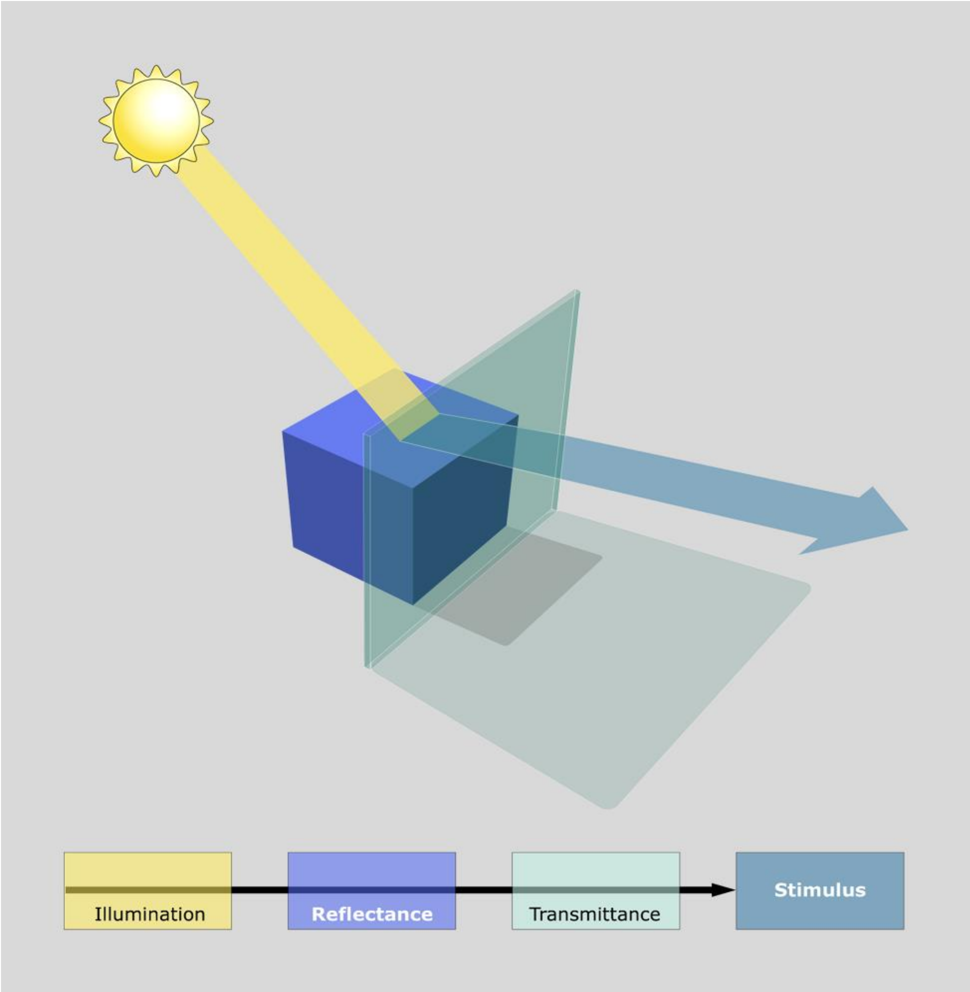
Spatial track: range acquisition modeling

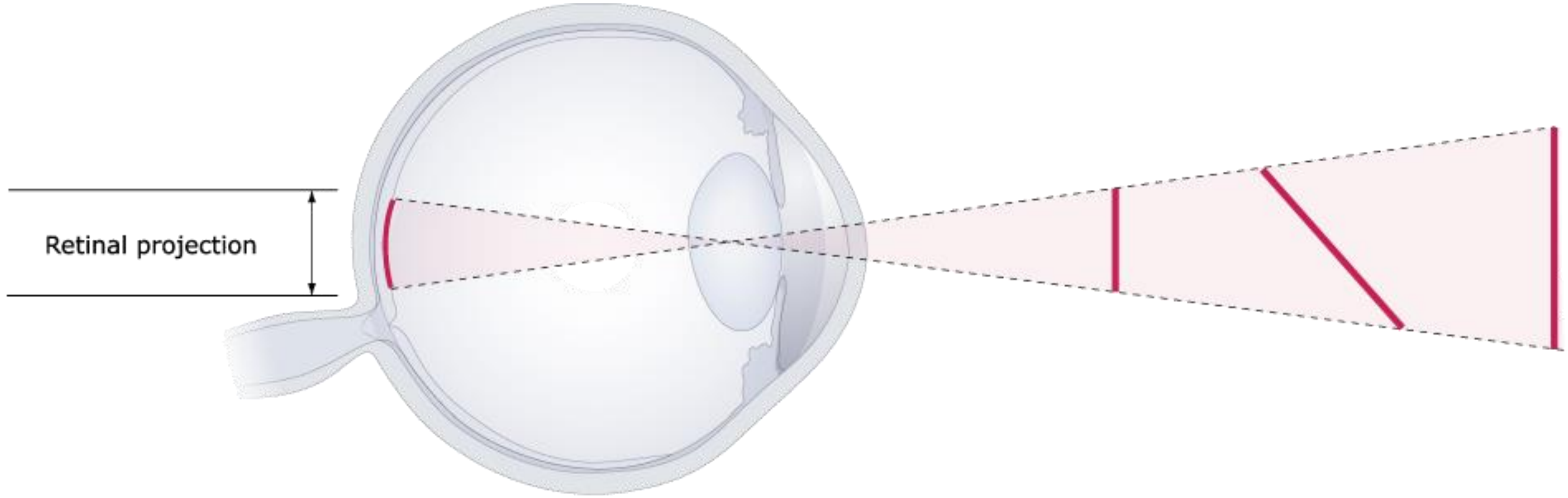
Virginio Cantoni

Laboratorio di Visione Artificiale
Università di Pavia
Via A. Ferrata 1, 27100 Pavia
virginio.cantoni@unipv.it
http://vision.unipv.it/va

The inverse problem



Physical space geometrical properties: distances in depth - the inverse problem



A basic problem in perception that provides a clue….

- The stimuli produced when energy interacts with sensory receptors cannot specify the real-world sources of that energy
- To survive, animals need to react successfully to the sources of the stimuli, not to the stimuli as such
- This quandary is called the inverse problem

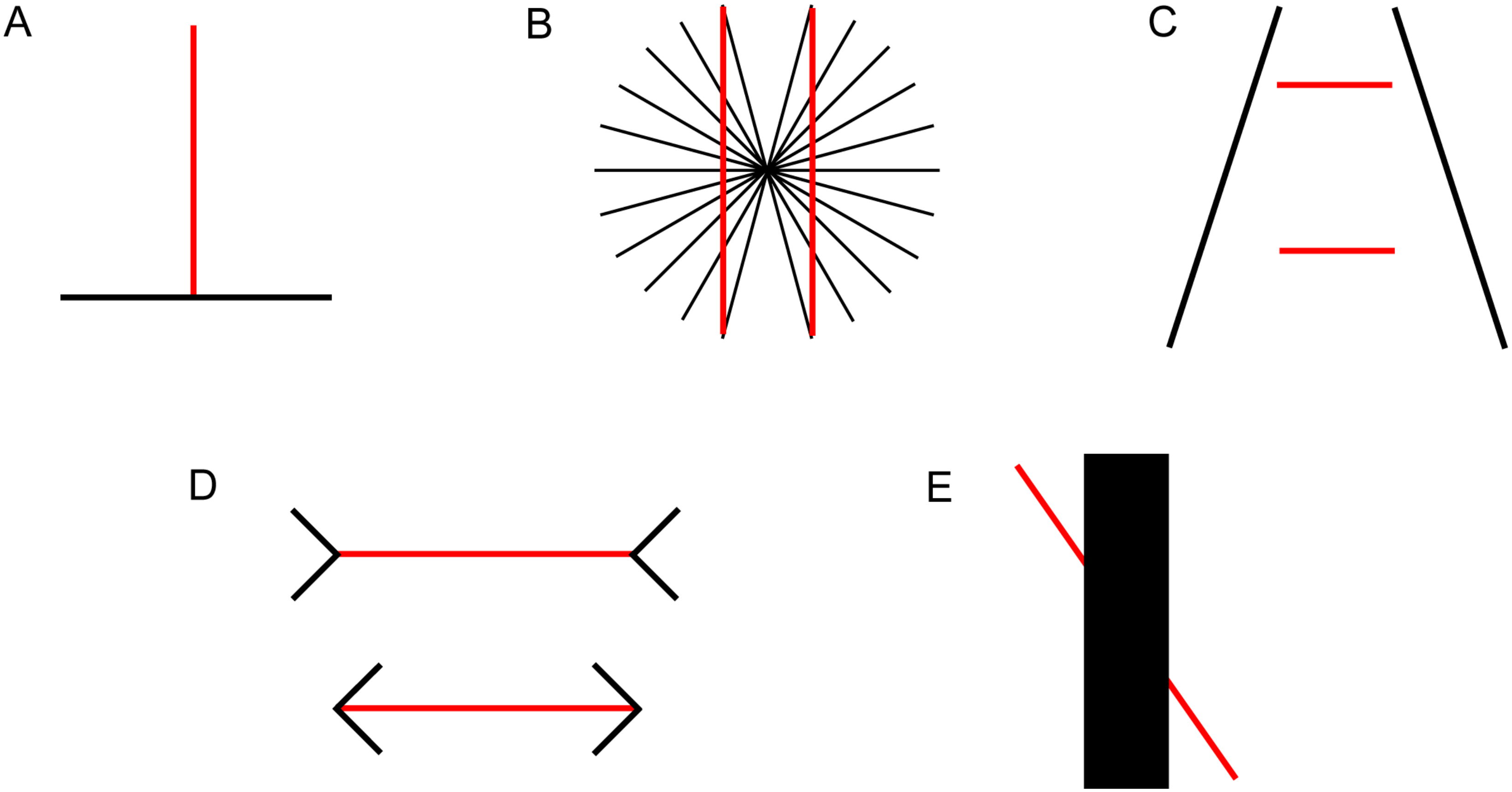
Explanation of Visual Processing and Percepts

- The basic problem understanding vision is that the real-world sources of light stimuli cannot be known directly
- The visual system generates percepts entirely on the basis of past experience, using stimulus patterns to trigger percepts as reflex responses that have been empirically successful.
- This strategy would contend with the inverse problem.

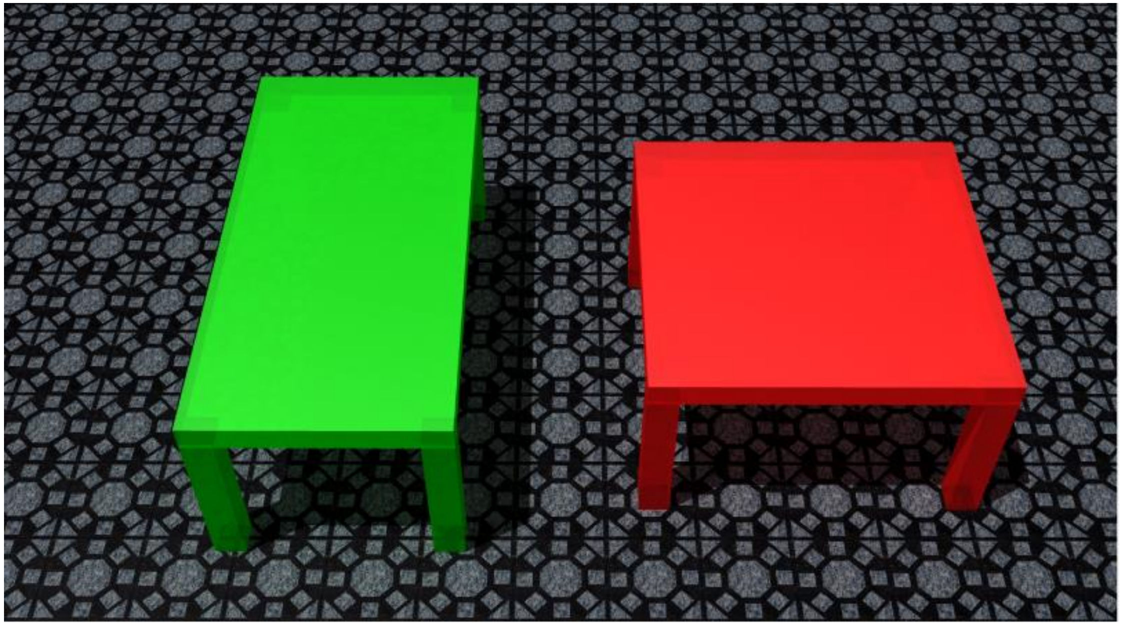
Explanation of Geometrical Percepts

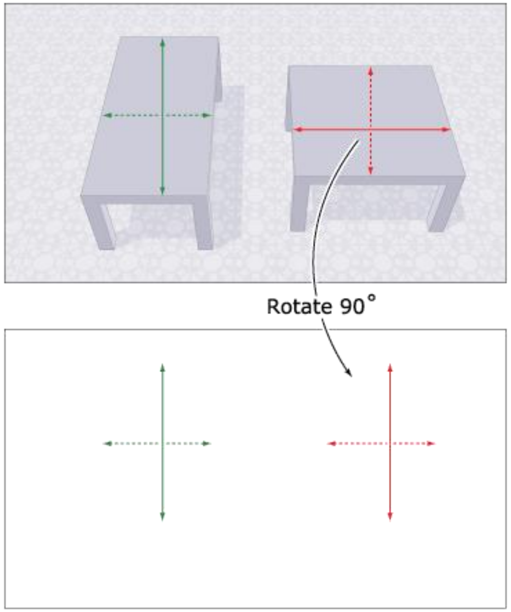
- Physical space is characterized by geometrical properties such as line lengths, angles, orientations and distances in depth
- Our intuition is that the subjective qualities arising from these properties should be a more or less direct transformation of physical space
- As in the domains of brightness and color, however, there are many discrepancies between measurements of physical space and the geometries people actually see

Physical space geometrical properties: line lengths



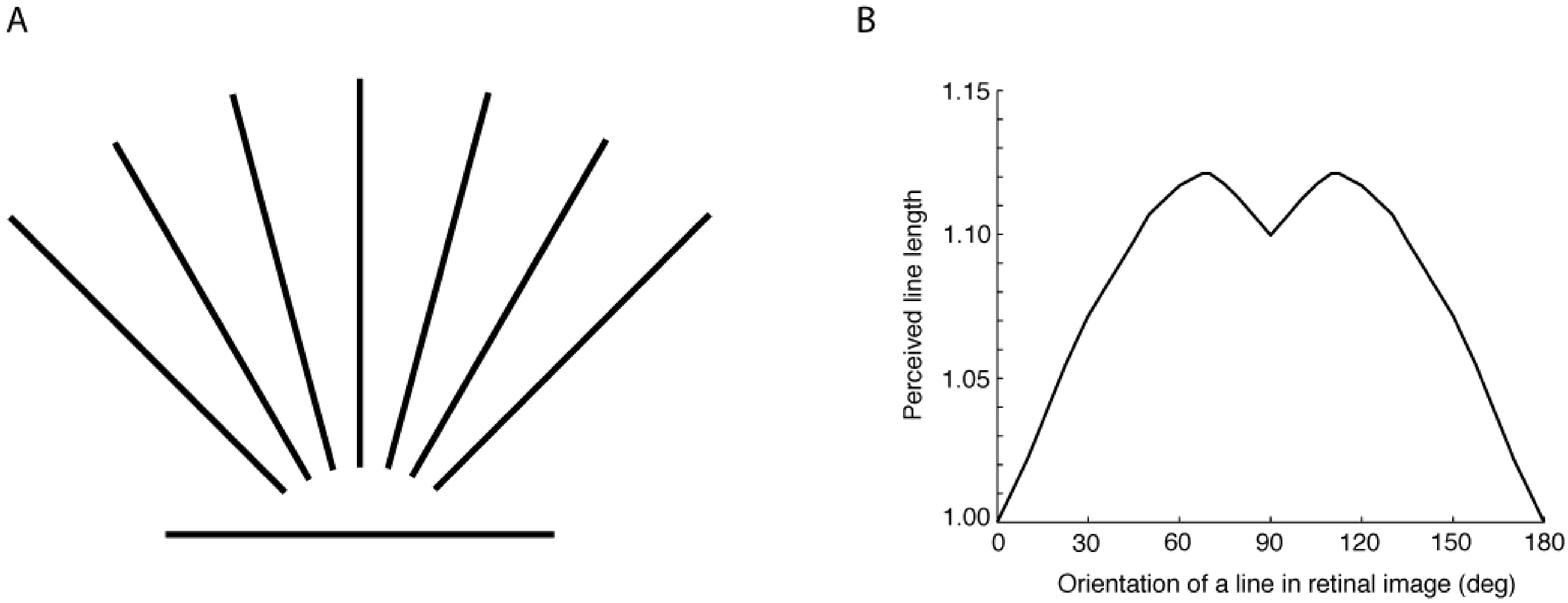
Physical space geometrical properties: orientation anisotropy





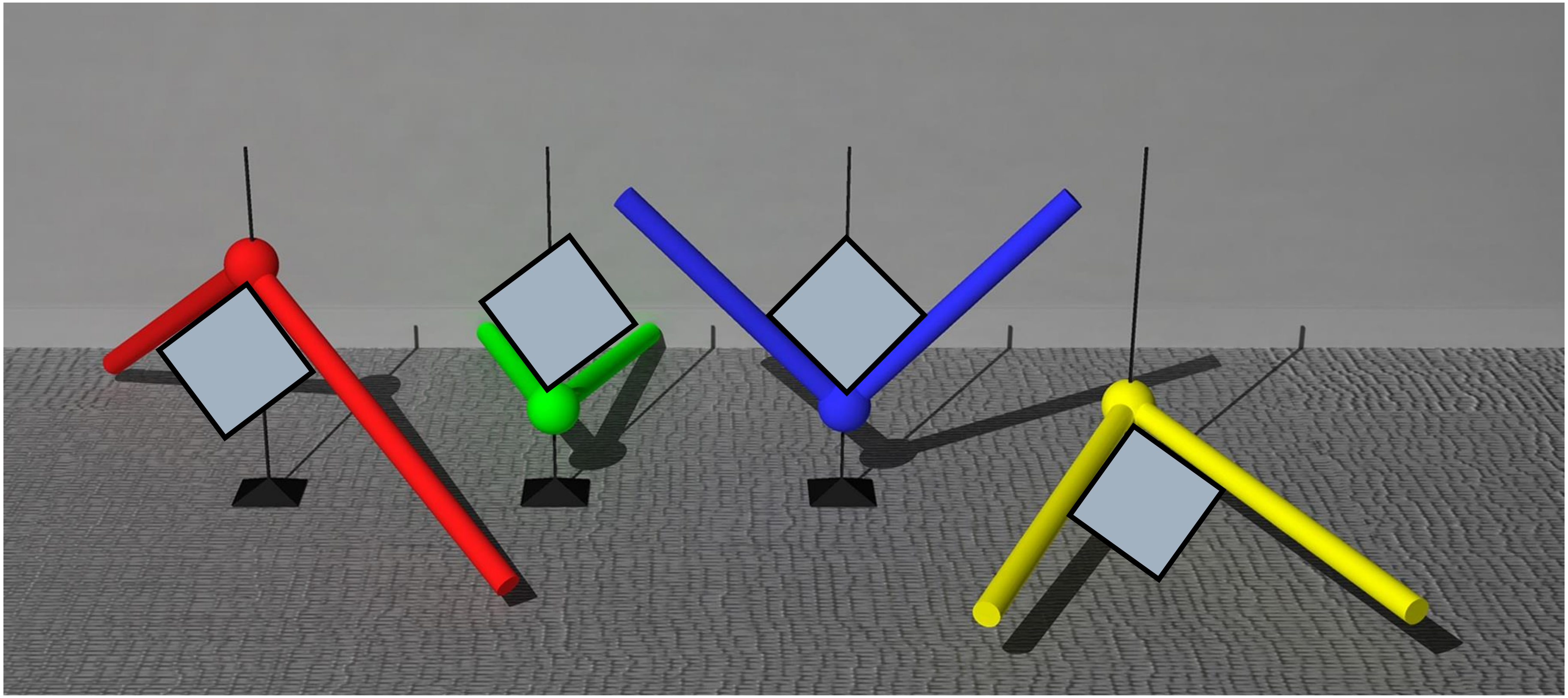
Dale Purves, Cognitive Neuroscience, Duke University

Physical space geometrical properties: line lengths

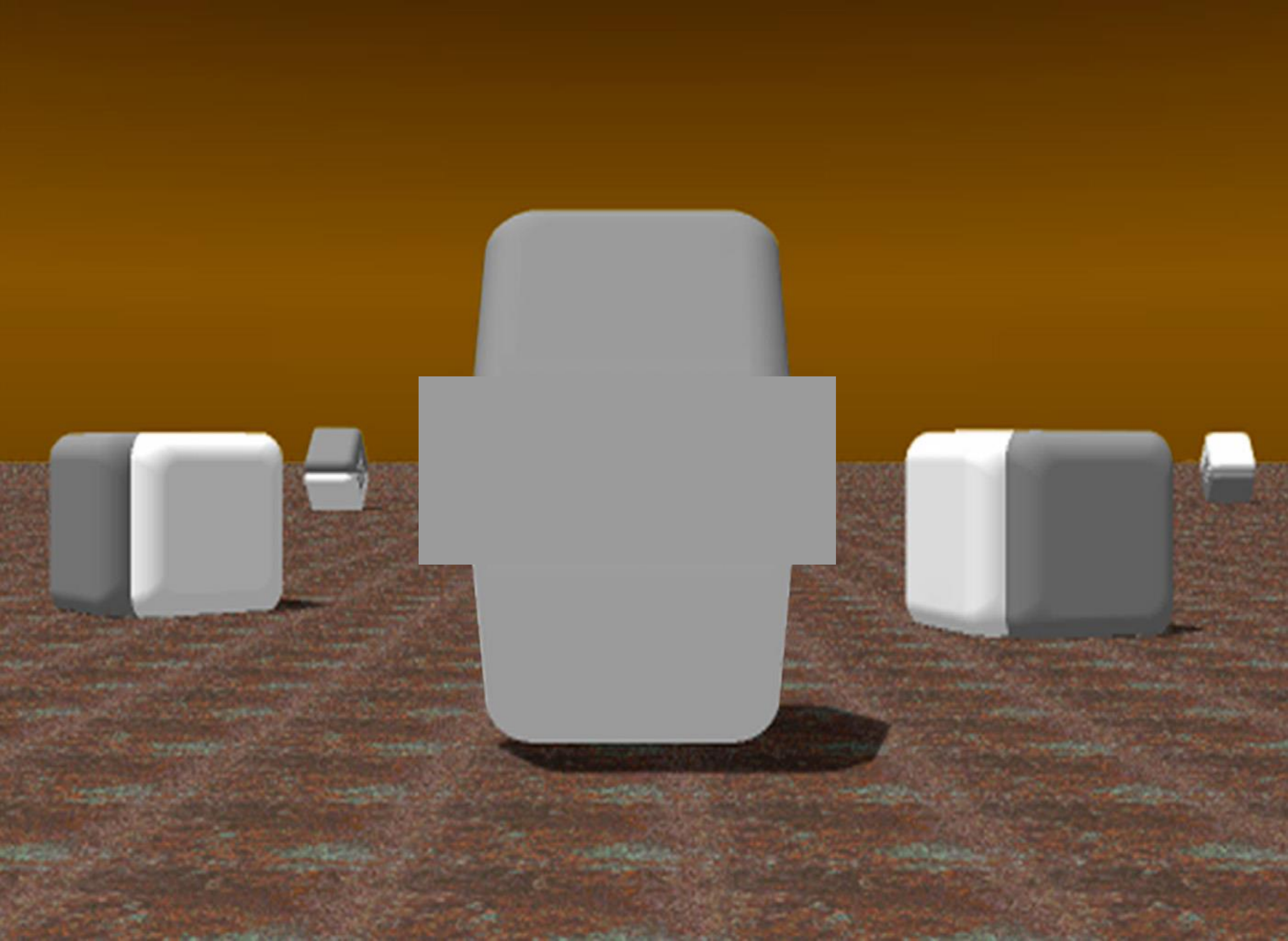


Dale Purves, Cognitive Neuroscience, Duke University

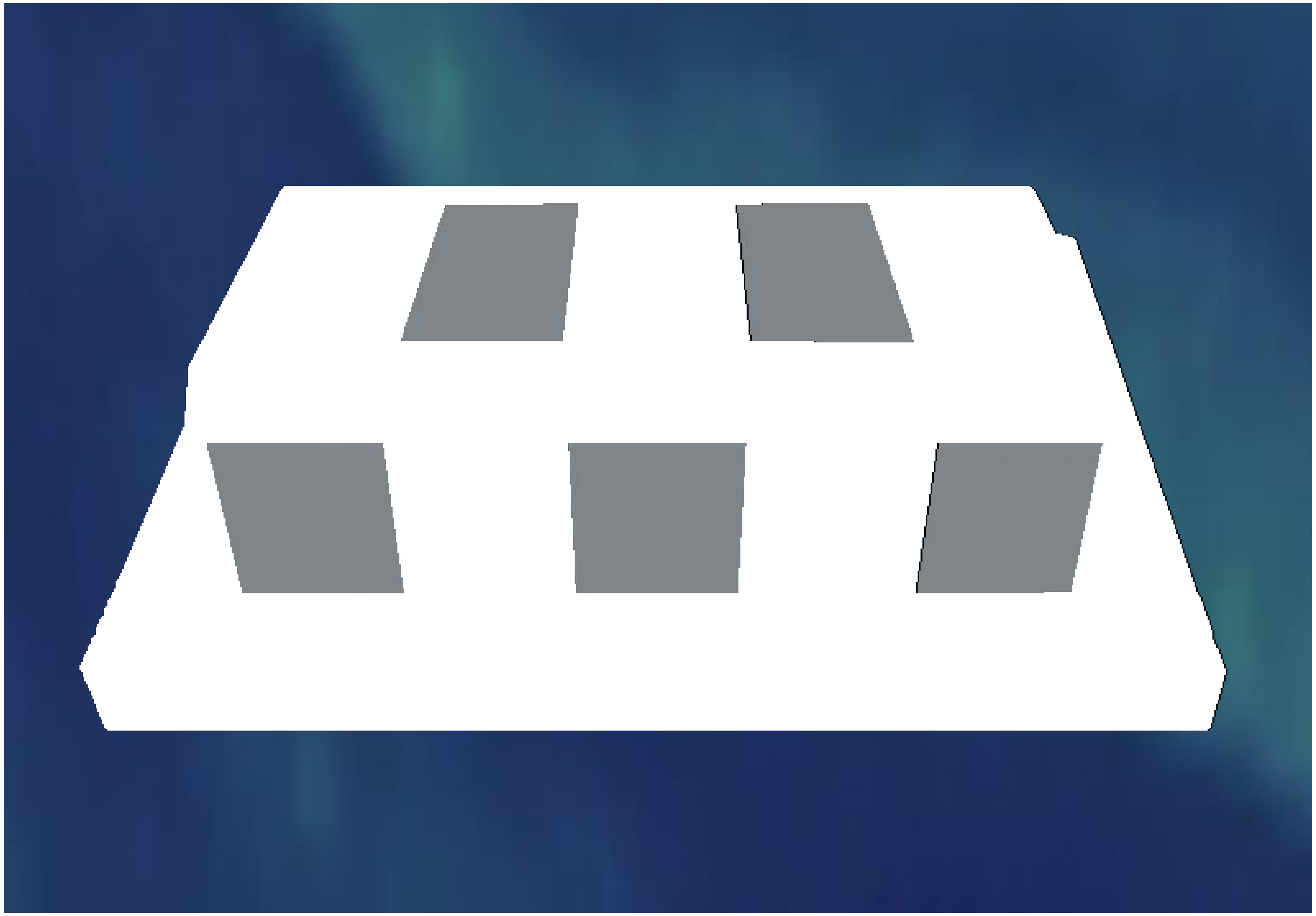
Physical space geometrical properties: angles



© Dale Purves and R. Beau Lotto 2002

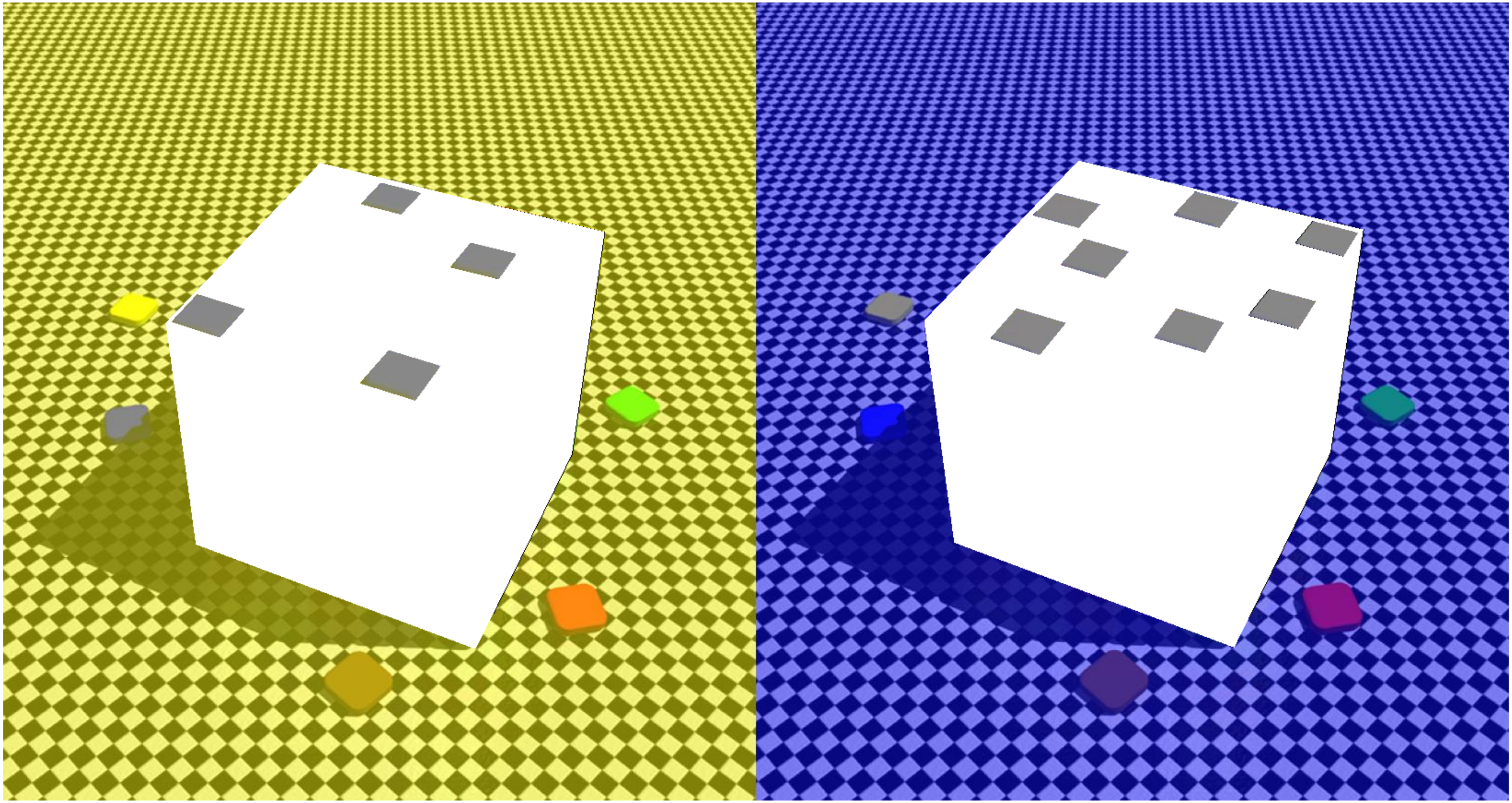


Optic illusions

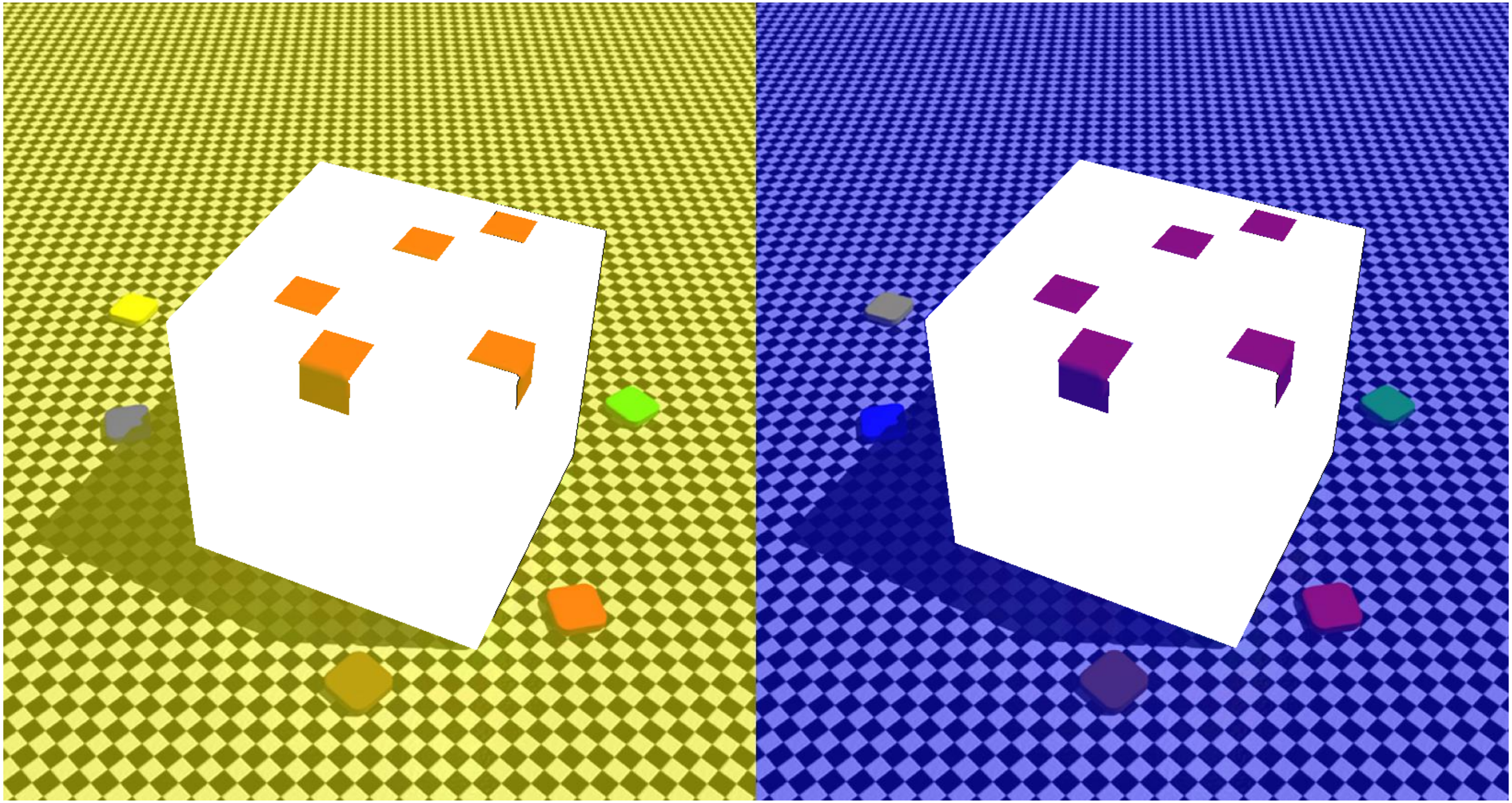


Dale Purves, Cognitive Neuroscience, Duke University

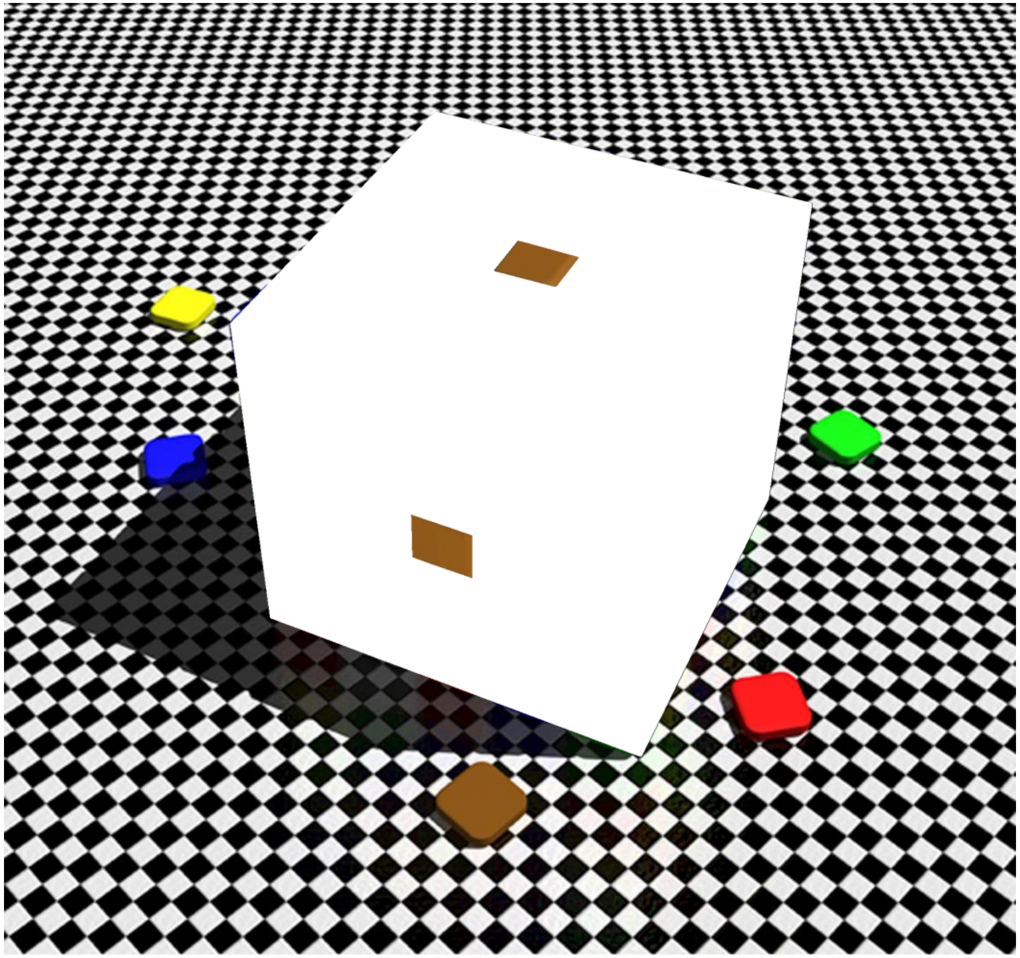
Optic illusions



Optic illusions



Optic illusions



Visual cues – The human headway

Overlapping objects

Quantized scenes

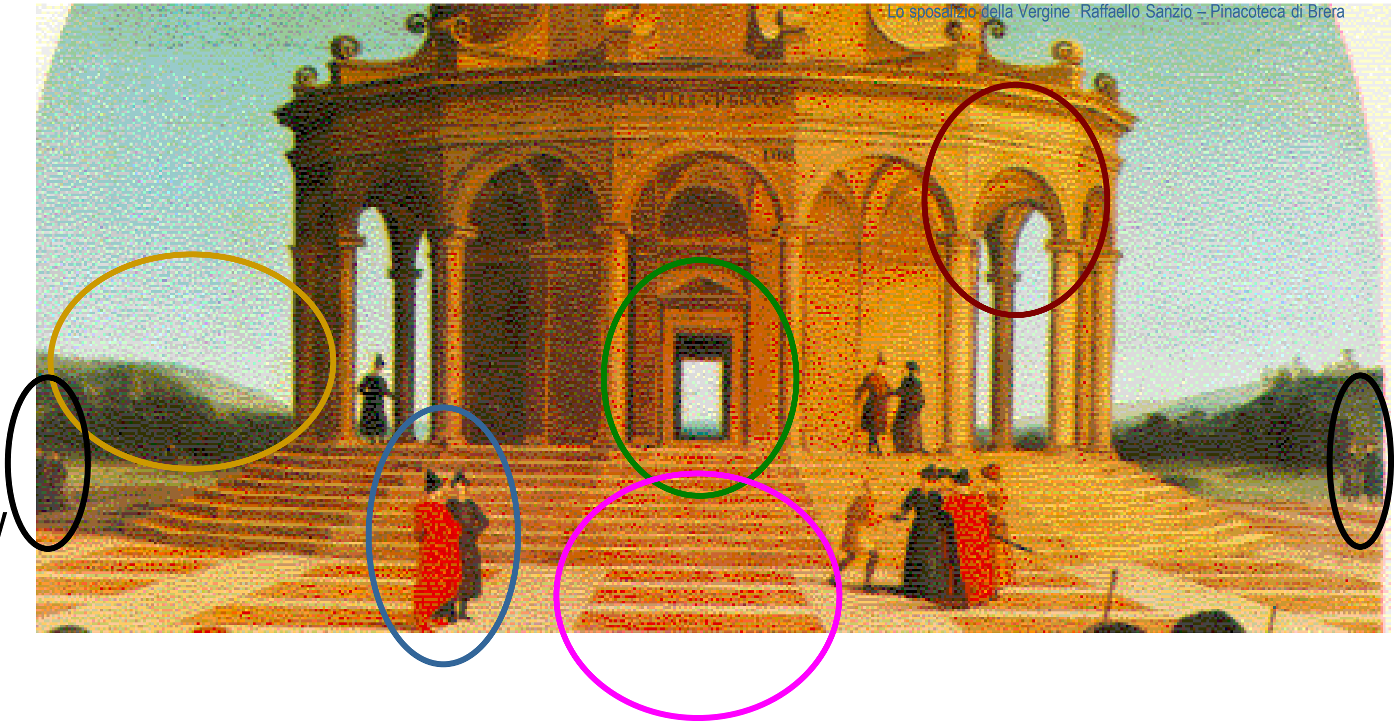
Perspective geometry

Depth from shading

Multi-presence

Depth from texture

Height in the field of view



Atmospheric perspective

- Based on the effect of air on the color and visual acuity of objects at various distances from the observer.
- Consequences:
  - Distant objects appear bluer
  - Distant objects have lower contrast.

