***Apurva Mithal***

***axm174531***

**Machine Learning Homework 3**

**Attempted Question 2: K-means clustering on images**

**Number of Iterations**: Fixed to 100.

**Question) What are the compression ratios for different values of K?**

**Answer)**

Compression Ratios:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Image** | **K** | **Original Image Size (KB)**  **SizeUncompressed** | **Compressed Image Size (KB)**  **SizeCompressed** | **Compression Ratio = SizeUncompressed : SizeCompressed** | **Compressed Percentage =**  **(SizeCompressed / SizeUncompressed) \* 100**  **(%)** |
| Penguins.jpg | 2 | 760 | 84 | 190:21 | 11.05263 |
| 5 | 760 | 105 | 152:21 | 13.81579 |
| 10 | 760 | 115 | 152:23 | 15.13158 |
| 15 | 760 | 114 | 20:3 | 15.00000 |
| 20 | 760 | 114 | 20:3 | 15.00000 |
| Koala.jpg | 2 | 763 | 128 | 763:128 | 16.77588 |
| 5 | 763 | 173 | 763:173 | 22.67366 |
| 10 | 763 | 161 | 109:23 | 21.10092 |
| 15 | 763 | 155 | 763:155 | 20.31455 |
| 20 | 763 | 152 | 763:152 | 19.92136 |

**Question) 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?**

**Answer)**

The compression method is usually [lossy](https://en.wikipedia.org/wiki/Lossy_compression), meaning that some original image information is lost and cannot be restored, possibly affecting image quality. To achieve high compression ratios, lossy compression methods are required. Lossy Compression provides a tradeoff between image quality and degree of compression, which further allows the compression algorithm to be customized based on the application requirement.

In general, a higher compression ratio results in poor image, but the results are highly dependent on the image. The appropriate level of compression depends on the use to which the image will be put.

A good value of k for the images will be:

|  |  |
| --- | --- |
| **Image** | **k-Value** |
| Penguins.jpg | 15 |
| Koala.jpg | 10 |