Notations

Note Title 10/21/2008

Let a line be defined by two 2D pts in homogeneous coordinates. $A = (x, y, t)$ & $B = (u, v, w)$.

$\therefore$ The line passes through $\left(\frac{x}{t}, \frac{y}{t}\right)$ & $\left(\frac{u}{w}, \frac{v}{w}\right)$.

Let m & c be the slope & offset of the eqn. of this line. Hence,

$$\frac{v}{w} = m \cdot \frac{u}{w} + c$$

$$\frac{y}{t} = m \cdot \frac{x}{t} + c$$

$$\therefore \left(\frac{v}{w} - \frac{y}{t}\right) = m\left(\frac{u}{w} - \frac{x}{t}\right)$$

$$\Rightarrow \frac{tv - yw}{wt} = m\left(\frac{tu - xw}{wt}\right)$$

$$\therefore m = \frac{tv - yw}{tu - xw}$$

$$\therefore c = \frac{v}{w} - \frac{tv - yw}{tu - xw} \cdot \frac{u}{w}$$

$$= \frac{\cancel{tuv} - xwv - \cancel{tvu} + ywu}{w(tu - xw)}$$

$$= \frac{\cancel{w}\,(yu - xv)}{\cancel{w}\,(tu - xw)}.$$

Note that any pt. P (3x1 vector) should satisfy the eq.
$$\begin{bmatrix} yw - tv \\ tu - xw \\ xv - yu \end{bmatrix} P = 0$$
to lie on the line. $\therefore$ This matrix signifies the line eqn. Note

$$\begin{bmatrix} yw - tv \\ tu - xw \\ xv - yu \end{bmatrix} = (u, v, w) \times (x, y, t) = B \times A$$

Now
$$\begin{bmatrix} yw - tv \\ tu - xw \\ xv - yu \end{bmatrix} = \begin{pmatrix} 0 & w & -v \\ -w & 0 & u \\ v & -u & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ t \end{pmatrix} = [B]_\times A$$

where B_m is this special matrix formed by the 3D homogeneous coordinates of a 2D pt.

Also
$$\begin{bmatrix} yw - tv \\ tu - xw \\ xv - yu \end{bmatrix} = A^T [B]_\times^T$$

Hence if A is a pt. on a line AB, the line eqn is given by $[B]_x A$ or $A^T [B]_x^T$.

