RC 1

L F12 solutions (2013)

Q1 (a) $S(x,y) = I(x,y) * G_\sigma(x,y)$ where $G_\sigma(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}$ /

$= I(x,y) * g_\sigma(x) * g_\sigma(y)$ where $g_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{x^2}{2\sigma^2}}$

$= \sum_{-n}^{n} \sum_{-n}^{n} I(u,v)\, g_\sigma(x-u)\, g_\sigma(y-v)$

$= \sum_{-n}^{n} \sum_{-n}^{n} I(x-u, y-v)\, g_\sigma(u)\, g_\sigma(v)$

(20%)

Show how to compute $g_\sigma(x)$ from $\frac{1}{\sqrt{2\pi}} e^{-x^2/2\sigma^2}$ curve eg. $2n+1 = 7$ for $\sigma = 1$. by sampling up to tails.

(b) $S(x, y, \sigma^2)$, family of blurred images with continuous σ

(see backwork overleaf)

$S(x, y, t) = G_\sigma(x,y) * I(x,y)$ for $t = \sigma^2$

Q1. (b) Image pyramids and blob-detection (scale-space)

- Need to generate a discrete set of images with different amount of blur. We sample $S(x, y, \sigma^2)$, logarithmically spaced;

$$\sigma_i = 2^{\frac{i}{s}} \sigma_0 \quad , \quad \sigma_{i+1} = 2^{\frac{i}{s}} \sigma_i \qquad (1)$$

with s images per octave (i.e. after s images, σ has doubled).

— apply incremental blur (gaussian σ_k) between images in octave to get images with increasing amount of blur.

$$G(\sigma_{i+1}) = G(\sigma_i) * G(\sigma_k) \qquad \sigma_{i+1}^2 = \sigma_i^2 + \sigma_k^2 \text{ and } \sigma_{i+1} = 2^{\frac{i}{s}} \sigma_i$$

$$\sigma_k = \sigma_i \sqrt{2^{\frac{2}{s}} - 1} \qquad (2)$$

— Each blur is performed as 2 1D convolutions (see (a) ii).

— After scale has doubled (i.e. s blurs) resize image by subsampling by 2 (i.e. $\frac{1}{4}$ size images). We can represent blurred images with fewer pixels without loss of information (Nyquist). [biggest saving] (3)

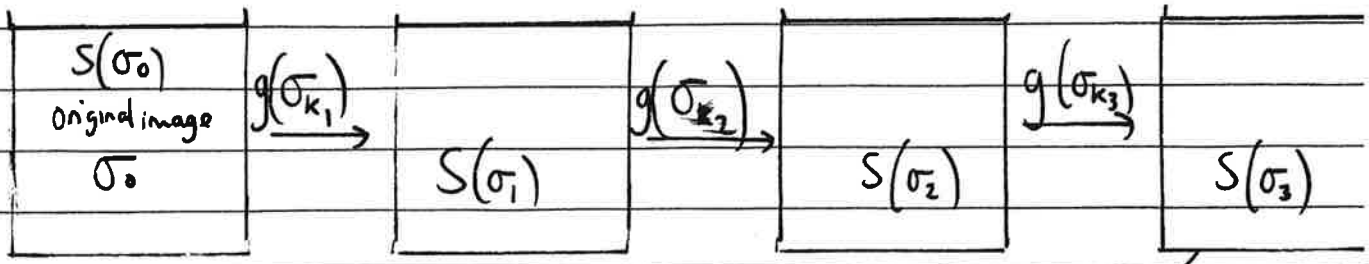
— store (small) incremental blur kernels used in each octave.

σ_{k_1}, σ_{k_2}, σ_{k_3} etc. on sub-sampled images,

but really representing with larger kernels, $2\sigma_{k_1}$ ..... $4\sigma_{k_1}$ ... $8\sigma_{k_1}$.

