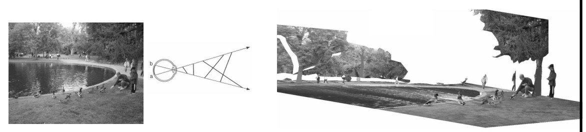
Depth Perception:
The inverse problem



First part: Monocular cues
Second part: Stereopsis

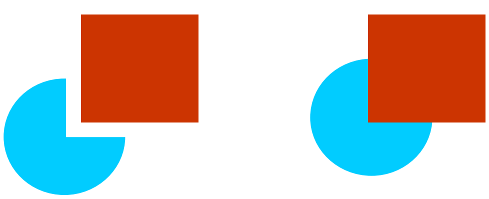
Monocular cues to depth

- **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, …)

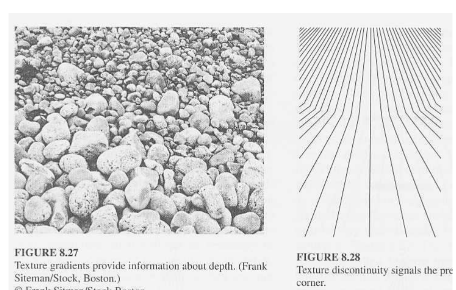
Relative depth cues

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

Interposition / occlusion



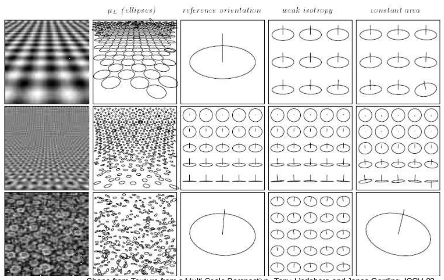
Texture Gradient



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

7

Texture Gradient

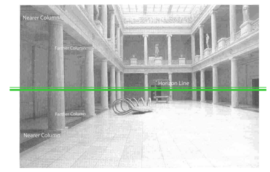


Shape from Texture from a Multi-Scale Perspective. Tony Lindeberg and Jonas Garding. ICCV 93

8

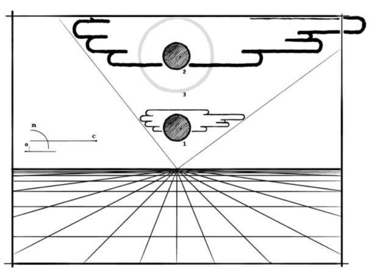
Distance from the horizon line

- Based on the tendency of objects to appear nearer the horizon line with greater distance to the horizon.
- 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.



9

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

10

Relative height

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

11

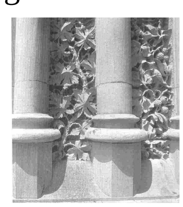
Illumination

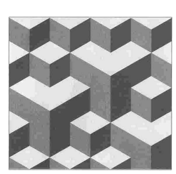
- Shading
- Shadows
- Inter-reflections

12

Shading

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

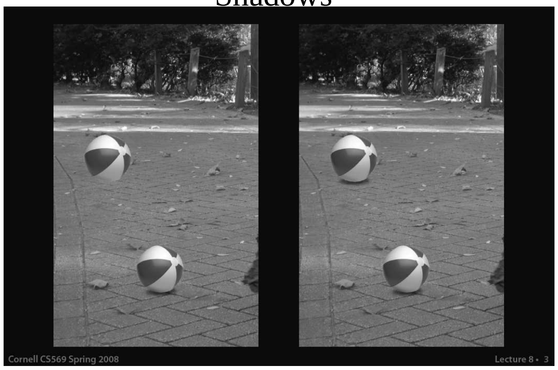




- Perception of depth through shading alone is always subject to the concave/convex inversion. The pattern shown can be perceived as stairsteps receding towards the top and lighted from above, or as an overhanging structure lighted from below.

13

Shadows

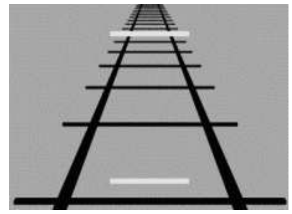


Slide by Steve Marschner

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

14

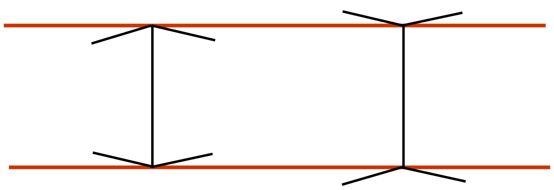
Linear Perspective



Ponzo's illusion

15

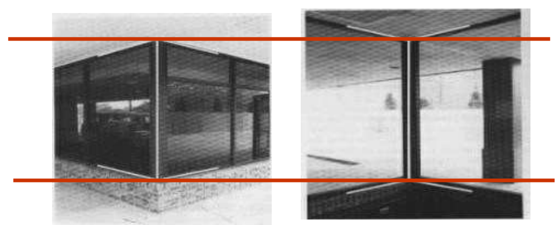
Linear Perspective



Muller-Lyer
1889

16

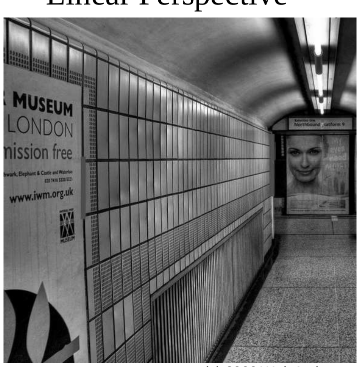
Linear Perspective



Muller-Lyer
1889

17

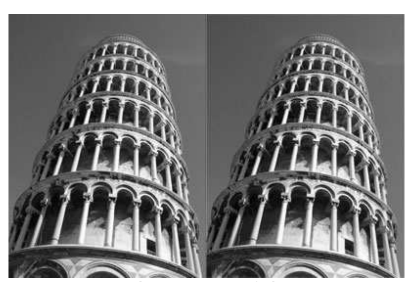
Linear Perspective



(c) 2006 Walt Anthony

18

3D drives perception of important object attributes



The two Towers of Pisa

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

19

True perspective

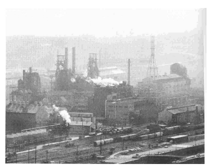
20

Expected perspective

21

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.



