Midterm Exam, COL 780 / JRL 780

Maximum Marks: 50. Duration 90 mins

Instructions: Attempt all questions. Do not ask for any clarifications during the exam. Make appropriate assumptions, if you think there is any ambiguity in the question, and write it clearly along with your answer

Q1 (3+5 Marks) What is the difference between image gradients and edges? Give a method to obtain edges from image gradients. → Describe any

Q2 (10 Marks) We saw in the class how Eigenvectors of the covariance matrix formed from a set of images can be used as its low dimensional representation. Here instead of using the actual image we used the reconstructed image composed using the eigenvectors and corresponding coefficients. We also discussed in the class that the same can be utilized to recognize objects, e.g. face recognition. Assume a gallery of 100 face images from as many subjects, and a probe image of one of the subjects (not identical to the one in the gallery), can you discover an algorithm to use eigen vectors to find out which gallery image matches the most with given probe image.

Q3 (10 Marks) Give the algorithm for Harris corner detection.

Q4 (5 Marks) Given a dataset of images, how can you obtain visual word vocabulary for large scale image instance recognition.

Q5 (7 Marks) Given "n" point correspondences between a pair of images, how can you estimate an affine transformation between the images using least squared error minimization.

calculus

Q6 (3 Marks). Describe "depth of field" in the image formation process.

Q7 (3 Marks). How do you make a patch descriptor rotationally invariant?

Q8 (4 Marks). What is the difference between the derivative of a Gaussian filter and the difference of Gaussians filter? application

Department of Computer Science and Engineering
Indian Institute of Technology Delhi
COL780 and JRL780

Instructions: Please attempt all questions. If you feel any question/statement is ambiguous, please write your assumptions clearly, and then answer as per your assumptions.

Time: 2 hours | **Major Exam** | Maximum Marks: 100

Pre Midterm (10 Marks)

1. [5 marks] Consider a checkerboard image, and we wish to detect corners of the checkerboard. Give atleast two techniques for detecting corners, which may lead to detected corners lying at subpixel locations.

2. [5 marks] Harris corners are dependent on the scale of the object in the image. Give an algorithm to detect features which can be reliably detected under wide variations in the visible scale.

Image Classification (30 Marks)

3. [10 marks] What is Vanishing Gradient problem in training deep neural networks. How does ResNet architecture helps mitigate it.

4. [10+10 marks] Describe how you can use Transformer architecture for Image classification, giving mathematical expressions to compute self-attention. What is the role of multi-head attention block in a Transformer architecture.

Object Detection (30 Marks)

5. [10 marks] How does FastRCNN achieve higher speed over RCNN. How does FasterRCNN achieve even higher speed over the FastRCNN?

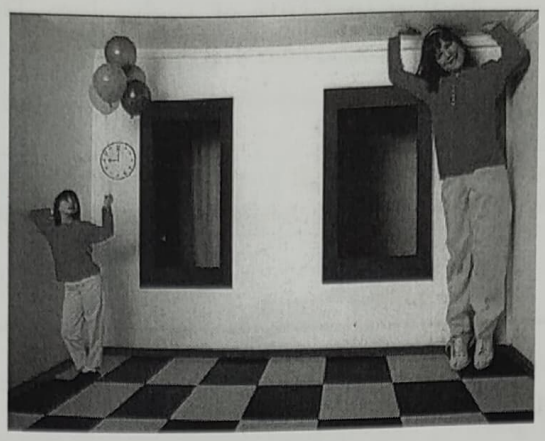
6. [5 marks] I build a naive object detector which predicts every possible rectangle in the image as the object. What will be the precision and recall of such a model? How do you compute mAP metric for benchmarking object detection techniques. Describe in a step-by-step manner.

7. [5 marks] Describe how is object detection a set prediction task. How is it handled differently in a CNN based, and Transformer based models.

8. [10 marks] How does cascaded architecture helps achieving extremely low false positive rate In a traditional object detection system.

Please go on to the next page...

Geometry (30 Marks)



9. [5 marks] The "Ames room" principle (as shown in the figure above) has been used widely in TV and movie productions for special effects when it was necessary to show actors in giant size next to actors in small size. For example, production of The Lord Of The Rings film trilogy used several Ames room sets in Shire sequences to make the heights of the diminutively-sized hobbits correct when standing next to the taller Gandalf. The illusion of an ordinary room is because most information about the true shape of the room does not reach the observer's eye. The geometry of the room is carefully designed, using perspective, so that, from the peephole, the image projected onto the retina of the observer's eye is the same as that of an ordinary room. One key aspect of preventing the observer from perceiving the true shape of the room is the peephole. It has at least three consequences (1) It forces the observer to be at the location where the image projected into his eye is of an ordinary room. From any other location, the observer would see the room's true shape. (2) It forces the observer to use one eye to look into the room, preventing him from getting any information about the real shape of the room from stereopsis, which requires two eyes. (3) It prevents the observer from moving his eye to a different location, preventing him from getting any information about real shape of the room from motion parallax. The key idea behind the Ames room is in the special structure of the room which is not rectangular or cuboidal. Can you think of the true 3D shape of the room which can create perspective like the one shown above?

10. [5 marks] Prove that the projection of parallel lines in 3D on the image plane intersect at a vanishing point.

11. [10 marks] Describe, with derivation, Zhang's method for camera calibration.

12. [10 marks] Consider a point P in world coordinate system, and its two projections p_1 and p_2, in two cameras with their intrinsic matrices as identity. Prove mathematically that, given p_1, and irrespective of the depth of P, p_2 will always lie on a certain line in the second image.

End of exam