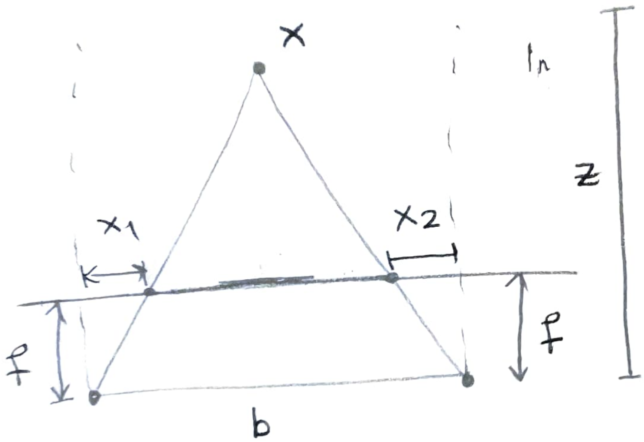
COMP 523 HW 1 BERK COSAR 69557

Pen-and-Paper Questions

1. (2 points) What is the difference between depth and disparity? How are they related to each other mathematically?

⇒ Disparity is the horizontal displacement of a point's projections between the left and the right image

Depth is the horizontal distance to the cameras.



$x_1 > 0$
$x_2 < 0$

In above figure, depth is z, and disparity $d = x_1 - x_2$

Using triangles, we can obtain $\frac{z-f}{b-d} = \frac{z}{b}$

$\Rightarrow zb - bf = zb - zd$

$\Rightarrow zd = bf \Rightarrow \boxed{d = \frac{bf}{z}}$

where $d \to$ disparity

$b \to$ distance between cameras, baseline

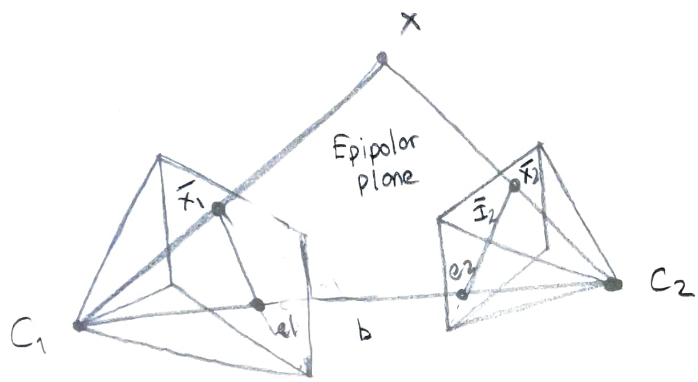
$f \to$ focal length

$z \to$ depth

2. (2 points) What do we assume to know in calibrated two-view geometry?

We assume we know
- relative position of cameras
- their internal parameters

3. (5 points) Define the following terms related to epipolar geometry.



- Epipolar Line: Epipolar lines shown in above figure with $\bar{I}_1$ and $\bar{I}_2$. It's the line connects the projection of 3D point X and the epipole.

- Epipolar Plane: The plane that formed by C_1, C_2 (the cameras) and 3D point X.

- Epipole: Epipole is the points that baseline intersects with image planes.

- Projection and Backprojection: Projection is finding the 2D virtual point corresponds to the 3D point. In the figure projecting means finding $\bar{x}_1$ and $\bar{x}_2$ of X. Backprojection is the reverse of it. Finding X from $\bar{x}_1$ and $\bar{x}_2$.

- Baseline: In figure, it is b. The line between C_1 and C_2, the cameras.

4. (4 points) Derive the matrices $M \in SE(3) \subset \mathbb{R}^{4\times4}$ representing the following transformations:

- Translation by the vector $T \in \mathbb{R}^3$
- Rotation by the rotation matrix $R \in \mathbb{R}^{3\times3}$
- Rotation by R followed by the translation T
- Translation by T followed by the rotation R

<u>Hint:</u> Remember that we can write the transformation matrix M for a given rotation matrix $R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$ and a translation vector $T = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$ as follows: $M = \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$

<u>Translation by the vector T</u>

$M = \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix}$, for just translation by T, we can set R to identity matrix I.

