Version RC/4

EGT3
ENGINEERING TRIPOS PART IIB

Thursday 28 April 2016 2 to 3.30

Module 4F12

COMPUTER VISION AND ROBOTICS

*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.

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 grey scale image, $I(x,y)$, is smoothed before image gradients are computed as part of the feature detection and matching process.

(i) What smoothing filter is used in practice? Give an expression for computing the intensity of a smoothed pixel, $S(x,y)$, with two discrete 1D convolutions. [20%]

(ii) How are the 2D image gradients computed? Show how differentiation can also be performed by two discrete 1D convolutions and identify the filter coefficients. [10%]

(iii) Show how different resolutions of the image can be represented efficiently in an *image pyramid*. Your answer should include details of the implementation of smoothing within an octave and sub-sampling of the images between octaves. [20%]

(b) Consider an algorithm to detect and match image features in a 2D image.

(i) Show how image features such as *blob-like* shapes can be localised in position and scale by *band-pass* filtering. How are these features localised efficiently using differences in the images of an image pyramid? [30%]

(ii) How is the dominant orientation of each feature determined? [20%]

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

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

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

(ii) A camera is to be calibrated by observing the images (u_i, v_i) of known reference points (X_i, Y_i, Z_i). Derive two linear equations in the unknown elements p_{ij} of the projection matrix. [20%]

(iii) Show how the unknown elements, p_{ij}, are estimated in practice. Give details of the optimisation techniques used when measurements are noisy and there are a large number of 3D reference points and their image correspondences. [30%]

(iv) Show how to recover the camera position, orientation and focal length from the projection matrix. [10%]

(b) A *weak perspective* projection comprises an *orthographic* projection onto a plane which is parallel to the image plane followed by perspective projection onto the image plane.

(i) Derive the projection matrix between a 3D point and its image under weak perspective projection. [20%]

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

3 A mobile phone camera is used to reconstruct the 3D shape of an unknown object by taking two photographs from distinct viewpoints. Corresponding points in the pair of photographs, (u, v) and (u', v'), are found by matching interest points extracted in each view.

(a) The SIFT descriptor is used to find correspondences.

(i) Describe the main steps in computing this descriptor for each interest point. [20%]

(ii) What properties of the neighbourhood of each feature is this descriptor encoding? [10%]

(iii) Give details of an algorithm used to find the correct correspondence. [10%]

(b) Assume that the mobile phone's camera (internal) parameters are known and represented by a camera calibration matrix **K**. The camera rotation and translation between viewpoints are unknown and can be represented by a rotation matrix, **R**, and translation vector, **T**, respectively.

(i) Show how to recover the relative orientation, **R**, and position, **T**, of the camera as it moves around the object from point correspondences. [30%]

(ii) Give expressions for the *projection* matrices of each viewpoint. [10%]

(iii) How are the 3D positions of the visible points computed? Give details and show how to make a more accurate and complete reconstruction. [20%]

4 A convolutional neural network is to be used for estimating the age of a person from an image of their face. A labelled dataset has been collected that contains N greyscale face images $\{Z^{(n)}\}_{n=1}^{N}$ and labels $\{\mathrm{t}^{(n)}\}_{n=1}^{N}$. The label is the age of the person in each image.

The network contains three stages. The first stage carries out a 2D convolution between the image pixels $Z_{i,j}^{(n)}$ and convolutional weights $W_{i,j}$,

$$a_{i,j}^{(n)} = \sum_{k,l} W_{k,l} Z_{i-k,j-l}^{(n)}.$$

The second stage applies a point-wise non-linearity $y_{i,j}^{(n)} = f\left(a_{i,j}^{(n)}\right)$.

The third stage applies a set of output weights $V_{i,j}$ in order to form the scalar output of the network,

$$x^{(n)} = \sum_{i,j} V_{i,j} y_{i,j}^{(n)}.$$

The network's weights will be trained using the following objective function,

$$G(V,W) = \frac{1}{2\sigma^2} \sum_{n=1}^{N} \left(\mathrm{t}^{(n)} - x^{(n)}\right)^2 + \frac{\alpha}{2} \sum_{i,j} V_{i,j}^2 + \frac{\beta}{2} \sum_{i,j} W_{i,j}^2.$$

(a) Provide an interpretation for the network's output in terms of probability distributions and use this to justify the form of the objective function. [20%]

(b) Describe how to train the network's convolutional weights W using gradient descent. Compute the derivative required to implement gradient descent. Simplify your expression and interpret the terms. [40%]

(c) Describe enhancements to the architecture of the network that might improve its ability to estimate the age of a person from an image of their face. [40%]

END OF PAPER

THIS PAGE IS BLANK