1 Pen and Paper

1.1 Epipolar Geometry

a) $\hat{E}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$

$\hat{E}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}$ $\quad \hat{E} = [t]_\times R = [t]_\times$

$\hat{E}_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

epipolar lines:

对于$\hat{E}_1$:

$\tilde{l}_2 = \hat{E}_1 \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ y \end{bmatrix}$ $\quad \tilde{l}_1 = \hat{E}_1^T \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -y \end{bmatrix}$

对于$\hat{E}_2$:

$\tilde{l}_2 = \hat{E}_2 \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -x \end{bmatrix}$ $\quad \tilde{l}_1 = \hat{E}_2^T \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ x \end{bmatrix}$

对于$\hat{E}_3$:

$\tilde{l}_2 = \hat{E}_3 \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} -y \\ x \\ 0 \end{bmatrix}$ $\quad \tilde{l}_1 = \hat{E}_3^T \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} y \\ -x \\ 0 \end{bmatrix}$

对于$\hat{E}_1$,对极线是水平的。

对于$\hat{E}_2$,对极线是垂直的。

对于$\hat{E}_3$,对极线是所有归一化平面上经过(0,0,1)的直线

因为对极点是另一个相机光心在本照相机的投影点,所以直接对t投影

$t_1 = (1,0,0)^T$,投影点 x轴无穷远点

$t_2 = (0,1,0)^T$, ... y轴 ...

$t_3 = (0,0,1)^T \Rightarrow$ 对极点就在原点

b) baseline 位于两个相机的光轴,

对极点位于 principal point

c) $F = K_2^{-T} \hat{E} K_1^{-1} = \hat{E}$

$\Rightarrow K_1 = K_2 = I \Rightarrow \begin{cases} f_x = 0 \\ f_y = 0 \\ s = 0 \end{cases}$ 且 $\begin{cases} c_x = 0 \\ c_y = 0 \end{cases}$ (中心在主点)

1.2 Triangulation

a) $P_1 = K_1[R_1|t_1] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

$P_2 = K_2[R_2|t_2] = \begin{bmatrix} -2 & 0 & 1 & -3 \\ 0 & -2 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$

考虑 $\bar{x}_i^s = \tilde{P}_i \tilde{x}_s^w \Rightarrow \bar{x}_i^s \times \tilde{P}_i \tilde{x}_s^w = 0$

令 $\tilde{p}_{ik}^T$ 表示 $\tilde{P}_i$ 的第k行,有:

$\begin{bmatrix} 0 & -1 & y_i^s \\ 1 & 0 & -x_i^s \\ -y_i^s & x_i^s & 0 \end{bmatrix} \begin{bmatrix} \tilde{p}_{i1}^T \\ \tilde{p}_{i2}^T \\ \tilde{p}_{i3}^T \end{bmatrix} \tilde{x}_s^w = 0$

即: $\underbrace{\begin{bmatrix} -\tilde{p}_{i2}^T + y_i^s \tilde{p}_{i3}^T \\ \tilde{p}_{i1}^T - \hat{x}_i^s \tilde{p}_{i3}^T \end{bmatrix}}_{A_i} \tilde{x}_s^w = 0$

可以写出A矩阵如下式:

$$A = \begin{bmatrix} 0 & -1 & 0.5 & 0 \\ 1 & 0 & -0.25 & 0 \\ 0 & 2 & -0.8 & -0.8 \\ -2 & 0 & 1.2 & -2.8 \end{bmatrix}$$

求解 $A\tilde{x}_s^w = 0$, SVD分解 or 直接算

$\begin{cases} x = 1 \\ y = 2 \\ z = 4 \end{cases} \Rightarrow \tilde{x}^w = \begin{pmatrix} 1 \\ 2 \\ 4 \\ 1 \end{pmatrix}$

1.3 Stereo Vision

a) $z = \frac{bf}{d} \qquad \frac{dz}{dd} = -\frac{bf}{d^2}$

代入 $d = \frac{bf}{z} \Rightarrow \frac{dz}{dd} = -\frac{z^2}{bf}$

$|\Delta z| \approx \frac{z^2}{bf} |\Delta d|$,假设视差误差为常数,那么 $\Delta z \propto z^2$

b) depth resolution:

$$|\Delta Z| \approx \frac{Z^2}{bf}|\Delta d|$$

让 $|\Delta Z|$ ↓, 即 b或f 增大

① b 增大

半遮挡区域增大,

特征匹配困难

② f 增大

视场角 fov 变小,

景深变浅

1.4 Block Matching

a) before: w_1 w_2 $\bar{w}_1$ $\bar{w}_2$

after: $\alpha_1 w_1+1\beta_1$ $\alpha_2 w_2+1\beta_2$ $\alpha_1\bar{w}_1+1\beta_1$ $\alpha_2\bar{w}_2+1\beta_2$

$$ZNCC_{before} = \frac{(w_1-\bar{w}_1)^T(w_2-\bar{w}_2)}{\|w_1-\bar{w}_1\|_2\,\|w_2-\bar{w}_2\|_2}$$

$$ZNCC_{after} = \frac{[(\alpha_1 w_1+1\beta_1)-(\alpha_1\bar{w}_1+1\beta_1)]^T[(\alpha_2 w_2+1\beta_2)-(\alpha_2\bar{w}_2+1\beta_2)]}{\|\alpha_1+w_1+1\beta_1-(\alpha_1\bar{w}_1+1\beta_1)\|_2\,\|\alpha_2 w_2+1\beta_2-(\alpha_2\bar{w}_2+1\beta_2)\|_2}$$

$$= \frac{\alpha_1(w_1-\bar{w}_1)^T\alpha_2(w_2-\bar{w}_2)}{|\alpha_1|\,\|w_1-\bar{w}_1\|_2\,|\alpha_2|\,\|w_2-\bar{w}_2\|_2}$$

$$= ZNCC_{before}$$

b) $SSD(x,y,0) = \|w_L(x,y)-w_R(x,y)\|_2^2$
$= 81$

$SSD(x,y,1) = \|w_L(x,y)-w_R(x-1,y)\|_2^2$
$= 108$

$SSD(x,y,2) = \|w_L(x,y)-w_R(x-2,y)\|_2^2$
$= 6$

基于WTA, 选SSD最小的情况, 即 $d=2$.

续b) 图中像素为5的点的真实视差应该是1, 但SSD算出来是2, 原因是发生的前景膨胀, 要算的是背景视差, 但前景的纹理特征强烈的多, 导致算出前景误差。

c) 一致性: $d_L(x,y) == d_R(x-d_L, y)$

Cyan: $d_L(x,y)=1$ $d_R(x-1,y)=1$ ✓

Green: $d_L(x,y)=2$ $d_R(x-2,y)=2$ ✓

Red: $d_L(x,y)=2$ $d_R(x-2,y)=2$ ✓

其中红色虽然通过, 但它的真实视差是1, 这是因为 Border Bleeding, 物体边缘处前景视差溢出到背景区域了

1.5 Learned stereo and end-to-end models

a) Layer IS OS

Conv2D(32,64,3) (32,128,128) (64,128,128)

#Trainable Parameters $= 32\times64\times3\times3 = 18432$

Memory $= (18432+64\times128\times128)$ MB/par $= 1067008$ M/par

≈ 4.07 MB

(假设每个参数为Float 32, 占用4个Byte.

conv2D(?,128,3) → (~~3~~64,128,128) → (128,128,128)

#Trainable Par $= 64\times128\times3\times3 = 73728$

Memory $= (73728+128\times128\times128)\times4$ Byte

≈ 8.28MB

conv3D(1,64,3) → (1,32,128,128) → (64,32,128,128)

#Trainable Par $= 1\times64\times3\times3\times3 = 1728$

Memory $= (1728+64\times32\times128\times128)\times4$Byte

≈ 128MB

conv3D(?,128,3) → (~~128~~ 64,32,128,128) → (128,32,128,128)

#Trainable Par $= 64\times128\times3\times3\times3 = 221184$

Memory $= (221184+128\times32\times128\times128)\times4$ Byte

≈ 256MB

b) for p_1:

$$\sigma(-C_\theta(d)) = [0.44, 0.06, 0.00, 0.06, 0.44]$$

$$E[d] = \sum_{d=0}^{4} \sigma(-C_\theta) \cdot d = 2$$

for p_2:

$$\sigma(-C_\theta) = [0.00, 0.21, 0.58, 0.21, 0.00]$$

$$E[d] = \sum_{d=0}^{4} \sigma(-C_\theta) \cdot d = 2$$

对于P_1，最高概率出现在两边，呈"U"形的双峰分布，对P_2，最高概率出现在中间，呈"∩"形单峰分布，它们加权求和结果都为2，这对P_2正确，但P_1在边缘处，2的概率最低，是平均结果的错误中间值。

$loss = |\hat{d} - d_{gt}|$ 鼓励单峰分布，对P_1 loss应当较大，结果会使边缘模糊。