ENGINEERING TRIPOS PART IIB

ELECTRICAL AND INFORMATION SCIENCES TRIPOS PART II

Monday 15 January 1996   2 to 3.30

Module I12

COMPUTER VISION AND ROBOTICS

Answer not more than **four** *questions.*

All questions carry the same number of marks.

(TURN OVER

1 (a) A 512×512 grey level image is to be smoothed by convolution with a 2D Gaussian kernel. Show how the convolution can be performed by two 1D convolutions. What computational saving does this achieve when the filter has a standard deviation of 2?

(b) A row of smoothed image pixel intensities is shown below.

48	50	53	56	64	79	98	115	126	132	133

Show that an approximation to the second-order spatial derivatives can be found by convolving with the following kernel:

1	-2	1

Use this kernel to localise any intensity discontinuities in the row of pixels to *sub-pixel* accuracy.

2 (a) A set of parallel planes are viewed under perspective projection. Show that they have a common vanishing line in the image. Derive the relationship between the orientation of the planes in the world and the equation of the vanishing line in the image.

(b) A camera comprises a lens of focal length 20mm and a 10mm$\times$10mm square CCD array. The CCD array is divided into 500×500 square pixels. The pixel at the top left corner of the CCD array has coordinates $(0, 0)$, and the optical axis intersects the CCD array at the pixel with coordinates $(200, 200)$.

For the camera and world coordinate systems shown in Fig. 1, find the 3×4 projection matrix which relates homogeneous pixel coordinates (su, sv, s) to world coordinates (X_w, Y_w, Z_w). To check your answer, use the projection matrix to find the equation of the ray (in *world coordinates*) corresponding to the image point (200,200).

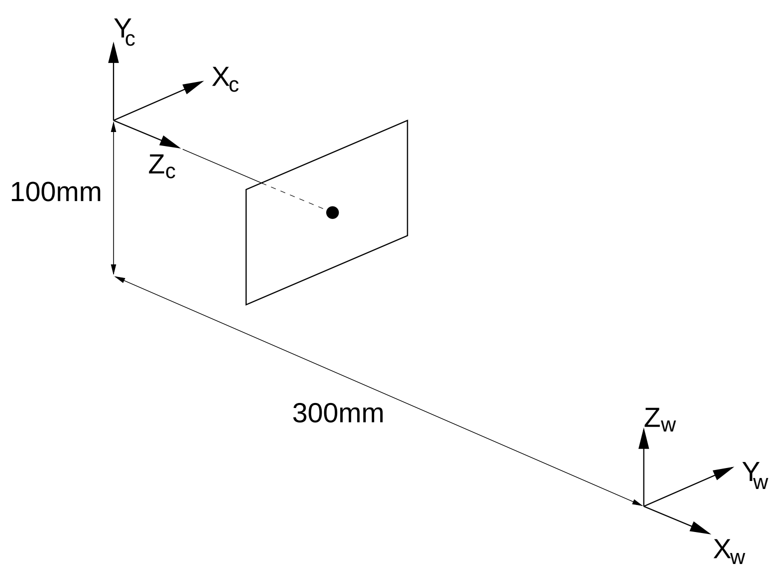


Fig. 1

3 (a) Describe the degrees of freedom and the geometric invariants of the following plane-to-plane transformations:

(i) Euclidean

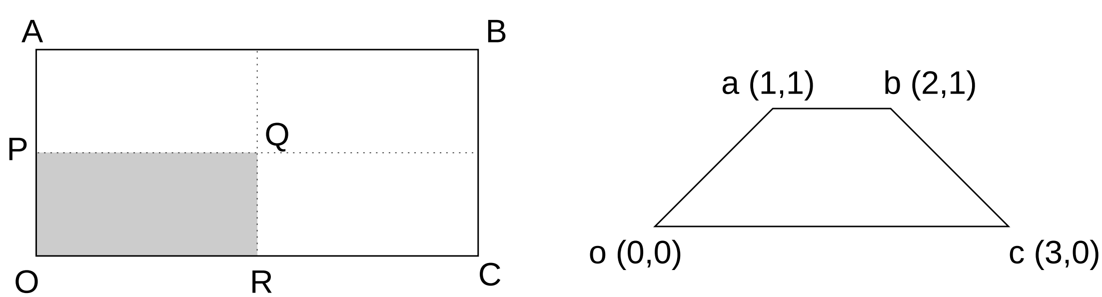
(ii) Similarity

(iii) Affine

(iv) Projective

(b) A video surveillance system views a rectangular car-park to ensure that only authorised vehicles park in the bottom left quadrant. The restricted area is shaded in the plan view of the car-park shown in Fig. 2(a). The boundaries of the car-park appear as a trapezium in the image. The image coordinates of the car park's corners are given in Fig. 2(b).

Find, by a geometric construction or otherwise, the image coordinates of the points P, Q and R.



(a) Plan view (b) Image

Fig. 2

4 (a) List four matching constraints which can be used to find point correspondences in stereo vision.

(b) Two cameras, each with focal length 15mm, are arranged symmetrically as shown in Fig. 3. The two optical axes are coplanar with the baseline and make an angle of 60°. The optical centres are 200mm apart. Measurements in the left and right camera coordinate systems are related by

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & -\sqrt{3}/2 \\ 0 & 1 & 0 \\ \sqrt{3}/2 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} 100\sqrt{3} \\ 0 \\ 100 \end{bmatrix}$$

(i) Sketch the regions in the plane OPO' where the ordering constraint is not obeyed.

(ii) A point in 3D space has left image coordinates (x, y). Derive a constraint (the *epipolar constraint*) on the image coordinates (x', y') of the corresponding point in the right image. Sketch the epipole and the family of epipolar lines in the left image.

(iii) Discuss the advantages and disadvantages of this setup compared with an alternative setup where the image planes are parallel.

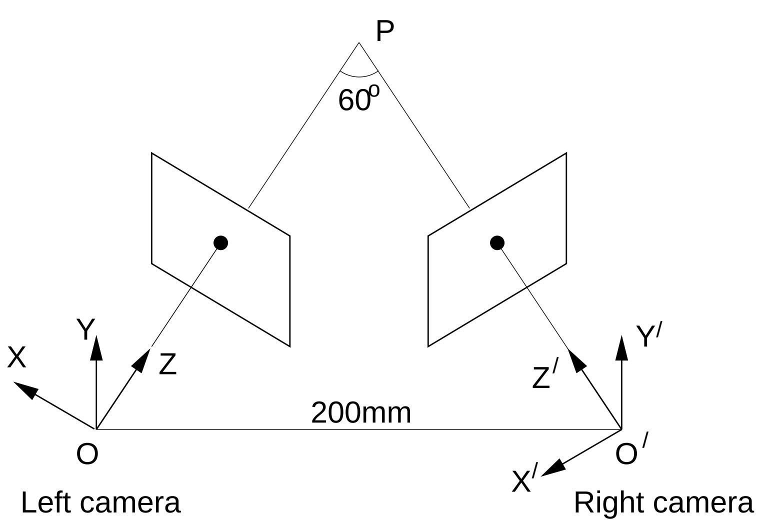


Fig. 3

5 (a) A video camera undergoes rigid body motion with translational velocity $\mathbf{U}$ and rotational velocity $\mathbf{\Omega}$. Show that, under planar perspective projection, a stationary point in the scene with image position $\mathbf{p} = (x, y, f)$ has image motion

$$\dot{\mathbf{p}} = -\frac{f\mathbf{U}}{Z} + \frac{(\mathbf{U}.\mathbf{k})\,\mathbf{p}}{Z} - \mathbf{\Omega} \wedge \mathbf{p} + \frac{[\mathbf{\Omega}, \mathbf{p}, \mathbf{k}]\,\mathbf{p}}{f}$$

where $\mathbf{k}$ is a unit vector along the optical axis, Z is the distance of the point along $\mathbf{k}$, $[\mathbf{\Omega}, \mathbf{p}, \mathbf{k}]$ is a triple scalar product and $\wedge$ denotes a vector product.

(b) An unmanned aircraft is visually guided towards its target. Figure 4 shows the image velocities of four points A, B, C and D, with image coordinates $(-f, 0)$, $(0, -f)$, $(f, 0)$ and $(0, f)$ respectively. The scene is viewed through a camera of focal length f.

An inertial sensor gives the angular velocity $\mathbf{\Omega}$ as $(0, -1/f, 0)$ rad/sec.

(i) Calculate the rotational components of image motion at the points A, B, C and D.

(ii) Hence, or otherwise, calculate the heading (the direction of translation) of the aircraft.

(iii) Which of the points lies on the horizon? Explain your reasoning.

