—— SOLUTIONS ——

# King's College London



**Degree Programmes** MSc, MSci

**Module Code** 7CCSMCVI
**Module Title** Computer Vision
**Examination Period** January 2015 (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. image processing

ii. epipolar constraint

iii. hyper-column

[6 marks]

Answer

i) image processing = signal processing applied to an image, with another image as the resulting output
ii) The epipolar constraint can be used to reduce the search space for possible matches when solving the correspondence problem: it uses geometrical information to reduce the search space to a line
iii) A hyper-column is a region of V1 that contains neurons covering the full range of RF types for a single spatial location.

Marking scheme

2 marks for each correct definition.

**b.** Below is shown a convolution mask, $H$ and an image, $I$.

$$H = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \qquad I = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Calculate the result of convolving this mask with this image, producing a result that is the same size as $I$.

[5 marks]

Answer

Rotated mask$= \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

$$H * I = \begin{bmatrix} 1.5 & 0.5 & 0 \\ 2.5 & 1 & 1 \\ 3 & 2 & 1 \end{bmatrix}$$

Marking scheme

5 marks

**c.** Briefly describe what is meant by a "well-posed, forward problem", and give a Computer Vision related example.

[5 marks]

Answer

A forward problem is one where we know the causes and want to predict or model the outcomes.

A well-posed problem is one which has one, unique, solution.

An example is a model of image formation.

Marking scheme

2 marks each for knowing what "well posed" and "forward" mean. 1 for example.

**d.** Briefly explain why vision is an "ill-posed, inverse problem".

[4 marks]

Answer

An inverse problem is one where we know the outcomes and want to infer the causes.

An ill-posed problem has multiple solutions (or no solution).

Vision is an inverse problem as we know the pixel intensities (the outcomes) and want to infer the causes (i.e. the objects in the scene, *etc.*). It is ill-posed as there are usually multiple solutions (i.e. multiple causes that could give rise to the same outcomes).

Marking scheme

4 marks.

**e.** A successful object recognition system needs to be sensitive to differences between images that are relevant to distinguishing one object (or category of object) from another, but needs to be insensitive (or tolerant) to other differences that are not relevant to object identity or category. List five causes of image changes that are not relevant to object recognition.

[5 marks]

Answer

- viewpoint (scale, rotation, translation)
- lighting
- non-rigid deformations
- within category variations
- clutter/occlusion

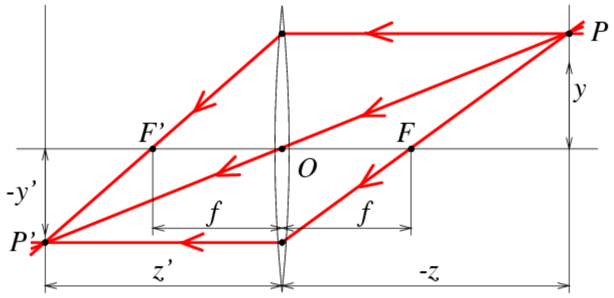
Marking scheme

1 mark each

**2.** **a.** Draw a cross-sectional diagram showing how a lens forms an image (P') of a point (P). Ensure that you label the optical centre (O), the focal point (F), and the coordinates of the world point (y,z) and the image point (y',z').

[5 marks]

Answer




Marking scheme

5 marks.

**b.** Derive the equation for the pinhole camera model of image formation relating the coordinates of a 3D point P(y,z) to the coordinates of its image P'(y',f'). Note that in the pinhole camera model the image plane is located at distance f' from the optical centre.

[4 marks]

Answer

From similar triangles:

$$\frac{y'}{z'} = \frac{y}{z} \qquad \Longrightarrow y' = \frac{z'y}{z}$$

substituting $z' = f'$ gives: $y' = \frac{f'y}{z}$

Marking scheme

4 marks.

**c.** Hence, write in matrix form the equation relating the image coordinates (x',y',z') to the world coordinates (x,y,z), where both sets of coordinates are measured relative to the optical centre of the camera.

[3 marks]

Answer

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \frac{f'}{z} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

Marking scheme

3 marks

**d.** Use the pinhole camera model to calculate the coordinates (x',y') of the image of a point in 3D space which has coordinates (0.4,0.5,2) measured, in metres, relative to the optical centre of the camera. Assume that the lens has a focal length of 35mm.

