



COMSATS University Islamabad (Lahore Campus)
Department of Computer Science

Terminal Exam- Fall 2023

Course Title:	Computer Vision	Course Code:	CSC455	Credit Hours:	3(3,0)
Course Instructor/s:	Dr. Allah Bux Sargana	Programme Name:	BSCS	Section	
Time Allowed:	3 Hours	Maximum Marks:	100		
Name	Registertaion No.				
Important Instructions / Guidelines:					
Attempt all question					

Question No 1. Write the brief answer to the following questions :

Marks: 5x3=15

CLO: <1>; Bloom Taxonomy Level: <Understanding>

- What are the essential components of the pinhole camera model? How does the pinhole camera model represent the mapping between the 3D world and the 2D image? Discuss the limitations of the pinhole camera model and how more complex camera models address these limitations.
- In a pinhole camera with a 25 mm focal length, if an object is positioned 2.5 meters away along the optical axis, calculate the size of the object's image on the image sensor. Consider a full-frame sensor with dimensions 30 mm x 20 mm.
- Explain the concept of Epipolar geometry. What is the Epipolar line, and how is it related to the fundamental matrix? How does the Epipolar constraint help in matching points between two images?

Question No 2. Solve the following problems.

Marks: 5x3=15

CLO: <2>; Bloom Taxonomy Level: <Applying>

- In an image classification task using BoVW, suppose you have a training dataset with 100 images, and each image is represented by a histogram with 200 bins (visual words). If you use a Support Vector Machine (SVM) classifier with 5000 parameters to be trained, calculate the total number of parameters to be learned in the SVM model.
- Two images have SIFT descriptors with Euclidean distances as follows: If a threshold for accepting matches is set at 5.0, determine which descriptor pairs would be considered valid matches.

Descriptor Pair	Euclidean Distance
1 to 2	6.5
1 to 3	8.2
1 to 4	4.0

- Given an original image with dimensions 512x512 pixels, construct a scale-space pyramid with three octaves and four scales per octave. If the original image is sampled at a base scale of 2, calculate the dimensions of the images at different scales in the pyramid.

Question No 3. Answer the following question.

Marks: 5x3=15

CLO: <3>; Bloom Taxonomy Level: <Understanding>

- Explain a step-by-step breakdown of how the Hough Line Transform algorithm detects lines in an image.
- Describe the main phases of Bag of Visual Words (BoVW), including feature extraction, codebook creation, and histogram representation.

- c. Explain the process of scale-space extrema detection in the SIFT algorithm. How does the construction of a scale-space pyramid contribute to the detection of key points?

Question No 4. Answer the following questions.

Marks: $14+7+7+7=35$

CLO: <4>; Bloom Taxonomy Level: <Analysis>

- a. Analyze the CNN architecture with the following layers: Provide answers and calculations for each part of the question. Better to write the answer in a tabular form.
- After each layer, calculate the output volume dimensions (width, height, and depth/channel). Show your calculations.
 - How many trainable parameters are there in this CNN architecture? Show your calculations.
 - If we have a training set of 10,000 images, how many forward and backward passes are needed to complete one training epoch? Assume a batch size of 64.

The architectural detail of the model

INPUT-CONV-POOL-CONV-POOL-FC-OUTPUT

- Input layer: Accepts grayscale images of size 64×64 pixels.
 - Convolutional layer: Applies 16 filters of size 3×3 with a stride of 1 and no padding.
 - Max pooling layer: Performs max pooling with a pool size of 2×2 and a stride of 2.
 - Convolutional layer: Applies 32 filters of size 5×5 with a stride of 1 and no padding.
 - Max pooling layer: Performs max pooling with a pool size of 2×2 and a stride of 2.
 - Fully connected layer: Consists of 256 neurons.
 - Output layer: Generates predictions for ten classes using softmax activation.
- b. Analyze the GoogleNet inception module with three parallel convolutional branches and filter sizes of 1×1 , 3×3 , and 5×5 , and each branch produces 64 output channels; what will be the total number of output channels in the concatenated output?
- c. Suppose you have two bounding boxes with coordinates (x_1, y_1, x_2, y_2) of $(50, 60, 150, 200)$ and $(70, 80, 180, 220)$. Calculate and compare the intersection over the union (IoU) of these two bounding boxes.
- d. Three bounding boxes with confidence scores (PC value) and IoU values are given as follows: Apply Non-Maximum Suppression and identify the final set of retained bounding boxes. Consider IoU threshold 0.5.
- Box 1: Confidence = 0.7, IoU = 0.5
 - Box 2: Confidence = 0.8, IoU = 0.5
 - Box 3: Confidence = 0.6, IoU = 0.7

Question No 5. Answer the following questions.

Marks: $4 \times 5 = 20$ Marks

CLO: <5>; Bloom Taxonomy Level: <Creating>

- Suppose you need to design a model based on ResNet architecture. Discuss the crucial innovations and benefits of using ResNet to train deep neural networks.
- Design a convolutional layer in ResNet with an input size of 128×128 , a kernel size of 5×5 , and a stride of 1, and write the output dimensions of this layer.
- Design an inception module, considering it part of a deep-learning model for image classification. Also, explain the significant benefits you expect from using the inception-based network for this problem.
- Design a Long Short-Term Memory (LSTM) unit. Discuss its key components, including the input gate, forget gate, memory cell, and output gate. Also, explain how the LSTM architecture addresses the vanishing gradient problem compared to traditional RNNs.