Computer Science and Software Engineering

EXAMINATION

Mid-year Examinations, 2020

COSC428-20S1 (C) Computer Vision

Examination Duration: 120 minutes

Exam Conditions:

This is an Open Book exam, but it is assumed that

you will not communicate with anyone else during it

Instructions to Students:

- This exam is worth a total of 100 marks
- Contribution to final grade: 50%
- Length: 10 questions
- Answer all questions – with text only
  (There is no expectation or need for you to write equations or diagrams.)
- Check carefully the number of marks allocated to each question. This suggests the degree of detail required in each answer and therefore amount of time to spend on it.

A few things we are doing at UC to assist whilst you are taking your assessment:

The ITS and eLearning teams will be monitoring all systems used for assessment. If an issue is identified with any of our IT systems, for example UC Learn, we will provide an update on the ITS Website, https://status.canterbury.ac.nz and we will also send a message via the UC Student Blog

What will happen if your connection drops part way through a test?

Quizzes on Learn save when you click to go to the "Next page" (or click "Previous page") – but it may be a good idea to type your answers into Word (or equivalent) with constant saving – and then just cut and paste your answers into the quiz as you go.

Both LEARN and the assessment server keep sessions for 2 hours, meaning that you can get back to where you were if you did lose connection.

If you experience an issue whilst taking an assessment please advise your Course Coordinator Richard Green immediately richard.green@canterbury.ac.nz (or 021331707).

Declaration

By entering my name below, I declare that this exam represents my own work in accordance with university regulations. I also understand that exam materials are to be considered confidential and I must not discuss the exam materials until the exam is over for all students. I also understand that failure to comply with these requirements may mean that the matter will be referred to the Head of Department, Dean or Proctor as appropriate for disciplinary action.

Answer:

1 [9 marks total)]

Name and describe the three stages of a convolutional neural network (deep learning) in the order that they operate from an input image.

2 (12 marks)

In order for a feature registration algorithm to work well it must be robust to common image transformations and distortions. List six such image transformations and distortions.

3 (8 marks)

a) **Spectral Resolution**. Humans can perceive 10 octaves of sound frequencies, from 20Hz to 20kHz. State the approximate spectral resolution (wavelength in nm) that humans can perceive. [2 marks]

b) **Dynamic Range**. The difference in intensity between the softest perceivable sound and the loudest sound that can be tolerated without pain is a ratio of 10^9:1. State the approximate visual dynamic range (as a ratio) that humans can perceive in regards to the difference between the lowest perceptible light intensity and the highest intensity we can tolerate without glare. [2 marks]

c) **Spatial Resolution**. State the approximate number of centimetres spatial resolution that humans can perceive at 20 metres. [2 marks]

d) **Radiometric Resolution**. Regardless of our spectral resolution and dynamic range, most humans can only reliably distinguish between a limited number of colours and shades of grey. State the approximate number of colours and the approximate number of shades of grey that humans can reliably distinguish between (when placed in a graduated scale on the same page). [2 marks]

4 (9 marks)

The Hough (pronounced “huff”) transform (HT) can detect a line using a “voting” scheme where points vote for a set of parameters describing a line. The more votes for a particular set, the more evidence that the corresponding line is present in the image. So it can detect MULTIPLE lines in one shot.

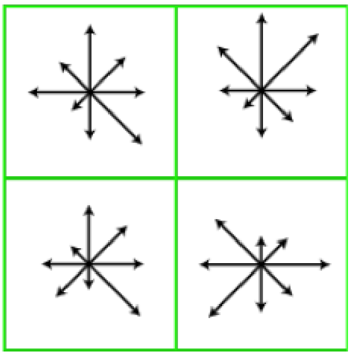
To find straight lines:

- A line in the image corresponds to a point in Hough space
- To go from image space to Hough space:
- given a set of points (x,y), find all (m,b) such that y = mx + b

Describe how this Hough transform can be generalised to detect curved lines in an image, even when the curve does not have a simple analytic form.

5 (8 marks)

The orientation of objects can be tracked from one frame to the next using the *scale-invariant feature transform* (SIFT) which extracts scale and rotation invariant features from images. SIFT is also used to recognise objects and match different views of a scene for stereo vision. In this algorithm, descriptors of key-points across an image are created as a set of 8 gradient orientations about a pixel as shown in the diagram to the right.