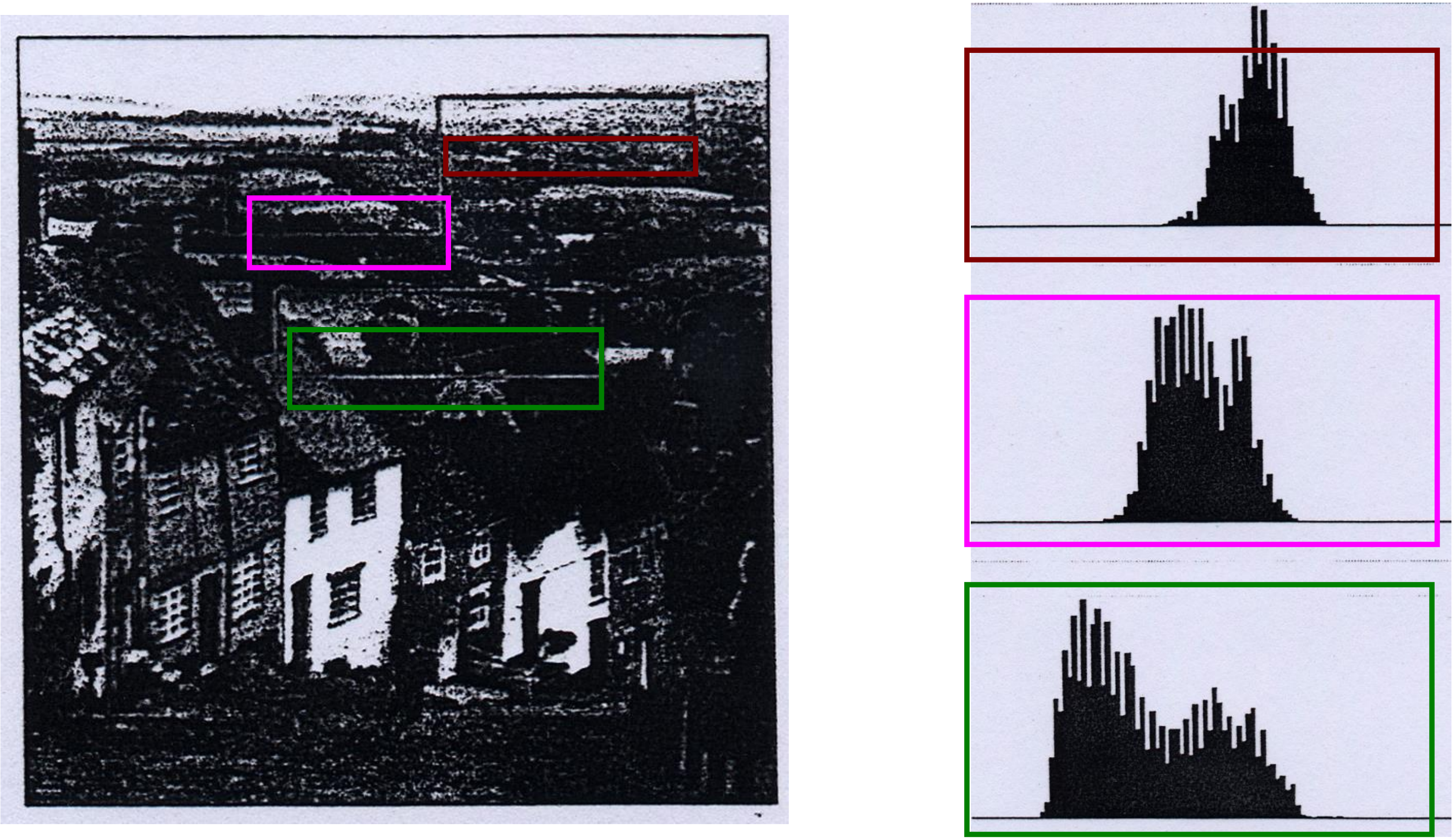
Atmospheric perspective

http://encarta.msn.com/medias_761571997/Perception_(psychology).html

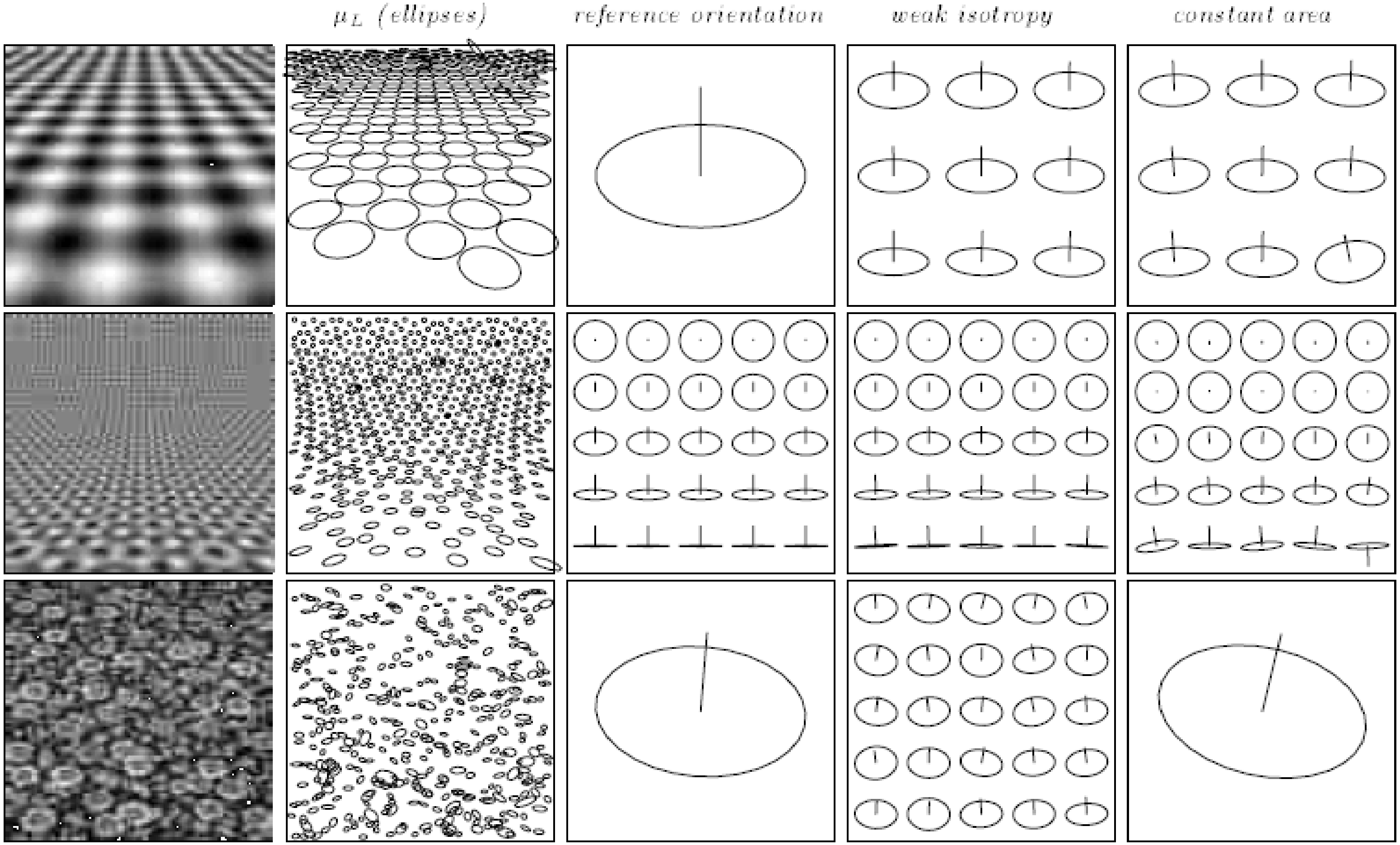
Atmospheric perspective

Claude Lorrain (artist)
French, 1600 - 1682
Landscape with Ruins, Pastoral Figures, and Trees, 1643/1655

Histogram

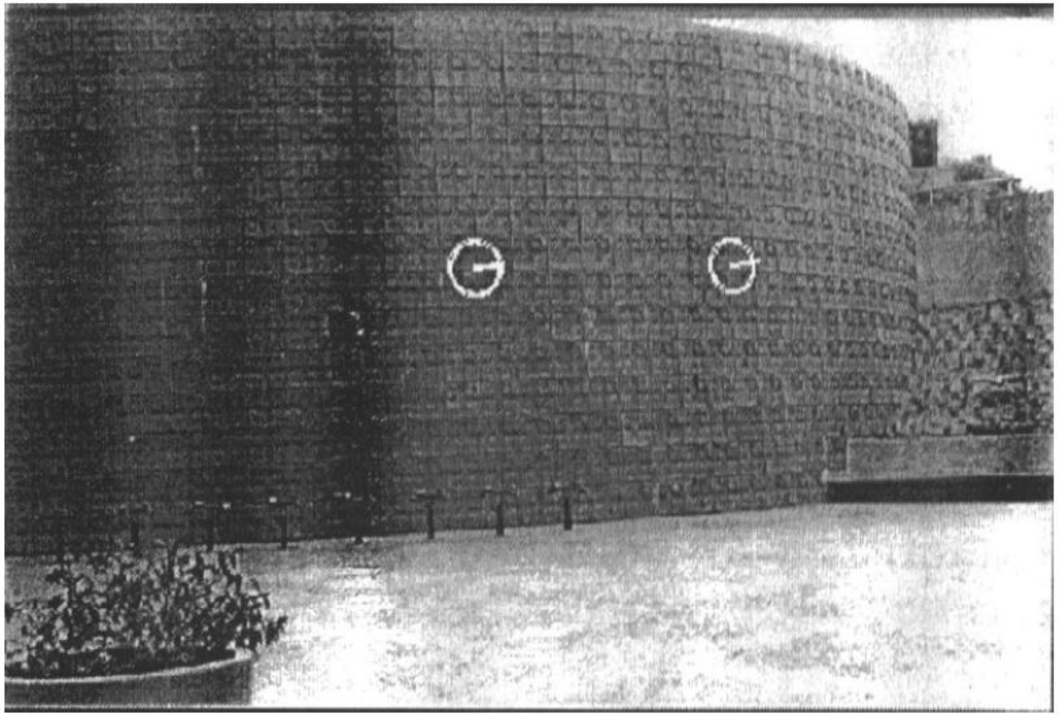


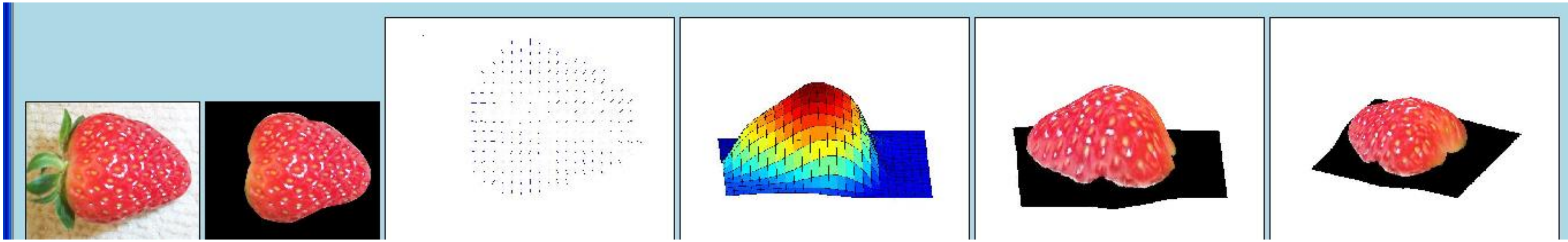
e Gradient



Shape from Texture from a Multi-Scale Perspective. Tony Lindeberg and Jonas Garding. ICCV 93

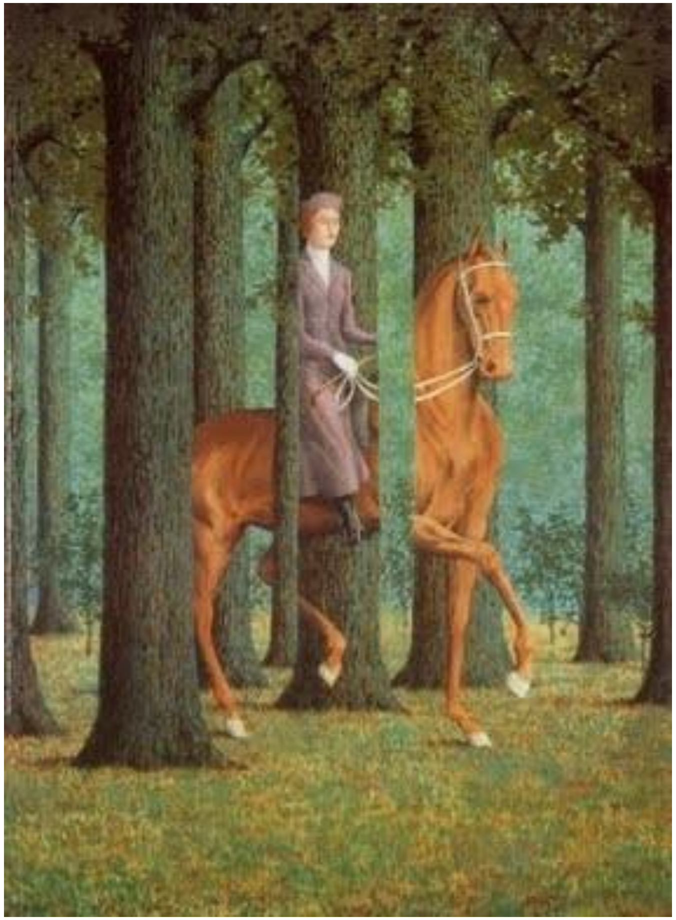
Texture





[From A.M. Loh. The recovery of 3-D structure using visual texture patterns. PhD thesis]

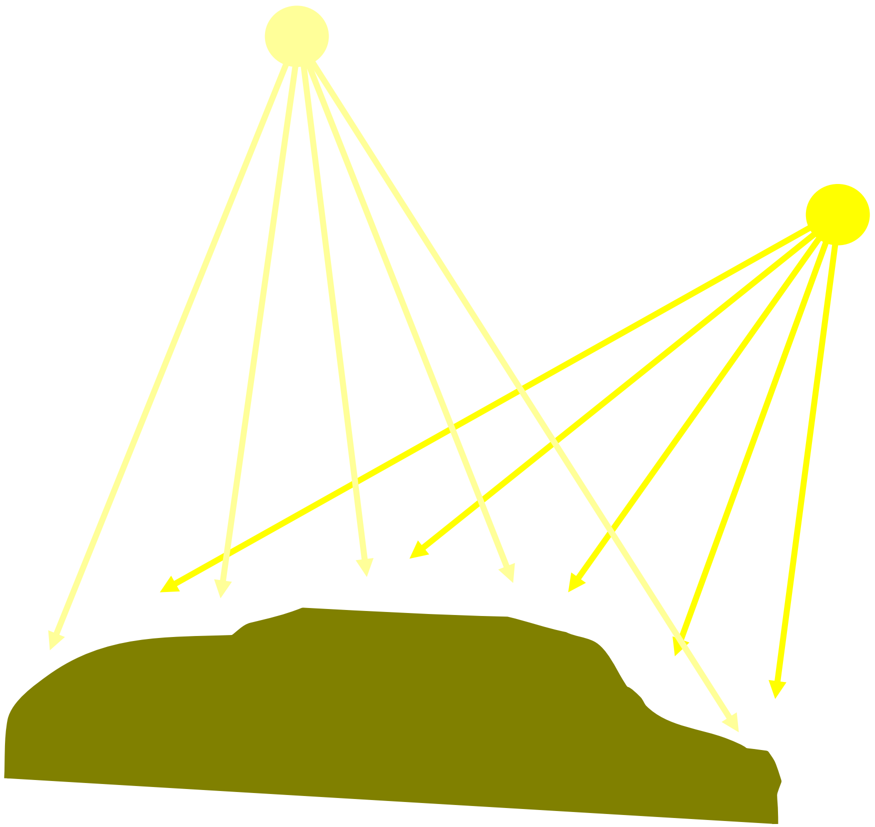
Occlusion

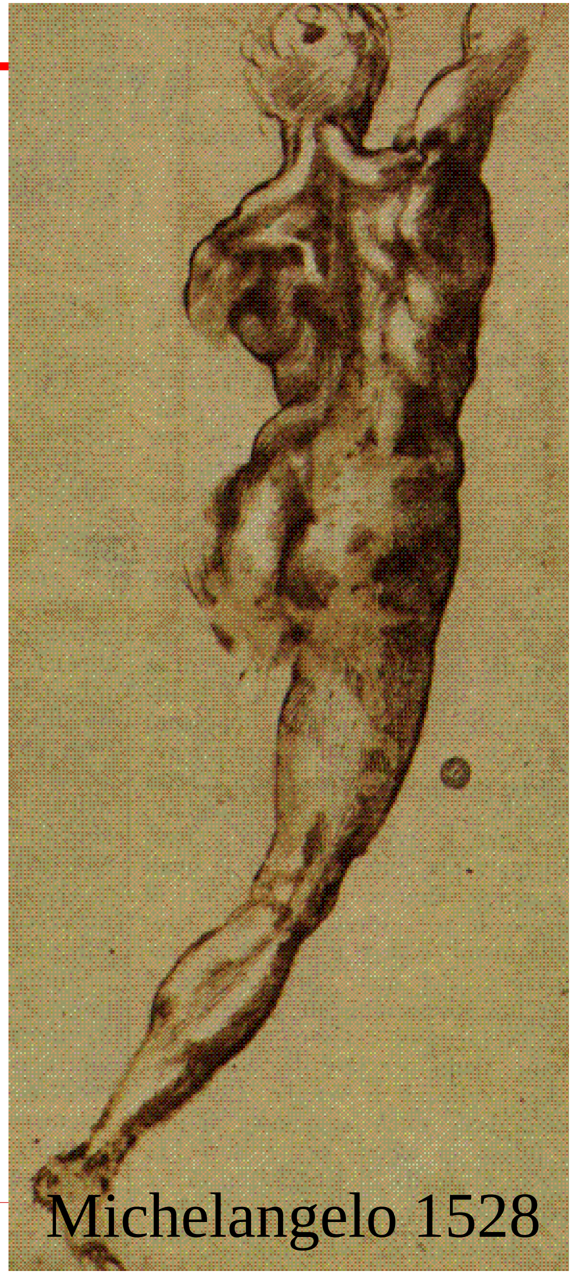


Rene Magritt'e famous painting *Le Blanc-Seing* (literal translation: "The Blank Signature") roughly translates as "free hand" or "free rein".

Shape from.....

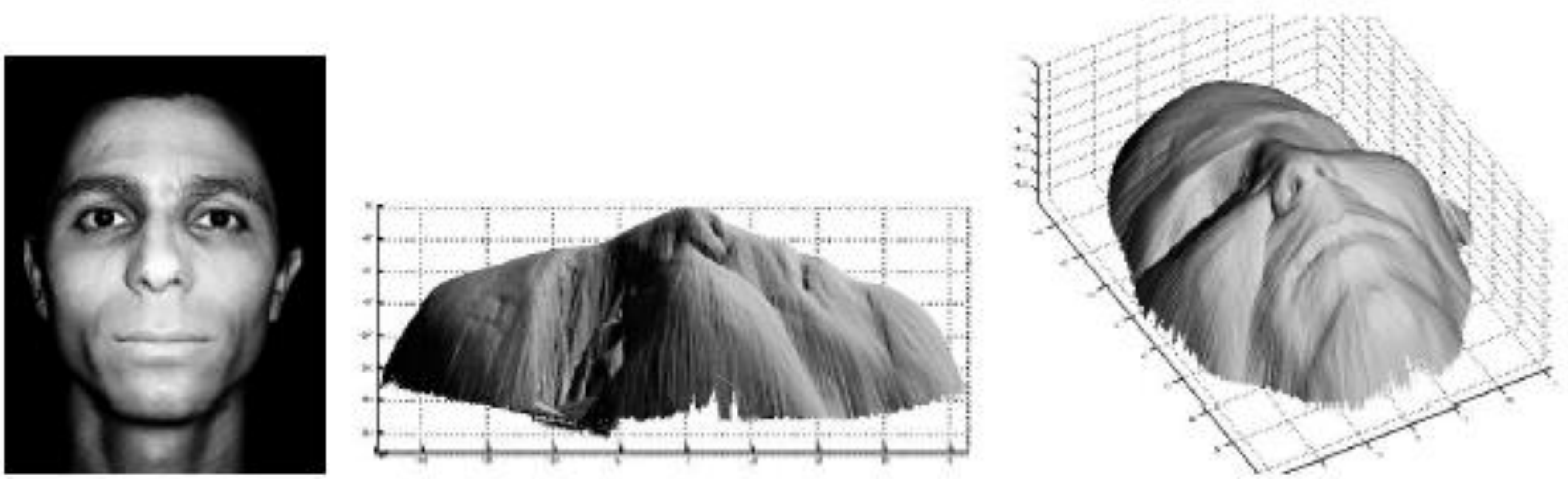
shadows





Michelangelo 1528

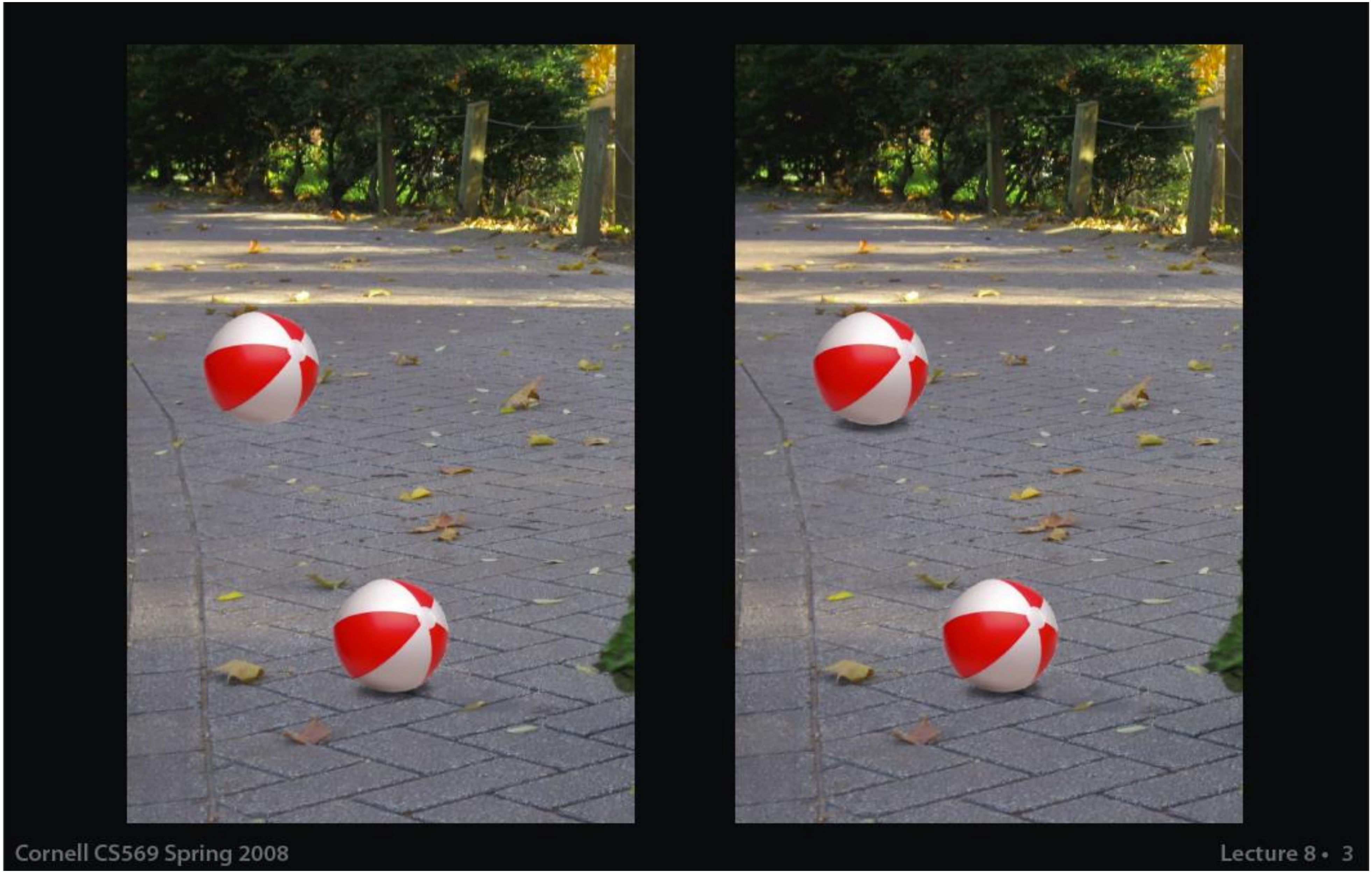
Shading



a) b) c)

[Figure from Prados & Faugeras 2006]

Shadows

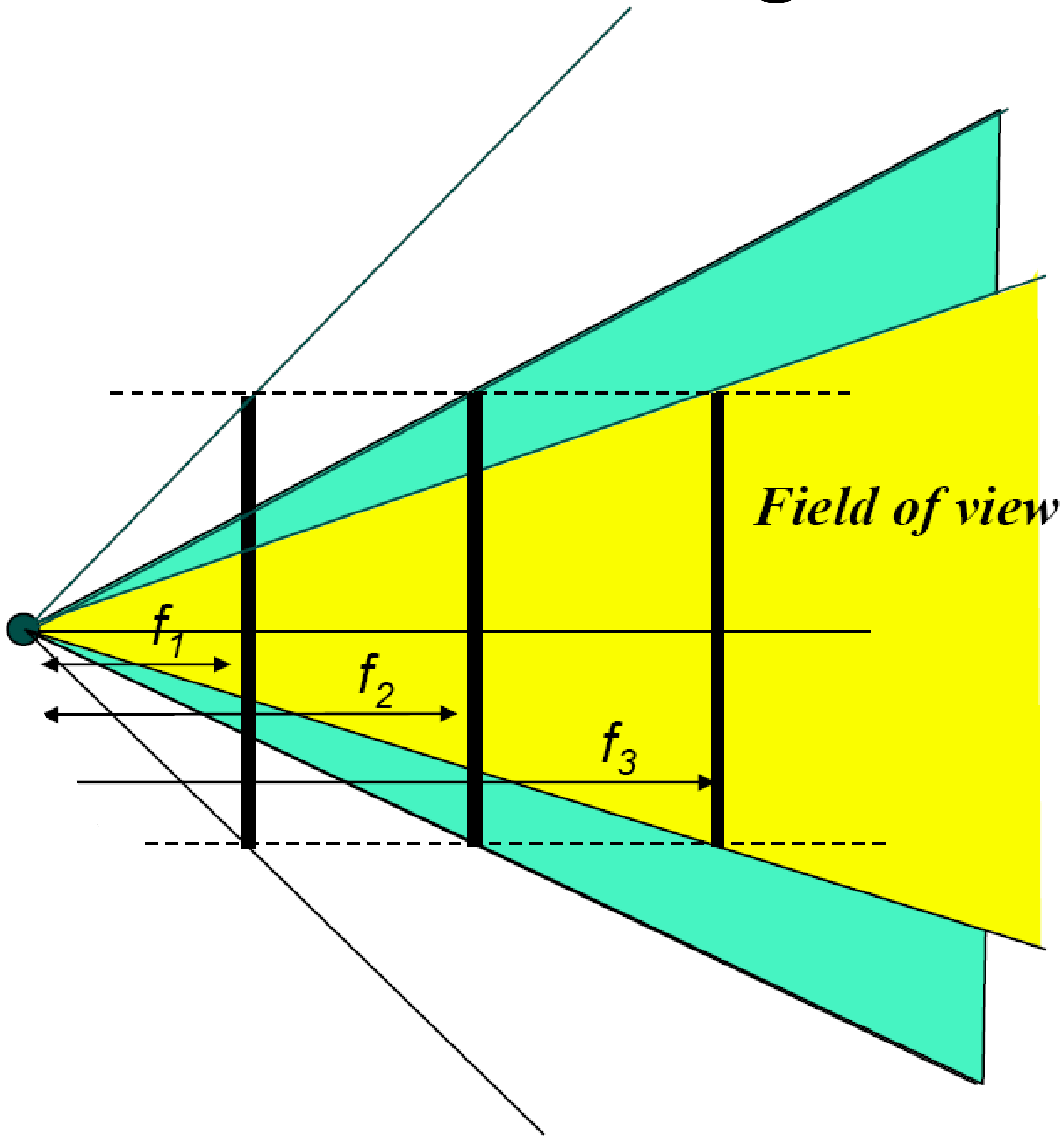


Slide by Steve Marschner

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

Field of view depends on focal length

- As ***f*** gets smaller, image becomes more *wide angle*
  - more world points project onto the finite image plane
- As ***f*** gets larger, image becomes more *telescopic*
  - smaller part of the world projects onto the finite image plane



from R. Duraiswami

Field of view

- Angular measure of portion of 3d space seen by the camera

28 mm lens, 65.5° × 46.4°

50 mm lens, 39.6° × 27.0°

70 mm lens, 28.9° × 19.5°

210 mm lens, 9.8° × 6.5°

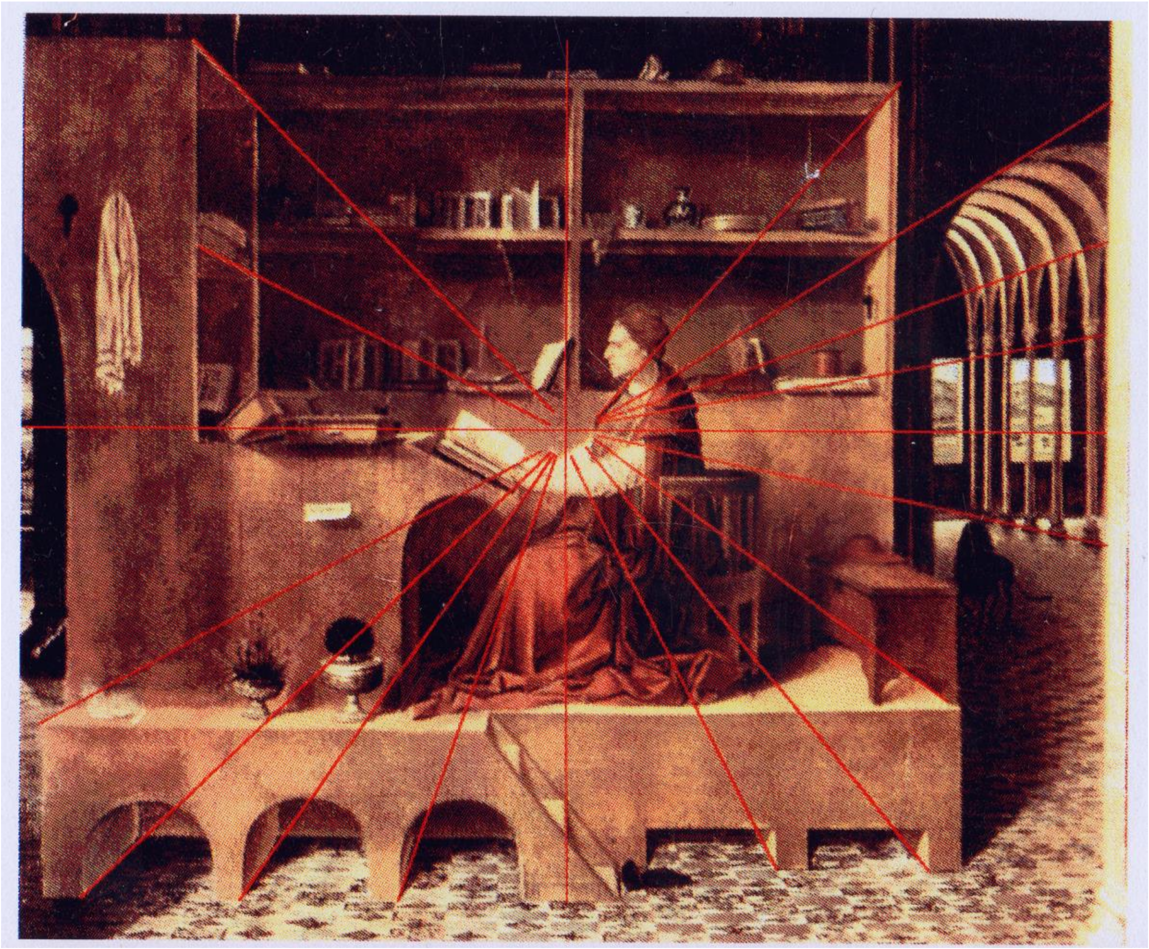
Images from http://en.wikipedia.org/wiki/Angle_of_view

K. Grauman

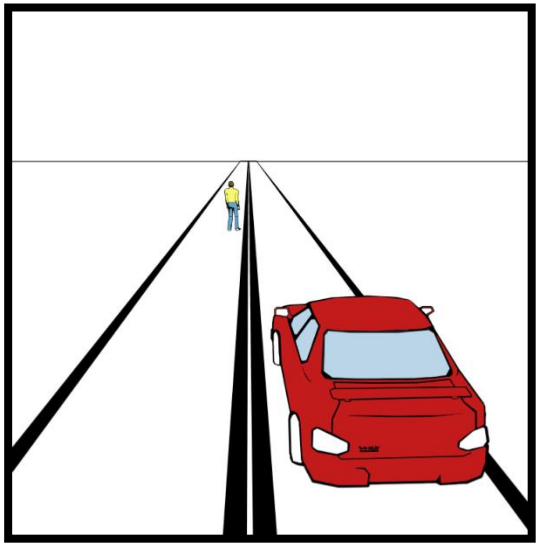
Perspective effects

Image credit: S. Seitz

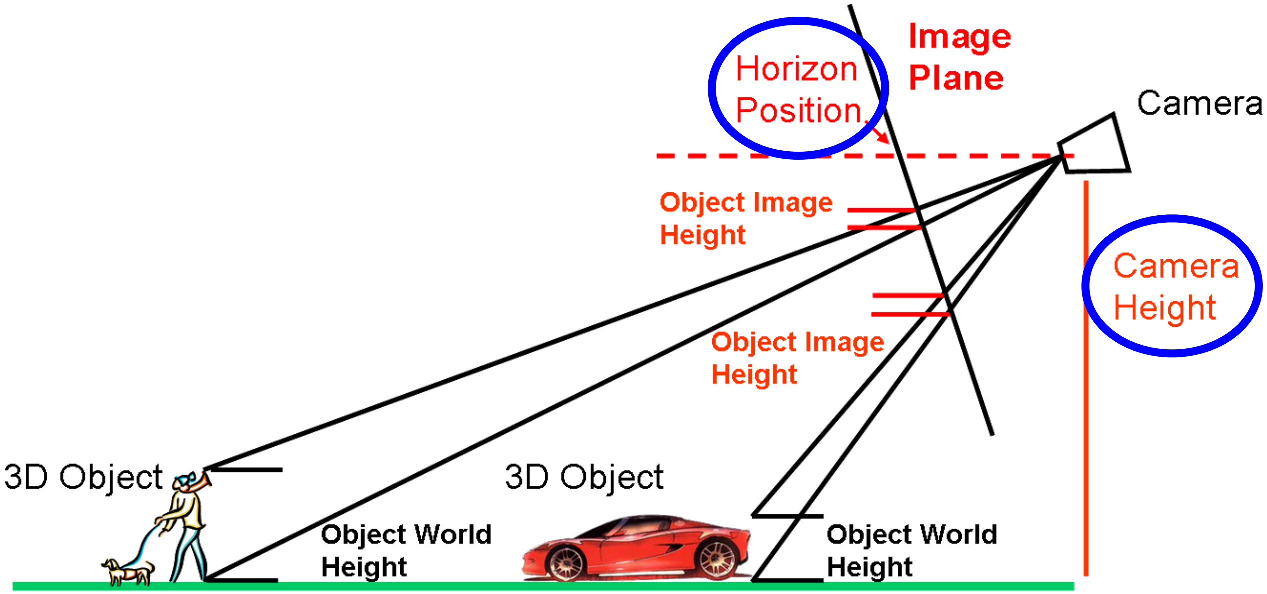
Perspective geometry



Object Size in the Image



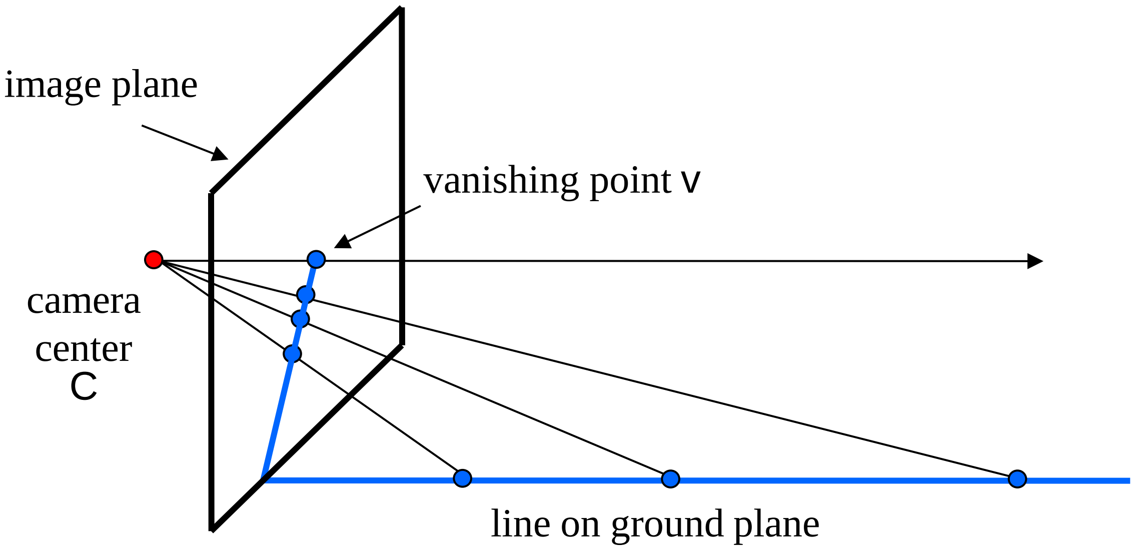
Image



World

Slide by Derek Hoiem

Vanishing points

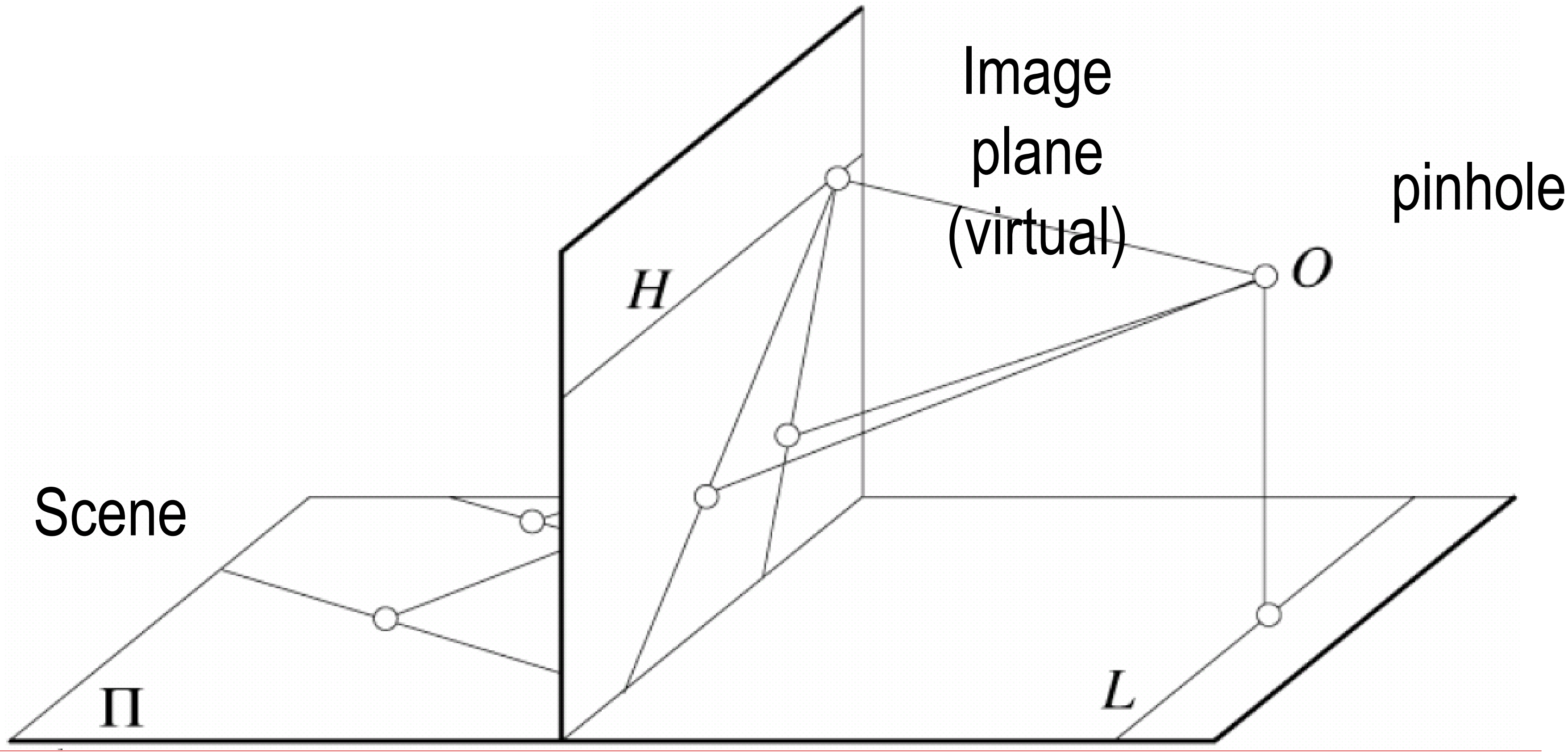


Vanishing point

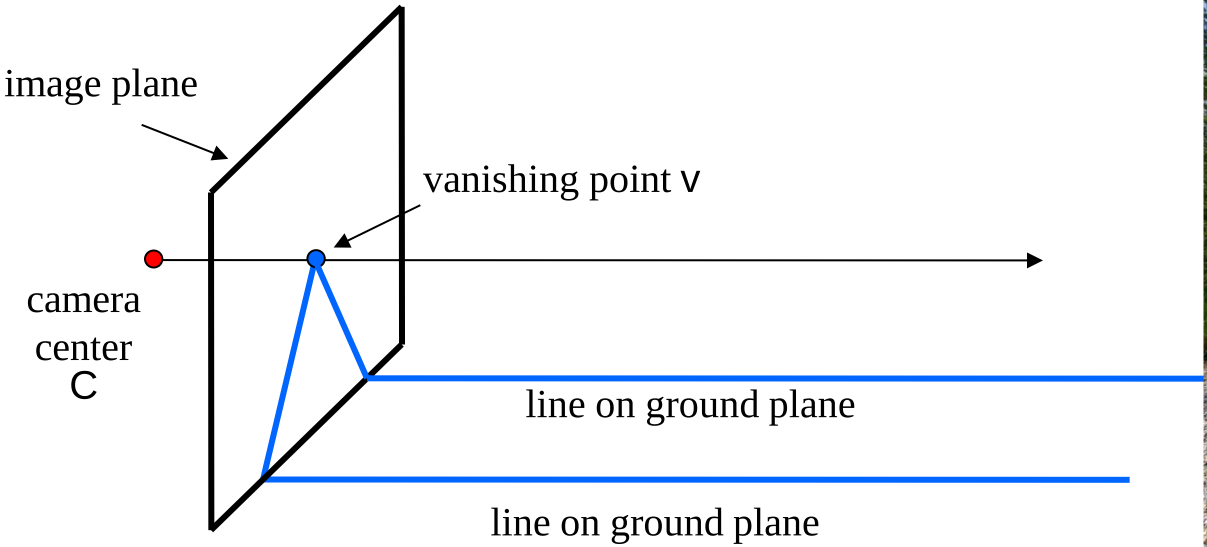
- projection of a point at infinity

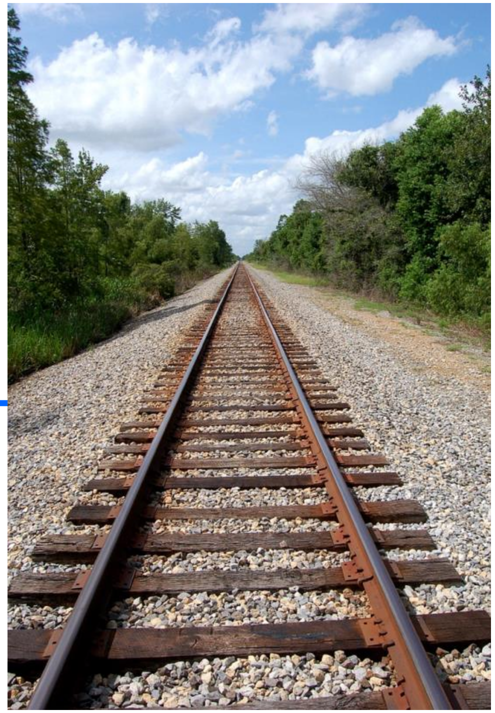
Perspective effects

- Parallel lines in the scene intersect in the image
- Converge in image on horizon line



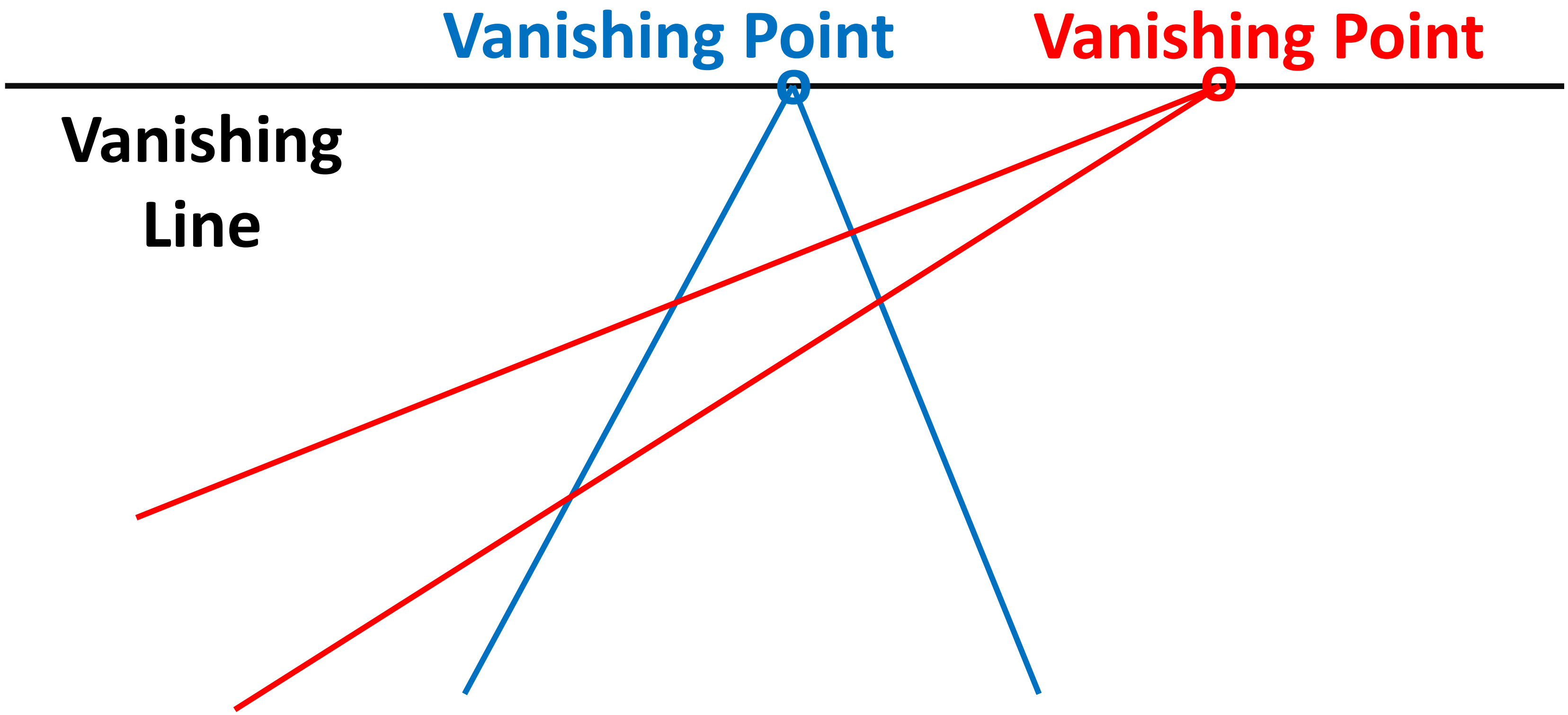
Vanishing points





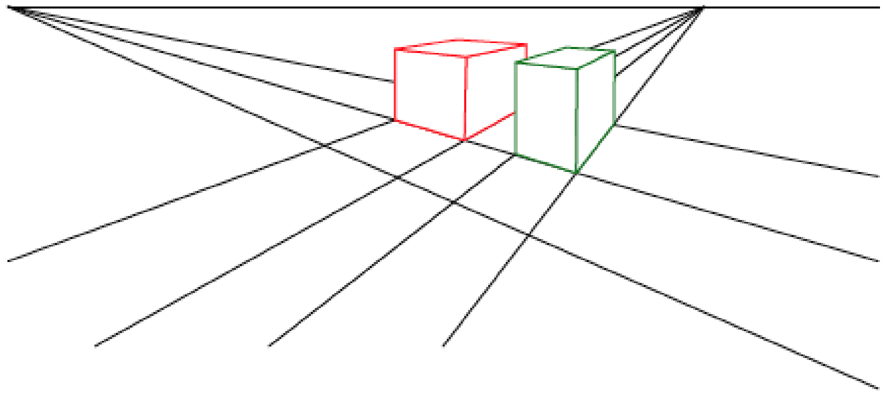
- Properties
  - Any two parallel lines have the same vanishing point v
  - The ray from C through v is parallel to the lines
  - An image may have more than one vanishing point
    - ✓ in fact every pixel is a potential vanishing point

Vanishing points and lines

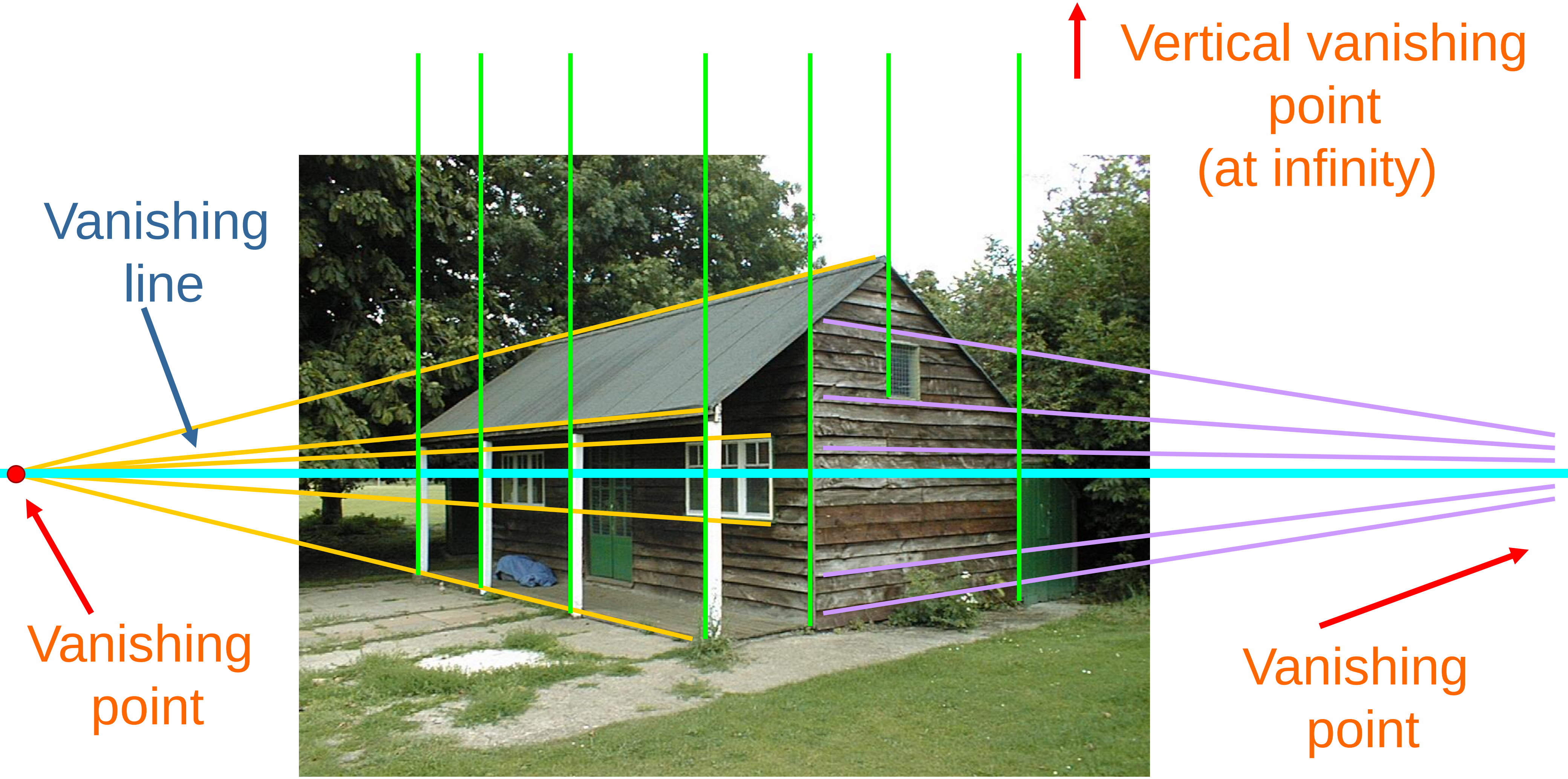


Vanishing points

- Each set of parallel lines (=direction) meets at a different point
  - The *vanishing point* for this direction
- Sets of parallel lines on the same plane lead to *collinear* vanishing points.
  - The line is called the *horizon* for that plane

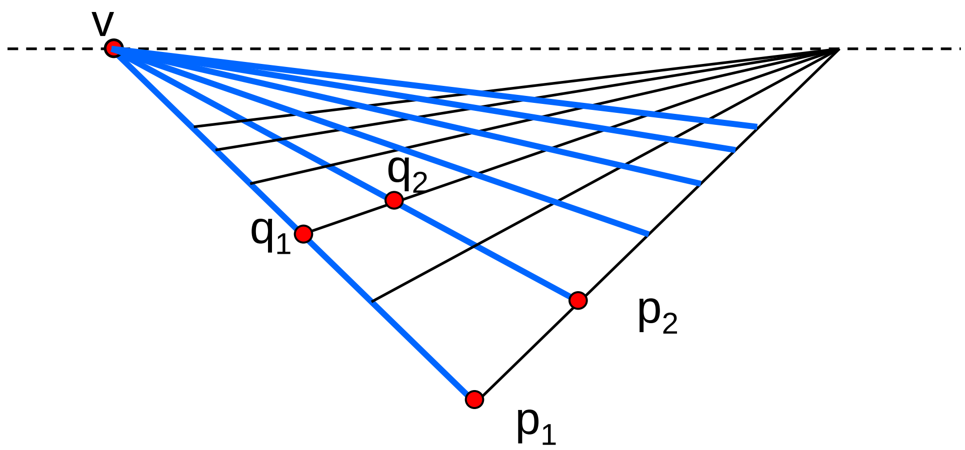


Perspective cues



Slide from Efros, Photo from Criminisi

Computing vanishing points (from lines)



- Intersect p_1q_1 with p_2q_2

$$v = (p_1 \times q_1) \times (p_2 \times q_2)$$

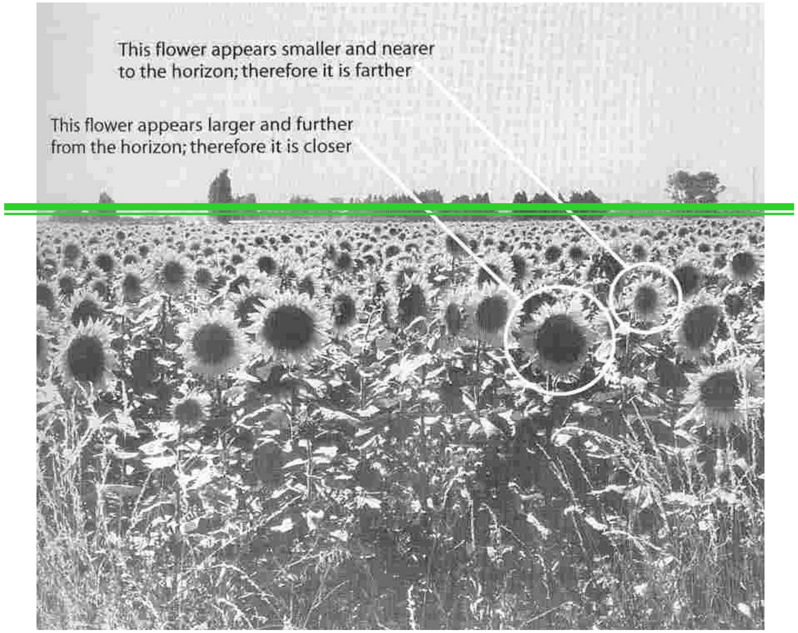
Least squares version

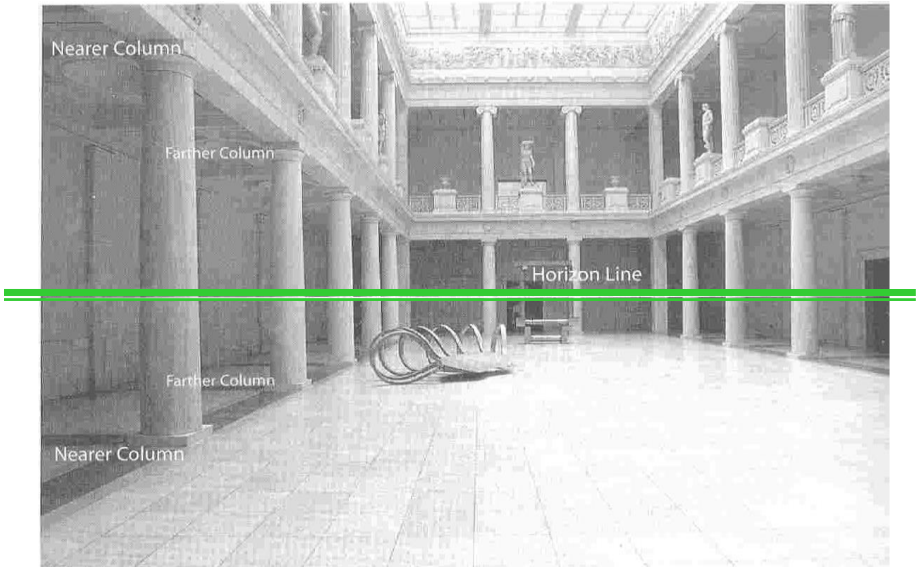
- Better to use more than two lines and compute the “closest” point of intersection
- See notes by Bob Collins for one good way of doing this:

http://www-2.cs.cmu.edu/~ph/869/www/notes/vanishing.txt

Distance from the horizon line

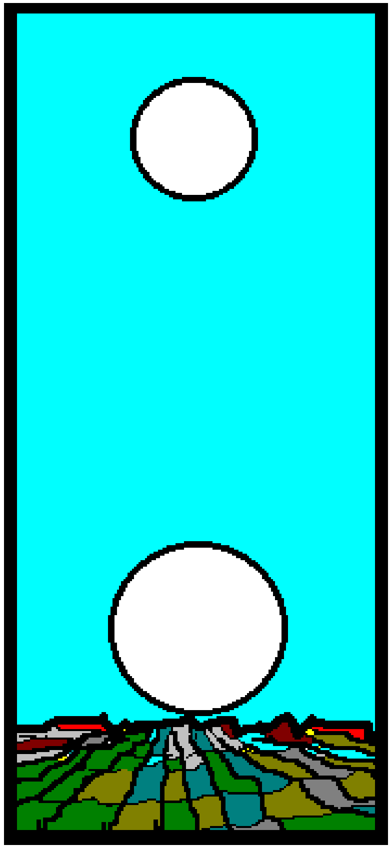
- Based on the tendency of objects to appear nearer the horizon line with greater distance to the horizon.
- Objects above the horizon that appear higher in the field of view are seen as being further away.
- Objects below the horizon that appear lower in the field of view are seen as being further away.



