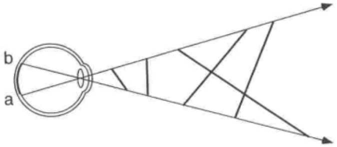
Computer Vision

Monocular & SfM

Jan van Gemert (J.C.vanGemert@uva.nl)

Depth Perception:

The inverse problem



First part: Monocular cues

Second part: Structure From Motion

Monocular cues to depth

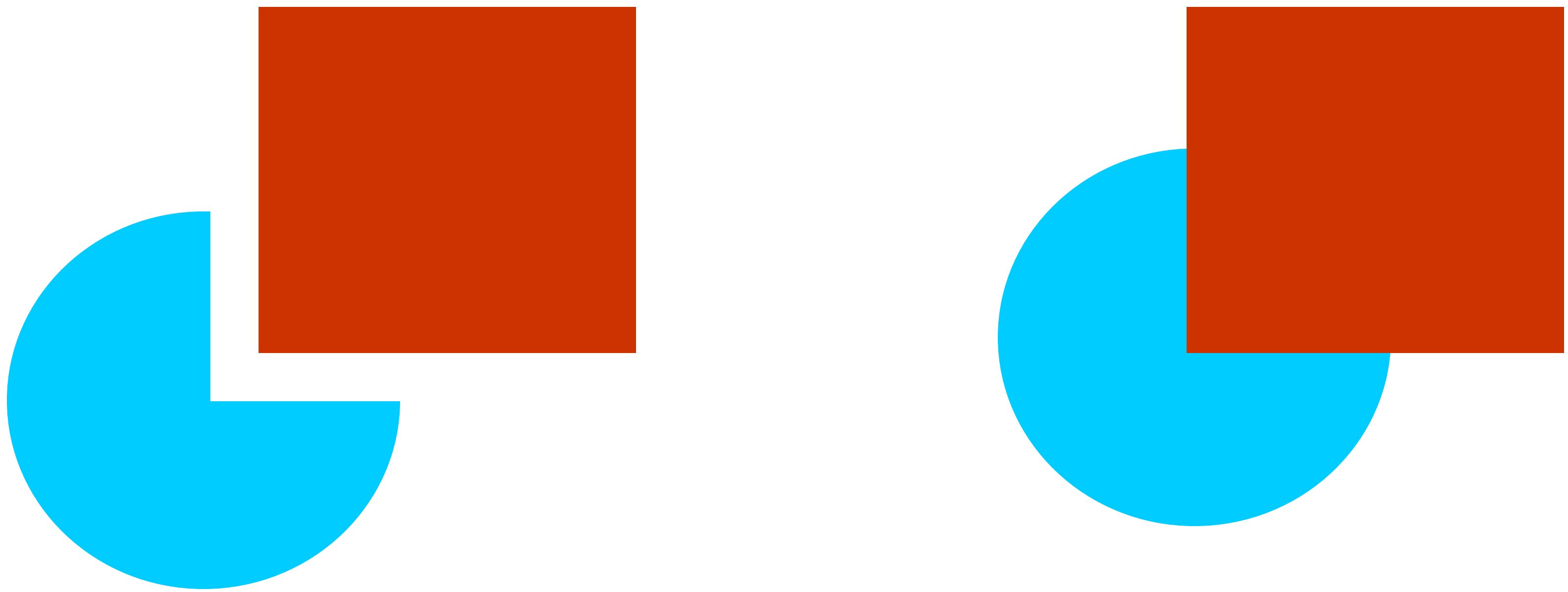
- **Absolute depth cues**: (assuming known camera parameters) these cues provide information about the absolute depth between the observer and elements of the scene
- **Relative depth cues**: provide relative information about depth between elements in the scene (this point is twice as far at that point, …)

Relative depth cues

Simple and powerful cue, but hard to make it work in practice…

What are issues here?

Interposition / occlusion



- Hard to tell if occluded or not
- Does not give ‘absolute’ depth, only an ordering

Texture Gradient

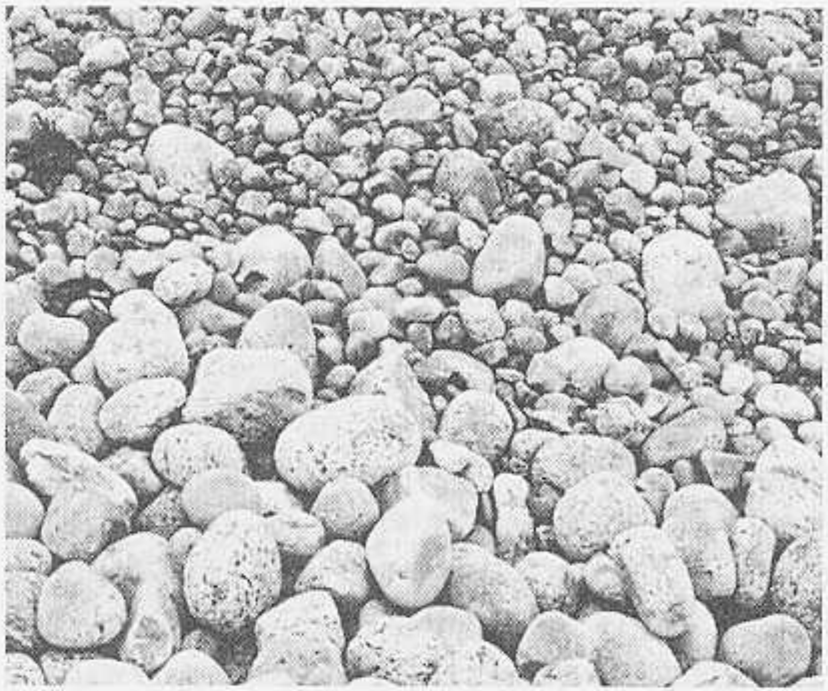


FIGURE 8.27
Texture gradients provide information about depth. (Frank Siteman/Stock, Boston.)

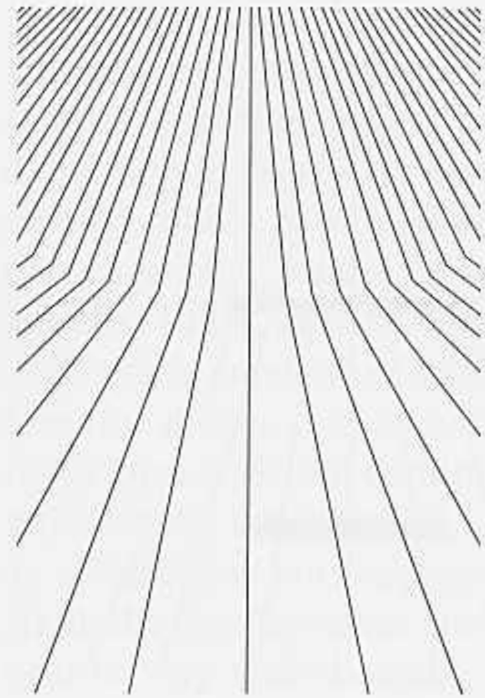
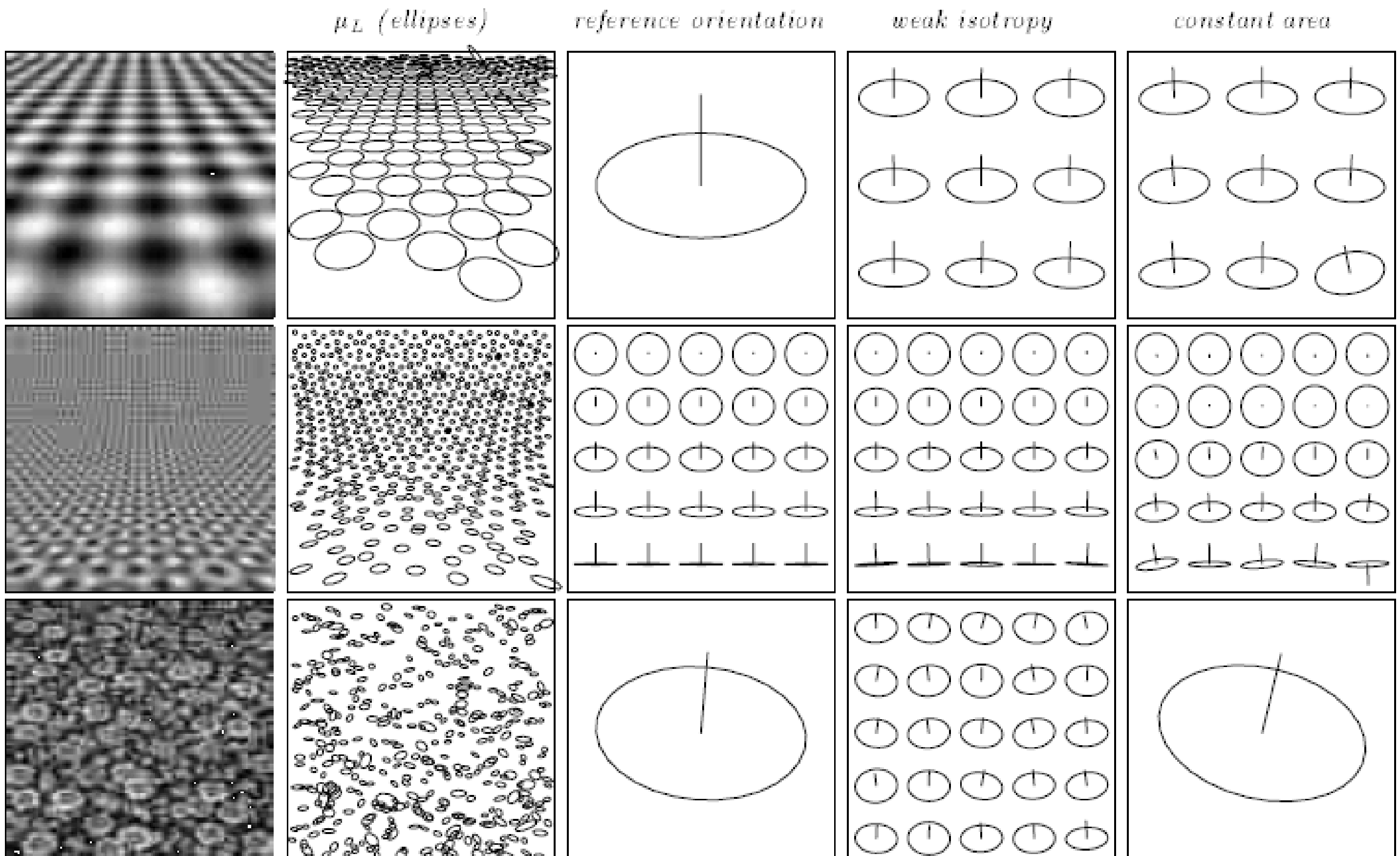


FIGURE 8.28
Texture discontinuity signals the pre
corner.

A Witkin**.** Recovering Surface Shape and Orientation from Texture (1981)

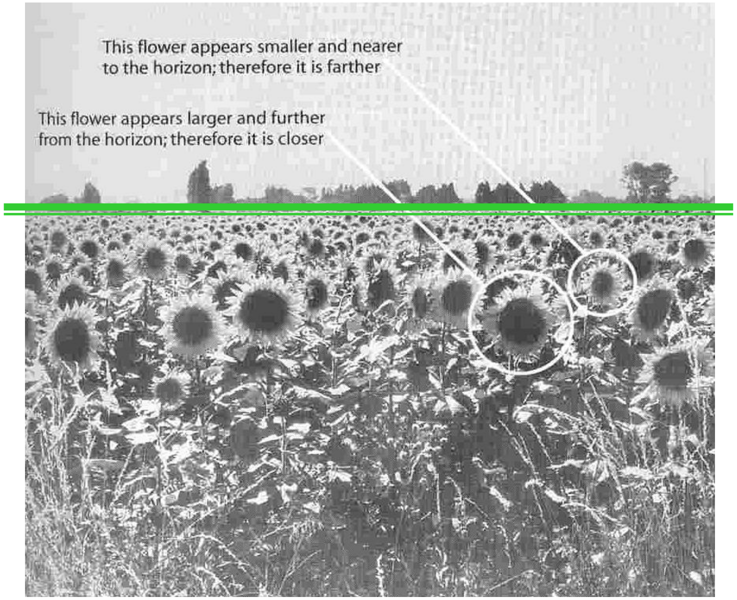
Texture Gradient



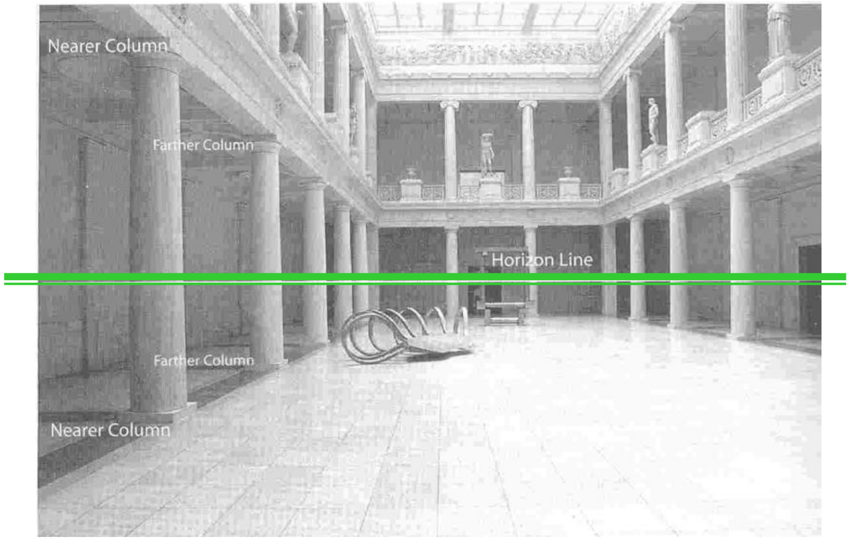
Shape from Texture from a Multi-Scale Perspective. Tony Lindeberg and Jonas Garding. ICCV 93

Distance from the horizon line

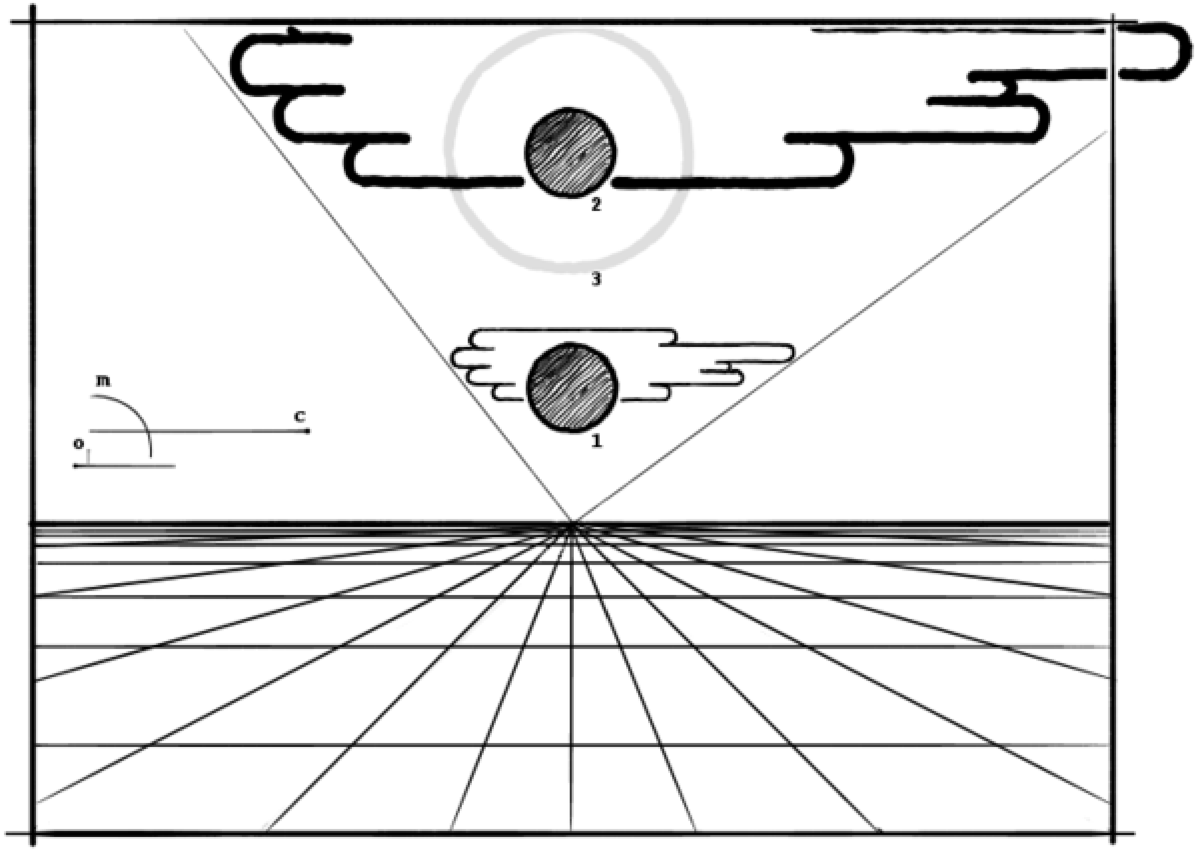
- Objects approach the horizon line with greater distance from the viewer. The base of a nearer column will appear lower against its background floor and further from the horizon line. Conversely, the base of a more distant column will appear higher against the same floor, and thus nearer to the horizon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



the object closer to the horizon is perceived as farther away, and the object further from the horizon is perceived as closer



the Moon illusion



1. Moon, near to the horizon
2. Moon, higher up in the sky
   (appears to be smaller than 1, actually having the same size)
3. the size Moon should have according to the perspective rules

o. observer
c. cloud's path (the distance between the observer and the cloud varies)
m. Moon's path (the distance between the observer and the Moon is almost the same)

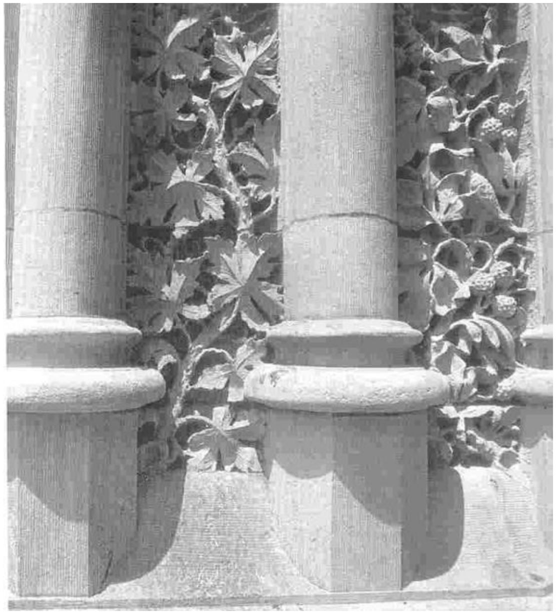
http://en.wikipedia.org/wiki/Moon_illusion

