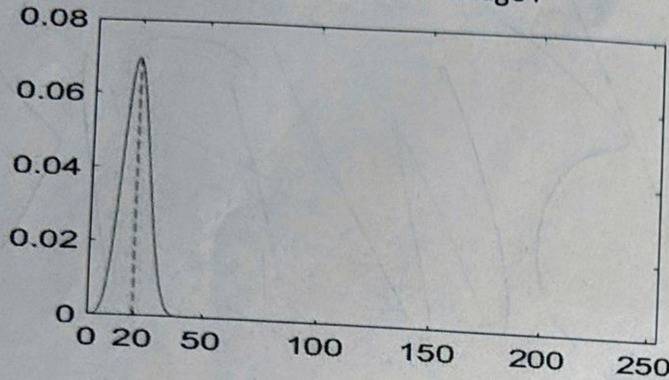
NAME: ____________________ MarkUS ID: ____________________

1. We have a single channel image *I* with the following histogram of pixel intensity values.

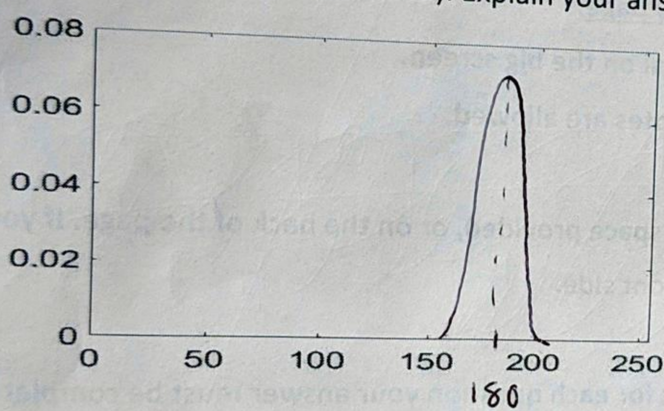
Histogram of image *I*



F =

0	0.5	0
0.5	1	0.5
0	0.5	0

[0.5]. We apply smoothing filter *F* (given above) to image *I* twice. Draw the histogram of the resulting image below (in the box provided). Explain your answer.



filter F sums up to 3.
So each time we apply it
to the image, intensity values
increas by a factor of 3

$(20 \times 3) \times 3 = 180$

2. In the Canny edge detector,
[0.5] what are the 3 adjustable parameters of the algorithm? Briefly explain the role of each parameter. (All three must be named and discussed correctly. No partial marks)

slides: lecture 3

1) Gaussian filter σ
- large σ → larger scale edges
- small σ → finer structures

(1.5) size of the Gaussian filter (e.g 3×3, 5×5, ...)
↳ how well a Gaussian filter is approximated)

2) high threshold (strong edges; ~~that~~ initiate hysteresis thresholding to connect edge fragments along the edge)

3) low threshold (weak edges; to continue connecting along the edge ...)

MarkUs ID: ______________________

3. The *p*-percentile filter maps a (single channel) input image *I* to an output image *J* of the same size and is defined as follows: in each neighborhood N_X centered at position *X* calculate the *p*-percentile value of all pixel values in that neighborhood. For example, if N_X is a 3x3 neighbourhood and p=33.33%, the nine pixel values in the neighbourhood will be sorted and the 3[rd] (0.3333 = 1/3 = 3/9) value will be picked.

[0.25] can this operation be performed through convolution? Explain your answer.

No.
Not a linear operation → cannot be implemented via a linear operator (convolution)

[0.75] If your answer is yes, write the filter and explain if it is separable.
If your answer is no, are there any values of *p* for which this would be possible? Explain your answer and list those values of *p* if any.

No values of p.
for all values of p the filter is non-linear

4. In SIFT,

(Lecture 8, Part 1)

[0.25] How is scale invariance achieved?

Your answer must refer to the Gaussian Pyramid, difference of Gaussians at multiple scales within each octave, and ~~a~~ selecting ~~peaks~~ key points that are peaks spatially (x, and y direction) and also in scale

[0.25] How is rotation invariance achieved?

- within a neighbourhood (16x16), form a histogram of gradient orientations (increments of 10°) (weight influence of orientation based on distance from centre)
- find the mode (highest peak) of the histogram → dominant orientation
- SIFT descriptor is a histogram of orientations. Subtract the dominant orientation from each

5. Consider the 1-D image I below. We want to up sample this image to size 5×n. The first step of our desired upsampling procedure is shown below.

