Course Outline

- Image Processing Basics
- Segmentation & Grouping
- Object Recognition
- Local Features & Matching
- Object Categorization
- 3D Reconstruction
  - Epipolar Geometry and Stereo Basics
  - Camera calibration & Uncalibrated Reconstruction
  - Multi-view Stereo
- Motion and Tracking

Recap: Part-Based Models for Categorization

- Fischler & Elschlager 1973
- Model has two components
  - parts (2D image fragments)
  - structure (configuration of parts)
- Looked at 3 classes of models

Recap: Bag-of-Words Approach

- Compute the word activation histogram for each image.
- Let each such BoW histogram be a feature vector.
- Use images from each class to train a classifier (e.g., an SVM).

Recap: Advantage of BoW Histograms

- Bag of words representations make it possible to describe the unordered point set with a single vector (of fixed dimension across image examples).
- Provides easy way to use distribution of feature types with various learning algorithms requiring vector input.
Recap: Constellation Model

\[ P(\text{Image} \mid \text{object}) = P(\text{appearance}, \text{shape} \mid \text{object}) \]
\[ = \sum P(\text{appearance} \mid h, \text{object}) P(\text{shape} \mid h, \text{object}) p(h \mid \text{object}) \]

- assignment of features to parts

Recap: Implicit Shape Model - Representation

- Learn appearance codebook
- Clustering appearance codebook
- Learn spatial distributions
- Match codebook to training images
- Record matching positions on object

Recap: Implicit Shape Model - Recognition

Interest Points

Matched Codebook Entries

Probabilistic Voting

"Generalized Hough Transform with backprojection"

3D Voting Space (continuous)

Backprojected Hypotheses

Backprojection of Maxima

Improving Detections Using Geometry...

Geometric vision

- Goal: Recovery of 3D structure
  - What cues in the image allow us to do this?
Visual Cues

- Shading

Merle Norman Cosmetics, Los Angeles

Visual Cues

- Shading
- Texture

The Visual Cliff, by William Vandiver, 1960

Visual Cues

- Shading
- Texture
- Focus

From The Art of Photography, Canon

Visual Cues

- Shading
- Texture
- Focus
- Perspective

Visual Cues

- Shading
- Texture
- Focus
- Perspective
- Motion

Our Goal: Recovery of 3D Structure

- We will focus on perspective and motion
- We need multi-view geometry because recovery of structure from one image is inherently ambiguous

Our Goal: Recovery of 3D Structure

- We will focus on perspective and motion
- We need multi-view geometry because recovery of structure from one image is inherently ambiguous

Our Goal: Recovery of 3D Structure

- We will focus on perspective and motion
- We need multi-view geometry because recovery of structure from one image is inherently ambiguous

Our Goal: Recovery of 3D Structure

- We will focus on perspective and motion
- We need multi-view geometry because recovery of structure from one image is inherently ambiguous
To Illustrate This Point...

- Structure and depth are inherently ambiguous from single views.

Stereo Vision

- Stereograms: Invented by Sir Charles Wheatstone, 1838

What Is Stereo Vision?

- Generic problem formulation: given several images of the same object or scene, compute a representation of its 3D shape

What Is Stereo Vision?

- Narrower formulation: given a calibrated binocular stereo pair, fuse it to produce a depth image.
  - Humans can do it
What Is Stereo Vision?

- Narrower formulation: given a calibrated binocular stereo pair, fuse it to produce a depth image.
  - Humans can do it

Autostereograms: www.magiceye.com

Historic Origin: Random Dot Stereograms

- Julesz 1960: Do we identify local brightness patterns before fusion (monocular process) or after (binocular)?
- To test: pair of synthetic images obtained by randomly spraying black dots on white objects

Random Dot Stereograms

- When viewed monocularly, they appear random; when viewed stereoscopically, see 3d structure.

