Bean Duran
00319930

Geiger CV HW02

1.1 Review on Monday

a) 2 cameras w/ intrinsic parameters & rotations: $K = R = I$

Given the following translation vectors:

$$t_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},\ t_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},\ t_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Essential matrix: skew-symmetric matrix

$$\hat{E} = [\hat{t}]_\times R \quad \text{with} \quad [\hat{t}]_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

$I = R$ is the 3x3 identity matrix

$$t_1 = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix},\ \hat{E} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Essential Matrices $\Rightarrow$

$$\boxed{\hat{E}_{t_1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix},\ \hat{E}_{t_2} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix},\ \hat{E}_{t_3} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}$$

Orientation of epipolar lines:

$$\hat{l}_{t_2} = \hat{E}\hat{x}_1,\ \hat{l}_1 = \hat{E}^T\hat{x}_2 \quad \text{with} \quad \hat{x}_1 = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$\Rightarrow$ Calculate $\hat{l}_{t_2}, \hat{l}_1$ for all t:

$$t_1:\ \hat{l}_{t_2} = \hat{E}_{t_1}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} -0 \\ 0 \\ y \end{pmatrix},\ \hat{l}_1 = \hat{E}_{t_1}^T\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ y \end{pmatrix} \Rightarrow \boxed{\begin{aligned} \hat{l}_{t_2} &= (0, -1, y)^T \\ \hat{l}_1 &= (0, 1, -y)^T \end{aligned}}$$

$$t_2:\ \hat{l}_{t_2} = \hat{E}_{t_2}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ x \end{pmatrix},\ \hat{l}_1 = \hat{E}_{t_2}^T\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ x \end{pmatrix} \Rightarrow \boxed{\begin{aligned} \hat{l}_{t_2} &= (1, 0, -x)^T \\ \hat{l}_1 &= (-1, 0, x)^T \end{aligned}}$$

$$t_3:\ \hat{l}_{t_2} = \hat{E}_{t_3}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} -y \\ x \\ 0 \end{pmatrix},\ \hat{l}_1 = \hat{E}^T\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} -y \\ x \\ 0 \end{pmatrix} \Rightarrow \boxed{\begin{aligned} \hat{l}_{t_2} &= (-y, x, 0)^T \\ \hat{l}_1 &= (y, -x, 0)^T \end{aligned}}$$

Epipolar lines for t_1, t_2, t_3

How do I interpret these coordinate triples?

Location of epipoles:

$\hat{e} = \hat{l}_1 \times \hat{l}_2$ and since $R = I$ we know that both cameras are facing the same direction. Thus the epipolar lines are parallel for t_1 and t_2 since they only translate x and y. Therefore $\hat{e}_{t_1}$ and $\hat{e}_{t_2}$ lie at infinity.

However, for $\hat{e}_{t_3}$: $\hat{l}_1 \times \hat{l}_2 = (0, 0, x_1y_2 - x_2y_1)^T \Rightarrow x$ and y are always 0, z is a scalar which we can set to: $\underline{\hat{e}_{t_3} = (0, 0, 1)^T} \Rightarrow$ In the image plane, the epipole is at $(0,0)$ i.e. the image center.

b) The baseline connects the 2 camera centers

$R = I \Rightarrow$ both cameras' optical axes are aligned with the z-axis

$\Rightarrow$

c)

1.2

a) Consider a 2-camera system.

$$K_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad K_2 = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad R_2 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$t_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \qquad t_2 = \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix}$$

Initial notes: $K_i = R_i = I$, t_i has no effect

Observations:

$$\tilde{x}_1' = \begin{pmatrix} 1/4 \\ 1 \end{pmatrix},\ \tilde{x}_2' = \begin{pmatrix} -1/3 \\ 1/3 \\ 1 \end{pmatrix}$$

Triangulate would send $\tilde{x}_w$ corresponding to these observations:

First compute $P = K[R|t]$ for both cameras:

$$P_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},\ P_2 = \begin{bmatrix} -2 & 0 & 1 & -3 \\ 0 & -2 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

