Discrimination of Moderate and Acute Drowsiness Based on Spontaneous Facial Expressions

Esra Vural*, Marian Bartlett*, Gwen Littlewort*, Mujdat Cetin [†], Aytul Ercil [†], Javier Movellan*

*Institute of Neural Computation
University of California San Diego, San Diego, USA
Email: vesra@ucsd.edu, marni@salk.edu, gwen@mplab.ucsd.edu, movellan@mplab.ucsd.edu

[†]Faculty of Engineering and Natural Science
Sabanci University, Istanbul, Turkey
Email: mcetin@sabanciuniv.edu, aytulercil@sabanciuniv.edu

Abstract

It is important for drowsiness detection systems to identify different levels of drowsiness and respond appropriately at each level. This study explores how to discriminate moderate from acute drowsiness by applying computer vision techniques to the human face. In our previous study, spontaneous facial expressions measured through computer vision techniques were used as an indicator to discriminate alert from acutely drowsy episodes. In this study we are exploring which facial muscle movements are predictive of moderate and acute drowsiness. The effect of temporal dynamics of action units on prediction performances is explored by capturing temporal dynamics using an overcomplete representation of temporal Gabor Filters. In the final system we perform feature selection to build a classifier that can discriminate moderate drowsy from acute drowsy episodes. The system achieves a classification rate of .96 A' in discriminating moderately drowsy versus acutely drowsy episodes. Moreover the study reveals new information in facial behavior occurring during different stages of drowsiness.

Keywords-**Moderate versus Acute Drowsiness Detection ;Facial Expression Recognition; Temporal Dynamics**

I. Introduction

The computer vision field has advanced to the point that we are now able to begin to apply automatic facial expression recognition systems to explore human facial behavior in the state of drowsiness. Most of the published research on computer vision approaches to detection of fatigue has focused on the analysis of blinks, yawns, and head movements. However the effect of drowsiness on other facial expressions have not been studied thoroughly. Gu & Ji presented one of the first fatigue studies that incorporates certain facial expressions other than blinks. Their study feeds action unit information as an input to a dynamic Bayesian network. The network was trained on subjects posing a state of fatigue [1]. In our work we mine datasets of real drowsiness to learn signals of fatigue. In our previous study we focused on detecting crash episodes versus alert episodes [2][3]. Yet a safety system would benefit from detecting finer levels of drowsiness then alert vs crash. Here we seperate moderate from acute drowsiness. The objective of this study is to discover which action units predict moderate and acute drowsiness. In this study, facial motion was analyzed automatically from video by using a fully automated facial expression analysis system based on Facial Action Coding System (FACS). We employ an overcomplete representation of Gabor filters to discriminate moderate from acute drowsiness.

II. Methods

A. Driving Task

Subjects drove a virtual car simulator on a Windows machine using a steering wheel [1] and an open

[1]ThrustMaster steering wheel

source multi-platform video game [2] . The simulator displayed the driver's view of a car through a computer terminal. The windows version of the video game was maintained such that at random times, a wind effect was applied that dragged the car to the right or left, forcing the subject to correct the position of the car. This type of manipulation had been used in monitoring the dynamics of alertness in a compensatory tracking task [4]. Subjects were asked to drive until falling asleep for a maximum of 4 hours. Driving speed was held constant. Video of the subjects' face was recorded using a DV camera for the entire session. Subjects drove in dim light from one 60 watt diffuse desk lamp. During this time subjects fell asleep multiple times thus crashing their vehicles. Episodes in which the car left the road (crash) were recorded.

Subject data were partitioned into 2 groups labeled as "moderately drowsy" (MD) and "acutely drowsy" (AD). The one minute preceding a sleep episode or a crash was identified as an acutely drowsy state. Moderately drowsy episodes were selected from the first 10 minutes of the driving task. The average time to initial crash had a mean of 25 minutes over subjects. [3] 5 one minute moderately drowsy episodes that are farthest away from a crash point were selected for each subject. Subjects had a mean of 46 acutely drowsy episodes ranging from 3 to 244.

