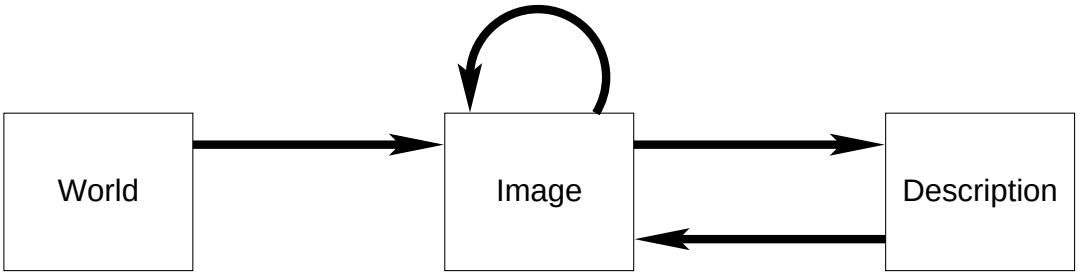
Introduction

1. Give a definition of “Computer Vision”.

2. The diagram below shows different representations of the same information, and possible transitions between them. Label each of the arrows with the name of the field of study or process that deals with the transition illustrated.



3. What type of information is extracted in each of the following applications:

- optical character recognition
- face tracking web-cam
- face recognition software / biometrics
- content based image retrieval
- driver assistance
- 3D modelling

4. Vision is often described as an “ill-posed, inverse problem”. Briefly describe what is meant by the terms “ill-posed” and “inverse problem” and their opposites “well-posed” and “forward problems”.

5. What is a “prior” and how does it help solve the ill-posed, inverse problem of vision? Give examples of priors.

Image Formation

1. What determines where a point in the 3D world appears on a 2D image?

2. What determines how bright the image of that point is?

3. Given the RGB values of a pixel, how can we tell the colour of the surface that is shown at that point in the image?

4. Briefly define what is meant by the terms “focus” and “exposure”.

5. Write down the thin lens equation, which relates the focal length of a lens to the depths of the image and object.

6. Derive the thin lens equation.

7. If a lens has a focal length of 35mm at what depth should the image plane be placed to bring an object 3m from the camera into focus? What if the object is at 0.5m?

8. If an object is at a distance of 3m from a *pinhole* camera, where should the image plane be placed to get a focused image?

9. Briefly compare the mechanisms used for focusing a camera and an eye.

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

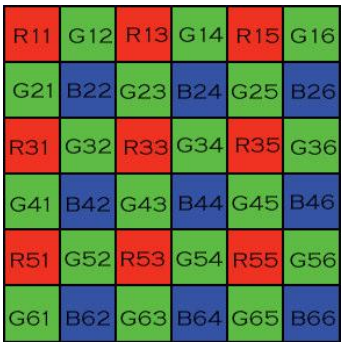
11. A point in 3D space has coordinates [10,10,500] mm relative to the camera reference frame. If the image principal point is at coordinates [244,180] pixels, and the magnification factors in the x and y directions are 925 and 740, then determine the location of the point in the image. Assume that the camera does not suffer from skew or any other defect.

How do the coordinates of the image change if (a) the world coordinates are [20,20,500] (i.e. object is twice as large), (b) the world coordinates are [10,10,250] (i.e. camera is half as far from the object), (c) the world coordinates are [10,10,500] but the CCD array doubles its resolution from 20 pixels/mm to 40 pixels/mm

12. The RGB channels for a small patch of image are shown below. Convert this image patch to a greyscale representation using an equal weighting for each channel and (a) 8 bits per pixel, (b) 2 bits per pixel.

$$I_R = \begin{bmatrix} 205 & 195 \\ 238 & 203 \end{bmatrix},\ I_G = \begin{bmatrix} 143 & 138 \\ 166 & 143 \end{bmatrix},\ I_B = \begin{bmatrix} 154 & 145 \\ 174 & 151 \end{bmatrix}$$

13. The image below shows the values of the red, green and blue pixels in a Bayer masked image. Calculate the RGB values for each pixel at the four central locations (*i.e.*, locations 33, 34, 43, and 44), using: a) bilinear interpolation; b) smooth hue transition interpolation; c) edge-directed interpolation.



14. Briefly describe the different characteristics of the fovea and periphery of the retina.

15. Describe the general receptive field structure of a retinal ganglion cell. Describe the full range of ganglion cell RFs found in the normally functioning human retina.

