ENGINEERING TRIPOS PART IIB

Tuesday 23 April 2013 2 to 3.30

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

rc03

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

(a) The image is first smoothed with a low-pass filter before image gradients are computed. What smoothing filter is used in practice? Give an expression for computing the intensity of a smoothed pixel efficiently. [20%]

(b) What is meant by a *scale-space* representation of an image? Show how different resolutions of the image can be represented efficiently in an *image pyramid*. Your answer should include details of the implementation of smoothing within an octave and subsampling of the image between octaves. [20%]

(c) How can *band-pass* filtering at different scales be implemented efficiently using the image pyramid? [10%]

(d) Show how image features such as *blob-like* shapes can be localized in both position and scale. How can the orientation of the image feature be assigned from image gradients in the neighbourhood of the feature? [30%]

(e) Hence show how the neighbourhood of each image feature can be normalised to a 16×16 patch of pixels and give details of a suitable descriptor for use with matching image features in different images. [20%]

2 (a) Under what assumptions can the relationship between a 3D world point (X,Y,Z) and its perspective projection with image coordinates (u,v) be written as follows:

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

and identify the internal (focal length, principal point, pixels per unit length) camera parameters. [20%]

(b) Show that, under perspective projection, parallel lines in space meet at *vanishing points* in the image, where the locations of the vanishing points depend only on the orientations of the lines in space. Why are the vertical lines in Renaissance paintings usually drawn without vanishing points? [20%]

(c) *Weak perspective* projection comprises an orthographic projection onto the plane $Z = Z_A$ followed by perspective projection onto the image plane.

(i) Show that the error $(u - u_A, v - v_A)$ introduced by the weak perspective approximation is given by

$$\left((u - u_0)\frac{\Delta Z}{Z_A}, \, (v - v_0)\frac{\Delta Z}{Z_A} \right)$$

where $\Delta Z \equiv Z_A - Z$. [30%]

(ii) Under what viewing conditions is weak perspective a good camera model? [10%]

(iii) Show how to calibrate the weak perspective projection matrix for a camera in an arbitrary position. [20%]

3 A planar object is viewed from multiple viewpoints with a mobile phone camera.

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

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

and identify the dependence of the transformation on the translation, **T**, and rotation, **R**, of the camera's movement between images and the camera internal parameters, **K**. [20%]

(b) The transformation is estimated from point correspondences.

(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 (Random Sample Consensus) algorithm. [20%]

(iii) How is the transformation estimated when a large number of consistent matches is available? [10%]

(c) What additional information is required to recover the position and orientation of the camera in each viewpoint? Give details on how to recover the motion. [20%]

(d) Consider a planar object with bilateral symmetry. What happens to the symmetry properties when the object is viewed by the mobile phone? [20%]

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 $\mathbf{X}' = \mathbf{RX} + \mathbf{T}$. The internal calibration parameter matrices of the left and right cameras are represented by matrices $\mathbf{K}$ and $\mathbf{K}'$, respectively.

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

(b) 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, respectively. [20%]

(c) Explain how the fundamental matrix can be estimated from point correspondences. [20%]

(d) Give algebraic expressions 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? [20%]

END OF PAPER