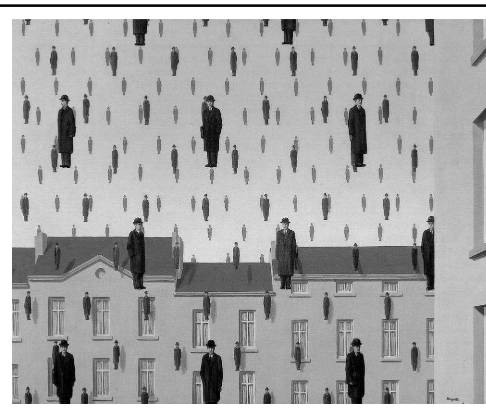
22

Atmospheric perspective

23

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

24



[*Golconde* Rene Magritte]

25

Absolute (monocular) depth cues

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

26

Familiar size

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

27

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 ?

28

Slide by Aude Oliva

29

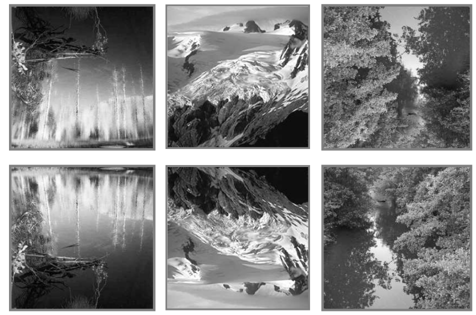
30

Slide by Aude Oliva

31

32

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

33

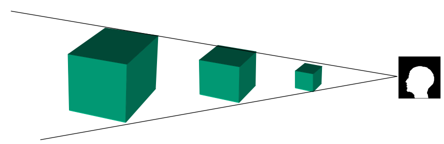
Depth Perception from Image Structure

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

34

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.

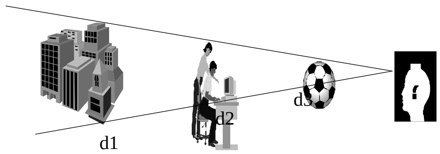


Stimulus ambiguity: the three cubes produce the same retinal image. Monocular information cannot give absolute depth measurements. Only relative depth information such as shape from shading and junctions (occlusions) can be obtained.

35

Depth Perception from Image Structure

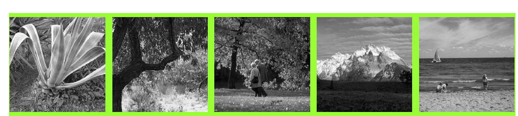
However, nature (and man) do not build in the same way at different scales.



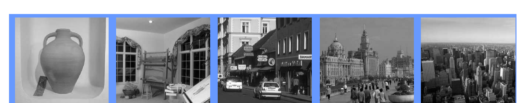
If d1>>d2>>d3 the structures of each view strongly differ. **Structure** provides monocular information about the scale (mean depth) of the space in front of the observer.

36

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

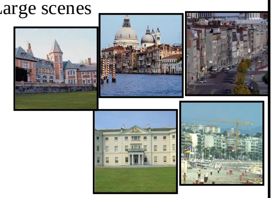
Slide by Aude Oliva

37

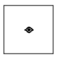
Image Statistics and Scene Scale

Close-up views

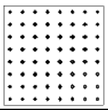
Large scenes



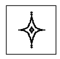
On average, low clutter



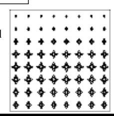
Point view is unconstrained



On average, highly cluttered



Point view is strongly constrained



38

Image Scale vs. Scene Scale

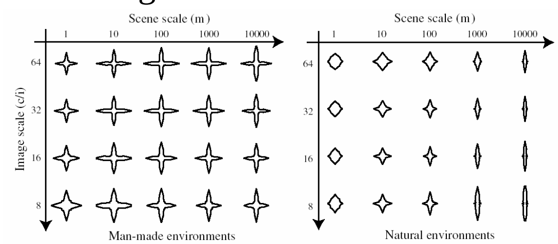


Figure 5. Polar plots of responses of multiscale oriented Gabor filters. The magnitude of each orientation corresponds to the total output energy averaged across the entire image. The energies are normalized across image scale by multiplying by a constant so that noise with $1/f$ amplitude spectrum has the same polar plots at all image scales.

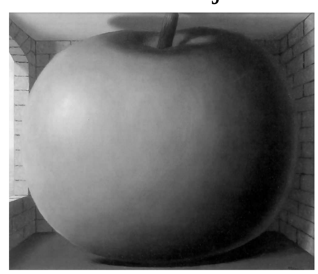
39

It is not all about objects

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

40

Scene vs. objects



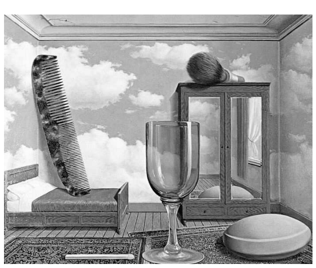
What do you see? A big apple or a small room?

