

COMP ENG 4SL4: Machine learning

Assignment 3 - k-Means Clustering for Image Segmentation and Compression

Instructor: Dr. Sorina Dumitrescu

TA: Ruilin Ji

Richard Qiu – qiur12 – 400318681

Simulation Result

1st Image

Original image:



K = 2 Init = 1
MSE = 620.97



K = 2 Init = 2
MSE = 627.23



K = 2 Init = 3
MSE = 707.14



K = 3 Init = 1
MSE = 337.90



K = 3 Init = 2
MSE = 356.93



K = 3 Init = 3
MSE = 504.40



K = 10 Init = 1
MSE = 93.49



K = 10 Init = 2
MSE = 102.57



K = 10 Init = 3
MSE = 87.38



K = 20 Init = 1
MSE = 65.86



K = 20 Init = 2
MSE = 47.02



K = 20 Init = 3
MSE = 41.23



K = 40 Init = 1
MSE = 33.68



K = 40 Init = 2
MSE = 30.93



K = 40 Init = 3
MSE = 25.79

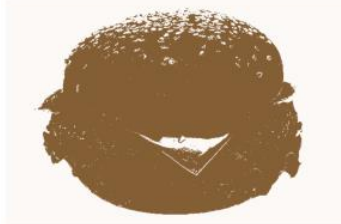


2nd Image

Original image:



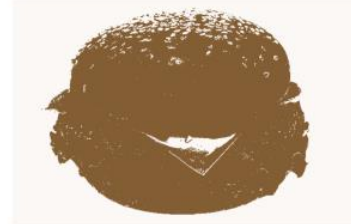
K = 2 Init = 1
MSE = 1349.85



K = 2 Init = 2
MSE = 1349.50



K = 2 Init = 3
MSE = 1349.76



K = 3 Init = 1
MSE = 1369.13



K = 3 Init = 2
MSE = 1375.83



K = 3 Init = 3
MSE = 576.66



K = 10 Init = 1
MSE = 173.67



K = 10 Init = 2
MSE = 189.05



K = 10 Init = 3
MSE = 101.15



K = 20 Init = 1
MSE = 83.89



K = 20 Init = 2
MSE = 166.03



K = 20 Init = 3
MSE = 56.11



K = 40 Init = 1
MSE = 43.48



K = 40 Init = 2
MSE = 40.48



K = 40 Init = 3
MSE = 28.80



Analysis Result

Duck			
K	Init	MSE	# iterations
2	1	620.97	13
2	2	627.23	2
2	3	707.14	2
3	1	337.9	5
3	2	356.93	3
3	3	504.4	3
10	1	93.49	6
10	2	102.57	5
10	3	87.38	52
20	1	65.86	5
20	2	47.02	15
20	3	41.23	21
40	1	33.68	6
40	2	30.93	8
40	3	25.79	9

CheeseBurger			
K	Init	MSE	# iterations
2	1	1349.85	3
2	2	1349.5	3
2	3	1349.76	6
3	1	1369.13	3
3	2	1375.83	2
3	3	576.66	2
10	1	173.67	2
10	2	189.05	2
10	3	101.15	8
20	1	83.89	6
20	2	166.03	2
20	3	56.11	11
40	1	43.48	8
40	2	40.48	8
40	3	28.8	23

These two tables show the corresponding results of each image analysis results with different K values and initialization strategies. For different K values, when init is equal to 1 and 2, the initialization strategy is randomly picked the centers from the data points. And when init is equal to 3, the initialization strategy (K means++) is firstly choosing a random center then find the min distance of each data point to the available centroids, the data point has the largest distance will be chosen for the next centroid. This way we can make sure that the initialized centroids are far from each other which could provide a better clustering result.

Discussion

Based on these two images analysis data and how pictures behave with different K values. We can see that in general, as K increases, the MSE between new image and original image would decrease and the picture would provide a better visual quality. And the second initialization strategy (K means++) would produce a lower MSE results when K is large enough (above 10) in these two image examples. But when K is small, the MSE for the second initialization strategy could larger than the first one. Overall, this behaviour is expected which each cluster could represent a color of the pixel, so when K is small, the color difference would be the largest when choose the second strategy which could usually lead a higher MSE and when K is becoming larger this strategy could hold more color pixels which provide a better clustering technique.