ENGINEERING TRIPOS PART IIB

Thursday 24 April 2008 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.

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

1 A commonly used 1D smoothing filter in computer vision and image processing is the Gaussian: $g_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$ where σ determines the size of the filter.

(a) Show that repeated convolutions with a series of 1D Gaussians, each with a particular standard deviation σ_i, is equivalent to a single convolution with a Gaussian of variance $\sum_i \sigma_i^2$. [30%]

(b) The first stage in localising *edges* and *blobs* at different scales in images is to low-pass filter an image with smoothing filters of different scales. Show how these computations can be performed efficiently. [40%]

(c) Explain how blobs are localised in the image and how the appropriate scale is chosen. [30%]

2 (a) An image is formed by perspective projection onto an image plane, as shown in Fig. 1. The image plane is sampled by a CCD array with k_u pixels per unit length in the X_c direction and k_v pixels per unit length in the Y_c direction. Show that the relationship between a point (X_c, Y_c, Z_c) and its image co-ordinates (u, v) (in pixels) can be expressed in homogeneous co-ordinates by

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} k_u f & 0 & u_0 & 0 \\ 0 & k_v f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$ [50%]

where f is the focal length and (u_0, v_0) is the principal point.

(b) A *weak perspective* projection comprises an orthographic projection onto the plane $Z_c = Z_A$ followed by perspective projection onto the image plane. Derive the homogeneous relationship between a point (X_c, Y_c, Z_c) and its image (u_A, v_A) under weak perspective projection. [30%]

(c) Under what viewing conditions is weak perspective a good camera model and what are the advantages of using this simpler model? [20%]

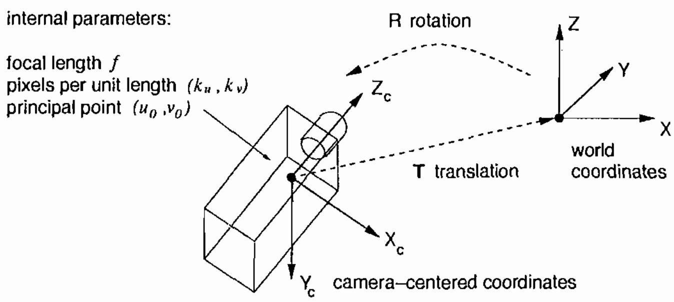

Fig. 1

(TURN OVER

3 A single camera views a world plane in the scene. The transformation between points (X, Y) in the plane and pixels (u, v) in the image is described by a 2D *projective* transformation.

(a) Under what viewing conditions will the mapping from world plane to image plane be a 2D linear (affine) transformation? How many degrees of freedom does this transformation have? Explain their geometric significance by showing their effects on the transformation of a unit square in the world plane. [20%]

(b) What do the additional degrees of freedom of the planar projective transformation specify? [20%]

(c) Explain how the transformation from world to image plane can be estimated in practice. Your answer should include details of the localisation of image features, suitable descriptors for matching and an algorithm to reject incorrect matches. [30%]

(d) For a camera that moves relative to the plane, derive an expression for the transformation between correspondences in two successive views and show how this depends on the camera motion. [30%]

4 (a) What matching constraints can be used to find point correspondences in stereoscopic vision? [30%]

(b) Explain what is meant by the *epipole* and describe the *epipolar constraint.* [20%]

(c) Derive expressions for the left (u, v) and right (u', v') pixel positions of a point in space viewed through *weak* perspective cameras. [20%]

By eliminating the world coordinates, show that the pixel coordinates are related by

$$[\, u' \;\; v' \;\; 1 \,]\, F_A \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0$$

where F_A is a matrix which has maximum rank 2 and can be expressed in the form:

$$F_A = \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & 1 \end{bmatrix}$$

[30%]

(TURN OVER

5 (a) A video sequence is taken with a single camera undergoing general 3D motion in an arbitrary scene. A set of image features and point correspondences from frame to frame are automatically extracted from the sequence. Describe an algorithm to recover the scene structure and camera motion from these image measurements. [50%]

(b) Give details of an algorithm to detect a known 2D pattern from an image taken with a mobile phone which is equipped with a camera. Discuss how the image features are extracted and matched and how the pattern is localised in the image. [50%]

END OF PAPER