who improved the process. His exposures only required half an hour and were quite sharp.

The Worlds First  
Known Photo?



# The anatomy of a modern camera

- Lens, Shutter, Diaphragm, Film, Focusing Control





# Optics

- Lenses are used to increase the light gathering power of the instrument
- The thin lens equation

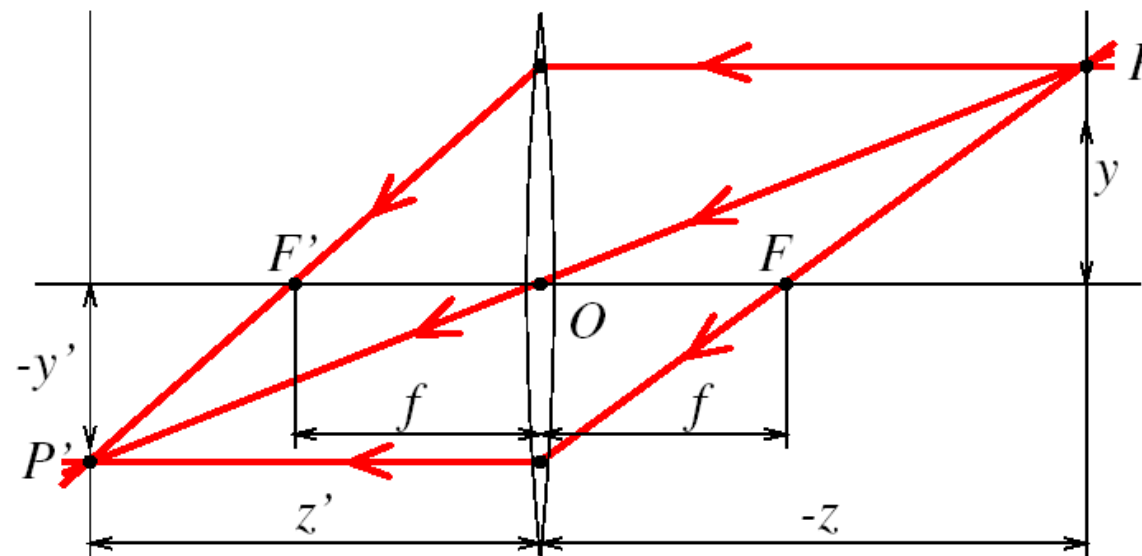


Figure 1.12. The focal points of a thin lens.

$$\frac{1}{z'} - \frac{1}{z} = \frac{1}{f}$$



# Fundamental Equation

- The fundamental equation relates scene radiance  $L$  to image irradiance
- Points to note
  - ▼  $E$  linearly related to  $L$
  - ▼ f-number ( $d/z'$ )
  - ▼  $\cos^4 \alpha$  term

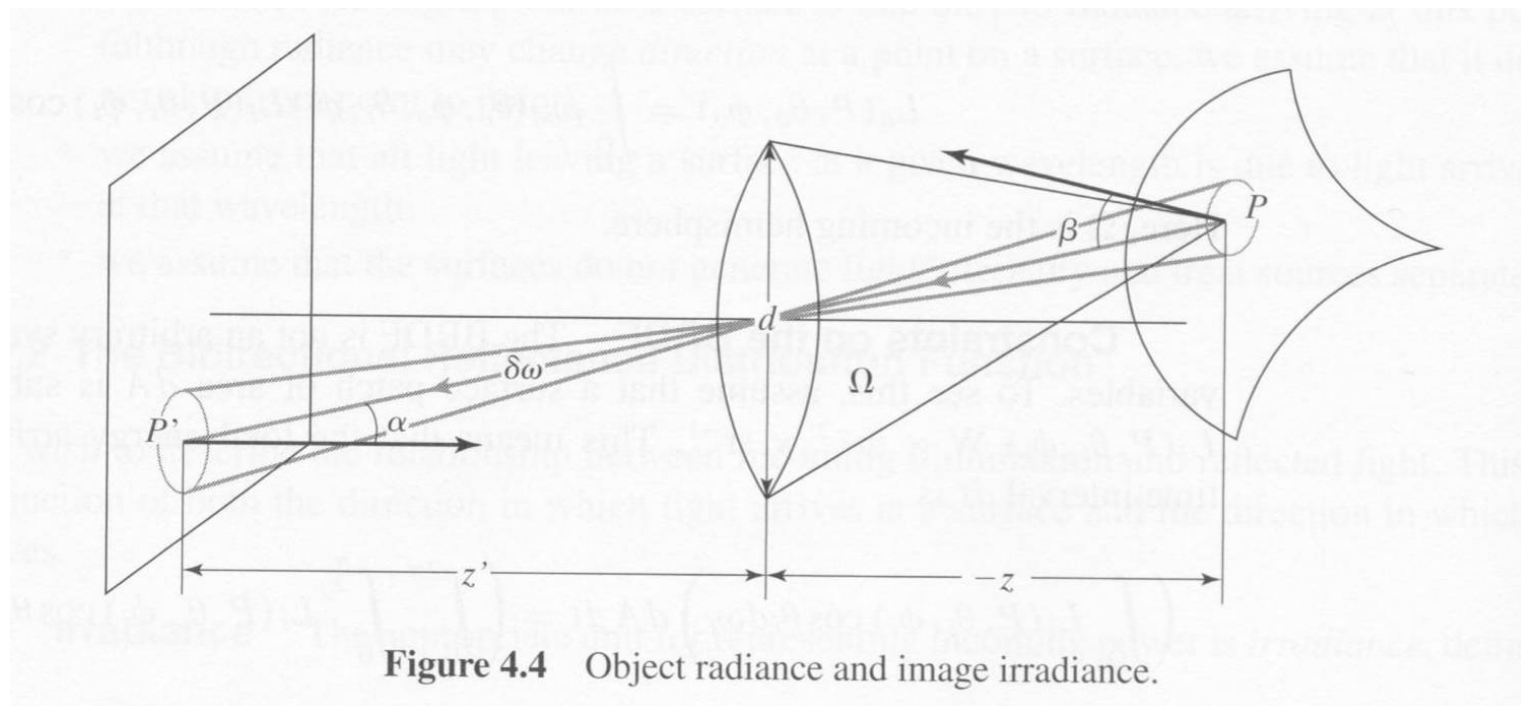
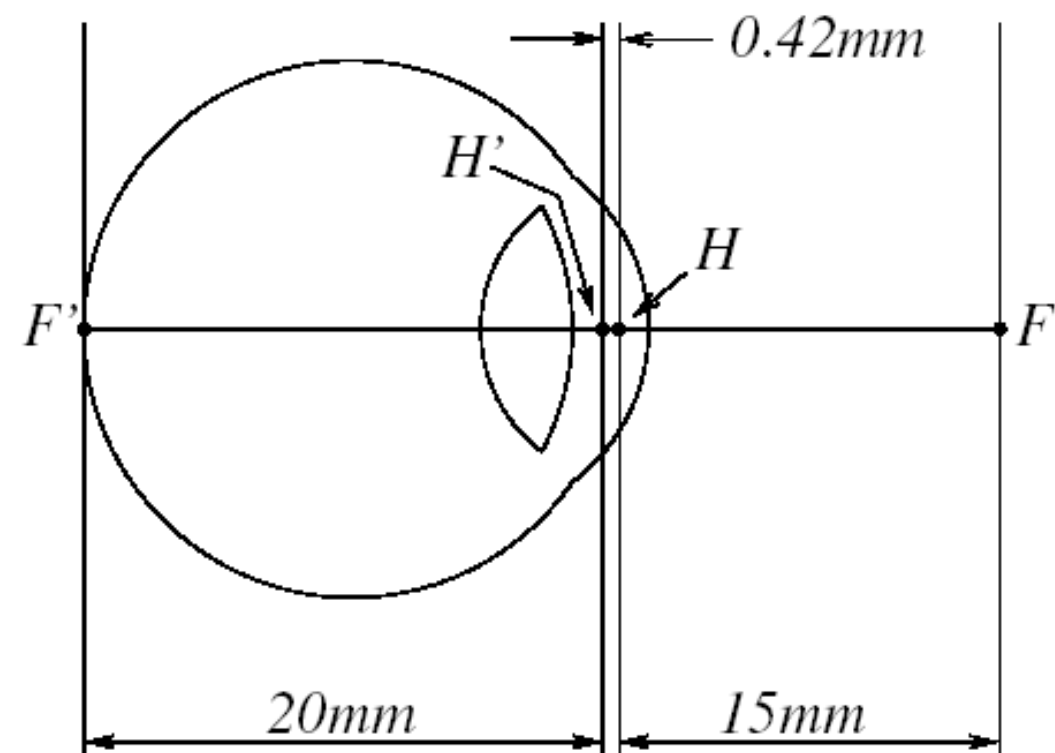
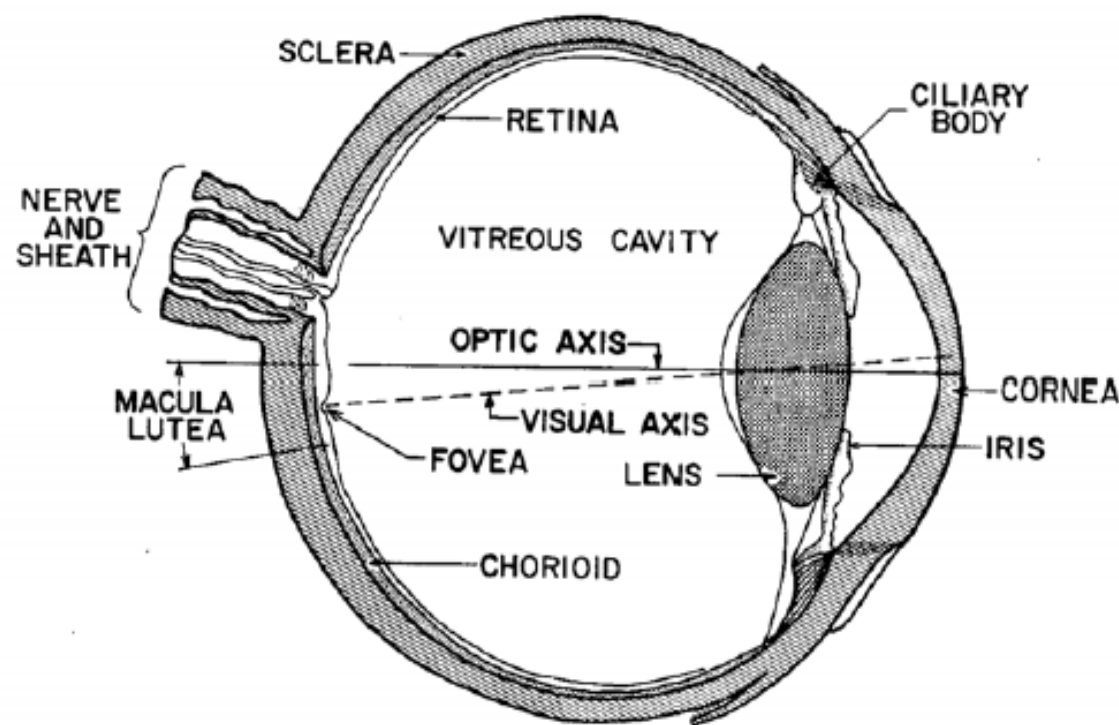


