UNIVERSITY OF BRISTOL

January 2023 Examination Period

Third Year Examination for the Degree of
Bachelor of Science and Master of Engineering

COMS30032J
Image Processing and Computer Vision

January 2023

TIME ALLOWED:
2 Hours

This paper contains *12* questions.
All answers will be used for assessment.
The maximum for this paper is *50 marks*.

Each question has exactly one correct answer.
You must select exactly one answer per question.

Answers

Do not turn over until told to start the exam.

Q1. Which of the statements A-D is INCORRECT? If you think none of them are incorrect, choose statement E.

A. Spatial aliasing artefacts in images and video can be mitigated by using a higher pixel density camera sensor.

B. Given a signal (e.g. an image) with a maximum frequency of w_{max} present, the signal can be reconstructed without loss when sampled above twice w_{max}.

C. Spatial aliasing artefacts in images and video can be mitigated by using a high pass filter in front of the camera's sensor.

D. Temporal aliasing artefacts in video can give the illusion of helicopter blades rotating backwards.

E. None of the statements above about aliasing is incorrect.

[2 marks]

Solution: C – All statements A-D are correct, except for C where a LOW PASS filter should be used to suppress the high frequencies.

Q2. In relation to K-means Clustering consider points 1 to 4 below:

1. We do not require a termination condition.
2. Outlier points may result in false cluster centroid positions.
3. We can choose any random initial centroids at the beginning of K-Means.
4. K-means performs best when clusters are of varying sizes and density.

Which set of the above are FALSE statements?

A. 1 and 4

B. 2 and 3

C. 2 only

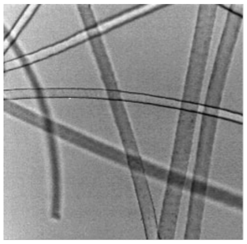
D. 1 only

E. 1, 2, and 4

[3 marks]

Solution: A 1 and 4 are false.

Q3. Given the original graylevel image below, we apply the convolution filters (1) to (5) with the resulting images given in a random order in images (a) to (e). Select which of the statements A to D below gives the correct pairing of filter and resulting image.



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

filter (1)

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

filter (2)

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

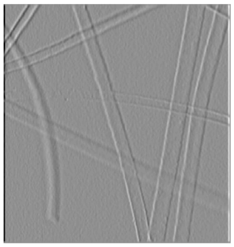
filter (3)

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

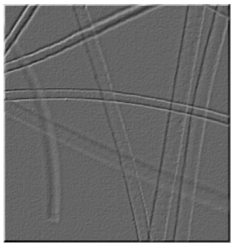
filter (4)

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

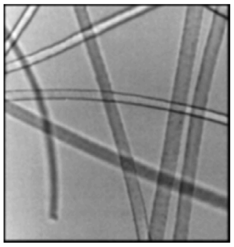
filter (5)



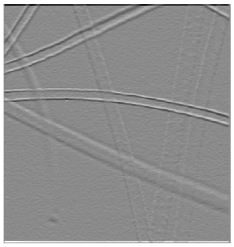
(a)



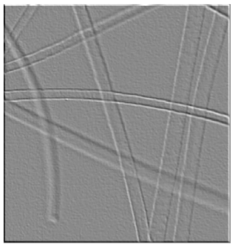
(b)



(c)



(d)



(e)

A. (1) with (a), (2) with (c), (3) with (b), (4) with (e), (5) with (d)

B. (1) with (d), (2) with (c), (3) with (b), (4) with (a), (5) with (e)

C. (1) with (d), (2) with (e), (3) with (a), (4) with (b), (5) with (c)

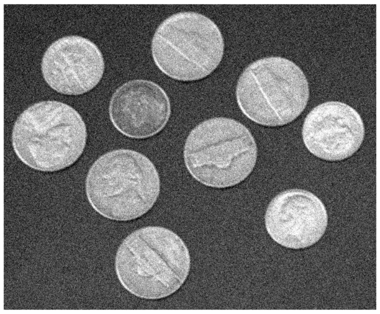
D. (1) with (d), (2) with (b), (3) with (a), (4) with (e), (5) with (c)

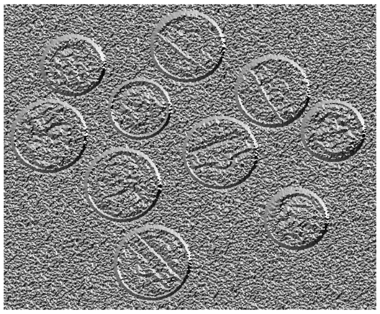
E. (1) with (e), (2) with (b), (3) with (d), (4) with (c), (5) with (a)

[6 marks]