$\Rightarrow M = \begin{pmatrix} I & T \\ 0 & 1 \end{pmatrix}$ where $I_{3\times3}$ identity matrix.

<u>Rotation by rotation matrix R</u>

$M = \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix}$, for just rotation by R, we can set $T = 0$

$\Rightarrow M = \begin{pmatrix} R & 0 \\ 0 & 1 \end{pmatrix}$

<u>Rotation by R followed by translation T</u>

$\begin{pmatrix} I & T \\ 0 & 1 \end{pmatrix}\begin{pmatrix} R & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix}$

$\Rightarrow$ <u>Translation by T followed by R</u>

$\begin{bmatrix} R & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} I & T \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} R & RT \\ 0 & 1 \end{bmatrix}$

5. (5 points) A classic ambiguity of the perspective projection is that one cannot tell an object from another object that is exactly twice as big but twice as far. Explain why this is true.

Hint: Let $P = (X, Y, Z)$ be a point on the smaller object and $P' = (X', Y', Z')$ a point on the larger object. Define $X' = 2X$, $Y' = 2Y$, $Z' = 2Z$ and perspective projection as a function $p = \pi(P)$. How does π transform the world coordinate P to image coordinate p according to perspective projection? Repeat the same for P' and p'

$p = \pi(P)$

$$x_s = f \cdot \frac{x_c}{z_c} \qquad \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \rightarrow \begin{pmatrix} f \cdot X/Z \\ f \cdot Y/Z \\ 1 \end{pmatrix}$$

$$y_s = f \cdot \frac{y_c}{z_c}$$

$p' = \pi(P')$

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} 2X \\ 2Y \\ 2Z \end{pmatrix} \rightarrow \begin{pmatrix} f \cdot 2X/2Z \\ f \cdot 2Y/2Z \\ 1 \end{pmatrix} = \begin{pmatrix} f\, X/Z \\ f\, Y/Z \\ 1 \end{pmatrix}$$

As it seems from above calculations, the perspective projections of P and P' are same. Therefore the statement is true.

6. (5 points)

(a) Given a 3D point $x_w \in \mathbb{R}^3$, such that $x_w = \begin{bmatrix} 3 \\ 2 \\ 6 \end{bmatrix}$, write x_w as an augmented vector $\hat{x}_w$ and as homogenous vector $\tilde{x}_w$

(b) Consider a camera C with pose defined by the rotation and translation matrices, $R = \begin{bmatrix} 0.38 & -0.82 & 0.42 \\ 0.87 & 0.17 & -0.45 \\ 0.29 & 0.54 & 0.78 \end{bmatrix}$ and $t = \begin{bmatrix} 1.3 \\ 2.0 \\ -1.5 \end{bmatrix}$

The camera C has focal lengths $(f_x, f_y) = (785, 786)$, skew $s=0$ and camera center $(c_x, c_y) = (680, 680)$. Project the point x_w to the image plane of camera C.

a/ $\hat{X}_w = \begin{bmatrix} 3 \\ 2 \\ 6 \\ 1 \end{bmatrix}$ we add 1 to make it augmented vector

this augmented vector is homogenous, so

$\tilde{X}_w = \begin{bmatrix} 3 \\ 2 \\ 6 \\ 1 \end{bmatrix}$

b/

$\tilde{x}_s = [K \; 0] \bar{x}_c = [K \; 0] \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \bar{x}_w$

$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$ $\tilde{x}_s \Rightarrow \begin{bmatrix} 785 & 0 & 680 & 0 \\ 0 & 786 & 680 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.38 & -0.82 & 0.42 & 1.3 \\ 0.87 & 0.17 & -0.45 & 2.0 \\ 0.29 & 0.54 & 0.78 & -1.5 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 6 \\ 1 \end{bmatrix}$

$\tilde{x}_s = \begin{bmatrix} 6094.6 \\ 5256.9 \\ 5.13 \end{bmatrix}$ we need to divide z.

