Assignment 1

CPS584 - Advanced Intelligent Systems and Deep Learning

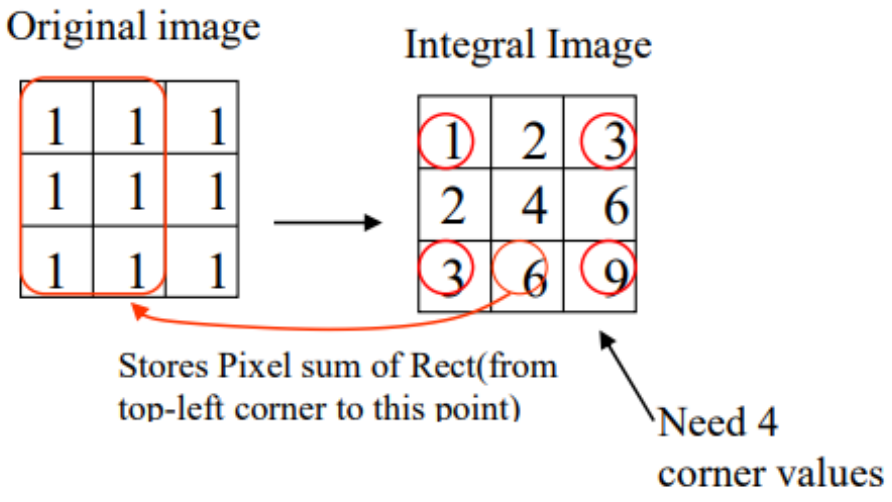
Released Date: 01/30/2023

Requirements

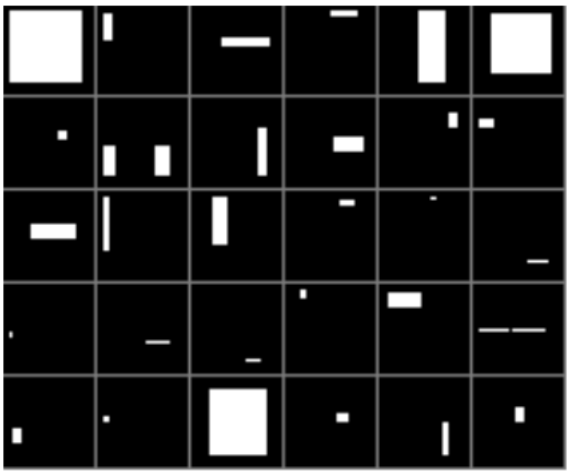
In this assignment, you will solve practical and interesting problems. By completing the project, you will gain valuable hands-on experience in the design, implementation and evaluation of classification algorithms. The details are listed as below.

You are provided with the "Flowers.zip" file which contains images of two classes: *Rose* and *Tulip*. For each class, 40 training images and 20 testing images are given.

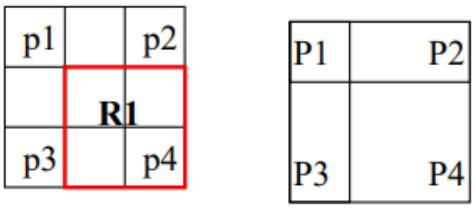
1. You need to resize (imresize function in MATLAB) the image to the size 50 x 50. Then, you write MATLAB code to extract the Haar-like features for each training and testing image. The details are as follows:
    a. You compute the integral image (Lecture "Handcrafted Features", slides 16-20).



b. By using the integral image, you compute the Haar-like features with 30 boxes below.



Note: use the integral image to compute the rectangle sum.



$$Pixel\_Sum(R1) = P4 - P2 - P3 + P1 = 4$$

c. You then extract Haar-like features for both training and testing images. Each training/testing image has a 30-dim Haar-like feature.

2. You train 2 classifiers: K-nearest neighbor (KNN) and Neural Networks (NN) on the Haar-like features of training images. Then, you test the trained models on the Haar-like features of the testing images. Please report the accuracy rate for each class. You are requested to try different K values: 1, 3, 5, and 7.
3. Manually collect additional 30 training images for both classes: Rose and Tulip. Note that those images **must not be identical** to any already given training/testing image. You may need to use any tool to resize and crop the newly collected images to the size 50 x 50.
4. Run 2 classifiers: K-nearest neighbor and Neural Networks on the testing images. Note that both classifiers (KNN and NN) are trained on **the new training data** – there are 40 (already provided) + 30 (newly collected) = 70 training images for each class. Again,

you are requested to try different K values: 1, 3, 5, and 7. Please report the accuracy rate for each class.

5. Discuss the accuracy rates in (2) and (4). Which one is better? Will more training data lead to a better performance?

What to Submit

1. A well-documented MATLAB program that implements the aforementioned problem in the Assignment 1. You must submit your program source code and the newly collected training data set.

2. A well-written, concise project report. It should include: (a) title and names of group members; (b) the analysis of each problem; (c) the issues during the implementation; (d) the solutions to overcome the issues in (c); (e) the contribution of each individual member; and (f) the powerpoint slides (maximum 20 slides) used in the Assignment presentation.
For each group, you must submit the files above in a single zipped folder. Your group will be required to **do a face-to-face evaluation** for the grading. One group member will submit the file on behalf of the group.

Important: Your submission will be thoroughly checked. If any plagiarism (from Internet, former students, or anywhere else) is found in this assignment, an F will be assigned to course grade and an academic dishonesty report will be given.

Submission Due: 11:55pm, February 22, 2023