We've never been eye to eye: a pupillometry pipeline for the detection of stress and negative affect in remote working scenarios

Alexander Heimerl, Linda Becker, Dominik Schiller, Tobias Baur, Fabian Wildgrube, Nicolas Rohleder, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Heimerl, Alexander, Linda Becker, Dominik Schiller, Tobias Baur, Fabian Wildgrube, Nicolas Rohleder, and Elisabeth André. 2022. "We've never been eye to eye: a pupillometry pipeline for the detection of stress and negative affect in remote working scenarios." In *PETRA '22: proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments, 29 June - 1 July 2022, Corfu, Greece*, edited by Fillia Makedon, 486–93. New York, NY: ACM. https://doi.org/10.1145/3529190.3534729.

We've never been eye to eye: A Pupillometry Pipeline for the Detection of Stress and Negative Affect in Remote Working Scenarios

Alexander Heimerl
alexander.heimerl@informatik.uni-augsburg.de
Lab for Human-Centered AI,
Augsburg University
Augsburg, Germany

Linda Becker
linda.becker@fau.de
Department of Psychology,
Friedrich-Alexander University
Erlangen-Nürnberg
Erlangen, Germany

Dominik Schiller
dominik.schiller@informatik.uni-augsburg.de
Lab for Human-Centered AI,
Augsburg University
Augsburg, Germany

Tobias Baur
tobias.baur@informatik.uni-augsburg.de
Lab for Human-Centered AI,
Augsburg University
Augsburg, Germany

Fabian Wildgrube
fabian.wildgrube@student.uni-augsburg.de
Lab for Human-Centered AI,
Augsburg University
Augsburg, Germany

Nicolas Rohleder
nicolas.rohleder@fau.de
Department of Psychology,
Friedrich-Alexander University
Erlangen-Nürnberg
Erlangen, Germany

Elisabeth André
andre@informatik.uni-augsburg.de
Lab for Human-Centered AI,
Augsburg University
Augsburg, Germany

ABSTRACT

In this paper, we present a processing pipeline for the analysis of stress and negative affect based on pupillometry. We were able to show that it is possible to extract meaningful pupil features from video data recorded by an infrared- (IR-) sensitive webcam and successfully trained a Support Vector Machine on the corresponding dataset. Further, we conducted a study that shows that the proposed pipeline is suitable for the assessment of stress as well as negative affect during stress eliciting situations in a digital environment.

CCS CONCEPTS

• **Human-centered computing**; • **Computing methodologies** → *Machine learning*; • **Hardware** → Sensor applications and deployments;

KEYWORDS

Pupillometry, Stress Detection, Remote Working, Affective Computing

PETRA '22, June 29-July 1, 2022, Corfu, Greece

ACM ISBN 978-1-4503-9631-8/22/06…$15.00
https://doi.org/10.1145/3529190.3534729

ACM Reference Format:
Alexander Heimerl, Linda Becker, Dominik Schiller, Tobias Baur, Fabian Wildgrube, Nicolas Rohleder, and Elisabeth André. 2022. We've never been eye to eye: A Pupillometry Pipeline for the Detection of Stress and Negative Affect in Remote Working Scenarios. In *The15th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '22), June 29-July 1, 2022, Corfu, Greece.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3529190.3534729

1 INTRODUCTION

As the pandemic continues to affect our daily lives, working from home has become a routine for many. While this shift brings some potential benefits to workers, such as less commuting and more flexibility in work schedules, it also comes with considerable challenges for an employee's health. A recent meta-study by Oakman et al. [16] identified a series of physical- and mental health-related issues related to working at home: pain, self-reported health, safety, well-being, stress, depression, fatigue, quality of life, strain, and happiness. Although an employee may be well aware of such problems, tracking down the often individual causes proves to be a non-trivial task. Therefore, an important first step in preventing the manifestation of circumstances that are detrimental to health is to create awareness. In this paper, we want to specifically address the problem of recognizing perceived stress and negative emotions for home office employees in order to provide them with helpful feedback. A promising technique that, to the best of our knowledge, has not been investigated in such a scenario so far, is pupillometry. This is a widely used method to measure conscious and unconscious emotional reactivity, or cognitive load [23] [17]. Changes in pupil size have been associated with negative emotions, such

as anxiety, as well as with perceived and physiological stress. An advantage of pupillometry over other physiological and psychological measures that are related to emotions (e.g., body posture or voice) is that pupil dilations are not voluntarily controlled [2]. Furthermore, pupil analysis is usually performed by using images from an optical sensor as input. This means that data collection is not only unobtrusively but can potentially also be done with traditional webcams that are already present in most home offices. Therefore, pupil responses are a promising means for uncovering stress and negative emotions when working on a PC. The goal of this paper is to develop a cost-efficient setup, including an open-source pipeline, for the assessment and analysis of pupillometric reactivity.

