Distinction of Stress and Non-Stress Tasks using Facial Action Units

Carla Viegas
Carnegie Mellon University
NOVA University of Lisbon
Pittsburgh, PA 15213, USA
cviegas@cs.cmu.edu

Roy Maxion
Carnegie Mellon University
Pittsburgh, PA 15213, USA
maxion@cs.cmu.edu

Shing-Hon Lau
Carnegie Mellon University
Pittsburgh, PA 15213, USA
shinghol@cs.cmu.edu

Alexander Hauptmann
Carnegie Mellon University
Pittsburgh, PA 15213, USA
alex@cs.cmu.edu

ICMI '18,, October 16–20, 2018, Boulder, CO, USA
ACM 978-1-4503-6002-9/18/10.
https://doi.org/10.1145/3281151.3281158

Abstract

Long-exposure to stress is known to lead to physical and mental health problems. But how can we as individuals track and monitor our stress? Wearables which measure heart variability have been studied to detect stress. Such devices, however, need to be worn all day long and can be expensive. As an alternative, we propose the use of frontal face videos to distinguish between stressful and non-stressful activities. Affordable personal tracking of stress levels could be obtained by analyzing the video stream of inbuilt cameras in laptops. In this work, we present a preliminary analysis of 114 one-hour long videos. During the video, the subjects perform a typing exercise before and after being exposed to a stressor. We performed a binary classification using Random Forest (RF) to distinguish between stressful and non-stressful activities. As features, facial action units (AUs) extracted from each video frame were used. We obtained an average accuracy of over 97% and 50% for subject dependent and subject independent classification, respectively.

Author Keywords

stress detection; affective computing; facial action units.

ACM Classification Keywords

I.5.4 [Pattern Recognition]: Applications; I.4.9 [Image Processing and Computer Vision]: Applications

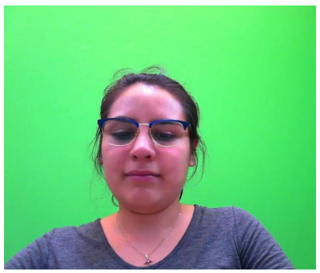

Figure 1: Example of a frame of one of the videos recorded under the same condition as the dataset used for analysis.

Introduction

Stress over longer time periods can lead to severe physical and psychological conditions. In a world where sleep quality [16] and depression can be tracked [21] tracking stress is one of the next steps to go.

There are several known techniques to detect stress, such as measuring stress hormones in saliva or blood and measuring heart rate variability. As those techniques can be intrusive, new ways of measuring stress are being investigated such as using voice features, eye gaze, and facial expressions [3]. Specially, mobile monitoring in combination with wearables has gained increased attention [5, 4]. However, different methods evaluated require wearing expensive wrist bands or chest bands all-day long which can be invasive.

In this work, we propose the detection of facial expressions on video as a potential method for stress detection. By analyzing video streams of webcams, personal monitoring of stress could be more affordable and less invasive than other methods. In [13, 7] cognitive stress during driving was detected through video recordings. Specific facial action units (AUs) [6] that occur in the presence of fear and anger were used as features. In contrast, we propose a less restricted analysis by evaluating 18 different AUs in each frame.

We were able to distinguish stress from non-stress situations with an average accuracy of over 97% and over 50% for person dependent and independent classification respectively. Besides personal tracking of stress, other applications can profit from stress detection through video such as security surveillance and interview training.

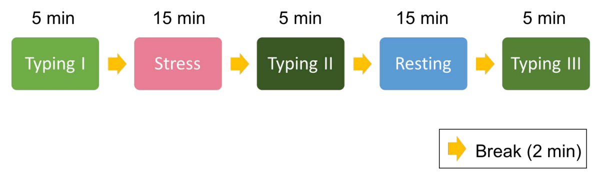

Figure 2: Overview of the different phases recorded for each subject.

Related Work

The main purpose of stress is to provide the human body with extra energy in situations of danger to increase chances of survival. However, when exposed to stress during a longer time period the human body is not resting properly which can lead to cardiovascular disorders, anxiety, depression, weakened immune system and obesity [1].

