

Programming Assignment 1: Data Understanding and Processing (10 points)

Shen-Shyang Ho (Dr.)

September 14, 2022

1. Download image dataset to your local drive.

(a) **Go to the following Google Drive link (Students must be logged in to their Rowan accounts):**

i. <https://drive.google.com/drive/folders/1GfmG8cNpqSLlkBBIG2qUTRBG8yPmFW8d?usp=sharing>

(b) **Download the 'Negatives.zip' file and the 'Weed-4class-(Your Dataset Number).zip' file from the Google Drive Link.**

(c) **Extract the contents of each zip file into the directory where your python scripts will be held.**

i. Each extraction should produce a folder of the same name.

ii. Inside each folder are the image files with a .csv file containing the labels.

(d) **Download the file 'getting-started.py' file from Canvas and place it in the same directory as the extract folders.**

i. This file requires the pandas and matplotlib python libraries which can be installed using the following command in your terminal: 'pip install pandas matplotlib'

(e) **Inside that script set the DATASET_NUMBER variable to your assigned number, then run that script. This will combine the two extracted folders into a single dataset.**

i. Everything will be combined into the 'Weed-4class-(Your Dataset Number)' folder

ii. If everything has been done correctly, a screen with 12 random images should be presented to you.

2. Use OpenCV to perform image processing and feature extraction

(Reference: https://docs.opencv.org/4.x/d6/d00/tutorial_py_root.html)

(a) **Histogram Equalization (Image Intensity Normalization)**

i. Choose 10 random images from your dataset (two from each class).

ii. Convert the color images to grayscale images (MUST use iteration; No points given if no iteration is used) (0.5 point)

iii. Plot 10 grayscale images with their corresponding pixel intensity histograms. (0.5 point)

iv. Perform histogram equalization on the 10 images. Plot the NEW intensity equalized grayscale images and their corresponding equalized pixel intensity histograms. (1.5 point)

v. Pick a grayscale image and its corresponding equalized image. Plot the 2 images next to each other. What did you observe? (0.5 point)

(b) RGB histogram

- i. Choose 5 random images from your dataset (one from each class and different from Question 2)
- ii. Plot the images with their corresponding RGB histograms (The three histograms MUST be in one image). (1 point)

(c) Histogram Comparison (Measures of Similarity and Dissimilarity)

- i. Pick 2 images from the same class and 1 image from another class.
- ii. Perform histogram comparison using the following metrics/measures.
 - Euclidean Distance
 - Manhattan Distance
 - Bhattacharyya distance
 - Histogram Intersection

For this task, you will compare histogram by computing the metrics/measures of (1) the 2 images from the same class, AND (2) 2 images from different classes. (2 points) (Note: You can also use other packages.)

(d) Image Feature Descriptor: SURF (Speeded-Up Robust Feature)

- i. Pick a random image and perform keypoint extraction using SURF feature descriptor with 128 dimensions.
- ii. Use a threshold value so that you obtain 40-50 keypoints in the image. What is the threshold you use? What is the number of keypoints you extracted? (1.5 points)
- iii. Plot the keypoints on the image. (0.5 point)

(e) Dimensionality reduction (using PCA)

- i. Randomly pick 2 classes.
- ii. Convert all the images to grayscale pixel intensity histograms. (These will be the vector representations of the images)
- iii. Perform Principal Component Analysis (PCA) dimensionality reduction on the set of histograms to 2 dimensions. (1 point) (Note: You should not use the class labels)
- iv. Plot the 2D points using 2 different colors for data from the 2 classes. Are the data from the two classes separable? (1 point)