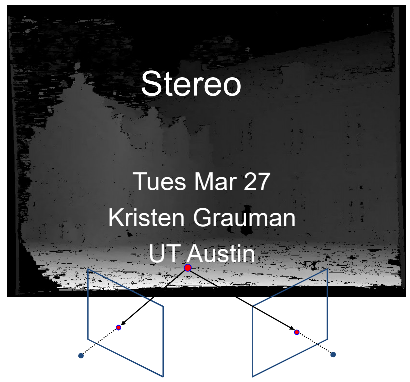
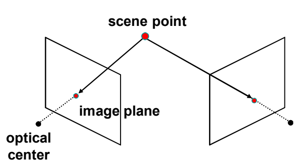


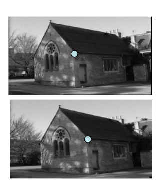
Outline

- Human stereopsis
- Epipolar geometry and the epipolar constraint
  - Case example with parallel optical axes
  - General case with calibrated cameras
- Stereo solutions
  - Correspondences
  - Additional constraints

Estimating depth with stereo

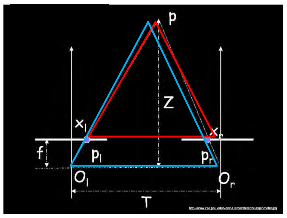
- **Stereo**: shape from "motion" between two views
- We'll need to consider:
  - Info on camera pose ("calibration")
  - Image point correspondences





Last time: geometry for a simple stereo system

- Assume parallel optical axes, known camera parameters (i.e., calibrated cameras). **What is expression for Z?**



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

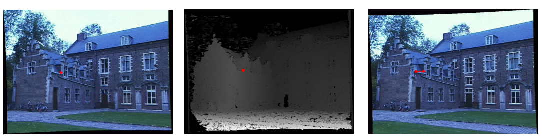
$$\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z}$$

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

disparity → $x_r - x_l$

Last time: Depth from disparity

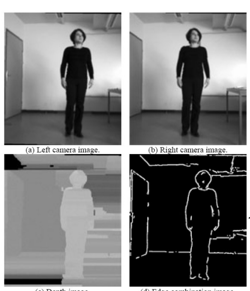
image I(x,y) Disparity map D(x,y) image I´(x´,y´)



(x´,y´)=(x+D(x,y), y)

So if we could find the **corresponding points** in two images, we could **estimate relative depth**…

Depth for segmentation



(a) Left camera image. (b) Right camera image. (c) Depth image. (d) Edge combination image.

Figure 3 Stereo video frames with computed depth map and edge combination result.

Edges in disparity in conjunction with image edges enhances contours found

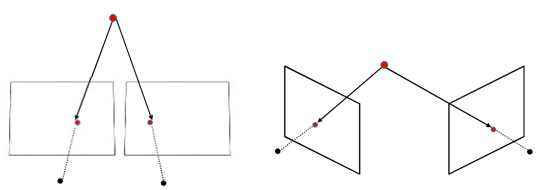
Danijela Markovic and Margrit Gelautz, Interactive Media Systems Group, Vienna University of Technology

Outline

- Human stereopsis
- Epipolar geometry and the epipolar constraint
  - Case example with parallel optical axes
  - General case with calibrated cameras
- Stereo solutions
  - Correspondences
  - Additional constraints

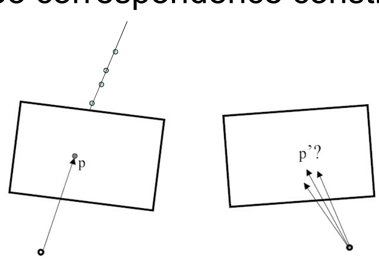
General case, with calibrated cameras

- The two cameras need not have parallel optical axes.



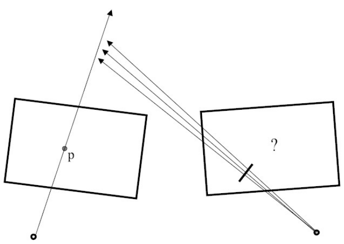
Vs.

Stereo correspondence constraints

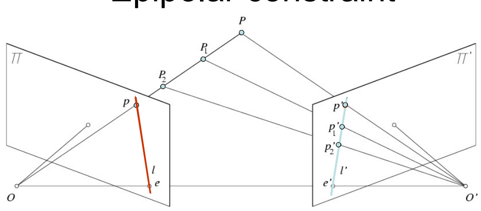


- Given p in left image, where can corresponding point p' be?

