CSE 152: Computer Vision
Manmohan Chandraker

Lecture 5: 3D Reconstruction
Recap
Homogeneous coordinates

- Converting to homogeneous coordinates

\[(x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}\]  
Homogeneous 2D point

\[(x, y, z) \Rightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}\]  
Homogeneous 3D point

- Converting from homogeneous coordinates

\[\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w)\]  
\[\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow (x/w, y/w, z/w)\]

- \((x, y, w)\) and \((kx, ky, kw)\) are the same point.
Modeling projection

\[ (x, y, z) \rightarrow \left(-d \frac{x}{z}, -d \frac{y}{z}\right) \]

- A matrix multiplication using homogeneous coordinates

\[
\begin{bmatrix}
-d & 0 & 0 & 0 \\
0 & -d & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z \\
1
\end{bmatrix}
= \begin{bmatrix}
-dx \\
-dy \\
-z \\
1
\end{bmatrix}
\rightarrow \left(-d \frac{x}{z}, -d \frac{y}{z}\right)
From image plane to pixel coordinates

\[
\begin{bmatrix}
-\alpha d & 0 & 0 \\
0 & -\alpha d & 0 \\
0 & 0 & 1
\end{bmatrix}
\quad \text{K (intrinsics)}
\quad \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\quad \text{Projection (converts from 3D rays in camera coordinate system to pixel coordinates)}
\]

In general, \( \mathbf{K} = \begin{bmatrix}
-\alpha d & s & c_x \\
0 & -\alpha d & c_y \\
0 & 0 & 1
\end{bmatrix} \) (upper triangular matrix)

\( \alpha \) : aspect ratio (1 unless pixels are not square)

\( S \) : skew (0 unless pixels are shaped like rhombi/parallelograms)

\((c_x, c_y)\) : principal point ((0,0) unless optical axis doesn’t intersect projection plane at origin)
A Tale of Two Coordinate Systems

Two important coordinate systems:

1. *World* coordinate system
2. *Camera* coordinate system
Extrinsics

• How do we get the camera to “canonical form”?  
  – (Center of projection at the origin, x-axis points right, y-axis points up, z-axis points backwards)
Extrinsics

- How do we get the camera to “canonical form”?
  - (Center of projection at the origin, x-axis points right, y-axis points up, z-axis points backwards)

Step 1: Translate by \(-\mathbf{c}\)

\[
T = \begin{bmatrix}
I_{3 \times 3} & -\mathbf{c} \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

How do we represent translation as a matrix multiplication?
Extrinsics

• How do we get the camera to “canonical form”? 
  – (Center of projection at the origin, x-axis points right, y-axis points up, z-axis points backwards)

Step 1: Translate by \(-c\)
Step 2: Rotate by \(R\)

\[
R = \begin{bmatrix} u^T \\ v^T \\ w^T \end{bmatrix}
\]
Projection matrix

$$\Pi = K \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} I_{3 \times 3} & -c \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Denote this by $t$. 
Structure from Motion (SFM)
Visual SLAM
Feature detection

Several images observe a scene from different viewpoints
Feature detection

Detect features using, for example, SIFT [Lowe, IJCV 2004]
Feature matching

Match features between each pair of images
Feature matching
Counting inliers
Counting inliers

Inliers: 3
Counting inliers

Inliers: 20
How do we find the best line?

• Unlike least-squares, no simple closed-form solution

• Hypothesize-and-test
  – Try out many lines, keep the best one
  – RANSAC: Random Sample Consensus

• Number of samples depends on
  – Outlier ratio
  – Probability of correct answer
  – Model size
RANSAC

• General version:
  1. Randomly choose $s$ samples
     • Typically $s = \text{minimum sample size to fit a model}$
  2. Fit a model (say, line) to those samples
  3. Count the number of inliers that approximately fit the model
  4. Repeat $N$ times
  5. Choose the model with the largest set of inliers
RANSAC

\[ N = \frac{\log(1 - p)}{\log(1 - (1 - \epsilon)^s)} \]

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Proportion of outliers $\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td>$s$</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>
Fundamental Matrix
Fundamental Matrix

\[ x_1 \leftrightarrow x_2 \]

\[ x_1^T F x_2 = 0 \]
RANSAC to Estimate Fundamental Matrix

• For $N$ times
  – Pick 8 points
  – Compute a solution for $F$ using these 8 points
  – Count number of inliers with $x_1^TFx_2$ close to 0
• Pick the one with the largest number of inliers
Estimating $F$

• Given just the two images, can we estimate $F$?

