EPIPOLAR GEOMETRY, THE ESSENTIAL & FUNDAMENTAL MATRICES

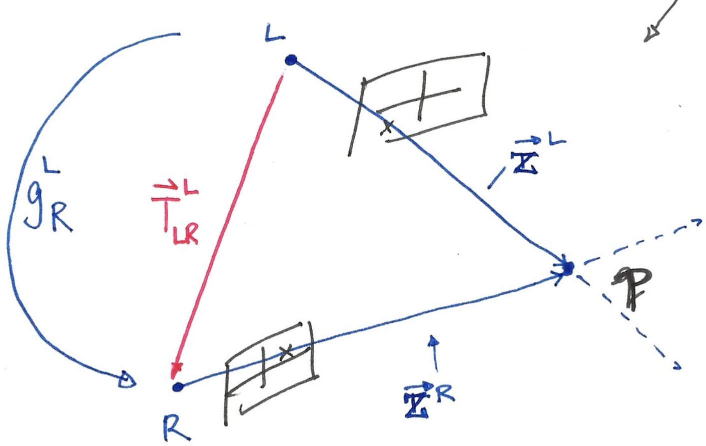
There are some nice geometric relationships that hold when viewing the world through two or more camera locations.

We can't explore it all since that would take a whole semester.

We can explore a little though.

The Essential Matrix

First, let's work out some basic math, then interpret and explain utility. Consider a two-camera setup. In fact let's assume it's a stereo camera for now:



Two cameras: L & R

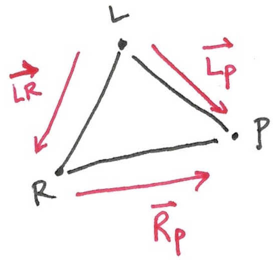
One point p (q in homogeneous form)

lies on two different rays

$\vec{Z}^L$ & $\vec{Z}^R$

$$g_R^L = \left[\begin{array}{c|c} R_R^L & T_{LR}^L \\ \hline 0 & 1 \end{array}\right]$$

The geometry of this setup creates a plane, defined by the L-R-p triangle where L-R-p means origin of L to origin of R to p (back to origin of L)



The vectors associated to the sides of a triangle satisfy a special triple-product equation.

It is

$$\underbrace{(\overrightarrow{LR} \times \overrightarrow{Lp})} \cdot \overrightarrow{Rp} = 0$$

also lies in the plane

cross product creates a vector orthogonal to the plane since $\overrightarrow{LR}$ & $\overrightarrow{Lp}$ lie in the plane.

inner product must vanish.

NOTE: Scale of the vectors doesn't matter. it is the direction of the vector that matters. (so having rays will do)

Now, we need to write the above in our language & in a consistent plane. What variables/quantities are available?

$\vec{Z}^L$ - ray starting at L passing through p, in frame L.
more completely: $\vec{Z}^L_{Lp}$.

$\vec{T}^L_{LR}$ - the vector $\overrightarrow{LR}$, in frame L.

$\vec{Z}^R$ - ray starting at R passing through p, in frame R.
more completely: $\vec{Z}^R_{Rp}$

↳ so, we just need to map $\vec{Z}^R$ to L frame.

Since a ray is a vector (up to scale), we can use homogeneous vector form

$$\begin{bmatrix} \vec{z}^{\,L}_{Rp} \\ \hline 0 \end{bmatrix} = g^L_R \begin{bmatrix} \vec{z}^{\,R}_{Rp} \\ \hline 0 \end{bmatrix} = \left[\begin{array}{c|c} R^L_R & T^L_{LR} \\ \hline 0 & 1 \end{array}\right] \begin{bmatrix} \vec{z}^{\,R}_{Rp} \\ \hline 0 \end{bmatrix}$$

zero because its a vector, not a point.

$$= \begin{bmatrix} R^L_R \, \vec{r}^{\,R}_{Rp} \\ \hline 0 \end{bmatrix}$$

$\Rightarrow$ Now we have

$\vec{z}^{\,L}_{Lp}$, $\vec{T}^{\,L}_{LR}$, & $\vec{z}^{\,L}_{Rp}$

RAY VECTOR RAY/VECTOR

$\Rightarrow$

dot product!

$$(\vec{T}^{\,L}_{LR} \times \vec{z}^{\,L}_{Lp}) \cdot \vec{z}^{\,L}_{Rp} = 0$$

scalar!