Stereo correspondence constraints



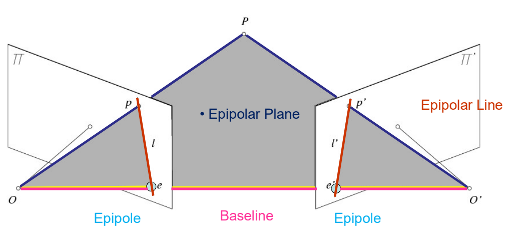
Epipolar constraint



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



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

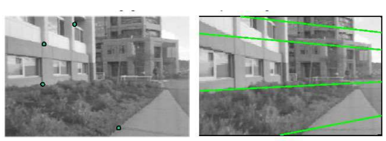
Why is the epipolar constraint useful?

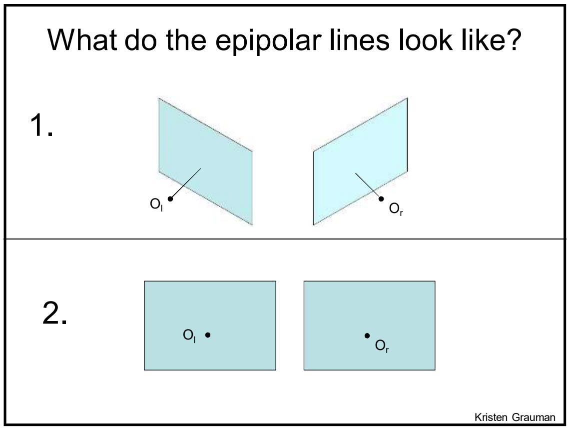
Epipolar constraint

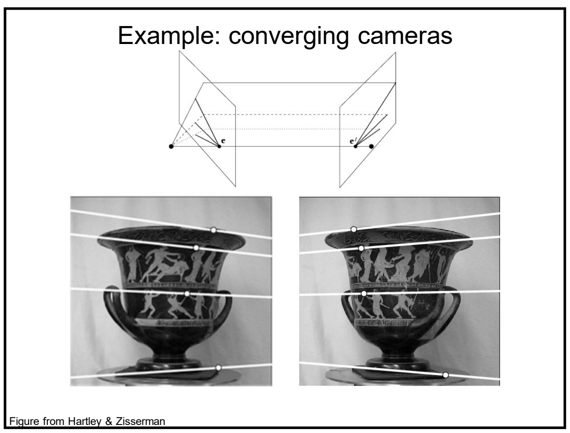
This is useful because it reduces the correspondence problem to a 1D search along an epipolar line.

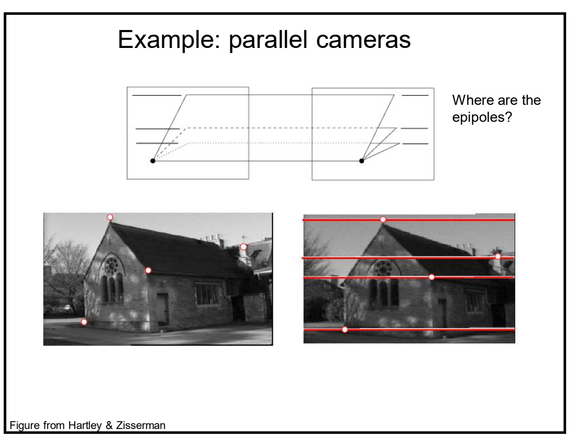
Image from Andrew Zisserman

Example



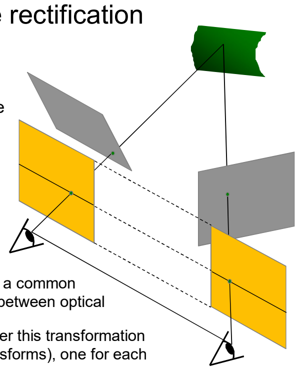

What do the epipolar lines look like?
1.
O_l
O_r
2.
O_l
O_r
Kristen Grauman


Example: converging cameras
Figure from Hartley & Zisserman


Example: parallel cameras
Where are the epipoles?
Figure from Hartley & Zisserman