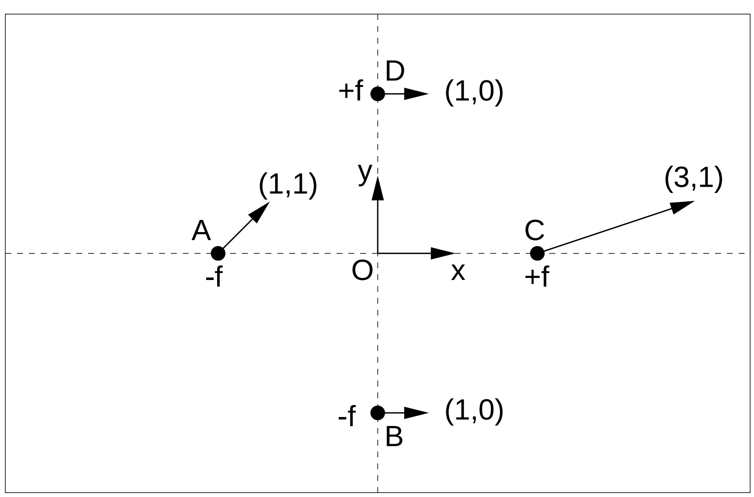


Fig. 4

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

(a) Describe an algorithm to measure image motion. Show how estimates of image motion can be used to reduce camera shake in a hand-held video camera.

(b) A fruit picking robot detects ripe fruit using a colour camera mounted at the end of its manipulator. The vision system then tracks the fruit's outline as the camera moves at constant speed towards the fruit. Describe how the change in the apparent area of the fruit can be used to guide the robot manipulator.

(c) Outline a model-based vision system to recognise planar, curvilinear industrial parts (such as spanners and wrenches). Explain how the models can be acquired from sample images, and how the recognition system can be made to work independent of lighting conditions and changes in viewpoint.

(d) 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.

END OF PAPER

ENGINEERING TRIPOS PART IIB

ELECTRICAL AND INFORMATION SCIENCES TRIPOS PART II

Friday 25 April 1997 2 to 3.30

Module I12

COMPUTER VISION AND ROBOTICS

Answer not more than **four** *questions.*

All questions carry the same number of marks.

(TURN OVER

1 (a) Why do many computer vision systems start by extracting edges from the raw images? When would corners be more appropriate?

(b) An image is represented as a two-dimensional intensity array $I(x, y)$. If I_n is the derivative of I in the direction $\mathbf{n}$, show that

$$I_n^2 = \frac{\mathbf{n}^T \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \mathbf{n}}{\mathbf{n}^T \mathbf{n}}$$

where $I_x \equiv \partial I / \partial x$ and $I_y \equiv \partial I / \partial y$. Hence show how corner features can be detected by examining the eigenvalues of the matrix A, where

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

and $\langle\rangle$ denotes a two-dimensional smoothing operation. How are the partial derivatives I_x and I_y computed in practice?

(c) Explain carefully why the smoothing stage is essential, even if the image is free of noise. What extra step might be necessary if the image is known to be noisy?

2 (a) Distinguish between *intrinsic* and *extrinsic* camera parameters.

(b) A vision system is used to control a car driving along a flat, straight road. A single camera is mounted behind the windscreen and looks straight ahead. At a particular instant, the camera captures the image in Fig. 1. The image shows the sides of the road vanishing at the point D on the horizon. Points A and C are road signs: A reads "Services 500m" and C reads "Services 100m". Point B is an emergency phone box. Assuming the signs are accurate, find the roadside distance between sign A and the phone box.

(c) Given the *single* image in Fig. 1 and the *intrinsic* camera parameters, outline a graphical construction to find the heading of the car (with respect to the road) and the distance from the camera to the phone box.

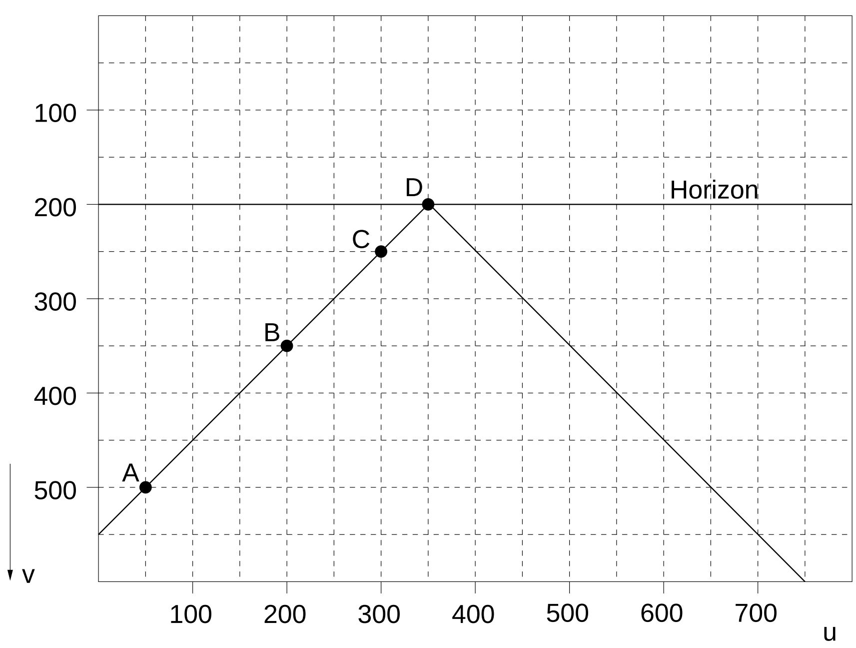


Fig. 1

3 (a) A camera views the world plane $Z = 0$. The transformation between points (X, Y) on the plane and pixels (u, v) in the image is described by the plane projective camera model:

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

(i) How many degrees of freedom does this transformation have? If the camera were to view a square, show, using sketches, how the square might appear in the image. Be sure to account for each degree of freedom of the transformation.

(ii) How many points are required to calibrate this camera? How could the calibration be improved if more points were available?

(iii) List the invariants of the plane projective transformation.

(b) An alternative camera model is the planar affine model:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

Under what conditions is this a good camera model? What are the advantages of using an affine model instead of a projective one?

(c) A camera views two planar industrial parts lying on a tabletop: the resulting image is shown in Fig. 2. Part A is known to be square-shaped.

(i) How is it possible to deduce that an affine camera model is appropriate in this case?

(ii) Using affine invariants or otherwise, find the relative lengths of the three sides of part B.

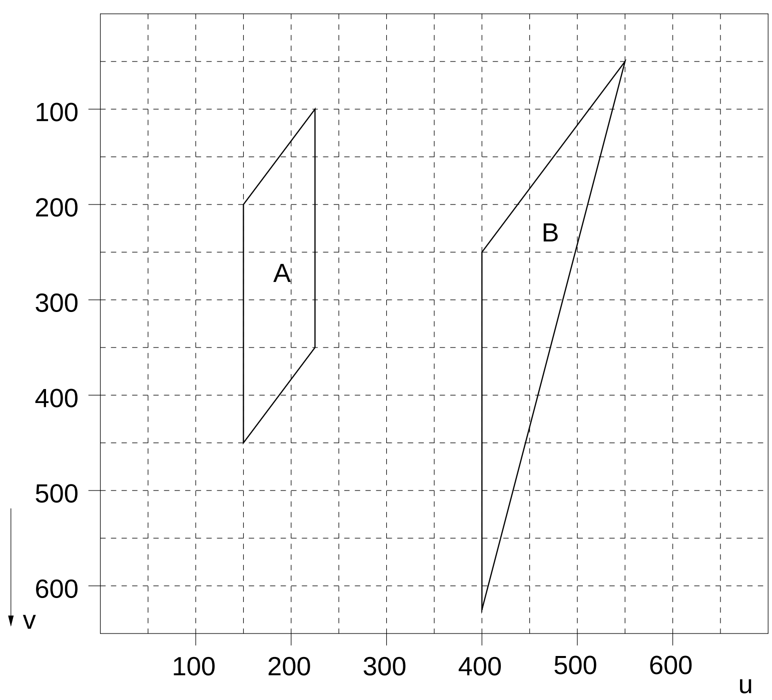


Fig. 2

(TURN OVER

4 (a) What is meant by *triangulation* in the context of stereo vision?

(b) A stereo vision rig comprises two cameras, both of focal length f, arranged symmetrically at right angles as shown in Fig. 3. The distance between the cameras' optical centres is $\sqrt{2}d$. Points $\mathbf{X}_c$ and $\mathbf{X}'_c$ in the two camera-centered coordinate systems are related by the rigid body transformation

$$\mathbf{X}'_c = \mathrm{R}\mathbf{X}_c + \mathbf{T}$$

(i) Write down the rotation matrix R and the translation vector $\mathbf{T}$.

