Depth Estimation (Sz 11)
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.
Why Stereo Vision?

\[ P = (X, Y, Z) \]

\[
x = f \frac{X}{Z} = f \frac{kX}{kZ}
\]

\[
y = f \frac{Y}{Z} = f \frac{kY}{kZ}
\]

**Fundamental Ambiguity:**

Any point on the ray \( OP \) has image \( p \)
Why Stereo Vision?

A second camera can resolve the ambiguity, enabling measurement of depth via triangulation.
WHAT CUES HELP US TO PERCEIVE DEPTH (AND 3D SHAPE IN GENERAL)?
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
Higher-level Depth Cues

Perspective
(vanishing points)
Higher-level Depth Cues

Similar sized objects appear smaller at a distance (this is also related to perspective)
Higher-level Depth Cues

Occluded contours (perceptual completion)
Motion

Figures from L. Zhang

http://www.brainconnection.com/teasers/?main=illusion/motion-shape
Stereo vision

Two cameras, simultaneous views

Single moving camera and static scene
Stereo Vision

• Not that important for humans, especially at longer distances. Perhaps 10% of people are stereo blind.

• Many animals don’t have much stereo overlap in their fields of view
Public Library, Stereoscopic Looking Room, Chicago, by Phillips, 1923
Teesta suspension bridge-Darjeeling, India
Woman getting eye exam during immigration procedure at Ellis Island, c. 1905 - 1920, UCR Museum of Phography
Yes, you can be stereoblind.
Random dot stereograms

• Julesz 1960:
  Do we identify local brightness patterns before fusion (monocular process) or after (binocular)?

• Think Pair Share – yes / no? how to test?
Random dot stereograms

- Julesz 1960:
  Do we identify local brightness patterns before fusion (monocular process) or after (binocular)?

- To test: pair of synthetic images obtained by randomly spraying black dots on white objects
Random dot stereograms
Random dot stereograms
1. Create an image of suitable size. Fill it with random dots. Duplicate the image.

2. Select a region in one image.

3. Shift this region horizontally by a small amount. The stereogram is complete.

Random dot stereograms

• When viewed monocularly, they appear random; when viewed stereoscopically, see 3d structure.
• Human binocular fusion not directly associated with the physical retinas; must involve the central nervous system (V2, for instance).
• Imaginary “cyclopean retina” that combines the left and right image stimuli as a single unit.
• High level scene understanding not required for stereo...but, high level scene understanding is arguably better than stereo.
Stereo attention is weird wrt. mind’s eye

[Li Zhaoping, Understanding Vision]
Autostereograms – ‘Magic Eye’

Exploit disparity as depth cue using single image.

(Single image random dot stereogram, Single image stereogram)
Autostereograms

Images from magiceye.com
Stereograms
Stereo Depth Estimation
Geometry for a simple stereo system

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

Two cameras, simultaneous views
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

So if we could find the corresponding points in two images, we could estimate relative depth…
Basic stereo matching algorithm

- 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 $x'$
    - Use individual pixels? Use patches! This should look familiar...
    - Compute disparity $x - x'$ and set $\text{depth}(x) = fB/(x - x')$
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
Correlation-based window matching
Correlation-based window matching

left image band \((x)\)

right image band \((x')\)
Correlation-based window matching

left image band (x)

right image band (x')

cross correlation

disparity = x' - x
Correlation-based window matching
Correlation-based window matching

Textureless regions are non-distinct; high ambiguity for matches.
Think-Pair-Share

WHEN WILL THIS WORK/NOT WORK?
Stereo reconstruction pipeline

What will cause errors?
- Camera calibration errors
- Poor image resolution
- Occlusions
- Violations of brightness constancy (specular reflections)
- Large motions
- Low-contrast image regions
Failures of correspondence search

Textureless surfaces

Occlusions, repetition

Non-Lambertian surfaces, specularities
Effect of window size

- Smaller window
  - More detail
  - More noise

- Larger window
  - Smoother disparity maps
  - Less detail

**Adaptive window approach**


Stereo results

- Data from University of Tsukuba
- Similar results on other images without ground truth

Scene

Ground truth
Results with window search

Window-based matching | Ground truth

Data
Correspondence problem

• There are “soft” constraints to help identify corresponding points
  • Similarity
  • Disparity gradient – depth doesn’t change too quickly.
  • Uniqueness
  • Ordering
Uniqueness constraint

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

Figure from Gee & Cipolla 1999
Problem: Occlusion

- Uniqueness says “up to match” per pixel
- When is there no match?

Occluded pixels
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.
Ordering constraint

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

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

Figures from Forsyth & Ponce
Stereo as energy minimization

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

Energy functions of this form can be minimized using graph cuts

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

- **data term**
- **smoothness term**
Stereo matching as energy minimization

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

\[ E_{\text{data}} = \sum_i (W_1(i) - W_2(i + D(i)))^2 \]

\[ E_{\text{smooth}} = \sum_{\text{neighbors } i, j} \rho(D(i) - D(j)) \]

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
Constraints 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/)
Depth from disparity

\[ \text{disparity} = x - x' = \frac{\text{baseline} \times f}{z} \]
Active stereo: project structured light onto object

Simplifies the correspondence problem

Li Zhang's one-shot stereo

Active stereo with structured light

Sequence of known expected appearance across surface
How does a depth camera work?

Microsoft Kinect v1

Intel laptop depth camera
Kinect: Structured infrared light

How does a depth camera work?

