—— 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 2014 (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

—— SOLUTIONS ——

1. Compulsory Question

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

i. computer vision

ii. exposure

iii. epipolar line

[6 marks]

Answer

i) computer vision = extracting information about the world from images.
ii) exposure = the time needed to allow sufficient light to reach the sensor in order to form an image.
iii) epipolar line = the image in one camera of the line joining a point in 3d space and the optical centre of the other camera in a stereo pair. Equivalently, the line where the epipolar plane (the plane formed by the 3d point and the two optical centres) intersects the image plane.

Marking scheme

2 marks for each correct definition.

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

$$I = \begin{bmatrix} 3 & 6 & 21 \\ 6 & 18 & 9 \\ 3 & 9 & 18 \end{bmatrix}$$

Calculate the result of convolving this image with a 3-by-3 pixel box mask (or mean filter). Pad the image with zeros where necessary to generate a 3-by-3 pixel filtered image.

[4 marks]

Answer

A 3-by-3 pixel box mask =

$$\begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix}$$

Filtered image =

$$\begin{bmatrix} (3+6+6+18)/9 & (3+6+21+6+18+9)/9 & (6+21+18+9)/9 \\ (3+6+6+18+3+9)/9 & (3+6+21+6+18+9+3+9+18)/9 & (6+21+18+9+9+18)/9 \\ (6+18+3+9)/9 & (6+18+9+3+9+18)/9 & (18+9+9+18)/9 \end{bmatrix}$$

$$= \begin{bmatrix} 33/9 & 63/9 & 54/9 \\ 45/9 & 93/9 & 81/9 \\ 36/9 & 63/9 & 54/9 \end{bmatrix} = \begin{bmatrix} 3.67 & 7 & 6 \\ 5 & 10.33 & 9 \\ 4 & 7 & 6 \end{bmatrix}$$

Marking scheme

2 marks for knowing a correctly normalised box mask. 2 marks for correctly applying convolution to the image.

c. Convert image I, as defined in question 1.b, to a binary image by applying a threshold of 8.

[3 marks]

Answer

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

Marking scheme

3 marks

d. List four scene properties that might cause an intensity discontinuity (or edge) in an image.

[4 marks]

Answer

- Depth discontinuity (due to surfaces at different distances)
- Orientation discontinuity (due to changes in the orientation of a surface)
- Reflectance discontinuity (due to change in surface material properties)
- Illumination discontinuity (shadows)

Marking scheme

One mark each.

e. List eight cues to depth.

[8 marks]

Answer

Any eight from:

- Stereo Disparity
- Accommodation
- Convergence
- Interposition/Occlusion
- Size familiarity
- Texture gradients
- Linear perspective
- Aerial perspective
- Shading
- Motion parallax
- Optic Flow
- Accretion and deletion
- Structure from motion (kinetic depth)

Marking scheme

One mark each.

2. a. Write down the thin lens equation, which relates the focal length of a lens to the depths of the image and object. Be sure to define each term in this equation.

[4 marks]

Answer

$\frac{1}{f} = \frac{1}{\|z\|} + \frac{1}{\|z'\|}$

where:

f = focal length of lens

z = distance of object from the lens

z' = distance of image from the lens

Marking scheme

2 marks for equation, 2 marks for correct definition of terms.

b. If a lens has a focal length of 50mm at what depth should the image plane be placed to bring an object 6m from the camera into focus?

[3 marks]

Answer

$\frac{1}{f} = \frac{1}{\|z\|} + \frac{1}{\|z'\|}$

$\frac{1}{50} = \frac{1}{\|6000\|} + \frac{1}{\|z'\|}$

$\frac{1}{\|z'\|} = \frac{1}{50} - \frac{1}{\|6000\|}$

$z' = 50.42mm$

Marking scheme

3 marks

c. For the situation described in question 2.b, where should the image plane be placed to form a focused image if the object is moved to be 3m from the camera?

[1 marks]

Answer

$\frac{1}{\|z'\|} = \frac{1}{50} - \frac{1}{\|3000\|}$

$z' = 50.85mm$

Marking scheme

1 mark

d. For the pinhole camera model of image formation, what would be the effect on the size of the image of decreasing the relative distance between the camera and the object from 6m to 3m?

