COMP 9517 Computer Vision

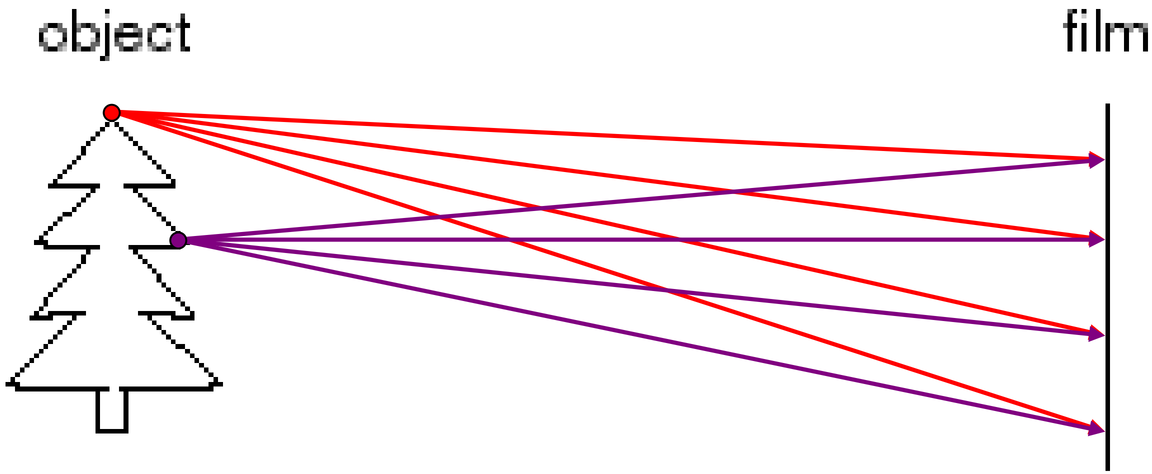
Image Formation

Geometry of Image Formation

Mapping between image and world coordinates

- Pinhole camera model
- Projective geometry
  - Vanishing points and lines
- Projection matrix

Image formation

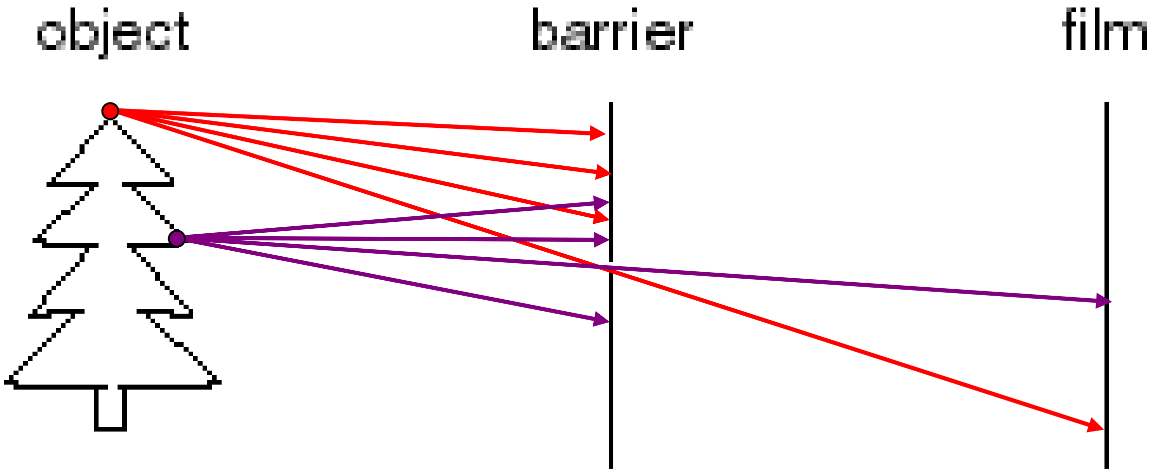


Let us design a camera

- Idea 1: put a piece of film in front of an object
- Do we get a reasonable image?

Slide source: Seitz

Pinhole camera

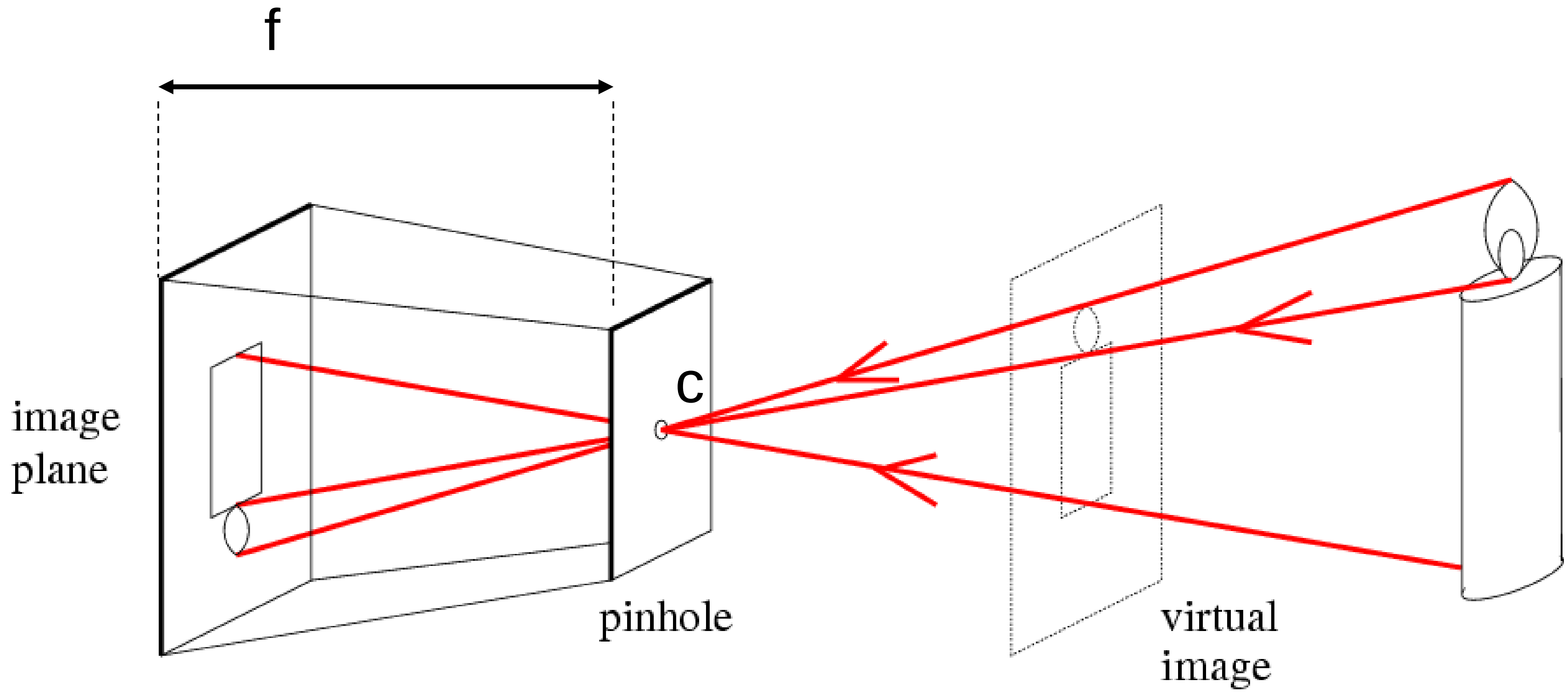


Idea 2: add a barrier to block off most of the rays

- This reduces blurring
- The opening known as the **aperture**

Slide source: Seitz

Pinhole camera



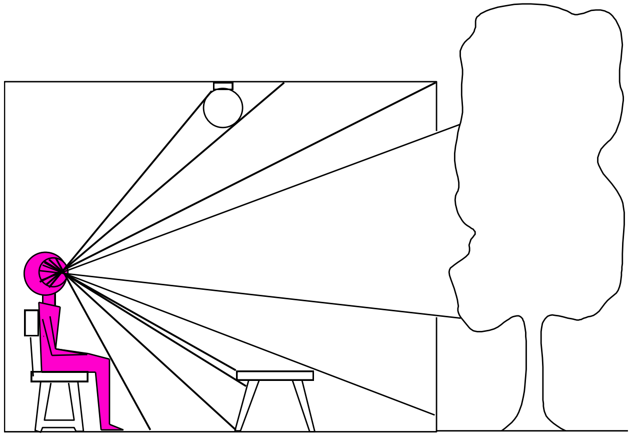
f = focal length
c = centre of the camera

