Task 1(a)

A developer intends to use matching the histograms of images as a way of performing content-based image retrieval. Discuss briefly whether this approach is likely to work.

This approach would work, as it is possible to characterise an image by the colours that it contains with a histogram. Therefore, if an image has a similar histogram to another image, then they will likely contain the same colours and therefore look the similar.

Task 1(b)

Outline briefly the principle of the FAST corner detector, and hence explain whether it is able to operate at video rates.

The FAST corner detector is based on the observation of a corner. By looking at a specific corner, it is possible to ascertain that more than half of the pixel will be light or dark based on how many pixels have the same value.

This would require some sophistication in order to work optimally, as FAST will produce some false positives, however due to its simplicity, it will easily operate at video rates.

Task 1(c)

Otsu's method is widely used in computer vision. State its purpose and give any shortcomings it may have.

Otsu's method is a method that allows for us to partition or segment objects of interest from their surrounding background. This method maximised the separation of the objects in the foreground and background, using the peaks in the histogram.

Otsu's method will function correctly when both the foreground and the background have a similar number of pixels. However, it is less effective when there are a few foreground pixels that are superimposed on a large, dark background.

Task 1(d)

Outline briefly a way of determining whether a region found in an image is circular.

If the image does not contain any colour, one method that can be used would be to convert the image into grey scale. We can then apply binary thresholding, and determine the contours of the object which can be used to generate a circle

The most common method used for images with colour would be to first isolate the desired object by creating a mask using lower and upper HSV colour values of the colour object. After the circle is isolated, a circular Hough transform can then be performed, which will calculate an appropriate circle for the isolated object.

Task 1(e)

A convolutional neural network is built up from five types of layers. Describe briefly what these layers do.

- **INPUT** - This layer will receive raw pixels of the data. For colour data, the red, green and blue (or HSV) values will be presented to different input neurons.
- **CONV** - The convolutional layer computes the output of neurons that are connected to local regions of the input. This layer performs a convolution with coefficients that are learned from the data.
- **RELU** – This layer applies an element-wise activation function which leaves the width and height of the network unchanged.
- **POOL** – This layer down-samples or averages the regions of its input to reduce the overall width and height of the network.
- **FC** - This layer computes the class scores, and then returns the calculated class of the pattern presented at the network's input.

Task 2

A local company, Wivenhoe Innovative Software Engineering (WISE), has appointed you as a consultant because of the extensive knowledge of computer vision you have gained from CE316. WISE is developing a vision-guided robot system for a warehouse to manipulate parcels for shipment. All the parcels in the warehouse are shipped in dark red cardboard boxes.

Task 2(a)

What advice would you give WISE regarding how the walls and shelving in the warehouse are painted and the lighting arranged?

Firstly, the colours used for the walls and the shelving should contrast so that they can be identified from each other. Furthermore, the colours used for the walls and shelves should be different from the colour used for the boxes in order to differentiate them from each other.

In regard to the lighting, the lights should be fairly bright and distributed evenly in order to prevent shadows and glare, as well as to ensure that the shades of colour are consistent throughout the warehouse. This will ensure that there is a consistent level of colour in the warehouse which will increase the accuracy of the object detection

Task 2(b)

If you were able to configure the vision system in the robot in any way you wish, describe two ways in which parcels might be segmented from the surroundings.

One method that could be used in order to segment the parcels from their surroundings in the warehouse is by using region labelling (providing that there is a sufficient gap between each parcel). This is where the pixels that belong to a specific object are identified, and a number is assigned to them. When the regions have been numbered in this way, it is possible to identify which of them belong to each object.

Another method that could be used to segment the parcels would be to make use of thresholding. This is where we would first process the image in order to generate a histogram. After this has been done, the histogram will have two peaks, one due to the background and the other to the objects of interest. It is then possible to set a threshold that will divide the parcels from background, which will be between the two peaks.

Task 2(c)

Describe a simple way in which the shape of features resulting from the segmentation process can be identified as being rectangular.

In order to identify that the shape is rectangular, we will first have to look at the main characteristics of a rectangle. The rectangle has 4 edges and corners, 2 of the edges are short and of equal length, and the other 2 edges are long and also equal length. Taking this into consideration, our algorithm will need to determine if the isolated region contains all of the listed properties. A different approach would be to count the number of pixels in each column of a shape, in order to give a vertical profile, a horizontal profile, and a diagonal profile. These can be used to match the shapes on their own.