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

[*The Listening Room* Rene Magritte]

41

Scene vs. objects

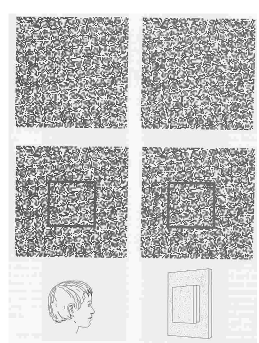


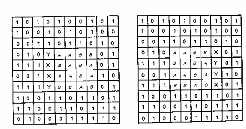
[*Personal Values* Rene Magritte]

42

Depth without objects

Random dot stereograms (Bela Julesz)





Julesz, 1971

Select a region in one image.

Shift this region horizontally by a small amount.

43

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

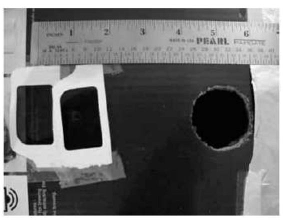
44

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

de credit: Kristen Grauman

45

Anaglyph pinhole camera



46

Autostereograms

Exploit disparity as depth cue using single image.

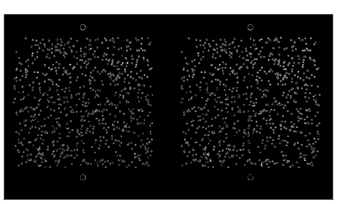
(Single image random dot stereogram, Single image stereogram)

Images from magiceye.com

Slide credit: Kristen Grauman

47

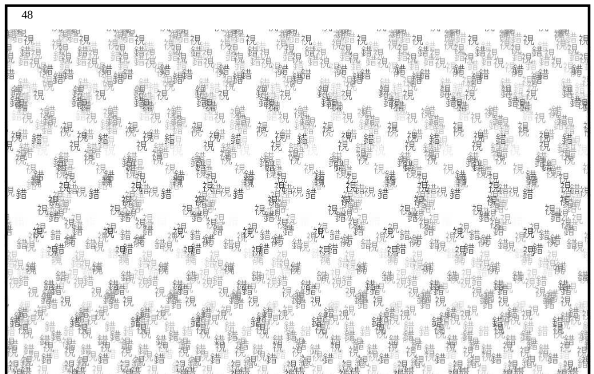
Cross-fusion



http://www.journalofvision.org/8/8/5/article.aspx

A typical disparity-defined stimulus from the experiment, showing a horizontally oriented half-cylinder. This figure is designed for cross-fusion, but in the experiment the stimuli were viewed through LCD-shuttered glasses and the large dots were not present.

48



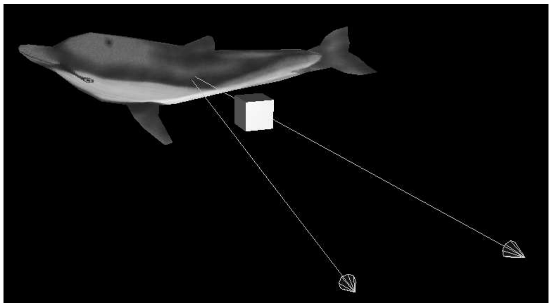
http://www.psy.ritsumei.ac.jp/~akitaoka/stereo3e.html

49

Questions?

50

Geometrical basis for stereopsis

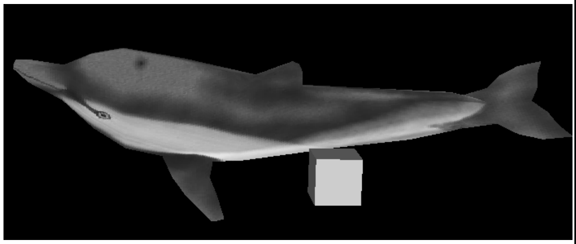


Two eyes look at the object of interest

http://en.wikipedia.org/wiki/Stereopsis#Geometrical_basis_for_stereopsis

51

Geometrical basis for stereopsis

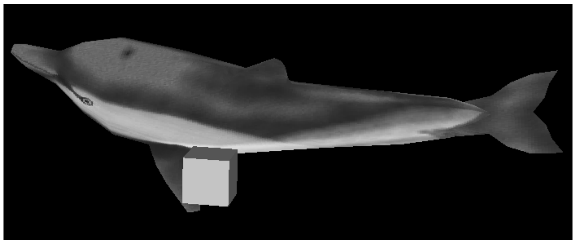


Which eye?

http://en.wikipedia.org/wiki/Stereopsis#Geometrical_basis_for_stereopsis

52

Geometrical basis for stereopsis

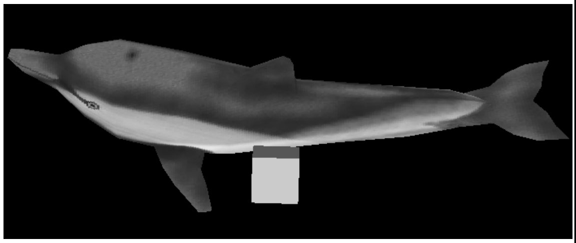


Which eye?

