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

Mid-year Examinations, 2018

COSC428-18S1 (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 is a blank page 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 [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]

For all depth cameras, reflective (e.g. wet) surfaces can cause noisy depth values.

Structured light:
- Cannot work in direct sunlight because the strong infra-red sunlight interferes with the low intensity projected infra-red camera light (low signal-to-noise ratio)
- Cannot work closer than 0.5m because the projected pattern of dots become too close together in the image.
- Cannot work further away than about 3.5m because the projected dots become too far apart and the intensity is too low.
- Motion blur occurs for fast motion because of the low intensity of the projected infra-red pattern of dots.
- Accuracy decreases with distance

Time of flight camera:
- Cannot work in direct sunlight because the strong infra-red sunlight interferes with the low intensity infra-red camera light (low signal-to-noise ratio)
- Limited range due to low intensity infra-red light
- Accuracy is independent of distance

Stereo camera:
- Potential for highest resolution
- Colour is also available for each pixel (as well as depth)
- Works well in direct sunlight.
- Noisy depth values in low ambient light.
- Works for motion (if well illuminated)
- Accuracy decreases with distance
- Many gaps in depth values in image regions without features (i.e. regions of uniform colour/intensity). Depth accuracy can be increased using higher resolution cameras.
- Depth accuracy over longer distances can be increased using a wider baseline.
- Cheap cameras (e.g. webcams) need extensive calibration for useful depth accuracy.

LIDAR:
- Good range (e.g. used for mapping ground from aircraft).
- Accuracy is independent of distance
- Works well in direct sunlight
- Low resolution
- Low frame rate
- Has moving parts (e.g. motor rotating mirror)
- Expensive

2 [8 marks total]

In the Canny edge detector, describe the impact of varying the

(a) σ (Gaussian kernel size) [4 marks]
(b) threshold [4 marks]

(a) σ (Gaussian kernel size)

- large detects large scale edges
- small detects fine features

(b) Hysteresis requires two thresholds – high and low: Begin by applying the high threshold to locate genuine edges. Then while tracing an edge, apply the low threshold to trace faint sections of edges. A threshold set too high can miss important information but a threshold set too low will falsely identify irrelevant information (such as noise).

3 [12 marks total]

Briefly describe the following morphological operators and explain what effect they have on an image and why they have such an effect:

(a) Erosion [3 marks]
(b) Dilation [3 marks]
(c) Open [3 marks]
(d) Close [3 marks]

(a) Erosion: **Removes outside pixels** of a region/blob (and internal holes/regions) usually using a convolution kernel/mask in an *and* operation (or subtracts the *convolution of the kernel with the image*). Removes small details such as thin lines, noise points and widens gaps. Shrinks a region (to a skeleton with successive erosions).

(b) Dilation: **Adds pixels to the outside** of a region/blob usually using a convolution kernel/mask in an *or* operation (or adds the *convolution of the kernel with the image*). Enlarges a region/blob, thickens lines, fills small holes.

(c) Open: **Erode then dilate** image. (i.e. dilates an eroded image.)

Removes small details such as thin lines, spurs and noise. Smoothes jagged edges **without changing the size of the original object**.

(d) Close: **Dilate then erode image**. (i.e. erodes a dilated image.)

Closes/fills in small gaps/holes and preserves thin lines **without changing the size of the original object**.

4 [12 marks]

Describe how correctly matched points in two images enable finding:

(a) depth values in a stereo pair of images [4 marks]
(b) optical flow points in two successive frames of video using the Lukas Kanade algorithm [4 marks]
(c) Describe how depth can be calculated from optical flow. [4 marks]

(a)

One image is **rectified** (aligned) with respect to the other (using the "essential matrix").

Points lying on a **horizontal line** in one image are **matched with corresponding points** on the same line in the other image (.e.g. using **least squares** of pixel values over a region around each point).

The "*x*" distance between a matching pair of points is called the **disparity**. The **larger** the disparity, the **closer** is that point to the camera based on triangulation (but this is not linear).

(b)

Lucas-Kanade method **integrates gradients over a patch** to find features good enough to track using the **Harris** detector.

A **constant velocity** is assumed for all pixels within an image patch.

Optical flow is the measure of the movement that feature points undergo in successive frames.

(c)

Relative depth can be calculated from the **velocity of optical flow points** – which is larger when the depth is less. So absolute depth could be determined if the **velocity** is known.

Even for a camera moving forwards or backwards with no rotation - as depth decreases, the "**focus of expansion**" velocity increases (and vice-versa).

