—— SOLUTIONS ——

King's College London

This paper is part of an examination of the College counting towards the award of a degree. Examinations are governed by the College Regulations under the authority of the Academic Board.

Degree Programmes MSc, MSci

Module Code 7CCSMCVI
Module Title Computer Vision
Examination Period January 2017 (Period 1)

Time Allowed Three hours
Rubric ANSWER QUESTION ONE AND ANY THREE OTHER QUESTIONS.

All questions carry equal marks. If more than four questions are answered, the answer to the first four questions in exam paper order will count.

Calculators Calculators may be used. The following models are permitted: Casio fx83 / Casio fx85.

Notes Books, notes or other written material may not be brought into this examination

PLEASE DO NOT REMOVE THIS PAPER FROM THE EXAMINATION ROOM

 DO NOT TURN OVER UNTIL INSTRUCTED TO DO SO

1. Compulsory Question

a. Give a brief definition of each of the following terms.

i. computer vision

ii. down-sampling

iii. transduction

[6 marks]

Answer

i) computer vision = extracting information about the world from images.
iii) down-sampling = reducing the sampling rate, or resolution, of an image
(iv) transduction = the transformation of one form of energy to another

Marking scheme

2 marks for each correct definition.

b. The array below shows the intensity values in a 3-by-3 pixel greyscale image.

$$I = \begin{bmatrix} 5 & 11 & 21 \\ 1 & 6 & 9 \\ 3 & 9 & 8 \end{bmatrix}$$

Calculate the 2-by-2 pixel image that would result if this image was convolved with a 2-by-2 pixel box mask (or mean filter).

[5 marks]

Answer

A 2-by-2 pixel box mask =

$$\begin{bmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{bmatrix}$$

Filtered image =

$$\begin{bmatrix} (5+11+1+6)/4 & (11+21+6+9)/4 \\ (1+6+3+9)/4 & (6+9+9+8)/4 \end{bmatrix}$$

$$= \begin{bmatrix} 23/4 & 47/4 \\ 19/4 & 32/4 \end{bmatrix} = \begin{bmatrix} 5.75 & 11.75 \\ 4.75 & 8 \end{bmatrix}$$

Assuming output, like original image, should also be in integer format: Filtered image

$$= \begin{bmatrix} 6 & 12 \\ 5 & 8 \end{bmatrix}$$

Marking scheme

2 marks for knowing a correctly normalised box mask. 2 marks for correctly applying convolution to the image. 1 mark for realising that image format is integer and result should be rounded.

c. Define a MATLAB function B = binarize(I, t) that will convert an image, I, to a binary image, B, by applying a threshold of t.

[4 marks]

<u>Answer</u>

```
function B = binarize(I, t)
B=I;
B(B>t)=1;
B(B<=t);
```

<u>Marking scheme</u>

4 marks.

d. The binary image that results from applying a threshold of 8 to image I, as defined in question 1.b, is:

$$B = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Write down the new binary image that results from performing dilation on B, assuming each pixel has 4 neighbours (horizontal and vertical).

[4 marks]

<u>Answer</u>

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

<u>Marking scheme</u>

4 marks.

e. Describe briefly in words, or using pseudo-code, each step required to create:

i. a Gaussian image pyramid,

ii. a Laplacian image pyramid.

[6 marks]

Answer

(i)

- For each level in the pyramid
  convolve current image with Gaussian mask
  subsample convolved image (this is pyramid image)

Marking scheme

3 marks.

(ii)

- For each level in the pyramid
  convolve current image with Gaussian mask
  subtract the smoothed image from the previous image (this is pyramid image)
  subsample convolved image

Marking scheme

3 marks.

2. The arrays below show the values of the red, green and blue pixels in a Bayer masked image.

$R =$

0.2		0.3	
0.4		0.3	

$G =$

	0.7		0.6
0.1		0.6	
	0.8		0.9
0.4		0.1	

$B =$

	0.2		0.3
	0.4		0.6

a. Calculate the RGB values for the pixel at coordinates (2,2) in the demosaiced image using bi-linear interpolation.

[4 marks]

Answer

(a)

$$R(2,2) = \frac{R(1,1) + R(3,1) + R(1,3) + R(3,3)}{4} = \frac{0.2 + 0.3 + 0.4 + 0.3}{4} = 0.3$$

$$G(2,2) = \frac{G(2,1) + G(1,2) + G(3,2) + G(2,3)}{4} = \frac{0.7 + 0.1 + 0.6 + 0.8}{4} = 0.55$$

B(2,2)=0.2 as given.