(ii) Show that rays $\mathbf{p}$ and $\mathbf{p}'$ in the two coordinate systems are related by

$$\left(\frac{Z_c}{f} \mathrm{R}\mathbf{p} + \mathbf{T} \right) \times \mathbf{p}' = \mathbf{0}$$

where $\mathbf{p} = [\ x \quad y \quad f\ ]^T$ and $\mathbf{p}' = [\ x' \quad y' \quad f\ ]^T$. Hence show that the epipolar constraint is

$$y(f + x') = y'(f - x)$$

and that depths can be recovered using

$$Z_c(f^2 + xx') = df(f + x')$$

(iii) A moving aircraft is observed by the stereo rig. In the left camera's image the aircraft appears stationary at the point $(x, y) = (0, 1)$ cms. In the right camera's image the aircraft's trajectory is given by

$$x' = t\ , \quad y' = 1 + 1.25t \quad (x' \text{ and } y' \text{ in cms, } t \text{ in seconds})$$

Find the cameras' focal length f. If $d = 100$ m, find also the aircraft's physical speed.

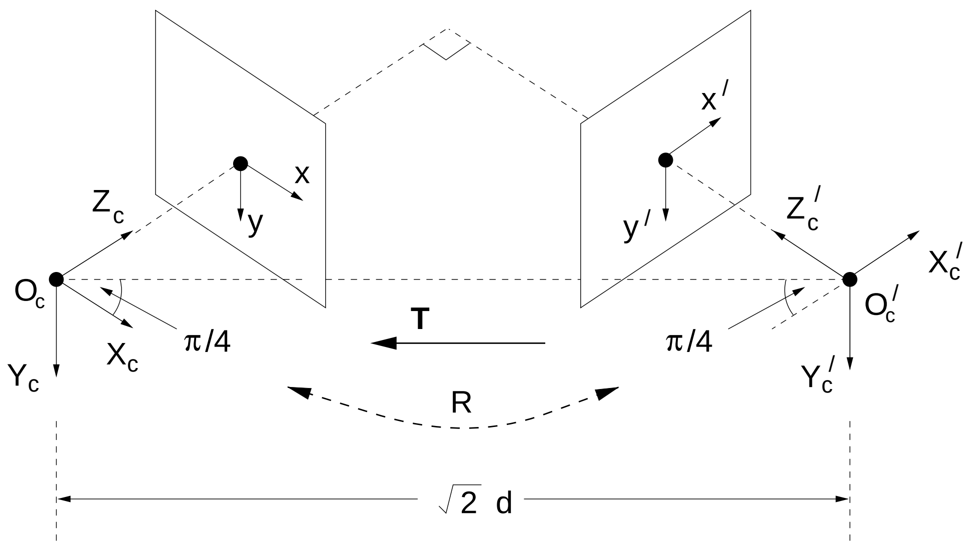


Fig. 3

5 A camera views a scene containing two planar, rectangular objects (A and B) and a spherical ball (C). Edges detected in an image taken at time $t = t_0$ are shown in Fig. 4(a). The camera translates with velocity $\mathbf{U} = (U_1, 0, 0)$, expressed in the camera-centered coordinate system shown in Fig. 4(b). An x-t slice through the resulting spatiotemporal image is shown in Fig. 5(a). The slice is taken at the position of the dotted line in Fig. 4(a).

(a) Show that $\dot{x}$, the x-component of image velocity, is given by

$$\dot{x} = -\frac{fU_1}{Z_c}$$

where f is the focal length of the camera and Z_c is the depth of the imaged feature. If $f = 50$mm, calculate the depths of the two rectangular objects (A and B) in terms of U_1.

(b) What can you deduce about the depth of the ball (C)? If the ball's radius is known to be 10cm, estimate a lower bound for the camera's speed U_1.

(c) Now consider the motion trajectories in Fig. 5(b). Explain why these trajectories could never be obtained in practice.

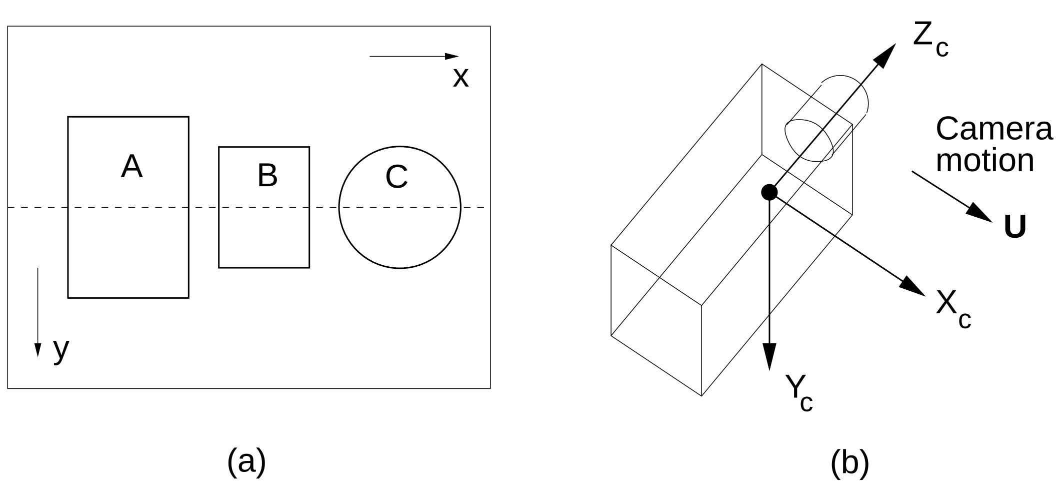


Fig. 4

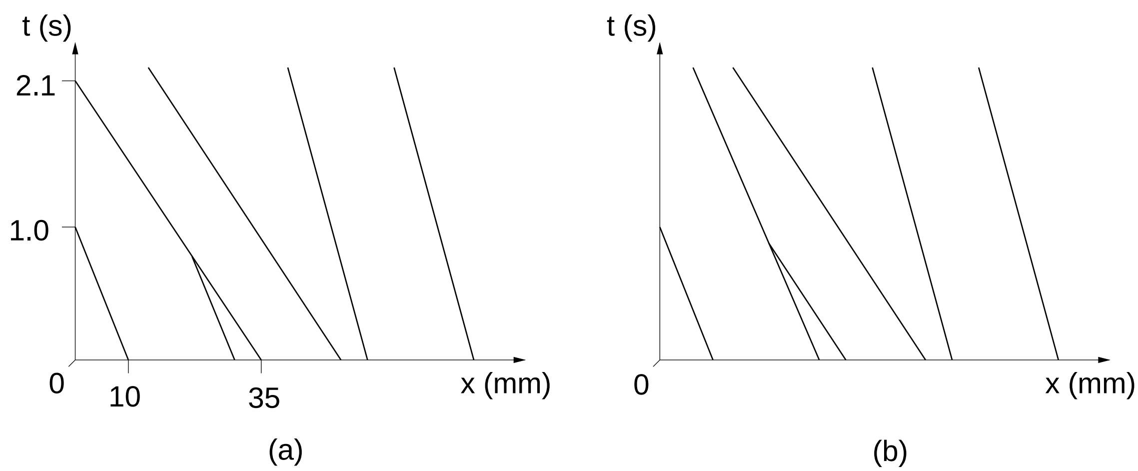


Fig. 5

(TURN OVER

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

(a) A 3D model of King's College Chapel is required for a virtual reality application. Data is gathered by somebody walking around the chapel with a standard camcorder. Describe how the 3D model could be constructed from the resulting image sequence.

(b) A terrestrial TV channel is to broadcast a football match across the entire country. The pitch is ringed with advertising hoardings of only local interest. The TV channel proposes to process the images at each transmitter, so that more appropriate advertisements appear on the hoardings. Describe how computer vision techniques could be used to achieve this.

(c) A computer vision system is to help a motorist decide when it's safe to overtake on a motorway. A camera looks through the car's rear window and observes oncoming vehicles in the fast lane. Describe how the change in the apparent area of the approaching vehicle can be used to decide whether it's safe to pull out.

(d) A tourist on top of the Empire State Building wishes to acquire a 180° panoramic image of the view to the north using a standard camcorder without a wide angle lens. The tourist stands still and pans the camcorder clockwise from west through north to east. How could the panoramic view be constructed from the resulting image sequence?

