CARDIFF UNIVERSITY
EXAMINATION PAPER

Academic Year:	2014/2015
Examination Period:	Autumn
Examination Paper Number:	CMT107
Examination Paper Title:	Visual Computing
Duration:	2 hours

Do not turn this page over until instructed to do so by the Senior Invigilator.

Structure of Examination Paper:

There are 6 pages.
This examination paper is divided into 2 sections.
There are 4 questions in total.
There are no appendices.
The maximum mark for the exam paper is 100%, and the mark obtainable for a question or part of a question is shown in brackets alongside the question.

Students to be provided with:

The following items of stationery are to be provided:
ONE answer book.

Instructions to Students:

Answer the COMPULSORY question in Section A and TWO questions from Section B.

Important note: if you answer more than the number of questions instructed, then answers will be marked in the order they appear only until the above instruction is met. Extra answers will be ignored. Clearly cancel any answers not intended for marking. Write clearly on the front of the answer book the numbers of the answers to be marked.

The use of translation dictionaries between English or Welsh and a foreign language bearing an appropriate school stamp is permitted in this examination.

SECTION A

Q1. **Compulsory Question**

(a) What are the inputs and outputs of a **basic graphics pipeline**? List the graphics pipeline tasks, and briefly describe each task. [10]

(b) A quadric is the solution of the following equation:

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0$$

Give a **matrix/vector representation** of this equation. [6]

(c) Fixing the reference frame and transforming an object (**object transformation**) can be equivalently represented by fixing the object and transforming the reference frame (**coordinate transformation**). Suppose we want to translate the object by $(-1, 2, -1)$, and then rotate it around z axis by $45°$. What is the **equivalent coordinate transformation**? [4]

(d) Given the following **perspective camera matrix**:

$$P = \begin{bmatrix} 1 & -3 & 0 & 3 \\ -4 & -1 & -2 & 10 \\ 2 & 0 & -2 & 3 \end{bmatrix}$$

and a 3D point in homogeneous coordinates $X = [4\ 0\ 6\ 2]^T$
– What are the **Cartesian coordinates** of the point X in 3D?
– What are the **Cartesian image coordinates**, (u, v) of the projection of X? [6]

(e) Explain the concept, steps of **unsharp masking techniques**. Give the mathematical representation of an **unsharp mask filter**. Can we use a **median filter** as an unsharp mask filter? Justify your answer. [8]

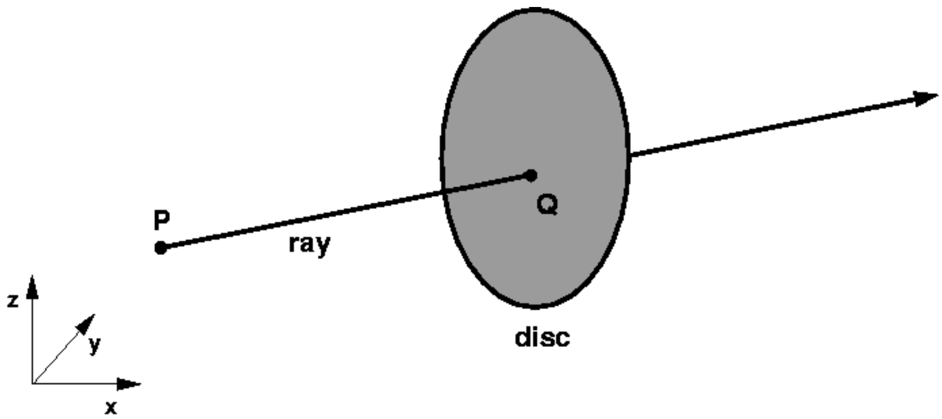
(f) **Feature extraction** has many applications. Briefly describe one example of its applications. [6]

SECTION B

Q2. **Ray Tracing and BSP Tree**

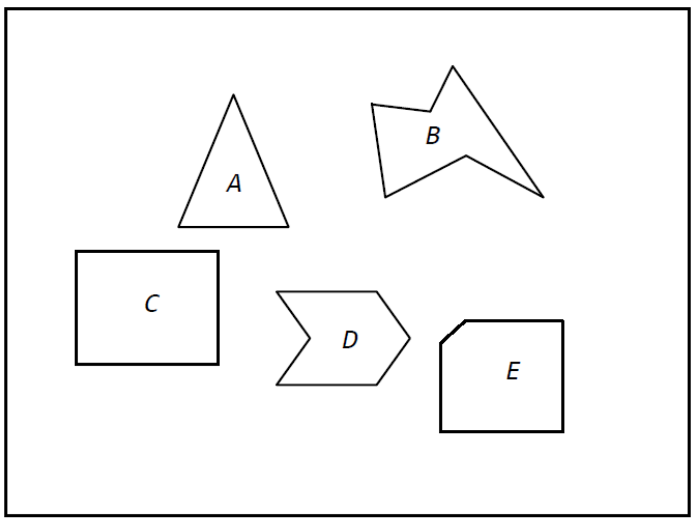
(a) In ray tracing, **Phong illumination** can be used to define the colour of a point on a surface. Qualitatively explain the different **light types** that have to be considered. [9]

