Applied Computer Vision

David Vernon
Carnegie Mellon University Africa

vernon@cmu.edu
www.vernon.eu

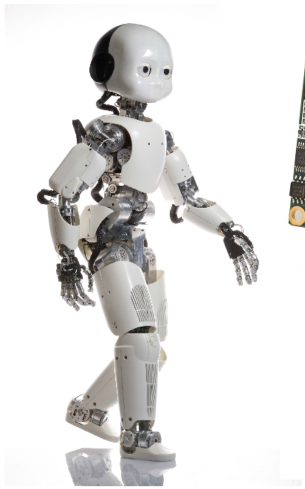
Lecture 25

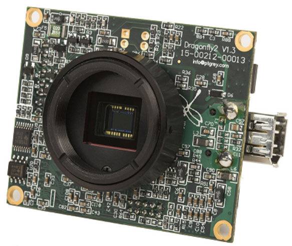
3D Vision II

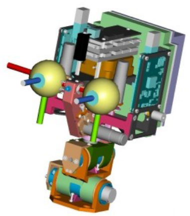
Stereopsis, stereo correspondence, epipolar geometry, depth cues, structured light

Stereopsis
Stereo Vision

We refer to stereo vision as the problem of inferring 3D information (structure and distances) from two or more images taken from different viewpoints

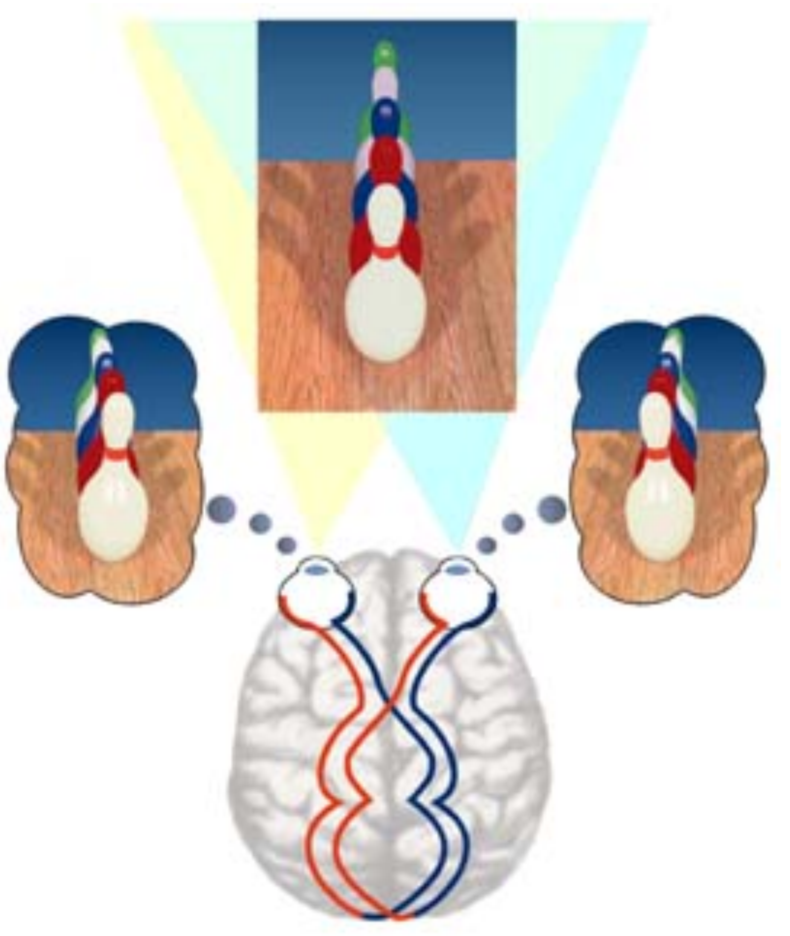






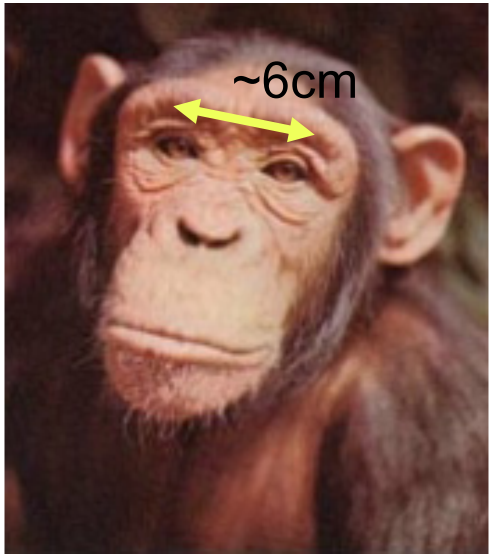
Credit: Francesca Odone, University of Genova

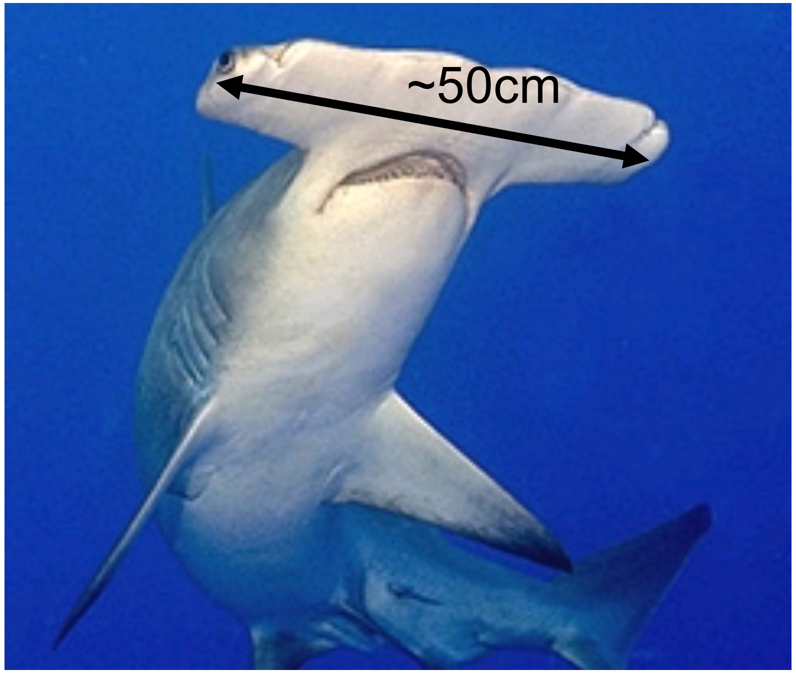
Left and Left Image



Credit: Markus Vincze, Technische Universität Wien

Stereo Vision in Nature





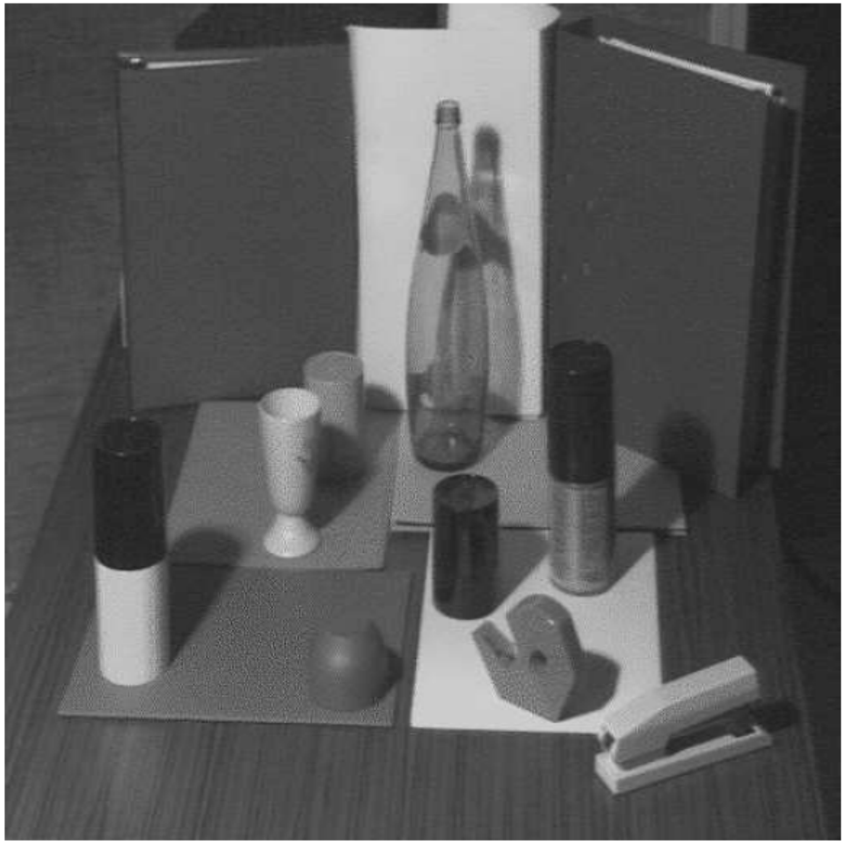
- Larger baseline increases useful range of depth estimated from stereo
- After a few meters disparity is quite small and depth from stereo is unreliable …

Left and Left Image

[Young]

Credit: Markus Vincze, Technische Universität Wien

Left and Right Image

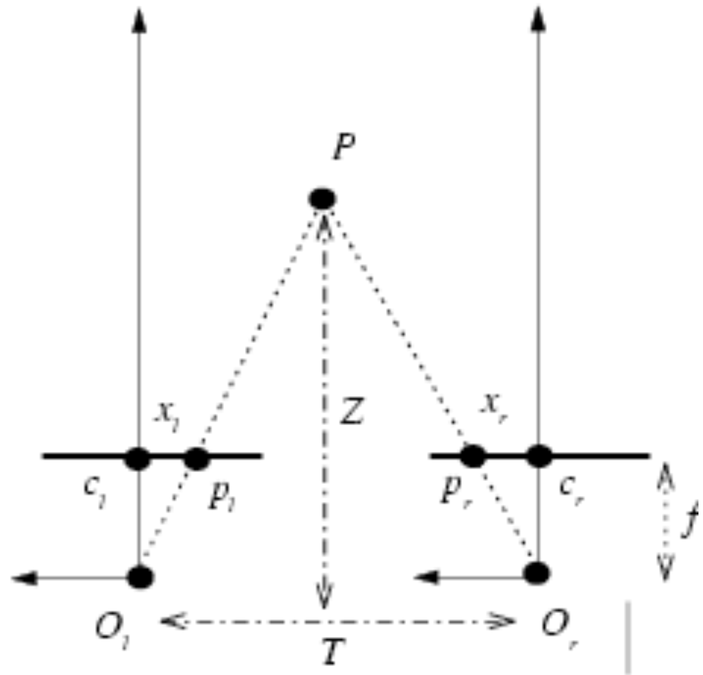
[Young]

Credit: Markus Vincze, Technische Universität Wien

Stereopsis – Stereo Vision

- Disparity d is the relative distance between corresponding points (on the image plane)
- Depth Z is the distance from a 3D point to the viewing system
- Depth is inversely proportional to disparity

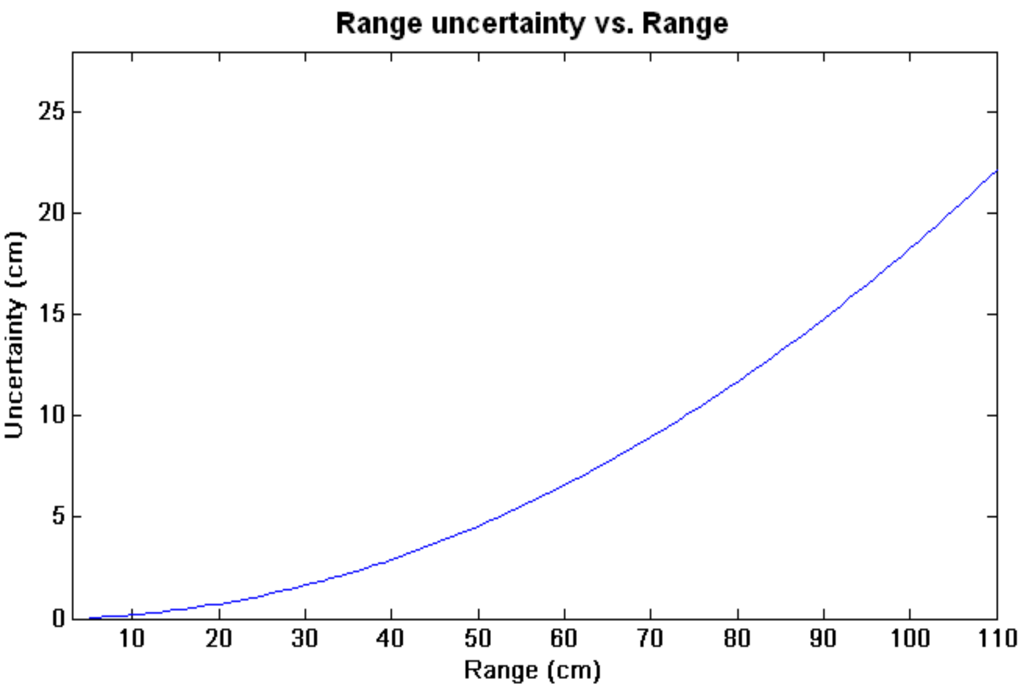
$$Z = \frac{fT}{x_r - x_l} = \frac{fT}{d}$$



Credit: Francesca Odone, University of Genova

Increasing Depth Uncertainty

- The more distant the object, the larger the depth uncertainty
  - Acute angle: disparity uncertainty grows non-linearly
  - Improve with large focal length and baseline distance
- Humans use stereo only up to arm length
  - Then relative and perspective cues dominate



[S. Ahuja]