$\Rightarrow \begin{bmatrix} 1188 \\ 1024.7 \\ 1 \end{bmatrix}$ so $\begin{pmatrix} 1188 \\ 1024.7 \end{pmatrix}$ is the projection

7) (6 points) We are dealing with a dual-camera setup where the primary camera is positioned to the left of the second camera. The matrices that describe their relative rotation and translation are specified as follows:

- The rotation matrix is identity matrix
- The translation vector is [1, 0, 0]

a) For $x_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ on image plane of right camera, identify the complete set of corresponding points on the image plane of the left camera that comply with Epipolar constraint

b) Determine whether the point $x_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ on the image plane of the right camera and the point $x_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$ on image plane of left camera satisfy the Epipolar constraint

c) Given N cameras with $\Pi = \{\pi_i\}$ their intrinsic and extrinsic parameters, $X_W = \{x_w^p\}$ with $x_w^p \in R^3$ denote a set of P 3D points in world coordinates and $X_S = \{x_i^p\}$ with $x_i^p \in R^2$ denote the image (screen) observations in all i cameras. Give an example of a scenario where the bundle adjustment error is 0?

Hint: In an ideal setting with exact camera parameters and no distortions, observed image points would perfectly align with projections from 3D points, resulting 0 bundle adjustment error.

a/ Epipolar constraint $\Rightarrow \tilde{x}_2^T \tilde{E} \tilde{x}_1 = 0$

$\tilde{E} = [t]_\times R$

$$[t]_\times = \begin{bmatrix} 0 & -tz & ty \\ tz & 0 & -tx \\ -ty & tx & 0 \end{bmatrix}$$

so $\tilde{E} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$

$\tilde{x}_2^T \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = 0 \Rightarrow x_2^T \begin{bmatrix} 0 \\ -1 \\ 3 \end{bmatrix} = 0$

$[a\ b\ c] \begin{bmatrix} 0 \\ -1 \\ 3 \end{bmatrix} = 0$

$-b + 3c = 0 \quad x_2 = \begin{bmatrix} a \\ 3k \\ k \end{bmatrix} = \begin{bmatrix} a/k \\ 3 \\ 1 \end{bmatrix} \rightarrow \begin{bmatrix} a/k \\ 3 \end{bmatrix}$ so $y = 3$

b/

we calculated $\tilde{E} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$

$\tilde{x}_2^T E \tilde{x}_1 = 0$? for $x_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ & $x_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$$\begin{bmatrix} 4 & 4 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 4 & 4 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \\ 3 \end{bmatrix} = 0 - 4 + 3 = -1 \neq 0$$

so does not satisfy epipolar constraint. Also we know from part a, it should be in form $\begin{pmatrix} x \\ 3 \end{pmatrix}$. y coordinate should be 3, to satisfy epipolar constraint.

---

c/

To get 0 bundle adjustment error, we should know camera intrinsics precisely, there should be no measurement noise, the scene geometry, camera positions, and orientations should be known perfectly without any ambiguity. There should no distortions. Computations should be in infinite precision.

8. (6 points) In a stereo camera setup, we aim to determine the disparity caused by variations in brightness between two cameras set at different exposure levels. We have a sequence of pixel intensity values from a single row in both the left and right camera images as follows:

red

19	0	12	2	13	22	20	26	left

11	19	18	23	18	18	25	24	right

consider red pixel, whose true disparity is 4. We would like to estimate the disparity of this pixel using block matching with a window size of 3 (1 x 3 as we consider 1D pixel row).

a) Is the true disparity recovered if we use the Sum of squared differences (SSD) similarity metric?

b) Which similarity metric would you use instead considering the brigthness changes between the left & right images? Show that the proposed metric can recover the true disparity for the pixel in red?

a/ the window spans 22 20 26 on left camera.

disp 5: 22 20 26 / 11 19 18 → $SSD = (22-11)^2 + (20-19)^2 + (26-18)^2 = 186$

disp 4: 22 20 26 / 19 18 23 → $SSD = (22-19)^2 + (20-18)^2 + (26-23)^2 = 22$

disp 3: 22 20 26 / 18 23 18 → $SSD = (22-18)^2 + (20-23)^2 + (26-18)^2 = 89$

disp 2: 22 20 26 / 23 18 18 → $SSD = (22-23)^2 + (20-18)^2 + (26-18)^2 = 69$

disp 1: 22 20 26 / 18 18 25 → $SSD = (22-18)^2 + (20-18)^2 + (26-25)^2 = 21$ → smallest disp = 1

disp 0: 22 20 26 / 18 25 24 → $SSD = (22-18)^2 + (20-25)^2 + (26-24)^2 = 45$

so disparity is 1 estimated did not recovered.

b/ lets use ZNCC as another metric.

