CARDIFF UNIVERSITY
EXAMINATION PAPER

Academic Year:	2016/2017
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 5 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) Briefly outline *one* application for *each* of the following *three sub-fields* in Visual Computing: *Graphics, Image Processing, and Computer Vision*. [9]

(b) List the *main operations* in the 3D graphics pipeline *in order*, and briefly describe each operation. [8]

(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 a transformation from an object coordinate system to a camera coordinate system consists of a rotation around z axis by $90°$, followed by a translation by $(1, 2, 0)$.

1) Describe its *equivalent object transformation*.

2) Give the *homogeneous matrix* representation of this transformation.

3) Given a point $(1, 0, -1)$ in the object coordinate system, compute its new coordinates in the camera coordinate system. [10]

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

(e) Give a definition of the *epipolar line*. Briefly describe a *basic stereo matching algorithm* using the epipolar line. [8]

SECTION B

Q2. **Computer Graphics**

(a) Briefly describe three types of *light reflections* and four types of *light sources*. [7]

(b) Briefly explain *flat, Gouraud, and Phong shading*. Compare their advantages and disadvantages. [12]

(c) A triangle surface with vertices p_0, p_1, p_2 is lit by a point light source with light intensity L and position p_L. The viewer position is p_V. Let p be the centroid of the triangle. State the formula for *Phong's reflection model* to compute the light reflected from the point p, using p_0, p_1, p_2, p_L, p_V and p to define any necessary intermediate quantities in the formula. It is assumed that the material properties of the surface are known, and you can use any symbols to represent the material properties in stating the formulae. [11]

Q3. **Image Processing**

(a) **[Image filtering]** The pixel values of an image and a filter kernel (denoted by g) are given in the two tables below.

2	3	3	8	7
6	4	2	0	3
3	5	**1**	3	4
5	6	8	7	2
4	1	9	0	5

The pixel values

0.25	0.25	0.25
-0.2	-0.2	0.25
-0.2	-0.2	-0.2

The filter kernel: g

(i) Compute the filtering results of the pixel in the centre of the image (with value 1) using:
1) *a* 3×3 *box filter*, 2) *a* 3×3 *median filter*, and 3) *the filter kernel* g. [6]

(ii) Compute the filtering results of the pixel at the top-left corner of the image (with value 2) using the box filter, based on the following boundary extension methods, separately:
1) *clip filter*, 2) *wrap around*, 3) *copy edge*, and 4) *reflect across edge*. [12]

(b) **[Morphology]** For the image and the structuring element given below, give the *dilation*, *erosion*, *closing* and *opening* results. Note that the points with '$\times$' are local origins with value 1.

0	1	0	0
1	1	1	0
0	1	1	1
0	0	0	0
$\times$	0	1	0

Image to be processed

0	1	0
1	$\times$	1
0	1	0

Structuring element

[12]

Q4. **Camera Calibration**

The relationship between a 3D point at world coordinates (x, y, z) and its corresponding 2D pixel at image coordinates (u, v) can be defined as a projective transformation using a 3×4 camera projection matrix P.

(a) Can the matrix P incorporate any *lens distortions* that might be in the camera? Briefly justify your answer. [5]

(b) Give two lists, one specifying the *intrinsic* camera parameters and the other giving the *extrinsic* camera parameters. [10]

(c) Show how P can be decomposed into a product of matrices that contain elements expressed in terms of the *intrinsic* and *extrinsic* camera parameters. [5]

(d) Give the main steps of an algorithm for computing the matrix P from a single image of a known 3D "*calibration object*". [10]

SECTION A

A1. (a) *3 points for each application example. Other appropriate application examples are also acceptable.*

- Graphics: Data Visualisation. Visualise real data, highlight the structure of the data, or show some special information.
- Image processing: Image filtering. Remove image noise, or enhance image features, so that the image features can be easily identified.
- Computer vision: Robot navigation. Use camera mounted on the robot to take images while it's moving, and use the images to reconstruct the 3D scene in the environment, to help the robot to avoid obstacles and reach its destination.

[9]

(b)

- The operations in the 3D graphics pipeline include transformations, lighting, 3D clipping, projection, 2D clipping, and rasterisation. (2 points – lose 1 point if not in correct order)
- 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 displaying. 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)

[8]

(c)

- The equivalent coordinate transformation is rotating the object around the z axis by $-90°$, and then translating it by $(-1, -2, 0)$. (4 points, 2 for the correct rotation angle and translation offset, and 2 for the correct order of transformations)
- The rotation matrix is

$$R = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(1 point)

- The translation matrix is

$$T = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(1 point)

- The combined translation matrix is

$$M = TR = \begin{bmatrix} 0 & 1 & 0 & -1 \\ -1 & 0 & 0 & -2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(2 points)

- The new coordinates are computed by

$$p' = Mp = M \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -3 \\ -1 \\ 1 \end{bmatrix}$$