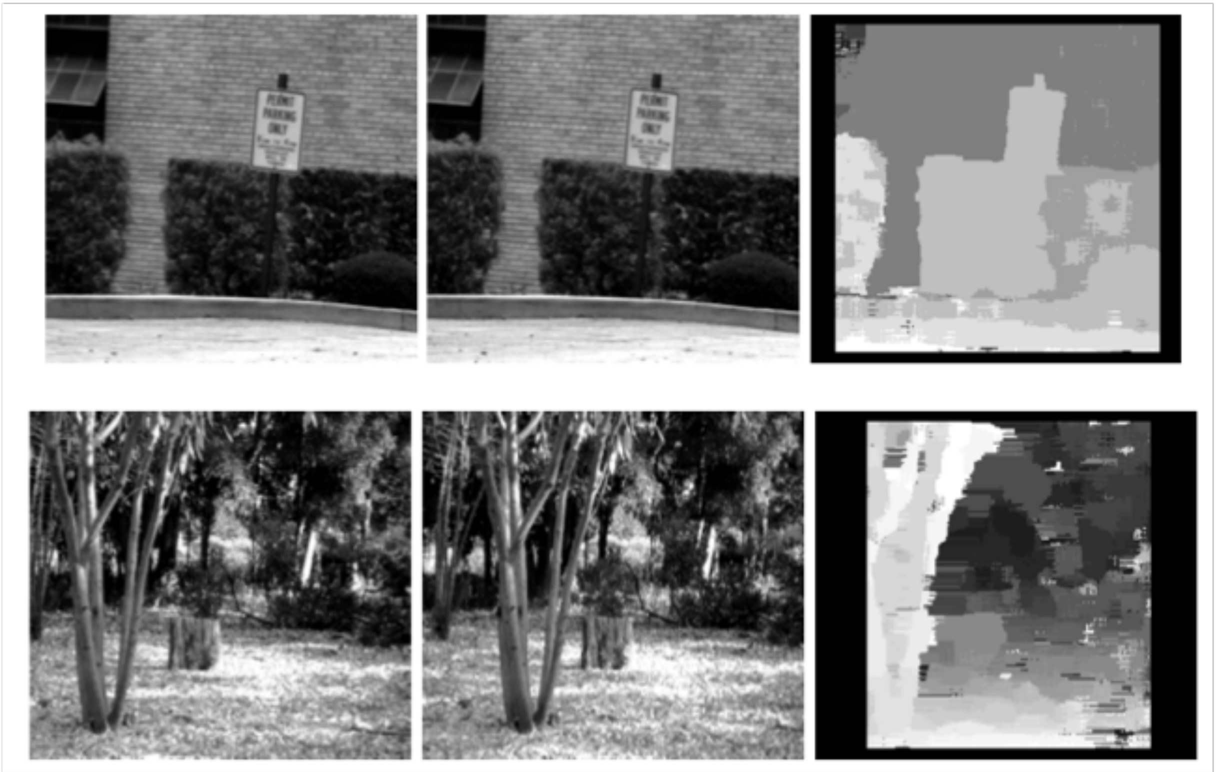
Stereopsis – Stereo Vision

- Dense stereo correspondence
- We assume we have two **rectified** images
  - where conjugate points lie on corresponding scanlines of the image ("rows")
- Our goal is to obtain a **disparity map** giving the relative displacement for each pixel

Credit: Francesca Odone, University of Genova

Stereopsis – Stereo Vision

Assuming a fixation point at infinity, disparity is proportional to the inverse of the distance $Z = \frac{fT}{x_r - x_l} = \frac{fT}{d}$

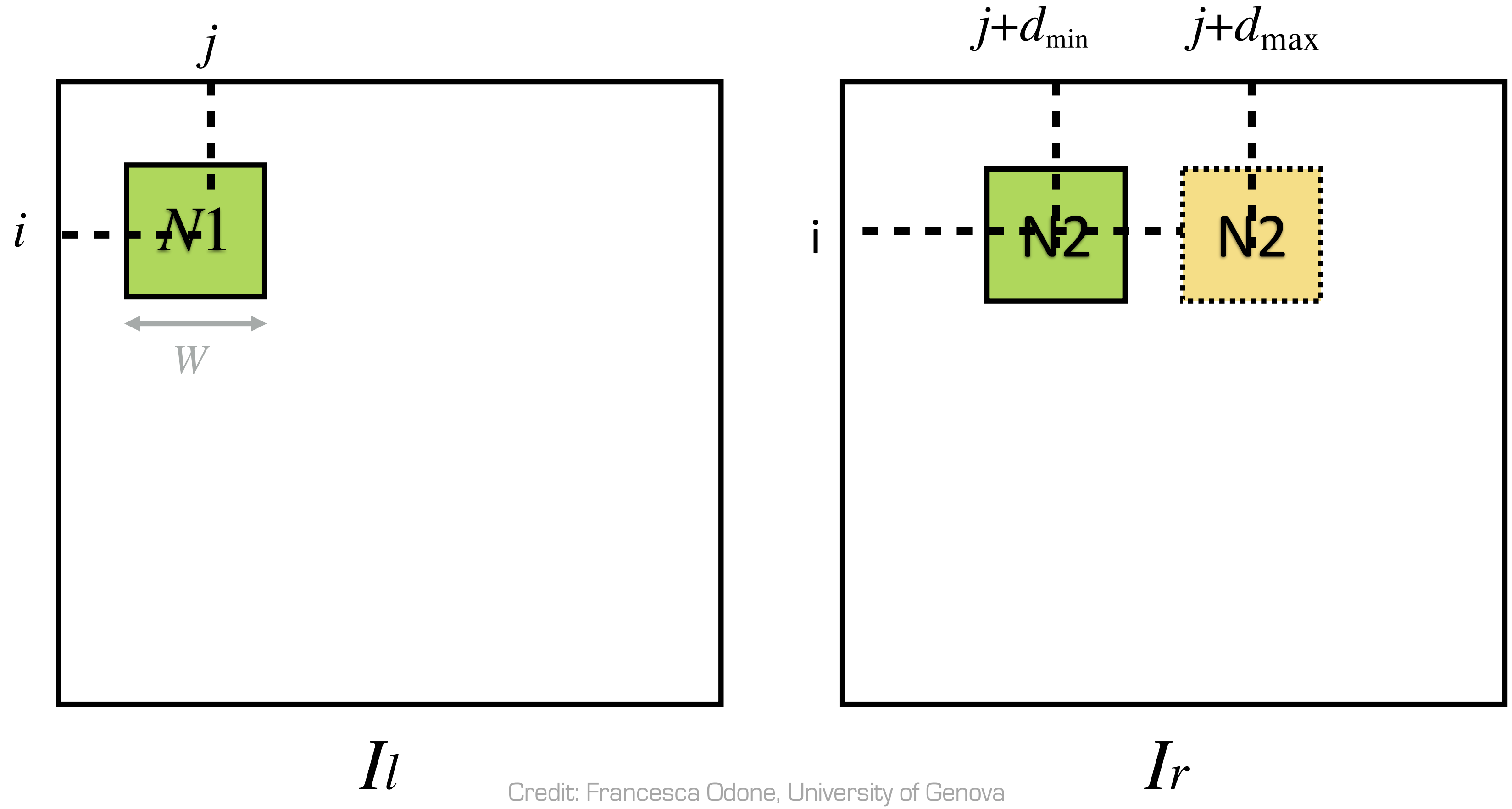


Credit: Francesca Odone, University of Genova

Stereopsis – Stereo Vision

Given a stereo pair of **rectified** images Il and Ir

- size of a correlation window W
- a search range $[d_{min}, d_{max}]$



Credit: Francesca Odone, University of Genova

Stereopsis – Stereo Vision

for each pixel (i, j) in Il

for each disparity d in the search range

estimate the similarity

$$c(d) = \phi(N1(i,j), N2(i,j+d))$$

the disparity of the pixel is

$$\bar{d} = \text{argmax}_{d \in [d_{min}, d_{max}]} \{c(d)\}$$

Credit: Francesca Odone, University of Genova

Stereopsis – Stereo Vision

Dense correspondences: left-right consistency

- Correspondences are made more difficult by occlusions (points with no counterpart on the other image)

- Let us compute

  D_{lr} disparity map from I_l to I_r

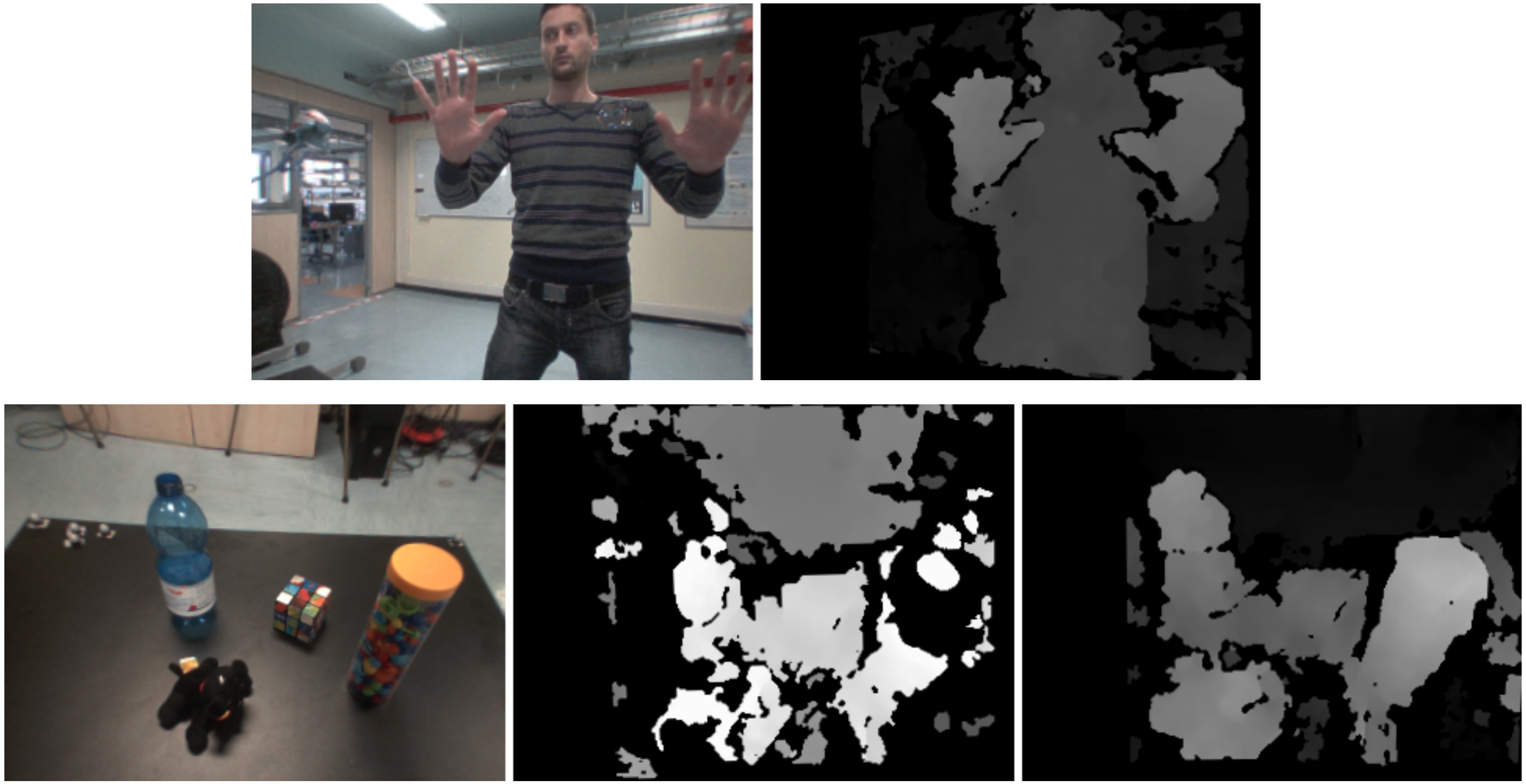
  D_{rl} disparity map from I_r to I_l

- then $D(i, j)=d$ iff $Dlr(i, j) = -Drl\ (i, j+d) = d$

Credit: Francesca Odone, University of Genova

Stereopsis – Stereo Vision

Dense correspondences: left-right consistency



Credit: Francesca Odone, University of Genova

Correspondence Problem

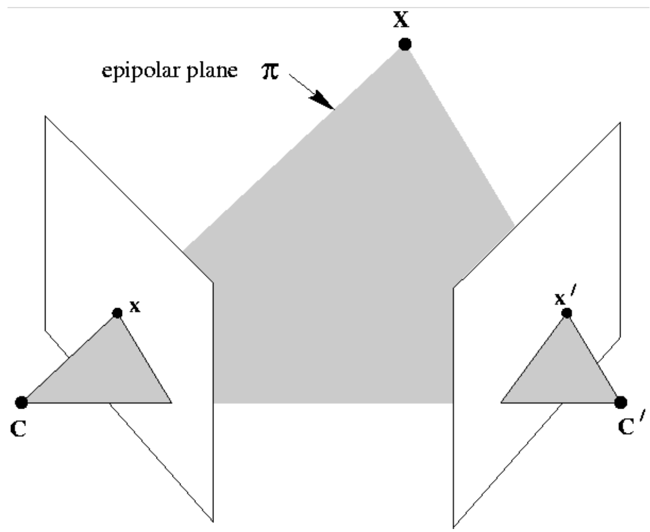
- Finding the *same* point in both images
- Search in entire image is very costly
- Geometry of cameras produces constraints: epipolar plane and epipolar line
  - Limits search to a line in the image
- Finding the same points
  - Correlation (region) or features (edge)