• Yes, with enough correspondences.
Estimating F: 8-point algorithm

• The fundamental matrix F is defined by

\[ x'^T F x = 0 \]

for any pair of matches \( x \) and \( x' \) in two images.

• Let \( x = (u,v,1)^T \) and \( x' = (u',v',1)^T \),

\[ F = \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \]

• Each match gives a linear equation:

\[ uu' f_{11} + vu' f_{12} + u' f_{13} + uv' f_{21} + vv' f_{22} + v' f_{23} + uf_{31} + vf_{32} + f_{33} = 0 \]
8-point algorithm

Given $n$ point correspondences, set up a system of equations:

\[
\begin{pmatrix}
    u_1 u'_1 & v_1 u'_1 & u'_1 & u_1 v'_1 & v_1 v'_1 & v'_1 & u_1 & v_1 & 1 \\
    u_2 u'_2 & v_2 u'_2 & u'_2 & u_2 v'_2 & v_2 v'_2 & v'_2 & u_2 & v_2 & 1 \\
    \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
    u_n u'_n & v_n u'_n & u'_n & u_n v'_n & v_n v'_n & v'_n & u_n & v_n & 1
\end{pmatrix}
\begin{pmatrix}
    f_{11} \\
    f_{12} \\
    f_{13} \\
    f_{21} \\
    f_{22} \\
    f_{23} \\
    f_{31} \\
    f_{32} \\
    f_{33}
\end{pmatrix} = 0
\]

- In reality, instead of solving $Af = 0$, we seek $f$ to minimize $\|Af\|$. 
Solving homogeneous systems

• In reality, instead of solving $\mathbf{A}\mathbf{f} = 0$, we seek $\mathbf{f}$ to minimize $\|\mathbf{A}\mathbf{f}\|$.

• Singular value decomposition:

\[
\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^\top
\]

$\mathbf{U}$, $\mathbf{V}$ are rotation matrices

\[
\Sigma = \begin{bmatrix}
  s_1 \\
  \vdots \\
  s_n
\end{bmatrix}
\]

• Solution $\mathbf{f}$ given by the last column of $\mathbf{V}$. 
8-point algorithm: Problem?

- \( \mathbf{F} \) should have rank 2
- To enforce that \( \mathbf{F} \) is of rank 2, \( \mathbf{F} \) is replaced by \( \mathbf{F}' \) that minimizes \( \| \mathbf{F}^\top \mathbf{F}' \| \) subject to the rank constraint.

- This is achieved by SVD. Let \( \mathbf{F} = \mathbf{U}\Sigma\mathbf{V} \), where

\[
\Sigma = \begin{pmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \end{pmatrix}.
\]

Let \( \Sigma' = \begin{pmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \)

then \( \mathbf{F}' = \mathbf{U}\Sigma'\mathbf{V} \) is the solution.
8-point algorithm

% Build the constraint matrix
A = [x2(1,:)'.*x1(1,:)' x2(1,:)'.*x1(2,:)' x2(1,:) ... 
x2(2,:)'.*x1(1,:)' x2(2,:)'.*x1(2,:)' x2(2,:) ... 
x1(1,:)' x1(2,:)'
ones(npts,1) ];

[U,D,V] = svd(A);

% Extract fundamental matrix from the column of V
% corresponding to the smallest singular value.
F = reshape(V(:,9),3,3)';

% Enforce rank 2 constraint
[U,D,V] = svd(F);
F = U * diag([D(1,1) D(2,2) 0]) * V';
8-point algorithm

• Pros: it is linear, easy to implement and fast
• Cons: susceptible to noise
Projective Geometry
Homogeneous coordinates

- Converting to homogeneous coordinates

\[(x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad \text{Homogeneous 2D point} \]

\[(x, y, z) \Rightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad \text{Homogeneous 3D point} \]

