Name: Mital Suresh Modha
NetID: msm160530

ASSIGNMENT – 3

Data Compression using K-means clustering

Image compression is carried out using k-means algorithm, using JAVA. The program is fed an image (Koala.jpg or Penguins.jpg), and a compressed image is produced, depending on the K value. The K values that are considered are 2, 5, 10, 15, 20 and the number of iterations are 100.

COMPRESSED IMAGE:

Koala.jpg

1. **K = 2**
   Compressed file size: 130163 Bytes
   Compression ratio: 0.16669804349468706

Name: Mital Suresh Modha
NetID: msm160530

2. **K = 5**
   Compressed file size: 175621 Bytes
   Compression ratio: 0.22491550668454505

Name: Mital Suresh Modha
NetID: msm160530

3. **K = 10**
   Compressed file size: 163505 Bytes
   Compression ratio: 0.20939870471331185

Name: Mital Suresh Modha
NetID: msm160530

4. **K = 15**
   Compressed file size: 156982 Bytes
   Compression ratio: 0.20104478433873654

Name: Mital Suresh Modha
NetID: msm160530

5. **K = 20**
   Compressed file size: 157328 Bytes
   Compression ratio: 0.20148790199159614

Name: Mital Suresh Modha
NetID: msm160530

Penguins.jpg

1. **K = 2**
   Compressed file size: 85013 Bytes
   Compression ratio: 0.109294387627196

Name: Mital Suresh Modha
NetID: msm160530

2. **K = 5**
   Compressed file size: 107571 Bytes
   Compression ratio: 0.13829539683866116

Name: Mital Suresh Modha
NetID: msm160530

3. **K = 10**
   Compressed file size: 117640 Bytes
   Compression ratio: 0.15124030160638183

Name: Mital Suresh Modha
NetID: msm160530

4. **K = 15**
   Compressed file size: 118302 Bytes
   Compression ratio: 0.15209138184833546

Name: Mital Suresh Modha
NetID: msm160530

5. **K = 20**
   Compressed file size: 115324 Bytes
   Compression ratio: 0.14826280637924497

Name: Mital Suresh Modha
NetID: msm160530

Compression ratio:

1. **Koala.jpg**

    a. K=2: Compression ratio: 0.16669804349468706
    b. K=5: Compression ratio: 0.22491550668454505
    c. K=10: Compression ratio: 0.20939870471331185
    d. K=15: Compression ratio: 0.20104478433873654
    e. K=20: Compression ratio: 0.20148790199159614

    Average: 0.20071
    Variance: 0.00045

2. **Penguins.jpg**
    a. K=2: Compression ratio: 0.109294387627196
    b. K=5: Compression ratio: 0.13829539683866116
    c. K=10: Compression ratio: 0.15124030160638183
    d. K=15: Compression ratio: 0.15209138184833546
    e. K=20: Compression ratio: 0.14887604697654386

    Average: 0.13996
    Variance: 0.00032

Average and Variance

Average and variance is calculated over each value of k, running the code 5 times for each image.

Koala.jpg

1. **K=2**
    [0.16668267525239136, 0.16668267525239136, 0.16758171742669029, 0.16668267525239136, 0.16669804349468706]
    Average: 0.16686555733571
    Variance: 1.2825674640881E-7

2. **K=5**
    [0.2260924579070247, 0.2261334398864799, 0.22491038393711316, 0.22492062943197697, 0.22492062943197697]
    Average: 0.22539550811891
    Variance: 3.433294614225E-7

3. **K=10**
    [0.20976626184155087, 0.20995708418338924, 0.2103950790888169, 0.20879422051634733, 0.21059358555180313]
    Average: 0.20990124623638
    Variance: 3.940098443628E-7

Name: Mital Suresh Modha
NetID: msm160530

4. **K=15**
   [0.20043901945491405, 0.2014059380326857, 0.202443294387646, 0.20367275377130262, 0.20372910399305355]
   Average: 0.20233802192792
   Variance: 1.6405382953381E-6

5. **K=20**
   [0.19886633599331993, 0.19817604577687106, 0.20034296794056589, 0.20218587632919288, 0.2001047601849824]
   Average: 0.19993519724499
   Variance: 1.8995326267929E-6

Penguins.jpg

1. **K=2**
   [0.10926996085288011, 0.109294387627196, 0.109294387627196, 0.10926996085288011, 0.10926996085288011]
   Average: 0.10927973156261
   Variance: 1.4320015283498E-10

2. **K=5**
   [0.13606098979860767, 0.14050409148469792, 0.13996027435124417, 0.13661380626996728, 0.13922232864296413]
   Average: 0.1384722981095
   Variance: 3.234640620882E-6

3. **K=10**
   [0.15099860510262458, 0.14916659702893287, 0.1517391220503063, 0.151458856955524, 0.14948543071474027]
   Average: 0.15056972237043
   Variance: 1.0972890876459E-6

4. **K=15**
   [0.14883233590671544, 0.1473037340824211, 0.14949957253144947, 0.14879119607628868, 0.1503930782235307]
   Average: 0.14896398336408
   Variance: 1.0265564101227E-6

5. **K=20**
   [0.14959727962871303, 0.150736338683654, 0.14684991032802586, 0.14930544395662318, 0.14960627896661888]
   Average: 0.14921905031273
   Variance: 1.6430911150464E-6

Name: Mital Suresh Modha
NetID: msm160530

TRADEOFF

As the value of **K** increases, the value of compression decreases. And Higher the compression ratio, lower the quality of the image that is being produced. This is observed from the compressed images that are produced.

Koala.jpg

Since the compressed image is similar for k = 10, k = 15 and k = 20, the optimum trade-off between compression and quality would be best for k = 10.

Penguin.jpg

Since the compressed image is similar for k = 15 and k = 20, the optimum trade-off between compression and quality would be best for k = 15.