[2 marks]

Answer

Halving the distance would double the size of the image.

Marking scheme

2 marks.

e. Briefly compare the mechanisms used for sampling the image in a camera and in an eye.

[6 marks]

Answer

Camera:

- Has sensing elements sensitive to three wavelengths (RGB).
- Sensing elements occur in a fixed ratio across the whole image plane.
- The sampling density is uniform across the whole image plane.

Eye:

- Has sensing elements sensitive to four wavelengths (RGBW).
- Sensing elements occur in a variable ratios across the image plane (cone density highest at fovea, rod density highest outside fovea).
- The sampling density is non-uniform across the image plane (density is highest at the fovea).

Marking scheme

3 marks for each part.

f. Briefly describe how a colour image is sampled by a camera using

i. Three CCD arrays,

ii. One CCD array.

[4 marks]

Answer

i. Three CCD arrays = image is split using a prism and separate arrays are used to detect the red, green, and blue wavelengths separately.
ii. One CCD array = CCD array is covered with an array of filters so that different pixels only receive light that is red, or green, or blue.

Marking scheme

2 marks for each part.

g. In order to locate red regions of the image it has been decided to model the response properties of a red on/green off (R+/G-) colour-opponent retinal ganglion cell. A simple difference of Gaussians model is to be used. Design a MATLAB function $R = f(I)$ that will calculate the response of a R+/G- cell at all locations in the image I. Assume I is a colour image in RGB format and that you have access to a function $g = gaussian(sigma, d)$ that returns a d-by-d pixel Gaussian filter with standard deviation sigma. Use a Gaussian with standard deviation 2 for the centre, and a Gaussian with standard deviation 3 for the surround.

[5 marks]

Answer

```
function R = f(I)
onFilter=gaussian(2,10);
offFilter=gaussian(3,15);
R=conv2(I(:,:,1),onFilter,'same')...
 -conv2(I(:,:,2),offFilter,'same');
```

Marking scheme

1 marks for knowing how to access red and green channels of image, 2 marks for knowing that colour channels should be convolved with Gaussians of the correct size, 2 marks for knowing result is difference between convolutions.

3. a. Convolution masks can be used to provide a finite difference approximation to first and second order directional derivatives. Write down the masks that approximate the following directional derivatives:

i. $-\frac{\delta}{\delta x}$

ii. $-\frac{\delta^2}{\delta x^2}$

iii. $-\frac{\delta^2}{\delta x^2} - \frac{\delta^2}{\delta y^2}$

[6 marks]

Answer

(i) $-\frac{\delta}{\delta x} \approx \begin{bmatrix} -1 & 1 \end{bmatrix}$

(ii) $-\frac{\delta^2}{\delta x^2} \approx \begin{bmatrix} -1 & 2 & -1 \end{bmatrix}$

(iii) $-\frac{\delta^2}{\delta x^2} - \frac{\delta^2}{\delta y^2} \approx \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} or \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$

Marking scheme

2 marks for each correct definition.

b. For edge detection, a difference mask is usually "combined" with a smoothing mask.

i. How are these masks "combined"?

ii. Why is this advantageous for edge detection?

iii. Name the mathematical function that is usually used to define the smoothing mask.

[7 marks]

Answer

(i) Masks are combined using convolution.

Marking scheme

2 marks

(ii) A difference mask is sensitive to noise as well as other intensity-level discontinuities.

A smoothing mask suppresses 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

3 marks

(iii) Gaussian.

Marking scheme

2 marks

c. An image, I, is to be convolved with a 1-dimensional mask, h_1, and the result is to be convolved with another 1-dimensional mask, h_2. Calculate the single, 2-dimensional, mask H that would produce the same result, when: $h_1 = [-1 \quad 1 \quad -1]$ and $h_2 = [1 \quad 2 \quad 1]^T$.

[4 marks]

Answer

$$M = h_1 * h_2 = h_2 \times h_1 = \begin{bmatrix} -1 & 1 & -1 \\ -2 & 2 & -2 \\ -1 & 1 & -1 \end{bmatrix}$$

Marking scheme

4 marks.

d. What is the advantage of sequentially performing two convolutions with 1D masks, rather than a single convolution with a 2D mask, to obtain the same result?