(b) A disc is a planar face with a circular boundary. Describe an algorithm to find the **intersection point** Q of an arbitrary light ray from the camera position P with an arbitrary disc in 3D (see figure). Clearly state suitable input parameters to define both the ray and the disc.



[9]

(c) A **BSP tree** is a hierarchical representation for a scene. Partition the scene shown below into regions with a **single convex polygon in each region**, and build a **BSP tree** based on your partitioning.



[12]

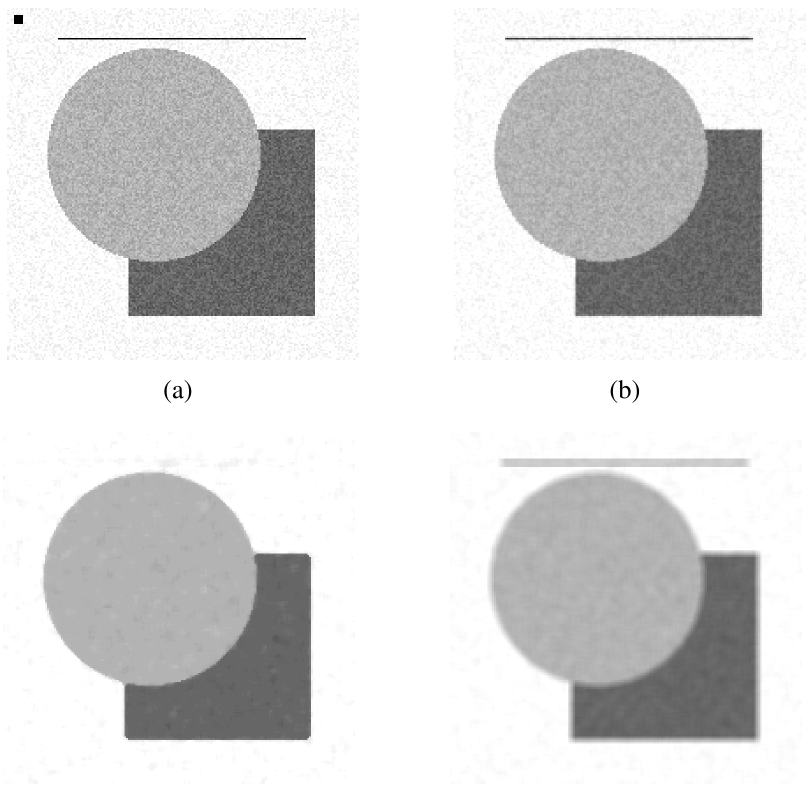
Q3. **Image Processing**

(a) **[Image filtering]** The images (b), (c), and (d) below are the filtering results of the image (a) using the following filters:

1) **box filter**, 2) **Gaussian low pass filter**, 3) **median filter**.

Which filters generate image (b), (c), and (d), separately? Justify your answer.

Note: the small black square on the top left corner of the original image shows the size of the mask that was used. It is not part of the original image.



(a) (b) (c) (d)

[9]

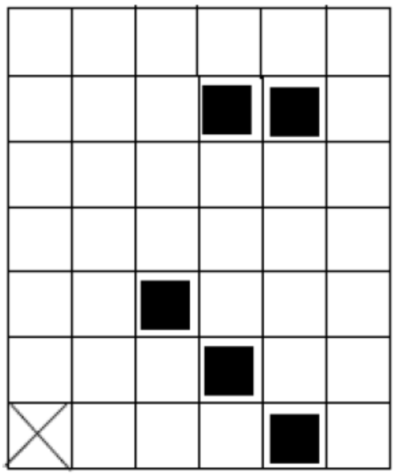
(b) **[Harris Detector]**

(i) Briefly describe the main steps of the **Harris corner detector**. [5]

