Assignment 5 Answers

1. a) In sparse stereo matching we do correspondence with just few points in the image or scene. We try to find the matching points for specific points.

Advantage: 1) Can be used with far views.
2) Efficient and unambigious

disadvantage: 1) may not find correspondence due to obscure or change in intensity.

In dense stereo matching we take many pixels and try to find the correspondence for each of the pixels.

Advantage: 1) Accurate matching.
2)

Disadvantage: 1) images has to be close views.
2) Can be ambiguous.

b) Normalized cross correlation (NCC), we take a window in each view and multiply corresponding elements/pixels in them.

$$\phi(w_1, w_2) = \sum_i \left(\frac{w_1(x_i, y_i) - M_1}{\sigma_1}\right) \cdot \left(\frac{w_2(x_i, y_i) - M_2}{\sigma_2}\right)$$

We first normalize each pixel with the centre of mean & variance in the window to overcome the problem of just considering the value of pixels by the approach of just using correlations. Because, the higher the value of NCC, the higher is the correspondence, just using the pixels values makes pixels with high intensity to be highly correlated.

Sum of Squared Distance (SSD), we take the difference of the corresponding pixels in the two windows in different views and square them.

$$\phi(w_1, w_2) = \sum_i (w_1(x_i, y_i) - w_2(x_i, y_i))^2$$

the lower the value (i.e, closer the 0) the more the correspondence

By allowing the search space to be the entire image, we are prone to mistake of making an incorrect correspondence between points and in turn making wrong reconstruction. And also, we are more likely to find ambiguous problem, where there are more than one suitable match in the other view for a point in current view.

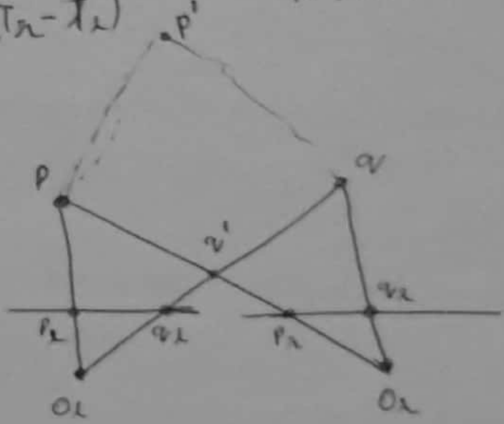
c) $x_l = (100, 200)$
$x_r = (103, 200)$
$f = 10$
$T(\text{baseline}) = 100$

$d = \sqrt{(103-100)^2 + (200-200)^2}$
$= \sqrt{3^2}$
$= 3$

$z = f\frac{T}{d} \Rightarrow z = \frac{10 \times 100}{3} = \frac{1000}{3} = 333.333$

e) Expressions for rotation: $R = R_l^T R_r$, R_r: Rotation in right image
R_l: " left "
R: overall rotation

Expression for translation $T = R_l^T(T_r - T_l)$



d) In this scenario, where there are 2 points p & q and there corresponding points p_l, p_r & q_l, q_r on left & right images.

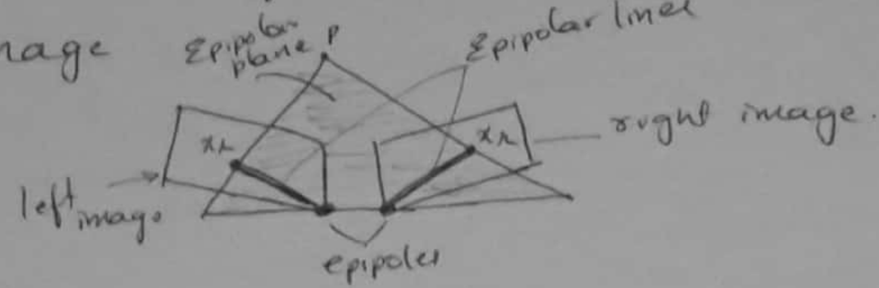
If we wrongly correspond p_l to q_r for p and p_r to q_l for q

We get reconstruction p at p' and q at q', which are totally wrong.

This is the when we choose wrong matches as there are many ambiguous points that match for a single point.

2 a) Epipole is the point where the baseline of the epipolar plane intersects the images.

Epipolar lines are formed when the epipolar plane intersects the image



b) Essential matrix, $E = R^T[T]_x$ where R is rotational matrix and $[T]_x$ is the skew symmetric matrix formed by cross product

$$T = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}, \text{ then } [T]_x = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$

Given points P_l & P_r in Camera coordinates,

Epipolar constraint equation is $P_r^T E P_l = 0$

c) $F = \left(\left(K_r^*\right)^{-1}\right)^T E \left(K_l^*\right)^{-1}$

where E is the Essential matrix, with external camera params and K_r^* & K_l^* are internal camera coordinates of left and right camera.

Epipolar Constraint using F : $P_r^T F P_l = 0$

d) Rank of both Essential matrix, E, and Fundamental matrix, F, is 2. Because, $E = R^T[T]_x$ & $F = K_r^{*-T} E K_l^{*-1}$, since $[T]_x$ has only 2 independent rows, $[T]$ is rank 2 matrix $\Rightarrow$ E & F are rank 2

e) given point P_l in left image and fundamental matrix, F, the Corresponding right epipolar line is given by : FP_l

f) given point P_r in right image and fundamental matrix, F, the corresponding left epipolar line is given by : $F^T P_r$

g) Weak calibration in stereo pair is an approach to find the Fundamental matrix, F, directly by using the 8-point correspondence algorithm and epipolar constraints.

