Computer Vision - Lecture 15

Epipolar Geometry & Stereo Basics

08.01.2009

Bastian Leibe
RWTH Aachen
http://www.umic.rwth-aachen.de/multimedia
leibe@umic.rwth-aachen.de
Announcements

- Seminar (combined with Graphics group)
  - “Current topics in Computer Graphics, Geometry Processing, and Computer Vision”

http://www.graphics.rwth-aachen.de/index.php?id=204
Announcements

• Seminar (combined with Graphics group)
  - “Current topics in Computer Graphics, Geometry Processing, and Computer Vision”

• Exams
  - Who is in one of the Master Programs and wants to take the exam in the semester break?
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
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
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

Slide credit: Steve Seitz
Visual Cues

- Shading
- Texture
- Focus
- Perspective
Visual Cues

- Shading
- Texture
- Focus
- Perspective
- Motion

Figures from L. Zhang

Slide credit: Steve Seitz, Kristen Grauman

http://www.brainconnection.com/teasers/?main=illusion/motion-shape
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

http://www.well.com/~jimg/stereo/stereo_list.html

Slide credit: Kristen Grauman
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?

- 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

Image 1  Image 2

Dense depth map

Slide credit: Svetlana Lazebnik, Steve Seitz
What Is Stereo Vision?

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

Stereograms: Invented by Sir Charles Wheatstone, 1838

Slide credit: Svetlana Lazebnik, Steve Seitz
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

Slide credit: Svetlana Lazebnik, Steve Seitz
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

Slide credit: Svetlana Lazebnik, Steve Seitz
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

Image Source: Forsyth & Ponce

Slide credit: Kristen Grauman
Random Dot Stereograms

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

Image Source: Forsyth & Ponce

Slide credit: Kristen Grauman
Random Dot Stereograms

Figure 5.3.8 A random dot stereogram. These two images are derived from a single array of randomly placed squares by laterally displacing a region of them as described in the text. When they are viewed with crossed disparity (by crossing the eyes) so that the right eye’s view of the left image is combined with the left eye’s view of the right image, a square will be perceived to float above the page. (See pages 210–211 for instructions on fusing stereograms.)
Application of Stereo: Robotic Exploration

**Nomad robot** searches for meteorites in Antarctica

Real-time stereo on Mars

Slide credit: Steve Seitz
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
Depth with Stereo: Basic Idea

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

Slide credit: Steve Seitz
Camera Calibration

Extrinsic parameters:
Camera frame ↔ Reference frame

Intrinsic parameters:
Image coordinates relative to camera ↔ Pixel coordinates

- **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):
Geometry for a Simple Stereo System

- Assume parallel optical axes, known camera parameters (i.e., calibrated cameras). We can triangulate via:

Similar triangles $(p_l, P, p_r)$ and $(O_l, P, O_r)$:

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

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

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.

Slide credit: Kristen Grauman, Steve Seitz
Stereo Correspondence Constraints

- Given $p$ in the left image, where can the corresponding point $p'$ in the right image be?
Stereo Correspondence Constraints

Given $p$ in the left image, where can the corresponding point $p'$ in the right image be?
Stereo Correspondence Constraints

\[ \text{Diagram showing stereo correspondence constraints.} \]

Slide credit: Kristen Grauman
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

- Epipolar Plane
- Epipoles
- Baseline
- Epipolar Lines

Slide adapted from Marc Pollefeys
Epipolar Geometry: Terms

- **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 \( l \).

http://www.ai.sri.com/~luong/research/Meta3DViewer/EpipolarGeo.html

Slide credit: Marc Pollefeys
Example
Example: Converging Cameras

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

Figure from Hartley & Zisserman
Example: Motion Parallel With Image Plane

Slide credit: Kristen Grauman

Figure from Hartley & Zisserman
Example: Forward Motion

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

Slide credit: Kristen Grauman

B. Leibe

Figure from Hartley & Zisserman
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:

$$R = R_x R_y R_z$$
3D Rigid Transformation

\[
\begin{bmatrix}
X' \\
Y' \\
Z'
\end{bmatrix} = \begin{bmatrix}
r_{11} & r_{12} & r_{13} \\
r_{21} & r_{22} & r_{23} \\
r_{31} & r_{32} & r_{33}
\end{bmatrix} \begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix} + \begin{bmatrix}
T_x \\
T_y \\
T_z
\end{bmatrix}
\]

\[X' = RX + T\]
Stereo Geometry With Calibrated Cameras

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

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

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

\[ \overrightarrow{a} \cdot \overrightarrow{c} = 0 \]

\[ \overrightarrow{b} \cdot \overrightarrow{c} = 0 \]

- Vector cross product takes two vectors and returns a third vector that’s perpendicular to both inputs.