5 [12 marks total]

Briefly describe the following four goals of deep learning applied to images:

(a) classification [3 marks]
(b) object detection [3 marks]
(c) dense segmentation [3 marks]
(d) instance segmentation [3 marks]

a) detect if an object is present or not in an image (I.e. not detecting where it is, but just detecting if such an object exists anywhere in an image.)

b) detect the location of an object in an image (returning region-of-interest/bounding-box coordinates)

c) label every pixel in an image as belonging to a class (such as grass pixels, sheep pixels)

d) label segmented pixels for each instance of a class (such as recognise which general sheep pixels in an image belong to which individual sheep for many overlapping sheep in a flock of sheep in an image)

6 [8 marks]

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

- Translation (includes x, y and z translations (scale change is just a z translation))
- Rotation (includes x, y and z translations (skew is a rotation))
- Illumination (variations including consequent colour shifts and shadows)
- Blur (motion blur or defocus blur)
- Non-rigid deformations include
  - radial distortion,
  - stretching,
  - warping,
  - intrinsic camera parameters
- Noise
- Partial Occlusion
- Camera gain changes
- Self similarity

7 [6 marks]

Describe how to remove noise from a 3D point cloud using PCL (Point Cloud Library).

Use the Statistical Outlier Removal (SOR) Filter which consists of two passes:

First pass: for each point, find the mean distance to k-neighbours

Second pass: remove outliers with high means

8 [6 marks]

List three advantages of fiducial marker tracking over natural feature tracking and list three advantages of natural feature tracking over fiducial marker tracking.

Fiducial: Tracking is less computationally efficient
Fiducial: More accurate 6 degree of freedom pose
Fiducial: Usually requires no database to be stored

NFT: Don't need markers in the scene
NFT: Natural feature targets catch the attention less
NFT: Natural feature targets work also if partially in view

9 [8 marks]

PyTorch and TensorFlow are two popular deep learning frameworks. Describe two advantages for each of these two frameworks.

PyTorch: More flexible for experiments

PyTorch: API is easier to use

TensorFlow: Runs on more devices

TensorFlow: Larger user community and trained networks

10 [16 marks]

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) "Wheelchair Docking" at a desk used an Intel Realsense D435 camera to locate a desk immediately in front of the wheelchair.

Gaussian Blur

Convolution

Thresholding

Opening and Closing

Contours

Deprojection

(b) "Navigation of Robotic Platform using a single webcam"

HSV

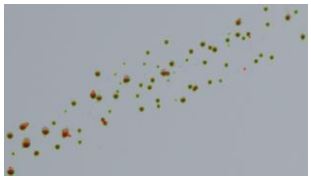
Binary

opening and closing

canny

Hough

(c) "Blood Spatter Segmentation"



Convert to HSV

Opening morphological operation

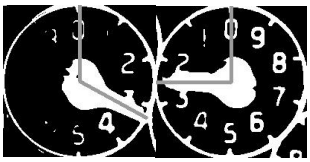
Otsu Thresholding

Hough Circle Transform

Contours Found with Suzki Border Algorithm

Fit ellipses

(d) "Automated Electricity Meter Dial Reading"



Blur Image with Gaussian

Hough Circles (to find ROI)

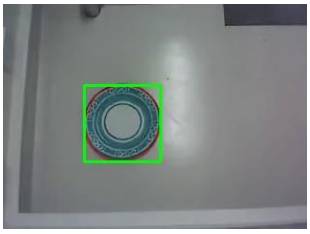
Convert to binary

Otsu's Algorithm for Thresholding

Bresenham's Line Algorithm

Convert angle to integer

(e) “Always Clean Kitchen” to detect dishes left behind



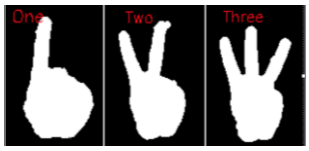
Background Subtraction and Thresholding

Morphological Transforms

Canny Edge Detection

Contour Detection and Size Comparison

(f) “REAL-time Hand Gesture Recognition Using Webcam”



HSV to filter skin colour

Haar Feature-based Cascade Classifier to detect and mask face

contour analysis to separate hand

Use Convolutional Neural Network to classify hand pose (number of fingers)

(g) "Book Call Number Detection"

Grayscale

Canny Edge

Dilation

Tesseract OCR

(h) "Crop-row Detection for Agricultural Robots"

HSV to threshold green

transform to birds-eye view

skeleton (morphology)

mean-shift clustering

fit lines

transform back into perspective view

… extra space …

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

End of Examination