I =

I(1)	I(2)	I(3)	...	...	...	I(n)

➔

I(1)	0	0	0	0	I(2)	0	0	0	0	I(3)	...

[0.25] Provide the filter that could be used for linear interpolation via convolution.

0.2	0.4	0.6	0.8	1	0.8	0.6	0.4	0.2

[0.25] Provide the filter that could be used for nearest neighbour interpolation via convolution.

1	1	1	1	1

6. In the Harris corner detector, M is a 2 × 2 second moment matrix computed from image gradients:

$$M = \sum_x \sum_y w(x,y) \begin{bmatrix} I_x^2 & I_x \cdot I_y \\ I_x \cdot I_y & I_y^2 \end{bmatrix}$$

where w(x,y) is a windowing function.

[0.5] What do you expect matrix M to look like for a point that is on a horizontal edge?

$$\approx \underbrace{C}_{\text{some constant}} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

[0.5] Harris corner detector applies a threshold on the value of *R*, where $\boxed{R = det(M) - \alpha\, trace(M)}$. Instead of *R*, what would applying a threshold on the value of $\boxed{trace(M)}$ achieve?

$trace(M) = \lambda_1 + \lambda_2$

trace ↑ ⟺ either λ_1 or λ_2 or both are high

⟹ edges

[1.0] Harris corner detector is applied to image *I* and also to its invert image *J*, where *J(x,y)* = 255 – *I(x,y)*. Describe how the corners found on *I* and *J* relate to each other.

Same corners will be detected

I_x and I_y will change sign

But I_x^2, I_y^2, and $I_x \cdot I_y$ will remain the same

7. The image on the left was convolved with a filter to produce the result on the right.
(credit: University of Cambridge, Department of Computer Science and Technology)

0	0	0
0	0	0
1	1	1
1	1	1
1	1	1
0	0	0
0	0	0

* ? =

1	1	1
-1	-1	-1
0	0	0
-1	-1	-1
1	1	1

$$\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

[0.25] Specify the filter (i.e. write down its numbers in an array) and explain what it accomplishes.

laplacian filter

edges → zero crossing

answer: 2

8. Similarity transformation can be written as $[\,s\mathbf{R} \mid \mathbf{t}\,]_{2\times3}$.

[0.25] how many pixel coordinate correspondences are needed to fit a similarity transformation?

4 degrees of freedom $\begin{cases} s: 1 \\ R: 1 \\ t: 2 \end{cases}$ $\Rightarrow$ $\boxed{\text{2 correspondences needed}}$

↳ each add 2 constraints

[0.5] Assuming a number of point matches, an initial estimate on the percentage of inliers (p), and a desired level of confidence to obtain the true transformation (e.g. $P \geq 0.99$), which would require fewer number of RANSAC iterations: fitting a similarity transformation or fitting an affine transformation? Explain your answer.

$$1 - P = (1 - p^k)^S \quad \text{or} \quad S = \frac{\log(1-P)}{\log(1-p^k)}$$

$k \uparrow \Rightarrow S \uparrow$

similarity : $k = 2$
affine : $k = 3$ $\Big\} \rightarrow$ similarity would require fewer iterations

9. Edges in images are cause by a variety of factors.

[0.5] Name these (4) factors (no partial marks)

- depth discontinuity
- surface normal discontinuity
- surface colour "
- illumination "

[0.5] Image gradients are used for edge detection. Edges formed by which of the above factors are detected by this operation? Briefly justify your answer.

all of the above.

They all show up as sharp changes in intensities.

So image gradient picks edges from all factors

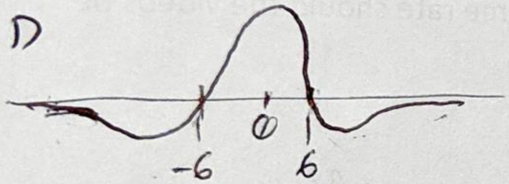
10. Laplacian of Gaussian is defined as

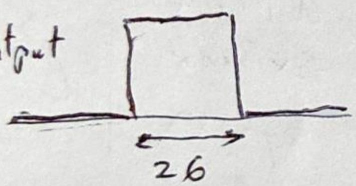
$$\nabla^2 g(x,y,\sigma) = \frac{\partial^2 g(x,y,\sigma)}{\partial x^2} + \frac{\partial^2 g(x,y,\sigma)}{\partial y^2} = -\frac{1}{\pi\sigma^4}\left(1 - \frac{x^2+y^2}{2\sigma^2}\right)\exp^{-\frac{x^2+y^2}{2\sigma^2}}$$

where g is a Gaussian defined as $g(x,y,\sigma) = \dfrac{1}{2\pi\sigma^2}\exp\left(-\dfrac{x^2+y^2}{2\sigma^2}\right)$

[1.0] Characteristic scale is defined as the scale that produces peak (minimum or maximum) of the Laplacian response. What value of **σ** maximises the magnitude of the response of the filter to an image of a black circle with diameter *D* on a white background? Justify your answer.