Figure from Forsyth

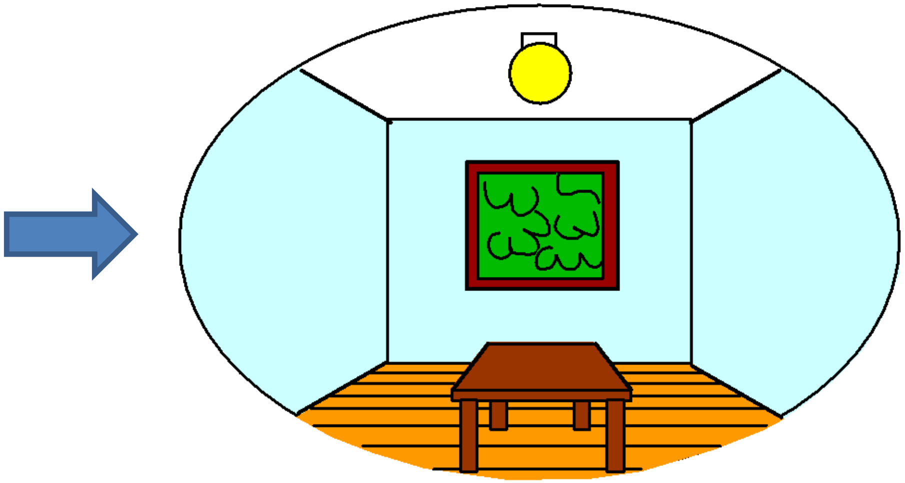
Dimensionality Reduction Machine (3D to 2D)

3D world

2D image



Point of observation



Slide source: Seitz

Projection can be tricky…

Projection can be tricky…

Length and area are not preserved

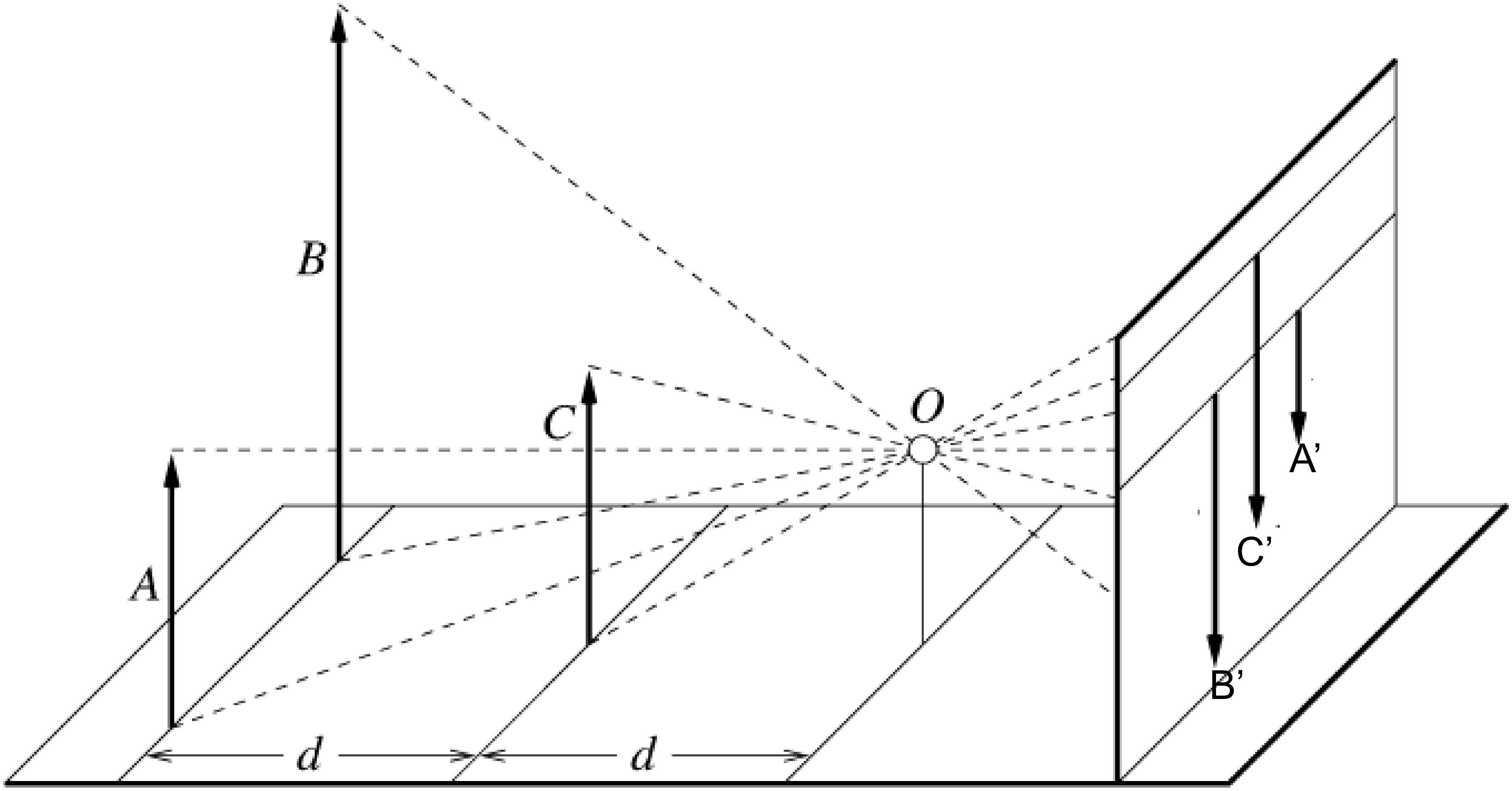
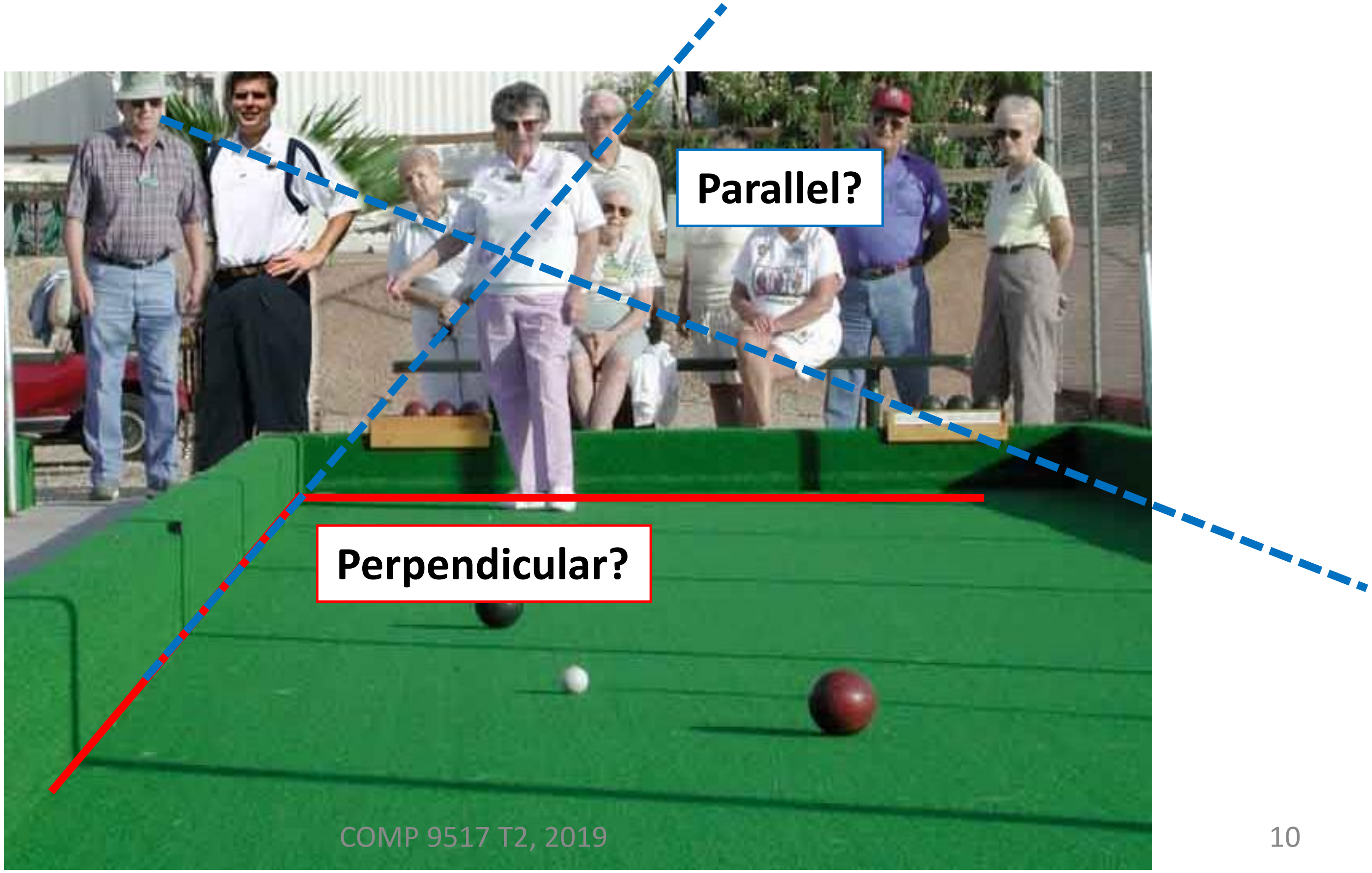


Figure by David Forsyth

Projective Geometry

What is lost?

