The Australian National University
College of Engineering and Computer Science
Final Examination, Second Semester 2020

ENGN6528 Computer Vision

Question Booklet

Reading time: 15 minutes
Writing time: 3 hours
Uploading time: 15 minutes

Instructions on next page

Allotted Time

You have 3 hours to complete the exam plus 15 minutes of reading time. Additional 15 minutes has been allowed to accommodate the additional task of uploading your completed exam to the final exam Turnitin submission portal on the ENGN6528 Wattle site. Thus you have 3 hours and 30 minutes to complete the exam. NO late exams will be accepted. You may begin the exam as soon as you download it.

Minimal requirements

You may attempt all questions.

You SHOULD NOT include an assignment cover sheet.

You must type your ANU student identification number at the top of the first page of your submission.

You must monitor your own time (i.e. there is no invigilator to tell you how many minutes are left).

Your answers must be clear enough that another person can read, understand and mark your answer. 11 or 12 point font with 1.5 spacing is preferred. Scanned images of handwritten equations or diagrams must be legible and of a suitable size.

Numbering questions

- You must specify the question number you are answering at the top of the page.
- Each question should begin on a new page.
- Multi-part questions (e.g. question 1 parts a and b) may be addressed on the same page but should be clearly labelled (e.g. 1a, 1b).
- Questions should be answered in order.

You must upload your completed answers **in a single document file** within the allotted time using a compatible file type for Turnitin (Preference: MS Word .doc or .docx format) **It is the student's responsibility to check that the file has uploaded correctly in Turnitin. No late exams will be accepted.**

Academic integrity

Students are reminded of the declaration that they agree to when submitting this exam paper via Turnitin. I declare that this work:

- upholds the principles of academic integrity as defined in the Academic Misconduct Rules;
- is original, except where collaboration (for example group work) has been authorised in writing by the course convener in the course outline and/or Wattle site;
- is produced for the purposes of this assessment task and has not been submitted for assessment in any other context, except where authorised in writing by the course convener;
- gives appropriate acknowledgement of the ideas, scholarship and intellectual property of others insofar as these have been used;
- in no part involves copying, cheating, collusion, fabrication, plagiarism or recycling.

There are 8 questions in total
(Q1-Q8)

Please name your submission as
ENGN6528_exam_u1234567.docx

Questions on the next page

Q1: Basic Concepts (12 marks)

Answer the following questions concisely. Each question must be answered in no more than 5 lines of text. Longer answers will be penalized.

a. According to David Marr, computer vision can be separated into three layers, low-level, mid-level and high-level vision. Give one example each for low-level and mid-level vision tasks. [1 marks]

b. Consider the HSV colour space. What does H, S, and V stand for? Describe the information that the H channel contains. [1 mark]

c. Define outliers of a given dataset. Give two examples of robust cost functions that can be used in the presence of outliers. [2 marks]

d. Let A be a matrix of k rows and n columns. Briefly describe the Singular Value Decomposition (SVD) of A and clearly explain the components of this decomposition along with their dimensions. [2 marks] Can the rank of matrix A be determined by SVD? If yes, how? [1 mark]

e. Briefly explain the terms precision, recall, and average precision for a binary classifier. [3 marks]

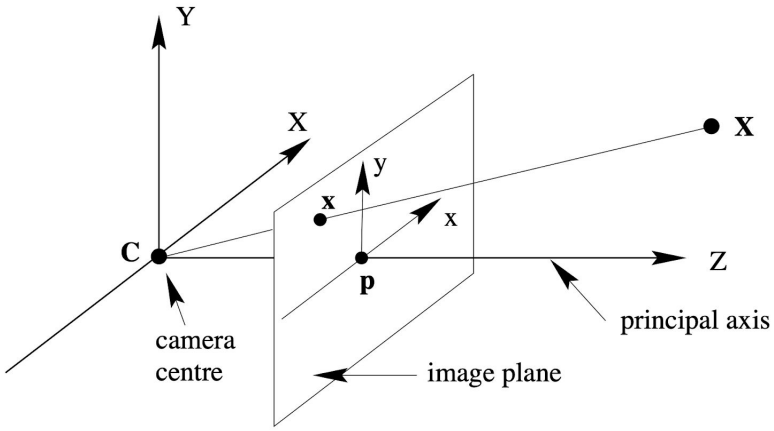
f. Explain the difference between generative models and discriminative models. Give an example application of generative models in computer vision. [2 marks]

Q2: 3D SFM and Image Formation (28 marks)

