Author(s): Sun, Zhaodong; Vedernikov, Alexander; Kykyri, Virpi-Liisa; Pohjola, Mikko; Nokia, Miriam; Li, Xiaobai

Title: Estimating Stress in Online Meetings by Remote Physiological Signal and Behavioral Features

Year: 2022

Version: Published version

Please cite the original version:

Sun, Z., Vedernikov, A., Kykyri, V.-L., Pohjola, M., Nokia, M., & Li, X. (2022). Estimating Stress in Online Meetings by Remote Physiological Signal and Behavioral Features. In UbiComp/ISWC '22 Adjunct : Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers (pp. 216-220). ACM. https://doi.org/10.1145/3544793.3563406

Estimating Stress in Online Meetings by Remote Physiological Signal and Behavioral Features

Zhaodong Sun
zhaodong.sun@oulu.fi
Center for Machine Vision and Signal
Analysis, University of Oulu
Oulu, Finland

Alexander Vedernikov
aleksandr.vedernikov@oulu.fi
Center for Machine Vision and Signal
Analysis, University of Oulu
Oulu, Finland

Virpi-Liisa Kykyri
virpi-liisa.e.kykyri@jyu.fi
Department of Psychology, University
of Jyväskylä
Jyväskylä, Finland

Mikko Pohjola
mikko.j.pohjola@jyu.fi
Department of Psychology, University
of Jyväskylä
Jyväskylä, Finland

Miriam Nokia
miriam.nokia@jyu.fi
Department of Psychology, University
of Jyväskylä
Jyväskylä, Finland

Xiaobai Li
xiaobai.li@oulu.fi
Center for Machine Vision and Signal
Analysis, University of Oulu
Oulu, Finland

ABSTRACT

Work stress impacts people's daily lives. Their well-being can be improved if the stress is monitored and addressed in time. Attaching physiological sensors are used for such stress monitoring and analysis. Such approach is feasible only when the person is physically presented. Due to the transfer of the life from offline to online, caused by the COVID-19 pandemic, remote stress measurement is of high importance. This study investigated the feasibility of estimating participants' stress levels based on remote physiological signal features (rPPG) and behavioral features (facial expression and motion) obtained from facial videos recorded during online video meetings. Remote physiological signal features provided higher accuracy of stress estimation (78.75%) as compared to those based on motion (70.00%) and facial expression (73.75%) features. Moreover, the fusion of behavioral and remote physiological signal features increased the accuracy of stress estimation up to 82.50%.

CCS CONCEPTS

• **Applied computing** → *Health care information systems.*

KEYWORDS

stress estimation, remote photoplethysmography, facial expression, head pose, eye gaze

1 INTRODUCTION

Nowadays, people are greatly affected by work stress. Early diagnosis and examination of stress levels can improve their well-being. Recent years have been marked by humans' life shift from offline to online mode due to the COVID-19 pandemic. Therefore, remote stress measurement is of high importance. Previous developments in Affective Computing (AC) made it possible to perform stress diagnostic remotely. Behavioral features (facial expression [6, 23], head motion [15, 16], and eye gaze [18, 19]) are common tools in AC for emotion recognition and stress estimation [6, 9, 23]. However, physiological signal features (e.g., heart rate variability) are more reliable than behavioral ones, as people might control or disguise behaviors, while physiological features are difficult to alter by will [8, 25]. Electrocardiography is the most accurate measure reflecting cardiac activity, although not feasible for online video meetings application. Photoplethysmography (PPG) is another measure that is used to examine cardiac activities closely related to humans' stress levels. There are contact PPG (cPPG) and remote PPG (rPPG). Although stress level estimation using cPPG has been previously studied [7], it is not a convenient tool for online video meetings. On the other hand, rPPG is an alternative as facial videos allow for tracking face color change induced by the blood volume variation [22].