Figure 4.4 Object radiance and image irradiance.

$$E = \left( \left( \frac{\pi}{4} \right) \left( \frac{d}{z'} \right)^2 \cos^4 \alpha \right) L$$

# The Camera of the Mind

- Components of the human eye
  - ▼ Pupil, Lens, ciliary muscle, retina, fovea, blind spot
- Note that unlike a camera, the eye does not have a shutter



# The Retina

- There are two types of photosensitive cells in the retina, rods and cones
- Cones come in three flavors which exhibit different sensitivities to different wavelengths of light, red green and blue.
- Rods are not sensitive to variations in wavelength but they are more sensitive than cones and can pick up much dimmer light
- The fovea is populated entirely by cones.

# More Cells

- Ganglion Cells

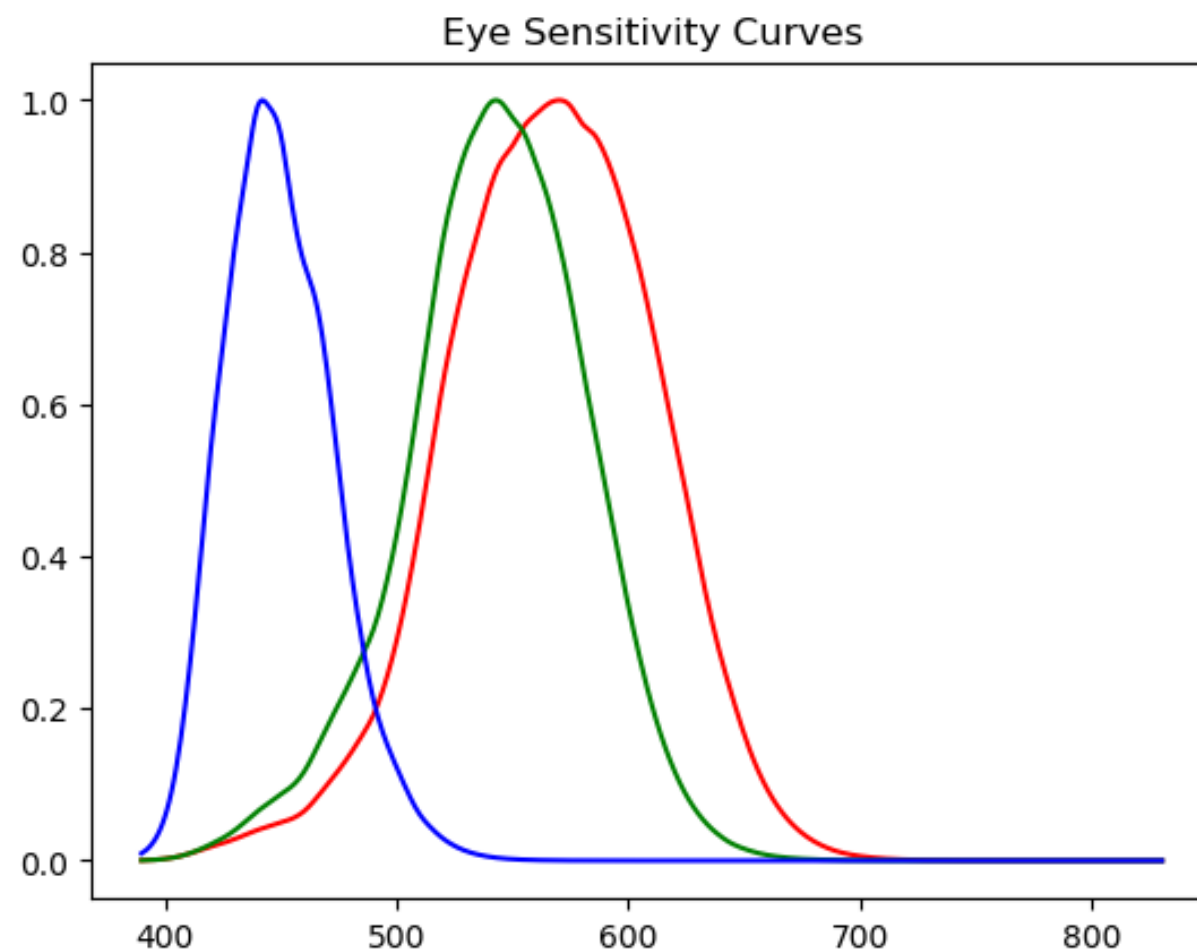
- ▼ The photosensitive cells transmit their information to ganglion cells which in turn transmit information to the brain via the optic nerve

- Numbers of cells

- ▼ There are approximately 6 million cone cells, 120 million rods and 1 million optic nerve fibers.

