ENGINEERING TRIPOS PART IIB
ELECTRICAL AND INFORMATION SCIENCES TRIPOS PART II

Thursday 24 April 2003 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.*

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) List five factors which influence the intensity $I(x, y)$ of a monochrome CCD image. Explain why edge detection is commonly used in computer vision applications. [30%]

(b) The Marr–Hildreth edge detector convolves the image with a discrete version of the Laplacian of a Gaussian and then localises edges at the resulting zero-crossings. In contrast, the Canny edge detector first establishes the orientation of the edge and then searches for a local maximum of the intensity gradient normal to the edge. What are the advantages and disadvantages of the Marr–Hildreth edge detector compared with the Canny edge detector? [30%]

(c) The spatial derivatives of an image, $I(x, y)$, are used in edge and *feature of interest* detection. Show why 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 $I_y \equiv \partial I/\partial y$. [40%]

2 (a) Derive the equation of the *vanishing line* of parallel planes with a normal $\mathbf{n} = (n_x, n_y, n_z)$ when viewed with a pin-hole camera under perspective projection. [30%]

(b) Outline an algorithm to recover the position, orientation and internal camera parameters of a CCD camera (see Fig. 1 for definitions of all unknowns) with no non-linear lens distortion from a single perspective image of a known 3D object. You should state clearly the number of image measurements required and how noisy image measurements are processed in practice. [50%]

(c) Under what viewing conditions is the relationship between the CCD image coordinates and the 3D object/world coordinates linear? [20%]

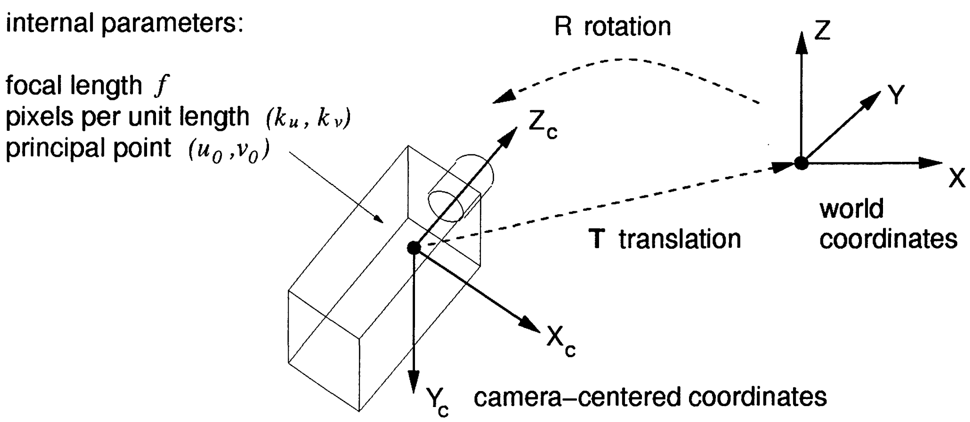

Fig. 1

(TURN OVER

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

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

(i) How many degrees of freedom does the 2D projective transformation have? [10%]

(ii) 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%]

(iii) Consider the conic $au^2 + buv + cv^2 + du + ev + f = 0$. Derive the equation of the conic after applying the projective transformation. [20%]

(b) A *mosaic* 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. [20%]

4 (a) What is meant by the *epipolar constraint* in stereo vision? What is the *fundamental matrix*? [20%]

(b) If 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, derive an expression for the fundamental matrix 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. [30%]

(c) Explain how the fundamental matrix can be estimated from point correspondences between the stereo views. What special property must the estimated matrix have? [20%]

(d) Give algebraic expressions which can be used for finding the left and right *epipoles* and the *epipolar line* for a point in the left image with pixel coordinates (u, v). [20%]

(e) What additional information is required in order to recover 3D positions from image correspondences from a pair of uncalibrated cameras? [10%]

(TURN OVER

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

(a) 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 the people and the visual tracking. [50%]

(b) A video camera is to be mounted on the front of a car to aid the driver to avoid obstacles. Outline a template-based vision system that can be used to detect pedestrians walking in front of it. 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 independent of lighting conditions and small changes in viewpoint. [50%]

(c) In a video-conferencing application, it is required to determine where on a screen a user is pointing. Show how this system can be implemented using an uncalibrated stereo pair of cameras. Outline briefly how the user's arm can be detected and tracked in the images. [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