

SEMESTER 2 FINAL ASSESSMENT 2020 - 2021

COMPUTER VISION

---

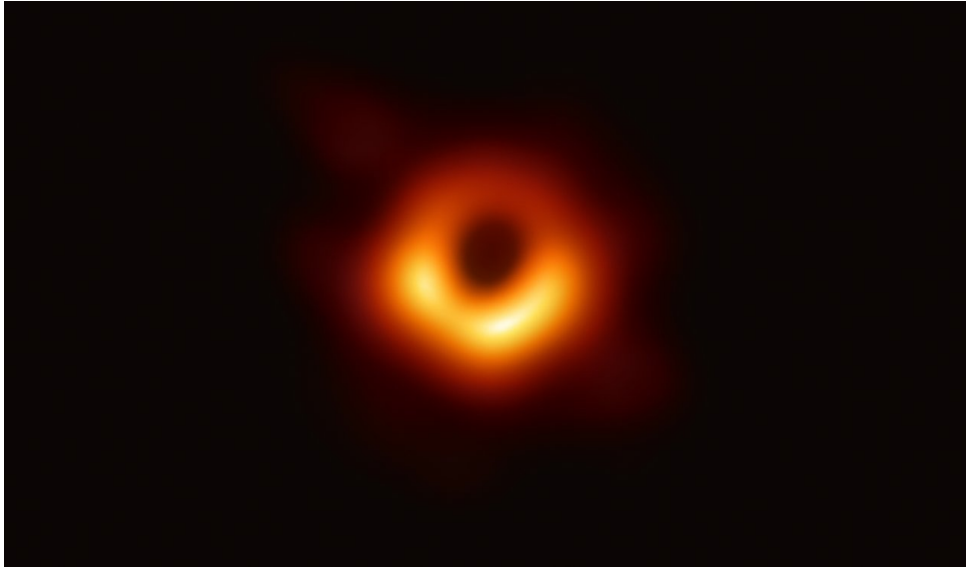
This paper contains 6 questions

Answer **THREE** questions.

An outline marking scheme is shown in brackets to the right of each question.

Each question is worth 33 marks. A maximum of 99 marks is available for the paper. We expect that you should spend no more than 30-40 minutes on any question.

9 page paper.

**Question 1.**

First Image of a Black Hole (<https://www.eso.org/public/images/>)

- (a) The above image is the first image of a black hole. As we can see, it looks quite blurry. **Provide** an explanation as to why it is so blurred. [5 marks]
- (b) **Design** an edge detection system to detect the edge of the black hole, and **explain** how you address the blurred edges within the image. [12 marks]
- (c) **Design** a shape detection system to detect as many shapes in the image as you can when you have *no template to use*. **Explain** how to *improve* your shape detection system if you are allowed to use templates. [16 marks]

**Question 2.**

- (a) **Explain** the difference between the spatial domain (that the above image is in) and the frequency domain after applying the Fourier transform on the above image. **List** some advantages and disadvantages of analysing the image in its frequency domain.

[8 marks]

- (b) Suppose you are not satisfied with the noise present in the image above. **Design** a system which can remove the noise using the *frequency* information.

[4 marks]

- (c) Suppose the noise type in the above image is Gaussian. **Design** a noise removal filter which works in the *spatial domain* of the image. When the filter size is large, **design** a way to improve the noise removal efficiency, and **explain** why this works.

[12 marks]

- (d) **Explain** how the image boundaries impact the size of the image with noise removed in the noise removal techniques you developed in (c).

[9 marks]

**TURN OVER**

**Question 3.**

- (a) The police received an image with a person (or maybe more) in it (see above), but it cannot help since the image is too dark to recognise the scene in it. **Design** a contrast enhancement system which will enable humans to see more details of the above image. You should **list** the details of your design in dealing with individual pixels.

[9 marks]

- (b) In order to see what the exact content in the above image, you threshold the image into binary images using different thresholds. Your police colleague says “histogram equalisation will help to get an image with better contrast, and then the thresholding will be more effective”. **Discuss** if you agree with them and give reasons. **Discuss**, giving reasons, if the automatic thresholding selection method – Otsu’s method – would be more effective.

[8 marks]

- (c) **Describe** the hysteresis thresholding method. Will it give a better result than the methods you tried in (b) and why?