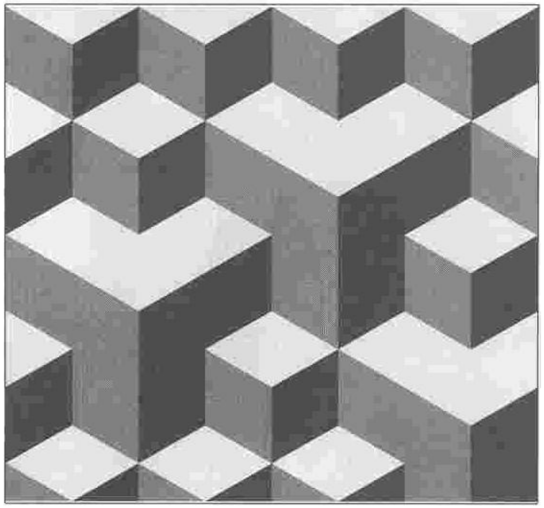
Illumination

- Shading
- Shadows
- Inter-reflections

Shading

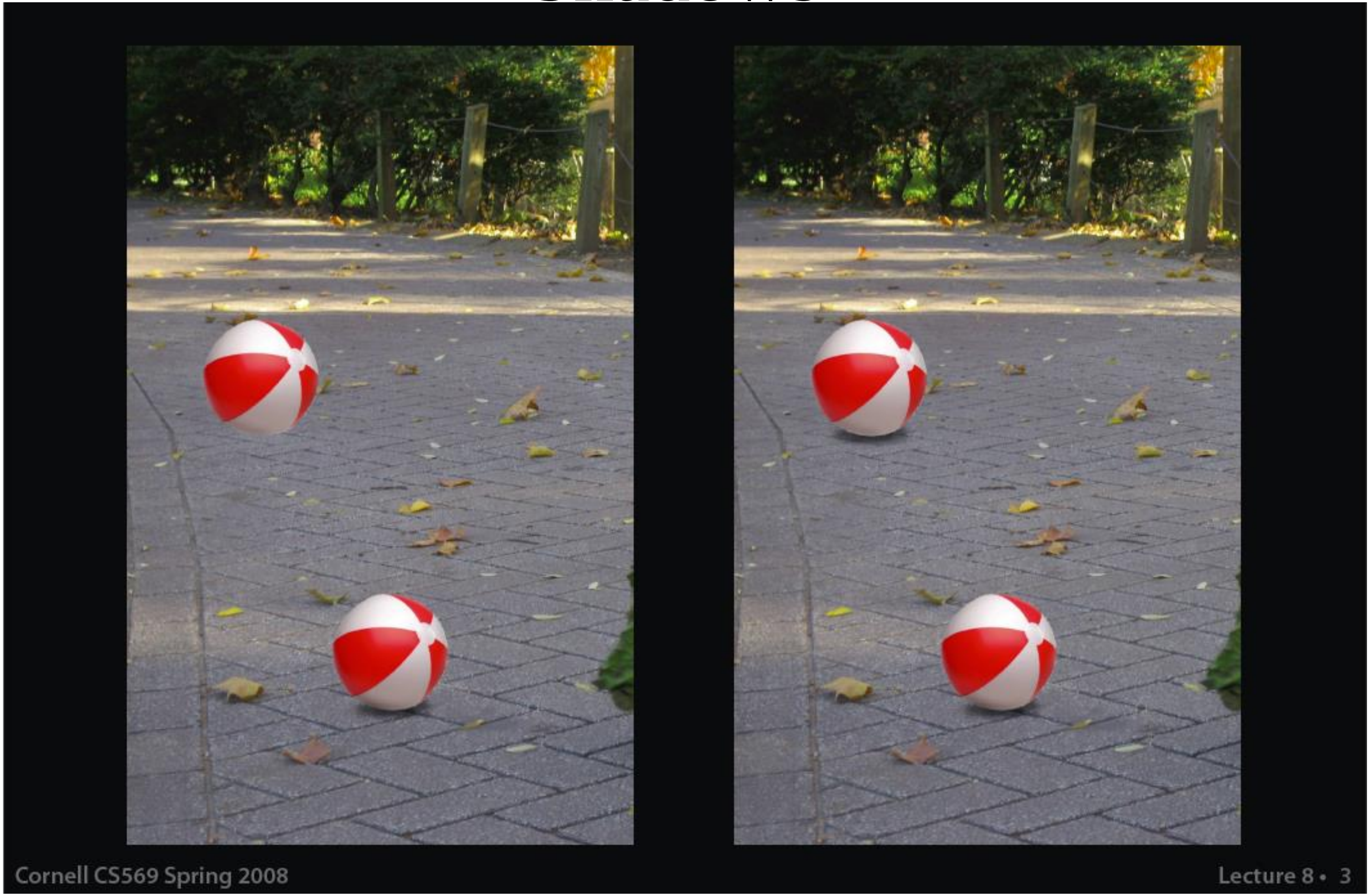
- Based on 3 dimensional modeling of objects in light, shade and shadows.





- Perception of depth through shading alone is always subject to the concave/convex inversion. The pattern shown can be perceived as stairsteps receding towards the top and lighted from above, or as an overhanging structure lighted from below.

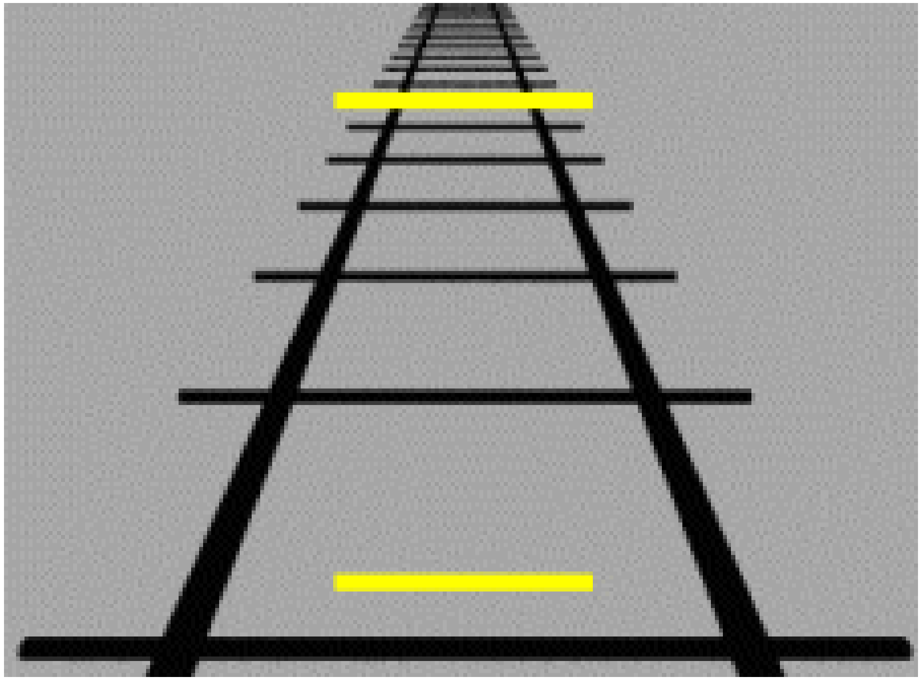
Shadows



Slide by Steve Marschner

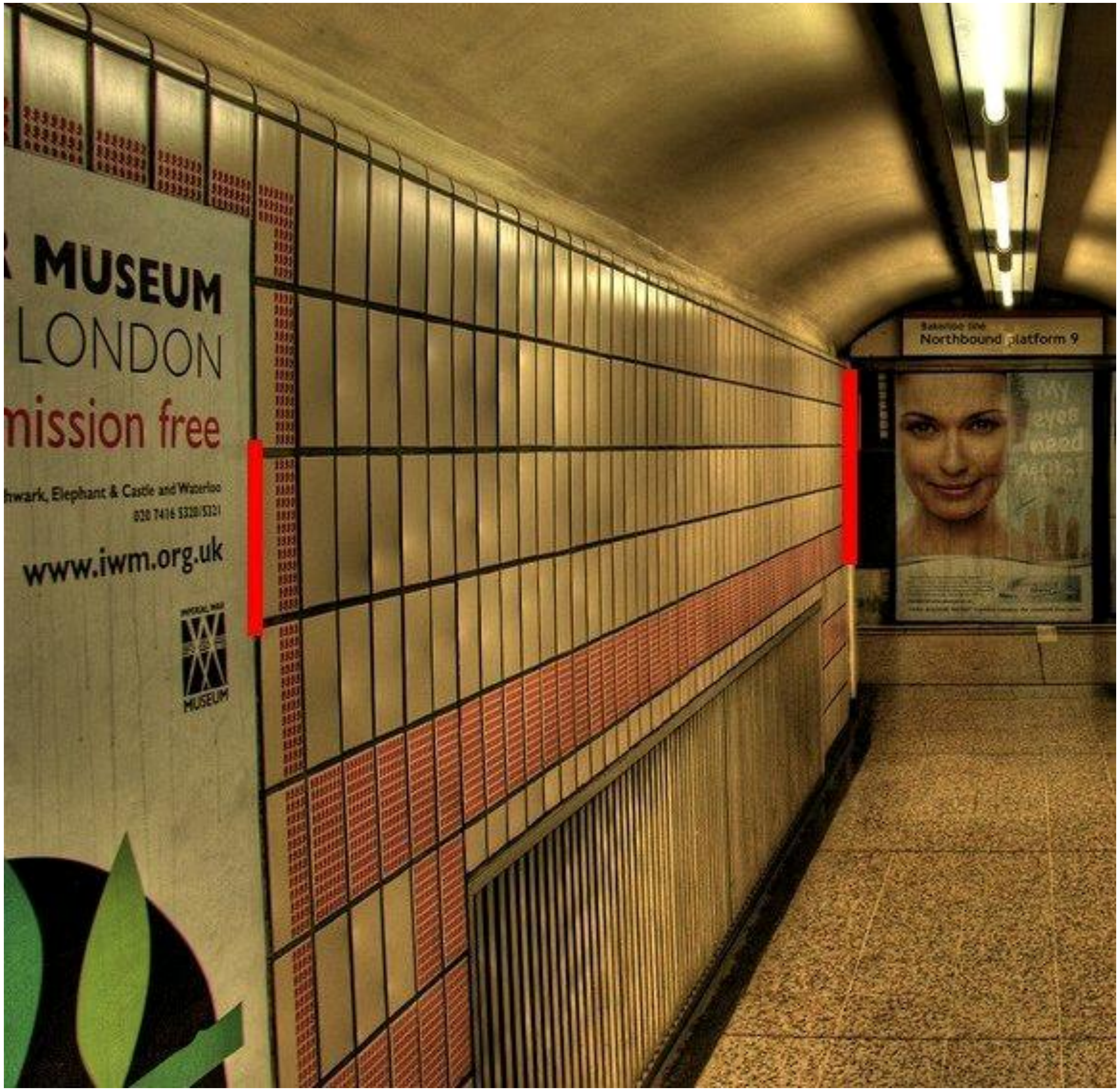
http://www.cs.cornell.edu/courses/cs569/2008sp/schedule.stm

