**Recap: What Is Stereo Vision?**
- Generic problem formulation: given several images of the same object or scene, compute a representation of its 3D shape.

**Recap: Depth with Stereo - Basic Idea**
- Basic Principle: Triangulation
  - Gives reconstruction as intersection of two rays
  - Requires
    - Camera pose (calibration)
    - Point correspondence

**Recap: Epipolar Geometry**
- Geometry of two views allows us to constrain where the corresponding pixel for some image point in the first view must occur in the second view.
- Epipolar constraint:
  - Correspondence for point \( p \) in \( \Pi \) must lie on the epipolar line \( l' \) in \( \Pi' \) (and vice versa).
  - Reduces correspondence problem to 1D search along conjugate epipolar lines.

**Recap: Stereo Geometry With Calibrated Cameras**
- Camera-centered coordinate systems are related by known rotation \( R \) and translation \( T \):
  \[
  X' = RX + T
  \]
Recap: Essential Matrix

\[ \begin{align*}
X' \cdot (T \cdot RX) &= 0 \\
X' \cdot (T \cdot RX) &= 0
\end{align*} \]

Let \( E = T \cdot R \)

\[ X' \cdot E \cdot X = 0 \]

- This holds for the rays \( p \) and \( p' \) that are parallel to the camera-centered position vectors \( X \) and \( X' \), so we have: \( p' \cdot E \cdot p = 0 \)
- \( E \) is called the essential matrix, which relates corresponding image points [Longuet-Higgins 1981]

Recap: Essential Matrix and Epipolar Lines

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

\[ E \cdot p \]

is the coordinate vector representing the epipolar line for point \( p \)

\[ E' \cdot p \]

is the coordinate vector representing the epipolar line for point \( p' \)

Recap: Stereo Image Rectification

- In practice, it is convenient if image scanlines are the epipolar lines.
- Algorithm
  - Reproject image planes onto a common plane parallel to the line between optical centers
  - Pixel motion is horizontal after this transformation
  - Two homographies (3x3 transforms), one for each input image reprojection

Recap: Dense Correspondence Search

- For each pixel in the first image
  - Find corresponding epipolar line in the right image
  - Examine all pixels on the epipolar line and pick the best match (e.g. SSD, correlation)
  - Triangulate the matches to get depth information
- This is easiest when epipolar lines are scanlines
  - Rectify images first

Topics of This Lecture

- Camera Calibration
  - Camera parameters
  - Calibration procedure
- Revisiting Epipolar Geometry
  - Triangulation
  - Calibrated case: Essential matrix
  - Uncalibrated case: Fundamental matrix
  - Weak calibration
  - Epipolar Transfer
- Improving Stereo Reconstruction
  - Dynamic Programming
  - MRFs
- Active Stereo
Recall: Pinhole Camera Model

\[
X, Y, Z \mapsto (fX/Z, fY/Z)
\]

\[
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix} \mapsto \begin{bmatrix}
fX \\
fY \\
1
\end{bmatrix}
\]

\[
x = PX
\]

Camera Coordinate System

- **Principal axis**: line from the camera center perpendicular to the image plane
- **Normalized (camera) coordinate system**: camera center is at the origin and the principal axis is the z-axis
- **Principal point** \((p)\): point where principal axis intersects the image plane (origin of normalized coordinate system)

Principal Point Offset

- **Camera coordinate system**: origin at the principal point
- **Image coordinate system**: origin is in the corner

\[
\begin{bmatrix}
fX \\
fY \\
Z
\end{bmatrix} = \begin{bmatrix} f & p_1 & 0 \\ f & p_1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}
\]

\[
P = \text{diag}(f, f, 1) \cdot [1, 0]
\]
In general, the camera coordinate frame will be related to the world coordinate frame by a rotation and a translation.

Camera Parameters: Degrees of Freedom

- Intrinsic parameters
  - Principal point coordinates
  - Focal length
  - Pixel magnification factors
  - Skew (non-rectangular pixels)
  - Radial distortion

- Extrinsic parameters
  - Rotation $R$
  - Translation $t$

How many degrees of freedom does $P$ have?

Summary: Camera Parameters

- Intrinsic parameters
  - Principal point coordinates
  - Focal length
  - Pixel magnification factors
  - Skew (non-rectangular pixels)
  - Radial distortion

- Extrinsic parameters
  - Rotation $R$
  - Translation $t$

Camera Parameters: Degrees of Freedom

- Intrinsic parameters
  - Principal point coordinates
  - Focal length
  - Pixel magnification factors
  - Skew (non-rectangular pixels)
  - Radial distortion

- Extrinsic parameters
  - Rotation $R$
  - Translation $t$

Camera projection matrix $P = K[R | t]$.
Calibrating a Camera

- 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 \( P = P_{\text{int}} P_{\text{ext}} \)

Camera Calibration

- Given \( n \) points with known 3D coordinates \( X_i \) and known image projections \( x_i \), estimate the camera parameters.