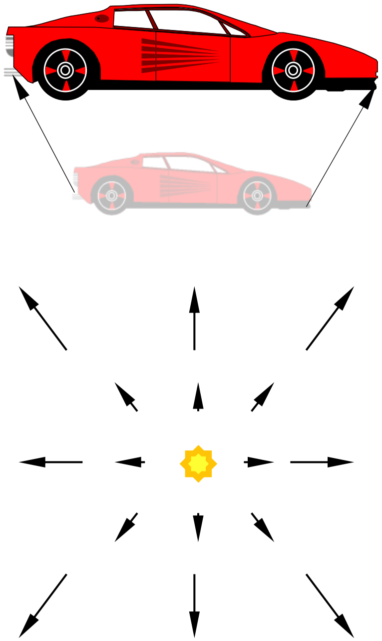


- Objects approach the horizon line with greater distance from the viewer.
- The base of a nearer column will appear lower against its background floor and further from the horizon line.
- Conversely, the base of a more distant column will appear higher against the same floor, and thus nearer to the horizon line.

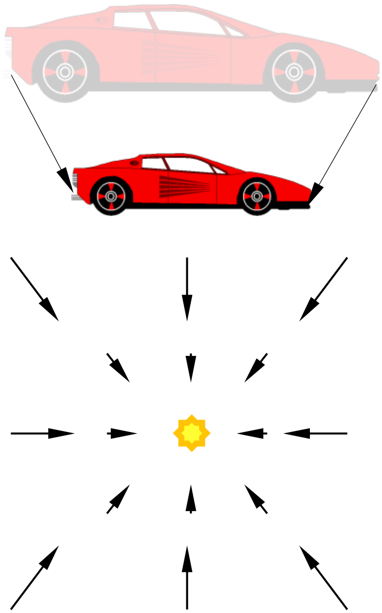
Moon illusion



Focus of expansion



Focus of contraction



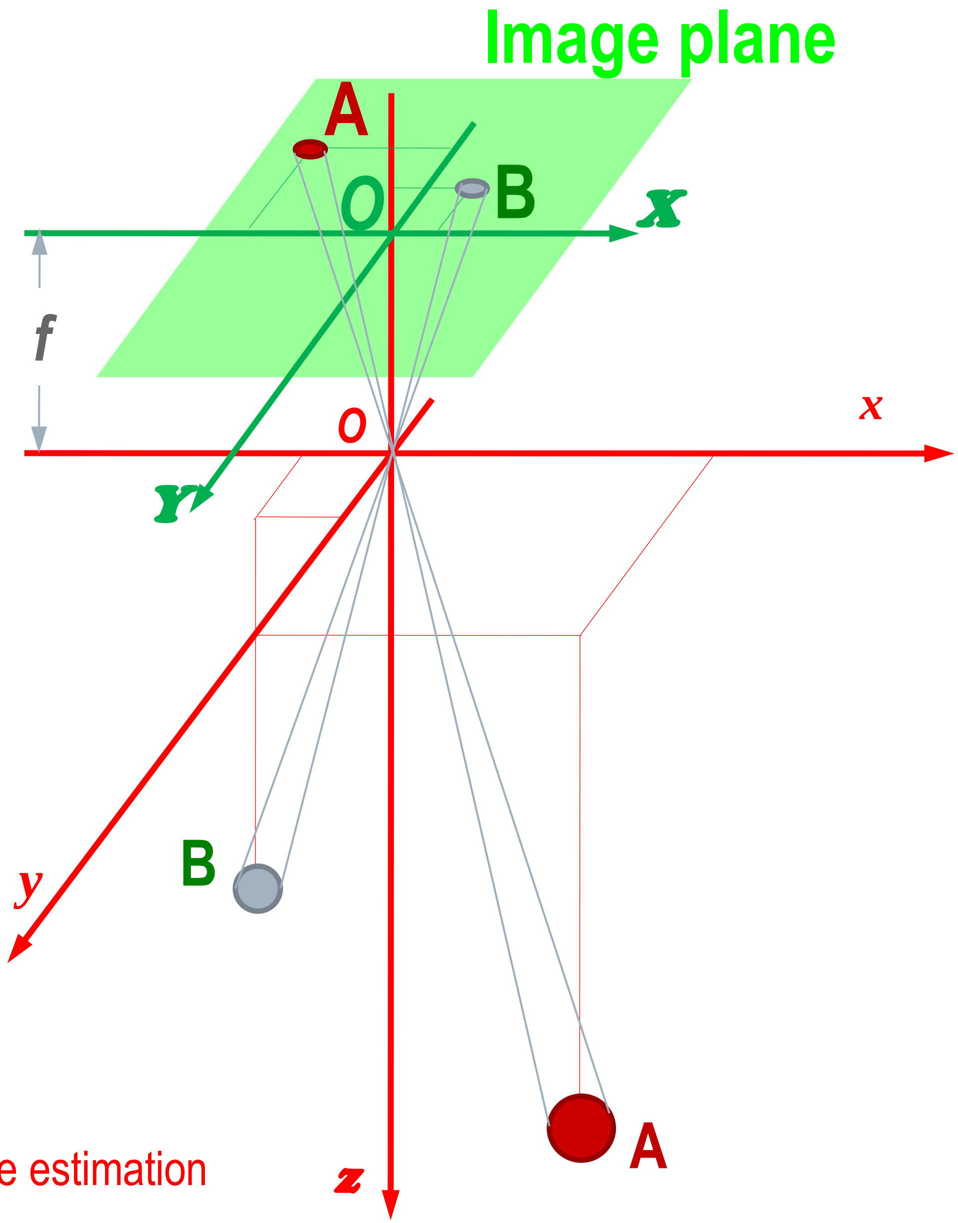
Shape from.....

Egomotion

$$\frac{Y}{y} = -\frac{f}{z}$$

$$\frac{\partial Y}{\partial z} = \frac{yf}{z^2} = -\frac{Y}{z}$$

$$z = -\frac{Y\,\partial z}{\partial Y}$$



Impact time estimation

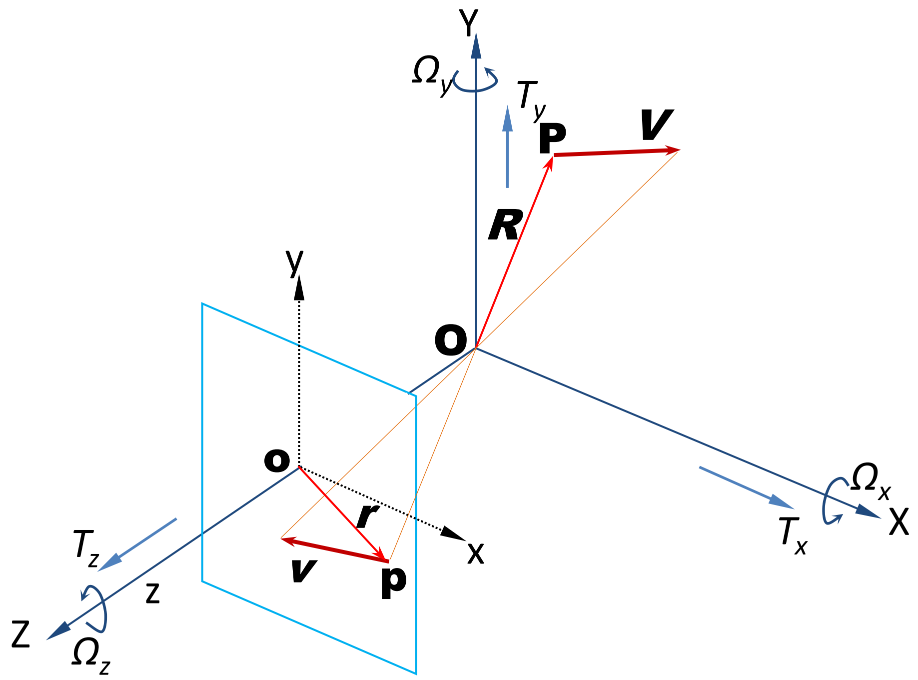
Camera and motion models

- *The egomotion makes all still objects in the scene to verify the same motion model defined by three translations **T** and three rotations **Ω**. Conversely, mobile obstacles pop out as not resorting to the former dominating model.*
- *Under such assumptions, the following classical equations hold:*

$$u_t = \frac{-fT_X + xT_Z}{Z},\; u_r = \frac{-xy}{f}\Omega_X - \left(\frac{-x^2}{f} + 1\right)\Omega_Y + y\Omega_Z$$

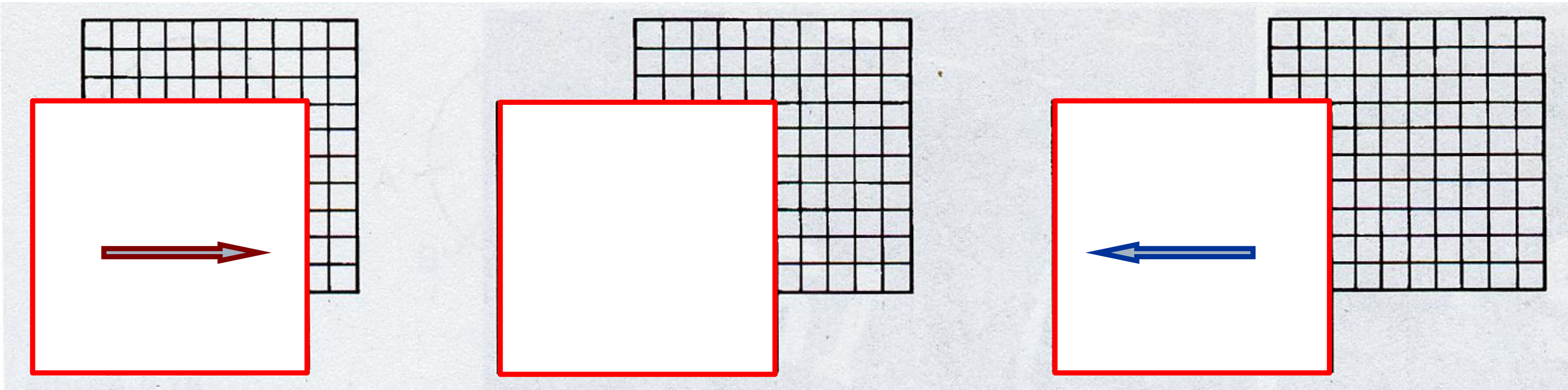
$$v_t = \frac{-fT_Y + yT_Z}{Z}, v_r = \frac{-xy}{f}\Omega_Y - \left(\frac{-y^2}{f} + 1\right)\Omega_X + x\Omega_Z$$

- *where* $\mathbf{w} = [u, v]^T = [u_t + u_r, v_t + v_r]^T$ *stands for the 2-D velocity vector of the pixel under the focal length **f**.*



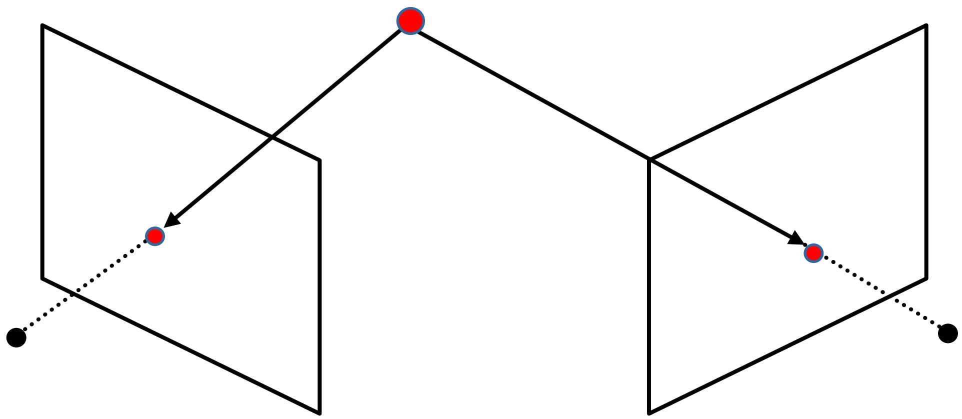
Motion occlusion and egomotion

Deletion and accretion occur when an observer moves in a direction not perpendicular to two surfaces that are at different depths. If an observer perceives the two surfaces as in the center and then moves to the left, deletion occurs so that the front object covers more that the back one, as shown on the left. Vice versa for the movement in the opposite direction as shown on the right



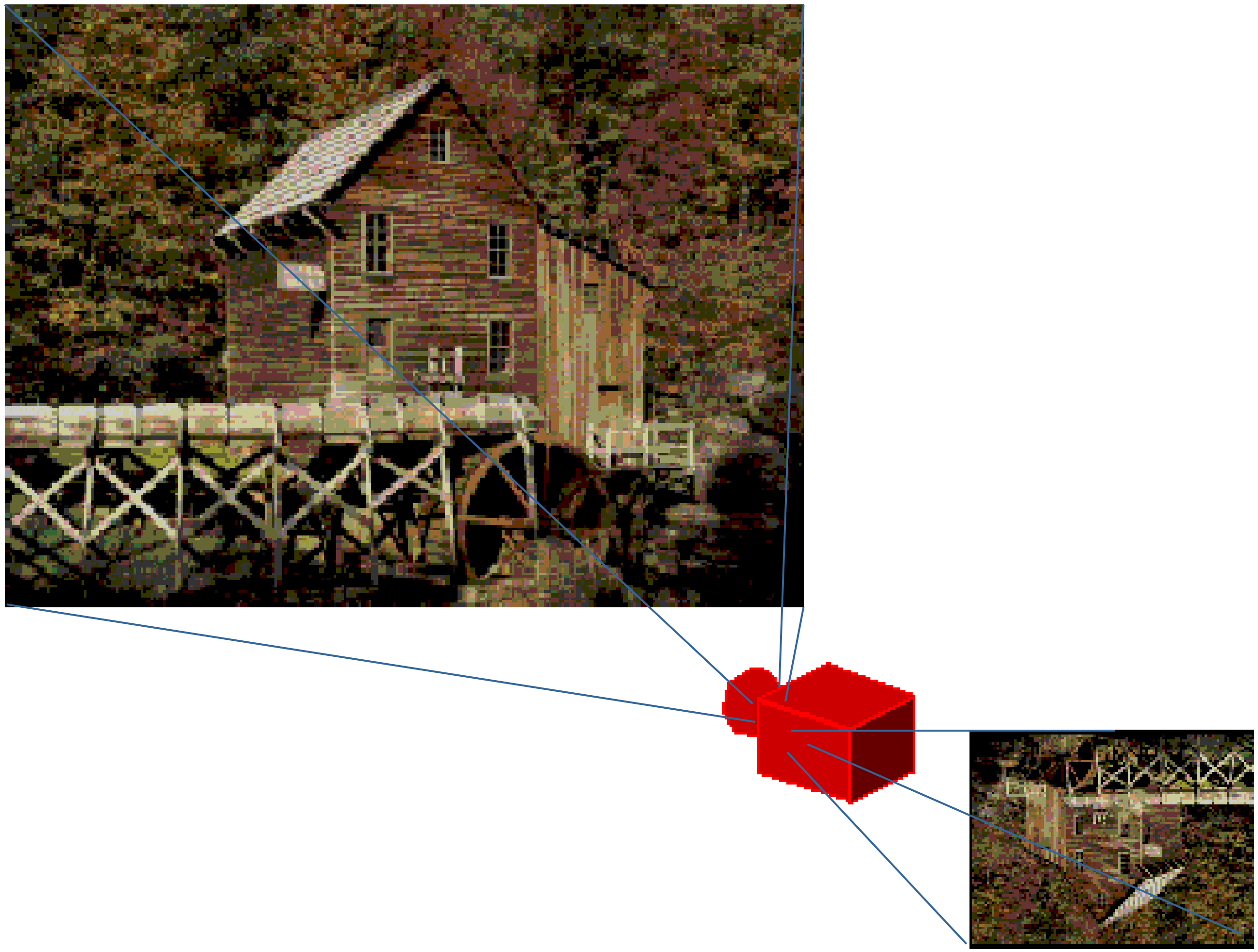
Deletion Initiale position Accretion

Stereo:
Epipolar geometry

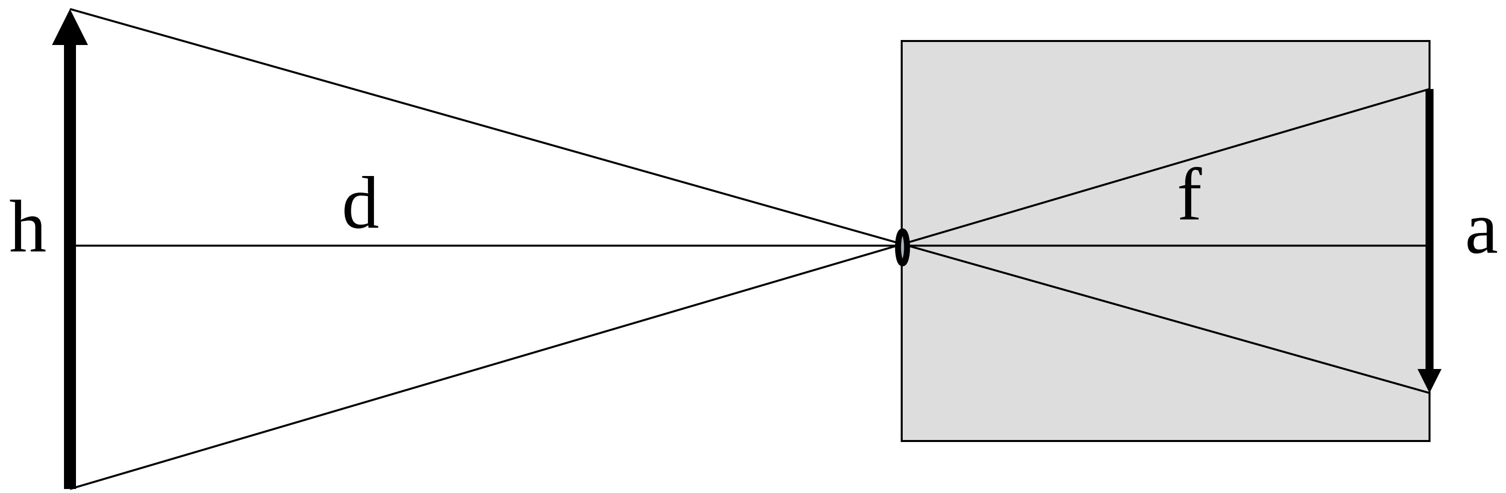


Slides by
Kristen Grauman

Pinhole camera model

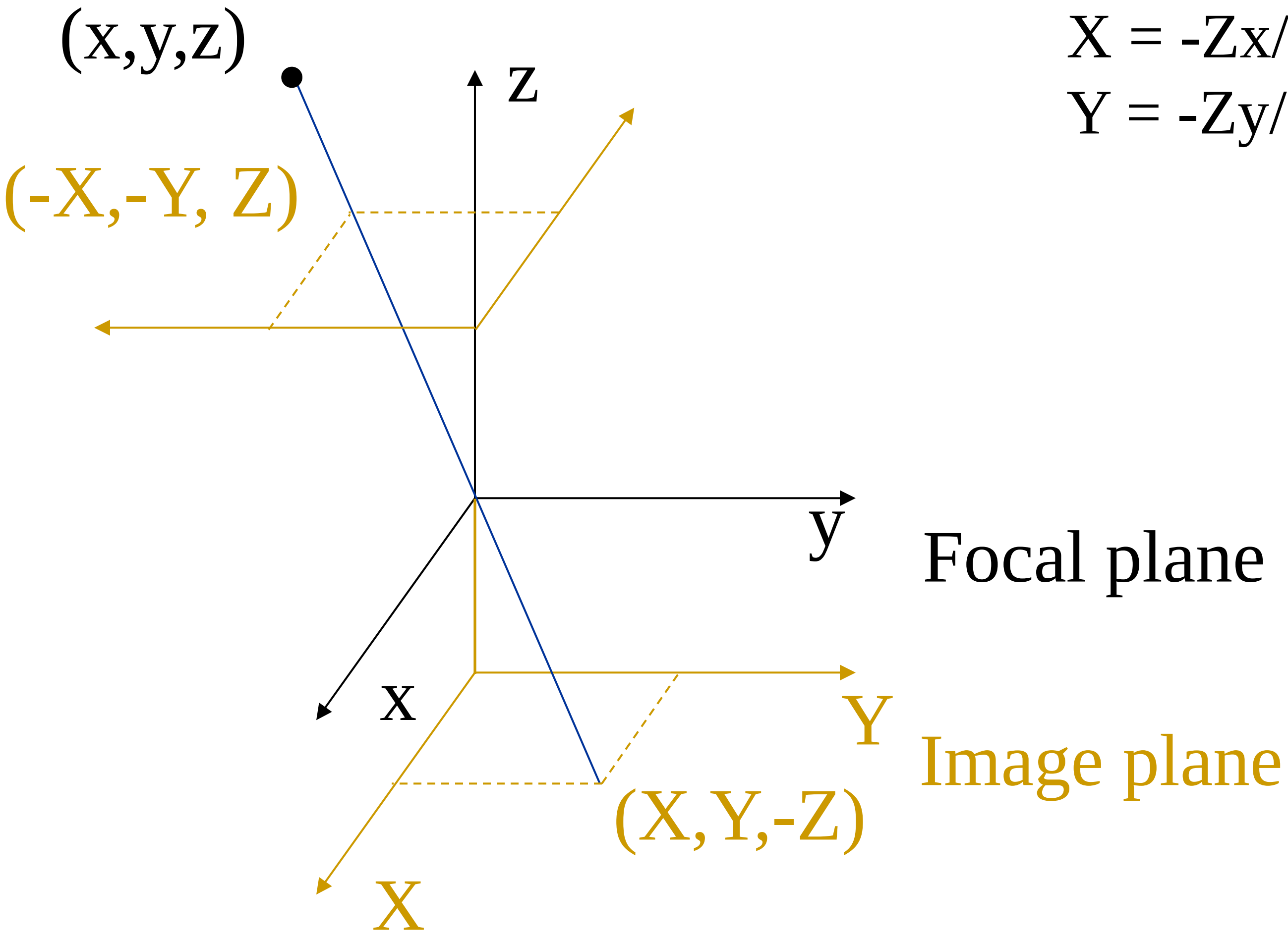


Pinhole camera model



$$h/d=a/f$$

Geometry of the camera



X = -Zx/z
Y = -Zy/z

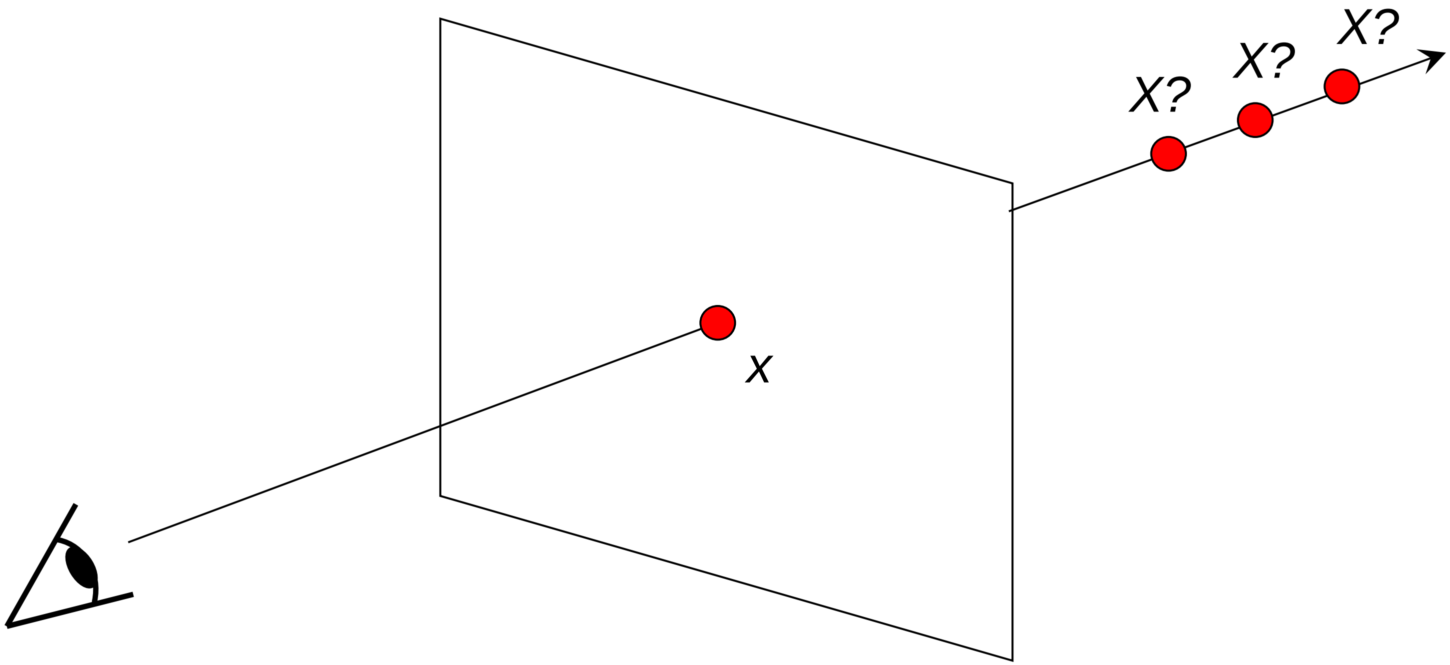
Why multiple views?

- Structure and depth are inherently ambiguous from single views.

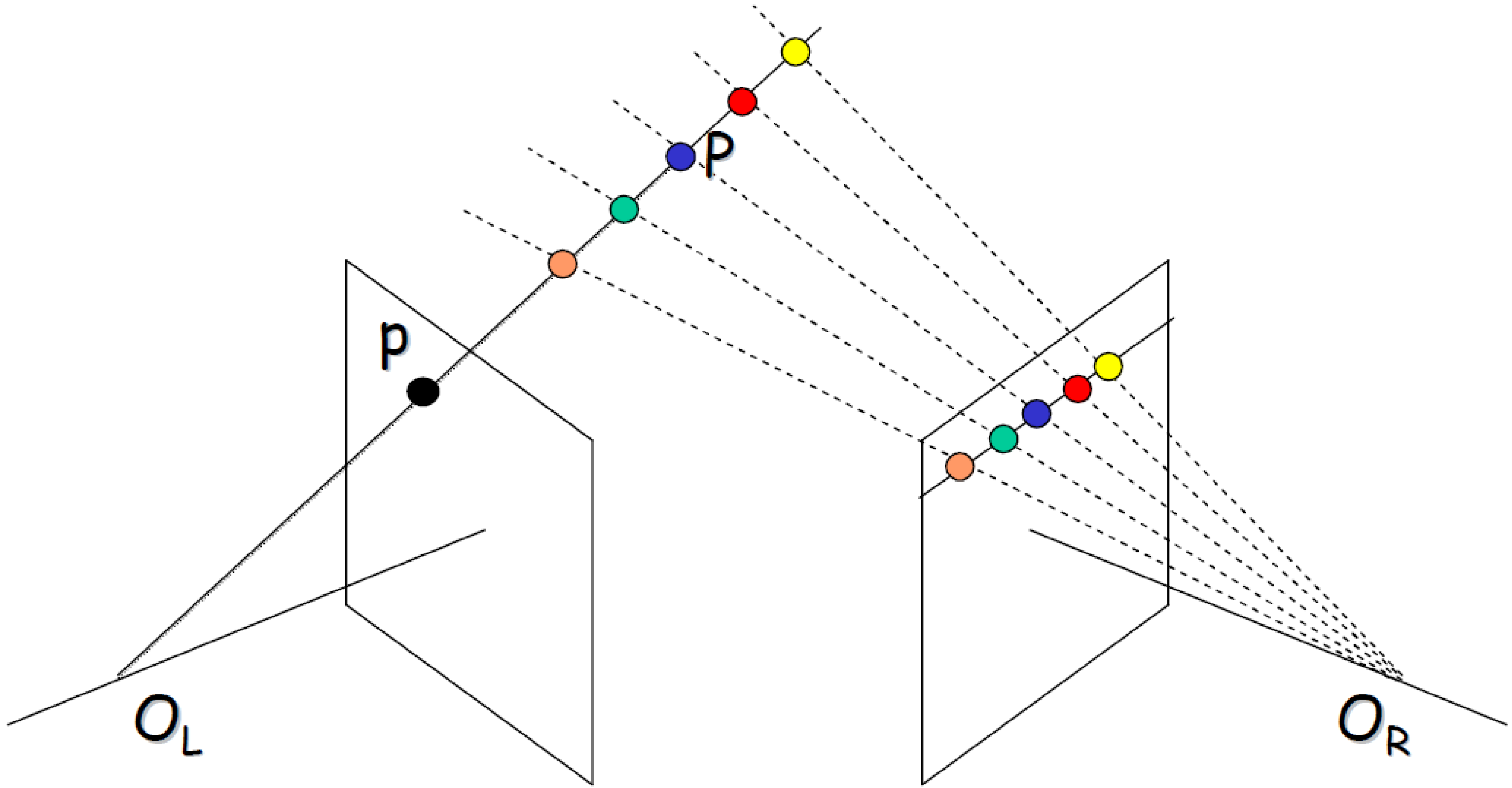
Images from Lana Lazebnik

Our goal: Recovery of 3D structure

- Recovery of structure from one image is inherently ambiguous



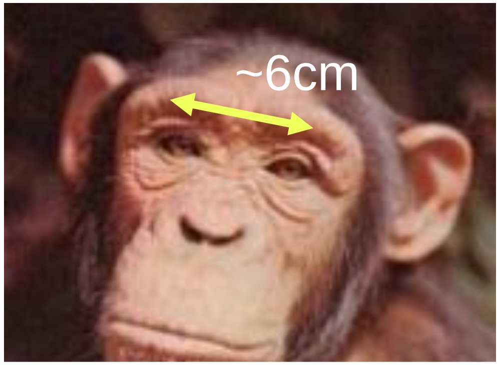
Why Stereo Vision?

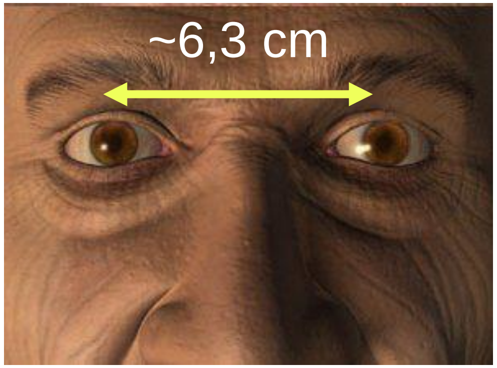


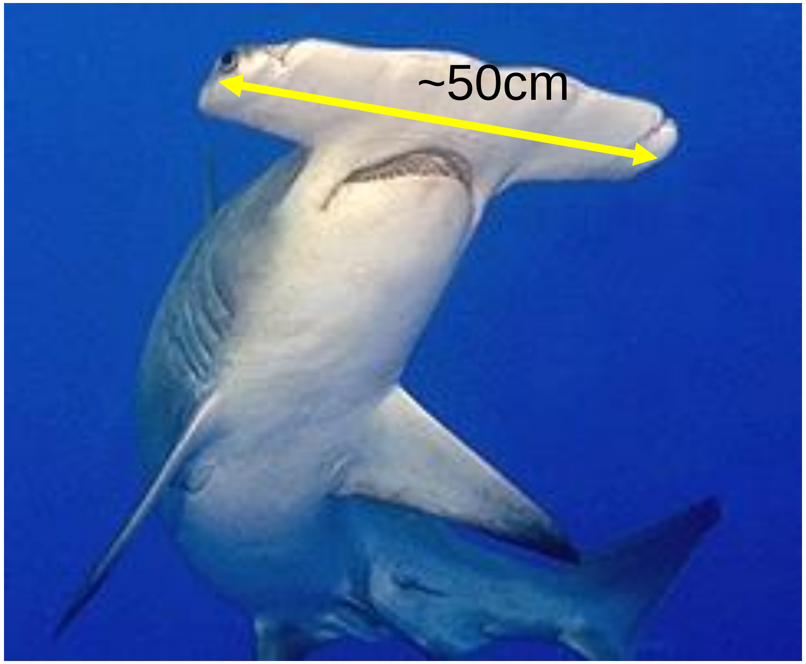
A second camera can resolve the ambiguity, enabling measurement of depth via triangulation.

Stereo vision

After 30 feet (10 meters) disparity is quite small and depth from stereo is unreliable…

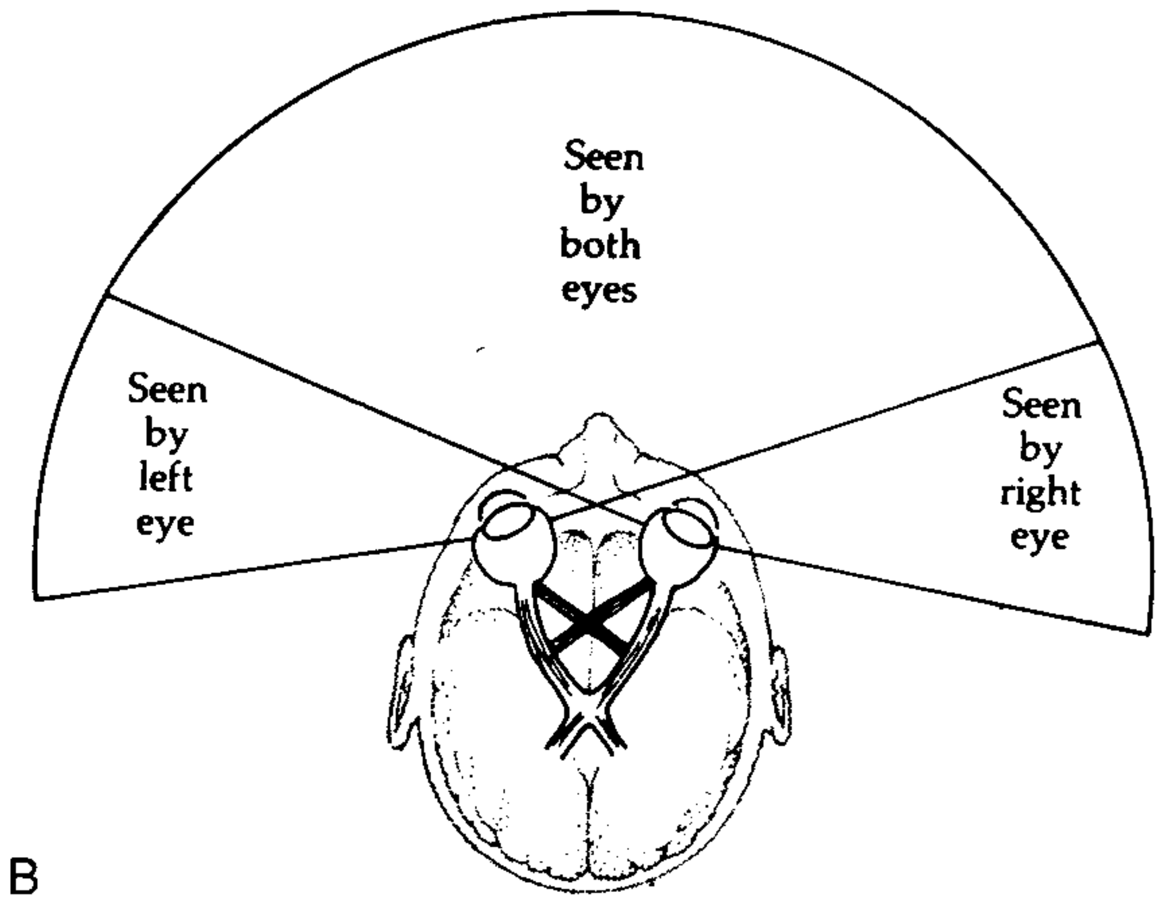




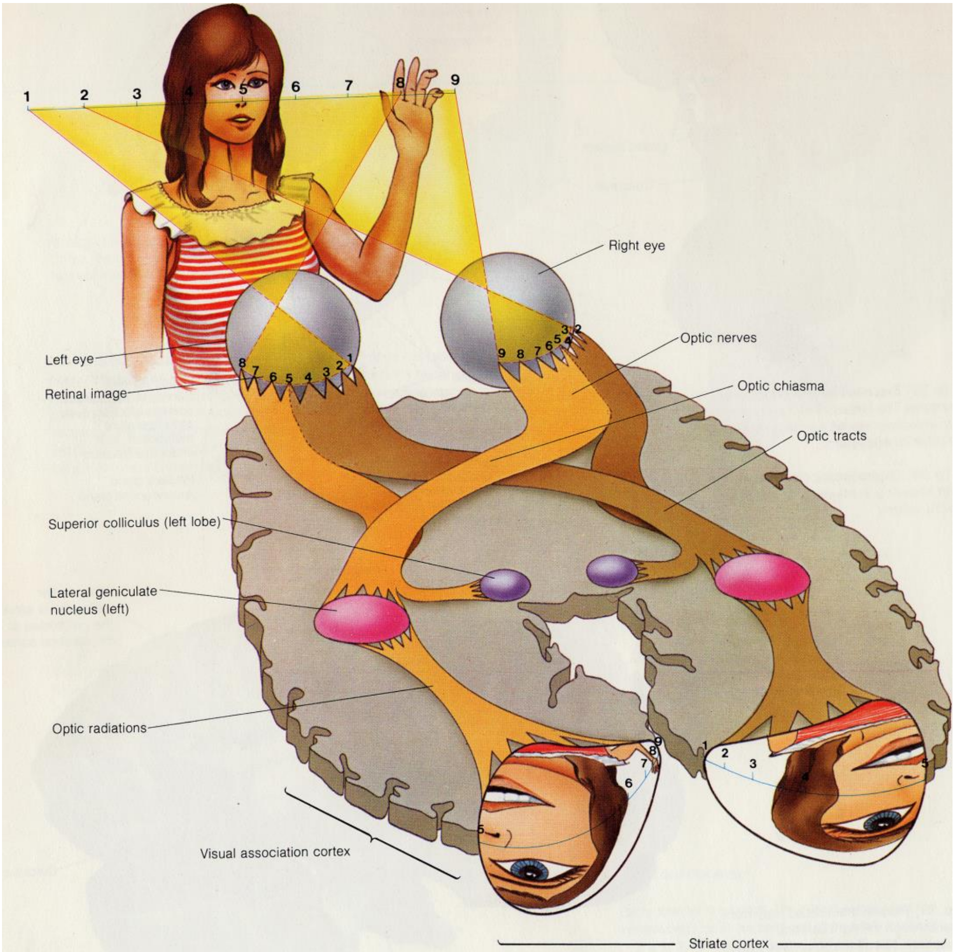


Monocular Visual Field: 160 deg (w) X 135 deg (h)

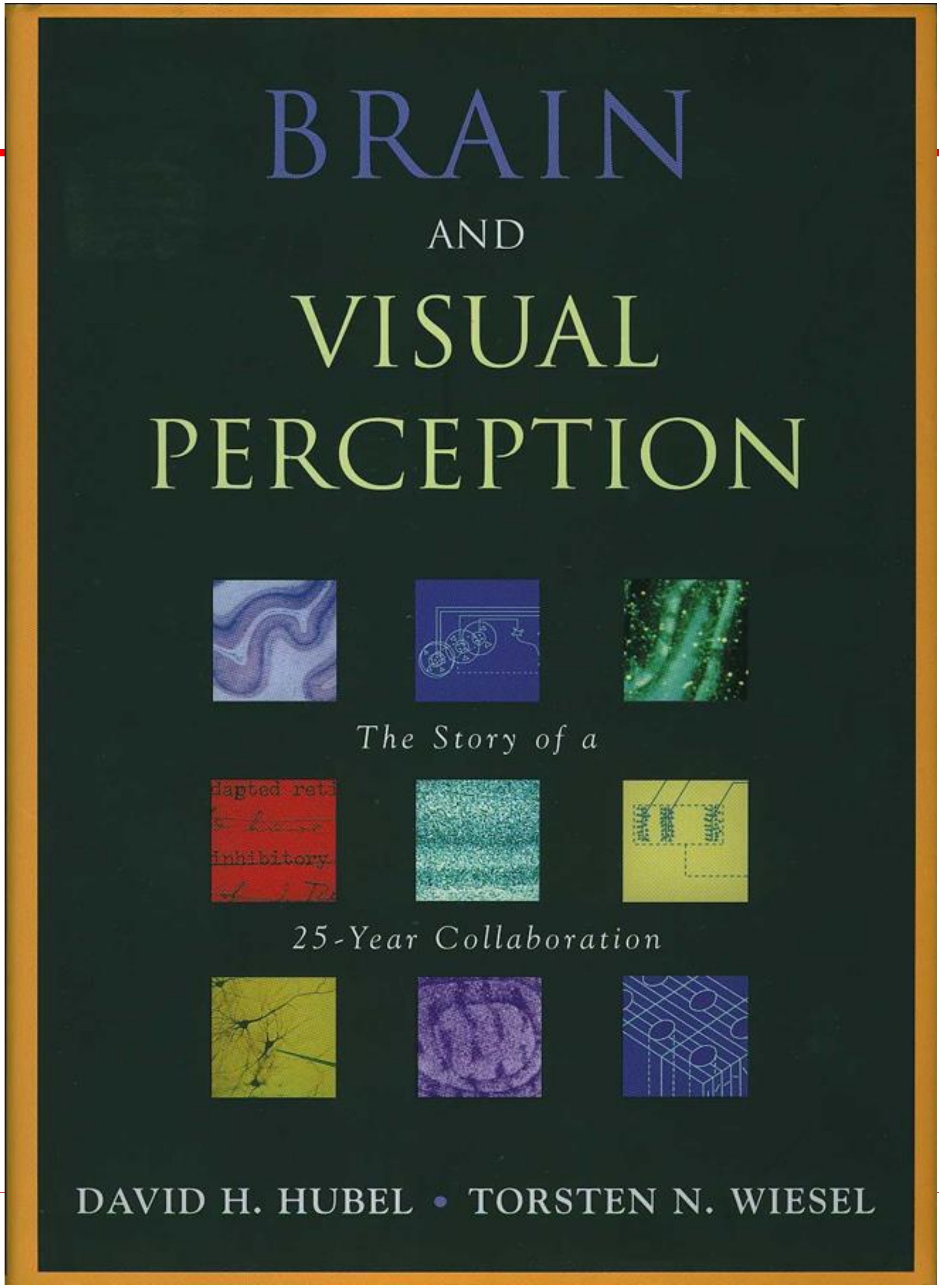
Binocular Visual Field: 200 deg (w) X 135 deg (h)



