Can Eye Gaze Improve Emotional State Detection on Off the Shelf Smart Devices

Jiwan Kim
Interactions Lab, UNIST
Ulsan, Republic of Korea
kjwan4435@gmail.com

Doyoung Lee
Digital Healthcare center, Hanyang University
Seoul, Republic of Korea
dylee.hci@gmail.com

Ian Oakley
Interactions Lab, UNIST
Ulsan, Republic of Korea
ian.r.oakley@gmail.com

***Abstract*—Smartphones and wearable technology have enabled new forms of digital healthcare thanks to the rich set of sensors and their data. Digital phenotyping is one typical approach in this area that primarily uses smartphone data to detect or recognize cognitive, behavioral, or affective states and traits. It is generally based on data from sensors (e.g., inertial or touch sensors), activity or use logs, and analysis of user-generated content. While these data sources are valuable, we argue they do not leverage the full potential of modern smart devices. In this paper, we propose an eye gaze feature for a new digital biomarker using the advanced capabilities of off-the-shelf smart devices that can support affect detection. To detect emotional state change, since Social Network Services(SNSs) evoke diverse affective experiences, we plan to conduct two studies in the Instagram use scenario. The first study improved our understanding of gaze features and explores the value of gaze for affect detection in a controlled setting. We will conduct the second study for a long-term period in-the-wild settings to assess our approach with larger populations. We believe that extending the scope of affective digital phenotyping to include these new sensing modalities will increase the reliability, and robustness, of the emotional state detection it enables.**

***Index Terms*—digital phenotyping, smartphones, gaze**

I. Introduction

Digital phenotyping, defined as the "moment-by-moment, in situ quantification of the individual-level human phenotype using data from personal digital devices" [1], has been used to detect or predict individual's inner states such as cognitive traits, affective state, mood disorder, and stress. This trend is accelerating in recent years with remarkable advances in current mobile sensing technology. For example, much research has explored different types of digital bio-markers [2] via data obtained from multiple sensor channels of mobile devices. These include GPS location, phone usage pattern, social media log, touch screen input, device motion, and voice. In this paper, we present an eye-gaze, a new digital bio-marker based on a newly emerging mobile sensing channel.

Eye gaze already has been recognized as a promising human bio-marker revealing the inner state of the user[ref]. Especially, many investigations have reported the relevance between emotion and the features with eye gaze such as pupil position [3], fixation duration [4], eye motion speed [5]. Although these studies have provided hints on the importance of gaze data in emotion recognition, mostly rely on high-end eye trackers [6] and have only conducted laboratory-scale experiments, leaving questions about whether these relationships are applicable in the real world. Thanks to the new capabilities of iOS devices with built-in TrueDepth [7] cameras, current mobiles can create the depth map of the user's face and it enables researchers to explore novel interactions using gaze features [8]. In this paper, we extend the strong potential of the gaze to detect emotional states in the real world by adapting Instagram use scenarios. To achieve this, we captured various sensor data (e.g., IMU, touch, head motion, face, image label, eye gaze) from an off-the-shelf smartphone, and built the affective state detection model with multi-modal machine learning.

II. Study

The study consists of two parts: studies in controlled and field settings. This paper includes a brief introduction to collecting sensor data (including gaze) from an off-the-shelf mobile device and future plan for the field study.

A. Lab Study

In the lab study, we sought to detect participants' momentary affective states in a controlled setting. While the study, participants were asked to use Instagram and we captured various sensor data using iPhone 12 pro, and then we labeled the data with participants' affective state via a simple survey.

1) Participants: 24 participants (12 male, mean age of 23.83 (SD 2.92)) were recruited from the local university through online community channels. We screened participants for their ownership of an iPhone to access the TrueDepth camera and also for their regular Instagram usage: at least 30 minutes of use on a daily basis. To prevent participants from consuming all the new posts on their feeds, we asked them not to use Instagram from 12 hours before the study.

2) Design: Figure 1 illustrates the overall procedure for this lab study. Participants first read the instructions and completes PHQ-9 survey [9] to measure the brief depression severity. To ensure gaze data are being collected by the TrueDepth camera, participants were asked to hold their smartphones with the front camera facing their faces during the experiment. For the data collection system, we built an Instagram web-wrapping application that can log any mobile sensor data at 60Hz. The study consists of 5 real Instagram sessions and 6 emotion induction sessions. Firstly, in the real Instagram session, participants signed in to Instagram wrapper with their own accounts and consumed their feeds. In this session,

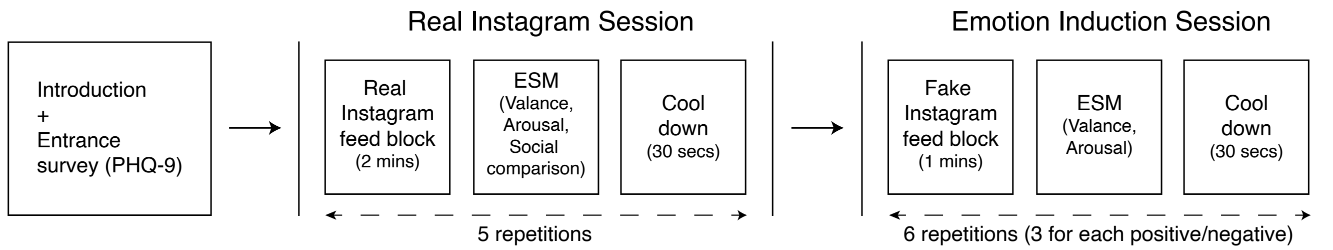

Fig. 1. Procedure for the lab study: During the real Instagram session, participants consumed their own feed, on the other hand, in the emotion induction session, they explored fake feeds consisting of positive or negative image sets 3 times each. After every Instagram feed block, participants are asked to self-report their feelings and have 30 secs cool-down by showing them a neutral image.

