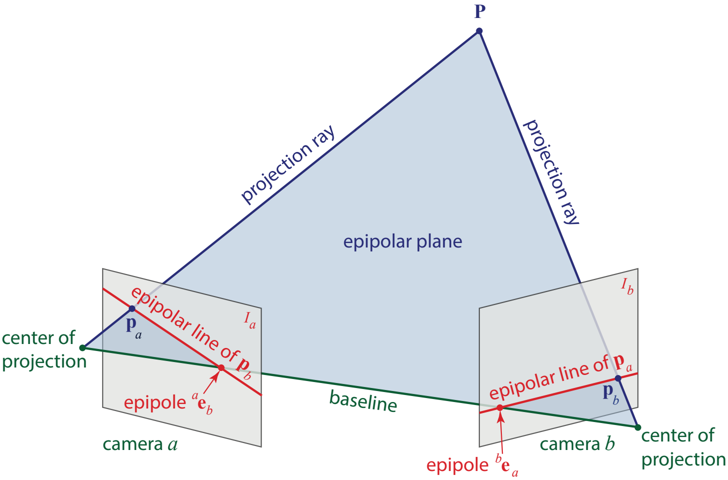
3D Reconstruction

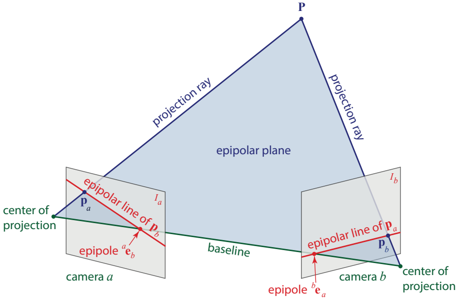
COMPSCI 527 — Computer Vision

Outline

The Epipolar Geometry of a Pair of Cameras

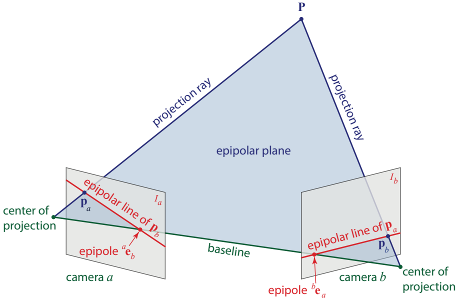


The Epipolar Constraint



- The point $\mathbf{p}_a$ in image *a* that corresponds to point $\mathbf{p}_b$ in image *b* is on the epipolar line of $\mathbf{p}_b$
  ... 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

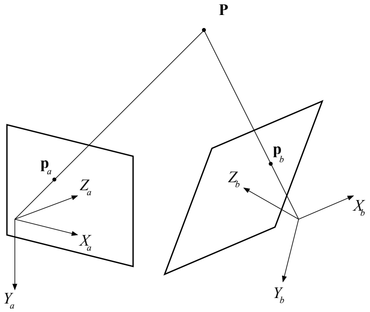
The Epipolar Constraint and 3D Reconstruction

- Relative position and orientation of the two cameras are *unknown*
- Given corresponding points ${}^a\mathbf{p}_a$, ${}^b\mathbf{p}_b$ (found, say, by tracking) we can write one algebraic constraint on ${}^aR_b$ and ${}^a\mathbf{t}_b$
- With enough pairs of corresponding points, we can write a system of equations in ${}^aR_b$ and ${}^a\mathbf{t}_b$ 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 ${}^aR_b$, ${}^a\mathbf{t}_b$
- Introduce a new 3×3 *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



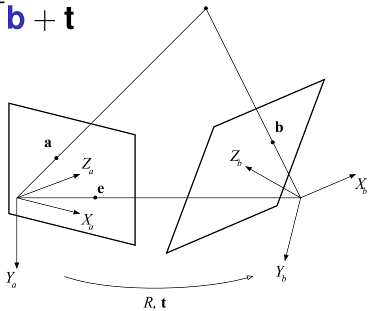
- (Known) image points as world points:

$$ {}^a\mathbf{p}_a = \begin{bmatrix} {}^a x_a \\ {}^a y_a \\ f \end{bmatrix} \quad \text{and} \quad {}^b\mathbf{p}_b = \begin{bmatrix} {}^b x_b \\ {}^b y_b \\ f \end{bmatrix} $$