Task 3

Face recognition is a well-known application of computer vision.

Task 3(a)

What is meant by enrolment in a face recognition system?

This is when an individual has been encountered by the facial recognition system for the first time, the image is stored as a record for later comparison. The face of the person is stored as reference for the individual. If a face is detected that is already in the system, the enrolment will not take place. Enrolment may also be used in other types of systems such as a biometric authentication, where a record is stored for a user that can be used for future reference.

Task 3(b)

When a person enrols in a face recognition system, it is normal to use several images of their face rather than a single one. Explain why this is the case.

This is the case as by storing multiple images of the person, and then using these as a reference, the facial recognition system will be much more accurate. This in turn will decrease the chances of a false positive, as the facial recognition system will be checking the person's face against a set of images as a reference. In a best case scenario, the system will always allow for correct facial recognition to take place.

Task 3(c)

In the particular case of the eigenfaces technique, what is stored when a person is enrolled?

In the case of the eigenfaces technique, the information that is stored for each person is the face of said person, and also a name that is used to identify the person. This information is stored in a database in the form of a 'cloud' of values.

Task 3(d)

When a person is compared against enrolled faces in eigenfaces, how is that comparison performed?

The person is compared against enrolled faces in eigenfaces, by a series of steps as detailed:

- Firstly, the database will display "cloud" of values.
- After this, the directions of the maximum spread of the data in the cloud (the Eigen decomposition) is identified and recorded.
- Each point of data within the cloud is then identified by its coordinates in the axes.
- Each point of data will represent an image of a person, in turn the coordinates are a compact way of representing that image.
- The nearby points will represent similar faces.

Task 4

A mobile robot is equipped with cameras at its front two corners; a schematic plan view is shown in Figure 5.1. Two identical cameras of focal length f are arranged so that their optical axes are perfectly parallel and separated by a distance B.

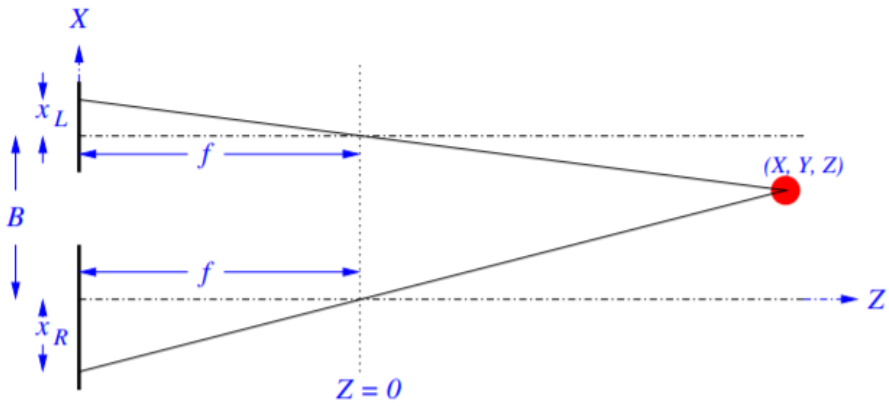

Figure 5.1

Task 4(a)

Show that the distance Z to an object visible in both camera images is given by:

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

where xL and xR are the x-locations of the object in the left and right images respectively.

Left Camera: xL / f = B – X / Z
Right camera: - xR / f = X / Z

Obtain: xL – Xr / f = B - X + X / Z = - B / Z
Equation: Z = f B / xL – xR

Task 4(b)

In the case where B = 80 mm, f = 50 mm, xL = 12 mm and xR = −12 mm, calculate the distance Z to the object.

B: 80mm
F: 50mm
xL: 12mm
xR: 12mm

Answer: Z = 50 * 80 / 12 - - 12 = 166.7 (1 s.f)

Task 4(c)

If a second pair of images were taken and it was found that xL = 8 mm and xR = −8 mm, what can you infer about the motion of the object?

B: 80
F: 50mm
xL: 8mm
xR: -8mm

Answer: Z = 50 * 80 / 8 – - 8 = 250mm

Disparity Before: 12 - - 12 = 24
Disparity Now: 8 - - 8 = 16

Disparity Change Factor: 24 / 16 = 1.5
Depth Change Factor: 250 / 166.7 = 1.5

The distance of the object from the camera is actually the same (the object has not moved), however the object does appear further away.