From Lec 03 slide 31:

p_1^k are the k-th row of the i-th camera's projection matrix

$$\tilde{x}_i' \times \tilde{P}_i \tilde{x}_w = 0 \Rightarrow \underbrace{\begin{bmatrix} x_i'p_3^T - p_1^T \\ y_i'p_3^T - p_2^T \end{bmatrix}}_{A} \tilde{x}_w = 0$$

$$A = \begin{bmatrix} -1 & 0 & 1/4 & 0 \\ 0 & -1 & 1/4 & 0 \\ 2 & 0 & -2/3 & 1/3 \\ 0 & 2 & -1/3 & -1/3 \end{bmatrix} \Rightarrow \text{Now solve } A\tilde{x}_w = 0 \Rightarrow \text{Use SVD: } A = U\Sigma V^T$$

$\tilde{x}_w$ is the last column of V

Found V last column w/ NumPy

$$\boxed{\tilde{x}_w = \begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix}}$$

1.3

a) Assuming disparity $d = x_1 - x_2$:

triangulated depth, focal length, baseline, disparity

$$\frac{z - f}{b - d} = \frac{z}{b} \Rightarrow zb - fb = zb - zd \Rightarrow z = \frac{fb}{d}$$

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

Deriving depth measurement error Δz:

assume f and b are accurate

$$\Delta z = \frac{fb}{d + \Delta d} \Rightarrow z_{est} - z_{true} = \frac{fb}{d + \Delta d} - \frac{fb}{d}$$

disparity error

$$= fb\left(\frac{1}{d + \Delta d} - \frac{1}{d}\right)$$

$$= fb\left(\frac{d}{d(d + \Delta d)} - \frac{d + \Delta d}{d(d + \Delta d)}\right)$$

$$= fb\left(-\frac{\Delta d}{d(d + \Delta d)}\right)$$

$$\Delta z = -fb\frac{\Delta d}{d^2 + d\Delta d}$$

Substitute d in terms of z:

$$\Delta z = -fb\frac{\Delta d}{\left(\frac{fb}{z}\right)^2 + \left(\frac{fb}{z}\right)\Delta d} \cdot \frac{z^2}{z^2}$$

$$\Delta z = -fb\frac{z^2\Delta d}{(fb)^2 + fb\Delta d z} \Rightarrow \Delta z = -\frac{z^2\Delta d}{fb + \Delta d z}$$

And if we assume Δd is negligible

$$\Rightarrow \boxed{\Delta z = -\frac{z^2}{fb}\Delta d}$$

Depth measurement error Δz grows quadratically w/ depth z

b) Depth resolution $\Leftrightarrow$ precision of depth measurement

We desired $\Delta z = -\frac{z^2}{fb}\Delta d$ so our depth error would be minimized by maximizing focal length f and baseline b

However, this would have significant consequences:

- Increasing f reduces the camera FOV (physical constraint)
- Increasing b means the cameras' centers move further apart which then increases disparity

Therefore, maximizing f and b results in a memory, disparity-sensitive setup in exchange for better depth resolution.

(i.e. nearsighted vs farsighted)

1.4

a) Consider two $K \times K$ windows of pixels flattened across vectors $w_1, w_2 \in \mathbb{R}^{K^2}$

Changes of brightness: $w_1' = \alpha_1 w_1 + \beta_1$, where $1 \in \mathbb{R}^{K^2}$ is a vector of all 1's

Show that ZNCC is invariant to changes in brightness in these windows:

$$ZNCC(x_1, y_1, d) = \frac{(w_L(x_1,y) - \bar{w}_L(x_1,y))^T (w_R(x - d_1, y) - \bar{w}_R(x - d_1, y))}{\|w_L(x_1,y) - \bar{w}_L(x_1,y)\|_2 \|w_R(x - d_1, y) - \bar{w}_R(x - d_1, y)\|_2}$$

where $\bar{w}$ denotes the mean of the vector w

$$w_1' = \alpha_1 w_1 + \beta 1_1 \Rightarrow w_1' = \alpha_1 w_1 + \beta_1$$

Mean of transformed window:

$$\bar{w}' = \frac{1}{K^2} 1^T w' = \frac{1}{K^2} 1^T (\alpha w + \beta 1) = \alpha\bar{w} + \beta \Leftrightarrow \text{The mean brightness is scaled by } \alpha \text{ and shifted by } \beta$$