2 BACKGROUND AND RELATED WORK

2.1 Automatic stress detection

Stress and its negative effects on health and well-being are impacting the lives of many humans around the world. In order to mitigate or negate the impact of stress, various approaches for automatic stress detection have been developed utilizing different modalities, e.g., physiological signals (EEG, ECG, GSR, etc.), video, audio, eye blinking, or keyboard typing behavior [5]. Moreover, there is an increasing interest in the automatic detection of stress using wearables. By now, wearables (e.g., smartphones, smartwatches, fitness trackers) are commonly used in the general population. Therefore, making the automatic detection of stress applicable in day-to-day life. Wearables are equipped with capable sensors to measure physiological signals like ECG, PPG, GSR, allowing to gain valuable insights on perceived stress and emotional arousal [22]. Can et al. [4] developed a system for automatic stress detection based on heart activity, skin conductance, and accelerometer data from Samsung Gear S smartwatches and Empatica E4 wristbands. They trained six different classifiers with the Weka machine learning toolkit to distinguish between three levels of stress. Their best classifier achieved an accuracy of 90.40%. Jebelli et al. [8] developed a mobile deep learning based framework to detect low and high stress in construction workers during their work at construction sites. The authors trained several fully connected deep neural network and convolutional neural network architectures on EEG signals obtained from mobile EEG sensors that have been fit into the workers' safety hats. For training the classifiers they collected EEG data from 10 construction workers. On their binary classification task, the best performing fully connected deep neural network achieved an accuracy of 86.62%, whereas the best convolutional neural network had a prediction accuracy of 64.20%. Reliable automatic detection of stress is an important pillar to mitigate the negative effects of stress, as it can be used to raise peoples' awareness of stressful situations in their everyday life. Raising awareness enables them to reflect on the reason for stress and potential stress eliciting environments. Schmidt et al. [21] developed a mobile assistant that helps people cope with their stress. Their coping assistant uses sensor data to warn the user of increased stress and also reports back why they might be currently stressed. Next, the assistant proposes targeted coping strategies on how to deal with the perceived stress. Finally, automated measures are performed to reduce the stress exposure, e.g., reducing interruptions by blocking notifications.

2.2 Pupillometry for stress and emotion recognition

Emotional arousal is associated with the activation of the autonomic nervous system (ANS), i.e., with an up-regulation of its sympathetic and a down-regulation of its parasympathetic branch [19]. ANS activity is usually assessed by means of electrophysiological measures such as heart rate (HR) or skin conductance (SC). However, a further promising ANS marker are pupil responses [23]. Overall, pupil dilations can be used as a readout of emotional arousal, because they are related to negative emotions such as anxiety as well as with the human stress response. Emotional arousing stimuli are associated with pupil dilation, i.e., with an increase in pupil diameter [23] [17]. Pupil responses have been successfully applied in several experiments in the field of fear conditioning (e.g., [11], [24]). Moreover, pupil responses are closely related to subjective and physiological stress responses (e.g., the activation of the hypothalamic-pituitary adrenal (HPA) axis; [26] [18]). Furthermore, a recent study has shown that pupillometry is a suitable tool to measure arousal during emotion regulation after an acute stressor [9] [18]. Importantly, pupil dilations are involuntary responses of the ANS and can, therefore, not be voluntarily controlled [17], which is an advantage over other ANS or psychological measures in the context of stress and emotional reactivity, as it enables to assess more reliable results, due to the mitigated influence of social and behavioral interdependencies. Therefore, pupillometry is a promising tool that might be superior over other ANS measures for the assessment of stress and emotions in several settings.

3 PUPILLOMETRY PIPELINE

In order to analyse stress and emotional arousal through pupillometry, we implemented an open-source processing pipeline that is implemented with the Social Signal Interpretation framework (SSI) [25]. SSI provides an infrastructure for the development of online recognition systems from multiple synchronized sensory devices. The SSI Framework already provides a variety of processing methods and feature extractors for physiological signals like ECG, PPG, SC. Therefore, we extended the SSI Framework by implementing a new SSI plugin named PupilTrackingCore, to extract meaningful pupil features. In this section, we focus on the details of this plugin.

In order to track the pupil size in an input video, produced by a video capture device, just like the webcam displayed in Figure 2, we combined different existing technologies in the PupilTrackingCore package. To detect the pupil diameter in an arbitrary video frame, a number of processing steps are performed. An overview of the architecture, inputs and outputs of the core pupil tracking package is displayed in Figure 1.

First, we extracted two regions of interest containing the left and right eye. For this purpose, we incorporated MediaPipe [14], a framework to build perception pipelines. Part of this framework is a pipeline to reliably track the human eyes [1]. Amongst other things, the pipeline returns the position of the iris center as well as the diameter of the iris. Based on the provided coordinates of both irises, we crop two regions of interest (R_{eye_left} and R_{eye_right}) from the original input frame containing the eyes. The cropped regions are rectangular with a side length of three times the iris diameter. This ensures that the entire eye is present in the extracted frames.