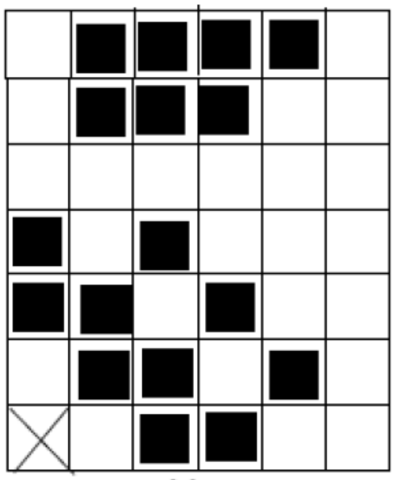
(ii) Suppose we want to generalise the Harris corner detector to detect corners in 3D volumetric images with an intensity value for each (x, y, z) voxel. Explain **which computations need to be generalised** in the original Harris corner detector. [6]

(c) **[Morphology]** Images X and Y are shown below with the crossing in the left bottom showing their correspondence. Design an appropriate **structuring element** such that Y can be obtained after **dilating** X. (black for 1 and white for 0)

Note: mark the centre of the structure element clearly.



X



Y

[10]

Q4. **Epipolar Geometry and Stereo**

(a) Draw a picture to show the **baseline**, **epipolar plane**, **epipoles** and **epipolar lines**. What is the **epipolar constraint** and how can it be used in **stereo matching**? [12]

(b) Consider the following top view of a stereo rig:

Draw the **front views** of the two 2D images and show (and clearly label) the **approximate positions of the epipoles and epipolar lines** for this configuration of cameras. [6]

(c) Given a **conjugate pair of points**, at pixel coordinates $p = (x_l, y_l)$ in the left image and at pixel coordinates $q = (x_r, y_r)$ in the right image of a stereo pair, give the equation that describes the relationship between these points when neither the **intrinsic** nor **extrinsic parameters** of the cameras are known. Also specify how many conjugate pairs of points are needed to solve for all of the unknowns in your equation. [6]

(d) Estimate the **essential matrix** between two consecutive images for a forward translating camera. What is the **equation of the epipolar line** for the point $p = [x\ y\ 1]^T$? [6]

SECTION A

A1. (a)
- The inputs are 3D models, light sources and camera, and the output is 2D pixels in a raster. (2 points)
- The graphics pipeline tasks include transformations, lighting, 3D clipping, projection, 2D clipping, and rasterisation. (2 points)
- The transformations include modelling transformations which convert model coordinates into world coordinates, and viewing transformations which convert world coordinates into camera coordiantes. (1 point)
- Lighting involves evaluate illumination model. (1 point)
- Clipping selects visible part of whole scene for dispalying. 3D clipping selects primitives inside viewing volume. 2D clipping cuts off 2D shapes outside a rectanglular area. (2 points)
- Projection projects 3D primitives onto plane to give 2D shapes. (1 point)
- Rasterisation converts 2D shapes into pixels. (1 point)

[10]

(b) *2 point for each item*

- $X^T M X + v^T X + j = 0$, where $X = [x\ y\ z]^T$
- $M = \begin{bmatrix} a & d/2 & e/2 \\ d/2 & b & f/2 \\ e/2 & f/2 & c \end{bmatrix}$
- $v = [g\ h\ i]^T$

[6]

(c) *2 points for the correct rotation angle and translation offset, and 2 points for the correct order of transformations:*
The equivalent coordinate transformation is rotating the reference frame around the z axis by $-45°$, and then translating it by $(1, -2, 1)$. [4]

(d) *2 points for each item*

- $(4/2, 0/2, 6/2) = (2, 0, 3)$
- $$\begin{bmatrix} 1 & -3 & 0 & 3 \\ -4 & -1 & -2 & 10 \\ 2 & 0 & -2 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \\ 6 \\ 2 \end{bmatrix} = \begin{bmatrix} 10 \\ -8 \\ 2 \end{bmatrix}$$
- $(u, v) = (10/2, -8/2) = (5, -4)$

[6]

(e) *2 points for each big item*

- The process of subtracting an unsharp (smoothed) version of an image from the original images to get sharpened images (high pass filtering).
- Steps:
  - Blur the original image.

- Subtract the blurred image from original (the resulting difference is called the mask)
- Add the mask to the original image

- Unsharp mask filtering can be expressed as:

$$\bar{f} = f + \alpha(f - f * g) = f * ((1+\alpha)e - \alpha g)$$

and the filter kernel is

$$\bar{g} = (1+\alpha)e - \alpha g$$

where f is the original image; $\bar{f}$ is the sharpened image; g is a low pass filter kernel, usually Gaussian kernel; e is a unit impulse (identity), and α is a positive coefficient.

- A median filter cannot be used as an unsharp mask filter as it not low pass or high pass filter.

[8]

(f)
- Example: Panorama Stitching
- Step 1: extract features of two or more images;
- Step 2: match the features;
- Step 3: Align the images according to matched features.