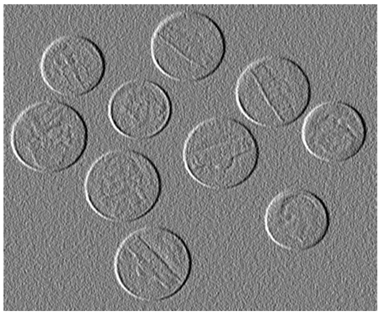
Solution: D – The filters from (1) to (4) detect, respectively, horizontal, +45 degrees, vertical, -45 degrees gradients. Since a line oriented in one of those filter directions also activates adjacent filters, we solve this problem by checking images where filters are not activated so much. The first pair that can be identified is (1) with (d), as (d) is the only image where the vertical lines don't appear, and it must therefore be because the filter used is detecting horizontal edges. Next, we notice that in (a) the horizontal edges are missing, which means that this filter must be detecting vertical edges instead: (3) is therefore associated with (a). In images (b) and (e) we can see only faint evidence evidence of lines oriented at around -45 degrees and +45 degrees respectively. We can thus associate (2) with (b) and (4) with (e). Finally, the averaging/smoothing filter (5) is associated with (c).

Q4. The coins in the noisy image below can be detected with circle Hough transform. We start with identifying edges using gradient extraction via filtering. However, to deal with noise, we need to design a new operator h that has an additional Gaussian smoothing h_g. A one-dimensional h_g is defined as h_g = [1 4 6 4 1], and a one-dimensional central difference is h_e = [-1 -2 0 2 1]. Then, we generate the image gradients and edge map, shown in (a)-(c), to be used to generate the circle Hough space.

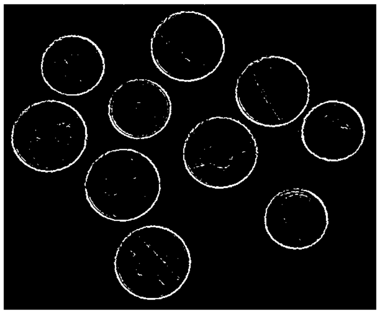




(a) I_a



(b) I_b



(c) I_c

Which of the following statements is CORRECT?

A. A kernel $h = h_e^T h_g$ is used to calculate I_b.

B. I_a can be used to limit the Hough space into three dimensions.

C. I_c is achieved by thresholding the magnitude of the gradients, $|I_{ab}| = \sqrt{I_a^2 + I_b^2}$.

D. A. and B. are both correct.

E. A., B. and C. are all correct.

[4 marks]

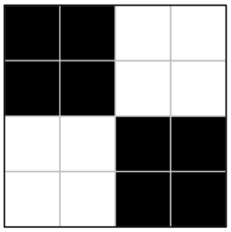
Solution: A is incorrect because $h_e^T h_g$ results vertical derivatives, but I_b shows horizontal derivatives. B is correct because the gradient vectors are used to calculate a centre of the circle. Otherwise 4-dimensional Hough space is needed as orientation is also considered: $H(x_0, y_0, r, \theta)$ C is incorrect because I_a shows gradient vectors, not horizontal derivatives.

Q5. Here is a region, taken from an integral image (**II**), as used in the Viola-Jones object detector.

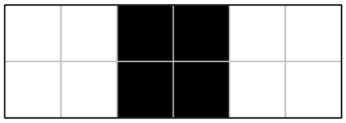
x

y	0	1	2	3	4	5	6	7	8	9
0	0	0	1	1	1	1	2	2	2	2
1	0	0	1	2	2	3	4	4	4	4
2	0	0	1	2	3	4	5	5	5	5
3	0	0	1	2	4	6	7	7	7	7
4	0	0	1	3	5	7	9	9	9	9
5	0	0	2	4	6	8	10	11	11	11
6	0	0	3	5	7	9	11	12	13	13
7	0	1	4	6	8	10	12	13	14	15
8	1	2	5	7	9	11	13	14	15	16
9	2	3	6	8	10	12	14	15	16	17

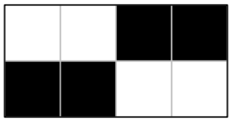
Here are four Haar-like features. In each case, the location at which the feature is evaluated in the integral image is given by the (x, y) coordinate as written under the feature. The location corresponds to the top-left position of the Haar mask. For this question, white areas of the features are positive, black are negative.



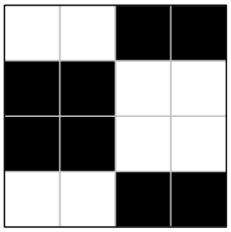
(a) x=0, y=0



(b) x=1, y=7



(c) x=4, y=3



(d) x=4, y=1

Which of the following statements has the CORRECT ascending order of the feature values when placing the mask at position (x,y)?

A. (a), (b), (d), (c)

B. (b), (a), (c), (d)

C. (c), (a), (b), (d)

D. (d), (a), (c), (b)

E. (d), (b), (a), (c).

[6 marks]

Solution: The feature value of the mask (a) is 2. The feature value of the mask (b) is 1. The feature value of the mask (c) is 3. The feature value of the mask (d) is ~~-1~~.