Marking scheme

4 marks.

b. Calculate the RGB values for the pixel at coordinates (2,2) in the demosaiced image using edge-directed interpolation. Assume that pixels at coordinates (1,1), (1,3), and (3,1) all have green values of 0.5.

[5 marks]

<u>Answer</u>

At (2,2) $\Delta H = \|0.6 - 0.1\| = 0.5$, $\Delta V = \|0.7 - 0.8\| = 0.1$
Therefore,

$$G(2,2) = \frac{G(2,1) + G(2,3)}{2} = \frac{0.7 + 0.8}{2} = 0.75$$

At (3,3) $\Delta H = \|0.8 - 0.9\| = 0.1$, $\Delta V = \|0.6 - 0.1\| = 0.5$
Therefore,

$$G(3,3) = \frac{G(2,3) + G(4,3)}{2} = \frac{0.8 + 0.9}{2} = 0.85$$

$$R(2,2) = G(2,2) \times \frac{1}{4}\left(\frac{R(1,1)}{G(1,1)} + \frac{R(3,1)}{G(3,1)} + \frac{R(1,3)}{G(1,3)} + \frac{R(3,3)}{G(3,3)}\right)$$

$$R(2,2) = 0.75 \times \frac{1}{4}\left(\frac{0.2}{0.5} + \frac{0.3}{0.5} + \frac{0.4}{0.5} + \frac{0.3}{0.85}\right) = 0.4037$$

B(2,2)=0.2 as given.

<u>Marking scheme</u>

5 marks.

c. The pinhole camera model of image formation relates the coordinates of a 3D point P with coordinates (X,Y,Z) relative to the camera reference frame to the coordinates of its image p=(x,y,z) via the following equation:

$$(x, y, z) = \left(\frac{fX}{Z}, \frac{fY}{Z}, f\right)$$

Given that two identical cameras are mounted so as to have coplanar image planes, such that one camera (the right one) is displaced distance B along the x-axis of the other (the left one), derive an equation for the depth of a 3D point P visible in both cameras.

[6 marks]

Answer

For left camera:

$$(x_L, y_L) = \left(\frac{fX_L}{Z_L}, \frac{fY_L}{Z_L}\right)$$

For right camera:

$$(x_R, y_R) = \left(\frac{fX_R}{Z_R}, \frac{fY_R}{Z_R}\right)$$

Because the x-axes of the camera are collinear, $y_L = y_R$, also because the image planes are coplanar $Z_L = Z_R$. Furthermore, the X-coordinates of the 3D point in the right camera is related to its position in the left camera by: $X_R = X_L - B$.

Hence,

$$(x_R, y_R) = \left(\frac{f(X_L - B)}{Z_L}, \frac{fY_L}{Z_L}\right)$$

Subtracting x-coordinates:

$$x_L - x_R = \frac{fX_L}{Z_L} - \frac{f(X_L - B)}{Z_L}$$

Hence,

$$Z_L = \frac{fB}{x_L - x_R} \qquad (= Z_R)$$

Marking scheme

6 marks.

d. Describe five constraints that might be applied in trying to determine the locations of the points in the left and right image that correspond to the same point in 3D space.

[10 marks]

Answer

epipolar constraint (corresponding points lie along the epipolar line)

maximum disparity (extent of search reduced by knowledge of baseline and minimum depth)

continuity (disparity varies smoothly assuming continuous surfaces)

uniqueness (one-to-one matches assuming surfaces are approximately perpendicular to camera and no occlusion)

ordering (points occur in same order in each image assuming no occlusion)

[similarity (locations should have similar features)]

Marking scheme

2 marks for each correct definition.

3. a. The array below shows a Laplacian mask, L.

$$L = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Write down a mathematical expression describing the effect of convolving an image I with L.

[2 marks]

<u>Answer</u>

$I' = I * L \approx -\left(\frac{\delta^2 I}{\delta x^2} + \frac{\delta^2 I}{\delta y^2}\right)$

<u>Marking scheme</u>

2 marks.

b. Write down the mask that approximates the following directional derivative: $-\dfrac{\delta^2}{\delta y^2}$

[2 marks]

<u>Answer</u>

$-\frac{\delta^2}{\delta y^2} \approx \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}$

<u>Marking scheme</u>

2 marks.

c. Use the following formula for a 2D Gaussian to calculate a 3-by-3 pixel numerical approximation to a Gaussian with standard deviation of 0.5 pixels, rounding values to two decimal places.