[6]

SECTION B

A2. (a)
- Ambient light:
  - comes from all directions and is reflected in all directions
  - intensity the same everywhere (3 points)
- Diffuse light:
  - comes from a specific direction and is reflected in all directions
  - intensity depends on the angle between the incoming light and the surface normal (3 points)
- Specular light:
  - comes from a specific direction and is reflected in a preferred direction
  - intensity depends on the angle between the direction of the perfectly reflected light and the direction to the viewer (3 points)

[9]

(b) INPUT:
  - Ray defined by camera position p and direction vector d
  - Disc defined by centre c, radius r and normal vector n (2 points)

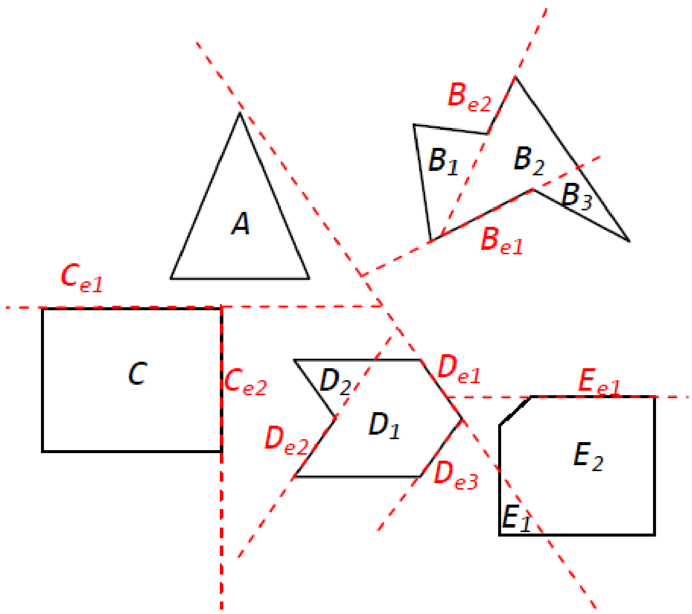
I. Compute intersection point between ray and disc plane:
Plane equation: $(x - c)^t n = 0$, $x \in R^3$
Ray parametrisation: $r(t) = p + td$, $t \geq 0$
Intersection: $(p + td - c)^t n = 0$, $td^t n = (c - p)^t n$ (3 points)

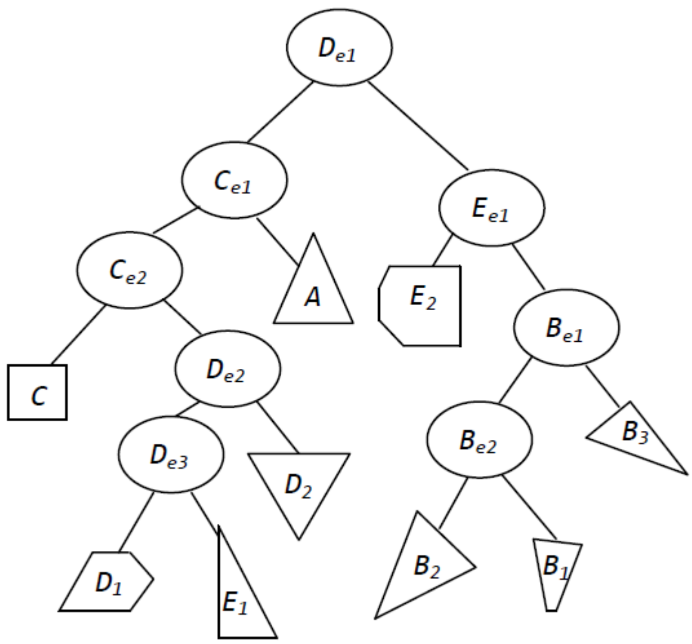
1. If $d^t n = 0$, return no intersection point (1 point)
2. $t = ((c - p)^t n)/d^t n$ (1 point)
3. If $t < 0$, return no intersection point
   else $q = p + td$ (1 point)

II. If $\|q - c\| \leq r$, return intersection point q
else return no intersection point (1 point)

[9]

(c) *6 points for correct partitioning, and 6 points for correct BSP tree. Note: the answer is not unique. The following two figures show an example answer of this question.*





[12]

A3. (a) *1 point for each correct selection of the filters, and 2 points for each justification:*

(i) (b) Gaussian low pass filter. The Gaussian filter can blur the image, but because its kernel has a higher weight on the centre element than those off-centre, the centre pixel has bigger effect on the final results. Figure (b) shows that the black thin line was blurred, but its centre is obviously different from the pixel around.