16. Briefly describe how Ganglion cell RFs give rise to: (a) efficient image coding, (b) invariance to illumination, (c) edge enhancement.

Low-Level Vision (Artificial)

1. Below is shown a convolution mask, H. Calculate the result of convolving this mask with (a) image I_1, and (b) image I_2.

$$H = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \qquad I_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad I_2 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

2. Calculate $H * I$, padding the image with zeros where necessary, when:

$$H = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad I = \begin{bmatrix} 0.25 & 1 & 0.8 \\ 0.75 & 1 & 1 \\ 0 & 1 & 0.4 \end{bmatrix}$$

3. Calculate $h * h^T$, where $h = [1, 0.5, 0.1]$. Show that this is equal to $h^T \times h$. Hence, calculate $H * I$, where:

$$H = \begin{bmatrix} 1 & 0.5 & 0.1 \\ 0.5 & 0.25 & 0.05 \\ 0.1 & 0.05 & 0.01 \end{bmatrix}, \qquad I = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

4. List the categories of image features that can produce intensity-level discontinuities in an image.

5. 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: (a) $-\frac{\delta}{\delta x}$, (b) $-\frac{\delta}{\delta y}$, (c) $-\frac{\delta^2}{\delta x^2}$, (d) $-\frac{\delta^2}{\delta y^2}$, (e) $-\frac{\delta^2}{\delta x^2} - \frac{\delta^2}{\delta y^2}$.

6. Convolve the mask $\begin{bmatrix} -1 & 1 \end{bmatrix}$ with itself.

7. Write down a mathematical expression describing the effect of convolving an image I with a Laplacian mask (i.e. $L = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$). Hence, write down a mathematical expression describing the effect of convolving an image I with the following mask: $L' = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix}$

8. For edge detection, a Laplacian mask is usually "combined" with a Gaussian mask to create a Laplacian of Gaussian (or LoG) mask. (a) How are these masks "combined"? (b) Why is this advantageous for edge detection? (c) What mathematical function is usually used to approximate a LoG mask?

9. To perform multiscale feature analysis, it would be possible to either (1) keep the image size fixed and vary the size of the mask, or (2) keep the mask size fixed and vary the size of the image. (a) Why is the latter preferred? (b) Give and explicit example of the advantage of method (2) assuming that we have a 100 by 100 pixel image and a 3 by 3 pixel mask and we want to detect features at this scale and at double this scale.

10. What is aliasing and how is this avoided when down-sampling images to create an image pyramid?

11. Briefly describe what is meant by (a) a Gaussian image pyramid, and (b) a Laplacian image pyramid.

Low-Level Vision (Biological)

1. List the range of image properties to which V1 cells show selectivity.

2. What is a hyper-column?

3. Briefly describe the stimulus selectivities of simple and complex cells in V1.

4. Describe the receptive field structure of (a) a simple cell, and (b) a complex cell and explain how these inputs give rise to the observed response properties.

5. Describe how simple cell RFs could be modelled using convolution.

6. Describe how complex cell RFs could be modelled using convolution.

7. Gabors functions are the components of natural images under the "sparsity" constraint. What is the sparsity constraint and how is this relevant to efficient coding?

8. Briefly describe what is meant by the classical receptive field and the non-classical receptive field.

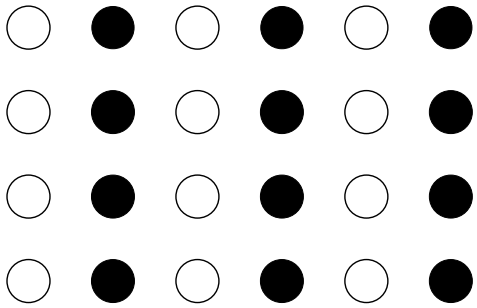
9. What is an "association field". Describe the association field for a V1 cell with an orientation preference.

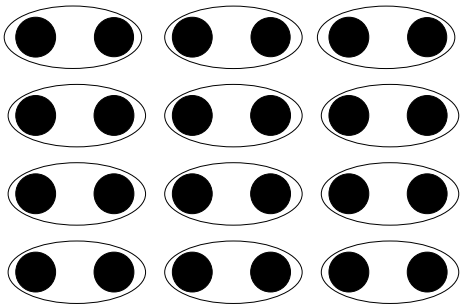
10. How do lateral connections in V1 give rise to (a) contour integration, (b) pop-out, (c) texture segmentation?