Stereo image rectification

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



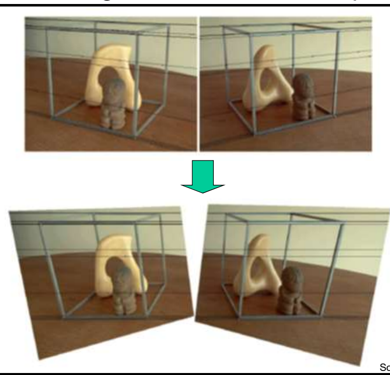
reproject image planes onto a common plane parallel to the line between optical centers

pixel motion is horizontal after this transformation

two homographies (3x3 transforms), one for each input image reprojection

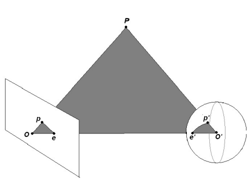
Slide credit: Li Zhang

Stereo image rectification: example



Source: Alyosha Efros

An audio camera & epipolar geometry



Spherical microphone array

Adam O' Donovan, Ramani Duraiswami and Jan Neumann Microphone Arrays as Generalized Cameras for Integrated Audio Visual Processing, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis, 2007

An audio camera & epipolar geometry

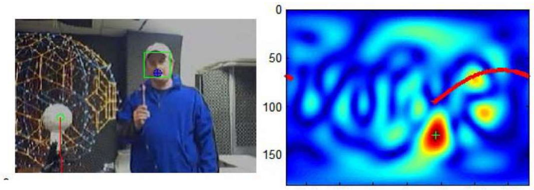


Figure 4. An example of the use of the system in speaker tracking with noise suppression. The bright red spot on the sound image (marked with a +) corresponds to the dominant source. The less dominant source however lies on the epipolar line in the sound image induced by the location of the mouth in the camera image, and this source is beamformed.

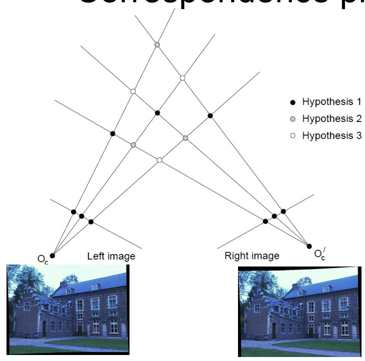
Summary so far

- Depth from stereo: main idea is to triangulate from corresponding image points.
- Epipolar geometry defined by two cameras
  – We've assumed known extrinsic parameters relating their poses
- Epipolar constraint limits where points from one view will be imaged in the other
  – Makes search for correspondences quicker
- **Terms**: epipole, epipolar plane / lines, disparity, rectification, baseline

Outline

- Human stereopsis
- Epipolar geometry and the epipolar constraint
  – Case example with parallel optical axes
  – General case with calibrated cameras
- Stereo solutions
  – Correspondences
  – Additional constraints

Correspondence problem



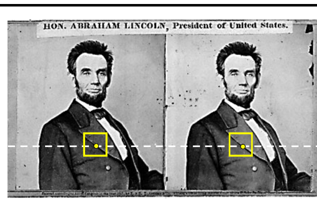
Multiple match hypotheses satisfy epipolar constraint, but which is correct?

Figure from Gee & Cipolla 1999

Correspondence problem

- Beyond the hard constraint of epipolar geometry, there are "soft" constraints to help identify corresponding points
  - Similarity
  - Uniqueness
  - Ordering
  - Disparity gradient
- To find matches in the image pair, we will assume
  - Most scene points visible from both views
  - Image regions for the matches are similar in appearance

Dense correspondence search



For each epipolar line

For each pixel / window in the left image

- compare with every pixel / window on same epipolar line in right image
- pick position with minimum match cost (e.g., SSD, correlation)

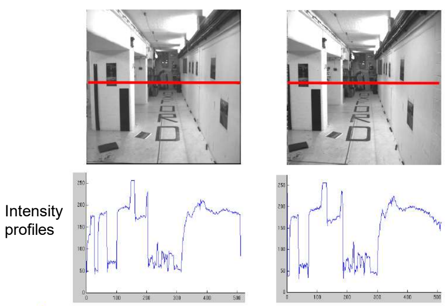
Adapted from Li Zhang

Correspondence problem

Parallel camera example: epipolar lines are corresponding image scanlines

Source: Andrew Zisserman

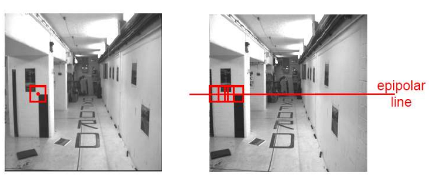
Correspondence problem



- Clear correspondence between intensities, but also noise and ambiguity

Source: Andrew Zisserman

Correspondence problem



Neighborhoods of corresponding points are similar in intensity patterns.