(ii) (c) median filter. Median filter eliminates small objects. The thin black line can only be removed by median filter.

(iii) (d) box filter. The kernel of a box filter has the same weight for every element, thus it blurs the images uniformly. In image (d), the thin black line was uniformly blurred around the original line.

[9]

(b) (i) *1 points for each item:*

- Compute Gaussian derivatives at each pixel
- Compute second moment matrix M in a Gaussian window around each pixel
- Compute corner response function R
- Threshold R
- Find local maxima of response function (non-maximum suppression)

[5]

(ii) *2 points for each item:*

- Gaussian derivatives at each voxel to be generalised with respect to x,y and z coordinates
- The second moment matrix M to be generalised to 3×3 matrix
- The corner response function will be defined as $R = det(M) - \alpha trace(M)^3$

[6]

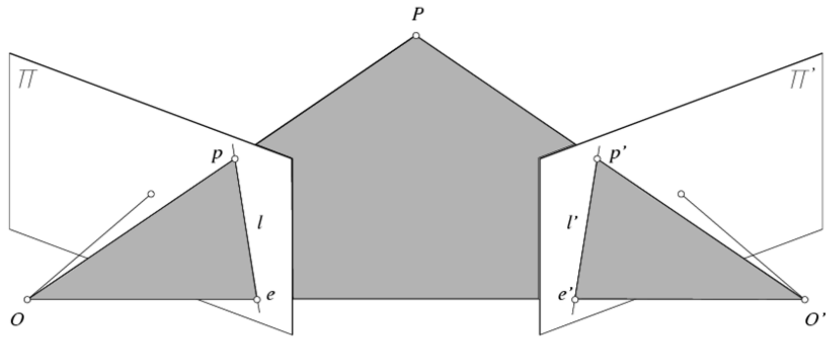
(c) The structuring element is a 2-by-3 matrix:

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & \times \end{bmatrix}$$

with $\times$ being the origin (and is zero) [10]

A4. (a)
- A picture is shown below, where O and O' are the camera centres; (2 points)
  - Line $O - O'$ is the baseline; (1 point)
  - $OO'P$ is on an epipolar plane; (1 point)

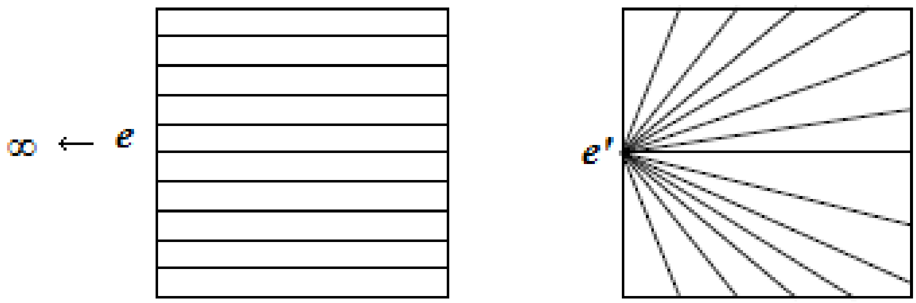
  - e and e' are epipoles; (1 point)
  - l and l' are epipolar lines (1 point)



- The epipolar constraint represents geometry of two cameras, reduces a correspondence problem to 1D search along an epipolar line. (2 points)
- A point in one view "generates" an epipolar line in the other view. The corresponding point lies on this line. (2 points)
- Epipolar geometry is a result of co-planarity between camera centres and a world point - all of them lie in the epipolar plane. (2 points)

[12]

(b) *3 points for each figure*
See the following:



[6]

(c) *2 points for each figure*

- $q^T F p = 0$ where F is the 3×3 Fundamental matrix containing 8 degrees of freedom.
- Each correspondence generates one linear constraint on the elements of F;
- hence at least 8 conjugate pairs (and no noise) are needed to compute it using a linear algorithm such as the 8-point algorithm.

[6]

(d) *2 points for each item*

- The essential matrix is calculated by $E = [t_\times]R$, where $[t_\times]$ is a skew-symmetric matrix related to translation vector. For a forward translating camera, suppose the translation in z direction is t_z), we have $R = I$, and

$$[t_\times] = \begin{bmatrix} 0 & -t_z & 0 \\ t_z & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- therefore

$$E = \begin{bmatrix} 0 & -t_z & 0 \\ t_z & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- From $l = Ep$, the epipolar line for point $p = [x\ y\ 1]^T$ is

$$l = \begin{bmatrix} 0 & -t_z & 0 \\ t_z & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} -t_z y \\ t_z x \\ 0 \end{bmatrix}$$

[6]