Ryerson University

Department of Electrical & Computer Engineering
Faculty of Engineering & Architectural Science

Course Title:	ELE: Intelligent Systems
Course Number:	888
Semester/Year (e.g.F2016)	W2021

Instructor:	Dr. Xiao-Ping Zhang

Assignment/Lab Number:	4
Assignment/Lab Title:	4

Submission Date:	April 11, 2021
Due Date:	April 11, 2021

Student LAST Name	Student FIRST Name	Student Number	Section	Signature*
D'Souza	Jude	07724	06	J.D

Reset Form

*By signing above you attest that you have contributed to this written lab report and confirm that all work you have contributed to this lab report is your own work. Any suspicion of copying or plagiarism in this work will result in an investigation of Academic Misconduct and may result in a "0" on the work, an "F" in the course, or possibly more severe penalties, as well as a Disciplinary Notice on your academic record under the Student Code of Academic Conduct, which can be found online at: http://www.ryerson.ca/senate/current/pol60.pdf

Objective

The purpose of the lab was to use unsupervised learning with natural clusters within a set of data for iterative learning processes. The data used was an image of a house, and the algorithm used was K-means algorithm for clustering the unlabelled data.

Theory

During unsupervised learning, unlabelled data samples with natural clusters can be identified. By using K-means algorithm, the unknown parameter values from the probability densities of the data can be calculated. K-means randomly initializes the mean vectors and computes the Euclidean distance between the nearest cluster. The means are the recomputed and once there are no observed changes in the means, the algorithm ends.

Observations and Analysis

Prelab

In the prelab, the dominant colors on an image “House.tiff” were calculated and plotted. The observed result can be seen below:

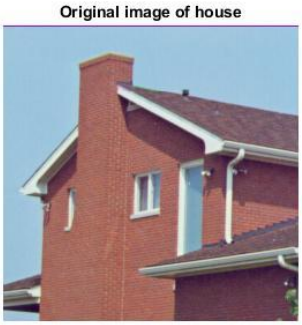

Figure 1: Original Image of House

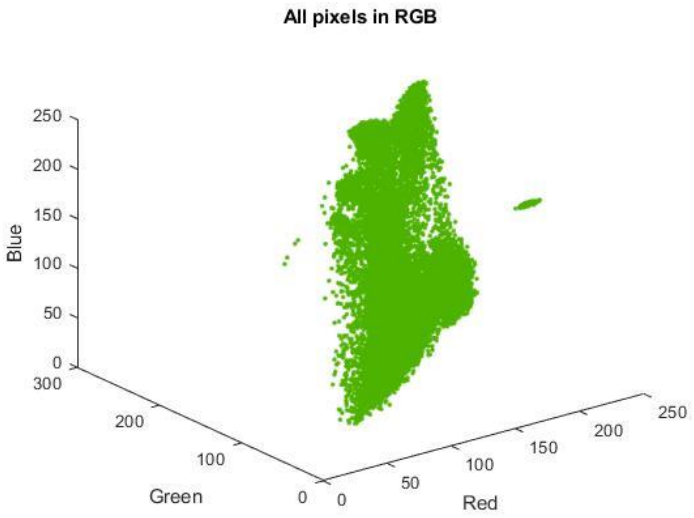

Figure 2: Clustering of colors on the original image

Part A

The starting values of the K-means were randomly assigned, and the algorithm was run to solve. The error criterion of J was calculated as the algorithm progressed which can be seen in the following image:

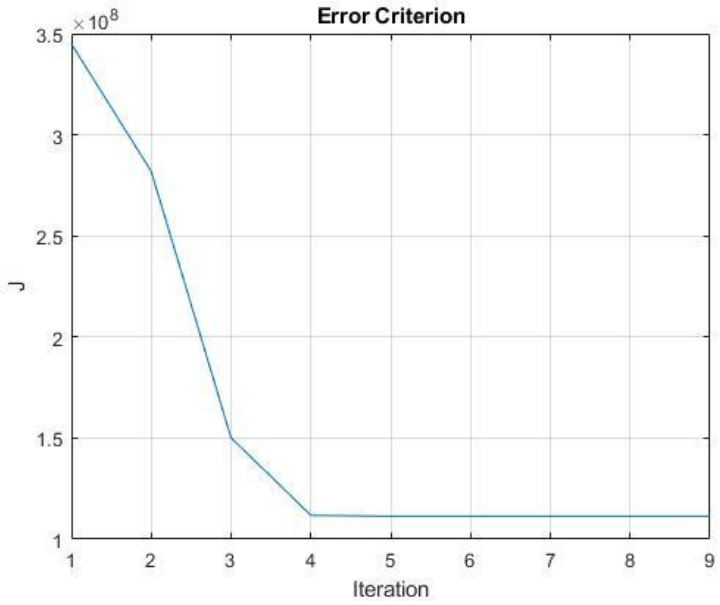

Figure 3: Error criterion of J as Iterations increase

From the graph above, Figure 3, the error criterion of J steadily approaches a smaller and smaller value as the number of iterations in the algorithm increase. This is because as the algorithm progresses the error must become smaller and smaller in order of the change of the means values to show no observed changes. Since the c value of the image is 2, there are only 2 available clustering's to be seen. The clustering of the means can be seen as well:

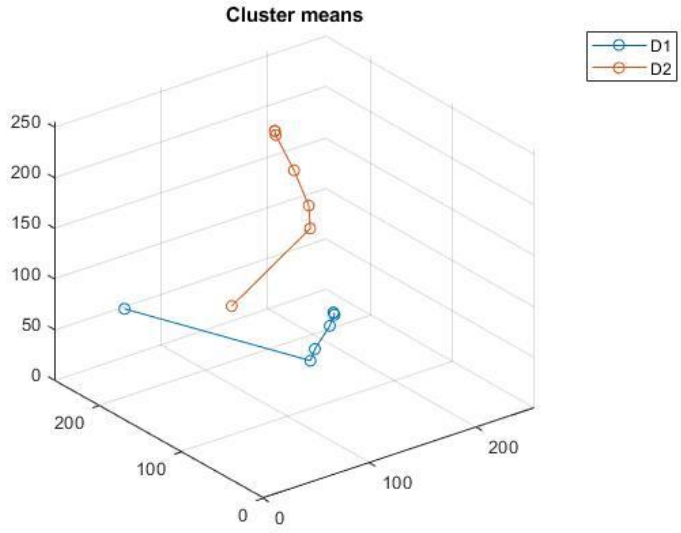

Figure 4: Clustering of the means as iterations increase

From Figure 4, as the number of iterations progress, the clustering of the means Euclidean distances changes and approaches a smaller number until there is no change. This is important because from the image, the first initial steps of the clustering are large but as the algorithm progress, the changes become smaller and smaller until the observed result has no change. The RGB from the unique clustering can be plotted as well using the mean values to represent the unique RGB value associated to the image. From the c value, 2 colors are used in the RGB space.

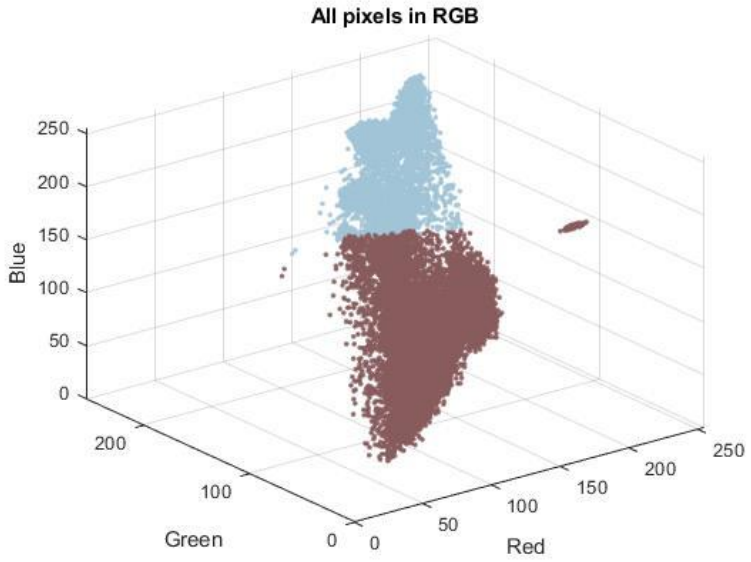

Figure 5: 2 Color Clustering of the original image in RGB

From Figure 5, the 2-color image clustering can be seen in the RGB color space. From the color space, the dominant mean colors of the image can be seen. The images are plotted in a 3d plane to show the colors in the RGB color space. A reconstruction of the image can also be seen:

Figure 6: Reconstruction of the original image

From figure 6, the reconstructed images colors can be seen. There are only the who mean colors, D1 and D2 which can be seen since the c value is only equal to two. This image lacks detail and clarity of that of the original image but has the outlines and general color shape very similar to the original.

Part B

Two c = 5 K-means were compared to each other in this section of the lab. The first iteration of when c = 5, is called Z. Z, similarity to Part A, consist of starting values of the K-means which were randomly assigned. After the K-means algorithm was run, the Z in the RGB space can be plotted.

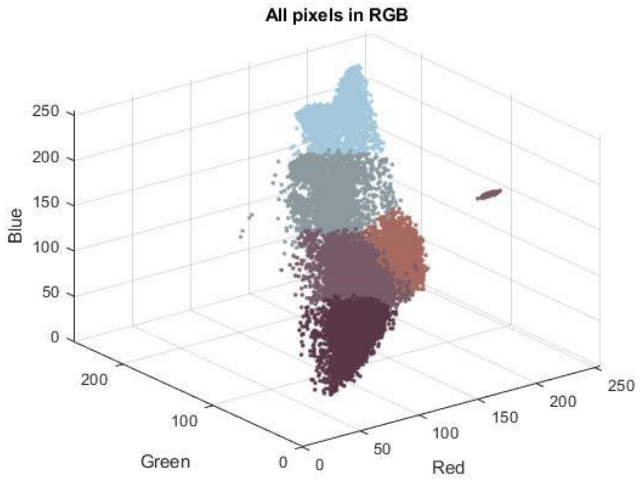

Figure 7: 5 Color Clustering of Z in the RGB space

From Figure 7, the color clustering of Z in the RGB space can be seen. Since the c in this portion of the lab is 5, the number of colors seen is also 5. This provides a more accurate assessment of the colors in the RGB space vs when c was 2 in Part A of the lab. By using a greater value in c, more accuracy can be drawn towards and image. The reconstruction of the image can also be seen:

Figure 8: Reconstruction of the original image using Z

From figure 8, the reconstructed images can be seen according to Z. Since there are now 5 colors, a far greater amount of detail can be seen compared to that of Part A when there were only 2 colors. The image is still not as detailed as the original but now certain portion of the image has much higher depth of detail compared to that of the Part A image. Another iteration of when c = 5, was created called W. W, similarity to Z, consist of starting values of the K-means which were randomly assigned. After the K-means algorithm was run, the W in the RGB space can be plotted.

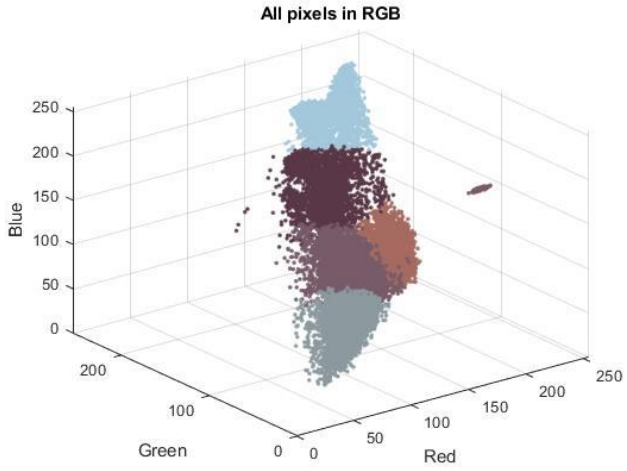

Figure 9: 5 Color Clustering of W in the RGB space

From Figure 9, the color clustering of W in the RGB space can be seen. Since the c in this portion of the lab is 5, the number of colors seen is also 5. This provides a more accurate assessment of the colors in the RGB space vs when c was 2 in Part A of the lab. By using a greater value in c, more accuracy can be drawn towards and image. The reconstruction of the image can also be seen:

Figure 10: Reconstruction of the original image using W

From figure 10, the reconstructed images can be seen according to W. Since there are now 5 colors, a far greater amount of detail can be seen compared to that of Part A when there were only 2 colors. The image is still not as detailed as the original but now certain portion of the image has much higher depth of detail compared to that of the Part A image. A comparison between W and Z can be seen below:

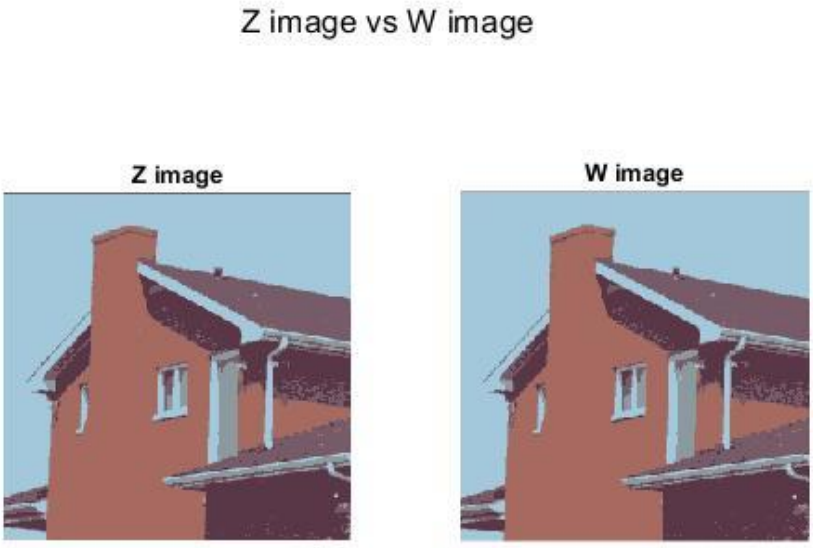

Figure 11: Z image vs W image

From Figure 11, it is important to note the similarities between both Z and W. Even though Z and W had both different starting parameters, the result is nearly identical with slight discrepancies between the two images.

Part C

The Xie-Beni (XB) of the Z and W were also calculated to assess the quality of the clustering solutions found in Part B. The values of the Xie-Beni can be seen below:

Xie-Beni (XB) of Z	Xie-Beni (XB) of W
0.2673	0.2670

Table 1: Xie-Beni (XB) of Z and W

There is a close relation between the Xie-Beni (XB) of Z and W as seen in table 1. This is because the quality of the clustering is very consistent across both images. The only major change discrepancy between the two images is if the c value will change. If the c value would change, the quality of the clustering could change which will in turn affect the Xie-Beni (XB).