Application of Stereo: Robotic Exploration

- Nomad robot searches for meteorites in Antarctica
- Real-time stereo on Mars
Topics of This Lecture

- Geometric vision
  - Visual cues
  - Stereo vision
- Epipolar geometry
  - Depth with stereo
  - Geometry for a simple stereo system
  - Case example with parallel optical axes
  - General case with calibrated cameras
- Correspondence & 3D Reconstruction
  - Correspondence search
  - Additional correspondence constraints
  - Possible sources of error
  - Applications

Depth with Stereo: Basic Idea

- Basic Principle: Triangulation
  - Gives reconstruction as intersection of two rays
  - Requires
    - Camera pose (calibration)
    - Point correspondence

Camera Calibration

- Extrinsic params: rotation matrix and translation vector
- Intrinsic params: focal length, pixel sizes (mm), image center point, radial distortion parameters

We’ll assume for now that these parameters are given and fixed.

Geometry for a Simple Stereo System

- First, assuming parallel optical axes, known camera parameters (i.e., calibrated cameras):

  Similar triangles ($p_l$, $O_l$, $O_r$) and ($O_l$, $P_l$, $O_r$):

  $T + x_r - x_l = T$ 
  
  $Z - f$ 
  
  $Z = \frac{T}{x_r - x_l}$ 

  disparity
**Depth From Disparity**

\[
(x', y') = (x + D(x, y), y)
\]

**General Case With Calibrated Cameras**

- The two cameras need not have parallel optical axes.

**Stereo Correspondence Constraints**

- Given \( p \) in the left image, where can the corresponding point \( p' \) in the right image be?

**Stereo Correspondence Constraints**

- 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: Why is this useful?
  - Reduces correspondence problem to 1D search along conjugate epipolar lines.
Epipolar Geometry

- Baseline: line joining the camera centers
- Epipole: point of intersection of baseline with the image plane
- Epipolar plane: plane containing baseline and world point
- Epipolar line: intersection of epipolar plane with the image plane

- All epipolar lines intersect at the epipole.
- An epipolar plane intersects the left and right image planes in epipolar lines.

Epipolar Constraint

- Potential matches for \( p \) have to lie on the corresponding epipolar line \( l' \).
- Potential matches for \( p' \) have to lie on the corresponding epipolar line \( l \).

Example

As position of 3D point varies, epipolar lines “rotate” about the baseline.

Example: Converging Cameras

Example: Motion Parallel With Image Plane
Example: Forward Motion

- Epipole has same coordinates in both images.
- Points move along lines radiating from e: “Focus of expansion”

For a given stereo rig, how do we express the epipolar constraints algebraically?

If the rig is calibrated, we know:
- How to rotate and translate camera reference frame 1 to get to camera reference frame 2.
  - Rotation: 3 x 3 matrix; translation: 3 vector.

Rotation Matrix

Express 3d rotation as series of rotations around coordinate axes by angles $\alpha$, $\beta$, $\gamma$

Overall rotation is product of these elementary rotations:

$$ R = R_z R_y R_x $$

Camera-centered coordinate systems are related by known rotation $R$ and translation $T$:

$$ X' = RX + T $$
**Cross Product**

\[ \mathbf{a} \times \mathbf{b} = \mathbf{c} \]

- Vector cross product takes two vectors and returns a third vector that's perpendicular to both inputs.
- So here, \( \mathbf{c} \) is perpendicular to both \( \mathbf{a} \) and \( \mathbf{b} \), which means the dot product = 0.

**From Geometry to Algebra**

\[
\begin{align*}
X' &= RX + T \\
X' \cdot (T \times X') &= 0 \\
T \times X &= T \times RX + T \times T \\
\text{Normal to the plane} &= T \times RX
\end{align*}
\]

**Essential Matrix**

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

Let \( \mathbf{E} = T \cdot R \)

\[ \mathbf{X'} \mathbf{E} \mathbf{X} = 0 \]

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

**Essential Matrix and Epipolar Lines**

\( \mathbf{p}' \mathbf{E} \mathbf{p} = 0 \)

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

\[ \mathbf{E}^T \mathbf{p} \] is the coordinate vector representing the epipolar line for point \( \mathbf{p} \)

\[ \mathbf{E}^T \mathbf{p}' \] is the coordinate vector representing the epipolar line for point \( \mathbf{p}' \)
**Essential Matrix: Properties**

- Relates image of corresponding points in both cameras, given rotation and translation.
- Assuming intrinsic parameters are known

\[ E = T \cdot R \]

**Essential Matrix Example: Parallel Cameras**

\[ R = I \]
\[ T = [-d, 0, 0]^T \]
\[ E = [T]_R = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & d \\ 0 & d & 0 \end{bmatrix} \]

\[ p^{T_1}E p = 0 \]

For the parallel cameras, image of any point must lie on same horizontal line in each image plane.