- Length
- Angles



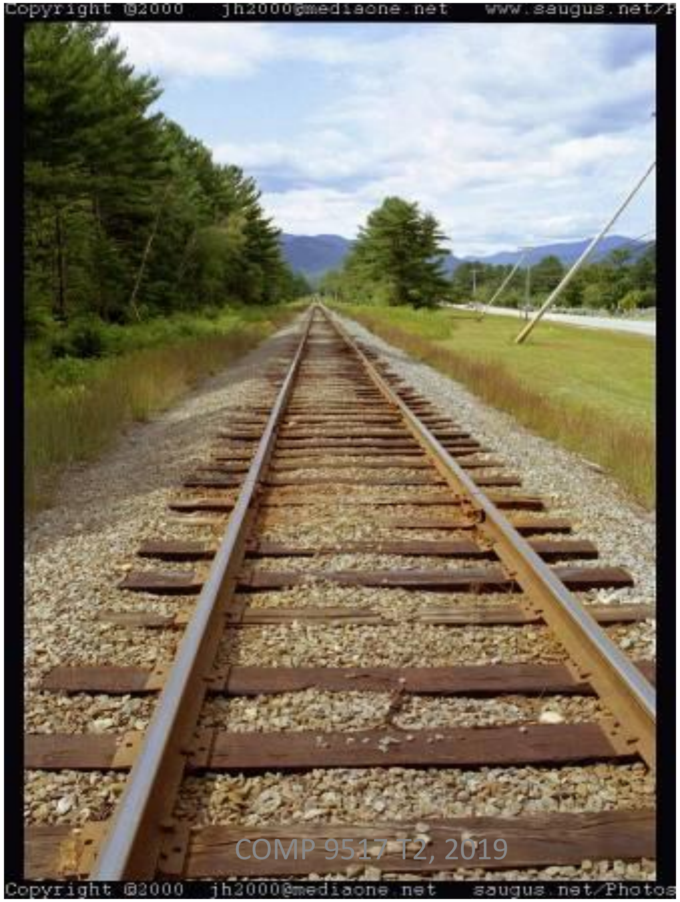
Projective Geometry

What is preserved?

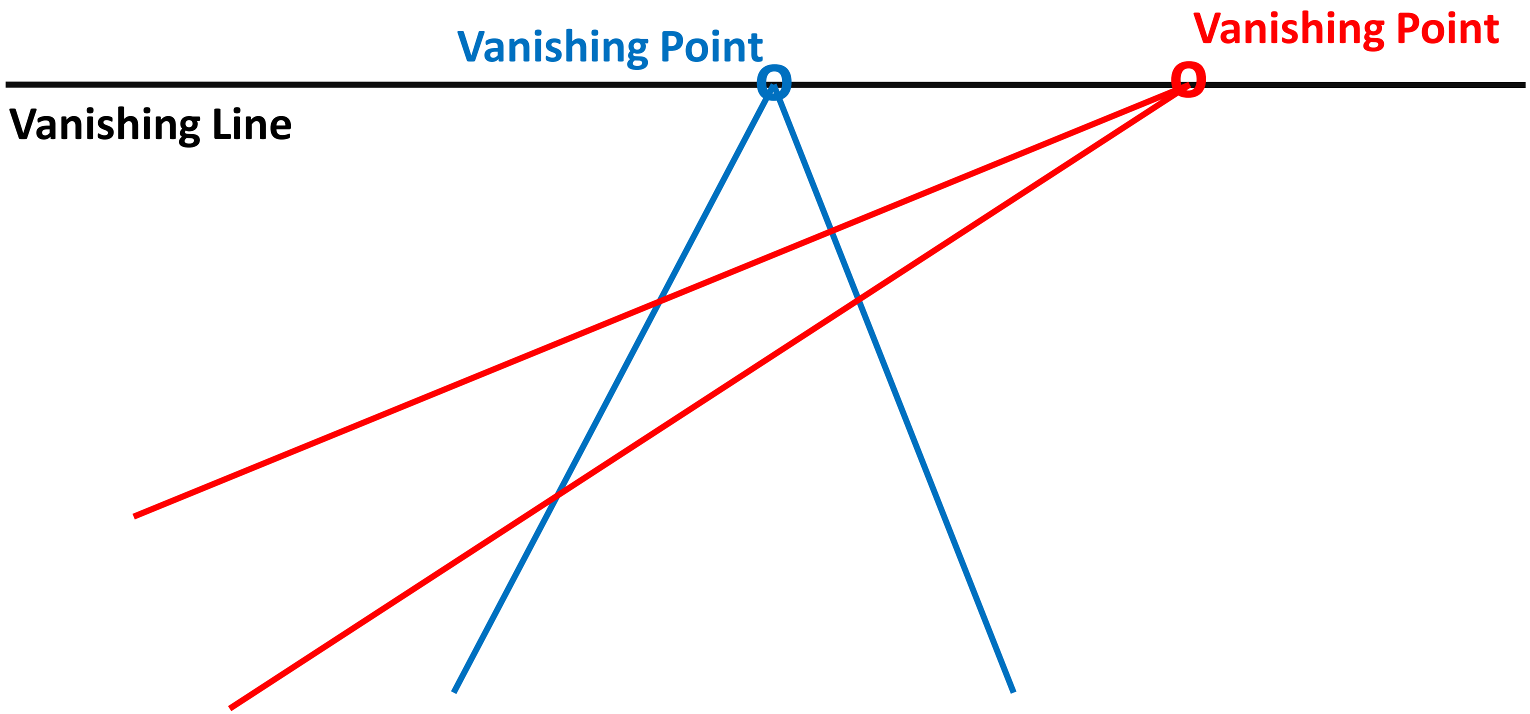
- Straight lines are still straight

Vanishing points and lines

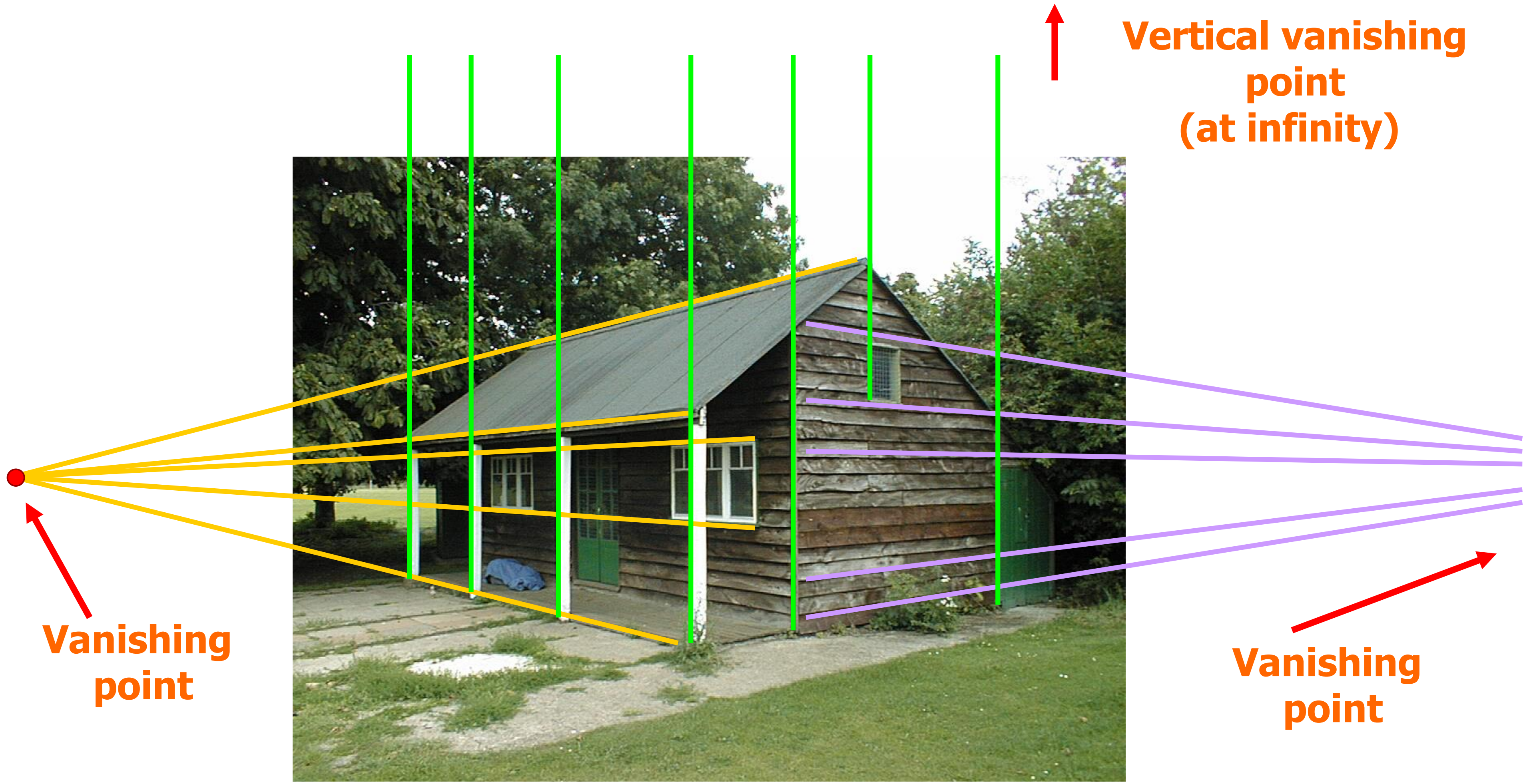
Parallel lines in the world intersect in the image at a "vanishing point"



