3D Reconstruction

COMPSCI 527 — Computer Vision
Outline

1. The Epipolar Geometry of a Pair of Cameras
2. The Essential Matrix
3. The Eight-Point Algorithm: $t, R$
4. Triangulation: $P_m$
5. Bundle Adjustment
The Epipolar Geometry of a Pair of Cameras

The epipolar plane
projection ray
projection ray
epipolar plane
epipolar line of camera a
epipolar line of camera b
epipole e_a
depipole e_b
center of projection
camera a
camera b

P
projection ray

The Epipolar Constraint

- The point \( p_a \) in image \( a \) that corresponds to point \( p_b \) in image \( b \) is on the epipolar line of \( p_b \)
  ... and *vice versa*
- This is the only general constraint between two images of the same scene; 3D reconstruction depends on it
- Epipolar lines come in corresponding pairs
- Two *pencils* of lines supported by the two epipoles
Another Way to State the Epipolar Constraint

The two projection rays and the baseline are *coplanar* for corresponding points.
The Epipolar Constraint and 3D Reconstruction

• Relative position and orientation of the two cameras are unknown
• Given corresponding points \( a^p a, b^p b \) (found, say, by tracking) we can write one algebraic constraint on \( a^R b \) and \( a^t b \)
• With enough pairs of corresponding points, we can write a system of equations in \( a^R b \) and \( a^t b \) and solve it
• We can then solve for the coordinates of the 3D points whose images we have
• Solving the system is 3D reconstruction
The Essential Matrix

• How to write the epipolar constraint algebraically?
• The constraint is nonlinear in $^aR_b$, $^at_b$
• Introduce a new $3 \times 3$ essential matrix $E$ that combines rotation and translation to make motion estimation a linear problem in $E$
• Computation sequence:
  • Find $E$ by solving a homogeneous linear system
  • Find rotation and translation from $E$
  • Find structure (3D points in the world) by intersecting projection rays
Coordinates

• (Known) image points as world points:
  \[ a\mathbf{p}_a = \begin{bmatrix} a^x_a \\ a^y_a \\ f \end{bmatrix} \quad \text{and} \quad b\mathbf{p}_b = \begin{bmatrix} b^x_b \\ b^y_b \\ f \end{bmatrix} \]

• Each camera measures a point \textit{in its own reference system}

• (Unknown) transformation: \[ b\mathbf{p} = aR_b(a\mathbf{p} - a\mathbf{t}_b) \]

• Inverse: \[ aR^T_b b\mathbf{p}_b + a\mathbf{t}_b \]
Writing all Quantities in System $a$

- Pose of camera $b$ in $a$ is specified by $^aR_b$, $^at_b$, both in $a$
- Image point $^ap_a$ is in $a$
- Image point $^bp_b$ is in $b$, need to transform to $^ap_b$
- Invert $^bp_b = ^aR_b(^ap_b - ^at_b)$ to obtain $^ap_b = ^aR_b^Tb^bp_b + ^at_b$
- Too many super/subscripts to keep track of. Define $a = ^ap_a$, $b = ^bp_b$, $R = ^aR_b$, $t = ^at_b$, $e = ^ae_b$
- $ab \overset{\text{def}}{=} ^ap_b = R^Tb + t$
Aside: Epipole and Translation

- The epipole of $b$ in $a$ is the same as $t$ up to norm
- Define: $e = a e_b$
- $e \propto t$
The Epipolar Constraint, Algebraically

\[ \mathbf{a} \mathbf{b} = R^T \mathbf{b} + \mathbf{t} \]