22 20 26 $\rightarrow$ mean $= \frac{22+20+26}{3} = 22.67$

11 19 18 $\rightarrow$ mean $= 16$ $\Rightarrow$ ZNCC $= \frac{(22-22.67)(11-16)+(20-22.67)(19-16)+(26-22.67)(18-16)}{\sqrt{0.67^2+2.67^2+3.33^2}\sqrt{5^2+3^2+2^2}}$

$= 0.08$

19 18 23 $\rightarrow$ $m = 20$ $\Rightarrow$ ZNCC $= \frac{(22-22.67)(19-20)+(20-22.67)(18-20)+(26-22.67)(23-20)}{\sqrt{0.67^2+2.67^2.333^2}\sqrt{1^2+2^2+3^2}} = 0.99$

18 23 18 $\rightarrow$ $m = 19.7$ $\Rightarrow$ ZNCC $= \frac{(-0.67)(-1.7)+(-2.67)(3.3)+(3.3)(-1.7)}{\sqrt{0.67^2+2.67^2+3.33^2}\sqrt{1.7^2+3.3^2+1.7^2}} = -0.75$

23 18 18 $\rightarrow$ $m = 19.7$ $\Rightarrow$ ZNCC $= \frac{(-0.67)(3.3)+(-2.67)(-1.7)+(3.3)(-1.7)}{\sqrt{0.67^2+2.67^2+3.33^2}\sqrt{3.3^2+1.7^2+1.7^2}} = -0.1$

18 18 25 $\rightarrow$ $m = 20.3$ $\Rightarrow$ ZNCC $= \frac{(-0.67)(-2.3)+(-2.67)(-2.3)+(3.3)(4.7)}{\sqrt{0.67^2+2.67^2+3.33^2}\sqrt{2.3^2+2.3^2+4.7^2}} = 0.94$

18 25 24 $\rightarrow$ $m = 22.3$ $\Rightarrow$ ZNCC $= \frac{(-0.67)(-4.3)+(-2.67)(2.7)+(3.3)(1.7)}{\sqrt{0.67^2+2.67^2+3.33^2}\sqrt{4.3^2+2.7^2+1.7^2}} = 0.06$

so with ZNCC, estimated disparity is 4.

disparity is recovered.

9. (5 points) Given the focal length f and baseline b for the left camera of a stereo camera setup and the disparity d, the depth can be calculated using the simple formula, $z = \frac{fb}{d}$. Show that for an error of ε_d in the estimated disparity, the corresponding error in the obtained depth ε_z is quadratic in depth,

$$\varepsilon_z = \frac{z^2}{bf}\varepsilon_d.$$

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

let's say our estimated disparity is $\hat{d} = d + \varepsilon_d$, so $\hat{z} = \frac{fb}{\hat{d}}$.

We can define ε_z as $|\hat{z} - z|$, so

$$\varepsilon_z = \frac{fb}{d} - \frac{fb}{\hat{d}} = fb\left(\frac{1}{d} - \frac{1}{\hat{d}}\right) = fb\left(\frac{\hat{d} - d}{d\hat{d}}\right)$$

we know $\hat{d} - d = \varepsilon_d$, so $\varepsilon_z = fb\frac{\varepsilon_d}{d\hat{d}}$

we know $\hat{d} = \varepsilon_d + d$, so $\varepsilon_z = \frac{fb\varepsilon_d}{d(\varepsilon_d + d)}$.

Now, we can say $\varepsilon_d + d \approx d$ because ε_d is small.

$$\Rightarrow \varepsilon_z = \frac{fb\varepsilon_d}{d^2} = \frac{fb\varepsilon_d}{(fb)^2/z^2} = \frac{z^2\varepsilon_d}{fb} \Rightarrow \boxed{\varepsilon_z = \frac{z^2}{bf}\varepsilon_d}$$

Therefore ε_z is quadratic in depth.