- Each camera measures a point *in its own reference system*
- (Unknown) transformation: ${}^b\mathbf{p} = {}^aR_b({}^a\mathbf{p} - {}^a\mathbf{t}_b)$
- Inverse: ${}^aR_b^T\,{}^b\mathbf{p}_b + {}^a\mathbf{t}_b$

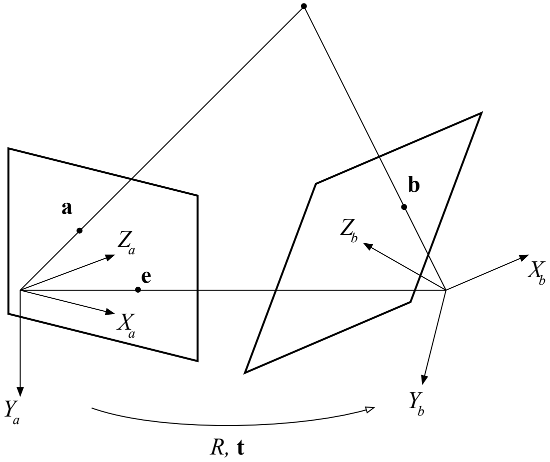
Writing all Quantities in System a

- Pose of camera b in a is specified by ${}^aR_b$, ${}^a\mathbf{t}_b$, both in a
- Image point ${}^a\mathbf{p}_a$ is in a
- Image point ${}^b\mathbf{p}_b$ is in b, need to transform to ${}^a\mathbf{p}_b$
- Invert ${}^b\mathbf{p}_b = {}^aR_b({}^a\mathbf{p}_b - {}^a\mathbf{t}_b)$ to obtain ${}^a\mathbf{p}_b = {}^aR_b^T\,{}^b\mathbf{p}_b + {}^a\mathbf{t}_b$
- Too many super/subscripts to keep track of. Define
  $\mathbf{a} = {}^a\mathbf{p}_a \quad , \quad \mathbf{b} = {}^b\mathbf{p}_b \quad , \quad R = {}^aR_b \quad , \quad \mathbf{t} = {}^a\mathbf{t}_b \quad , \quad \mathbf{e} = {}^a\mathbf{e}_b$
- ${}^a\mathbf{b} \stackrel{\text{def}}{=} {}^a\mathbf{p}_b = R^T\mathbf{b} + \mathbf{t}$

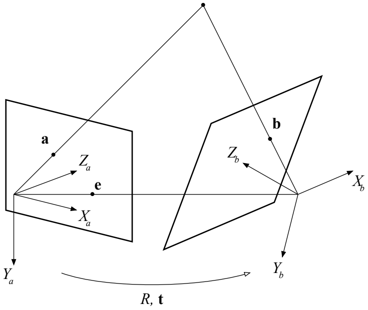


Aside: Epipole and Translation

- The epipole of b in a is the same as $\mathbf{t}$ up to norm
- Define: $\mathbf{e} = {}^a\mathbf{e}_b$
- $\mathbf{e} \propto \mathbf{t}$



The Epipolar Constraint, Algebraically



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

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

The Essential Matrix

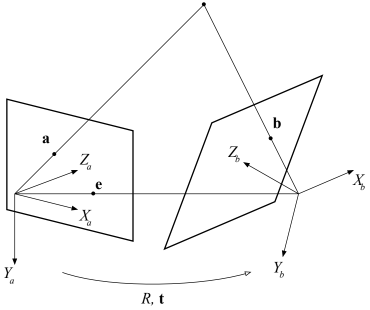
$$(R^T\mathbf{b})^T(\mathbf{t}\times\mathbf{a}) = 0$$
$$\mathbf{b}^T R\,(\mathbf{t}\times\mathbf{a}) = 0$$
$$\mathbf{b}^T R\,[\mathbf{t}]_\times\mathbf{a} = 0$$

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

$\mathbf{b}^T E\,\mathbf{a} = 0 \quad$ where $\quad E = R\,[\mathbf{t}]_\times$

- This equation is the *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 *essential matrix*
- The epipolar constraint is linear in E but not in R and $\mathbf{t}$