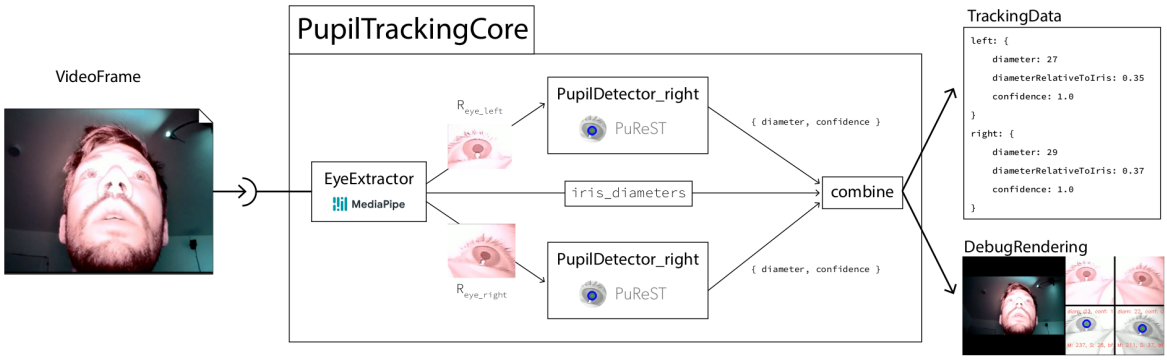

Figure 1: The architecture, inputs and outputs of the core pupil tracking package. A bitmap image with a human face is passed in, the eyes are detected, their respective pupils are tracked, and the tracking data as well as a debug video rendering are output.

Each R_{eye_*} is automatically optimized by converting it to a gray scale image and increasing brightness and contrast. To optimize the contrast between pupil and iris, we sample the brightness of the pupil and of the surrounding parts of the iris by averaging the brightness of six cells with a fixed size. One cell covers the pupil while the other five cover neighboring cells of the pupil. Cells that might be covered by the eyelid are omitted. Finally, the average brightness of each cell is calculated and the contrast in the image is boosted inversely to the difference between the brightness of the pupil's cell and the average of all other cells.

Each optimized R_{eye_*} is then passed to an instance of the pupil-tracker PuReST[20], which returns the pupil diameter in pixels as well as a confidence rating (0..1). In addition to that, we calculate a normalized pupil diameter (0..1) by dividing the pupil diameter by the iris diameter. This way, the distance between the capturing device and the face does not affect the returned size of the pupils. Whenever the tracking fails, -1 is returned for all dimensions. The component accepts video frames served from any video capture device via a Transmission Control Protocol (TCP) socket connection. Establishing the connection and synchronizing the output with other components is handled by SSI. This way, it is possible to use the PupilTrackingCore alongside multiple sensory devices in multimodal recording and processing pipelines. Additionally, the PupilTrackingCore package provides the possibility to return a separate MP4 file for manual inspection of the source video and tracking results (see Figure 2). In addition to that, we also developed a stand-alone command-line tool to process pre-recorded videos. [1]

4 PILOT STUDY

To assess the feasibility of our processing pipeline and our PupilTrackingCore component we conducted a pilot study. The study setting had to fulfil two main criteria. First, the setup had to elicit stress and emotional arousal in the participants. Second, in order to be comparable to a remote-working environment, the setting had to take place in virtual space. Therefore, we decided on a virtual job-interview scenario conducted via an online meeting tool. Performing online job-interviews has become a common procedure in modern working environments. Moreover, job interviews are by their nature a complex stressful social scenario where different aspects of human interaction and perception collude. Previous research has shown that psychosocial stress also occurs in mock job interviews [3, 6]. This makes the study setting not only meet our criteria but also depict a realistic real-world remote scenario. Furthermore, the physiological response to the stressor can be translated to other virtual settings, due to the fact that pupil dilations are involuntary responses of the ANS and cannot be voluntarily controlled [17]. Therefore, it can be assumed that stress and emotional arousal eliciting situations produce comparable pupil responses. In order to have a benchmark for our generated pupil features we also decided to consider two well-established physiological modalities in the context of automatic stress detection, namely HR and SC [22].

4.1 Participants

Participants were eight German-speaking young adults (2 male, 28.8 ± 5.7 years), who all had experience with job interviews. Mean number of previous job interviews was 5.4 ± 2.4 (range 3 to 10). Seven participants were students and one participant was a full-time employee. Highest educational degrees were secondary school ($n = 1$) in one case, general qualification for university entrance ($n = 4$), Bachelor degree ($n = 2$), and Ph.D. ($n = 1$). Participants were recruited from local universities via flyers and mailing lists. All participants provided written and informed consent.