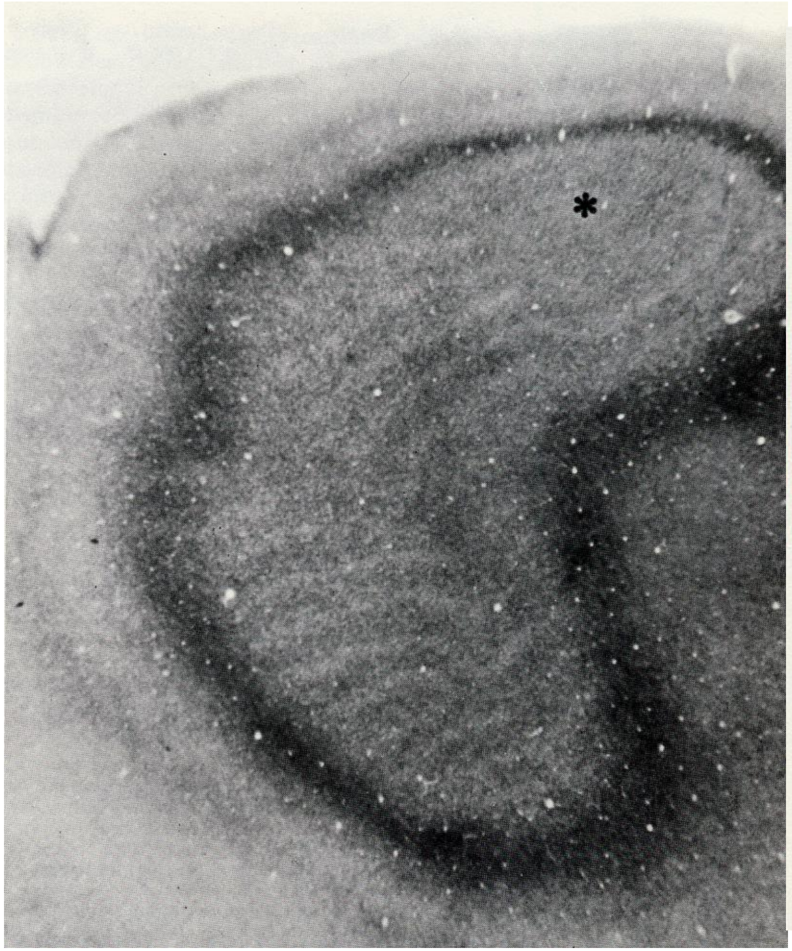
Schema of the two human visual pathways

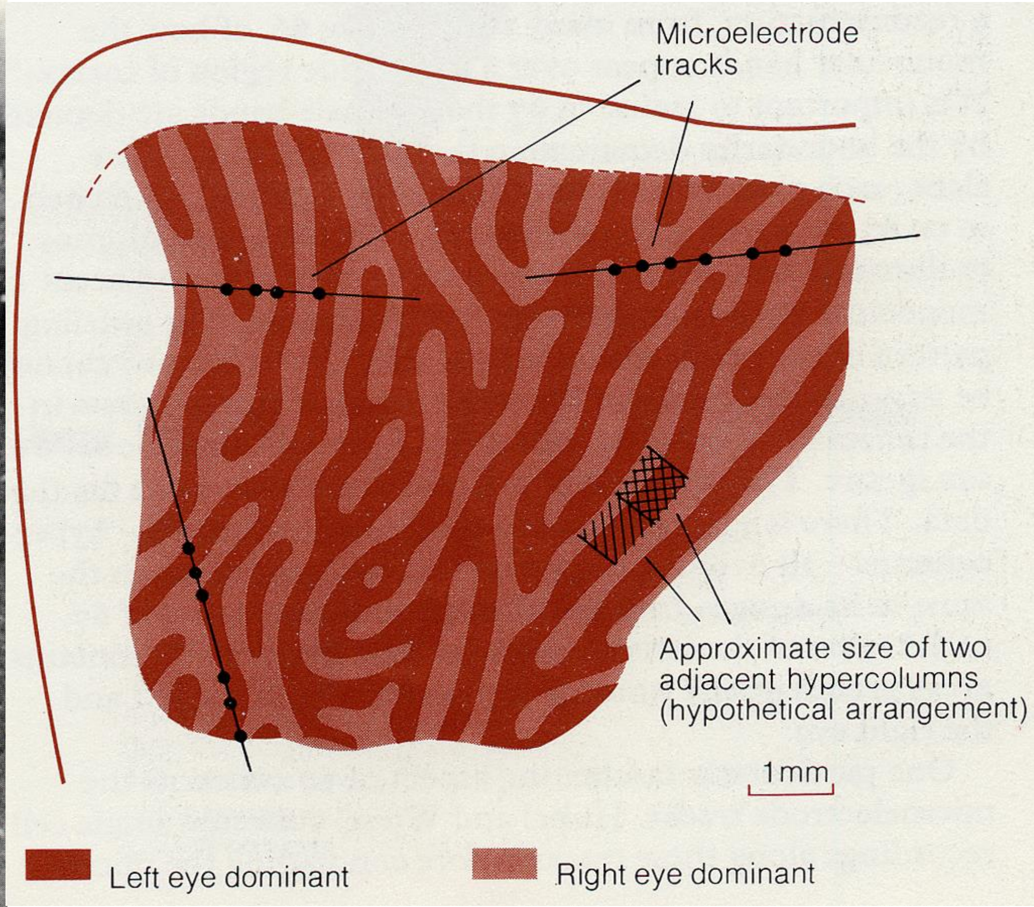


Illusion, Brain and Mind, John P. Frisby


BRAIN
AND
VISUAL
PERCEPTION
The Story of a
25-Year Collaboration
DAVID H. HUBEL • TORSTEN N. WIESEL

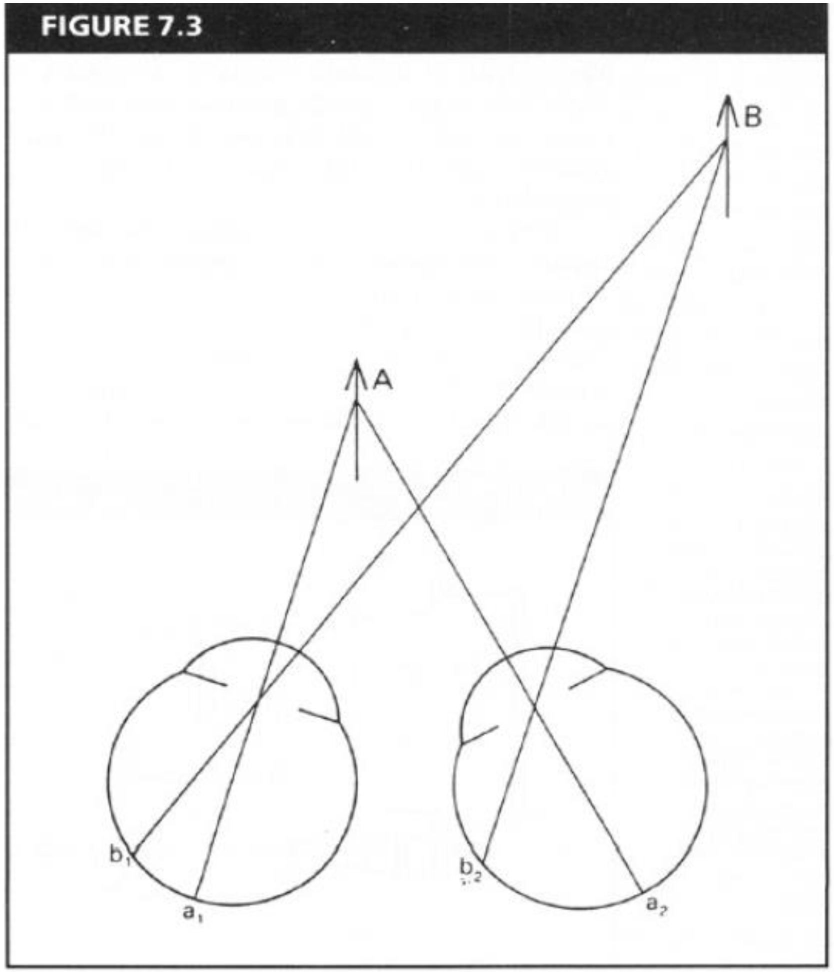
Section of striate cortex: schematic diagram of dominant band cells





Illusion, Brain and Mind, John P. Frisby

Human stereopsis: disparity

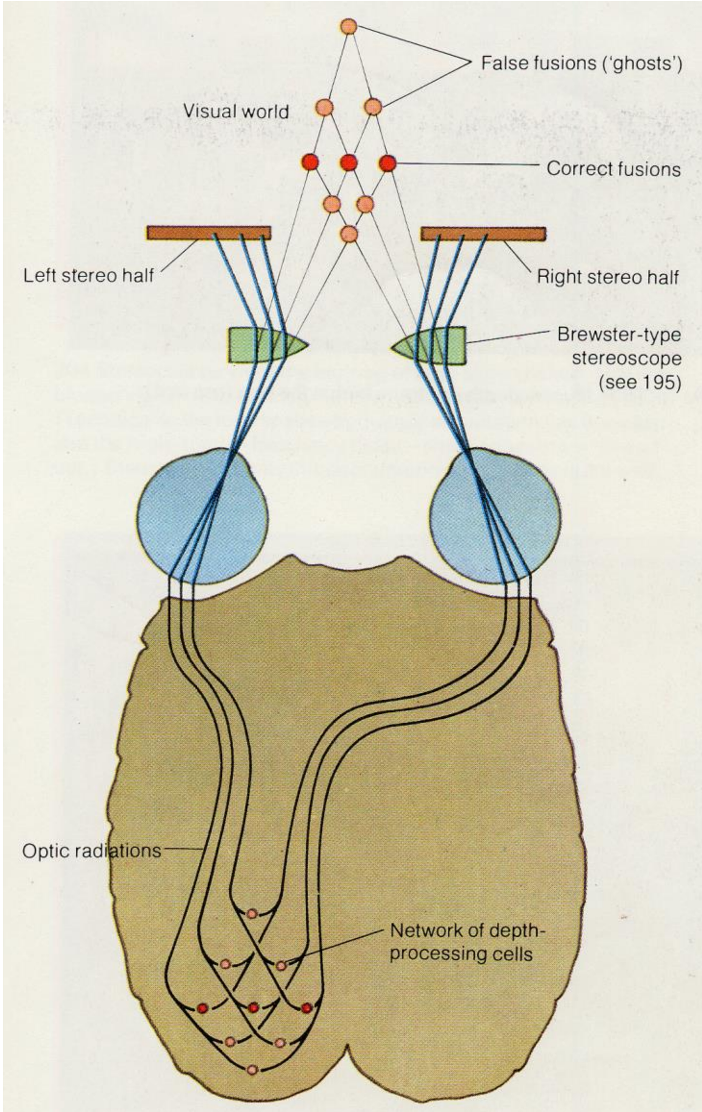


From Bruce and Green, Visual Perception, Physiology, Psychology and Ecology

- Human eyes **fixate** on point in space – rotate so that corresponding images form in centers of fovea.
- **Disparity** occurs when eyes fixate on one object; others appear at different visual angles

Adapted from David Forsyth, UC Berkeley

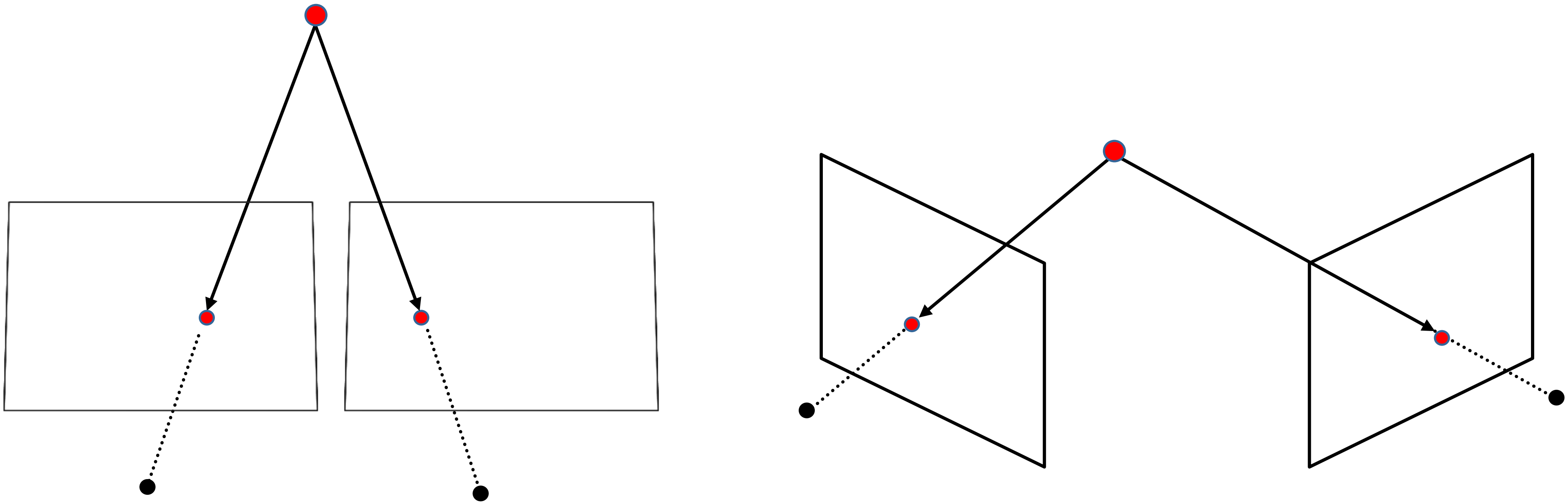
The problem of global stereopsis



Illusion, Brain and Mind, John P. Frisby

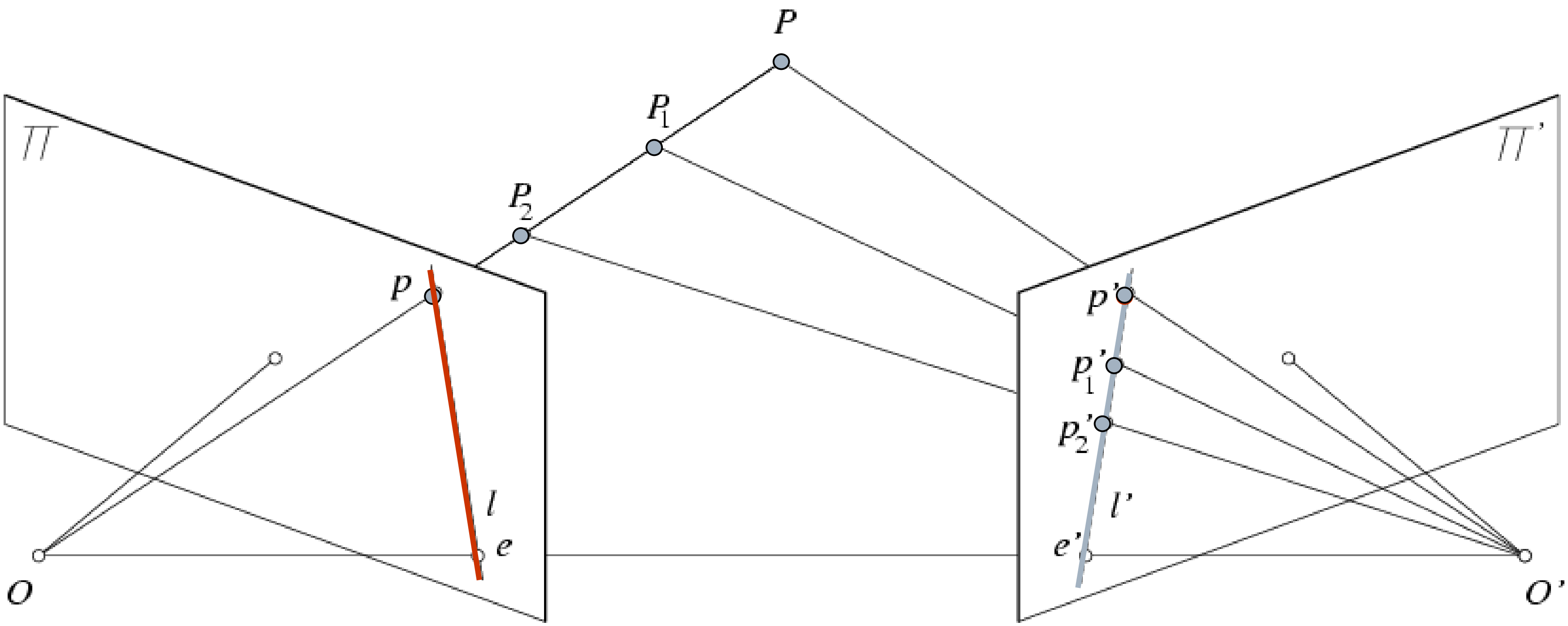
General case, with calibrated cameras

- The two cameras need not have parallel optical axes.



Vs.

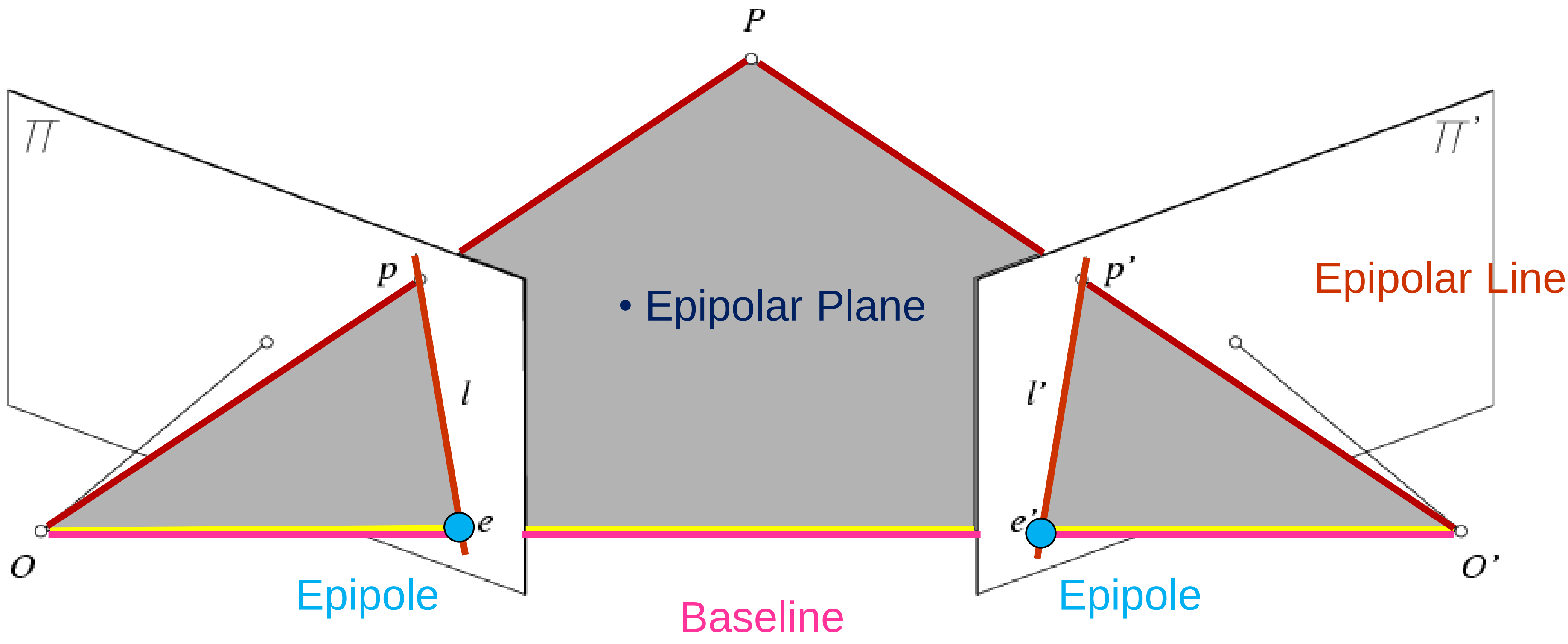
Epipolar constraint



Geometry of two views constrains where the corresponding pixel for some image point in the first view must occur in the second view.

- It must be on the line carved out by a plane connecting the world point and optical centers.

Epipolar geometry



http://www.ai.sri.com/~luong/research/Meta3DViewer/EpipolarGeo.html

Epipolar geometry: terms

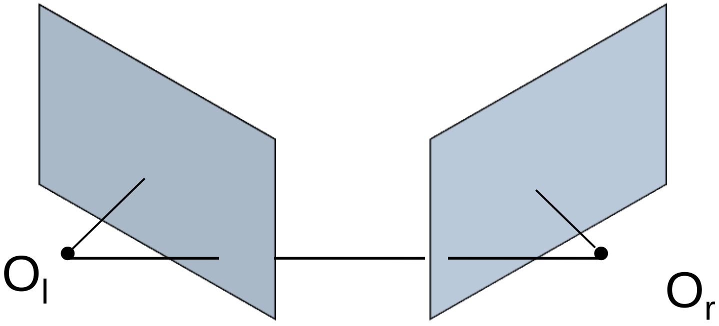
- **Baseline**: line joining the camera centers
- **Epipole**: point of intersection of baseline with image plane
- **Epipolar plane**: plane containing baseline and world point
- **Epipolar line**: intersection of epipolar plane with the image plane

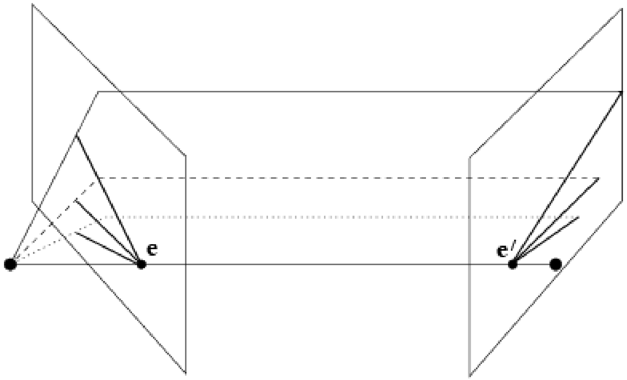
- All epipolar lines intersect at the epipole
- An epipolar plane intersects the left and right image planes in epipolar lines

Why is the epipolar constraint useful?

Example: converging cameras

What do the epipolar lines look like?





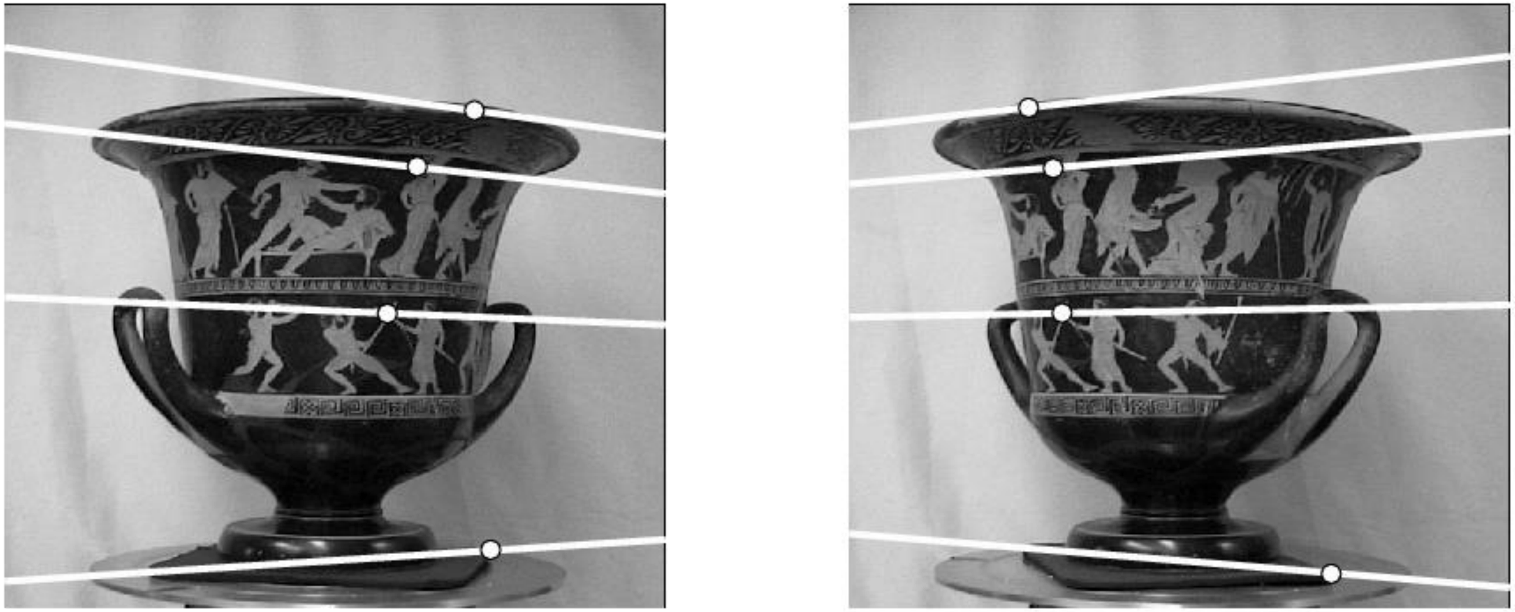
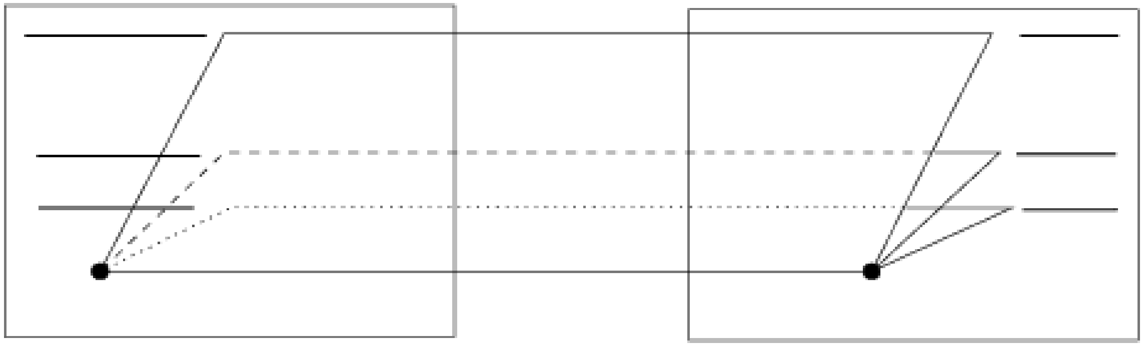


Figure from Hartley & Zisserman

Example: parallel cameras



Where are the epipoles?

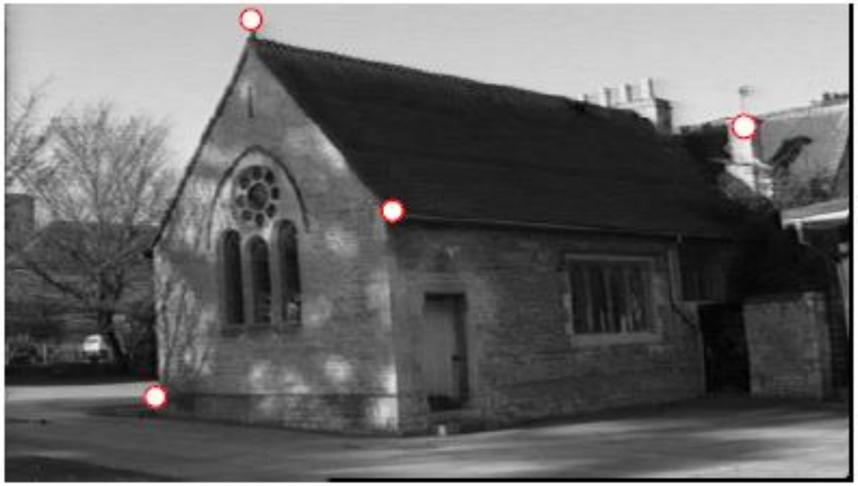
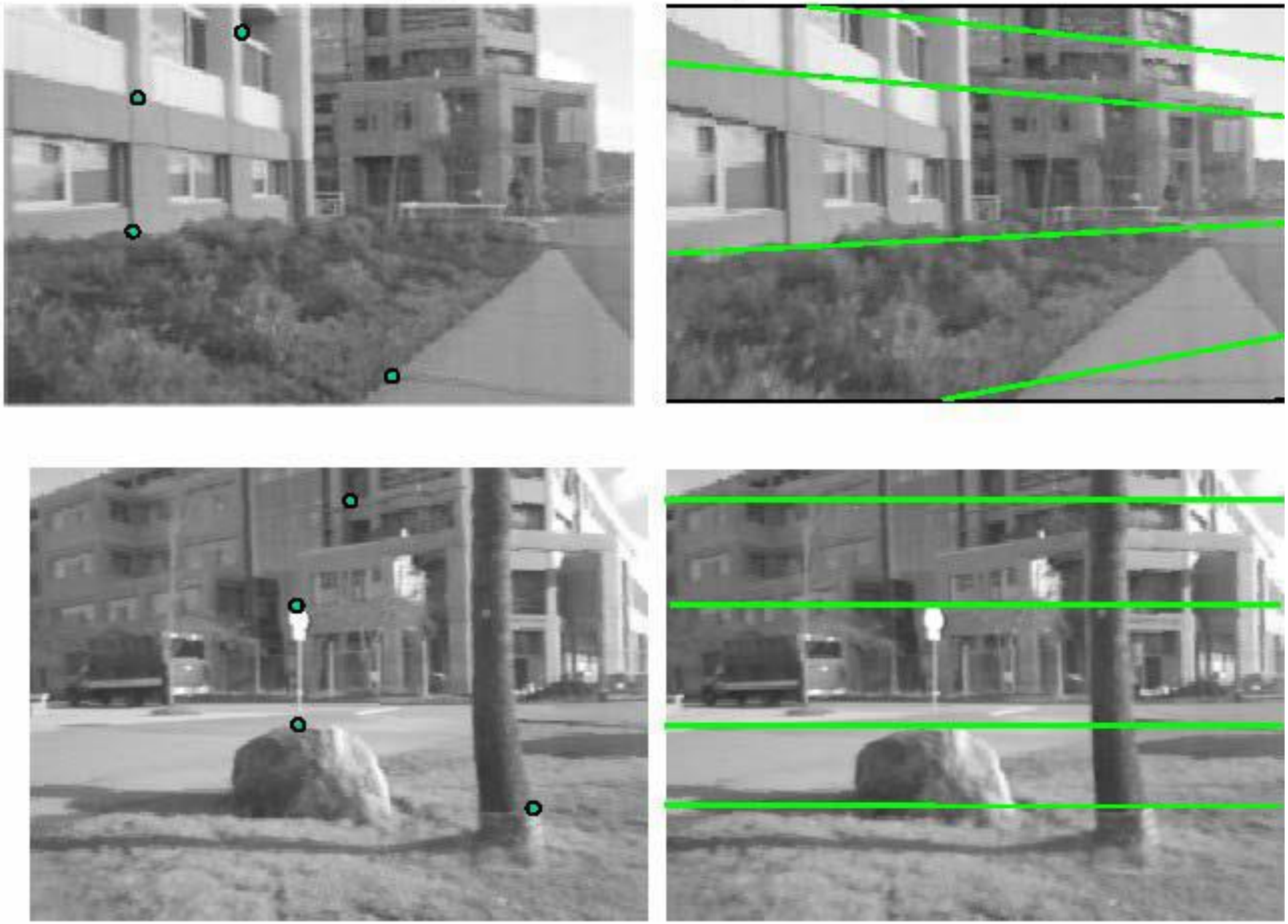
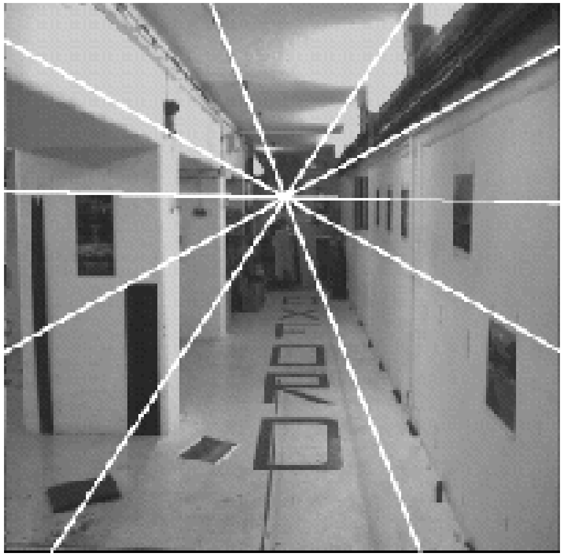


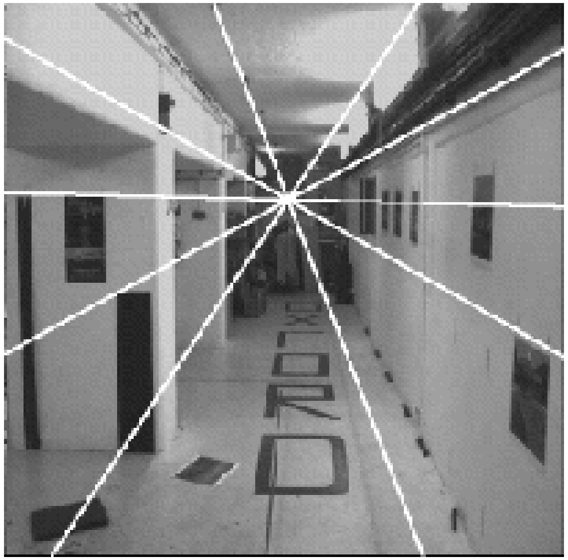
Figure from Hartley & Zisserman

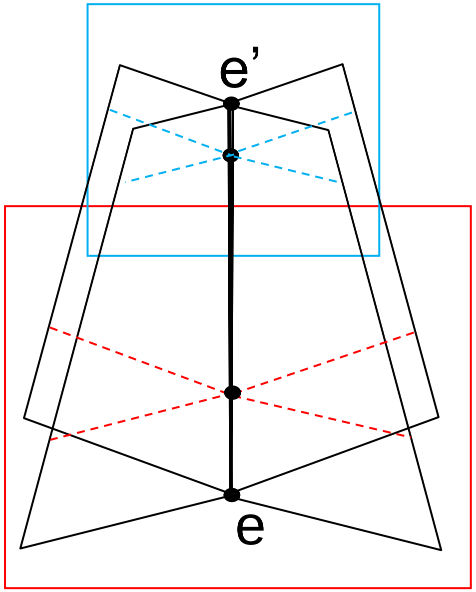
Epipolar constraint example



Example: Forward motion



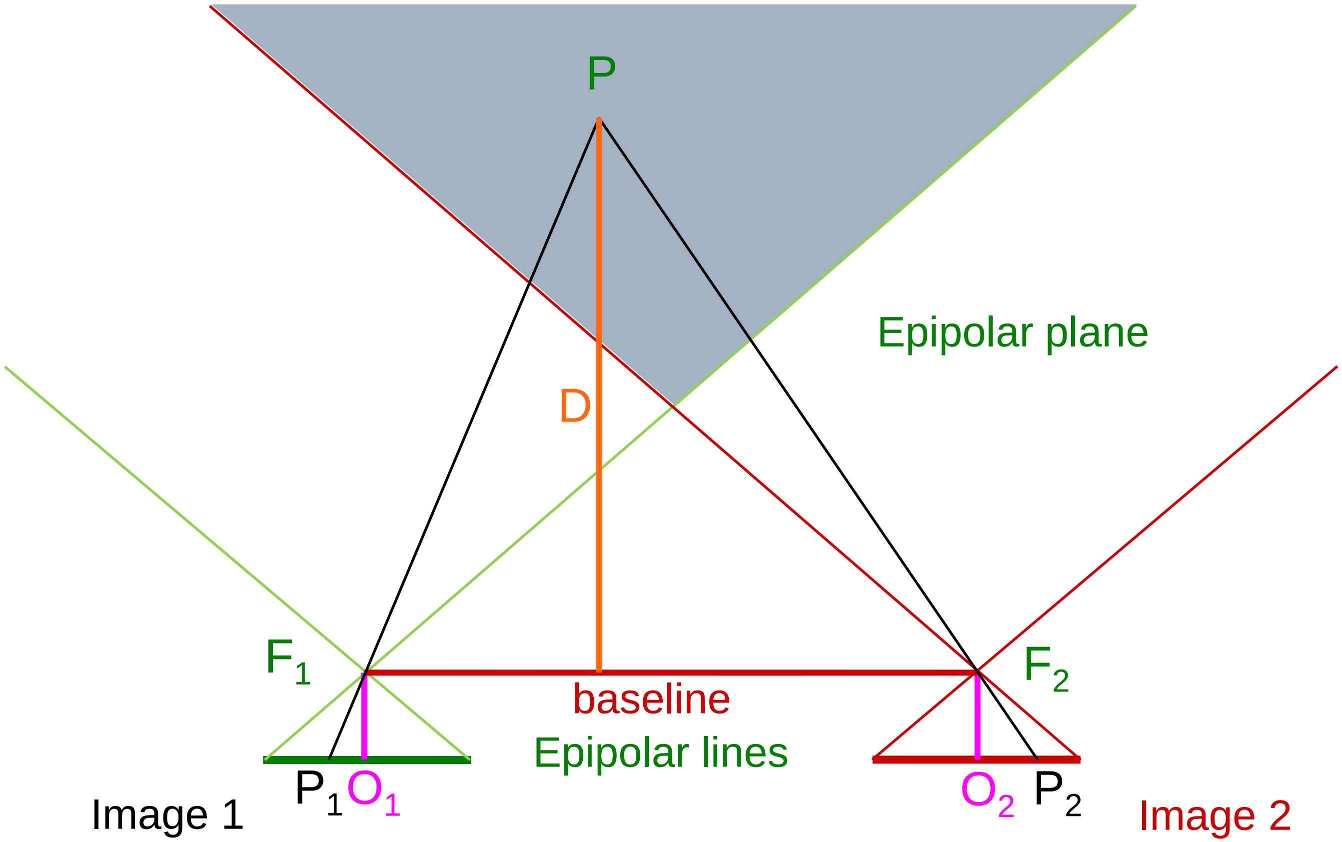




Epipole has same coordinates in both images.
Points move along lines radiating from e:
“Focus of expansion”

Correspondences – homologous points

- Stereo vision geometry: the light gray zone corresponds to the two view-points image overlapping area



Finding the D value

$$\frac{\overleftarrow{P_1 O_1}\ \overrightarrow{O_2 P_2}}{B} = \frac{f}{D}$$

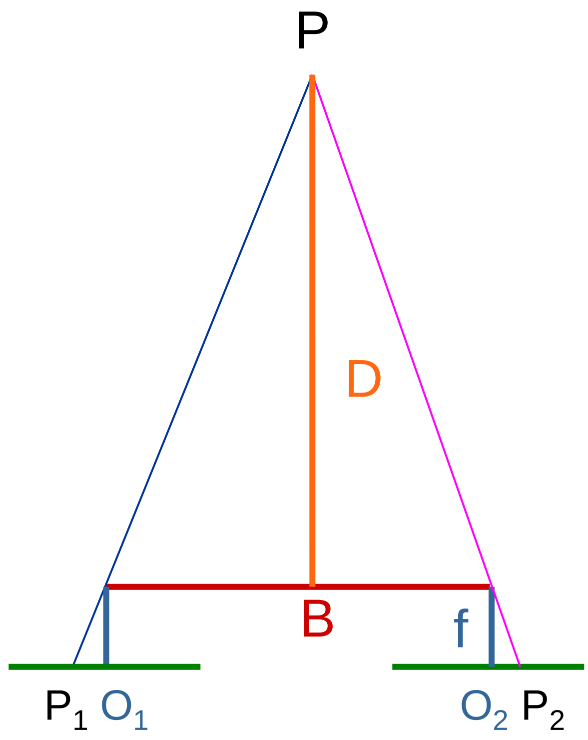
$$D = \frac{f\ B}{\Delta_1 + \Delta_2}$$

$\Delta_1 + \Delta_2$ displacements on the epipolar lines

- The influence of the distance D on the error of the computed $\Delta = \Delta_1 + \Delta_2$ is evidenced by mere derivation:

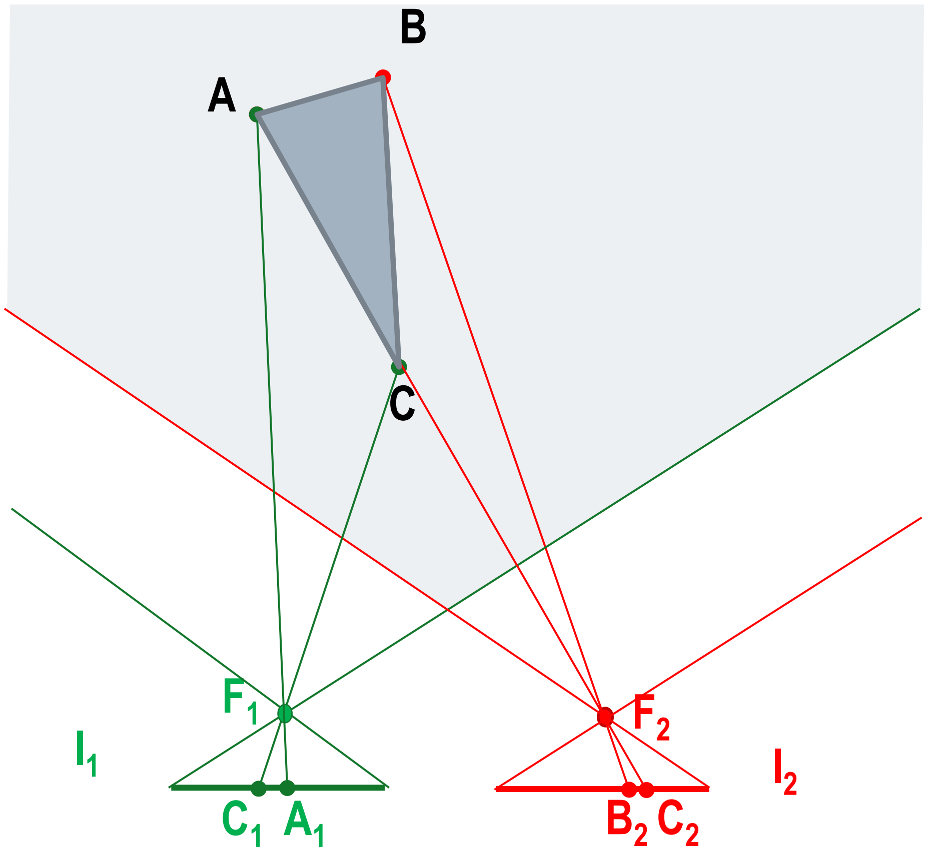
$$\frac{\partial D}{\partial \Delta} = -\frac{D}{\Delta}$$

- Note that the error increases linearly with the depth and is amplified in case of small Δ values.

