Single-view geometry

Odilon Redon, Cyclops, 1914
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

Focus

From *The Art of Photography*, Canon

Slide credit: S. Seitz
Visual cues

Perspective
Visual cues

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

![Diagram showing perspective and motion in multi-view geometry](image-url)
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
Recall: Pinhole camera model

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

\[
\begin{pmatrix}
X \\
Y \\
Z \\
1
\end{pmatrix} \mapsto \begin{pmatrix}
fX \\
fY \\
fZ \\
1
\end{pmatrix} = \begin{bmatrix}
f & 0 & 0 \\
f & 0 & 0 \\
1 & 0 & 0
\end{bmatrix} \begin{pmatrix}
X \\
Y \\
Z \\
1
\end{pmatrix} \quad x = PX
\]
Pinhole camera model

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

\[x = PX \quad P = \text{diag}(f, f, 1)[I | 0]\]
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 is at the principal point
- Image coordinate system: origin is in the corner

principal point: \((p_x, p_y)\)
Principal point offset

principal point: \((p_x, p_y)\)

\[(X, Y, Z) \mapsto (f \frac{X}{Z} + p_x, f \frac{Y}{Z} + p_y)\]

\[
\begin{pmatrix}
X \\
Y \\
Z \\
1
\end{pmatrix}
\mapsto
\begin{pmatrix}
fX + Zp_x \\
fY + Zp_y \\
Z
\end{pmatrix}
= \begin{bmatrix}
f & p_x & 0 \\
f & p_y & 0 \\
1 & 0 & 1
\end{bmatrix}
\begin{pmatrix}
X \\
Y \\
Z \\
1
\end{pmatrix}
\]
Principal point offset

\[
\begin{pmatrix}
 fX + Zp_x \\
 fY + Zp_y \\
 Z
\end{pmatrix}
= \begin{bmatrix}
 f & p_x & 1 \\
 f & p_y & 1 \\
 1 & 1 & 0
\end{bmatrix}
\begin{pmatrix}
 X \\
 Y \\
 Z
\end{pmatrix}
\]

\[K = \begin{bmatrix}
 f & p_x \\
 f & p_y \\
 1 & 1
\end{bmatrix}\]

principal point: \((p_x, p_y)\)

calibration matrix

\[P = K[I | 0]\]
Pixel coordinates

Pixel size: \( \frac{1}{m_x} \times \frac{1}{m_y} \)

\( m_x \) pixels per meter in horizontal direction,
\( m_y \) pixels per meter in vertical direction

\[
K = \begin{bmatrix}
  m_x & m_y & 1 \\
  f & p_x & \alpha_x \\
  f & p_y & \beta_x
\end{bmatrix}
= \begin{bmatrix}
  \alpha_y & \beta_y \\
  \alpha_x & \beta_x
\end{bmatrix}
\]

pixels/m × m × pixels
Camera rotation and translation

- In general, the camera coordinate frame will be related to the world coordinate frame by a rotation and a translation.

\[
\tilde{X}_{\text{cam}} = R \left( \tilde{X} - \tilde{C} \right)
\]

- coords. of a point in world frame (nonhomogeneous)
- coords. of camera center in world frame
- coords. of point in camera frame
Camera rotation and translation

In non-homogeneous coordinates:

$$\tilde{X}_{\text{cam}} = R(\tilde{X} - \tilde{C})$$

$$X_{\text{cam}} = \begin{bmatrix} R & -R\tilde{C} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{X} \\ 1 \end{bmatrix} = \begin{bmatrix} R & -R\tilde{C} \\ 0 & 1 \end{bmatrix} X$$

$$x = K \begin{bmatrix} I & 0 \end{bmatrix} X_{\text{cam}} = K \begin{bmatrix} R & -R\tilde{C} \end{bmatrix} X \quad P = K \begin{bmatrix} R & t \end{bmatrix}, \quad t = -R\tilde{C}$$

Note: C is the null space of the camera projection matrix (PC=0)
Camera parameters