4.2 Materials

In order to record the different modalities during the remote job-interview simulation, we utilized different sensors and video-capture devices. We used the IOM-biofeedback sensor to collect SC and blood volume pulse. The participant's audio was recorded using a studio microphone with a sample rate of 16,000 Hz. The virtual job interviews were conducted by using an online meeting tool (ZOOM). Moreover, we utilized different kinds of video capture

[1] Our implementation is available at https://hcai.eu/git/alexanderh/pupil_tracking.git.

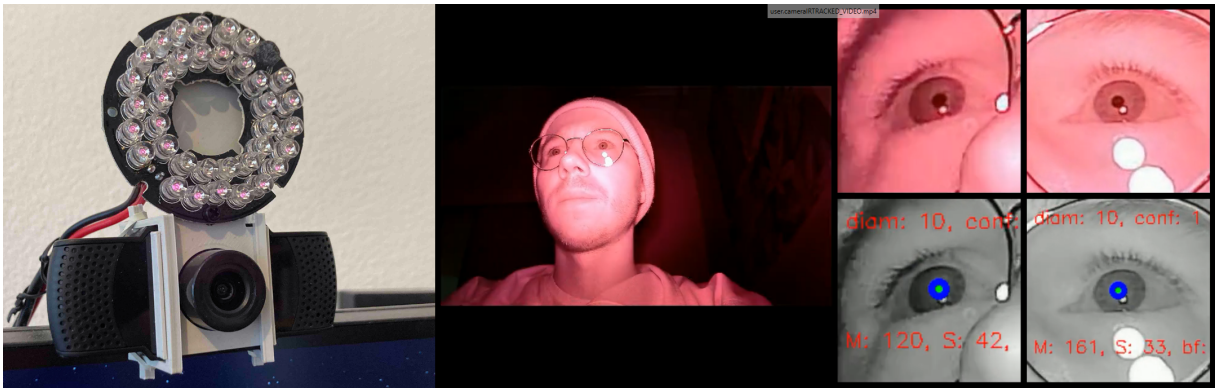

Figure 2: Screenshot of the video returned by the PupilTrackingCore. The middle frame displays the original video recorded by the IR-sensitive webcam (left). On the right side the cropped regions for the left and right eye are displayed. The bottom row shows additional tracking information.

devices. For the ZOOM meeting itself, a common webcam was used. For recording the pupil data, we decided to use two different devices to be able to investigate the applicability of our approach, given different setups. We used a Pupil Labs wearable eyetracker as well as a modified full HD webcam (see the left frame in Figure 2). Since CMOS sensors used in most low-cost digital cameras are capable of recording infrared light, we converted a ordinary webcam to record in the infrared spectrum. We opted to use a "CONCEPTRONIC AMDIS 1080p Webcam" on which we removed the infrared filter. This modification isn't challenging and can be done in a few minutes with just a screwdriver. In addition to that, we equipped the webcam with a ring-light of 36 infrared LEDs to increase the contrast of the pupils.

4.3 Study design and procedure

Participants were invited to the laboratory and were told that physiological reactions during an online job interview will be recorded. In advance, participants sent their curriculum vitae (CV) to the experimenter and filled-out an online survey, in which demographic variables and experiences with job interviews were assessed. After arrival in the laboratory, they were asked about their dream job and were equipped with HR and SC sensors. Then, they had ten minutes to prepare for the interview. The actual interview took place online via ZOOM through a computer that was placed in front of the participants. Participant and interviewer were sitting in separate rooms during the mock interview. The interviewer tried to ask critical questions to stress the applicant and to induce negative emotions. Contents of the interviews included questions about strengths and weaknesses of the applicant, dealing with difficult situations in the job, salary expectations, willingness to work overtime, as well as inconsistencies in the CV. In addition, tasks related to logical thinking were asked as well as questions about basic knowledge in the areas of mathematics and language. After the job interviews, participants were asked about their emotions during the interview and rated whether they perceived the situation rather as a threat or as a challenge. After this, participants reported whether they felt stressed at any time point during the interviews. Afterwards, participants were instructed to describe as precisely as possible in which specific situations during the job interviews they felt stress. This procedure (rating and assignment to specific situations) was repeated for all of the reported negative emotional states.

4.4 Analysis/Annotation

Figure 3: An instance of NOVA From top downwards: Infrared video containing the cropped regions for both eyes with and without tracking information. In the lower part, several sensor signals, e.g. audio, heart rate, skin conductance, and pupil diameter are displayed. The bottom row shows a discrete annotation tier.