Source: Andrew Zisserman

Normalized cross correlation

subtract mean: $A \leftarrow A - \langle A \rangle, B \leftarrow B - \langle B \rangle$

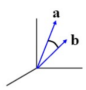
$$\text{NCC} = \frac{\sum_i \sum_j A(i,j) B(i,j)}{\sqrt{\sum_i \sum_j A(i,j)^2} \sqrt{\sum_i \sum_j B(i,j)^2}}$$

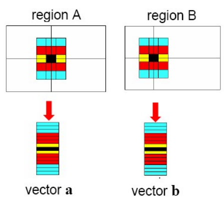
Write regions as vectors

$$\text{A} \to \mathbf{a}, \ \text{B} \to \mathbf{b}$$

$$\text{NCC} = \frac{\mathbf{a}.\mathbf{b}}{|\mathbf{a}||\mathbf{b}|}$$

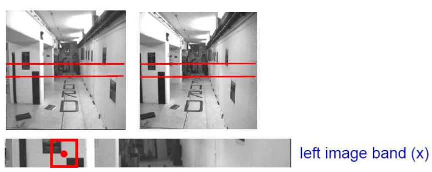
$$-1 \leq \text{NCC} \leq 1$$





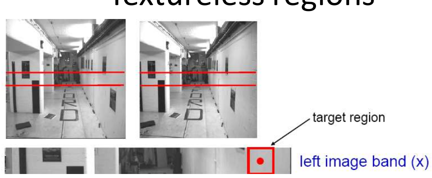
Source: Andrew Zisserman

Correlation-based window matching



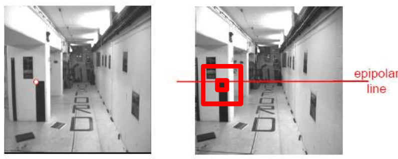
Source: Andrew Zisserman

Textureless regions



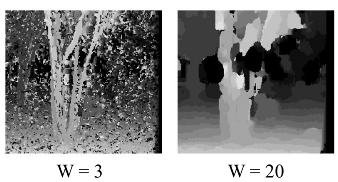
Source: Andrew Zisserman

Effect of window size?



Source: Andrew Zisserman

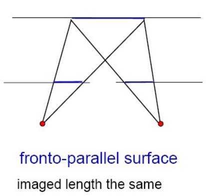
Effect of window size



Want window large enough to have sufficient intensity variation, yet small enough to contain only pixels with about the same disparity.

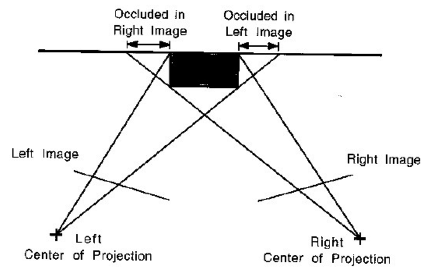
Figures from Li Zhang

Foreshortening effects



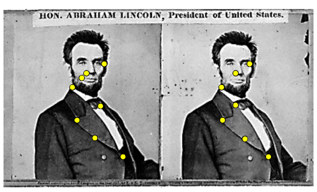
Source: Andrew Zisserman

Occlusion



Slide credit: David Kriegman

Sparse correspondence search



- Restrict search to sparse set of **detected features** (e.g., corners)
- Rather than pixel values (or lists of pixel values) use *feature descriptor* and an associated *feature distance*
- Still narrow search further by epipolar geometry

Tradeoffs between dense and sparse search?

Correspondence problem

- Beyond the hard constraint of epipolar geometry, there are "soft" constraints to help identify corresponding points
  - Similarity
  - Uniqueness
  - Disparity gradient
  - Ordering

Uniqueness constraint

- Up to one match in right image for every point in left image

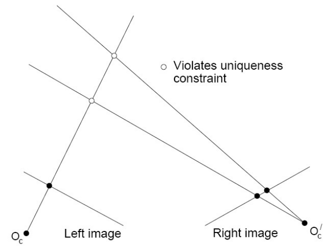
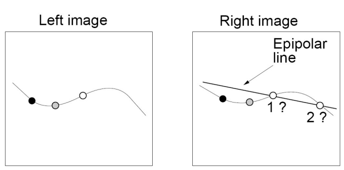


Figure from Gee & Cipolla 1999

Disparity gradient constraint

- Assume piecewise continuous surface, so want disparity estimates to be locally smooth



Given matches ● and ◎, point ○ in the left image must match point 1 in the right image. Point 2 would exceed the disparity gradient limit.