A variety of physiological signals have been studied for stress detection [9]. Of them the most used has been Galvanic Skin Response as it provides direct insides into the autonomous emotion regulation [18, 10, 20]. Also speech has been used for stress detection, using Mel-frequency Cepstral Coefficients and pitch [12, 17]. Hernandez et al. used a pressure-sensitive keyboard and a capacitive mouse to detect stress [11]. His results show that participants increase their typing pressure and the contact surface with the mouse when exposed to stress. Although wearables would be ideal to track stress continuously, average accuracies vary from 71% to 75% for binary classification of high and low stress conditions for user-specific models [8, 19].

As stress is related to emotions, facial expressions have been used to detect cognitive stress during driving. In [7] the presence of anger and fear was used to determine whether the driver is stressed and in [22] the correlation be-

Table 1: Action Units (AUs) used as features. Most of them provide intensity information which vary between 0-5. AU7, AU23 and AU45 are binary values.

AU	Description
AU1	Inner brow raiser
AU2	Outer brow raiser
AU4	Brow lowerer
AU5	Upper lid raiser
AU6	Cheek raiser
AU7	Lid tightener
AU9	Nose wrinkler
AU10	Upper lip raiser
AU12	Lip corner puller
AU14	Dimpler
AU15	Lip corner depr.
AU17	Chin raiser
AU20	Lip stretcher
AU23	Lip tightener
AU25	Lips part
AU26	Jaw drop
AU28	Lip suck
AU45	Blink

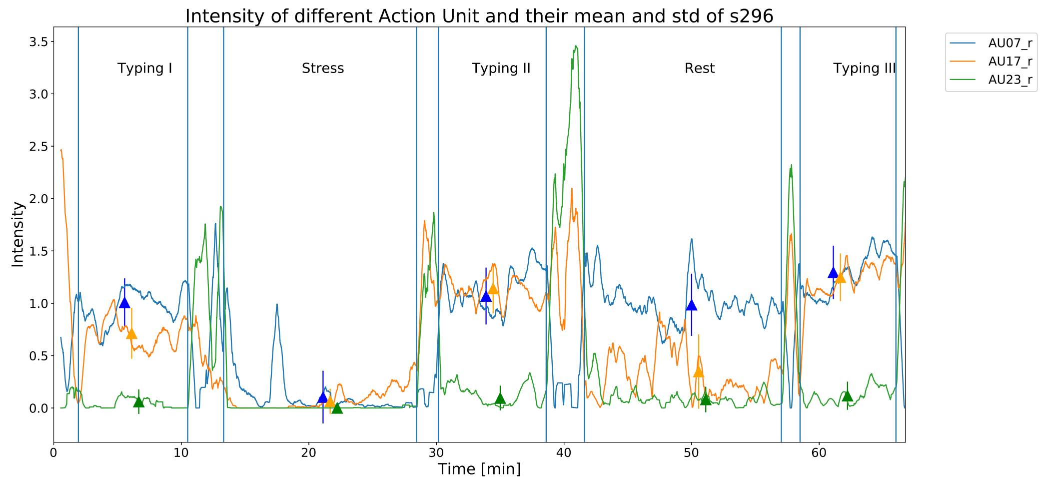

Figure 3: Plot showing three different AUs over the entire one-hour video of one subject. The mean intensity and their standard deviation are shown as triangle markers.

tween AUs was used for stress detection. In this work, we use 18 AUs as features to distinguish between stress and non-stress activities. AUs were chosen as features because they are more interpretable than low-level features.

Methodology

Data

The dataset consists of 114 one-hour long videos, each of a different subject [15]. The camera used was a Microsoft Life Studio Pro webcam with 1080p resolution at 30 frames per second. The webcam was positioned below the monitor used by the participants, recording the frontal face of the subject (see Fig. 1) during the entire experiment. Fig. 2 shows the order of the different actions. There are three typing phases: 1) before stressor, 2) after stressor, and 3) after relaxation. The stressor used was a multitasking exercise with social evaluation. for 15 min. In between the different phases, the subjects were asked to fill out questionnaires. As during those periods, it is not guaranteed to obtain frontal face data, those segments are disregarded. To ensure the stress state of the participants caused by the stressor, blood pressure and ECG was collected [15].