B. Facial Action Coding

The facial action coding system (FACS) [5] is one of the most widely used methods for coding facial expressions in the behavioral sciences. The system describes facial expressions in terms of 46 component movements, which roughly correspond to the individual facial muscle movements. An example is shown in Figure 1. FACS provides an objective and comprehensive way to analyze expressions into elementary components, analogous to decomposition of speech into phonemes. Researchers have been developing methods for fully automating the facial action coding system [6][7]. In this paper we apply a computer vision system trained to automatically detect FACS to data mine facial behavior under driver fatigue.

C. The Computer Expression Recognition Toolbox (CERT)

CERT, developed by researchers at Machine Perception Laboratory UCSD [6], is a user independent fully automatic system for real time recognition of facial

---

[2]Torcs

[3]In our previous study average time to first crash after an alert episode was 60 minutes
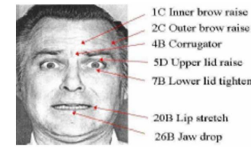

Figure 1. Example facial action decomposition from the Facial Action Coding System [5].

actions from the Facial Action Coding System (FACS). This study uses the output of CERT as an intermediate representation to study fatigue and drowsiness. The system automatically detects frontal faces in the video stream and codes each frame with respect to 20 Action units.

III. Results

11 subjects who were able to fall asleep and recorded under dim light conditions were selected for the analysis. Our initial analysis focused on the prediction power of individual action units in discriminating moderate versus acute drowsiness.

A. Prediction Power of Individual Action Units in Discriminating Moderate versus Acute Drowsiness

Here our goal is to explore the prediction power of individual action units in discriminating moderate versus acute drowsy episodes. Each one minute MD and AD episode is partitioned into 6 non-overlapping 10 second patches. The 10 second patches that contain face occlusions or false alarms in face detection occurring in more than 30 video frames (1 second) were eliminated. Subjects had a mean of 27 MD patches and 177 AD patches. In our first analysis we focus on discriminating the AD and MD patches by employing raw action unit output.

Raw Action Unit Output : Averages of raw action unit outputs were computed over 10 second patches of individual CERT action unit outputs. The mean intensity of each of the 20 AU's comprised the input to an MLR, which was trained to predict moderately versus acutely drowsy. We tested generalization to novel subjects with leave-one-out cross validation training procedure. Performance was first tested for each AU individually. The performance measure was Area Under the ROC curve (A'). Individual action unit discriminability measure is estimated by averaging over all test subject A's. Using this method we can highlight the action units that are informative for a person independent system. This analysis was repeated

for all the action units. Table I displays the individual action unit mean A' estimates for most informative 5 action units. The 5 most informative units were (1) Eye Closure (AU45) with an A' of 0.83 (2) Lip Puckerer (AU18) with an A' of 0.82 (3) Head Roll with an A' of 0.77 (4) Lid Tightener (AU7) with an ROC of 0.71 (5) Nose Wrinkle (AU9) with an A' of 0.69. Previous approaches to drowsiness detection primarily associate drowsiness with blink rate, eye closure, yawning and head nodding (downward movement). This study shows that in addition to eye closure, head roll (sideways movement), Lid Tightener, Lip Pucker and Nose Wrinkle are strong predictors of moderate versus acute drowsiness.

Action Unit	A'	Standard Error
AU 45	0.8346	0.0587
AU 18	0.8247	0.0367
Head Roll	0.7761	0.0723
AU 7	0.7175	0.0884
AU 9	0.6951	0.0702

Table I. A' performance results for the output of the raw action unit outputs over individual action units.

