Tamara Buch

Perry Huang

CS 498-DAF

December 14, 2012

Final Project

Using K-means Clustering for Image Classification

In our project we attempt to classify images depicting a season of the year into an appropriate category (Winter, Spring, Summer and Autumn). The metric we used was colour, as images depicting each season tend to be very distinct in the primary colours used. Similar to in Homework 5, we took images from each season and created 8x8x8 histograms by extracting the RGB data per pixel, however, this time we split each image up into 400 small blocks and created the histograms from each of these. We had 25 pictures from each season and created a total of 10,000 “picture pieces” per season (40,000 total picture pieces). Unwrapping each histogram into 512x1 vectors allowed us to use the kmeans function in Matlab to cluster the picture pieces. We used the kmeans function to cluster each image set individually, as well as the sets concatenated. For the individual sets we used the value of k = 100, and for all the images combined we used k = 200. Although it was suggested to use a larger number of clusters, we found many of the clusters to be empty when using values higher than this. On the following pages are the clustering histograms for each individual season followed by the histogram for the full image set.

An interesting thing to note about the histograms is, as we predicted, each has a significant colour. This is actually a small colour range, as we are truncating the 256-colour spectrum into 8-colours.

Using the histogram for all the images we trained an SVM classifier (libsvm) by supplying it with a vector of the labels (0 = Winter, 1 = Spring, 2 = Summer, 3 = Autumn). So we had a k-by-100 matrix where each row represented an image, and each column represented the number of picture pieces of that image were in that cluster. We trained the classifier with blah blah and blah parameters. Finally, to test the classifier we gave it 5 new images from each season. Our classifier determined blah of blah Winter images were indeed Winter images with a blah accuracy. For Spring, it determined blah of blah images belonged to the Spring label with blah accuracy. For Summer, blah/blah and blah accuracy. Autumn finished with blah/blah and blah. Our Matlab workspace, images, and any code, etc used for this project can be found at <https://github.com/perryh/cs498-daf>.





