David Skoda

CS 435

Assigment 1 Write Up

1. See hw.py for the code I wrote. In order to run this code, please run it in a python 3 environment. I use functions that do not work in python 2.x. There are 3 csv files as exported from the excel sheet: i1.csv, i2.csv, i3.csv. These are the files that work with the code. Command line usage is as follows:

python3 hw.py <file.csv> <bucket\_size>

This will produce a histogram based on the bucket size specified in the command line arguments.

1. Observations on the output
   1. General observations on all three files- For a small bucket size, more data needs to be stored, as there is more buckets. This however allows you to better see how the data is laid out in terms of frequency of pixel values. The first and second images share a much more similar layout of data, whereas the third one has a differing layout. I think this is because of the darkness in the pictures. In the third picture the trees are significantly darker shaded than anywhere in the first two pictures. This contributes to the higher spike of values between 0-60 for the third picture, whereas the first and second share more lighter values, so they are more concentrated between the 80-200 values, with sharp spikes around 90. The first and second are also spread out over the larger range of 80-200 with only that narrow spike around 90, while the third image has the bulk of values between 0-60 and a thick spike around 30.
   2. For i1.csv: Bucket sizes observed- 3, 8, 15
      1. In all three instances of the first image, there are clearly 3 spikes observable. They are occurring around 95, 145, and 190. The total range is from 57-244. The frequency of values is much higher on a per bucket basis as you go from a bucket size of 3->8->15. With a bucket size of 3, the bucket containing the peak around value 95 only has a frequency of about 6.5k, while at bucket sizes of 8 and 15, it has a frequency of 16k and 24k, respectively. This is because the buckets absorb surrounding buckets as they join into larger and larger buckets, thus losing overall data about what the image looks like but saving space.
      2. In the second image, the range is a little larger, from 24-245. This is because of that two very dark spots towards the bottom left corner, and also the sort of bottom right. It also differs from the first image in that the middle of the 3 peaks is shifted to higher values slightly
      3. The third image has the largest range, going from 0-255. The bucket frequencies get very low after about 145, until 255 however. This image is significantly different from the first two in that most of the values are in the 0-70 ranges, then it levels out until 132 wherein it peaks again. As stated above, this appears to be because the darker values fit into the lower numbered buckets.
2. I don’t have any data on representing the same object under rotations in 3D space like in I4, I5, and I6, but my guess is that enough data could be lost that you might not even realize they are the same object. It really depends on what color the different views of the object are. If an object is the same color or a similar color all around, then the histograms for rotations of it will look similar. The histograms might be misleading and show different data if the rotation of the object reveals drastically differing sides of the object, or different sides of the object show drastically different colors.