Features

Several AUs from the Facial Action Coding System (FACS) are used to classify stressed from non-stressed activities. To obtain the AUs per frame, the toolbox OpenFace [2] was used which can detect 18 different AUs (see Table 1). Only the frames in which OpenFace was able to detect successfully all AUs were used for classification. The intensity of AUs are continuous values between 0-5 except for AU7, AU23, and AU45 which are binary (present/absent). To

Acknowledgements: This work was supported by a fellowship from the Portuguese Foundation for Science and Technology through the CMU Portugal Program and the Bio-VisualSpeech project (grant CMUP-ERI/TIC/0033/2014).

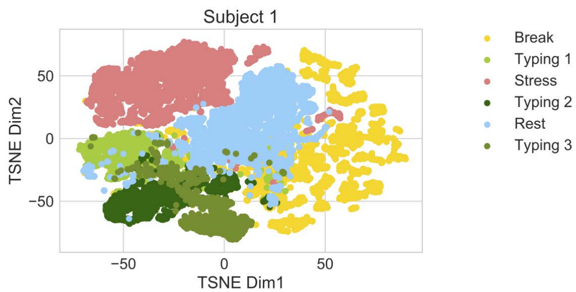

Figure 4: t-SNE plot showing the reduced feature representation for each frame for one subject. Each color shows the correspondence to the phase of the data point.

account for person-specific differences, the features were standardized per subject.

Data Visualization

Fig. 3 shows the behaviour of three AUs for one subject during the entire recording. The AUs behave visibly different during stress than during the remaining phases. However, the behaviour of the shown AUs is different from person to person. For different subjects other AUs have higher correlation with stress. To see if it is possible to distinguish within subject the different phases, all data points of one subject are visualized in Fig. 4. Each point represents one frame. The dimensions of the features representing one frame are reduced from 18 to 2 dimensions using t-SNE (t-distributed Stochastic Neighbor Embedding) [14]. In Fig. 4 we see how frames during stress are clustered together. We also see how the different typing phases partially overlap.

Classification

Two types of classification problems were addressed: 1) subject dependent and 2) subject independent. For the first, 20% of each phase were used for testing and the remaining for training. In the second case, leave-one-subject-out (LOSO) was performed: the data of 113 subjects was used for training and the data of one subject was used solely for testing. Besides performing binary classification between Stress and Resting, Typing 1 and Typing 2, a classifier was trained to distinguish all five phases (breaks in between phases were excluded). The classifier chosen was RF.

Labels	*Subj. Dep.*	*Subj. Indep.*
All classes	0.95 ± 0.04	0.30 ± 0.09
Stress vs Rest	0.99 ± 0.01	0.50 ± 0.13
Typ.1 vs Typ.2	0.97 ± 0.03	0.51 ± 0.23

Table 2: Average accuracies and standard deviations of the results obtained using RF as well as using all 5 classes.

Results

The results in Table 2 show that RF achieves over 95% average accuracy for the classification of all phases as well as for binary classification between stressful and non-stressful activities. The results for subject independent classification are in contrast not satisfactory.

Discussion

Distinguishing the different activities of the videos using solely AUs was possible using RF for the subject dependent case. However, this is not the case for the person independent classification. The classification results as well as visual analysis of the AUs over time suggest that subjects react differently to stress. The next step is to evaluate if smaller groups of subjects behave similarly. If that is the case, subject independent classification could be improved by using a model for classification that was only trained on data from subjects that behave similarly.

