UNIVERSITY OF BRISTOL

January 2024 Examination Period

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

COMS30032J
Image Processing and Computer Vision

January 2024

TIME ALLOWED:
2 Hours

This paper contains *19* 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 following statements about segmentation is CORRECT?

A. Under-segmentation: Pixels belonging to the same object are classified as belonging to different segments.

B. In the Merge step of the Split-Merge method, all subregions that pairwise satisfy an inhomogeneity condition are merged together.

C. K-Means is insensitive to the initial cluster centres chosen at the start of the method.

D. Over-segmentation: Pixels belonging to the different objects are classified as belonging to the same segment.

E. None of the above are correct.

[2 marks]

Solution: E – None of the statements are correct.

Q2. Which of the following statements about convolution is INCORRECT?

A. Convolution can be implemented by applying the forward Fourier Transform to both an image and a kernel, then multiplying the two responses element by element, before applying the reverse Fourier Transform to the result.

B. In case of a kernel NOT being symmetrical with respect to a 180 degree rotation, convolution is then equivalent to correlation.

C. Convolution in the frequency domain can be implemented by applying the reverse Fourier Transform to both the Fourier Transforms of an image and a kernel, then multiplying the two responses element by element, before applying the forward Fourier Transform to the result.

D. The Convolution Theorem states that multiplication in the spatial domain is equivalent to convolution in the frequency domain and vice versa.

E. None of the above, all given statements about convolution are correct.

[2 marks]

Solution: B – Correlation and convolution are equivalent when the kernel is symmetric wrt 180 degrees rotation.

Q3. In relation to K-means Clustering consider points 1 to 5 below:

1. A termination condition is required.
2. Outlier points do not affect the convergence to the true cluster centroid positions.
3. Converges to the local maximum of within-cluster squared error
4. We can choose any reasonable number of random initial centroids at the beginning of K-Means.
5. K-means is more likely to detect more spherical clusters.

Which set of the above are FALSE statements?

(cont.)

A. 2 and 5

B. 3 and 4

C. 2 and 3

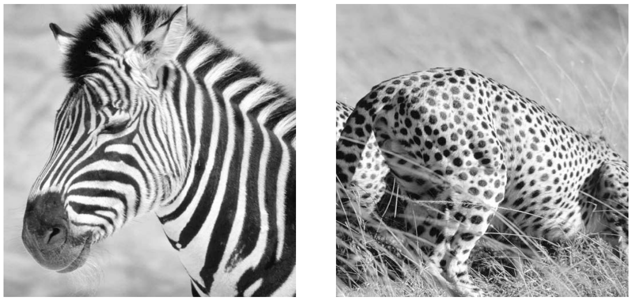
D. 5 only

E. 3 and 5

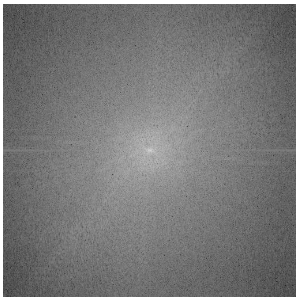
[3 marks]

Solution: C – 2 and 3 are false.

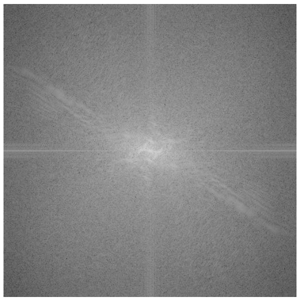
Q4. Consider these graylevel images of a Zebra (left) and a Cheetah (right):



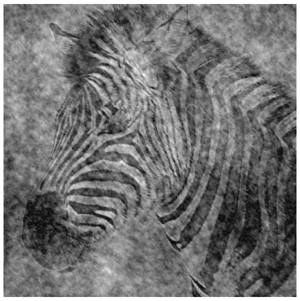
The results W, X, Y, and Z below, labelled *IN NO PARTICULAR ORDER*, are the outcomes of applying the Fourier transform to the above images to obtain their Fourier magnitudes (W and X) and phases (not shown), as well as the results after applying the inverse Fourier transform (Y and Z), but mistakenly mixing up the phase values of one image with another:



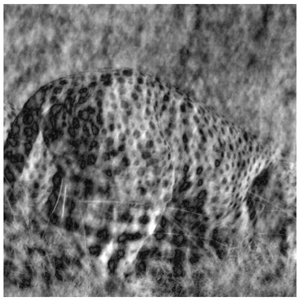
W



X



Y



Z

Select which of the statements A to D is TRUE:

A. W shows magnitudes of Zebra image, X shows magnitudes of Cheetah image, Y is magnitudes of Zebra with phases of Cheetah, Z is magnitudes of Cheetah with phases of Zebra

B. W shows magnitudes of Zebra image, X shows magnitudes of Cheetah image, Z is magnitudes of Zebra with phases of Cheetah, Y is magnitudes of Cheetah with phases of Zebra

(cont.)

C. X shows magnitudes of Zebra image, W shows magnitudes of Cheetah image, Y is magnitudes of Zebra with phases of Cheetah, Z is magnitudes of Cheetah with phases of Zebra

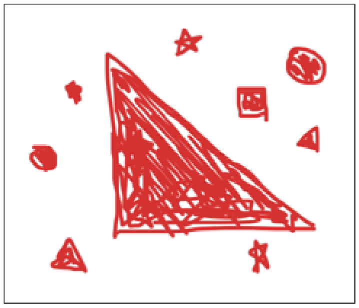
D. X shows magnitudes of Zebra image, W shows magnitudes of Cheetah image, Z is magnitudes of Zebra with phases of Cheetah, Y is magnitudes of Cheetah with phases of Zebra

E. All of the above choices are FALSE

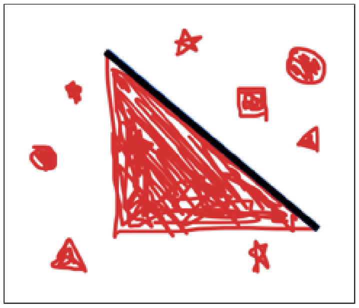
[4 marks]

Solution: D – For example, the strong direction of the magnitudes perpendicular to the stripes on the Zebra indicate that X belong to the Zebra image. As the phase records the structure of the image, then Y includes the phases of Zebra (and correspondingly Z includes the phases of the Cheetah image).

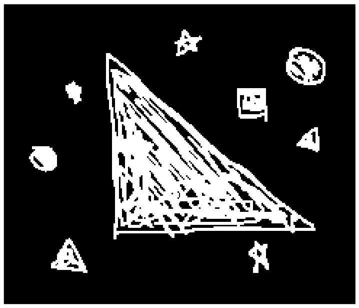
Q5. We have a hand-drawn colour RGB image I_a (see a.) and want to detect the longest side of the large triangle (the blue line as seen in b.). We begin by preprocessing the image to obtain the binary image I_b (c.), followed by the cleaned image I_c (d.), and finally, the edge image I_e (e). Next, we employ Hough transform to detect the target line as shown in (b).



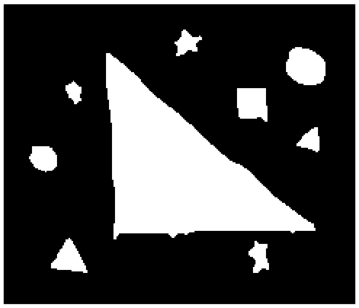
(a) Hand-drawn image I_a



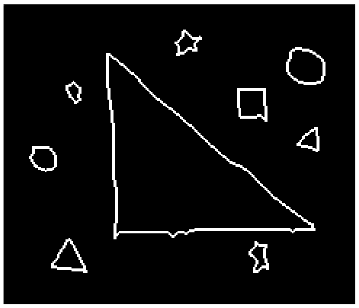
(b) Target line (black)



(c) Binary image I_b



(d) Cleaned image I_c



(e) Edge image I_e

Using a threshold T=100, what process can be used to obtain the binary image I_b from the image I_a, where I_a is in RGB (Red, Green, Blue) format with 8 bits per pixel per channel?