END OF PAPER

ENGINEERING TRIPOS PART IIB

ELECTRICAL AND INFORMATION SCIENCES TRIPOS PART II

Friday 16 January 1998 2 to 3.30

Module I12

COMPUTER VISION AND ROBOTICS

Answer not more than **four** *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.*

(TURN OVER

1 (a) A 512×512 grey scale image is smoothed by convolution with a 2D Gaussian kernel:

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} \exp - \left(\frac{x^2 + y^2}{2\sigma^2} \right)$$

(i) Write down an expression showing how the 2D convolution can be performed as two 1D convolutions.

(ii) What is the typical size of the kernel for $\sigma = 1$? Hence calculate the number of computations required for the 2D and 1D schemes. [40%]

(b) A discrete approximation to the second-order spatial derivative d^2I/dx^2 can be obtained by convolving $I(x, y)$ with the following kernel:

1	-2	1

Use this result to derive a 3 × 3 kernel that can be used to compute a discrete approximation to the Laplacian of an image:

$$\nabla^2 I = \frac{d^2 I}{dx^2} + \frac{d^2 I}{dy^2}$$ [30%]

(c) 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%]

2 (a) An image is formed by perspective projection onto an image plane, as shown in Fig. 1. If the image plane is sampled by a CCD array with k_u pixels per unit length in the u direction and k_v pixels per unit length in the v direction, show that the relationship between a point (X_c, Y_c, Z_c) and its image (u, v) (in pixels) is given by

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

[30%]

(b) A *weak perspective* projection comprises an orthographic projection onto the plane $Z_c = Z_A$ followed by perspective projection onto the image plane.

(i) Derive the homogeneous relationship between a point $(X_c, Y_c, Z_c, 1)$ and its image (su_A, sv_A, s) under weak perspective projection.

(ii) 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 $\Delta Z \equiv Z_A - Z_c$.

(iii) Under what viewing conditions is weak perspective a good camera model? What are its advantages?

[70%]

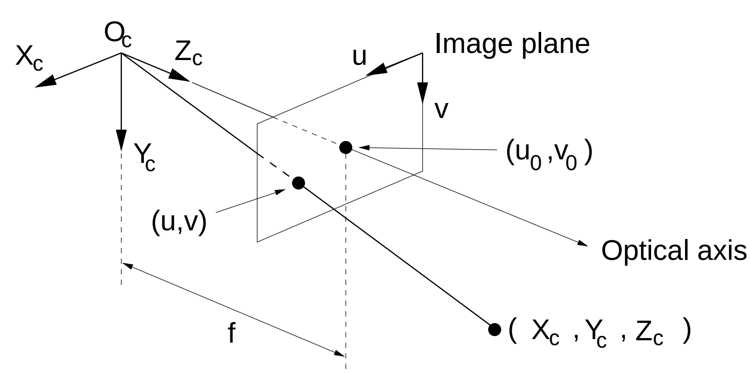


Fig. 1

3 (a) A camera views a point (X, Y, Z) in 3D space. The image of the point is at the pixel with coordinates (u, v). The relationship between (X, Y, Z) and (u, v) can be written as

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

(i) Under what assumptions is this relationship valid?

(ii) Comment on the algebraic and geometric significance of s and the elements p_{ij}.

(iii) Explain in detail how the elements p_{ij} are estimated in practice. [50%]

(b) An overhead surveillance camera looks down on a traffic junction. The relationship between points (X, Y) (in metres) on the world plane and corresponding points (u, v) (in pixels) in the image is known to be

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} 10.8 & 8.2 & 230.5 \\ 5.4 & -5.6 & 142.0 \\ 0.001 & 0.001 & 1.0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

(i) If the visible portion of the road junction lies within the quadrilateral with (X, Y) vertices $(-23, 3)$, $(5, 30)$, $(45, -24)$ and $(15, -48)$, calculate bounds on s. What can you deduce about the viewing conditions?

(ii) A moving car is observed in the image. The centre of the car is at pixel $(100, 100)$ and the car's image velocity is $(180, 90)$ pixels/s. Estimate the world position and physical velocity of the car. [50%]

4 (a) A stereo vision rig comprises two cameras, both of focal length f, arranged as in Fig. 2. Points $\mathbf{X_C}$ and $\mathbf{X'_C}$ in the two camera-centered coordinate systems are related by the rigid body transformation

$$\mathbf{X'_C} = \mathrm{R}\mathbf{X_C} + \mathbf{T}$$

If a point projects onto the left camera's image plane at (x, y) and the right camera's image plane at (x', y'), show that the following constraint is satisfied:

$$\begin{bmatrix} x' & y' & f \end{bmatrix} [\mathrm{E}] \begin{bmatrix} x \\ y \\ f \end{bmatrix} = 0$$

What is the rank of the 3×3 matrix E? [40%]

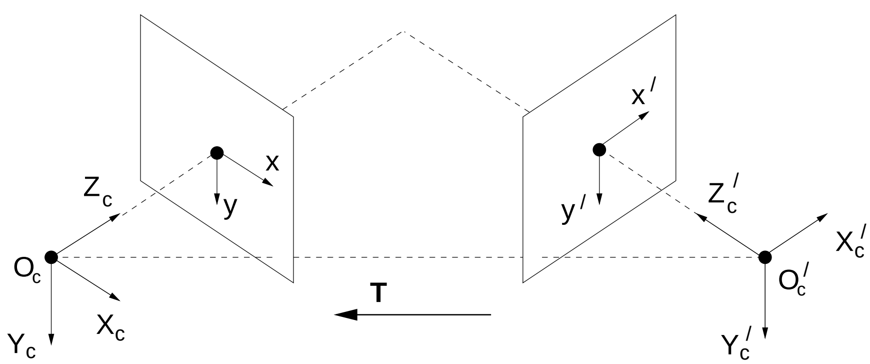


Fig. 2

(b) A camera with focal length f acquires an image of a scene, undergoes a pure translation and then acquires a second image of the same scene. The relationship between corresponding points (x, y) and (x', y') in the two views is given by

$$\begin{bmatrix} x' & y' & f \end{bmatrix} \begin{bmatrix} 0 & +c & 0 \\ -c & 0 & +a \\ 0 & -a & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ f \end{bmatrix} = 0 \quad (a \text{ and } c \text{ are positive})$$

(i) Find the locations of the epipoles in the two views.

(ii) What can you deduce about the translation undergone by the camera?

(iii) Sketch and comment on the epipolar line structure in the two views.

(iv) Sketch the image trajectory of a point induced by the camera motion. [60%]

(TURN OVER

5 (a) Describe how *optical flow* can be estimated from spatial and temporal derivatives of image intensities. Be sure to include reference to the aperture problem. [30%]

(b) A video camera with focal length f undergoes pure translation along its optical axis with speed U, as shown in Fig. 3.

(i) Show that the image velocity of a point (X_c, Y_c, Z_c) is given by

$$\dot{\mathbf{p}} = \begin{bmatrix} \frac{xU}{Z_c} & \frac{yU}{Z_c} & 0 \end{bmatrix}^T$$

where $\mathbf{p} = [x\ y\ f]^T$ is the projection of the point onto the image plane. Sketch the image velocity field.

(ii) If the camera is viewing points on the plane $Z_c = Z_0 + pX_c + qY_c$, show that

$$\begin{aligned} \dot{x} &= \frac{U}{Z_0}\left(x - \frac{px^2}{f} - \frac{qxy}{f}\right) \\ \dot{y} &= \frac{U}{Z_0}\left(y - \frac{pxy}{f} - \frac{qy^2}{f}\right) \end{aligned}$$

(iii) Find the divergence of the image velocity field at the principle point $\mathbf{p} = [0\ 0\ f]^T$. How can the divergence be used to estimate the time to contact with the plane? [70%]

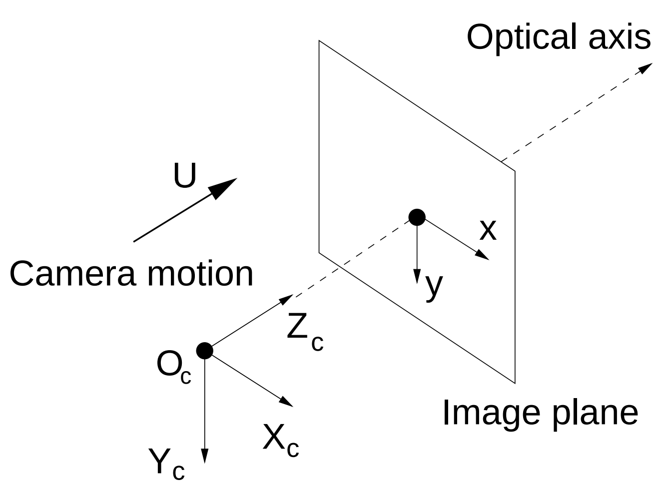


Fig. 3

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

(a) What are the drawbacks of using edge and corner detectors for tracking moving objects in images? Describe how B-spline snakes can be used instead. What are the advantages and disadvantages of B-spline snakes? [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, the image processing required to detect the people and the visual tracking. [50%]

(c) Two images of a scene are acquired by a stereo pair of cameras with parallel image planes and optical axes perpendicular to the baseline. Explain how to synthesise an image from a third, intermediate viewpoint. Include details of the correspondence problem and any assumptions made. [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

ENGINEERING TRIPOS PART IIB

ELECTRICAL AND INFORMATION SCIENCES TRIPOS PART II

Friday 15 January 1999 2 to 3.30

Module I12

COMPUTER VISION AND ROBOTICS

Answer not more than **four** *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.*

(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. [20%]

(b) The filter kernel $\frac{1}{A}[\ 1\ \ 4\ \ 6\ \ 4\ \ 1\ ]$, which is a binomial approximation to the 1D Gaussian function, is used to smooth a row of pixels $I(x)$.