Linear Perspective



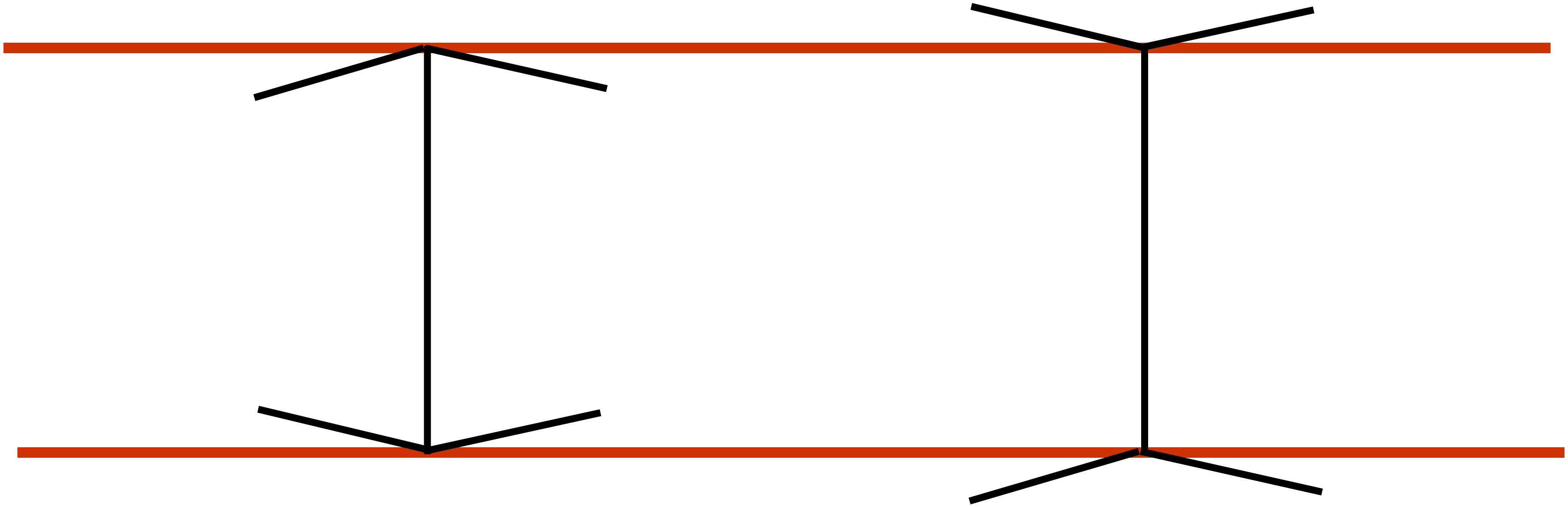
Ponzo's illusion

Linear Perspective



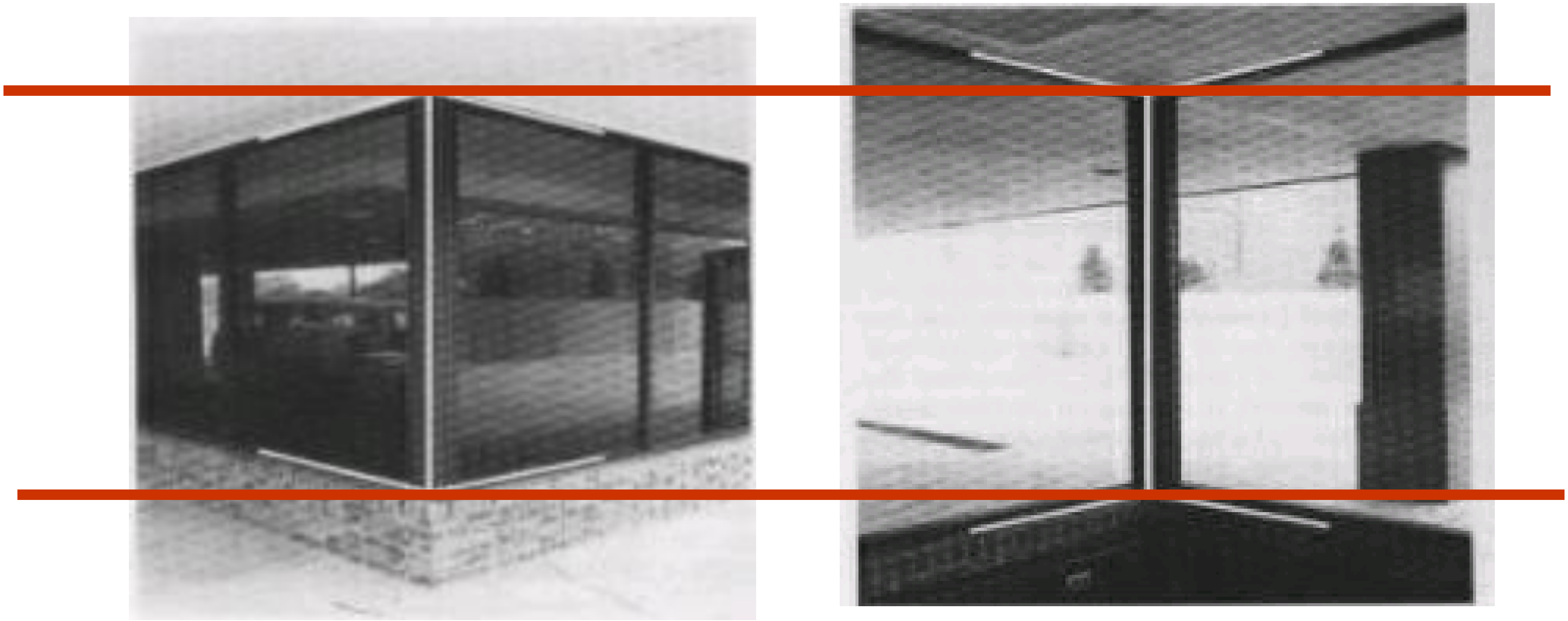
(c) 2006 Walt Anthony

Linear Perspective



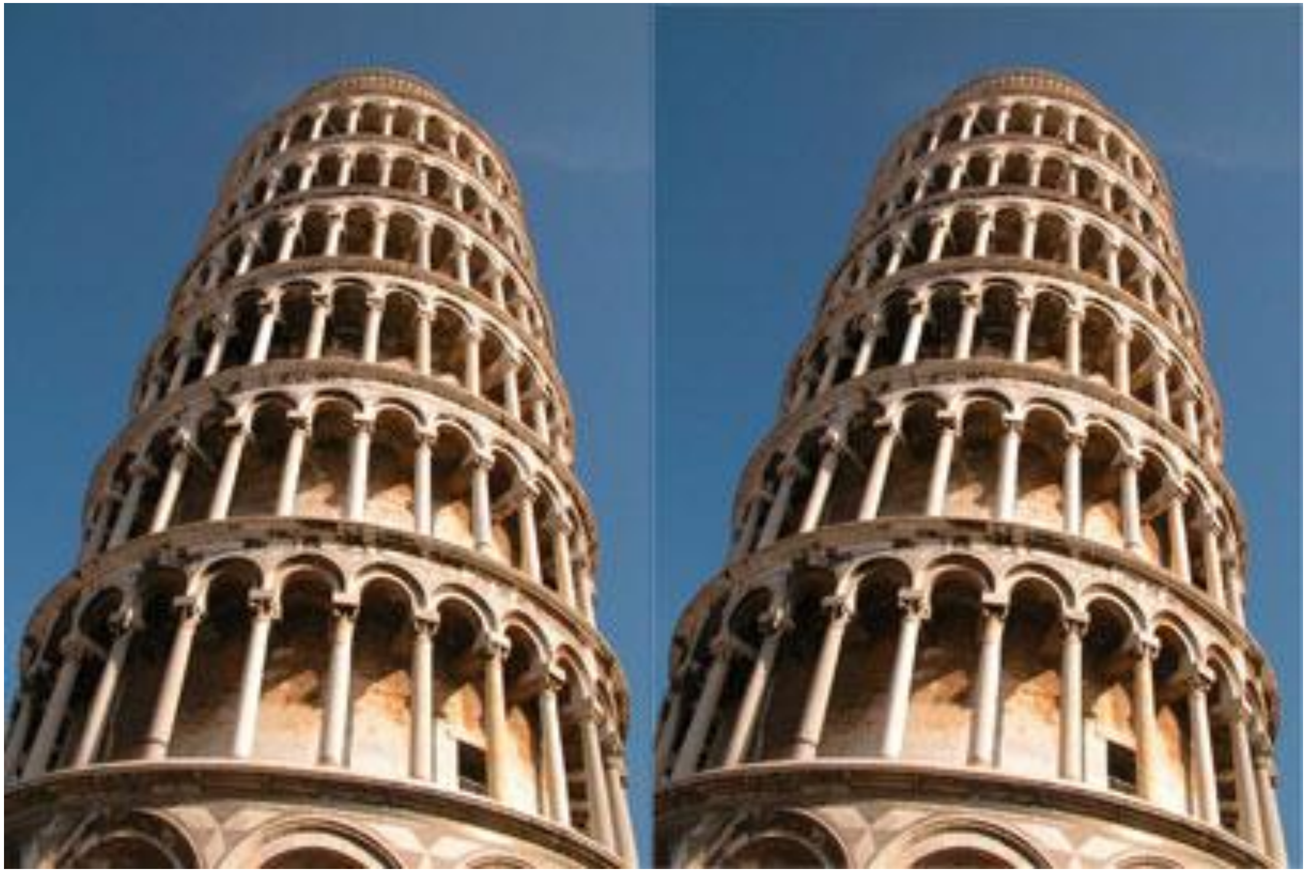
Muller-Lyer
1889

Linear Perspective



Muller-Lyer
1889

3D drives perception of important object attributes



The two Towers of Pisa

Frederick Kingdom, Ali Yoonessi and Elena Gheorghiu of McGill Vision Research unit.

True perspective

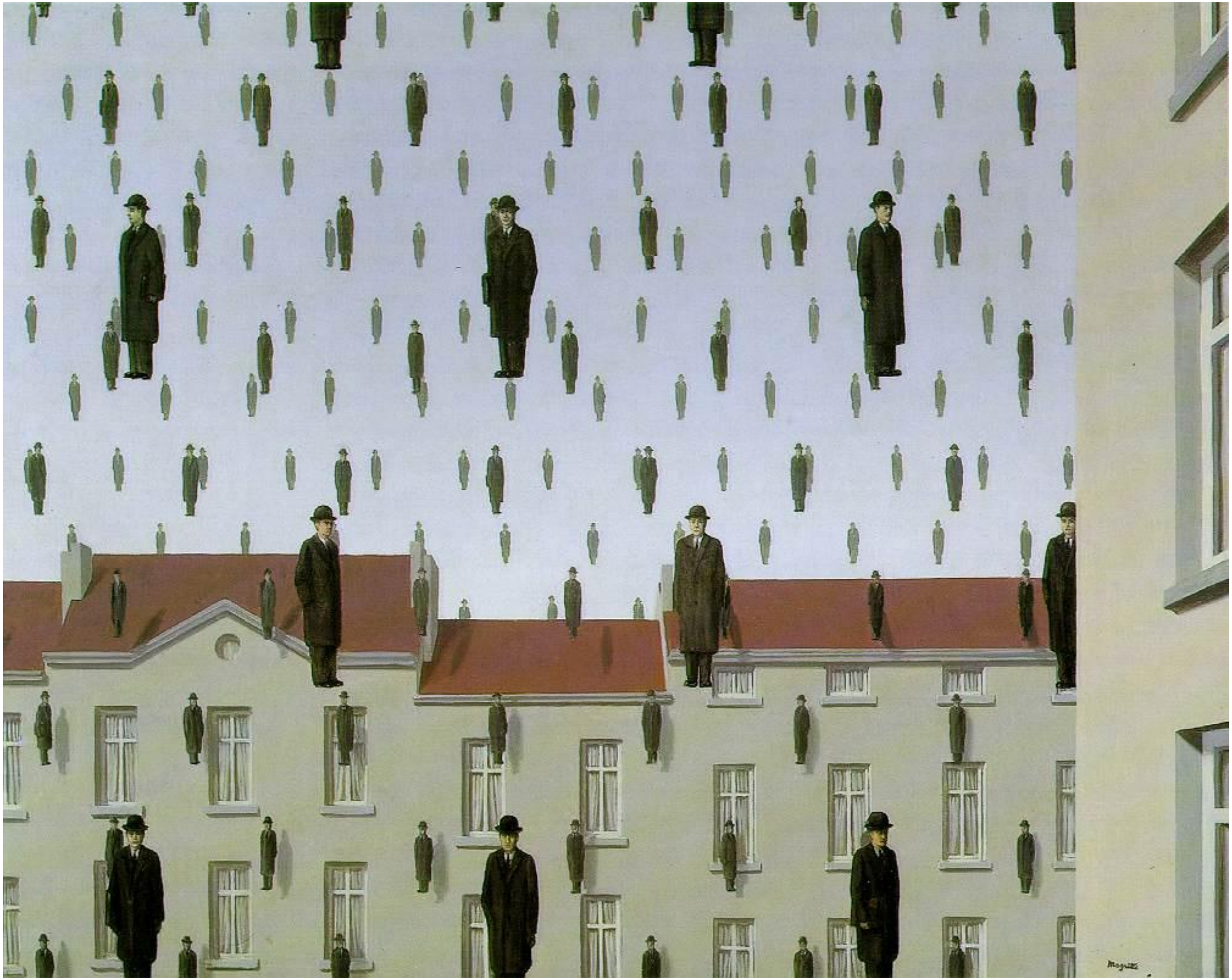
Expected perspective

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

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



Which cues do you see here?

[*Golconde* Rene Magritte]

Questions?

Absolute (monocular) depth cues

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

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 ?

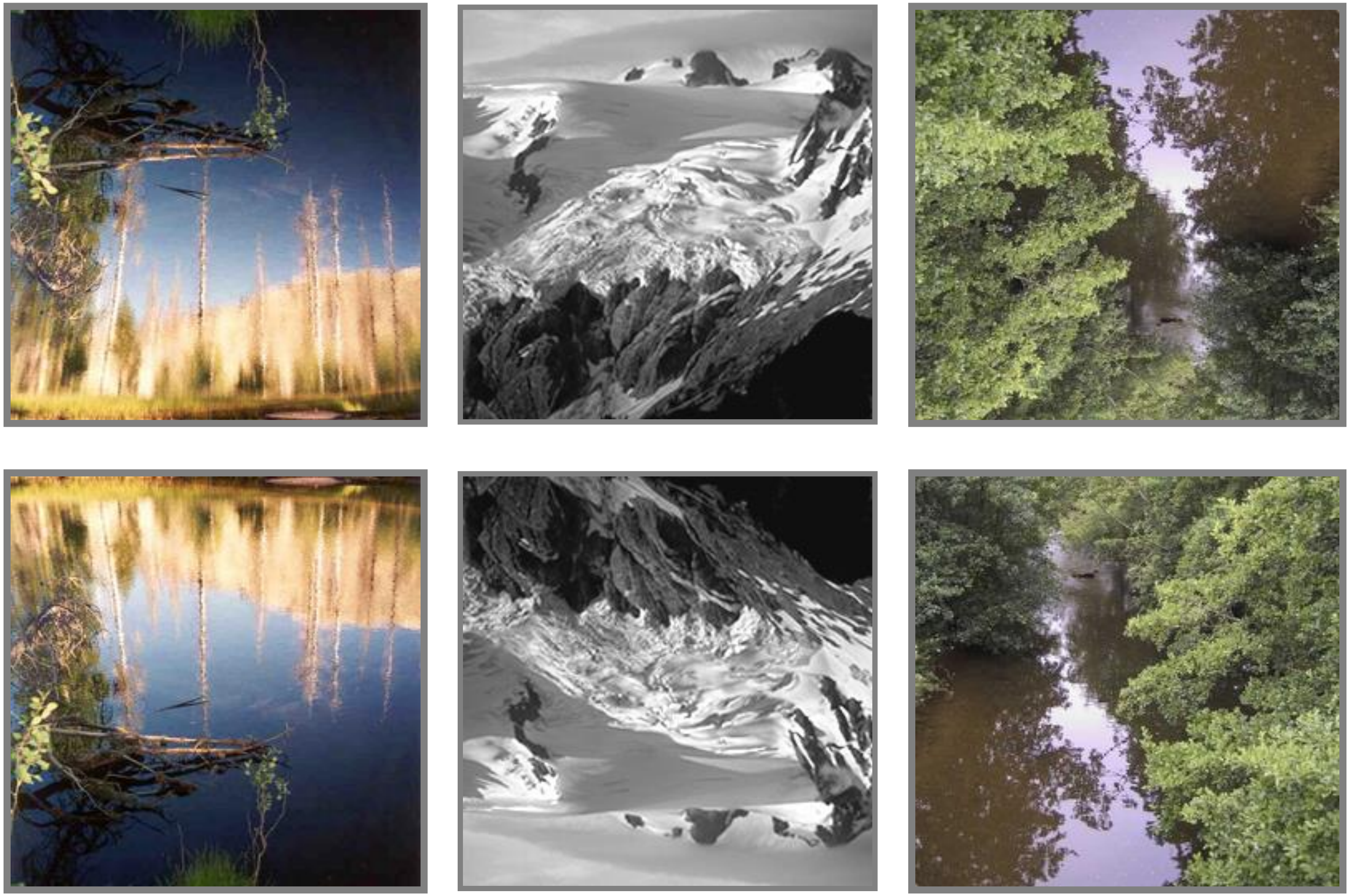
I’m going to quickly show 2 images

Which one is closer?

Slide by Aude Oliva

Which one is closer?

Textured surface layout influences depth perception



Learned that a specific distribution of features is correlated with a volume.
The interpretation of objects is different: sky and water, reflection and trees, branches, rocks

Torralba & Oliva (2002, 2003)

Slide by Aude Oliva

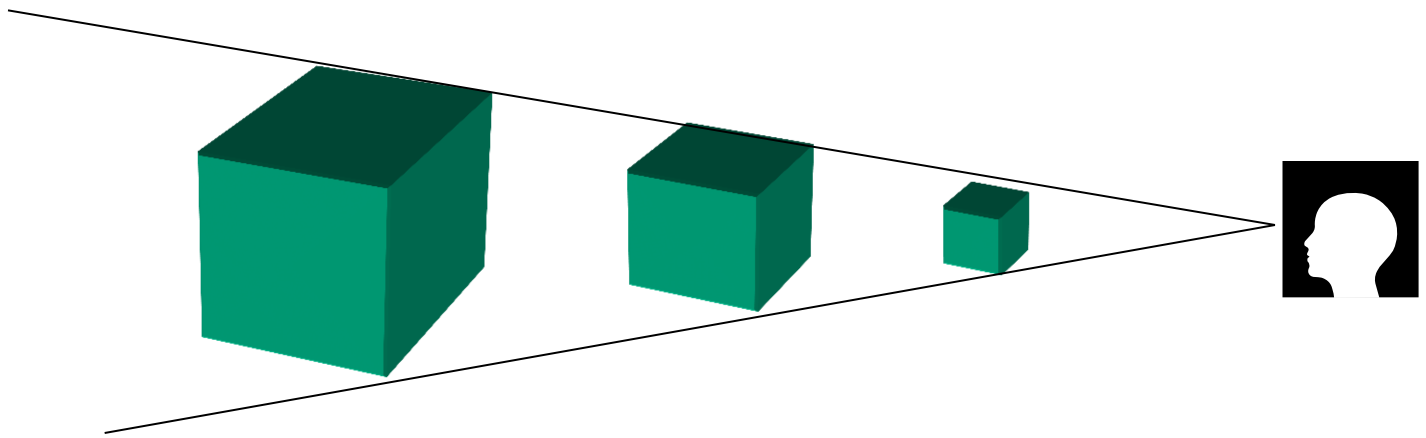
Depth Perception from Image Structure

We got wrong:

- 3D shape (mainly due to assumption of light from above)
- The absolute scale (due to the wrong recognition).

Depth Perception from Image Structure

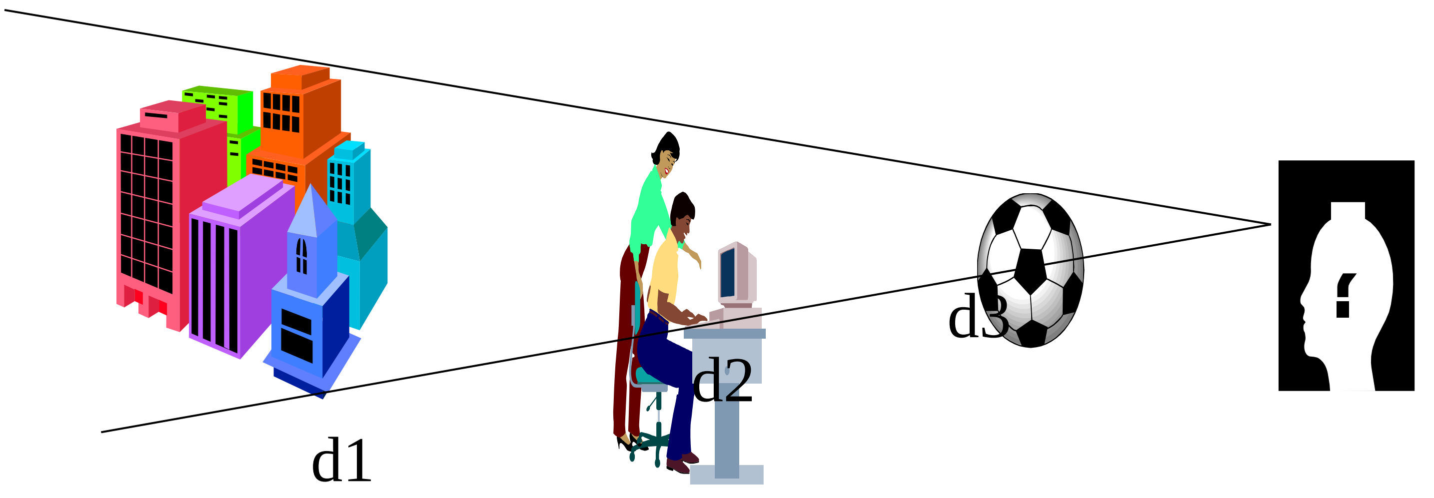
Mean depth refers to a global measurement of the mean distance between the observer and the main objects and structures that compose the scene.



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.

Depth Perception from Image Structure

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



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

Statistical Regularities of Scene Volume

When increasing the size of the space, natural environment structures become larger and smoother.

For man-made environments, the clutter of the scene increases with increasing distance: close-up views on objects have large and homogeneous regions. When increasing the size of the space, the scene “surface” breaks down in smaller pieces (objects, walls, windows, etc).