Only three studies [11–13] have investigated the application of rPPG signal features for stress estimation. These articles are devoted to stress temporally imposed by experimental tasks, which differs from stress arising in a realistic environment. Moreover, the videos in these works were recorded in a laboratory environment with well-controlled light and setup. However, in realistic scenes, the participants are in various uncontrolled environments with different recording setups. Whether rPPG signal features (obtained from facial videos recorded in realistic scenes) can be utilized for stress estimation remains unknown. Moreover, in [11–13], facial expression and motion features have not been used for stress estimation. Research has not yet determined whether fusing additional features with rPPG signal features can impact stress estimation accuracy. Therefore, it is of high importance to address mentioned limitations for the further promotion of rPPG signal features application in a realistic environment.

Dataset of online video meetings was collected, and the obtained facial videos were further used for the stress estimation. Experiments were carried out to answer the following questions: 1) Is it feasible to use rPPG signals measured from facial videos for stress estimation? 2) What is the accuracy difference of stress estimation based on rPPG and cPPG features? 3) Does fusion of behavioral features (facial expression and motion) with physiological signal (rPPG or cPPG) features improve stress estimation?

2 METHOD

2.1 Dataset Collection

A dataset of 1.5-hours long online video meetings was collected. All in all, 24 online video meetings were recorded. From seven to nine participants and one consultant took part in each online video meeting. Totally, two consultants and 32 participants (29 females and three males) were involved. Participants are social services employees who work with clients with mental health problems and substance abuse. During an online video meeting, participants were asked to share their experiences and make reflections about their working daily life. The consultant aimed to guide the reflection process and help participants to find ways to reduce the strains. At the end of each session, participants completed a questionnaire providing feedback about the meeting and their feelings. Participants report their stress levels with a scale of 1 to 10, that 1 indicates the lowest and 10 indicates the highest stress level. The online video meeting is a simulation and reflection of participants' daily work, which can reveal their work stress levels. Therefore, self-reported work stress levels were used as labels. More details about data collection are shown in the auxiliary materials.

2.2 Feature Extraction

rPPG and behavioral (facial expression and motion) features were extracted from the recorded facial videos. cPPG features were extracted from pulse oximeter signals and used as a reference. Various combinations of the mentioned features were used for the work stress estimation. The method workflow is shown Fig. 1.

rPPG Features. The facial videos were processed by OpenFace library [2] for the face landmark detection. It allows to tackle head motion problems and track facial landmarks accurately, thus, solving the issue of uncontrolled environment recordings and participants' movements. Facial landmarks were used to define regions of interest (ROI) of bare skin areas (forehead, cheeks, and chin) and to average the pixel values in the ROI for each color channel for the raw RGB signals acquisition. Then, the Plane-Orthogonal-to-Skin (POS) algorithm [24] was used to extract the rPPG waveform from the raw RGB signals. Each facial video was divided into non-overlapped 5-min clips [17] to obtain 5-min rPPG segments. Then, the signal to noise ratio (SNR) [5] was calculated for each rPPG segment. The rPPG segments with SNR below 0.5 were removed to ensure that only high-quality rPPG segments were kept. Subsequently, heart rate variability (HRV) features were extracted by Neurokit2 library [10] from the rPPG segments by locating the systolic peaks and deriving the inter-beat interval (IBI) curves. 20 widely used HRV features (described in the auxiliary materials) in time, spectral and geometrical domains were utilized [17]. HRV feature vectors were calculated for 5-min rPPG segments. Finally,

the obtained vectors were averaged to get a 20-dimensional HRV feature vector for each facial video.

Facial Expression Features. Action units (AUs) were used to compute facial expression features [21]. The OpenFace library [2] was utilized to detect and track 16 key types of AUs (mentioned in the auxiliary materials) on each video frame. The chosen list covers the most frequently occurring AUs related to emotions as per Facial Action Coding System (FACS) Investigator's Guide [1]. The intensity of AUs, measured from 0 and 5, was used as an output (Fig. 1). The mean, median, standard deviation, minimum, maximum, and range along the time dimension of the AUs were computed [3]. Finally, a 96-dimensional facial expression feature vector was obtained for each facial video.