(i) What is the correct value for A? [10%]

(ii) Give mathematical expressions showing how to obtain each smoothed pixel and localise intensity discontinuities. [20%]

(iii) Show how successive applications of the kernel can be used to smooth a 2D image $I(x, y)$. [20%]

(c) Show how 1D smoothing with the kernel in (b) is equivalent to successively averaging neighbouring pixels (repeated convolution of the image with the kernel $[\ \frac{1}{2}\ \ \frac{1}{2}\ ]$). How many times must neighbouring pixels be averaged? [30%]

2 (a) Show how the overall projection matrix between a point (X, Y, Z) in 3D space and its image position (u, v) (in pixels) can be represented by a 3×4 matrix:

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

See Fig. 1 for a definition of the relevant coordinate systems. [30%]

(b) Explain how the matrix elements p_{ij} can be estimated by observing the images of known 3D points. How can the *intrinsic* and *extrinsic* camera parameters be derived from the overall projection matrix? [40%]

(c) Using the projection matrix above, find the vanishing point (u, v) of lines which are parallel to the world X-axis. [30%]

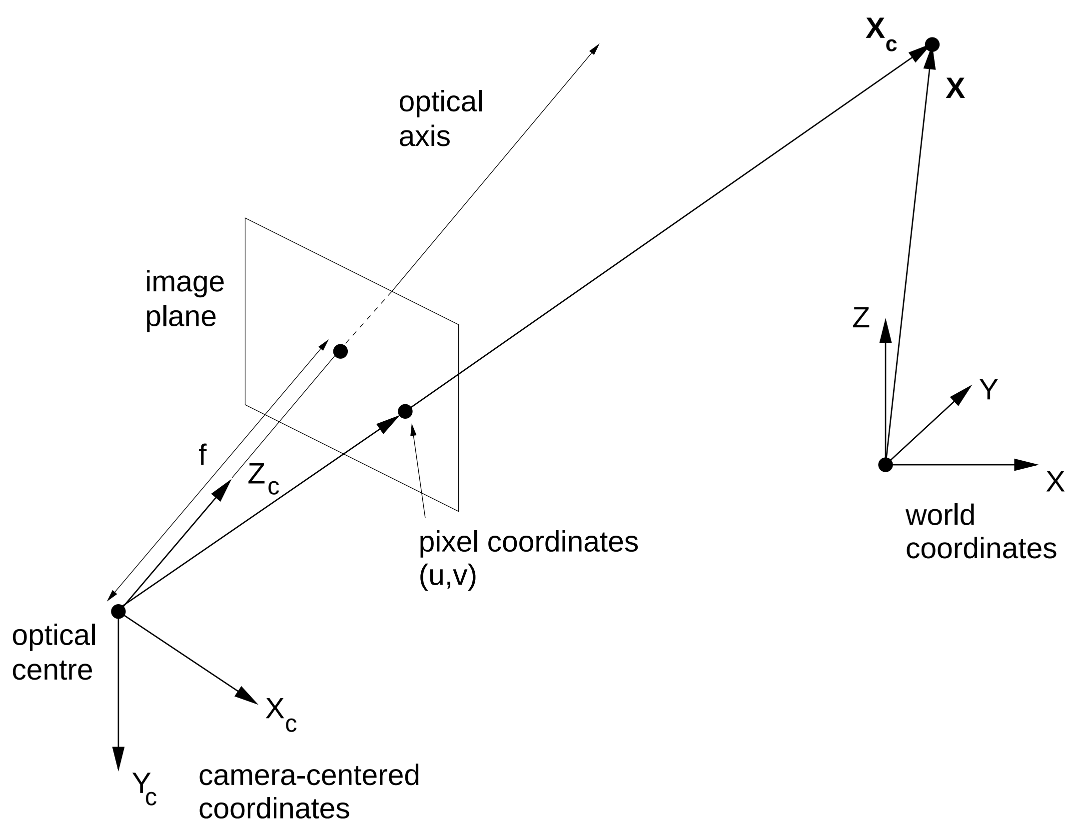


Fig. 1

3 (a) Under perspective projection, the relationship between points X on a world line and the corresponding points l on the image line is given by

$$\begin{bmatrix} sl \\ s \end{bmatrix} = \begin{bmatrix} p & q \\ r & 1 \end{bmatrix} \begin{bmatrix} X \\ 1 \end{bmatrix}$$

Show that the cross-ratio

$$\frac{(X_d - X_a)(X_c - X_b)}{(X_d - X_b)(X_c - X_a)}$$

of four collinear points a, b, c, and d is invariant under perspective projection. [40%]

(b) A biscuit production plant uses an automated inspection system to measure the thickness Z of pastry. A point mark is projected onto the pastry surface and observed using a fixed CCD camera — see Fig. 2. As the thickness of pastry changes, show that the image of the projected point moves along a line. How can the system be calibrated and how can the thickness be recovered? [60%]

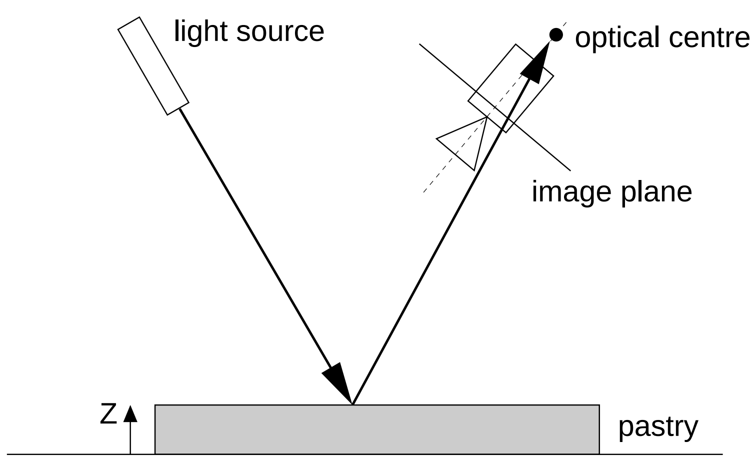


Fig. 2

4 (a) What is meant by the *fundamental matrix* in stereo vision? Explain how the fundamental matrix can be estimated from point correspondences. [30%]

(b) For a particular pair of images, the fundamental matrix is estimated as

$$\mathrm{F} = \begin{bmatrix} 0 & 1 & -200 \\ 1 & 0 & -2700 \\ -200 & 2300 & 80000 \end{bmatrix}$$

Show that the location of the epipole in the left image ($\tilde{\mathbf{w}}_e$ in homogeneous pixel coordinates) satisfies $\mathrm{F}\tilde{\mathbf{w}}_e = \mathbf{0}$. State the corresponding result for the right image, and calculate the positions of the left and right epipoles. [30%]

(c) For the case in (b), it is known that both images were acquired using the same camera which has a focal length of 50 mm and a 10 mm $\times$ 10 mm square CCD array. The CCD array is divided into 500 $\times$ 500 square pixels. The pixel at the top left corner of the CCD array has coordinates $(0, 0)$, and the optical axis intersects the CCD array at the pixel with coordinates $(200, 200)$. *Without calculating the essential matrix*, deduce as much as you can about the relative position and orientation of the left and right cameras. What can you *not* deduce about the relative position? [40%]

5 A video camera undergoes rigid body motion with translational velocity $\mathbf{U}$ and rotational velocity $\mathbf{\Omega}$. Under planar perspective projection, a stationary point in the scene with image position $\mathbf{p} = (x, y, f)$ has image motion

$$\dot{\mathbf{p}} = -\frac{f\mathbf{U}}{Z_c} + \frac{(\mathbf{U}.\mathbf{k})\,\mathbf{p}}{Z_c} - \mathbf{\Omega} \times \mathbf{p} + \frac{[(\mathbf{\Omega} \times \mathbf{p}).\mathbf{k}]\,\mathbf{p}}{f}$$

where f is the focal length of the camera, $\mathbf{k}$ is a unit vector along the optical axis and Z_c is the $\mathbf{k}$ component of the point's displacement from the optical centre.