(b)

1st octave $\sigma_0 \rightarrow 2\sigma_0$

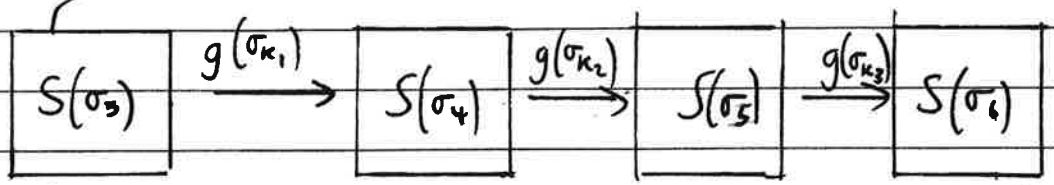


$\sigma_3 = 2\sigma_0$

Simage $\sigma_i = 2^{1/s}\sigma_0$

sub-sample to $\frac{1}{4}$ size

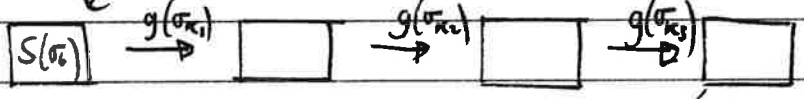
2nd octave $2\sigma_0 \rightarrow 4\sigma_0$



$2\sigma_6$ $\sigma_6 = 4\sigma_0$

subsample

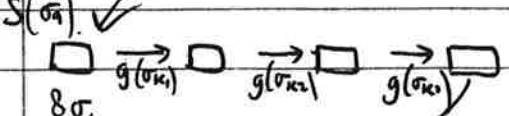
3rd octave $4\sigma_0 - 8\sigma_6$



$4\sigma_6$ $\sigma_9 = 8\sigma_0$

subsample

4th octave. $8\sigma_0 - 16\sigma_0$



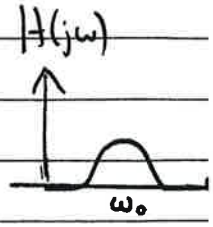
subsample

5th octave · $16\sigma_0 - 32\sigma_0$

$16\sigma_0$

16 pixel images

6(c) Band-pass filtering



- tuned to a small band of spatial frequencies (difference of high pass and low pass filtered o/p)

$$-\nabla^2 G_{\sigma_i} * I = \nabla^2 S(\sigma_i^2) \simeq S(\sigma_{i+1}^2) - S(\sigma_i^2)$$

i.e. subtract neighbouring images in same octave.

- generate a pyramid of DOG images by subtraction in same octave (2
(neighbours in "same" octave).

(c) Blobs are localised at max/min of $\nabla^2(G_\sigma * I)$ response.
Need to search over $\nabla^2 S(x, y, \sigma_i^2)$ for (max/min) local.

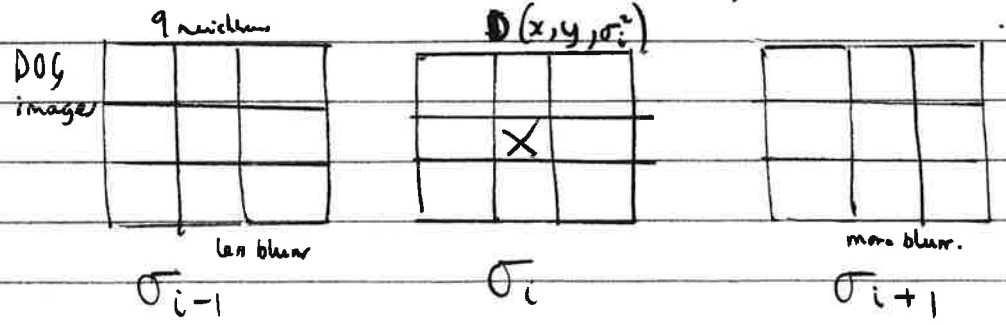
- More efficient to search over difference of gaussian (DOG) pyramid

$$\nabla^2 S \approx D(x, y, \sigma_i^2) \approx S(x, y, \sigma_{i+1}^2) - S(x, y, \sigma_i^2)$$

(10%)

for max/min in x, y and σ.