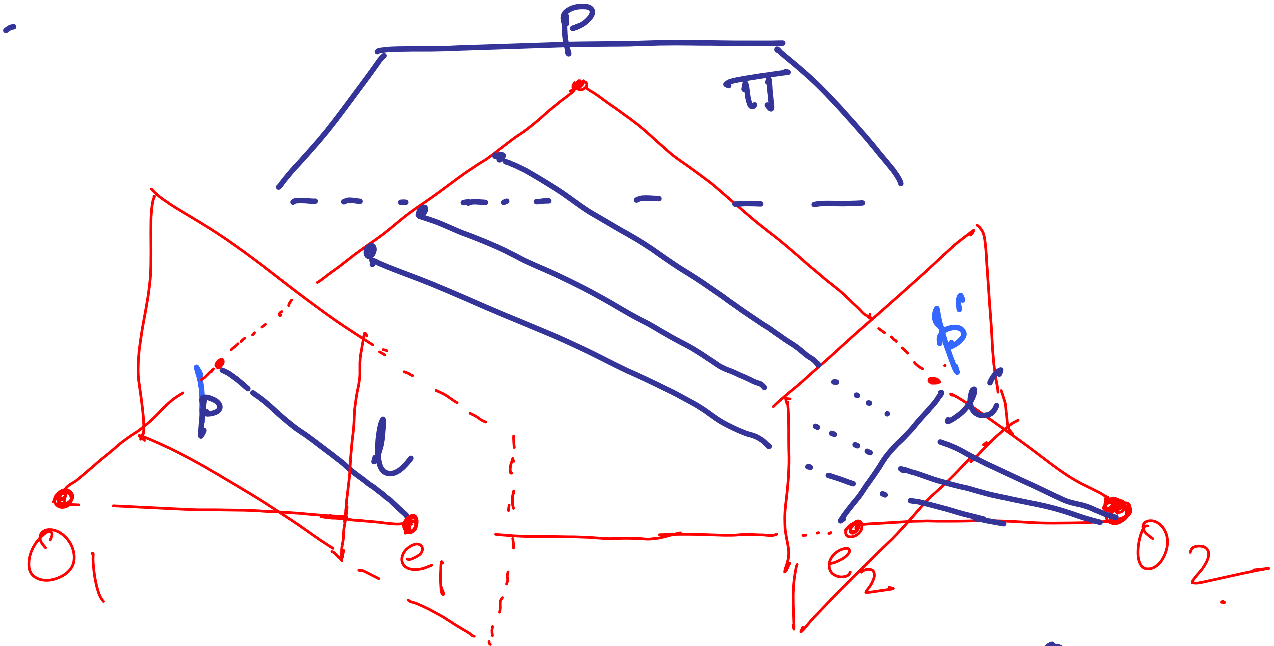
Now if P satisfies this eqn then

$$\underset{1\times 3}{P^T} \left(\underset{3\times 3}{[B]_x} \underset{3\times 1}{A} \right) = \left(\underset{1\times 3}{A^T} \underset{3\times 3}{[B]_x^T} \right) \underset{3\times 1}{P} = 0$$

Note $\det([B]_x) = 0$, all 2x2 submatrices have $\det = 0$. $\therefore$ Rank 2 matrix.

EPIPOLAR GEOMETRY

It deals with several constraints & invariants when considering a pair of camera. This helps in several problem like depth reconstruction & motion estimation.



Consider two cameras. O_1 & O_2 are the COP of the two cameras.
O_1O_2 is often called the baseline, especially considering O_1 & O_2 as a stereo pair.

1. Consider a 3D pt. P. Let p be its image in camera O_1 & p' be its image in O_2.
PO_1O_2 defines a plane. Note that as P changes this plane changes but rotates about O_1O_2. This defines a pencil of planes rooted at O_1O_2.

2. Note that the image of any pt. on the ray O_1P forms a line l' on

c_2 & vice versa (l).

3. The line joining $O_1 O_2$ intersects image plane of O_1 & O_2 at pts. e_1 & e_2 respectively. These are called the epipoles of O_1 & O_2 respectively.

4. The lines $e_1 p = l$, $e_2 p' = l'$ are the epipolar lines. Note that as the plane $P O_1 O_2$ changes, since O_1 & O_2 are fixed, the epipoles don't change, hence all the epipolar lines pass through the epipole of the image.

Why is this important?

Assume calibrated cameras & stereo depth reconstruction. First, it reduces the search space for correspondence. If we detect feature p in c_1, then we need to search for its correspondence on the line $e_2 p'$ instead of the whole image. $\therefore$ Reduces the search space from a 2D plane to 1D line.

FUNDAMENTAL MATRIX

If $P = (x, y, t)$ & $q = (u, v, w)$

$\therefore$ The line l is given by the matrix

$$l = \begin{pmatrix} 0 & w & -v \\ -w & 0 & u \\ v & -u & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ t \end{pmatrix} = Lp$$

Note that l' & l are coplanar. So there exists a 2D affine transformation to map one to another. Let this be denoted by a 3X3 matrix A.

$$l' = A l$$

$$l' = A L \begin{pmatrix} x \\ y \\ t \end{pmatrix}$$

$$= F \begin{pmatrix} x \\ y \\ t \end{pmatrix} = F p$$

Now since $p' = (x', y', t')$ lies on this line we will get

$$p'^T F p = 0$$

$\therefore$ p' satisfies the line eqn.

F is called the fundamental matrix.

$\therefore$ Point p defines a line l' on which p' lies.

Estimating The Fundamental matrix

$$
\begin{bmatrix} x' & y' & 1 \end{bmatrix}
\begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix}
\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0
$$

$$
\therefore\ xx'f_1 + xy'f_2 + xf_3 + yx'f_4 + yy'f_5 + yf_6 + x'f_7 + y'f_8 + f_9 = 0
$$

$\therefore$ With multiple correspondences you can solve for F.

What is the minimum number of points?

$\|f\| = 1$ since correct upto a scale factor. $\therefore$ Need at least eight pts.

Properties

1. F is a rank 2 matrix with 7 degrees of freedom

   2 rotations
   2 translations
   3 params of L.

2. $p'^T F p = 0$

3. $l' = F p$

   $l = F^T p'$

4. $F e_1 = 0$

   $F^T e_2 = 0$

Difference from homography

∴ You plug in a pt & get a corresponding pt. in the second image.

But in this case, the constraint

does not allow you to find the pt. If you are given p, you can find Fp. But this is not p'. p' is a pt. which falls on the line defined by Fp & hence you have to search that line. (Normalized coordinates provides a well-conditioned system).