A. $I_b = R > T$, where R is the red channel of I_a

B. $I_b = G < T$, where G is the green channel of I_a

C. $I_b = (R < T) and (B > T)$, where R and B are the red and blue channels of I_a

D. $I_b = (R < T) and (G < T)$, where R and G are the red and green channels of I_a

E. $I_b = (R < T) and (G > T) and (B > T)$, where R, B and G are the red, blue and green channels of I_a

[2 marks]

Solution: The red area is where the value of red is high, whilst the values of blue and green are low. The white colour has (R,G,B) = (255, 255, 255)

Q6. Following the previous question, what morphological operations can be used to obtain the cleaned image I_c from the binary image I_b ?

A. Erosion

B. Dilation

C. Closing

D. Opening

E. B. and C. are correct.

[1 mark]

Solution: The closing operation dilates an image and then erodes the dilated image, using the same structuring element for both operations. Morphological closing is useful for filling small holes in an image while preserving the shape and size of large holes and objects in the image.

Q7. Following the previous question, what process can be used to obtain the edge image I_e from the cleaned image I_c ?

A. Compute the image gradient based on central differences, and then the edge exists if the magnitude of the image gradient is greater than a threshold.

B. Apply the Sobel operator, and then the edge exists if the magnitude of the filtered output is greater than a threshold.

C. Apply the morphological erosion, then subtract the eroded image from the cleaned image I_c

D. A. and B. are correct.

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

[2 marks]

Solution: See lecture week 3: edge detection for choice A. and B. The morphological operation can also be used to obtain in bounary of the object (edge), by using a small seed like one with a size of 3×3 pixels.

Q8. Following the previous question, now we apply Hough transform to the the edge image I_e and the Hough space is shown below. We identify lines from local maxima. Which peak is our target line (the longest side of the large red triangle). Note that the coordinate origin is the top-left corner of the image.

A. Peak A.

B. Peak B.

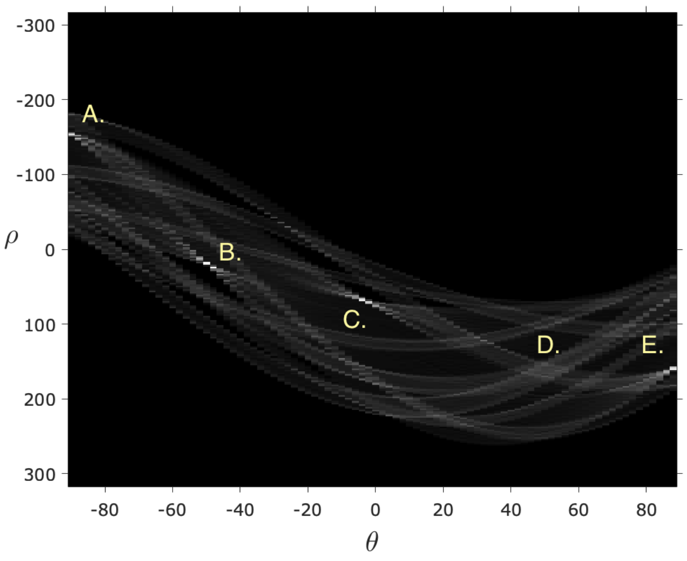
C. Peak C.

D. Peak D.

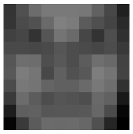
E. Peak E.

[2 marks]

Solution: Peak B is the result of the line with approximate -45 degree and the distance to the origin is very short.



Q9. We want to detect faces using the Viola-Jones object detector. A region of a face image shown in (a) has an integral image (**II**) shown in (b).