- Evaluate 26 neighbours of $D(x, y, \sigma_i^2)$ to see if local max/min



(30%)

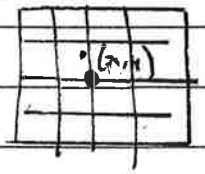
- Local max/min is blob location (x, y); scale (size of feature) is σ_i

- SIFT descriptor looks at 16×16 pixels sampled from $S(x, y, \sigma_i^2)$ in in correct octave.

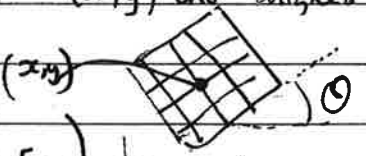
(d) cont Orientation: Look at gradients in 11x16 patch and bin (histogram at 10° intervals). Find dominant orientation after smoothing histogram.

Q12 SIFT, matching and classification

(a) - For each interest point at location (x, y) and scale σ estimate the dominant orientation θ, by looking at histogram of edge gradients from $\nabla S(x, y, \sigma^2)$.
Bins 10° apart + smoothed by gaussian. (36 bins)

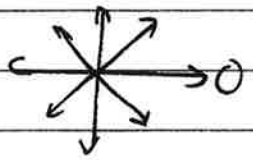


(b) - Sample 16x16 gradients from (x, y) and aligned with θ at image of scale, σ



- Smooth these gradients with $g(1.5\sigma)$ to emphasise gradients at interest point.

- produce orientation histogram (HOG) for each 4x4 block (cell)
Each bin records grad. magnitude interpolated over 8 dir^s (quadratic) only.



- concatenate 16 histograms of gradients (HOG's) to vectors 128D

- normalize to unit length; truncate to 0.2 to avoid illumination effects and normalize to 1.

- SIFT encodes 2D shape - invariant to lighting by using edges + normalisation step + to exact 2D position by histogram / bin effects [pooling]
Edges taken around a blob in centre.

(20/24)

Q2 (a) Consider $\underline{x}_i = \frac{f \underline{X}_i}{Z_i}$ and let $\underline{X}_i = \underline{a}_i + \lambda_i \underline{b}$

In lim. as $Z_i \to \infty$

$$\underline{x}_{i\infty} = \frac{f \underline{X}_i}{Z_i} = f\left(\frac{a_{1i} + \lambda_i b_1, a_{2i} + \lambda_i b_2}{a_{3i} + \lambda_i b_3} \right)$$

$$= f\left(\frac{b_1}{b_3}, \frac{b_2}{b_3} \right) \quad \text{and indep. of } \underline{a}_i$$

i.e. All have common vanishing pt if //. (20%)

2 (Q) i) $\underline{X}_c = R\underline{X} + \underline{T} = \left[\begin{array}{c|c} R & T \\ \hline 0 & 1 \end{array}\right] \left[\underline{X}\right]$ In this example $R = I$, $T = 0$

ii) perspective $\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$ S is used to represent homogeneous co-ordinates

iii) CCD scaling $u = u_0 + k_u \frac{f X_c}{Z_c}$, $v = v_0 + k_v \frac{f Y_c}{Z_c}$ $\cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix}$

Combine to get projection matrix.

Assumptions: pin-hole, no-non linear distortion, planar perspective (20%)

2(c)(i)

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

$$u = \frac{k_u f X_c}{Z_c} + u_0$$

$$v = \frac{k_v f Y_c}{Z_c} + u_0$$

while $$\begin{bmatrix} s_A u_A \\ s_A v_A \\ s_A \end{bmatrix} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 0 & Z_A \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

$$\therefore u_A = \frac{k_u f X_c}{Z_A} + u_0$$

$$v_A = \frac{k_v f X_c}{Z_A} + u_0$$

(3%)

$$\therefore (u - v_A) = \frac{k_u f X_c}{Z_c}\left(1 - \frac{Z_c}{Z_A}\right) = \frac{k_u f X_c}{Z_c}\left(\frac{Z_A - Z_c}{Z_A}\right)$$

$$v - v_A = \frac{k_v f Y_c}{Z_c}\left(1 - \frac{Z_c}{Z_A}\right) = \frac{\Delta Z}{Z_A}(u - u_0)$$

$$= \frac{\Delta Z}{Z_A}[v - v_0]$$

(10%)

(iii) Good approximation at centre of image and for small variation in depth, ΔZ (20%)

(iii) Calibrate from 4 pts by linear least-squares, (if more non-coplanar pts)

(3).