[3 marks]

Answer

$$x' = \frac{f'x}{z} \qquad \Longrightarrow x' = \frac{35 * 400}{2000} = 7mm$$
$$y' = \frac{f'y}{z} \qquad \Longrightarrow y' = \frac{35 * 500}{2000} = 8.75mm$$

Marking scheme

1 mark each, plus 1 additional mark for getting the units correct.

**e.** The camera moves so that the point in 3D space now has coordinates (0.6,0.5,2) measured, in metres, relative to the optical centre of the camera. Use the pinhole camera model to calculate the new coordinates of the image of this point.

[2 marks]

Answer

$$x' = \frac{f'x}{z} \qquad \Longrightarrow x' = \frac{35 * 600}{2000} = 10.5mm$$
$$y' = \frac{f'y}{z} \qquad \Longrightarrow y' = \frac{35 * 500}{2000} = 8.75mm$$

Marking scheme

2 marks

**f.** Between the situation described in part 2.d and part 2.e the camera moved at $-1.6ms^{-1}$ along its x-axis and the two images were taken $0.125s$ apart. Given this knowledge, and the *image* coordinates calculated in answer to part 2.d and part 2.e, calculate the depth of the point in 3D space relative to the optical centre of the camera. Assume that the lens has a focal length of 35mm.

[4 marks]

Answer

The depth is given by: $Z = -\frac{fV_x}{\dot{x}}$.

The velocity of the image point is $\dot{x} = \frac{10.5-7}{0.125} = 28mm/s$

Hence, the depth is $Z = -\frac{0.035\times-1.6}{0.028} = 2m$.

Marking scheme

2 marks for equation, 2 marks for correct application.

**g.** The arrays below show the values of the red, green and blue pixels in a Bayer masked image. Calculate the RGB values for the pixel at coordinates (2,2) in the demosaiced image using bi-linear interpolation.

$R =$

| | | | |
|---|---|---|---|
| 0.9 | | 0.3 | |
| | | | |
| 0.4 | | 0.3 | |
| | | | |

$G =$

| | | | |
|---|---|---|---|
| | 0.7 | | 0.6 |
| 0.1 | | 0.8 | |
| | 0.8 | | 0.9 |
| 0.4 | | 0.1 | |

$B =$

| | | | |
|---|---|---|---|
| | | | |
| | 0.2 | | 0.3 |
| | | | |
| | 0.4 | | 0.6 |

[4 marks]

Answer

$R(2,2) = \frac{R(1,1)+R(1,3)+R(3,1)+R(3,3)}{4} = \frac{0.9+0.3+0.4+0.3}{4} = 0.475$

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

$B(2,2) = 0.2$ as given.

Marking scheme

4 marks

**3.** **a.** Convolve the mask $\begin{bmatrix} -1 & 1 \end{bmatrix}$ with itself to produce a 1 by 3 pixel output.

[4 marks]

Answer

The "image" padded with zeros is $\begin{bmatrix} 0 & -1 & 1 & 0 \end{bmatrix}$ and the rotated mask is $\begin{bmatrix} 1 & -1 \end{bmatrix}$.

Hence, the result is:

$$\begin{bmatrix} (0\times 1 + -1 \times -1) & (-1\times 1 + 1 \times -1) & (1 \times 1 + 0 \times -1) \end{bmatrix} = \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}$$

Marking scheme

4 marks

**b.** 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.

**c.** Use the answer to part 3.b to calculate a 2 by 3 pixel numerical approximation to the first derivative of a Gaussian in the x-direction.

[4 marks]

Answer

$$\begin{bmatrix} 0.01 & 0.09 & 0.01 \\ 0.09 & 0.64 & 0.09 \\ 0.01 & 0.09 & 0.01 \end{bmatrix} * \begin{bmatrix} -1 & 1 \end{bmatrix} = \begin{bmatrix} -0.08 & 0.08 \\ -0.55 & 0.55 \\ -0.08 & 0.08 \end{bmatrix}$$

Marking scheme

4 marks

**d.** Derivatives of Gaussian masks (in the x and y directions) are used by the Canny edge detector. Describe briefly in words, or using pseudo-code, each step performed by the Canny edge detection algorithm.