Credit: Markus Vincze, Technische Universität Wien

Stereopsis – Stereo Vision

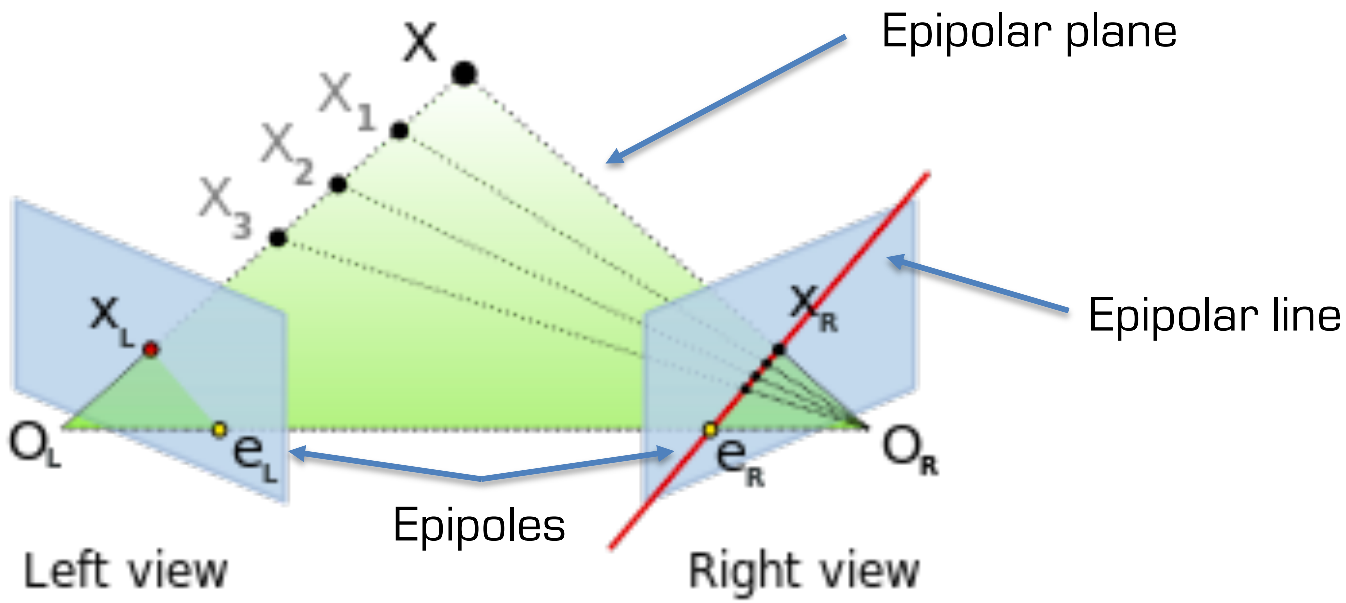
- The geometry of a stereo-system is called epipolar geometry

- It provides a geometrical prior to the algorithms



Credit: Francesca Odone, University of Genova

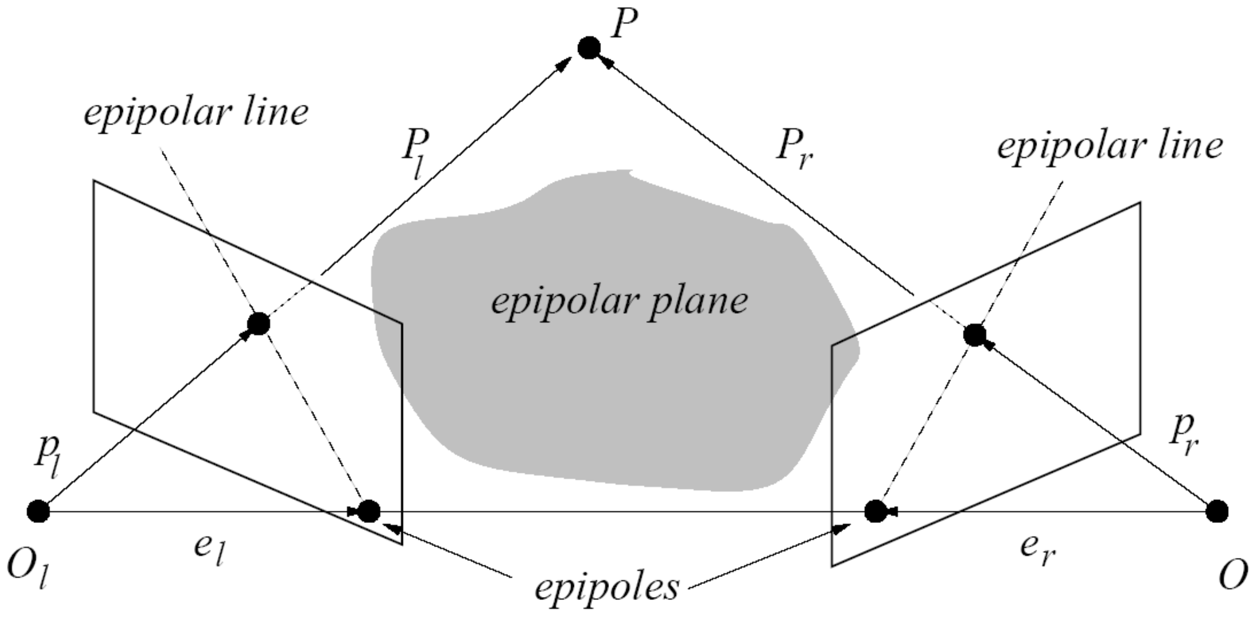
Epipolar Geometry



https://en.wikipedia.org/wiki/Epipolar_geometry

Epipolar Geometry

- Epipolar plane: plane of the two visible rays
- Pre-condition: known camera geometry, calibration

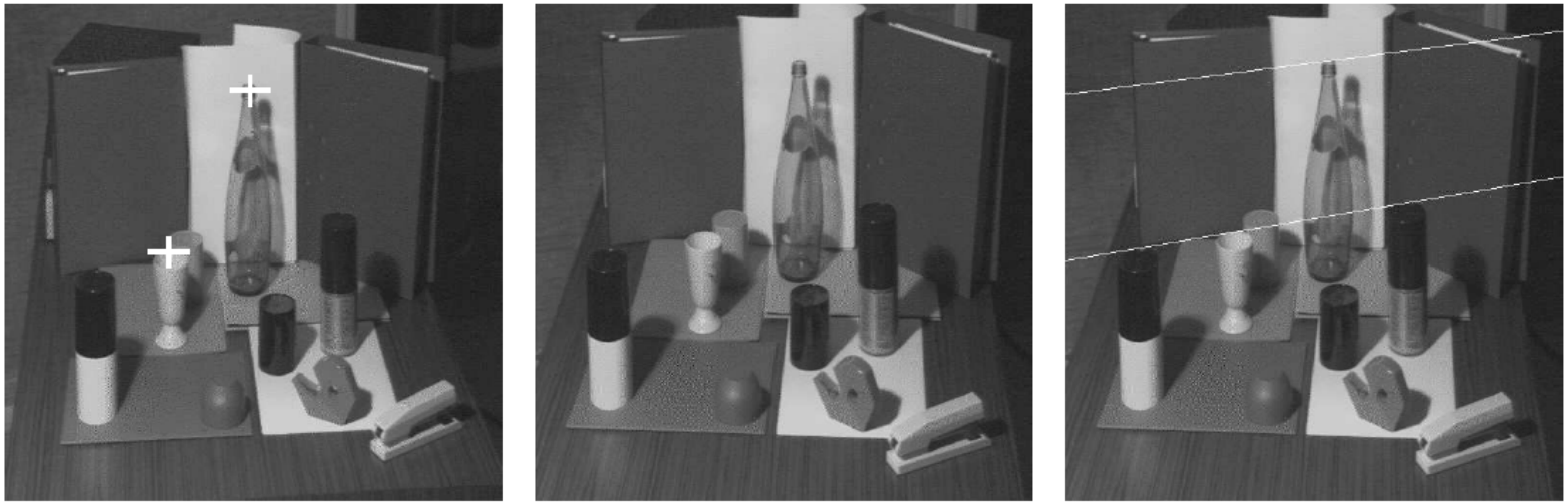


[Young]

Credit: Markus Vincze, Technische Universität Wien

Epipolar Lines

- Each point left defines epipolar line right
- → 1D search for the same feature
- Simplifies correspondence problem



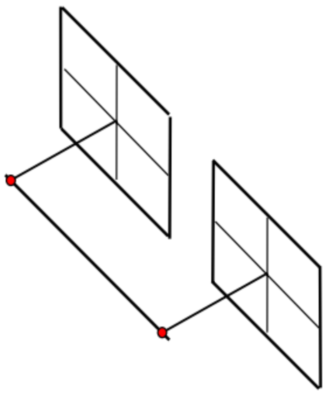
[Young]

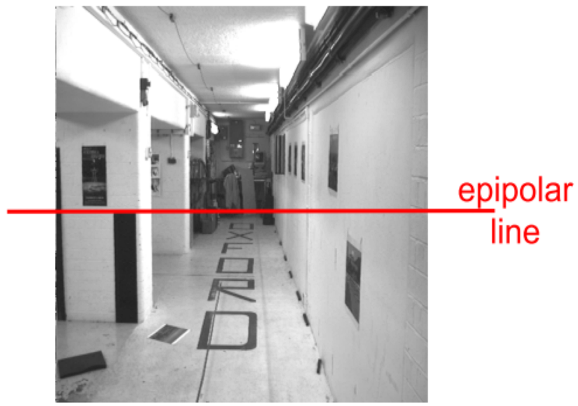
Credit: Markus Vincze, Technische Universität Wien

Horizontal Epipolar Lines

- Simple case: parallel cameras (fronto-parallel stereo)
  - In practice not obtainable accurately

- Rectification (calibration and elimination of distortions) of the images to obtain epipolar lines on the pixel array of the camera

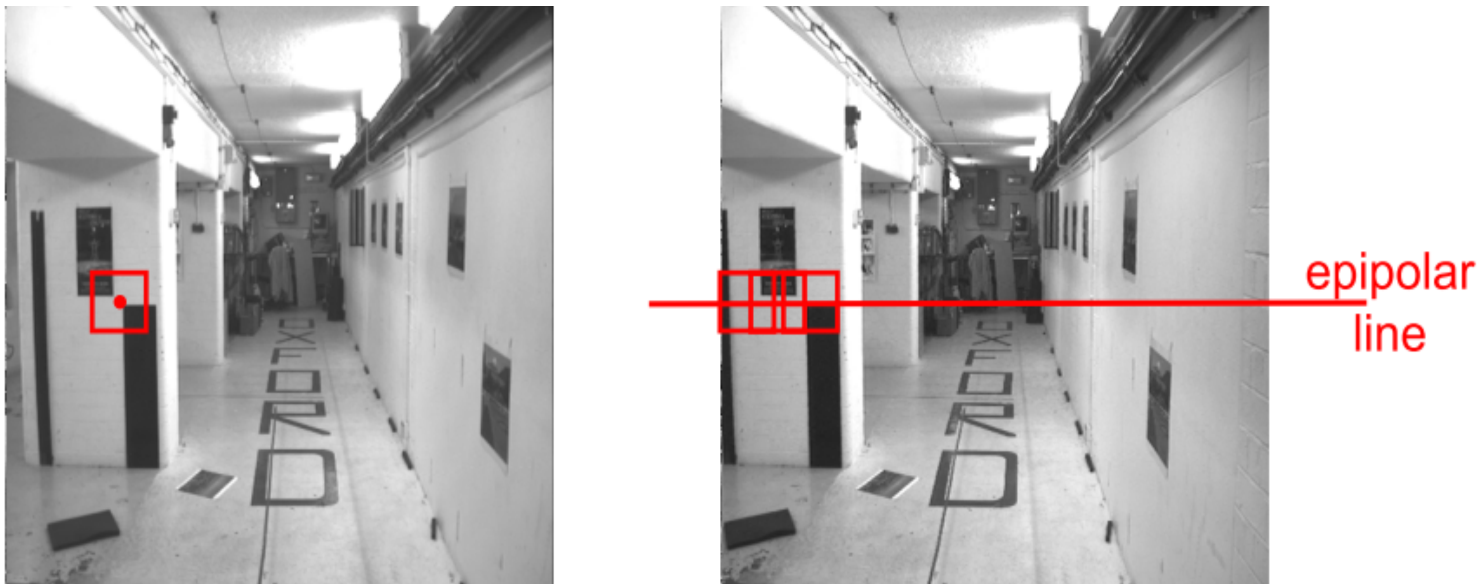




Credit: Markus Vincze, Technische Universität Wien

Correspondence along a Line

- Search for left image point in the right image
- Dense depth image: correspondence for every point
- Sparse depth image: only distinctive points



Credit: Markus Vincze, Technische Universität Wien

Other Cues to Depth

Absolute depth cues (assuming known camera parameters)

- information about the absolute depth between the observer and elements of the scene

Relative depth cues

- relative information about depth between elements in the scene (this point is twice as far at that point, …)