participants were allowed to browse only feed (e.g., newsfeed and individual feed), but no story or direct message to restrict users' interaction diversity. In the emotion induction session, we induced participants' emotions by showing them fake Instagram feeds consisting of positive or negative images three times each. To quantify participants' emotions, we conducted a simple survey after every Instagram feed block. Each survey includes valence and arousal for both real and fake Instagram sessions, and social and appearance comparisons for the real Instagram session only. After each survey, a neutral image appears on the screen to mitigate the effect of current feelings to the next session.

3) Data Collection: We collected sensor data at 60Hz for the gaze, device motion (3 axes for each motion-related feature including acceleration, attitude, rotation, and orientation), touch (position and radius), and facial expression (52 movements of specific facial features for eyes, mouth, jaw, eyebrows, cheeks, nose, and tongue). Additionally, we also recorded image labels on feed images using Google ML kit whenever the user made a touch. This enables us to collect a few explaining words (out of 400 words) for each feed image without privacy issues. In total, we ran 11 sessions for each participant, resulting in 264 sessions of sensor data. We captured sensor data at 60Hz during the experiment, on average, we collected about 63646 (SD 30) device-motion data, 44617 (SD 12325) face and gaze data, and 30776 (SD: 9289) touch data, and 1080 (SD 242) labels of feed images.

4) Data Preprocessing: For preprocessing, we merged all the data captured in different modalities using interpolation based on timestamp and conducted zero padding for missing data such as touch data during no touch. We then divided the merged data into windows of 2-seconds with 1-second overlapping. For each window, we calculated features in the form of summary statistics (mean, median, minimum, maximum, standard deviation, and variance).

5) Emotional Status Prediction: In this study, we sought to predict participants' emotional status using a simple Random Forest classifier. Since there are substantial individual difference variations in affective experience, we built 24 individual models for each user. Due to the face recognition problem, we fail to capture the gaze and face data for one participant (P3), and partially lost data for 5 participants (P2, P10, P17, P21, P23). Our model predicts participants' affective state with 76% accuracy and 73% F1 scores in emotion induction sessions. Now, we plan to conduct gaze feature engineering (e.g., fixation, saccade), and predict emotional status in real Instagram sessions using an improved model.

III. Future Plan

In the near future, we will evaluate our affect detection system over two weeks in field settings. Additionally, we will design and evaluate interventions such as behavior recommendations that use outcomes of digital phenotyping models. For example, increased awareness of affective states may support better mental well-being and alleviate stress, anxiety and non-clinical depression.

In conclusion, we argue that gaze features available on smart devices will improve the quality of predictions to monitor affective states. Our system will be able to assess user's emotional states during daily life activities, a capability we suggest has strong potential to improve both individual and societal well-being.

References

[1] K. Huckvale, S. Venkatesh, and H. Christensen, "Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety," *npj Digital Medicine*, vol. 2, p. 88, Sep 2019.

[2] P. Dagum, "Digital biomarkers of cognitive function," *npj Digital Medicine*, vol. 1, p. 10, Mar 2018.

[3] C. Aracena, S. Basterrech, V. Snáel, and J. Velásquez, "Neural networks for emotion recognition based on eye tracking data," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2632–2637, IEEE, 2015.

[4] V. Tsang, "Eye-tracking study on facial emotion recognition tasks in individuals with high-functioning autism spectrum disorders," *Autism*, vol. 22, no. 2, pp. 161–170, 2018.

[5] V. Raudonis, G. Dervinis, A. Vilkauskas, A. Paulauskaitė-Tarasevičienė, and G. Keršulytė-Raudonė, "Evaluation of human emotion from eye motions," *International journal of advanced computer science and applications*, vol. 4, no. 8, pp. 79–84, 2013.

[6] S. Alghowinem, M. AlShehri, R. Goecke, and M. Wagner, *Exploring Eye Activity as an Indication of Emotional States Using an Eye-Tracking Sensor*, pp. 261–276. Cham: Springer International Publishing, 2014.

[7] Apple, "Truedepth camera." https://support.apple.com/en-us/HT208108, 2021. Online;.

[8] A. Kong, K. Ahuja, M. Goel, and C. Harrison, "Eyemu interactions: Gaze+ imu gestures on mobile devices," in *Proceedings of the 2021 International Conference on Multimodal Interaction*, pp. 577–585, 2021.

[9] K. Kroenke, R. L. Spitzer, and J. B. Williams, "The phq-9: validity of a brief depression severity measure," *Journal of general internal medicine*, vol. 16, no. 9, pp. 606–613, 2001.