Stereo in infrared.
Laser scanning

Optical triangulation
- 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

Digital Michelangelo Project
http://graphics.stanford.edu/projects/mich/
Laser scanned models

The Digital Michelangelo Project, Levoy et al.
Laser scanned models

*The Digital Michelangelo Project*, Levoy et al.
Laser scanned models

The Digital Michelangelo Project, Levoy et al.
Laser scanned models

The Digital Michelangelo Project, Levoy et al.
Time of Flight (Kinect V2)

- Depth cameras in HoloLens use *time of flight*
  - “SONAR for light”
  - Emit light of a known wavelength, and time how long it takes for it to come back
Time of Flight

Humans can learn to echolocate!
Break here.
How do we see the world?

Let’s design a camera

- Idea 1: put a piece of film in front of an object
- Do we get a reasonable image?
Pinhole camera

Add a barrier to block off most of the rays

- This reduces blurring
- The opening known as the **aperture**
- How does this transform the image?
Pinhole camera model

Pinhole model:

- Captures **pencil of rays** – all rays through a single point
- The point is called **Center of Projection (COP)**
- The image is formed on the **Image Plane**
- **Effective focal length** \( f \) is distance from COP to Image Plane
Dimensionality Reduction Machine (3D to 2D)

What have we lost?

- Angles
- Depth, lengths
Funny things happen...
Lengths can’t be trusted...

Figure by David Forsyth
...but humans adopt!

Müller-Lyer Illusion

We don’t make measurements in the image plane

http://www.michaelbach.de/ot/sze_muelue/index.html
Modeling projection

The coordinate system

- We will use the pin-hole model as an approximation
- Put the optical center (Center Of Projection) at the origin
- Put the image plane (Projection Plane) in front of the COP
  - Why?
- The camera looks down the negative z axis
  - we need this if we want right-handed-coordinates
Modeling projection

Projection equations

- Compute intersection with PP of ray from \((x,y,z)\) to COP
- Derived using similar triangles

\[
(x, y, z) \rightarrow \left( -d \frac{x}{z}, -d \frac{y}{z}, -d \right)
\]

- We get the projection by throwing out the last coordinate:

\[
(x, y, z) \rightarrow \left( -d \frac{x}{z}, -d \frac{y}{z} \right)
\]
Homogeneous coordinates

Is this a linear transformation?
  • no—division by z is nonlinear

Trick: add one more coordinate:

\[
(x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad \text{homogeneous image coordinates}
\]

\[
(x, y, z) \Rightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad \text{homogeneous scene coordinates}
\]

Converting *from* homogeneous coordinates

\[
\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow \left( \frac{x}{w}, \frac{y}{w} \right)
\]

\[
\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow \left( \frac{x}{w}, \frac{y}{w}, \frac{z}{w} \right)
\]

Slide by Steve Seitz
Perspective Projection

Projection is a matrix multiply using homogeneous coordinates:

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & -1/d & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z \\
1
\end{bmatrix} = \begin{bmatrix}
x \\
y \\
z/d \\
-1
\end{bmatrix} \Rightarrow \left( -\frac{d}{z}, \frac{y}{z} \right)
\]

divide by third coordinate

This is known as **perspective projection**

- The matrix is the **projection matrix**
- Can also formulate as a 4x4

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & -1/d & 0
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z \\
1
\end{bmatrix} = \begin{bmatrix}
x \\
y \\
z \\
-z/d
\end{bmatrix} \Rightarrow \left( -\frac{d}{z}, \frac{y}{z} \right)
\]

divide by fourth coordinate
Orthographic Projection

Special case of perspective projection

• Distance from the COP to the PP is infinite

• Also called “parallel projection”
• What’s the projection matrix?

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z \\
1
\end{bmatrix} =
\begin{bmatrix}
x \\
y \\
1
\end{bmatrix} \Rightarrow (x, y)
\]
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.
Estimating depth with stereo

**Stereo**: shape from “motion” between two views

We’ll need to consider:
- Info on camera pose (“calibration”)
- Image point correspondences
General case, with calibrated cameras

- The two cameras need not have parallel optical axes.

Vs.
Given $p$ in left image, where can corresponding point $p'$ be?

Stereo correspondence constraints
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

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
Think Pair Share

Where are the epipoles?
What do the epipolar lines look like?

a) ![Diagram a](image)
b) ![Diagram b](image)
c) ![Diagram c](image)
d) ![Diagram d](image)

- red dot = camera center
Example: converging cameras

Figure from Hartley & Zisserman
Example: parallel cameras

Where are the epipoles?
Example: Forward motion

Epipole has same coordinates in both images.
Points move along lines radiating from e:
“Focus of expansion”
Example: Forward motion

Epipole has same coordinates in both images.
Points move along lines radiating from e: “Focus of expansion”
**Fundamental matrix**

Let $x$ be a point in left image, $x'$ in right image

**Epipolar relation**
- $x$ maps to epipolar line $l'$
- $x'$ maps to epipolar line $l$

Epipolar mapping described by a 3x3 matrix $F$:

$$l' = Fx$$
$$l = F^T x'$$

It follows that: $$x'Fx = 0$$
Fundamental matrix

This matrix $F$ is called

- the “Essential Matrix”
  - when image intrinsic parameters are known
- the “Fundamental Matrix”
  - more generally (uncalibrated case)

Can solve for $F$ from point correspondences

- Each $(x, x')$ pair gives one linear equation in entries of $F$

$$x'Fx = 0$$

- $F$ has 9 entries, but really only 7 degrees of freedom.
- With 8 points it is simple to solve for $F$, but it is also possible with 7. See Marc Pollefey’s notes for a nice tutorial
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
Summary: Key idea: Epipolar constraint

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

• 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
Stereo reconstruction pipeline

• Steps
  – Calibrate cameras
  – Rectify images
  – Compute disparity
  – Estimate depth