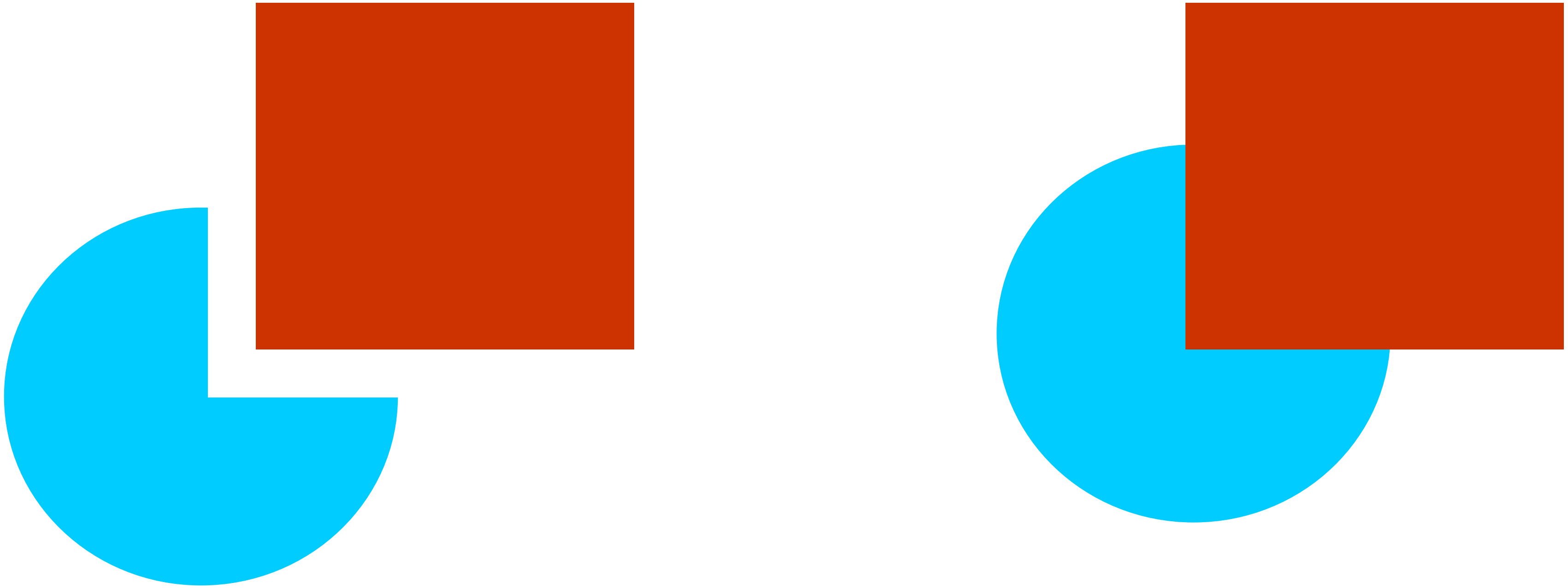
Credit: Markus Vincze, Technische Universität Wien

Relative Depth Cues

- Simple and powerful cue, but hard to make it work in practice
- Requires scene understanding

Credit: Markus Vincze, Technische Universität Wien

Interposition / Occlusion



Credit: Markus Vincze, Technische Universität Wien

Texture Gradient

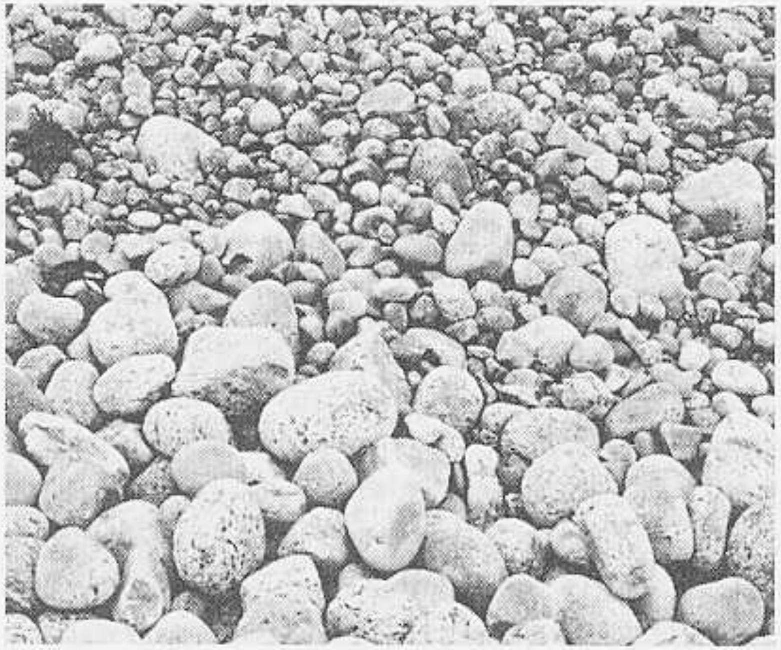


FIGURE 8.27
Texture gradients provide information about depth. (Frank Siteman/Stock, Boston.)
© Frank Sitman/Stock Boston

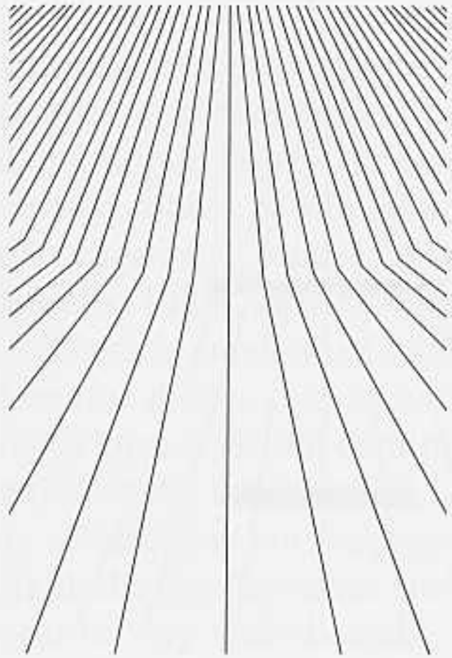
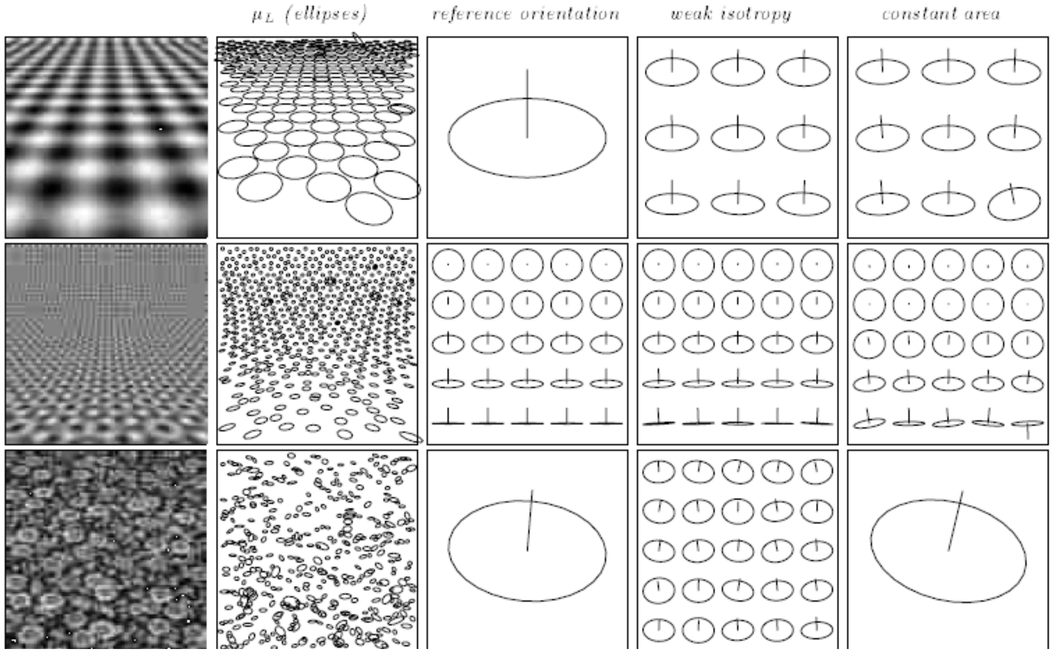


FIGURE 8.28
Texture discontinuity signals the pre
corner.

[A. Witkin: Recovering Surface Shape and Orientation from Texture, 1981]

Credit: Markus Vincze, Technische Universität Wien

Texture Gradient

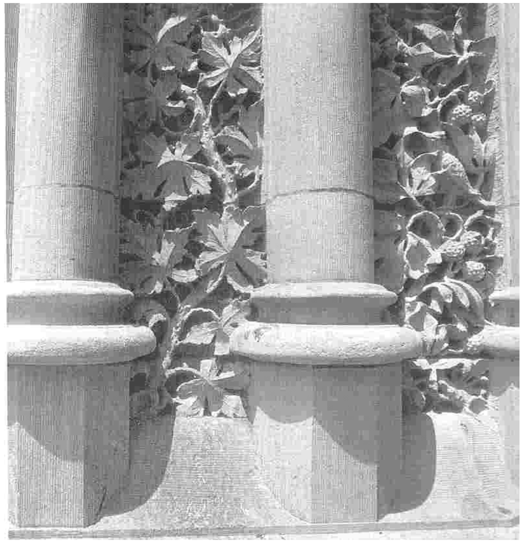


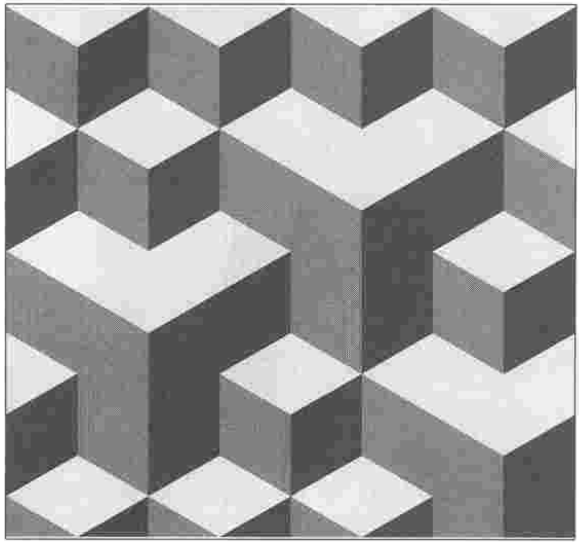
T. Lindeberg, J. Garding: Shape from Texture from a Multi-Scale Perspective, ICCV 93]

Credit: Markus Vincze, Technische Universität Wien

Illumination: Shading

Based on 3 dimensional modeling of objects in light, shade and shadows.

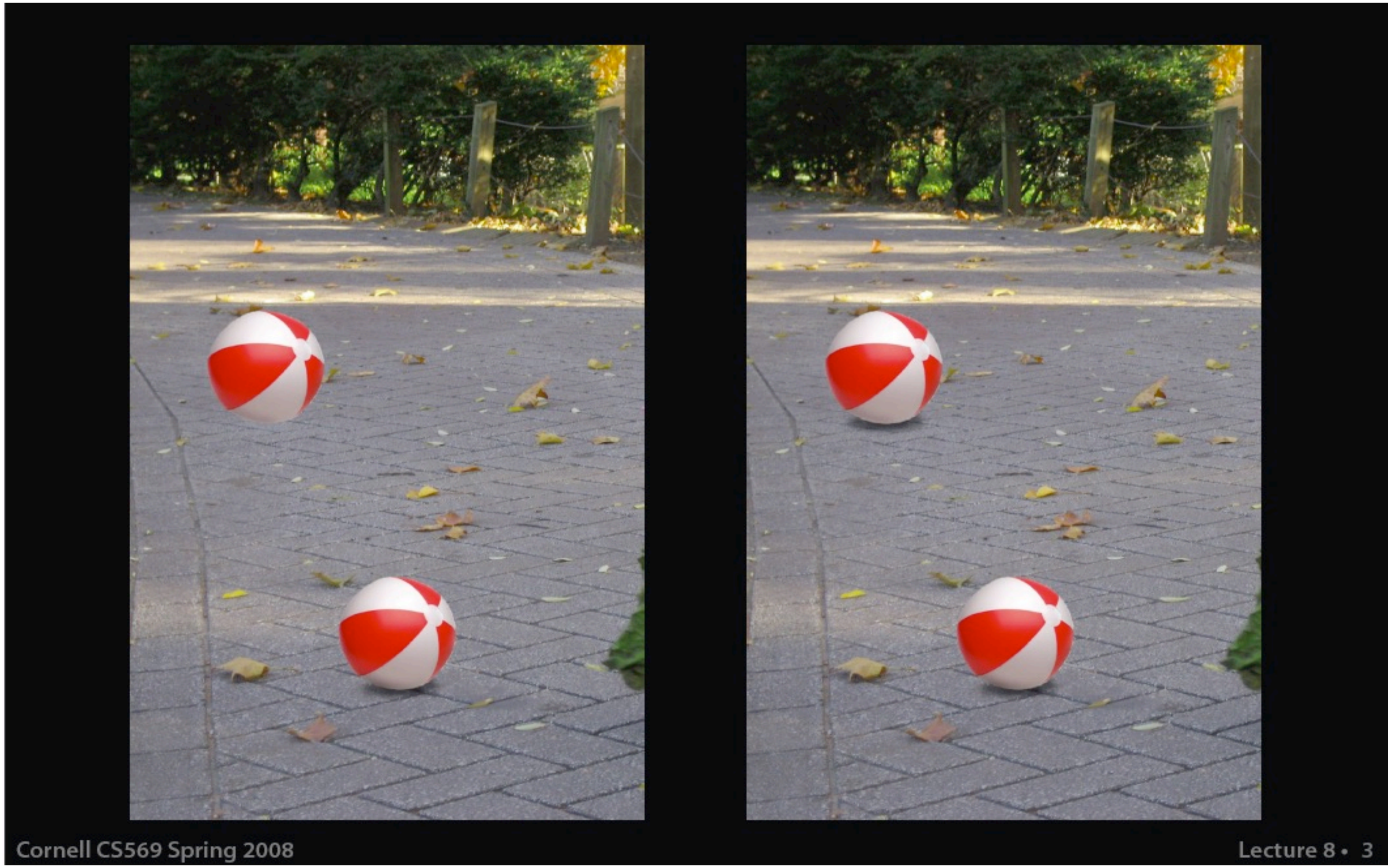




- Perception of depth through shading alone is always subject to the concave/convex inversion
- Stairs receding towards the top and lighted from above, or
- Overhanging structure lighted from below?

