Family Name _____

First Name _____

Student Number |__|__|__|__|__|__|__|__|

Venue _____

Seat Number _____

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

No electronic/communication devices are permitted.

No exam materials may be removed from the exam room.

Computer Science and Software Engineering

EXAMINATION

End-of-year Examinations, 2020

COSC428-20S2 (C) Computer Vision

Examination Duration: 120 minutes

Exam Conditions:

Closed Book exam: Students may not bring in anything apart from writing instruments.

No calculators are permitted

Materials Permitted in the Exam Venue:

None

Materials to be Supplied to Students (if needed)**:**

- Extra sheets of write-on question paper (or answer book)

Instructions to Students:

- **Write your name and student ID above**
- This exam is worth a total of 100 marks
- Contribution to final grade: 40%
- Length: 10 questions
- Answer all questions.
- 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.
- The amount of space provided also indicates the amount of detail expected.
- **Write strictly in the spaces allocated to each answer**. Do not write close to the margins, as the answer books will be scanned, and writing very close to the margin may not be picked up. If you require extra room, there are blank pages at the end of this booklet. You may also use additional sheets of paper; these must be fastened securely to your answer booklet. You should clearly indicate in the appropriate space that the answer is continued/provided elsewhere.

For Examiner Use Only

Question	Mark

Total _____

Questions Start on Page 3

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.

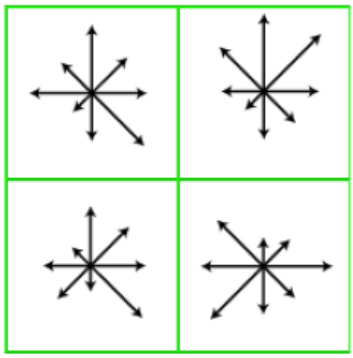
To find straight lines:

- 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.



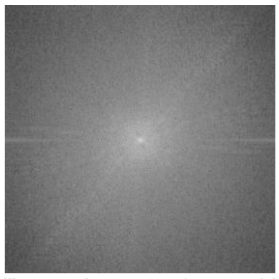
Name and describe the gradient operators you would use to create such a set of **eight** orientations as shown in the diagram to the right and also describe how you would use them to create these orientations.

(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.)

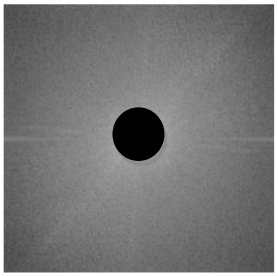
6 (6 marks)

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:

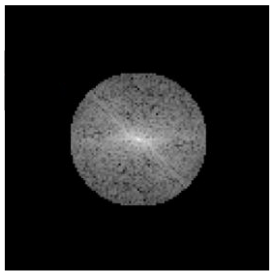
Original cheetah image



(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.

Δ_1

0	1
-1	0

Δ_2

1	0
0	-1

(a)

Δ_1

-1	0	1
-1	0	1
-1	0	1

Δ_2

1	1	1
0	0	0
-1	-1	-1

(b)

Δ_1

-1	0	1
-2	0	2
-1	0	1

Δ_2

1	2	1
0	0	0
-1	-2	-1

(c)

Δ_1

-3	-1	1	3
-3	-1	1	3
-3	-1	1	3
-3	-1	1	3

Δ_2

3	3	3	3
1	1	1	1
-1	-1	-1	-1
-3	-3	-3	-3

(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) the algorithmic steps, **naming the algorithms** used in the order they were used.

Do not select your own project.

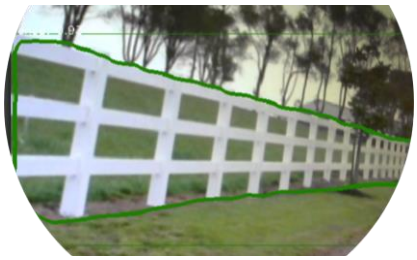
(a) Recognising Settlers of Catan tokens

(b) Tracking barbell and squat pose

(c) Detecting sheep

(d) Blackjack trainer

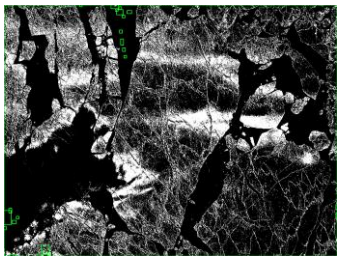
(e) Recognising farm fences

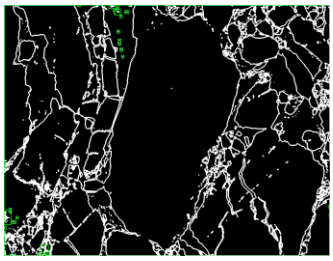


(f) Crossword scanner and solver

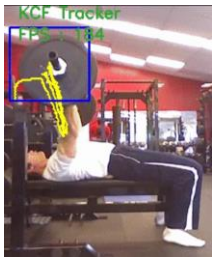
```
['#', 'T', '#', 'L', '#', 'o', '#', '#', '#', 'B', '#', ' ', '#', 's', '#']
['H', 'A', 'R', 'a', 'C', 'O', 'T', '#', 'V', 'A', 'R', 'a', 'o', 'U', 'S']
['#', 'O', '#', 'O', '#', 'U', '#', 'M', '#', 'T', '#', 'M', '#', 'K', '#']
[' ', 'N', ' ', 'o', ' ', 'T', '#', 'A', 'C', 'H', 'a', ' ', 'L', 'E', 'S']
['#', 'E', '#', 'T', '#', 'H', '#', 'S', '#', 'U', '#', 'E', '#', 'B', '#']
['x', 'R', 'A', 'Y', '#', 'S', 'A', 'T', 'U', 'R', 'A', ' ', 'I', 'O', 'O']
['#', 'Y', '#', 'P', '#', '#', '#', 'E', '#', 'S', '#', '#', '#', 'X', '#']
['#', '#', 'D', 'E', 'M', 'O', 'C', 'R', 'A', 'T', 'I', 'Z', 'E', '#', '#']
['#', 'F', '#', '#', '#', 'V', '#', 'C', '#', '#', '#', 'E', '#', 'Q', '#']
['W', 'a', 'O', 'E', 'C', 'E', 'L', 'L', 'A', 'R', '#', 'P', 'L', 'U', 'G']
['#', 'R', '#', 'A', '#', 'R', '#', 'A', '#', 'E', '#', 'P', '#', 'L', '#']
['O', 'E', 'U', 'R', 'a', 'T', ' ', 'S', '#', 'C', 'H', 'E', 'E', 'R', 'y']
['#', 'M', '#', 'W', '#', 'A', '#', 'S', '#', 'E', '#', 'L', '#', 'R', '#']
['K', 'A', 'R', 'A', 'o', 'K', 'E', '#', 'U', 'O', 'A', ' ', 'D', 'E', 'D']
['#', 'O', '#', 'x', '#', 'E', '#', '#', '#', 'T', '#', 'O', '#', 'L', '#']
```

(g) Automatic extraction of ice mesh





(h) Barbell motion tracking



… extra space …

If you use this page, please refer to it from the original question.

… extra space …

If you use this page, please refer to it from the original question.

End of Examination