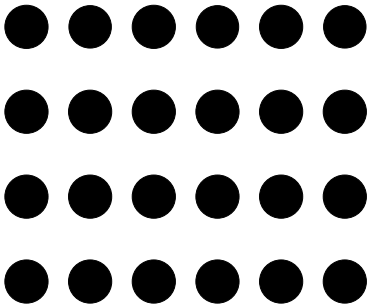
Mid-Level Vision: Segmentation (Biological)

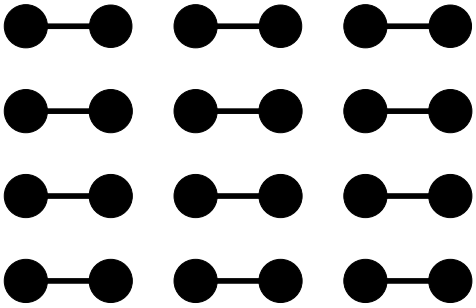
1. Briefly describe the difference between bottom-up and top-down influences on grouping.

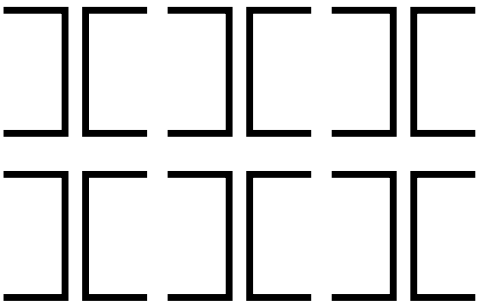
2. For each of the following images identify the "Gestalt Law" that gives rise to the observed grouping.

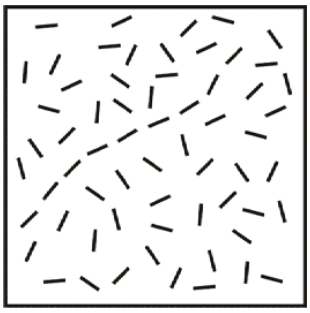
(a) 

(b) 

(c) 

(d) 

(e) 

(f) 

3. Explain how lateral connections in V1 give rise to the Gestalt biases of similarity and continuity.

4. Explain what is meant by border ownership?

5. Explain how lateral connections in V2 could give rise to border ownership.

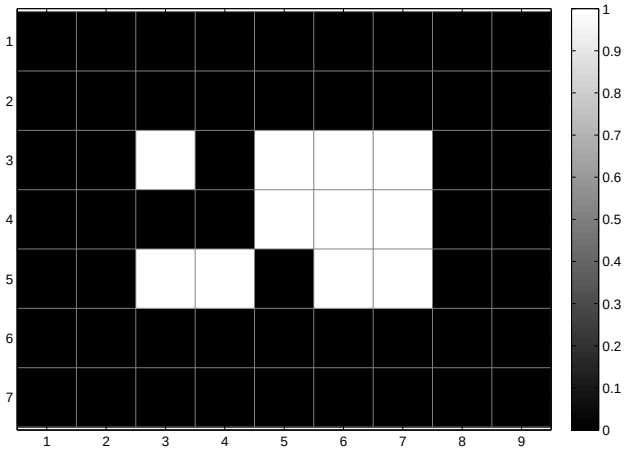
6. Give a definition of the Helmholtz Likelihood Principle

Mid-Level Vision: Segmentation (Artificial)

1. Briefly describe the role of mid-level vision.

2. One simple method of segmentation is thresholding. (a) Briefly explain how an appropriate threshold might by obtained by using an intensity histogram. (b) Explain the procedure used when applying hysteresis thresholding.

3. The figure shows a 7 by 9 pixel binary image.



(a) Draw the image that results from performing erosion followed by dilation on this image, assuming each pixel has 4 (horizontal and vertical) neighbours.
(b) Draw the image that results from performing dilation followed by erosion on this image, assuming each pixel has 8 (horizontal, vertical and diagonal) neighbours.

4. Write pseudo-code for the region growing algorithm.

5. The array below shows feature vectors for each pixel in a 3 by 3 image.

$$\begin{bmatrix} (5,10,15) & (10,15,30) & (10,10,25) \\ (10,10,15) & (5,20,15) & (10,5,30) \\ (5,5,15) & (30,10,5) & (30,10,10) \end{bmatrix}$$

Apply the region growing 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 elements are similar is that the SAD is less than 12, (3) the seed pixel is the top-left corner, (4) pixels have horizontal, vertical and diagonal neighbours.