Credit: Markus Vincze, Technische Universität Wien

Shadows



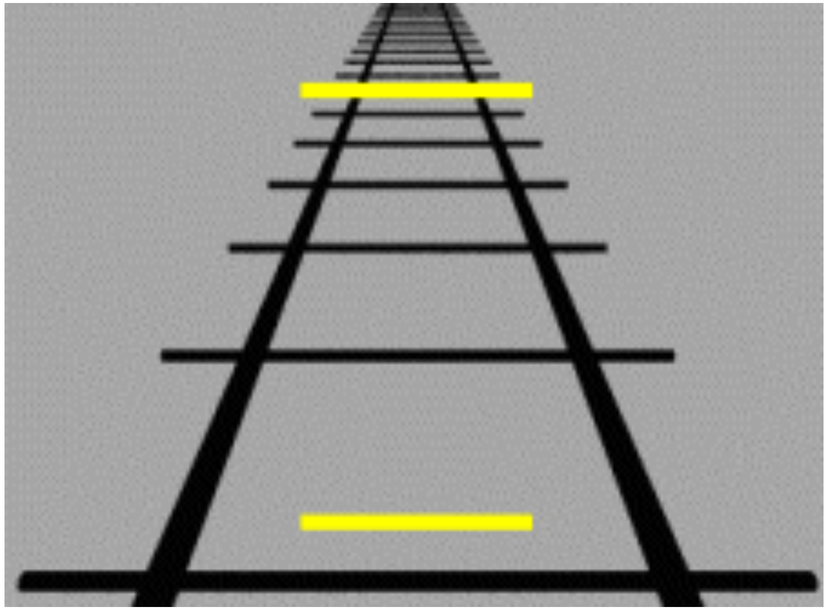
Slide by Steve Marschner

http://www.cs.cornell.edu/courses/cs569/2008sp/schedule.stm

Credit: Markus Vincze, Technische Universität Wien

Linear Perspective

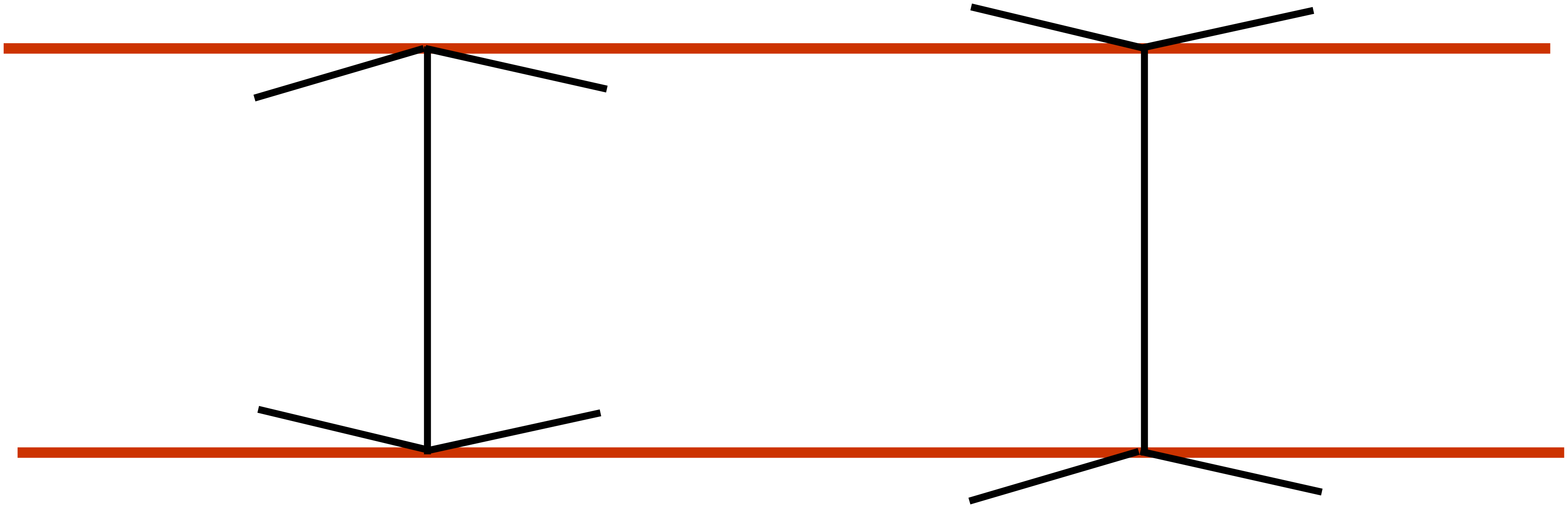
- Perspective strong cue for depth
- A representations specific to one individual
- Changes with space and time