[2 marks]

Answer

It is computationally more efficient. Each element of the result is generated by 3 multiplications + 3 multiplications by the 1st method, but by 9 multiplications with the second method.

Marking scheme

2 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

4. a. Briefly describe what is meant by "grouping and segmentation" in mid-level vision.

[4 marks]

Answer

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

Marking scheme

4 marks.

b. One method of image segmentation is the split and merge algorithm. Write pseudo-code for the split and merge algorithm.

[7 marks]

Answer

1. Label every pixel with the same region label
2. For each region
    3. If all pixels are not similar
        4. Assign the four quadrants different region labels
5. Repeat from 2 until all regions homogeneous
6. For each region
    7. Compare the region's properties with those of its neighbours
    8. For all regions with properties which are within the similarity threshold
        9. Re-label them with the region label of the chosen region
        10. Calculate properties of the merged region as mean of all its constituents
11. Repeat from step 6 until no more regions can be merged.

Marking scheme

7 marks

c. The array below shows feature vectors for each pixel in a 2-by-2 pixel image.

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

Apply the split and merging algorithm to assign each pixel to a region. Assume that (1) the method used to assess similarity is the sum of absolute differences (SAD), (2) the criteria for deciding if regions are similar is that the SAD is less than 12, (3) regions have horizontal, vertical and diagonal neighbours.

[6 marks]

Answer

Give each pixel the same label. Determine if SAD value for any pair of pixels within the the region exceeds 12:

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

Split the region into quadrants:

$$\begin{bmatrix} (10,15,5):1 & (15,15,15):2 \\ & \\ (5,15,10):3 & (20,10,15):4 \end{bmatrix}$$

Can not split any further. So perform merge operation for region 1:

$$\begin{bmatrix} (10,15,5):1 & \mathbf{15} \rightarrow & (15,15,15):2 \\ \mathbf{10} \downarrow & \mathbf{25} \searrow & \\ (5,15,10):3 & & (20,10,15):4 \end{bmatrix}$$

Calculate properties of merged region:

$$\begin{bmatrix} (7.5,15,7.5):1 & (15,15,15):2 \\ & \\ (7.5,15,7.5):1 & (20,10,15):4 \end{bmatrix}$$

Continue merging region 1:

$$\begin{bmatrix} (7.5, 15, 7.5) : 1 & \mathbf{15} \rightarrow & (15, 15, 15) : 2 \\ & \mathbf{25} \searrow & \\ (10, 15, 10) : 1 & & (20, 10, 15) : 4 \end{bmatrix}$$

Region 1 can not be merged further, mark as final and try merging region2:

$$\begin{bmatrix} 1 & (15, 15, 15) : 2 \\ & \mathbf{10} \downarrow \\ 1 & (20, 10, 15) : 4 \end{bmatrix}$$

Merge regions 2 and 4. Can not merge any further. Final result: $\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$

Marking scheme

6 marks

d. In the Hough Transform, a point with coordinates (x,y) in the image-plane is mapped onto a curve in the (θ, r)-parameter space of the accumulator array. Write down the equation for r in terms of x, y, and θ.

[2 marks]

Answer

$r = ycos(\theta) - xsin(\theta)$

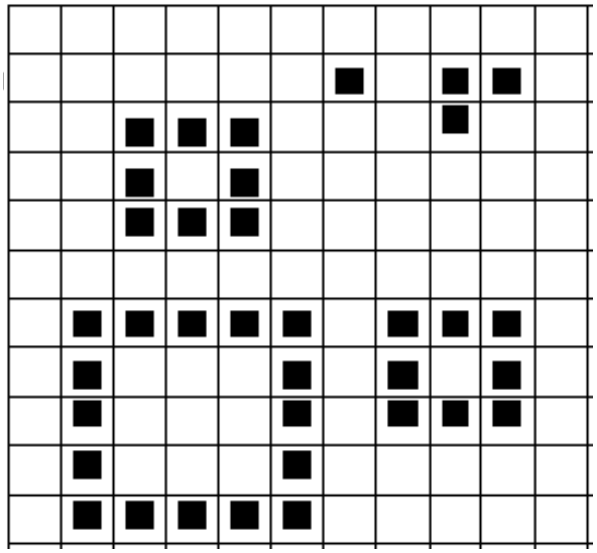
Marking scheme

2 marks

e. Below is an 11x11 pixel binary image in which the edge pixels are indicated by the black squares. If we apply the Hough transform to this image, using the following values for θ: [0,45,90,135],

i. what would be the maximum value in the accumulator array?

