# Intensity Measurement of Spontaneous Facial Actions: Evaluation of Different Image Representations

Nazanin Zaker[1], Mohammad H. Mahoor[1], Whitney I. Mattson[2], Daniel S. Messinger[2] and Jeffrey F. Cohn[3]

[1]Department of Electrical and Computer Engineering, University of Denver, Denver, CO 80208
[2]Department of Psychology, University of Miami, Coral Gables, FL 331462
[3]Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260

*Abstract*— **Intensity measurements of infant facial expressions are central to understand emotion-mediated interactions and emotional development. We evaluate alternative image representations for automatic measurement of the intensity of spontaneous facial action units related to infant emotion expression. Twelve infants were video-recorded during face-to-face interactions with their mothers. Facial features were tracked using active appearance models (AAMs) and registered to a canonical view. Three representations were compared: shape and grey scale texture, Histogram of Oriented Gradients (HOG), and Local Binary Patten Histograms (LBPH). To reduce the high dimensionality of the appearance features (grey scale texture, HOG, and LBPH), a non-linear algorithm was used (Laplacian Eigenmaps). For each representation, support vector machine classifiers were used to learn six gradations of AU intensity (0 to maximal). The target AUs were those central to positive and negative infant emotion. Shape plus grey scale texture performed best for AUs that involve non-rigid deformations of permanent facial features (e.g., AU 12 and AU 20). These findings suggest that AU intensity detection may be maximized by choosing feature representations best suited for specific AU.**

## I. INTRODUCTION

Face-to-face communication is a salient developmental issue during the first year of life. Infants and parents experience emotional engagement with one another by learning to coordinate variation in emotion intensity. To understand emerging patterns of synchrony and dys-synchrony, precise measurement of emotion expression is needed. Previous efforts have relied on manual coding of discrete infant and parent behaviors or ordinal scaling of predefined affective engagement states. Both methods rely on labor-intensive manual categorization of ongoing behavior streams.

The most comprehensive measurement approach to quantify facial expressions is the Facial Action Coding System (FACS) [1]. FACS enables powerful description of nearly all visually detectable changes in facial movement in terms of 44 Action Units (AUs). To represent intensity, AU may vary on a 6-point ordinal scale from absent to maximal. FACS requires considerable training to master (100+ hours), is labor intensive (one hour or more to code a minute of video), and difficult to standardize over time and across research groups. Automated FACS coding is an emerging alternative, and considerable progress has been made in automated coding of posed and to some extent unposed facial expression in adults. Ours is the only effort to attempt automatic AU detection in infants. Infants are especially challenging because of their unique face shape and appearance and wider range of facial actions relative to adults [8].

Almost all approaches to automatic AU detection seek to detect onset and offset of AU rather than changes in AU intensity. In a pilot study, we demonstrated the feasibility of automated detection of AU and AU intensity in two infants [9]. In the current work, we concentrate on a larger number of infants, explore alternative appearance representations, and identify the most informative representations for the goal of detecting graded changes in AU intensity.

We explore three feature representations. One is shape and grey scale texture from Active Appearance Model (AAM) [3]. Shape is represented by the position of non-rigid facial landmarks (e.g., mouth); texture (aka appearance) is represented by grey scale values of pixels within the face region bounded by face landmarks. Shape features may be best for AU related to deformations of permanent facial features (e.g., lips or eyes). Appearance representations may perform best for less well-defined features (e.g., cheeks) or when registration error is moderate. The latter refers to error in registering face images to a canonical view, which occurs from mild to moderate changes in face pose and self-occlusion. In addition to grey scale texture, two biologically inspired appearance representations are LBPH [4] and HOG [5]. We compare the representational power of each in measuring the intensity of three facial action units: cheek raiser (AU6), lip corner puller (AU12), and lip stretcher (AU20) in infants' facial expression during infant-parent interaction. The AU chosen include deformations of both well-defined features (e.g., lips) and ones that can be less well defined (e.g., cheeks) and are central to positive and negative emotion in infants.

## II. PROPOSED METHOD

Our proposed framework consists of four phases: preprocessing, feature extraction, dimensionality reduction and classification. In preprocessing, 66 facial landmark points are extracted using AAM [3]. These landmark points are used in registering face images via Procrustes analysis. Next, three different features are extracted from each image. These features are a) grey scale texture (aka appearance) combined with the similarity normalized shape features, b) LBPH features, and c) HOG features. The latter are briefly explained in the following subsections.

After extracting facial features by each method, manifold learning (Laplacian Eigenmap) decreases the dimensionality of each feature vector. This method is based on the assumption that the feature data lie on a low dimensional manifold embedded in a high dimensional space. Finally, Support Vector Machine classifiers (SVM) are used to classify images into 6 ordinal intensities for each AU.

### A. Active Appearance Model

An AAM matches a statistical model of object shape and texture to a new image [3]. AAM are learned from training data, which consist of manually labeled images (about 3% of images). The model then automatically extracts facial landmark points in the image sequence. The shape features extracted by AAM are used to register facial images to a canonical view and with texture represent face variation due to different facial expressions.