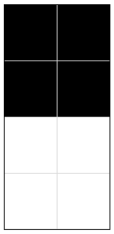


(a) face

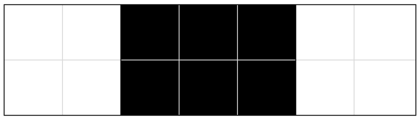
y	0	1	2	3	4	5	6	7	8	9
0	52	132	240	361	482	603	722	819	890	922
1	105	260	449	676	925	1169	1379	1555	1694	1768
2	170	395	635	947	1312	1662	1942	2169	2386	2508
3	243	541	852	1238	1689	2116	2470	2767	3060	3239
4	329	737	1157	1621	2160	2664	3110	3518	3916	4161
5	417	949	1486	2022	2657	3242	3776	4309	4823	5135
6	493	1145	1798	2434	3165	3846	4486	5139	5763	6124
7	551	1311	2056	2769	3585	4348	5066	5824	6545	6931
8	580	1435	2277	3076	3979	4829	5635	6496	7292	7685
9	585	1498	2431	3337	4346	5303	6212	7153	7986	8379

(b) integral image of (a)

Here are two 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=2



(b) x=2, y=0

What are the feature values of the mask (a) and (b)?

A. 51, -40

B. 91, -8

(cont.)

C. **127, 28**

D. 1083, 1404

E. None of the above is correct.

[3 marks]

Solution: The feature value of the mask (a) is 127. The feature value of the mask (b) is 28

Q10. Following the previous question, which of the following statements is CORRECT?

A. Both Haar-like features in (a) and (b) are designed to detect the eyes.

B. The Haar-like features in (a) and (b) are designed to detect the eyes and the nose, respectively.

C. The Haar-like features in (a) and (b) are designed to detect the mouth and the forehead, respectively.

D. The Haar-like features in (a) and (b) are designed to detect the eyebrow and the forehead, respectively.

E. All of the above are correct.

[2 marks]

Solution: The eyes, eyebrows and mouth form darker pixels. The forehead and cheeks form lighter pixels. The nose forms darker pixels than cheeks.

Q11. Which of the following statements about ADABOOST algorithm is INCORRECT?

A. In Viola-Jones algorithm, one classifier is trained by one Haar-like feature.

B. The algorithm prioritises misclassified samples and searches for the best available classifier based on these reweighted samples.

C. The final strong classifier is a nonlinear combination of the classifiers that achieve the lowest detection error from each iteration round.

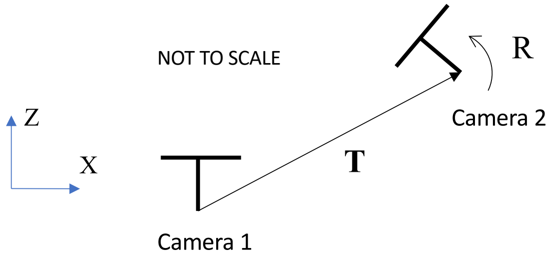
D. The drawback of ADABOOST is that it cannot be parallelised, which increases computational time.

E. None of the above is incorrect.

[2 marks]

Solution: The final strong classifier is a **linear** combination of the classifiers that achieve the lowest detection error from each iteration round.

Q12. This question and the following two questions relate to the scenario shown in the figure below in which two cameras, 1 and 2, are viewing a scene in 3-D space. The figure shows a plan view looking down onto the X-Z plane. The relative position and orientation of the cameras are defined by a vector **T** = (10, 0, 5) and a rotation matrix R, which is given below and defines a rotation about the Y-axis in the direction shown by the arrow in the figure. The focal length of the cameras is 1 and you should assume perspective projection.



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

A 3-D point P in the scene is given by the vector $\mathbf{P}_2 = (-1, -1, 8)$ defined in the coordinate system of camera 2. Determine which of the following is closest to the x-coordinate of the projection of P into the image plane of camera 1.