6. Write pseudo-code for the region merging algorithm.

7. The array below shows feature vectors for each pixel in a 3 by 3 image.

$$\begin{bmatrix} (5,10,15) & (10,15,30) & (10,10,25) \\ (10,10,15) & (5,20,15) & (10,5,30) \\ (5,5,15) & (30,10,5) & (30,10,10) \end{bmatrix}$$

Apply the region 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) the first chosen region is the top-left corner, (4) regions have horizontal, vertical and diagonal neighbours.

8. Write pseudo-code for the split and merge algorithm.

9. The array below shows feature vectors for each pixel in a 3 by 3 image.

$$\begin{bmatrix} (5,10,15) & (10,15,30) & (10,10,25) \\ (10,10,15) & (5,20,15) & (10,5,30) \\ (5,5,15) & (30,10,5) & (30,10,10) \end{bmatrix}$$

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.

10. Write pseudo-code for the k-means clustering algorithm.

11. The array below shows feature vectors for each pixel in a 3 by 3 image.

$$\begin{bmatrix} (5,10,15) & (10,15,30) & (10,10,25) \\ (10,10,15) & (5,20,15) & (10,5,30) \\ (5,5,15) & (30,10,5) & (30,10,10) \end{bmatrix}$$

Apply the k-means clustering 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) there are two clusters, and the cluster centres are initially positioned at these positions in feature space: (5, 10, 15) and (10, 10, 25).

12. Write pseudo-code for the agglomerative hierarchical clustering algorithm.

13. The are a number of methods for calculating the distance between clusters in the agglomerative hierarchical clustering algorithm. Briefly describe each of the following methods. (a) single-link clustering, (b) complete-link clustering, (c) group-average clustering, (d) centroid clustering.

14. The array below shows feature vectors for each pixel in a 3 by 3 image.

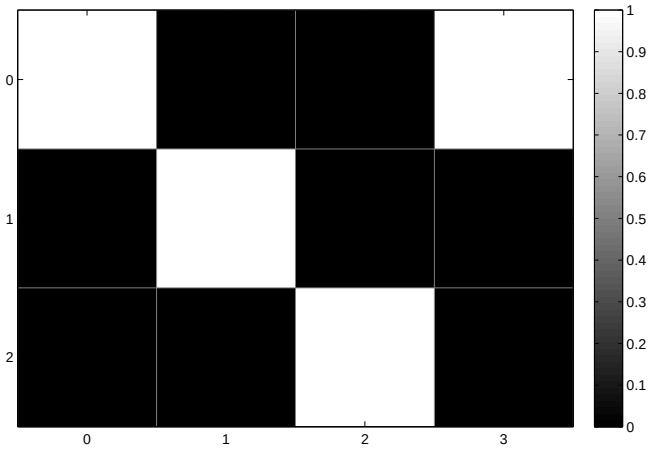
$$\begin{bmatrix} (5,10,15) & (10,15,30) & (10,10,25) \\ (10,10,15) & (5,20,15) & (10,5,30) \\ (5,5,15) & (30,10,5) & (30,10,10) \end{bmatrix}$$

Apply the agglomerative hierarchical clustering algorithm to assign pixels in to three regions. Assume that (1) the method used to assess similarity is the sum of absolute differences (SAD), (2) centroid clustering is used to calculate the distance between clusters.

15. Write pseudo-code for the Normalised Cuts algorithm.

16. Write pseudo-code for the Hough transform for straight lines

17. Following some initial processing, a 4 by 3 pixel region of an image has been converted into the following binary image:



Perform the Hough transform for straight lines on this image region, using the following values for θ: [0,30,60,90,120,150], and quantizing r to the nearest whole number.