Q6. We want to detect the InfinityXYZ logo below on real outdoor images using an object detector that utilises Haar-like features. We build an object detector as follows.

1. Create the positive training data set of the InfinityXYZ logo from the single prototype image (the image (a) in the previous question).
2. Prepare the negative training data set from a thousand street images.
3. During the training, the detector via Adaboost, the boosting procedure considers all the positive images and employs sampled patches from the negative images to learn.

Which of the statements A-D is INCORRECT? If none of them are incorrect, choose option E.

A. The positive training data set, created in step 1, should be randomly changing viewing angle and contrast to reflect the possible variability of viewing parameters. However, features selected later in the Adaboost process tend to have higher error rates.

B. The number of negative training images in step 2 should be reduced, because the current classifier gives biased results toward the negative samples, leading to many false negatives.

C. During step 3, each weak classifier tested by Adaboost only uses one feature, and the strong classifier is a weighted linear combination of these weak classifiers.

D. To speed up the detection process, the strong classifier is built in several parts to form an attentional cascade.

E. None of the above statements is incorrect.

Solution: B – In contrast, more negative images are needed to reduce false positives.

[4 marks]

Q7. This question and the following two questions relate to the following scenario. Two views of a scene are taken by a single camera for stereo analysis. The first view is taken with the camera's principal axis pointing towards the centre of the scene. Before taking the second view, the camera is translated to a new position defined by the vector $\mathbf{T} = (10, 0, 0)$ and then rotated about its centre of projection (COP), as defined by the rotation matrix R given below. The focal length of the camera is 1. In this and the following questions you should assume perspective projection.

$$R = \frac{1}{2}\begin{bmatrix} \sqrt{3} & 0 & -1 \\ 0 & 2 & 0 \\ 1 & 0 & \sqrt{3} \end{bmatrix}$$

Consider a 3-D point $\mathbf{P} = (2, 0, 10)$ in the scene, defined w.r.t the camera coordinate system in the first view. Determine which of the following is closest to the magnitude of the disparity between the projections of **P** into each view.

A. 0.55

B. 0.45

C. 0.65

D. 0.25

E. 0.35

[4 marks]

Solution: 3-D point **P** projects to $p_1 = (0.2, 0, 1)$ in first view (perspective projection). Need to transform **P** to coordinate system of second view to find projection. $P_2 = R^T(P_1 - T)$, note use of R^T, not R, which is needed for coordinate transformation.

$$P_2 = \frac{1}{2}\begin{bmatrix} \sqrt{3} & 0 & 1 \\ 0 & 2 & 0 \\ -1 & 0 & \sqrt{3} \end{bmatrix}\begin{bmatrix} -8 \\ 0 \\ 10 \end{bmatrix} = \begin{bmatrix} 10 - 8\sqrt{3} \\ 0 \\ 8 + 10\sqrt{13} \end{bmatrix}$$

Hence $p_2 = ((10 - 8\sqrt{3})/(8 + 10\sqrt{13}), 0, 1) \approx (-0.15, 0, 1)$ and so magnitude of disparity is $d = 0.2 + 0.15 = 0.35$, hence E is correct.

Q8. *Same scenario as* Q7: Consider the following 3 points in the second view image: (a) $(0.1, 0.2)$; (b) $(-0.2, 0.3)$; and (c) $(-0.6, 0.1)$. Use epipolar geometry to determine which is most likely to be the corresponding point for the point $(0.2, 0.2)$ in the first view image. **Note**: the cross product between two 3-D vectors $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v}$ can be expressed as $A\mathbf{v}$, where A is given by:

$$A = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix}$$

A. (a)

B. (b)

C. (c)

D. (a) and (b) are equally likely

E. (a) and (c) are equally likely

[7 marks]

Solution: Need to compute essential matrix E and test for which point minimises $\mathbf{p}_2^T E \mathbf{p}_1$. $E = R^T S$, where S is the cross product matrix made from components of **T** and note use of R^T, as above.

$$E = R^T S = \frac{1}{2}\begin{bmatrix} \sqrt{3} & 0 & 1 \\ 0 & 2 & 0 \\ -1 & 0 & \sqrt{3} \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -10 \\ 0 & 10 & 0 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 0 & 10 & 0 \\ 0 & 0 & -20 \\ 0 & 2\sqrt{3} & 0 \end{bmatrix}$$