http://en.wikipedia.org/wiki/Stereopsis#Geometrical_basis_for_stereopsis

53

Geometrical basis for stereopsis

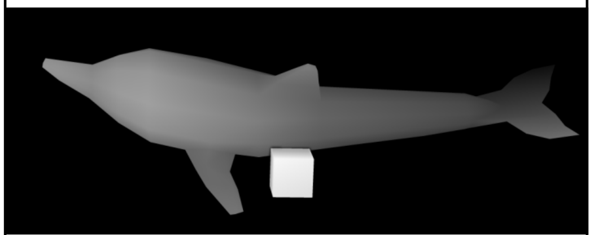


A single composite image is created

http://en.wikipedia.org/wiki/Stereopsis#Geometrical_basis_for_stereopsis

54

Geometrical basis for stereopsis

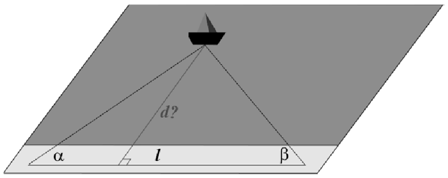


And a depth map (lighter is closer)

http://en.wikipedia.org/wiki/Stereopsis#Geometrical_basis_for_stereopsis

55

Triangulation



Distance d can be computed when alpha, beta and l is known

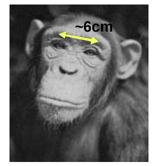
$$l = \frac{d}{\tan\alpha} + \frac{d}{\tan\beta}$$

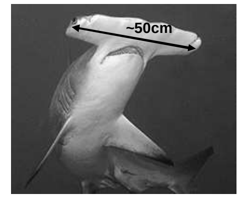
$$d = l/(\frac{1}{\tan\alpha} + \frac{1}{\tan\beta}$$

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

56

Stereo vision





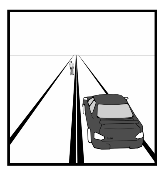
After 10 meter, disparity is quite small and depth from stereo is unreliable…

57

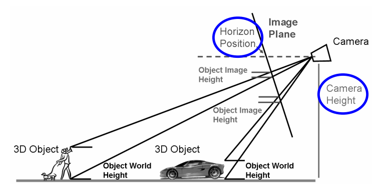
Questions?

58

Object Size in the Image



Image



World

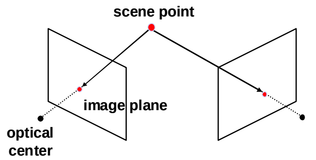
If you know camera parameters: height and position of the camera, then we know real depth

Slide by Derek Hoiem

59

Estimating depth with stereo

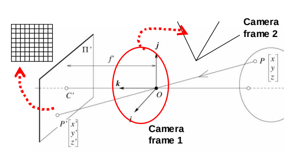
- **Stereo**: shape from “motion” between two views
- We’ll need to consider:
- Info on camera pose (“calibration”)
- Image point correspondences



Slide credit: Kristen Grauman

60

Camera parameters



Extrinsic parameters:
Camera frame 1 ←→ Camera frame 2

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

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

We’ll assume for now that these parameters are given and fixed.

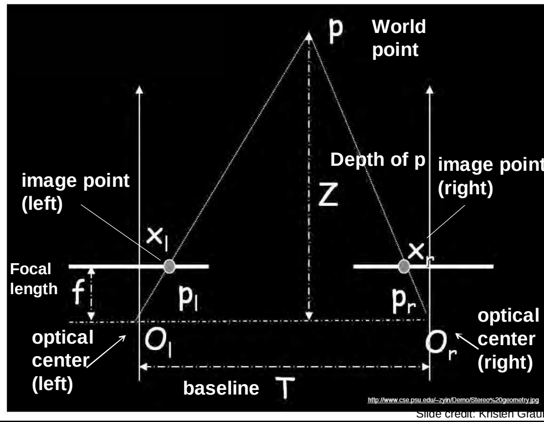
Slide credit: Kristen Grauman

61

Geometry for a simple stereo system

- First, assuming parallel optical axes, known camera parameters (i.e., calibrated cameras):

Slide credit: Kristen Grauman

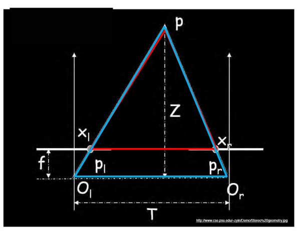


Slide credit: Kristen Grauman

63

Geometry for a simple stereo system

- Assume parallel optical axes, known camera parameters (i.e., calibrated cameras). We can triangulate via:



Similar triangles (p_l, P, p_r) and (O_l, P, O_r):

$$\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z}$$

$$Z = f \frac{T}{x_r - x_l}$$

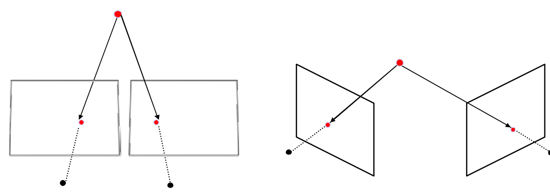
disparity

Slide credit: Kristen Grauman

64

General case, with calibrated cameras

- The two cameras need not have parallel optical axes.

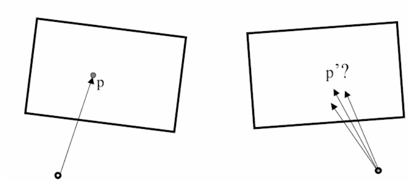


Vs.

Slide credit: Kristen Grauman

65

Stereo correspondence constraints

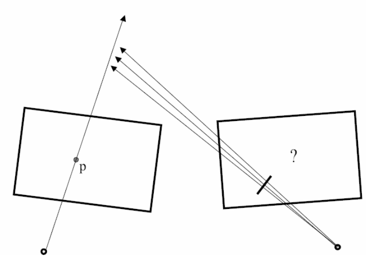


- **Given p in left image, where can corresponding point p' be?**

Slide credit: Kristen Grauman

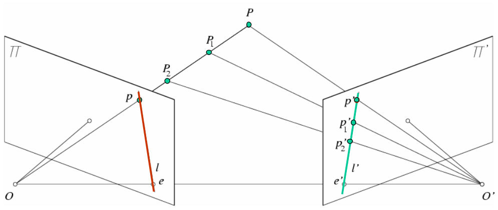
66

Stereo correspondence constraints



Slide credit: Kristen Grauman

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:

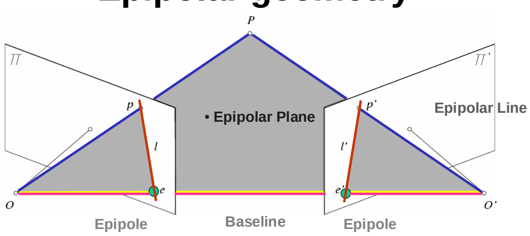
- **It must be on the line carved out by a plane connecting the world point and optical centers.**

Why is this useful?