18. Repeat the previous question using the following values for θ: [0,45,90,135]

19. Briefly describe the two terms that form the energy function that an active contour attempts to minimise.

Mid-Level Vision: Multiview

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

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

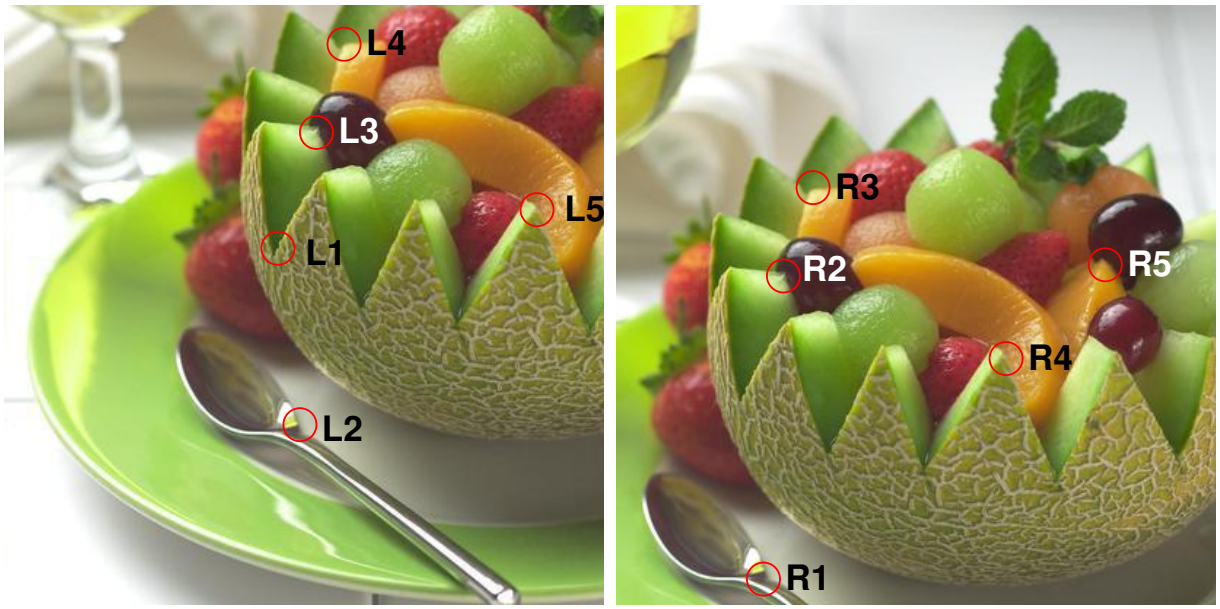
3. For feature-based methods of solving the correspondence problem, briefly explain what is meant by a "detector" and a "descriptor"

4. The two arrays below show the intensity values for each pixel in a stereo pair of 4 by 3 pixel images.

$$left: \begin{bmatrix} 4 & 7 & 6 & 7 \\ 3 & 4 & 5 & 4 \\ 8 & 7 & 6 & 8 \end{bmatrix} \qquad right: \begin{bmatrix} 7 & 6 & 7 & 5 \\ 4 & 5 & 4 & 5 \\ 7 & 6 & 8 & 7 \end{bmatrix}$$

Calculate the similarity of the pixel at coordinates (2,2) in the left image, to all pixel location in the right image, and hence, calculate the disparity at that point. Repeat this calculation for the pixel at coordinates (3,2) in the left image. Assume that (a) a 3 by 3 pixel window is used, (b) similarity is measured using the Sum of Absolute Differences (SAD), (c) the image is padded with zeros to allow calculation of similarity at the edges, (d) the cameras have coplanar image planes, (e) disparity is calculated as the translation from right to left.

5. Below is a pair of images showing different views of the same scene.



The locations of interest points are indicated on each image, and vectors of features values for each of these interest point is given below:

Point	Feature Values	Point	Feature Values
L1	(10, 4)	R1	(3, 7)
L2	(3, 8)	R2	(1, 1)
L3	(0, 2)	R3	(5, 7)
L4	(6, 9)	R4	(8, 0)
L5	(9, 1)	R5	(1, 2)

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

6. The coordinates of the interest points in Question 5, are as follows:

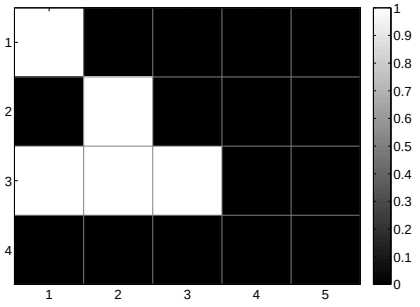
Point	Coordinates	Point	Coordinates
L1	(187, 168)	R1	(101, 394)
L2	(203, 290)	R2	(115, 186)
L3	(215, 87)	R3	(135, 128)
L4	(234, 28)	R4	(269, 243)
L5	(366, 142)	R5	(336, 178)

Calculate the disparity at each point in the left image. Assume that (a) the cameras have coplanar image planes (although not collinear x-axes), (b) disparity is calculated as the translation from right to left.