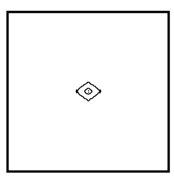
Torralba & Oliva. (2002). Depth estimation from image structure. IEEE Pattern Analysis and Machine Intelligence

Slide by Aude Oliva

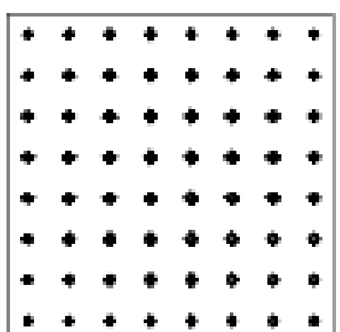
Image Statistics and Scene Scale

Close-up views

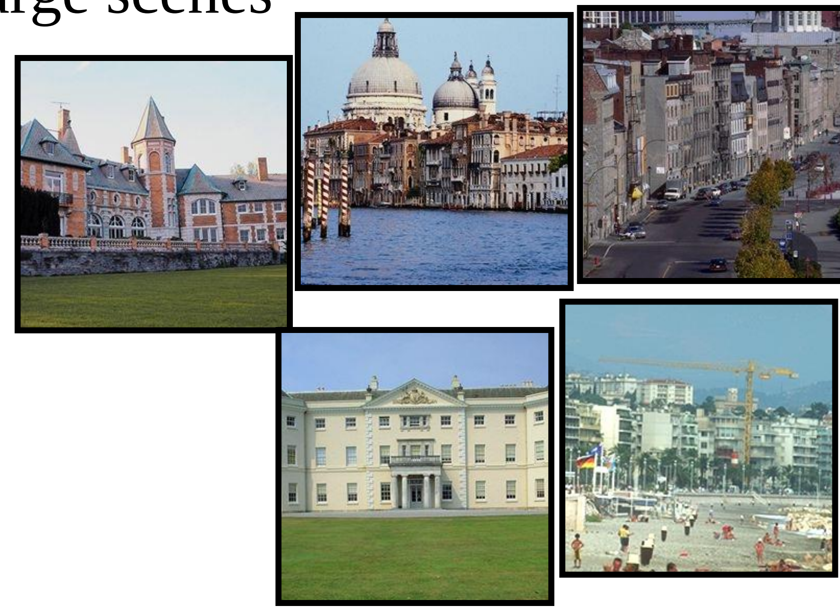
On average, low clutter



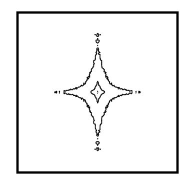
Point view is unconstrained



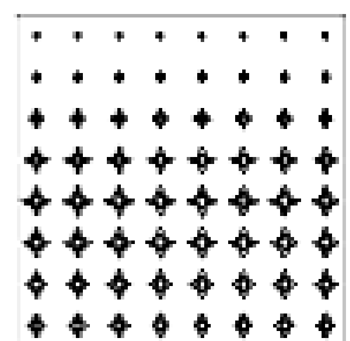
Large scenes



On average, highly cluttered



Point view is strongly constrained



It is not all about objects

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

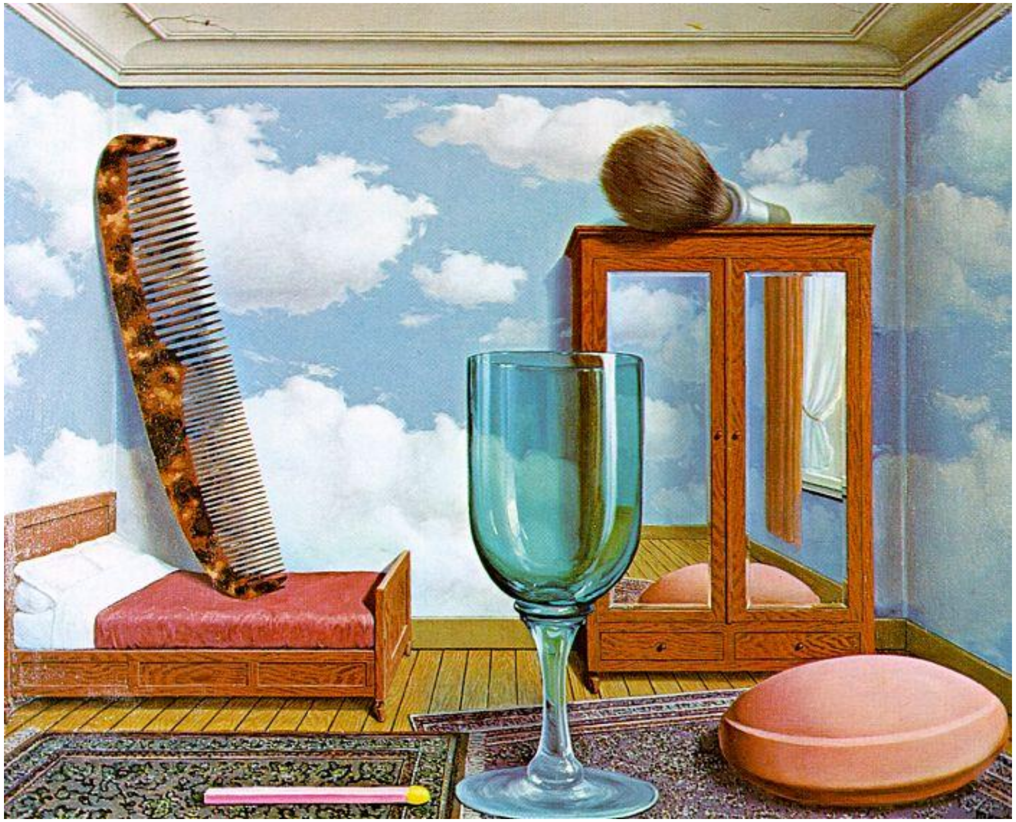
Scene vs. objects

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?

[*The Listening Room* Rene Magritte]

Scene vs. objects



[*Personal Values* Rene Magritte]

Depth without objects

Random dot stereograms (Bela Julesz)

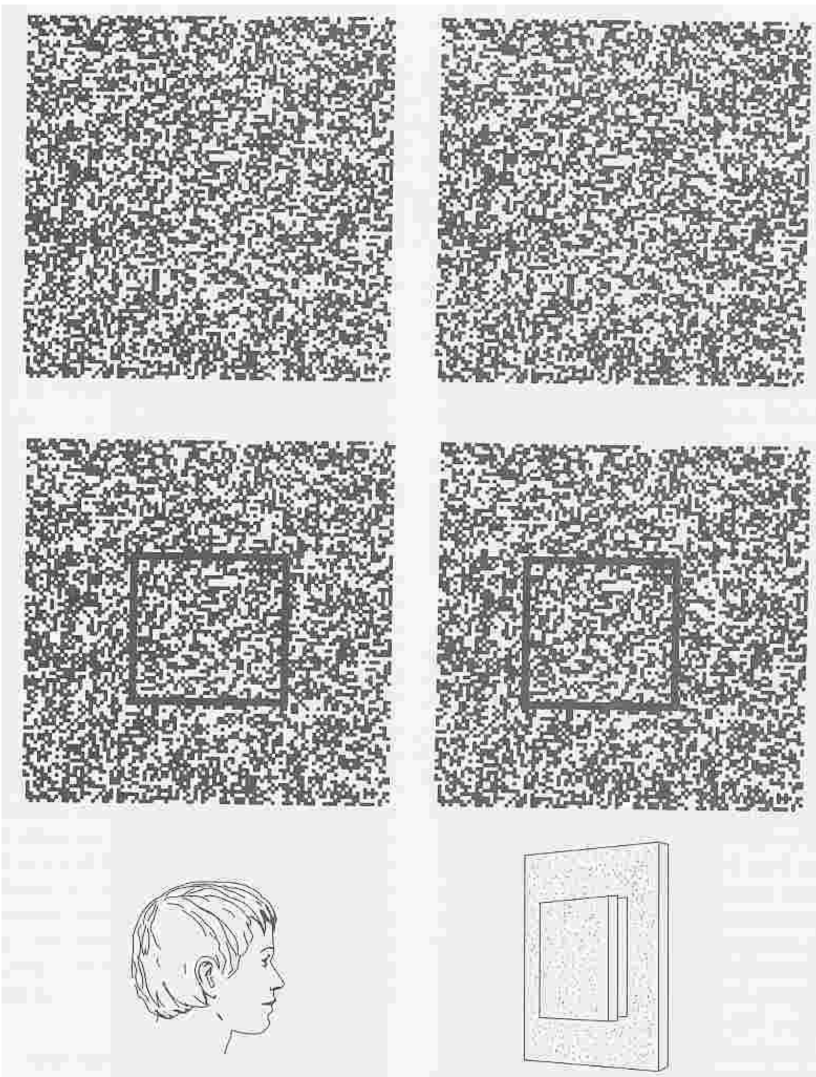


FIGURE 8.13

1	0	1	0	1	0	0	1	0	1
1	0	0	1	0	1	0	1	0	0
0	0	1	1	0	1	1	0	1	0
0	1	0	Y	A	A	B	B	0	1
1	1	1	X	B	A	B	A	0	1
0	0	1	X	A	A	B	A	1	0
1	1	1	Y	B	B	A	B	0	1
1	0	0	1	1	0	1	1	0	1
1	1	0	0	1	1	0	1	1	1
0	1	0	0	0	1	1	1	1	0

1	0	1	0	1	0	0	1	0	1
1	0	0	1	0	1	0	1	0	0
0	0	1	1	0	1	1	0	1	0
0	1	0	A	A	B	B	X	0	1
1	1	1	B	A	B	A	Y	0	1
0	0	1	A	A	B	A	Y	1	0
1	1	1	B	B	A	B	X	0	1
1	0	0	1	1	0	1	1	0	1
1	1	0	0	1	1	0	1	1	1
0	1	0	0	0	1	1	1	1	0

Julesz, 1971

Select a region in one image.

Shift this region horizontally by a small amount.

Stereo photography and stereo viewers

Take two pictures of the same subject from two slightly different viewpoints and display so that each eye sees only one of the images.

Invented by Sir Charles Wheatstone, 1838

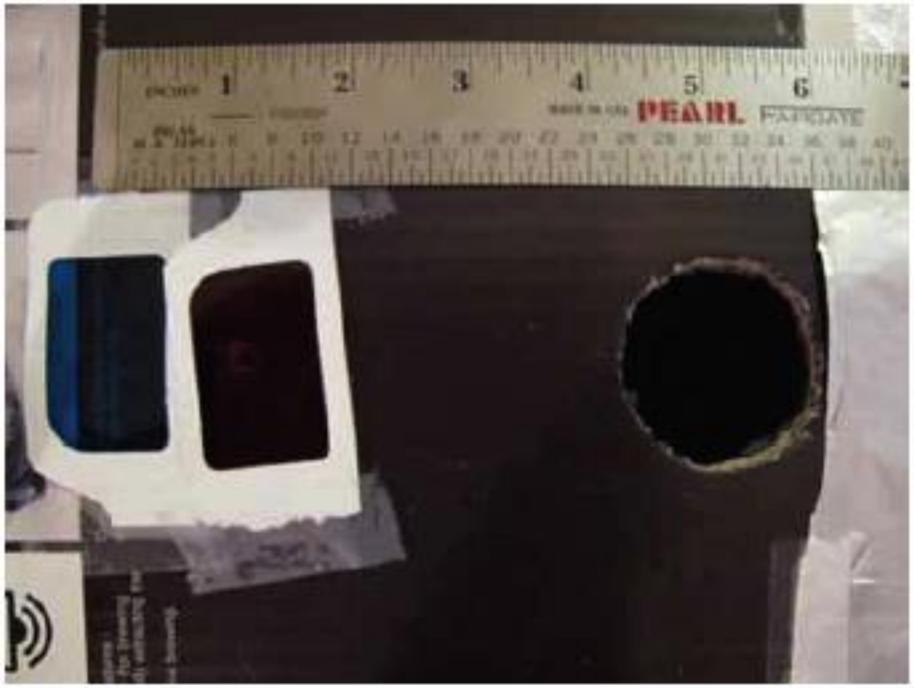
Image courtesy of fisher-price.com

Slide credit: Kristen Grauman

Public Library, Stereoscopic Looking Room, Chicago, by Phillips, 1923

de credit: Kristen Grauman

Anaglyph pinhole camera



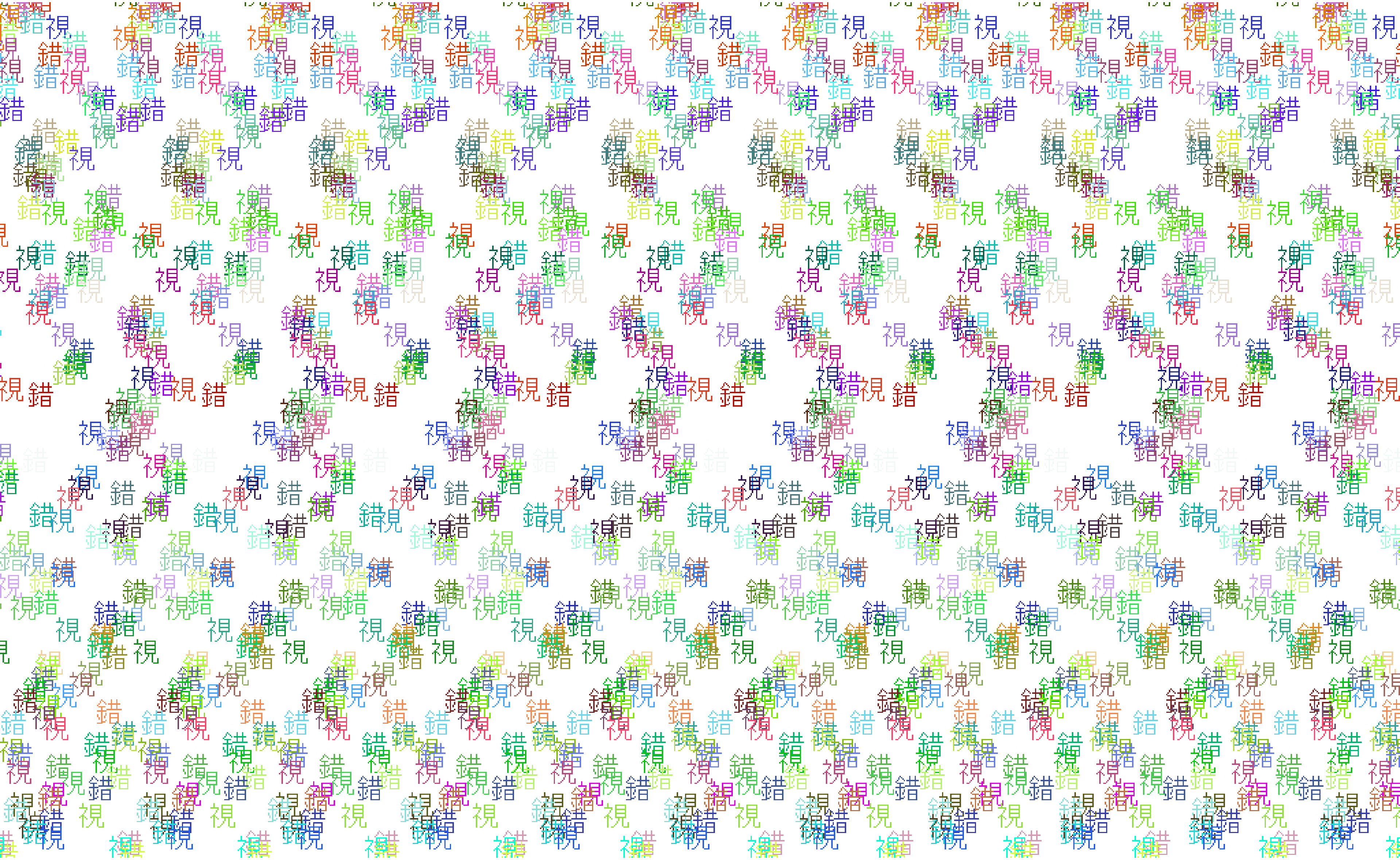
Autostereograms

Exploit disparity as depth cue using single image.

(Single image random dot stereogram, Single image stereogram)

Images from magiceye.com

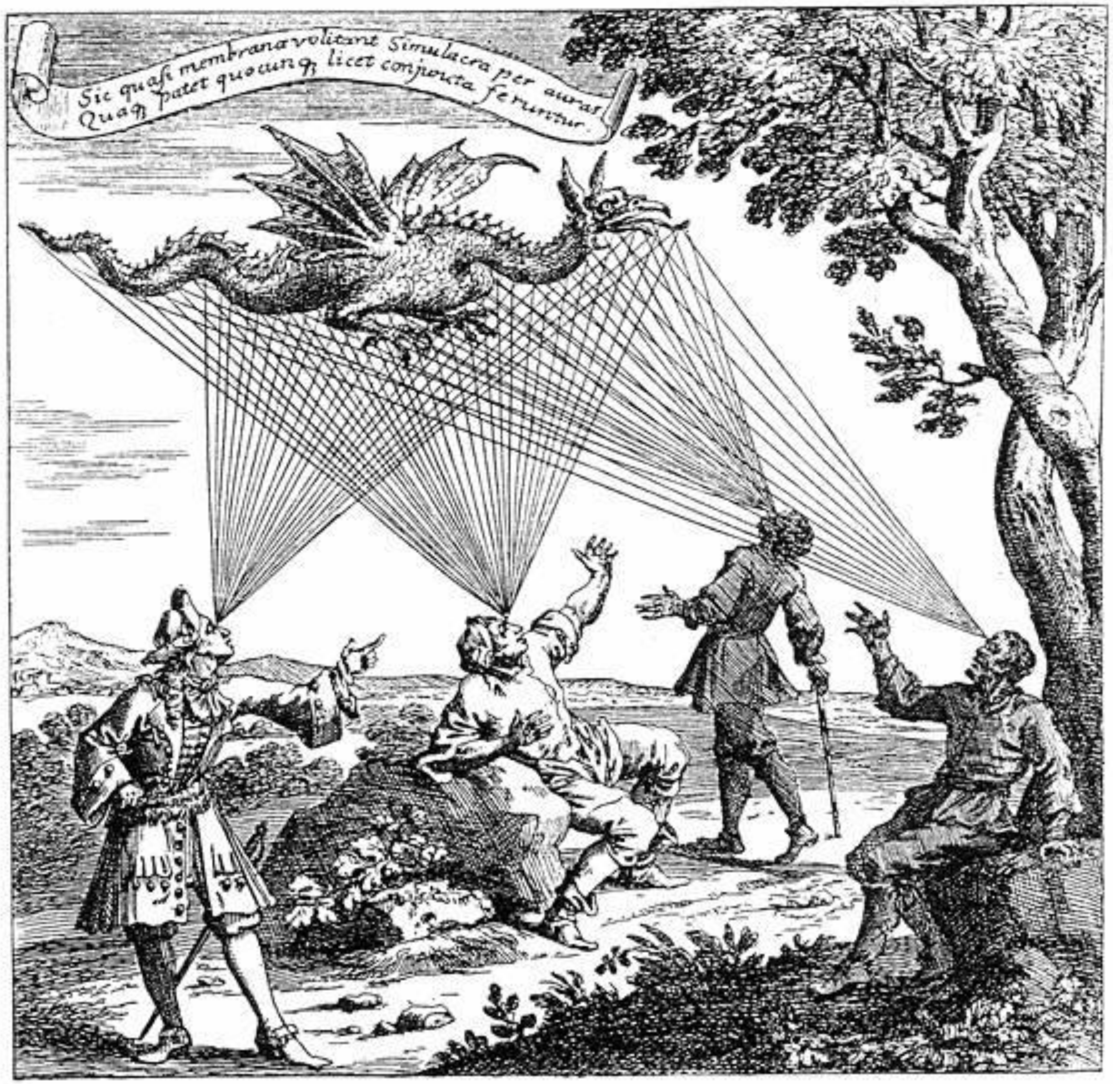
Slide credit: Kristen Grauman



http://www.psy.ritsumei.ac.jp/~akitaoka/stereo3e.html

Questions?

Structure from motion



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

Multiple-view geometry questions

- **Scene geometry (structure):** Given 2D point matches in two or more images, where are the corresponding points in 3D?