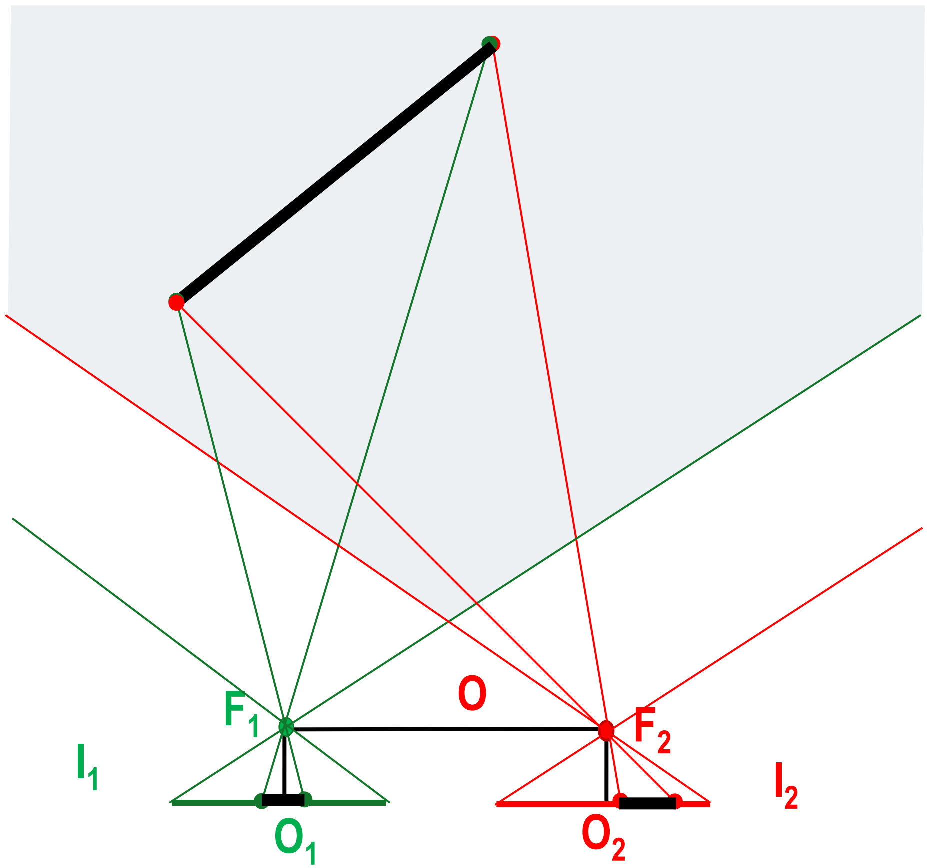


Looking for the tie point

Occlusions : B is occluded in I_1, while A in I_2

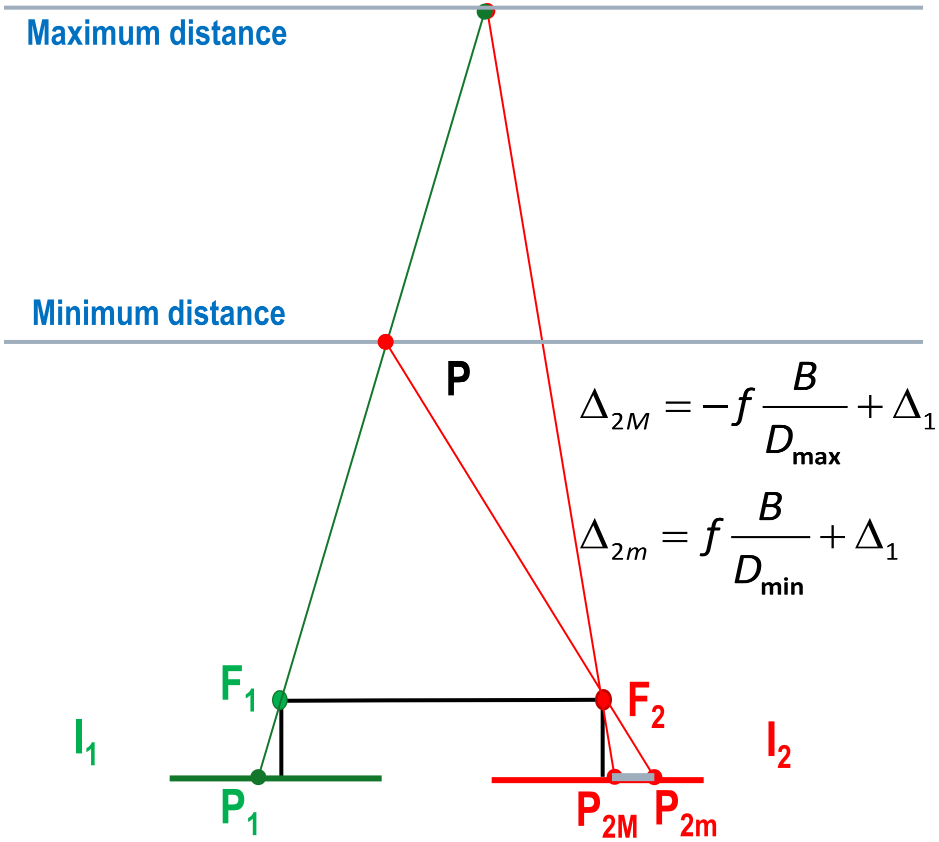


Distorted views due to different projections



Looking for the tie point

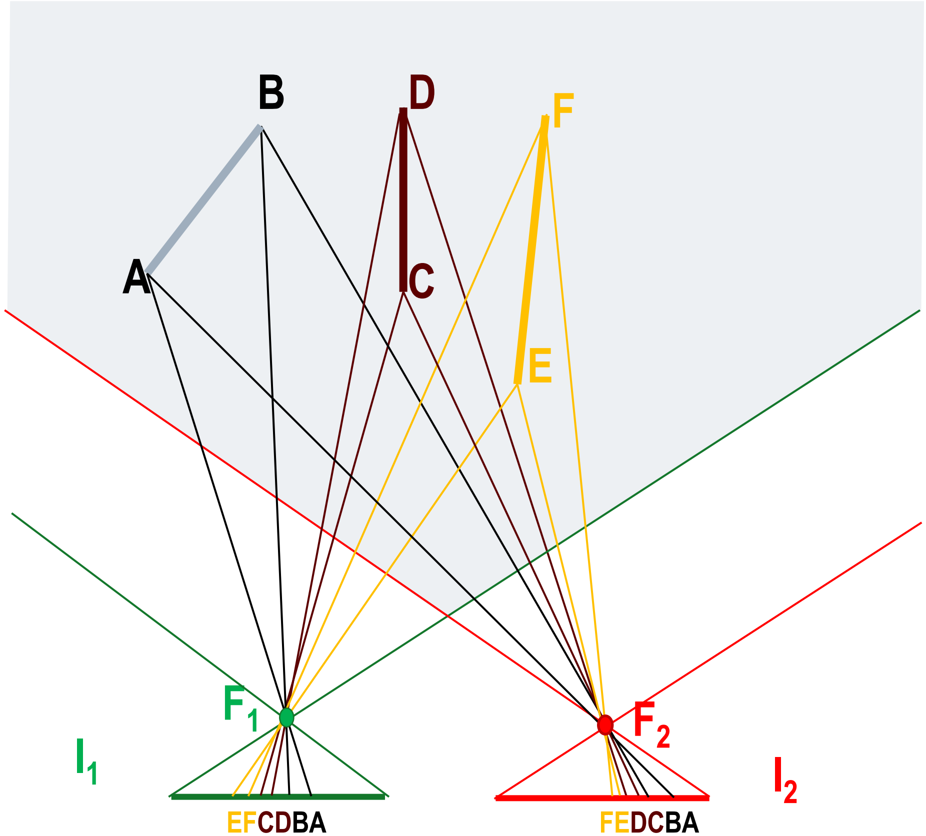
The epipolar segment P_{2M} P_{2m}



$$\Delta_{2M} = -f\frac{B}{D_{max}} + \Delta_1$$

$$\Delta_{2m} = f\frac{B}{D_{min}} + \Delta_1$$

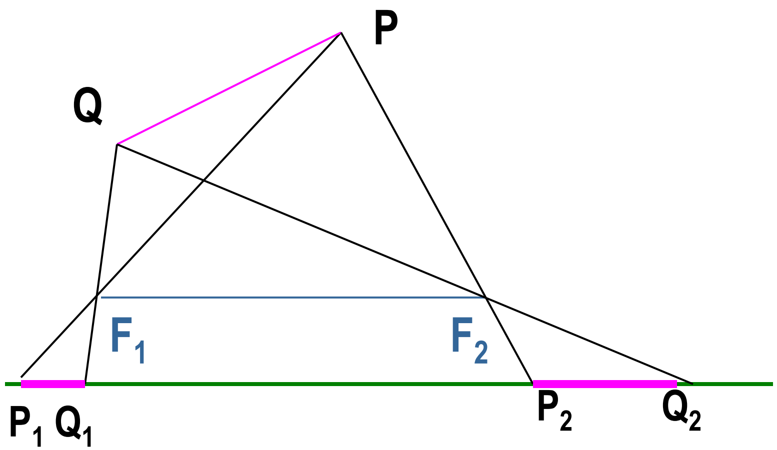
The ordering problem as seen by the letter sequence on each image

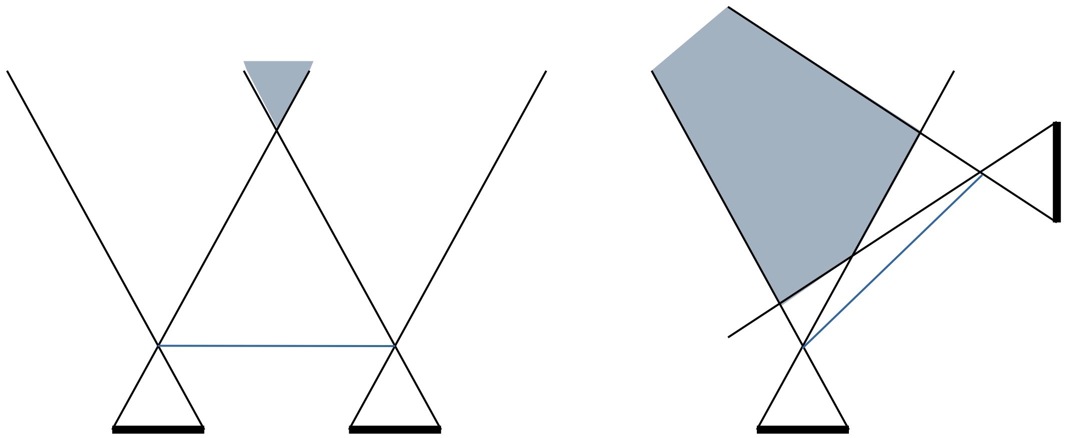


Looking for the tie point

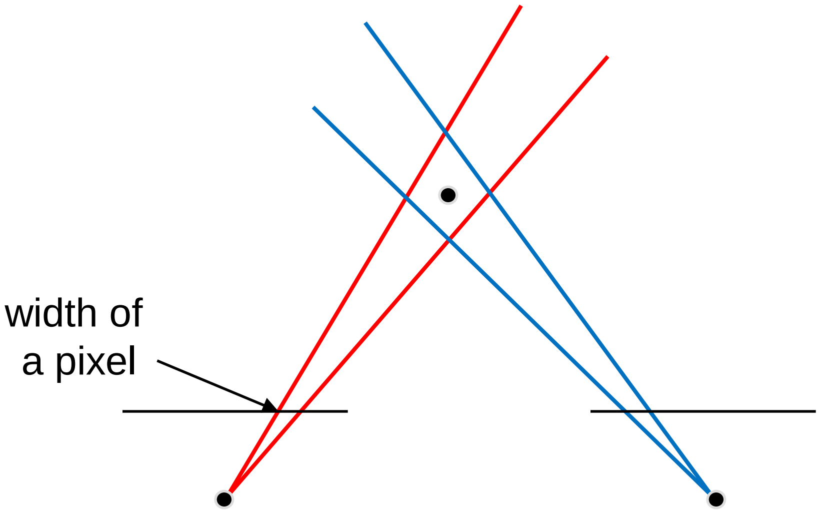
The higher the baseline the higher the deformation and the lower the overlapping

To obtain an extended overlapping area it is often necessary to tilt the camera axis

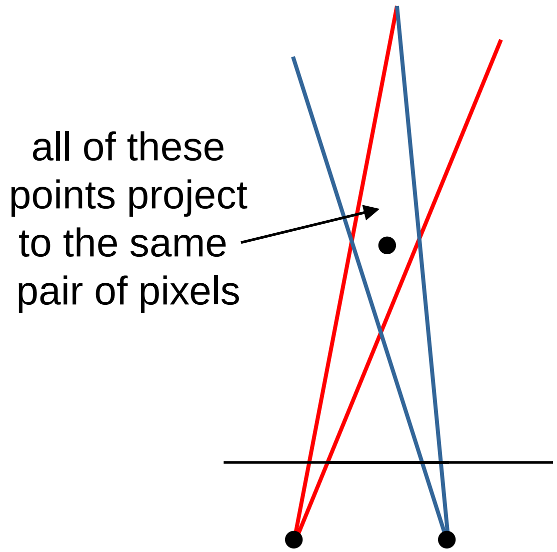




Choosing the stereo baseline



Large Baseline



Small Baseline

- What's the optimal baseline?
  - Too small: large depth error
  - Too large: difficult search problem

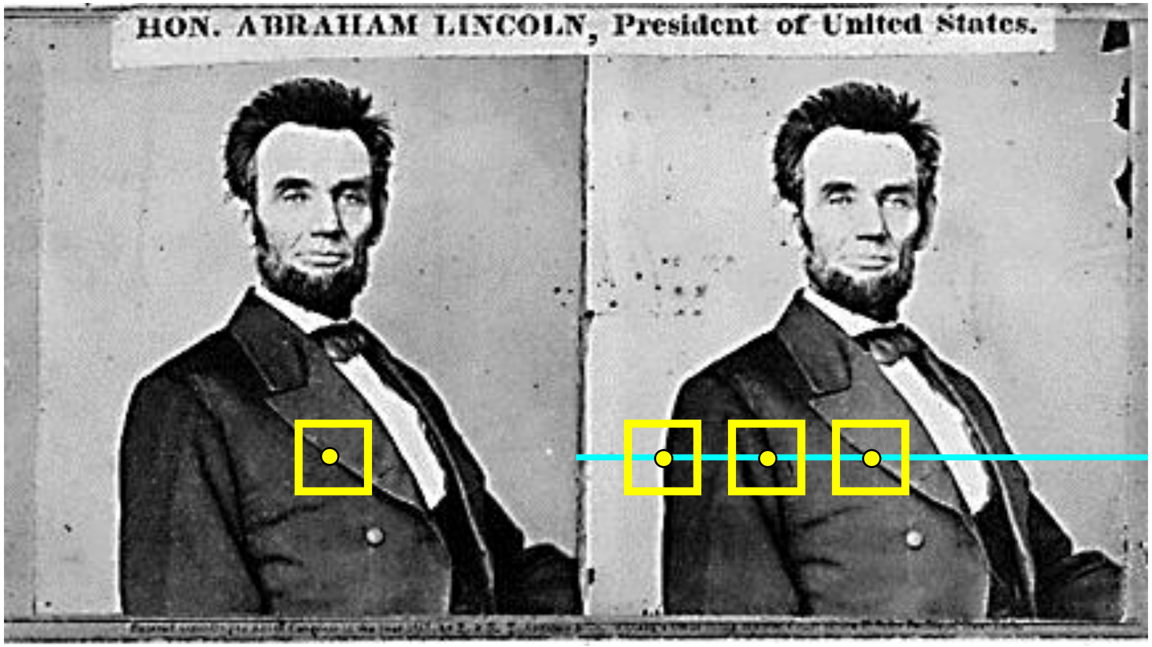
Homologous points

- The simplest ways to determine if a given pixel (p, q) on one image I_1 is a good candidate, is to evaluate the gray level variance in a limited neighborhood of such pixel.
- If its value exceeds a given threshold, then a neighborhood (2n+1)x(2m+1) is considered and *correlated* with candidate regions on image I2.
- Candidate regions are selected on the epipolar line; in order to compute the correlation between regions of both images the following formula may be used:

$$C(i,j)=\sum_{r=-n}^{n}\sum_{s=-m}^{m}\left[I_2(i+r,j+s)-I_1(p+r,q+s)\right]^2$$

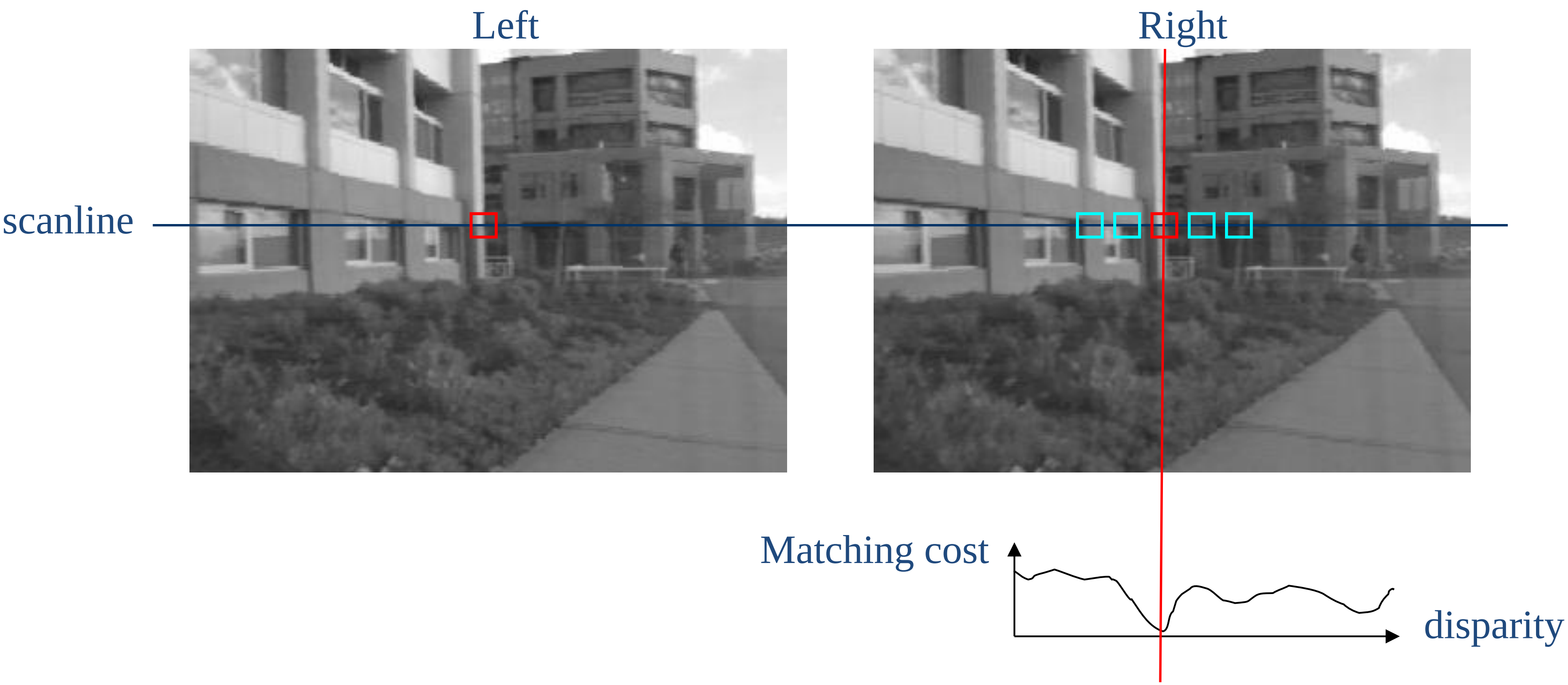
- If cameras are parallel and at the same height, the searching homologous tie points are positioned onto the horizontal epipolar lines with same coordinate. In practical applications only a *calibration* phase and *image registration* guarantee such properties.
- A cross check can be applied: if P is obtained from Q, Q must correspond be obtained from P

Basic stereo matching algorithm



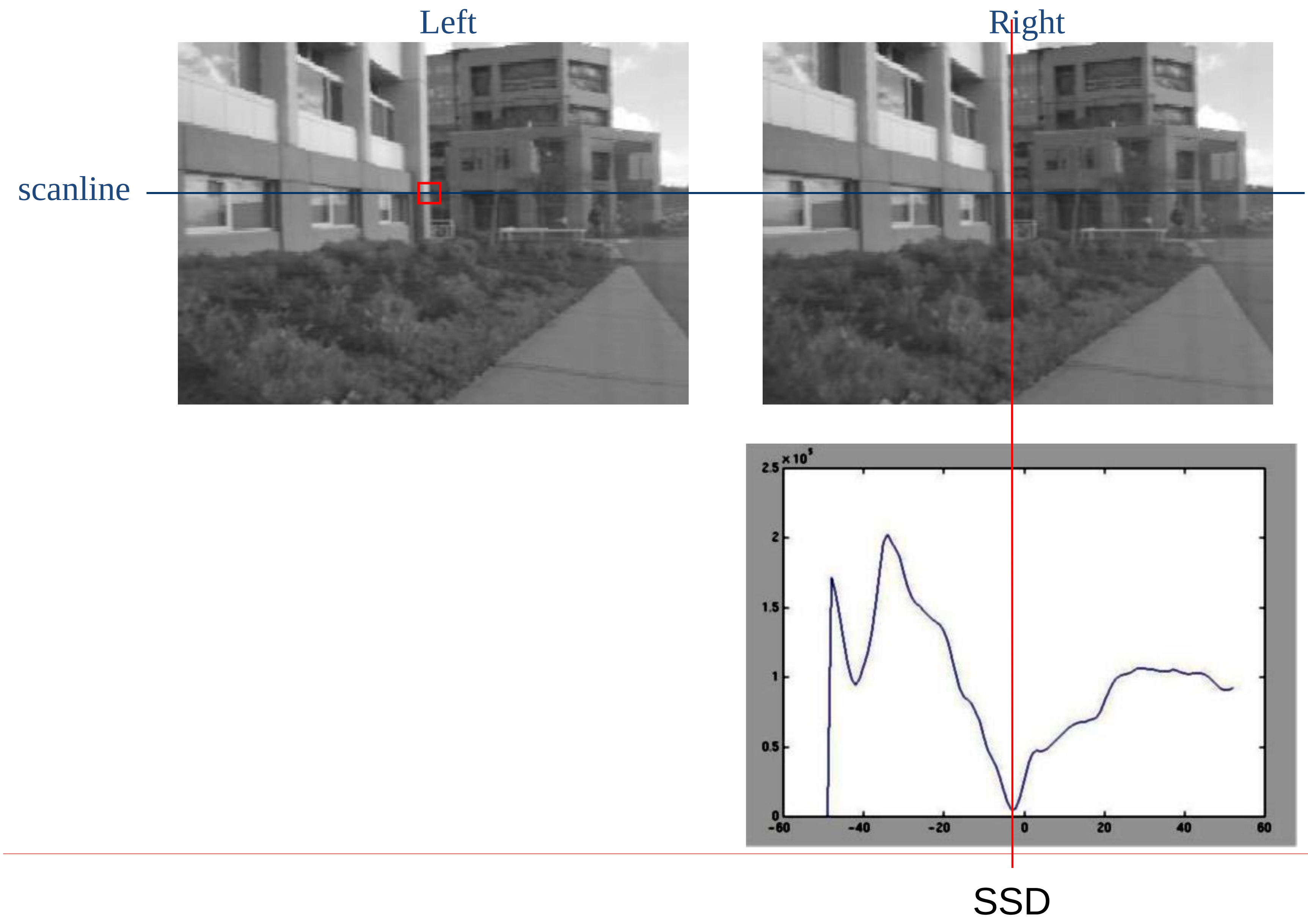
- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- For each pixel x in the first image
  - Find corresponding epipolar scanline in the right image
  - Examine all pixels on the scanline and pick the best match x'
  - Compute disparity x-x' and set depth(x) = fB/(x-x')

Correspondence search



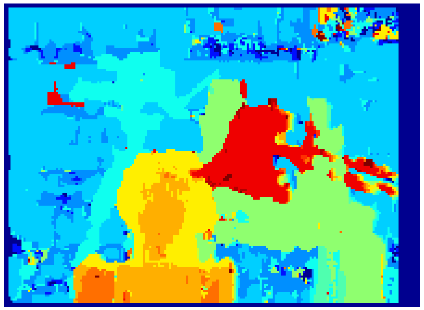
- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

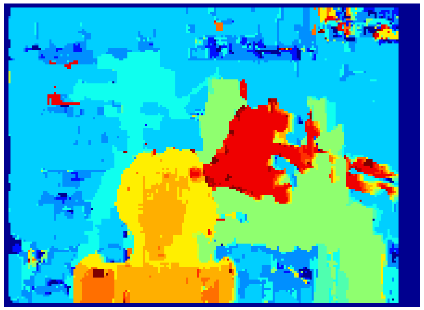
Correspondence search

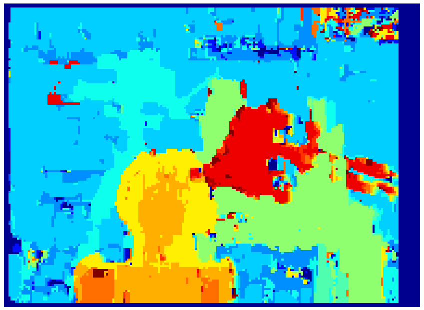


Matching windows

Similarity Measure	Formula
Sum of Absolute Differences (SAD)	$\sum_{(i,j)\in W} \lvert I_1(i,j) - I_2(x+i, y+j) \rvert$
Sum of Squared Differences (SSD)	$\sum_{(i,j)\in W} \left(I_1(i,j) - I_2(x+i, y+j)\right)^2$
Zero-mean SAD	$\sum_{(i,j)\in W} \lvert I_1(i,j) - \bar{I}_1(i,j) - I_2(x+i,y+j) + \bar{I}_2(x+i,y+j) \rvert$
Locally scaled SAD	$\sum_{(i,j)\in W} \lvert I_1(i,j) - \frac{\bar{I}_1(i,j)}{\bar{I}_2(x+i,y+j)} I_2(x+i,y+j) \rvert$
Normalized Cross Correlation (NCC)	$\frac{\sum_{(i,j)\in W} I_1(i,j).I_2(x+i,y+j)}{\sqrt[2]{\sum_{(i,j)\in W} I_1^2(i,j).\sum_{(i,j)\in W} I_2^2(x+i,y+j)}}$


SAD

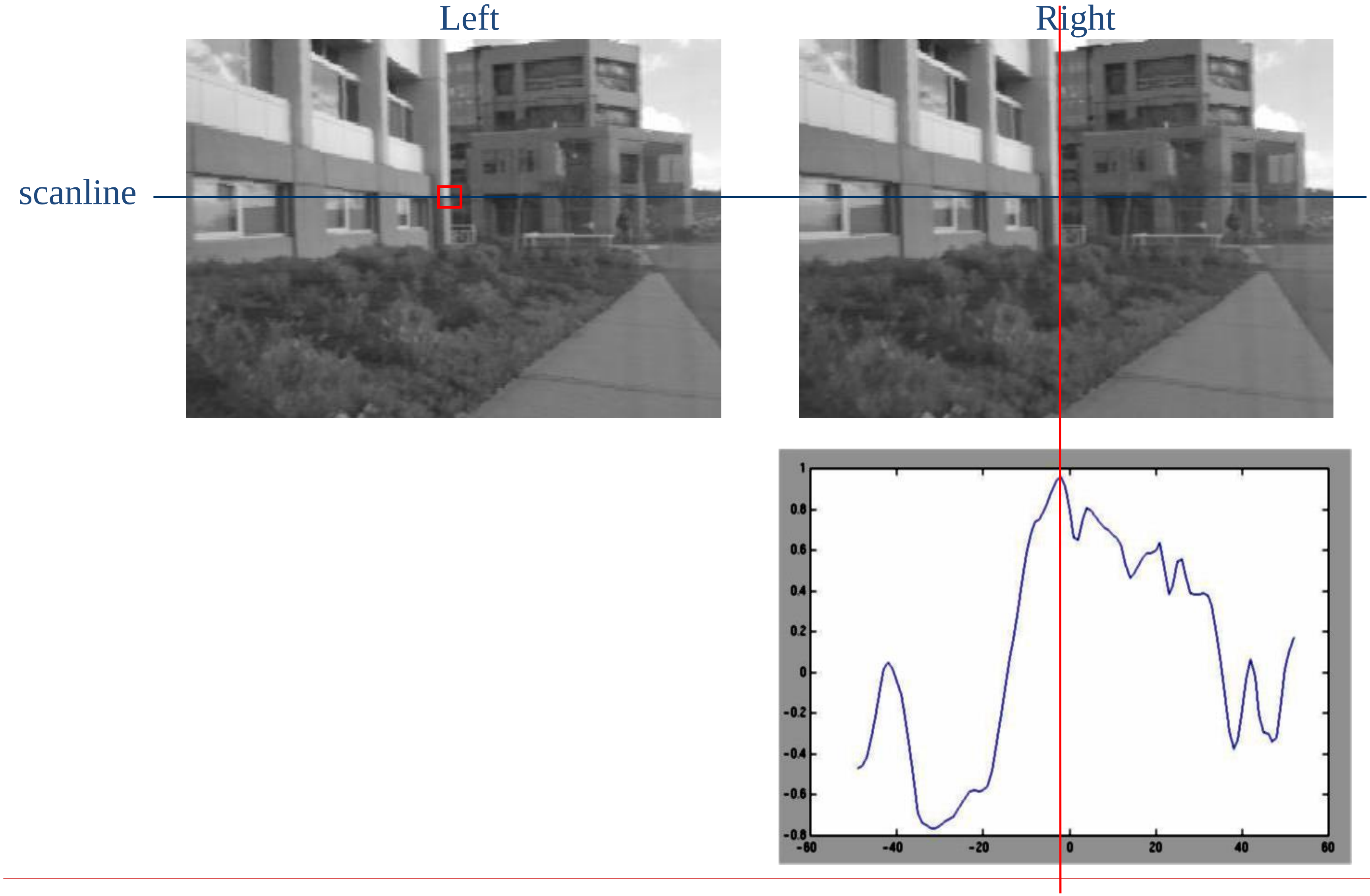

SSD


NCC


Ground truth

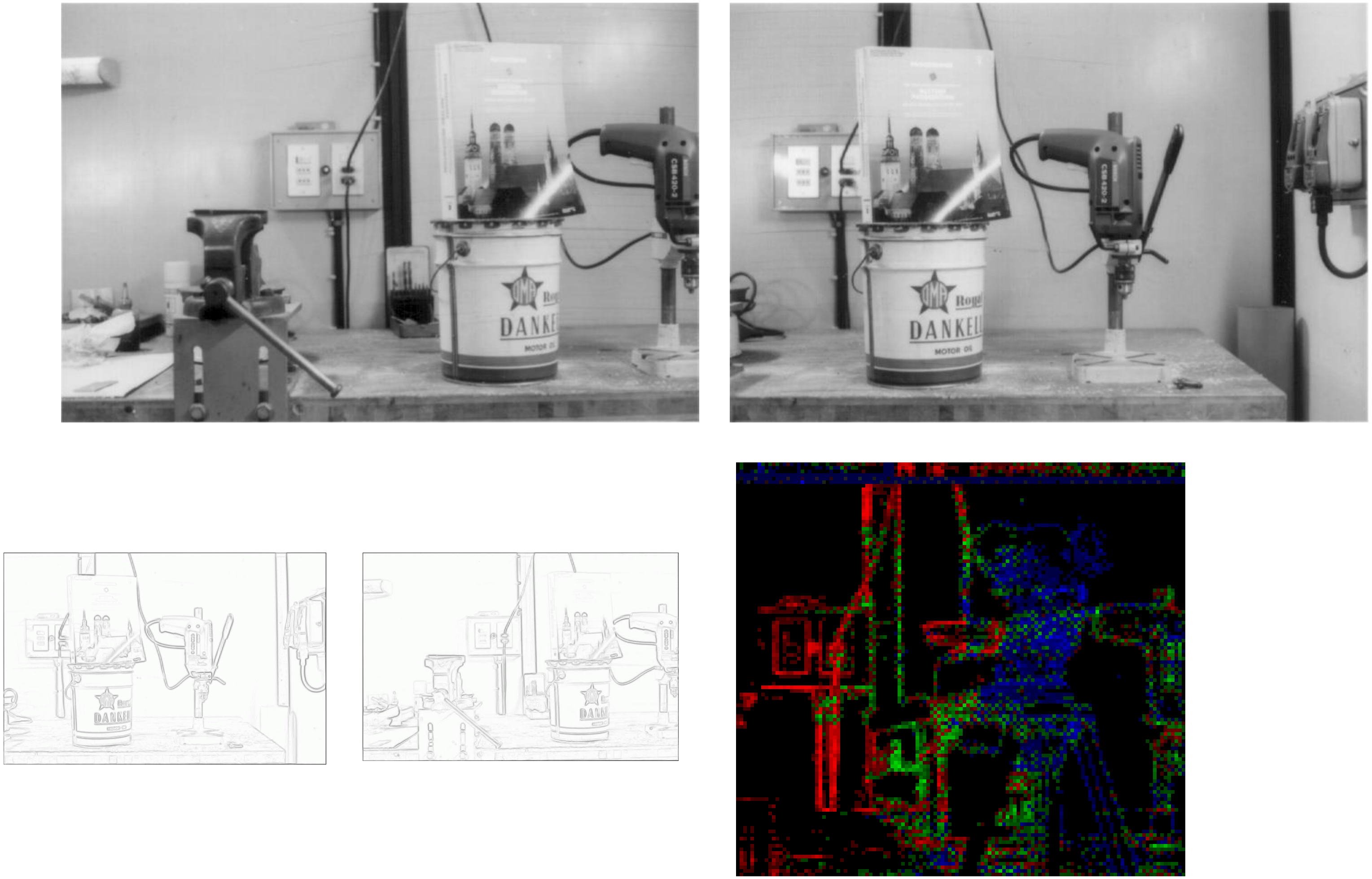
http://siddhantahuja.wordpress.com/category/stereo-vision/

Correspondence search

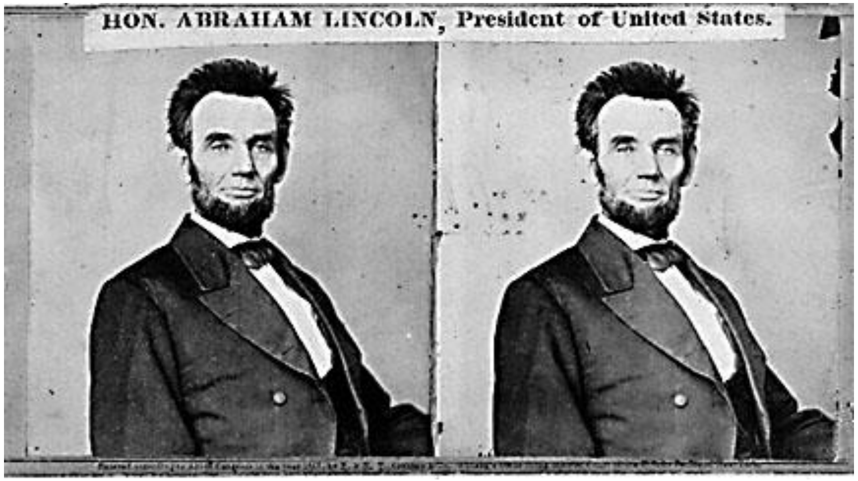


Norm. corr

Exemple



Failures of correspondence search



Textureless surfaces

Occlusions, repetition

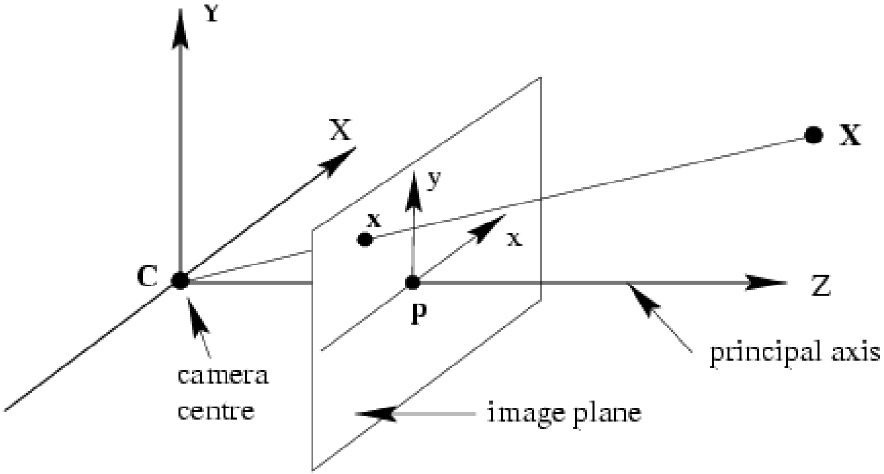
Non-Lambertian surfaces, specularities

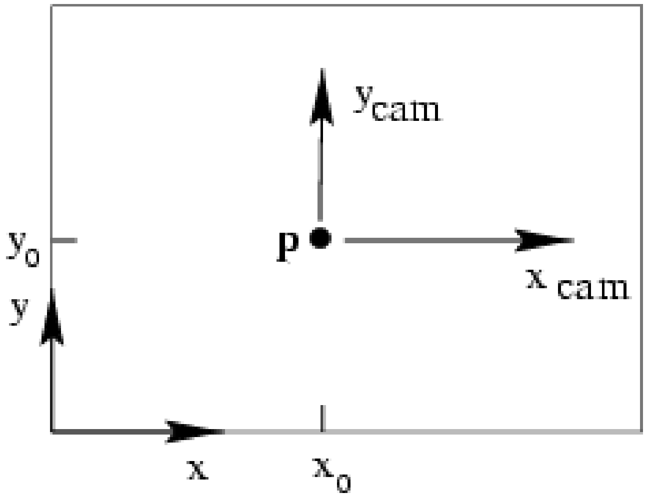
Implementation aspects

The search can be done in four steps:

- Selection of interesting points (through a threshold S_1 applied to the variance in the neighborhood or to the result of an *edge detector*)
- For each point selected, finding if exists the tie point (with a cross-check and a threshold S_2 of cross-similarity)
- Evaluation of the distance on the basis of the extracted homologous points
- Experimentation of the best solution, considering that:
  - augmenting S_1 the number of tie points is reduced but the reliability increases
  - augmenting S_2 increases the number of homologous couples but it is reduced the reliability

Principal point

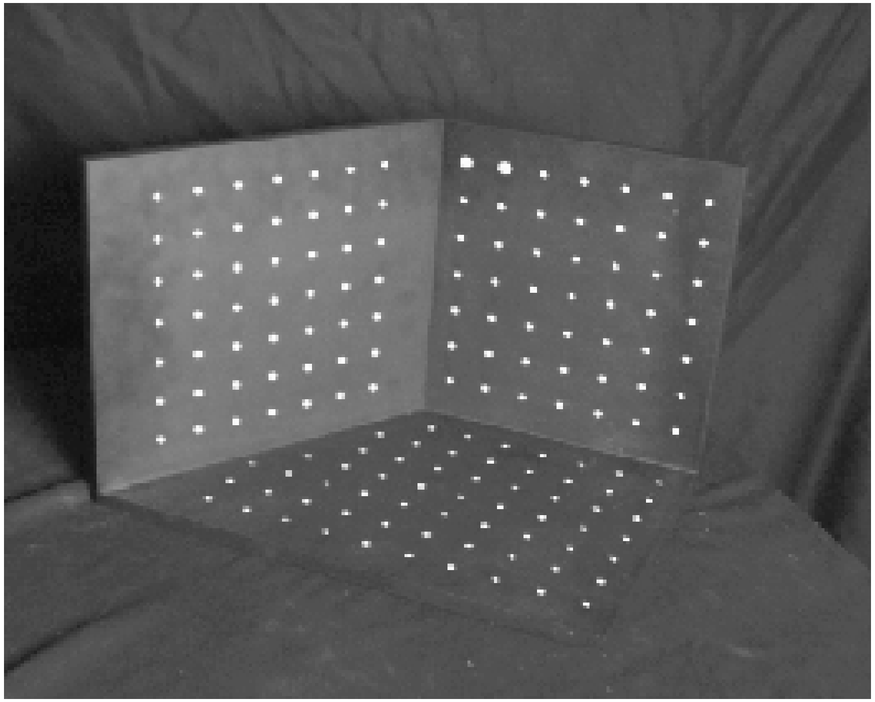


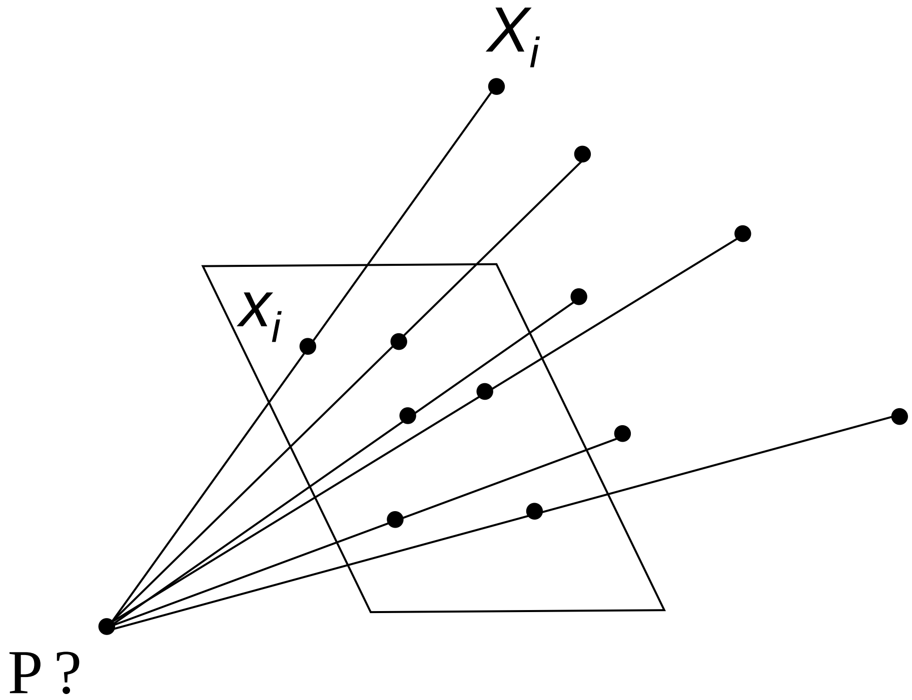


- **Principal point (p):** point where principal axis intersects the image plane (origin of normalized coordinate system)
- Normalized coordinate system: origin is at the principal point
- Image coordinate system: origin is in the corner
- How to go from normalized coordinate system to image coordinate system?

Camera calibration

- Given n points with known 3D coordinates X_i and known image projections x_i, estimate the camera parameters

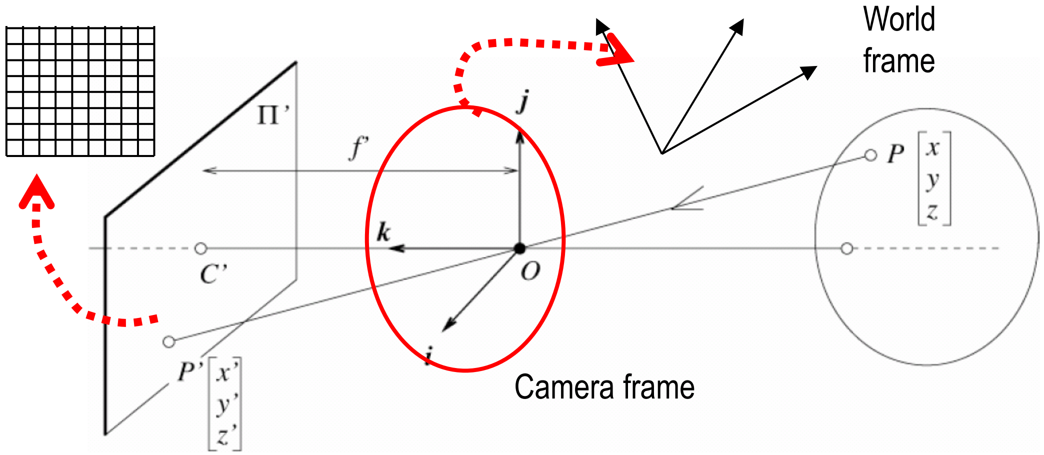




Camera parameters

- Intrinsic parameters
  - Principal point coordinates
  - Focal length
  - Pixel magnification factors
  - *Skew (non-rectangular pixels)*
  - *Radial distortion*
- Extrinsic parameters
  - Rotation and translation relative to world coordinate system

Camera calibration

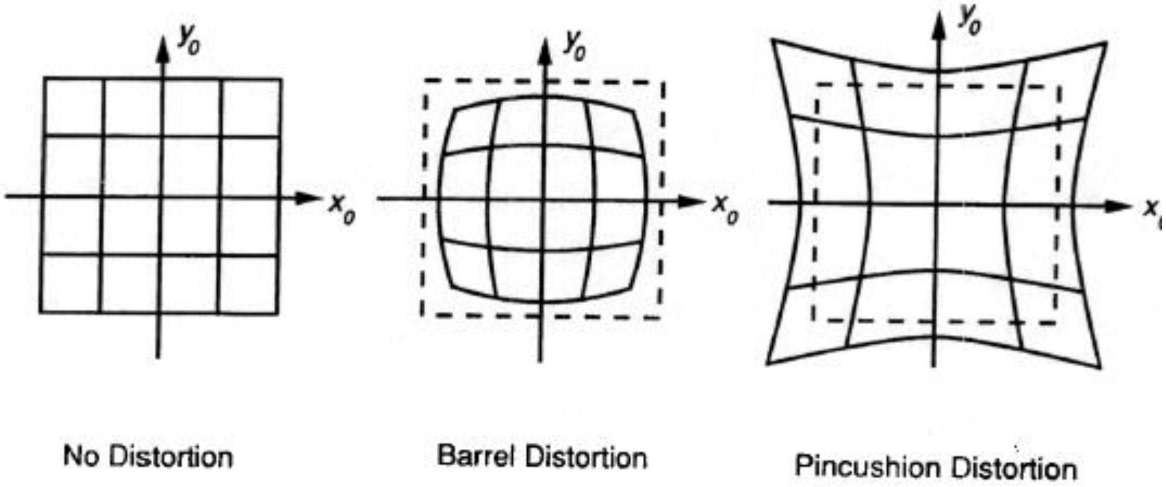


Extrinsic parameters:
Camera frame ←→ Reference frame

Intrinsic parameters:
Image coordinates relative to camera ←→ Pixel coordinates

- *Extrinsic* parameters: rotation matrix and translation vector
- *Intrinsic* parameters: focal length, pixel sizes (mm), image center point, radial distortion parameters

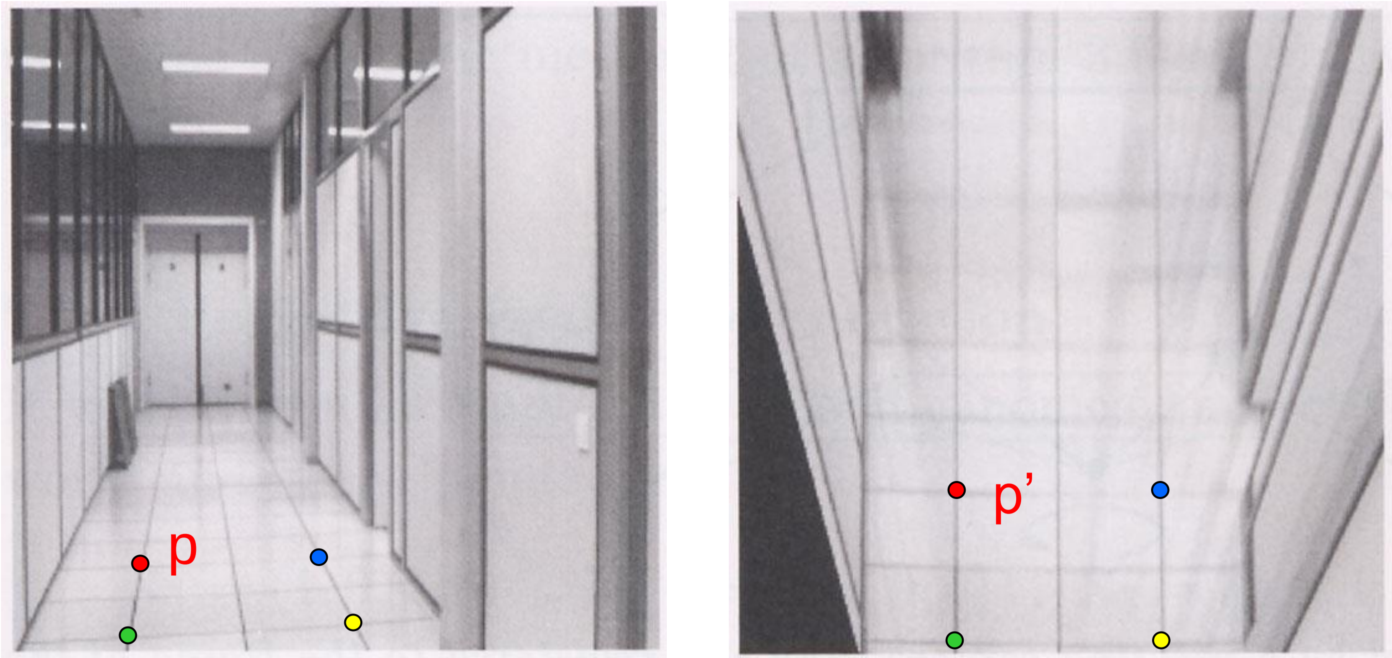
Beyond Pinholes: Radial Distortion



Corrected Barrel Distortion

Image from Martin Habbecke

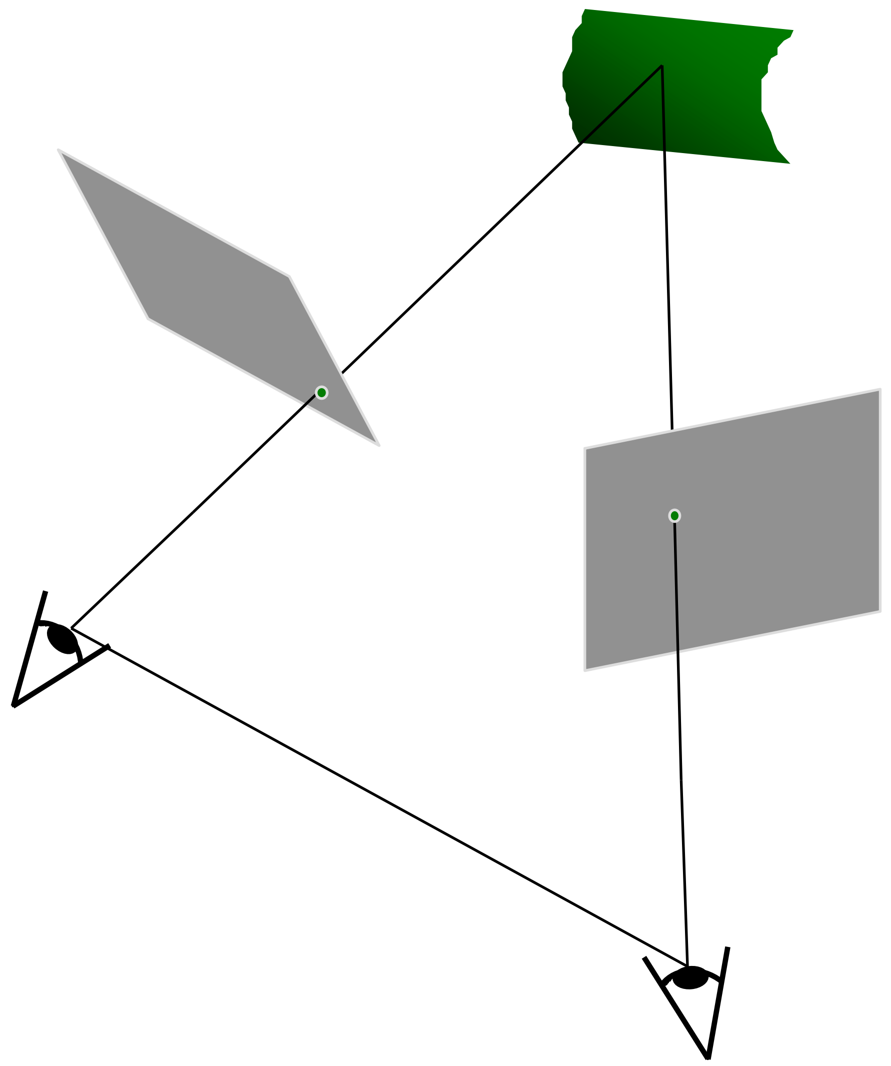
Image rectification



To unwarp (rectify) an image

- solve for homography H given p and p’
- solve equations of the form: wp’ = Hp
    - linear in unknowns: w and coefficients of H
    - H is defined up to an arbitrary scale factor
    - how many points are necessary to solve for H?

