Computer Vision - Lecture 15
Camera Calibration & 3D Reconstruction
14.01.2014

Bastian Leibe
RWTH Aachen
http://www.vision.rwth-aachen.de
leibe@vision.rwth-aachen.de

Announcements

- Seminar in the summer semester
  - “Current Topics in Computer Vision and Machine Learning”
  - Block seminar, presentations in 1st week of semester break
  - You can sign up for the seminar here: https://www.graphics.rwth-aachen.de/apse
  - Quick poll: Who would be interested in that?

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
  - Structure-from-Motion
- Motion and Tracking

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

- \( X' = (T \times RX) = 0 \)
- \( X' = (T \cdot RX) = 0 \)
- Let \( E = TR \)
  \[
  X' \cdot EX = 0
  \]
- This holds for the rays \( p \) and \( p' \) that are parallel to the camera-centered position vectors \( X \) and \( X' \), so we have:
  \[
  p' \cdot Ep = 0
  \]

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.
  \[
  l' = Ep
  \]
- \( l' = Ep \) is the coordinate vector representing the epipolar line for point \( p \) (i.e., the line is given by: \( l' \cdot x = 0 \))
- \( l = E' p' \) 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 \((3 \times 3 \) 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

Recap: A General Point

- Equations of the form
  \[
  Ax = 0
  \]
- How do we solve them? (always!)
  - Apply SVD
  - Singular values of \( A \) are 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 \( A \).
### Summary: Stereo Reconstruction

- **Main Steps**
  - Calibrate cameras
  - Rectify images
  - Compute disparity
  - Estimate depth
- **So far, we have only considered calibrated cameras...**
- **Today**
  - Uncalibrated cameras
  - Camera parameters
  - Revisiting epipolar geometry
  - Robust fitting

### 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
- **Active Stereo**
  - Laser scanning
  - Kinect sensor

### Recall: Pinhole Camera Model

\[
\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \mapsto \begin{pmatrix} fX/Z \\ fY/Z \end{pmatrix}
\]

\[
x = PX
\]

### Pinhole Camera Model

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

\[
x = PX
\]

### 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
**Intrinsic parameters**

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

**Camera Rotation and Translation**

In non-homogeneous coordinates:

\[ \hat{X}_{\text{cam}} = R(\hat{X}-\hat{C}) \]

\[ X_{\text{cam}} = R \hat{X} \]

**Summary:** Camera Parameters

- Intrinsic parameters
- Principal point coordinates
- Focal length
- Pixel magnification factors
- Skew (non-rectangular pixels)
- Radial distortion
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
\[ \mathbf{P} = \mathbf{K} \mathbf{R} \mathbf{t} \]

For best results, it is important that the calibration object is placed correctly. For the best results, it is important that the calibration object is placed correctly. For the best results, it is important that the calibration object is placed correctly. For the best results, it is important that the calibration object is placed correctly.

Radial distortion
Skew (non-rectangular pixels)
Pixel magnification factors

How many degrees of freedom does \( \mathbf{P} \) have?

General pinhole camera:
- Number of constraints should exceed number of unknowns by a factor of five.
- If sufficient care is taken, the points can then be obtained with localization accuracy < 1/10 pixel.

Camera Calibration: DLT Algorithm

\[ \mathbf{A} \mathbf{x}_i = \mathbf{P} \mathbf{x}_i \]

\[ \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \end{bmatrix} \begin{bmatrix} X_{x,i} \\ X_{y,i} \\ 1 \end{bmatrix} = \begin{bmatrix} P'_x \\ P'_y \\ P'_z \end{bmatrix} \]

Only two linearly independent equations

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 \( \mathbf{P} \), at least 28 points should be used.
Camera Calibration: DLT Algorithm

\[ \begin{bmatrix} 0^T & X_i^T & -y_iX_i^T \\ X_i^T & 0^T & -x_iX_i^T \\ \vdots & \vdots & \vdots \\ 0^T & X_n^T & -y_nX_n^T \\ X_n^T & 0^T & -x_nX_n^T \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{bmatrix} = 0 \quad \text{Ap} = 0 \]

Solve using SVD!