The Epipolar Line in Image *a*



- Think of $\mathbf{b}$ as fixed
- What points $\mathbf{x}$ in image *a* satisfy the epipolar constraint? $\mathbf{b}^T E \mathbf{x} = 0$
- Let $\boldsymbol{\lambda}^T = \mathbf{b}^T E$, a row vector
- $\boldsymbol{\lambda}^T \mathbf{x} = 0$, a line!
- $\mathbf{a}$ satisfies this *homogeneous* equation (epipolar constraint)
- So does $\mathbf{t}$: $\boldsymbol{\lambda}^T \mathbf{t} = \mathbf{b}^T E \mathbf{t} = \mathbf{b}^T R [\mathbf{t}]_\times \mathbf{t} = 0$

  ... and therefore $\mathbf{e}$
- So the line is the epipolar line of $\mathbf{b}$

Two Key Problems

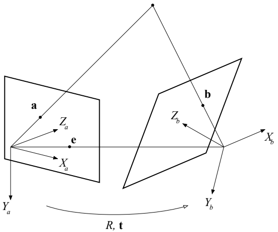
- How to find E given many pairs of corresponding points
  - Easy because $\mathbf{b}^T E \mathbf{a} = 0$ is linear and homogeneous in E
  - Let us postpone the details
- How to break up E into R and $\mathbf{t}$
  - A bit trickier because $E = R\,[\mathbf{t}]_\times$ is nonlinear in R and $\mathbf{t}$
  - Let us do this first

The Structure of $E = R\,[\mathbf{t}]_\times$: Rank and Null Space

- E has rank 2 and $\text{null}(E) = \text{span}(\mathbf{t}) = \text{span}(\mathbf{e})$
- Geometry:
  - The epipole $\mathbf{e}$ is in the epipolar line $\mathbf{b}^T E\mathbf{x} = 0$ for every $\mathbf{b}$
  - Therefore, $\mathbf{b}^T E\mathbf{e} = 0$ for all $\mathbf{b}$
  - Therefore $E\mathbf{e} = \mathbf{0}$, so $\mathbf{e} \in \text{null}(E)$
- Algebra:

  $[\mathbf{t}]_\times \mathbf{t} = \mathbf{t} \times \mathbf{t} = \mathbf{0}$

  $\mathbf{t} \times \mathbf{v} \neq \mathbf{0}$ if $\mathbf{v}$ is not parallel to $\mathbf{t}$
  - Therefore, the rank of $[\mathbf{t}]_\times$ is 2 for $\mathbf{t} \neq \mathbf{0}$
  - Since R is full rank, the solutions of $[\mathbf{t}]_\times \mathbf{x} = \mathbf{0}$ and $E\mathbf{x} = \mathbf{0}$ (*i.e.*, $R[\mathbf{t}]_\times \mathbf{x} = \mathbf{0}$) are the same
  - Therefore, $\text{rank}(E) = 2$ for nonzero $\mathbf{t}$ and $\text{null}(E) = \text{span}(\mathbf{t})$



- Either way, $\text{null}(E) = \text{span}(\mathbf{e}) = \text{span}(\mathbf{t}) = \text{baseline}$

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 $\mathbf{v} \in$ row space$([\mathbf{t}]_\times)$
    because null$(E) =$ span$(\mathbf{t})$
  - Therefore, all unit-norm vectors
    $\mathbf{v} \in$ row space$([\mathbf{t}]_\times)$ are mapped to a circle
  - Therefore $[\mathbf{t}]_\times$ has two equal singular values
  - Third is zero because $\mathbf{t} \in$ 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 $\mathbf{b}^T E \mathbf{a} = 0$ is homogeneous in E
- Therefore, we cannot tell the magnitude of E, or of $\mathbf{t}$ in $E = R\,[\mathbf{t}]_\times$
- 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 $\|\mathbf{t}\| = 1$
- Measure everything in units of inter-camera distance

How to Find E?

$$\mathbf{b}^T E \mathbf{a} = 0$$

- 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

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 $(\mathbf{a}_1, \mathbf{b}_1), \ldots (\mathbf{a}_n, \mathbf{b}_n)$ (tracking)
  - Write one epipolar constraint equation $\mathbf{b}_m^T E \, \mathbf{a}_m = 0$ per pair
  - Solve *linear* system $\mathbf{b}_1^T E \, \mathbf{a}_1 = 0, \ \ldots, \mathbf{b}_n^T E \, \mathbf{a}_n = 0$ for E
  - Solve $E = R \, [\mathbf{t}]_\times$ for $\mathbf{t}$, R
  - Compute the 3D structure (points $\mathbf{P}_m$) from $\mathbf{a}_m$, $\mathbf{b}_m$, $\mathbf{t}$, R
- The last step is called *triangulation*

