









• **Structure:** Given projections of the same 3D point in two or more images, compute the 3D coordinates of that point



• Motion: Given a set of corresponding points in two or more images, compute the camera parameters



• Stereo correspondence: Given a point in one of the images, where could its corresponding points be in the other images?



• **Optical flow:** Given two images, find the location of a world point in a second close-by image with no camera info.



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 I

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 <u>Marc Pollefey's notes</u> 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
- C. Loop and Z. Zhang. <u>Computing</u> <u>Rectifying Homographies for Stereo</u> <u>Vision</u>. IEEE Conf. Computer Vision and Pattern Recognition, 1999.



Rectification example



SIFT + Fundamental Matrix + RANSAC + Sparse correspondence

Photo Tourism Exploring photo collections in 3D

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

SIGGRAPH 2006

SIFT + Fundamental Matrix + RANSAC + dense correspondence

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 3×3 matrix *F*, the *Fundamental matrix*, such that corresponding points x_{ij} and x_{ik} (represented in homogeneous coordinates) in two images *j* and *k* satisfy¹⁰:

$$\boldsymbol{x}_{ij}^{\top} F \boldsymbol{x}_{ij} = \boldsymbol{0}. \tag{3}$$

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)⁶ and is used in many computer vision problems.

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

SIFT + Fundamental Matrix + RANSAC + 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

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

Correspondence problem



Multiple match hypotheses satisfy epipolar constraint, but which is correct?

Figure from Gee & Cipolla 1999

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)

Correspondence problem



Clear correspondence between intensities, but also noise and ambiguity

Correspondence problem





Neighborhoods of corresponding points are similar in intensity patterns.

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.

Stereo – Tsukuba test scene (now old)



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

Slide credit: Y. Boykov

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



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

$$E_{\text{data}} = \sum_{i} \left(W_1(i) - W_2(i + D(i)) \right)^2 \qquad E_{\text{smooth}} = \sum_{\text{neighbors } i, j} \rho \left(D(i) - D(j) \right)$$

Energy functions of this form can be minimized using graph cuts.

Y. Boykov, O. Veksler, and R. Zabih, <u>Fast Approximate Energy</u> <u>Minimization via Graph Cuts</u>, PAMI 2001

Source: Steve Seitz

Better results...



Graph cut method Boykov et al., <u>Fast Approximate Energy Minimization via Graph Cuts</u>, International Conference on Computer Vision, September 1999.

Ground truth

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

Challenges

- Low-contrast 'textureless' image regions
- Occlusions
- Violations of brightness constancy
 - 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



L. Zhang, B. Curless, and S. M. Seitz. <u>Rapid Shape Acquisition Using Color Structured</u> <u>Light and Multi-pass Dynamic Programming.</u> *3DPVT* 2002

Kinect: Structured infrared light



http://bbzippo.wordpress.com/2010/11/28/kinect-in-infrared/

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