$\Rightarrow$ substitute for $\vec{r}^{\,L}_{Rp}$

$$(\vec{T}^{\,L}_{LR} \times \vec{z}^{\,L}_{Lp}) \cdot (R^L_R \, \vec{z}^{\,R}_{Rp}) = 0$$

$\Rightarrow$ use hat operator

$$((\vec{T}^{\,L}_{LR})^{\wedge} \, \vec{z}^{\,L}_{Lp}) \cdot (R^L_R \, \vec{z}^{\,R}_{Rp}) = 0$$

$\Rightarrow$ $\vec{a} \cdot \vec{b} = \vec{a}^T \vec{b}$

$$((\vec{T}^{\,L}_{LR})^{\wedge} \, \vec{z}^{\,L}_{Lp})^T (R^L_R \, \vec{z}^{\,R}_{Rp}) = 0$$

$\Rightarrow$

$$\left(\vec{Z}^{L}_{Lp}\right)^{T} \left(\left(\vec{T}^{L}_{LR}\right)^{\wedge}\right)^{T}\left(R^{L}_{R}\, \vec{Z}^{R}_{Rp}\right) = 0$$

$\Rightarrow$ $(\vec{a}^{\wedge})^{T} = -\vec{a}^{\wedge}$

$$-\left(\vec{Z}^{L}_{Lp}\right)^{T} \left(\vec{T}^{L}_{LR}\right)^{\wedge} R^{L}_{R}\, \vec{Z}^{R}_{Rp} = 0$$

1×3 (after transpose) $\quad 3\times 3 \quad 3\times 3 \quad 3\times 1 \quad = 1\times 1$

sign doesn't matter

$\Rightarrow$

$$\left(\vec{Z}^{L}_{Lp}\right)^{T} \underbrace{\left[\left(\vec{T}^{L}_{LR}\right)^{\wedge} R^{L}_{R}\right]}_{3\times 3} \vec{Z}^{R}_{Rp} = 0$$

$1\times 3 \qquad 3\times 3 \qquad 3\times 1$

So, in the coordinates of the world, the rays connecting the camera origins to the observed point satisfy a special equation that is determined by the relative Rotation and translation of the two cameras (here camera R relative to camera L).

This equation is the epipolar equation (in world units). The matrix in the equation is called the essential matrix and is denoted by E. Here I will write it down with the frames indicated

$$E^{L}_{R} = \left(\vec{T}^{L}_{LR}\right)^{\wedge} R^{L}_{R} \qquad \text{given } g^{L}_{R} = \left[\begin{array}{c|c} R^{L}_{R} & \vec{T}^{L}_{LR} \\ \hline 0 & 1 \end{array}\right]$$

Now, of course when we measure things using a camera, the measurements are in pixels not world units. It's possible to create a similar kind of equation using pixels since the pixel coordinates are related to the world coordinates via the intrinsic camera matrix:

$$\vec{r} \sim \Psi \vec{z}$$

$$\Rightarrow \quad \vec{z} \sim \Psi^{-1} \vec{r}$$

so

$$(\vec{z}^{L})^{T} E^{L}_{R} \vec{z}^{R} = 0$$

can be rewritten

$$(\Psi_L^{-1} \vec{r}^{L})^{T} E^{L}_{R} (\Psi_R^{-1} \vec{r}^{R}) = 0$$

$$\Rightarrow \quad \vec{r}^{T} \Psi_L^{-T} E^{L}_{R} \Psi_R^{-1} \vec{r}^{R} = 0$$

NOTE $(A^{-1})^{T}$ IS ALSO WRITTEN AS A^{-T} FOR BREVITY.

invert then transpose.

$$\Rightarrow \quad \underset{1\times 3}{\vec{r}^{T}} \underbrace{\left[\Psi_L^{-T} E^{L}_{R} \Psi_R^{-1} \right]}_{3\times 3} \underset{3\times 1}{\vec{r}^{R}} = 0$$

Now we have the relationship using pixel values, which is what we directly measure.

The matrix we got in the epipolar equation for pixel coordinates is the fundamental matrix, denoted by F,

$$F^L_R = \Psi_L^{-1} E^L_R \Psi_R^{-1} = \Psi_L^{-1} (\vec{T}^L_{LR})^\wedge R^L_R \Psi_R$$

where I also included the frames for clarity:

$$(\vec{r}^L)^T F^L_R (\vec{r}^R) = 0$$

obtained from pixel coordinates of the point in the left & right cameras.

UTILITY OF THE EPIPOLAR EQUATION.

- If a ray in one of the frames is known, then projecting the ray into the other frame/image gives a line (or better put gives a set of points living on a line). That line is called the epipolar line.

$$\underbrace{(\vec{z}^{\,L}_{Lp})^T}_{\text{pretend known}} \underbrace{E^L_R}_{\text{known since } g^L_R \text{ known}} \vec{z}^{\,R}_{Rp} = 0$$

$\Rightarrow$

$$\underbrace{\left[\underset{1\times 3}{(\vec{z}^{\,L}_{Lp})^T} \underset{3\times 3}{E^L_R} \right]}_{1\times 3} \vec{z}^{\,R}_{Rp} = 0$$

let $\xi = (\vec{z}^{\,L}_{Lp})^T E^L_R = [\xi_1\ \xi_2\ \xi_3]$

let $\vec{z}^{\,R}_{Rp} = \alpha \begin{bmatrix} z^1 \\ z^2 \\ 1 \end{bmatrix}$ ← recall α can be anything. whatever is correct for your problem works.