- So here, \( \overrightarrow{c} \) is perpendicular to both \( \overrightarrow{a} \) and \( \overrightarrow{b} \), which means the dot product = 0.
From Geometry to Algebra

\[ \text{Normal to the plane} \]

\[ \mathbf{X}' = \mathbf{R} \mathbf{X} + \mathbf{T} \]

\[ \mathbf{T} \times \mathbf{X}' = \mathbf{T} \times \mathbf{R} \mathbf{X} + \mathbf{T} \times \mathbf{T} \]

\[ = \mathbf{T} \times \mathbf{R} \mathbf{X} \]

\[ \mathbf{X}' \cdot (\mathbf{T} \times \mathbf{X}') = \mathbf{X}' \cdot (\mathbf{T} \times \mathbf{R} \mathbf{X}) = 0 \]

Slide credit: Kristen Grauman
Matrix Form of Cross Product

\[
\vec{a} \times \vec{b} = \vec{c} \quad \vec{a} \cdot \vec{c} = 0 \\
\vec{b} \cdot \vec{c} = 0
\]

\[
\begin{bmatrix}
0 & -a_z & a_y \\
a_z & 0 & -a_x \\
-a_y & a_x & 0
\end{bmatrix}
\]

\[
\vec{a} \times \vec{b} = [a_x] \vec{b}
\]
From Geometry to Algebra

\[ X' = RX + T \]
\[ T \times X' = T \times RX + T \times T \]
\[ \text{Normal to the plane} \]
\[ = T \times RX \]

\[ X' \cdot (T \times X') = X' \cdot (T \times RX) = 0 \]

Slide credit: Kristen Grauman
**Essential Matrix**

\[
X' \cdot (T \times RX) = 0
\]

\[
X' \cdot (T_x \ RX) = 0
\]

Let \( E = T_x R \)

\[
X'^T E X = 0
\]

- This holds for the rays \( p \) and \( p' \) that are parallel to the camera-centered position vectors \( X \) and \( X' \), so we have:

\[
p'^T E p = 0
\]

- \( E \) is called the essential matrix, which relates corresponding image points [Longuet-Higgins 1981]

Slide credit: Kristen Grauman
Essential Matrix and Epipolar Lines

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

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

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

\( \mathbf{E} p' \) is the coordinate vector representing the epipolar line for point \( p' \)
Essential Matrix: Properties

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

\[ E = TxR \]
Essential Matrix Example: Parallel Cameras

\[
R = \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
1 & 0 & 0
\end{bmatrix}
\]

\[
T = \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix}
\]

\[
E = [T_x]R = \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{bmatrix}
\]

\[
p^T E p = 0
\]

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

Slide credit: Kristen Grauman
Essential Matrix Example: Parallel Cameras

\[ \mathbf{R} = \mathbf{I} \]
\[ \mathbf{T} = [-d, 0, 0]^T \]
\[ \mathbf{E} = [\mathbf{T}_x] \mathbf{R} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & d \\ 0 & -d & 0 \end{pmatrix} \]

\[ \mathbf{p}'^T \mathbf{E} \mathbf{p} = 0 \]

\[ \begin{bmatrix} x' & y' & f \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & d \\ 0 & -d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ f \end{bmatrix} = 0 \]

\[ \Leftrightarrow \begin{bmatrix} x' & y' & f \end{bmatrix} \begin{bmatrix} 0 \\ df \\ -dy \end{bmatrix} = 0 \]

\[ \Leftrightarrow y = y' \]

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

Slide credit: Kristen Grauman
More General Case

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

*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
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

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

Slide credit: Kristen Grauman

Figure from Gee & Cipolla 1999
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

adapted from Svetlana Lazebnik, Li Zhang
Example: Window Search

- Data from University of Tsukuba

Scene

Ground truth

Slide credit: Kristen Grauman

B. Leibe
Example: Window Search

- Data from University of Tsukuba

Window-based matching (best window size)  Ground truth

Slide credit: Kristen Grauman
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.

Slide credit: Kristen Grauman

B. Leibe

Figures from Li Zhang
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?

Slide credit: Kristen Grauman
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

Slide credit: Kristen Grauman
Difficulties in Similarity Constraint

Untextured surfaces

Occlusions

Slide credit: Kristen Grauman
Additional Correspondence Constraints

- Similarity
- Uniqueness
- Ordering
- Disparity gradient
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

Slide credit: Kristen Grauman
Disparity Gradient

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

Given matches ● and ○, point ○ in the left image must match point 1 in the right image. Point 2 would exceed the disparity gradient limit.
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.
Possible Sources of Error?

- Low-contrast / textureless image regions
- Occlusions
- Camera calibration errors
- Violations of brightness constancy (e.g., specular reflections)
- Large motions
Application: Depth for Segmentation

(a) Left camera image.  (b) Right camera image.
(c) Depth image.  (d) Edge combination image.

Edges in disparity in conjunction with image edges enhances contours found

Figure 3 Stereo video frames with computed depth map and edge combination result.
Depth for Segmentation

Original image with snake initialization.

(b) Final snake on original image.

Final snake on edge combination image.

(f) Original image with snake from (e) overlaid.

(c) Final snake on depth image.

(d) Original image with snake from (c) overlaid.

Image Source: Danijela Markovic and Margrit Gelautz

Slide credit: Kristen Grauman
Application: View Interpolation

Right Image

Slide credit: Svetlana Lazebnik
Application: View Interpolation

Left Image

Slide credit: Svetlana Lazebnik
Application: View Interpolation

Disparity

Slide credit: Svetlana Lazebnik
Application: View Interpolation
Applications: Virtual Viewpoint Video


Slide credit: Svetlana Lazebnik

Figure 6: Sample results from stereo reconstruction stage: (a) input color image; (b) color-based segmentation; (c) initial disparity estimates \( d_{ij} \); (d) refined disparity estimates; (e) smoothed disparity estimates \( d_i(x) \).
Virtual Viewpoint Video

Massive Arabesque

http://research.microsoft.com/IVM/VVV/

Slide credit: Svetlana Lazebnik  
B. Leibe
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