(a) For the case of pure translation, explain what is meant by the *focus of expansion*. Show that the image position of the focus of expansion is given by

$$(x, y) = \left(f\frac{U_1}{U_3}, f\frac{U_2}{U_3} \right)$$ [25%]

(b) A camera is mounted behind the windshield of a light aircraft. Four out of five points being tracked in the image happen to come into instantaneous alignment, so that two points are coincident at A and another two points are coincident at B. The fifth point is at image position C. The image positions and velocities of the five points are shown in Fig. 3: the units are mm for positions and mm/sec for velocities.

(i) Derive a pair of *motion parallax vectors* from the image velocities. [10%]

(ii) Deduce as much as you can about the aircraft's linear and angular velocities and the depths Z_c of each of the five points. Describe qualitatively the manoeuvre the aircraft is performing. [40%]

(c) In practice, it is rare that trackable points come into instantaneous alignment in the image. Describe briefly how else it is possible to generate motion parallax vectors. [25%]

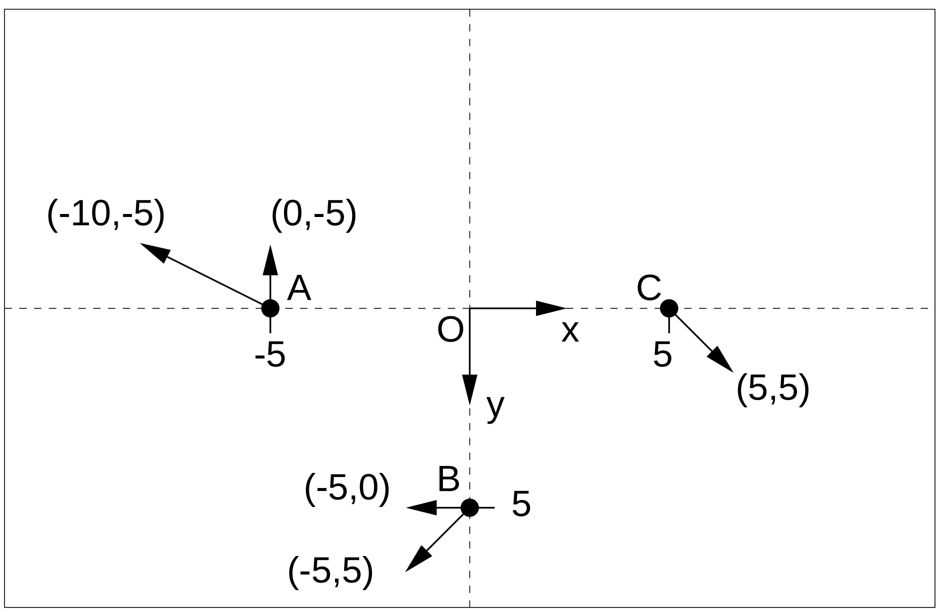


Fig. 3

(TURN OVER

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

(a) A *mosaic* of a scene is acquired by rotating a camera about its optical centre. Explain how the transformation between two successive images can be estimated. Include details of the localisation and matching of image features. [50%]

(b) The European Prometheus project is developing an autonomous visually-guided car. Part of the project involves being able to recognise road signs. The engineers on the project have decided to describe the imaging process using a *weak perspective* model. Is this a suitable model? The square and triangular borders of the signs can be detected by grouping straight line segments. Outline a method to compare the image of a sign from an arbitrary viewpoint to a database of road signs. [50%]

(c) A moving planar target is viewed under weak perspective. Show that the target's image undergoes a 2D affine transformation as the target moves. Explain how the target can be tracked using an affine template or *snake*. [50%]

(d) A student working on the Part 1B Integrated Design Project proposes to use a CCD camera looking at the play area to determine the position of the mobile robot. Show how the position of the robot in the image can be converted to a corresponding position in the play area. What image processing would be required to localise the mobile robot? [50%]

END OF PAPER

ENGINEERING TRIPOS PART IIB

ELECTRICAL AND INFORMATION SCIENCES TRIPOS PART II

Friday 21 January 2000 2 to 3.30

Module I12

COMPUTER VISION AND ROBOTICS

Answer not more than ***four*** *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.*

(TURN OVER

1 (a) What is meant by an image *edge*? Describe, in detail, the filter kernels that are commonly used for smoothing and differentiation as part of the edge detection process. Include an expression for computing the intensity of a smoothed pixel. [50%]

(b) What is meant by an image *corner* or *feature of interest*? Show how the spatial derivatives of an image can be used to detect a corner, describing clearly all the computational stages. [50%]

2 The relationship between a 3D world point (X, Y, Z) and its 2D image pixel coordinates (u, v) can be written as follows:

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

(a) Under what assumptions is this relationship valid? [10%]

(b) The relationship is to be calibrated by observing the images (u_i, v_i) of known reference points (X_i, Y_i, Z_i). Derive linear equations in the unknown elements p_{ij} of the projection matrix. How many reference points are required to estimate all the elements p_{ij}? [30%]

(c) How should the relationship be modified to represent points at infinity in both the image plane and the world? [20%]

(d) Find the vanishing points in the image of lines parallel to the world X, Y and Z axes. Show that the vanishing points do not depend on the position of the camera. [40%]

3 A static scene is observed twice by the same camera, producing a pair of images with pixel correspondences (u, v) and (u', v'). Under certain conditions, the two images are related by a 2D projective transformation:

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

(a) Show that the above relationship holds:

(i) when the camera is viewing a plane; [20%]

(ii) when the camera undergoes a pure rotation about the optical centre. [20%]

(b) How many degrees of freedom does the 2D projective transformation have? If an object appears as a square in the first image, describe, using sketches, how it might appear in the second image. Be sure to account for each degree of freedom of the 2D projective transformation. [20%]

(c) Consider the line $l_1 u + l_2 v + l_3 = 0$ in the first image. The line can be represented in homogeneous coordinates by the vector $\mathbf{l} = [\ l_1 \ \ l_2 \ \ l_3\ ]^T$. Derive an equation for the corresponding line in the second image. [20%]

(d) Repeat part (c) for the conic $au^2 + buv + cv^2 + du + ev + f = 0$. [20%]

(TURN OVER

4 (a) List four matching constraints which can be used to find point correspondences in stereo vision. Outline an algorithm which uses these constraints to match a large number of features between left and right images. [30%]

(b) For a virtual reality modelling application, the fundamental matrix is estimated from point correspondences using linear algebra techniques.

(i) One pair of images yields the following estimate:

$$\begin{bmatrix} 0 & 1 & -200 \\ 1 & 0 & -2695 \\ -200 & 2295 & 80005 \end{bmatrix}$$

Explain carefully why this estimate is evidently in error, how the error could have arisen, and how the estimate might be improved. [30%]

(ii) Another pair of images yields a fundamental matrix estimate of the following form:

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

Describe the possible camera configurations which could have resulted in this estimate. [20%]

(iii) Explain briefly how the 3D structure of objects in the scene can be recovered. [20%]

5 A video camera undergoes rigid body motion with translational velocity $\mathbf{U}$ and rotational velocity $\mathbf{\Omega}$. Under planar perspective projection, a stationary point in the scene with image position $\mathbf{p} = (x, y, f)$ has image motion

$$\dot{\mathbf{p}} = -\frac{f\mathbf{U}}{Z_c} + \frac{(\mathbf{U}.\mathbf{k})\,\mathbf{p}}{Z_c} - \mathbf{\Omega} \times \mathbf{p} + \frac{[(\mathbf{\Omega} \times \mathbf{p}).\mathbf{k}]\,\mathbf{p}}{f}$$

where f is the focal length of the camera, $\mathbf{k}$ is a unit vector along the optical axis and Z_c is the $\mathbf{k}$ component of the point's displacement from the optical centre.