So, the new coordinates are $(-1, -3, -1)$ (2 point)

[10]

(d) *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]

(e)
- The epipolar line is the intersection of the epipolar plane with the image plane. (2 point)
- A basic stereo matching algorithm consists of the following steps: For each pixel in the first image
  - (i) find corresponding epipolar line in the right image (2 point)
  - (ii) examine all pixels on the epipolar line and pick the best match (2 point)
  - (iii) triangulate the matches to get depth information (2 point)

[8]

SECTION B

A2. (a)
- Ambient light reflection: comes from all directions and is reflected in all directions; intensity is the same everywhere. (1 points)
- Diffuse light reflection: comes from a specific direction and is reflected in all directions; intensity depends on the angle between the incoming light and the surface normal. (1 points)

- Specular light reflection: 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. (1 points)
- Ambient light source: light from the environment. (1 points)
- Directional light source: light from infinite distance in a specified direction. (1 points)
- Point light source: light from a single point. (1 points)
- Spot light source: light emitted in a cone. (1 points)

[7]

(b)
- Flat shading calculates illumination per polygon. Each pixel in the polygon is assigned the same colour. (2 points)
- Gouraud shading compute illumination for vertices of polygon using vertex normals. The colours are linearly interpolated between vertices. (2 points)
- Phong shading compute illumination per pixel using normal of the corresponding point. The normals are linearly interpolated across the polygon using vertex normals. (2 points)
- Flat shading are easy to calculate, but have obvious artefact, except for polygon objects. (2 points)
- Both Gouraud and Phong shading are time-consuming, but can generate smoother effect than flat and facet shading (2 points)
- Phong shading is slower than Gouraud shading, but the former has smoother effect the latter. (2 points)

[12]

(c)
- The normal of the triangle is calculated by $n' = (p_1 - p_0) \times (p_2 - p_1)$, and $n = \frac{n'}{|n'|}$ (3 points)
- The centroid of the triangle is $p = (p_1 + p_2 + p_3)/3$ (1 point)
- The light direction is calculated by $l = \frac{(p_L - p)}{|p_L - p|}$ (1 point)
- The viewer direction is calculated by $v = \frac{(p_v - p)}{|p_v - p|}$ (1 point)
- The perfect reflection direction is calculated by $r = 2(n \cdot l)n - l$ (3 points)
- Let the material ambient, diffuse, and specular properties be R_a, R_d, R_s, respectively, and let the shininess exponent be σ, then according to the Phong's reflection model, the reflected light from point p is $I = R_a L + R_d(n \cdot l)L + R_s(r \cdot v)^\sigma L$ (2 points)

[11]

A3. (a) (i) *2 points for each correct answer.*
1) 4, 2) 4, 3) 4.5 (1 point if the answer is -3.15). [6]

(ii) *3 points for each correct answer.*
1) 5/3, 2) 35/9, 3) 10/3, 4) 10/3. [12]

(b) See next page. 3 points for each correct answer

1	1	1	0
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1

Dilation

0	0	0	0
0	1	0	0
0	0	0	0
0	0	0	0
0	0	0	0

Erosion

0	1	0	0
1	1	1	0
0	1	1	1
0	1	1	0
1	0	1	0

Closing

0	1	0	0
1	1	1	0
0	1	0	0
0	0	0	0
0	0	0	0

Opening

[12]

A4. (a) No, because lens distortions are usually highly nonlinear, which cannot be represented in P. [5]

(b) (i) The five intrinsic parameters are the focal length, f, principal point coordinates, $(u0, v0)$, and the pixel size scaling parameters, ku, kv. (5 points)

(ii) The six extrinsic parameters are the three translational and three rotational parameters that define the rigid body motion between the world coordinate frame and the camera coordinate frame. (5 points)

[10]

(c)

$$P = K[R|T] = \begin{bmatrix} fk_u & 0 & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}$$

[5]

(d) (i) Since P is a 3×4 matrix with 11 unknowns (the overall scale of P does not matter), observing a 3D scene containing at least 6 known points (in a non-degenerate configuration) is sufficient. . (2 points)

(ii) Each point gives a pair of equations of the form

$$u = \frac{su}{s} = \frac{P_{11}X + P_{12}Y + P_{13}Z + P_{14}}{P_{31}X + P_{32}Y + P_{33}Z + P_{34}}$$

$$v = \frac{sv}{s} = \frac{P_{21}X + P_{22}Y + P_{23}Z + P_{24}}{P_{31}X + P_{32}Y + P_{33}Z + P_{34}}$$

(2 points)

(iii) Rewrite these 12 equations in the form $Ap = 0$, where p is a 12×1 vector of unknowns, and A is a 12×12 matrix of coefficients. (3 points)

(iv) Use least squares to solve for p by computing the eigenvector corresponding to the smallest eigenvalue of $A^T A$. Note: This is only an approximate solution and often this is then used as a starting point for a nonlinear optimization. (3 points)

[10]