2D



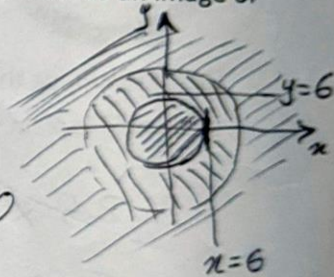


max output when input pulse has width 2σ

3D: $\nabla^2 g(x,y,\sigma) = 0$

$\Rightarrow \left(1 - \dfrac{x^2+y^2}{2\sigma^2}\right) = 0$

$\Rightarrow x^2 + y^2 = 2\sigma^2$ $\left\{\begin{array}{l} x=\sigma \\ y=\sigma \end{array}\right.$

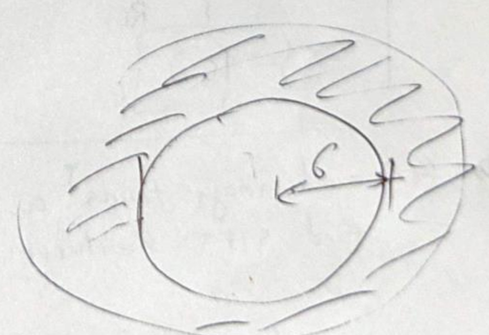


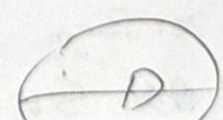
diameter $D \rightarrow$ radius $= \dfrac{D}{2}$

$\Rightarrow \sigma = \dfrac{D}{2}$

[0.5] What value of **σ** should we use if we want to instead detect a white circle of the same size on a black background?

same σ





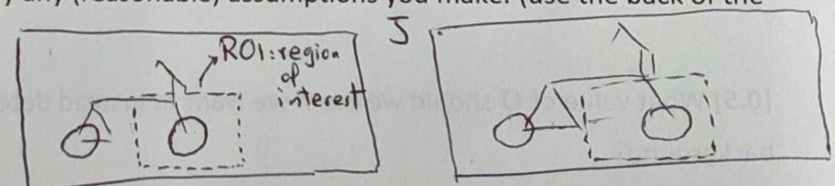
11. We want to measure the rotational speed (RPM: revolutions per minute) of the front wheel of a bicycle using a computer vision system. We have a camera on the sidewalk that records a short videos clip as the bicycle passes by. The bike belongs to a child who likes penguins and has placed several penguin stickers on the front wheel.

[0.25] Assuming that the child pedals in the 10-25 RPM range, at what frame rate should the videos be captured? Explain your answer.

$25 \times 2 = 50 \rightarrow$ 50 frames per minute

$\frac{50}{60} \approx 1$ frames per second

[3.25] Given a short video segment, assuming a sufficient frame rate, and assuming that the bounding box of the front wheel is known on the first frame of the video clip, using topics learned in this class, list ~5 algorithmic steps to implement such a system. Describe the purpose of each step, and also details of how it should be implemented. List and justify any (reasonable) assumptions you make. (use the back of the page for more space)



* For each frame I and subsequent frame J, do these steps:

- generate SIFT features ~~in~~ ROI on I
- 〃 〃 〃 〃 〃 〃 J
- match SIFT features
- perform RANSAC to fit an affine transform (A) or a similarity transform (S)
- calculate the rotation matrix from A (or from S) $\left(\text{e.g. } S = [sR \,\vdots\, \vec{t}\,] \Rightarrow R = \cdots \right.$ or $\left. A = \begin{bmatrix} s_x \cos\alpha & -s_x \sin\alpha & \vdots & t \\ s_y \sin\alpha & s_y \cos\alpha & \vdots & t \end{bmatrix} \Rightarrow R = \cdots \right)$
- from R, calculate the rotation angle α
- move the ROI to the new location

* average (or take the median of) all α values