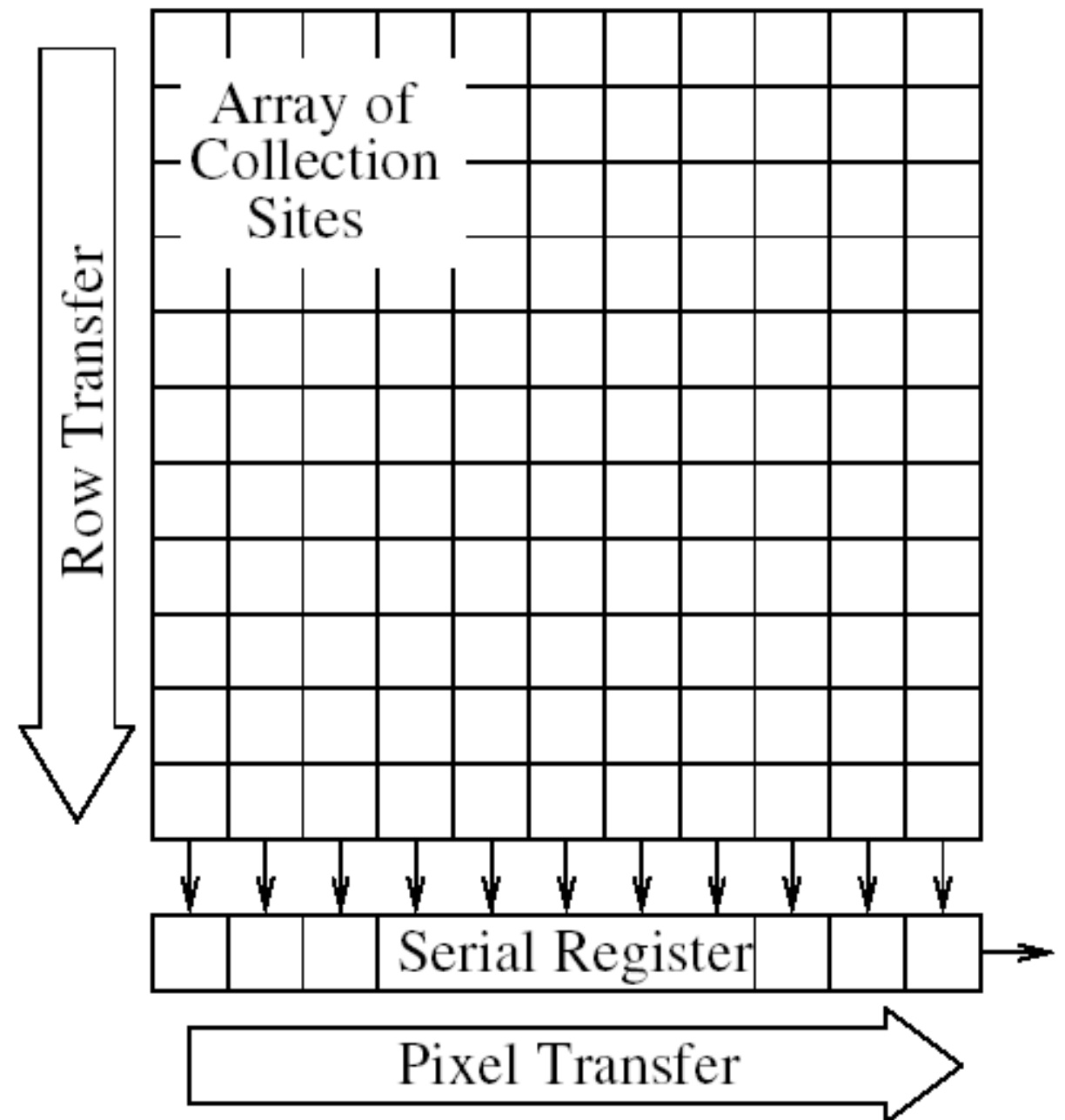
# Seeing Color

- The trichromacy theory of human due to Young (1802) and Helmholtz (1851) states the perception associated with all of the colors in the rainbow can be produced by combinations of three lights (RGB)
- Humans appear to measure the *integrals* of the incident light power spectrum with the tristimulus curves.
- Colors that appear indistinguishable to a human are said to be metameric



# Image Sensors

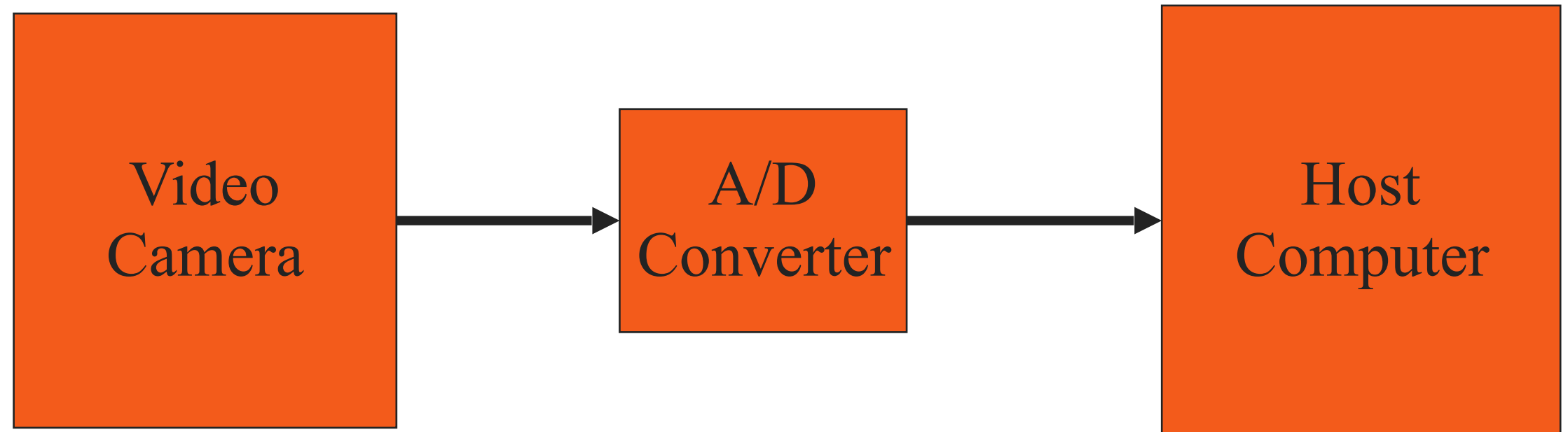
- Images are formed by the interaction of the incident image irradiance with light sensitive elements on the image plane
- Light sensitive elements
  - ▼ Film
  - ▼ Charge Coupled Device (CCD)
  - ▼ CMOS Imaging element



**Figure 1.22.** A CCD Device.

# Digital Imaging Systems

- CCD or CMOS imaging array
  - ▼ When light falls on the cells in these arrays a charge accumulates which is proportional to the incident light energy
- Analog to digital conversion unit
- Host Computer





# Digital Snapshots

- A digital image is an array of numbers indicating the image irradiance at various points on the image plane
- Image intensities are spatially sampled
  - ▼ The Image irradiance function across the retinal plane is sampled to obtain the digital image
  - ▼ The spacing of the image elements limits the resolution of the image
  - ▼ The frequency content of the irradiance function is limited by the effective aperture of the camera
- Intensity values are quantized (8-bits, 10-bits, 12-bits)
- Video Imagery
  - ▼ For a video camera, Images are taken sequentially by opening and closing the shutter 30 times per second

# Sensing Color

- In a 3 CCD video camera the light path is split into three components which are passed through colored light filters and then imaged
- As a result - a color image contains three channels of information; red, green and blue image intensities
- In a 1-CCD color camera color information is obtained by covering the individual elements with a spatially varying pattern of filters, RGB

G1	R2	G3	R4	G5
B6	G7	B8	G9	B10
G11	R12	G13	R14	G15
B16	G17	B18	G19	B20
G21	R22	G23	R24	G25

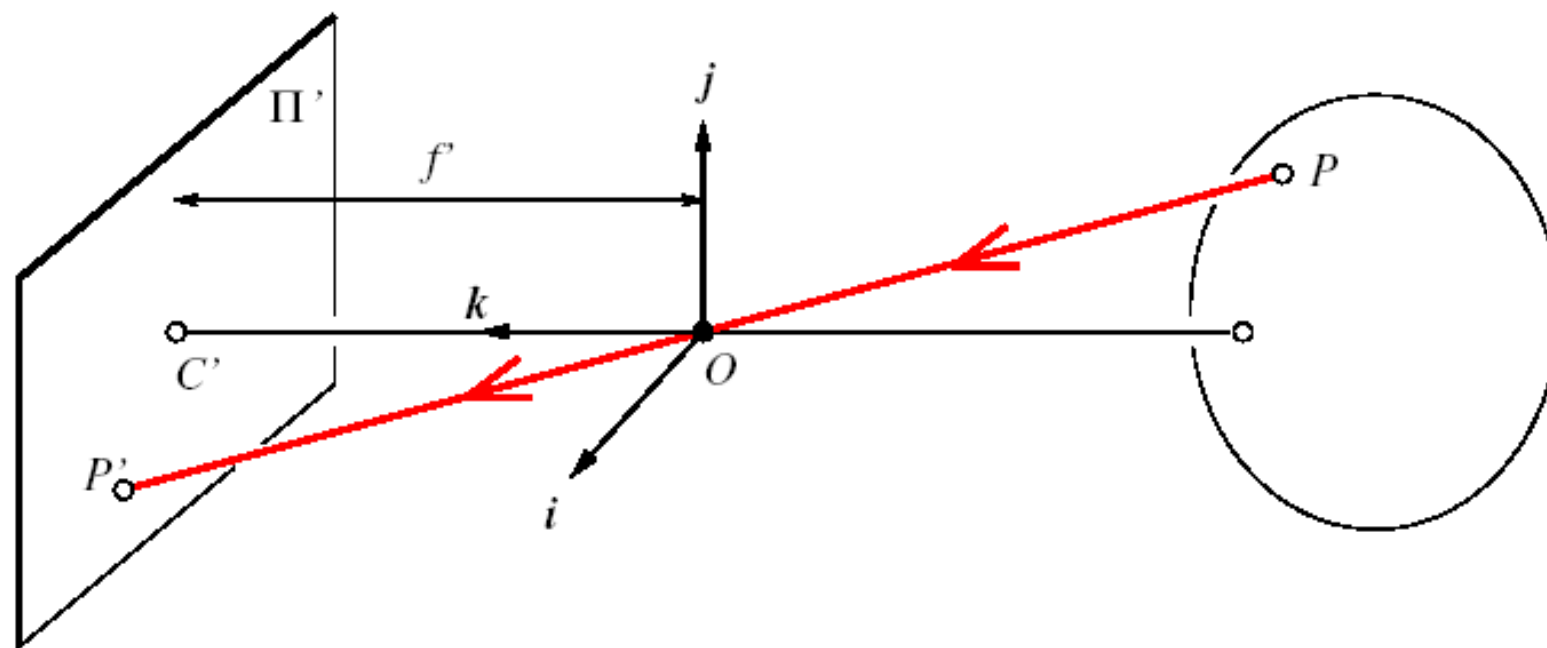
Bayer pattern used to capture color images on a single imaging surface

# Representing Color

- RGB
- CMYK
- CIE
- CIE LAB
- HSV

# Basic Camera Modeling

# The Pinhole Camera Model



**Figure 1.4.** The perspective projection equations are derived in this section from the colinearity of the point  $P$ , its image  $P'$  and the pinhole  $O$ .

