Nalawade, Rahul – rahul.nalawade@utdallas.edu

The University of Texas at Dallas
Machine Learning – Prof. Anjum Chida
Assignment #03

For this assignment submit implementation of ***ONE** of the following 4 problems.

*2. K-means clustering on images

In this problem, you will use K-means clustering for image compression. We have provided you with two images.

- Display the images after data compression using K-means clustering for different values of K (2, 5, 10, 15, 20).
- What are the compression ratios for different values of K?
- Is there a tradeoff between image quality and degree of compression? What would be a good value of K for each of the two images?

We have provided you java template KMeans.java which implements various image input/output operations. You have to implement the function k-means in the template. See the file for more details. **Note** that your program must compile and we should be able to replicate your results. Otherwise no credit will be given.

What to turn in:

1. Your code and datasets
2. A README for your compiling/using your code
3. A report (pdf or doc file) containing answers to the questions posed.

Ans. Images – ***Koala.jpg*** *(780.8kB)*, ***Penguins.jpg*** *(777.8kB)*, ***rahul.JPG*** *(219.1kB)*

Please see the attached *<filename><K>out.png* files after KMeans compression for K = {2, 3, 5, 7, 10, 15, 20}.

Compression Ratio = Original Size / Compressed Size

K	Koala.jpg		Penguins.jpg		rahul.JPG	
	Compressed Size (kB)	Compression Ratio	Compressed Size (kB)	Compression Ratio	Compressed Size (kB)	Compression Ratio
2	45.7	17.09	26.6	29.24	37.3	5.87
3	65.2	11.98	35.3	22.03	58.5	3.75
5	111.4	7.01	62.3	12.48	83.9	2.61
7	150.4	5.19	68.1	11.42	97.2	2.25
10	188.6	4.14	104.6	7.44	128.0	1.71
15	233.0	3.35	143.4	5.42	152.2	1.44
20	259.5	3.01	164.4	4.73	169.4	1.29

Observation:

As we reduce the K value, images are drastically compressed. More the K value, compression is lesser. Hence, more the compression, more the information loss, lower the image quality.

Thus, for K=20 we can see the images almost retain their qualities, and for K=2 we can see the images are highly compressed.

Comparison for rahul.JPG VS rahul<K>out.png

Original Image

K = 2

K = 5

K = 10

K = 15

K = 20

Nalawade, Rahul – rahul.nalawade@utdallas.edu

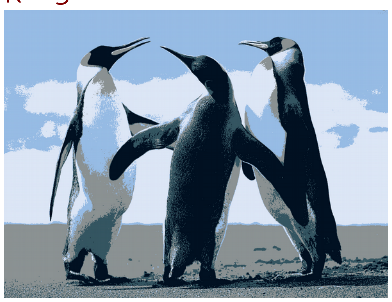
Comparison for Penguins.jpg VS Penguins<K>out.png

Original Image

K = 2

K = 3

K = 5

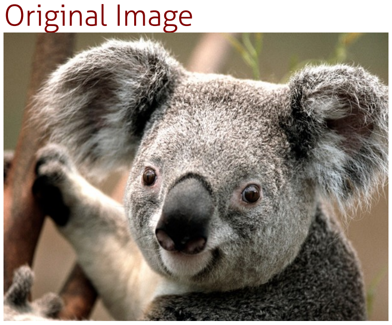


K = 10

K = 15

Comparison for Koala.jpg VS Koala<K>out.png

Original Image



K = 2

K = 3

K = 5

K = 7

K = 10