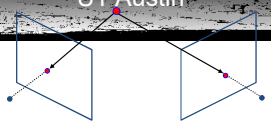
Stereo

Thurs Mar 30

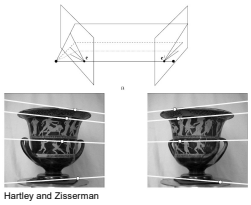
Kristen Grauman
UT Austin



Outline

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

Multiple views



Hartley and Zisserman

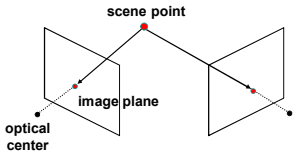
Multi-view geometry, matching, invariant features, stereo vision

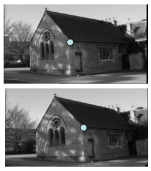
Lowe

Kristen Grauman

Estimating depth with stereo

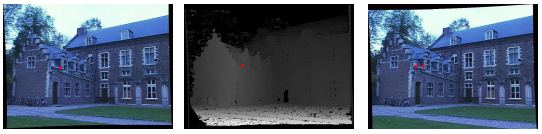
- **Stereo**: shape from "motion" between two views
- We'll need to consider:
  - Info on camera pose ("calibration")
  - Image point correspondences





Recall: Depth from disparity

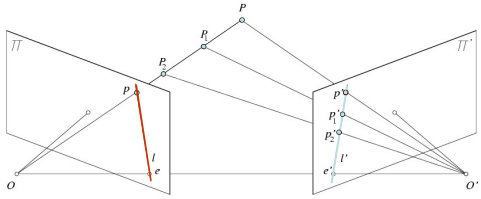
image I(x,y) | Disparity map D(x,y) | image I´(x´,y´)



(x´,y´)=(x+D(x,y), y)

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

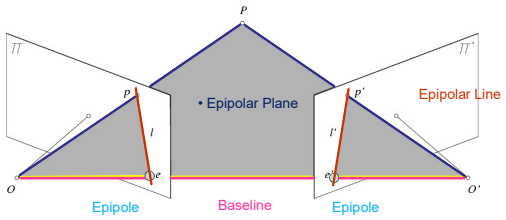
Recall: Epipolar constraint



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.

Recall: Epipolar geometry



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

Epipolar constraint

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

Image from Andrew Zisserman

Review questions

- Why perform rectification for stereo?
- What are the “extrinsic” camera parameters relating two stereo cameras?

What is the relationship between the two cameras?

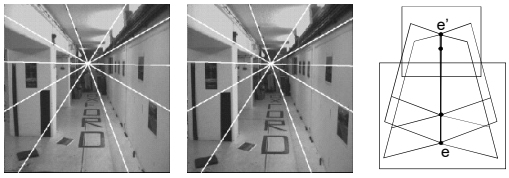
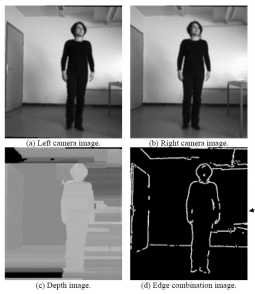


Figure from Hartley & Zisserman

Review questions

- Why perform rectification for stereo?
- What are the "extrinsic" camera parameters relating two stereo cameras?
- What's the result of convolving a disparity map with [-1 1]?

Depth for segmentation



(a) Left camera image. (b) Right camera image.

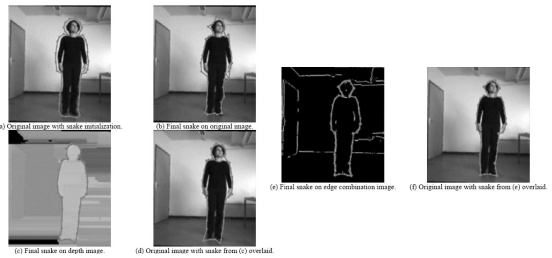
(c) Depth image. (d) Edge combination image.