A. 0.42

B. 0.36

C. 0.32

D. -0.25

E. 0.28

[4 marks]

Solution: Representing all vectors as column vectors, $\mathbf{P}_2 = R^T(\mathbf{P}_1 - T)$ and $\mathbf{P}_1 = R\mathbf{P}_2 + T$. Note that R^T is rotation which brings camera 2 back into alignment with camera 1. $R\mathbf{P}_2 + T = \frac{1}{\sqrt{2}}(-9, -\sqrt{2}, 7) + (10, 0, 5) \approx (3.63, -1, 9.94)$. Using perspective projection, $x = X/Z \approx 3.63/9.94 = 0.36$. Hence B is correct.

Q13. For the above stereo system determine the essential matrix and hence determine which of the following is closest to the slope of the epipolar line in the image plane of camera 2 corresponding to the point $(1, 2)$ in the image plane of camera 1. **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 $\mathbf{v}$ is a column vector and 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. 0.7

B. 1.1

C. 1.4

D. 0.6

E. 1.6

[4 marks]

Solution: Need to compute essential matrix E and equation of line is then given by $\mathbf{p}_2^T(E\mathbf{p}_1) = 0$.

$$E = R^T S = \frac{1}{\sqrt{(2)}}\begin{bmatrix} 1 & 0 & 1 \\ 0 & \sqrt{2} & 0 \\ -1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & -5 & 0 \\ 5 & 0 & -10 \\ 0 & 10 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 5/\sqrt{2} & 0 \\ 5 & 0 & -10 \\ 0 & 15/\sqrt{2} & 0 \end{bmatrix}$$

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

$$\mathbf{u} = \begin{bmatrix} 0 & 5/\sqrt{2} & 0 \\ 5 & 0 & -10 \\ 0 & 15/\sqrt{2} & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5\sqrt{2} \\ -5 \\ 15\sqrt{2} \end{bmatrix}$$

Hence epipolar line is $5\sqrt{2}x_2 - 5y_2 + 15\sqrt{2} = 0$ and so slope is $m = 5\sqrt{2}/5 = \sqrt{2} \approx 1.4$. Hence C is correct.

Q14. A second 3-D point Q projects to the principal point in the image plane of each of the cameras. Determine which of the following pairs are closest to the distance of Q from each of the centres of projection, where the first number refers to the distance w.r.t camera 1 and the second number to the distance w.r.t camera 2.

A. 12,13

B. 16,16

C. 15,17

D. 15,14

E. 16,14

[4 marks]

Solution: Wrt camera 1, let 3-D point be $a\mathbf{p}_1 = a(0,0,1)$, and wrt camera 2, $b\mathbf{p}_2 = b(0,0,1)$ (principal point is at 0,0 in image plane). Transforming into coordinate system of camera 2 gives $a\mathbf{p}_1 = bR\mathbf{p}_2 + T$ (note use of R, not R^T, as above), which gives

$$a\begin{bmatrix}0\\0\\1\end{bmatrix} = b\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 0 & -1\\0 & \sqrt{2} & 0\\1 & 0 & 1\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix} + \begin{bmatrix}10\\0\\5\end{bmatrix} = \begin{bmatrix}-b/\sqrt{2}+10\\0\\b/\sqrt{2}+5\end{bmatrix}$$

Equating terms then gives $a = b/\sqrt{2}+5$ and $b = 10\sqrt{2}$, giving $a = 15$ and $b \approx 14.14$, making D correct.

Q15. Consider the following statements:

1. Region based matching is preferred to point based matching because it is significantly faster.
2. Using the Harris corner detector enables points to be compared to see if they are corresponding points.

Select one of the following:

A. Both statements are true.

B. 1 is true and 2 is false.

C. 1 is false and 2 is true.

D. Both statements are false.

[2 marks]

Solution: 1. FALSE - if anything, region based matching is slower; 2. FALSE - Harris detects points which would be good for matching but doesn't provide a way to compare points. Hence D is correct.