(a) With reference to the above equation, explain clearly what is meant by the *speed-scale ambiguity*. [10%]

(b) In practice, $\dot{\mathbf{p}}$ is estimated by observing changes in the image over time. Describe two different ways in which this can be done. [40%]

(c) Suppose a uniform image motion field $\dot{\mathbf{p}} = [\,5\ \ 0\ \ 0\,]^T$ mm/sec is measured at all points in the camera's field of view. If the camera is known to be undergoing pure translation, deduce as much as you can about the camera, its motion and the scene. [20%]

(d) Repeat part (c) for the case that the camera is known to be undergoing pure rotation. [30%]

(TURN OVER

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

(a) A *mosaic* of a scene is acquired by rotating a camera about its optical centre. Explain how the transformation between two successive images can be estimated. Is there sufficient information in the images to estimate the angle of rotation and the camera's intrinsic parameters? [50%]

(b) An underwater robot is to use a camera mounted on its front to help it dock onto a planar target surface. Describe an algorithm that can be used to track the apparent shape of the target in the image, giving details of the image processing required, and show how the apparent area and its temporal derivatives can be used to generate a control signal to aid the robot in docking. State any assumptions you make. [50%]

(c) Outline a model-based vision system to recognise planar, curvilinear industrial parts (such as spanners and wrenches). Explain how the models can be acquired from sample images, and how the recognition system can be made to work independent of lighting conditions and changes in viewpoint. [50%]

(d) Describe how to interactively build a 3D VRML (Virtual Reality Modelling Language) model from two uncalibrated photographs of a building. Include details of the camera calibration, how the 3D coordinates are calculated and how the texture maps are represented. [50%]

END OF PAPER

ENGINEERING TRIPOS PART IIB
ELECTRICAL AND INFORMATION SCIENCES TRIPOS PART II

Friday 19 January 2001 2 to 3.30

Module I12

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.*

(TURN OVER

1 (a) A 512×512 grey level image, $I(x, y)$, is to be smoothed by convolution with a discrete approximation to the 2D Gaussian kernel, $G_\sigma(x, y)$, of size $(2n+1) \times (2n+1)$ pixels.

(i) Give an expression for computing the intensity of a smoothed pixel, $S(x, y)$. [10%]

(ii) Show how the convolution can be performed by two discrete 1D convolutions, and comment on the computational saving this achieves. [20%]

(iii) Determine the kernel size and coefficients for $\sigma = 1$. State clearly any assumptions and approximations used. [30%]

(b) Describe and compare the Marr–Hildreth and Canny algorithms for detecting and localizing intensity edges in grey level images. [40%]

2 (a) A number of parallel planes are viewed under perspective projection.

(i) Show that the planes have a common *horizon* or vanishing line in the image. You should derive the relationship between the orientation of the planes in the world and the equation of the horizon line in the image. [30%]

(ii) Verify that the vanishing point of any line parallel to these planes lies on the horizon. [20%]

(b) Figure 1 shows a circle inscribed in a square and the image of the square under perspective projection (the image of the circle is missing).

(i) Describe a *geometric* technique to find the image of the circle. Your technique should make use of vanishing points, cross-ratios and the image of the circumscribing square. There is no need to actually compute the image of the circle. [30%]

(ii) Now describe an *algebraic* method to compute the image of the circle. Once again, your method should exploit the image of the circumscribing square. [20%]

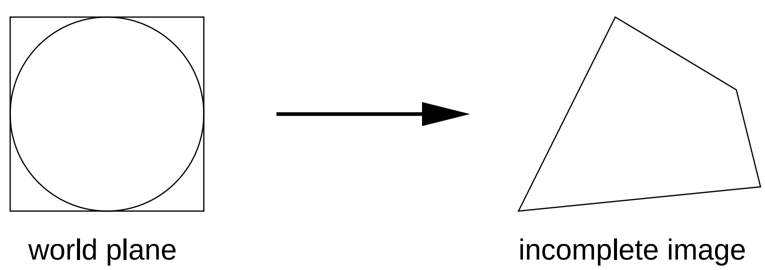


Fig. 1

3 (a) Describe how to calibrate a CCD camera from a set of image positions with pixel coordinates (u, v) corresponding to known 3D positions (X, Y, Z). Your answer should include details of:

(i) the relationship between the CCD pixel coordinates and the 3D coordinates of the point, and any assumptions made;

(ii) the minimum number of calibration points needed and why these are insufficient in practice;

(iii) the derivation of linear equations in the unknown calibration parameters;

(iv) how the set of linear equations is solved;

(v) how and why the linear solution can be improved;

(vi) how the position, orientation and focal length of the camera can be recovered. [60%]

(b) A video surveillance camera is used to determine the speed of cars entering a tunnel. Assuming that the road is flat and that the lane markings are clearly visible, describe how the image position of a car can be used to estimate its position in the lane, and how this information can then be processed to give an estimate of the car's speed. State clearly any assumptions made. [40%]

4 (a) In stereo vision, what is the *essential matrix*? If a point has coordinates $\mathbf{X_C}$ in the left camera's coordinate system and $R\mathbf{X_C} + \mathbf{T}$ in the right camera's coordinate system, derive an expression for the essential matrix in terms of the rotation matrix R and translation vector $\mathbf{T}$. [30%]

(b) In a system which builds 3D models from stereo pairs of images, point correspondences are used to estimate the fundamental matrix.

(i) Explain what further information is required to find the essential matrix and write down a formula relating the two matrices. [15%]

(ii) For one pair of images, the essential matrix is found to be

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}.$$

The system attempts to recover R and $\mathbf{T}$ from the essential matrix. Show that there are multiple solutions for the rotation R and the direction of translation $\hat{\mathbf{T}}$. Suggest how the ambiguity might be resolved in practice. [30%]

(iii) What effect does the unknown baseline length $\|\mathbf{T}\|$ have on the 3D reconstruction? Justify your answer. [25%]

(TURN OVER

5 A video camera undergoes rigid body motion with translational velocity $\mathbf{U}$ and zero rotational velocity. Under perspective projection, a stationary point in the scene with image position $\mathbf{p} = (x, y, f)$ has image motion

$$\dot{\mathbf{p}} = -\frac{f\mathbf{U}}{Z_c} + \frac{(\mathbf{U}.\mathbf{k})\,\mathbf{p}}{Z_c} \ ,$$

where f is the focal length of the camera, $\mathbf{k}$ is a unit vector along the optical axis and Z_c is the $\mathbf{k}$ component of the point's displacement from the optical centre.

(a) Explain what is meant by the *focus of expansion* of the image velocity field. Derive an expression for the focus of expansion in terms of $\mathbf{U}$, and show that the image velocity field is radial with respect to the focus of expansion. [30%]

(b) A camera is mounted on a mobile robot to facilitate accurate navigation. The camera has a focal length of 8 mm and a 10 mm $\times$ 10 mm square CCD array. The CCD array is divided into 500 $\times$ 500 square pixels. The pixel at the top left corner of the CCD array has coordinates $(0, 0)$, and the optical axis intersects the CCD array at the pixel with coordinates $(200, 200)$.

The robot moves across a horizontal ground plane with constant linear velocity and zero angular velocity. Initially, the focus of expansion is observed at the pixel with coordinates $(10, 180)$. The robot then changes direction, without rotating, and the new focus of expansion is observed at the pixel with coordinates $(490, 185)$.

(i) Find the angle between the robot's original and new headings. [20%]

(ii) Find the angle between the camera's optical axis and the horizontal. [20%]

(iii) Tracking the focus of expansion is not the best way to measure the angle of a turn. Identify potential problems and suggest a superior, vision-based technique for turning the robot through a fixed angle. [30%]

END OF PAPER

ENGINEERING TRIPOS PART IIB
ELECTRICAL AND INFORMATION SCIENCES TRIPOS PART II

Friday 18 January 2002 2 to 3.30

Module I12

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

1 (a) A grey scale image, $I(x, y)$, is to be smoothed and differentiated as part of the edge detection process.

(i) Why is smoothing required? Describe the smoothing filter used in practice, and explain the effects of increasing the size of the filter. [20%]

(ii) Give an expression for computing the intensity of a smoothed pixel in terms of two discrete 1D convolutions. [20%]

(iii) Give expressions for computing spatial derivatives after smoothing. [10%]

(b) The spatial derivatives of an image are also used in image *corner* or *feature of interest* detection.