Camera Calibration: Obtaining the Points

- For best results, it is important that the calibration points are measured with subpixel accuracy.
- How this can be done depends on the exact pattern.
- Algorithm for checkerboard pattern
  1. Perform Canny edge detection.
  2. Fit straight lines to detected linked edges.
  3. Intersect lines to obtain corners.
     - If sufficient care is taken, the points can then be obtained with localization accuracy \( 1/10 \) pixel.
- Rule of thumb
  - Number of constraints should exceed number of unknowns by a factor of five.
  - For 11 parameters of \( P \), at least 28 points should be used.

Camera Calibration: DLT Algorithm

\[
\lambda x_i = P X_i
\]

For best results, it is important that the calibration points are measured with subpixel accuracy. The algorithm for checkerboard pattern involves:
1. Performing Canny edge detection.
2. Fitting straight lines to detected linked edges.
3. Intersecting lines to obtain corners.
- If sufficient care is taken, the points can then be obtained with localization accuracy \( 1/10 \) pixel.
- Rule of thumb
  - Number of constraints should exceed number of unknowns by a factor of five.
  - For 11 parameters of \( P \), at least 28 points should be used.

Camera Calibration: DLT Algorithm

\[
\begin{bmatrix}
0^T & X_i^T & -y_i X_i^T \\
X_i^T & 0^T & -x_i X_i^T \\
\vdots & \vdots & \vdots \\
0^T & X_n^T & -y_n X_n^T \\
X_n^T & 0^T & -x_n X_n^T
\end{bmatrix}
\begin{bmatrix}
P_1 \\
P_2 \\
P_3 \\
P_4 \\
P_5
\end{bmatrix} = 0
\quad \text{Ap} = 0
\]

- \( P \) has 11 degrees of freedom (12 parameters, but scale is arbitrary).
- One 2D/3D correspondence gives us two linearly independent equations.
- Homogeneous least squares (similar to homography est.)
- 5 \( \frac{1}{2} \) correspondences needed for a minimal solution.
Camera Calibration

- Once we've recovered the numerical form of the camera matrix, we still have to figure out the intrinsic and extrinsic parameters.
- This is a matrix decomposition problem, not an estimation problem (see F&P sec. 3.2, 3.3).

Camera Calibration: Some Practical Tips

- For numerical reasons, it is important to carry out some data normalization.
  - Translate the image points $x_i$ to the (image) origin and scale them such that their RMS distance to the origin is $\sqrt{2}$.
  - Translate the 3D points $X_i$ to the (world) origin and scale them such that their RMS distance to the origin is $\sqrt{2}$.
- (This is valid compact point distributions on calibration objects).
- The DLT algorithm presented here is easy to implement, but there are some more accurate algorithms available (see H&Z sec. 7.2).
- For practical applications, it is also often needed to correct for radial distortion. Algorithms for this can be found in H&Z sec. 7.4, or F&P sec. 3.3.

Topics of This Lecture

- Camera Calibration
  - Camera parameters
  - Calibration procedure
- Revisiting Epipolar Geometry
  - Triangulation
  - Calibrated case: Essential matrix
  - Uncalibrated case: Fundamental matrix
  - Weak calibration
  - Epipolar Transfer
- Improving Stereo Reconstruction
  - Dynamic Programming
  - ADD's
- Active Stereo

Two-View Geometry

- Scene geometry (structure): Given corresponding points in two or more images, where is the pre-image of these points in 3D?
- Correspondence (stereo matching): Given a point in just one image, how does it constrain the position of the corresponding point $x'$ in another image?
- Camera geometry (motion): Given a set of corresponding points in two images, what are the cameras for the two views?

Revisiting Triangulation

- Given projections of a 3D point in two or more images (with known camera matrices), find the coordinates of the point $X$.

Revisiting Triangulation

- We want to intersect the two visual rays corresponding to $x_1$ and $x_2$, but because of noise and numerical errors, they will never meet exactly.
Triangulation: Geometric Approach

- Find shortest segment connecting the two viewing rays and let \( X \) be the midpoint of that segment.

\[
\begin{align*}
O_1 & \quad x_1 \quad \text{a}_1x_2 \quad O_2 \\
X & \quad x' \quad \text{a}_x
\end{align*}
\]

Triangulation: Linear Algebraic Approach

\[
\begin{align*}
\lambda_1 x_1 &= P_1 X \quad x_1 \times P_1 X = 0 \quad [x_1] P_1 X = 0 \\
\lambda_2 x_2 &= P_2 X \quad x_2 \times P_2 X = 0 \quad [x_2] P_2 X = 0
\end{align*}
\]

Cross product as matrix multiplication:

\[
\begin{bmatrix}
0 & -a_z & a_x \\
-a_z & 0 & -a_y \\
a_x & a_y & 0
\end{bmatrix}
\begin{bmatrix}
b_x \\
b_y \\
b_z
\end{bmatrix} = [\text{a}_x] b
\]

Two independent equations each in terms of three unknown entries of \( X \)

- This approach is often preferable to the geometric approach, since it nicely generalizes to multiple cameras.

Triangulation: Nonlinear Approach

- Find \( X \) that minimizes

\[
d^2 (x_1, P_1 X) + d^2 (x_2, P_2 X)
\]

- This approach is the most accurate, but unlike the other two methods, it doesn’t have a closed-form solution.