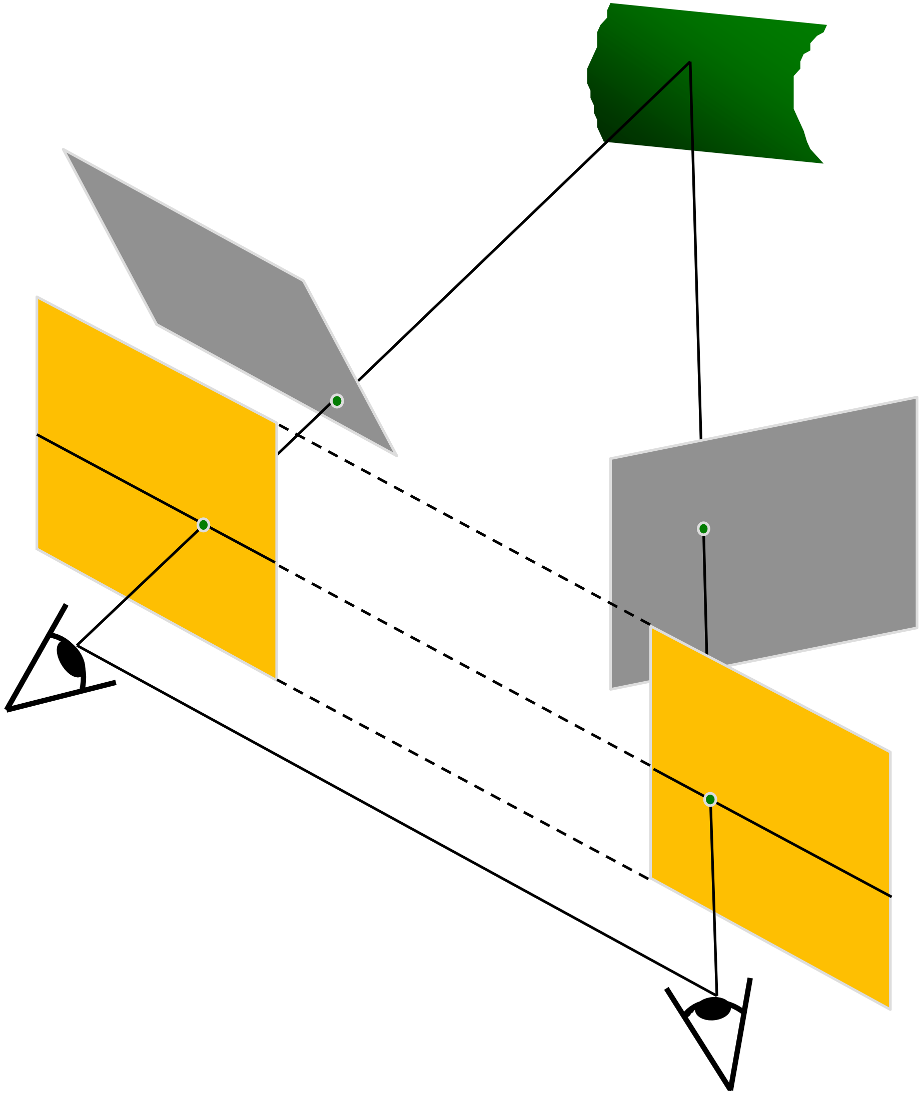
Stereo image rectification



Stereo image rectification

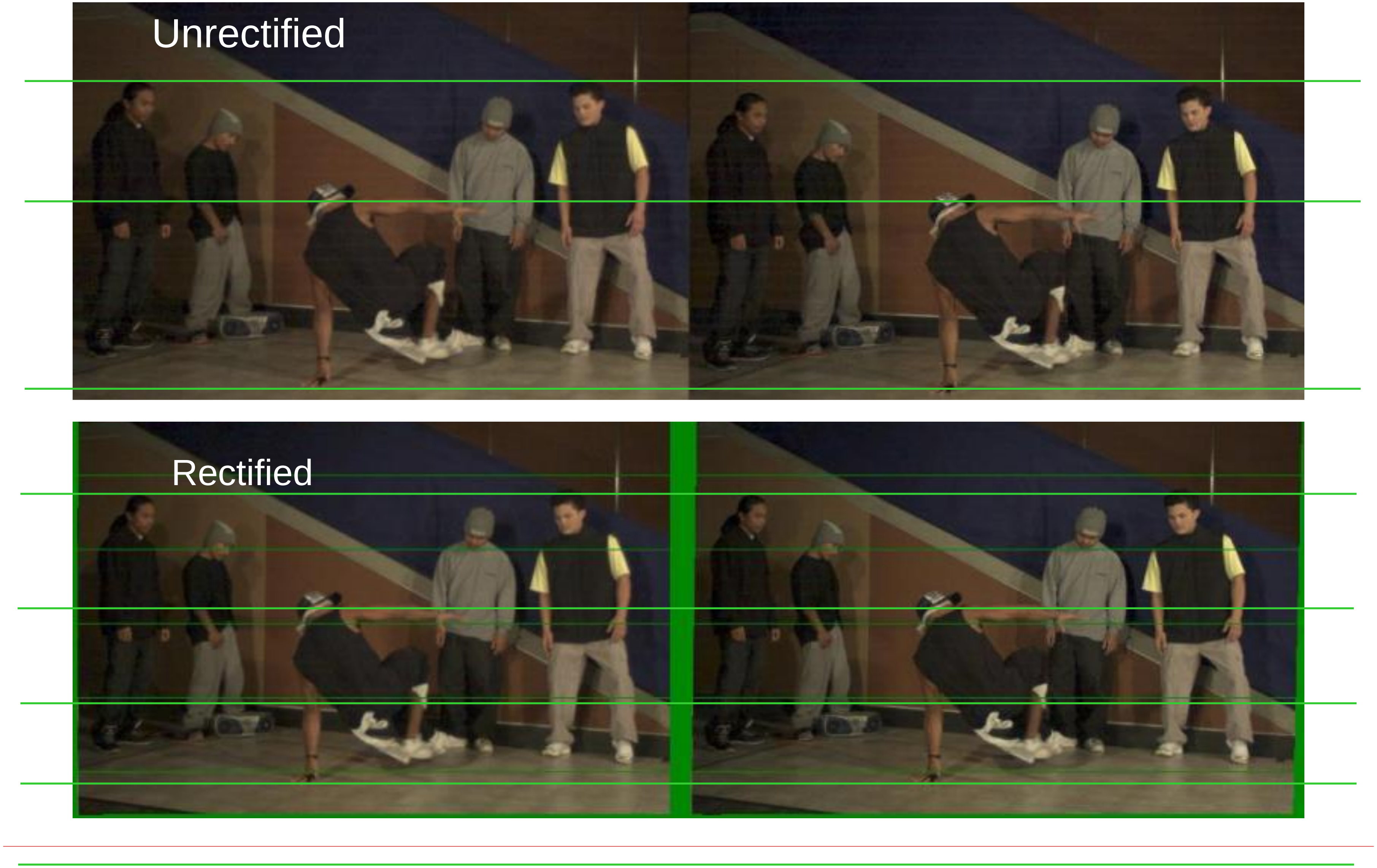
- Reproject image planes onto a common plane parallel to the line between camera centers
- Pixel motion is horizontal after this transformation
- Two homographies (3x3 transform), one for each input image reprojection

➢ C. Loop and Z. Zhang. Computing Rectifying Homographies for Stereo Vision. IEEE Conf. Computer Vision and Pattern Recognition, 1999.

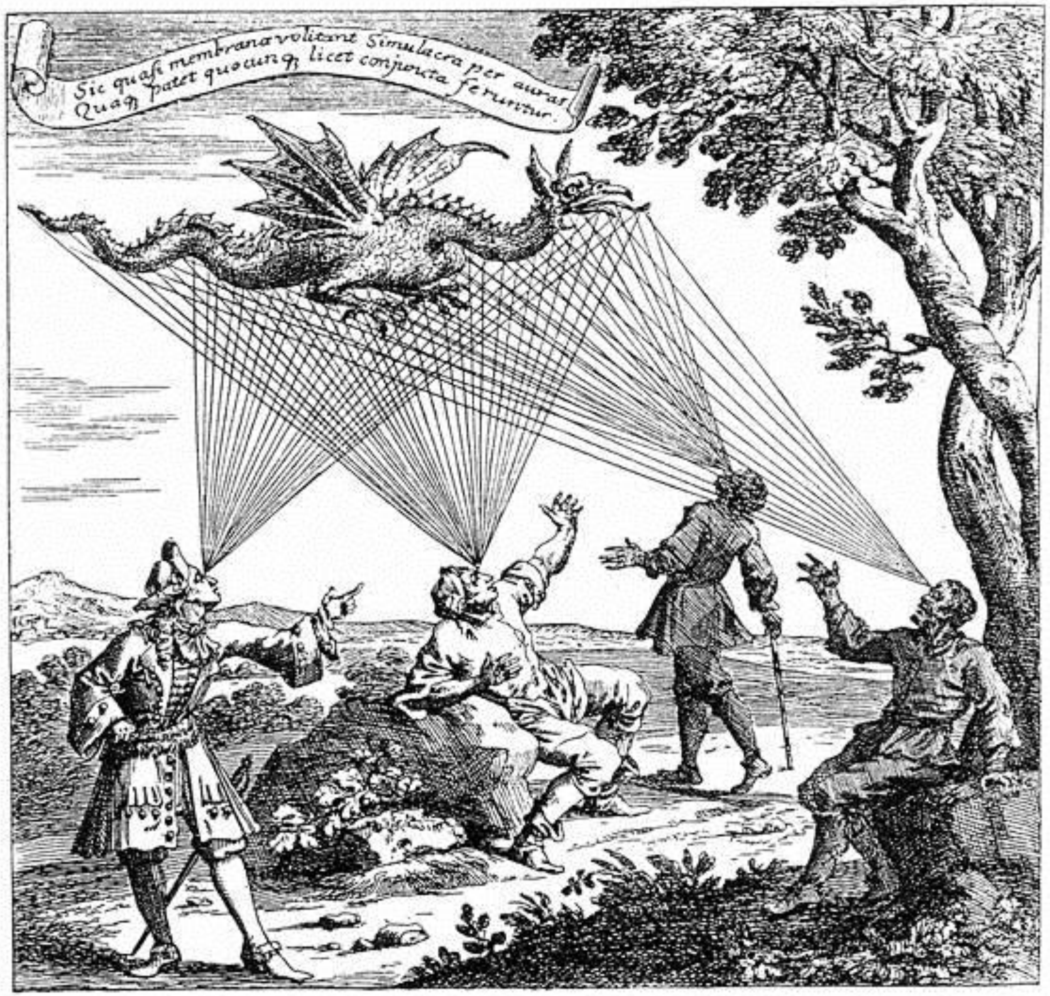


Rectification example

Example



Multi-view Stereo

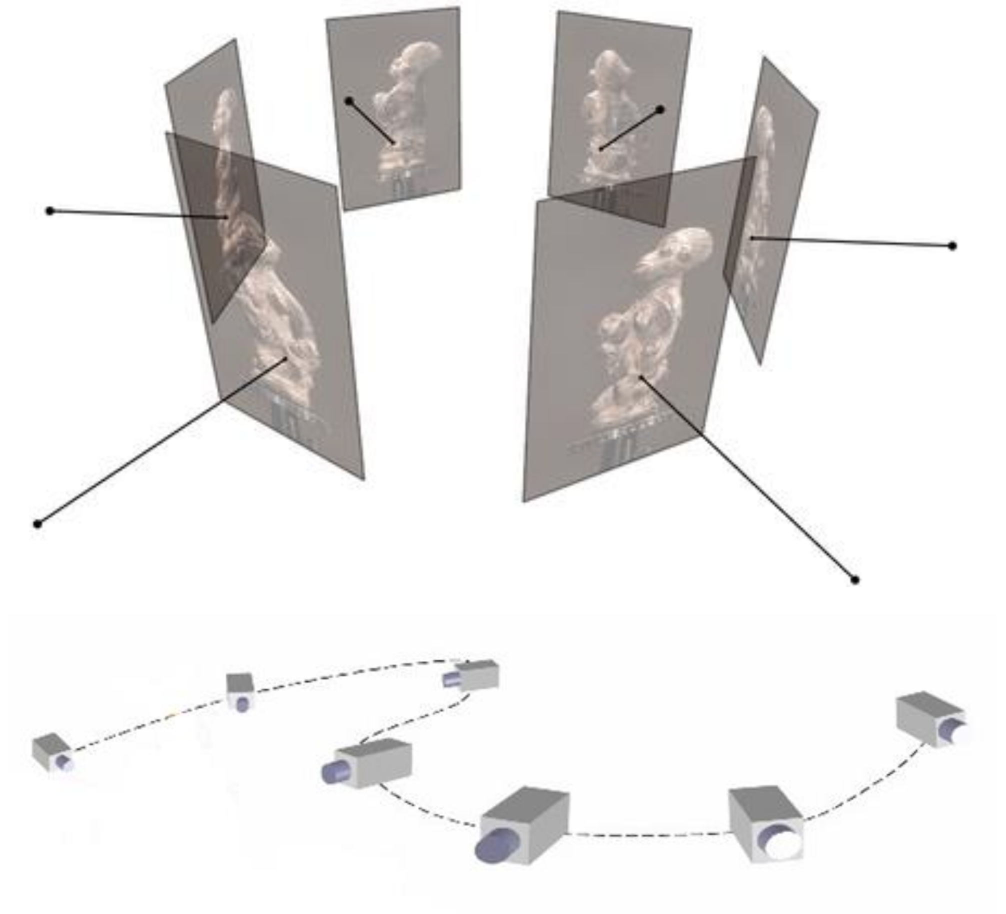


Драконъ, видимый подъ различными углами зрѣнія
По гравюрѣ на мѣди изъ „Oculus artificialis teledioptricus" Цана. 1702 года.

Multi-view Stereo

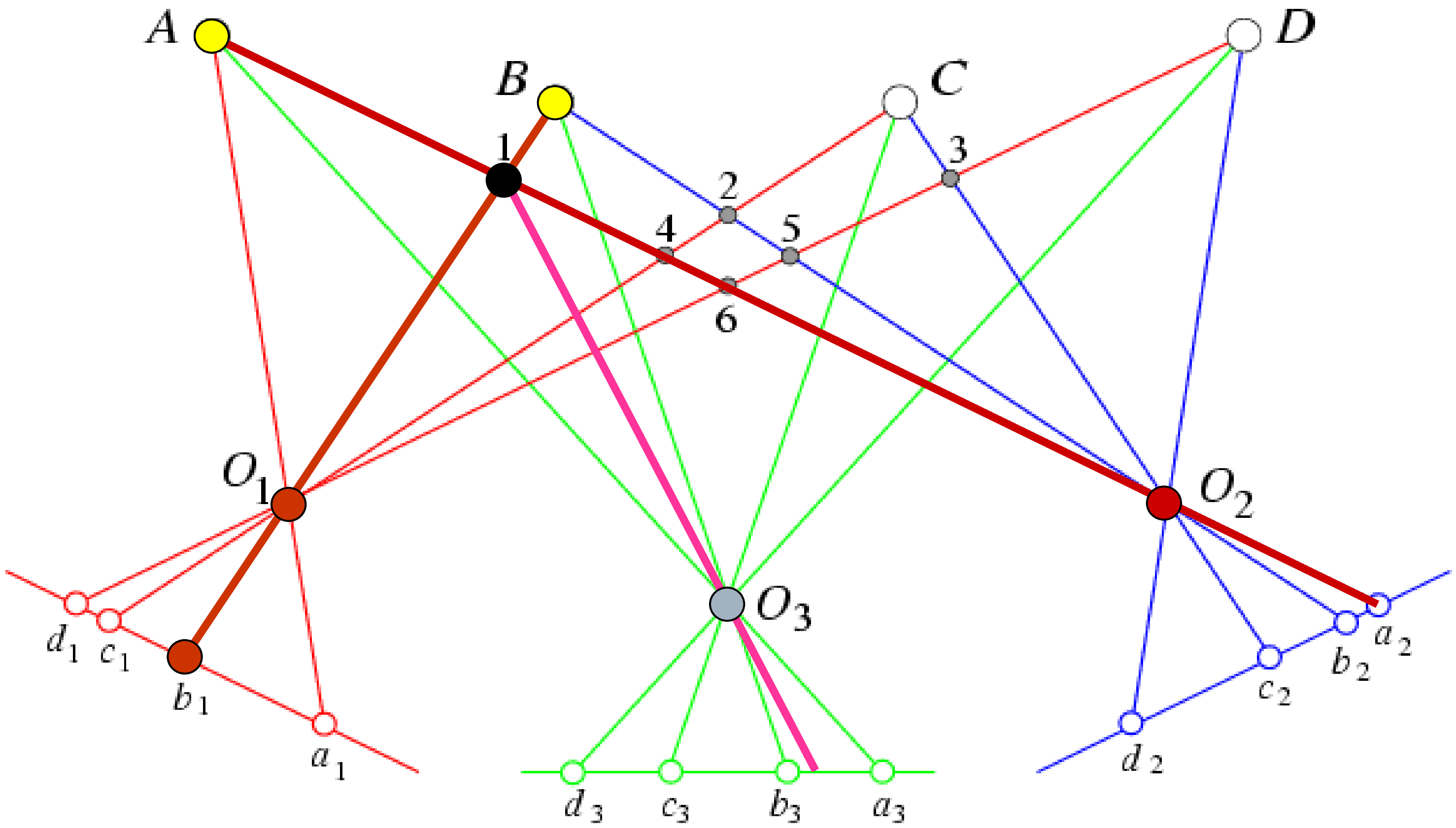
Input: calibrated images from several viewpoints

Output: 3D object model



[Seitz]

Beyond two-view stereo



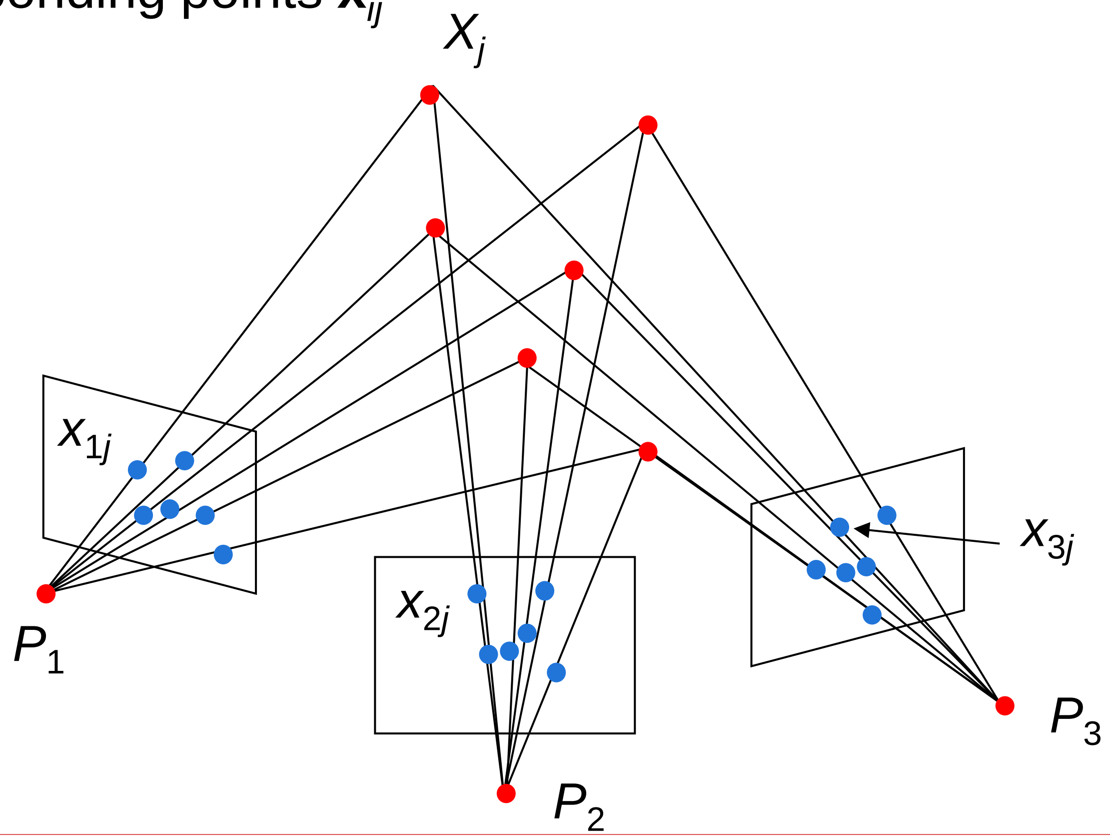
The third view can be used for verification

Projective structure from motion

- Given: *m* images of *n* fixed 3D points

$$\mathbf{x}_{ij} = \mathbf{P}_i \mathbf{X}_j , \qquad i = 1,\dots, m, \quad j = 1, \dots, n$$

- Problem: estimate *m* projection matrices $\mathbf{P}_i$ and *n* 3D points $\mathbf{X}_j$ from the *mn* corresponding points $\mathbf{x}_{ij}$

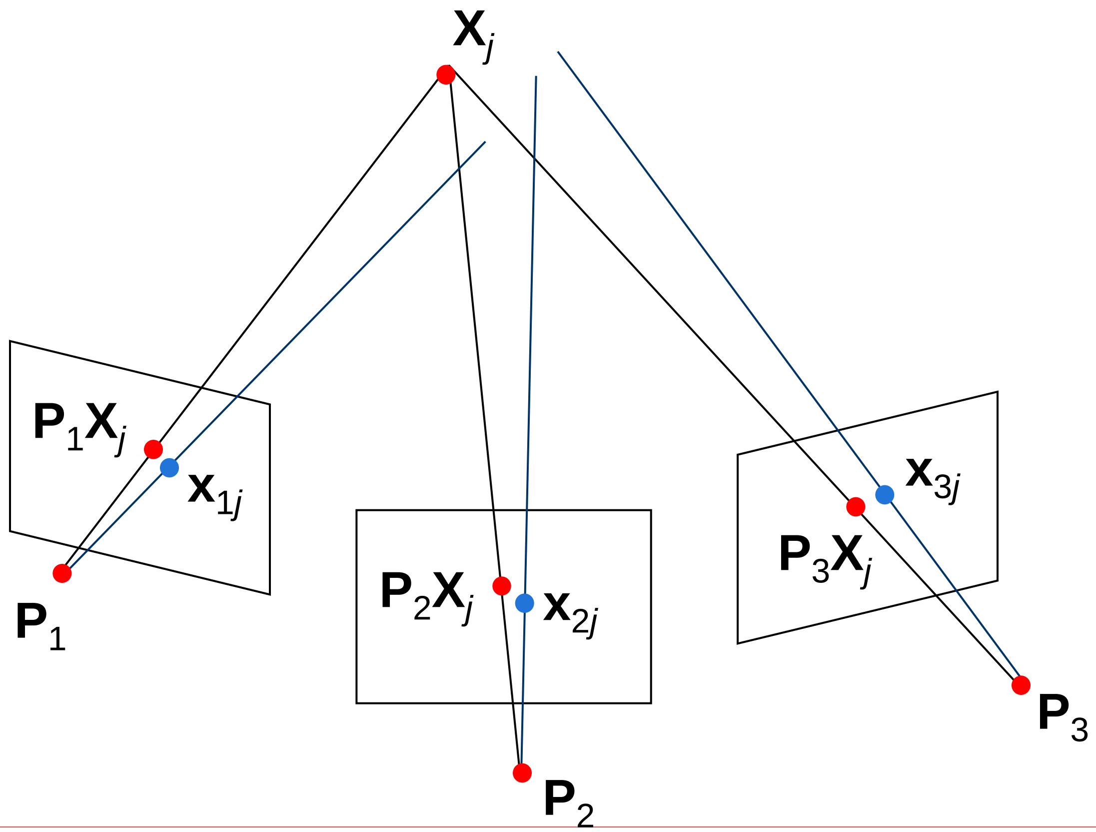


Slides from Lana Lazebnik

Bundle adjustment

- Non-linear method for refining structure and motion
- Minimizing reprojection error

$$E(\mathbf{P},\mathbf{X}) = \sum_{i=1}^{m}\sum_{j=1}^{n} D\left(\mathbf{x}_{ij}, \mathbf{P}_i\mathbf{X}_j\right)^2$$



Multiple-baseline stereo

- Pick a reference image, and slide the corresponding window along the corresponding epipolar lines of all other images, using inverse depth relative to the first image as the search parameter

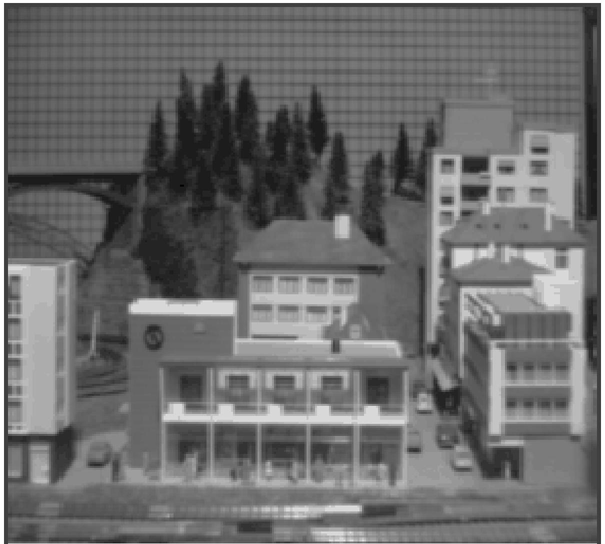
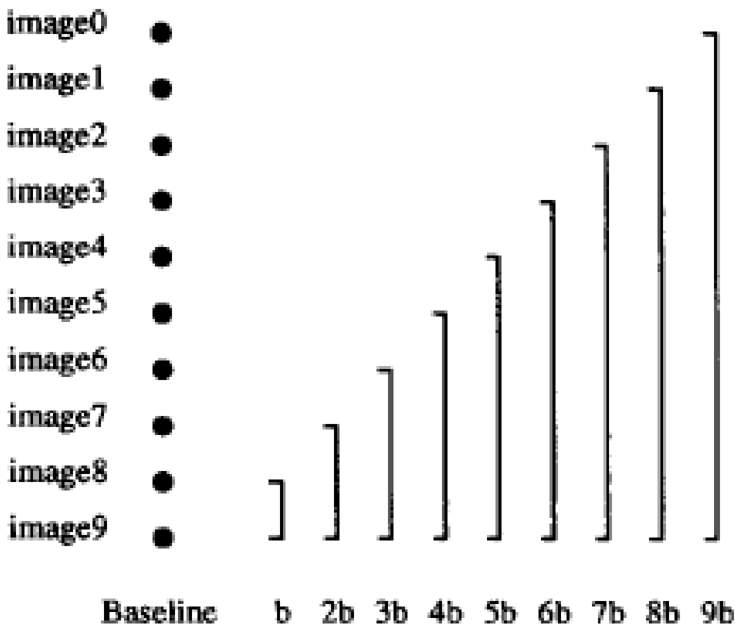


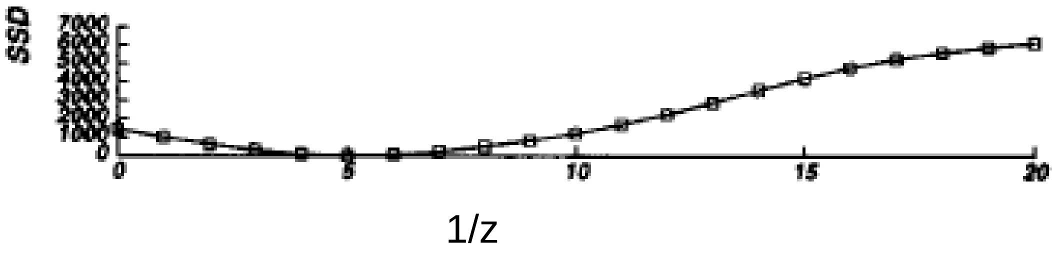
Figure 2: An example scene. The grid pattern in the background has ambiguity of matching.

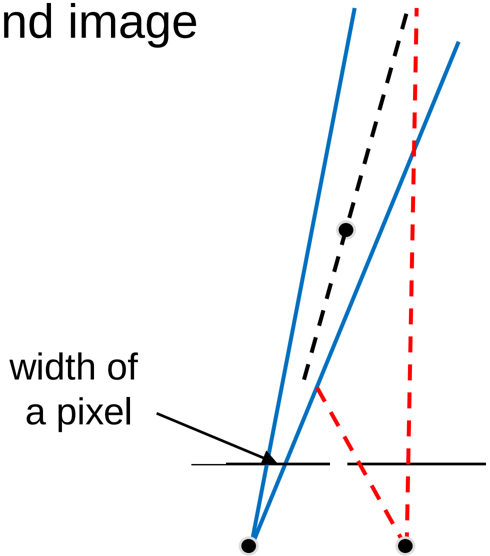


M. Okutomi and T. Kanade, [“A Multiple-Baseline Stereo System,”] IEEE Trans. on Pattern Analysis and Machine Intelligence, 15(4):353-363 (1993).

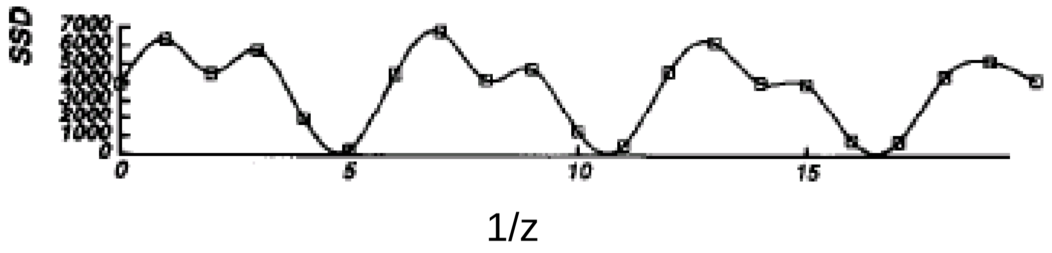
Multiple-baseline stereo

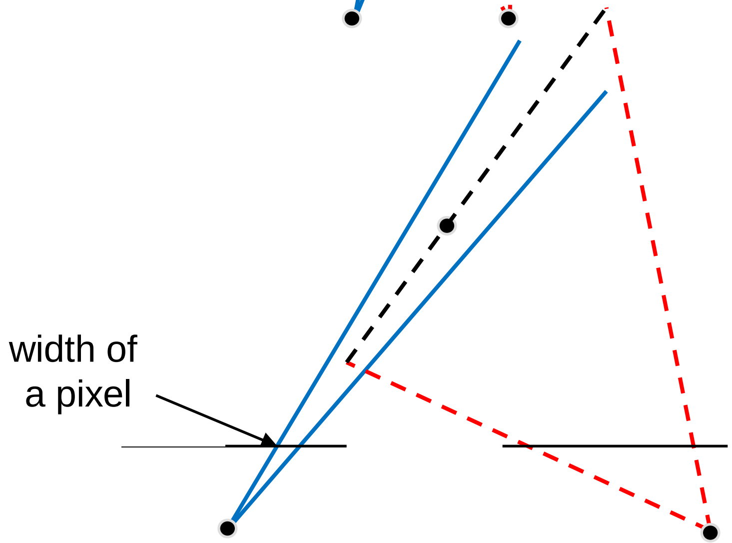
- For larger baselines, must search larger area in second image





pixel matching score





Multiple-baseline stereo

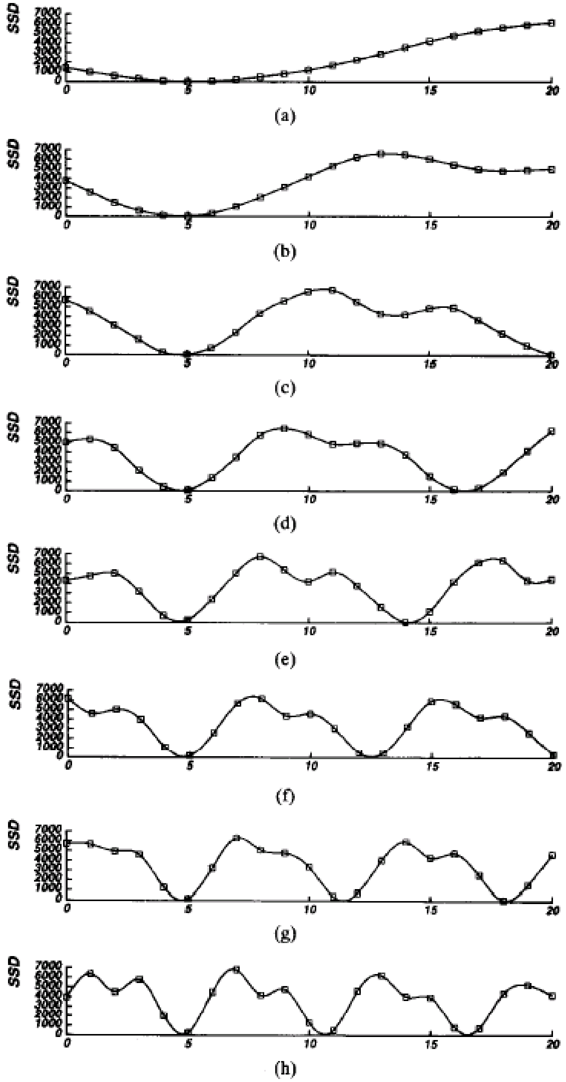


Fig. 5. SSD values versus inverse distance: (a) $B = b$; (b) $B = 2b$; (c) $B = 3b$; (d) $B = 4b$; (e) $B = 5b$; (f) $B = 6b$; (g) $B = 7b$; (h) $B = 8b$. The horizontal axis is normalized such that $8bF = 1$.

Use the sum of SSD scores to rank matches

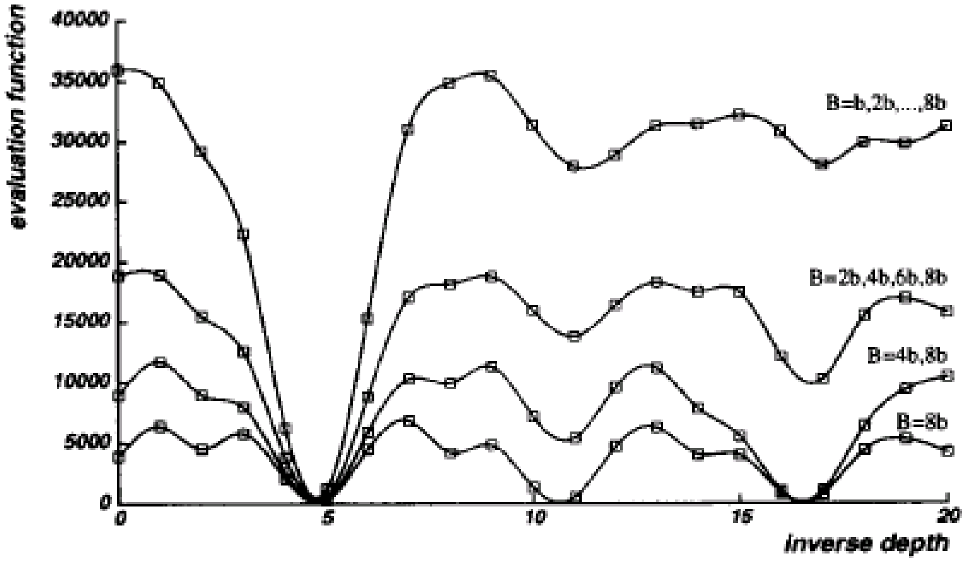
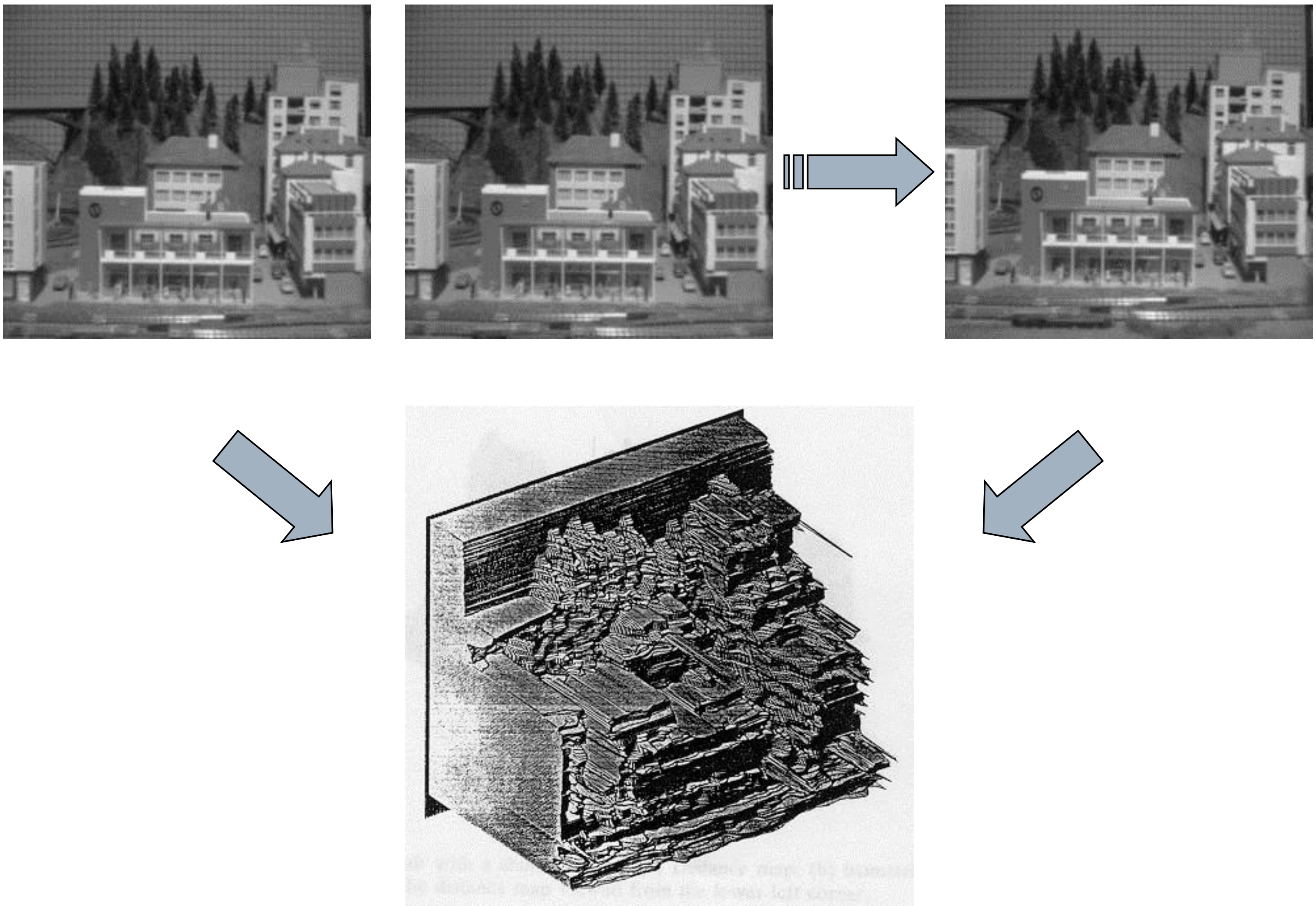


Fig. 7. Combining multiple baseline stereo pairs.

Multiple-baseline stereo results



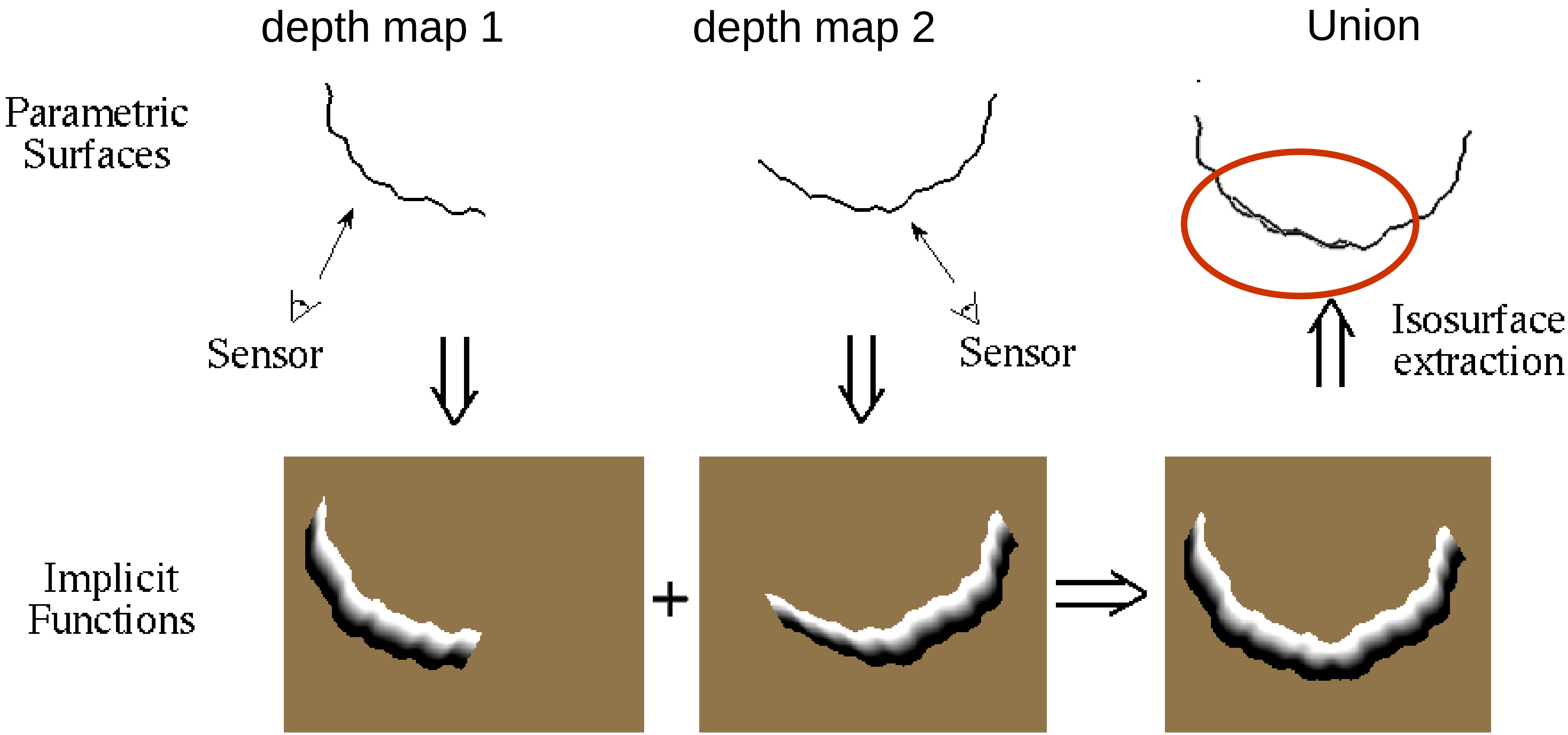
M. Okutomi and T. Kanade, “A Multiple-Baseline Stereo System,” IEEE Trans. on Pattern Analysis and Machine Intelligence, 15(4):353-363 (1993).