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

[3 marks]

Answer

Gaussian mask

$$= \begin{bmatrix} 0.01 & 0.09 & 0.01 \\ 0.09 & 0.64 & 0.09 \\ 0.01 & 0.09 & 0.01 \end{bmatrix}$$

Marking scheme

1 mark each for the 3 different values.

d. Using the Laplacian defined in question 3.a and the Gaussian calculated in answer to question 3.c calculate a 3-by-3 pixel Laplacian of Gaussian (or LoG) mask.

[5 marks]

Answer

$$LoG = \begin{bmatrix} -0.74 & -0.12 & -0.74 \\ -0.12 & 4.72 & -0.12 \\ -0.74 & -0.12 & -0.74 \end{bmatrix}$$

Marking scheme

2 marks for knowing that L and G need to be convolved together. 1 mark each for the 3 different values.

e. What advantage does a LoG mask have over a Laplacian for edge detection?

[2 marks]

Answer

The Laplacian is sensitive to noise as well as other intensity-level discontinuities.

The Gaussian is a smoothing mask that suppress noise.

The combination of the two produces a mask that is sensitive to intensity-level discontinuities that are image features rather than noise.

Marking scheme

2 marks.

f. An alternative method of performing edge detection is to simulate the responses of orientation tuned neurons in primary visual cortex (V1). Name the mathematical function that is used to model the receptive fields of cortical simple cells, and write down the equation for this function.

[4 marks]

Answer

The Gabor function.

$$Gabor(x,y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos(2\pi x' f + \psi)$$

Marking scheme

2 marks for name. 2 marks for equation.

g. Describe how the responses of V1 simple cells can be modelled using convolution.

[4 marks]

Answer

Convolving the image with a Gabor mask will simulate the response of all simple cells selective for the same parameters across all hyper-columns.

Repeating the convolution with Gabor masks with different parameters (e.g. orientation, spatial frequency, phase, aspect ratio) will simulate the responses of all different types of V1 simple cell.

Marking scheme

4 marks.

h. Describe how the responses of V1 complex cells could be modelled using convolution.

[3 marks]

Answer

A complex cell can be modelled by combining the outputs of two or more simple cells.

For example, the response from a quadrature pair of Gabor functions (two Gabors with a phase difference of $\pi/2$) can be used as input to a model of a complex cell. These inputs need to be combined by taking the square-root of the sum of the squares of the two inputs.

Alternatively, the half-wave rectified response from four Gabor functions differing in phase by $\pi/2$ can be used as input to a model of a complex cell. These inputs are combined by squaring and taking the sum.

Hence, complex cell responses can be modelled by performing multiple convolutions with different Gabor masks to simulate simple cell responses (as described in the answer to the previous question), and subsequently combining (as described above) those responses for Gabor masks with identical parameters except with different phases.

Marking scheme

3 marks.

4. a. Briefly describe the role of mid-level vision.

[1 mark]

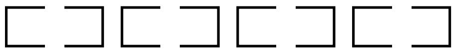
Answer

To group together image elements that belong together, and to segment them from all other image elements.

Marking scheme

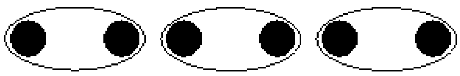
1 mark.

b. Below are four simple images. For each image identify the "Gestalt Law" that accounts for the observed grouping of the image elements.

i. 

ii. 

iii. 

iv. 

[8 marks]

Answer

(i) closure
(ii) proximity
(iii) similarity
(iv) common region

Marking scheme

2 marks for each correct answer.

c. One method of image segmentation is region growing. Write pseudo-code for the region growing algorithm.

[5 marks]

Answer

1. While not all pixels have been assigned a region label
2. Choose an unlabelled pixel at random
3. Assign chosen pixel the next unused region label
4. For all neighbouring pixels which haven't been assigned a label

    4.1 Calculate similarity to chosen pixel

    4.2 For each pixel within the similarity threshold

    4.2.1 Give that pixel the same region label as the chosen pixel

    4.2.2 Make that pixel the chosen pixel and repeat from step 4

5. Repeat from step 1

Marking scheme

5 marks.

d. The array below shows 3 element long feature vectors for each pixel in a 2-by-3 pixel image, I_F.

$$I_F = \begin{bmatrix} (20,10,5) & (10,20,15) \\ (15,5,5) & (5,5,20) \\ (15,15,15) & (20,15,10) \end{bmatrix}$$

Apply the region growing algorithm to image I_F in order to assign each pixel to a region. Assume that (i) the method used to assess similarity is the sum of absolute differences (SAD), (ii) the criterion for deciding if elements are similar is that the SAD is less than 12, (iii) the seed pixel is the top-left corner, (iv) pixels have horizontal, vertical and diagonal neighbours.