Edges in disparity in conjunction with image edges enhances contours found

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

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

Depth for segmentation

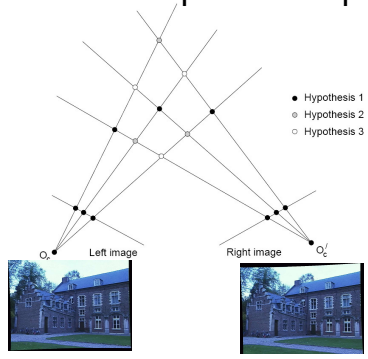


Danijela Markovic and Margrit Gelautz, Interactive Media Systems Group, Vienna University of Technology

Outline

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

Correspondence problem



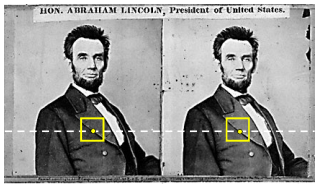
Multiple match hypotheses satisfy epipolar constraint, but which is correct?

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)

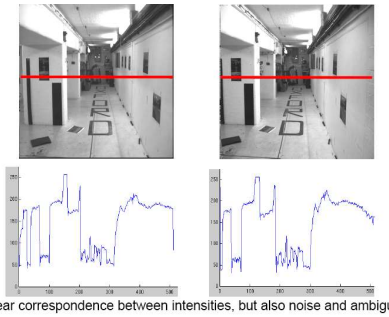
Adapted from Li Zhang

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: $A \leftarrow A - < A >, B \leftarrow B - < B >$

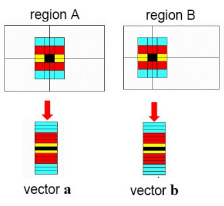
$$\text{NCC} = \frac{\sum_i \sum_j A(i,j)B(i,j)}{\sqrt{\sum_i \sum_j A(i,j)^2}\sqrt{\sum_i \sum_j B(i,j)^2}}$$

Write regions as vectors

$\mathrm{A} \rightarrow \mathbf{a}, \ \mathrm{B} \rightarrow \mathbf{b}$

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

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

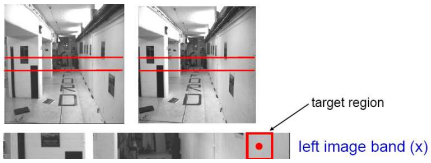


Source: Andrew Zisserman

Correlation-based window matching

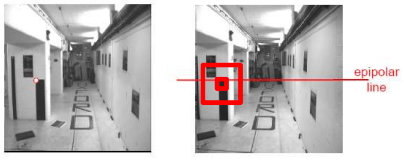
Source: Andrew Zisserman

Textureless regions



Source: Andrew Zisserman

Effect of window size?



Source: Andrew Zisserman

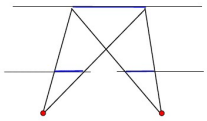
Effect of window size

W = 3 W = 20

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

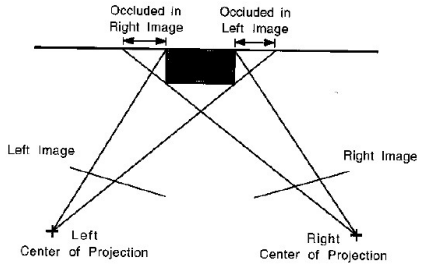


fronto-parallel surface

imaged length the same

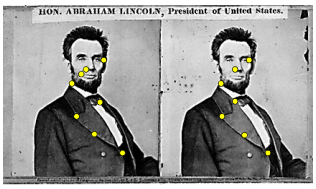
Source: Andrew Zisserman

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 for opaque objects

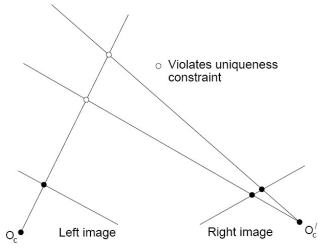
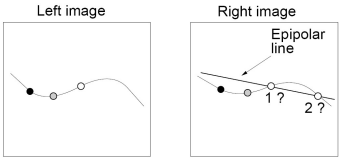