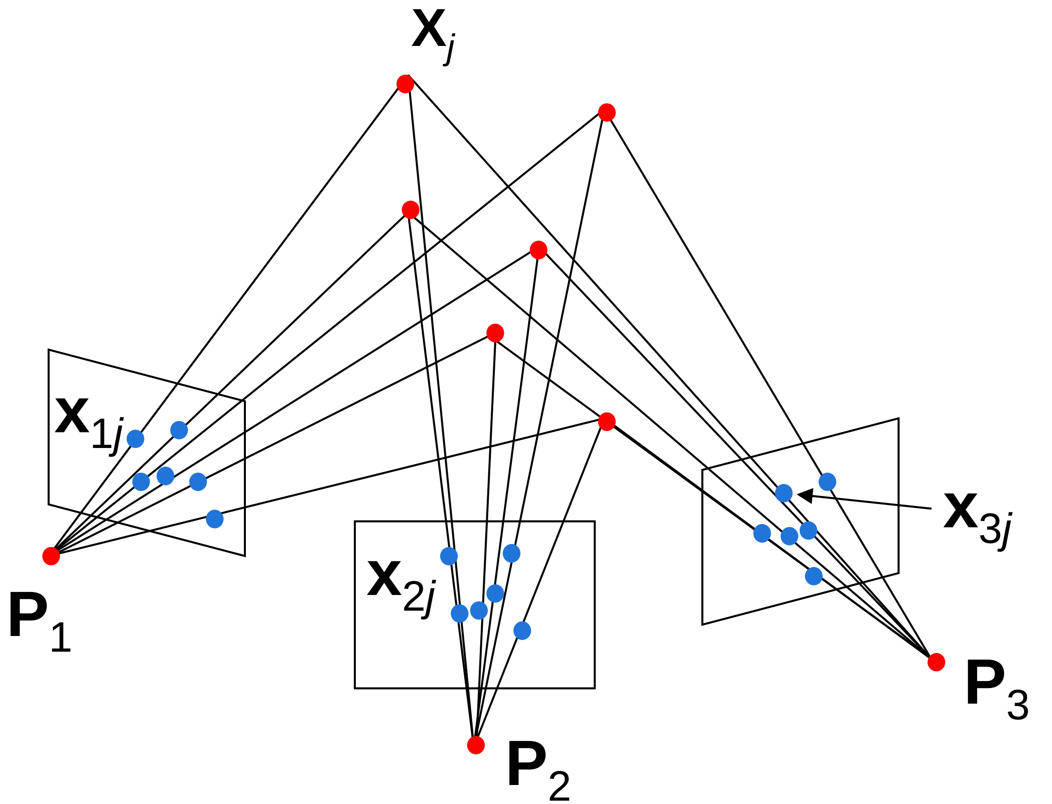
- **Correspondence (stereo matching):** Given a point in just one image, how does it constrain the position of the corresponding point in another image?

- **Camera geometry (motion):** Given a set of corresponding points in two or more images, what are the camera matrices for these views?

Slide: S. Lazebnik

Structure from motion

- Given: m images of n fixed 3D points
  - $\mathbf{x}_{ij} = \mathbf{P}_i \mathbf{X}_j\,,\ \ i = 1, \ldots, m, \quad j = 1, \ldots, n$
- Problem: estimate m projection matrices $\mathbf{P}_i$ and n 3D points $\mathbf{X}_j$ from the mn correspondences $\mathbf{x}_{ij}$



Slide: S. Lazebnik

Structure from motion ambiguity

- If we scale the entire scene by some factor *k* and, at the same time, scale the camera matrices by the factor of 1/*k*, the projections of the scene points in the image remain exactly the same:

$$\mathbf{x} = \mathbf{P}\mathbf{X} = \left(\frac{1}{k}\mathbf{P}\right)(k\mathbf{X})$$

It is impossible to recover the absolute scale of the scene!

Does this only hold for scaling?

Slide: S. Lazebnik

Structure from motion ambiguity

- If we scale the entire scene by some factor *k* and, at the same time, scale the camera matrices by the factor of 1/*k*, the projections of the scene points in the image remain exactly the same

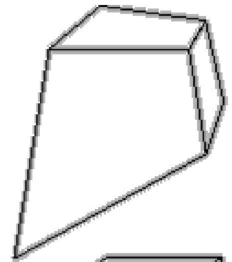
- More generally: if we transform the scene using a transformation **Q** and apply the inverse transformation to the camera matrices, then the images do not change

$$\mathbf{x} = \mathbf{P}\mathbf{X} = \left(\mathbf{P}\mathbf{Q}^{-1}\right)\left(\mathbf{Q}\mathbf{X}\right)$$

Types of ambiguity

Projective
15dof

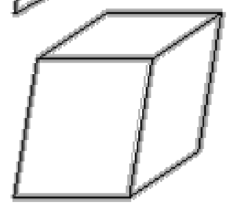
$$\begin{bmatrix} A & t \\ v^T & v \end{bmatrix}$$



Preserves intersection and tangency

Affine
12dof

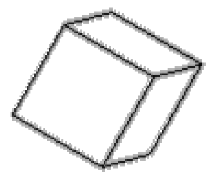
$$\begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix}$$



Preserves parallellism, volume ratios

Similarity
7dof

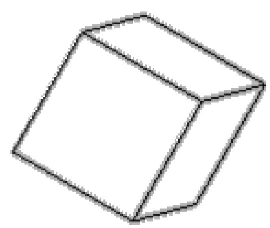
$$\begin{bmatrix} s\,R & t \\ 0^T & 1 \end{bmatrix}$$



Preserves angles, ratios of length

Euclidean
6dof

$$\begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$$

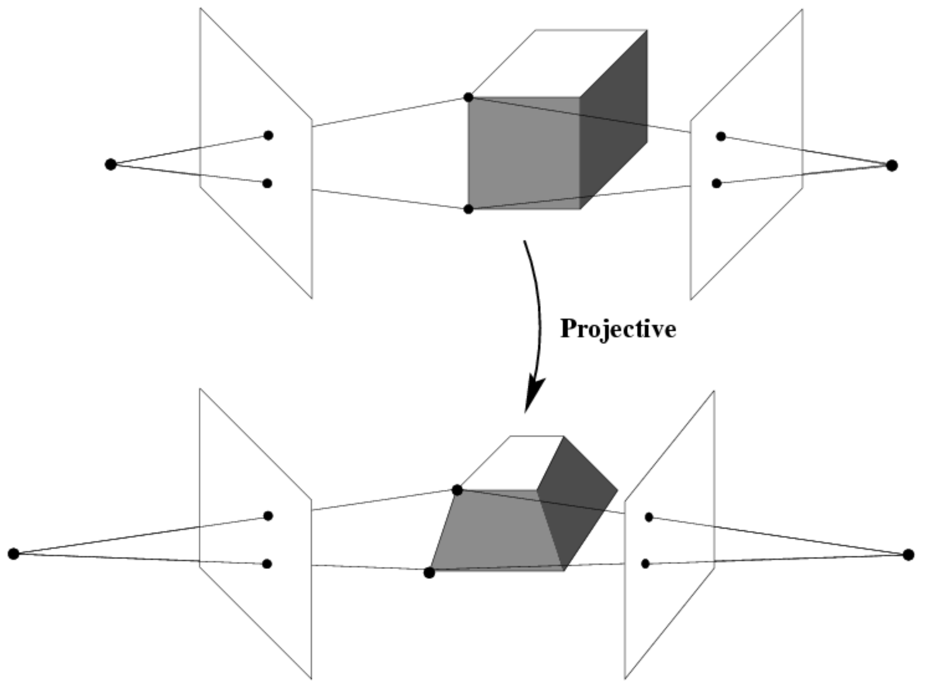


Preserves angles, lengths

- With no constraints on the camera matrix or on the scene, we get a *projective* reconstruction
- Need additional information to *upgrade* the reconstruction to affine, similarity, or Euclidean

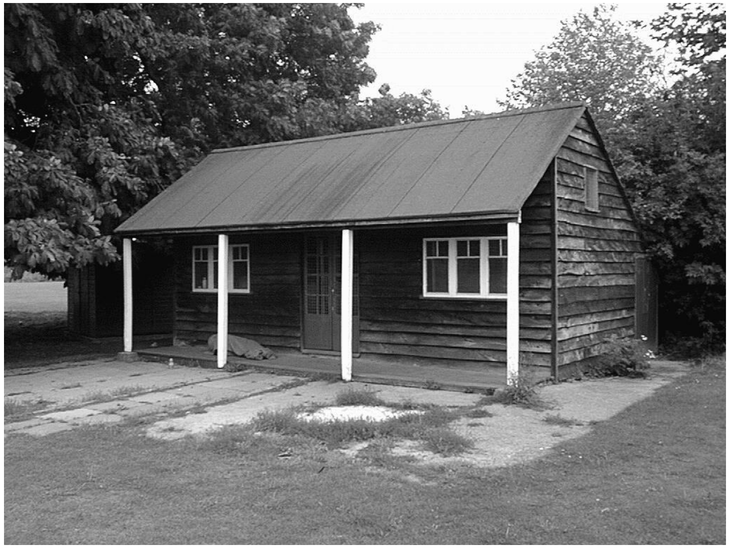
Slide: S. Lazebnik

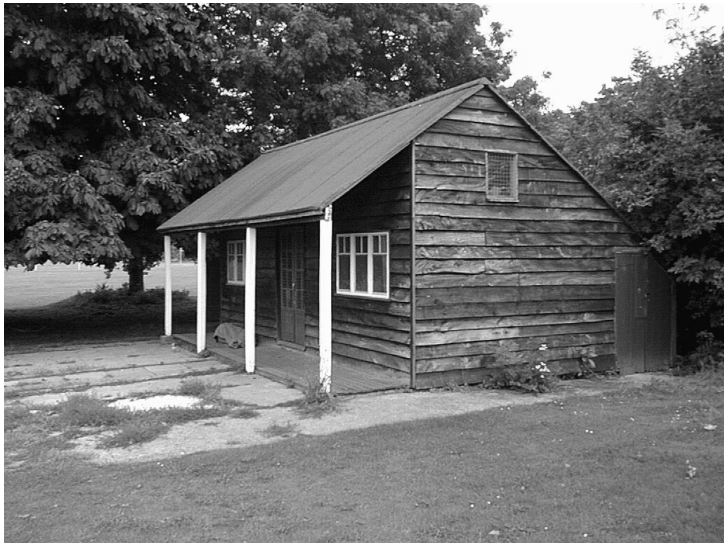
Projective ambiguity

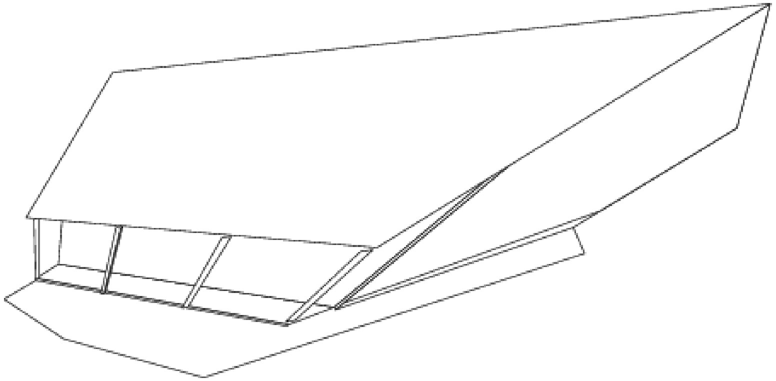


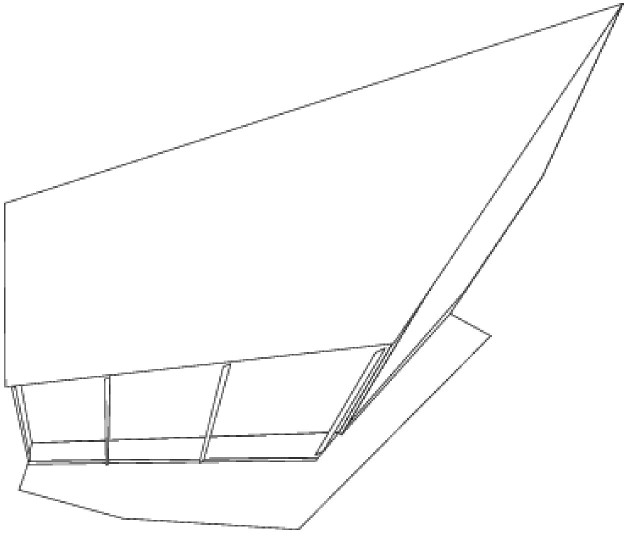
$$\mathbf{x} = \mathbf{P}\mathbf{X} = \left(\mathbf{P}\mathbf{Q}_{\mathbf{P}}^{-1}\right)\left(\mathbf{Q}_{\mathbf{P}}\,\mathbf{X}\right)$$

Projective ambiguity

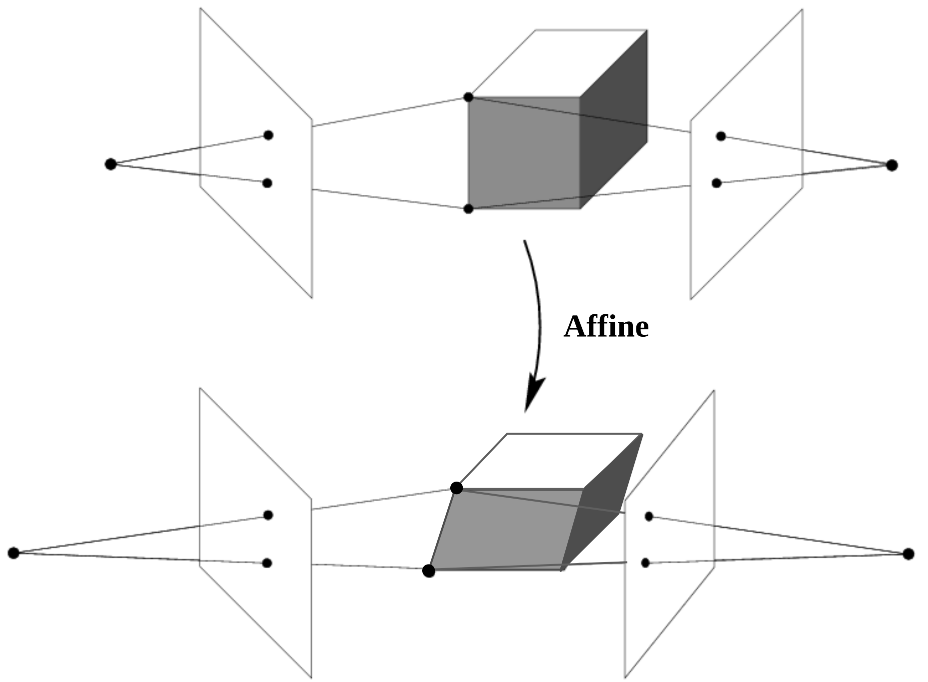






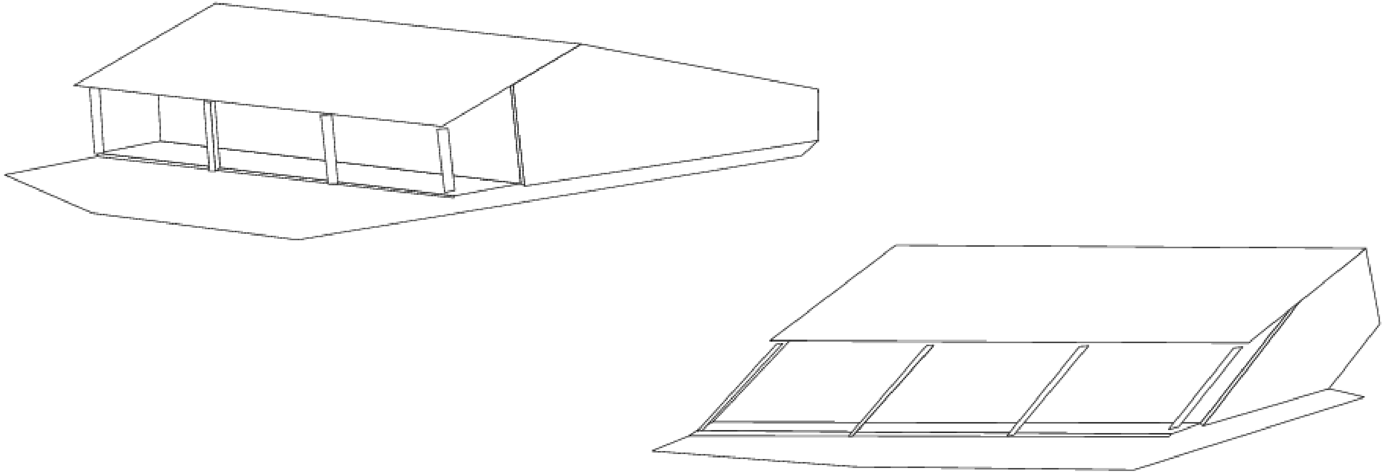


Affine ambiguity

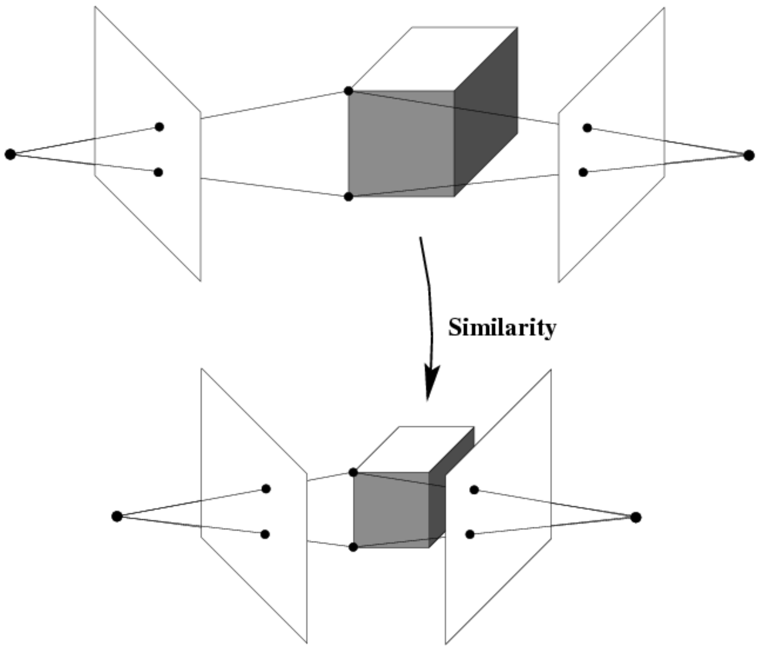


$$\mathbf{x} = \mathbf{P}\mathbf{X} = \left(\mathbf{P}\mathbf{Q}_{\mathrm{A}}^{-1}\right)\left(\mathbf{Q}_{\mathrm{A}}\,\mathbf{X}\right)$$