Merging depth maps

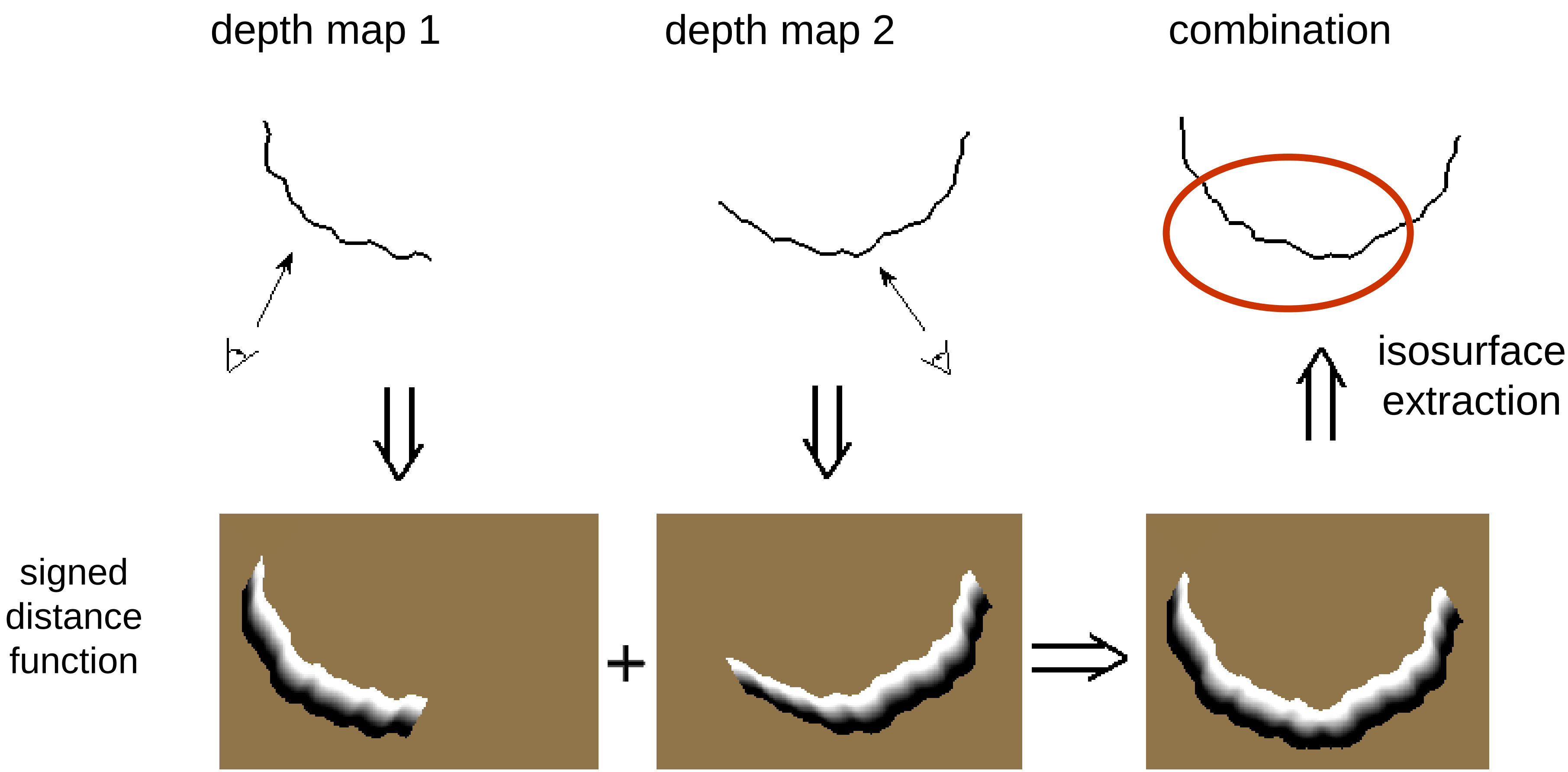
Naïve combination (union) produces artifacts

Better solution: find "average" surface

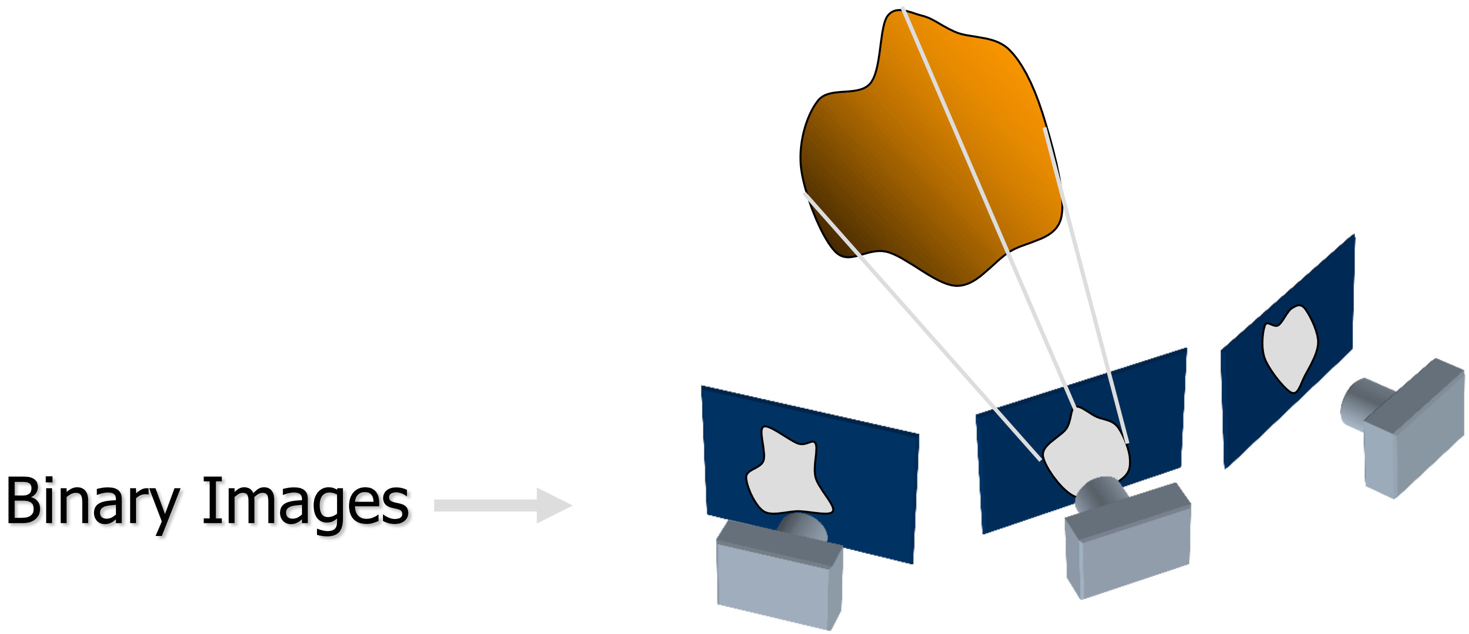
- Surface that minimizes sum (of squared) distances to the depth maps



VRIP [Curless & Levoy 1996]



Reconstruction from Silhouettes (C = 2)

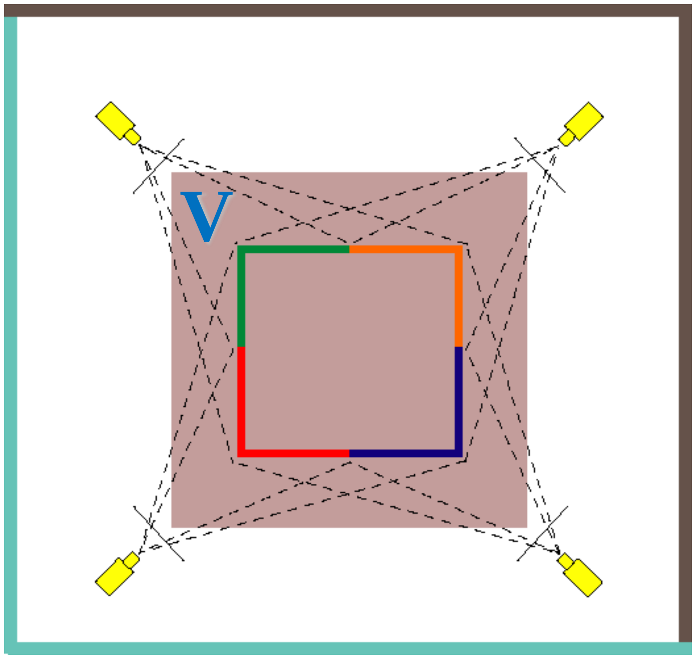


Approach:

- *Backproject* each silhouette
- Intersect backprojected volumes

Which shape do you get?

- The Photo Hull *is the UNION of all photo-consistent scenes in V*
  - It is a photo-consistent scene reconstruction
  - Tightest possible bound on the true scene



True Scene

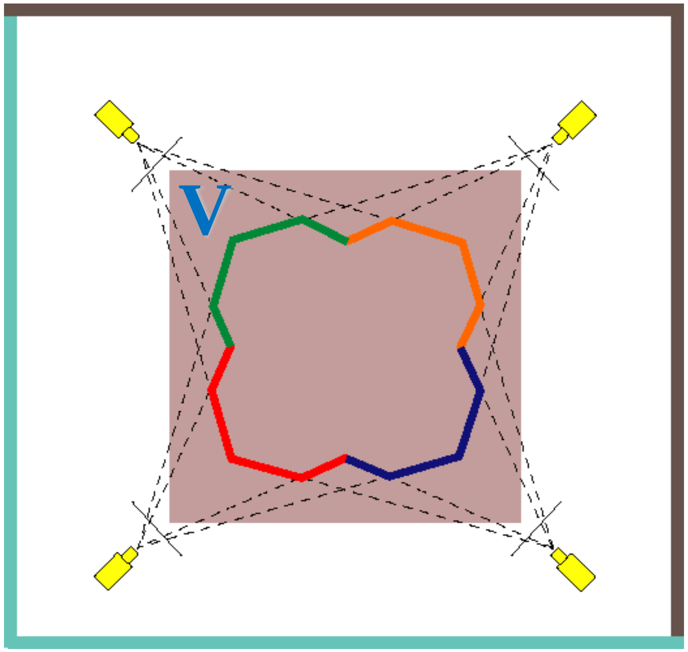
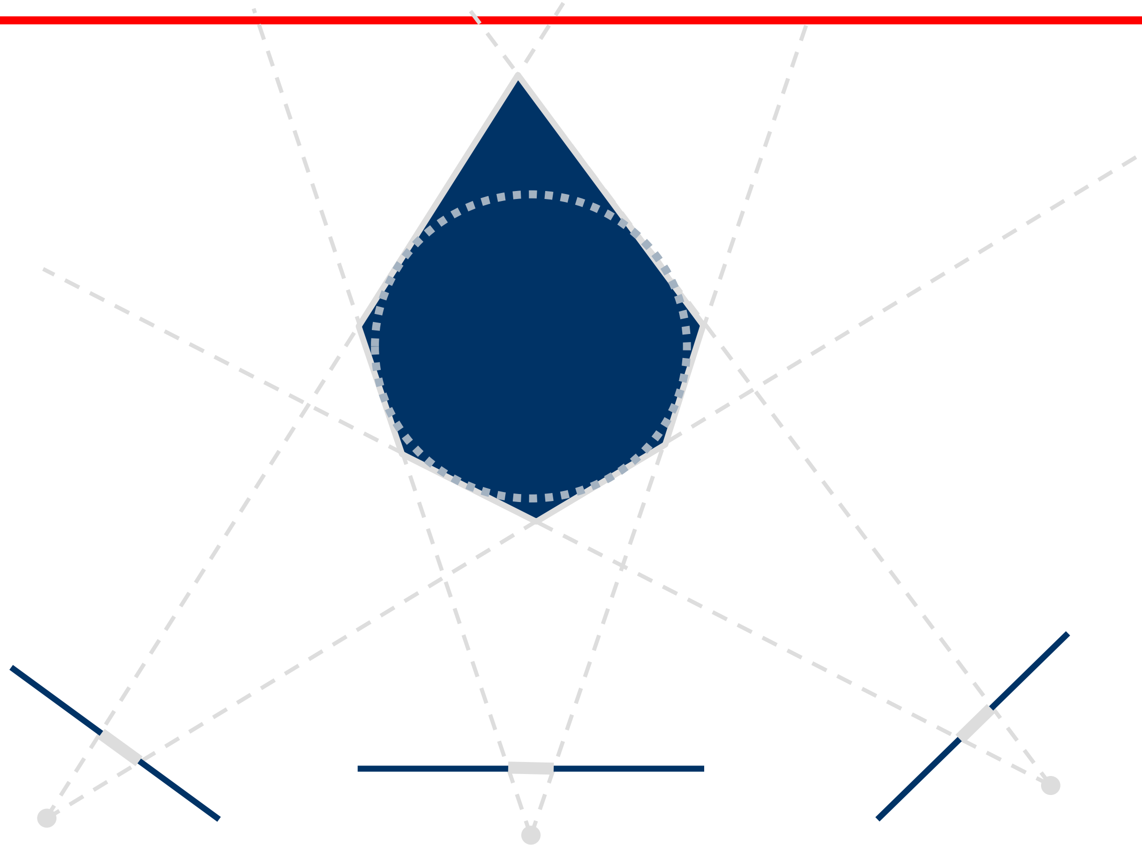


Photo Hull

Source: S. Seitz

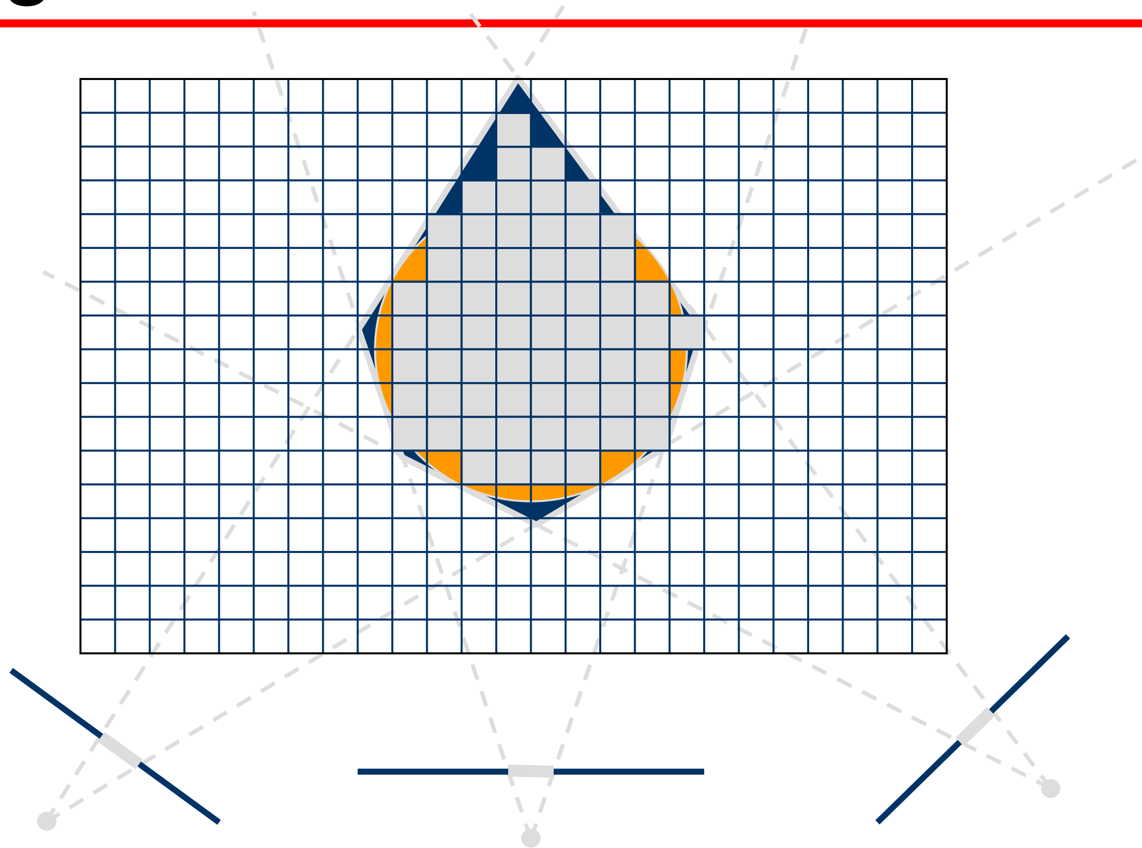
Volume intersection



Reconstruction Contains the True Scene

- But is generally not the same
- In the limit (all views) get *visual hull*
  - ✓ Complement of all lines that don't intersect S

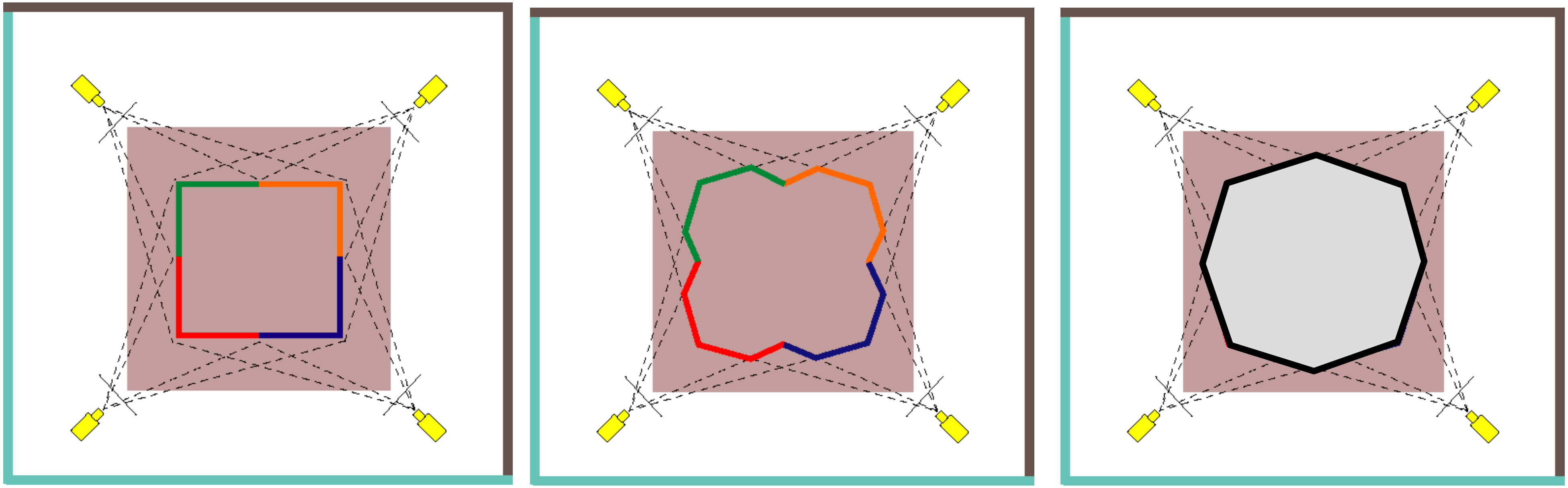
Voxel algorithm for volume intersection



Color voxel black if on silhouette in every image

- O(?), for M images, N^3 voxels O(MN^3)
- Don't have to search 2^{N^3} possible scenes!

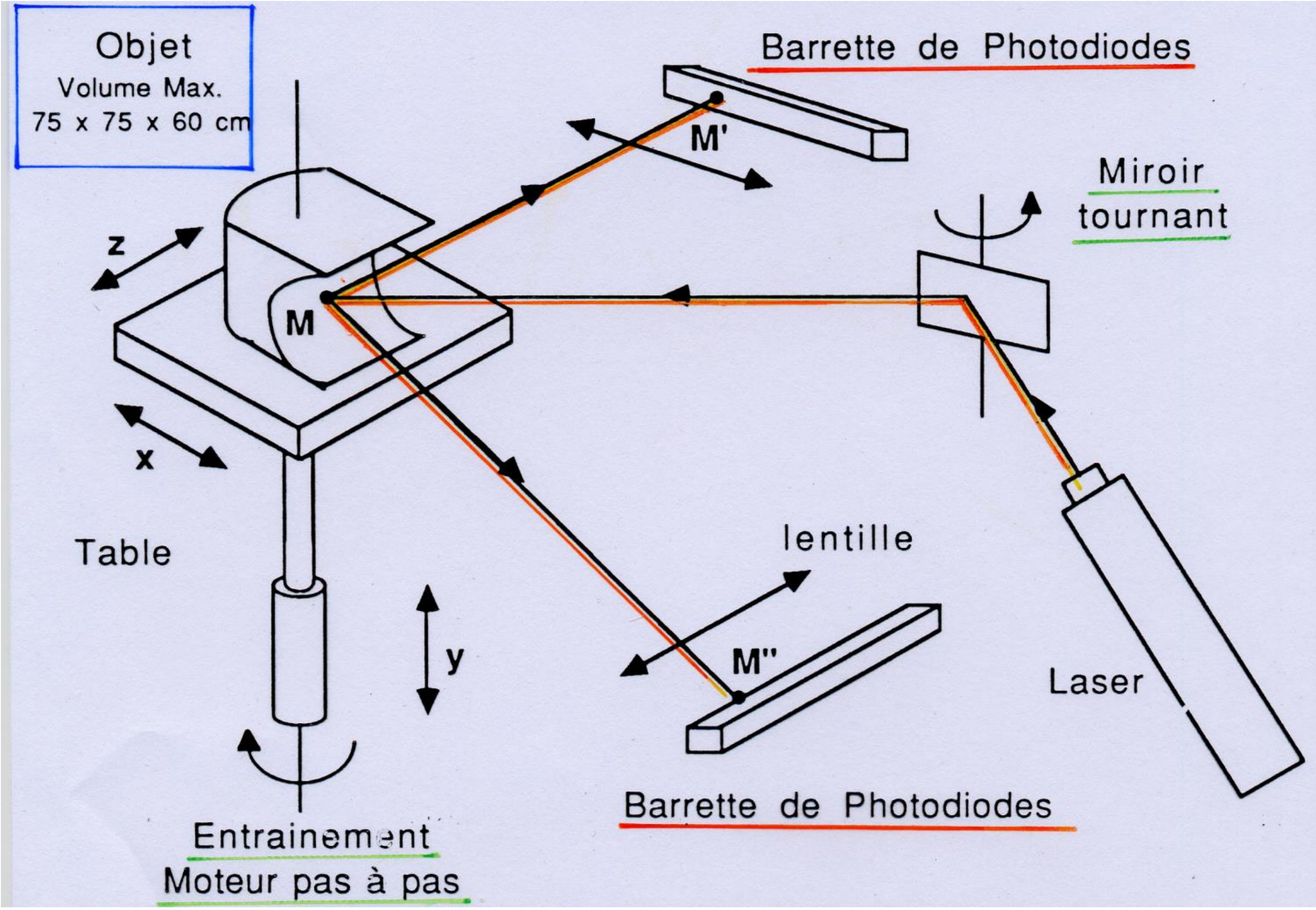
Photo-consistency vs. silhouette-consistency



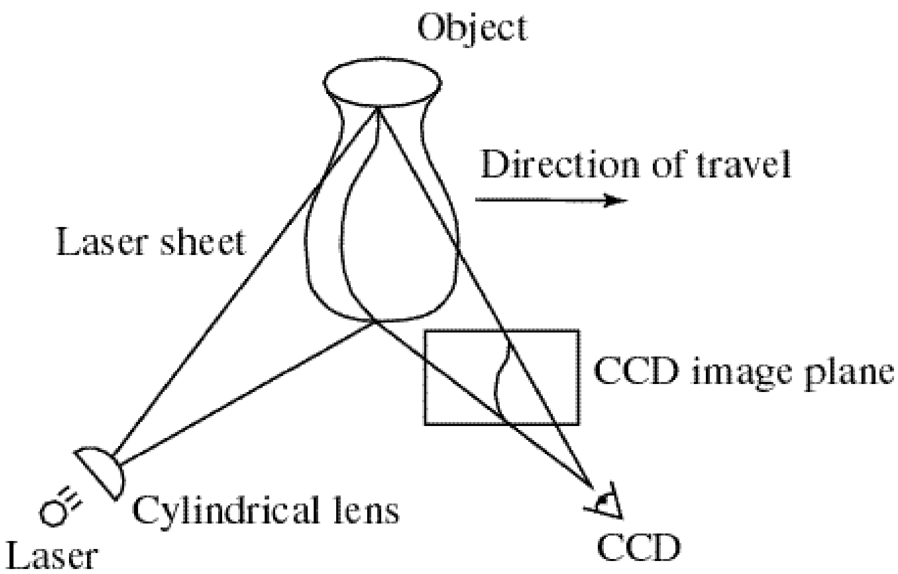
True Scene | Photo Hull | Visual Hull

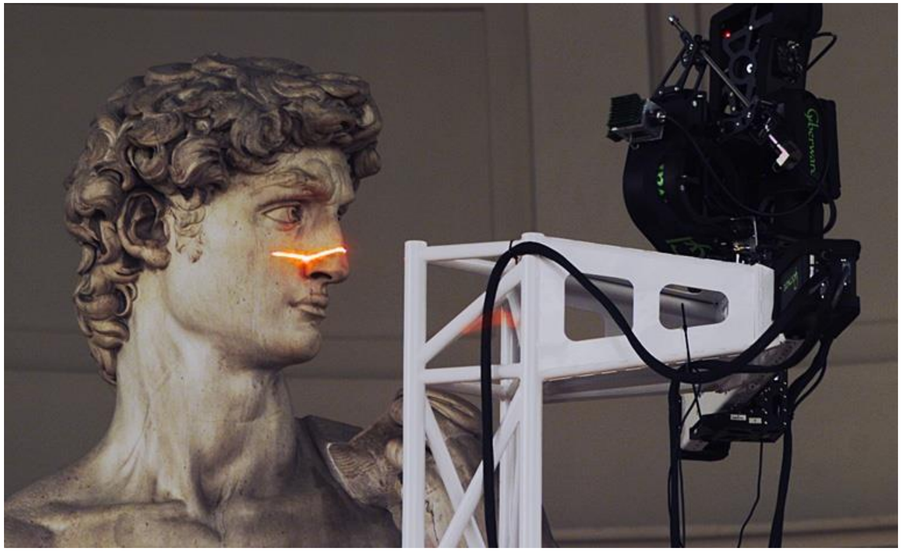
Structured light: point

- Point
- Plane
- Grid



Laser scanning





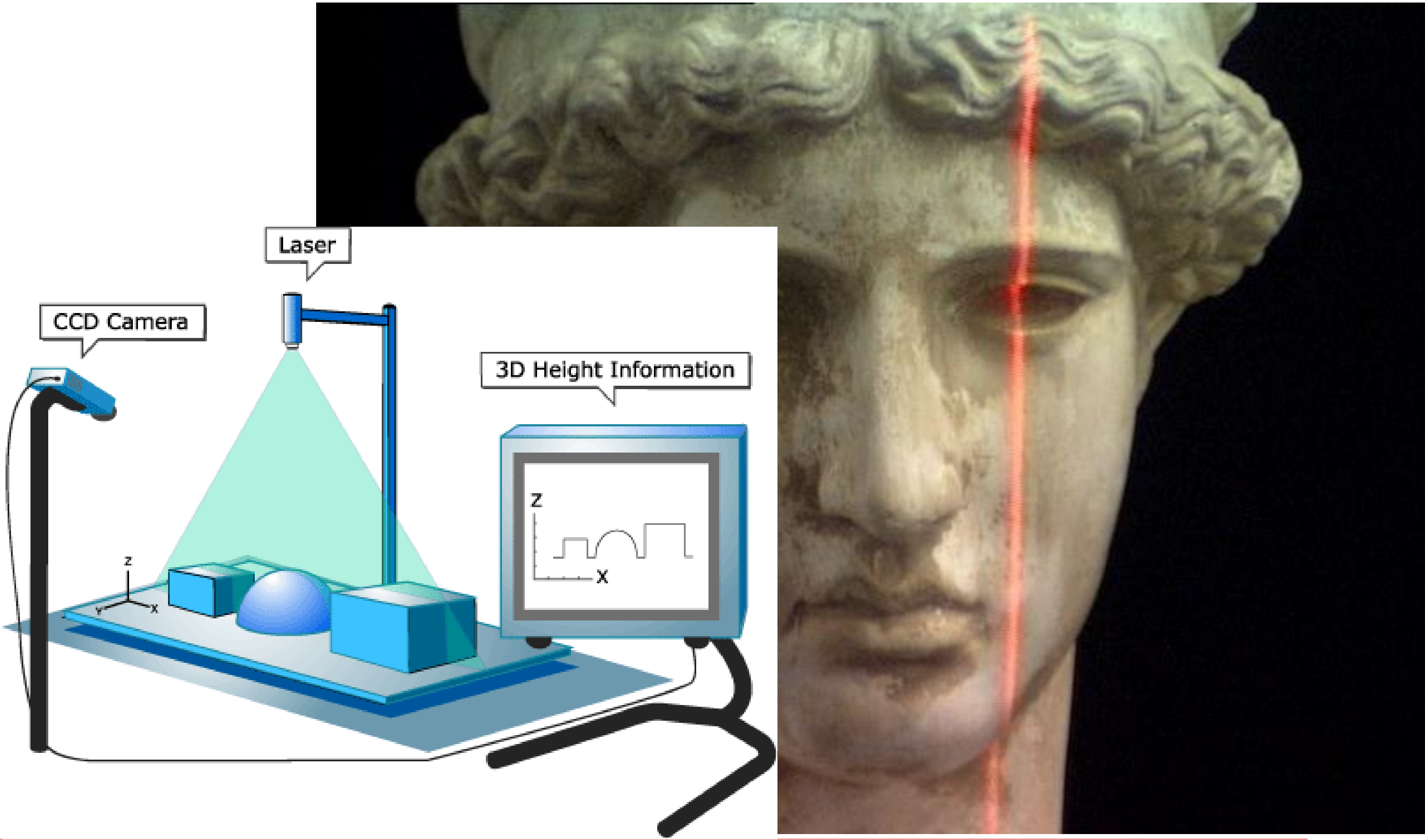
Digital Michelangelo Project
http://graphics.stanford.edu/projects/mich/

- Optical triangulation
  - Project a single stripe of laser light
  - Scan it across the surface of the object
  - This is a very precise version of structured light scanning

Source: S. Seitz

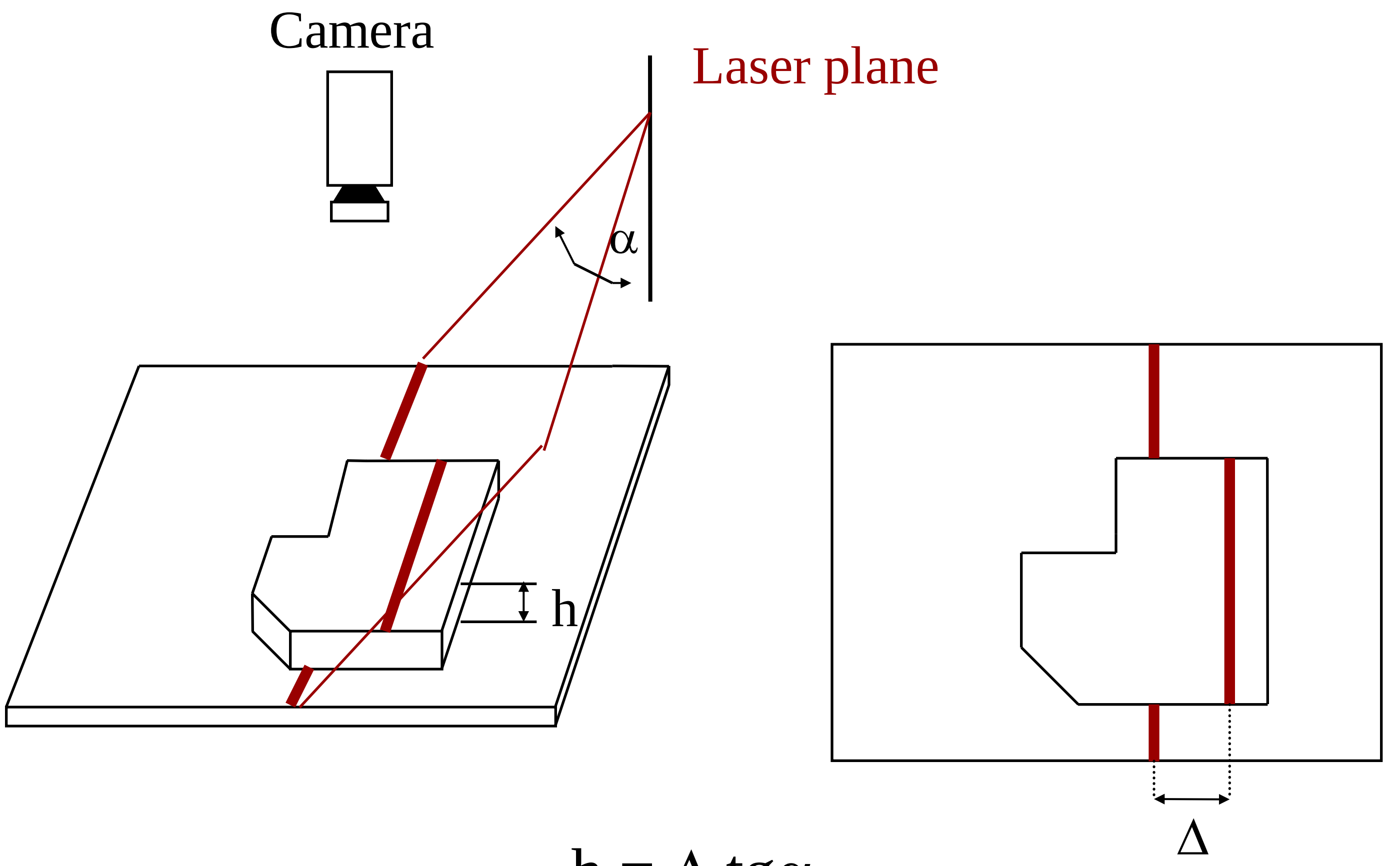
Structured light: plane

- Point
- Plane
- Grid



Structured light: plane

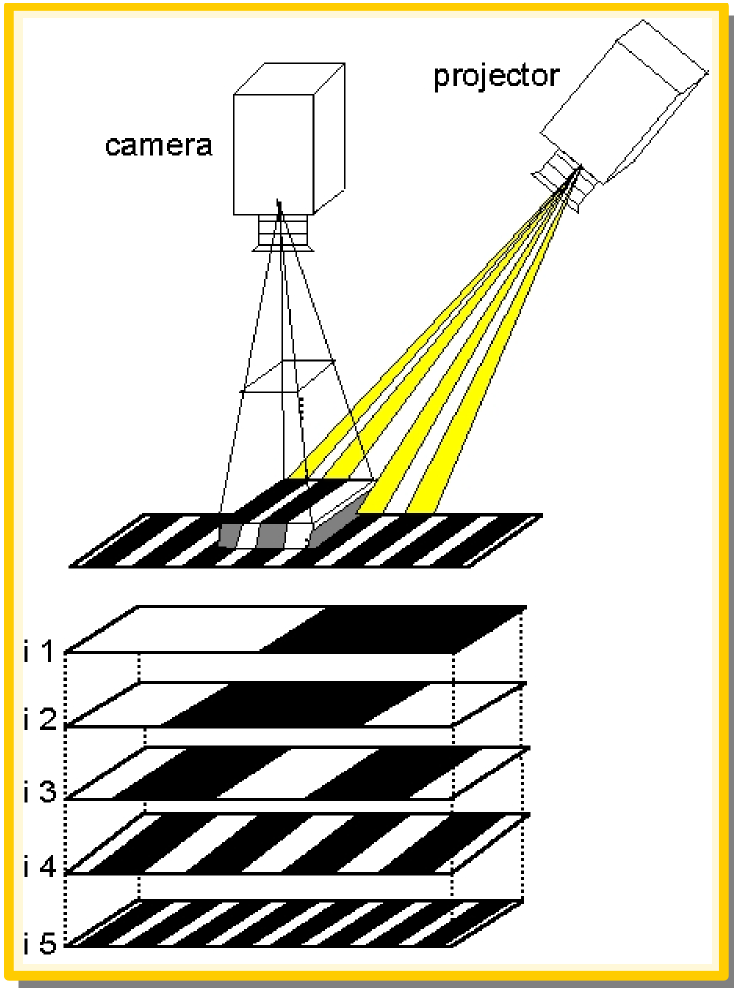
- Point
- Plane
- Grid

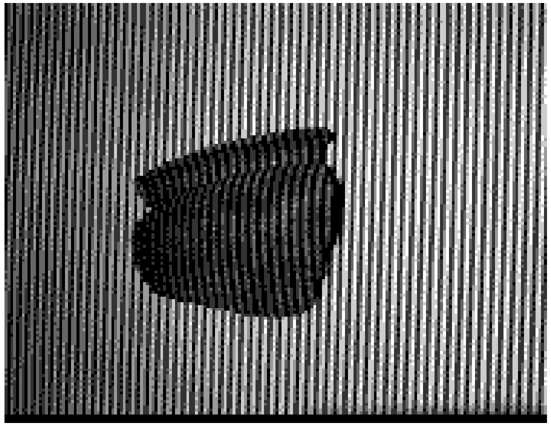


$$h = \Delta \, tg\alpha$$

Structured light: grid

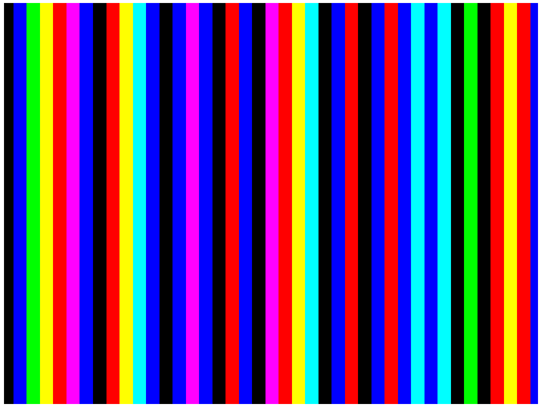
- Point
- Plane
- Grid

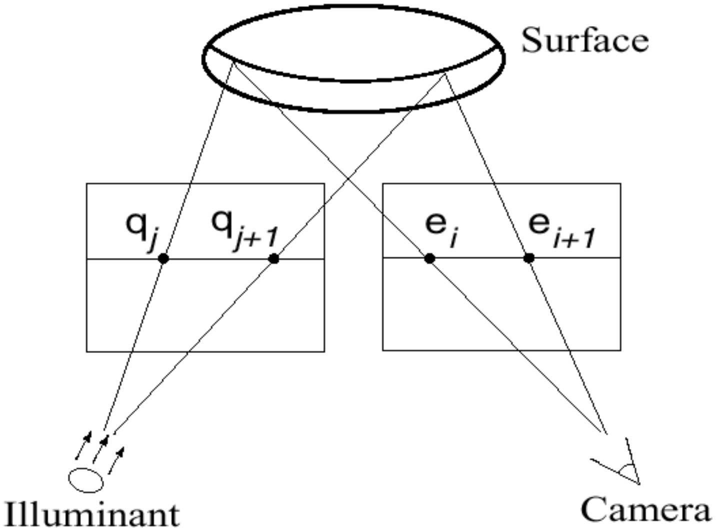


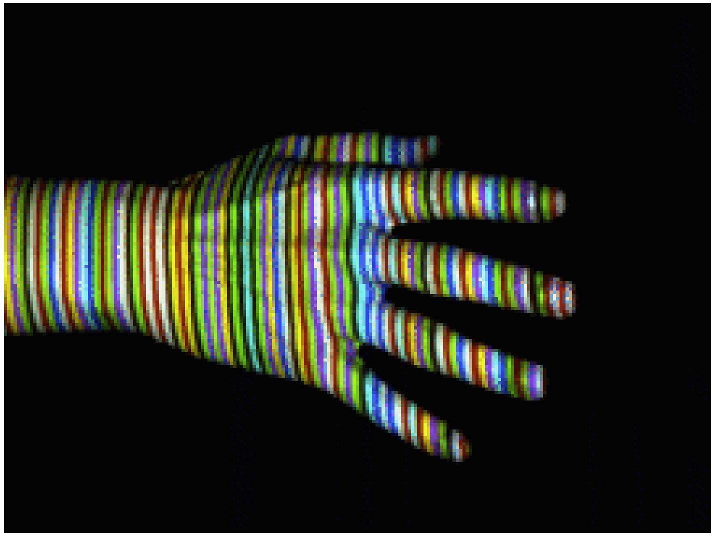


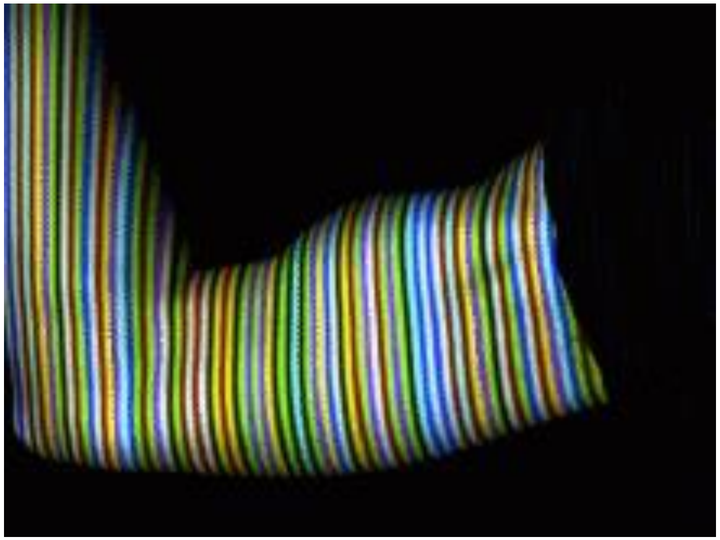
Structured light: plane

- Point
- Plane
- Grid



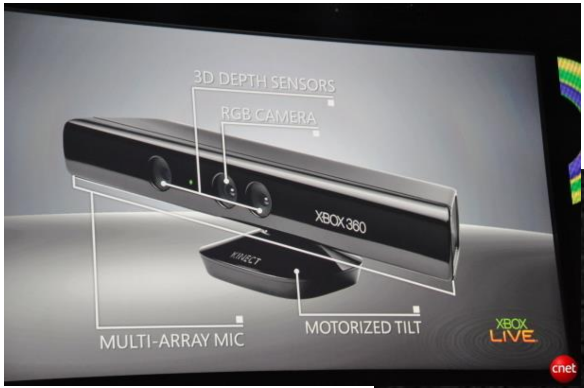






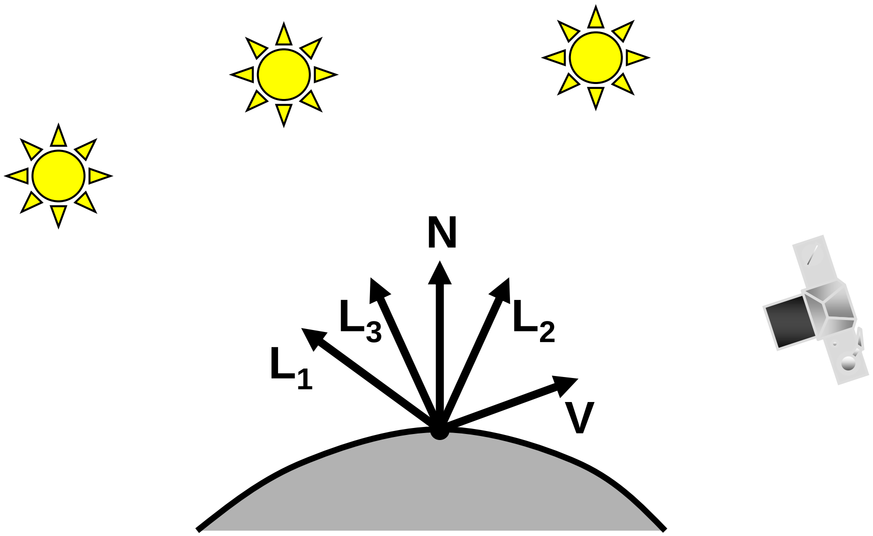
L. Zhang, B. Curless, and S. M. Seitz. Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming. *3DPVT* 2002

Kinect: Structured infrared light



http://bbzippo.wordpress.com/2010/11/28/kinect-in-infrared/

Photometric stereo



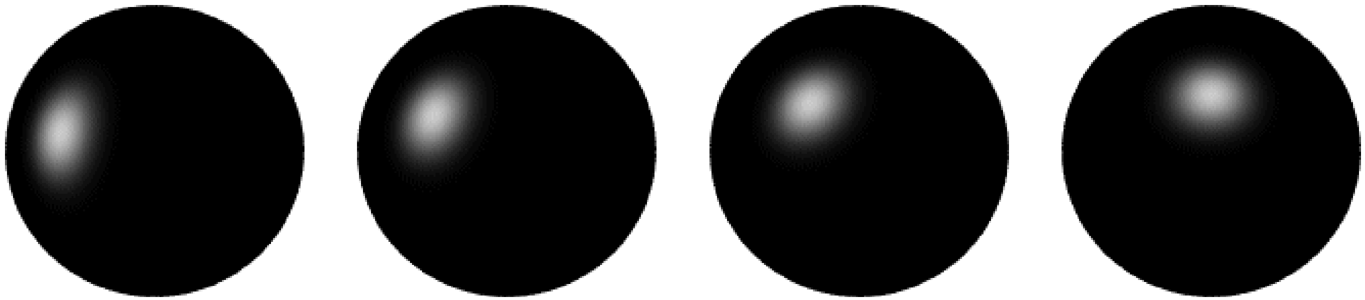
$$I_1 = k_d \mathbf{N} \cdot \mathbf{L_1}$$
$$I_2 = k_d \mathbf{N} \cdot \mathbf{L_2}$$
$$I_3 = k_d \mathbf{N} \cdot \mathbf{L_3}$$

Can write this as a matrix equation:

$$\begin{bmatrix} I_1 & I_2 & I_3 \end{bmatrix} = k_d \mathbf{N}^T \begin{bmatrix} \mathbf{L_1} & \mathbf{L_2} & \mathbf{L_3} \end{bmatrix}$$

Computing light source directions

- Trick: place a chrome sphere in the scene



  - the location of the highlight tells you where the light source is

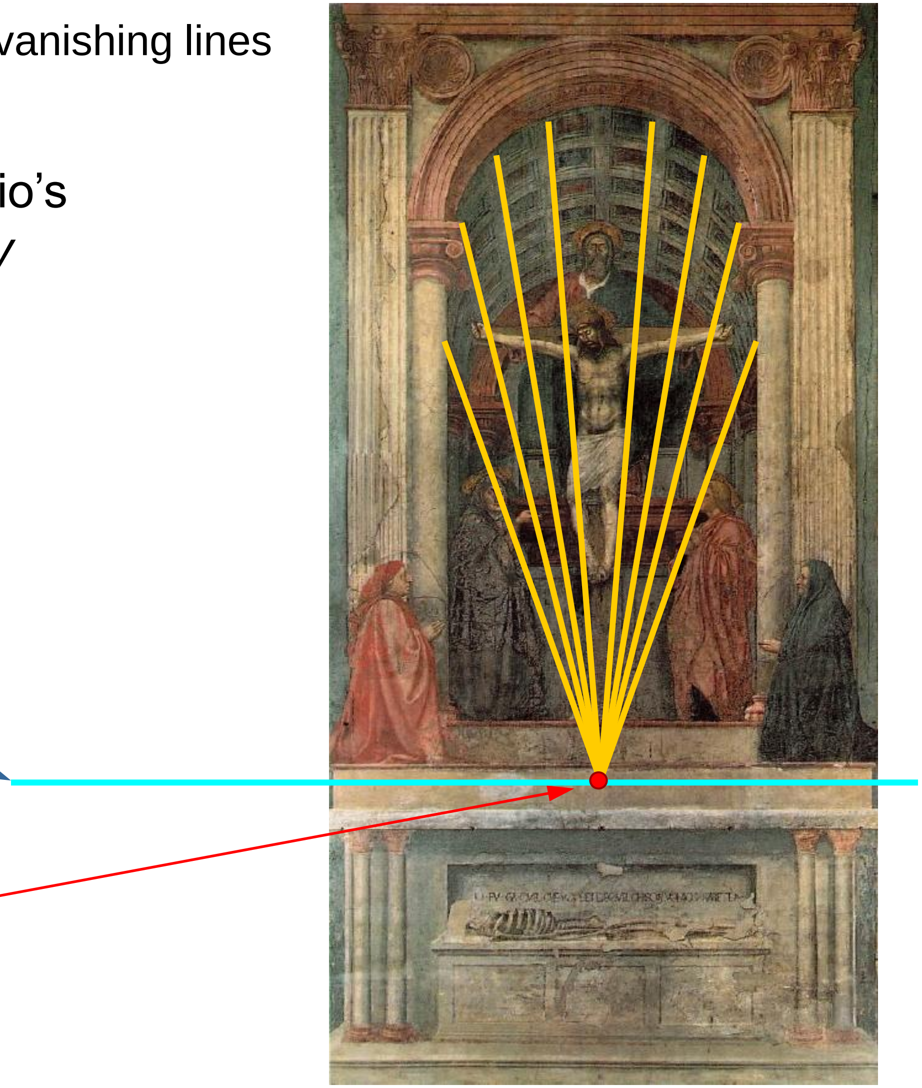
Single View Metrology

Three-dimensional reconstruction from single views

Single-View Reconstruction

- Geometric cues: Exploiting vanishing points and vanishing lines
- Interactive reconstruction process

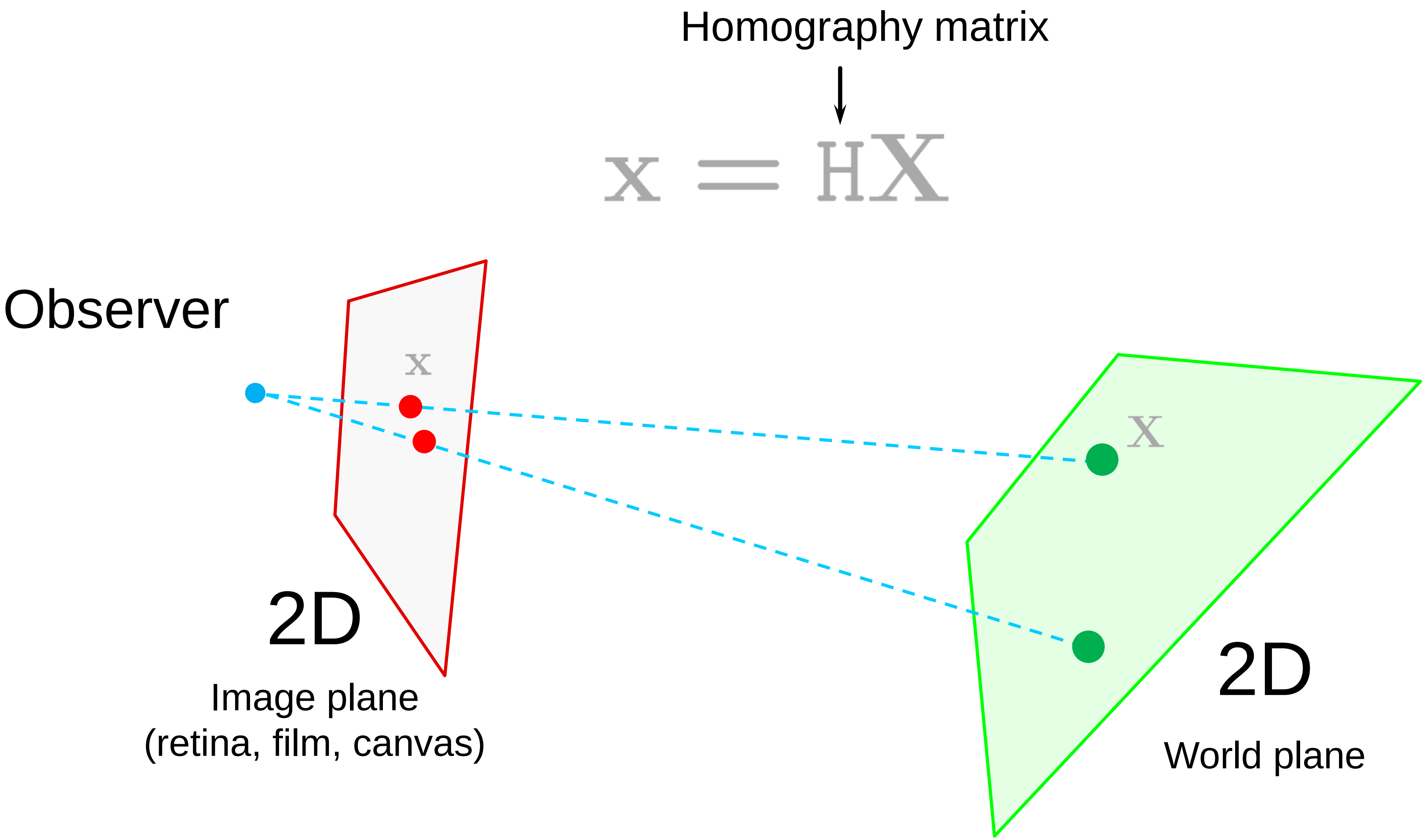
Masaccio’s
Trinity



Vanishing line (horizon)