REFERENCES

1. American Psychology Association. 2014. Understand Chronic Stress. (2014). Retrieved May 16, 2018 from `https://www.apa.org/helpcenter/understanding-chronic-stress.aspx`.
2. Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. 2016. Openface: an open source facial behavior analysis toolkit. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*. IEEE, 1–10.
3. Davide Carneiro, Paulo Novais, Juan Carlos Augusto, and Nicola Payne. 2017. New methods for stress assessment and monitoring at the workplace. *IEEE Transactions on Affective Computing* (2017).
4. Shanice Clarke, Luis G Jaimes, and Miguel A Labrador. 2017. mStress: A mobile recommender system for just-in-time interventions for stress. In *Consumer Communications & Networking Conference (CCNC), 2017 14th IEEE Annual*. IEEE, 1–5.
5. Božidara Cvetković, Martin Gjoreski, Jure Šorn, Pavel Maslov, Michał Kosiedowski, Maciej Bogdański, Aleksander Stroiński, and Mitja Luštrek. 2017. Real-time physical activity and mental stress management with a wristband and a smartphone. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM, 225–228.
6. Paul Ekman and Wallace V Friesen. 1978. *Facial Action Coding System: Investigatoris Guide*. Consulting Psychologists Press.
7. Hua Gao, Anil Yüce, and Jean-Philippe Thiran. 2014. Detecting emotional stress from facial expressions for driving safety. In *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 5961–5965.
8. Enrique Garcia-Ceja, Venet Osmani, and Oscar Mayora. 2016. Automatic stress detection in working environments from smartphones' accelerometer data: a first step. *IEEE journal of biomedical and health informatics* 20, 4 (2016), 1053–1060.
9. Jennifer A Healey and Rosalind W Picard. 2005. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on intelligent transportation systems* 6, 2 (2005), 156–166.
10. Javier Hernandez, Rob R Morris, and Rosalind W Picard. 2011. Call center stress recognition with person-specific models. In *International Conference on Affective Computing and Intelligent Interaction*. Springer, 125–134.
11. Javier Hernandez, Pablo Paredes, Asta Roseway, and Mary Czerwinski. 2014. Under pressure: sensing stress of computer users. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 51–60.
12. Hindra Kurniawan, Alexandr V Maslov, and Mykola Pechenizkiy. 2013. Stress detection from speech and galvanic skin response signals. In *Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on*. IEEE, 209–214.
13. Jennifer S Lerner, Ronald E Dahl, Ahmad R Hariri, and Shelley E Taylor. 2007. Facial expressions of emotion reveal neuroendocrine and cardiovascular stress responses. *Biological psychiatry* 61, 2 (2007), 253–260.

14. Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, Nov (2008), 2579–2605.

15. Roy A. Maxion and Shing-hon Lau. 2018. *Facial Expressions During Stress and Non-Stress Conditions: A Benchmark Video Collection*. Technical Report CMU-CS-18-110. Computer Science Department, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA 15213.

16. Rebecca Robbins, Paul Krebs, David M Rapoport, Girardin Jean-Louis, and Dustin T Duncan. 2018. Examining Use of Mobile Phones for Sleep Tracking Among a National Sample in the USA. *Health communication* (2018), 1–7.

17. Olympia Simantiraki, Giorgos Giannakakis, Anastasia Pampouchidou, and Manolis Tsiknakis. 2016. Stress Detection from Speech Using Spectral Slope Measurements. In *Pervasive Computing Paradigms for Mental Health*. Springer, 41–50.

18. Feng-Tso Sun, Cynthia Kuo, Heng-Tze Cheng, Senaka Buthpitiya, Patricia Collins, and Martin Griss. 2010. Activity-aware mental stress detection using physiological sensors. In *International Conference on Mobile Computing, Applications, and Services*. Springer, 282–301.

19. Elena Vildjiounaite, Johanna Kallio, Jani Mäntyjärvi, Vesa Kyllönen, Mikko Lindholm, and Georgy Gimel'farb. 2017. Unsupervised Stress Detection Algorithm and Experiments with Real Life Data. In *Portuguese Conference on Artificial Intelligence*. Springer, 95–107.

20. María Viqueira Villarejo, Begoña García Zapirain, and Amaia Méndez Zorrilla. 2012. A stress sensor based on Galvanic Skin Response (GSR) controlled by ZigBee. *Sensors* 12, 5 (2012), 6075–6101.

21. Rui Wang, Weichen Wang, Alex daSilva, Jeremy F. Huckins, William M. Kelley, Todd F. Heatherton, and Andrew T. Campbell. 2018. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 43 (March 2018), 26 pages. DOI:http://dx.doi.org/10.1145/3191775

22. Anıl Yüce, Hua Gao, Gabriel L Cuendet, and Jean-Philippe Thiran. 2017. Action units and their cross-correlations for prediction of cognitive load during driving. *IEEE Transactions on Affective Computing* 8, 2 (2017), 161–175.