The self-reports of the specific situations in which the emotions were experienced were used as a basis for the annotations. An experienced psychologist annotated the recordings based on the participants' reports and the content of the interviews. Categories for the annotations were the categories from the questionnaire (i.e., stress as well as the reported negative emotions shame, anxiety, anger) and a neutral state. In total, 301 minutes of data were annotated frame by frame. There were no disagreements between the psychologist's ratings and the participants' self-reports, i.e., for

every situation that was assigned to stress or an emotion by the participants, a time window could be assigned by the psychologist and a corresponding annotation could be created. The NOVA tool [7] was used for annotation. A screenshot of a loaded recording session from our study is shown in Figure 3.

NOVA complements the SSI framework as it allows to directly visualize and annotate streams recorded with the framework, including our PupilTracking Core plugin. Since SSI is primarily designed to build online recognition systems, a trained model can then be directly used to detect either social cues or as in our case, stress and negative emotions, in real-time.

4.5 Results

4.5.1 Perceived emotions during the mock job interviews. All participants (100%) reported that they felt stressed and uncomfortable during the job interviews as well as in the preparation phase. Most of the participants ($n = 5$, 62.5%) reported that they perceived the situation as a challenge rather than as a threat. One participant reported that she felt completely uncomfortable and that she had a blackout during solving the maths problems. However, no one had the need to cancel the interview. The other $n = 3$ participants (37.5%) rated the situation as rather threatening than challenging. Further negative emotions that were reported were shame, anxiety, and anger. Moreover, all participants reported that they forgot that they were part of an experiment during the interviews. They mainly attributed this to the virtual setting.

4.5.2 Automatic detection of stress and negative emotions using pupillometry. During stress and emotional events, pupil size increased. The largest changes in pupil diameter we found were approximately 33%. Figure 4 shows an example of pupil dilation in response to an emotional event.

Figure 4: Captured change in pupil diameter before and after an emotional event.

Moreover, we compared the tracking results when utilizing video data generated by the Pupil Labs eyetracker in contrast to an IR-sensitive webcam. Figure 5 displays example frames for successful and failed tracking. The tracking with the webcam often failed due to a misalignment of the participant towards the camera. Tracking with the eyetracker usually failed when participants were looking downwards or were blinking. In section 3 we described that every calculated sample also contains information about whether the tracking was successful or not and how confident the system is in the tracking result. Based on this information we calculated the tracking performance. We considered samples with a confidence value below 0.8 as unsuccessfully tracked. With the data generated

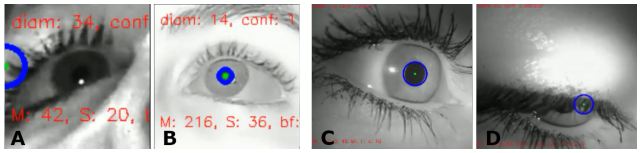

Figure 5: Comparison of the tracking performance between the modified webcam (images A and B) and the Pupil Labs eyetracker (images C and D). Images A and D display example frames where the tracking failed, whereas images B and C show examples where the tracking was successful.

from the pupil labs wearable eyetracker 75% of all frames could be tracked correctly, with the data from the modified webcam 53% of the frames were tracked successfully. The distance between the webcam and the participants was approximately 40 cms, whereas the distance between the camera of the eyetracker and the participants' eye was around 3 cms.

In order to investigate whether our proposed pipeline is capable of providing meaningful features to detect stress and negative emotions, we trained five One-Vs-All Support Vector Machines (SVM) to recognize the five different emotional states (stress, shame, anxiety, anger, neutral). The features used in training the SVM were extracted for each frame from the video provided by the modified webcam. In addition to that, we utilized HR and SC provided by the IOM-biofeedback sensor as a benchmark for our pupil features. The resulting feature set consisted of eight-dimensional vectors (six pupil features + heart rate + skin conductance). To improve the stability of the system we removed all samples where the confidence of the pupil tracker for a correct prediction fell below the threshold of 0.8. Subsequently, we split all samples randomly into fixed sets for training, validation and test using a ratio of 60 / 20 / 20 for the respective sets. This left us with an overall amount of 18,912 samples for train and 6,304 samples for validation and 6,305 samples for test. Since the class distribution in the training data is strongly unbalanced (anger: 758, shame: 1,935, anxiety: 2,519, stress: 9,390, neutral: 4,092), we first randomly removed samples from the stress-class to match the number of samples in the neutral class. We then used Synthetic Minority Oversampling to increase the number of samples from all other classes to match the number of samples in the neutral class. Therefore, the resulting training set consists of 20,460 samples evenly distributed over all classes. Before the training, all features have been scaled to a range between 0 and 1 over the whole dataset. For the final classifier, we first determined the optimal complexity parameter of the SVM on the validation set. The final classifier was then trained on both, the training- as well as on the validation set and evaluated on the test set. We repeated this experiment three times using different feature sets. The first feature set consists of only the pupil features described in section 3 for each eye. This feature set achieved an unweighted average recall of (UAR) **0.4208** on the validation set and **0.4335** on the test set. The respective distribution of predictions for the model trained on the pupil feature vectors is shown in Figure 6 and Figure 7. The second set consists of only the skin conductance feature as well as the heart rate and managed to achieve an UAR of **0.4302** for validation and **0.4303** for test. In the third set we fused both feature

vectors (pupil features + heart rate + skin conductance) to achieve scores of **0.5123** for validation and **0.5224** for test.