Now compute $\mathbf{u} = E\mathbf{p}_1$

$$\mathbf{u} = \frac{1}{2}\begin{bmatrix} 0 & 10 & 0 \\ 0 & 0 & -20 \\ 0 & 2\sqrt{3} & 0 \end{bmatrix}\begin{bmatrix} 0.2 \\ 0.2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -10 \\ \sqrt{3} \end{bmatrix}$$

and finally the dot product of **u** with the 3 possible corresponding points: $\mathbf{p}_{2a} = (0.1, 0.2, 1)$, $\mathbf{p}_{2b} = (-0.2, 0.3, 1)$ and $\mathbf{p}_{2c} = (-0.6, 0.1, 1)$, which gives: $\mathbf{u}.\mathbf{p}_{2a} = -0.17$; $\mathbf{u}.\mathbf{p}_{2b} = -1.47$; $\mathbf{u}.\mathbf{p}_{2c} = 0.13$. Thus $\mathbf{p}_{2c}$ is most likely since it is closest to the epipolar line and minimises $\mathbf{p}_2^T E \mathbf{p}_1$. Hence C is correct.

Q9. *Same scenario as* Q7: If the point $(-0.2, 0)$ in the first view and the principal point in the second view are corresponding points, which of the following is closest to the distance between the associated 3-D point and the centre of projection (COP) of the second view?

A. 26

B. 34

C. 32

D. 30

E. 28

[5 marks]

Solution: Wrt first view, let 3-D point be $a\mathbf{p}_1 = a(-0.2, 0, 1)$, and wrt second view, $b\mathbf{p}_2 = b(0, 0, 1)$ (scaled principal point). Transforming into first coordinate system gives $a\mathbf{p}_1 = bR\mathbf{p}_2 + T$ (note use of R, not R^T, as above), which gives

$$a\begin{bmatrix}-0.2\\0\\1\end{bmatrix} = b\begin{bmatrix}-0.5\\0\\\sqrt{3}/2\end{bmatrix} + \begin{bmatrix}10\\0\\0\end{bmatrix}$$

Equating terms then gives $a = \sqrt{3}b/2$ and $b = 10/(0.5 - 0.2\sqrt{3}/2) = 30.6$, making D correct.

Q10. Indicate which one of the following statements is NOT TRUE when considering the detection of corresponding points in a two camera stereo system.

A. The RANSAC algorithm can be used to identify likely and unlikely pairs of corresponding points

B. The descriptors used in the Scale-Invariant Feature Transform (SIFT) are derived from histograms of spatial gradients.

C. Epipolar lines always change if one of the cameras is translated to a new position

D. The fundamental matrix can be used to reduce the chance of mismatches

E. The Harris corner detector selects points which might be good for matching

[2 marks]

Solution: A. TRUE; B. TRUE; C. FALSE (translating COP along T vector keeps epipolar lines the same); D. TRUE (F can be used to form epipolar lines and so eliminate outliers). E. TRUE

Q11. A camera moves in the direction of its principal axis. To estimate motion in the image plane, the 2-D motion field within a region R is assumed to be given by $\mathbf{v} = (ax, ay)$, where x and y are the horizontal and vertical image coordinates, respectively, and a is the motion parameter. Estimates of a are determined by minimising the deviation from the optical flow equation within a region. If I_x, I_y and I_t are spatial and temporal intensity gradients, which of the following is a valid formula for a?

A. $a = -\sum_R I_t(I_x + I_y)/\sum_R(I_x + I_y)^2$

B. $a = -\sum_R I_t(I_x x + I_y y)/\sum_R(I_x x + I_y y)^2$

C. $a = -\sum_R I_t(I_x + I_y)/\sum_R(I_x x + I_y y)^2$

D. $a = -\sum_R I_t(I_x + I_y)/\sum_R(I_x + I_y)$

E. $a = -\sum_R I_t(I_x x + I_y y)/\sum_R(I_x + I_y)$

[5 marks]

Solution: Need to minimise $\sum_R(I_x ax + I_y ay + I_t)^2$. Differentiating wrt a gives $\sum_R 2(I_x ax + I_y ay + I_t)(I_x x + I_y y)$ and then setting to zero gives $\sum_R a(I_x x + I_y y)^2 = -\sum_R I_t(I_x x + I_y y)$ and so $a = -\sum_R I_t(I_x x + I_y y)/\sum_R(I_x x + I_y y)^2$, making B correct.

Q12. Indicate which one of the following statements is NOT TRUE when considering the estimation of the 2-D motion field produced by a moving camera viewing a static scene

A. If the camera only rotates about its centre of projection (COP), the motion field will be independent of depth in the scene

B. If the camera only translates in directions parallel to the image plane, the motion field is proportional to the focal length

C. The optical flow equation is based on the assumption that the motion field is constant within a local region

D. If the spatial gradients in a region are all equal and non-zero, only the normal flow can be estimated.

E. If the camera only translates along its principal axis, the motion field is inversely proportional to depth

[2 marks]

Solution: A. TRUE ; B. TRUE; C. FALSE; D. TRUE; E. TRUE

END OF PAPER