Temporal Dynamics of Action Units : Averaging the AU outputs over 10 second segments may lose important information about dynamics. Consider for example, the data displayed in Figure 2. The figure displays the output of AU 45 (eye closure) signal for an AD and MD patch. In the first signal the subject closes his eyes towards the end whereas in the second case the subject is constantly blinking. The averages of the eye closure signal for these two patches are equal. Thus the raw action unit output approach (mean filtering) would not be able to differentiate these two episodes. However the temporal analysis of action units can bring additional information that could help to discriminate these two patches.
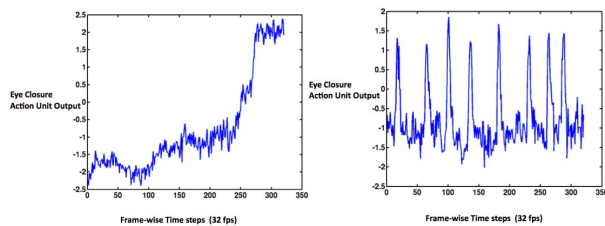

Figure 2. This figure displays a case where temporal dynamics plays an important role in discriminating two cases.

In order to capture the temporal dynamics, we pass the AU outputs through a bank of temporal Gabor Filters. Gabor filters are sine or cos gratings modulated by a Gaussian. A set of Gabor energy (Magnitude Gabor), Gabor cosine carrier (Real Gabor) and Gabor

sine (Imaginary Gabor) filters were convolved with 10 second patches of CERT action unit outputs. Here the real and imaginary Gabor filters are linear filters whereas the magnitude Gabor filter is a nonlinear filter, outputting the energy over the temporal signal consisting of the root sum square of the sine and cosine filters. The bank of filters consisted of 306 frequency and bandwidth combinations with 18 frequencies and 17 bandwidths. The Gabor filter frequencies used for the analysis have the following values : 8.0, 7.0, 6.0, 5.0, 4.0, 3.0, 2.25, 1.6875, 1.2656, 0.9492, 0.7119, 0.5339, 0.4005, 0.3003, 0.2253, 0.1689, 0.01 0. The bandwidths have the same values excluding zero frequency. For each 10 second AD or MD patch 918 filter outputs were obtained: (306 filters, 18 frequencyx 17 bandwidth) for each of the Magnitude, Real and Imaginary Gabor filters. Mean filter output over each 10 second clip comprise the input to an MLR classifier for each action unit. There were 918 features for each 10 second AU signal. Next we performed feature selection to find out relevant features for each action unit.

Feature Selection : The goal for this analysis is to select relevant features from the 918 Gabor filter possibilities for an individual action unit. An iterative feature selection scheme was followed: First the feature with the best performance was selected and then the next feature that achieves the best performance combined with the previously chosen feature is selected and the iteration continues in this fashion. Performance of a feature set was estimated by generalization to novel subjects using cross validation. The average of novel test subject A's gave the discriminability power performance for this feature set. We tried 1 to 10 features. The 5 most dicriminant action units found in the raw action unit analysis were considered for the evaluation of the performance in the temporal dynamics approach.

The A' performance obtained with an L2 regularized MLR model using different number of features are displayed for eye closure action unit (AU45) in Figure 3. As the performance saturates with 10 features we stopped after picking 10 features. The highest discriminability for eye closure was obtained with regularization constant 0 and using 10 features. Notice that the raw AU output was able to obtain an average A' of 0.83 for eye closure (AU45) whereas we obtain 0.9, better performance with temporal Gabors.

Figure 4 displays the performances of 5 most predictive action units with the raw action unit output (blue) versus the best Gabor model (red) for that action unit.
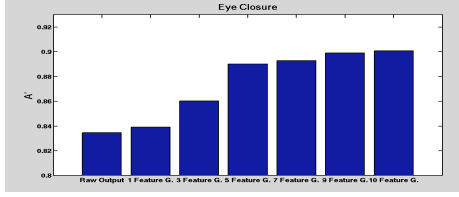
Figure 3. A' performance for Action Unit 45 (eye closure) with Raw Output (1 feature) and Gabor Models with features from 1 to 10.
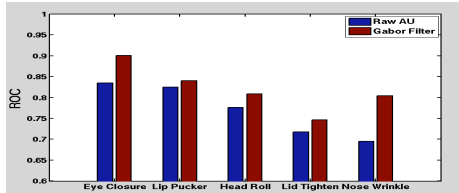

Figure 4. Bar graph displaying the performances for 5 best performing action units with the raw action unit output and the best model of Gabor Filter outputs.