Let  $P$  denote a scene point with coordinates  $(x, y, z)$  and  $P'$  denote its image with coordinates  $(x', y', z')$ . Since  $P'$  lies in the image plane, we have  $z' = f'$ . Since the three points  $P$ ,  $O$  and  $P'$  are colinear, we have  $\overrightarrow{OP'} = \lambda \overrightarrow{OP}$  for some number  $\lambda$ , so

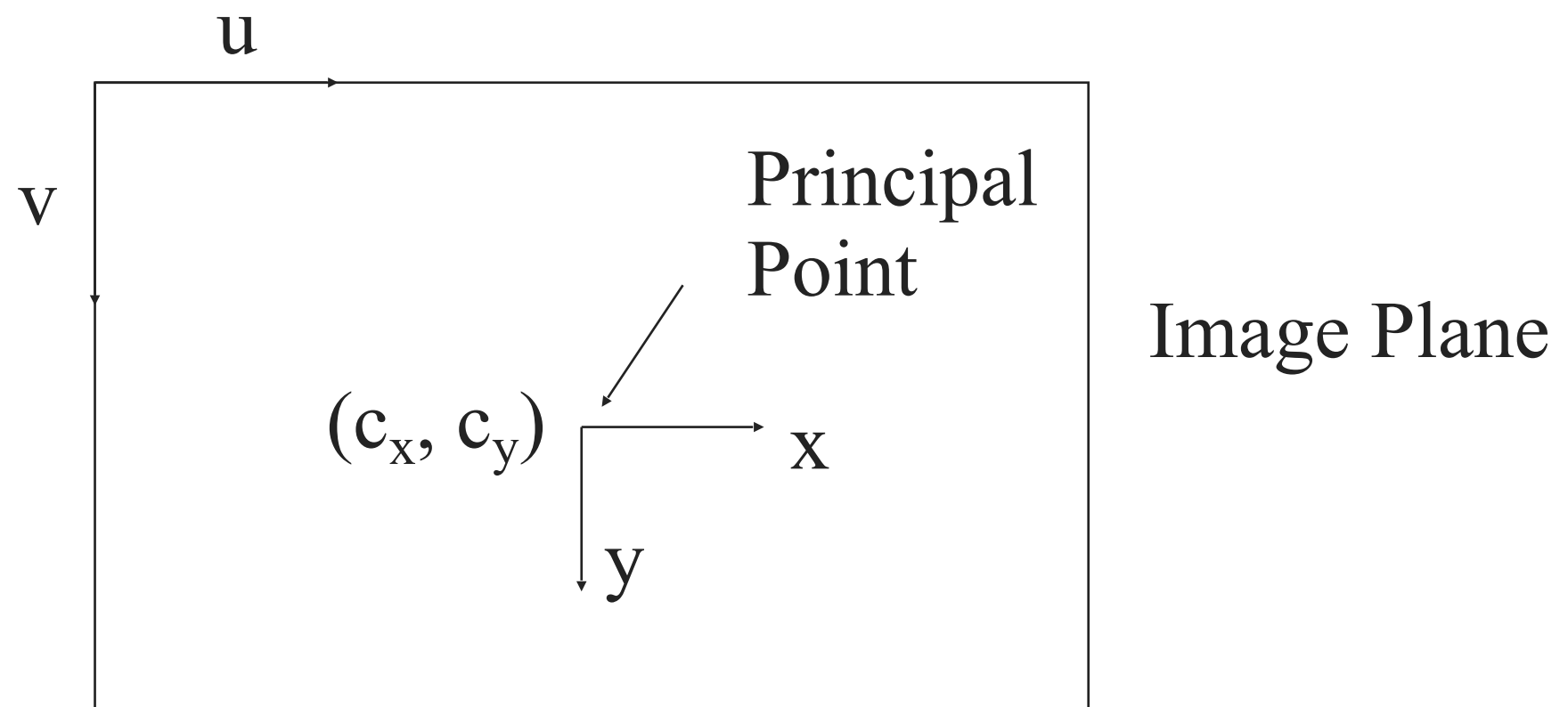
$$\begin{cases} x' = \lambda x \\ y' = \lambda y \\ f' = \lambda z \end{cases} \iff \lambda = \frac{x'}{x} = \frac{y'}{y} = \frac{f'}{z},$$

and therefore

$$\begin{cases} x' = f' \frac{x}{z}, \\ y' = f' \frac{y}{z}. \end{cases} \quad (1.1.1)$$

# Coordinate Systems

- Camera frame
- Image plane coordinates
- Pixel coordinates



# Pixel Coords to Image Plane

- There is an affine relationship between image plane coordinates (x,y) and pixel coordinates (u, v)

$$x = (u - c_x) s_x$$

$$y = (v - c_y) s_y$$



# Image Plane to Pixels

- There is an affine relationship between image plane coordinates (x,y) and pixel coordinates (u, v) which can easily be inverted

$$u = \frac{x}{s_x} + c_x = \left( \frac{f}{s_x} \right) \frac{X}{Z} + c_x$$
$$v = \frac{y}{s_y} + c_y = \left( \frac{f}{s_y} \right) \frac{Y}{Z} + c_y$$

# Issues with real lenses

- Depth of field
- Focus
- Lens aberrations
  - ▼ Chromatic aberration
  - ▼ spherical aberration

# Radial Distortion

- Note how straight lines are distorted into curves by radial distortion in this image



# Radial Distortion cont'd

- Corresponds to a dilation of the image
- It is most pronounced in optical systems with a wide field of view

$$x = x_d (1 + k_1 r^2 + k_2 r^4)$$

$$y = y_d (1 + k_1 r^2 + k_2 r^4)$$

$$r^2 = x_d^2 + y_d^2$$

# Homogenous Coordinates Revisited

# Homogenous Coordinates

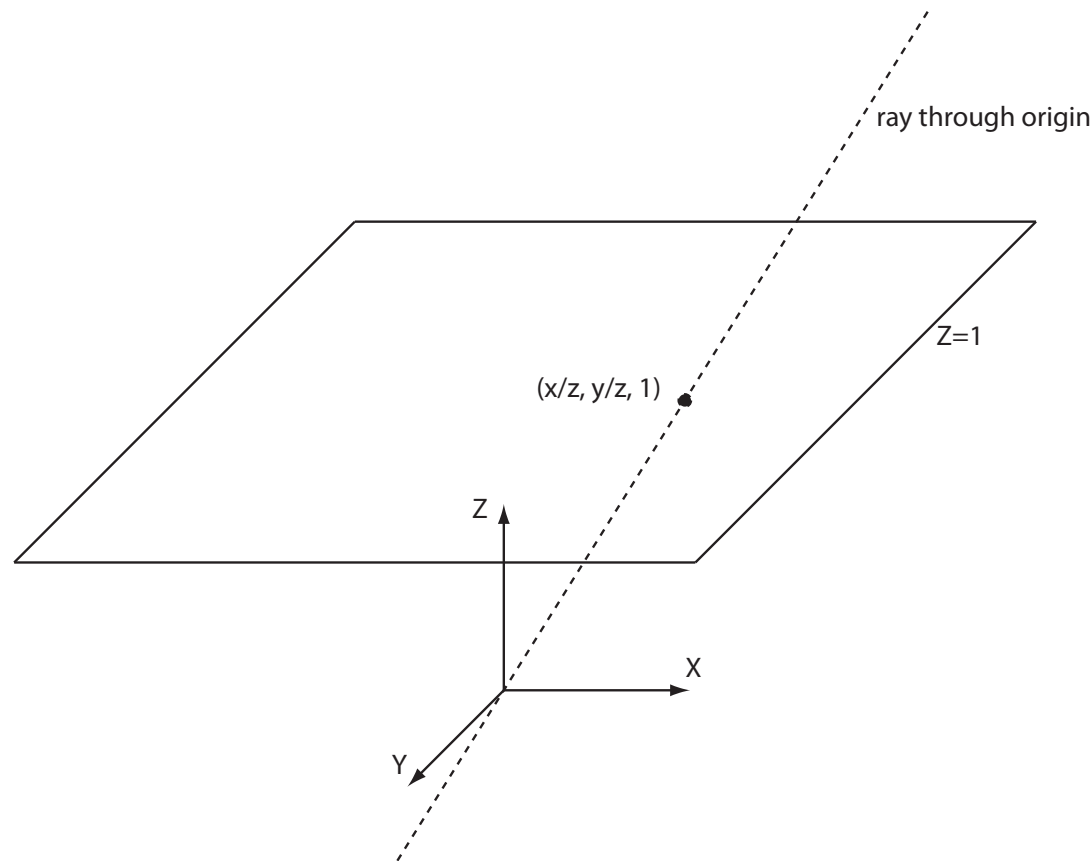
- Let's begin by considering the set of all rays through the origin in  $\mathbb{R}^3$ . (Note that we have gone up a dimension from  $\mathbb{R}^2$  to  $\mathbb{R}^3$ ). This set is referred to as the  $RP(2)$ , the real projective plane.