Notes
- 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 \( X \) to the (world) origin and scale them such that their RMS distance to the origin is \( \sqrt{3} \).
  - (This is valid for 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
- Active Stereo
  - Laser scanning
  - Kinect sensor

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

\[ X ? \]

Triangulation: 1) Geometric Approach

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

\[ X \]

Triangulation: 2) Linear Algebraic Approach

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

Cross product as matrix multiplication:

\[ \{0, -a_z, a_y\} \times \{b_x, 0, -a_z\} = [a \times b] \]

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

\[ \Rightarrow \text{Stack them and solve using SVD!} \]

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

Triangulation: 3) Nonlinear Approach

- Find \( X \) that minimizes

\[ d^2(x_1, P_1 X) + d^2(x_2, P_2 X) \]
**Triangulation: 3) Nonlinear Approach**

- Find $X$ that minimizes $d^2(x_1, P_1X) + d^2(x_2, P_2X)$
- 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**

- The vectors $x$, $t$, and $Rx’$ are coplanar
- $E x’$ is the epipolar line associated with $x’$ ($l = E x’$)
- $E x$ is the epipolar line associated with $x$ ($l’ = E’x$)
- $E e’ = 0$ and $E^T e = 0$
- $E$ is singular (rank two)
- $E$ has five degrees of freedom (up to scale)

**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:
  $$\hat{x}^T E \hat{x}’ = 0$$
  $$x = K \hat{x}, \quad x’ = K’ \hat{x}’$$
How can we estimate \( F \) from an image pair?

\[
\hat{x}^T E \hat{x} = 0 \quad \Rightarrow \quad \hat{x}^T F \hat{x'} = 0 \quad \text{with} \quad F = K^{-T} E K'^{-T}
\]

\( x = K \hat{x} \)

\( x' = K' \hat{x}' \)

**Fundamental Matrix**

(Faugeras and Luong, 1992)

---

**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...

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

\[
A = \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix}
\]

- Taking 8 correspondences:
- Solve using... SVD!

\[
A \hat{F} = 0
\]

This minimizes:

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

---

**Excursion: Properties of SVD**

- Frobenius norm
  - Generalization of the Euclidean norm to matrices
    \[
    \|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{\sum_{i=1}^{\min(m,n)} \sigma_i^2}
    \]

- Partial reconstruction property of SVD
  - Let \( \sigma_i \) be the singular values of \( A \).
  - Let \( A_{p} = U_p D_p V_p^T \) be the reconstruction of \( A \) when we set \( \sigma_{p+1} \) to zero.
  - Then \( A_{p} = U_p D_p V_p^T \) is the best rank-\( p \) 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.
  - \( \Rightarrow \) There will be no epipoles through which all epipolar lines pass!

- Enforce the rank-2 constraint using SVD

\[
SVDF = UD\Phi^T = U \begin{bmatrix} d_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_{n2} \end{bmatrix} \begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nn} \end{bmatrix}
\]

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

- As we have just seen, this provides the best least-squares approximation to the rank-2 solution.
Problem with the Eight-Point Algorithm

- In practice, this often looks as follows:

\[
\begin{bmatrix}
    x_1' & x_2' & x_3' & x_4' & x_5' & x_6' & x_7' & x_8' \\
    y_1' & y_2' & y_3' & y_4' & y_5' & y_6' & y_7' & y_8'
\end{bmatrix}
\]

- Use the eight epipolar lines to compute \( F \) from the normalized points.

- Enforce the rank-2 constraint using SVD.

\[
F = U \Sigma V^T
\]

- Transform fundamental matrix back to original units: if \( T \) and \( T' \) are the normalizing transformations in the two images, then the fundamental matrix in original coordinates is \( T^T F T' \).

\[
\begin{array}{c|c|c}
\text{Sample Estimation} & \text{Normalized 8-point} & \text{Nonlinear least squares} \\
\hline
\text{Av. Dist. 1} & 2.32 pixels & 0.80 pixel & 0.80 pixel \\
\text{Av. Dist. 2} & 2.18 pixels & 0.80 pixel & 0.90 pixel \\
\end{array}
\]

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.

\[
\sum_{i=1}^{N} (x_i^2 F x'_i)^2
\]

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

The Eight-Point Algorithm

- Meaning of error \( \sum_{i=1}^{N} (x_i^2 F x'_i)^2 \):

- Sum of Euclidean distances between points \( x_i \) and epipolar lines \( F x'_i \), (or points \( x'_i \) and epipolar lines \( F^T x_i \)), multiplied by a scale factor

- Nonlinear approach: minimize

\[
\sum_{i=1}^{N} \left( d^2(x_i, F x'_i) + d^2(x'_i, F^T x_i) \right)
\]

- Similar to nonlinear minimization approach for triangulation.
- Iterative approach (Gauss-Newton, Levenberg-Marquardt, ...)

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.
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)
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?

\[ l_{31} = F_{12} x_1 \]
\[ l_{32} = F_{22} x_2 \]

When does epipolar transfer fail?

Extension: 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?

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
- Active Stereo
  - Laser scanning
  - Kinect sensor

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...

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/
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 $F$. 

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