Clustering Facial Emotion Using Unsupervised Learning

Dylan Grosz, dgrosz@stanford.edu
Sam Kwong, samkwong@stanford.edu
Winton Yee, winton@stanford.edu

Abstract

The goal of our project was to create an unsupervised learning system that could accurately detect emotion in human facial expressions in a generalized fashion. To achieve this goal, we extracted features from face images using computer vision techniques and then clustered them. Possible applications of this project include mass labeling, general dataset exploration, and aiding supervised learning.

Data

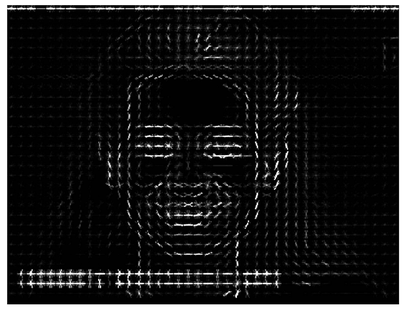
We used the Extended Cohn-Kanade dataset, an image set containing around 100 different subjects in various posed facial expressions representing six emotions: happiness, sadness, surprise, fear, disgust, and anger.

Model

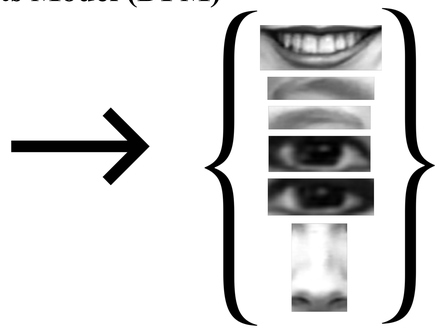
- We modeled our problem mainly as a clustering task. In the learning stage of our task, we decided to use the k-means algorithm to group together images given the algorithm's efficiency and ease of adjusting the number of clusters.
- Based on the features we decided to extract, k-means would cluster together images that have similar features. We then naively label each cluster as the mode of the emotions within the cluster.

Features

Histogram of Gradients (HoG)



Deformable Parts Model (DPM)



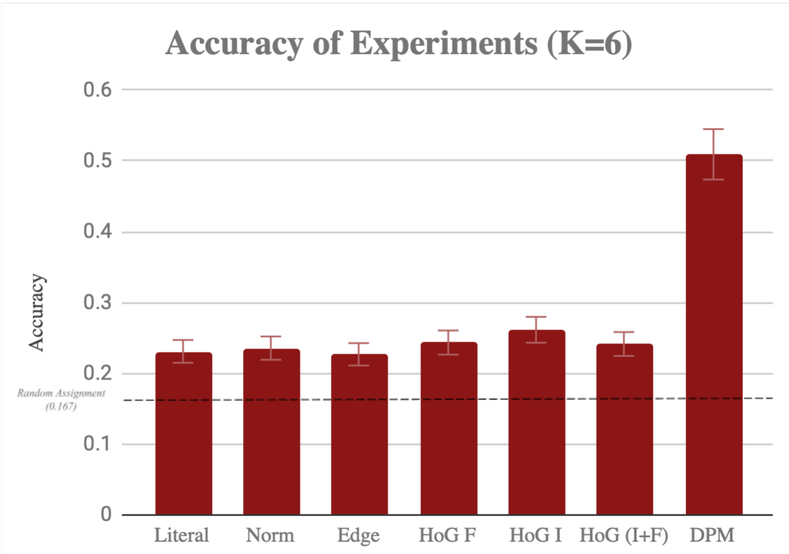
Experiments

We ran experiments corresponding to the use of a particular combination of features.

- Using only the literal black and white pixels
- Normalizing the literal black and white pixels to adjust for lighting
- Applying edge detection to the pixels
- Applying Histograms of Oriented Gradients (HoG) to the pixels and using the resulting HoG descriptor
- Applying HoG to the pixels and using the resulting visualization of the HoG image (shown above)
- Applying HoG to the pixels and using both HoG descriptor and visualization
- Using Detectable Parts Model (DPM) to extract and focus on just the mouth, eyebrows, eyes and nose as the features (shown above)

Results & Analysis

- The baseline for our experiment would just be randomly guessing one of the six possible emotions, which is 16.7%. We want to improve on this percentage.
- The oracle for our assignment would be human labeling of data, which should be near 100% accuracy.



Future

- We will attempt to combine these different feature extraction methods together with different weightings.
- Cater DPM feature extraction to better detect emotions (e.g. combine eye & eyebrow features)
- We will assign confidence of a label, which should be proportional to how far an image feature vector is from its assigned labeled centroid.

References

- *Histograms of Oriented Gradients for Human Detection* by Navneet Dalal and Bill Triggs
- *Object Detection with Discriminatively Trained Part Based Model*s by Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester and Deva Ramanan
- *A Computational Approach to Edge Detection* by John Canny