- The two projection rays and the baseline are coplanar.
- The triple product of \( \mathbf{a} \mathbf{b}, \mathbf{t}, \) and \( \mathbf{a} \) is zero: \( \mathbf{a} \mathbf{b}^T (\mathbf{t} \times \mathbf{a}) = 0 \)

\[ (R^T \mathbf{b} + \mathbf{t})^T (\mathbf{t} \times \mathbf{a}) = 0, \text{ but } \mathbf{t}^T (\mathbf{t} \times \mathbf{a}) = 0 \text{ so that} \]

\[ (R^T \mathbf{b})^T (\mathbf{t} \times \mathbf{a}) = 0 \]
The Essential Matrix

\[(R^Tb)^T(t \times a) = 0\]
\[b^TR(t \times a) = 0\]
\[b^TR[t]_\times a = 0\]

where \(t = (t_x, t_y, t_z)^T\) and \([t]_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}\)

\[b^T E a = 0 \quad \text{where} \quad E = R \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}\]

- This equation is the \textit{epipolar constraint}, written in algebra
- Holds for any corresponding \(a, b\) in the two images (as world vectors in their reference systems)
- \(E\) is the \textit{essential matrix}
- The epipolar constraint is linear in \(E\) but not in \(R\) and \(t\)
The Epipolar Line in Image $a$

- Think of $b$ as fixed
- What points $x$ in image $a$ satisfy the epipolar constraint? $b^T E x = 0$
- Let $\lambda^T = b^T E$, a row vector
- $\lambda^T x = 0$, a line!
- $a$ satisfies this *homogeneous* equation (epipolar constraint)
- So does $t$: $\lambda^T t = b^T Et = b^T R [t] \times t = 0$
  
  ... and therefore $e$
- So the line is the epipolar line of $b$
Two Key Problems

• How to find $E$ given many pairs of corresponding points
  • Easy because $b^T E a = 0$ is linear and homogeneous in $E$
  • Let us postpone the details
• How to break up $E$ into $R$ and $t$
  • A bit trickier because $E = R [t]_\times$ is nonlinear in $R$ and $t$
  • Let us do this first
The Structure of $E = R \ [t] \times$: Rank and Null Space

- $E$ has rank 2 and $\text{null}(E) = \text{span}(t) = \text{span}(e)$
- Geometry:
  - The epipole $e$ is in the epipolar line $b^T Ex = 0$ for every $b$
  - Therefore, $b^T Ee = 0$ for all $b$
  - Therefore $Ee = 0$, so $e \in \text{null}(E)$
- Algebra:
  - $[t] \times t = t \times t = 0$
  - $t \times v \neq 0$ if $v$ is not parallel to $t$
  - Therefore, the rank of $[t] \times$ is 2 for $t \neq 0$
  - Since $R$ is full rank, the solutions of $[t] \times x = 0$ and $Ex = 0$ (i.e., $R[t] \times x = 0$) are the same
  - Therefore, $\text{rank}(E) = 2$ for nonzero $t$ and $\text{null}(E) = \text{span}(t)$
  - Either way, $\text{null}(E) = \text{span}(e) = \text{span}(t) = \text{baseline}$
The Structure of $E$: Singular Values

- $E$ has two equal singular values and one zero singular value

**Proof**

- Let $\mathbf{v}$ be perpendicular to $\mathbf{t}$.
  Then $\| [\mathbf{t}] \times \mathbf{v} \| = \| \mathbf{t} \| \| \mathbf{v} \|$
  (geometric definition of cross product)
- Let $\| \mathbf{v} \| = 1$. Then $\| [\mathbf{t}] \times \mathbf{v} \| = \| \mathbf{t} \|$
- $\mathbf{v} \perp \mathbf{t}$ means that $\mathbf{v} \in \text{row space}([\mathbf{t}] \times)$
  because $\text{null}(E) = \text{span}(\mathbf{t})$
- Therefore, all unit-norm vectors $\mathbf{v} \in \text{row space}([\mathbf{t}] \times)$ are mapped to a circle
- Therefore $[\mathbf{t}] \times$ has two equal singular values
- Third is zero because $\mathbf{t} \in \text{null}([\mathbf{t}] \times)$
- Ditto for $E$, since $E = R[\mathbf{t}] \times$ and $R$ is orthogonal
- Therefore $\mathbf{v}_3 \propto \mathbf{e} \propto \mathbf{t}$ and $\sigma_1 = \sigma_2 = \sigma$

- If we have $E$, we can find camera translation $\mathbf{t}$ by SVD!
A Fundamental Ambiguity

• The equation $b^T E a = 0$ is homogeneous in $E$
• Therefore, we cannot tell the magnitude of $E$, or of $t$ in $E = R \, [t]_x$
• Absolute scale cannot be determined from images alone
• This ambiguity is general, has nothing to do with the specifics of the formulation
• Cameras fundamentally measure angles, not distances
• This ambiguity is often exploited in movie special effects
• W.l.o.g., let $\|t\| = 1$
• Measure everything in units of inter-camera distance
How to Find $E$?