- Consider the intersection of the ray defined by the equation  $\mathbf{P} = \alpha \begin{pmatrix} x \\ y \\ z \end{pmatrix}$  with the plane  $z = 1$ .

- In general, the ray will intersect the plane at  $\begin{pmatrix} (x/z) \\ (y/z) \\ 1 \end{pmatrix}$

- The exceptions are rays of the form  $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$  which do not intersect the plane  $z = 1$ .

- We say there are two kinds of rays: rays that intersect the plane  $z = 1$  with coordinates of the form  $\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$  these rays can be associated with points in the plane  $\mathbb{R}^2$ , and rays that do not intersect the plane and have coordinates of the form  $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$  these can be thought of as corresponding to directions in the plane  $\mathbb{R}^2$  or points at infinity.

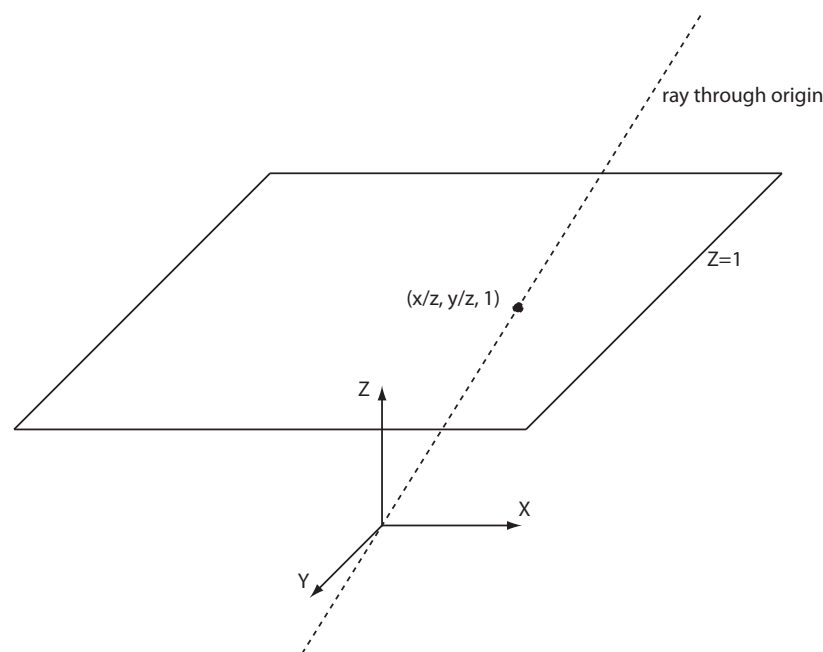


# Homogenous Coordinates

- Importantly when we are talking about projective coordinates we are referring to rays in space so absolute scale doesn't matter. As an example the following vectors all denote the same ray in space or projective point in  $RP(2)$ .

$$\begin{pmatrix} 4 \\ 2 \\ -3 \end{pmatrix} \propto \begin{pmatrix} 8 \\ 4 \\ -6 \end{pmatrix} \propto \begin{pmatrix} -12 \\ -6 \\ 9 \end{pmatrix}$$

- This interpretation of rays in space is very useful for modeling **cameras** in computer vision and graphics since cameras effectively measure the intersection of light rays with an image plane.





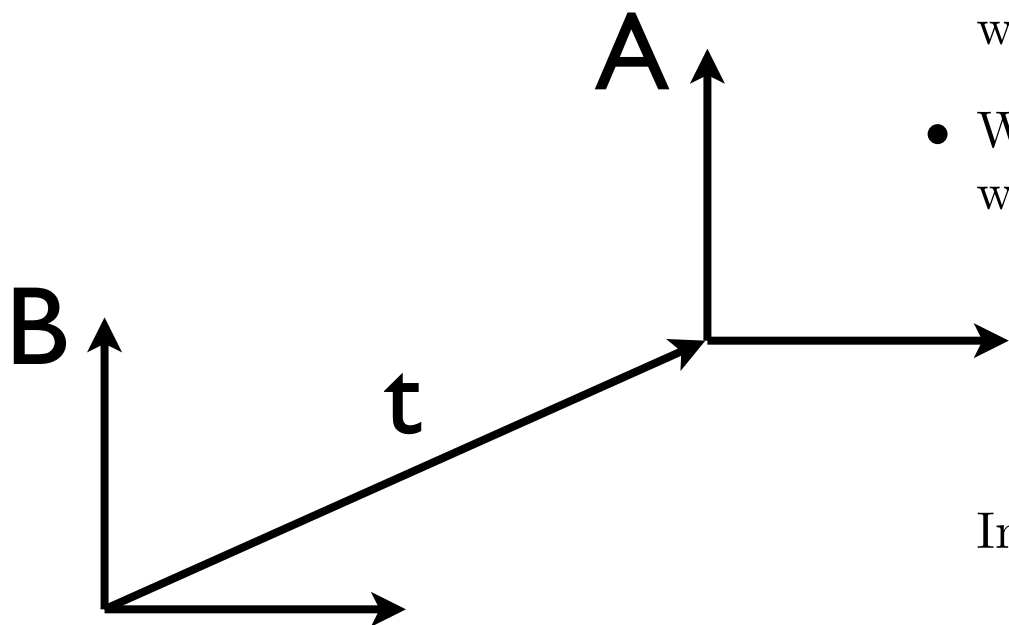
# Modeling Translation

- Let  $\mathbf{P}_A = \begin{pmatrix} x_A \\ y_A \\ 1 \end{pmatrix}$  denote the homogenous coordinates of a point wrt frame  $A$  and  $\mathbf{P}_B = \begin{pmatrix} x_B \\ y_B \\ 1 \end{pmatrix}$  denote the homogenous coordinates of the same point wrt frame  $B$ .
- We can represent the translation that relates these two coordinate frames with a  $3 \times 3$  matrix as follows:

$$\begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_A \\ y_A \\ 1 \end{pmatrix} = \begin{pmatrix} x_A + t_x \\ y_A + t_y \\ 1 \end{pmatrix} = \begin{pmatrix} x_B \\ y_B \\ 1 \end{pmatrix}$$

In other words:

$$\mathbf{P}_B = g_{BA} \mathbf{P}_A, \quad g_{BA} = \begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix}$$



# Modeling Rotation

- Similarly we can model the coordinate transformation associated with a rotation using an appropriate  $3 \times 3$  matrix.

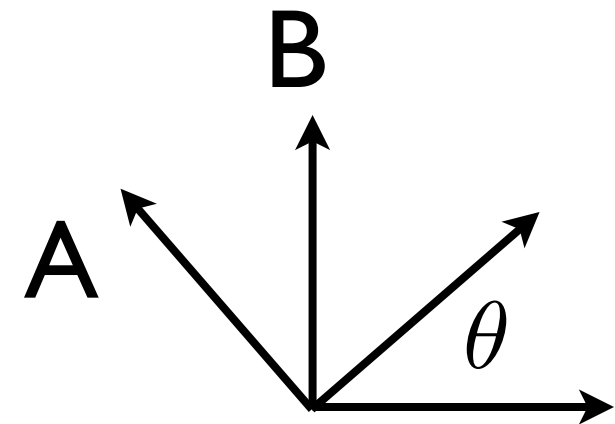
- Let  $\mathbf{P}_A = \begin{pmatrix} x_A \\ y_A \\ 1 \end{pmatrix}$  denote the homogenous coordinates of a point wrt frame  $A$  and  $\mathbf{P}_B = \begin{pmatrix} x_B \\ y_B \\ 1 \end{pmatrix}$  denote the homogenous coordinates of the same point wrt frame  $B$ .

- We can represent the rotation that relates these two coordinate frames with a  $3 \times 3$  matrix as follows:

$$\begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_A \\ y_A \\ 1 \end{pmatrix} = \begin{pmatrix} x_B \\ y_B \\ 1 \end{pmatrix}$$

In other words:

$$\mathbf{P}_B = g_{BA} \mathbf{P}_A, \quad g_{BA} = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$



# Modeling Scaling

- It is a simple matter to model scaling using homogenous coordinates.

$$\begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha x \\ \beta y \\ 1 \end{pmatrix}$$