Given $\{P_i'\}_{i=1}^{n} \leftrightarrow \{P_i\}_{i=1}^{n}$ & $n \geq 8$

where P_i' are the points from left image and

P_i are the points from right image.

h) $\begin{bmatrix} x_1 x_1' & x_1 y_1' & x_1 & y_1 x_1' & y_1 y_1' & y_1 & x_1' & y_1' & 1 \\ \vdots \\ \vdots \end{bmatrix} \Rightarrow \begin{bmatrix} 5000 & 1000 & 50 & 10000 & 20000 & 100 & 100 & 200 & 1 \\ \vdots \\ \vdots \end{bmatrix}$

$x_1' = 100 \qquad x_1 = 50$

$y_1' = 200 \qquad y_1 = 100$

i) We normalize the points in 8-point algorithm to the centre of the mean & variance of the points as shown below

Given $\{P_i'\}_{i=1}^{n} \leftrightarrow \{P_i\}_{i=1}^{n}$

We ~~find~~ convert to $\{q_i'\}_{i=1}^{n} \leftrightarrow \{q_i\}_{i=1}^{n}$

where $q_i' = \dfrac{P_i' - M_{p'}}{\sigma_{p'}}$, $M_{p'}$ is the mean of points in left image

& $\sigma_{p'}$ is the variance — " —

& $q_i = \dfrac{P_i - M_P}{\sigma_P}$, M_P is the mean of points in ~~left~~ right image

σ_P — " — variance — " —

Normalization is required for the 8-point algorithm to work well; otherwise the algorithm is not consistent.

Normalization above, can be converted to matrix operations, as,

$$q_i = \underbrace{\begin{bmatrix} 1/\sigma & 0 & 0 \\ 0 & 1/\sigma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -M_x \\ 0 & 1 & -M_y \\ 0 & 0 & 1 \end{bmatrix}}_{M} P_i$$

where $M_x \to$ mean of x coords & $M_y \to$ mean of y coords.

then $q_i = M P_i$ and similarly $q_i' = M' P_i'$

So, the Fundamental matrix (F) for original points can be recovered from the fundamental matrix (F') for normalized points as $\boxed{F = (M')^T F' M}$

3) All epipolar lines must pass through the epipole ~~and the~~
So the epipolar Constraint $P_r^T F P_l = 0$ becomes
$e_r^T F P_l = 0 \; \forall P_l$ where e_r is the right epipole
this is true only when $e_r^T F = 0 \Rightarrow F^T e_r = 0$

Then, we can do Singular Value Decomposition on F^T and ~~the~~ right epipole is the last column ~~(or left null~~ of U
(or left null space of F)

Similarly, left epipole, e_l, is the right null space of F

3 a) Stereo pair can be rectified using below steps:

1. align right ~~left~~ image with left image.
2. align both images with baseline.
3. make the images coplanar.

After rectification, the corresponding points in the left and right images are aligned horizontally in the same line.

b) different approaches for reconstruction are:

1. Absolute reconstruction (or Complete reconstruction) is performed when both intrinsic & extrinsic Camera parameters are known.
2. Euclidean reconstruction, upto unknown scale, is performed when we only know the intrinsic camera parameters.
3. Reconstruction upto unknown 3D projective map is performed when none of the parameters are known.

c) $\underbrace{\begin{bmatrix} P_l & -RP_r & P_l \times RP_r \end{bmatrix}}_{M}$ is the matrix to find coefficients (a, b, c)
and solve $M \begin{bmatrix} a \\ b \\ c \end{bmatrix} = T \Rightarrow \begin{bmatrix} a \\ b \\ c \end{bmatrix} = M^{-1} T$

d) $P = a P_l + \frac{1}{2} c w$ or $P = \frac{1}{2}(a P_l + b R P_r + T)$

where $w = P_l \times R P_r$

P_l : point in left image coordinates

P_r : point in right — " ——

R : Rotation of right image w.r.t left image

T : Translation — " ——

e) In Euclidean reconstruction, since we only know the intrinsic parameters and not the extrinsic parameters, we don't know the baseline (the distance between optical centres) and so, we don't know the scale of the image.

The ~~unkno~~ unknown scale can be removed by

1. Estimate Fundamental matrix, F, by weak calibration ^Using 8-pts algorithm
2. Find E from F as we know intrinsic parameters.
3. Then normalize E to get $\hat{E} = \frac{2}{tr(E^T E)} E$

and then the baseline in $\hat{E}$ has length 1 i.e, $\|T\| = 1$

then find T & R upto unknown scale.

f) Essential matrix, E, can be normalized to have baseline of 1 using the trace of $E^T E$ as $\hat{E} = \frac{2}{tr(E^T E)} E$

g) Unknown signs of rotation ~~of~~ and translation can be determined by extensive search when using Euclidean reconstruction. As possible signs of T, R are (+,+), (+,-), (-,+) & (-,-). We reconstruct using all these 4 options and the choose the reconstruction where all Z coordinates are positive.