$\Rightarrow$

$$\xi \vec{z}^{\,R}_{Rp} = [\xi_1\ \xi_2\ \xi_3] \cdot \alpha \begin{bmatrix} z^1 \\ z^2 \\ 1 \end{bmatrix} = 0$$

$\Rightarrow$

$$\underbrace{\xi_1 z^1 + \xi_2 z^2 + \xi_3 = 0}$$

GIVES EQUATION FOR LINE.

(IT'S A FAMILY OF RAYS IN MORE MATHY PARLANCE)

LINE EQUATIONS:
$y = mx + b$
$ax + by + c = 0$

Doing the same with the fundamental matrix gives:

$$(\vec{r}^{\,L}_{Lp})^T F^L_R \, \vec{r}^{\,R}_{Rp} = 0$$

again, suppose that $\vec{r}^{\,L}_{Lp}$ is known

and let $\xi = (\vec{r}^{\,L}_{Lp})^T F^L_R$

$\Rightarrow$

$$\xi \, \vec{r}^{\,R}_{Rp} = 0$$

$\Rightarrow$

$$[\xi_1 \;\; \xi_2 \;\; \xi_3] \begin{bmatrix} r^1 \\ r^2 \\ 1 \end{bmatrix}^R = 0$$

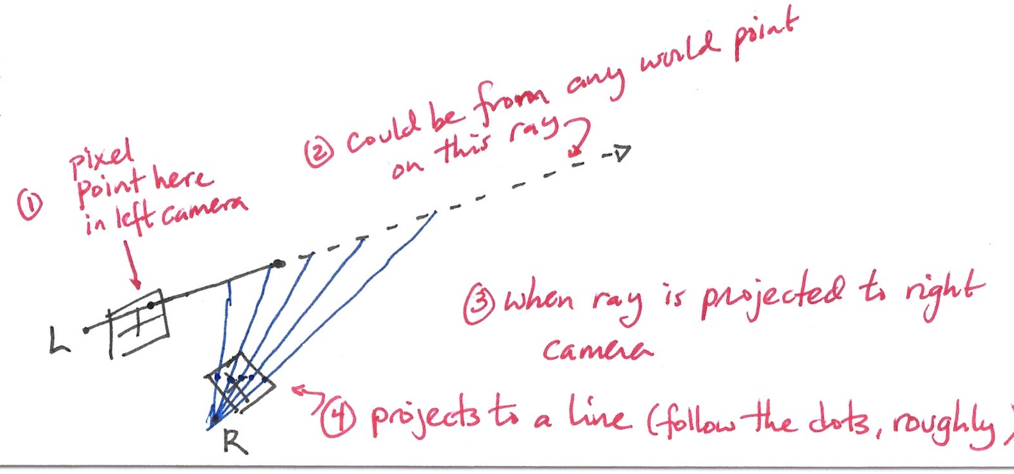
$\Rightarrow$

$$\Big[\underbrace{\xi_1 r^1 + \xi_2 r^2 + \xi_3 = 0}_{\text{Equation for a line in the right image.}} \Big]^R$$

IN R FRAME / IMAGE

The same point p seen in the left image at $\vec{r}^{\,L}$ must be seen in the right image on this line (if it is visible in the right image).

VISUALIZATION:



That line is the epipolar line (of the left image point in the right camera).

* The epipolar line and the fundamental matrix can be used to aid in the search for objects/points of interest across two views.

Suppose a face were found at a specific point in one image (left image). If there were many faces in the right image, the matching face must live on the epipolar line. Any faces far from the line are most likely not the matching face (strictly speaking they aren't, but we should be tolerant to a little error since nothing measured is really perfect).

* The fundamental and essential matrices can be used to estimate the rotation and translation (up to length scale) of a camera given two views and known common points across the views (AKA, point correspondences). To do so, use the same trick as for camera calibration; e.g., pull out E or F as a vector then used SVD to solve followed by some linear algebra to get R & T. T will be off by scale.

VALUE OF ALL THIS STUFF

	COVERED IN CLASS / HW
1) CALIBRATION	✓
2) CAMERA EGO-MOTION (ASSUMING STATIC WORLD)	✗
3) RIGID BODY MOTION (ASSUMING STATIC CAMERA)	✗
4) STEREO CALIBRATION	✓~ (KINDA)
5) STEREO EGO-MOTION	✗
6) STEREO DEPTH OR 3D-POINT RECOVERY.	✓