Name and describe the gradient operators you would use to create such a set of **<u>eight</u>** orientations as shown in the diagram to the right and also describe how you would use them to create these orientations.

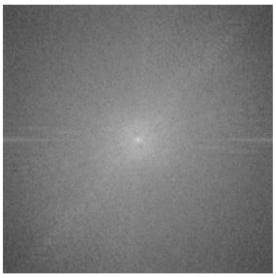
(Hint: If a gradient operator is only centred on a key-point, then the only result is a gradient in one direction equal to the gradient in the opposite direction – in contrast to orientations shown in the diagram to the right.)

6 (6 marks)

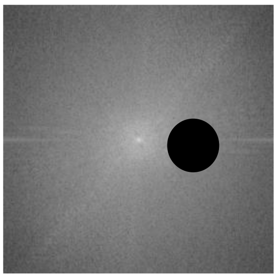
Images can be filtered by applying an inverse Fourier transform to a Fourier transformed image.

Given the following image and Fourier transforms of that original image:

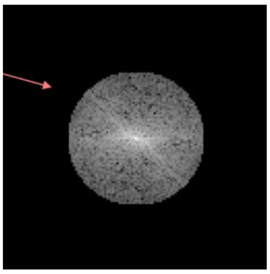
Original cheetah image



(i) Fourier transform of image



(ii) Fourier transform of image with centre blacked out



(iii) Fourier transform of image with all but the centre zeroed

(a) Is image (i) the magnitude transform or phase transform of the original image? Explain the appearance of image (i) [2 marks]

(b) If an inverse Fourier transform is applied to images (ii) and (iii), describe the resulting images and also explain why they would appear as described [4 marks]

7 (8 marks)

Name each pair of filters labelled (a) to (d) below and describe what each filter accomplishes.

$$\Delta_1 \quad \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix} \qquad \Delta_2 \quad \begin{matrix} 1 & 0 \\ 0 & -1 \end{matrix}$$

(a)

$$\Delta_1 \quad \begin{matrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{matrix} \qquad \Delta_2 \quad \begin{matrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{matrix}$$

(b)

$$\Delta_1 \quad \begin{matrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{matrix} \qquad \Delta_2 \quad \begin{matrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{matrix}$$

(c)

$$\Delta_1 \quad \begin{matrix} -3 & -1 & 1 & 3 \\ -3 & -1 & 1 & 3 \\ -3 & -1 & 1 & 3 \\ -3 & -1 & 1 & 3 \end{matrix} \qquad \Delta_2 \quad \begin{matrix} 3 & 3 & 3 & 3 \\ 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 \\ -3 & -3 & -3 & -3 \end{matrix}$$

(d)

8 (12 marks total)

When segmenting a moving object from a static background:

- "***Background subtraction***" usually refers to the first frame, or some derivative of it, being the reference frame.
- "***Difference***" algorithm usually refers to the difference between two adjacent frames where in this case, the previous frame is the reference frame.
- "***Ghosting***" refers to a second image of the moving object appearing as an artefact of a difference algorithm.
- "***Foreground aperture***" refers to a hole appearing in the moving object as an artefact of a difference algorithm.

In the case of a ball not visible in the first frame - and then it rolls into view of a stationary camera, compare the visible differences between:

(a) using the previous frame as a reference frame [4 marks]
(b) using the first frame as a reference frame [4 marks]
(c) using the double difference algorithm [4 marks]

In answering each part of this question above, explain the consequences of

i. the ball moving at a speed where there is a separation of one ball diameter between the position of the ball in consecutive frames,
ii. the ball moving so fast that it only appears in one single frame,
iii. the ball moving so slowly that the ball overlaps half of the ball in the previous frame,
iv. when the ball stops moving.

Answer these 12 questions (1 mark each) in the sequence: (a) i,ii,iii,iv (b) i,ii,iii,iv (c) i,ii,iii,iv

9 [12 marks total]

Briefly describe advantages and/or disadvantages of the following four different types of camera technologies **for acquiring image depth values**. [1 mark for each advantage *or* disadvantage cited]

(a) structured light camera [3 marks]
(b) time-of-flight camera [3 marks]
(c) stereo camera [3 marks]
(d) LIDAR (Light Detection and Ranging) [3 marks]

10 [16 marks total]

You are to briefly describe **only four of the following** class projects [for 4 marks each] by just listing (one per line) at least four algorithmic steps, **naming the algorithms** used in the order they were used.