(a) $\underset{3\times1}{\begin{bmatrix} su \\ sv \\ s \end{bmatrix}} = \underset{3\times3}{[K]} \underset{3\times4}{[R \mid T]} \underset{4\times1}{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}$ in general

Assume $Z=0$, $\underset{3\times1}{\begin{bmatrix} su \\ sv \\ s \end{bmatrix}} = \underset{3\times3}{[K]} \underset{3\times3}{\begin{bmatrix} r_1 & r_2 & T \end{bmatrix}} \underset{3\times1}{\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}}$ for planar object

3×3 (20%)

(b).(i) 4 (10%)

(ii) RANSAC. — random sample of 4 pts.
— estimate t_{ij}
— check for inliers by Euclidean distance of $u' = Hu$ (20%)
— accept as correct matches if large # inliers

(iii). Set up $2N$ equations from N correspondences

$$u' = \frac{h_{11}u + h_{12}v + h_{13}}{h_{31}u + h_{32}v + h_{33}}$$

$$v' = \frac{h_{21}u + h_{22}v + h_{23}}{h_{31}u + h_{32}v + h_{33}}$$ (10%)

Normalize data to unit variance and zero mean.

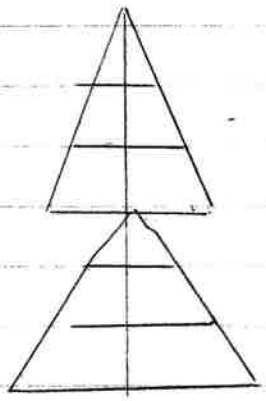
Set up $A\bar{h} = 0$ and solve by SVD.

3(c) Need to have $\underline{K}$ so as to recover R and t. (20%)

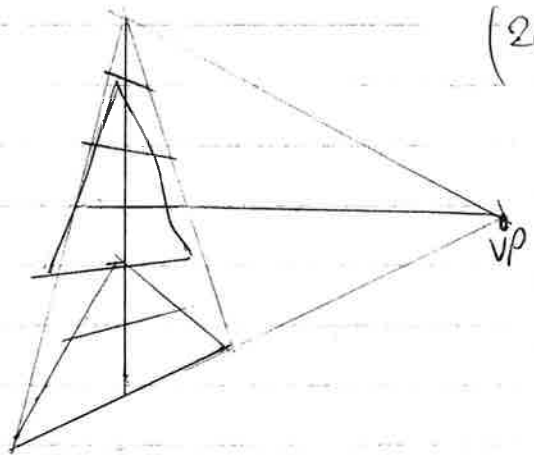
(d) Consider lines of symmetry as parallel lines eg. $y=0$, $y=a$ etc
Under perspective consider lines/pts at ∞

$$\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} h_{11} \\ h_{21} \\ h_{31} \end{bmatrix}$$

$$u'_{VP} = h_{11}/h_{31}$$
$$v'_{VP} = h_{21}/h_{31}$$



bilateral symmetry lost



(20%)

4a) The geometry of a stereo camera constrains each point-feature in one image to lie on a corresponding epipolar line in the other image. Epipolar lines meet at the epipole: this is the image of one camera's optical centre in the other camera's image plane. There are two epipoles – one for each image.

The fundamental matrix relates points in the left & right images of a stereo pair:

$$\underline{w}'^{T}\,\underline{\underline{F}}\,\underline{w} = 0 \qquad \text{where } \underline{w}^{T} = [u, v, 1]$$

$\underline{w}$ = points pixel coordinates in the left image and $\underline{w}'$ are the corresponding coordinates in the right image