Motion Features. The eye gaze and head pose were extracted using OpenFace library [2]. Pitch and yaw of eye gaze as well as pitch, yaw, and roll of head pose, were extracted (Fig. 1). The mean value of each feature along the time dimension was subtracted from the original feature signal to offset the different camera positions. Then, the median, standard deviation, minimum, maximum, and range of each motion feature along the time dimension were calculated to yield the final statistical features [3]. Finally, a 25-dimensional motion feature vector was obtained for each facial video.

cPPG Features. Like rPPG, each cPPG signal was divided into non-overlapped 5-min cPPG segments. Then, the SNR was calculated for each 5-min cPPG segment. The cPPG segments with SNR below 0.5 were removed to ensure that only high-quality cPPG segments were kept [5]. All in all, 20 HRV features, the same as used for rPPG, were extracted from 5-min cPPG segments and averaged. Finally, a 20-dimensional cPPG feature vector was obtained for each cPPG signal.

2.3 Stress Classification

Stress classification was based on logistic regression which suits better for tasks with small amount of training data. A threshold value, separating the reported values of stress levels into low-stress and high-stress groups, was used to get the binary stress labels [13]. The threshold value of 7 was utilized, which is the median of the obtained stress levels among the participants. The classification scores from different classifiers were averaged via decision-level fusion to get the final classification results (Fig. 1).

3 RESULTS AND DISCUSSION

rPPG Measurement Results. The average HR was computed for each 30-sec rPPG clip and compared with the corresponding HR obtained from cPPG. The Mean Absolute Error (MAE) between HR from rPPG and cPPG was the evaluation metric [4, 20, 26]. MAE of 5.13 bpm for the entire dataset was obtained, which is very promising considering the uncontrolled facial video recordings during online video meetings. A single example of HR curves computed from both rPPG and cPPG signals is shown in Fig. 2(b). For the sake of comparison, the SOTA performance [13] on data recorded in controlled labs achieved similar MAE values of 3.55, 9.26, and 5.99 bpm for three different settings.