Affine ambiguity

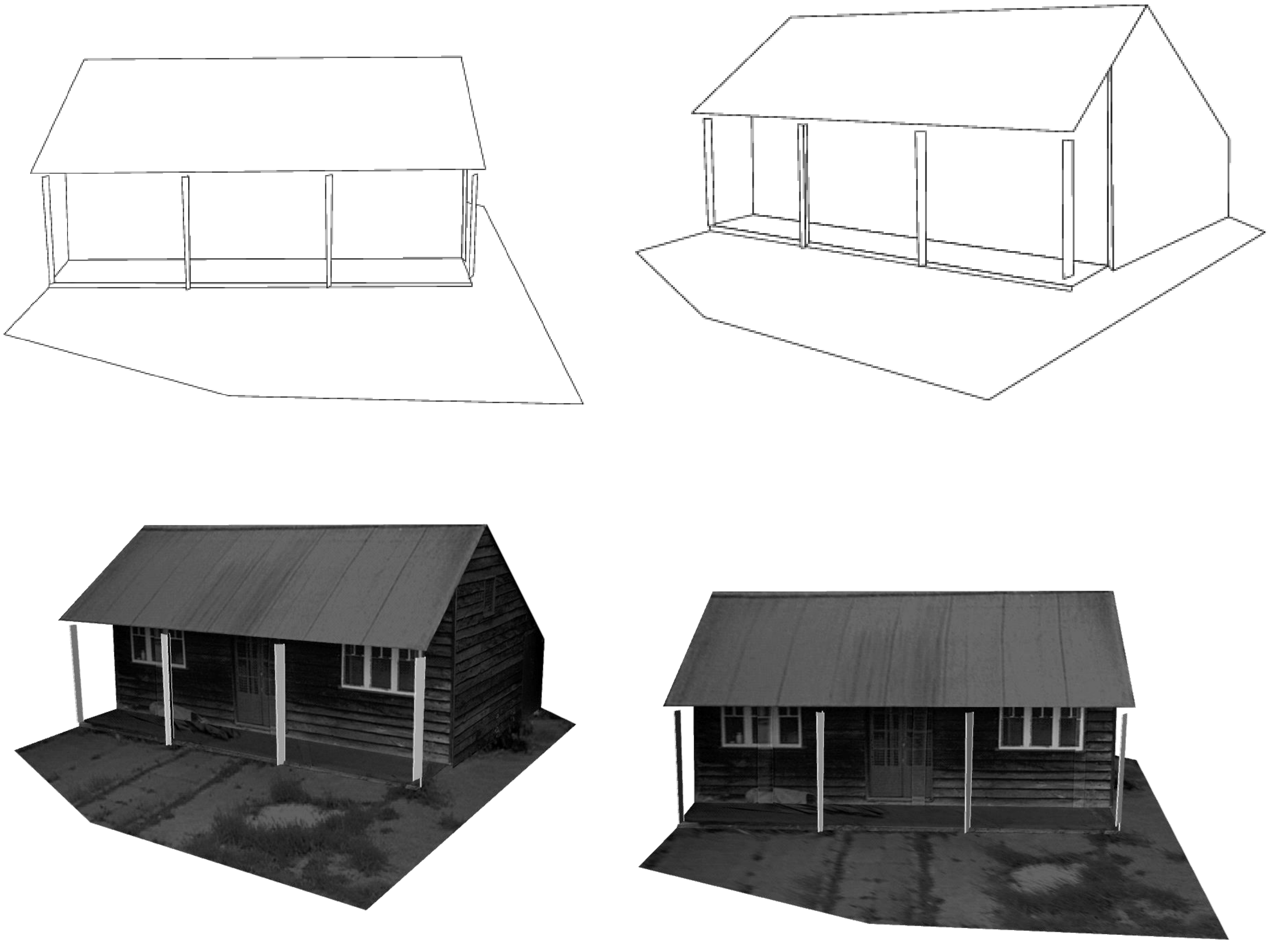


Similarity ambiguity

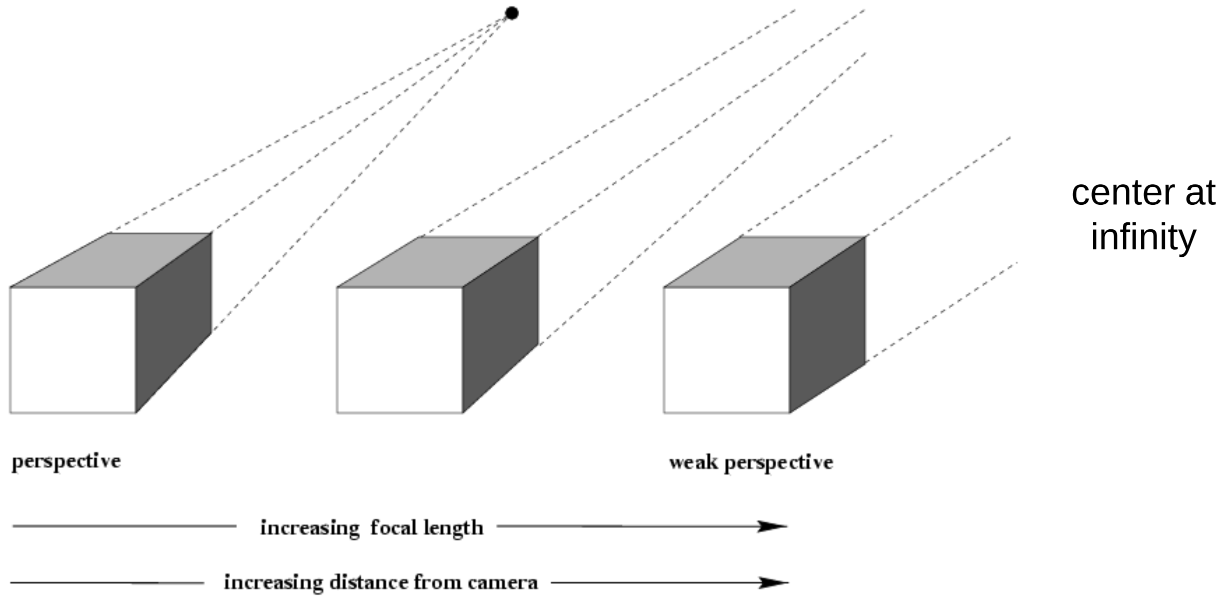


$$\mathbf{x} = \mathbf{P}\mathbf{X} = \left(\mathbf{P}\mathbf{Q}_S^{-1}\right)\left(\mathbf{Q}_S\mathbf{X}\right)$$

Similarity ambiguity

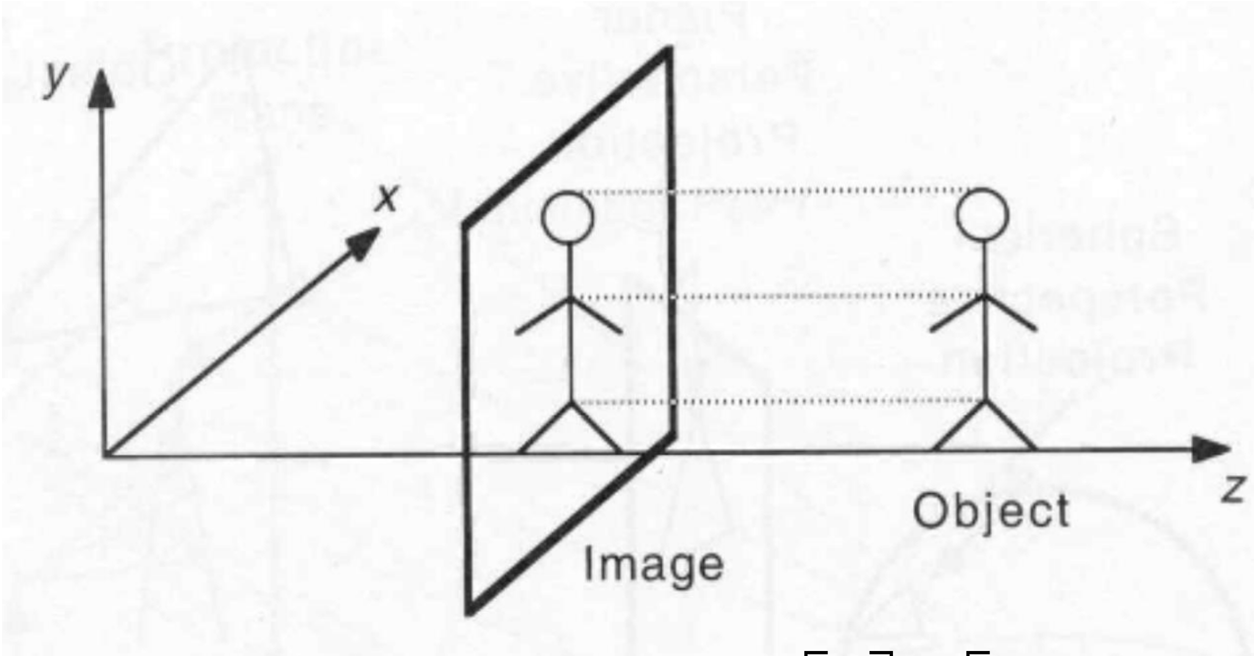


Affine Structure from motion



Recall: Orthographic Projection

- Special case of perspective projection
  - Distance from the COP to the image plane is ‘infinite’



  - Also called “parallel projection”
  - Projection matrix:

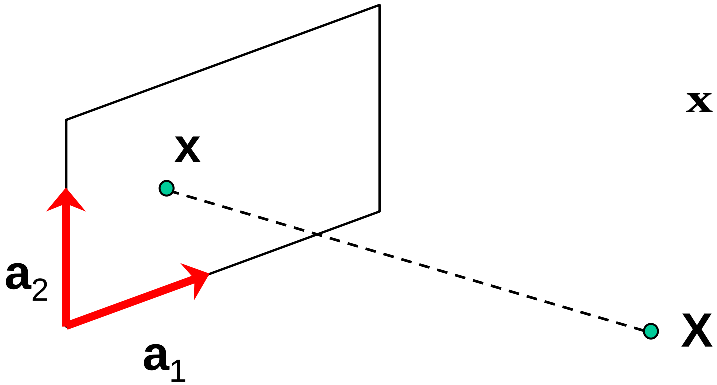
$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Affine cameras

- A general affine camera combines the effects of an affine transformation of the 3D space, orthographic projection, and an affine transformation of the image:

$$\mathbf{P} = [3\times 3\,\text{affine}]\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}[4\times 4\,\text{affine}] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$$

Affine projection is a linear mapping + translation in inhomogeneous coordinates



$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \mathbf{AX} + \mathbf{b}$$

Projection of world origin

Affine structure from motion

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \mathbf{AX} + \mathbf{b}$$

- Centering: subtract the centroid of the image points (removes translation)

$$\hat{\mathbf{x}}_{ij} = \mathbf{x}_{ij} - \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_{ik} = \mathbf{A}_i\mathbf{X}_j + \mathbf{b}_i - \frac{1}{n}\sum_{k=1}^{n}(\mathbf{A}_i\mathbf{X}_k + \mathbf{b}_i)$$

$$= \mathbf{A}_i\left(\mathbf{X}_j - \frac{1}{n}\sum_{k=1}^{n}\mathbf{X}_k\right) = \mathbf{A}_i\hat{\mathbf{X}}_j$$

For simplicity, assume that the origin of the world coordinate system is at the centroid of the 3D points

- After centering, each normalized point $\mathbf{x}_{ij}$ is related to the 3D point $\mathbf{X}_i$ by

$$\hat{\mathbf{x}}_{ij} = \mathbf{A}_i\mathbf{X}_j$$

Affine structure from motion

- Let's create a $2m \times n$ data (measurement) matrix:

$$\begin{array}{cc} \mathbf{D} = \begin{bmatrix} \hat{\mathbf{x}}_{11} & \hat{\mathbf{x}}_{12} & \cdots & \hat{\mathbf{x}}_{1n} \\ \hat{\mathbf{x}}_{21} & \hat{\mathbf{x}}_{22} & \cdots & \hat{\mathbf{x}}_{2n} \\ & & \ddots & \\ \hat{\mathbf{x}}_{m1} & \hat{\mathbf{x}}_{m2} & \cdots & \hat{\mathbf{x}}_{mn} \end{bmatrix} & \Big\downarrow \text{cameras } (2m) \\ \xrightarrow{\text{points } (n)} & \end{array}$$

C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *IJCV*, 9(2):137-154, November 1992.

Affine structure from motion

- Let's create a $2m \times n$ data (measurement) matrix:

$$\mathbf{D} = \begin{bmatrix} \hat{\mathbf{x}}_{11} & \hat{\mathbf{x}}_{12} & \cdots & \hat{\mathbf{x}}_{1n} \\ \hat{\mathbf{x}}_{21} & \hat{\mathbf{x}}_{22} & \cdots & \hat{\mathbf{x}}_{2n} \\ & & \ddots & \\ \hat{\mathbf{x}}_{m1} & \hat{\mathbf{x}}_{m2} & \cdots & \hat{\mathbf{x}}_{mn} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_m \end{bmatrix}}_{\text{Cameras } (2m \times ?)} \underbrace{\begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_n \end{bmatrix}}_{\text{points } (? \times n)}$$

How many dimensions?

C. Tomasi and T. Kanade. [Shape and motion from image streams under orthography: A factorization method.](#) *IJCV*, 9(2):137-154, November 1992.

Affine structure from motion

- Let's create a $2m \times n$ data (measurement) matrix:

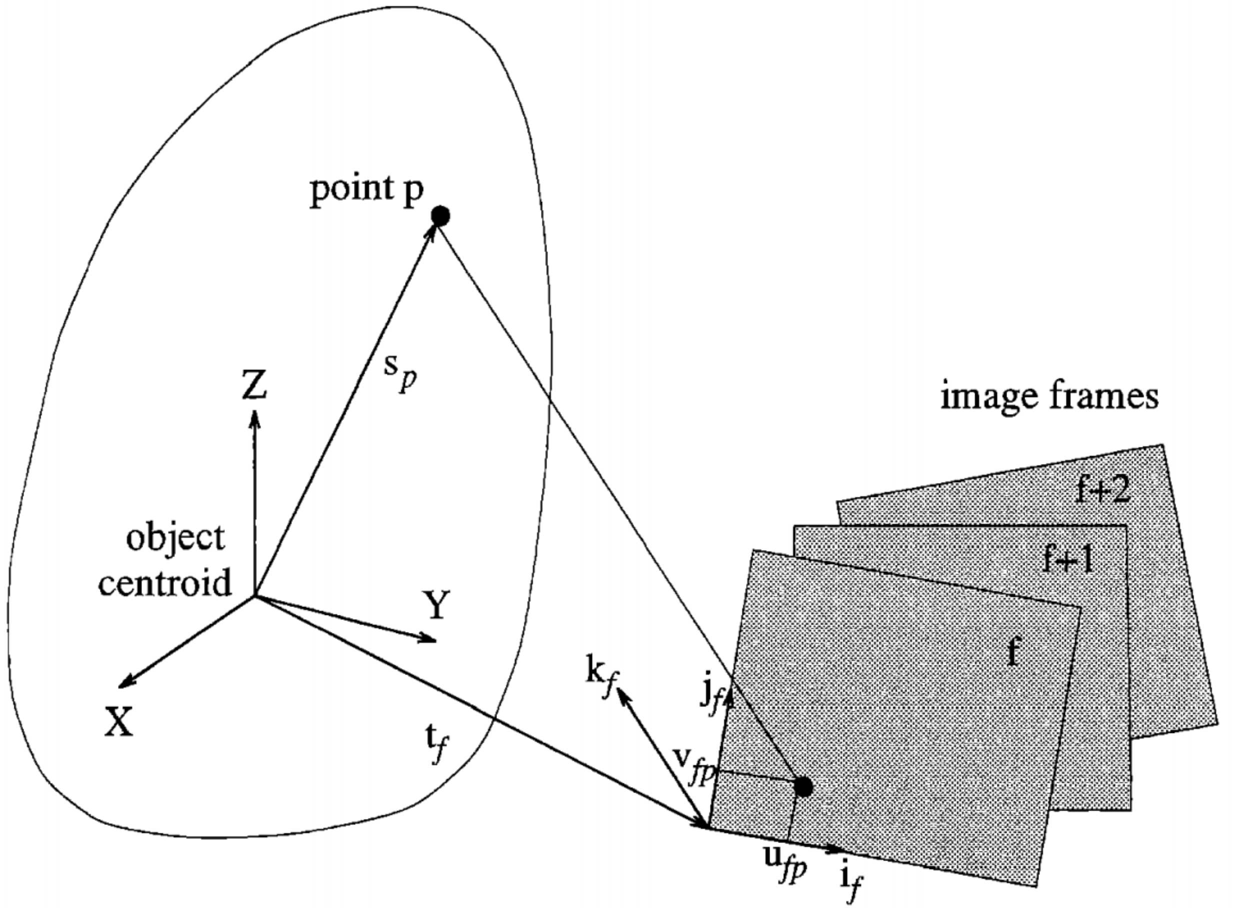
$$\mathbf{D} = \begin{bmatrix} \hat{\mathbf{x}}_{11} & \hat{\mathbf{x}}_{12} & \cdots & \hat{\mathbf{x}}_{1n} \\ \hat{\mathbf{x}}_{21} & \hat{\mathbf{x}}_{22} & \cdots & \hat{\mathbf{x}}_{2n} \\ & & \ddots & \\ \hat{\mathbf{x}}_{m1} & \hat{\mathbf{x}}_{m2} & \cdots & \hat{\mathbf{x}}_{mn} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_m \end{bmatrix}}_{\text{Cameras } (2m \times 3)} \underbrace{\begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_n \end{bmatrix}}_{\text{points } (3 \times n)}$$

The measurement matrix **D** = **MS** must have rank 3!