Epipolar constraint

This is useful because it reduces the correspondence problem to a 1D search along an epipolar line.

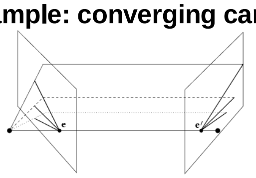
Epipolar geometry

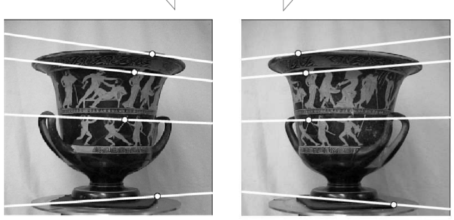


- **Epipolar plane**: plane containing baseline and world point
- **Epipole**: point of intersection of baseline with the image plane
- **Epipolar line**: intersection of epipolar plane with the image plane
- **Baseline**: line joining the camera centers
- All epipolar lines intersect at the epipole
- An epipolar plane intersects the left and right image planes in epipolar lines

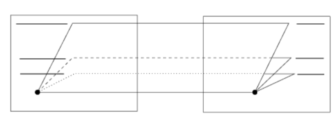
Example

Example: converging cameras





Example: parallel cameras



Where are the epipoles?

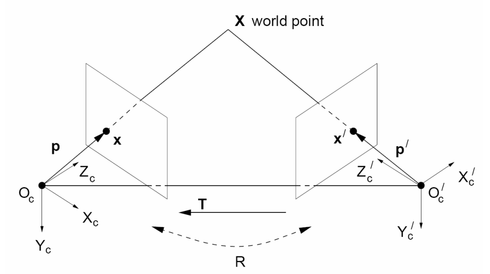
73

- So far, we have the explanation in terms of geometry.
- Now, how to express the epipolar constraints algebraically?

Slide credit: Kristen Grauman

74

Stereo geometry, with calibrated cameras

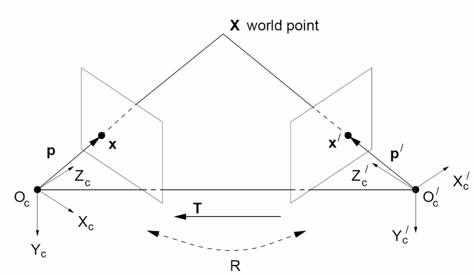


Main idea

Slide credit: Kristen Grauman

75

Stereo geometry, with calibrated cameras

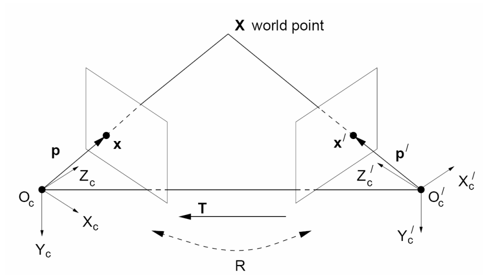


If the stereo rig is calibrated, we know :
how to rotate and translate camera reference frame 1 to get to camera reference frame 2.
Rotation: 3 x 3 matrix R; translation: 3 vector T.

Slide credit: Kristen Grauman

76

Stereo geometry, with calibrated cameras



If the stereo rig is calibrated, we know :
how to rotate and translate camera reference frame 1 to get to camera reference frame 2.

$$\mathbf{X}'_c = \mathbf{R}\mathbf{X}_c + \mathbf{T}$$

Slide credit: Kristen Grauman

77

An aside: cross product

$$\vec{a} \times \vec{b} = \vec{c} \qquad \vec{a} \cdot \vec{c} = 0 \qquad \vec{b} \cdot \vec{c} = 0$$

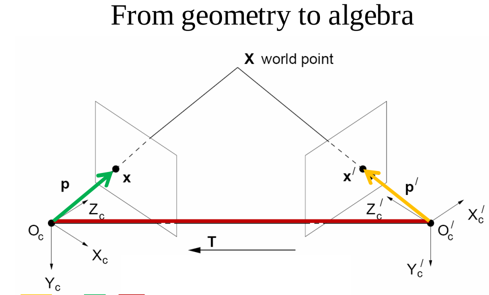
Vector cross product takes two vectors and returns a third vector that's perpendicular to both inputs.

So here, c is perpendicular to both a and b, which means the dot product = 0.

