Epipolar geometry & stereo vision  
Tuesday, Oct 21  
Kristen Grauman  
UT-Austin

Recap: Features and filters  
Transforming and describing images; textures, colors, edges

Recap: Grouping & fitting  
Clustering, segmentation, fitting; what parts belong together?

Now: Multiple views  
Multi-view geometry, matching, invariant features, stereo vision

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

Images from Lana Lazebnik
• What cues help us to perceive 3d shape and depth?

Shading

[Figure from Prados & Faugeras 2006]

Focus/Defocus

[Figure from H. Jin and P. Favaro, 2002]

Texture

[From M. L. Luo, The recovery of 3D structure using visual texture patterns, PhD thesis]

Perspective effects

[Image credit: S. Seitz]

Motion

Figures from L. Zhang http://www.brainconnection.com/wesley/visualillusionshasbeta
Estimating scene shape

- Shape from X: Shading, Texture, Focus, Motion...
- Stereo:
  - shape from "motion" between two views
  - infer 3d shape of scene from two (multiple) images from different viewpoints

Today

- Human stereopsis
- Stereograms
- Epipolar geometry and the epipolar constraint
  - Case example with parallel optical axes
  - General case with calibrated cameras
- Stereopsis
  - Finding correspondences along the epipolar line

Fixation, convergence

Human stereopsis: disparity

Disparity occurs when eyes fixate on one object; others appear at different visual angles

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

• When viewed monocularly, they appear random; when viewed stereoscopically, see 3d structure.

• Conclusion: human binocular fusion not directly associated with the physical retinas; must involve the central nervous system

• Imaginary “cyclopean retina” that combines the left and right image stimuli as a single unit

Autostereograms

Exploit disparity as depth cue using single image

(Single image random dot stereogram, Single image stereogram)
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

http://www.johnsonshawmuseum.org

Depth with stereo: basic idea

Source: Steve Seitz
Depth with stereo: basic idea

Basic Principle: Triangulation
- Gives reconstruction as intersection of two rays
- Requires
  - camera pose (calibration)
  - point correspondence

Source: Steve Seitz

Camera calibration

Extrinsic parameters: Camera frame ↔ Reference frame
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.

Today

- Human stereopsis
- Stereograms
- Epipolar geometry and the epipolar constraint
  - Case example with parallel optical axes
  - General case with calibrated cameras
- Stereopsis
  - Finding correspondences along the epipolar line

Geometry for a simple stereo system

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

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 \((O_l, P, O_r)\):

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

\[
Z = f \frac{T}{x_l - x_r}
\]

disparity
**Depth from disparity**

Image \( I(x,y) \)  
Disparity map \( D(x,y) \)  
Image \( I'(x',y') \)  

\((x',y') = (x + D(x,y), y)\)

---

**General case, with calibrated cameras**

- The two cameras need not have parallel optical axes.

---

**Stereo correspondence constraints**

Given \( p \) in left image, where can corresponding point \( p' \) be?

---

**Epipolar constraint**

Why is this useful?

- Reduces correspondence problem to 1D search along conjugate epipolar lines

---

Adapted from Steve Seitz
Epipolar geometry

- **Baseline**: line joining the camera centers
- **Epipole**: point of intersection of baseline with the 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

---

Epipolar constraint

- Potential matches for \( p \) have to lie on the corresponding epipolar line \( l' \).
- Potential matches for \( p' \) have to lie on the corresponding epipolar line \( l \).

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

Source: M. Pollefeys

---

Example

- Example: converging cameras
- Example: motion parallel with image plane

Figure from Hartley & Zisserman

---
Example: forward motion

Epipole has same coordinates in both images.
Points move along lines radiating from e: "Focus of expansion"

Figure from Hartley & Zisserman

For a given stereo rig, how do we express the epipolar constraints algebraically?

Stereo geometry, with calibrated cameras

If the 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; translation: 3 vector.

Rotation matrix

Express 3d rotation as series of rotations around coordinate axes by angles $\alpha, \beta, \gamma$

Overall rotation is product of these elementary rotations:

$R = R_z R_y R_x$

3d rigid transformation

$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$

$X' = RX + T$

Stereo geometry, with calibrated cameras

Camera-centered coordinate systems are related by known rotation $R$ and translation $T$:

$X' = RX + T$
Cross product

\[ \vec{a} \times \vec{b} = \vec{c} \quad \vec{a} \cdot \vec{c} = 0 \quad \vec{b} \cdot \vec{c} = 0 \]

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

So here, \( \vec{c} \) is perpendicular to both \( \vec{a} \) and \( \vec{b} \), which means the dot product = 0.

