EN. 520. 665 Final exam

1. The camera matrix C derived in previous section. has a null space which is spanned by the vector, $n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$, it's also the homogeneous representation of the 3D point which has cordinates (0, 0, 0), that is the camera center. This means that the camera center cannot be mapped to a point in the image plane by the camera

$$P^{3\times4} = U^{3\times4} D^{4\times4} V^{4\times4^T} \longrightarrow P^{3\times4} V^{4\times4}$$

$$= U^{3\times4} D^{4\times4} \rightarrow PV^4$$

$$= 0 \qquad \nwarrow \qquad PC = 0$$

2. For the way ~~we~~ talk about in the factorization, we are tracking the moving points build the matrix $W = [\frac{U}{V}]$, factorization and derive to the true matrix $\hat{W} = [\frac{\hat{U}}{\hat{V}}]$ and finally put in the training and draw the chain. for this part with some un-seen points, we can directly use the factorization method. However, there is usually sufficient information in the stream to determine all the camera positions and all the three-dimensional feature points coordinates. so we can not only solve the shape and motion recovery problem from the incomplete measurement matrix W, but also can even hallucinate the unknown entries of W by projecting the computed 3D feature coordinates onto the computed camera position. There is a detail method on paper shape-and-motion about the solution of Noise-Free image to talk about it.

3. $\because g(bx-ay)$, $\therefore \frac{dx}{dg} = g(b)$ $\frac{dy}{dg} = g(a)$

suppose $g(s)$ derivatives of x, y are equal

$z = f(x,y)$ and $z = f(x,y) + g(bx-ay)$ have the same derive of $R = ap + bq + c$, so they will have the same silhouette.

4. a) $L = D - W$ is also called the laplician matrix, is known to be positive semidefinite it's detail talked in paper "Partitioning Sparse Matrix withe Eigenvectors of Graphs", here gives a simple explained about: if B is the incidence matrix of an orientation of G, then $L = BB^T$, $x^T L x = \|Bx\|^2 \geq 0$ for all x

the matrix has rows indexed by the vertices columns by edges and the ij entry is 1 if the i-th vertex is the head of the j-th edge, -1 if it's the tail and 0 otherwise. $L = D - W$, D diagonal, W symmetric

$\downarrow$

L symmetric

$f^T L f \geq 0$, $\forall f \in R^n$

b) $$L = D - W = \begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n \end{pmatrix} - \begin{pmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=1}^{n} w_{1j} - w_{11} & \cdots & -w_{1n} \\ \vdots & \ddots & \\ -w_{n1} & \cdots & \sum_{j=1}^{n} w_{nj} - w_{nn} \end{pmatrix}$$

0 is an eigenvalue and its corresponding eigenvector is constant vector

since each eigenvalue is ≥ 0, it is the smallest eigenvalue

$$\Rightarrow L_1 = \begin{pmatrix} \sum_{j=1}^{n} w_{1j} - w_{11} & \cdots & -w_{1n} \\ \vdots & \ddots & \vdots \\ -w_{n1} & \cdots & \sum_{j=1}^{n} w_{nj} - w_{nn} \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=1}^{n} w_{1j} - w_{11} - \sum_{j=2}^{n} w_{1j} \\ \vdots \\ \sum_{j=1}^{n-1} (-w_{nj}) - \sum_{j=1}^{n} w_{nj} - w_{nn} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

4. continue : $\because$ L is symmetric, we can get L has n real eigenvalues

$\because$ L is positive-semidefinite

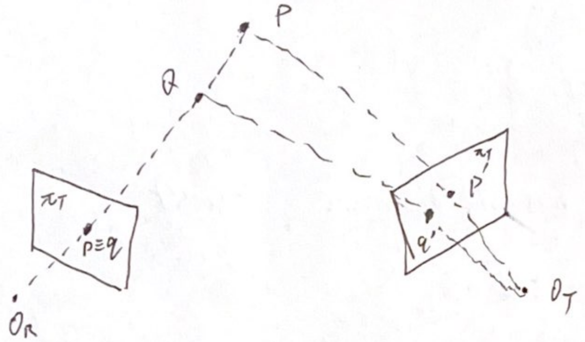
$$0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$$

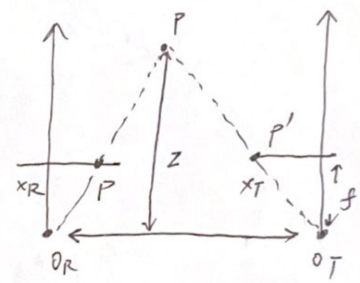
$\because$ The smallest eigenvalue of L is 0

$$\lambda_1 = 0$$

$$\therefore 0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$$

5. a) two cameras, just as we talk in the Lectures about the stereo cameras.





with two cameras we can infer depth, by means of triangulation, if we are able to find corresponding homologous points in the two images, and we can also use the epipolar constraints to make sure the image plane π_T of target image two cameras is the least cameras to make show the disparity and depth. as the graph shows. with the stereo rig in standard form and by considering similar triangles $\frac{b}{z} = \frac{(b+x_T)-x_R}{z-f} \Rightarrow Z = \frac{b \cdot f}{x_R - x_T} = \frac{b \cdot f}{d}$, $x_R - x_T$ is the disparity.

b) for the algorithm, for example we can just use the Marr - Poggio - Grimson - stereo algorithms. It's detailed talk on the paper you give us, here just a brief explain. First image Filtering $\nabla^2 G(x, y) = \left[\frac{x^2+y^2}{\sigma^2} - 2\right] \exp\left\{\frac{-(x^2+y^2)}{2\sigma^2}\right\}$, then test the original implementation $\left[\frac{n^2}{2^k cw}\right]^{n(2v+1)/2^k cw}$, then modify the algorithm $L_{cw}(x, y) = \nabla^2 G(w) * L$
$R_{cw}(x, y) = \nabla^2 G(w) * R$
as the step Loop-over → convolutions → zero crossings → Loop over fixation position → matching → Loop → Disambiguation → consistency.

for the parameters like the disparity and depth is necessary, and the points on these matrix points are also need, other detail parameter write in the paper, I can't list them all.

in real coding work: just a simple word: processing ⟶ feature tracking → modelling → find a package algorithm to do it
nearly every steps has their packaged function, like the SIFT tracking, the Flann matching

6. a) Activation volume dimensions: $62 \times 62 \times 32$

number of parameters: $(3 \times 3 \times 8 + 1) \times 32 = 2336$

b) Avd: $31 \times 31 \times 32$

np: 0

c) Avd: $31 \times 31 \times 32$

np: 2×32

d, Fliter size: 5

padding: 0

Activation volume dimensions: $12 \times 12 \times 4$

e, pool size : 2

padding = 0

Activation volume dimensions: $6 \times 6 \times 4$

f, Activation volume dimension: 1×10

g) 1, appropriate

2, 7 is vertically flipped and is unreadable, ∴ not appropriate

3, not appropriate

4) appropriate