5 DISCUSSION

We were able to show that remote mock job interviews are a suitable use case for the induction of stress and negative emotions such as shame, anger, and anxiety. Moreover, our results confirm that the pupil diameter is a valid and reliable measure of negative emotions and stress that are both associated with autonomic arousal. In accordance with the literature [23] [17], we found that pupil size increased when the participants felt stressed or during an emotional event.

Further, we investigated the applicability of an IR-sensitive webcam in regard to tracking results and compared it to a professional wearable eyetracker. Both sensors come with their advantages and disadvantages. The biggest benefit of utilizing a wearable eyetracker is that the distance between the eye and camera lens is very small and constant. Therefore, the tracking is rarely compromised by the movements of the participant. In our experiment, we were able to achieve 75% successfully tracked frames. The missing 25% are most likely due to blinking and the participants looking down. In comparison to that, we were able to achieve 53% correctly tracked frames when utilizing the data generated by our modified webcam. However, it is important to note that we started recording the participants during the preparation phase where some of them were still wearing masks, due to Covid-19 pandemic safety regulations. Wearing masks compromised the facial tracking of MediaPipe, which resulted in falsely recognized face regions. Therefore, it might be that the actual tracking scores for the modified webcam approach are higher. The downside of wearable devices like the Pupil Labs eyetracker is that they are always invasive to some extent. Often they are much bulkier than regular glasses and wearing them for an extended amount of time during a workday might disturb users. In contrast to that, utilizing an IR-sensitive webcam for the detection of stress and negative emotions is non-invasive and allows the user to move and act freely. However, the camera has a limited field of view and it is possible that the users accidentally leave the recordable area. Also, depending on the distance and angle between the camera and the user the tracking results may be compromised. However, this can be mitigated by utilizing a high-resolution webcam with a wide-angle lens. Also, given that the use case for the proposed pipeline is a remote working environment the distance and position of the users towards their PC most likely won't vary that much and might even be almost static for the majority of the day. Another advantage of utilizing a modified webcam over a professional eyetracker is that the webcam only costs a fraction of the price of an eyetracker. Further, it is important to point out that some of the big business laptop manufacturers like Dell or Lenovo already equip their latest laptops with IR-cameras. That means depending on the device at hand users might not even need to buy any additional hardware to utilize the stress and negative emotion detection pipeline.

In our preliminary machine learning experiment, using the pipeline that is described in section 3, we found that the classifier solely trained with the pupil features achieved an unweighted average recall of 0.4335 on the test. As a comparison, the benchmark classifier trained with the heart rate and skin conductance features had an unweighted average recall of 0.4303 on the test set. The pupil features classifier slightly outperformed our benchmark classifier on the recorded dataset. This indicates that the pupil features hold valuable information for the task that can be utilized by the classifier. Looking at the confusion matrices for the validation and test set (Figure 6 and Figure 7) we can observe that the classes anger, stress, and neutral are distinguished best in both cases. When looking at the misclassified samples, we can see that the classes shame and anxiety are most often confused with stress. The misclassifications of shame and anxiety as stress is an interesting observation as both are heavily related to stress [12] [15]. Furthermore, the distribution of the predictions in the two sets is rather similar, which hints at a good generalization capability of the classifier. Finally, the results indicate that it is feasible to use a modified version of an ordinary webcam to extract reliable features for the detection of stress and negative emotions in a remote environment.

However, our results are limited to the detection of stress and associated negative emotions. In future research, it should be investigated, whether our approach also works for positive emotions which have also been reported in stressful situations (e.g., so-called eustress, [10]).

Using the combination of features (pupil features + heart rate + skin conductance) we were able to classify stress and negative emotions of participants with an unweighted average recall of 0.5224. Those results are in line with related research, e.g., Lu et al. [13] achieved a classification accuracy of 87.59 % considering three classes (positive, neutral, negative). Also, Zheng et al. [27] trained several machine learning models for the recognition of four emotion categories. Their best performing model was a SVM which achieved an accuracy of 57.05%. Even though related research reported higher accuracy scores it should be kept in mind that our model was trained on five classes, which increased the complexity of the classification task. Moreover, the fusion of different modalities, like heart rate, skin conductance and pupillometry for the automatic detection of stress and negative emotions can also be applicable for a remote working environment. Our SSI pupillometry pipeline can easily be extended by additional modalities, as the SSI framework provides an infrastructure for the development of online recognition systems from multiple synchronized sensory devices. In addition to that, SSI already includes support for a large variety of sensors. Therefore, users could potentially use their already existing wearables that are capable of sensing physiological signals and utilize them alongside an IR-sensitive camera. Our proposed pipeline can be applied on any device running Windows as an operating system. Starting and stopping the pipeline is handled by the SSI framework. Moreover, the recognition of stress and negative affect of our pupillometry pipeline may be used as a foundation to provide users with coping strategies. The recognition results can be put into relation with additional context information, e.g., elapsed time since the last break. Based on the combined information adequate recommendations, like taking a brake and going for a short walk can be suggested to mitigate the exposure to stress and negative emotions.