Do not select your own or similar project (e.g. face recognition projects - do not select other face recognition projects, etc).

(a) Pedestrian safety

Gaussian blur (to reduce noise)

Histogram of Oriented Gradients (feature descriptor)

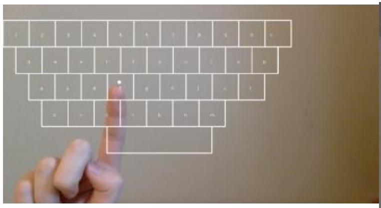
Support Vector Machine classifier (SVN) (for pedestrian detection)

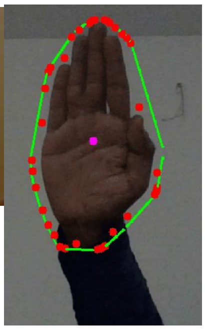
Non Maximum Suppression (removing overlapping bounding boxes)

Intersection-over-Union (IoU)

Background subtraction

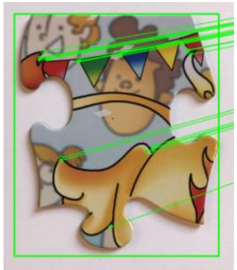
(b) Virtual keyboard





skin colour histogram
background subtraction
furtherest points convexity defects
key mapping

(c) Jigsaw puzzle piece location detection



Greyscale

Threshold (simple thresholding, **mean adaptive thresholding** or Gaussian adaptive thresholding)

Bounding box

ORB feature detector

FLAM feature matcher

(… or find contours, draw contours, probalistic hough line, brute force matcher …)

(d) Recognising text in images for censoring

- Greyscale
- Gaussian Blur
- Otsu’s Threshold
- Open (morphology - erosion and then dilation)
- Inversion
- OCR (Tesseract 4 using neural net (LSTM))

(e) Bicycle geometry identification

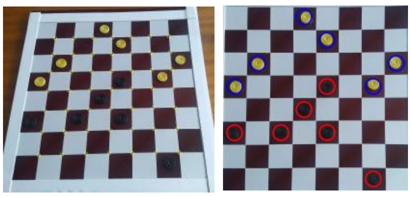
Gaussian blur, greyscale conversion

Hough circles (for identification of wheels)

Template matching (for stem, crank, saddle)

Image pyramid (for template matching different scale images)

(f) Checkers position recognition.



```
[['__' 'BM' '__' 'BM' '__' 'BM' '__' 'BM']
 ['BM' '__' 'BM' '__' 'BM' '__' 'BM' '__']
 ['__' 'BM' '__' 'BM' '__' 'BM' '__' 'BM']
 ['__' '__' '__' '__' '__' '__' '__' '__']
 ['__' '__' '__' '__' '__' '__' '__' '__']
 ['WM' '__' 'WM' '__' 'WM' '__' 'WM' '__']
 ['__' 'WM' '__' 'WM' '__' 'WM' '__' 'WM']
 ['WM' '__' 'WM' '__' 'WM' '__' 'WM' '__']]
```

Corner detection (openCV findChessboardCorners)

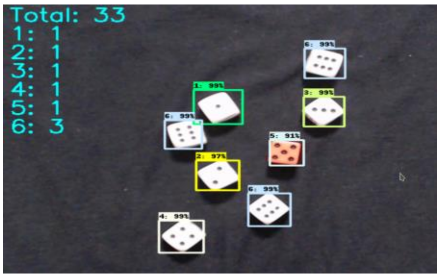
Canny edge detection

Morphological operations (erosion, closing)

K-means

Neural network classification

(g) Dice recognition and classification



(four of the following)
Greyscale
Otsu's thresholding
Morphological open
Gaussian Blur
Canny edge detection
Find contours and reject invalid shapes
Die face extraction (threshold: darken surrounding background and lighten die)
Single Shot Multibox object detection with inception feature extraction

(or Gaussian Filter, Erosion and Dilation, Binary Thresholding, Ramer-Douglas Peucker)

(h) Fire detection

Gaussian Blur

Conversion to binary black/white

Yolov3 (for speed & accuracy)

HSV colour threshold

End of Examination