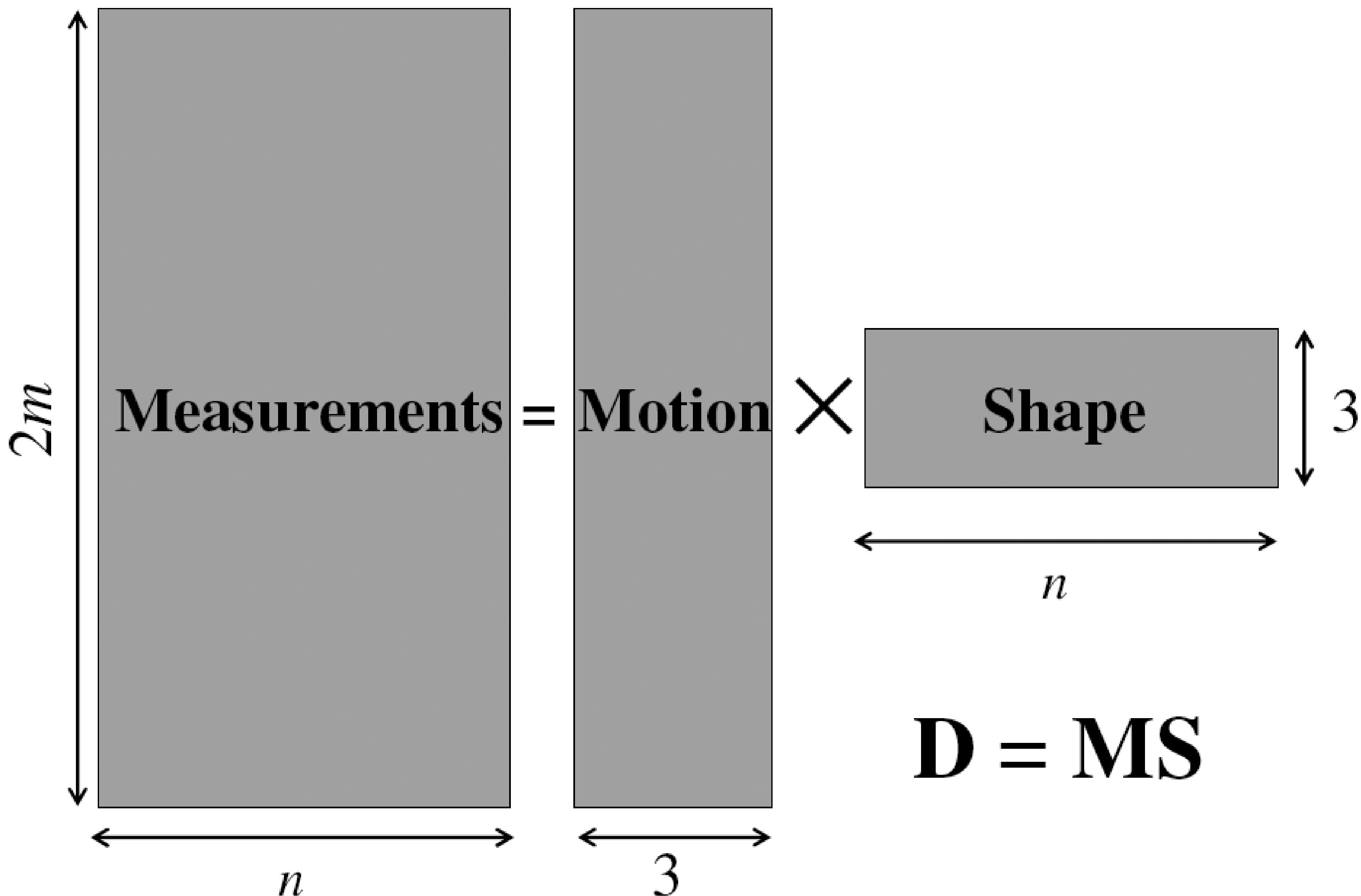
C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *IJCV*, 9(2):137-154, November 1992.

Affine structure from motion

Main idea: You only need M camera orientations and N points



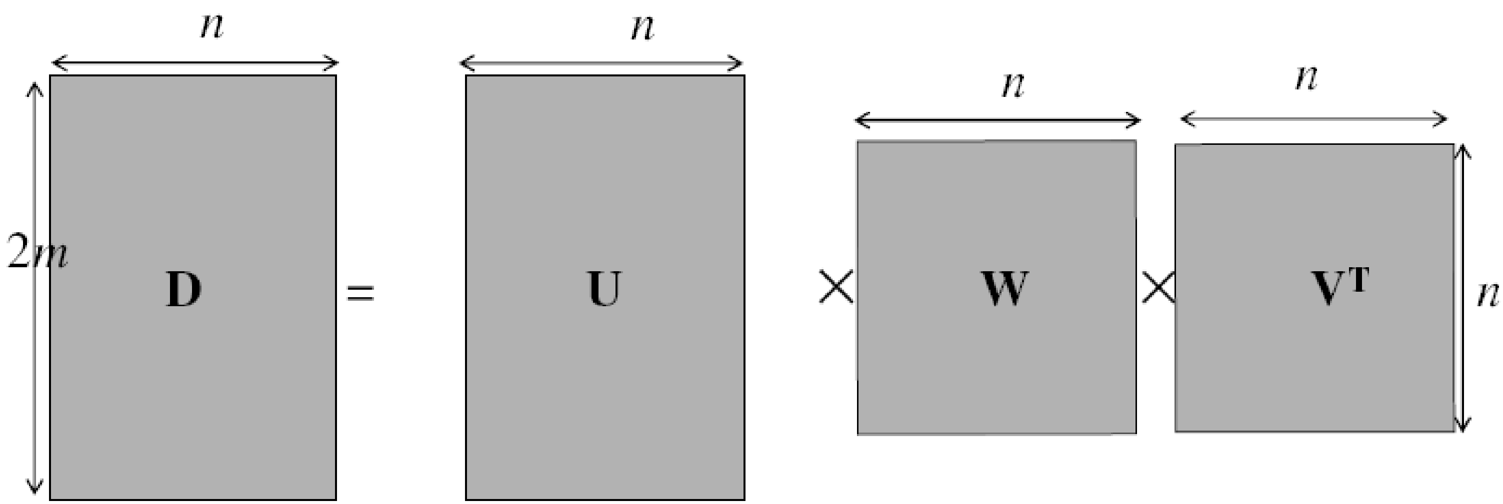
Factorizing the measurement matrix



Source: M. Hebert

Factorizing the measurement matrix

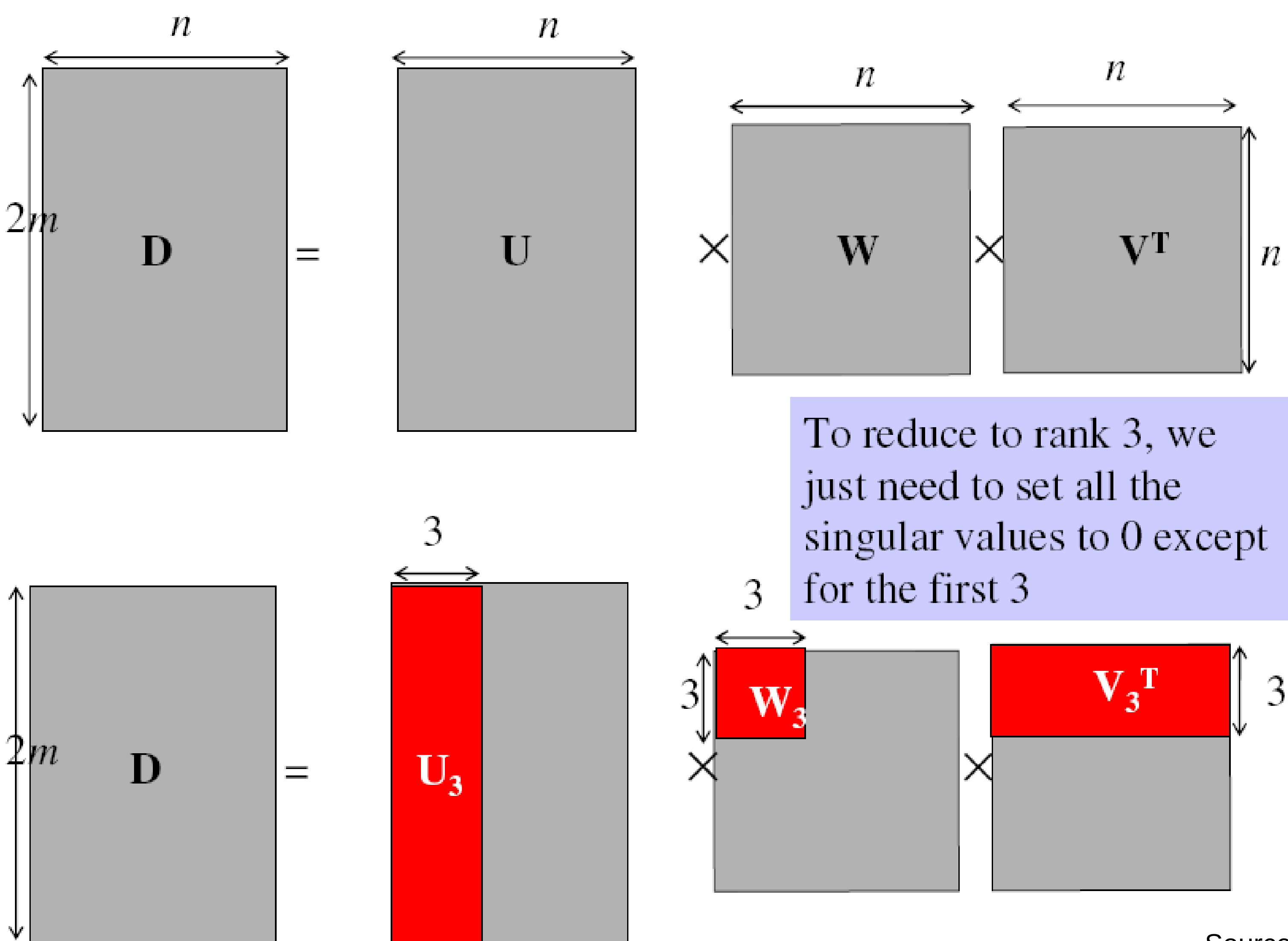
- Singular value decomposition of D:



How would you enforce the rank constraint?

Factorizing the measurement matrix

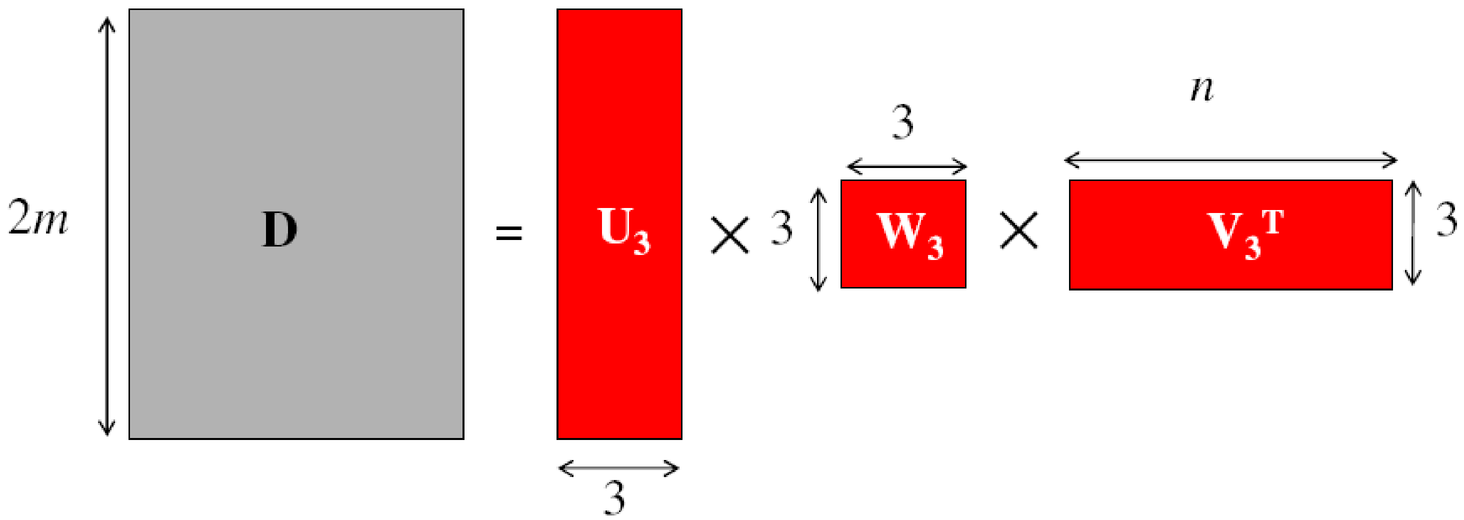
- Singular value decomposition of D:



Source: M. Hebert

Factorizing the measurement matrix

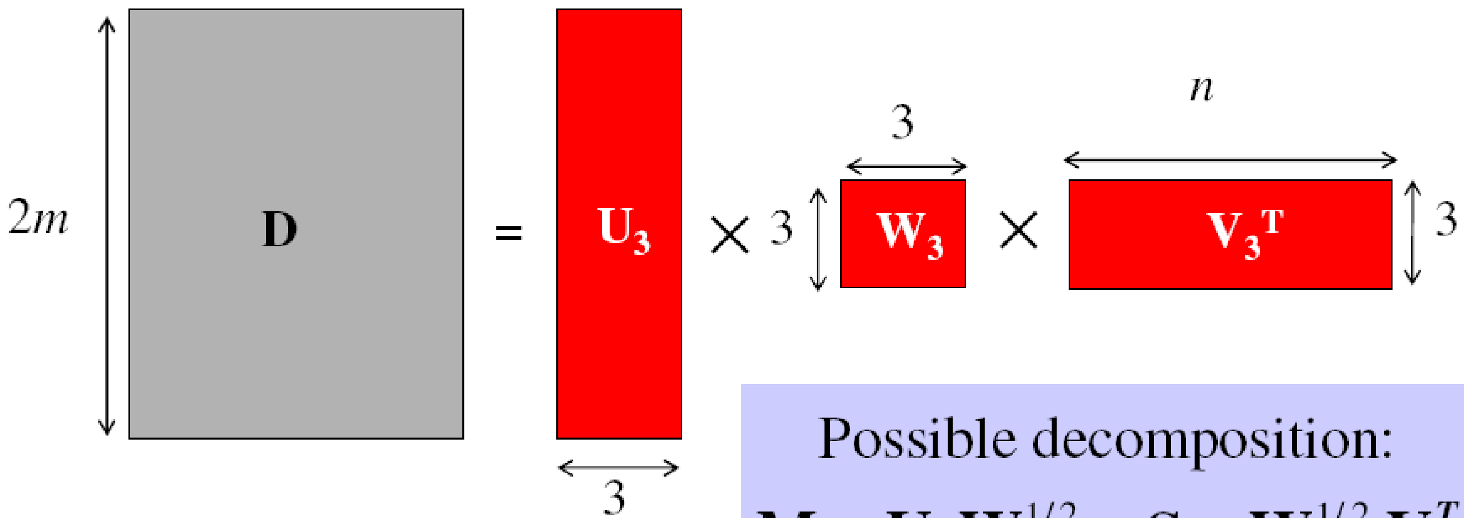
- Obtaining a factorization from SVD:



Source: M. Hebert

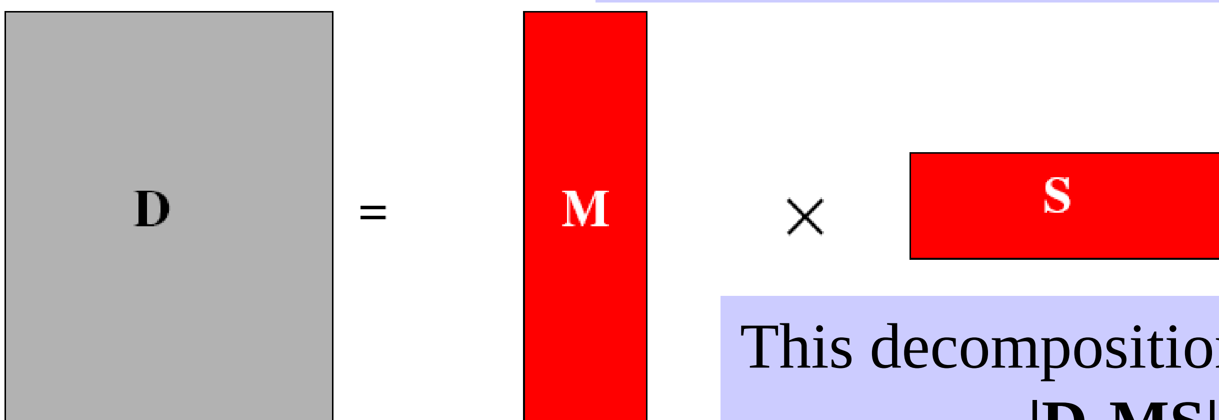
Factorizing the measurement matrix

- Obtaining a factorization from SVD:



Possible decomposition:

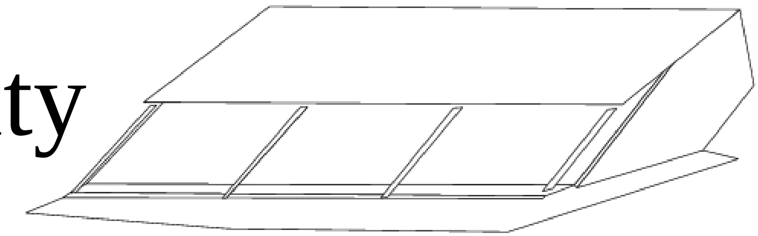
$$\mathbf{M} = \mathbf{U}_3 \mathbf{W}_3^{1/2} \qquad \mathbf{S} = \mathbf{W}_3^{1/2}\, \mathbf{V}_3^T$$