7. Write pseudo-code for the RANSAC algorithm.

8. Apply the RANSAC algorithm to find the true correspondence between the two images in Question 5. Assume (a) that the images are related by a pure translation in the x-y plane, (b) that t (the threshold for comparing the model's prediction with the data) is 20 pixels, (c) 3 trials are performed and these samples are chosen in the order L1, L2, L3 rather than being randomly chosen.

9. Below is shown a simple 5 by 4 pixel binary image. The two arrays show the derivatives of the image intensities in the x and y directions.



$$I_x = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad I_y = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Given the x any y derivatives of the image intensities shown above, calculate the response of the Harris corner detector at each of the six central pixels, assuming (a) a value of k=0.05, (b) that products of derivatives are summed over an equally weighted, 3 by 3 pixel, window around each pixel.

10. For the Harris corner detector, describe what type of image feature will give rise to the following values of R.
(a) $R \approx 0$
(b) $R < 0$
(c) $R > 0$.

11. Two cameras with identical focal lengths are set up so that their image planes are coplanar and their x-axes are collinear. Derive the relationship between the distance (Z) of a point from the cameras and the distance (B) that separates the origins of the two cameras.

12. Comment on the accuracy with which an object's depth can be measured with (a) changing distance, (b) changing baseline.

13. Briefly explain what is mean by the Epipolar constraint on the stereo correspondence problem.

14. List other constraints applied to solving the stereo correspondence problem, and note circumstances in which they fail.

15. List constraints typically applied to solving the video correspondence problem, and note circumstances in which they fail.

16. In a stereo vision system, the baseline between the camera centres is 400mm and the angle of convergence of the z-axes of the cameras is 60^o. Assume the z-axes of each camera make an equal angle with the baseline (*i.e.*, 60^o in this case). If the line-of-sight of a scene point makes angles α_L and α_R with the z-axes of the left and right cameras respectively, then what is the distance of the point from the horopter (a) $\alpha_L = \alpha_R = 15^o$, (b) $\alpha_L = +15^o$ and $\alpha_R = -15^o$, and (c) $\alpha_L = -15^o$ and $\alpha_R = +15^o$.

17. Give two reasons why the recovery of depth information is important for object recognition.

18. Briefly describe two oculomotor cues to depth.

19. Briefly describe four monocular cues to depth.

20. Briefly describe two motion induced cues to depth.

21. Define what is meant by the "aperture problem" and suggests how this problem can be overcome.

22. Consider a traditional barber's pole as shown in this image:

When the pole rotates on its axis in which direction is the (a) motion field, (b) optic flow? How might this be explained by the aperture problem?

23. Two frames in a video sequence were taken at times t and t+1s. The point (50,50,t) in the first image has been found to correspond to the point (25,50,t+1) in the second image. Given that the camera is moving at $0.1ms^{-1}$ along the camera x-axis, the focal length of the camera is 35mm, and the pixel size of the camera is 0.1mm/pixel, calculate the depth of the identified scene point.

24. Two frames in a video sequence were taken at times t and t+1s. The point (50,70,t) in the first image has been found to correspond to the point (45,63,t+1) in the second image. Given that the camera is moving at $0.1ms^{-1}$ along the optical axis of the camera (*i.e.*, the z-axis), and the centre of the image is at pixel coordinates (100,140), calculate the depth of the identified scene point.

25. Give an equation for the time-to-collision of a camera and a scene point which does not require the recovery of the depth of the scene point.

Using this equation, calculate the time-to-collision of the camera and the scene point in the previous question, assuming the camera velocity remains constant.

26. The arrays below show the pixel intensities in the same 1 by 5 pixel patch taken from four frames of a greyscale video. In order to segment any moving object from the background, calculate the result of performing (a) image differencing, (b) background subtraction. In both cases assume that the threshold is 50 and in (b) that the background is calculated using a moving average which is initialised to zero everywhere and which is updated using the formula $B(x,y) = (1-\beta)B(x,y) + \beta I(x,y,t)$ where $\beta = 0.5$.