Stress Classification Results. A 10-fold subject-independent cross-validation protocol was used. The participants were divided
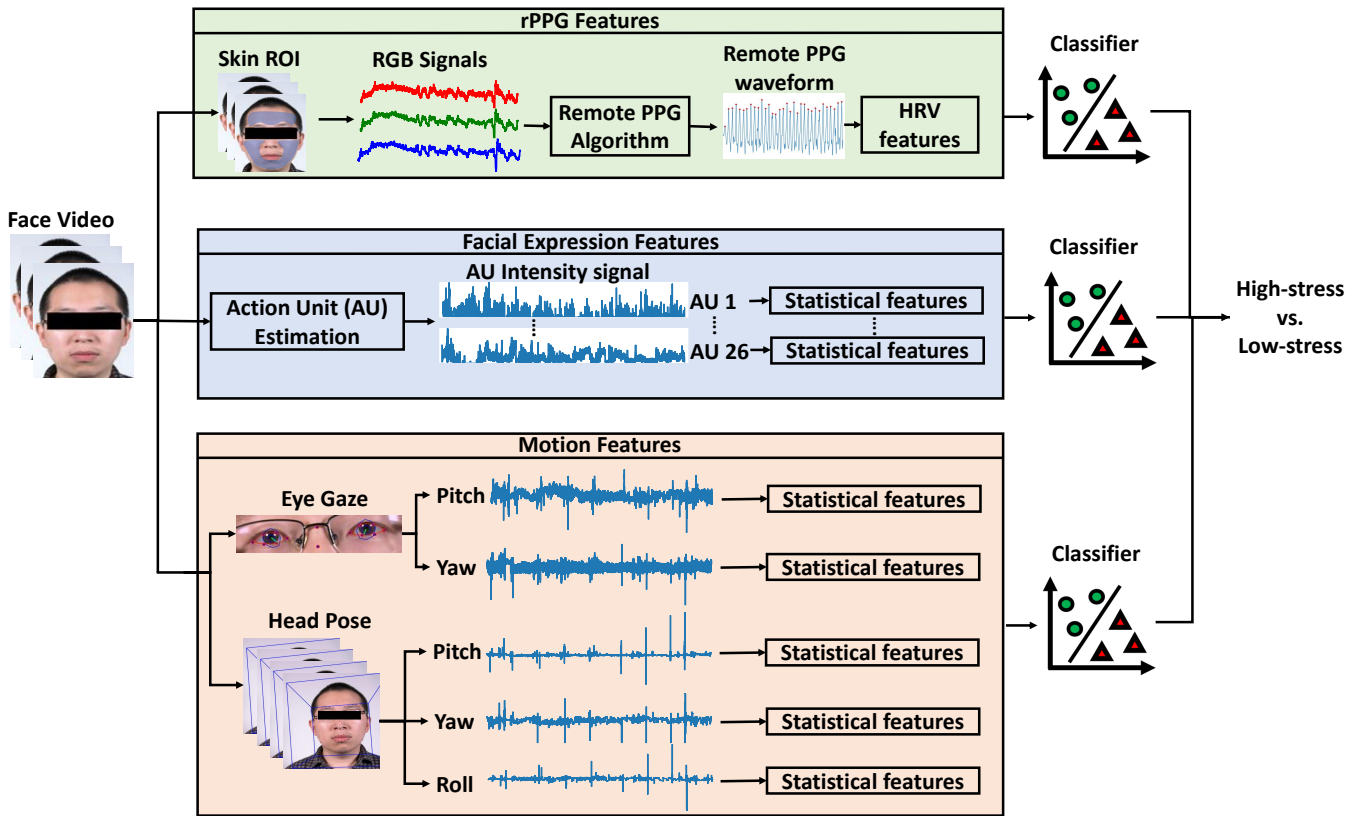
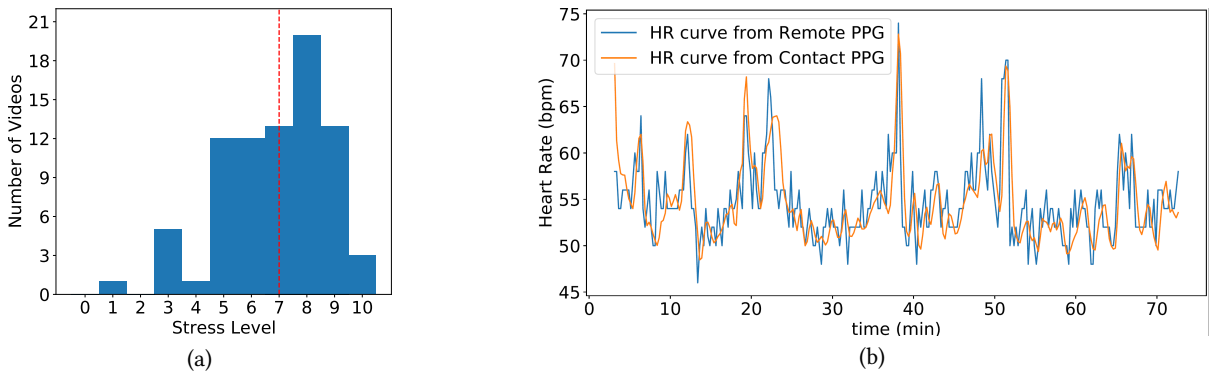Figure 1: Stress estimation from facial videos.

(a)

(b)

Figure 2: Stress classification results: (a) Histogram of reported stress levels (the red dotted line corresponds to the threshold); (b) HR curves computed from rPPG and cPPG signals.

into ten groups, each including about an equal number of participants. Nine groups were used for training and one group for testing and rotation. The accuracy and the area under the curve (AUC) were used as evaluation metrics. The performance of the stress classification algorithm was first evaluated based on single rPPG, cPPG, motion, and facial expression features, then the fusion of the features was performed. All in all, 36 high-stress and 44 low-stress samples were obtained during this experiment (Fig. 2a). The results

of a stress classification task based on a single set of features are presented in the first column of Table 1. It shows that stress levels predicted with cPPG had the highest accuracy (81.25%) and AUC (0.80). The performance of the stress classification task based on rPPG features was slightly lower and achieved 78.75% accuracy and 0.78 AUC. The facial expression features provided 73.75% accuracy and 0.71 AUC. Finally, motion features resulted in the worst performance and led to 70.00% accuracy and 0.70 AUC. The feature fusion

Table 1: Stress classification results for a single set of features (left), fusion of rPPG with facial expression and motion features (center), and fusion of cPPG with facial expression and motion features (right).