ii. what would be the (θ, r) coordinates of this location?



[4 marks]

Answer

i) 8

ii) (0,6)

Marking scheme

2 marks for each correct answer. The correct value of r requires knowing that the origin is at the top-left corner, as is the convention for images.

f. Briefly describe the two terms that form the energy function that an active contour (or "snake") attempts to minimise.

[2 marks]

Answer

Internal energy is a function of the shape of the contour, it is reduced if the curve is short and smooth.
External energy is a function of the image features near the contour, it is reduced if the intensity gradient is high.

Marking scheme

1 mark each.

5. Two images (A and B) of the same scene are taken from different viewpoints. Interest points are located in each image and a two element descriptor is calculated for each interest point. The image coordinates and feature vectors for each interest point are given in the table below:

Image A			Image B		
Point	Coordinates	Feature Values	Point	Coordinates	Feature Values
A1	(277, 168)	(10, 4)	B1	(111, 394)	(3, 7)
A2	(193, 290)	(3, 8)	B2	(125, 186)	(1, 1)
A3	(205, 87)	(0, 2)	B3	(145, 128)	(5, 7)
A4	(224, 28)	(6, 9)	B4	(90, 270)	(8, 0)

a. For each interest point in image A, find the best matching interest point in image B assuming that similarity is measured using the sum of absolute differences (SAD).

[8 marks]

Answer

For A1, SAD:

B1: 10; B2: 12; B3: 8; B4: 6

Therefore best match is B4.

For A2, SAD:

B1: 1; B2: 9; B3: 3; B4: 13

Therefore best match is B1

For A3, SAD:

B1: 8; B2: 2; B3: 10; B4: 10

Therefore best match is B2

For A4, SAD:

B1: 5; B2: 13; B3: 3; B4: 11

Therefore best match is B3

Marking scheme

8 marks

b. Write pseudo-code for the RANSAC algorithm.

[6 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

6 marks

c. Apply the RANSAC algorithm to find the true correspondence between the two images. Assume (1) that the images are related by a pure translation in the x-y plane, (2) that the threshold for comparing the model's prediction with the data is 20 pixels, (3) 3 trials are performed and these samples are chosen in the order A1, A2, A3 rather than being randomly chosen.

[11 marks]

Answer

Choose A1. Translation from B4 to A1 is (277,168)-(90, 270) = (187,-102). Hence, model is a translation of (187,-102).

Locations of matching points predicted by this model are:

For A2; (193, 290)-(187,-102) = (6, 392)

actual match is at (111, 394) hence this is an outlier for this model.

For A3; (205, 87)-(187,-102) = (18, 189)

actual match is at (125, 186) hence this is an outlier for this model.

For A4; (224, 28)-(187,-102) = (37, 130)

actual match is at (145, 128) hence this is an outlier for this model.

Hence, consensus set = 0.

Choose A2. Translation from B1 to A2 is (193, 290)-(111, 394) = (82, -104). Hence, model is a translation of (82, -104).

Locations of matching points predicted by this model are:

For A1; (277, 168)-(82, -104) = (195, 272)

actual match is at (90, 270) hence this is an outlier for this model.

For A3; (205, 87)-(82, -104) = (123, 191)

actual match is at (125, 186) hence this is an inlier for this model.

For A4; (224, 28)-(82, -104) = (142, 132)

actual match is at (145, 128) hence this is an inlier for this model.

Hence, consensus set = 2.

Choose A3. Translation from B2 to A3 is (205, 87)-(125, 186) = (80, -99). Hence, model is a translation of (80, -99).