Ponzo's illusion

Credit: Markus Vincze, Technische Universität Wien

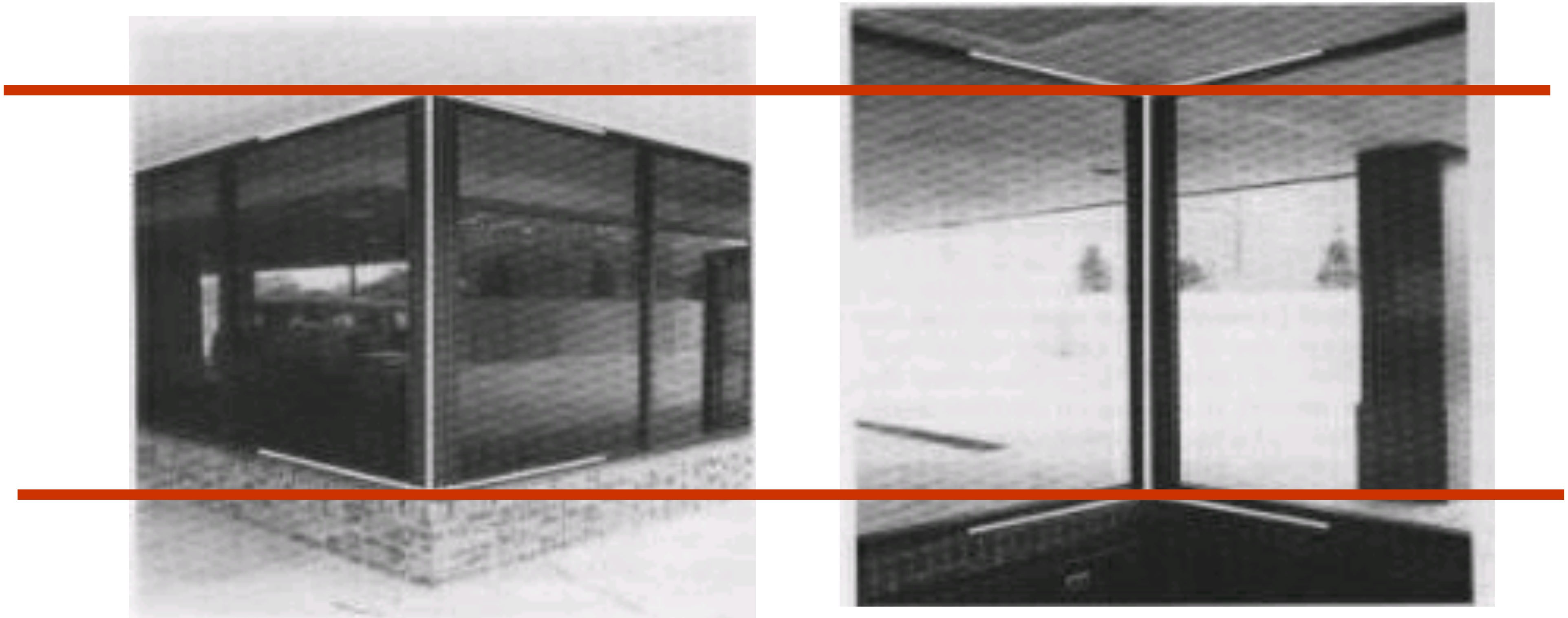
Linear Perspective



Muller-Lyer 1889

Credit: Markus Vincze, Technische Universität Wien

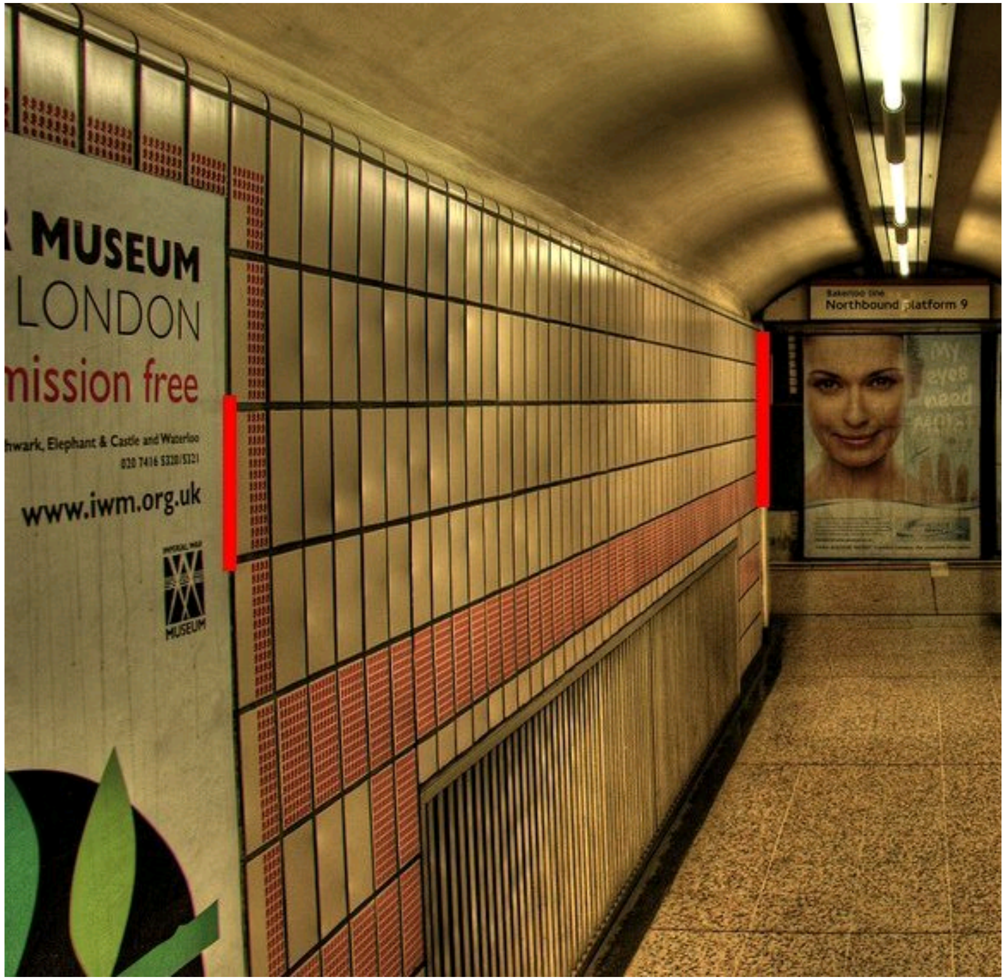
Linear Perspective



Muller-Lyer 1889

Credit: Markus Vincze, Technische Universität Wien

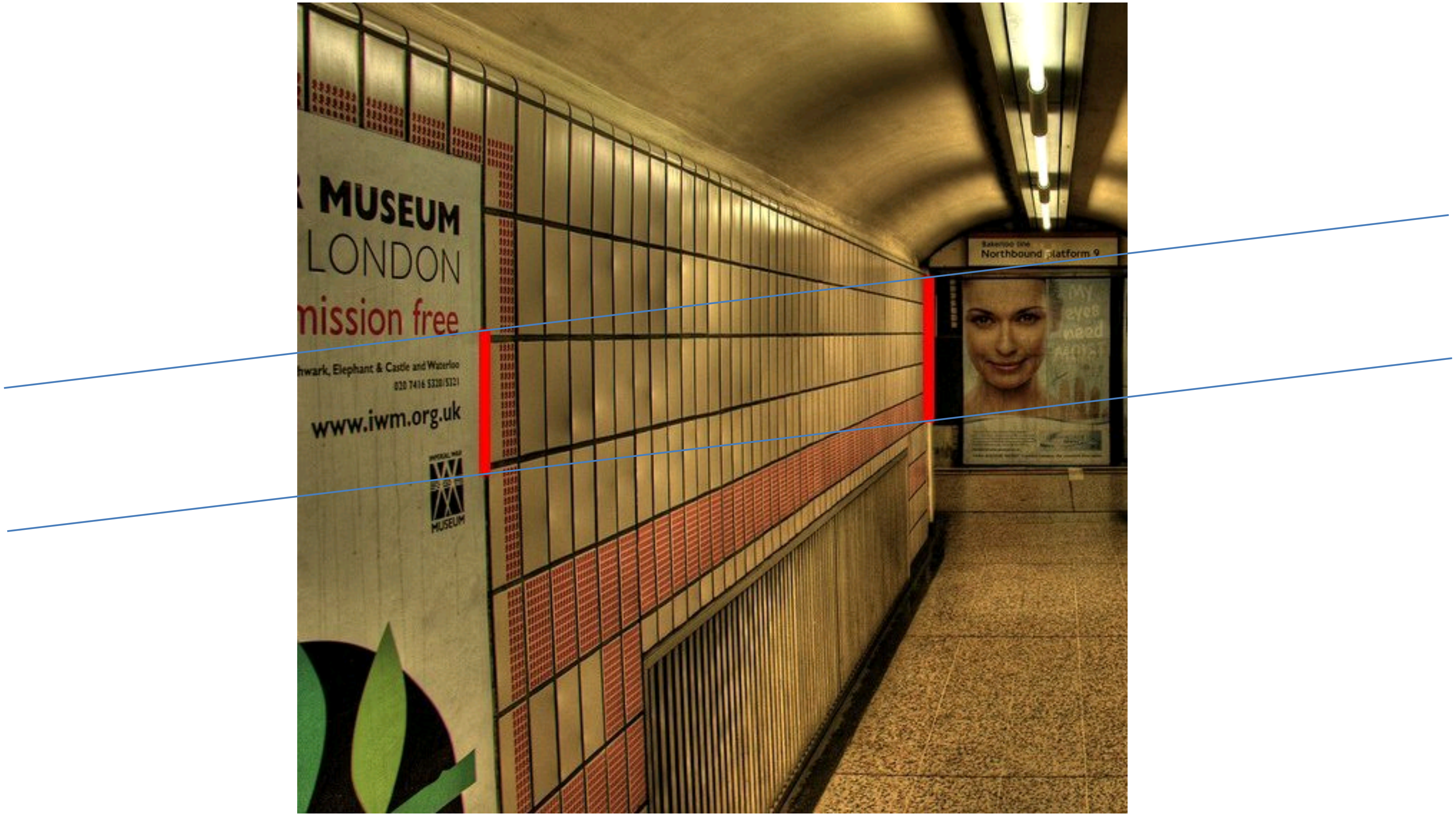
Linear Perspective



(c) 2006 Walt Anthony

Credit: Markus Vincze, Technische Universität Wien

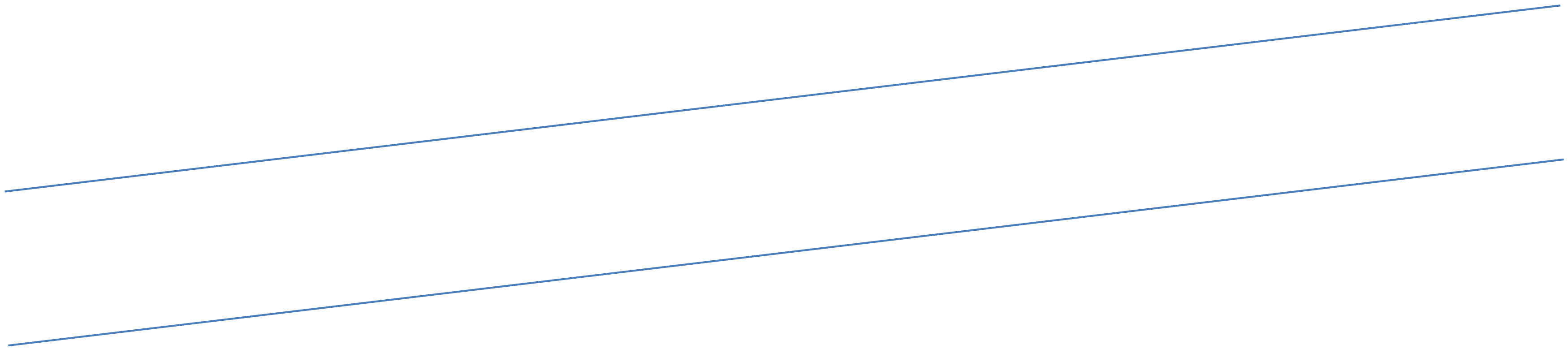
Linear Perspective



(c) 2006 Walt Anthony

Credit: Markus Vincze, Technische Universität Wien

Linear Perspective



Credit: Markus Vincze, Technische Universität Wien

Absolute (Monocular) Depth Cues

Are there any monocular cues that can give us absolute depth from a single image?

Credit: Markus Vincze, Technische Universität Wien

Familiar size

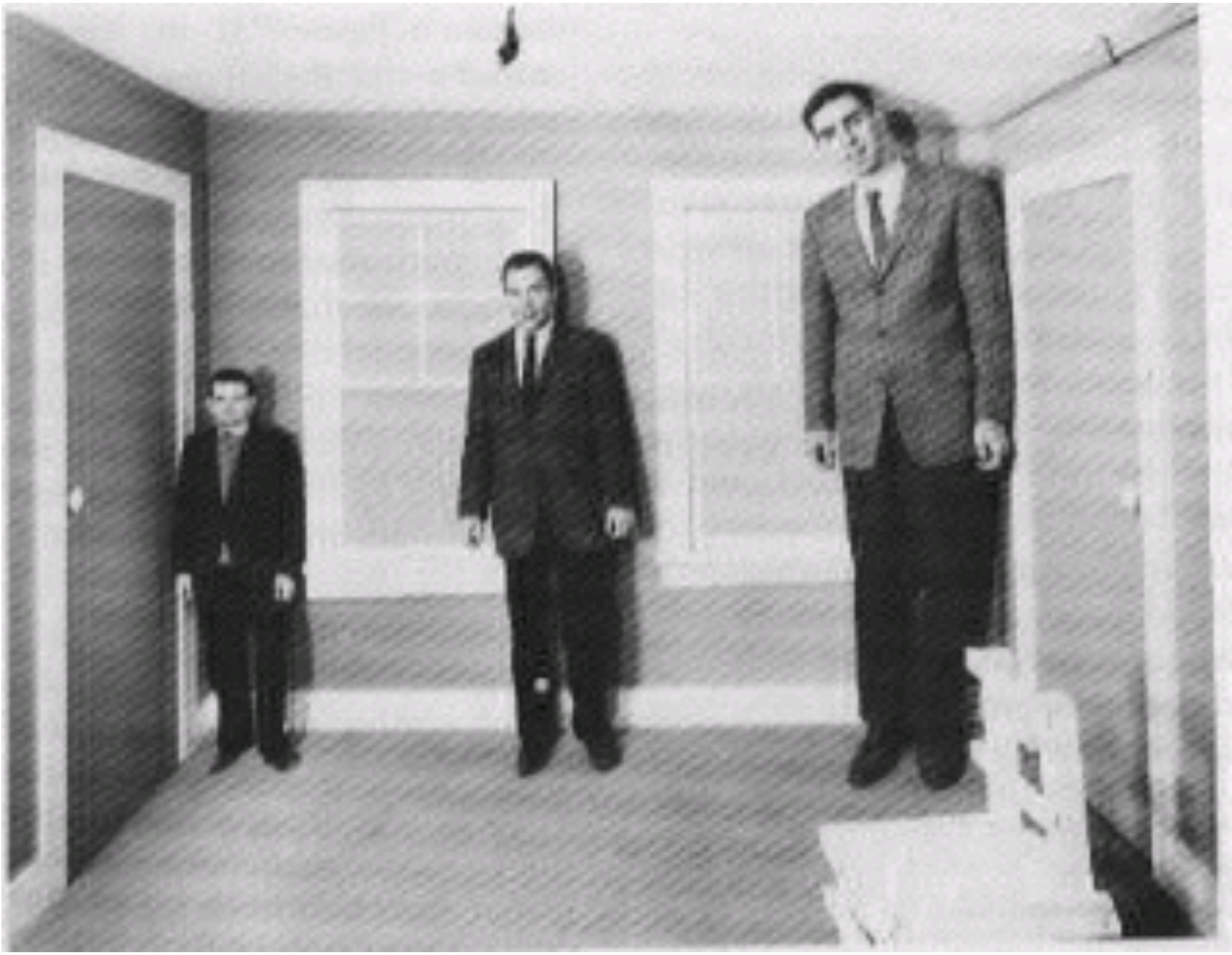
Which “object” is closer to the camera?
How close?

Familiar Size

- Apparent reduction in size of objects at a greater distance from the observer
- Size perspective is thought to be conditional, requiring knowledge of the objects
- But, material textures also get smaller with distance, so possibly, no need of perceptual learning?

Credit: Markus Vincze, Technische Universität Wien

Perspective vs. Familiar Size



3D percept is driven by the scene,
which imposes its ruling to the objects

Credit: Markus Vincze, Technische Universität Wien

Scene vs. objects

[*The Listening Room*
Rene Magritte]

What do you see? A big apple or a small room?

I see a big apple and a normal room.
The scene seems to win again?

Credit: Markus Vincze, Technische Universität Wien

Scene vs. objects

[*Personal Values* Rene Magritte]

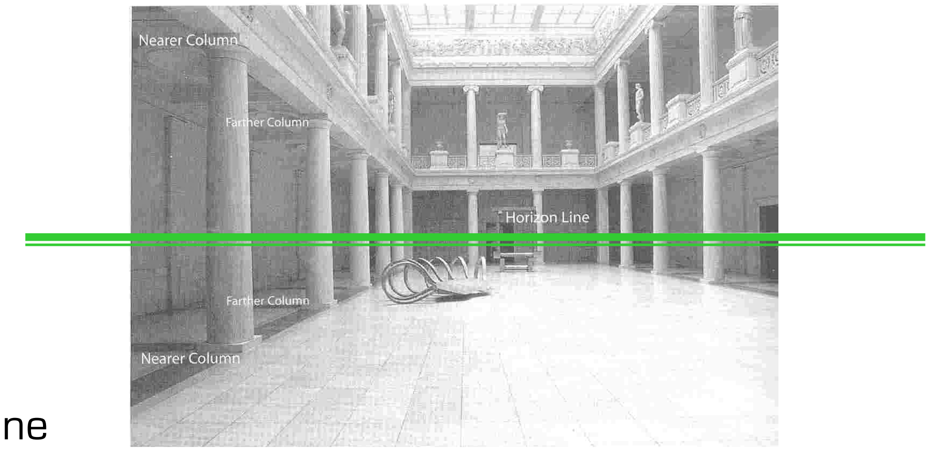
Credit: Markus Vincze, Technische Universität Wien

Importance of Horizon Line

Tendency of objects to appear nearer horizon line with greater distance to the horizon

Objects approach horizon line with greater distance from viewer

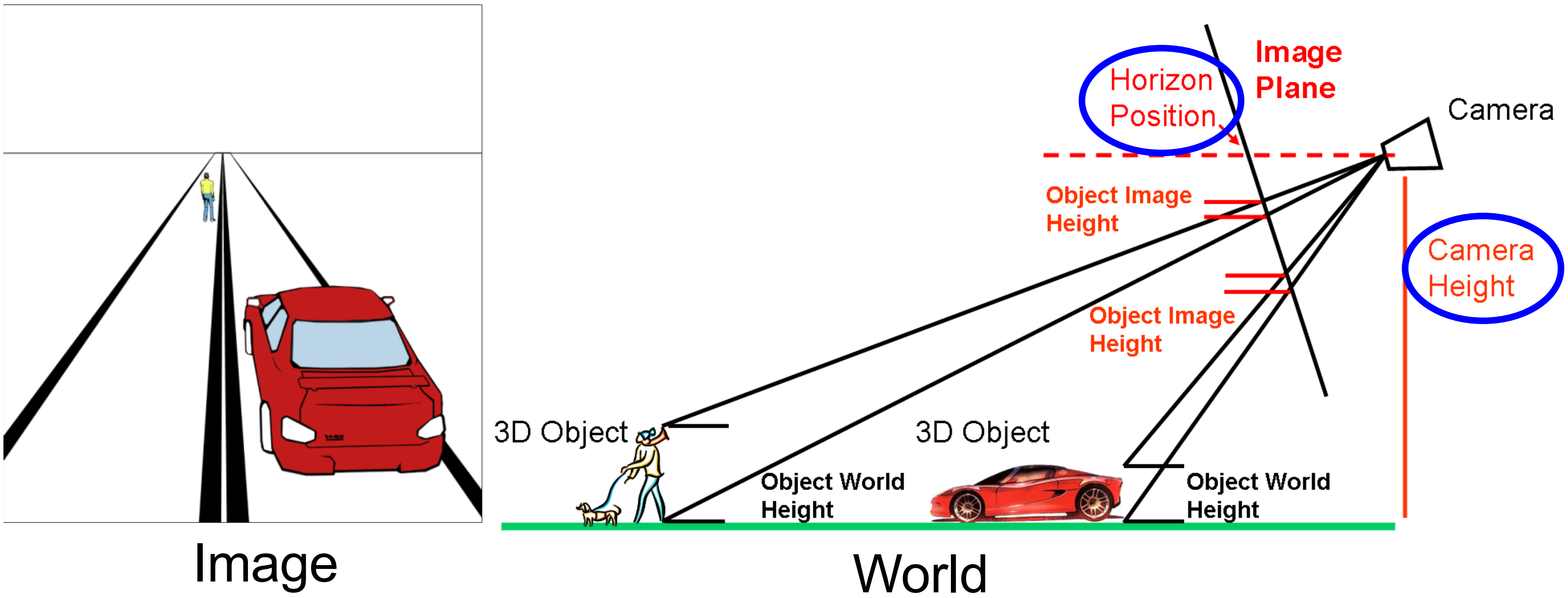
- Near column: base lower, lower than background floor, further from horizon
- Distant column: base higher against the same floor, and thus nearer to horizon line



Credit: Markus Vincze, Technische Universität Wien

Object Size in the Image

If you know camera parameters: height of the camera, then we know real depth



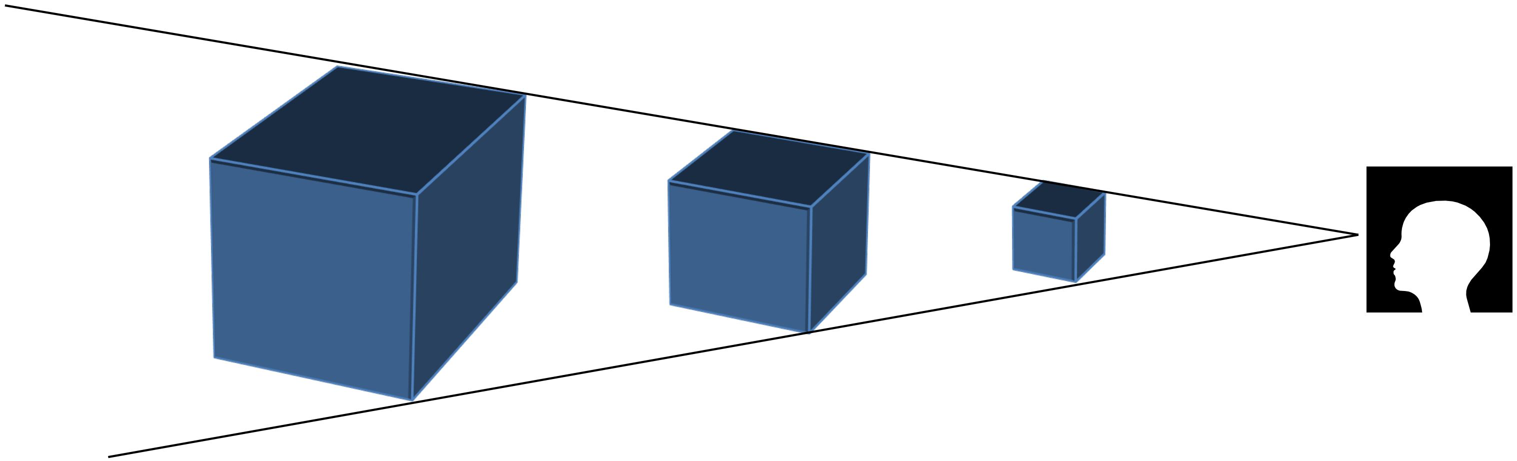
[Derek Hoiem]

