Kei Nakagawa
Capstone Abstract
CLPS1520

Multiclass Classification of Human Actions with Shape vs Motion Detection

Our main goal was to compare the accuracy of detecting specific human actions based on shape and motion flow field from videos. Classification of video feeds' actions and information is a growing issue as the ease of capturing video increases - sites like Youtube need automated tagging services for content, and surveillance applications would be ameliorated by the ability to automate action detection.

By shape, we mean the outlining contour of human bodies taking a specific action, and motion flow fields track the trajectory of movement of a specific focus point. Broadly speaking, the separation of visual information into shape and motion cues is mirrored by how information is encoded in video files and our brain's ventral "form" pathway and dorsal "motion" pathway.

One of the pieces of literature we relied on heavily for both inspiration and support code was "Efficient feature extraction, encoding and classification for action recognition (Kantorov and Laptev, 2014)," which was able to provide the "cvpr2014" package so we could efficiently extract video features.

To extract spatial features, we applied spatial gabor filters to the frames of the videos.

We conducted this experiment by gathering a set of .gif files showing handshakes, hugs, drinking, and kissing actions, applying both spatial and spatiotemporal feature extraction methods on those videos, then trying classification of those videos once each via SVM, KNN, and KMeans clustering algorithms.

Some of our key findings were that classification using cvpr extracted features resulted in much higher accuracy rates across classification methods than gabor filters, and that certain motions are better classified by motion-based features than form-based features, while others exhibit the opposite behavior; For example, the kiss and hug actions were better detected by shape.

On the technical side, this project involved writing a classification pipeline in MATLAB, writing scripts to automate the video conversion / feature extraction process in python, gaining familiarity with using the Brown CCV systems, and using Apache Spark library for the machine learning components.