.

Q16. Consider the following statements:

1. Using epipolar lines to find correspondences increases the likelihood of finding correct matches.
2. Using RANSAC to find good correspondences relies on having a method for estimating the fundamental matrix from a small number of potential corresponding points.

Select one of the following:

A. Both statements are true.

B. 1 is true and 2 is false.

C. 1 is false and 2 is true.

D. Both statements are false.

[2 marks]

Solution: 1. TRUE - Confining search to epipolar lines reduces likelihood of incorrect matches; 2. TRUE - F is computed from a small sample of possible correspondences and support is then tested by counting how many correspondences satisfy the epipolar constraint. Hence A is correct.

.

Q17. A motion estimation algorithm assumes the 2-D motion $\mathbf{v} = (v_x, v_y)$ within regions of a video frame can be expressed as $v_x = ax + b$ and $v_y = cy + d$, where a, b, c and d are motion parameters. Estimates of the parameters are obtained by minimising the deviation from the optical flow equation within a region R. This can be expressed in the form $\mathbf{p} = A^{-1}\mathbf{u}$, where A is a 4×4 matrix, $\mathbf{p}$ is a 4x1 column vector containing the parameters in the order given above and $\mathbf{u}$ is also a 4x1 column vector. If I_x, I_y and I_t denote spatial and temporal gradients, which of the following is a valid formula for the components of $\mathbf{u}$?

A. $-2\sum_R I_t(x^2 I_x, xI_x, y^2 I_y, yI_y)$

B. $-\sum_R I_t(xI_x, I_x, yI_y, I_y)$

C. $-\sum_R I_t[I_x, I_x, I_y, I_y)$

D. $-2\sum_R I_t[I_x, xI_x, I_y, yI_y)$

E. $-\sum_R I_t[xI_x, yI_x, yI_y, xI_y)$

[5 marks]

Solution: Need to minimise $\sum_R(I_x(ax + b) + I_y(cy + d) + I_t)^2$. Differentiating wrt a gives $\sum_R 2(I_x(ax + b) + I_y(cy + d) + I_t)xI_x$ and then setting to zero gives $\sum_R(x^2 I_x^2)a + (xI_x^2)b + (xyI_xI_y)c + (xI_xI_y)d = -\sum_R xI_xI_t$, where the coefficients of a-d become the first row of A and the RHS becomes the first element of $\mathbf{u}$. Repeating for the other parameters gives $\mathbf{u} = -\sum_R I_t(xI_x, I_x, yI_y, I_y)$, making B correct.

Q18. Consider the following statements:

1. The optical flow in a video frame will ALWAYS depend ONLY on the relative movement between the camera and the scene.
2. The optical flow equation assumes that the brightness of a moving point remains the same between frames.

Select one of the following:

A. Both statements are true.

B. 1 is true and 2 is false.

C. 1 is false and 2 is true.

D. Both statements are false.

[2 marks]

Solution: 1. FALSE - except when the 3-D motion is only a rotation about the COP, the motion field will depend on depth in the scene; 2. TRUE - the OFE is based on the brightness consistency constraint. Hence C is correct.

Q19. Consider the following statements:

1. The aperture problem refers to the fact that there has to be sufficient variation in spatial gradients within a region to get good motion estimates.
2. Normal flow refers to the average optical flow observed across the whole of a frame.

Select one of the following:

A. Both statements are true.

B. 1 is true and 2 is false.

C. 1 is false and 2 is true.

D. Both statements are false.

[2 marks]

Solution: 1. TRUE - aperture problems occur if the window is too small and there is insufficient variation in spatial gradient to get good motion estimates; 2. FALSE - normal flow refers to the 2-D motion parallel to the direction of the spatial gradient. hence B is correct

END OF PAPER

The following pages are left blank for your rough workings. They will not be collected or marked. You must enter your answers on the provided answer sheet only.