Stereo

11/02/2012
CS129, Brown
James Hays
Multiple views

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

Hartley and Zisserman

Lowe
Why multiple views?

- Structure and depth are inherently ambiguous from single views.
Why multiple views?

- Structure and depth are inherently ambiguous from single views.
• What cues help us to perceive 3d shape and depth?
Shading

[Figure from Prados & Faugeras 2006]
Focus/defocus

Images from same point of view, different camera parameters

3d shape / depth estimates

[figs from H. Jin and P. Favaro, 2002]
Texture

Perspective effects
Motion

Figures from L. Zhang

http://www.brainconnection.com/teasers/?main=illusion/motion-shape
Estimating depth with stereo

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

Two cameras, simultaneous views

Single moving camera and static scene
Camera parameters

**Extrinsic** parameters:
Camera frame 1 $\leftrightarrow$ Camera frame 2

**Intrinsic** parameters:
Image coordinates relative to camera $\leftrightarrow$ 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.
Geometry for a simple stereo system

• First, assuming parallel optical axes, known camera parameters (i.e., calibrated cameras):
image point (left)

Focal length

optical center (left)

baseline

image point (right)

Depth of $p$

optical center (right)

World point

http://www.cse.psu.edu/~zyin/Demo/Stereo%20geometry.jpg
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
Depth from disparity

\[
\text{image } I(x,y) \quad \text{Disparity map } D(x,y) \quad \text{image } I´(x´,y´)
\]

\[
(x´, y´) = (x + D(x,y), y)
\]

So if we could find the corresponding points in two images, we could estimate relative depth…
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
Geometry of two views constrains where the corresponding pixel for some image point in the first view must occur in the second view.

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

http://www.ai.sri.com/~luong/research/Meta3DViewer/EpipolarGeo.html
**Epipolar geometry: terms**

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

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

*Why is the epipolar constraint useful?*
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.

2.
Example: converging cameras

Figure from Hartley & Zisserman
Example: parallel cameras

Where are the epipoles?
Example: Forward motion

What would the epipolar lines look like if the camera moves directly forward?
Example: Forward motion

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

- Reproject image planes onto a common plane parallel to the line between camera centers.

- Pixel motion is horizontal after this transformation.

- Two homographies (3x3 transform), one for each input image reprojection.

Rectification example
The correspondence problem

- Epipolar geometry constrains our search, but we still have a difficult correspondence problem.
Basic stereo matching algorithm

- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- For each pixel $x$ in the first image
  - Find corresponding epipolar scanline in the right image
  - Examine all pixels on the scanline and pick the best match $x'$
  - Compute disparity $x - x'$ and set $\text{depth}(x) = \frac{f_B}{x - x'}$
Correspondence search

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

Left

scanline

Right

SSD
Correspondence search

scanline

Left

Right

Norm. corr
Effect of window size

- Smaller window
  + More detail
  - More noise

- Larger window
  + Smoother disparity maps
  - Less detail
Failures of correspondence search

Textureless surfaces

Occlusions, repetition

Non-Lambertian surfaces, specularities
Results with window search

Data

Window-based matching

Ground truth
What defines a good stereo correspondence?

1. Match quality
   - Want each pixel to find a good match in the other image

2. Smoothness
   - If two pixels are adjacent, they should (usually) move about the same amount
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
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

\[ E(D) = \sum_{i} \left( W_1(i) - W_2(i + D(i)) \right)^2 + \lambda \sum_{\text{neighbors } i, j} \rho(D(i) - D(j)) \]

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

Many of these constraints can be encoded in an energy function and solved using graph cuts.


For the latest and greatest: [http://www.middlebury.edu/stereo/](http://www.middlebury.edu/stereo/)
Active stereo with structured light

- Project “structured” light patterns onto the object
  - Simplifies the correspondence problem
  - Allows us to use only one camera

Laser scanning

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

Optical triangulation

Digital Michelangelo Project
http://graphics.stanford.edu/projects/mich/
Kinect: Structured infrared light

Potential matches for $x$ have to lie on the corresponding line $l'$. 

Potential matches for $x'$ have to lie on the corresponding line $l$. 

Summary: Key idea: Epipolar constraint
Summary

• Epipolar geometry
  – Epipoles are intersection of baseline with image planes
  – Matching point in second image is on a line passing through its epipole
  – Fundamental matrix maps from a point in one image to a line (its epipolar line) in the other
  – Can solve for F given corresponding points (e.g., interest points)

• Stereo depth estimation
  – Estimate disparity by finding corresponding points along scanlines
  – Depth is inverse to disparity