$I(x,y,t1) = [190,200,90,110,90]$
$I(x,y,t2) = [110,170,160,70,70]$
$I(x,y,t3) = [100,60,170,200,90]$
$I(x,y,t4) = [90,100,100,190,190]$

High-Level Vision (Artificial)

1. Below is shown a template T and an image I. Calculate the result of performing template matching on the image, and hence, suggest the location of the object depicted in the template assuming that there is exactly one such object in the image. Use the following similarity measures (a) normalised cross-correlation (b) sum of absolute differences.

$$T = \begin{bmatrix} 100 & 150 & 200 \\ 150 & 10 & 200 \\ 200 & 200 & 250 \end{bmatrix}, \qquad I = \begin{bmatrix} 60 & 50 & 40 & 40 \\ 150 & 100 & 100 & 80 \\ 50 & 20 & 200 & 80 \\ 200 & 150 & 150 & 50 \end{bmatrix}$$

2. A computer vision system uses template matching to perform object recognition. The system needs to detect 20 different objects each of which can be seen from 12 different viewpoints, each of which requires a different template. If an image is 300 by 200 pixels, and templates are 11 by 11 pixels, how many floating-point operations are required to process one image if cross-correlation is used as the similarity measure?

3. Below are shown three binary templates T_1, T_2 and T_3 together with a patch I of a binary image. Determine which template best matches the image patch using the following similarity measures (a) cross-correlation, (b) normalised cross-correlation, (c) correlation coefficient, (d) sum of absolute differences.

$$T_1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \qquad T_2 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \qquad T_3 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \qquad I = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

4. Below is shown an edge template T and a binary image I which has been pre-processed to extract edges. Calculate the result of performing edge matching on the image, and hence, suggest the location of the object depicted in the edge template assuming that there is exactly one such object in the image. Calculate the distance between the template and the image as the average of the minimum distances between points on the edge template (T) and points on the edge image (I).

$$T = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \qquad I = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

5. One method of object recognition is comparison of intensity histograms. Briefly describe two advantages and two disadvantages of this method.

6. In a very 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	(20,5)	(1,6,10)
	A2	(10,40)	(7,8,15)
	A3	(40,25)	(2,9,3)
B	B1	(20,10)	(6,1,12)
	B2	(30,5)	(13,4,8)
	B3	(30,45)	(3,8,4)

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

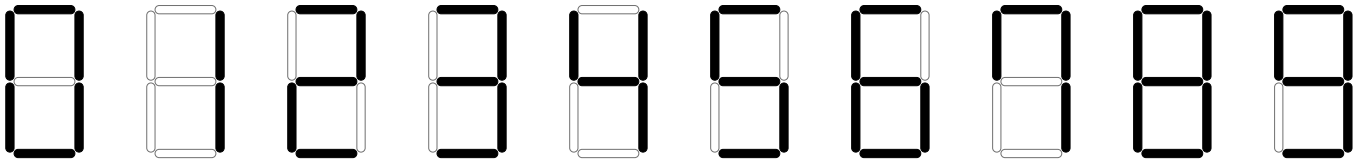
Keypoint Number	Coordinates (pixels)	Feature Vector
N1	(16,50)	(5,8,15)
N2	(25,14)	(2,6,11)
N3	(30,31)	(12,3,8)
N4	(40,45)	(5,2,11)
N5	(44,34)	(2,8,3)

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.

It is known that objects in different images are related by a pure translation in the image plane. 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.

7. In a simple bag-of-words object recognition system images are represented by histograms showing the number of occurrences of 10 "codewords". The number of occurrences of the codewords in three training images are given below:
ObjectA = (2,0,0,5,1,0,0,0,3,1)
ObjectB = (0,0,1,2,0,3,1,0,1,0)
ObjectC = (1,1,2,0,0,1,0,3,1,1)
A new image is encoded as follows:
New = (2,1,1,0,1,1,0,2,0,1)
Determine the training image that best matches the new image by finding the cosine of the angle between the codeword vectors.

8. A computer vision system is to be developed the can read digits from an 7-segment LCD display (like that on a standard calculator). On such a display, the numbers 0 to 9 are generated by turning on specific combinations of segments, as shown below.



A simple bag-of-words object recognition system is to be used. The codeword dictionary consists of two features: (1) a vertical line, and (2) a horizontal line.

(a) How would the digits 0 to 9 be encoded?

