Version RC/4

EGT3
ENGINEERING TRIPOS PART IIB

Wednesday 2 May 2018 2 to 3.40

Module 4F12

COMPUTER VISION

*Answer not more than **three** questions.*

All questions carry the same number of marks.

*The **approximate** percentage of marks allocated to each part of a question is indicated in the right margin.*

*Write your candidate number **not** your name on the cover sheet.*

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed
Engineering Data Book

10 minutes reading time is allowed for this paper at the start of the exam.

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

1 (a) A greyscale image, $I(x,y)$, is to be smoothed and then differentiated as part of the feature detection process.

(i) Explain why smoothing is necessary. [10%]

(ii) Give an expression for computing the intensity of a smoothed image, $S(x,y)$, in terms of two discrete 1-D convolutions. [20%]

(iii) Show how the image gradients, $\partial S/\partial x$ and $\partial S/\partial y$, can also be computed by two discrete 1-D convolutions and identify the filter coefficients. [10%]

(b) Consider an algorithm to detect and localise *corner* features in a 2-D image by evaluating the *auto-correlation matrix*, A, at each pixel. The 2×2 matrix of weighted (smoothed) intensity gradients, A, is defined as follows:

$$\mathrm{A} \equiv \begin{bmatrix} \langle S_x^2 \rangle & \langle S_x S_y \rangle \\ \langle S_x S_y \rangle & \langle S_y^2 \rangle \end{bmatrix}$$

where $S_x \equiv \partial S/\partial x$, $S_y \equiv \partial S/\partial y$ and $\langle\ \rangle$ denotes a 2-D weighting (smoothing) operation.

(i) The auto-correlation matrix, A, at a point in the image can be derived by considering the intensity differences between two patches of pixels in the smoothed image. The first image patch, W, is centred at pixel $\mathbf{x}$ so that $S(x,y) = S(\mathbf{x})$. The second patch is defined by displacing the centre of the first patch by a small amount, $\mathbf{n}$. Show how the auto-correlation matrix, A, is related to the weighted sum of squared differences, $C(\mathbf{x},\mathbf{n})$, between the two patches of pixels of the smoothed image:

$$C(\mathbf{x},\mathbf{n}) = \sum_{\mathbf{x}\in W} w(\mathbf{x})(S(\mathbf{x}+\mathbf{n}) - S(\mathbf{x}))^2 \approx \sum_{\mathbf{x}\in W} w(\mathbf{x})(\nabla S(\mathbf{x}).\mathbf{n})^2$$

[20%]

(ii) How are the 2-D weighted (smoothed) values obtained in practice? Give details of the weighting function and its size. [15%]

(iii) Explain how the determinant and trace of matrix A can be analysed to detect corner features. Give details of the Harris corner detection algorithm. [25%]

2 The relationship between a 3-D world point $\mathbf{X} = (X, Y, Z)^\top$ and its corresponding pixel at image co-ordinates, (u, v), under perspective projection can be written using *homogeneous* co-ordinates by a *projection* matrix:

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

(a) Under what assumptions is this relationship valid? [10%]

(b) For a *calibrated* camera the elements of the projection matrix are known. Show how a measured image point, (u, v), can be used to determine a ray in 3-D which is defined by the intersection of two planes. Give algebraic expressions for these two planes. [20%]

(c) For an *uncalibrated* camera the elements of the projection matrix, p_{jk}, are unknown. Show how a set of N known reference 3-D points, $\{\mathbf{X}_i\}_{i=1}^N$, and their corresponding image points, $\{(u_i, v_i)\}_{i=1}^N$, can be used to estimate the projection matrix.
Include details of the optimisation techniques used when the measurements are noisy and there are a large number of 3-D reference points. [30%]

(d) *Weak perspective* projection consists of an *orthographic* projection onto a plane which is parallel to the image plane followed by a perspective projection onto the image plane.

(i) Show that under weak perspective, the scaling due to depth can be assumed to be constant. [10%]

(ii) Hence derive the projection matrix between a 3-D point and its image under weak perspective projection. [20%]

(iii) Under what viewing conditions is weak perspective a good camera model? What are its advantages? [10%]

3 Consider multiple views of a static scene which have been taken with a single camera. Corresponding points in a pair of images, (u,v) and (u',v'), are found by matching interest points extracted in each view.

(a) When viewing a planar object the correspondences in the two views can be described by the 2-D *projective transformation* given below.

$$\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

(i) Describe a method for finding possible matches. [10%]

(ii) How many point correspondences are required to estimate the transformation? [10%]

(iii) Describe the RANSAC (Random Sample Consensus) algorithm for finding consistent matches in the presence of incorrect or outlier measurements. [10%]

(iv) How is the transformation estimated when a large number of consistent matches is available? [10%]

(b) In stereo vision the correspondences in the two views will be described by the *fundamental matrix* shown below.

$$\begin{bmatrix} u' & v' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0$$

(i) What is meant by the *epipolar constraint*? Give an algebraic expression for the epipolar line corresponding to a point in the left image with pixel coordinates (u,v). [20%]

(ii) What additional constraints must the elements of the fundamental matrix satisfy? [10%]

(iii) Describe an algorithm to recover the camera motion (rotation and translation) between the two views from a fundamental matrix if the internal camera parameters, $\mathbf{K}$, are known. [20%]

(iv) How are the 3-D co-ordinates of the visible points computed? [10%]

4 (a) The basic building block of a convolutional neural network comprises three stages: a convolutional stage, a non-linear stage, and a pooling stage. Define each stage mathematically and explain the rationale behind their design. [30%]

(b) A convolutional neural network has been trained to return the probability, $x(Z,\theta) = p(t=1|Z,\theta)$, that there is a pedestrian in an image. Here x is the scalar valued output of the network, Z denotes the input image, θ are the parameters of the network and $t=1$ indicates that a pedestrian is present. Experiments on test data show that these probabilistic predictions are poorly calibrated: the network sometimes returns highly confident predictions that are wrong.

In order to improve the calibration of the network the trained network's logistic output layer is augmented with a parameter γ to return a modified probabilistic output

$$x'(Z,\theta,\gamma) = p'(t=1|Z,\theta,\gamma) = \frac{1}{1+\exp(-\gamma\,\mathbf{w}^\top \mathbf{h}(Z;\theta))}$$

where $\mathbf{w}$ are the output weights of the unmodified network and $\mathbf{h}(Z;\theta)$ the final hidden layer activations.

(i) Describe what happens to the output of the new network as γ is swept from zero to infinity and therefore argue how an appropriate setting might improve calibration. [20%]

(ii) A separate dataset comprising M unseen images $\{Z^{(m)}\}_{m=1}^{M}$ and binary labels $\{\mathrm{t}^{(m)}\}_{m=1}^{M}$ will be used to train γ, whilst the original parameters are fixed. Write down the log-likelihood for γ and derive the gradients required for training. [40%]

(iii) The poor calibration of the original network is thought to be result of overfitting during the first training stage. Explain why the second stage of training is likely to improve the calibration of the network. [10%]

END OF PAPER

THIS PAGE IS BLANK