ENGINEERING TRIPOS PART IIB

Monday 7 May 2012 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

rc03

1 Consider an algorithm to detect interest points (features of interest) in a 2D image for use in matching.

(a) Show that the weighted sum of squared differences (SSD) between a patch of pixels (window W) in image $I(x,y) = I(\mathbf{x})$ and another patch of pixels taken by shifting the window by a small amount in the direction $\mathbf{n}$, can be expressed approximately by:

$$C(\mathbf{n}) = \sum_{\mathbf{x}\in W} w(\mathbf{x})(I(\mathbf{x}+\mathbf{n}) - I(\mathbf{x}))^2 \approx \sum_{\mathbf{x}\in W} w(\mathbf{x}) I_n^2$$

where I_n is the intensity gradient in direction $\mathbf{n}$. [20%]

(b) Hence show that the weighted SSD can be represented by:

$$C(\mathbf{n}) = \mathbf{n}^T \mathrm{A}\mathbf{n}$$

where A is a matrix of smoothed intensity gradients defined as follows:

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

where $I_x \equiv \partial I/\partial x$, $I_y \equiv \partial I/\partial y$ and $\langle\rangle$ denotes a 2D smoothing operation. [20%]

(c) How are the directional derivatives, I_x and I_y, computed? Give details of the 1D convolutions required for smoothing and differentiation. [20%]

(d) How are the 2D smoothed (weighted) values obtained? [20%]

(e) Show how A should be analysed to detect corner features and give details of the Harris-Stephens corner detection algorithm. [20%]

2 (a) Show how the relationship between a 3D world point (X,Y,Z) and its corresponding pixel at image coordinates (u,v) under perspective projection can be written using *homogeneous* co-ordinates as follows:

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \left[\begin{array}{ccc|c} & \mathbf{R} & & \mathbf{T} \\ \hline 0 & 0 & 0 & 1 \end{array}\right]$$

and identify the internal (focal length, principal point, pixels per unit length) and external (position and orientation) camera parameters. [40%]

(b) The camera is to be calibrated from a single perspective image of a known 3D object from the image measurements (u_i, v_i) of known reference points (X_i, Y_i, Z_i). Show how each reference point we observe provides the following pair of constraints:

$$u_i = \frac{p_{11}X_i + p_{12}Y_i + p_{13}Z_i + p_{14}}{p_{31}X_i + p_{32}Y_i + p_{33}Z_i + p_{34}}$$

$$v_i = \frac{p_{21}X_i + p_{22}Y_i + p_{23}Z_i + p_{24}}{p_{31}X_i + p_{32}Y_i + p_{33}Z_i + p_{34}}$$

which can be expressed with 12 unknown parameters, p_{jk}. [15%]

(c) Describe desirable properties of the calibration object and show how to recover the unknown camera parameters. [25%]

(d) How should the relationship be modified to model *weak perspective* and under which viewing conditions will this be appropriate? What are the advantages of using the weak perspective projection model? [20%]

3 A *mosaiced panorama* of a scene is acquired by taking multiple images with a mobile phone camera which is rotated about its optical centre.

(a) Show that the transformation between point correspondences in successive images can be expressed as a 2D projective transformation:

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

and identify the dependence of the transformation on the rotation of the camera between images, R, and the camera internal parameters, K. [20%]

(b) A large number of points of interest (*blob-like* features) are detected in each image and potential matches from one image to the other are found by comparing their SIFT descriptors.

(i) How many point correspondences are required to estimate the transformation? [10%]

(ii) How are consistent matches obtained in the presence of incorrect or outlier measurements? Give details of the RANSAC algorithm. [20%]

(iii) How is the transformation estimated when a large number of consistent matches is available? [20%]

(c) What happens to the shape of an image conic under the transformation? Derive the equation of the curve obtained by warping a circle in the first view. [30%]

4 In stereo vision a point has 3D coordinates $\mathbf{X}$ and $\mathbf{X}'$ 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 $\mathbf{X}' = \mathbf{R}\mathbf{X} + \mathbf{T}$ and the internal calibration parameter matrices of the left and right cameras are represented by matrices $\mathbf{K}$ and $\mathbf{K}'$ respectively.

(a) Give expressions for the 3×4 projection matrices for the left and right views. For convenience you should align the left camera and world coordinate systems. [20%]

(b) Show how to recover algebraically the 3D position of a point visible in both views with image coordinates (u, v) and (u', v') in the left and right view respectively. [30%]

(c) What additional geometric constraint needs to be satisfied for consistency and explain what is meant by the *epipolar constraint* for point correspondences. [20%]

(d) Show how the *epipolar constraint* can be expressed algebraically by the equation:

$$\begin{bmatrix} u' & v' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0$$

and identify the relationship between the elements of the matrix and the camera projection matrices. [30%]

END OF PAPER