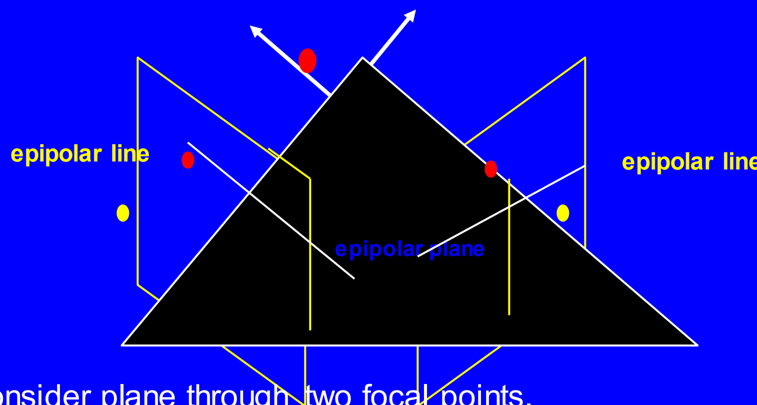
Structure-from-Motion: Perspective

- We'll just consider two images.
- Known cameras, perspective projection.
- Known corresponding points in each image.
- Want to recover relative position of cameras, 3D position of points.
- Given camera position, point recovery is just like stereo.

Strategy: Focus on Epipolar Geometry

- Recall from stereo, images have matching lines.
  - All points on line in one image match some point on matching line in other image.
  - Which lines match depends on relative camera positions.
- We use matching points to find matrix, *E,* that encodes epipolar lines.
  - This can also be useful to find more matching points.
- Once we have *E,* we can find relative camera positions.

So first, recall epipolar geometry



Consider plane through two focal points.

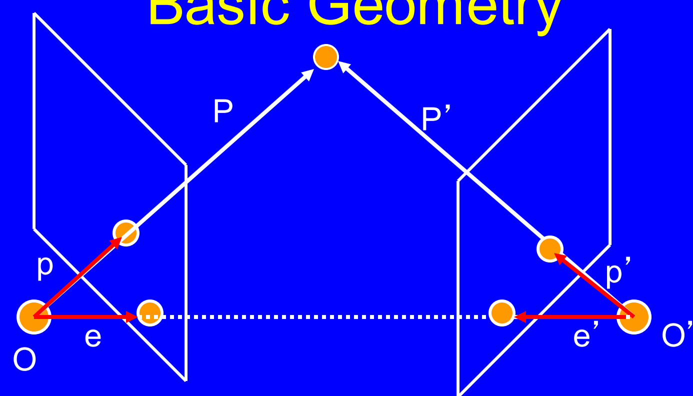
Projects to line in each image.

All points in world are on one of these planes.

Epipole

- Consider line through focal points.
  - All these planes contain this line.
  - This line intersects each image plane in epipole.
  - Every pair of epipolar lines has to go through these epipoles.
- Epipoles depend on direction of translation.

Basic Geometry



Relationship between images is translation, T, and rotation, R.

$T = (O' - O)$ $P' = R^T(P - T)$

Key Point about Epipoles

- The *Essential* matrix, E
  - $p'^T E p = 0$
  - Given either point, we have a linear equation in the other.
  - Questions are: how exactly do we get this, and what is relation between E and motion?

Digression: Representing coplanarity

Suppose U and V are unit vectors. Let W be a unit vector that is perpendicular to both of them. We write:

W = U x V. We can write:

$$W = \begin{pmatrix} 0 & -U_z & U_y \\ U_z & 0 & -U_x \\ -U_y & U_x & 0 \end{pmatrix} V$$

To verify this, consider $V^TW=V_x(-U_zV_y+U_yV_z) + V_y(U_zV_x-U_xV_z)+V_z(-U_yV_x+U_xV_y)=0$

$U^TW=U_x(-U_zV_y+U_yV_z) + U_y(U_zV_x-U_xV_z)+U_z(-U_yV_x+U_xV_y)=0$

So

p, T, and Rp' are coplanar.

$$s = \begin{pmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{pmatrix}$$

p (S*R) p' = 0. Define E = (S*R). And we have pEp' =0.

The Eight Point Algorithm

- $p^T E p'=0$. Each correspondence gives a linear equation in the unknowns of the matrix E.
- Scale of E doesn't matter, so it has only 8 unknowns.
- So with 8 correspondences we can determine E linearly.

Some subtleties when data isn't perfect

- E has rank 2 (because S does).
- With noise, we'll get an estimate of E with full rank. Find best rank 2 approximation.
- Use more points. Write overconstrained system as Ae=0 (e entries of E in vector form). Normalize values appropriately. Then SVD of A. Vector that is closest to its kernel gives entries of E. Then use SVD to enforce E has rank 2.

Some practice problems. In all these, there is no rotation, so S=E.

Suppose O = (0,0,0) and O' = (3,2,4), with a focal length of 1 and no rotation. Then the matrix E is:
(0 -4 2
4 0 -3
-2 3 0)

If p and p' are corresponding points, show pEp'=0.
Example: Suppose P = (16, 24, 8). p = (2,3,1). To find p' we can translate everything so O' is the origin and P would be at (13,22,4). Then we'd have p'=(3.25, 5.5,1). Then $p^T E$ = (10, -5, -5). And we'd have $p^T Ep'$=32.5-27.5-5=0.

If p, p' and T are coplanar, then if $p^T Ep'$=0.

$$E = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$

First, let's take an example.
Simplest way for them to be
Coplanar is if all have z values of 0. So then
E = (0 0 Ty; 0 0 Tx; -Ty Tx 0)
Then Ep' = (0, 0, something), and pEp' = 0.

$$E = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$

Suppose p1 = (3,7), p1'=(1,7), p2=(5,2), p2'=(2,2), p3=(6,4), p3'=(1,4), where p1 and p1' are corresponding points in the first and second images. What is E?

p1Ep1'=0 => (3,7,1)(-7Tz+Ty,Tz-Tx,-Ty+7Tx)^T = 0

-21Tz+3Ty+7Tz-7Tx-Ty+7Tx=0

2Ty-14Tz=0

p2Ep2'=0 =>(5,2,1)(-2Tz+Ty,2Tz-Tx,-2Ty+2Tx)=0

-10Tz+5Ty+4Tz-2Tx-2Ty+2Tx=0

3Ty-6Tz=0

Solving these two equations we get Ty=0, Tz=0.

This means the last equation is:

p3Ep3'=0 => (6,4,1)(0,-Tx,4Tx)=0

-4Tx+4Tx=0. So we can't know Tx.

$$E = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$

Suppose we have a camera with a focal length of 1 and focal point of (0,0,0), which translates by (1,1,1)

Suppose P1=(0,0,2), p1=(0,0,1), p1'=(-1,-1,1).

P2=(2,0,2),p2=(1,0,1), p2'=(1,-1,1).

P3=(0,2,2),p3=(0,1,1), p3'=(-1,1,1).

Can we recover the translation using pEp'=0?

p1Ep1'=0 => (0,0,1)(Tz+Ty,-Tz-Tx,Ty-Tx)=0, Ty-Tx=0

Ty=Tx

p2Ep2'=0 => (1,0,1)(Tz+Ty,Tz-Tx,-Ty-Tx)=0, Tz+Ty-Ty-Tx=0, Tz=Tx.

p3Ep3'=0 => (0,1,1)(…,-Tz-Tx,Ty+Tx)=0, -Tz+Ty=0.

We get Tx=Ty=Tz, but we can recover the overall magnitude.