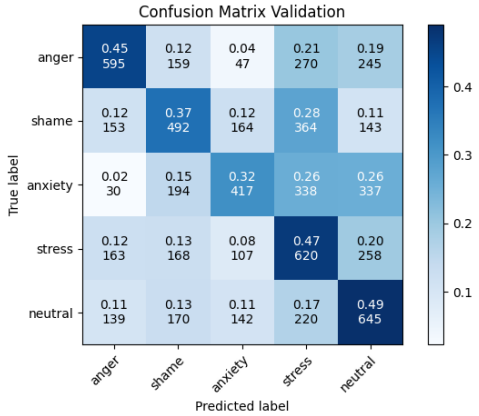

Figure 6: Confusion matrix for the validation set using only the train set for training.

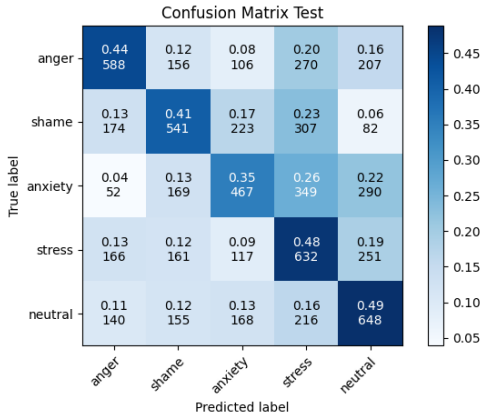

Figure 7: Confusion matrix for the test set using train and validation for training.

6 CONCLUSION

We presented a processing pipeline for the assessment of stress and negative emotions in a remote environment. Further, we employed the SSI real-time processing framework for which we implemented a plugin to run the PupilTrackingCore. We showed that it is possible to extract meaningful pupil features from video data recorded by a non-invasive, cost-efficient sensor device that is based on a modified webcam. Based on the extracted features and annotations we were able to successfully train a SVM to predict stress and negative emotions on our recorded dataset. The results of this preliminary study show that the proposed system can achieve performance that is in line with other state of the art approaches for the task. Further, the features extracted with the proposed pipeline hold additional relevant information over bio-signals. Moreover, the provided pipeline for pupillometry may be employed alongside multiple modalities (e.g., HR and SC). Therefore, it is applicable to a large variety of use cases. Further, our results indicate that measuring pupil dilation is a promising candidate for detecting stress and negative emotions in remote working scenarios.

ACKNOWLEDGMENTS

This work presents and discusses results in the context of the research project ForDigitHealth. The project is part of the Bavarian Research Association on Healthy Use of Digital Technologies and Media (ForDigitHealth), funded by the Bavarian Ministry of Science and Arts.

REFERENCES

[1] Artsiom Ablavatski, Andrey Vakunov, Ivan Grishchenko, Karthik Raveendran, and Matsvei Zhdanovich. 2020. Real-time Pupil Tracking from Monocular Video for Digital Puppetry. arXiv:2006.11341 [cs.CV]

[2] FD Bremner. 2004. Pupil assessment in optic nerve disorders. *Eye* 18, 11 (2004), 1175–1181.

[3] Jay Campisi, Yesika Bravo, Jennifer Cole, and Kyle Gobeil. 2012. Acute psychosocial stress differentially influences salivary endocrine and immune measures in undergraduate students. *Physiology & Behavior* 107, 3 (2012), 317–321.

[4] Yekta Said Can, Niaz Chalabianloo, Deniz Ekiz, and Cem Ersoy. 2019. Continuous Stress Detection Using Wearable Sensors in Real Life: Algorithmic Programming Contest Case Study. *Sensors* 19, 8 (2019).

[5] Sami Elzeiny and Marwa Qaraqe. 2018. Machine Learning Approaches to Automatic Stress Detection: A Review. In *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*. 1–6.

[6] Patrick Gebhard, Tobias Baur, Ionut Damian, Gregor Mehlmann, Johannes Wagner, and Elisabeth André. 2014. Exploring interaction strategies for virtual characters to induce stress in simulated job interviews. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. 661–668.

[7] A. Heimerl, T. Baur, F. Lingenfelser, J. Wagner, and E. André. 2019. NOVA - A tool for eXplainable Cooperative Machine Learning. In *2019 8th int. Conference on Affective Computing and Intelligent Interaction (ACII)*. 109–115.

