Automatic song suggestion using facial affect on mobile devices

Affective computing mini-project proposal

C. T. Hewitt (*cth40*), Trinity Hall

Overview

I intend to develop a mobile application which uses the front facing camera to detect the user's affect from their facial expression. Based on the classification made from the video feed (or individual frames within) a list of songs will be recommended to match the current users mood.

Feature Extraction

A number of options are available for feature extraction; OpenFace [1] can extract a number of facial landmarks from the input image data which can then be used for classification. Alternatively Affectiva's SDK [2] can extract a number of action unit based features.

Classification

Classification will be carried out using an existing machine learning library. Weka [3] can be used on desktop to help determine the optimal feature set and technique, but is not generally suitable for mobile devices. Therefore, an alternative library will be used for the app itself, such as libsvm [4] or iOS's built in machine learning library, CoreML [5]. Affectiva's SDK also includes some classification capabilities which could be used.

Training

Training will be performed using a large, openly available dataset which provides appropriate annotations (ideally arousal and valence). There are a number of options available: [6], [7], [8]. Alternatively a pre-trained model may be employed if a suitable option is available.

Recommendation

Song recommendations will be made using an external system, most likely the Spotify recommendations API [9].

Evaluation

The app will be evaluated via a user study. Participants will be asked to use the app while simulating a number of given facial expressions, then rate how well the app performs. Direct affect output (e.g. name of emotion or numeric ratings for valence/arousal etc.) will be given as well as song recommendations so that the affective aspect of the app and song recommendations can be evaluated independently.

Required Hardware

- Laptop for development. I plan to use my own 2013 MacBook Pro, 2.8GHz i7, 16GB RAM.
- Mobile phone with front facing camera. I plan to use my own iPhone 6s.

External Software

- Spotify recommendations API, takes as input various parameters such as *danceability*, *energy* and *loudness* [9].
- Libraries for feature extraction and classification of video data. This may include, but not be limited to: OpenFace [1], Affectiva SDK [2], Weka [3], LibSVM [4], iOS/MacOS SDK [5].

References

[1] OpenFace: an open source facial behavior analysis toolkit, Baltrušaitis et al. 2016. `http://www.cl.cam.ac.uk/research/rainbow/projects/openface/`

[2] Emotion SDK, Affectiva. `https://www.affectiva.com/product/emotion-sdk/`

[3] WEKA, E. Frank, M. Hall, and I. Witten, University of Waikato. 2016. `https://www.cs.waikato.ac.nz/ml/weka/`

[4] LibSVM, C. Chang and C. Lin. `https://www.csie.ntu.edu.tw/~cjlin/libsvm/`

[5] Core ML, Apple Inc. `https://developer.apple.com/documentation/coreml`

[6] Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild, S. Li and W. Deng. 2017. `http://www.whdeng.cn/RAF/model1.html`

[7] AR Face Database, A. Martinez, Ohio State University. `http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html`

[8] DEAP: A Database for Emotion Analysis using Physiological Signals, Koelstra et al. 2012. `http://www.eecs.qmul.ac.uk/mmv/datasets/deap/`

[9] Spotify API, Spotify Inc. `https://developer.spotify.com/web-api/get-recommendations/`