Slide credit: Kristen Grauman

78

From geometry to algebra



$$\boxed{\mathbf{X}'} = \mathbf{R}\boxed{\mathbf{X}} + \boxed{\mathbf{T}}$$

$$\underbrace{\mathbf{T} \times \mathbf{X}'}_{\text{Normal to the plane}} = \mathbf{T} \times \mathbf{R}\mathbf{X}$$

$$\mathbf{X}' \cdot (\mathbf{T} \times \mathbf{X}') = \mathbf{X}' \cdot (\mathbf{T} \times \mathbf{R}\mathbf{X}) = 0$$

Slide credit: Kristen Grauman

79

Another aside:
Matrix form of cross product

$$\vec{a}\times\vec{b} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \vec{c}$$

$$\vec{a}\cdot\vec{c} = 0$$
$$\vec{b}\cdot\vec{c} = 0$$

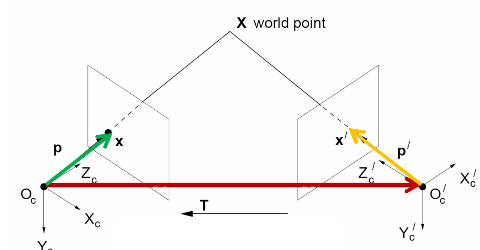
Can be expressed as a matrix multiplication.

$$[a_x] = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \qquad \boxed{\vec{a}\times\vec{b} = [a_x]\vec{b}}$$

Slide credit: Kristen Grauman

80

From geometry to algebra



$$\mathbf{X'} = \mathbf{RX} + \mathbf{T}$$

$$\underbrace{\mathbf{T}\times\mathbf{X'}}_{\text{Normal to the plane}} = \mathbf{T}\times\mathbf{RX} + \mathbf{T}\times\mathbf{T}$$
$$= \mathbf{T}\times\mathbf{RX}$$

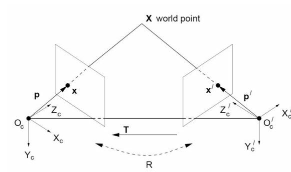
$$\boxed{\mathbf{X'}\cdot(\mathbf{T}\times\mathbf{X'}) = \mathbf{X'}\cdot(\mathbf{T}\times\mathbf{RX}) = 0}$$

Slide credit: Kristen Grauman

81

Essential matrix

$$\boxed{\mathbf{X'}\cdot(\mathbf{T}\times\mathbf{RX}) = 0}$$
$$\mathbf{X'}\cdot(\mathbf{T}_x\ \mathbf{RX}) = 0$$



Let $\mathbf{E} = \mathbf{T}_x\mathbf{R}$

$$\mathbf{X'}^T\mathbf{E}\mathbf{X} = 0$$

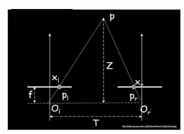
E is called the essential matrix, and it relates corresponding image points between both cameras, given the rotation and translation.

If we observe a point in one image, its position in other image is constrained to lie on line defined by above.

Note: these points are in camera coordinate systems.

82

Essential matrix example: parallel cameras



$\mathbf{R} =$

$\mathbf{T} =$

$\mathbf{E} = [\mathbf{T}_x]\mathbf{R} =$

$\mathbf{p} = [x, y, f]$

$\mathbf{p'} = [x', y', f]$

$$\mathbf{p'}^T\mathbf{E}\mathbf{p} = 0$$

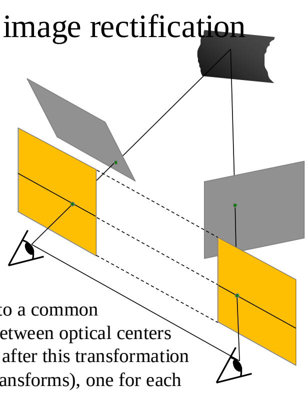
For the parallel cameras, image of any point must lie on same horizontal line in each image plane.

Slide credit: Kristen Grauman

83

Stereo image rectification

In practice, it is convenient if image scanlines (rows) are the epipolar lines.



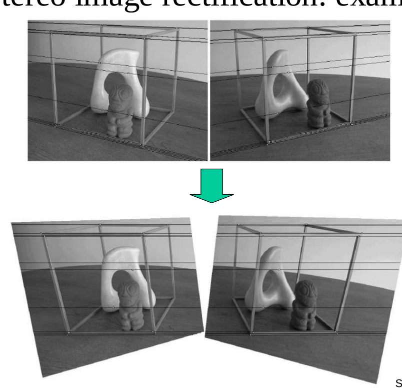
- reproject image planes onto a common plane parallel to the line between optical centers
- pixel motion is horizontal after this transformation
- two homographies (3x3 transforms), one for each input image reprojection

Adapted from Li Zhang

Slide credit: Kristen Grauman

84

Stereo image rectification: example



Source: Alyosha Efros

85

Questions?

86

Uncalibrated case

- What if we don't know the camera parameters?

Two possibilities:

1. Calibrate with a calibration object
2. Weak calibration

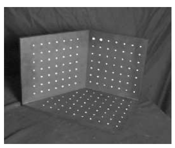
87

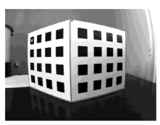
Calibrating a camera

- Compute intrinsic and extrinsic parameters using observed camera data

Main idea

- Place "calibration object" with known geometry in the scene
- Get correspondences
- Solve for mapping from scene to image

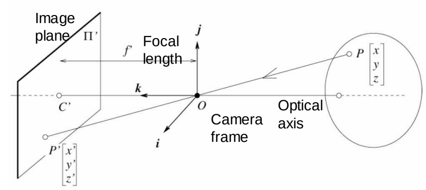




The Opti-CAL Calibration Target Image

88

Perspective projection



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

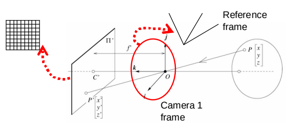
Scene point $\longrightarrow$ Image coordinates

Thus far, in **camera's** reference frame only.