- Converting from homogeneous coordinates

\[\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w) \quad \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow (x/w, y/w, z/w) \]

\[(x, y, w)^\top \quad \text{and} \quad (kx, ky, kw)^\top \quad \text{are the same point.} \]
Homogeneous points and lines

- Consider a line in 2D image:
  \[ ax + by + c = 0 \]

- Condition for 2D point \((x, y, 1)^\top\) to lie on the line is:
  \[ (a, b, c)^\top \cdot (x, y, 1)^\top = 0 \]

- Equation of the line in homogeneous coordinates:
  \[ \mathbf{l} = (a, b, c)^\top \]

- Condition for point \(\mathbf{x}\) to lie on line \(\mathbf{l}\) is:
  \[ \mathbf{l}^\top \mathbf{x} = 0 \]

Note: by default, we will denote any vector as a column vector.
Homogeneous points and lines

• Consider two homogeneous 2D points \( \mathbf{x} \) and \( \mathbf{x}' \).
• What is the line passing through these two points?
• Consider two homogeneous 2D points $\mathbf{x}$ and $\mathbf{x}'$.
• What is the line passing through these two points?
• Consider the vector $\mathbf{l} = \mathbf{x} \times \mathbf{x}'$
Homogeneous points and lines

- Consider two homogeneous 2D points $x$ and $x'$.
- What is the line passing through these two points?
- Consider the vector $l = x \times x'$
- We have:
  $$l^T x = 0 \quad l^T x' = 0$$
- Both $x$ and $x'$ lie on the line $l$ and there is a unique line between two points!
Consider two homogeneous 2D points \( x \) and \( x' \).

What is the line passing through these two points?

Consider the vector \( l = x \times x' \)

We have:

\[
    l^\top x = 0 \quad l^\top x' = 0
\]

Both \( x \) and \( x' \) lie on the line \( l \) and there is a unique line between two points!

Consider two lines \( l \) and \( l' \) in the 2D image

What is the point of intersection of these two lines?
Homogeneous points and lines

- Consider two homogeneous 2D points $\mathbf{x}$ and $\mathbf{x}'$.
- What is the line passing through these two points?
- Consider the vector $\mathbf{l} = \mathbf{x} \times \mathbf{x}'$
- We have:
  \[ \mathbf{l}^\top \mathbf{x} = 0 \quad \mathbf{l}^\top \mathbf{x}' = 0 \]
  - Both $\mathbf{x}$ and $\mathbf{x}'$ lie on the line $\mathbf{l}$ and there is a unique line between two points!

- Consider two lines $\mathbf{l}$ and $\mathbf{l}'$ in the 2D image
- What is the point of intersection of these two lines?
  \[ \mathbf{x} = \mathbf{l} \times \mathbf{l}' \]
- Do you notice a duality?
Ideal points
Ideal points

*Jesus before Caiaphas*,
Giotto di Bondone, 1305

*The School of Athens*,
Raphael Sanzio, 1510
Ideal points and the line at infinity

• Consider two parallel lines in the 2D image:
  \[ ax + by + c = 0 \]
  \[ ax + by + c' = 0 \]

• In homogeneous coordinates, the lines are:
  \[ l = (a, b, c)^\top \text{ and } l' = (a, b, c')^\top \]
Ideal points and the line at infinity

- Consider two parallel lines in the 2D image:
  \[ ax + by + c = 0 \]
  \[ ax + by + c' = 0 \]

- In homogeneous coordinates, the lines are:
  \[ l = (a, b, c)^\top \text{ and } l' = (a, b, c')^\top \]

- Their point of intersection is given by:
  \[ x_{\infty} = l \times l' = (c - c')(b, a, 0)^\top \sim (-b, a, 0)^\top \]
Ideal points and the line at infinity

• Consider two parallel lines in the 2D image:
  \[ ax + by + c = 0 \]
  \[ ax + by + c' = 0 \]

• In homogeneous coordinates, the lines are:
  \[ l = (a, b, c)^\top \text{ and } l' = (a, b, c')^\top \]

• Their point of intersection is given by:
  \[ x_\infty = l \times l' = (c - c')( -b, a, 0 )^\top \sim ( -b, a, 0 )^\top \]