Matrix form of cross product

\[
\begin{bmatrix}
0 & -a_z & a_y \\
-a_x & 0 & a_z \\
a_x & -a_y & 0
\end{bmatrix} \begin{bmatrix}
a_x \\
a_y \\
a_z
\end{bmatrix} = \begin{bmatrix}
\vec{a} \\
\vec{b}
\end{bmatrix} = \vec{c}
\]

Can be expressed as a matrix multiplication.

```latex
\[ \vec{a} \times \vec{b} = [a_i] \vec{b} \]
```

Essential matrix

\[
\begin{align*}
X' &= RX + T \\
X' &= T \times RX + T \times T \\
&= T \times RX
\end{align*}
\]

Let \( E = T \cdot R \)

\[ X'^T EX = 0 \]

This holds for the rays \( p \) and \( p' \) that are parallel to the camera-centered position vectors \( X \) and \( X' \), so we have:

\( E \) is called the **essential matrix**, which relates corresponding image points [Longuet-Higgins 1981]

Essential matrix and epipolar lines

\[ p^T E p = 0 \]

Epipolar constraint: if we observe point \( p \) in one image, then its position \( p' \) in second image must satisfy this equation.

\( E' \) is the coordinate vector representing the epipolar line associated with point \( p' \)
Essential matrix: properties

- Relates image of corresponding points in both cameras, given rotation and translation
- Assuming intrinsic parameters are known

\[ E = T \cdot R \]

Essential matrix example: parallel cameras

\[
R = \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

\[ T = \begin{bmatrix}
s & 0 & 0 \\
0 & s & 0 \\
0 & 0 & 1
\end{bmatrix}^T \]

\[ E = [T_1]R = \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{bmatrix} \]

\[ p^T E p = 0 \]

\[
\begin{bmatrix}
x' \\
y' \\
df
\end{bmatrix} = 0
\]

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

Image \((x,y)\)  Disparity map \(D(x,y)\)  Image \(I(x',y')\)

\((x',y') = (x+D(x,y), y)\)

What about when cameras’ optical axes are not parallel?

Stereo image rectification

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

Stereo image rectification: example

Today

- Human stereopsis
- Stereograms
- Epipolar geometry and the epipolar constraint
  - Case example with parallel optical axes
  - General case with calibrated cameras
- Stereopsis
  - Finding correspondences along the epipolar line
Stereo reconstruction: main steps

- Calibrate cameras
- Rectify images
- Compute disparity
- Estimate depth

Correspondence problem

- 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

Additional correspondence constraints

- Similarity
- Uniqueness
- Ordering
- Disparity gradient

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)

Example: window search

Data from University of Tsukuba
Example: window search

![Window search images](image1)

Effect of window size

![Window size images](image2)

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

Sparse correspondence search

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

What would make good features?

Dense vs. sparse

- Sparse
  - Efficiency
  - Can have more reliable feature matches, less sensitive to illumination than raw pixels
  - ...But, have to know enough to pick good features; sparse info
- Dense
  - Simple process
  - More depth estimates, can be useful for surface reconstruction
  - ...But, breaks down in textureless regions anyway, raw pixel distances can be brittle, not good with very different viewpoints

Difficulties in similarity constraint

- Untextured surfaces
- Occlusions

Uniqueness

- For opaque objects, up to one match in right image for every point in left image

Figure from Gee & Cipolla 1999
Ordering

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

Disparity gradient

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

Additional correspondence constraints

- Similarity
- Uniqueness
- Ordering
- Disparity gradient

Epipolar lines constrain the search to a line, and these appearance and ordering constraints further reduce the possible matches.

Possible sources of error?

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

Stereo reconstruction: main steps

- Calibrate cameras
- Rectify images
- Compute disparity
- Estimate depth

Stereo in machine vision systems

Left: The Stanford cart sports a single camera moving in discrete increments along a straight line and providing multiple snapshots of outdoor scenes.
Right: The INRIA mobile robot uses three cameras to map its environment.
Depth for segmentation

Edges in disparity in conjunction with image edges enhances contours found.

Virtual viewpoint video


Next

- Uncalibrated cameras
- Robust fitting