Credit: Markus Vincze, Technische Universität Wien

Depth Perception from Image Structure

Stimulus ambiguity: the three cubes produce the same retinal image

Monocular information cannot give absolute depth measurements

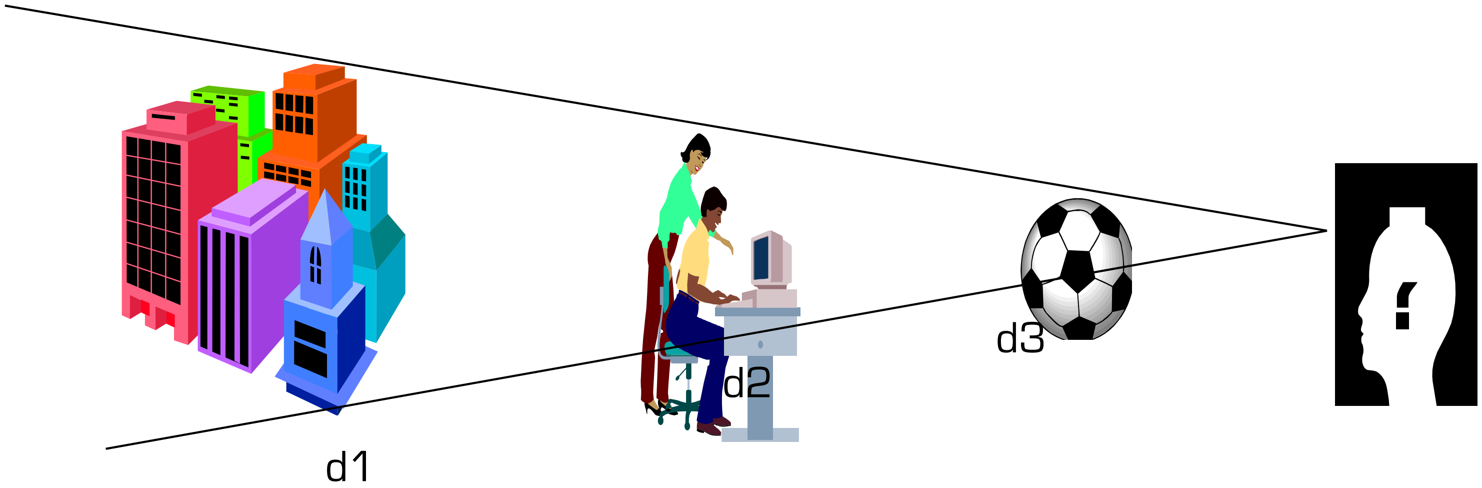


Only relative depth information such as shape from shading and junctions (occlusions) can be obtained

Credit: Markus Vincze, Technische Universität Wien

Depth Perception from Image Structure

However, nature (and man) do not build in the same way at different scales



If d1>>d2>>d3 structures of each view strongly differ. Structure provides monocular information about the scale (mean depth) of the space in front of the observer

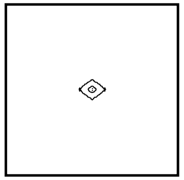
Credit: Markus Vincze, Technische Universität Wien

Image Statistics and Scene Scale

Close-up views

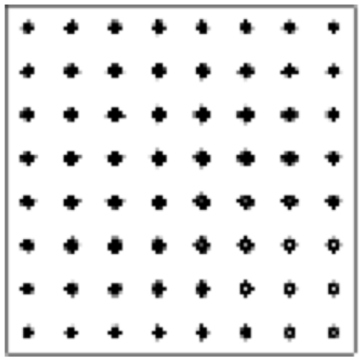
Large scenes

On average, low clutter

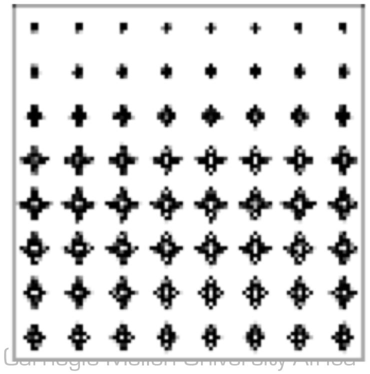


On average, highly cluttered

Point of view is unconstrained



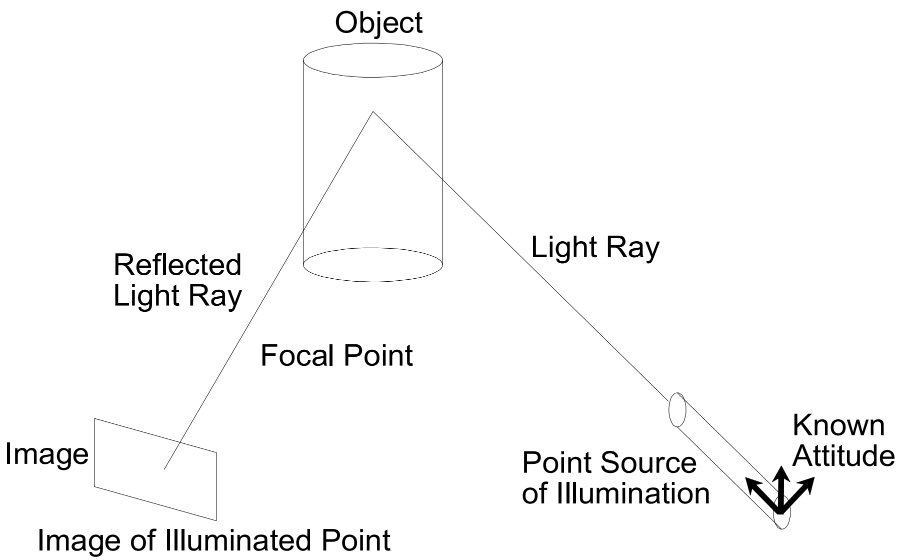
Point of view is strongly constrained



Structured Light

Scanning an object or scene with a point source of illumination, e.g. laser

- scan the surface with the dot of light, computing the range at each point
- or illuminate more than one point at the same time.

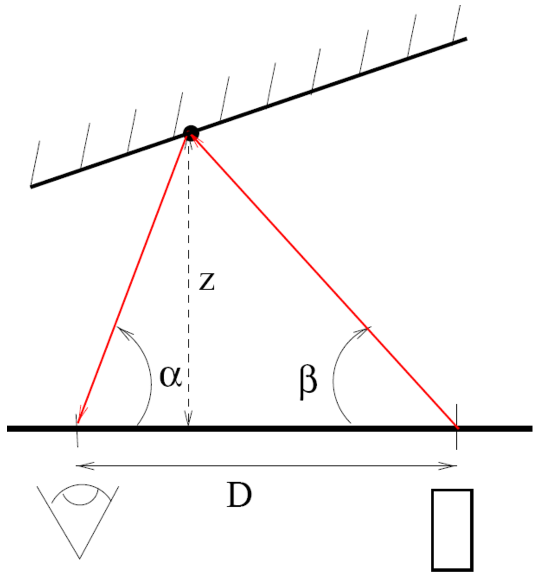


Credit: Markus Vincze, Technische Universität Wien

Structured Light

Scanning an object or scene with a point source of illumination, e.g. laser

- scan the surface with the dot of light, computing the range at each point
- or illuminate more than one point at the same time.

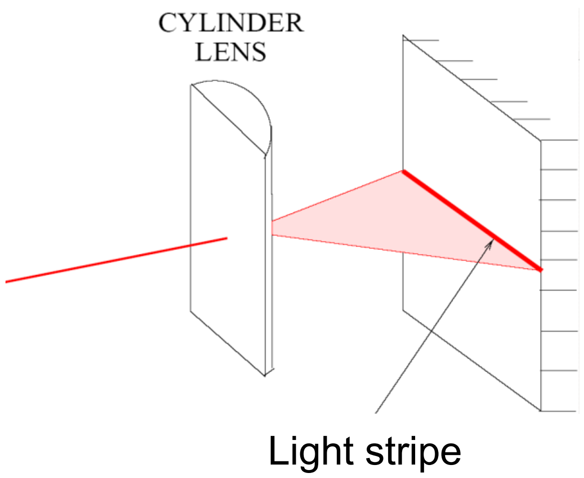


[Young]

Credit: Markus Vincze, Technische Universität Wien

Structured Light

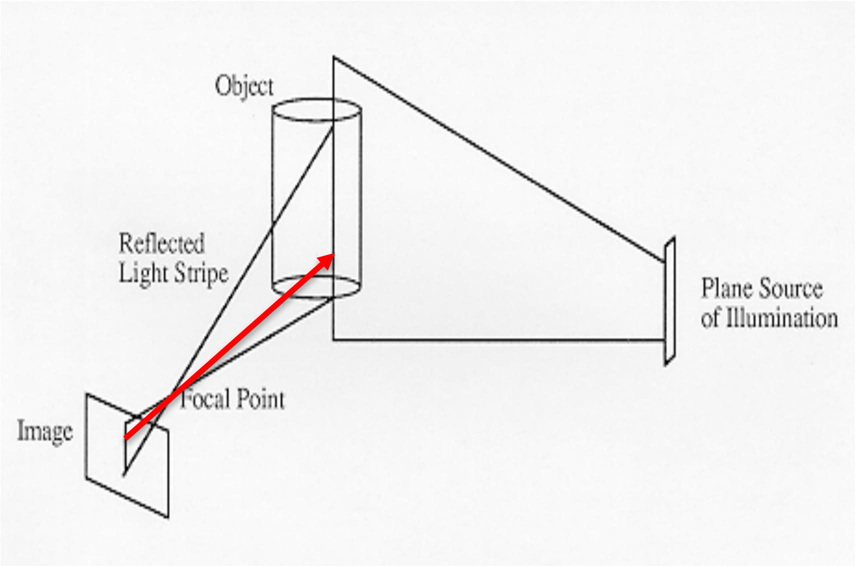
- Fanning out a point laser source with cylinder lens
- Relatively cheap yet most accurate method to obtain depth image
- Scanning is required



Credit: Markus Vincze, Technische Universität Wien

Structured Light

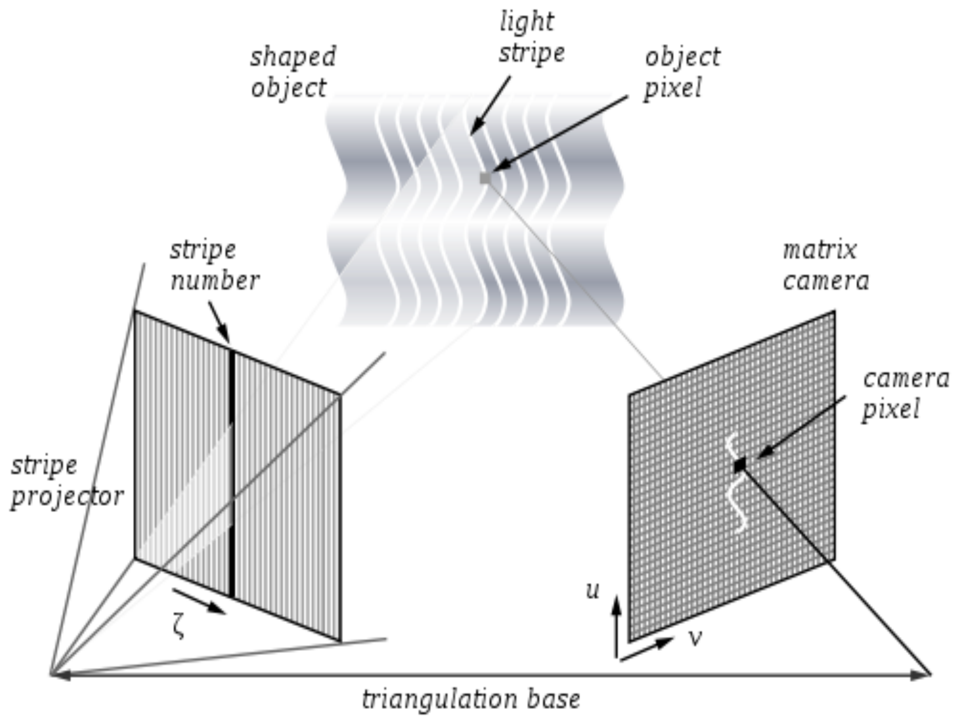
Depth is given by the intersection of the line of sight and the plane of light