Features	Acc (%)	AUC
Motion	70.00	0.70
Facial Expression	73.75	0.71
cPPG	**81.25**	**0.80**
rPPG	78.75	0.78

Features	Acc (%)	AUC
rPPG	78.75	0.78
rPPG+Motion	78.75	0.78
rPPG+ Facial Expression	81.25	0.81
rPPG+Motion +Facial Expression	**82.50**	**0.81**

Features	Acc (%)	AUC
cPPG	81.25	0.80
cPPG+Motion	81.25	0.80
cPPG+ Facial Expression	83.75	0.82
cPPG+Motion +Facial Expression	**85.00**	**0.83**

results are shown in the second and the third columns of Table 1. Fusion of rPPG, motion, and facial expression features resulted in slightly increased accuracy and AUC. The same phenomenon was observed when fusing cPPG with motion and facial expression features.

Is it feasible to use rPPG signals measured from facial videos for stress estimation? The accuracy of stress estimation based on rPPG (78.75%) features surpasses those based on motion (70.00%) and facial expression (73.75%) features. Stress estimation based on motion and facial expression features is less accurate for two reasons. Firstly, participants show neutral facial expressions and small motions during the online video meeting, which makes the facial expressions and motions subtle, short, and sparse [14]. Secondly, the facial expressions and motions in the stress state can be self-controlled while the physiological signals cannot [8, 25]

What is the accuracy difference of stress estimation based on rPPG and cPPG features? The accuracy of stress estimation based on rPPG (78.75%) features is close to those based on cPPG (81.25%) features. However, rPPG signals are more convenient for stress estimation as only facial videos are typically available in online video meetings. Using of cPPG signals for stress estimation requires pulse oximeters, which is hard to implement in practice.

Does fusion of behavioral features (facial expression and motion) with physiological signal (rPPG or cPPG) features improve stress estimation? The accuracy of stress estimation can be slightly improved by fusing behavioral features with physiological signal features (Table 1 center and right columns). Although the behavioral features are less effective for stress estimation if used separately, it is still beneficial to perform fusion as different features might complement each other for the stress estimation.

4 CONCLUSION

Stress estimation has been performed based on remote physiological signal features (rPPG) and behavioral features (facial expression and motion) obtained from facial videos recorded during online video meetings. The accuracy of stress estimation based on remote physiological signal features was higher than those based on behavioral features. The fusion of remote physiological signal and behavioral features increases the accuracy of stress estimation.

