Recap: A General Point

- Equations of the form
  \[ Ax = 0 \]
- How do we solve them? (always!)
  - Apply SVD
  \[ A = UDV^T = U \begin{bmatrix} d_{11} & \cdots & d_{1m} \\ \vdots & \ddots & \vdots \\ d_{m1} & \cdots & d_{mm} \end{bmatrix} V^T \]
  Singular values Singular vectors
  - Singular values of \( A \) = square roots of the eigenvalues of \( A^T A \).
  - The solution of \( Ax = 0 \) is the nullspace vector of \( A \).
  - This corresponds to the smallest singular vector of \( A \).

Recap: Camera Parameters

- Intrinsic parameters
  - Principal point coordinates
  - Focal length
  - Pixel magnification factors
  - Skew (non-rectangular pixels)
  - Radial distortion
- Extrinsic parameters
  - Rotation \( R \)
  - Translation \( t \) (both relative to world coordinate system)
- Camera projection matrix
  \[ P = K[R|t] \]
  \( \Rightarrow \) General pinhole camera: 9 DoF
  \( \Rightarrow \) CCD Camera with square pixels: 10 DoF
  \( \Rightarrow \) General camera: 11 DoF

Recap: Calibrating a Camera

Goal
- Compute intrinsic and extrinsic parameters using observed camera data.

Main idea
- Place “calibration object” with known geometry in the scene
- Get correspondences
- Solve for mapping from scene to image: estimate \( P = P_{\text{intr}} P_{\text{extr}} \)

Recap: Camera Calibration (DLT Algorithm)

\[
\begin{bmatrix}
0' & X_1' & -y_1X_1' \\
X_1' & 0' & -y_1X_1' \\
\vdots & \vdots & \vdots \\
X_m' & 0' & -y_1X_m'
\end{bmatrix}
\begin{bmatrix}
P_1 \\
P_2 \\
P_3 \\
P_4
\end{bmatrix} = 0
\]

- \( P \) has 11 degrees of freedom.
- Two linearly independent equations per independent 2D/3D correspondence.
- Solve with SVD (similar to homography estimation)
- Solution corresponds to smallest singular vector.
- \( 5 \frac{1}{2} \) correspondences needed for a minimal solution.
Recap: Triangulation - Linear Algebraic Approach

\[ \lambda_1 x_1 = P_1 X \]
\[ x_1 \times P_1 X = 0 \]
\[ [x_{1o}] P_1 X = 0 \]
\[ \lambda_2 x_2 = P_2 X \]
\[ x_2 \times P_2 X = 0 \]
\[ [x_{2o}] P_2 X = 0 \]

• Two independent equations each in terms of three unknown entries of \( X \).
• Stack equations and solve with SVD.
• This approach nicely generalizes to multiple cameras.

Topics of This Lecture

• Revisiting Epipolar Geometry
  - Calibrated case: Essential matrix
  - Uncalibrated case: Fundamental matrix
  - Weak calibration
  - Epipolar Transfer
• Active Stereo
  - Kinect sensor
  - Structured Light sensing
  - Laser scanning

Revisiting Epipolar Geometry

Let’s look again at the epipolar constraint
- For the calibrated case (but in homogenous coordinates)
- For the uncalibrated case

Epipolar Geometry: Calibrated Case

\[ E \times x' = 0 \]
\[ \text{with } E = [t_r] R \]

Essential Matrix (Longuet-Higgins, 1981)

Epipolar Geometry: Calibrated Case

\[ x \times t \times (R x') = 0 \]
\[ x^T E x' = 0 \]
\[ \text{with } E = [t_r] R \]

- \( E x' \) is the epipolar line associated with \( x' \) (\( l = E x' \))
- \( E^T x' \) is the epipolar line associated with \( x \) (\( l' = E^T x \))
- \( E e' = 0 \) and \( E^T e = 0 \)
- \( E \) is singular (rank two)
- \( E \) has five degrees of freedom (up to scale)
The calibration matrices $K$ and $K'$ of the two cameras are unknown.

We can write the epipolar constraint in terms of unknown normalized coordinates:

$$x^T E \hat{x}' = 0 \quad x = K \hat{x}, \quad x' = K' \hat{x}'$$

Let $F$ be the fundamental matrix.