$\Rightarrow$ When substituting back into ZNCC:

$$\left.\begin{aligned} w' - \bar{w}'1 &= (\alpha w + \beta 1) - (\alpha\bar{w} + \beta)1 \\ &= (\alpha w + \beta 1) - \alpha\bar{w}1 - \beta 1 \\ &= \alpha w - \alpha\bar{w}1 \end{aligned}\right\} \boxed{\beta \text{ term cancel out}} \text{ \& this holds when we substitute for both vectors.}$$

b) $SAD(x, y, d) = |w_L(x,y) - w_R(x - d, y)|$

5x7 Stereo Images:

	1	2	3	10	5	6	7
	1	2	3	10	5	6	7
	1	2	3	10	5	6	7
	1	2	3	10	5	6	7
	1	2	3	10	5	6	7

L

	2	10	4	5	6	7	8
	2	10	4	5	6	7	8
	2	10	4	5	6	7	8
	2	10	4	5	6	7	8
	2	10	4	5	6	7	8

R

occlusion

SAD w/ 3x3 window determine WTA disparity $d \in \{0, 1, 2\}$ for the cyan pixel:

10	5	6
10	5	6
10	5	6

L

$\Downarrow$

10	4	5	6	7
10	4	5	6	7
10	4	5	6	7

R

$$SAD(x, y, d) = \sum_{i,j \in 3\times 3} |w_L(x+i, y+j) - w_R(x+i-d, y+j)|$$

$$d = 0 \Rightarrow SAD(3,3,0) = 3 \cdot (\overset{5}{|10-2|} + \overset{1}{|5-4|} + \overset{1}{|6-5|}) = 3 \cdot 7 = \underline{21}$$

$$d = 1 \Rightarrow SAD(3,3,1) = 3 \cdot (\overset{6}{|10-10|} + \overset{0}{|5-4|} + \overset{0}{|6-5|}) = 3 \cdot 6 = \underline{18}$$

$$d = 2 \Rightarrow SAD(3,3,2) = 3 \cdot (\overset{8}{|10-10|} + \overset{1}{|5-4|} + \overset{1}{|6-5|}) = 3 \cdot 2 = \underline{6}$$

According to SAD, $\boxed{d = 2}$ is the optimal disparity

However, looking at the 2 images we can see that the background behind the thin occluding object only moved 1 pixel $\Rightarrow$ "fronto-parallel" assumption violated

c) Disparity maps for previous question

Check answers

0	0	2	2	2	1	0
0	1	2	2	2	0	0
0	1	2	0	2	1	0
0	1	2	2	2	1	0
0	1	0	2	2	1	0

L → R

2	2	2	2	1	0	0
2	2	2	2	1	0	0
2	2	2	2	1	0	0
2	2	2	2	1	0	0
2	2	2	2	1	0	0

R → L

$x' = x - d_{LR}(x, y)$, $y' = y$

$d_{LR}(x, y) = d_{RL}(x', y') \Rightarrow$ Pass

$d_{LR}(x, y) \neq d_{RL}(x', y') \Rightarrow$ Fail

L–R Consistency test for the colored pixels:

Orange: $d_{LR}(2,1) = 1$, $d_{RL}(\overset{0}{2-1}, 1) = 2 \Rightarrow$ Fail

Red: $d_{LR}(2,5) = 2$, $d_{RL}(\overset{0}{2-2}, 5) = 0 \Rightarrow$ Fail

Green: $d_{LR}(3,3) = 2$, $d_{RL}(\overset{0}{3-2}, 3) = 2 \Rightarrow$ Pass

Cyan: $d_{LR}(6,2) = 1$, $d_{RL}(\overset{0}{5-3}, 2) = 1 \Rightarrow$ Pass

Passed

1.5

a)

Layer	Input Shape	Output Shape	# Trainable Parameters	Memory (MB)
CONV2D(6,32,3)				