Credit: Markus Vincze, Technische Universität Wien

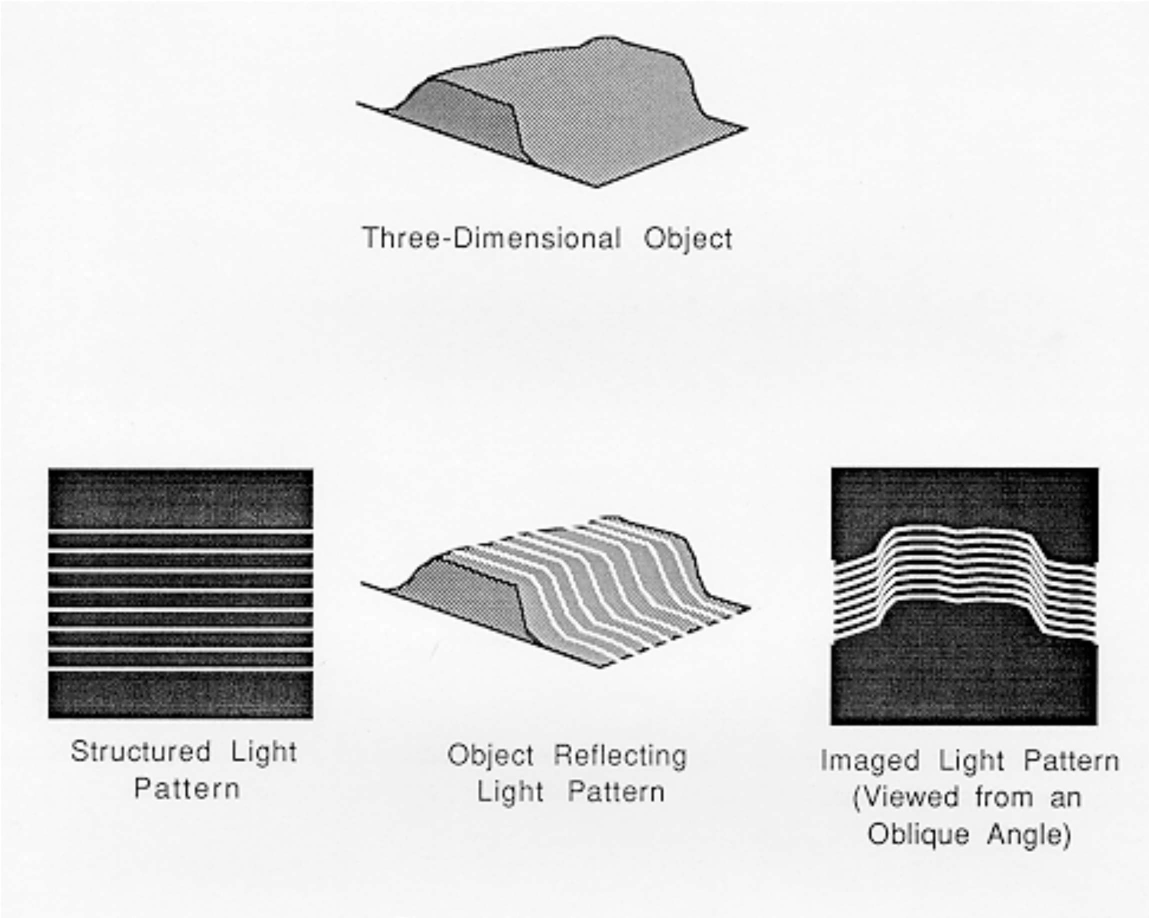
Structured Light

- Active transmission of structured light
  - Coded light
  - Moiré-pattern
  - Dot pattern
- One shot for a full scene
- Regular lines are not unique



Credit: Markus Vincze, Technische Universität Wien

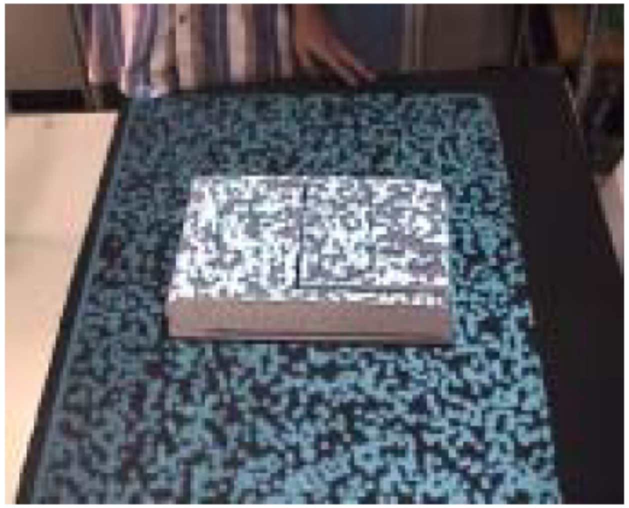
Structured Light



Structured Light

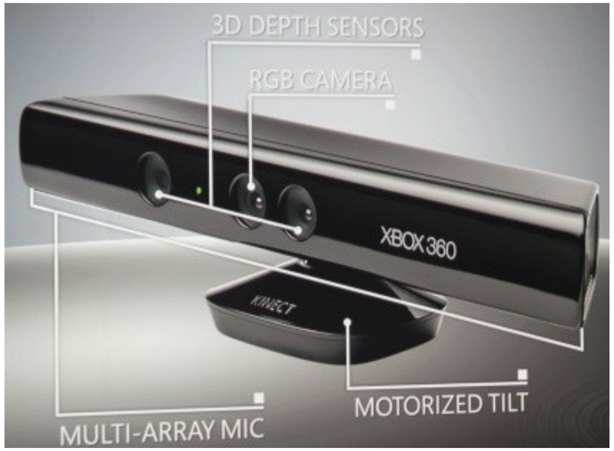
Dot patterns

- Regular
- Random



Example: Kinect

- RGB colour camera
- 3D Depth sensor = infrared projector and CMOS sensor
- Depth from triangulation
- 640x480 Resolution, 30 Hz
- Depth calculation onboard

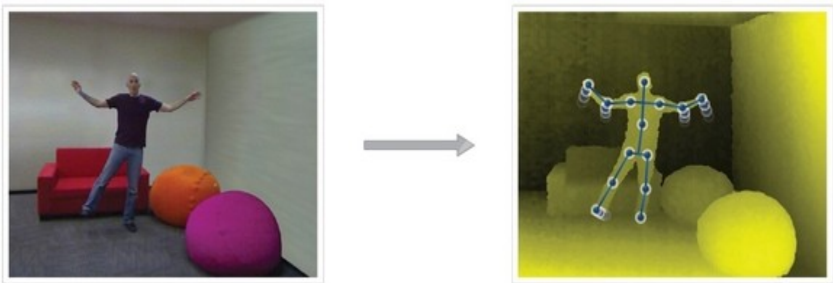


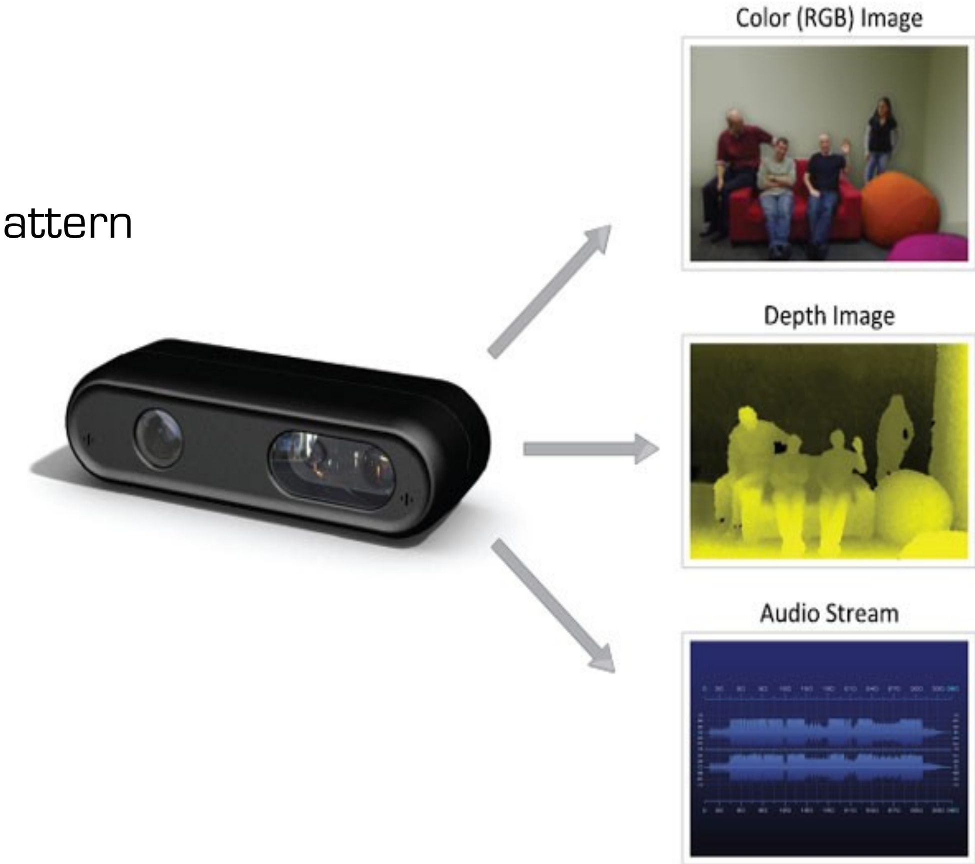
Colour and depth image [Gadget Lab]

Credit: Markus Vincze, Technische Universität Wien

Example: Sensor by PrimeSense

- Distance: 0.8 – 3.5 m
- Spatial x/y resolution 3mm @ 2m
- Depth z resolution: 10mm @ 2m
- Sun light reduces visibility of the infrared pattern

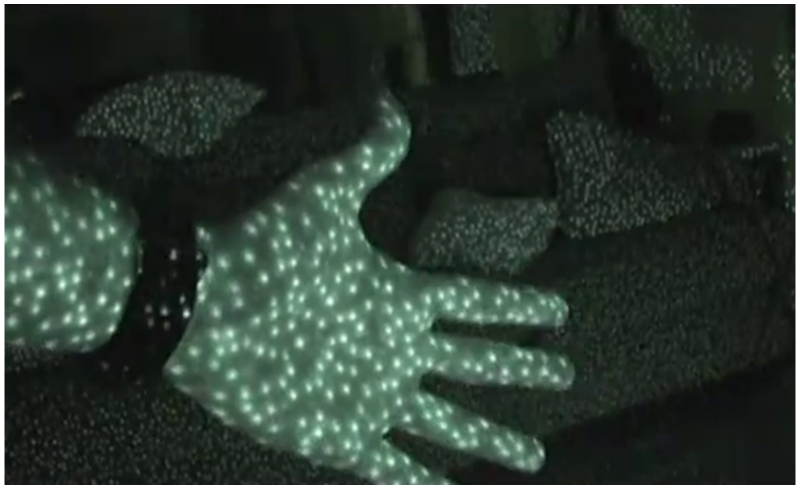


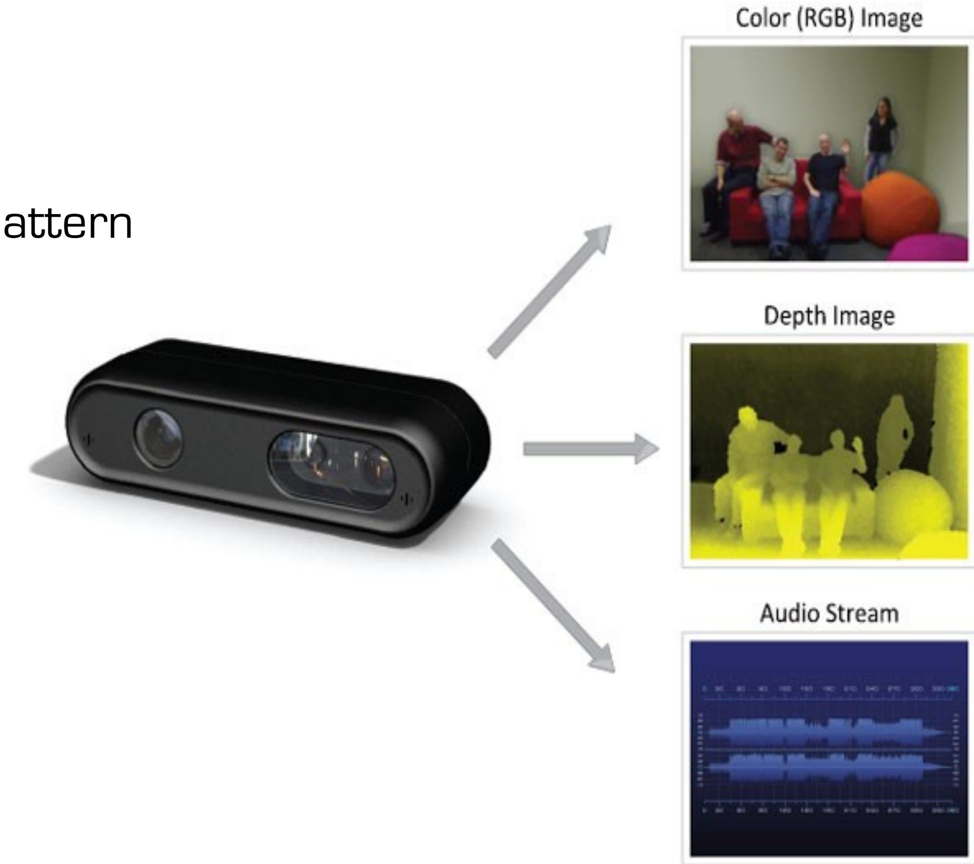


Credit: Markus Vincze, Technische Universität Wien

Sensor by PrimeSense

- Distance: 0.8 – 3.5 m
- Spatial x/y resolution 3mm @ 2m
- Depth z resolution: 10mm @ 2m
- Sun light reduces visibility of the infrared pattern





Credit: Markus Vincze, Technische Universität Wien

Reading

R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2010.

Section 12.1 Shape from X

Section 12.2 Active rangefinding