IV. Combining Multiple Action Units

The features of best performing fi ve action units in the single AU Gabor output models were combined to build a person independent drowsiness detector. Except Head Roll achieving the best performance with 8 features other four action units had a set of 10 best features resulting in a total of 48 features. An iterative feature selection procedure was performed. Up to 10 features with regularization constants were explored with an L2 regularized MLR model. MLR model was trained by leaving one subject out at a time. The results for the selected best feature sets from a set of 48 possible features was displayed in Figure 5. The highest discriminability A' performance of .96 was obtained for the combined action units with temporal Gabors with 10 features and regularization constant 0.01. Note that the highest A' performance of .96 cuts the error in half when compared with the highest A' performance of .90 for eye closure action unit (AU 45). Hence combining other action units helped to build a more accurate person independent drowsiness detector.
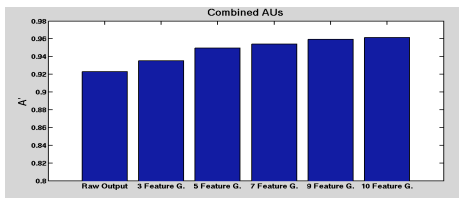

Figure 5. A' performance for the combined action units with Raw Output (1 feature) and best Gabor features from 3 to 10 .

V. Conclusion

In this study we found that temporal analysis with an overcomplete Gabor representation improves prediction performance for all action units. Here the markers for different levels of drowsiness change relative to our previous study. In our previous study comparing alert states to acute drowsy, we found that the Nose Wrinkle (AU9), Eye Closure (AU 45), Eye Brow Raise (AU2), Chin Raise (AU17), Yawn (AU26), Head Roll (Sideways Movement) were some of the most discriminative action units (Vural et al. 2007). In the present study comparing moderately drowsy to acutely drowsy, Eye Closure, Nose Wrinkle and Head Roll also differentiate. However there were also some differences : Lip Pucker (AU18), Lid Tightener (AU7) could differentiate acute versus moderate drowsiness. Finally this work shows that automated expression measurement can not only discriminate widely different levels of drowsiness (alert from acutely drowsy), but can also make more the operationally relevant distinction between moderately and acutely drowsy. Temporal dynamics of facial action carries crucial information for this distinction.

References

[1] H. Gu and Q. Ji, "An automated face reader for fatigue detection." in *FGR*, 2004, pp. 111–116.

[2] M. S. Bartlett, G. Littlewort, E. Vural, K. Lee, M. Çetin, A. Erçil, and J. R. Movellan, "Data mining spontaneous facial behavior with automatic expression coding," in *COST 2102 Workshop (Patras)*, 2007, pp. 1–20.

[3] E. Vural, M. Çetin, A. Erçil, G. Littlewort, M. S. Bartlett, and J. R. Movellan, "Drowsy driver detection through facial movement analysis," in *ICCV-HCI*, 2007, pp. 6–18.

[4] K. F. V. Orden, T.-P. Jung, and S. Makeig, "Combined eye activity measures accurately estimate changes in sustained visual task performance," *Biological Psychology*, 2000 Apr;52(3):221-40.

[5] P. Ekman and W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA: Consulting Psychologists Press, 1978.

[6] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Automatic recognition of facial actions in spontaneous expressions." *Journal of Multimedia.*, 1(6) p. 22-35.

[7] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Classifying facial actions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974–989, 1999.