89

Camera parameters

- Extrinsic: location and orientation of camera frame with respect to reference frame
- Intrinsic: how to map pixel coordinates to image plane coordinates



90

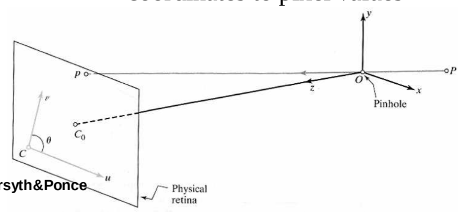
Camera calibration

Use the camera to tell you things about the world:

- Relationship between coordinates in the world and coordinates in the image: *geometric camera calibration, see* Szeliski, section 6.2, 6.3 for references
- (Relationship between intensities in the world and intensities in the image: *photometric image formation*, see Szeliski, sect. 2.2.)

91

Intrinsic parameters: from idealized world coordinates to pixel values



Forsyth&Ponce

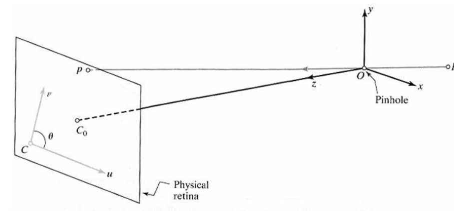
Perspective projection

$$u = f\frac{x}{z}$$

$$v = f\frac{y}{z}$$

92

Intrinsic parameters



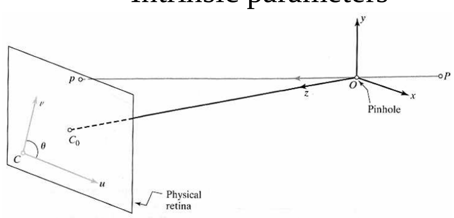
But “pixels” are in some arbitrary spatial units

$$u = \alpha\frac{x}{z}$$

$$v = \alpha\frac{y}{z}$$

93

Intrinsic parameters



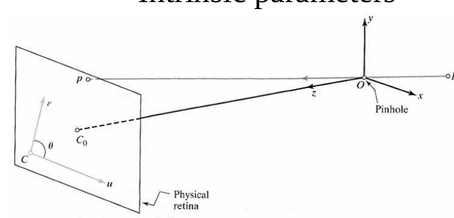
Maybe pixels are not square

$$u = \alpha\frac{x}{z}$$

$$v = \beta\frac{y}{z}$$

94

Intrinsic parameters



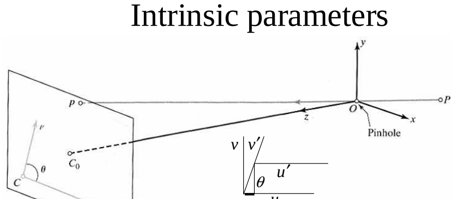
We don’t know the origin of our camera pixel coordinates

$$u = \alpha\frac{x}{z} + u_0$$

$$v = \beta\frac{y}{z} + v_0$$

95

Intrinsic parameters



$$v'\sin(\theta) = v$$

$$u' = u - \cos(\theta)v' = u - \cot(\theta)v$$

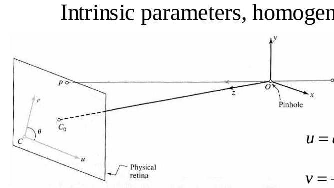
May be skew between camera pixel axes

$$u = \alpha\frac{x}{z} - \alpha\cot(\theta)\frac{y}{z} + u_0$$

$$v = \frac{\beta}{\sin(\theta)}\frac{y}{z} + v_0$$

96

Intrinsic parameters, homogeneous coordinates



$$u = \alpha\frac{x}{z} - \alpha\cot(\theta)\frac{y}{z} + u_0$$

$$v = \frac{\beta}{\sin(\theta)}\frac{y}{z} + v_0$$

Using homogenous coordinates, we can write this as:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha & -\alpha\cot(\theta) & u_0 & 0 \\ 0 & \frac{\beta}{\sin(\theta)} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

or:

In pixels

$$\vec{p} = \mathrm{K}\ {}^{C}\vec{p}$$

In camera-based coords

97

Extrinsic parameters: translation and rotation of camera frame

$${}^{C}\vec{p} = {}^{C}_{W}R\; {}^{W}\vec{p} + {}^{C}_{W}\vec{t}$$

Non-homogeneous coordinates

$$\begin{pmatrix} \\ {}^{C}\vec{p} \\ \\ \\ \end{pmatrix} = \left(\begin{array}{ccc|c} - & - & - & | \\ - & {}^{C}_{W}R & - & {}^{C}_{W}\vec{t} \\ - & - & - & | \\ 0 & 0 & 0 & 1 \end{array}\right) \begin{pmatrix} \\ {}^{W}\vec{p} \\ \\ \\ \end{pmatrix}$$

Homogeneous coordinates

98

Combining extrinsic and intrinsic calibration parameters, in homogeneous coordinates

pixels

$$\vec{p} = K\; {}^{C}\vec{p}$$

Intrinsic

Camera coordinates

World coordinates

$$\begin{pmatrix} \\ {}^{C}\vec{p} \\ \\ \\ \end{pmatrix} = \left(\begin{array}{ccc|c} - & - & - & | \\ - & {}^{C}_{W}R & - & {}^{C}_{W}\vec{t} \\ - & - & - & | \\ 0 & 0 & 0 & 1 \end{array}\right) \begin{pmatrix} \\ {}^{W}\vec{p} \\ \\ \\ \end{pmatrix}$$

Extrinsic

$$\vec{p} = K \underbrace{\left({}^{C}_{W}R \quad {}^{C}_{W}\vec{t} \right)}_{M} {}^{W}\vec{p}$$

$$\vec{p} = M\; {}^{W}\vec{p}$$

Forsyth&Ponce

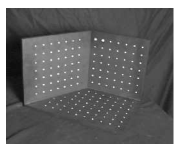
99

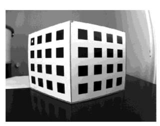
Calibrating a camera

- Compute intrinsic and extrinsic parameters using observed camera data

Main idea

- Place “calibration object” with known geometry in the scene
- Get correspondences
- Solve for mapping from scene to image: estimate $\mathbf{M} = \mathbf{M}_{int}\mathbf{M}_{ext}$





The Opti-CAL Calibration Target Image

100

When would we calibrate this way?