Vanishing points and lines

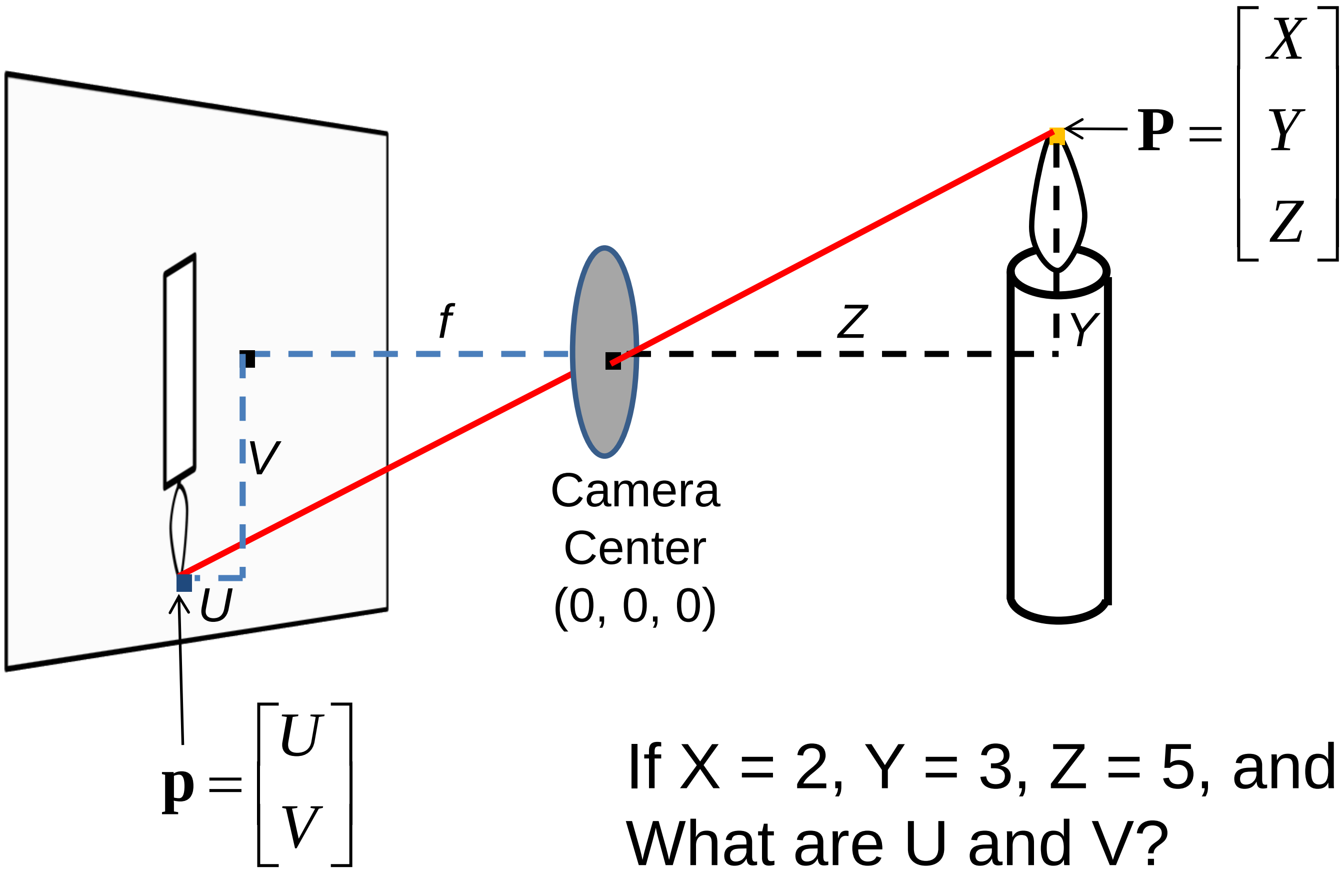


Vanishing points and lines



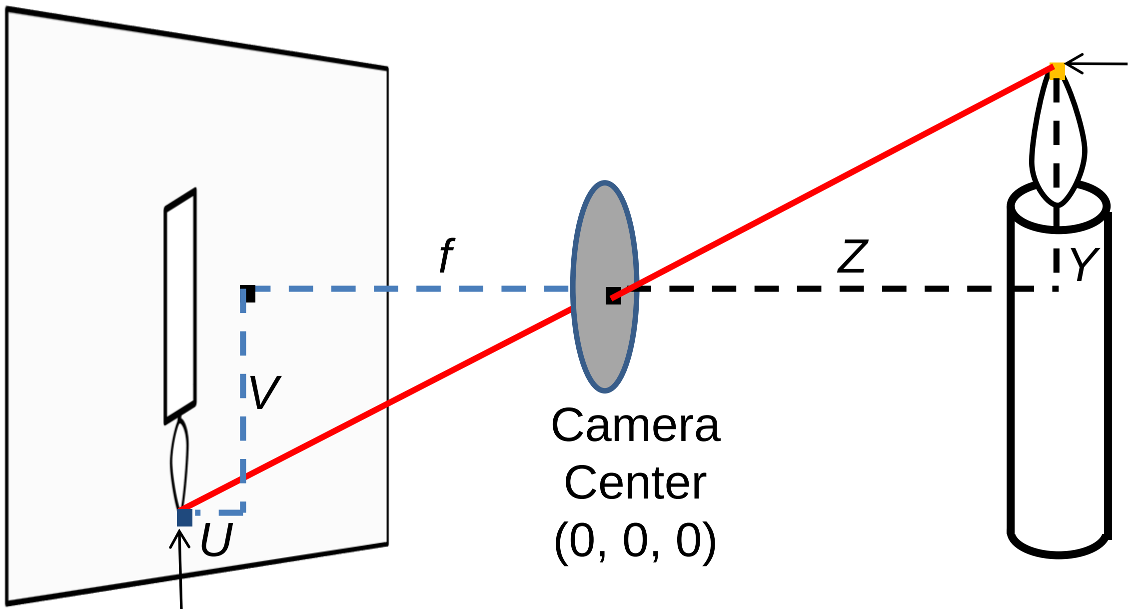
Slide from Efros, Photo from Criminisi

Projection: world coordinates→image coordinates



If X = 2, Y = 3, Z = 5, and f = 2
What are U and V?

