ENGINEERING TRIPOS PART IIB

Wednesday 21 April 2010 9 to 10.30

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.*

There are no attachments.

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS
Engineering Data Book
CUED approved calculator allowed

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

rc04

1 (a) List five factors which influence the intensity $I(x,y)$ of a monochrome CCD image. [10%]

(b) The filter kernel $\frac{1}{a}[\ 1\ ,\ 4\ ,\ 6\ ,\ 4\ ,\ 1\ ]$ is used to approximate the 1D Gaussian function for smoothing a row of pixels $I(x)$.

(i) What is the correct value for a? [10%]

(ii) Give an expression for computing the intensity of a smoothed pixel, $S(x)$. [20%]

(iii) Show how 1D smoothing with this kernel is equivalent to successively averaging neighbouring pixels (i.e., repeated convolution of the image with the kernel $[\ \frac{1}{2}\ ,\ \frac{1}{2}\ ]$). How many times must neighbouring pixels be averaged? [30%]

(c) Describe an algorithm to localise the position and to recover the orientation of an *edge* in an image $I(x,y)$ using spatial derivatives of the smoothed image intensities, $S(x,y)$. [30%]

2 The relationship between a 3D world point (X, Y, Z) and its corresponding pixel at image coordinates (u, v) can be written using a 3×4 camera *projection matrix* as follows:

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

(a) Under what assumptions is this relationship valid? Explain the algebraic and geometric significance of s and W. [20%]

(b) Describe how to calibrate a camera from a single perspective image of a 3D object from N image measurements (u_i, v_i) of known reference points (X_i, Y_i, Z_i) where $i = 1, 2, \ldots N$. Your answer should include details of:

(i) the minimum number of calibration points needed and why more are preferred in practice;

(ii) the derivation of linear equations in the unknown projection parameters p_{jk};

(iii) how the set of linear equations is solved;

(iv) how and why the linear solution can be improved;

(v) how the position, orientation and focal length of the camera can be recovered.

[50%]

(c) Find the *vanishing point* (projection in the image plane) of a set of parallel lines with direction orientation (b_1, b_2, b_3). [30%]

3 A static scene is observed twice by the same camera, producing a pair of images with pixel correspondences (u, v) and (u', v').

(a) Under what conditions will correspondences in the two views be described by a 2D *projective transformation*:

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

[30%]

(b) How many degrees of freedom does the 2D projective transformation have? An object appears as a square in the first image. Describe, using sketches, how it might appear in the second image. Account for each degree of freedom of the 2D projective transformation. [30%]

(c) Outline an algorithm for detecting and localising *interest points* in both images by *band-pass* filtering. Include details of the assignment of a scale and orientation to each feature and describe a suitable descriptor which can be used to match the image features between the two views. [40%]

4 In stereo vision a point has 3D coordinates $\mathbf{X}_c$ and $\mathbf{X}'_c$ in the left and right camera coordinate systems respectively. The rotation and translation between the two coordinate systems are represented by a matrix $\mathbf{R}$ and vector $\mathbf{T}$ with $\mathbf{X}'_c = \mathbf{R}\mathbf{X}_c + \mathbf{T}$

(a) What is meant by the *epipolar geometry* between the two views and derive an expression for the *fundamental matrix* in terms of the rotation matrix $\mathbf{R}$ and translation vector $\mathbf{T}$ and internal calibration parameter matrices of the left and right cameras, $\mathbf{K}$ and $\mathbf{K}'$, respectively.

Your answer should include an algebraic represenation of the *epipolar line* for a point in the left image with pixel coordinates (u, v) in terms of the fundamental matrix. [50%]

(b) Explain how the fundamental matrix can be estimated from point correspondences between the stereo views. How do you ensure that the matrix has the desired rank? [30%]

(c) The fundamental matrix can also be computed between two views taken with a single moving camera. What additional information is required to recover the camera motion? How can the 3D positions of points in the scene be recovered? [20%]

5 Consider the following two applications of computer vision technology for 3D shape acquisition and human-machine interaction.

(a) A single digital camera is to be used to acquire the 3D shape of a small object. Assuming the object can be placed on a calibration pattern and viewed from many viewpoints describe a practical system to acquire an accurate 3D model. Your answer should outline the key steps in image acquisition, calibration, feature correspondence and 3D shape output. [50%]

(b) A video camera is mounted on top of a computer in order to detect the user's hand position and orientation. Describe a vision system that uses *templates* to detect the hand. Your answer should explain how the templates can be acquired from sample images, and how hypotheses can be evaluated efficiently using a suitable distance measure computation and preprocessing of the images. The recognition system is required to work independently of lighting conditions and small changes in viewpoint. [50%]

END OF PAPER