**More General Case**

Image \( I(x, y) \)

Disparity map \( D(x, y) \)

Image \( I'(x', y') \)

\( (x', y') = (x + D(x, y), y) \)

What about when cameras’ optical axes are not parallel?

**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

**Stereo Image Rectification: Example**

Source: Alyosha Efros
Topics of This Lecture

- Geometric vision
  - Visual cues
  - Stereo vision
- Epipolar geometry
  - Depth with stereo
  - Geometry for a simple stereo system
  - Case example with parallel optical axes
  - General case with calibrated cameras
- Stereopsis & 3D Reconstruction
  - Correspondence search
  - Additional correspondence constraints
  - Possible sources of error
  - Applications

Stereo Reconstruction

- Main Steps
  - Calibrate cameras
  - Rectify images
  - Compute disparity
  - Estimate depth

Correspondence Problem

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

Additional Correspondence Constraints

- Similarity
- Uniqueness
- Ordering
- Disparity gradient

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
Example: Window Search

- Data from University of Tsukuba

Effect of Window Size

Want window large enough to have sufficient intensity variation, yet small enough to contain only pixels with about the same disparity.

Dense vs. Sparse

- Sparse
  - Efficiency
  - Can have more reliable feature matches, less sensitive to illumination than raw pixels
  - But...
    - Have to know enough to pick good features
    - Sparse information
- Dense
  - Simple process
  - More depth estimates, can be useful for surface reconstruction
  - But...
    - Breaks down in textureless regions anyway
    - Raw pixel distances can be brittle
    - Not good with very different viewpoints

Sparse Correspondence Search

- Restrict search to sparse set of detected features
- Rather than pixel values (or lists of pixel values) use feature descriptor and an associated feature distance
- Still narrow search further by epipolar geometry

What would make good features?

Difficulties in Similarity Constraint

- Untextured surfaces
- Occlusions
**Additional Correspondence Constraints**

- Similarity
- Uniqueness
- Ordering
- Disparity gradient

- Epipolar lines constrain the search to a line, and these appearance and ordering constraints further reduce the possible matches.

**Uniqueness**

- For opaque objects, up to one match in right image for every point in left image

**Ordering**

- Points on *same surface* (opaque object) will be in same order in both views

**Disparity Gradient**

- Assume piecewise continuous surface, so want disparity estimates to be locally smooth

**Possible Sources of Error?**

- Low-contrast / textureless image regions
- Occlusions
- Camera calibration errors
- Violations of brightness constancy (e.g., specular reflections)
- Large motions
Edges in disparity in conjunction with image edges enhances contours found.

Application: View Interpolation

Right Image

Left Image

Disparity
Application: Free-Viewpoint Video

http://www.liberovision.com

Summary: Stereo Reconstruction

- Main Steps
  - Calibrate cameras
  - Rectify images
  - Compute disparity
  - Estimate depth

- So far, we have only considered calibrated cameras...

- Next lecture
  - Uncalibrated cameras
  - Camera parameters
  - Revisiting epipolar geometry
  - Robust fitting

References and Further Reading

- Background information on epipolar geometry and stereopsis can be found in Chapters 10.1-10.2 and 11.1-11.3 of
  D. Forsyth, J. Ponce,
  Computer Vision - A Modern Approach.
  Prentice Hall, 2003

- More detailed information (if you really want to implement 3D reconstruction algorithms) can be found in Chapters 9 and 10 of
  R. Hartley, A. Zisserman
  Multiple View Geometry in Computer Vision
  2nd Ed., Cambridge Univ. Press, 2004