2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Simultaneous estimation of drowsiness, focus and stress from facial videos captured by smartphones

Terumi Umematsu, Masanori Tsujikawa, and Hideyuki Sawada

***Abstract*— We demonstrate simultaneous estimation of drowsiness from blinking, focus from gaze, and stress from heart rate, by using facial videos captured by a smartphone. The three mental states are estimated in parallel with parallel processing each frame, and we confirm that the overall system are operated in real time with processing delay of only 100-200[ms].**

I. INTRODUCTION

In recent years, there has been growing interest in the non-contact, simpler, and more convenient measurement of psychological states for mental health care and checking mental states in a daily life. Conventionally, relatively low-burden methods estimate focus from a glasses-type sensor [1] and drowsiness from a camera [2]. To the best of our knowledge, there is no existing method that can easily measure multiple mental states using a non-contact sensor such as smartphone camera.

In this paper, we demonstrate a system that simultaneously estimates drowsiness from blink, focus from gaze, and stress from heart rate, only using facial videos taken by a smartphone. We also demonstrate that this simultaneous estimation of these three mental states work in real time on the smartphone using parallel processing. This system suits to be used to check mental states daily and to understand the relationship among multiple mental states.

II. MENTAL STATE ESTIMATIONS AND DEVELOPED SYSTEM

We have developed a facial video-based system consisting of three parts: drowsiness estimation from blink, focus estimation from gaze, and stress estimation from heart rates. Figure 1 shows the overview of the system. Each estimation method is described below.

[Drowsiness estimation] The drowsiness level is estimated from facial videos using eyelid variability features [2]. As eyelid variability involves slow motions, drowsiness can be accurately estimated even using low frame rates. The correlation between the ground truth and the estimated score was shown 0.53 at 3 fps and 0.55 at 30 fps.

[Focus estimation] The focus level at each time *t* shown by Eq. (1), which is defined by the duration of watching the screen, is calculated based on the number of frames in which the face is located inside the defined region, and the gaze and head motion similar to the method described in [3]. Here, *gaze*(*t*) is the average of angle at which the user is looking at, and the *head motion*(*t*) is the average of the pan, tilt, and roll angles of a head.

$$Focus\ level(t) = framein(t) \times \{1 - (w_1 * head\ motion(t) + w_2 * gaze(t))\} \quad (1)$$

$$w_1, w_2: weight$$

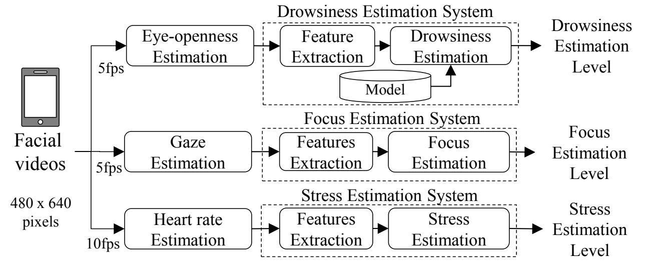

Figure 1. System overview: estimation of drowsiness, focus, and stress

[Stress Estimation] Pulse wave is extracted from the facial videos, and the heart rate is estimated from the pulse wave. Further, stress level is calculated from the normalized standard deviation of R-R intervals (RRI) given by Eq. (2). RRI is calculated by the heart rate estimation method, which is robust to head motion by referring to the fluctuations of facial feature points [4]. The estimation accuracy was reported as the root mean square error of 7.82[bpm] using the benchmark data.

$$Stress\ level(t) = 1 - (\overline{\mathrm{std.}(\mathrm{RRI}(t))}) \quad (2)$$

III. EVALUATION

We confirmed that the estimated levels of drowsiness, focus, and stress, all changed properly as expected, when a person felt drowsier, less focus, and more stressed during the math task. We implemented the system on a smartphone and confirmed its successful real-time simultaneous operation. To confirm with inexpensive smartphone, we used ASUS Zenfone Max Pro (M2) with Qualcomm Snapdragon 660 CPU, 6GB memory, Android 8.1 OS, and 13 million pixels front camera. All images were resampled to a size of 480x640 pixels. The drowsiness and focus estimations were performed at 5 fps, and the stress estimation was performed at 10 fps.

The image processing time was 200-300[ms] per 1 frame. By estimating three mental states in parallel, and by image processing each frame in parallel, we confirmed that the overall system operated in real time with only a constant processing delay of 100-200[ms]. In the future work, we will collect more videos under various conditions, evaluate estimation accuracy, and improve the accuracy for applications that can be used in daily life.

REFERENCES

[1] Japan Patent, JINS HOLDINGS, Eyewear, P2018-209615.
[2] Masanori Tsujikawa et al., "Drowsiness Estimation from Low-Frame-Rate Facial Videos using Eyelid Variability Features," EMBC, pp.5203-5206, 2018.
[3] Komiya Rinko et al., "Image-based Attention Level Estimation of Interaction Scene by Head Pose and Gaze Information", IEEE ACIS, pp.497-501, 2018.
[4] Utkarsh Sharma et al., "Adaptive Heart Rate Estimation from Face Videos", EMBC, 2018.

*Research supported by NEC Corporation.
Terumi Umematsu and Masanori Tsujikawa are with Biometrics Research Laboratories, NEC Corporation, 1753, Shimonumabe, Nakahara-Ku, Kawasaki, Kanagawa 211-8666, Japan (e-mail: terumi@nec.com, tujikawa@nec.com). Hideyuki Sawada is with Department of Applied Physics, Faculty of Science and Engineering, Waseda University, Japan (e-mail: sawada@waseda.jp). Terumi Umematsu is also with Waseda University.