[8] Houtan Jebelli, Mohammad Mahdi Khalili, and SangHyun Lee. 2019. Mobile EEG-Based Workers' Stress Recognition by Applying Deep Neural Network. In *Advances in Informatics and Computing in Civil and Construction Engineering*, Ivan Mutis and Timo Hartmann (Eds.). Springer International Publishing, Cham, 173–180.

[9] Katja Langer, Oliver T Wolf, and Valerie L Jentsch. 2021. Delayed effects of acute stress on cognitive emotion regulation. *Psychoneuroendocrinology* 125 (2021), 105101.

[10] Mark Le Fevre, Jonathan Matheny, and Gregory S Kolt. 2003. Eustress, distress, and interpretation in occupational stress. *Journal of managerial psychology* (2003).

[11] Laura Leuchs, Max Schneider, Michael Czisch, and Victor I Spoormaker. 2017. Neural correlates of pupil dilation during human fear learning. *Neuroimage* 147 (2017), 186–197.

[12] B. Leuner and T.J. Shors. 2013. Stress, anxiety, and dendritic spines: What are the connections? *Neuroscience* 251 (2013), 108–119.

[13] Yifei Lu, Wei-Long Zheng, Binbin Li, and Bao-Liang Lu. 2015. Combining Eye Movements and EEG to Enhance Emotion Recognition.. In *IJCAI*, Vol. 15. Citeseer, 1170–1176.

[14] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. 2019. MediaPipe: A Framework for Building Perception Pipelines. arXiv:1906.08172 [cs.DC]

[15] Sarah B Lupis, Natalie J Sabik, and Jutta M Wolf. 2016. Role of shame and body esteem in cortisol stress responses. *Journal of behavioral medicine* 39, 2 (2016), 262–275.

[16] Jodi Oakman, Natasha Kinsman, Rwth Stuckey, Melissa Graham, and Victoria Weale. 2020. A rapid review of mental and physical health effects of working at home: how do we optimise health? *BMC Public Health* 20, 1 (2020), 1–13.

[17] Timo Partala and Veikko Surakka. 2003. Pupil size variation as an indication of affective processing. *Int. journal of human-computer studies* 59, 1-2 (2003), 185–198.

[18] Marco Pedrotti, Mohammad Ali Mirzaei, Adrien Tedesco, Jean-Rémy Chardonnet, Frédéric Mérienne, Simone Benedetto, and Thierry Baccino. 2014. Automatic stress classification with pupil diameter analysis. *Int. Jour. of Human-Computer Interaction* 30, 3 (2014), 220–236.

[19] Ebony R Samuels and Elemer Szabadi. 2008. Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part I: principles of functional organisation. *Current neuropharmacology* 6, 3 (2008), 235–253.

[20] Thiago Santini, Wolfgang Fuhl, and Enkelejda Kasneci. 2018. PuReST: Robust Pupil Tracking for Real-Time Pervasive Eye Tracking. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications* (Warsaw, Poland) *(ETRA '18)*. Association for Computing Machinery, New York, NY, USA, Article 61, 5 pages.

[21] Marco Schmidt, Michelle Berger, Lea Görl, Stefanie Lahmer, and Henner Gimpel. 2022. Towards a Mobile Coping Assistant. In *Proceedings of the 55th Hawaii International Conference on System Sciences (HICSS)*. Honolulu, HI : University of Hawai'i at Manoa, Hamilton Library.

[22] Philip Schmidt, Attila Reiss, Robert Dürichen, and Kristof Van Laerhoven. 2018. Wearable affect and stress recognition: A review. *CoRR* abs/1811.08854 (2018). arXiv:1811.08854

[23] Sylvain Sirois and Julie Brisson. 2014. Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science* 5, 6 (2014), 679–692.

[24] Renée M Visser, H Steven Scholte, Tinka Beemsterboer, and Merel Kindt. 2013. Neural pattern similarity predicts long-term fear memory. *Nature neuroscience* 16, 4 (2013), 388.

[25] Johannes Wagner, Florian Lingenfelser, Tobias Baur, Ionut Damian, Felix Kistler, and Elisabeth André. 2013. The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time. In *Proceedings of the 21st ACM int. conference on Multimedia*. ACM, 831–834.

[26] Adriana A Zekveld, Johanna AM van Scheepen, Niek J Versfeld, Enno CI Veerman, and Sophia E Kramer. 2019. Please try harder! The influence of hearing status and evaluative feedback during listening on the pupil dilation response, saliva-cortisol and saliva alpha-amylase levels. *Hearing research* 381 (2019), 107768.

[27] Lim Jia Zheng, James Mountstephens, and Jason Teo. 2020. Four-class emotion classification in virtual reality using pupillometry. *Journal of Big Data* 7, 1 (2020), 1–9.