Locations of matching points predicted by this model are:

For A1; (277, 168)-(80, -99) = (197, 267)

actual match is at (90, 270) hence this is an outlier for this model.

For A2; (193, 290)-(80, -99) = (113, 389)

actual match is at (111, 394) hence this is an inlier for this model.

For A4; (224, 28)-(80, -99) = (144, 127)

actual match is at (145, 128) hence this is an inlier for this model.

Hence, consensus set = 2.

Therefore the true correspondence is given by the matches for A2, A3, and A4. The best estimation of the model is $\frac{1}{3}[(82, -104)+(80, -99)+(79, -100)] = (80.33, -101)$

Marking scheme

9 marks for correct application of RANSAC. 2 marks for recalculating transformation using all inliers.

6. 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, 3, 0, 1)
2	A	(3, 3, 1, 3)
3	B	(1, 2, 3, 2)
4	B	(0, 1, 3, 0)

A new object, of unknown class, has a feature vector (2, 2, 2, 2). Using Euclidean distance as the similarity measure, determine the classification of the new object using:

a. a nearest mean classifier,

[5 marks]

Answer

Prototype of class A = $\left(\frac{2+3}{2}, \frac{3+3}{2}, \frac{0+1}{2}, \frac{1+3}{2}\right) = (2.5, 3, 0.5, 2)$

Prototype of class B = $\left(\frac{1+0}{2}, \frac{2+1}{2}, \frac{3+3}{2}, \frac{2+0}{2}, \right) = (0.5, 1.5, 3, 1)$

Distance of new object from prototypes:

From prototype of class A: $\sqrt{(2-2.5)^2 + (2-3)^2 + (2-0.5)^2 + (2-2)^2} = 1.87$

From prototype of class B: $\sqrt{(2-0.5)^2 + (2-1.5)^2 + (2-3)^2 + (2-1)^2} = 2.12$

Hence, new object is class A.

Marking scheme

5 marks

b. 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)	$\sqrt{(2-2)^2+(2-3)^2+(2-0)^2+(2-1)^2} = 2.45$
2	A	(3,4,1,3)	$\sqrt{(2-3)^2+(2-3)^2+(2-1)^2+(2-3)^2} = 2.0$
3	B	(1,2,4,2)	$\sqrt{(2-1)^2+(2-2)^2+(2-3)^2+(2-2)^2} = 1.41$
4	B	(0,1,3,0)	$\sqrt{(2-0)^2+(2-1)^2+(2-3)^2+(2-0)^2} = 3.16$

The closest exemplar is object 3.

Since object 3 is of class B, the new object is also class B.

Marking scheme

5 marks

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

[4 marks]

Answer

The three closest exemplars are 1, 2, and 3.

Objects 1 and 2 are class A.

Object 3 is class B.

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

Marking scheme

4 marks

d. 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 = (2.5, 3.5, 0.5, 2)
ClassB = (0.5, 1.5, 3.5, 1)
A new image is encoded as follows:
New = (1, 2, 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:

$$\frac{(2.5 \times 1) + (3.5 \times 2) + (0.5 \times 2) + (2 \times 1)}{\sqrt{2.5^2 + 3.5^2 + 0.5^2 + 2^2}\sqrt{1^2 + 2^2 + 2^2 + 1^2}} = 0.829$$

Similarity between New and ClassB is:

$$\frac{(0.5 \times 1) + (1.5 \times 2) + (3.5 \times 2) + (1 \times 1)}{\sqrt{0.5^2 + 1.5^2 + 3.5^2 + 1^2}\sqrt{1^2 + 2^2 + 2^2 + 1^2}} = 0.916$$

Hence, the new image is most similar to ClassB.

Marking scheme

5 marks

e. Describe briefly in words, or using pseudo-code, each step required to train the Implicit Shape Model (ISM) for object recognition.

[6 marks]

Answer

- For each image in the training dataset:
  - Locate interest points
  - Collect 2D image patches centred at each interest point
- Cluster image patches to form an appearance codebook
- For each codebook entry:
  - Match to training images
  - Record location of object centre relative to matched location

Marking scheme

6 marks