Let $F_{11} \neq 0$.

Let $F_{11}$ be the reconstruction of $F$.

Estimating the Fundamental Matrix

- The fundamental matrix defines the epipolar geometry between two uncalibrated cameras.
- How can we estimate $F$ from an image pair?
  - We need correspondences...

The Eight-Point Algorithm

Let $x = (u, v, 1)^T$, $x' = (u', v', 1)^T$.

$$\sum_{i=1}^N (x_i^T F x_i') = 0$$

This minimizes:

$$\sum_{i=1}^N (x_i^T F x_i')^2$$

Solve using... SVD

Excursion: Properties of SVD

- Frobenius norm
  - Generalization of the Euclidean norm to matrices
    \[ \|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} = \sqrt{\sum_{i,j} \sigma_i^2} \]
  - Partial reconstruction property of SVD
    - Let $\sigma_i$ be the singular values of $A$.
    - Let $A_\sigma = U \Sigma V^T$ be the reconstruction of $A$ when we set $\sigma_{i\geq i}$ to zero.
    - Then $A_\sigma$ is the best rank-$\sigma$ approximation of $A$ in the sense of the Frobenius norm
      (i.e. the best least-squares approximation).
The Eight-Point Algorithm

- Problem with noisy data
  - The solution will usually not fulfill the constraint that \( F \) only has rank 2.
  \[ \Rightarrow \text{There will be no epipoles through which all epipolar lines pass!} \]
- Enforce the rank-2 constraint using SVD
  \[ F = U D V^T = U \begin{bmatrix} d_{i1} & \cdots & d_{i3} \end{bmatrix} \begin{bmatrix} v_1 & \cdots & v_3 \end{bmatrix} \]
- As we have just seen, this provides the best least-squares approximation to the rank-2 solution.

In practice, this often looks as follows:

- Poor numerical conditioning
- Can be fixed by rescaling the data

Set \( d_{i3} \) to zero and reconstruct \( F \)

The Normalized Eight-Point Algorithm

1. Center the image data at the origin, and scale it so the mean squared distance between the origin and the data points is 2 pixels.
2. Use the eight-point algorithm to compute \( F \) from the normalized points.
3. Enforce the rank-2 constraint using SVD.
   \[ F = U D V^T \begin{bmatrix} d_{i1} & \cdots & d_{i3} \end{bmatrix} \begin{bmatrix} v_1 & \cdots & v_3 \end{bmatrix} \]
4. Transform fundamental matrix back to original units: if \( T \) and \( T' \) are the normalizing transformations in the two images, than the fundamental matrix in original coordinates is \( T^T F T' \).

Set \( d_{i3} \) to zero and reconstruct \( F \)

Comparison of Estimation Algorithms

<table>
<thead>
<tr>
<th>Method</th>
<th>B-point</th>
<th>Normalized B-point</th>
<th>Nonlinear least squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av. Dist. 1</td>
<td>2.33 pixel</td>
<td>0.52 pixel</td>
<td>0.86 pixel</td>
</tr>
<tr>
<td>Av. Dist. 2</td>
<td>2.15 pixel</td>
<td>0.89 pixel</td>
<td>0.80 pixel</td>
</tr>
</tbody>
</table>

Slide adapted from Svetlana Lazebnik

Slide credit: Svetlana Lazebnik

B. Leibe
3D Reconstruction with Weak Calibration

- Want to estimate world geometry without requiring calibrated cameras.
- Many applications:
  - Archival videos
  - Photos from multiple unrelated users
  - Dynamic camera system
- Main idea:
  - Estimate epipolar geometry from a (redundant) set of point correspondences between two uncalibrated cameras.