### B. Local Binary Pattern Histogram

LBPH is a powerful texture descriptor that has been successful in face recognition [4]. In LBPH, the image is divided into blocks of desirable size. Each pixel, say $p$, in a block is compared to its neighbors in radius $r$. If $p$'s value is greater than the neighbor's value, "1" is coded, otherwise "0". As a result an $n$-digit binary number is coded as a label of pixel, where $n$ is the number of neighbors. Then, the histogram of the frequency of each label is computed over all blocks. In our implementation, the registered images are segmented into 35 blocks defined by the position of shape landmark points. A mapping then decreased the size of each histogram from 256 bins to 59 bins.

### C. Histogram of Oriented Gradients

HOG descriptors characterize local object appearance and shape from the distribution of image gradients or edge directions [5]. To implement HOG descriptors, the image is first segmented into 30 small regions based on the landmark points' positions. Then, for each region, the histogram of gradient directions (edge orientations) is computed. The concatenation of these histograms results in HOG descriptors, which are mapped onto a 64-bin histogram. The full image is represented by a feature vector consisting of 1920 features.

### D. Manifold learning and SVM classifier

The extracted features are given as inputs to a manifold learning module to decrease large dimensionality of the features. This step results in a feature vector of 29 dimensions for each image. These outputs are then input to the SVM classifiers [7]. For each AU, SVM classifiers are trained to classify images based on the 6-point ordinal intensity scale defined by FACS (0-5 metric) [1]. An expert FACS coder manually coded the intensity of AU6, AU12, and AU20 in all video frames for use in training and testing the system. A leave-one-subject-out technique was used to cross-validate the SVM classifiers. For training, we randomly selected 2% of all frames from 11 subjects and tested the learned model on all images from the omitted subject. This method was repeated for all the subjects.

## III. Experimental Results and Conclusions

Participants were ethnically diverse (16.7% African American, 16.7% Asian American, 33.3% Hispanic American, and 33.3% European American) six-month-old infants (M = 6.20, SD = 0.43, 66.7% male). Video was gathered from an observational procedure (the Face-to-Face/Still-Face) that emulates naturalistic interaction and a common infant stressor (maternal unresponsiveness).

To compare the power of the aforementioned feature representations (i.e. shape + texture, LBPH and HOG) for AU intensity measurement) we used intra-class Correlation Coefficient (ICC) [6] and F1 score to compare predicted AU intensities and those provided by manual FACS coding.

Table 1. Results for AU 6, AU 12, and AU 20 intensity for each representation.

| | **Shape + texture** | **HOG** | **LBPH** | **#Events** | **#Frames** |
|---|---|---|---|---|---|
| **AU 6** | | | | | |
| F1 | 0.81 | 0.83 | 0.80 | 69495 | 96974 |
| ICC | 0.59 | 0.60 | 0.52 | 69495 | 96974 |
| **AU 12** | | | | | |
| F1 | 0.68 | 0.47 | 0.24 | 26147 | 96974 |
| ICC | 0.63 | 0.37 | 0.13 | 26147 | 96974 |
| **AU 20** | | | | | |
| F1 | 0.69 | 0.65 | 0.63 | 43750 | 96974 |
| ICC | 0.57 | 0.50 | 0.47 | 43750 | 96974 |

For each set of features, moderate to high AU detection was found (Table 1). Specific features appeared best suited for specific AU. Shape and grey scale texture proved best for AU 12 and AU 20, for which deformation of permanent facial features (lips) is prominent. For AU 6, similar results were found for that representation and for HOG. LPBH was the least effective for all three AUs. In conclusion, AU intensity detection may be best achieved by choosing representations specific to each AU.

## References


[1] P. Ekman, W. V. Friesen, and J. C. Hager, "Facial Action Coding System", The Manual on CD ROM. 2002.

[2] J. F. Cohn, & T. Kanade, "Automated facial image analysis for measurement of emotion expression", in the handbook of emotion elicitation and assessment. Oxford University Press Series in Affective Science, 2007.

[3] I. Matthews and S. Baker. "Active appearance models revisited", International Journal of Computer Vision, 60(2):135–164, Nov. 2004.

[4] T. Ahonen et al., "Face Description with Local Binary Patterns: Application to Face Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, V. 28 Issue 12, 2006.

[5] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection", CVPR, pp. 886-893, 2005.

[6] J. M. Bland, and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement", Lancet, i, pp. 307–310. 1986.

[7] M. H. Mahoor, S. Cadavid, D. S. Messinger, and J. F. Cohn, "A Framework for Automated Measurement of the Intensity of Non-Posed Facial Action Units", CVPR4HB workshop, Miami Beach, June 2009.

[8] H. Oster, "Baby FACS, Facial action coding system for infants and young children", New York, NY: New York University., 2004.

[9] D. S. Messinger, M. H. Mahoor, S. Chow, J. F. Cohn, "Automated Measurement of Facial Expression in Infant–Mother Interaction: A Pilot Study", Infancy, Vol. 14, Iss. 3, 2009.