Projection: world coordinates→image coordinates



$$\mathbf{P} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$$\mathbf{p} = \begin{bmatrix} U \\ V \end{bmatrix}$$

$$U = -X * \frac{f}{Z}$$

$$V = -Y * \frac{f}{Z}$$

$$U = -2 * \frac{2}{5}$$

$$V = -3 * \frac{2}{5}$$

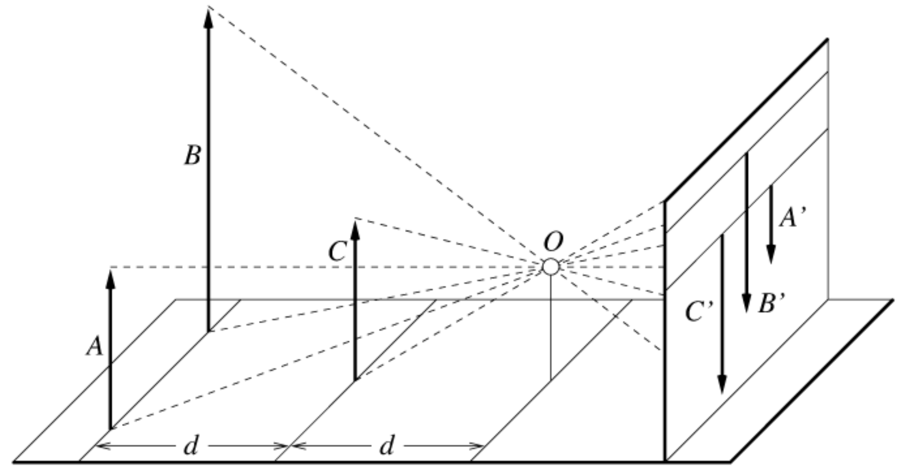
Perspective Projection

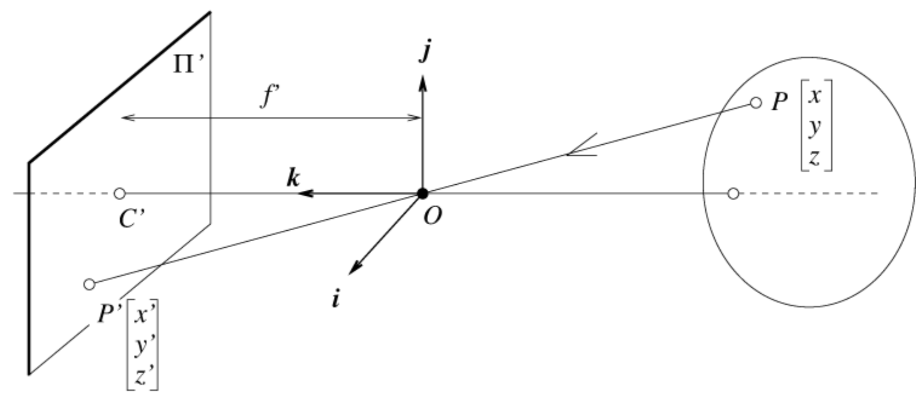
- Apparent size of object depends on its distance: far objects appear smaller
- By similar triangles

$$(x', y', z') \rightarrow (f\frac{x}{z}, f\frac{y}{z}, -f)$$

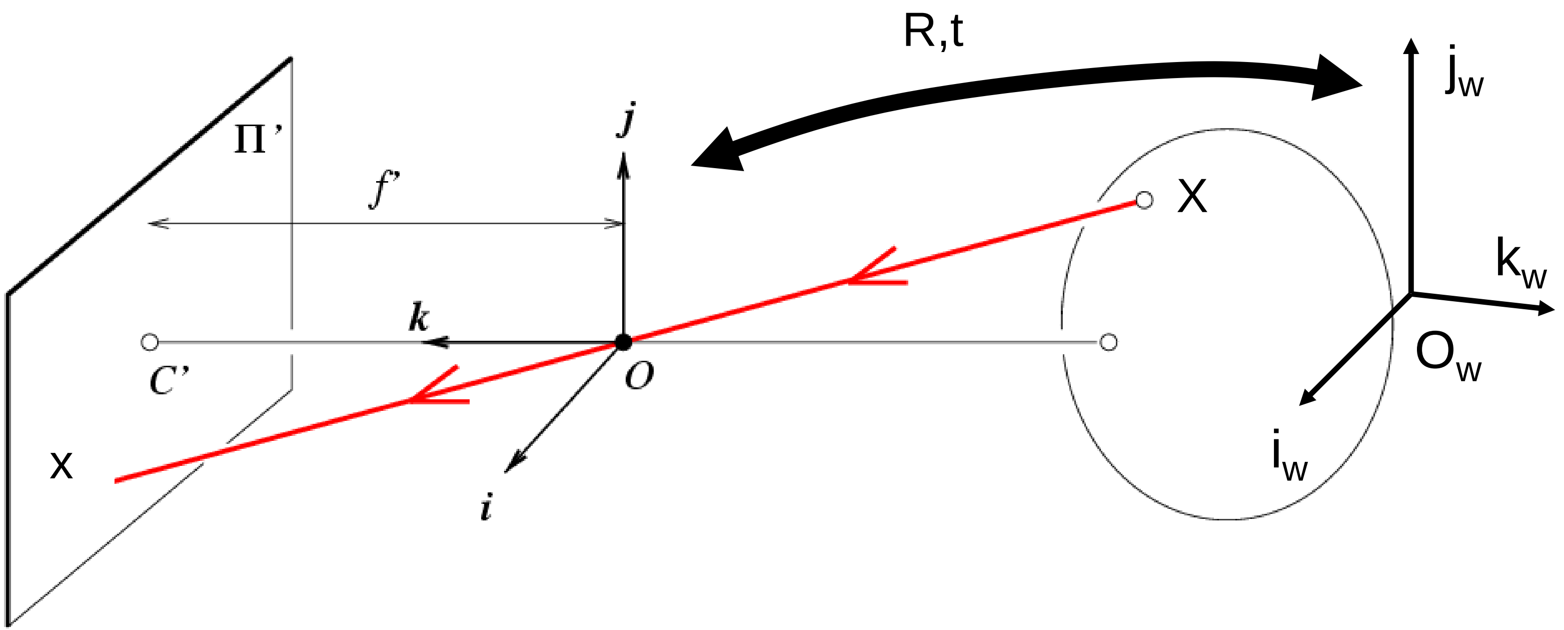
- Ignore the third coordinate, and get

$$(x', y') \rightarrow (f\frac{x}{z}, f\frac{y}{z})$$





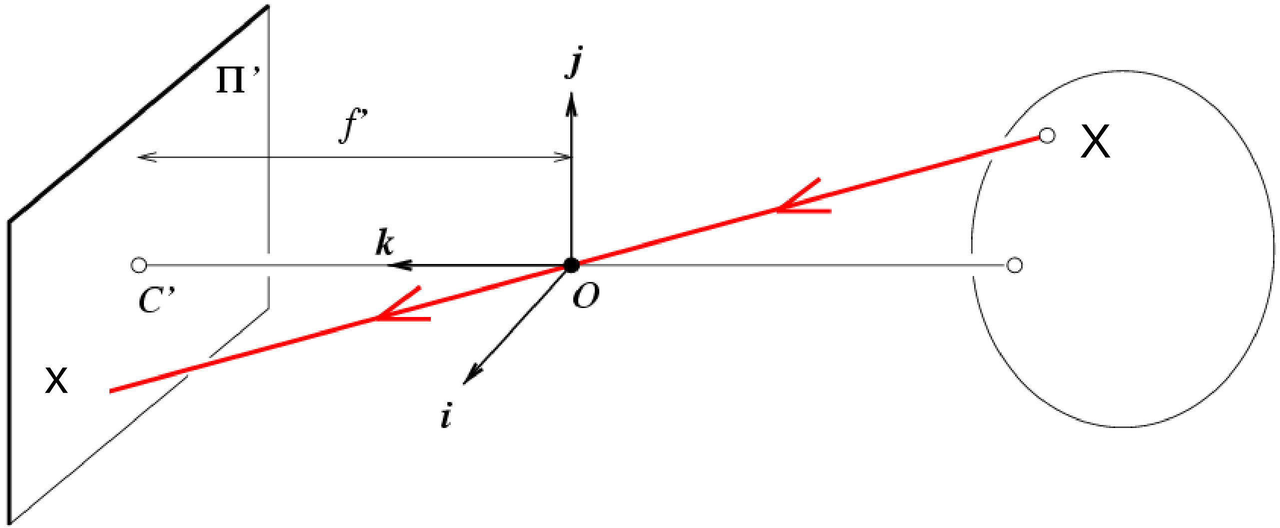
Projection matrix



$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}\mathbf{X}$$

x: Image Coordinates: (u,v,1)
K: Intrinsic Matrix (3x3)
R: Rotation (3x3)
t: Translation (3x1)
X: World Coordinates: (X,Y,Z,1)

Pinhole Camera Model



Intrinsic Assumptions

- Unit aspect ratio
- Optical center at (0,0)
- No skew

Extrinsic Assumptions

- No rotation
- Camera at (0,0,0)

