4F12 Computer Vision Solutions

1(a)(i) Low pass filter to reduce effect of image noise (high frequencies) on gradient operation since noise is amplified by differentiation.

(ii) Common filter kernel used is the Gaussian (2D)

$$G_\sigma(x,y) = \frac{1}{2\pi\sigma} e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$ where σ determines scale (cut-off frequency)

and smoothed image is produced by:

(iii) $$S(x,y) = G_\sigma(x,y) * I(x,y)$$
$$= g_\sigma(x) * g_\sigma(y) * I(x,y)$$

where $$g_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

(iv) This is more efficient since each pixel is smoothed by 2 1D convolutions and hence:
$2N$ multiplications instead of N^2 for 2D convolution.

(v) For $\sigma = 1$, $N = 7$; taps are obtained by sampling $g_\sigma(x)$ until values are less than $1/1000$ peak.

1(b). Interest point ("corner", feature of interest) is localised where distinct gradients exist in neighbourhood. The image autocorrelation has a peak.

Look for a point where min and max of $(\text{gradient})^2$ is large.

$\therefore$ Evaluate $C(x,y) = G(x,y) * I_n^2$ in a neighbourhood

$$= \frac{n^T A^T A n}{n^T n} \qquad \text{where } A = \begin{bmatrix} \langle I_x^2 \rangle & \langle I_x I_y \rangle \\ \langle I_x I_y \rangle & \langle I_y^2 \rangle \end{bmatrix}$$

and check max and min are both large since:

$$\lambda_1 < C(x,y) = \frac{n^T A^T A n}{n^T A} < \lambda_2$$

$$\lambda_1 \leq C(x,y) \leq \lambda_2$$

we check $\det A - {}^{0.04}\text{trace}^2 A$ is above a threshold and look for <u>local</u> maxima.

1(c). After localising in each image compare by

(i) NxN patch of pixels compared using <u>normalized cross-correlation</u> (after subtracting mean intensity, normalize by variance)

(ii) Produce a descriptor such as SIFT based on histograms of gradients. Match by comparing descriptors using nearest-neighbour with distance between feature descriptors.

2(a) Consider lines $\underline{X}_i = \underline{a}_j + \lambda_i \underline{b}$ under perspective projection

$$\underline{x}_i = \frac{f \underline{X}_i}{Z_i} \quad \text{where } \underline{x} = \begin{pmatrix} x \\ y \\ f \end{pmatrix} \text{ and } \underline{X} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

For parallel lines as $Z_i \to \infty$

$$\underline{x}_{v.p} = \lim_{Z \to \infty} \frac{f \underline{X}}{Z} = \left. \frac{f(\underline{a}_j + \lambda_i \underline{b})}{(a_{3j} + \lambda_i b_3)} \right|_{\lambda \to \infty}$$

$$= f\left(\frac{\underline{b}}{b_3}\right)$$

and indep of a_j and hence position of line.

$$\underline{x}_{v.p} = \begin{pmatrix} f\, b_1/b_3 \\ f\, b_2/b_3 \\ f \end{pmatrix}$$

In renaissance camera $b_Z = 0$ and hence v.p.s at ∞.

(b) See crib of Q2 (2005)

Key pts
- 3x4 matrix ✓
- 11 dof ✓
- 6 pts give 2 eqns each. ✓ non-cop
- Solve $A\underline{p} = 0$ ✓
- Non-linear optimisation for result. ✓

(c) Weak perspective when $\frac{\Delta Z}{Z} \ll 1$. The projection matrix can be approximated by linear equations in matrix form:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Q3.