Stereo Pipeline with Weak Calibration

- So, where to start with uncalibrated cameras?
  - Need to find fundamental matrix \( F \) and the correspondences (pairs of points \((u', v') \leftrightarrow (u, v))\).
- Procedure
  1. Find interest points in both images
  2. Compute correspondences
  3. Compute epipolar geometry
  4. Refine

Putative Matches based on Correlation Search

- Many wrong matches (10-50%), but enough to compute \( F \)

RANSAC for Robust Estimation of \( F \)

- Select random sample of correspondences
- Compute \( F \) using them
  - This determines epipolar constraint
- Evaluate amount of support - number of inliers within threshold distance of epipolar line
- Choose \( F \) with most support (#inliers)
Putative Matches based on Correlation Search

- Many wrong matches (10-50%), but enough to compute $F$

Pruned Matches

- Correspondences consistent with epipolar geometry

Resulting Epipolar Geometry

Epipolar Transfer

- Assume the epipolar geometry is known
- Given projections of the same point in two images, how can we compute the projection of that point in a third image?

\[
l_{31} = F^T_{13}x_1 \\
l_{32} = F^T_{23}x_2
\]

When does epipolar transfer fail?

Topics of This Lecture

- Revisiting Epipolar Geometry
  - Calibrated case: Essential matrix
  - Uncalibrated case: Fundamental matrix
  - Weak calibration
  - Epipolar Transfer

- Active Stereo
  - Kinect sensor
  - Structured Light sensing
  - Laser scanning
Microsoft Kinect - How Does It Work?

- Built-in IR projector
- IR camera for depth
- Regular camera for color

Recall: Optical Triangulation

- Principle: 3D point given by intersection of two rays.
  - Crucial information: point correspondence
  - Most expensive and error-prone step in the pipeline...

Active Stereo with Structured Light

- Idea: Replace one camera by a projector.
  - Project “structured” light patterns onto the object
  - Simplifies the correspondence problem

What the Kinect Sees...

3D Reconstruction with the Kinect

SIGGRAPH Talks 2011
KinectFusion:
Real-Time Dynamic 3D Surface Reconstruction and Interaction

Shahram Izadi 1, Richard Newcombe 2, David Kim 3,4, Ottmar Hilliges 1, shield Maburyaweni 1,4, Polumart Kahl 1, Janele Botton 1, Steve Hodges 1, Dustin Freeman 5, Andrew Davison 2, Andrew Fitzgibbon 1

1 Microsoft Research Cambridge 2 Imperial College London 3 Newcastle University 4 Lancaster University 5 University of Toronto
Active Stereo with Structured Light

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

- The Kinect uses one such approach (“structured noise”)
  - What other approaches are possible?

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

Laser Scanned Models

The Digital Michelangelo Project, Levoy et al.
Multi-Stripe Triangulation

- To go faster, project multiple stripes
- But which stripe is which?
- Answer #1: assume surface continuity

Laser Scanned Models

- The Digital Michelangelo Project, Levoy et al.

Multi-Stripe Triangulation

- To go faster, project multiple stripes
- But which stripe is which?
- Answer #2: colored stripes (or dots)


Multi-Stripe Triangulation

- To go faster, project multiple stripes
- But which stripe is which?
- Answer #3: time-coded stripes


Time-Coded Light Patterns

- Assign each stripe a unique illumination code over time [Posdamer 82]

Better codes...

- Gray code
  - Neighbors only differ one bit
Phase-Shift Structured Light Scanning

- Faster procedure by projecting continuous patterns
  - Project 3 sinusoid grating patterns shifted by 120° in phase.
  - For each pixel, compute relative phase from 3 intensities.
  - Recover absolute phase by adding a 2nd camera.

A High-Speed 3D Scanner

- Dense 3D reconstructions at 30 fps (on the GPU)

Problems with Dynamically Moving Objects

- Moving objects lead to artifacts
  - Measurements correspond to different 3D points!
- Derived a geometric model for the error
  - Designed a motion compensation method
    ⇒ Result: Cleaned-up geometry + motion estimate!

Effect of Motion Compensation

- Hand Gestures

Application: Online Model Reconstruction

- Poor Man's Scanner
  - The idea
Slightly More Elaborate (But Still Cheap)

Software freely available from Robotics Institute TU Braunschweig
http://www.david-laserscanner.com/

References and Further Reading

- Background information on camera models and calibration algorithms can be found in Chapters 6 and 7 of
  R. Hartley, A. Zisserman
  Multiple View Geometry in Computer Vision
  2nd Ed., Cambridge Univ. Press, 2004

- Also recommended: Chapter 9 of the same book on Epipolar geometry and the Fundamental Matrix and Chapter 11.1-11.6 on automatic computation of $F$. 