Rewriting the Epipolar Constraint

$$\mathbf{b}^T E \mathbf{a} = 0$$

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

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

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

$$\mathbf{c}^T \boldsymbol{\eta} = 0$$

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

Solving for E

$$\mathbf{c}_m^T \boldsymbol{\eta} = 0 \text{ for } m = 1, \ldots, n$$
$$C\boldsymbol{\eta} = \mathbf{0} \text{ where } C \text{ is } n \times 9$$

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

- Repackage η into 3×3 matrix E

Solving for **t**

- We have E now

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

Solving for R

- We have both E and $T = \pm[\mathbf{t}]_\times$
  $E = R\,[\mathbf{t}]_\times$
- 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* $T = \pm[\mathbf{t}]_\times$

Eight-Point Algorithm So Far

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

Triangulation

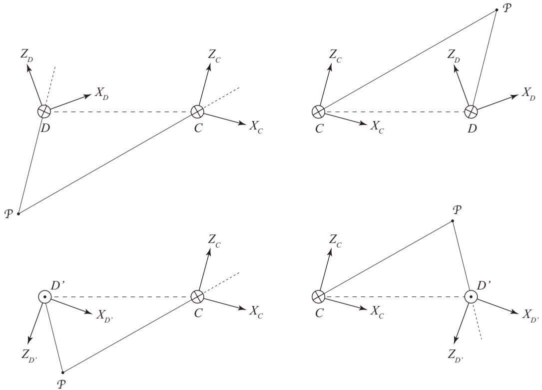
- For simplicity, divide $\mathbf{a}' = \begin{bmatrix} a'_1 \\ a'_2 \\ f \end{bmatrix}$ by f so that now $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ 1 \end{bmatrix}$
- Let $\boldsymbol{\alpha} \stackrel{\text{def}}{=} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ (coordinates in canonical image reference system)
- Ditto for $\mathbf{b}$, $\boldsymbol{\beta}$
- Projection equations in each camera reference frame: $\mathbf{A}$ is $\mathbf{P}$ in frame a

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

  Plug $\mathbf{B} = R(\mathbf{A} - \mathbf{t})$ 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

$$(\mathbf{t}, R_1)\,,\ (-\mathbf{t}, R_1)\,,\ (-\mathbf{t}, R_2)\,,\ (\mathbf{t}, R_2)$$



- 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 $(\mathbf{a}_m, \mathbf{b}_m)$ for $m = 1, \ldots, n$
- Solve $n \times 9$ linear homogeneous system $\mathbf{b}_m^T E\, \mathbf{a}_m = 0$ for E
- Compute $\pm\mathbf{t}$ as the third right singular vector of $\pm E$
- Solve $\pm E = R\ \pm [\mathbf{t}]_\times$ for R by Procrustes (linear problem with orthogonality constraint) to obtain R_1, R_2
- Triangulate scene points $\mathbf{P}_m$ from $\mathbf{a}_m$, $\mathbf{b}_m$, $\mathbf{t}$, R and for all four combinations of $\mathbf{t}$ and R

  (n separate problems, one per point pair)
- Choose the one combination of $\mathbf{t}$, R that places world points in front of both cameras
- Keep the corresponding triangulated scene points $\mathbf{P}_m$
- Everything is found up to a single, global scale factor

Bundle Adjustment

- Let π be the perspective projection function. We are after

$$\arg\min_{\mathbf{t}, R, \mathbf{A}_1, \ldots, \mathbf{A}_n} \frac{1}{2n} \sum_{m=1}^{n} \underbrace{\left[\|\mathbf{a}_m - \pi(\mathbf{A}_m)\|^2 + \|\mathbf{b}_m - \pi(R(\mathbf{A}_m - \mathbf{t}))\|^2 \right]}_{\textit{reprojection error}}$$

$$\arg\min_{\mathbf{t}, R, \mathbf{A}_1, \ldots, \mathbf{A}_n} \rho(\mathbf{t}, R, \mathbf{A}_1, \ldots, \mathbf{A}_n)$$

- Eight-point algorithm solves this single optimization problem in multiple steps
- This greedy approach leads to a suboptimal solution
- Use solution $\mathbf{t}, R, \mathbf{P}_1, \ldots, \mathbf{P}_n$ 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*