- Given pairs $(\mathbf{a}_1, \mathbf{b}_1), \ldots (\mathbf{a}_n, \mathbf{b}_n)$
- Write one epipolar constraint equation per pair
- Linear and homogeneous in $E$

\[
\mathbf{b}^T E \mathbf{a} = 0
\]
The Eight-Point Algorithm

- H. C. Longuet-Higgins, *Nature*, 293:133–135, 1981
- Needs *at least* 8 corresponding point pairs
- Preferably many more
- Overview:
  - Given pairs \((a_1, b_1), \ldots, (a_n, b_n)\) (tracking)
  - Write one epipolar constraint equation \(b_m^T E a_m = 0\) per pair
  - Solve linear system \(b_1^T E a_1 = 0, \ldots, b_n^T E a_n = 0\) for \(E\)
  - Solve \(E = R [t]_\times\) for \(t, R\)
  - Compute the 3D structure (points \(P_m\)) from \(a_m, b_m, t, R\)
- The last step is called *triangulation*
Rewriting the Epipolar Constraint

\[ \mathbf{b}^T E \mathbf{a} = 0 \]

\[
\begin{bmatrix}
  b_1 & b_2 & b_3 \\
\end{bmatrix}
\begin{bmatrix}
  e_{11} & e_{12} & e_{13} \\
  e_{21} & e_{22} & e_{23} \\
  e_{31} & e_{32} & e_{33} \\
\end{bmatrix}
\begin{bmatrix}
  a_1 \\
  a_2 \\
  a_3 \\
\end{bmatrix} = 0
\]

\[ e_{11} a_1 b_1 + e_{12} a_2 b_1 + e_{13} a_3 b_1 +
  e_{21} a_1 b_2 + e_{22} a_2 b_2 + e_{23} a_3 b_2 +
  e_{31} a_1 b_3 + e_{32} a_2 b_3 + e_{33} a_3 b_3 = 0 \]

\[
\begin{bmatrix}
  a_1 b_1 & a_2 b_1 & a_3 b_1 & a_1 b_2 & a_2 b_2 & a_3 b_2 & a_1 b_3 & a_2 b_3 & a_3 b_3 \\
\end{bmatrix}
\begin{bmatrix}
  e_{11} \\
  e_{12} \\
  e_{13} \\
  e_{21} \\
  e_{22} \\
  e_{23} \\
  e_{31} \\
  e_{32} \\
  e_{33} \\
\end{bmatrix} = 0
\]

\[ \mathbf{c}^T \eta = 0 \]

- With \( n \) point pairs, \( \mathbf{c}_m^T \eta = 0 \) for \( m = 1, \ldots, n \)
Solving for $E$

$c_m^T \eta = 0$ for $m = 1, \ldots, n$

$C \eta = 0$ where $C$ is $n \times 9$

- Because of the scale ambiguity, we cannot tell the norm of $\eta$
- Set $\| \eta \| = 1$
- Homogeneous, least squares problem on the unit sphere
- We know how to solve that!