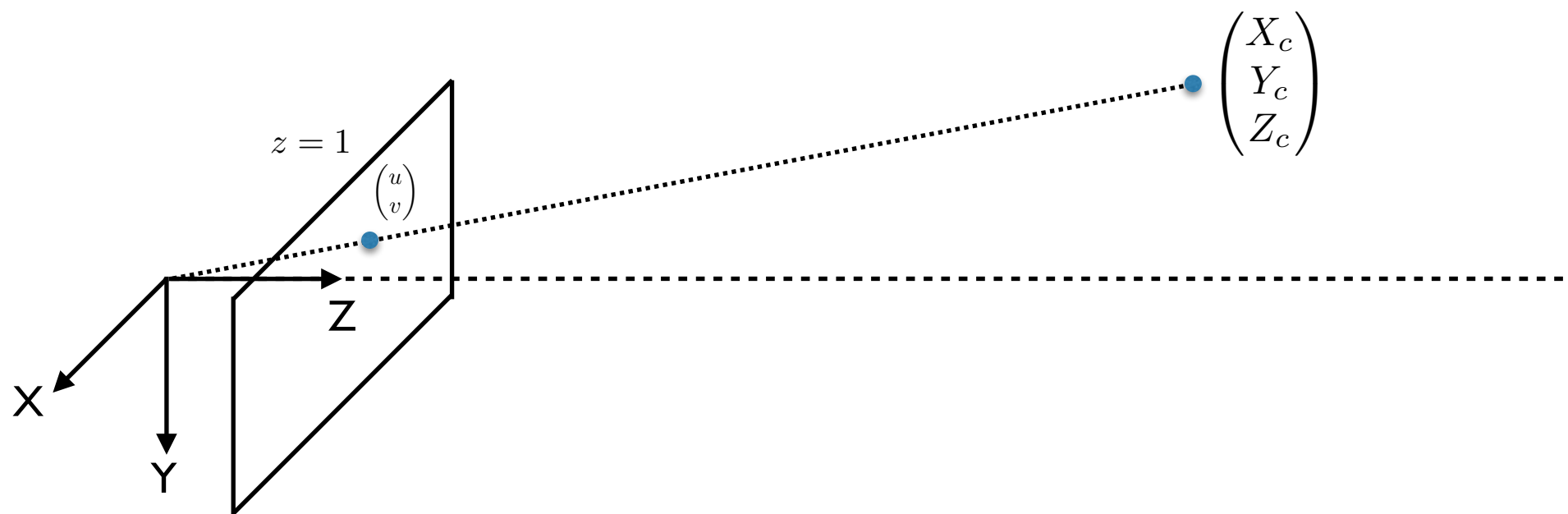
# Homogenous Coordinates

- Homogenous coordinates allow us to represent a variety of coordinate transformations including, scaling, rotation and translation using matrices.
- This is convenient for a range of applications since it then allows us to compose transformations via matrix multiplication.

# Perspective Camera Model

- In the *Perspective Projection* camera model, points in the scene are projected via straight lines through the origin onto an image plane, in this case the plane  $z = 1$ .
- In this drawing we can see from similar triangles that a point with coordinates  $(X_c, Y_c, Z_c)$  with respect to the cameras frame of reference will project onto the image plane at coordinates  $(u, v)$  where:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} X_c/Z_c \\ Y_c/Z_c \end{pmatrix} \quad (1)$$

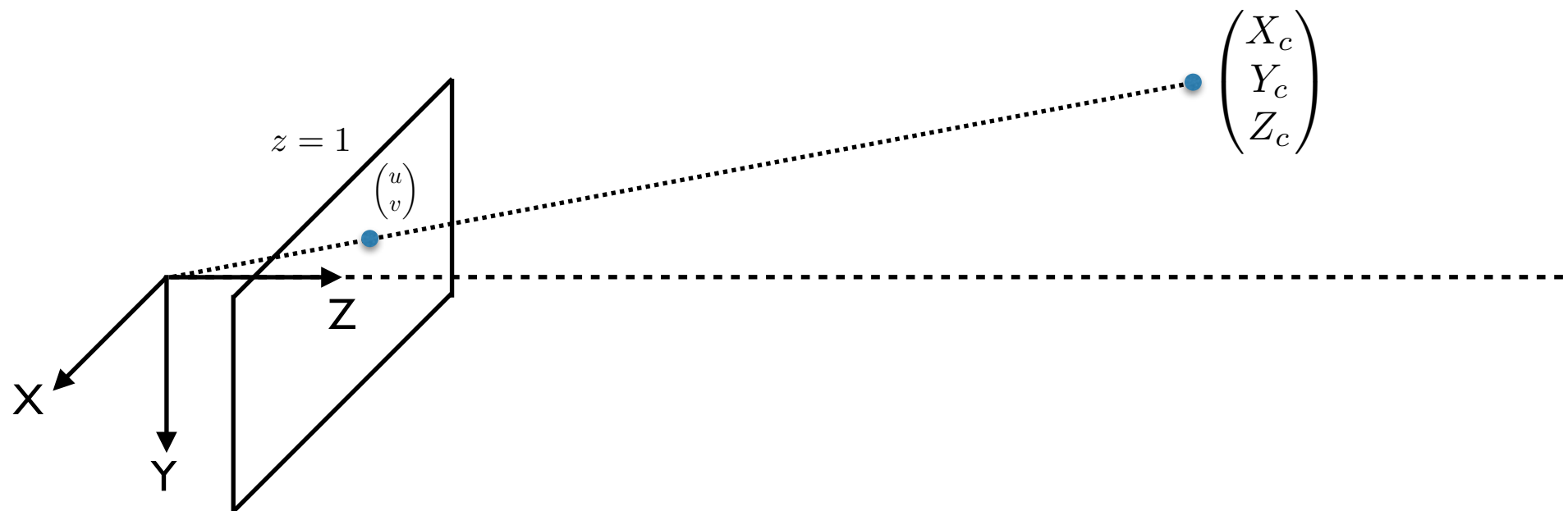


# Perspective Camera Model

- Note that this operation of perspective projection is *not* linear, however, in Computer Vision and Graphics we typically model points in 3D as homogenous coordinates in  $\mathbb{RP}(3)$  then we model perspective projection as a mapping from  $\mathbb{RP}(3)$  to  $\mathbb{RP}(2)$  exploiting the fact that projective coordinates are only defined up to a scale factor.

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} X_c/Z_c \\ Y_c/Z_c \end{pmatrix} \quad (1)$$

$$\Rightarrow \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \propto \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} \quad (2)$$

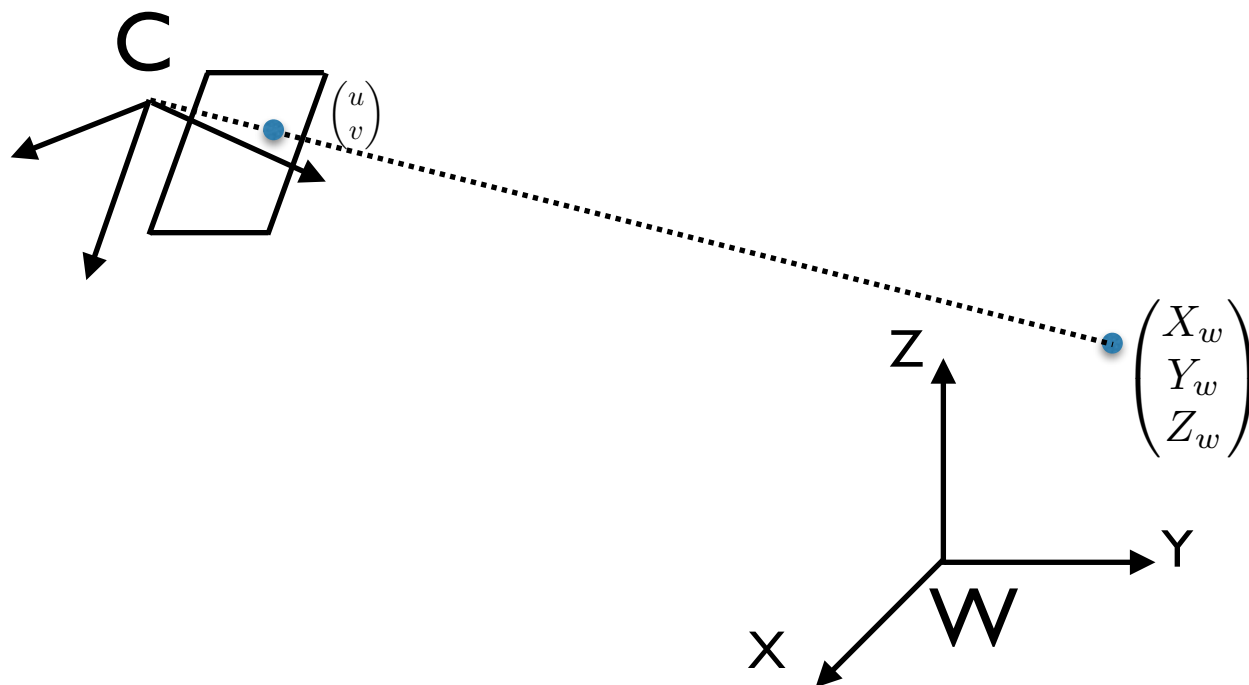


# Perspective Camera Model

- To employ the camera projection model in computer vision and graphics code what we typically do is construct a  $4 \times 4$  coordinate transformation matrix to transform vertex coordinates from the world frame,  $W$ , to the camera frame,  $g_{CW} \in \mathbb{R}^{4 \times 4}$ . We then apply the projection matrix to compute the homogenous coordinates.

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \propto \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} g_{CW} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \quad (1)$$

- Remember that in order to get the final image coordinates,  $(u, v)$ , you would need to divide the homogenous coordinates by the third component to turn the proportionality into an equality.