Say $\underset{3\times3}{F}\,\underset{3\times1}{p}$ is $\begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}$

This is a line whose slope is given by $\frac{f_1}{f_3}$ & offset by $\frac{f_2}{f_3}$.

Let us assume that C_1 & C_2 are the two camera calibration matrix.

$$\therefore \quad C_2 O_1 = e_2$$

Let the 3D point P is

$$p = C_1 P$$

$$\therefore P = C_1^+ p$$ where C_1^+ is a pseudo inverse.

$$\therefore p' = C_2 C_1^+ p$$

$\therefore$ The line $e_2 p'$ is given by $\underset{3\times 1}{C_2 O_1} \times \underset{3\times 1}{C_2 C_1^+ p}$

or $e_2 \times C_2 C_1^+ p$

or $[e_2]_\times C_2 C_1^+ p$

$= Fp$

Now note that

$$F = \underset{3\times 3}{[e_2]_\times}\ \underset{3\times 4}{C_2}\ \underset{4\times 3}{C_1^+}$$

$\therefore C_2 C_1^+$ is a 3×3 matrix. This is exactly the homography via the plane Π defined by P, O_1 & O_2.

rank 2 rank 3

rank 2 $\therefore F = [e_2]_\times H_\Pi$

Let us consider a calibrated stereo rig.

$C_1 = K_1[I|0]$ $\quad C_2 = K_2[R_1|t] = [K_2R_1|K_2t]$

$C_1^+ = \begin{bmatrix} K_1^{-1} \\ 0 \end{bmatrix}$ $\quad C_2 C^+ = K_2 R_1 K_1^{-1}$

Since $O_1 = (0,0,0)$, $\therefore C_2 O_1 = K_2 t$

$\therefore F = [e_2]_\times C_2 C_1^+$

$= [e_2]_\times K_2 R_1 K_1^{-1}$

$= [C_2 O_1]_\times K_2 R_1 K_1^{-1} = [K_2 t]_\times K_2 R_1 K_1^{-1}$

$\therefore$ If given calibrated camera, you can easily find the fundamental matrix.

How does it help?

Say two camera positions of the same camera obtained after a pure translation.

$\therefore C_1 = K[I \mid 0]$ $\qquad \because R_1 = I = R_2$

$C_2 = K[I \mid t]$ $\qquad K_1 = K_2 = K$

$\therefore F = [e_2]_\times K K^{-1}$

$= [e_2]_\times K K^{-1}$

If the t is parallel to x axis, then epipole is on x axis at infinity.

$$\therefore e_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\therefore [e_2]_\times = F = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\therefore [x' \ y' \ 1] \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0$$

or $[0 \ \ 1 \ \ -y'] \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0$

or $y - y' = 0$

$\therefore y = y'$

$\therefore$ Epipolar lines are raster lines (lines parallel to x axis) in each image. $\therefore$ Very easy to find correspondences.

Essential Matrix

If you have a camera whose K is identity. Then

$$C = (R|t)P$$

This is called a normalized camera. This is achieved when the camera coordinates are normalized & hence the name.

Now if we consider two normalized cameras, then their fundamental matrix is called an essential matrix.

$\therefore$ Two normalized cameras satisfy,

$$\hat{p}'^{T} E \hat{p} = 0$$

where $\hat{p}$ indicates normalized camera coordinates, E is essential matrix.

For general cameras, the essential matrix is given by

$E = K_2^T F K_1$ [Proof not included]

Estimating Essential Matrix

Apply the same method as estimating fundamental matrix but using normalized camera coordinates. This assumes that these are normalized cameras.