[6 marks]

<u>Answer</u>

Label top-left pixel 1. Calculate SAD values for pixels neighbouring the top-left pixel:

$$\begin{bmatrix} (20,10,5):1 & \mathbf{30}\rightarrow & (10,20,15) \\ \mathbf{10}\downarrow & \mathbf{35}\searrow & \\ (15,5,5):1 & & (5,5,20) \\ & & \\ (15,15,15) & & (20,15,10) \end{bmatrix}$$

Calculate SAD values to unlabelled pixels from the middle-left pixel:

$$\begin{bmatrix} (20,10,5):1 & & (10,20,15) \\ & \nearrow\mathbf{30} & \\ (15,5,5):1 & \mathbf{25}\rightarrow & (5,5,20) \\ \mathbf{20}\downarrow & \searrow\mathbf{20} & \\ (15,15,15) & & (20,15,10) \end{bmatrix}$$

Region 1 cannot grow further. Choose another seed pixel: top-right:

$$\begin{bmatrix} 1 & (10,20,15):2 \\ & \mathbf{25}\downarrow \\ 1 & (5,5,20) \\ & \\ (15,15,15) & (20,15,10) \end{bmatrix}$$

Region 2 cannot grow further. Choose another seed pixel: bottom-left:

$$\begin{bmatrix} 1 & & 2 \\ & & \\ 1 & & (5,5,20) \\ & \mathbf{25}\nearrow & \\ (15,15,15):3 & \mathbf{10}\rightarrow & (20,15,10):3 \end{bmatrix}$$

Calculate SAD values to unlabelled pixels from right-bottom-right pixel:

$$\begin{bmatrix} 1 & 1 \\ & \\ 1 & (5,5,20) \\ & \mathbf{35}\uparrow \\ (15,15,15):3 & (20,15,10):3 \end{bmatrix}$$

Region 3 cannot grow further. Last pixel has no unlabelled neighbours.

So final region labels: $\begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 3 & 3 \end{bmatrix}$

<u>Marking scheme</u>

6 marks.

e. An alternative method of image segmentation is Normalised Cuts. Write pseudo-code for the Normalised Cuts algorithm for image segmentation.

[5 marks]

Answer

1. Form a graph such that each vertex represents an image element and each edge, e_{pq} represents the similarity between the connected elements (p and q)

2. Partition the graph into two sets of vertices A and B

3. Calculate:

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(B, V)}$$

Where:

$$cut(A, B) = \sum_{p \in A, q \in B} e_{pq}$$

$$assoc(A, V) = \sum_{p \in A, q \in V} e_{pq}$$

4. Repeat from step 2 for every possible combination of nodes in A and B

5. Cut the edges between the sets A and B for which $Ncut(A, B)$ is minimum

6. Repeat from step 2 for all subgraphs until minimum value of $Ncut(A, B)$ exceeds a threshold for each subgraph or maximum number of segments have been reached.

Marking scheme

5 marks.

5. a. Briefly describe the methodology used in correlation-based and feature-based methods of solving the stereo correspondence problem.

[4 marks]

Answer

Correlation-based methods attempt to establish a correspondence by matching image intensities between windows of pixels extracted from both images.

Feature-based methods attempt to establish a correspondence by matching image descriptors extracted from around a sparse sets of image locations (usually corners) in each image .

Marking scheme

2 marks for correct description.

b. The two arrays below show the intensity values for each pixel in a stereo pair of 5-by-3 pixel images.

$$I_{left} = \begin{array}{|c|c|c|c|c|} \hline 40 & 60 & 40 & 20 & 50 \\ \hline 10 & 50 & 80 & 80 & 30 \\ \hline 70 & 10 & 70 & 60 & 90 \\ \hline \end{array} \qquad I_{right} = \begin{array}{|c|c|c|c|c|} \hline 20 & 70 & 70 & 20 & 50 \\ \hline 30 & 20 & 50 & 10 & 50 \\ \hline 50 & 70 & 40 & 80 & 70 \\ \hline \end{array}$$

Calculate the similarity of the pixel at coordinates (2,2) in I_{left}, to locations in I_{right}, and hence, calculate the disparity at that point. Assume that (i) a 3-by-3 pixel window is used, (ii) similarity is measured using the Sum of Absolute Differences (SAD), (iii) that corresponding points will occur along corresponding scan lines (*i.e.*, on the same row in both images), (iv) disparity is calculated as the translation from right to left.