$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\mathbf{X} \Rightarrow w\begin{bmatrix}u\\v\\1\end{bmatrix} = \begin{bmatrix}\boxed{\begin{matrix}f & 0 & 0\\0 & f & 0\\0 & 0 & 1\end{matrix}} & \begin{matrix}0\\0\\0\end{matrix}\end{bmatrix}\begin{bmatrix}x\\y\\z\\1\end{bmatrix}$$

K

Slide Credit: Savarese

Homogeneous coordinates

- Line equation: ax + by + c = 0

$$line_i = \begin{bmatrix} a_i \\ b_i \\ c_i \end{bmatrix}$$

- Append 1 to pixel coordinate to get homogeneous coordinate

$$p_i = \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix}$$

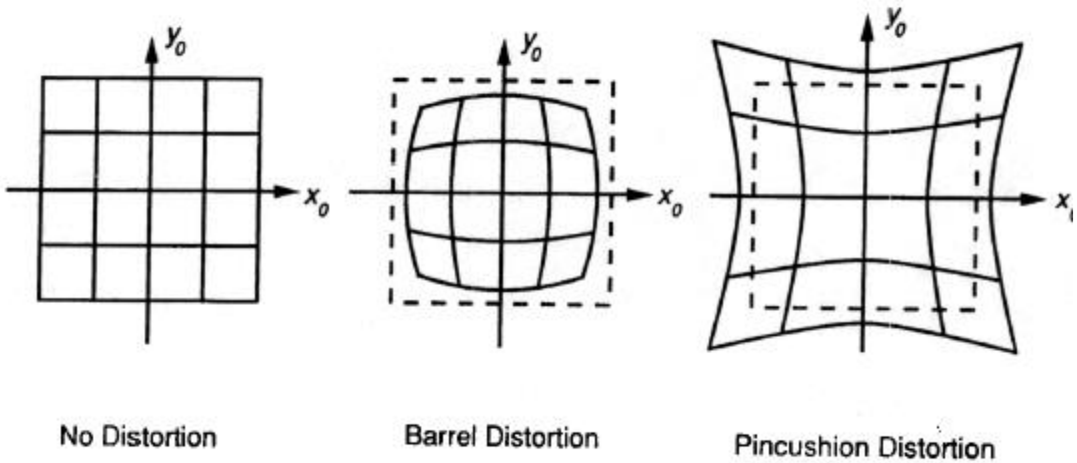
- Line given by cross product of two points

$$line_{ij} = p_i \times p_j$$

- Intersection of two lines given by cross product of the lines

$$q_{ij} = line_i \times line_j$$

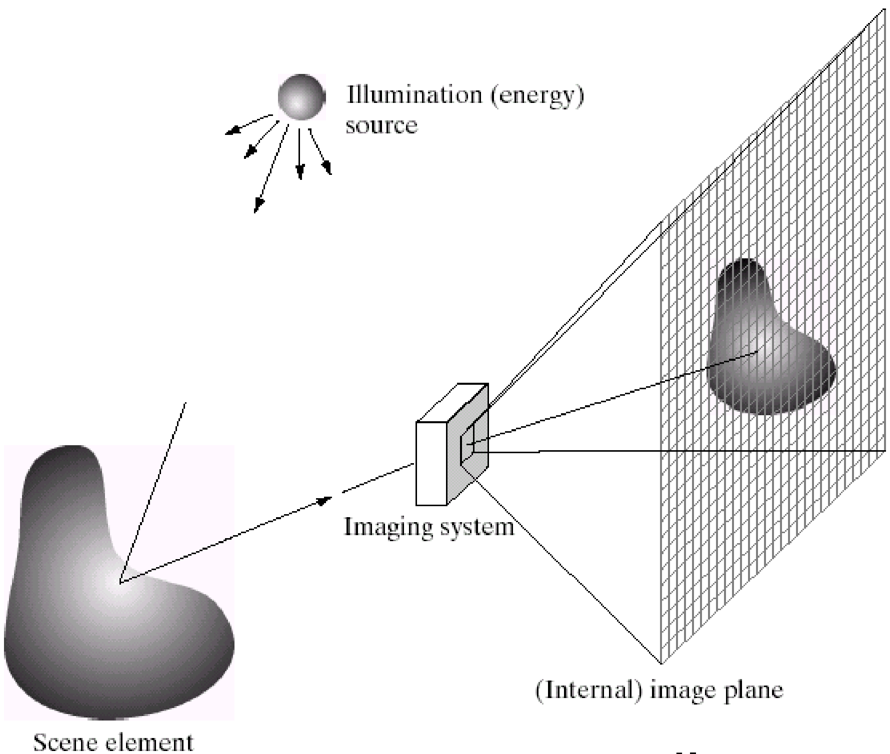
Beyond Pinholes: Radial Distortion



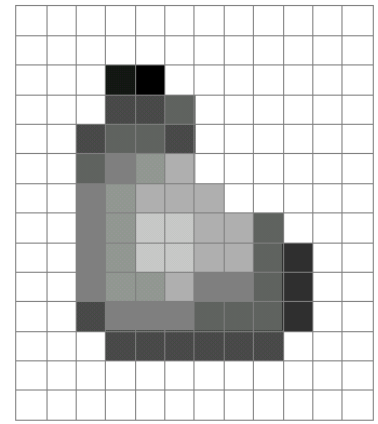
Corrected Barrel Distortion

Image from Martin Habbecke

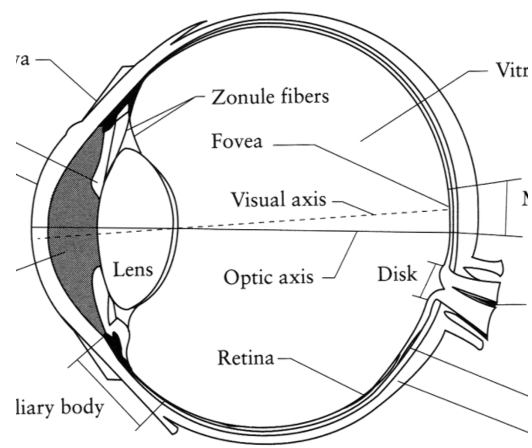
Image Formation



Film



Digital Camera

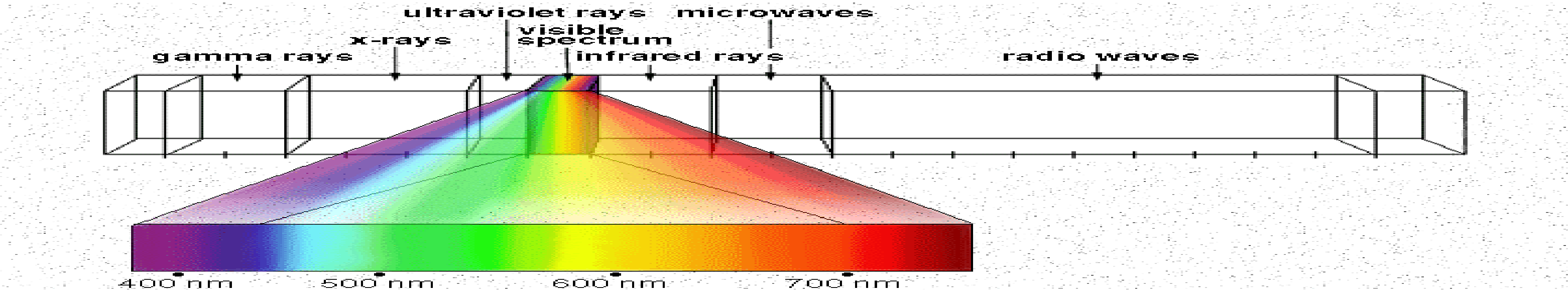


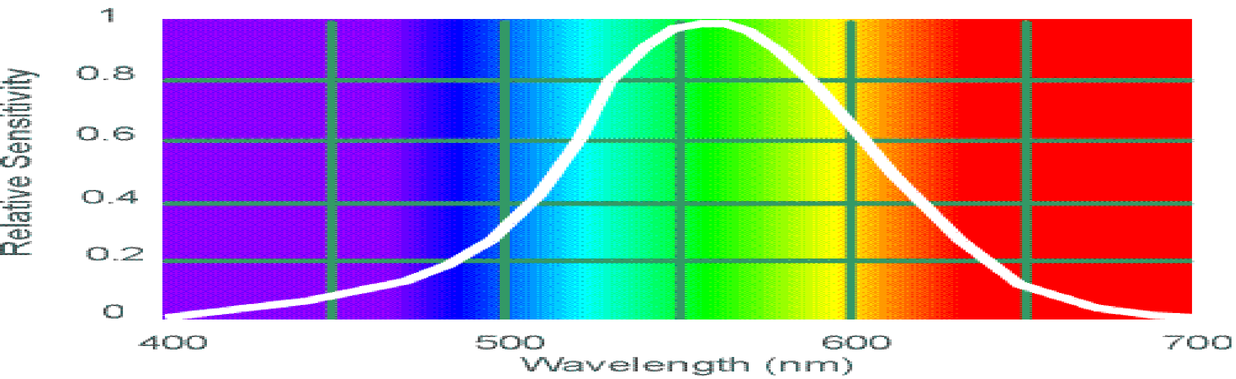
The Eye

Why do we care about human vision?