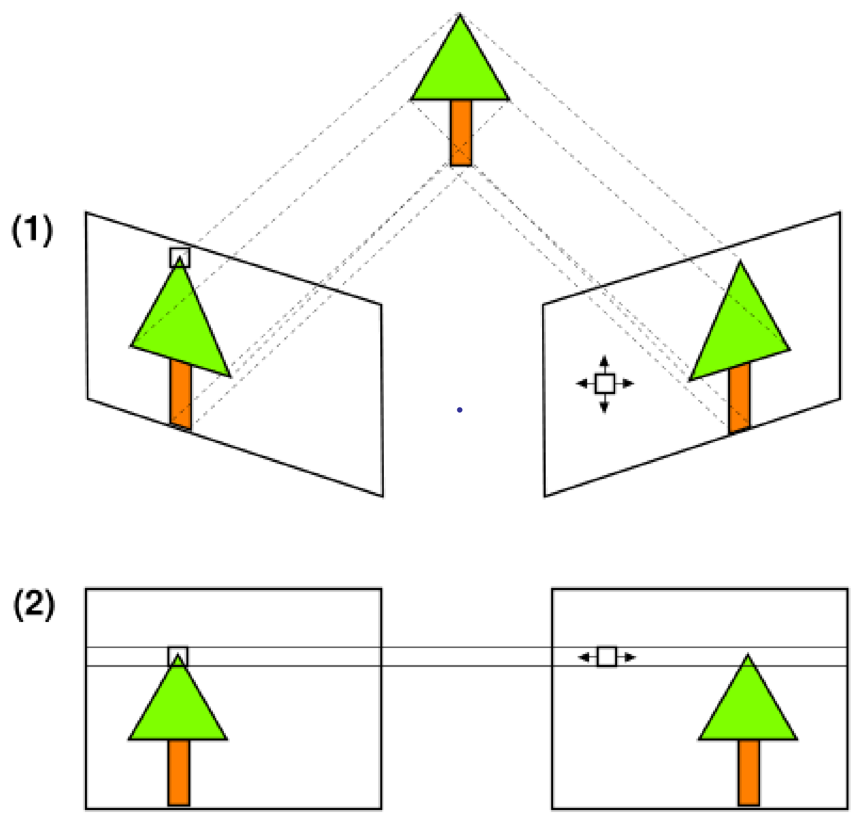
It can be shown that for normalized cameras

$E = R[t]_x$

where R & t are the transformation

required to the image planes of two cameras parallel when $y = y'$ — correspondences lie on raster lines.

So if we can estimate R & t, then we can apply it to an image & get the correspondences easily searchable on raster lines. This is called <u>Rectification.</u>



$\therefore$ How do you estimate R & t?

$E = U \Sigma V^T$ [Proof not included]

U & V are 3×3 orthogonal matrices.

$$\Sigma = \begin{pmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

By internal constraints on E.

Define orthonormal

$$W = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$W^{-1} = W^T = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\therefore [t]_\times = V W \Sigma V^T \quad \text{--- } (1)$$

$$R = U W^{-1} V^T$$

<u>Proof</u>: $R[t]_\times = U W^{-1} V^T V W \Sigma V^T$

$= U W^T I W \Sigma V^T$

$$= U I \Sigma V^T$$
$$= U \Sigma V^T = E$$

Actually 4 sol$\underline{\underline{n}}$

$$R = U W V^T$$
& $$[t]_x = V W^{-1} \Sigma V^T \quad \text{—— } 2$$

will also work.

$$R = U W V$$
$$[t]_x = V^T W^{-1} \Sigma V^T \quad \text{—— } 3$$

or

$$R = U W^{-1} V$$
$$[t]_x = V^T W \Sigma V^T \quad \text{—— } 4$$

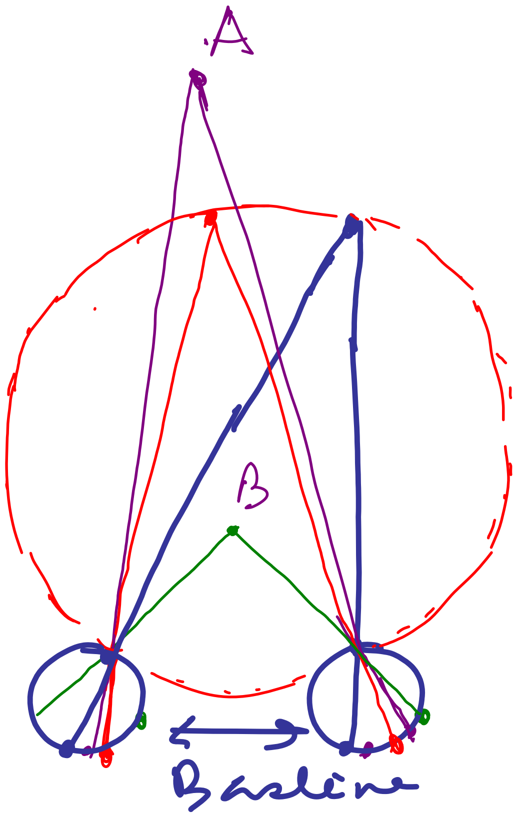
$\therefore$ Four solutions to this problem.