Figure from Gee & Cipolla 1999

Ordering constraint

- Points on **same surface** (opaque object) will be in same order in both views

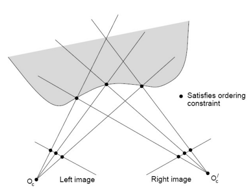
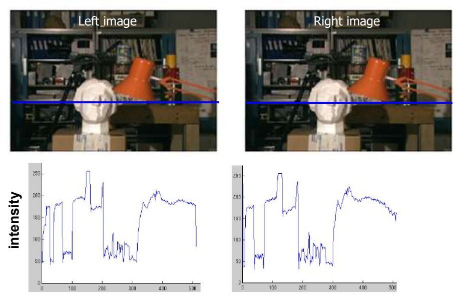


Figure from Gee & Cipolla 1999

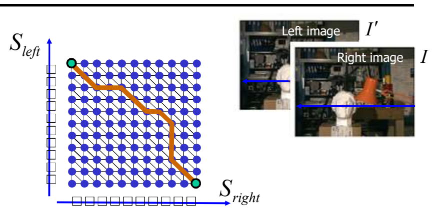
- Beyond individual correspondences to estimate disparities:
- Optimize correspondence assignments jointly
  - Scanline at a time (DP)
  - Full 2D grid (graph cuts)

Scanline stereo

- Try to coherently match pixels on the entire scanline
- Different scanlines are still optimized independently



"Shortest paths" for scan-line stereo

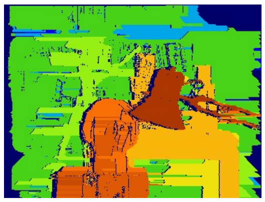


Can be implemented with dynamic programming
Ohta & Kanade '85, Cox et al. '96

Slide credit: Y. Boykov

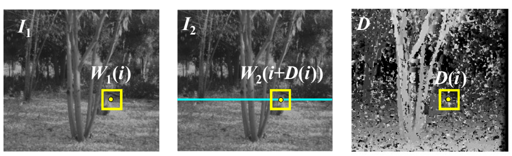
Coherent stereo on 2D grid

- Scanline stereo generates streaking artifacts



- Can't use dynamic programming to find spatially coherent disparities/ correspondences on a 2D grid

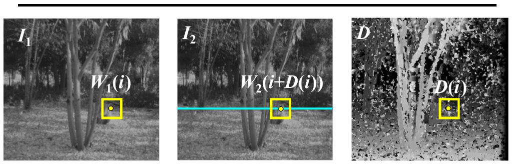
Stereo matching as energy minimization



$$E = \alpha E_{\text{data}}(I_1, I_2, D) + \beta E_{\text{smooth}}(D)$$

$$E_{\text{data}} = \sum_i \left(W_1(i) - W_2(i + D(i))\right)^2 \qquad E_{\text{smooth}} = \sum_{\text{neighbors } i,j} \rho\left(D(i) - D(j)\right)$$

Stereo matching as energy minimization



$$E = \alpha E_{\text{data}}(I_1, I_2, D) + \beta E_{\text{smooth}}(D)$$

$$E_{\text{data}} = \sum_i \left(W_1(i) - W_2(i + D(i))\right)^2 \qquad E_{\text{smooth}} = \sum_{\text{neighbors } i,j} \rho\left(D(i) - D(j)\right)$$

- Energy functions of this form can be minimized using *graph cuts*

Y. Boykov, O. Veksler, and R. Zabih, Fast Approximate Energy Minimization via Graph Cuts, PAMI 2001

Source: Steve Seitz

Error sources

- Low-contrast ; textureless image regions
- Occlusions
- Camera calibration errors
- Violations of *brightness constancy* (e.g., specular reflections)
- Large motions

Virtual viewpoint video

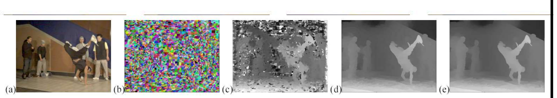


Figure 6: *Sample results from stereo reconstruction stage: (a) input color image; (b) color-based segmentation; (c) initial disparity estimates $\hat{d}_{ij}$; (d) refined disparity estimates; (e) smoothed disparity estimates $d_i(x)$.*

d) A depth-matted object from earlier in the sequence is inserted into the video.

C. Zitnick et al, High-quality video view interpolation using a layered representation, SIGGRAPH 2004.

Summary

- Depth from stereo: main idea is to triangulate from corresponding image points.
- Epipolar geometry defined by two cameras
  - We've assumed known extrinsic parameters relating their poses
- Epipolar constraint limits where points from one view will be imaged in the other
  - Makes search for correspondences quicker
- To estimate depth
  - Limit search by epipolar constraint
  - Compute correspondences, incorporate matching preferences

Coming up

- Instance recognition
  - Indexing local features efficiently
  - Spatial verification models