- Iterative algorithm

  - Initialize with linear estimate.
  - Optimize with Gauss-Newton or Levenberg-Marquardt (see F&P sec. 3.1.2 or H&Z Appendix 6).

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

Camera matrix: $[1|0]

$X = (u, v, w, 1)^T$

$X = (u, v, w)^T$

The vectors $x$, $t$, and $R x'$ are coplanar.

Epipolar Geometry: Uncalibrated Case

- The calibration matrices $K$ and $K'$ of the two cameras are unknown.
- We can write the epipolar constraint in terms of unknown normalized coordinates:
  $\tilde{x}^T E \tilde{x}' = 0$
  $x = K \tilde{x}, \quad x' = K' \tilde{x}'$

Fundamental Matrix

(Faugeras and Luong, 1992)
The Eight-Point Algorithm

\[ x = (u, v, 1)^T, \quad x' = (u', v', 1)^T \]

\[
\begin{bmatrix}
F_{11} & F_{12} & F_{13} \\
F_{21} & F_{22} & F_{23} \\
F_{31} & F_{32} & F_{33}
\end{bmatrix}
\begin{bmatrix}
u \\
v' \\
1
\end{bmatrix} = 0
\]

Minimize:

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

under the constraint

\[
|F| = 1
\]

Problem with the Eight-Point Algorithm

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

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 (for example, take SVD of \( F \) and throw out the smallest singular value).
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 \( T^T F T' \).

(Hartley, 1995)

Comparison of Estimation Algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Avg. Dist. 1</th>
<th>Avg. Dist. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-point</td>
<td>2.33 pixels</td>
<td>3.19 pixels</td>
</tr>
<tr>
<td>Normalized 8-point</td>
<td>0.92 pixel</td>
<td>0.89 pixel</td>
</tr>
<tr>
<td>Nonlinear least squares</td>
<td>0.85 pixel</td>
<td>0.85 pixel</td>
</tr>
</tbody>
</table>
Weak Calibration

- Want to estimate world geometry without requiring calibrated cameras
  - 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?

$$l_{12} = F^{21}x_3$$

Topics of This Lecture

- Camera Calibration
  - Camera parameters
  - Calibration procedure
- Revisiting Epipolar Geometry
  - Trigmentation
  - Calibrated case: Essential matrix
  - Uncalibrated case: Fundamental matrix
  - Weak calibration
  - Epipolar Transfer
- Improving Stereo Reconstruction
  - Dynamic Programming
  - MRFs
- Active Stereo
Recap: Dense Correspondence Search
- For each pixel in the first image:
  - Find corresponding epipolar line in the right image
  - Examine all pixels on the epipolar line and pick the best match (e.g., SSD, correlation)
  - Triangulate the matches to get depth information
- This is easiest when epipolar lines are scanlines ⇒ Rectify images first

Recap: Window Search
- Results on Tsukuba test scene
- How can we enforce additional constraints to improve the reconstruction?

Recap: Non-Local Constraints
- Uniqueness
  - For any point in one image, there should be at most one matching point in the other image
- Ordering
  - Corresponding points should be in the same order in both views

Scanline Stereo
- Try to coherently match pixels on the entire scanline
- Different scanlines are still optimized independently

“Shortest Path” for Scanline Stereo
- Can be implemented with dynamic programming

Coherent Stereo on 2D Grid
- Scanline stereo generates streaking artifacts
Dynamic Programming

- Can we apply this trick in 2D as well?

- No: \( d_{x,y} \) and \( d_{x-1,y} \) may depend on different values of \( d_{x-1,y-1} \)

⇒ Can’t use dynamic programming to find spatially coherent disparities/correspondences on a 2D grid!

⇒ But... we’ve already seen methods for performing inference on 2D Markov Random Fields.

Stereo Matching as Energy Minimization

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

\[
E_{\text{data}} = \sum \left( W_i(i) - W_j(i + D(i)) \right)^2, \quad E_{\text{smooth}} = \sum \rho(D(i) - D(j))
\]

- Energy functions of this form can be minimized using Graph Cuts.


Graph Cuts Results


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

The Role of the Baseline

- Small baseline: large depth error
- Large baseline: difficult search problem

Problem for Wide Baselines: Foreshortening

- Matching with fixed-size windows will fail!
- Possible solution: adaptively vary window size
- Another solution: model-based stereo
Model Based Stereo


Applications: The Campanile Movie

Video from SIGGRAPH'97 Animation Theatre
http://www.debevec.org/Campanile/#movie

Topics of This Lecture

- Camera Calibration
  - Camera parameters
  - Calibration procedure
- Revisiting Epipolar Geometry
  - Triangulation
  - Calibrated case: Essential matrix
  - Uncalibrated case: Fundamental matrix
  - Weak calibration
  - Epipolar Transfer
- Improving Stereo Reconstruction
  - Dynamic Programming
  - MRFs
- Active Stereo

Active Stereo with Structured Light

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

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/
Multi-Stripe Triangulation

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

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

Time-Coded Light Patterns

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

Better codes...

• Gray code
  Neighbors only differ one bit

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.