Stereo Matching
Fundamental matrix

Let \( p \) be a point in left image, \( p' \) in right image

Epipolar relation

- \( p \) maps to epipolar line \( l' \)
- \( p' \) maps to epipolar line \( l \)

Epipolar mapping described by a 3x3 matrix \( F \)

\[
l' = Fp
\]
\[
l = p'F
\]

It follows that

\[
p'Fp = 0
\]
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 \((p, p')\) pair gives one linear equation in entries of F

\[ p' F p = 0 \]

- F has 9 entries, but really only 7 or 8 degrees of freedom.
- With 8 points it is simple to solve for F, but it is also possible with 7. See Marc Pollefeys’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
The correspondence problem

• Epipolar geometry constrains our search, but we still have a difficult correspondence problem.
Photo Tourism
Exploring photo collections in 3D

Noah Snavely    Steven M. Seitz    Richard Szeliski
University of Washington    Microsoft Research

SIGGRAPH 2006
The Visual Turing Test for Scene Reconstruction
Supplementary Video

Qi Shan⁺  Riley Adams⁺  Brian Curless⁺
Yasutaka Furukawa*  Steve Seitz**

⁺University of Washington  *Google

3DV 2013
Despite their scale invariance and robustness to appearance changes, SIFT features are local and do not contain any global information about the image or about the location of other features in the image. Thus feature matching based on SIFT features is still prone to errors. However, since we assume that we are dealing with rigid scenes, there are strong geometric constraints on the locations of the matching features and these constraints can be used to clean up the matches. In particular, when a rigid scene is imaged by two pinhole cameras, there exists a $3 \times 3$ matrix $F$, the Fundamental matrix, such that corresponding points $x_{ij}$ and $x_{ik}$ (represented in homogeneous coordinates) in two images $j$ and $k$ satisfy:

$$x_{ij}^T F x_{ij} = 0. \quad (3)$$

A common way to impose this constraint is to use a greedy randomized algorithm to generate suitably chosen random estimates of $F$ and choose the one that has the largest support among the matches, i.e., the one for which the most matches satisfy (3). This algorithm is called Random Sample Consensus (RANSAC) and is used in many computer vision problems.
Sparse to Dense Correspondence

Building Rome in a Day
By Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz, Richard Szeliski
Communications of the ACM, Vol. 54 No. 10, Pages 105-112
Structure from motion (or SLAM)

- Given a set of corresponding points in two or more images, compute the camera parameters and the 3D point coordinates.

Slide credit: Noah Snavely
Structure from motion ambiguity

- If we scale the entire scene by some factor $k$ and, at the same time, scale the camera matrices by the factor of $1/k$, the projections of the scene points in the image remain exactly the same:

$$x = PX = \left( \frac{1}{k} P \right) (kX)$$

It is impossible to recover the absolute scale of the scene!
How do we know the scale of image content?
Bundle adjustment

- Non-linear method for refining structure and motion
- Minimizing reprojection error

\[ E(P, X) = \sum_{i=1}^{m} \sum_{j=1}^{n} D(x_{ij}, P_i X_j)^2 \]
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, normalized correlation)

Adapted from Li Zhang
Correspondence search with similarity constraint

• 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 with similarity constraint
Correspondence search with similarity constraint

Left

Right

scanline

Norm. corr
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.
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

target region

left image band \((x)\)

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

Textureless regions are non-distinct; high ambiguity for matches.
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.
Results with window search

Window-based matching
(best window size)

Ground truth
Better solutions

• Beyond individual correspondences to estimate disparities:
  • 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, Intille & Bobick, ‘01
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 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
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
Better results…

Graph cut method

For the latest and greatest: http://www.middlebury.edu/stereo/
Challenges

• Low-contrast; textureless image regions
• Occlusions
• Violations of brightness constancy (e.g., specular reflections)
• Really large baselines (foreshortening and appearance change)
• Camera calibration errors
Active stereo with structured light

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

Kinect: Structured infrared light

iPhone X
The scale of algorithm name quality

better

RANSAC
SIFT
Deep Learning
Optical Flow
Hough Transform
Neural Networks
Essential and Fundamental Matrix

worse

Dynamic Programming
Computer Vision

Motion and Optical Flow

Many slides adapted from S. Seitz, R. Szeliski, M. Pollefeys, K. Grauman and others…
Video

- A video is a sequence of frames captured over time
- Now our image data is a function of space \((x, y)\) and time \((t)\)
Motion and perceptual organization

Gestalt psychology
(Max Wertheimer, 1880-1943)
Motion and perceptual organization

- Sometimes, motion is the only cue

Gestalt psychology (Max Wertheimer, 1880-1943)
Motion and perceptual organization

• Sometimes, motion is the only cue
Motion and perceptual organization

- Sometimes, motion is the only cue
Motion and perceptual organization

- Sometimes, motion is the only cue
Motion and perceptual organization

• Even “impoverished” motion data can evoke a strong percept
Motion and perceptual organization

- Even “impoverished” motion data can evoke a strong percept
Motion and perceptual organization

Animation from:
An experimental study of apparent behavior.
American Journal of Psychology, 57, 243-262.

Courtesy of:
Department of Psychology,
University of Kansas, Lawrence.

Experimental study of apparent behavior.
Fritz Heider & Marianne Simmel. 1944
More applications of motion

- Segmentation of objects in space or time
- Estimating 3D structure
- Learning dynamical models – how things move
- Recognizing events and activities
- Improving video quality (motion stabilization)
Motion estimation techniques

- **Feature-based methods**
  - Extract visual features (corners, textured areas) and track them over multiple frames
  - Sparse motion fields, but more robust tracking
  - Suitable when image motion is large (10s of pixels)

- **Direct, dense methods**
  - Directly recover image motion at each pixel from spatio-temporal image brightness variations
  - Dense motion fields, but sensitive to appearance variations
  - Suitable for video and when image motion is small
Motion estimation: Optical flow

*Optic flow* is the *apparent* motion of objects or surfaces.

Will start by estimating motion of each pixel separately. Then will consider motion of entire image.
To be continued…