PRIFYSGOL ABERYSTWYTH UNIVERSITY

SEMESTER ONE 2025–2026

COMPUTER SCIENCE

CSM4220 : COMPUTER VISION

Time allowed: **2 HOURS**

Answer all **THREE** questions

All questions carry equal marks.

YOU MAY BRING THE FOLLOWING INTO THE EXAMINATION:
UP TO 3 SHEETS OF NOTES ON A4 PAPER (WHICH MAY BE DOUBLE-SIDED). THERE ARE NO ADDITIONAL RESTRICTIONS ON CONTENT OR FORMATTING OF THESE NOTES.
NO ELECTRONIC DEVICES.

CALCULATORS ARE PERMITTED (CASIO FX991, FX83 OR FX85 MODELS ONLY)

STUDENTS MAY SUBMIT ANSWERS TO THIS PAPER IN EITHER WELSH OR ENGLISH

Once seated, students may complete the front cover of the answer book(s) and the attendance slip ONLY.

DO NOT TURN THE EXAMINATION QUESTION PAPER OVER UNTIL INSTRUCTED TO DO SO.

The Scenario

This exam paper uses the following scenario as a context for the questions:

Computer Vision on the Aberystwyth Prom. The three questions (provided below) each cover areas related to the Aberystwyth Prom, where computer vision-based solutions have been requested for different problems, and you have been hired as a consultant to advise on them.

1. **Object detection.** The town council has installed a number of low-cost cameras to monitor the usage of the promenade footpath. Promenade users include walkers, cyclists, people on scooters, roller-skaters, skateboarders, dog-walkers, and joggers. They would like to use these cameras to estimate the number of people in each of these groups during daylight hours i.e. for each hour of the day, the system should record an estimate of the number of people in each group.



Figure 1: Example of people on a prom with some detected.

a) Suppose that the processing power available is just about sufficient to use some of the more lightweight deep learning approaches. Suggest a suitable deep learning network for object detection and classification (as shown in Figure 1) from those you have studied on the module. In your answer, briefly describe what kind of network it is, how it works, and explain why you have chosen that network. [12 marks]

b) Outline a computer vision pipeline that might be able to provide reasonable estimates on these low-powered devices. Provide reasoning for your answer and indicate any assumptions and the likely limitations. [12 marks]

c) In testing it becomes apparent that the variable weather and lighting can cause the performance to drop at certain times. Suggest three basic image pre-processing operations and explain how they might be useful (or not) in this context. Also, suggest which image processing operations among the three would be the best and why? [9 marks]

2. **Image matching.** The Wales Coastal Monitoring Centre (WCMC) has setup CoastSnap sites around the Welsh coast to allow people to take pictures with their mobile phones from certain positions / directions. The public can upload their images (an example of 'Citizen Science') and WCMC would like to use these images to monitor coastal erosion. In particular, they would like to measure the height of the beach at the sea wall, as shown in Figure 2.



Figure 2: Two images from coastsnap registered, with one warped and overlaid on the other.

a) Repeatability and distinctiveness are two of the six key attributes we look for in a good feature. Name the other four attributes and briefly explain the importance of each of those four attributes. [6 marks]

b) The first step is to align different images from the same site. Give 3 reasons (with an explanation) why the images would need alignment. [6 marks]

c) Select an algorithm that would be suitable for detecting and matching points across pairs of images. Briefly describe how it works and why that algorithm was selected. [9 marks]

d) The next stage is to warp one image to align with the other. Given a set of candidate point matches (where some may be incorrect) describe a process and algorithms needed to reliably distort one image to align with another. Your answer

should include the following terms and explain their meanings: epipolar; homography; RANSAC; and affine. [12 marks]

3. **Segmentation.** Following on from the previous question, the next stage in the Wales Coastal Monitoring Centre's (WCMC) CoastSnap project is to perform *image semantic segmentation*. They are also interested in applying segmentation algorithms to 3D data they have captured using photogrammetry.



Figure 3: Example of segmentation of two aligned images (green is beach and blue is the seawall).

a) Following on from the previous question, the next stage in the Wales Coastal Monitoring Centre's (WCMC) CoastSnap project is to perform *image semantic segmentation*. Explain what image semantic segmentation is, and how it could be used in this project for important elements such as the beach, sea and seawall (see Figure 3). [4 marks]

b) You have seen various semantic segmentation neural networks in the module, including CNNs, UNets and Transformer based. Describe in outline the structure of **one** of these networks and how it is trained. This should include the basic structure of the network, the format of the output, and the choice of loss function. Which algorithm for image semantic segmentation would you choose for this

problem, and why do you think this approach is appropriate? [12 marks]



Figure 4: Example of a 3D model of the Welsh coast constructed using photogrammetry.

c) The WCMC have used a drone to capture a sequence of images along the coastline, and have used some existing software to create 3D models of parts of the coast (e.g. Figure 4). They are excited by the possibility of a new method VGGT [1] (that was provided to you earlier) and have asked for your opinion on it. Outline the method described in the paper, and discuss the suitability of the method for their 3D reconstruction problem. [14 marks]

d) Once they have captured and created the 3D models, WCMC would like to apply segmentation to these, to label the 3D points that are sea, beach, wall, cliff etc.. Discuss the suitability of the VGGT method for this task, or how it could be made suitable. Explain your reasoning. [3 marks]

END OF PAPER

[1] Wang, J., Chen, M., Karaev, N., Vedaldi, A., Rupprecht, C. and Novotny, D., 2025. VGGT: Visual geometry grounded transformer. In Proceedings of the Computer Vision and Pattern Recognition Conference (pp. 5294-5306).