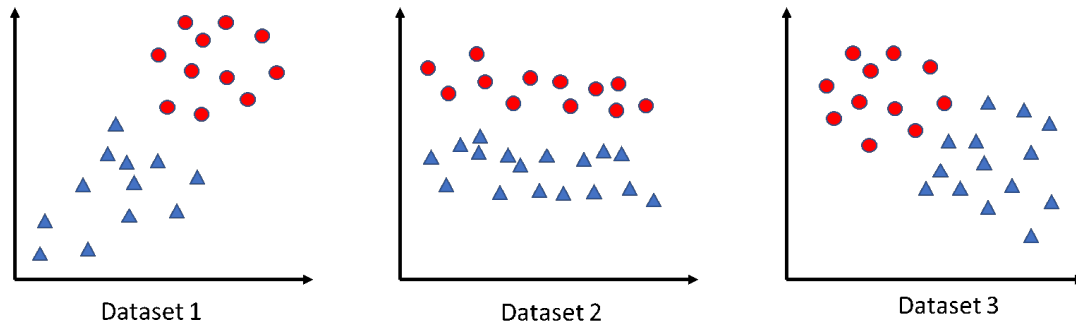
[7 marks]

- (d) Suppose you are motivated by the hysteresis thresholding method. Can you use more than two thresholds to **develop** a more effective thresholding method than using two thresholds? If yes, give the details; if not, give the reason.

[9 marks]

**Question 4.**

- (a) You have 3 training datasets of data points  $p$  belonging to two classes on a 2D space  $(x,y)$  with the following distributions.



**For each dataset**, answer the questions below:

- Draw** a picture illustrating all the principal axes.
- Describe** if it is possible to correctly classify the dataset by using a threshold after projecting onto *any* principal component? If your answer is *yes*, **define** which principal component they should be projected onto. If your answer is *no*, **explain** why it is not possible in 1 or 2 sentences.

[9 marks]

- (b) You want to apply the following transforms against the origin  $(0,0)$  in the following order:

- Rotate counter-clockwise (CCW)  $45^\circ$
- Translate by  $(2,1)$
- Scale by  $0.5$

The Affine transform is defined as:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

**Find** the Affine transform matrix which performs the 3 transforms in the given order. **Ensure** you show all working.

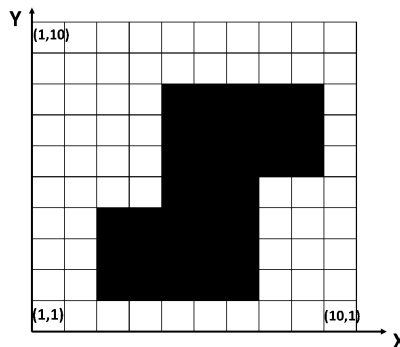
[11 marks]

**TURN OVER**

- (c) You want to take two photos of a static object from two slightly different directions with two cameras to recover dense 3D geometry (a 3D point cloud) of this object from the photos. The final 3D points should have real-world scale with the unit of their  $(x, y, z)$  coordinates in metres or millimetres. **Design and describe** the whole process. *You do not need to mention any specific algorithm or equation.* [13 marks]

**Question 5.**

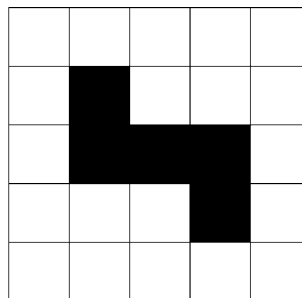
- (a) Consider the shape depicted by black pixels as follows on the (x,y) pixel domain. The image shows some examples of pixel coordinates.



**Calculate** the compactness  $C(s)$  of the above shape, Irregularities  $I(s)$  and  $IR(s)$ , considering the outer border and performing calculations in the “pixel” domain. *Do not consider sub-pixels (e.g., half- or quarter-pixels with floating-point values).* **Ensure** you show all working.

[16 marks]

- (b) Consider the connected area component depicted by the solid black pixels.



Hu's first invariant moment is given by  $M1 = \eta_{20} + \eta_{02}$ . **Compute**  $M1$  and **show** this value is invariant to rotation and scale by comparing with the **90° rotated component** (rotation about the centre of mass) and **scaled component by a factor of 3**. **Ensure** you show all working.