[4 marks]

Answer

SAD =

310	250	180	290	290

Hence, best match is at location (3,2) in the right image.

Disparity is left-right = (2,2)-(3,2) = (-1,0).

Marking scheme

4 marks.

c. Write down the equation that defines the measure R used by the Harris corner detector.

[4 marks]

Answer

$$R = \left[\sum I_x^2 \sum I_y^2 - \left(\sum I_x I_y\right)^2\right] - k\left[\sum I_x^2 + \sum I_y^2\right]^2$$

Marking scheme

4 marks.

d. The two arrays below show the derivatives of the image intensities in the x and y directions for a small 3-by-3 pixel image.

$$I_x = \begin{array}{|c|c|c|}\hline 1 & 0 & -2 \\ \hline 0 & -1 & 0 \\ \hline 0 & -3 & 1 \\ \hline \end{array} \qquad I_y = \begin{array}{|c|c|c|}\hline 2 & 1 & -3 \\ \hline 0 & 0 & 0 \\ \hline 0 & -2 & 1 \\ \hline \end{array}$$

Calculate the measure R of the Harris corner detector at each pixel, assuming (i) a value of k=0.05, (ii) that products of derivatives are summed over an equally weighted, 3-by-3 pixel, window around each pixel, padding with zeros when necessary.

[4 marks]

Answer

$$I_x^2 = \begin{array}{|c|c|c|}\hline 1 & 0 & 4 \\ \hline 0 & 1 & 0 \\ \hline 0 & 9 & 1 \\ \hline \end{array} \qquad I_y^2 = \begin{array}{|c|c|c|}\hline 4 & 1 & 9 \\ \hline 0 & 0 & 0 \\ \hline 0 & 4 & 1 \\ \hline \end{array} \qquad I_xI_y = \begin{array}{|c|c|c|}\hline 2 & 0 & 6 \\ \hline 0 & 0 & 0 \\ \hline 0 & 6 & 1 \\ \hline \end{array}$$

$$\sum I_x^2 = \begin{array}{|c|c|c|}\hline 2 & 6 & 5 \\ \hline 11 & 16 & 15 \\ \hline 10 & 11 & 11 \\ \hline \end{array} \qquad \sum I_y^2 = \begin{array}{|c|c|c|}\hline 5 & 14 & 10 \\ \hline 9 & 19 & 15 \\ \hline 4 & 5 & 5 \\ \hline \end{array} \qquad \sum I_xI_y = \begin{array}{|c|c|c|}\hline 2 & 8 & 6 \\ \hline 8 & 15 & 13 \\ \hline 6 & 7 & 7 \\ \hline \end{array}$$

$$R = \left[\sum I_x^2 \sum I_y^2 - (\sum I_x I_y)^2\right] - k\left[\sum I_x^2 + \sum I_y^2\right]^2$$

Hence,

$$R = \begin{array}{|c|c|c|}\hline 3.55 & 0 & 2.75 \\ \hline 15.00 & 17.75 & 11.00 \\ \hline -5.80 & -6.80 & -6.80 \\ \hline \end{array}$$

Marking scheme

4 marks.

e. For the Harris corner detector, describe what type of image feature will give rise to the following values of R.
(i) $R \approx 0$
(ii) $R < 0$
(iii) $R > 0$.

[6 marks]

Answer

(i) $R \approx 0$ occurs where intensity values are unchanging
(ii) $R < 0$ occurs at edges
(iii) $R > 0$ occurs at corners.

Marking scheme

2 for each correct answer.

f. Briefly explain what role the Harris corner detector might play in solving the stereo correspondence problem.

[3 marks]

Answer

The Harris corner detector can be used to locate interest points in both images which will be matched, *i.e.*, it can be used as a detector in a feature-based method of solving the stereo correspondence problem.

Marking scheme

3 marks.