[6 marks]

Answer

1. convolve the image with each derivative of Gaussian mask, to generate $I_x$ and $I_y$.
2. calculate the magnitude and direction of the intensity gradient ($M = \sqrt{I_x^2 + I_y^2}, D = tan^{-1}\left(\frac{I_y}{I_x}\right)$).
3. perform non-maximum suppression (thin multi-pixel wide edges down to a single pixel by setting M to zero for all pixels that have a neighbour, perpendicular to the direction of the edge, with a higher magnitude).
4. perform hysteresis thresholding (pixels above high thresholds set to one, pixels below low threshold set to zero, pixels with values between low and high thresholds set to one if they are connected to a pixel with a magnitude over the high threshold, and set to zero otherwise).

Marking scheme

6 marks

**e.** 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.

**f.** Describe how the responses of 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 hypercolumns.

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

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

[6 marks]

<u>Answer</u>

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

<u>Marking scheme</u>

6 marks

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

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

**c.** 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:

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

Calculate SAD values to unlabelled pixels from the top-right pixel:

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

Region 1 cannot grow further. Choose another seed pixel: middle-left:

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

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

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

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

$$\begin{bmatrix} 1 & 1 \\ & \\ (10,20,15):2 & (5,5,20) \\ & \mathbf{35}\uparrow \\ (15,15,15):2 & (20,15,10):2 \end{bmatrix}$$

Region 2 cannot grow further. Last pixel has no unlabelled neighbours.

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

<u>Marking scheme</u>

6 marks

**d.** An alternative segmentation algorithm is region merging. Write pseudo-code for the region merging algorithm.

[6 marks]

Answer

1. Give every pixel a unique region label
2. Randomly choose a region which has not been marked final
    3. Compare the chosen region's properties with those of its neighbours
    4. For all regions with properties which are within the similarity threshold
        5. Re-label them with the region label of the chosen region
        6. Calculate properties of the merged region as mean of all its constituents
    7. If no regions can be merged with chosen region, mark it as final
8. Repeat from step 2 until all image regions are marked final.

Marking scheme

6 marks

**e.** Apply the region merging 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 regions are similar is that the SAD is less than 12, (iii) the first chosen region is the top-left corner, (iv) regions have horizontal, vertical and diagonal neighbours.

[6 marks]

<u>Answer</u>

Give each pixel a unique label. Calculate SAD values for regions neighbouring the top-left region:

$$\begin{bmatrix} (20,10,5):1 & \mathbf{10} \rightarrow & (15,5,5):2 \\ \mathbf{30} \downarrow & \mathbf{35} \searrow & \\ (10,20,15):3 & & (5,5,20):4 \\ & & \\ (15,15,15):5 & & (20,15,10):6 \end{bmatrix}$$

Merge similar regions, and calculate properties of the merged region:

$$\begin{bmatrix} (17.5,7.5,5):1 & (17.5,7.5,5):1 \\ & \\ (10,20,15):3 & (5,5,20):4 \\ & \\ (15,15,15):5 & (20,15,10):6 \end{bmatrix}$$

Randomly choose region 1, and calculate SAD values for neighbouring regions:

$$\begin{bmatrix} (17.5,7.5,5):1 & (17.5,7.5,5):1 \\ \mathbf{30} \downarrow & \mathbf{30} \downarrow \\ (10,20,15):3 & (5,5,20):4 \\ & \\ (15,15,15):5 & (20,15,10):6 \end{bmatrix}$$

Mark region 1 as final. Randomly choose region 3 and calculate SAD values for neighbouring regions:

$$\begin{bmatrix} 1 & & 1 \\ \\ (10,20,15):3 & \mathbf{25}\rightarrow & (5,5,20):4 \\ \mathbf{10}\downarrow & \mathbf{20}\searrow & \\ (15,15,15):5 & & (20,15,10):6 \end{bmatrix}$$

Merge similar regions, and calculate properties of the merged region:

$$\begin{bmatrix} 1 & 1 \\ \\ (12.5,17.5,15):3 & (5,5,20):4 \\ \\ (12.5,17.5,15):3 & (20,15,10):6 \end{bmatrix}$$

Randomly choose region 3, and calculate SAD values for neighbouring regions:

$$\begin{bmatrix} 1 & & 1 \\ \\ (12.5,17.5,15):3 & \mathbf{25}\rightarrow & (5,5,20):4 \\ \\ (12.5,17.5,15):3 & \mathbf{15}\rightarrow & (20,15,10):6 \end{bmatrix}$$