(a) In general for perspective projection from 3D pt (X, Y, Z) to image (u, v)

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \underbrace{\begin{bmatrix} p_{ij} & \cdots\cdots \\ & \end{bmatrix}}_{3\times 4} \begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix}$$

WLOG let $Z = 0$

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} t_{ij} \end{bmatrix} \begin{bmatrix} X \\ Y \\ W \end{bmatrix}$$

in homogeneous co-ordinates

(b) Consider 2 V.Ps corresponding with lines // to X-axis and Y-axis.

$$\underline{x}_{V.P_x} = \begin{bmatrix} t_{ij} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} t_{11} \\ t_{21} \\ t_{31} \end{bmatrix} \quad \text{or} \quad \left(\frac{t_{11}}{t_{31}}, \frac{t_{21}}{t_{31}} \right)$$

$$\underline{x}_{V.P_y} = \begin{bmatrix} t_{ij} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} t_{12} \\ t_{22} \\ t_{32} \end{bmatrix} \quad \text{or} \quad \left(\frac{t_{12}}{t_{32}}, \frac{t_{22}}{t_{32}} \right)$$

Both vanishing pts lie on horizon $\underline{L}$. In homogeneous co-ordinates

$\underline{L}\,\underline{x} = 0$ for pts on a line

$$\therefore \underline{L} = \underline{x}_{VP_x} \times \underline{x}_{VP_y} \quad \begin{vmatrix} i & j & k \\ t_{11} & t_{21} & t_{31} \\ t_{12} & t_{22} & t_{32} \end{vmatrix}$$

$$\therefore (t_{21} t_{32} - t_{31} t_{22})\,x + y\,(-t_{11} t_{32} + t_{31} t_{12}) + w\,(t_{11} t_{22} - t_{12} t_{21}) = 0$$

Q3(c)

$$VP_{\underline{x}} = K\underline{r}_1$$

$$VP_{\underline{x}} = K\underline{r}_2$$

Hence $\underline{r}_1 = K^{-1}\, VP_x$

$\underline{r}_2 = K^{-1}\, VP_y$

and $\underline{r}_3 = \pm\, \underline{r}_1 \times \underline{r}_2$

The orientation $\underline{R} = \begin{bmatrix} \underline{r}_1 & \underline{r}_2 & \underline{r}_3 \end{bmatrix}$ is recovered upto an ambiguity in sign.

Q4 (a) Each point in each image gives 2 equations in unknowns (X, Y, Z) of 3D pt.

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \overset{3\times4}{P} \begin{bmatrix} X \\ \hline 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} u' \\ v' \\ w' \end{bmatrix} = \overset{3\times4}{P'} \begin{bmatrix} X \\ \hline 1 \end{bmatrix}$$

Left camera right camera.

We have $\overset{(4\text{ planes})}{4}$ equations in 3 unknowns and can solve by Least-squares

$$\overset{4\times4}{\begin{bmatrix} A \end{bmatrix}} \begin{bmatrix} X \\ \hline 1 \end{bmatrix} = 0$$

(b) Essential matrix $E = [t_\times] R$ where $[t_\times] = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$

and $\underline{x}'^T E \underline{x} = 0$ where $\underline{x}$ and $\underline{x}'$ are vectors to image pt. $\begin{pmatrix} x \\ y \\ f \end{pmatrix}$.

$F = K'^{-T} E K^{-1}$ and $m'^T F m = 0$ where $\underline{m}$ is pixel loc.

and R = 3x3 matrix rep orientation of camera 2 rel to camera 1

$\underline{t}$ = translation bet. optical centres

K = 3x3 matrix rep internal parameters of camera (L)

K' = 3x3 matrix rep internal parameters of camera (R).

4(c) Stereo constraints
- epipolar geometry
- ordering
- figural continuity
- disparity gradient for smooth surfaces

4(d) Epipolar constraint

$\underline{u}'^T F \underline{u} = 0$ — since pt $\underline{u}'$ lies on epipolar line $\underline{l}'$

and $\underline{u}'^T \underline{l}' = 0$ — $\therefore$ since $\underline{u}' \cdot \underline{l}' = 0$

hence $\underline{l}' = F\underline{u}$

pt in left view $\underline{u}$

line in right view corresponding to $\underline{u}$ on which $\underline{u}'$ must lie.

(e). From $\underline{E}$ and SVD we can recover R and $\underline{t}$ and resolve any ambiguity.

From $\underline{F}$ we also need K and K' or euclidean knowledge of 5 pts in 3D world.