ACKNOWLEDGMENTS

REFERENCES

[1] [n.d.]. Facial Action Coding System. https://www.paulekman.com/facial-action-coding-system/. Accessed: 2022-08-24.
[2] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. IEEE, 59–66.
[3] Nigel Bosch and Sidney D'Mello. 2019. Automatic detection of mind wandering from video in the lab and in the classroom. *IEEE Transactions on Affective Computing* (2019).
[4] Weixuan Chen and Daniel McDuff. 2018. Deepphys: Video-based physiological measurement using convolutional attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 349–365.
[5] Gerard De Haan and Vincent Jeanne. 2013. Robust pulse rate from chrominance-based rPPG. *IEEE Transactions on Biomedical Engineering* 60, 10 (2013), 2878–2886.
[6] Hua Gao, Anil Yüce, and Jean-Philippe Thiran. 2014. Detecting emotional stress from facial expressions for driving safety. In *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 5961–5965.
[7] Giorgos Giannakakis, Dimitris Grigoriadis, Katerina Giannakaki, Olympia Simantiraki, Alexandros Roniotis, and Manolis Tsiknakis. 2022. Review on Psychological Stress Detection Using Biosignals. *IEEE Transactions on Affective Computing* 13, 1 (2022), 440–460. https://doi.org/10.1109/TAFFC.2019.2927337
[8] Jukka Kortelainen, Suvi Tiinanen, Xiaohua Huang, Xiaobai Li, Seppo Laukka, Matti Pietikäinen, and Tapio Seppänen. 2012. Multimodal emotion recognition by combining physiological signals and facial expressions: a preliminary study. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 5238–5241.
[9] Hitoshi Kusano, Yuji Horiguchi, Yukino Baba, and Hisashi Kashima. 2020. Stress Prediction from Head Motion. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 488–495.
[10] Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and S. H. Annabel Chen. 2021. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods* 53, 4 (feb 2021), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y
[11] Daniel McDuff, Sarah Gontarek, and Rosalind Picard. 2014. Remote measurement of cognitive stress via heart rate variability. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2957–2960.
[12] Daniel McDuff, Sarah Gontarek, and Rosalind W. Picard. 2014. Improvements in Remote Cardiopulmonary Measurement Using a Five Band Digital Camera. *IEEE Transactions on Biomedical Engineering* 61, 10 (2014), 2593–2601. https://doi.org/10.1109/TBME.2014.2323695
[13] Rita Meziatisabour, Yannick Benezeth, Pierre De Oliveira, Julien Chappe, and Fan Yang. 2021. UBFC-Phys: A Multimodal Database For Psychophysiological Studies Of Social Stress. *IEEE Transactions on Affective Computing* (2021).
[14] Rajitha Navarathna, Peter Carr, Patrick Lucey, and Iain Matthews. 2017. Estimating audience engagement to predict movie ratings. *IEEE Transactions on Affective Computing* 10, 1 (2017), 48–59.
[15] Atanu Samanta and Tanaya Guha. 2017. On the role of head motion in affective expression. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2886–2890.
[16] Atanu Samanta and Tanaya Guha. 2020. Emotion sensing from head motion capture. *IEEE Sensors Journal* 21, 4 (2020), 5035–5043.
[17] Fred Shaffer and Jay P Ginsberg. 2017. An overview of heart rate variability metrics and norms. *Frontiers in public health* (2017), 258.
[18] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. 2011. A multimodal database for affect recognition and implicit tagging. *IEEE transactions*

on *affective computing* 3, 1 (2011), 42–55.

[19] Mohammad Soleymani, Maja Pantic, and Thierry Pun. 2011. Multimodal emotion recognition in response to videos. *IEEE transactions on affective computing* 3, 2 (2011), 211–223.

[20] Zhaodong Sun and Xiaobai Li. 2022. Contrast-Phys: Unsupervised Video-based Remote Physiological Measurement via Spatiotemporal Contrast. *arXiv preprint arXiv:2208.04378* (2022).

[21] Y-I Tian, Takeo Kanade, and Jeffrey F Cohn. 2001. Recognizing action units for facial expression analysis. *IEEE Transactions on pattern analysis and machine intelligence* 23, 2 (2001), 97–115.

[22] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. 2008. Remote plethysmographic imaging using ambient light. *Optics express* 16, 26 (2008), 21434–21445.

[23] Carla Viegas, Shing-Hon Lau, Roy Maxion, and Alexander Hauptmann. 2018. Distinction of stress and non-stress tasks using facial action units. In *Proceedings of the 20th international conference on multimodal interaction: Adjunct.* 1–6.

[24] Wenjin Wang, Albertus C den Brinker, Sander Stuijk, and Gerard De Haan. 2016. Algorithmic principles of remote PPG. *IEEE Transactions on Biomedical Engineering* 64, 7 (2016), 1479–1491.

[25] Zitong Yu, Xiaobai Li, and Guoying Zhao. 2021. Facial-Video-Based Physiological Signal Measurement: Recent advances and affective applications. *IEEE Signal Processing Magazine* 38, 6 (2021), 50–58.

[26] Zitong Yu, Wei Peng, Xiaobai Li, Xiaopeng Hong, and Guoying Zhao. 2019. Remote heart rate measurement from highly compressed facial videos: an end-to-end deep learning solution with video enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 151–160.