Answer the following questions concisely. Write down working, and if you are unsure about some part along the way, state your best assumption and use it for the remaining parts. Similarly, if you think some aspect is ambiguous, state your assumption and write the answer as clearly as you can.

a. Given two calibrated cameras, C_1 and C_2, with a focal length of 500 in x and 375 in y directions and the cameras have resolution 1024 x 1024 with no skew. Suppose the camera centre of C_1 projected to image is at (499, 501) and the camera centre of C_2 projected to the image is at (500, 500). Write down the calibration matrix K_1 and K_2 for C_1 and C_2 respectively. [2 marks]

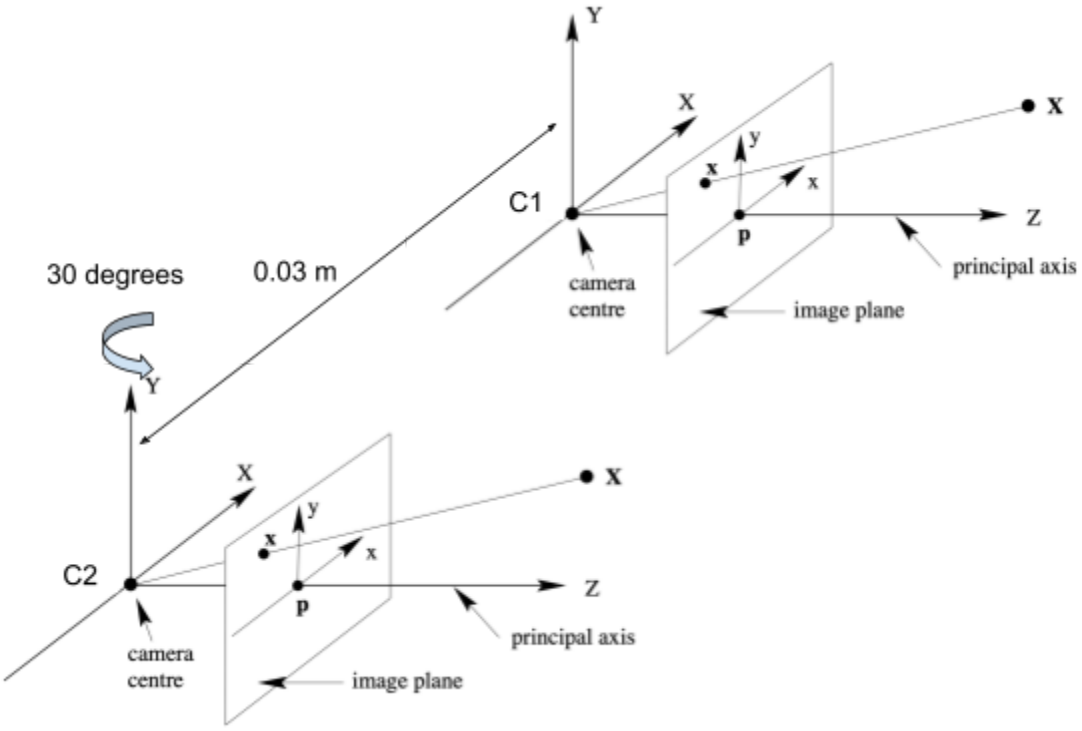
b. Suppose that a 3D world coordinate system (X, Y, Z) is aligned with the camera coordinate system of C_1. More specifically, the world origin is at the camera centre of C_1, the Z axis is aligned with the optical axis and the X and Y world coordinate systems aligned parallel with the x and y axes of the image of C_1. Write down the matrices $K[R|t]$ which define the projection of a point in the world coordinate system to the image of C_1. [2 marks]



c. Suppose that the scene has a point, P_1, that in the world coordinate system defined above that lies at (0.235, 0.24, 1.0). Note that the points in the world coordinate system are measured in m. What location (to the nearest pixel) will that world point (P_1) map to in the image of C_1? [2 marks]

d. Suppose that with respect to the world coordinate system that is aligned with camera C_1, camera C_2 is rotated by 30 degrees about its vertical axis (Y axis) and the centre of C_2 is translated by 0.03 m to the left of C_1 (along the negative direction of X axis of C_1). Both the camera centres

remain on the same (X, Z) plane. The arrangement is shown below. Write down the matrices $K[R|t]$, which define the projection of points in the world system to the image of C_2. [3 marks]



e. What is the location (to the nearest pixel) that P_1 given above maps to in the image of Camera C_2? [2 marks]

f. For camera C_1, there is an epipole (or epipolar point) that relates to camera C_2 which is the image of C_2 in camera C_1. For the above two-camera set up what is the position of the epipole in camera C_1 of camera C_2? [2 marks]