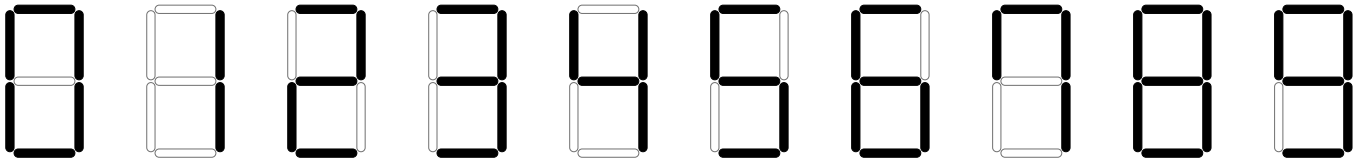
(b) Does this system succeed in recognising all 10 digits?

(c) Suggest an alternative object recognition method that might work better?

(d) If the camera capturing images of the LCD display gets rotated 180 degrees around the optical axis, what effect does this have on the bag-of-words solution and your alternative method?

(e) If the LCD display shows multiple digits simultaneously, what effect does this have on the bag-of-words solution and your alternative method?

9. A computer vision system is to be developed the can read digits from an 7-segment LCD display (like that on a standard calculator). On such a display, the numbers 0 to 9 are generated by turning on specific combinations of segments, as shown below.



A simple bag-of-words object recognition system is to be used. The SIFT feature detector has been used to locate features in the 10 training digits in order to create a codeword dictionary. Due to the rotation invariance of the SIFT descriptor only three distinct features are identified: (1) an "L" shaped corner (at any orientation), (2) a "T" shaped corner (at any orientation), (3) a line termination, or end point, (at any orientation).

(a) How would the digits 0 to 9 be encoded?

(b) Does this system succeed in recognising all 10 digits?

10. Projective geometry does not preserve distances or angles. However, the cross-ratio (which is a ratio of ratios of distances) is preserved. Given four collinear points p_1, p_2, p_3, and p_4, the cross-ratio is defined as:

$$Cr(p_1, p_2, p_3, p_4) = \frac{\Delta_{13}\Delta_{24}}{\Delta_{14}\Delta_{23}}$$

Where Δ_{ij} is the distance between two points p_i and p_j.

Four co-linear points are at the following 3D coordinates relative to the camera reference frame: $p_1 = [40, -40, 400]$, $p_2 = [23.3, -6.7, 483.3]$, $p_3 = [15, 10, 525]$, $p_4 = [-10, 60, 650]$.
(a) Calculate the cross-ratio for these points in 3D space.
(b) Calculate the cross-ratio for these points in the image seen by the camera. The image principal point is at coordinates

[244,180] pixels, and the magnification factors in the x and y directions are 925 and 740. Assume that the camera does not suffer from skew or any other defect.

High-Level Vision (Biological)

1. Give an example of a superordinate, a basic, and a subordinate category in the domain of (a) food, (b) furniture.

2. What is the difference between the "viewer-centred" and "object-centred" approach to object recognition?

3. What is a geon? What is a structural description?

4. An object recognition system encodes objects using 2-element feature vectors. Four objects from two classes are encoded as follows:

Object	Class	Feature Vector
1	A	(7,7)
2	A	(7,4)
3	B	(3,4)
4	B	(1,4)

A new object, of unknown class, has a feature vector (3,7). Determine the classification of the new object using a (1) nearest mean classifier, (2) nearest neighbour classifier, (3) k-nearest neighbour classifier, with k=3. Use the Euclidean distance as the similarity measure.

5. How do the following properties of cortical neurons change when moving along the ventral pathway from more peripheral areas to higher areas: (1) receptive field size, (2) sensitivity to stimulus location, (3) complexity of preferred stimulus.

6. Describe the mathematical operation that is performed by neurons in the HMAX model in the (1) simple cells, (2) complex cells.

7. Write down Bayes' theorem, and explain the interpretation of each term in relation to a computer vision system.

8. A production line produces two objects (objA and objB) which are sorted into separate bins using a computer vision system controlling a robot arm. The two objects have distinct shapes from most viewpoints. However, if objA happens to lie at a particular orientation (oriA), and objB lies at oriB, then the images of the two objects are indistinguishable.

It is known that the production line produces three times as many of objA than objB. It is also known that the probability of objA lying at oriA is 0.1, while the probability of objB lying at oriB is 0.2.

Using Bayes' theorem determine into which bin the robot should sort an object which could be either objA at oriA or objB at oriB in order to minimise the number of errors.