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

\[ K = \begin{bmatrix} m_x & f & p_x \\ m_y & f & p_y \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & \beta_x \\ \alpha_y & \beta_y \\ 1 & 1 \end{bmatrix} \]
Camera parameters

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

• Extrinsic parameters
  • Rotation and translation relative to world coordinate system
Camera calibration

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

\[ \lambda x_i = PX_i \]

\[ \lambda \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} P_1^T \\ P_2^T \\ P_3^T \end{bmatrix} X_i \]

\[ x_i \times PX_i = 0 \]

\[
\begin{bmatrix}
0 & -X_i^T & y_iX_i^T \\
X_i^T & 0 & -x_iX_i^T \\
-y_iX_i^T & x_iX_i^T & 0
\end{bmatrix}
\begin{bmatrix}
P_1 \\
P_2 \\
P_3
\end{bmatrix} = 0
\]

Two linearly independent equations
Camera calibration

\[
\begin{bmatrix}
0^T & X_1^T & -y_1X_1^T \\
X_1^T & 0^T & -x_1X_1^T \\
\vdots & \vdots & \vdots \\
0^T & X_n^T & -y_nX_n^T \\
X_n^T & 0^T & -x_nX_n^T \\
\end{bmatrix}
\begin{pmatrix}
P_1 \\
P_2 \\
P_3 \\
\end{pmatrix} = 0 \quad \text{Ap} = 0
\]

- 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
- 6 correspondences needed for a minimal solution
Camera calibration

\[
\begin{bmatrix}
0^T & X_1^T & -y_1X_1^T \\
X_1^T & 0^T & -x_1X_1^T \\
\vdots & \vdots & \vdots \\
0^T & X_n^T & -y_nX_n^T \\
X_n^T & 0^T & -x_nX_n^T \\
\end{bmatrix}
\begin{pmatrix}
P_1 \\
P_2 \\
P_3 \\
\end{pmatrix} = 0 \quad \text{Ap} = 0
\]

- Note: for coplanar points that satisfy \(\Pi^TX=0\), we will get degenerate solutions \((\Pi,0,0)\), \((0,\Pi,0)\), or \((0,0,\Pi)\)
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).
Two-view geometry

- **Scene geometry (structure):** Given projections of the same 3D point in two or more images, how do we compute the 3D coordinates of that point?

- **Correspondence (stereo matching):** Given a point in just one image, how does it constrain the position of the corresponding point in a second image?

- **Camera geometry (motion):** Given a set of corresponding points in two images, what are the cameras for the two views?
Triangulation

• Given projections of a 3D point in two or more images (with known camera matrices), find the coordinates of the point
Triangulation

- We want to intersect the two visual rays corresponding to $x_1$ and $x_2$, but because of noise and numerical errors, they don’t meet exactly.
Triangulation: Geometric approach

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

\[ \lambda_1 x_1 = P_1 X \quad x_1 \times P_1 X = 0 \quad [x_1 \times] P_1 X = 0 \]

\[ \lambda_2 x_2 = P_2 X \quad x_2 \times P_2 X = 0 \quad [x_2 \times] P_2 X = 0 \]

Cross product as matrix multiplication:

\[
\mathbf{a} \times \mathbf{b} = \begin{bmatrix}
0 & -a_z & a_y \\
- a_z & 0 & -a_x \\
-a_y & a_x & 0
\end{bmatrix}
\begin{bmatrix}
b_x \\
b_y \\
b_z
\end{bmatrix} = [\mathbf{a} \times] \mathbf{b}
\]
Triangulation: Linear approach

\[ \lambda_1 x_1 = P_1 X \quad x_1 \times P_1 X = 0 \quad [x_{1x}]P_1 X = 0 \]

\[ \lambda_2 x_2 = P_2 X \quad x_2 \times P_2 X = 0 \quad [x_{2x}]P_2 X = 0 \]

Two independent equations each in terms of three unknown entries of X
Triangulation: Nonlinear approach

Find $X$ that minimizes

$$d^2(x_1, P_1X) + d^2(x_2, P_2X)$$