- Makes sense when geometry of system is not going to change over time

…when would it change?

101

Weak calibration

- Want to estimate world geometry without requiring calibrated cameras
  - Archival videos
  - Photos from multiple unrelated users
  - Dynamic camera system
- **Main idea**:
  - Estimate epipolar geometry from a (redundant) set of point correspondences between two uncalibrated cameras

102

From before: Projection matrix

- This can be rewritten as a matrix product using homogeneous coordinates:

$$\begin{bmatrix} wx_{im} \\ wy_{im} \\ w \end{bmatrix} = \mathbf{M}_{\text{int}} \mathbf{M}_{ext} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

where:

$$\mathbf{M}_{\text{int}} = \begin{bmatrix} -f/s_x & 0 & o_x \\ 0 & -f/s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{M}_{ext} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & -\mathbf{R}_1^{\mathrm{T}}\mathbf{T} \\ r_{21} & r_{22} & r_{23} & -\mathbf{R}_2^{\mathrm{T}}\mathbf{T} \\ r_{31} & r_{32} & r_{33} & -\mathbf{R}_3^{\mathrm{T}}\mathbf{T} \end{bmatrix}$$

103

From before: Projection matrix

- This can be rewritten as a matrix product using homogeneous coordinates:

$$\begin{bmatrix} wx_{im} \\ wy_{im} \\ w \end{bmatrix} = \mathbf{M}_{int}\mathbf{M}_{ext} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

$$\mathbf{p}_{im} = \mathbf{M}_{int}\underbrace{\mathbf{M}_{ext}\mathbf{P}_w}_{\mathbf{p}_c}$$

$$\mathbf{p}_{im} = \mathbf{M}_{int}\mathbf{p}_c$$

104

Uncalibrated case

For a given camera: $\mathbf{p}_{im} = \mathbf{M}_{int}\mathbf{p}_c$

So, for **two** cameras (left and right):

$$\mathbf{p}_{c,left} = \mathbf{M}^{-1}_{int,left}\,\mathbf{p}_{im,left}$$

$$\mathbf{p}_{c,right} = \underbrace{\mathbf{M}^{-1}_{int,right}}_{\text{Internal calibration matrices, one per camera}}\,\mathbf{p}_{im,right}$$

105

$$\mathbf{p}_{c,left} = \mathbf{M}^{-1}_{int,left}\,\mathbf{p}_{im,left}$$
$$\mathbf{p}_{c,right} = \mathbf{M}^{-1}_{int,right}\,\mathbf{p}_{im,right}$$

Uncalibrated case

$\boxed{\mathbf{p}_{c,right}{}^{\mathrm{T}}\mathbf{E}\mathbf{p}_{c,left} = 0}$ From before, the **essential** matrix **E.**

$$\left(\mathbf{M}^{-1}_{int,right}\mathbf{p}_{im,right}\right)^{\mathrm{T}}\mathbf{E}\left(\mathbf{M}^{-1}_{int,left}\mathbf{p}_{im,left}\right) = 0$$

$$\mathbf{p}^{\mathrm{T}}_{im,right}\underbrace{\left(\mathbf{M}^{-\mathrm{T}}_{int,right}\mathbf{E}\mathbf{M}^{-1}_{int,left}\right)}_{\mathbf{F}}\mathbf{p}_{im,left} = 0$$

"Fundamental matrix"

$$\boxed{\mathbf{p}^{\mathrm{T}}_{im,right}\mathbf{F}\mathbf{p}_{im,left} = 0}$$

106

Computing F from correspondences

Each point correspondence generates one constraint on F $\quad \mathbf{p}^{\mathrm{T}}_{im,right}\mathbf{F}\mathbf{p}_{im,left} = 0$

$$\begin{bmatrix} u' & v' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0$$

Collect n of these constraints

$$\begin{bmatrix} u'_1u_1 & u'_1v_1 & u'_1 & v'_1u_1 & v'_1v_1 & v'_1 & u_1 & v_1 & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = \mathbf{0}$$

Solve for f , vector of parameters.

107

Fundamental matrix

- Relates **pixel coordinates** in the two views
- More general form than essential matrix: we remove need to know intrinsic parameters
- If we estimate fundamental matrix from correspondences in *pixel coordinates*, can reconstruct epipolar geometry without intrinsic or extrinsic parameters.

108

Stereo pipeline with weak calibration

- So, where to start with uncalibrated cameras?
  - Need to find fundamental matrix F **and** the correspondences (pairs of points (u',v') ↔ (u,v)).

- 1) Find interest points in image
- 2) Compute correspondences
- 3) Compute epipolar geometry
- 4) Refine

Example from Andrew Zisserman

109

Stereo pipeline with weak calibration

1) Find interest points

110

Stereo pipeline with weak calibration

2) Match points within proximity to get putative matches

111

Stereo pipeline with weak calibration

3) Compute epipolar geometry -- robustly with RANSAC

Select random sample of putative correspondences

Compute **F** using them
- determines epipolar constraint

Evaluate amount of support
- inliers within threshold distance of epipolar line

Choose F with most support (inliers)

Using correlation search to get putative matches: noisy, but enough to compute F using RANSAC

Pruned matches: those consistent with epipolar geometry

113

Questions?