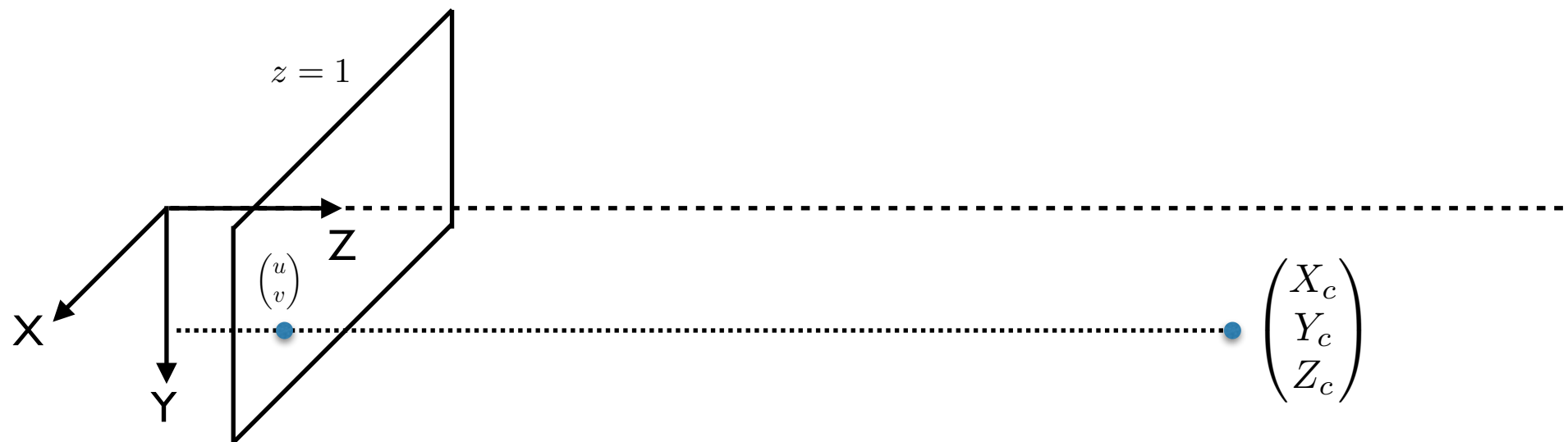


# Orthographic Camera Model

- In the *Orthographic Projection* camera model, points in the scene are projected via straight lines parallel to the optical axis onto the image plane, in this case the plane  $z = 1$
- The operation of orthographic projection is linear and can be captured by the following projection equations:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} X_c \\ Y_c \end{pmatrix} \quad (1)$$

$$\Rightarrow \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \propto \begin{pmatrix} X_c \\ Y_c \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} \quad (2)$$

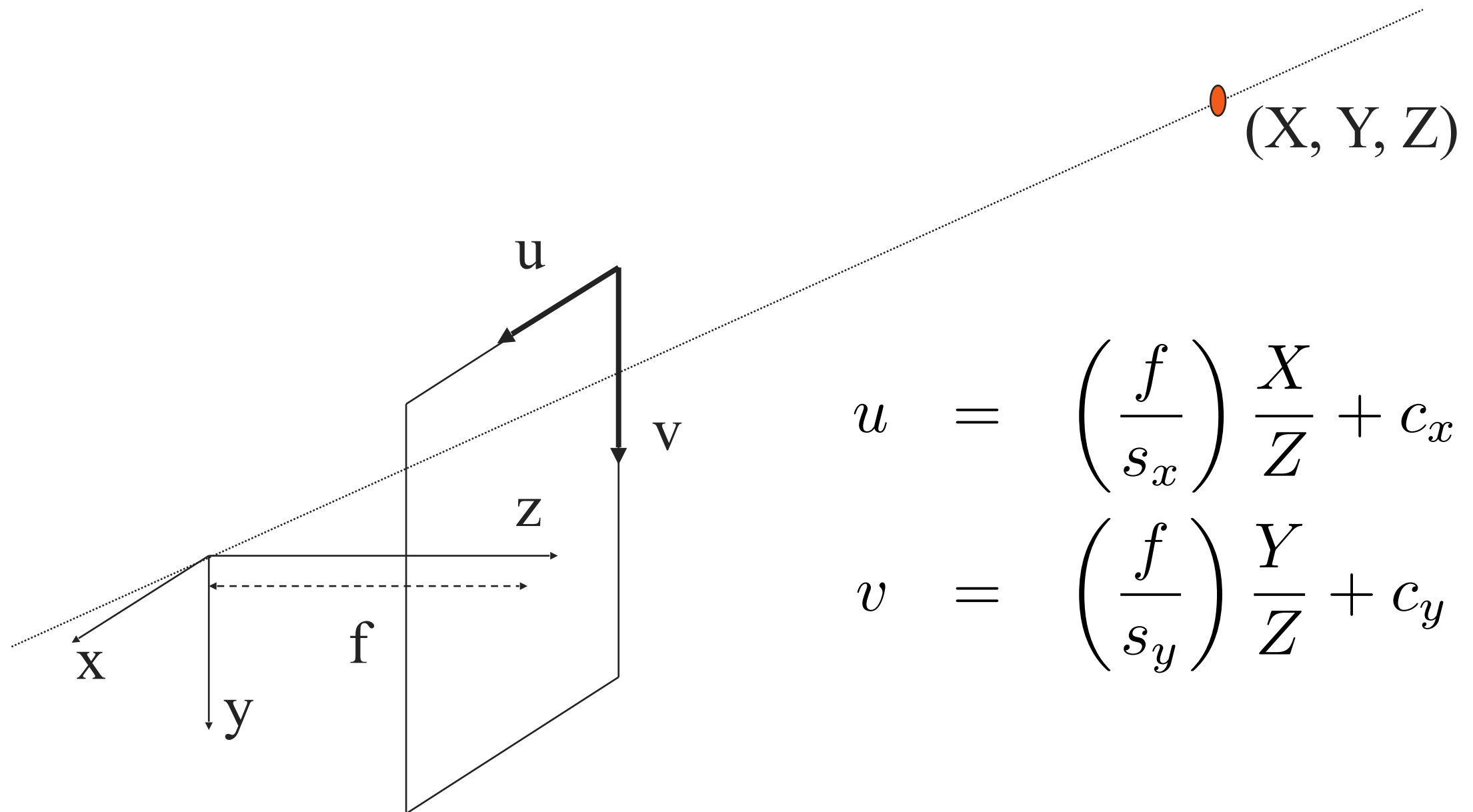


# Camera Modeling

## Continued

# Modeling Cameras

- Perspective Projection



# Modeling Cameras

- The mapping between the coordinates of a point wrt to the camera's frame of reference,  $(X, Y, Z)$ , and the coordinates of its projection in the image can be viewed as a projective transformation
- The upper triangular matrix  $A$  is referred to as the matrix of *intrinsic parameters*.

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \propto A \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}, \quad A = \begin{pmatrix} \frac{f}{s_x} & 0 & c_x \\ 0 & \frac{f}{s_y} & c_y \\ 0 & 0 & 1 \end{pmatrix}$$
$$A \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} \left(\frac{f}{s_x}\right) X + c_x Z \\ \left(\frac{f}{s_y}\right) Y + c_y Z \\ Z \end{pmatrix} \propto \begin{pmatrix} \left(\frac{f}{s_x}\right) \frac{X}{Z} + c_x \\ \left(\frac{f}{s_y}\right) \frac{Y}{Z} + c_y \\ 1 \end{pmatrix}$$

# Modeling Cameras

- Note that you can also use this model to relate the pixel coordinates of a point  $(u,v)$ , to the corresponding ray in space  $(X, Y, Z)$
- Note that the resulting ray is a projective point so it is only defined up to a scale factor.

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \propto A^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}, \quad A = \begin{pmatrix} \frac{f}{s_x} & 0 & c_x \\ 0 & \frac{f}{s_y} & c_y \\ 0 & 0 & 1 \end{pmatrix}$$

# Extrinsic Parameters

- In many cases the coordinates of the points are given wrt to a coordinate system other than the one fixed to the camera
- In this case we must apply a transformation in  $SE(3)$  to map coordinates from this frame to the cameras frame
- The parameters,  $\mathbf{R}$  in  $SO(3)$  and  $\mathbf{T}$  in  $\mathbb{R}^3$ , which capture the relationship between the extrinsic reference frame and the cameras frame are referred to as **extrinsic parameters**

$$g_{CW} = \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix}$$

# The projection matrix

- The relationship between the homogenous coordinates of a point wrt the extrinsic frame and the homogenous coordinates of its projection in the image can be modeled as a projective transformation from  $\mathbb{RP}^3$  to  $\mathbb{RP}^2$
- This transformation can be represented with a 3 by 4 matrix - the projection matrix,  $P$

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \propto P \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}, \quad P \in \mathbb{R}^{3 \times 4}$$

$$P \propto A \begin{pmatrix} R & T \end{pmatrix}$$

# The projection matrix

- Note that in general when we apply a projection matrix to a point we could end up with a point at infinity ( $w = 0$  in the equation below)
- The general procedure is to apply the projection matrix to the point to produce a projective point, then divide through by the last component to get the pixel coordinates if possible.