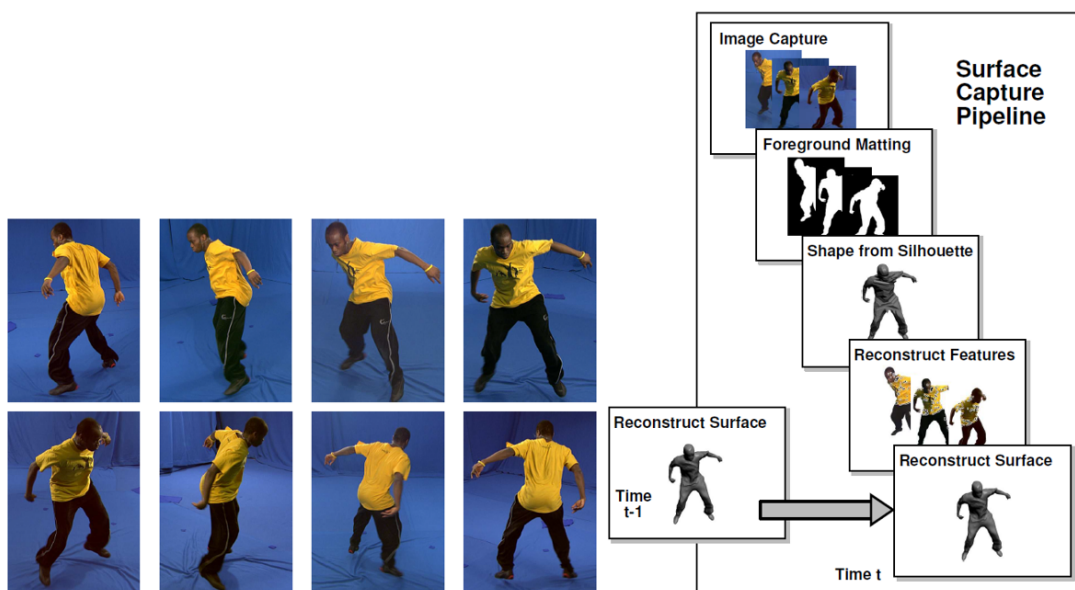
Note: Be aware that Hu's moments are only invariant within some tolerance, which means the values may not be exactly same.

[17 marks]

**TURN OVER**

### Question 6.

You want to build a 3D capture studio using a wide-baseline multiple view stereo method to provide full 3D geometry model with 360° appearance (texture) for a human model from 8 cameras. The following figures show the captured 8 camera views and its processing pipeline for dynamic 3D human model reconstruction. This pipeline includes human region segmentation, reprojection of the 2D silhouette into the 3D space, feature extraction and matching. The matching can be performed between views (View  $i$  and View  $i + 1$ ) and also between frames in individual videos (Time  $t - 1$  and Time  $t$ ) for optimal surface reconstruction.



- (a) **Define** all hardware constraints (including ones in the environment) you can consider for this studio under the goal of achieving most accurate 3D human surface model reconstruction. For each constraint, **provide** justification and invariants required for robustness.

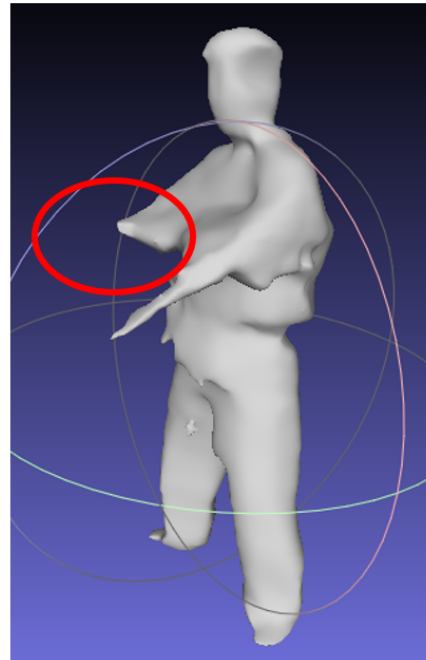
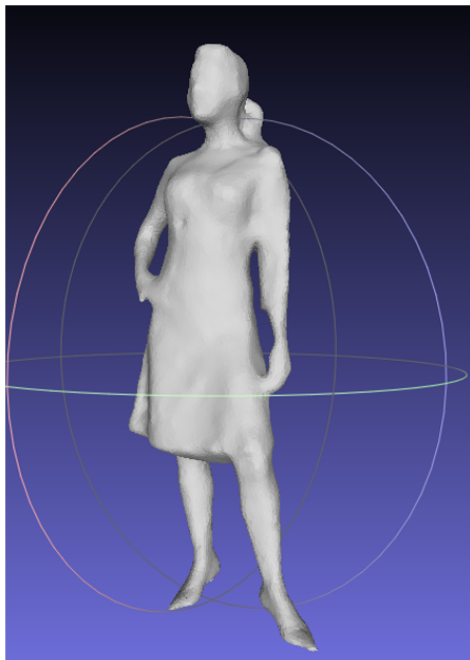
[15 marks]

- (b) **Design** a system to automatically match corresponding points between viewpoints and also points between frames in individual videos for 3D model reconstruction. This can be a combination of several methods. **Clarify** the advantages and limitations of the methods you propose. *Do not consider any algorithm which was not dealt in the lectures.*

[10 marks]



- (c) You expected a complete 3D model like left image below from the system, but got an incomplete model like right one which failed at reconstructing the right arm. Assuming that the hardware settings were fine, **analyse** where we can find problems in the whole software pipeline. **List** all possible problems which could induce this result.



[8 marks]

**END OF PAPER**