6. a. A simple object recognition system encodes objects using 4-element feature vectors. Four objects from two different classes (A and B) are encoded as follows:

Object	Class	Feature Vector
1	A	(2, 2, 0, 1)
2	A	(3, 3, 1, 1)
3	B	(1, 2, 3, 2)
4	B	(2, 1, 3, 0)

A new object, of unknown class, has a feature vector (1, 1, 1, 1). Using Euclidean distance as the similarity measure, determine the classification of the new object using:

i. a nearest mean classifier,

[5 marks]

Answer

Prototype of class A = $\left(\frac{2+3}{2}, \frac{2+3}{2}, \frac{0+1}{2}, \frac{1+1}{2}\right) = (2.5, 2.5, 0.5, 1)$
Prototype of class B = $\left(\frac{1+2}{2}, \frac{2+1}{2}, \frac{3+3}{2}, \frac{2+0}{2},\right) = (1.5, 1.5, 3, 1)$
Distance of new object from prototypes:
From prototype of class A: $\sqrt{(1-2.5)^2+(1-2.5)^2+(1-0.5)^2+(1-1)^2} = 2.18$
From prototype of class B: $\sqrt{(1-1.5)^2+(1-1.5)^2+(1-3)^2+(1-1)^2} = 2.12$
Hence, new object is class B.

Marking scheme

5 marks.

ii. a nearest neighbour classifier,

[5 marks]

Answer

Distance of new object from exemplars:

Obj.	Class	Features	Distance to (2,2,2,2)
1	A	(2,3,0,1)	$\sqrt{(1-2)^2+(1-2)^2+(1-0)^2+(1-1)^2} = 1.73$
2	A	(3,3,1,1)	$\sqrt{(1-3)^2+(1-3)^2+(1-1)^2+(1-1)^2} = 2.83$
3	B	(1,2,3,2)	$\sqrt{(1-1)^2+(1-2)^2+(1-3)^2+(1-2)^2} = 2.45$
4	B	(2,1,3,0)	$\sqrt{(1-2)^2+(1-1)^2+(1-3)^2+(1-0)^2} = 2.45$

The closest exemplar is object 1.

Since object 1 is of class A, the new object is also class A.

Marking scheme

5 marks.

iii. a k-nearest neighbour classifier, with k=3.

[4 marks]

Answer

The three closest exemplars are 1, 3, and 4.

Object 1 is in class A;

Objects 2 and 3 are in class B.

The majority are class B, so the new object is classified as B.

Marking scheme

4 marks.

b. Describe briefly in words, or using pseudo-code, each step required to train a bag-of-words object recognition system, and then apply it to classifying a new image.

[6 marks]

Answer

Training:

- Extract local feature descriptors around all chosen interest points in all training images.
- Apply k-means clustering to feature descriptors in order to generate a codeword dictionary.
- Remove codewords from dictionary that occur in most training images.

Classifying:

- Encode new image using a histogram showing the frequency of appearance of each codeword in that image.
- Compare histogram to those generated for each training image to find best match.

Marking scheme

6 marks.

c. In a simple bag-of-words object recognition system classes are represented by histograms showing the number of occurrences of 4 "codewords". The number of occurrences of the codewords in two training images are given below:

ClassA = (2, 3.5, 0.5, 2)

ClassB = (0.5, 0.75, 3.5, 1)

A new image is encoded as follows:

New = (2, 1, 2, 1)

Determine the training image that best matches the new image by finding the cosine of the angle between the codeword vectors.

[5 marks]

Answer

Similarity $= cos(\theta) = \frac{\sum_i A(i)N(i)}{\sqrt{\sum_i A(i)^2}\sqrt{\sum_i N(i)^2}}$ (*i.e.*, the normalised cross-correlation).

Similarity between New and ClassA is:

$$\frac{(2 \times 2) + (3.5 \times 1) + (0.5 \times 2) + (2 \times 1)}{\sqrt{2^2 + 3.5^2 + 0.5^2 + 2^2}\sqrt{2^2 + 1^2 + 2^2 + 1^2}} = 0.733$$

Similarity between New and ClassB is:

$$\frac{(0.5 \times 2) + (0.75 \times 1) + (3.5 \times 2) + (1 \times 1)}{\sqrt{0.5^2 + 0.75^2 + 3.5^2 + 1^2}\sqrt{2^2 + 1^2 + 2^2 + 1^2}} = 0.822$$

Hence, the new image is most similar to Class B.

Marking scheme

5 marks.