- Cameras necessarily imitate the frequency response of the human eye, so we should know that much.
- Computer vision probably would not get as much attention if biological vision (especially human vision) had not proved that it was possible to make important judgements from 2D images.

Electromagnetic Spectrum





Human Luminance Sensitivity Function

http://www.yorku.ca/eye/photopik.htm

Colour Image

Images represented as a matrix

column

row

R

0.92	0.93	0.94	0.97	0.62	0.37	0.85	0.97	0.93	0.92	0.99
0.95	0.89	0.82	0.89	0.56	0.31	0.75	0.92	0.81	0.95	0.91
0.89	0.72	0.51	0.55	0.51	0.42	0.57	0.41	0.49	0.91	0.92
0.96	0.95	0.88	0.94	0.56	0.46	0.91	0.87	0.90	0.97	0.95
0.71	0.81	0.81	0.87	0.57	0.37	0.80	0.88	0.89	0.79	0.85
0.49	0.62	0.60	0.58	0.50	0.60	0.58	0.50	0.61	0.45	0.33
0.86	0.84	0.74	0.58	0.51	0.39	0.73	0.92	0.91	0.49	0.74
0.96	0.67	0.54	0.85	0.48	0.37	0.88	0.90	0.94	0.82	0.93
0.69	0.49	0.56	0.66	0.43	0.42	0.77	0.73	0.71	0.90	0.99
0.79	0.73	0.90	0.67	0.33	0.61	0.69	0.79	0.73	0.93	0.97
0.91	0.94	0.89	0.49	0.41	0.78	0.78	0.77	0.89	0.99	0.93

G

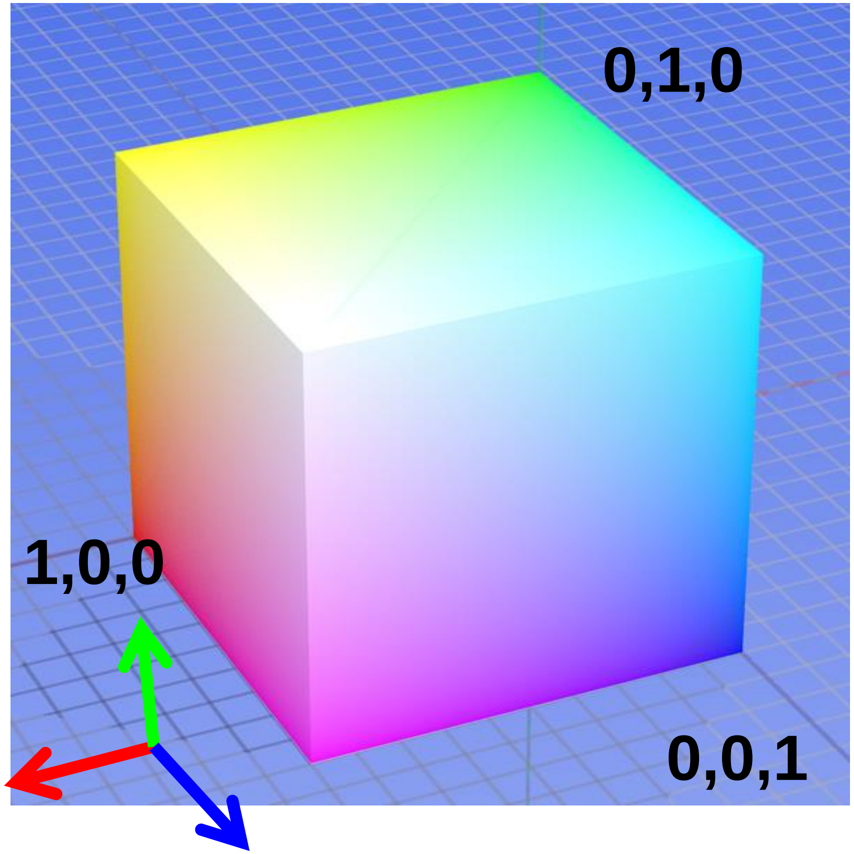
									0.92	0.99
									0.95	0.91
									0.91	0.92
									0.97	0.95
									0.79	0.85
									0.45	0.33
									0.49	0.74
									0.82	0.93
									0.90	0.99
0.79	0.73	0.90	0.67	0.33	0.61	0.69	0.79	0.73	0.93	0.97
0.91	0.94	0.89	0.49	0.41	0.78	0.78	0.77	0.89	0.99	0.93

B

									0.92	0.99
									0.95	0.91
									0.91	0.92
									0.97	0.95
									0.79	0.85
									0.45	0.33
									0.49	0.74
									0.82	0.93
									0.90	0.99
0.79	0.73	0.90	0.67	0.33	0.61	0.69	0.79	0.73	0.93	0.97
0.91	0.94	0.89	0.49	0.41	0.78	0.78	0.77	0.89	0.99	0.93

Colour spaces: RGB

Default colour space



R
(G=0,B=0)

G
(R=0,B=0)

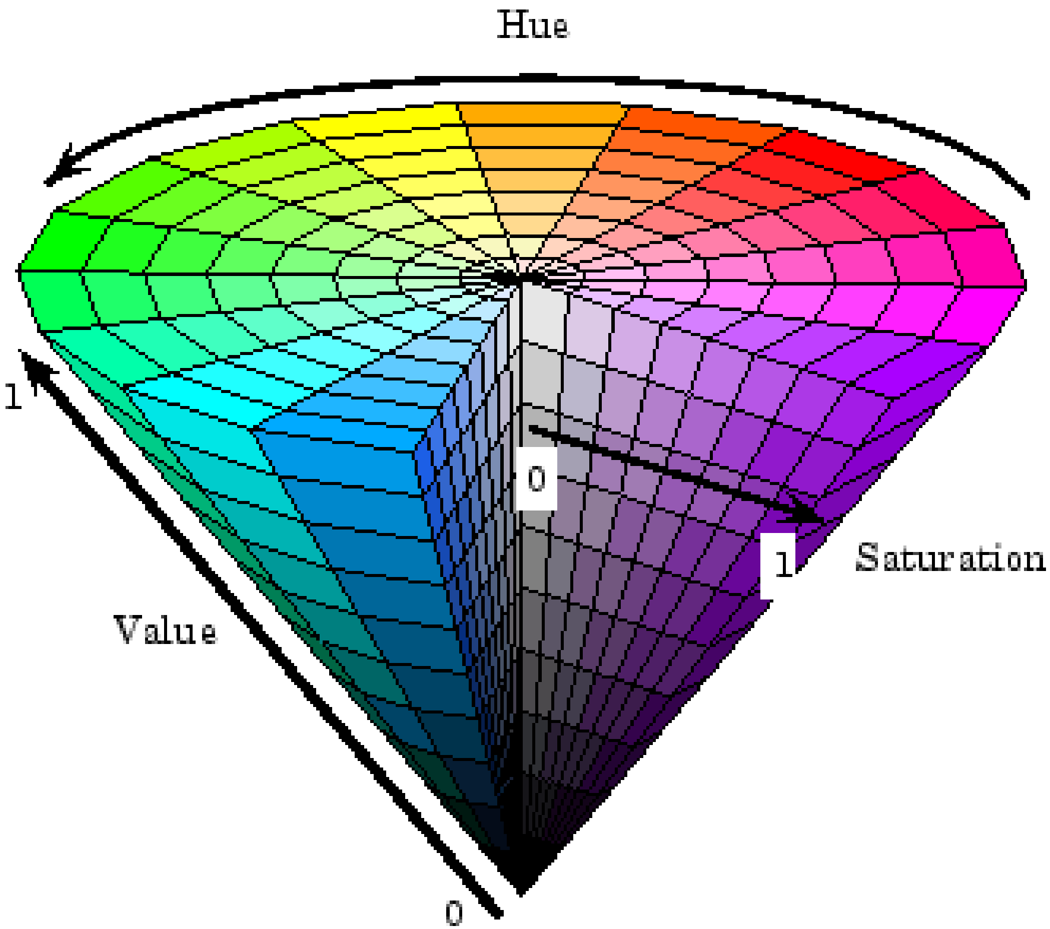
B
(R=0,G=0)