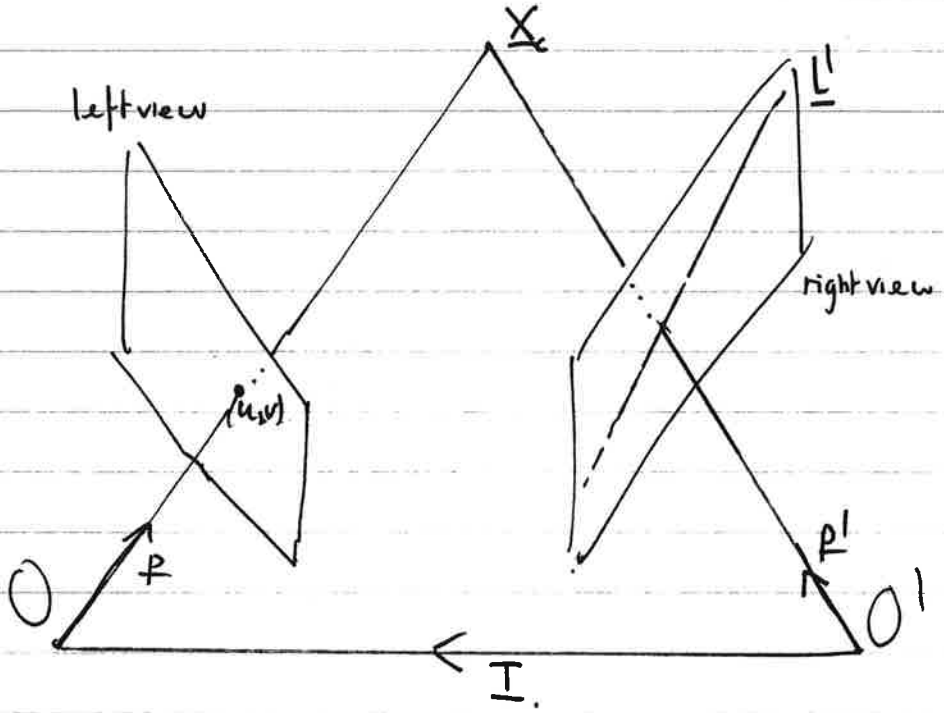
$\underline{\underline{F}}$ has zero determinant & can be determined only up to a scale factor.

Epipolar lines can be derived from the fundamental matrix.

20%

4(b)

$$\underline{X}'_{(c)} = R\,\underline{X}_{(c)} + \underline{T}$$



$\underline{\omega} = K\,\underline{p}$ and $\underline{\omega}' = K'\,\underline{p}'$

$\underline{X}' = R\underline{X} + T$ $\quad\quad$ $\underline{X}' \cdot (T_x R\,\underline{X}) = 0$ by coplanarity of $[\underline{X}', \underline{X}, \underline{T}]$

$\underline{p}'^T \; T_x R \; \underline{p} = 0.$

or $\underline{\omega}'^T \; K'^{-T} T_x R \; K^{-1} \; \underline{\omega} = 0$

$\underline{\omega}'^T F \underline{\omega} = 0$ where $\underline{F = K'^{-T} T_x R K^{-1}}$

$\therefore \underline{L' = K'^{-T} T_x R K^{-1} \omega}$

20%

4(c). A pt on an epipolar line in right view corresponding to $\underline{w}$ in left lies on a line $\underline{l}'$ such that

$$\underline{w}' \cdot \underline{l}' = 0 \qquad \text{where line is represented by } \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} \text{ and } \underline{w}'$$

But $\underline{w}' \cdot F\underline{w} = 0$ and hence $\underline{l}' = F\underline{w}$.

where $\underline{l}' = K'^{-T} T_x R K^{-1} \underline{w}$ (2 pts)

(d). Each pair of correspondence gives 1 equation in unknown F element f_{ij}

$\therefore Af = 0$ and solve by least-squares

Matrix F has max rank 2 ($\det F = 0$). Enforce by finding nearest matrix F^* to estimate $\hat{F}$ by minimizing Frobenius norm.

(2 pts)

(e) (*) E can be decomposed into $\underline{e}'_x [M]_*$ where $\underline{e}'$ is an epipole in right vi

or if intrinsic cameras are known it can be decomposed into $\underline{t}$ and $\underline{R}$

$P = K[I|0]$ and $\underline{P}' = K'[R|t]$ (2 pts)

$\therefore$ need to know K and K' of cameras.