(i) Give an expression for I_n, the directional derivative of I in the direction $\mathbf{n}$. Show that

$$I_n^2 = \frac{\mathbf{n}^T \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \mathbf{n}}{\mathbf{n}^T \mathbf{n}}$$

where $I_x \equiv \partial I / \partial x$ and $I_y \equiv \partial I / \partial y$. [20%]

(ii) Hence show how corners, or features of interest, can be detected and localised by examining the values of $\langle I_x^2 \rangle$, $\langle I_x I_y \rangle$ and $\left\langle I_y^2 \right\rangle$ at each pixel, where $\langle \ \rangle$ denotes a 2D smoothing operation. Explain why smoothing is required. [30%]

2 The relationship between a 3D world point (X, Y, Z) and its 2D image pixel coordinates (u, v) can be written as follows:

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

(a) Under what assumptions is this relationship valid? Explain the algebraic and geometric significance of s. [20%]

(b) Show how the projection matrix can be decomposed into the product of an upper diagonal matrix of internal camera parameters, and a matrix representing the position and orientation of the camera. Express the elements of these matrices in terms of the internal and external parameters (see Fig. 1). [20%]

(c) Outline an algorithm to recover the elements of the projection matrix from a single image of a known 3D object. State clearly the number of image measurements required and how noisy image measurements are processed in practice. Explain how to recover the camera's position, orientation and internal parameters. [40%]

(d) How can the projection matrix be simplified when viewing a distant and compact object. What are the advantages of using this simplified relationship? [20%]

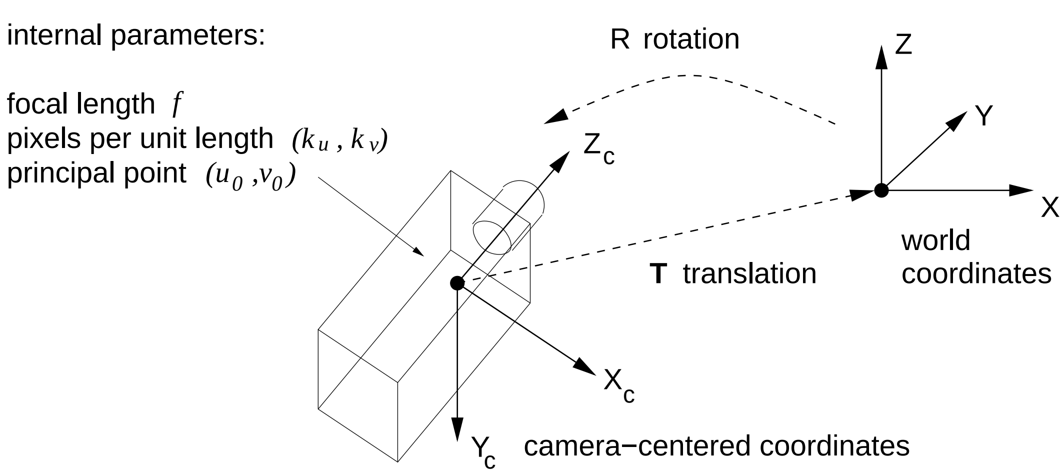


Fig. 1

3 The planar facade of a building is observed by a camera, producing images with pixel positions (u, v) corresponding to positions (X, Y) on the facade.

(a) Show that the relationship between the corresponding points is given by a 2D projective transformation:

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

[20%]

(b) How many degrees of freedom does the 2D projective transformation have? Describe, using sketches, how a square window might appear in any of the images. Be sure to account for each degree of freedom of the 2D projective transformation. [20%]

(c) The horizon (the vanishing line of the plane) is a line $l_1 u + l_2 v + l_3 = 0$ in the image. Lines can be represented in homogeneous coordinates by vectors $\mathbf{l} = [\ l_1 \ \ l_2 \ \ l_3\ ]^T$.

(i) Find two vanishing points which lie on the horizon of the planar facade. Hence find the equation of the horizon in terms of the elements of the projective transformation. [20%]

(ii) Show how the horizon, or vanishing points, can be used to recover the orientation of the camera if the internal parameters are known. [10%]

(d) Derive the relationship between lines on the building's facade (represented in homogeneous coordinates) and their correspondences in the image in terms of the elements t_{ij} of the projective transformation. Show how line correspondences can be used to recover the projective transformation. [30%]

4 (a) What is meant by the *fundamental matrix* in stereo vision? Explain how the fundamental matrix can be estimated from point correspondences. [30%]

(b) For a particular pair of images, the fundamental matrix is estimated to be

$$\mathrm{F} = \begin{bmatrix} 0 & 1 & -300 \\ 1 & 0 & -541.4214 \\ -300 & -58.5786 & 18 \times 10^4 \end{bmatrix} .$$

Given that the epipoles lie in the null spaces of F and F^T, explain why a valid fundamental matrix has maximum rank 2. Check the validity of the matrix F. [20%]

(c) In a three-camera configuration, cameras A and B are related by the fundamental matrix F above (A is left, B is right), while cameras B and C are related by the same matrix F (B is left, C is right).

(i) A point is observed at pixel coordinates $(300, 500)$ in A's image and $(376, 700)$ in C's image. Use the fundamental matrices to deduce the point's location in B's image. Do not calculate any 3D world coordinates. [30%]

(ii) The fundamental matrices cannot always be used to transfer points into B's image. Explain why this is so, and characterise the set of points for which the transfer will fail. [20%]

(TURN OVER

5 A camera views a scene containing a single object. Boundary edges detected in an image at time $t = t_0$ form a square, with one vertex at the origin of the (x, y) coordinate system, as shown in Fig. 2(a). The camera translates with velocity $\mathbf{U} = (U_1, 0, 0)$, expressed in the camera-centered coordinate system shown in Fig. 2(b).

(a) Show that $\dot{x}$, the x-component of image velocity, is given by

$$\dot{x} = -\frac{fU_1}{Z_c}$$

where f is the focal length of the camera and Z_c is the depth of the imaged feature. [20%]

(b) An x-t slice through the resulting spatiotemporal image is shown in Fig. 3. The slice is taken at the position of the dotted line in Fig. 2(a), though other x-t slices at neighbouring positions reveal identical trajectories.

(i) Explain the significance of x-t slices through the spatiotemporal image, as opposed to slices at other orientations. [10%]

(ii) Calculate the depths of the left and right sides of the object. Express your answers in terms of U_1. [20%]

(iii) Deduce as much as you can about the shape of the object. In particular, state whether the object is planar or not. [25%]

(c) Discuss the relative advantages and disadvantages of spatiotemporal image analysis and real-time feature tracking for structure-from-motion applications. [25%]

(cont.

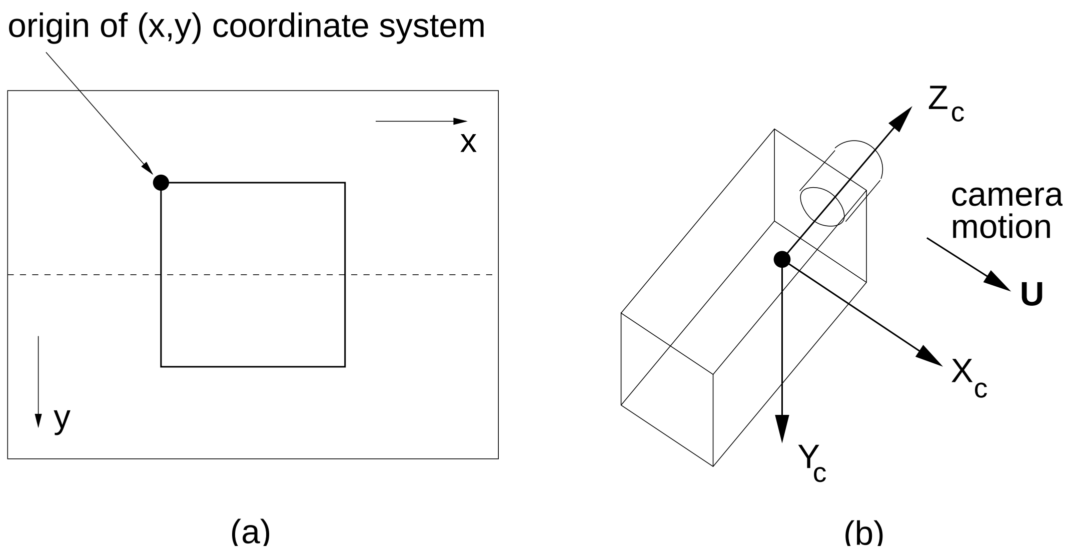


Fig. 2

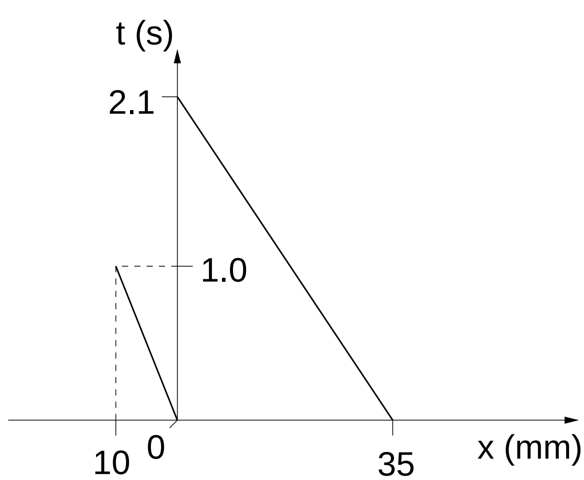


Fig. 3

END OF PAPER