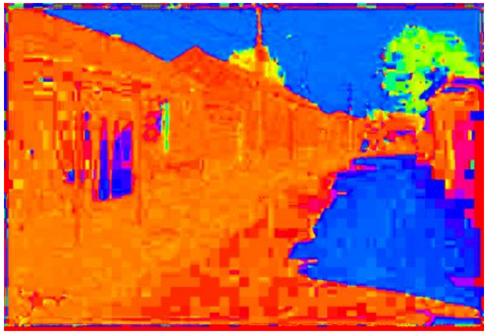
Some drawbacks
- Strongly correlated channels
- Non-perceptual

Image from: http://en.wikipedia.org/wiki/File:RGB_color_solid_cube.png

Colour spaces: HSV

Intuitive colour space





H
(S=1,V=1)

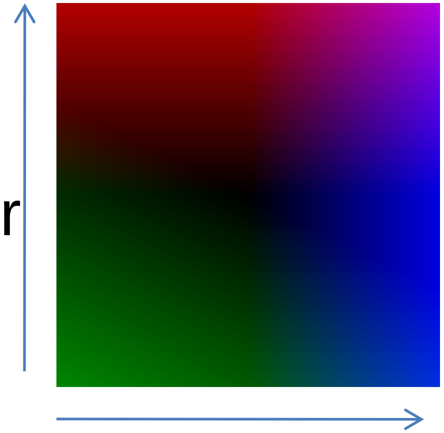
S
(H=1,V=1)

V
(H=1,S=0)

Colour spaces: YCbCr

Fast to compute, good for compression, used by TV

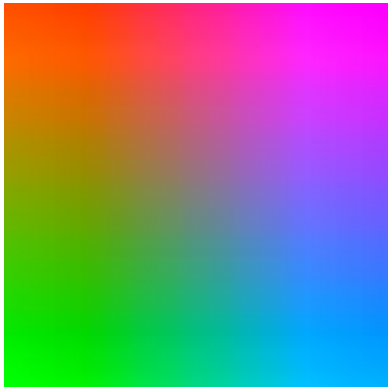
Y=0



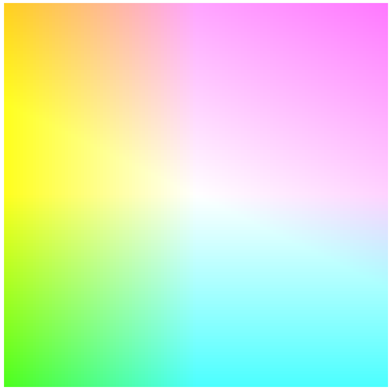
Cr

Cb

Y=0.5



Y=1



Y

(Cb=0.5,Cr=0.5)

Cb

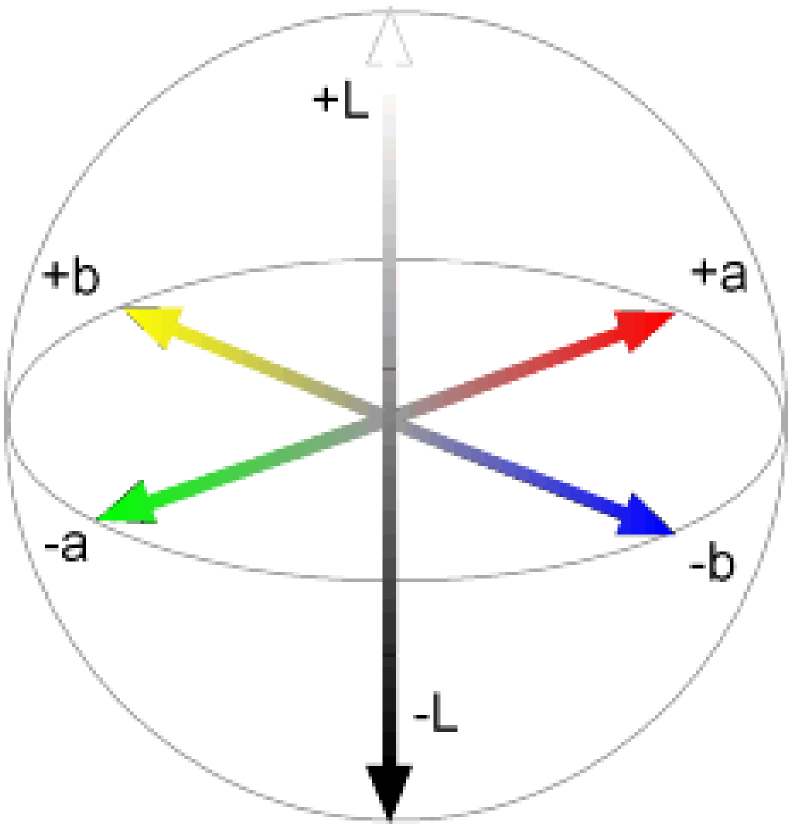
(Y=0.5,Cr=0.5)

Cr

(Y=0.5,Cb=05)

Colour spaces: L*a*b*

"Perceptually uniform"* colour space



L
(a=0,b=0)

a
(L=65,b=0)

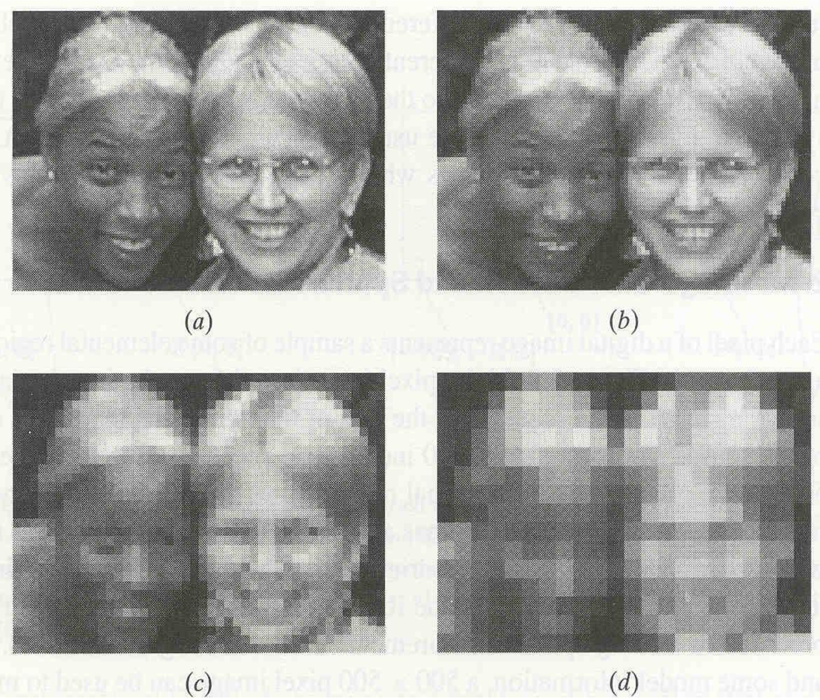
b
(L=65,a=0)

Digitisation and Sampling

- Digitisation: converts analog image to digital image
- Sampling digitises the coordinates x and y:
  - *spatial discretisation* of picture function F (x,y)
  - use a grid of sampling points, normally rectangular: image sampled at points $x = j\,\Delta x$, $y = k\,\Delta y$, $j = 1 \ldots M$, $k = 1 \ldots N$.
  - Δx, Δy called the **sampling intervals.**

Spatial Resolution

- Spatial Resolution: number of pixels per unit of length
- Resolution decreases by one half- see right
- Human faces can be recognized at 64 x 64 pixels per face



- Appropriate resolution is essential:
  - too little resolution, poor recognition
  - too much resolution, slow and wastes memory

Quantisation

- Quantisation digitises the intensity or amplitude values, ie F (x, y)
  - called intensity or gray level quantisation
  - Gray-level resolution:
    - usually has 16, 32, 64, ..., 128, 256 levels
    - number of levels should be high enough for human perception of shading details - human visual system requires about 100 levels for a realistic image.

For Reading

- Szeliski, Chapter 2
- Shapiro and Stockman, Chapter 2

Acknowledgement

- Slides from Derek Hoiem, Alexei Efros, Steve Seitz, and David Forsyth
- Image sources credited where possible
- Some material, including images and tables, were drawn from the referenced textbooks and associated online resources.