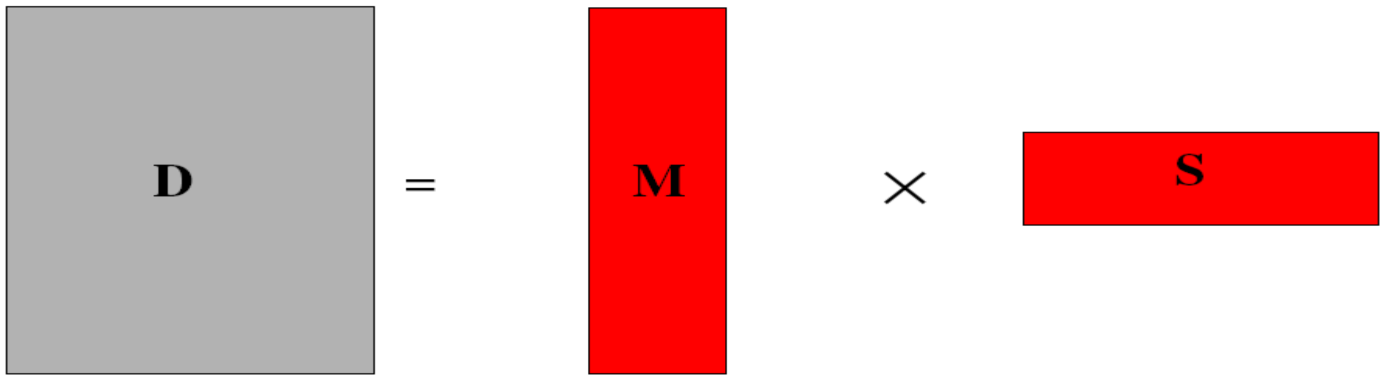


This decomposition minimizes $|\mathbf{D}\text{-}\mathbf{MS}|^2$

Source: M. Hebert

Affine ambiguity



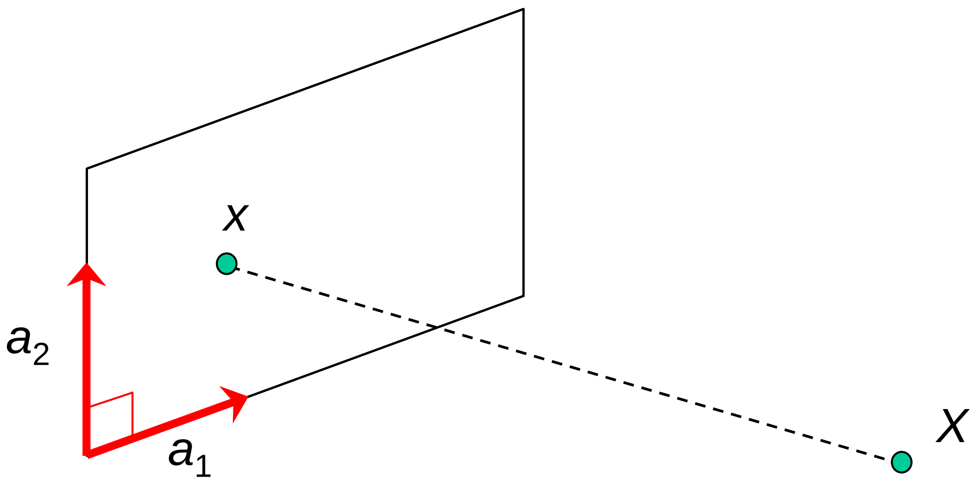


- The decomposition is not unique. We get the same **D** by using any 3×3 matrix **C** and applying the transformations
  $\mathbf{M} \rightarrow \mathbf{MC},\ \mathbf{S} \rightarrow \mathbf{C}^{-1}\mathbf{S}\ ;\quad \mathbf{D} \rightarrow \mathbf{MCC}^{-1}\mathbf{S}$
  (The true M and S are a linear transformation of M, S)
- That is because we have only an affine transformation and we have not enforced any Euclidean constraints (like forcing the image axes to be perpendicular, for example)

Source: M. Hebert

Eliminating the affine ambiguity

- Orthographic: image axes are perpendicular and scale is 1



$\mathbf{a}_1 \cdot \mathbf{a}_2 = 0$

$|\mathbf{a}_1|^2 = |\mathbf{a}_2|^2 = 1$

$$\tilde{\mathbf{a}}_{i1}^T \mathbf{C}\mathbf{C}^T \tilde{\mathbf{a}}_{i1}^T = 1$$

$$\tilde{\mathbf{a}}_{i2}^T \mathbf{C}\mathbf{C}^T \tilde{\mathbf{a}}_{i2}^T = 1$$

$$\tilde{\mathbf{a}}_{i1}^T \mathbf{C}\mathbf{C}^T \tilde{\mathbf{a}}_{i2}^T = 0$$

- This translates into $3m$ equations in $\mathbf{L} = \mathbf{CC}^\mathbf{T}$:

$$\mathbf{A_i\, L\, A_i^T = Id}, \qquad i = 1, \ldots, m$$

  - Solve for **L**
  - Recover **C** from **L** by Cholesky decomposition: $\mathbf{L} = \mathbf{CC}^\mathbf{T}$
  - Update **M** and **S**: $\mathbf{M} = \mathbf{MC}$, $\mathbf{S} = \mathbf{C}^{-1}\mathbf{S}$

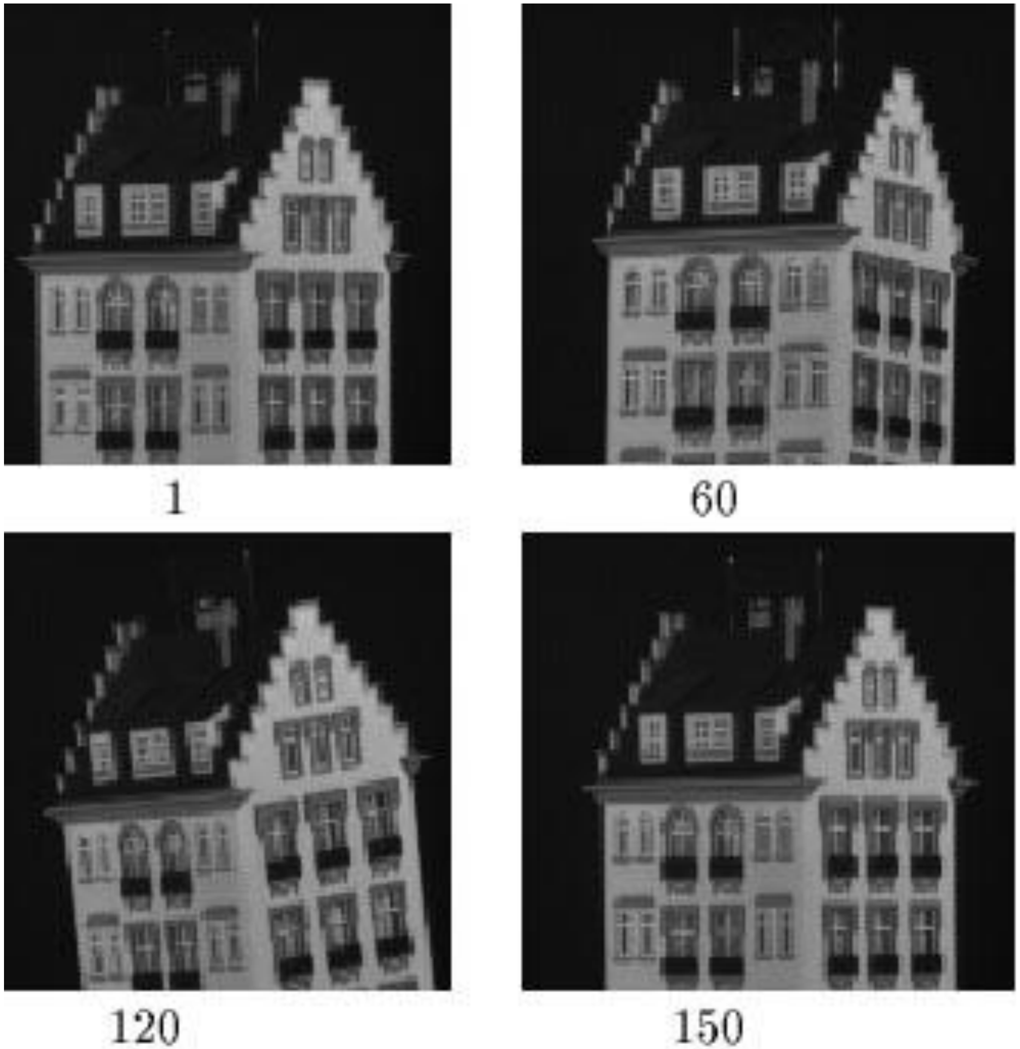
Source: M. Hebert

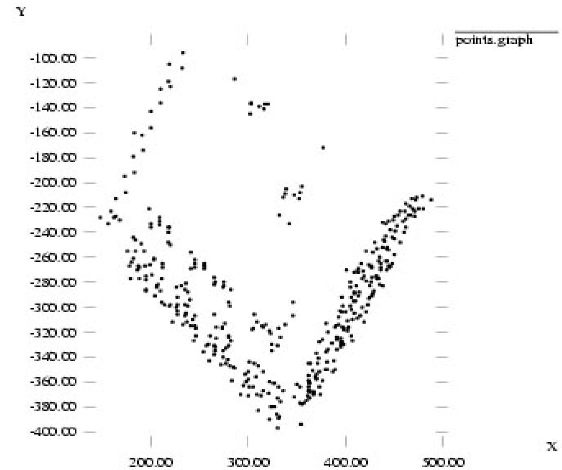
Algorithm summary

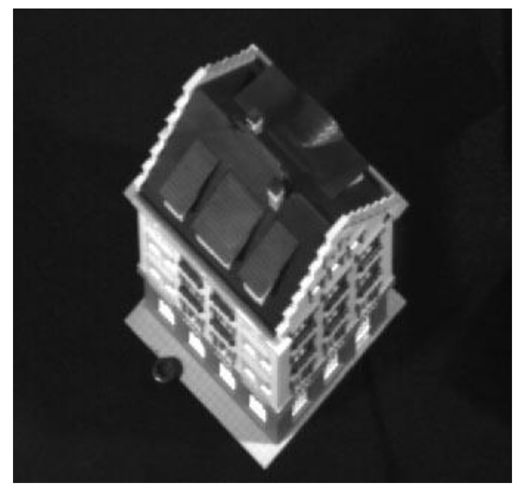
- Given: m images and n features $\mathbf{x}_{ij}$
- For each image i, center the feature coordinates
- Construct a $2m \times n$ measurement matrix $\mathbf{D}$:
  - Column j contains the projection of point j in all views
  - Row i contains one coordinate of the projections of all the n points in image i
- Factorize $\mathbf{D}$:
  - Compute SVD: $\mathbf{D} = \mathbf{U}\ \mathbf{W}\ \mathbf{V}^{\mathbf{T}}$
  - Create $\mathbf{U}_3$ by taking the first 3 columns of $\mathbf{U}$
  - Create $\mathbf{V}_3$ by taking the first 3 columns of $\mathbf{V}$
  - Create $\mathbf{W}_3$ by taking the upper left 3×3 block of $\mathbf{W}$
- Create the motion and shape matrices:
  - $\mathbf{M} = \mathbf{U}_3\mathbf{W}_3^{½}$ and $\mathbf{S} = \mathbf{W}_3^{½}\ \mathbf{V}_3^{T}$ (**or** $\mathbf{M} = \mathbf{U}_3$ and $\mathbf{S} = \mathbf{W}_3\mathbf{V}_3^{T}$)
- Eliminate affine ambiguity

Source: M. Hebert

Reconstruction results







C. Tomasi and T. Kanade. [Shape and motion from image streams under orthography: A factorization method.]() *IJCV*, 9(2):137-154, November 1992.

Results

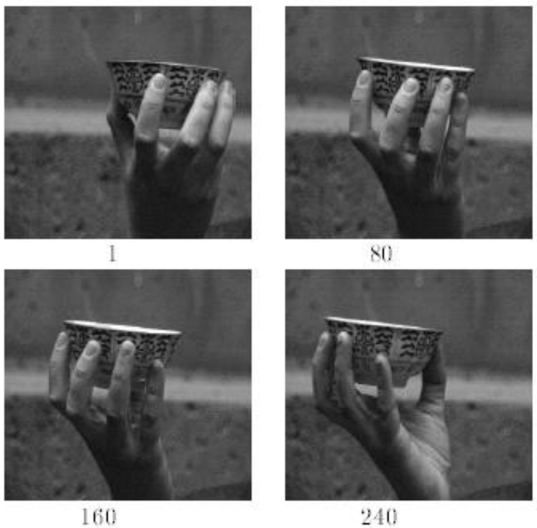


Figure 6.20: Four out of the 240 frames of the cup image stream.

Figure 6.23: A front view of the cup and fingers, with the original image intensities mapped onto the resulting surface.

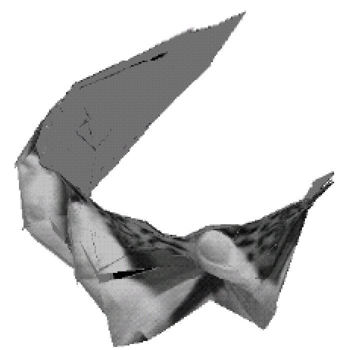
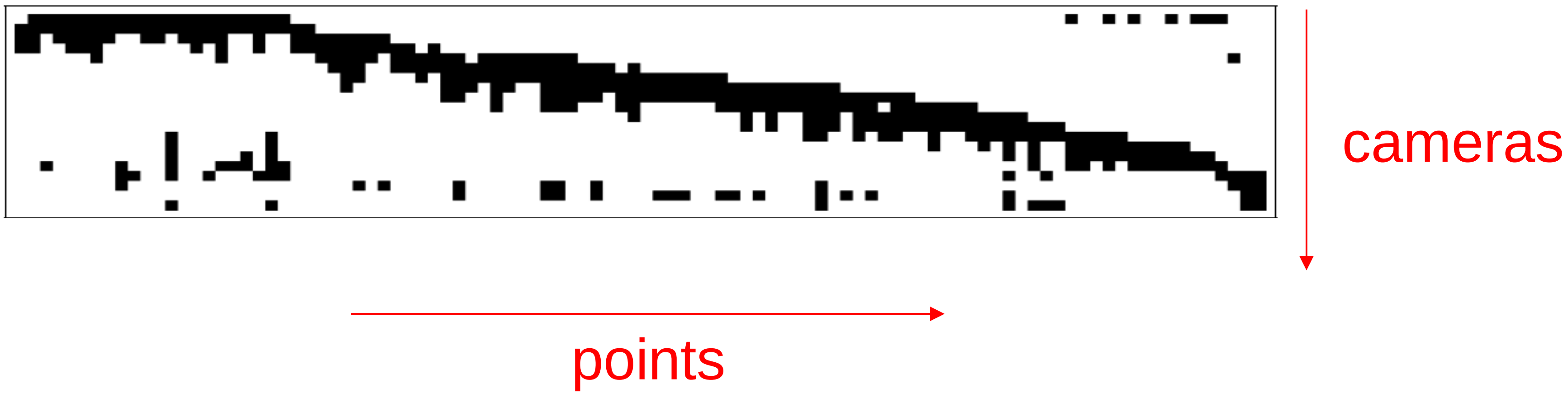


Figure 6.24: A view from above of the cup and fingers with image intensities mapped onto the surface.

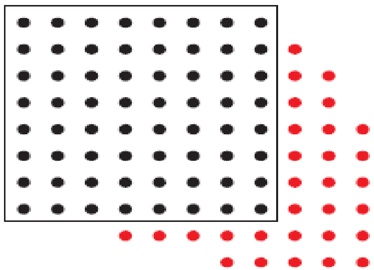
Dealing with missing data

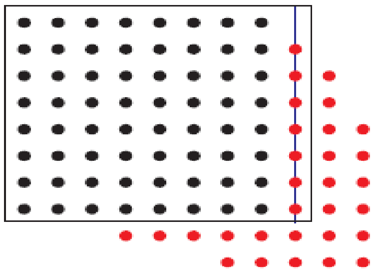
- So far, we have assumed that all points are visible in all views
- In reality, the measurement matrix typically looks something like this:

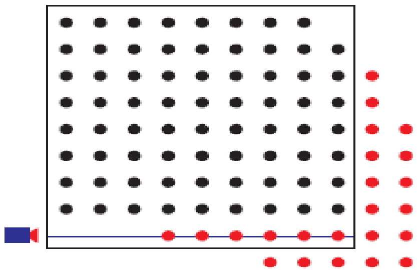


Dealing with missing data

- Possible solution: decompose matrix into dense sub-blocks, factorize each sub-block, and fuse the results







(1) Perform factorization on a dense sub-block

(2) Solve for a new 3D point visible by at least two known cameras (linear least squares)

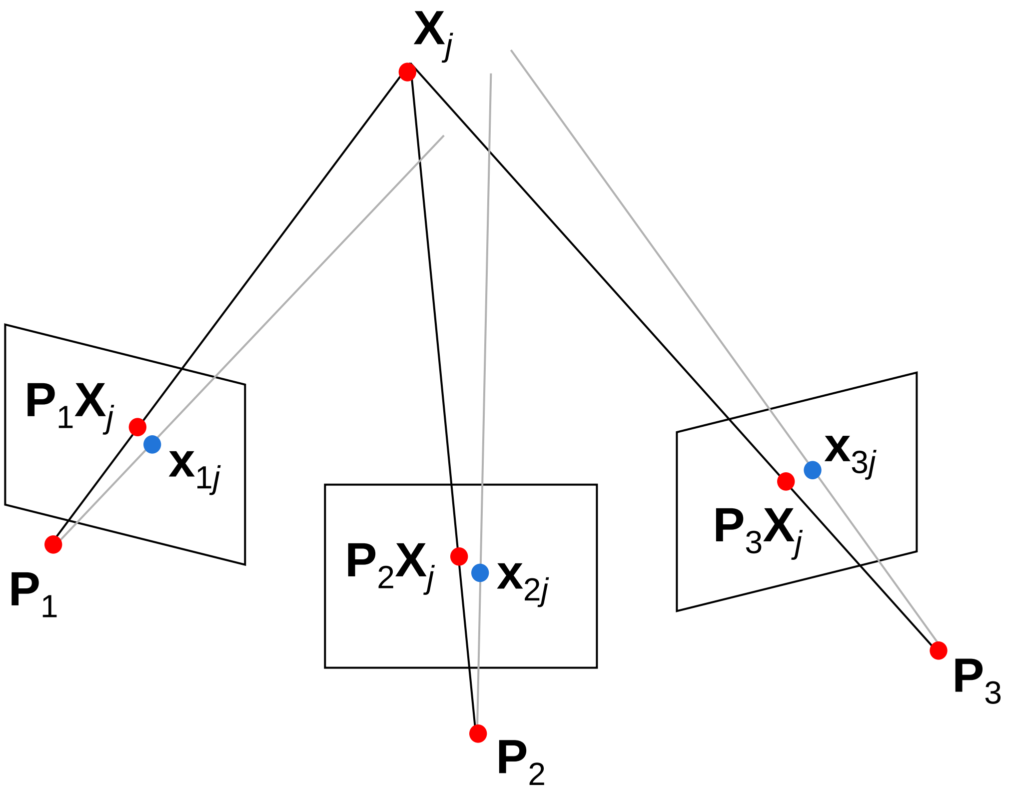
(3) Solve for a new camera that sees at least three known 3D points (linear least squares)

F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce. [Segmenting, Modeling, and Matching Video Clips Containing Multiple Moving Objects.] PAMI 2007.

Bundle adjustment

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

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



Further Factorization work

Factorization with uncertainty
(Irani & Anandan, IJCV'02)

Factorization for indep. moving objects
(Costeira and Kanade '94)

Factorization for articulated objects
(Yan and Pollefeys '05)

Factorization for dynamic objects
(Bregler et al. 2000, Brand 2001)

Perspective factorization
(Sturm & Triggs 1996, …)

Factorization with outliers and missing pts.
(Jacobs '97 (affine), Martinek & Pajdla'01 Aanaes'02 (perspective))

Questions?