• To de-homogeneize involves a division by 0
  • This is a point at “infinity”, called an ideal point
Ideal points and the line at infinity

- Consider two parallel lines in the 2D image:
  \[ ax + by + c = 0 \]
  \[ ax + by + c' = 0 \]

- In homogeneous coordinates, the lines are:
  \[ l = (a, b, c)^\top \text{ and } l' = (a, b, c')^\top \]

- Their point of intersection is given by:
  \[ x_\infty = l \times l' = (c - c')(b, a, 0)^\top \sim (-b, a, 0)^\top \]

- To de-homogeneity involves a division by 0
  - This is a point at “infinity”, called an ideal point

- Which line contains all ideal points \[ x_\infty = (x, y, 0)^\top \]?
  Line at infinity: \[ l_\infty = (0, 0, 1)^\top \].
Two-View Reconstruction
Cross-product as linear operator

**Useful fact:** Cross product with a vector \( \mathbf{t} \) can be represented as multiplication with a \((\text{skew-symmetric})\) 3x3 matrix

\[
[t]_\times = \begin{bmatrix}
0 & -t_z & t_y \\
t_z & 0 & -t_x \\
-t_y & t_x & 0
\end{bmatrix}
\]

\[
\mathbf{t} \times \mathbf{p} = [\mathbf{t}]_\times \mathbf{p}
\]

What is the rank of \([\mathbf{t}]_\times\)?
Two-view geometry

Corresponding point in other image is constrained to lie on a line, called the *epipolar line*.

![Diagram showing epipolar lines and planes in two views.](image)
Epipoles

Two special points: $e_1$ and $e_2$ (the *epipoles*): projection of one camera into the other.
Epipolar lines

Two special points: $e_1$ and $e_2$ (the epipoles): projection of one camera into the other.

All of the epipolar lines in an image pass through the epipole.
Essential matrix

• Assume calibrated cameras with $K_1 = K_2 = I_{3x3}$.
• Let camera 1 be $[I, 0]$ and camera 2 be $[R, t]$.
• In camera 1 coordinates, 3D point $X$ is given by $X_1 = \lambda_1 p$.
• In camera 2 coordinates, 3D point $X$ is given by $X_2 = \lambda_2 q$.
• Since camera 2 is related to camera 1 by rigid-body motion $[R, t]$
  \[
  X_2 = RX_1 + t \\
  \lambda_2 q = \lambda_1 Rp + t
  \]
Essential matrix

- We have: $\lambda_2 q = \lambda_1 R p + t$
- Take cross-product with respect to $t$:
  $$\lambda_2 [t] \times q = \lambda_1 [t] \times Rp$$
- Take dot-product with respect to $q$:
  $$0 = \lambda_1 q^\top [t] \times Rp$$
Essential matrix

- We have: \( q^\top [t] \times R p = 0 \)
- Define: \( E = [t] \times R \)
- Then, we have: \( q^\top E p = 0 \)

How many degrees of freedom does \( E \) have?
Relax the assumption of calibrated cameras.

Then, $p$ and $q$ are in metric coordinates and pixel counterparts are:

$$p' = K_1 p \quad q' = K_2 q$$

Recall essential matrix constraint:

$$q^\top E p = 0$$

Substituting, we have:

$$\left(K_2^{-1} q'\right)^\top E \left(K_1^{-1} p'\right) = 0$$
- Essential matrix constraint in pixel space: \((K^{-1}_2 q')^\top E (K^{-1}_1 p') = 0\).
- Rearranging:
  \[ q'^\top (K^{-\top}_2 E K^{-1}_1) p' = 0 \]
- Define: \(F = K^{-\top}_2 E K^{-1}_1\)
- Then, we have: \(q'^\top F p' = 0\)

How many degrees of freedom does \(F\) have?
Properties of the Fundamental Matrix

- $F_p'$ is the epipolar line associated with $p'$
- $F^\top q'$ is the epipolar line associated with $q'$
- $F e'_1 = 0$ and $F^\top e'_2 = 0$
- $F$ is rank 2.
Results (ground truth)

Ground truth with standard stereo calibration