Project 4 - Color Image Compression Using Unsupervised Learning (Clustering)

Sam Neyhart
sneyhart@vols.utk.edu
ECE 471 - Professor Qi

April 23, 2017

Abstract

This project serves as an introduction to unsupervised learning methods such as k-means, winner-take-all, Kohonen maps, and means-shift. In order to gain a better understanding of how these algorithms work, the algorithms are implemented in order to cluster the colors of the image. These images are then compared in order to gain insight into how the various clustering algorithms perform.

Introduction

This project performed for ECE 471 is an exercise in unsupervised learning. By experimenting with various unsupervised learning methods, the colors of the *flowers.ppm* image are clustered into various numbers of clusters. To form the clusters, first the k-means algorithm is used then the winner-take-all algorithm. The algorithms are implimented in python using the Python Image Library (Pillow) package to handle image input and output.

Overall the project served its purpose and was an opportunity to learn much about the concepts of unsupervised learning. For this lab the PIL library was very helpful as it greatly simplified the input and output of image data. The algorithm used for k-means was incredible time consuming.

Technical Approach

K-means The *kmeans.py* file impliments the k-means algorithm using basic python and the numpy library. It will perform k-means clustering on the data and then subsequently output the mean values for the requested number of clusters. This implementation is incredible slow however so in order to speed the process of obtaining the various versions of the *flowers.ppm* image required, the Project4.py file was created.

Project4.py This project makes use of the file *Project4.py* which contains instructions that perform k-means clustering on the image file using the PIL Image module which contains the function quantize. The quantize function performs kmeans and then outputs a copy of the *flowers.png* image consisting only of the reduced color set. The PIL image function is incredible fast compared to the k-means function I implimented. After verifying that the two methods produce the same results for k=4, I chose to use the Project4.py file in order to perform the clustering.

Winner-take-all This project makes use of the file *winner.py* which contains instructions which implement winner-takes-all clustering. This algorithm is much faster than the k-means algorithm. The winner-take-all clustering algorithm is initialized with random cluster centers which can result in differing final values.

Comparison Metric In order to compare the methods, several comparison metrics were devised. The file size of the *.png* images are compared, which provides some insight into how the reduced color set allows for more compression. In addition to these to metrics it seems that simple visual comparison is one of the most powerful ways to understand the differences in the images.

Experiments and Results

The original image *flowers.ppm* is displayed above. All experiments are performed on this image. The first expiriment is the kmeans clustering of colors in *flowers.png*. The resulting images can be seen on the following page. To generate these images k-means clustering was performed with [1, 2, 4, 8, 16 , 32, 64, 128, 256] clusters respectively.

The k-means images are seen above. The visual differences on these images are easy to see for the first couple images but by k=16 it becomes harder to find difference between the original image and the processed image. A list of the colors included in the each of these images is found in the apendix.

Discussion

This lab was very interesting to perform. Using python was a good choice because it greatly simplified the k-means part of the lab. The lab certainly increased my understanding of clustering algorithms and gave me greater insight to how and why they are used. The whole concept of unsupervised learning is very interesting and it was fascinating to see it in action here.

Discussion

This lab was very interesting to perform. Using python was a good choice because it greatly simplified the k-means part of the lab. The lab certainly increased my understanding of clustering algorithms and gave me greater insight to how and why they are used. The whole concept of unsupervised learning is very interesting and it was fascinating to see it in action here.

Appendix

Project4.py

kmeans.py

winner.py

kmeans-colors.txt

wta-colors.txt