Mark region 3 as final. Randomly choose region 4 and calculate SAD values for neighbouring regions:

$$\begin{bmatrix} 1 & 1 \\ \\ 3 & (5,5,20):4 \\ & \mathbf{35}\downarrow \\ 3 & (20,15,10):6 \end{bmatrix}$$

Mark region 4 as final. Last pixel has no unlabelled neighbours. So final region labels: $\begin{bmatrix} 1 & 1 \\ 3 & 4 \\ 3 & 6 \end{bmatrix}$

Marking scheme

6 marks.

**5. a.** Describe what is meant by the "correspondence problem" and briefly describe three scenarios which might require a solution to this problem.

[6 marks]

Answer

The correspondence problem is the problem of finding the same 3D location in two or more images.

Marking scheme

3 marks

This problem might arise:

1. When using multiple cameras to obtain two, or more, images of the same scene from different locations - solving the correspondence problem would enable the 3D structure of the scene to be recovered.

2. When using a single camera to obtain two, or more, images of the same scene at different times - solving the correspondence problem would enable estimation of camera and object motion.

3. When comparing an image with one or more stored images - solving the correspondence problem would enable the similarity between the images to be determined and hence allow object recognition.

Marking scheme

1 mark each.

**b.** Two identical cameras are mounted so as to have coplanar image planes and such that one camera (the right one) is displaced distance B along the x-axis of the other (the left one). Write down the equation for the depth of a 3D point P visible in both cameras.

[4 marks]

Answer

$$Z = \frac{fB}{x_L - x_R}$$

Marking scheme

4 marks

**c.** For the stereo vision system described in question 5.b, comment on the accuracy with which the depth of a 3D point can be measured as the baseline distance (B) is increased.

[2 marks]

Answer

As the baseline, B, increases, the disparity increases. Hence, accuracy increases.

Marking scheme

2 marks

**d.** The two arrays below show the intensity values for each pixel in a pair of 5 by 3 pixel images which have been taken with the stereo vision system described in question 5.b.

$I_L =$

| | | | | |
|---|---|---|---|---|
| 40 | 60 | 40 | 20 | 50 |
| 10 | 50 | 80 | 80 | 30 |
| 70 | 10 | 70 | 60 | 90 |

$I_R =$

| | | | | |
|---|---|---|---|---|
| 70 | 50 | 10 | 20 | 50 |
| 40 | 80 | 80 | 10 | 50 |
| 20 | 80 | 70 | 80 | 70 |

Calculate the similarity of the pixel at coordinates (3,2) in $I_L$ (*i.e.*, the central pixel), to locations in $I_R$, 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.

[5 marks]

Answer

SAD =

| | | | | |
|---|---|---|---|---|
| 290 | 70 | 230 | 310 | 350 |

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

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

Marking scheme

5 marks.

**e.** Using the answers to part 5.b and part 5.d calculate the depth of the point shown at the pixel with coordinates (3,2) in $I_L$. Assume that the baseline distance (B) is 0.5m, the two camera lenses have focal lengths of 30mm, and the pixel size of each camera is 0.1mm/pixel.

[4 marks]

Answer

disparity=1 pixel = 0.1mm

$$Z = \frac{fB}{x_L - x_R}$$

$$Z = \frac{0.030 \times 0.5}{0.1 \times 10^{-3}} = 150m$$

Marking scheme

4 marks

**f.** An alternative way to solve the stereo correspondence problem is to use a feature-based method. Briefly explain what is meant by a "detector" and a "descriptor" in a feature-based method.

[4 marks]

Answer

The detector is the method used to locate image features (or interest points) which are suitable for matching.

The descriptor is an array of feature values associated with each interest point. These descriptors are compared to determine which points match.

Marking scheme

2 marks for each correct definition.