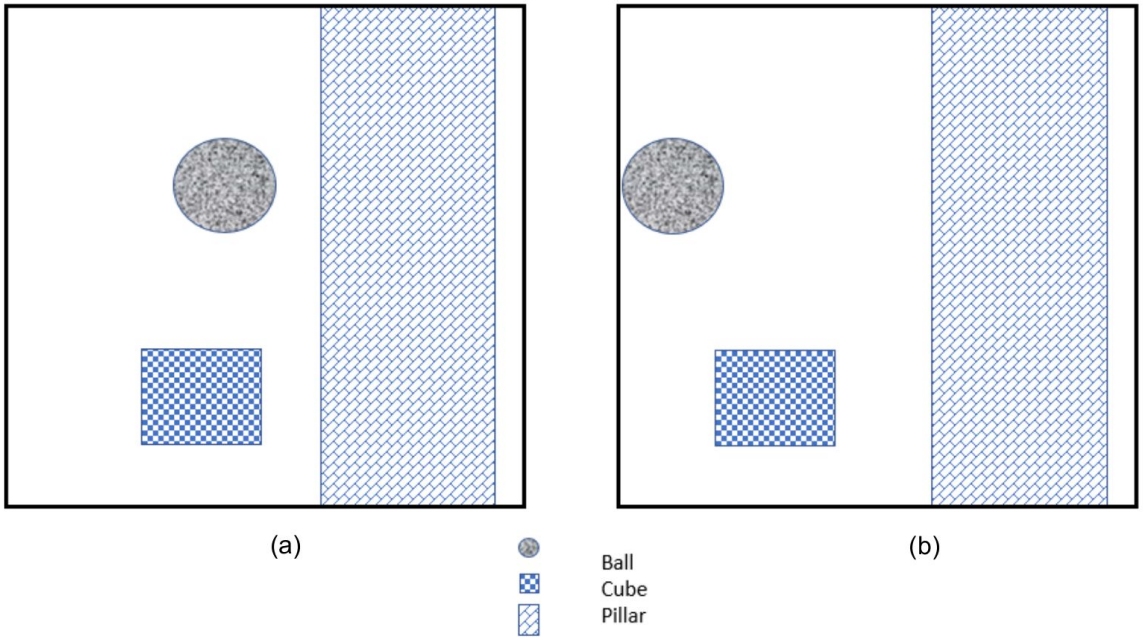
g. Determine the fundamental matrix in the above two camera setup. [3 marks]

h. Given a point P_2 that appears in camera C_1 at image location (x_1, y_1), and in camera C_2 at image location (x_2, y_2). How would you find the world coordinates of point P_2? [4 marks]

i. The normalization step in the 8-point algorithm is crucial in estimating the fundamental matrix from point correspondences. Briefly describe the normalization step and explain why it is crucial. [3 points]

j. The fundamental matrix has rank = 2. However, if this constraint is not enforced when estimating it, what would be the effect in the resulting epipolar lines? [2 marks]

k. The images (a and b) shown below are the left and the right image of an ideal stereo pair, taken with two identical cameras (A and B) mounted at the same horizontal level and with their optical axes parallel.

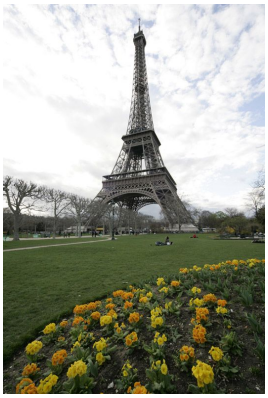


Draw a planar-view (i.e., a top-down bird-eye's view) of the scene showing roughly what the spatial arrangements of the three objects are. Only relative (rather than accurate) positions are required. [3 marks]

Q3: Histogram Modification (5 marks)

Answer the following questions concisely.

a. Consider a histogram modification function $M(v) = v^{\gamma}$ where $v \in [0, 1]$. Suppose that we apply the above transformation to the histogram of the following image with $\gamma = 2.5$ and $\gamma = 0.4$ respectively. Describe the effect on the output image for both of the cases. [2 marks]



b. Briefly explain the steps of the histogram equalization algorithm. [2 marks]

c. What is the effect of applying histogram equalization to an image twice? Justify your answer. [1 mark]

Q4: Image Filtering (10 marks)

Answer the following questions concisely.

a. Consider the following 4 x 5 image (I) with pixel grey values as indicated in the cells and the 3 x 3 filter (f).

