CS4670 / 5670: Computer Vision
Kavita Bala

Lec 22: Stereo
Road map

• What we’ve seen so far:
  – Low-level image processing: filtering, edge detecting, feature detection
  – Geometry: image transformations, panoramas, single-view modeling Fundamental matrices

• What’s next:
  – Finishing up geometry: multi view stereo, structure from motion
  – Recognition
  – Image formation
Announcements

• Wed: photometric stereo

• No class Dragon Day
Fundamental matrix result

\[ q^T F p = 0 \]

(Longuet-Higgins, 1981)
Properties of the Fundamental Matrix

- $Fp$ is the epipolar line associated with $p$
- $F^Tq$ is the epipolar line associated with $q$
- $Fe_1 = 0$ and $F^Te_2 = 0$
- $F$ is rank 2
Estimating $\mathbf{F}$

• If we don’t know $K_1$, $K_2$, $R$, or $t$, can we estimate $\mathbf{F}$ for two images?

• Yes, given enough correspondences
Estimating F – 8-point algorithm

• The fundamental matrix $F$ is defined by

$$x'^{T}Fx = 0$$

for any pair of matches $x$ and $x'$ in two images.

• Let $x=(u,v,1)^{T}$ and $x'=(u',v',1)^{T}$,

$$F = \begin{bmatrix}
    f_{11} & f_{12} & f_{13} \\
    f_{21} & f_{22} & f_{23} \\
    f_{31} & f_{32} & f_{33}
\end{bmatrix}$$

each match gives a linear equation

$$uu'f_{11} + vu'f_{12} + u'f_{13} + uv'f_{21} + vv'f_{22} + v'f_{23} + uf_{31} + vf_{32} + f_{33} = 0$$
8-point algorithm

\[
\begin{bmatrix}
u_1 u_1' & v_1 u_1' & u_1' & u_1 v_1' & v_1 v_1' & v_1' & u_1 & v_1 & 1 \\
u_2 u_2' & v_2 u_2' & u_2' & u_2 v_2' & v_2 v_2' & v_2' & u_2 & v_2 & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
u_n u_n' & v_n u_n' & u_n' & u_n v_n' & v_n v_n' & v_n' & u_n & v_n & 1 \\
\end{bmatrix}
\begin{bmatrix}
f_{11} \\
f_{12} \\
f_{13} \\
f_{21} \\
f_{22} \\
f_{23} \\
f_{31} \\
f_{32} \\
f_{33} \\
\end{bmatrix}
= 0
\]

- In reality, instead of solving \( Af = 0 \), we seek \( f \) to minimize \( \| Af \| \), least eigenvector of \( A^TA \).
8-point algorithm – Problem?

• \( F \) should have rank 2

• To enforce that \( F \) is of rank 2, \( F \) is replaced by \( F' \) that minimizes \( \|F - F'\| \) subject to the rank constraint.

• This is achieved by SVD. Let \( F = U\Sigma V^T \), where

\[
\Sigma = \begin{bmatrix}
\sigma_1 & 0 & 0 \\
0 & \sigma_2 & 0 \\
0 & 0 & \sigma_3 \\
\end{bmatrix}, \quad \Sigma' = \begin{bmatrix}
\sigma_1 & 0 & 0 \\
0 & \sigma_2 & 0 \\
0 & 0 & 0 \\
\end{bmatrix}
\]

then \( F' = U\Sigma'V^T \) is the solution.
8-point algorithm

- Pros: it is linear, easy to implement and fast
- Cons: susceptible to noise

- Normalized 8-point algorithm: Hartley
What about more than two views?

• The geometry of three views is described by a $3 \times 3 \times 3$ tensor called the *trifocal tensor*

• The geometry of four views is described by a $3 \times 3 \times 3 \times 3$ tensor called the *quadrifocal tensor*

• After this it starts to get complicated...
  – Structure from motion
Stereo reconstruction pipeline

• Steps
  – Calibrate cameras
  – Rectify images
  – Compute correspondence (and hence disparity)
  – Estimate depth
Correspondence algorithms

Algorithms may be classified into two types:

1. Dense: compute a correspondence at every pixel
2. Sparse: compute correspondences only for features
Example image pair – parallel cameras
Dense correspondence algorithm

Search problem (geometric constraint): for each point in left image, corresponding point in right image lies on the epipolar line (1D ambiguity)

Disambiguating assumption (photometric constraint): the intensity neighborhood of corresponding points are similar across images

Measure similarity of neighborhood intensity by cross-correlation
Intensity profiles

- Clear correspondence, but also noise and ambiguity
Normalized Cross Correlation

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

regions as vectors

\[ A \rightarrow a, \quad B \rightarrow b \]

\[ NCC = \frac{a \cdot b}{|a||b|} \]

\[ -1 \leq NCC \leq 1 \]
Cross-correlation of neighborhood

regions A, B, write as vectors $a, b$

translate so that mean is zero

$a \rightarrow a - \langle a \rangle, \ b \rightarrow b - \langle b \rangle$

cross correlation $= \frac{a \cdot b}{|a||b|}$

Invariant to $I \rightarrow \alpha I + \beta$
Why is cross-correlation such a poor measure?

1. The neighborhood region does not have a “distinctive” spatial intensity distribution

2. Foreshortening effects

- **fronto-parallel surface**: imaged length the same
- **slanting surface**: imaged lengths differ
Limitations of similarity constraint

Textureless surfaces

Occlusions, repetition

Non-Lambertian surfaces, specularities
Other approaches to obtaining 3D structure
Active stereo with structured light

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

Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming.
3DPVT 2002
Active stereo with structured light


Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming. 3DPVT 2002
Microsoft Kinect
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/

Source: S. Seitz
Laser scanned models

The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz
Laser scanned models

*The Digital Michelangelo Project*, Levoy et al.

Source: S. Seitz
The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz
The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz
The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz
Aligning range images

• A single range scan is not sufficient to describe a complex surface
• Need techniques to register multiple range images

• … which brings us to multi-view stereo
Quiz