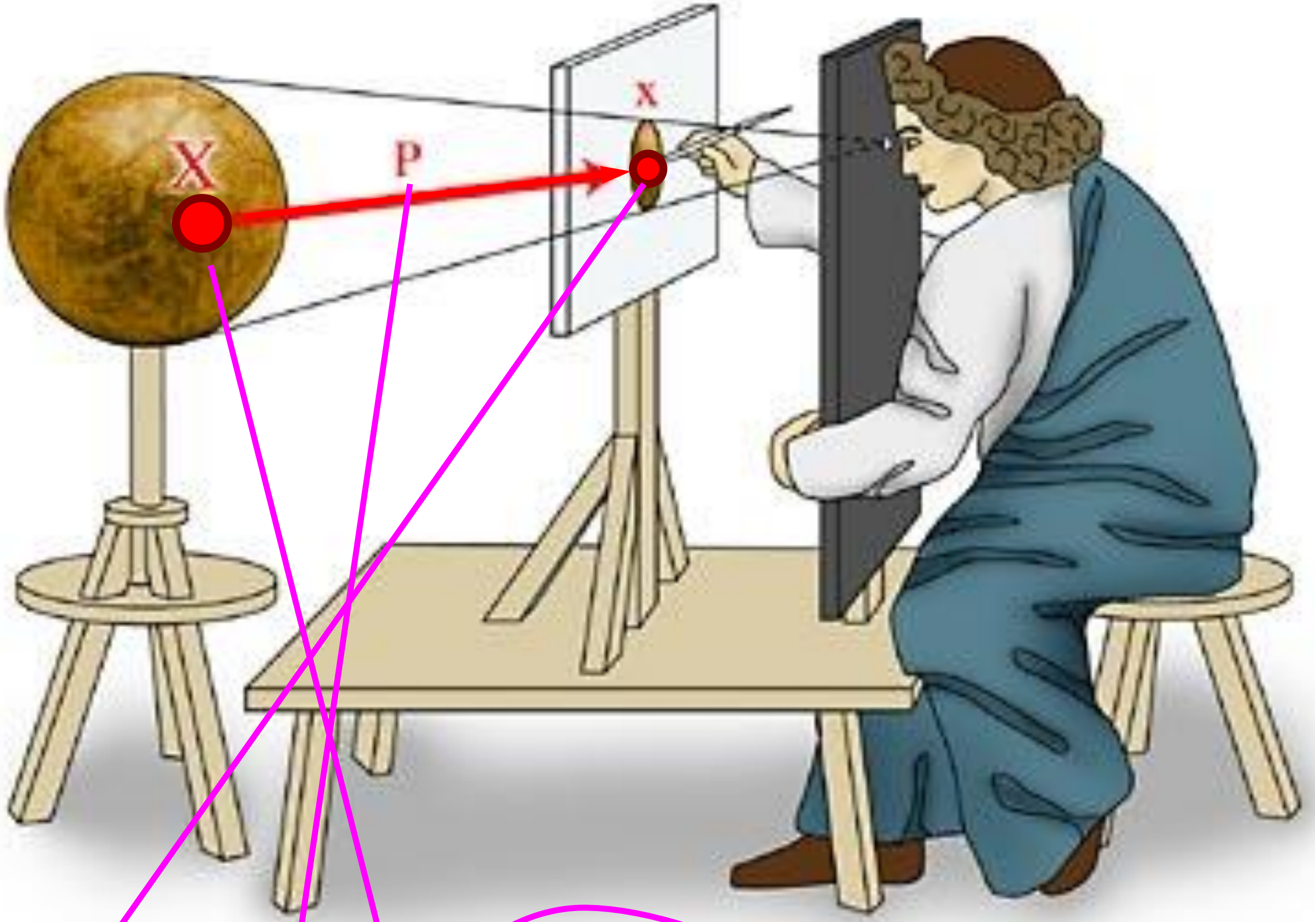
Vanishing point

A special case, planes



H: a plane to plane projective transformation

3D-2D Projective mapping

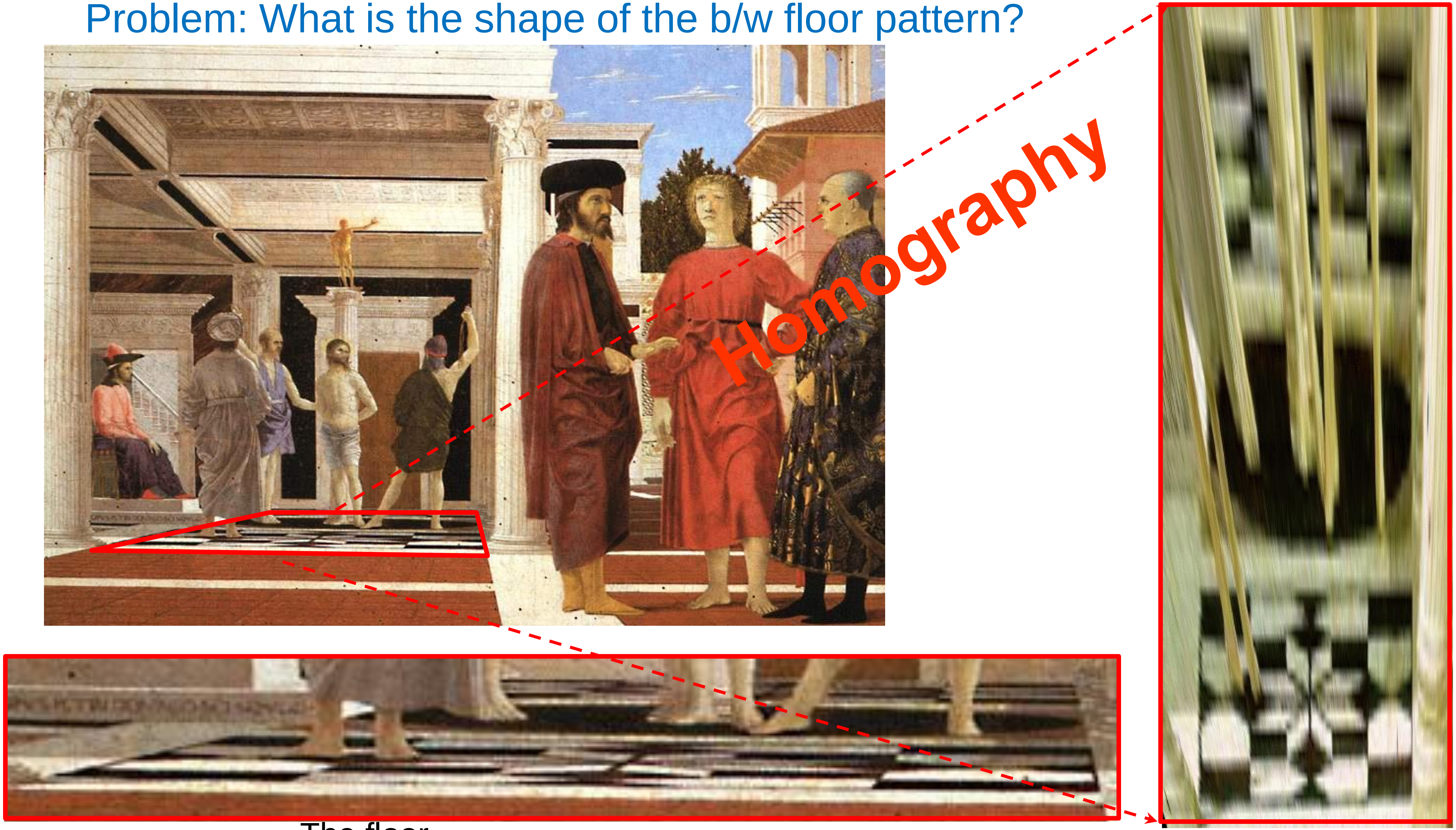


$$\mathbf{x} = \mathbf{P}\,\mathbf{X}$$

Projection Matrix (3x4)

Analysing patterns and shapes

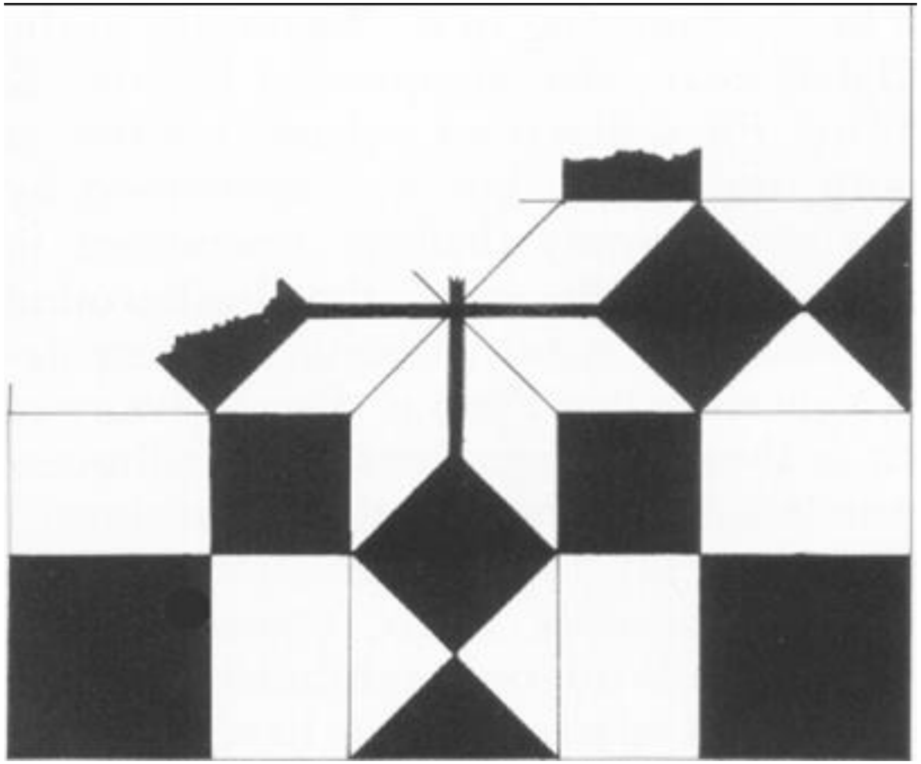
Problem: What is the shape of the b/w floor pattern?



The floor

Automatically rectified floor

Analysing patterns and shapes

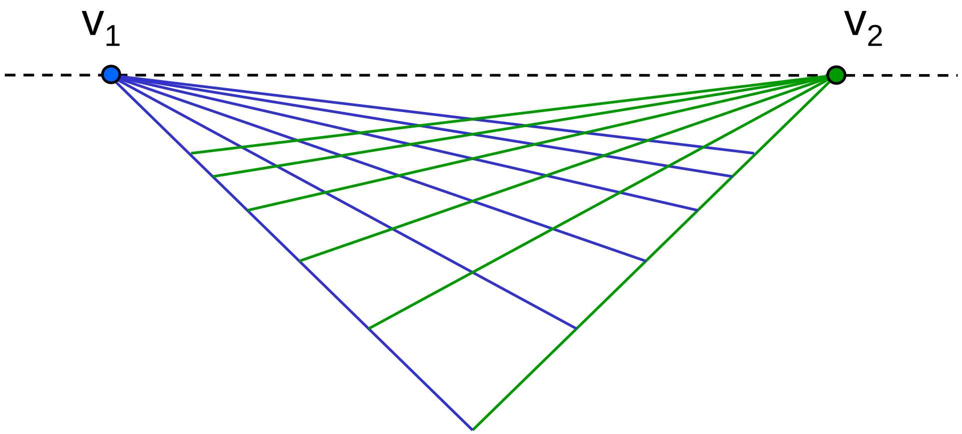


From Martin Kemp *The Science of Art*
(manual reconstruction)

automatic rectification

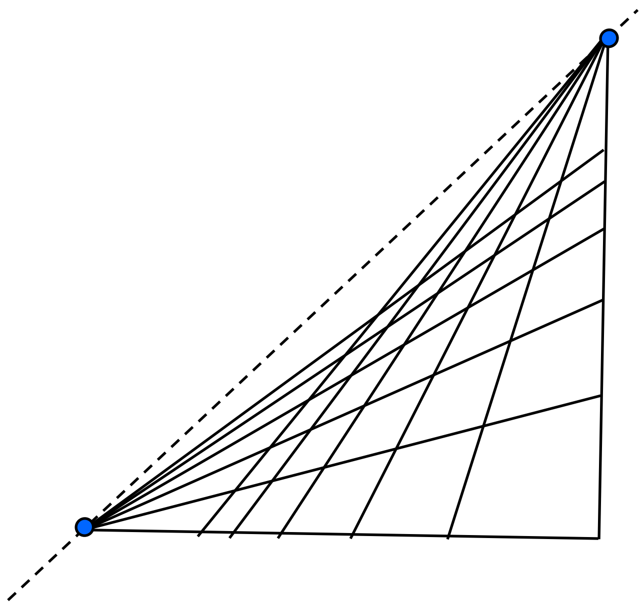
2 patterns have been discovered !

Vanishing lines



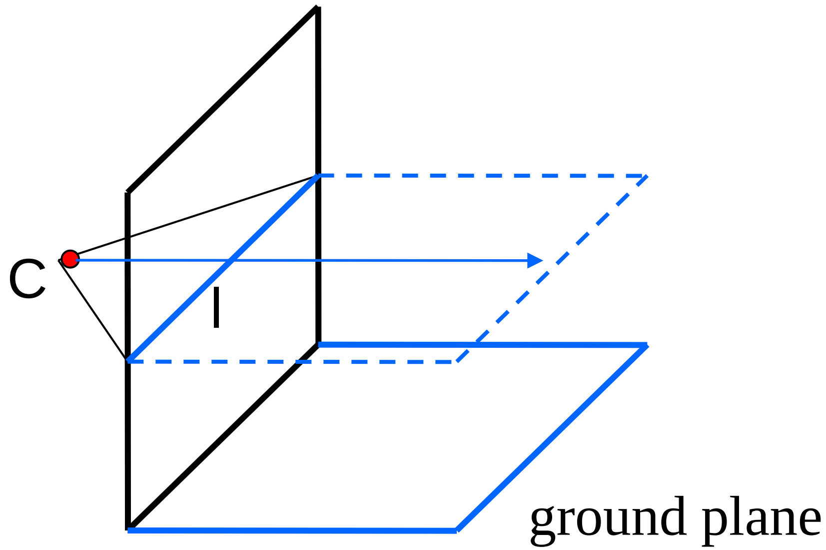
- Multiple Vanishing Points
  - Any set of parallel lines on the plane define a vanishing point
  - The union of all of vanishing points from lines on the same plane is the vanishing line
    - ✓ For the ground plane, this is called the horizon

Vanishing lines



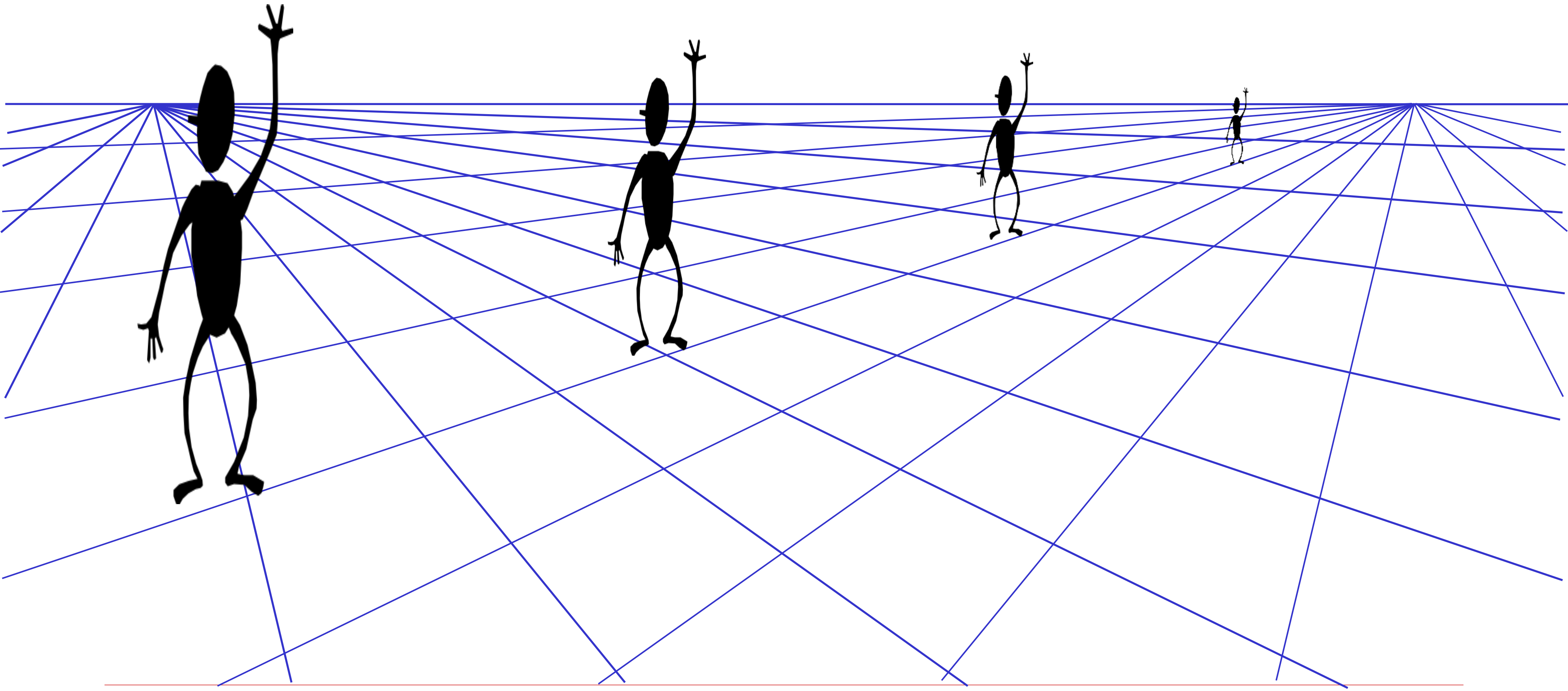
- Multiple Vanishing Points
  - Different planes define different vanishing lines

Computing the horizon

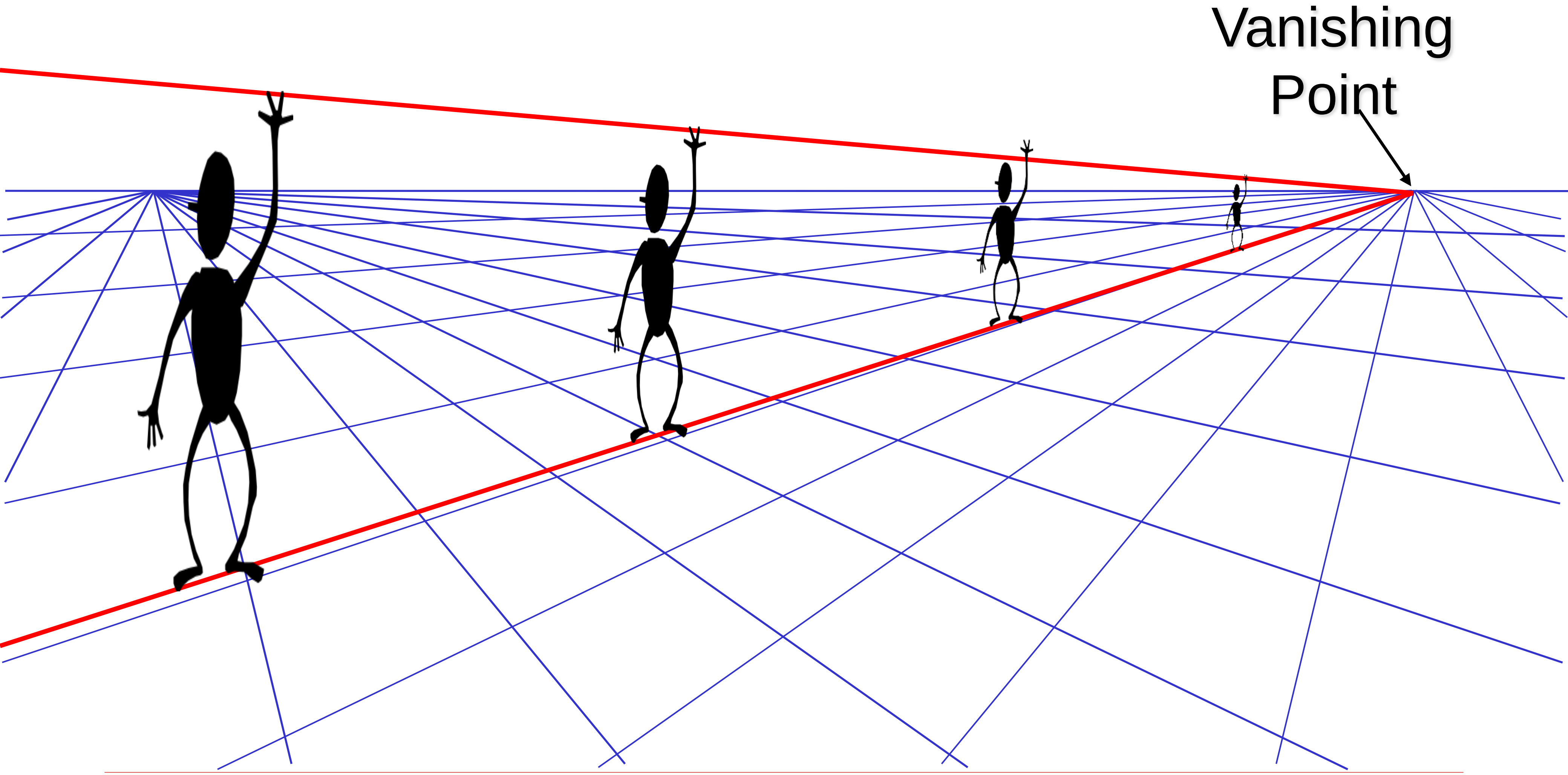


- Properties
  - l is intersection of horizontal plane through C with image plane
  - Compute l from two sets of parallel lines on ground plane
  - All points at same height as C project to l
  - Provides way of comparing height of objects in the scene

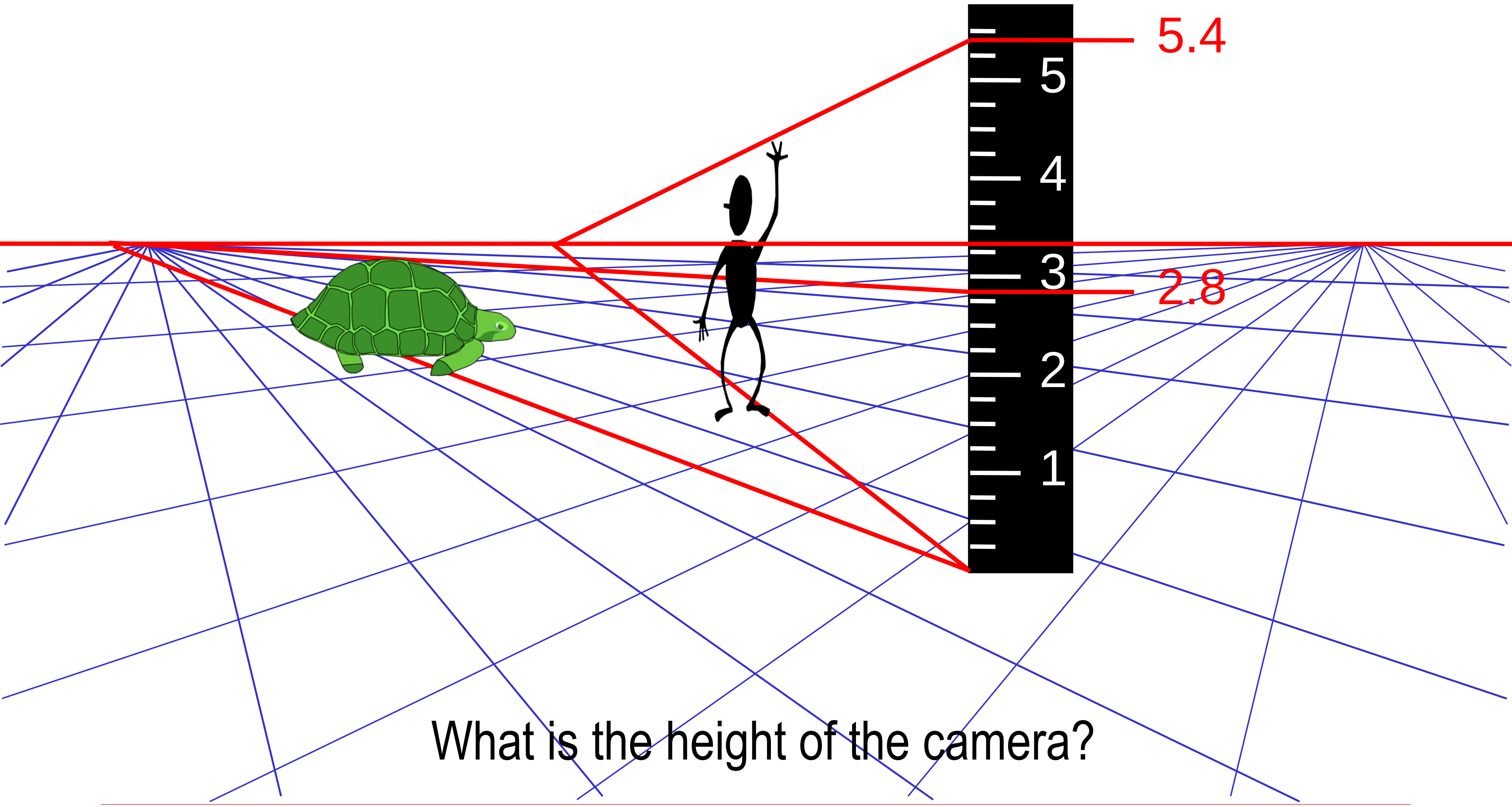
Are these guys the same height?



Comparing heights

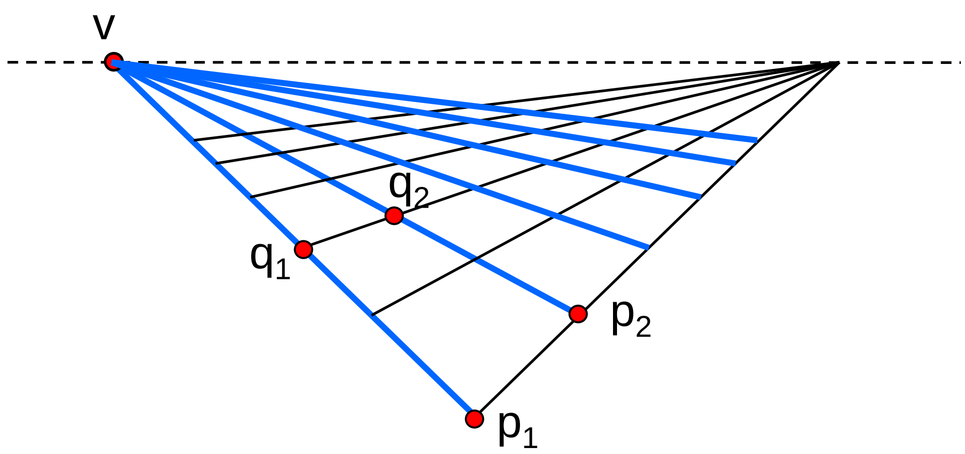


Measuring height



What is the height of the camera?

Computing vanishing points (from lines)

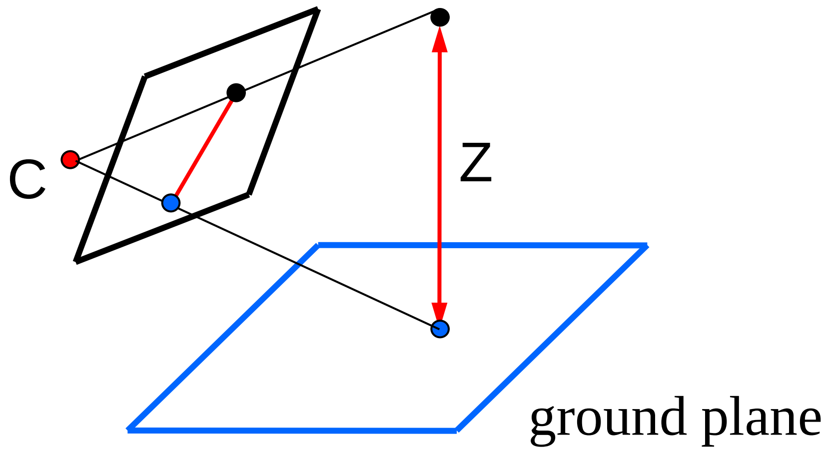


- Intersect p_1q_1 with p_2q_2

$$v = (p_1 \times q_1) \times (p_2 \times q_2)$$

- Least squares version
- Better to use more than two lines and compute the “closest” point of intersection
- See notes by Bob Collins for one good way of doing this:
  - http://www-2.cs.cmu.edu/~ph/869/www/notes/vanishing.txt

Measuring height without a ruler



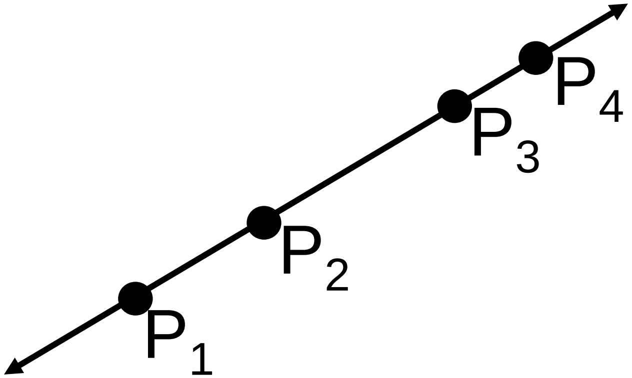
Compute Z from image measurements

- Need more than vanishing points to do this

The cross ratio

- A Projective Invariant
  - Something that does not change under projective transformations (including perspective projection)

The cross-ratio of 4 collinear points



$$\frac{\|\mathbf{P}_3 - \mathbf{P}_1\| \, \|\mathbf{P}_4 - \mathbf{P}_2\|}{\|\mathbf{P}_3 - \mathbf{P}_2\| \, \|\mathbf{P}_4 - \mathbf{P}_1\|}$$

$$\mathbf{P}_i = \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}$$

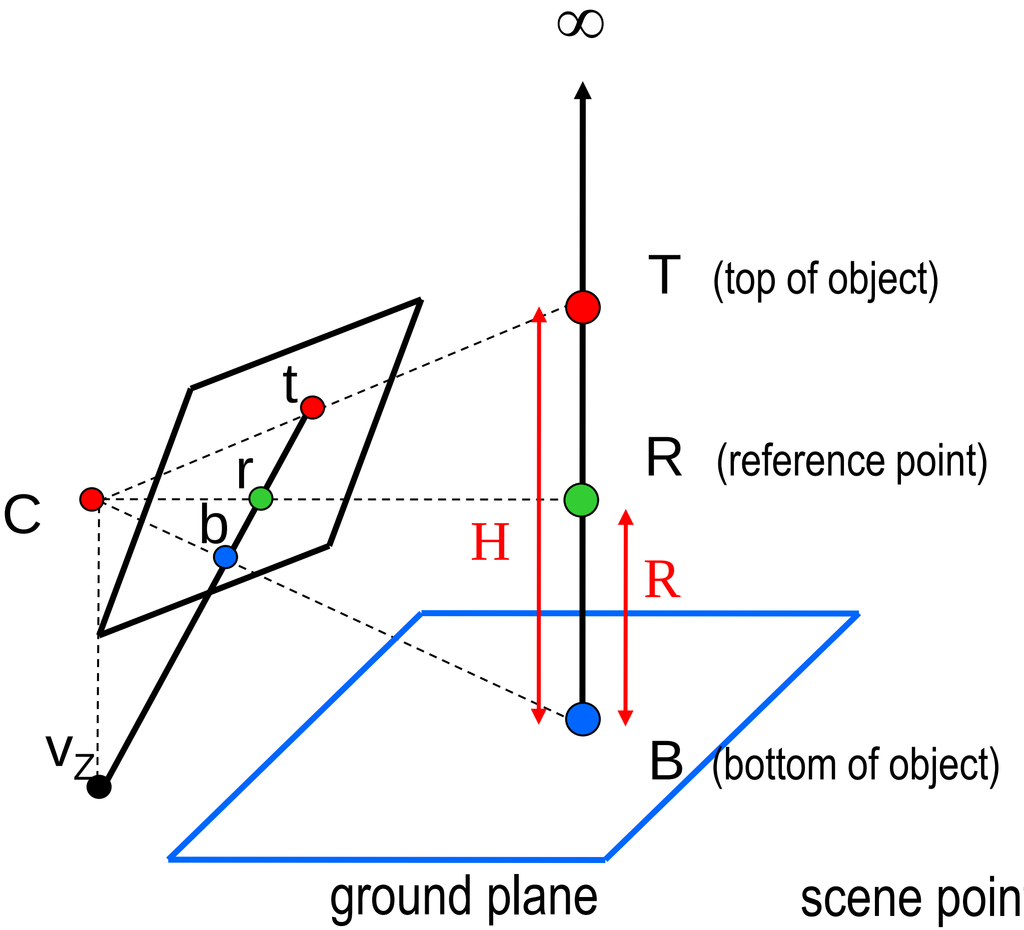
$$\frac{\|\mathbf{P}_1 - \mathbf{P}_3\| \, \|\mathbf{P}_4 - \mathbf{P}_2\|}{\|\mathbf{P}_1 - \mathbf{P}_2\| \, \|\mathbf{P}_4 - \mathbf{P}_3\|}$$

Can permute the point ordering

- 4! = 24 different orders (but only 6 distinct values)

This is the fundamental invariant of projective geometry

Measuring height



$$\frac{\|\mathbf{T}-\mathbf{B}\|\,\|\infty-\mathbf{R}\|}{\|\mathbf{R}-\mathbf{B}\|\,\|\infty-\mathbf{T}\|} = \frac{H}{R}$$

scene cross ratio

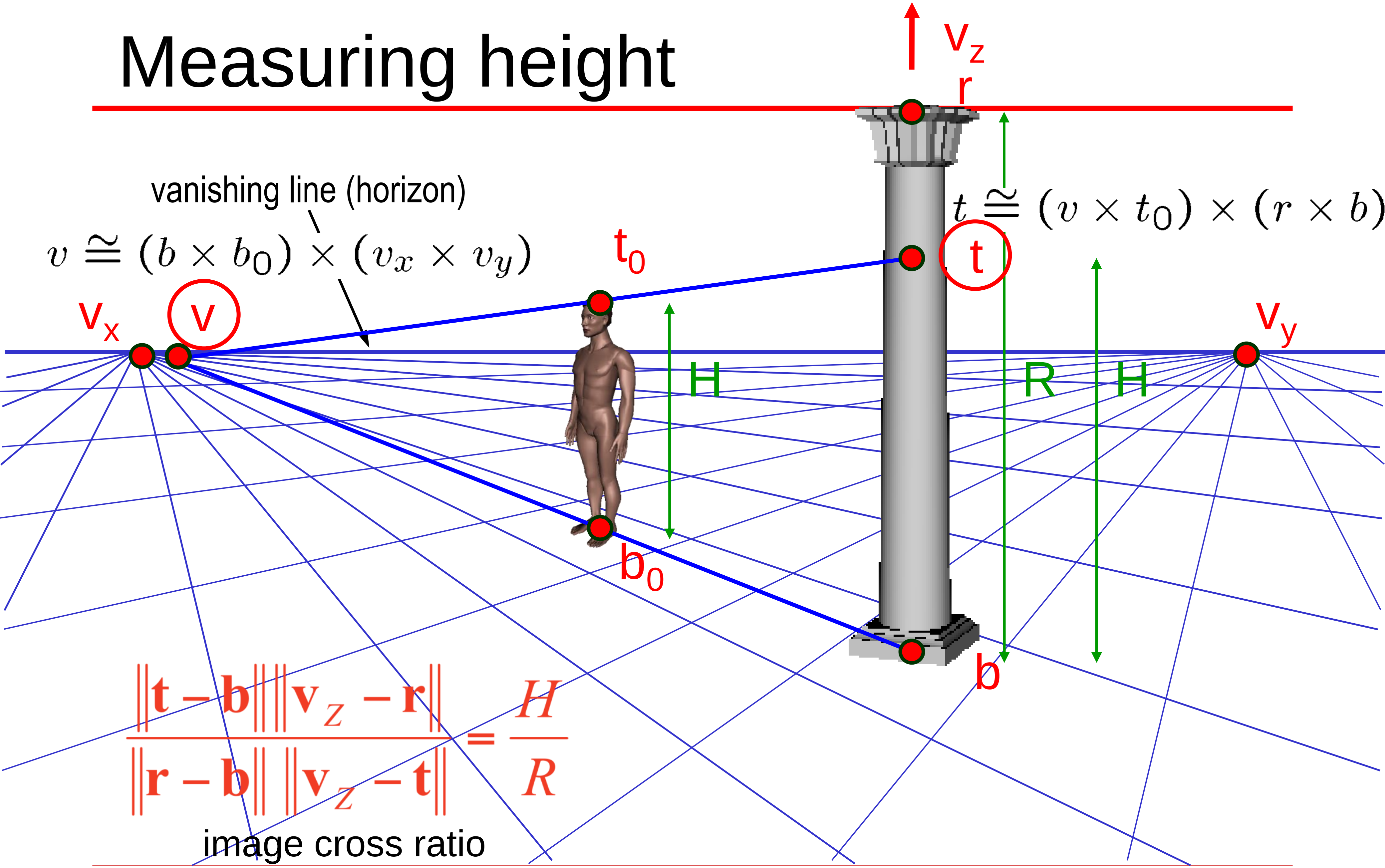
$$\frac{\|\mathbf{t}-\mathbf{b}\|\,\|\mathbf{v}_Z-\mathbf{r}\|}{\|\mathbf{r}-\mathbf{b}\|\,\|\mathbf{v}_Z-\mathbf{t}\|} = \frac{H}{R}$$

image cross ratio

image points as $\mathbf{p} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$

scene points represented as $\mathbf{P} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$

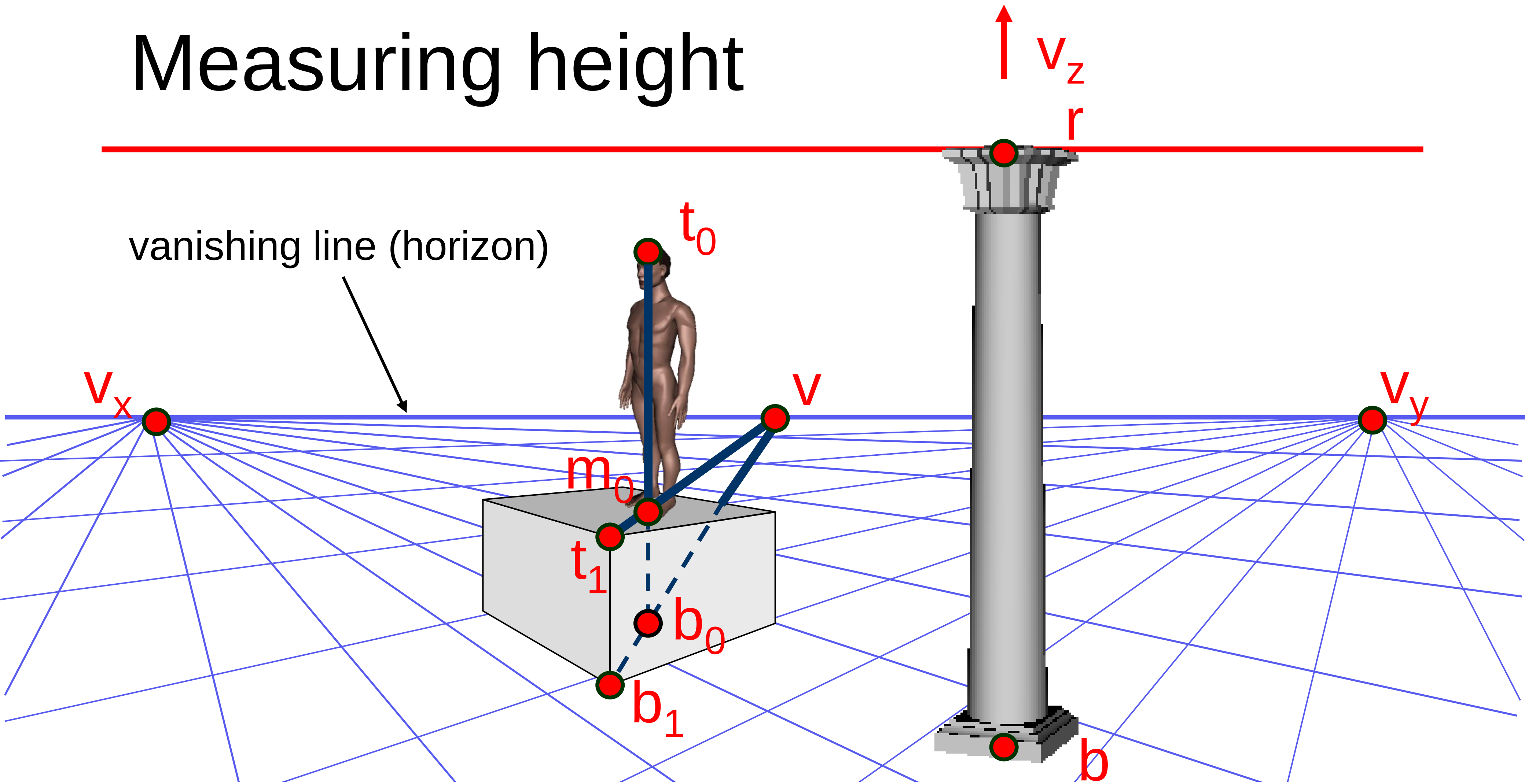
Measuring height



$$\frac{\|\mathbf{t}-\mathbf{b}\|\,\|\mathbf{v}_Z-\mathbf{r}\|}{\|\mathbf{r}-\mathbf{b}\|\,\|\mathbf{v}_Z-\mathbf{t}\|}=\frac{H}{R}$$

image cross ratio

Measuring height



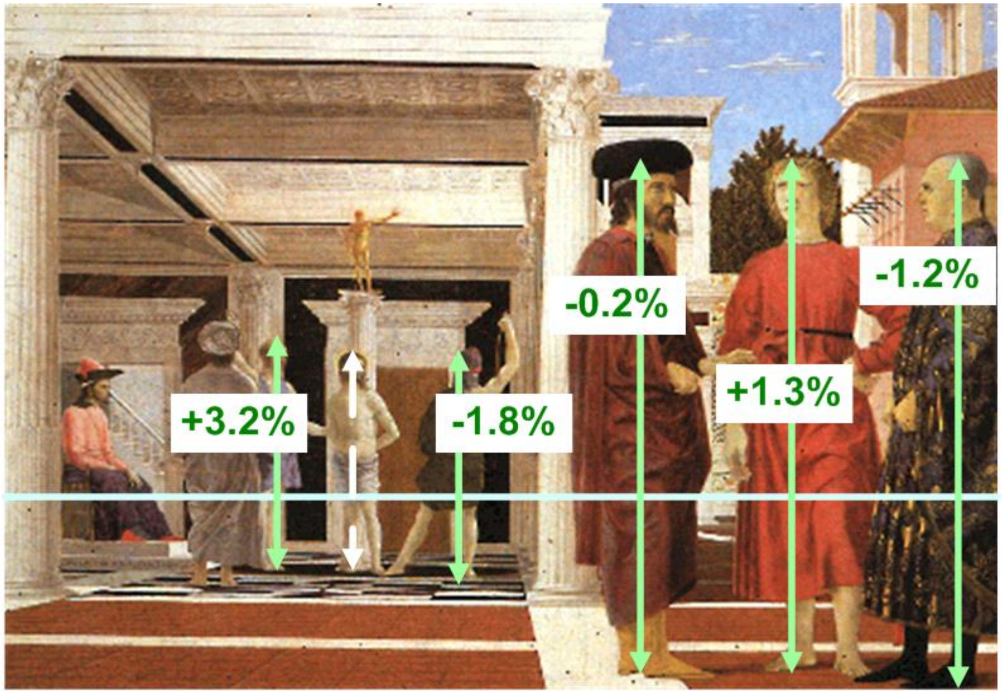
What if the point on the ground plane b_0 is not known?

- Here the guy is standing on the box
- Use one side of the box to help find b_0 as shown above

Assessing geometric accuracy

Problem: Are the heights of the two groups of people consistent with each other?

Piero della Francesca,
Flagellazione di Cristo,
c.1460, Urbino

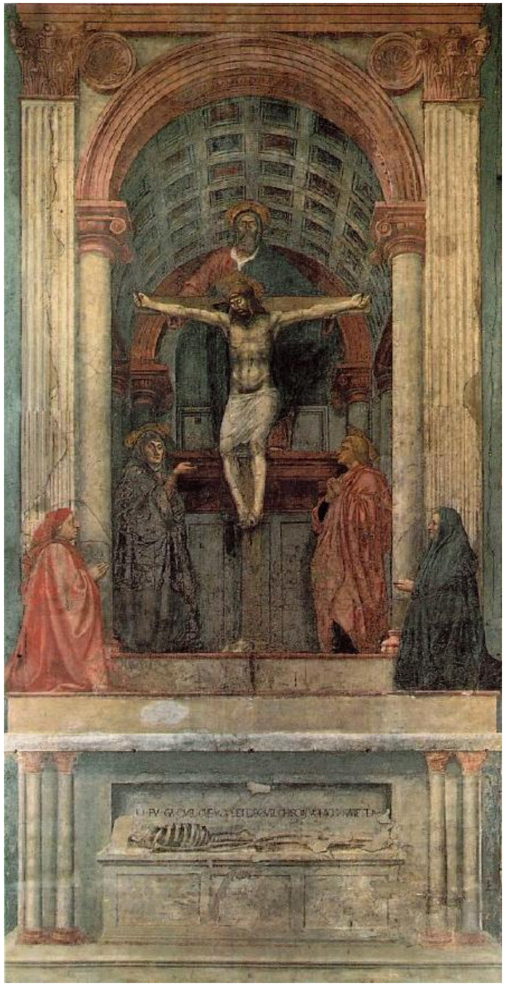


Measuring relative heights

Single-View Metrology

Complete 3D reconstructions from single views

Example: The Virtual Trinity



Masaccio, *Trinità*,
1426, Florence

Complete 3D reconstruction

Example: The Virtual Flagellation

Piero della Francesca,
Flagellazione di Cristo,
c.1460, Urbino

Complete 3D reconstruction

Example: The Virtual St. Jerome

Henry V Steenwick,
St.Jerome in His Study,
1630, The Netherlands

Complete 3D reconstruction

Example: The Virtual Music Lesson

J. Vermeer,
The Music Lesson,
1665, London

Complete 3D reconstruction

Example: A Virtual Museum @ Microsoft

The Image-Based Realities team @ Microsoft Research

Why do we perceive depth?