

Automatic Action Unit Recognition Using Computer Vision

Project Proposal

Derek Schultz

October 31, 2013

1 Introduction

Facial expressions are one of the primary methods humans use to convey and perceive emotions. In order for machines to be able to properly detect the emotional state of human users, facial expression recognition is an important capability. Using Paul Ekman's action unit system^[1], emotions can be classified based on the actions of facial muscle groups. Automatic detection of these action units by means of computer vision is the goal of this project.

2 Methods

2.1 Face Detection

In order to access the computer's webcam to obtain a video feed and for an abstracted interface to pixel data, OpenCV^[2] will be used in this project. OpenCV's provided cascade classifiers will be used to locate face in the image and provide basic manipulation functionality to crop and convert the image to grayscale, etc.

2.2 Facial Landmark Detection

A heuristic approach will be used to divide the face into a set of regions of interest. The eyes and mouth can be located first and used as landmarks to calculate the locations of other regions of interest.

Within each region, the specific landmark points will be identified using feature patch templates. The features will be based upon the gray level intensities and properties of Gabor filters^[3]. The Gabor filters help to eliminate noise from uncertainties such as color and lighting conditions^[4]. Classification will be done by a boosting algorithm such as AdaBoost. The training data will come from the Cohn-Kanade facial expression database^[5]. The labeled points and its immediate surrounding points will be used as positive samples, and negative samples will be chosen randomly from other points in the region of interest. During classification, all pixels in the region of interest are checked, and the pixel with the highest response is determined to be the landmark.

2.3 Facial Landmark Tracking

The initial facial landmark detection will not be performant enough to sustain in a real time video feed^[4]. The positions of each landmark can instead be tracked using Particle Filtering with Factorized Likelihoods (PFFL). From frame to frame, a color difference is calculated for each pixel. PFFL creates a particle based representation of the probability of a state, given observations of past states. The PFFL will be trained with the MMI-Facial Expression Database^[6]. Color differences from each frame of video can be compared to the PFFL to determine the movement of the facial landmarks.

2.4 Action Unit Predictions

Once facial landmarks are established, there exists a set of data that can be used to recognize the action units, which is the core purpose of the program. There exist perhaps twenty or so facial landmarks, discovered in parts 2.2 and 2.3. Because the data is coming from a video feed, there are $20n$ data points, where n is the number of frames captured. For every frame of video, features are created based on the difference in intensity of each pixel compared to the intensity of the pixel in the first frame. And for each pixel in each frame, the L_2 norm for all pairs of points and all pairs of points compared to the first frame are also added. Put more simply,

$$f_1(p_i) = p_{i,y,n} - p_{i,y,1} \quad (1)$$

$$f_2(p_i) = p_{i,x,n} - p_{i,x,1} \quad (2)$$

$$f_3(p_i, p_j) = \| p_i - p_j \| \quad (3)$$

$$f_4(p_i, p_j) = f_3(p_i, p_j) - \| p_i, 1 - p_j, 1 \| \quad (4)$$

From these features, a boosting algorithm (for example, AdaBoost) selects the best performing ones. A support vector machine (or possibly another machine learning algorithm) is run for each possible action unit. The MMI-Facial Expression Database^[6] will be used to train these SVMs. This data set provides video sequences tagged with action units.

With this classification, a list will be produced of the active action units, hopefully accomplishing the intended goal.

3 References

- [1] Ekman, Paul, "FACS". <http://www.paulekman.com/facs/>
- [2] OpenCV. <http://opencv.org/>
- [3] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, T.J. Sejnowski, "Classifying Facial Actions", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 21, No. 10, pp. 974-989, 1999.
- [4] M.Pantic, M.F. Valstar, "Fully Automatic Facial Action Unit Detection and Temporal Analysis", *In Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, 2006
- [5] P. Ekman, W.V. Friesen and J.C. Hager, "The Facial Action Coding System: A Technique for the Measurement of Facial Movement", *San Francisco: Consulting Psychologist*, 2002
- [6] M.Pantic, M.F. Valstar, R. Rademaker and L. Maat, "Web-based database for facial expression analysis", *In ICME'05*, pp. 317-321, 2005