Name(s)	Project Number
Preetha S. Krishnamurthy	**S1414**

Project Title

Compression Based Data Mining for Assessing Similarity in Object Textures in Multiple Images

Abstract

Objectives/Goals
The goal of this project is to analyze if a given image is similar to any of the images in a large dataset, and if so, how similar. While comparing two images, we seek to also account for distinctions in texture in the objects captured in the images. The approach must be simple (low complexity) and should not require extensive training (as in machine learning).

Methods/Materials
The code was written and executed using Matlab and its tool suite, on a Lenovo laptop computer. The images considered were from datasets of (i) toolmarks on metal obtained from the Internet and (ii) a dataset that we created by cutting up two types of wood into blocks and taking pictures.

Results
The results from this project demonstrate that the CK algorithm can quickly analyze a large image dataset and pick out those images that are most similar (best matches) to a target image with an accuracy of 100%. It is resistant to distortions in the image; specifically it can handle cropping, erasures, and noise/speckles and yields an accuracy of between 80 to 100 % with such distortions. It significantly outperforms an algorithm that uses the Euclidean distance measure as a similarity assessment to compare two images.

Conclusions/Discussion
In this project, we seek to tackle the hard problem of examining if an image is similar to any of the images in a large data set via an automated technique. The motivating application is criminal forensics. We perform a clever application of what is called the Campana-Keogh (CK) algorithm to quickly, and yet accurately obtain the aforementioned similarity assessments. The algorithm exploits a key property of MPEG encoding. Specifically, if the images to be compared are concatenated and encoded as a video, similar images will lead to smaller encoded sizes. We conduct extensive experiments to evaluate our approach with two datasets (a metal dataset that we obtained from the Internet, and a wood dataset that we developed). We show that the algorithm provides accuracies of 80 % to a 100% with both datasets even when the images are subject to various forms of distortion.

Summary Statement

My project seeks to compare a given image with a large set of images in a dataset with high accuracy but low complexity, using compression based data mining.

Help Received

I recieved guidance from Professor Eamonn Keogh at UCR, who taught me the concepts behind the CK algorithm as well as video compression. He also provided guidance on how to develop the code. The supporting algorithms were built by me.