I =

3	9	5	2	9
2	4	2	4	7
7	3	8	6	3
2	0	2	6	0

f =

4	2	3
1	2	7
5	6	3

Perform correlation with the filter to the image. Note that to avoid problems at boundaries of the image you only need to calculate the filtered values for the 2 x 3 centred region. [4 marks]

b. Is the following filter separable? If yes, write down the separate components of the following filter. Otherwise, explain why it is not separable. [4 marks]

-3	-6	-9
2	4	6
1	2	3

c. What is the effect of applying each of the following filters to a given image? [2 marks]

f_1 =

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

f_2 =

1/7	1/7	1/7	1/7	1/7	1/7	1/7

Q5: Basic Algorithms (15 marks)

Answer the following questions concisely.

a. The Integral Image is a part of the Viola-Jones face detector that makes feature computation more efficient. Consider the following 4 x 5 image with pixel intensity values as indicated in the cells. Write down the corresponding integral image. [4 marks]

3	9	5	2	9
2	4	2	4	7
7	3	8	6	3
2	0	2	6	0

b. Given the integral image, how can one compute the sum of pixel values in the shaded area in the above image? Show your steps. [2 marks]

c. The Histogram of Oriented Gradients detector applies a window to determine if a pedestrian occurs at a particular location in an image. This is broken up into 15 x 7 individual cells. What information does the descriptor include for each of these cells? [2 marks]

d. Briefly describe the procedure to obtain a set of visual words for a given image. [2 marks]

e. Consider the RANSAC algorithm for fitting a line for a set of points. Let the probability of a point being an outlier $\varepsilon = 0.19$ and the number of points in a sample be $s = 2$. Determine the required number of sampling iterations to ensure that at least one of the sample is free from outliers with probability $p = 0.98$. [2 marks]

f. Briefly describe the Expectation Maximization (EM) algorithm and discuss its differences compared to the K-means clustering. [3 marks]

Q6: Deep Learning (12 marks)

Answer the following questions concisely. Write down your working, and if you are unsure about some part along the way (or some aspect is ambiguous), state your best assumption and use it for the remaining parts. Note that for some questions multiple answers are possible and in such situations state any assumptions you make.

a. Consider that you have given the task of designing a classifier to distinguish 10 different breeds of dog. Briefly discuss the main steps to design such a machine learning system. Give the necessary details and design choices of each of these steps. [4 marks]

b. Represent the following function as a computation graph with elementary nodes such as addition, multiplication, min, and tanh. [2 marks]

$$f(\mathbf{w}, \mathbf{x}) = (1.5 \tanh(\min(w_0 x_0, w_1 x_1)) + w_2))^2$$

where,

$$\tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1} \quad \text{and} \quad \frac{d \tanh(x)}{dx} = 1 - \tanh^2(x)$$

c. Suppose $w_0 = -1$, $w_1 = -2$, $w_2 = 0.75$, $x_0 = 3$, $x_1 = 1$. Calculate the value of f by performing the forward pass and calculate the derivatives of f with respect to w_0, w_1, w_2 via the backpropagation. Show your steps. [4 marks]

d. Briefly describe the residual block with a diagram, which is the building block for a ResNet. [2 marks]

Q7: Algorithm Design - 1 (8 marks)

Answer the following questions concisely and write down your working. If you are unsure about some part along the way (or some aspect is ambiguous), state your best assumption and use it for the remaining parts. Note that for some questions multiple answers are possible and in such situations state any assumptions you make.

a. Consider that you have given a dataset of images D and a set of query images Q. No overlap between D and Q except that they have common objects/scenes. Design an image based search engine. The objective is to find a set of matching images (or closely related images) from the dataset for any given query image. Clearly explain the key steps of your approach and state any assumptions you make. **Hint:** You can use classical approaches or deep learning methods but provide all the necessary information. [6 marks]

b. For a given query image, provide the memory requirement and running time complexity of your method. [2 marks]

Q8: Algorithm Design - 2 (10 marks)

Answer the following questions concisely and write down your working. If you are unsure about some part along the way (or some aspect is ambiguous), state your best assumption and use it for the remaining parts. Note that for some questions multiple answers are possible and in such situations state any assumptions you make.

a. Turn your phone into a GPS in an art museum or a library. GPS usually does not work well in an indoor environment. The goal of designing this algorithm is to localise your position by taking a few images around you in the museum. Please Briefly describe the key steps of your method.

========== END of ALL QUESTIONS in the EXAM ==========