**6.** In a simple feature-matching object recognition system each keypoint has x,y-coordinates and a 3 element feature vector. Two training images, one of object A and the other of object B, have been processed to create a database of known objects, as shown below:

| Object | Keypoint Number | Coordinates (pixels) | Feature Vector |
|---|---|---|---|
| A | A1 | (25, 2) | (1, 6, 10) |
| | A2 | (40, 25) | (2, 9, 3) |
| | A3 | (15, 37) | (7, 8, 15) |
| B | B1 | (14, 11) | (6, 1, 12) |
| | B2 | (30, 45) | (3, 8, 4) |
| | B3 | (24, 6) | (13, 4, 8) |

The keypoints and feature vectors extracted from a new image are as follows:

| Keypoint Number | Coordinates (pixels) | Feature Vector |
|---|---|---|
| N1 | (21, 47) | (5, 8, 15) |
| N2 | (30, 11) | (2, 6, 11) |
| N3 | (24, 32) | (12, 3, 8) |
| N4 | (34, 46) | (5, 2, 11) |
| N5 | (44, 37) | (2, 8, 3) |

**a.** Perform feature matching using the sum of absolute differences as the distance measure and applying the following criterion for accepting a match: that the ratio of distance to first nearest descriptor to that of second is less than 0.4.

[7 marks]

<u>Answer</u>

SAD distance between each keypoint in the training image database and each keypoint in the new image:

| Feature Vectors | N1(5,8,15) | N2(2,6,11) | N3(12,3,8) | N4(5,2,11) | N5(2,8,3) |
|---|---|---|---|---|---|
| A1(1,6,10) | 11 | 2 | 16 | 9 | 10 |
| A2(2,9,3) | 16 | 11 | 21 | 18 | 1 |
| A3(7,8,15) | 2 | 11 | 17 | 12 | 17 |
| B1(6,1,12) | 11 | 10 | 12 | 3 | 20 |
| B2(3,8,4) | 13 | 10 | 18 | 15 | 2 |
| B3(13,4,8) | 19 | 16 | 2 | 13 | 20 |
| Ratio 1st to 2nd best: | 2/11 = 0.18 | 2/10 =0.2 | 2/12=0.17 | 3/9=0.33 | 1/2=0.5 |

Hence, the following matches are found:

N1 - A3

N2 - A1

N3 - B3

N4 - B1

Marking scheme

7 marks

**b.** Write pseudo-code for the RANSAC algorithm.

[5 marks]

Answer

1. Randomly choose a minimal subset (a sample) of data points necessary to fit the model
2. Fit the model to this subset of data
3. Test all the other data points to determine if they are consistent with the fitted model (i.e. if they lie within a distance t of the model's prediction).
4. Count the number of inliers (the consensus set). Size of consensus set is model's support
5. Repeat from step 1 for N trials

After N trials select the model parameters with the highest support and re-estimate the model using all the points in this subset.

Marking scheme

5 marks

**c.** It is known that the object in the new image is related by a pure translation in the image plane to the images of both the known objects. Hence, use the RANSAC algorithm to assess the consistency of the matched points and so determine which of the two training objects is present in the new image. Apply RANSAC exhaustively to all matches, rather than to a subset of matches chosen at random, and assume the threshold for comparing the model's prediction with the data is 3 pixels.

[9 marks]

Answer

Choose N1. Model is a translation of (21,47)-(15,37) = (6,10)
Locations matching points predicted by the model are:
For N2: (30,11)-(6,10)=(24,1)
actual match is at (25,2), hence, this is an inlier for this model.
Hence, consensus set =1.

Marking scheme

3

Choosing N2, will also predict a translation consistent with N1.

Marking scheme

1

Choose N3. Model is a translation of (24,32)-(24,6) = (0,26)
Locations matching points predicted by the model are:
For N4: (34,46)- (0,26) = (34,20)
actual match is at (14,11), hence, this is an outlier for this model.
Hence, consensus set =0.

Marking scheme

3

Choosing N4, will predict a translation inconsistent with N3.

Marking scheme

1

The new image contains object A, as this is the only object for which the locations of the matching keypoints are consistent.

Marking scheme

1

**d.** A number of other standard metrics exist that could be used to compare feature vectors. Write down the formulae for (i) cross-correlation, and (ii) correlation coefficient, for comparing two vectors $a$ and $b$.

[4 marks]

Answer

(i) cross-correlation $= \sum_i a_i b_i$
(ii) correlation coefficient $= \frac{\sum_i (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_i (a_i - \bar{a})^2} \sqrt{\sum_i (b_i - \bar{b})^2}}$

Marking scheme

2 marks for each correct equation