Figure from Gee & Cipolla 1999

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** will typically be in same order in both views

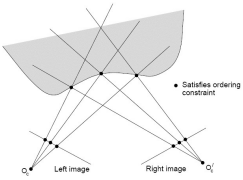
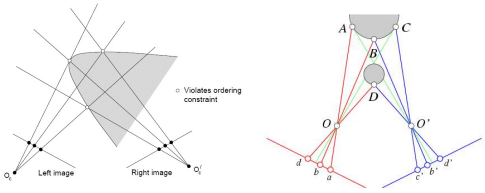


Figure from Gee & Cipolla 1999

Ordering constraint

- Won't always hold, e.g. consider transparent object, or an occluding surface



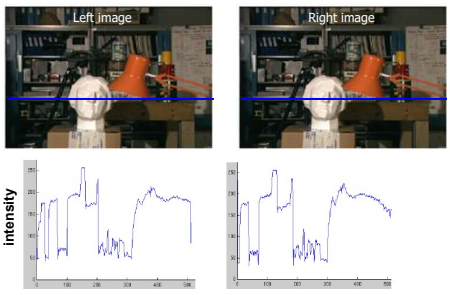
Figures from Forsyth & Ponce

Beyond matched pairs

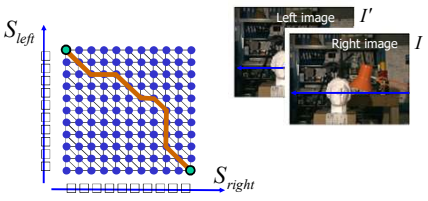
- Optimize correspondence assignments jointly
  - Scanline at a time (DP)
  - Full 2D grid (graph cuts)

Scanline stereo

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



"Shortest paths" for scan-line stereo

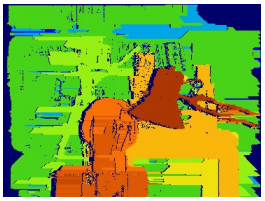


Can be implemented with dynamic programming
Ohta & Kanade '85, Cox et al. '96

Slide credit: Y. Boykov

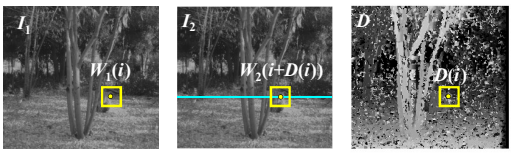
Coherent stereo on 2D grid

- Scanline stereo generates streaking artifacts



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

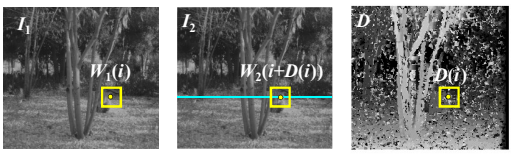
Stereo matching as energy minimization



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

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

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

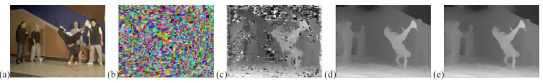


Figure 6: *Sample results from stereo reconstruction stage: (a) input color image; (b) color-based segmentation; (c) initial disparity estimates $\hat{d}_{xy}$; (d) refined disparity estimates; (e) smoothed disparity estimates $d_x(x)$.*
(f) A depth-matted object from earlier in the sequence is inserted into the video.

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

Virtual viewpoint video

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

http://research.microsoft.com/IVM/VVV/

Video examples

- https://www.youtube.com/watch?v=sz0UbHvEttI
- https://www.youtube.com/watch?v=kelirXrRb1k
- https://www.youtube.com/watch?v=1dT9Gwx1gVM
- https://www.youtube.com/watch?v=cizgVZ8rjKA

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