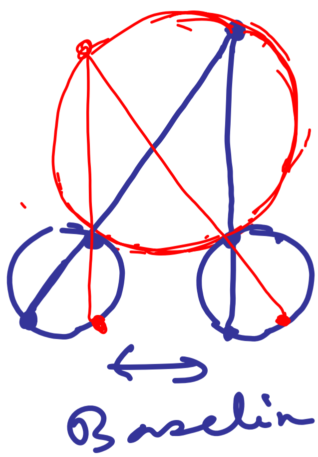
Only one of The 4 results are possible in practice. Other will generate 3D pts. behind 1st, 2nd or both cameras.

Rectification is very common in stereo matching procedure to reduce complexity of correspondence search.

After rectification images are very close to The two eyes. i.e. Two cameras with parallel image plane separated by a translation.

Let us investigate the eye situation.





← location of Horopter depends on the baseline.

Smaller baseline, ∴ radius of horopter is smaller.

$\therefore z_H \propto b$

1. A point inside Horopter moves outward
2. A point outside horopter moves inward.
3. The shift depends on the ratio of the depth of the pt. with horopter

Shift is called Disparity.

$$d = x' - x$$

$$z = \frac{b}{d}$$

Note that this is a relative depth.

Let us assume that same camera moves along the principal axis.

$$\therefore\ C_1 = k[I \mid 0]$$

$$C_2 = k[I \mid t_z]$$ where $t_z = \begin{pmatrix} 0 \\ 0 \\ t \end{pmatrix}$

Let $k = \begin{pmatrix} k_1 & k_2 & k_3 \\ 0 & k_4 & k_5 \\ 0 & 0 & k_6 \end{pmatrix}$

$$\therefore\ p = C_1 \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} k_1 X + k_2 Y + k_3 Z \\ k_4 Y + k_5 Z \\ k_6 Z \end{pmatrix}$$

$$p' = C_2 \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} k_1 X + k_2 Y + k_3 Z + k_3 t \\ k_4 Y + k_5 Z + k_5 t \\ k_6 Z + k_6 t \end{pmatrix}$$

Let $p = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$ & $p' = \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix}$

$$\therefore x = \frac{k_1}{k_6} \cdot \frac{x}{z} + \frac{k_2}{k_6} \frac{y}{z} + \frac{k_3}{k_6}$$

$$y = \frac{k_4}{k_6} \cdot \frac{y}{z} + \frac{k_5}{k_6}$$

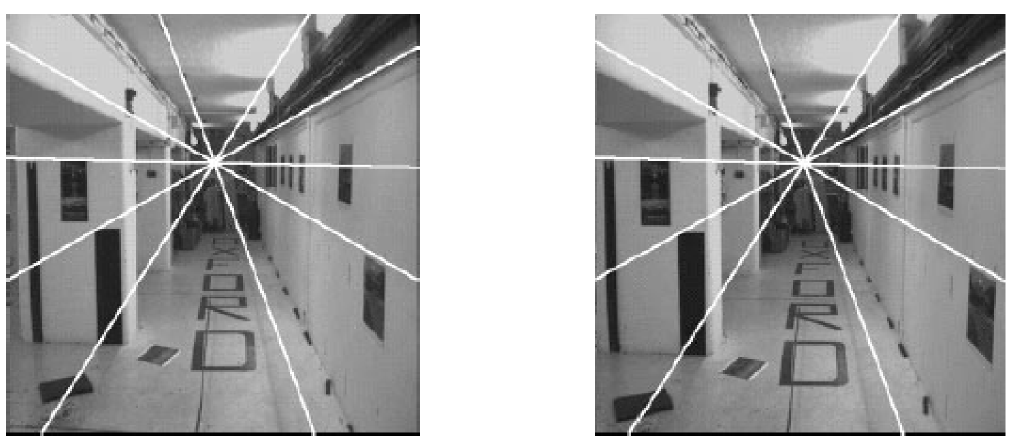
Similarly, you can find (x', y').
Then you will find this eqn:

$$\left.\begin{aligned} x - x' &= \frac{t}{z}\left(x' - \frac{k_3}{k_6}\right) \\ y - y' &= \frac{t}{z}\left(y' - \frac{k_5}{k_6}\right) \end{aligned}\right\}$$

Now note if $z = \infty$ then $p = p'$. This point is called the focus of expansion & it is the $e_1 = e_2$

Second, if z decreases, the movement of the 3D pt. on the camera increases. $\therefore$ More displacement for closer pts than further pt. Also with increase in t, the displacement is more. $\boxed{\text{OPTICAL FLOW}}$

Note (x', y') is related to (x, y) by a linear equation. Hence all corresponding pts will fall on a line.



So, think about the opposite thing, if you have been able to figure out these lines and given the motion, just by analysing shifting of same pts, you can recover depth.

Structure from motion.