ENGINEERING TRIPOS PART IIB

Thursday 28 April 2005 2.30 to 4

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.

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

(TURN OVER

1 (a) Why do many computer vision systems start by extracting edges from the raw images? When would detecting corners, or features of interest, be more appropriate? [20%]

(b) Describe how intensity discontinuities in an image, $I(x, y)$, can be detected and localised by convolving the image with a discrete version of the Laplacian of a Gaussian. [40%]

(c) Show how corners, or features of interest, can be detected and localised by examining the eigenvalues of the 2×2 matrix

$$\begin{bmatrix} \langle I_x^2 \rangle & \langle I_x I_y \rangle \\ \langle I_x I_y \rangle & \langle I_y^2 \rangle \end{bmatrix}$$

evaluated at each pixel, where $\langle\,\rangle$ denotes a 2D smoothing operation and where $I_x \equiv \partial I / \partial x$ and $I_y \equiv \partial I / \partial y$. You should also state the algorithm for labelling a pixel as a corner or feature of interest. [40%]

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? Comment on the algebraic and geometric significance of s and W. [15%]

(b) Derive expressions for the vanishing point in the image plane of lines parallel to the X axis. [15%]

(c) Show how the projection matrix can be decomposed into a product of matrices that contain elements expressed in terms of the internal (focal length, principal point, pixels per unit length) and external (position and orientation) camera parameters. [20%]

(d) A camera is to be calibrated from a single perspective image of a known 3D object from the image measurements (u_k, v_k) of known reference points (X_k, Y_k, Z_k). By first deriving linear equations in the unknown elements p_{ij} of the projection matrix, outline an algorithm to recover the projection matrix. How many reference points, $k = 1 \ldots N$, are required to estimate all the elements p_{ij}? State clearly how noisy image measurements are processed in practice. [50%]

3 (a) A 2D projective transformation can be described algebraically with homogeneous coordinates by:

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

(i) Describe, using sketches, how a square might appear after the transformation. Be sure to account for each degree of freedom of the 2D projective transformation. [30%]

(ii) List the invariants of the 2D projective transformation. [20%]

(b) A *mosaiced panorama* of a scene is acquired by rotating a camera about its optical centre.

(i) Give an expression for the transformation between correspondences in two successive images and show how this transformation depends on the camera rotation between the two views. [20%]

(ii) Explain how this transformation can be estimated in practice. Include details of the localisation and matching of image features. [30%]

4 In stereo vision a point has 3D coordinates $\mathbf{X_C}$ in the left camera's coordinate system and $\mathrm{R}\mathbf{X_C} + \mathbf{T}$ in the right camera's coordinate system.

(a) Derive an expression for the *epipolar constraint* in terms of the rotation matrix R and translation vector $\mathbf{T}$ and internal calibration parameter matrices of the left and right cameras, K and K', respectively. [20%]

(b) Derive an algebraic expression for the *epipolar line* for a point in the left image with pixel coordinates (u, v) in terms of the *fundamental matrix*. [20%]

(c) Explain how the fundamental matrix can be estimated from point correspondences between the stereo views. What special property must the estimated matrix have? [30%]

(d) How can the left and right camera projection matrices be recovered from a fundamental matrix? What additional information is required in order to recover 3D positions from image correspondences from a pair of uncalibrated cameras? [30%]

(TURN OVER

5 Answer any **two** of the following four parts.

(a) Describe an algorithm to find stereo correspondences in two images. You may assume that the images have been rectified. [50%]

(b) Describe how a single surveillance camera can be used to detect and track people moving in a room. Include details of the calibration and the image processing required to detect and track the people. [50%]

(c) A video camera is mounted on top of a computer in order to detect the user's hand position and orientation. Outline a template-based vision system that can be used to detect the hand. 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 should be made to work independently of lighting conditions and small changes in viewpoint. [50%]

(d) Describe how a stereo pair of uncalibrated cameras can be used to guide a robot manipulator as it attempts to pick up objects in its workspace. You may assume that the cameras are placed at some distance from the workspace. Include details of calibration and the visual tracking of the robot's gripper. [50%]

END OF PAPER