- Repackage $\eta$ into $3 \times 3$ matrix $E$
Solving for $t$

- We have $E$ now
  \[ E = R \ [t] \times \]
- We saw that $\text{null}(E) = \text{span}(t)$
- So we know how to find $t$ with $\|t\| = 1$, \textit{up to a sign}
- $\pm t$ (and also $\pm [t] \times$)
Solving for $R$

- We have both $E$ and $T = \pm [t]_\times$
  \[ E = R [t]_\times \]
- Linear system in $R$, but with the constraints $R^T R = I$ and $\det(R) = 1$
- Linear, constrained LSE optimization problem: The Procrustes problem, $\arg \min_{R^T R = I} \| E - RT \|_F$
- Appendix in the notes gives a solution based on the SVD
- Since $T$ has rank 2, it turns out that the there are two solutions, $R_1$ and $R_2$ for each choice of sign in $T = \pm [t]_\times$
Eight-Point Algorithm So Far

- Given \( n \geq 8 \) image point pairs \((a_m, b_m)\) for \( m = 1, \ldots, n\)
- Solve \( n \times 9 \) linear homogeneous system \( b_m^T E a_m = 0 \) for \( E \)
- Compute \( \pm t \) as the third right singular vector of \( \pm E \)
- Solve \( \pm E = R \pm [t]_x \) for \( R \) by Procrustes (linear problem with orthogonality constraint) to obtain \( R_1, R_2 \)
- We obtain two translations \( \pm t \) and two rotations \( R_1, R_2 \)
- Four combinations: \((t, R_1), (−t, R_1), (−t, R_2), (t, R_2)\)
- Which is the right one?
- Let us first *triangulate*: Reconstruct the world points given one solution \((t, R)\)
- Only one of the four sets of world points will make sense
Triangulation

- For simplicity, divide $a' = \begin{bmatrix} a'_1 \\ a'_2 \\ f \end{bmatrix}$ by $f$ so that now $a = \begin{bmatrix} a_1 \\ a_2 \\ 1 \end{bmatrix}$

- Let $\alpha \overset{\text{def}}{=} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ (coordinates in canonical image reference system)

- Ditto for $b, \beta$

- Projection equations in each camera reference frame: $A$ is $P$ in frame $a$

  - $\alpha = \frac{1}{A_3} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$ and $\beta = \frac{1}{B_3} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$

- Rewrite as $\alpha A_3 = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$ and $\beta B_3 = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$

  - Plug $B = R(A - t)$ into second set of equations

- All equations are linear. Four equations, 3 unknowns

- Solve in the LSE sense, get a modicum of noise rejection
The Fourfold Ambiguity

\((t, R_1), (-t, R_1), (-t, R_2), (t, R_2)\)

- Only one solution places all world points in front of both cameras
- Try all four solutions, and reconstruct world points by triangulation
- Pick the one solution that makes sense
Summary of Eight-Point Algorithm

- Given \( n \geq 8 \) image point pairs \((a_m, b_m)\) for \( m = 1, \ldots, n\)
- Solve \( n \times 9 \) linear homogeneous system \( b_m^T E a_m = 0 \) for \( E \)
- Compute \( \pm t \) as the third right singular vector of \( \pm E \)
- Solve \( \pm E = R \pm [t]_\times \) for \( R \) by Procrustes (linear problem with orthogonality constraint) to obtain \( R_1, R_2 \)
- Triangulate scene points \( P_m \) from \( a_m, b_m, t, R \) and for all four combinations of \( t \) and \( R \)
  (\( n \) separate problems, one per point pair)
- Choose the one combination of \( t, R \) that places world points in front of both cameras
- Keep the corresponding triangulated scene points \( P_m \)
- Everything is found up to a single, global scale factor
Bundle Adjustment

Let $\pi$ be the perspective projection function. We are after

$$\arg \min_{t, R, A_1, \ldots, A_n} \frac{1}{2n} \sum_{m=1}^{n} \left[ \| a_m - \pi(A_m) \|^2 + \| b_m - \pi(R(A_m - t)) \|^2 \right]$$

reprojection error

$$\arg \min_{t, R, A_1, \ldots, A_n} \rho(t, R, A_1, \ldots, A_n)$$

Eight-point algorithm solves this single optimization problem in multiple steps

This greedy approach leads to a suboptimal solution

Use solution $t, R, P_1, \ldots, P_n$ from 8-point algorithm to initialize a gradient-descent search for an optimal solution to the full problem

This fine-tuning step is called bundle adjustment