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} \propto P \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}, \quad P \in \mathbb{R}^{3 \times 4}$$

$$P \propto A \begin{pmatrix} R & T \end{pmatrix}$$



# Orthographic projection

- Orthographic projections can also be modeled with projection matrices

$$u = k_x X_c + c_x$$

$$v = k_y Y_c + c_y$$

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \propto \begin{pmatrix} k_x & 0 & 0 & c_x \\ 0 & k_y & 0 & c_y \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

# Orthographic Projection cont'd

- Matrices corresponding to orthographic projections have a particular form as shown below:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \propto \begin{pmatrix} k_x & 0 & 0 & c_x \\ 0 & k_y & 0 & c_y \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \propto \begin{pmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

# Some facts about projection

- If three points in the world are collinear then their projections in the image will also be collinear
  - ▼ Corollary: If two lines in the scene intersect then the projection of their intersection will be the same as the intersection of the projections of the two lines
- N.B. These rules also apply to points at infinity

# Vanishing Points

- Under perspective projection parallel lines in the image may seem to verge to a point of intersection in the image. This is a consequence of the fact that points at infinity in the scene can project to ‘regular points’ in the image
- This fact has been used by artists for centuries
- Under perspective projection, points at infinity in the world can project to arbitrary locations in the image

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} \propto A \begin{pmatrix} R & T \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 0 \end{pmatrix} = AR \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix}$$

# Vanishing Points

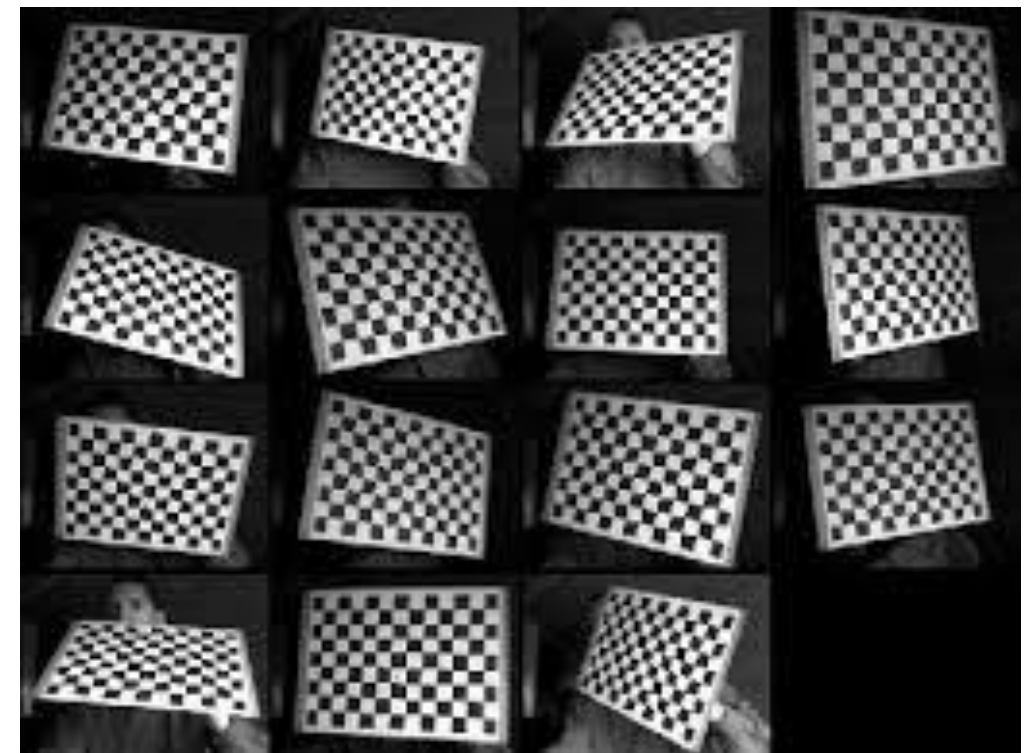
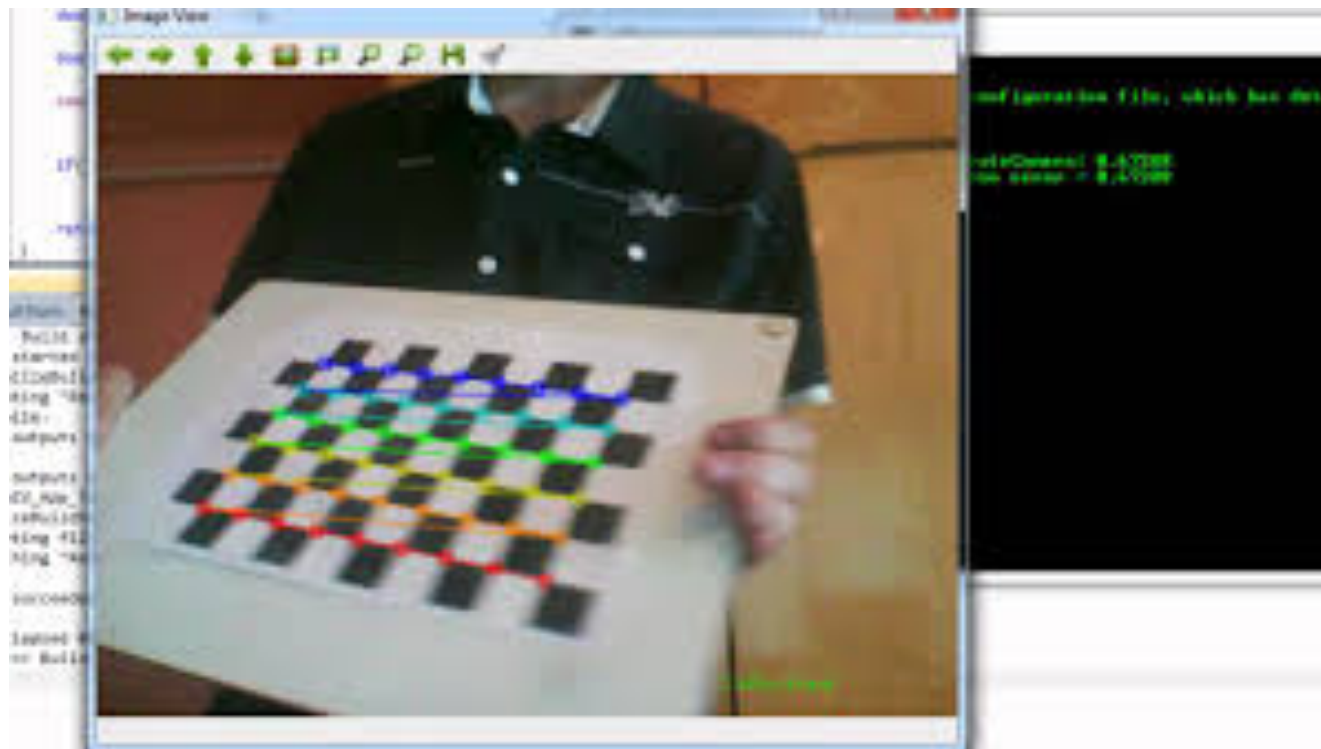
- Under orthographic/weak perspective projection points at infinity in the world project to points at infinity in the image which means that parallel lines in the world will also appear parallel in the image

$$\begin{pmatrix} u \\ v \\ 0 \end{pmatrix} \propto \begin{pmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 0 \end{pmatrix}$$

# Camera Calibration

# Camera Calibration

- The goal of camera calibration procedures is to produce an estimate for the projection matrix,  $P$ , and/or the intrinsic and extrinsic parameters,  $A$ ,  $R$  and  $T$
- Given the correspondences between a set of point features in the world and their projections in an image compute the coefficients of the transformation



# Common Approach

- A common approach to calibration is to view it as a parameter estimation problem
- General formulation
  - ▼ Given some measurements,  $\mathbf{x}_i$  and  $\mathbf{y}_i$ , a parameter vector  $\mathbf{a}$  and a function which relates them find the parameter vector that best matches the data

$$\min_{\mathbf{a}} \sum_i \|\mathbf{y}_i - f(\mathbf{a}, x_i)\|^2$$



# Parameter Estimation

- The parameter estimation framework is frequently employed in situations where:
  - ▼ There is no easy way to invert the function  $f$
  - ▼ There are multiple noisy measurements which overconstrain the system

# Camera Calibration

- Camera calibration can be viewed as a parameter estimation problem where:
  - ▼  $y = (u, v)$
  - ▼  $x = (X_w, Y_w, Z_w)$
  - ▼  $a = (\text{focal\_length}, \text{aspect\_ratio}, c_x, c_y, R, T)$
  - ▼  $f$  is the function that relates the coordinates of a point in the world to the coordinates of its projection in the image

# Intrinsic Parameters

- In many situations it is appropriate to take advantage of a priori knowledge about the cameras intrinsic parameters
- Some common simplifying assumptions
  - ▼ Assume that the skew angle is zero
  - ▼ Often the aspect ratio ( $s_y/s_x$ ) is known. In most digital images this is engineered to be 1
  - ▼ In many cases we assume that the principal point ( $c_x, c_y$ ) is located in the center of the image
- It can be difficult to recover estimates for some of the parameters because their effects are confounded in the projection matrix egs the ( $c_x$  and  $c_y$ ) are usually confounded with ( $T_x$  and  $T_y$ ) while  $f_x$  and  $f_y$  are usually confused with  $T_z$