What You Wear Know How You Feel: An Emotion Inference System with Multi-modal Wearable Devices

Dan Wang[†], Haibo Lei[†], Haozhi Dong, Yunshu Wang, Yongpan Zou, Kaishun Wu
{wangdan9,leihaibo2019,2018152044}@email.szu.edu.cn;{yongpan,wu}@szu.edu.cn
College of Computer Science and Software engineering, Shenzhen University

ABSTRACT

Emotions show high significance on human health. Automatic emotion recognition is helpful for monitoring psychological disorders, mental problems and exploring behavioral mechanisms. Existing approaches adopt costly and bulky specialized hardware such as EEG/ECG helmet, possess privacy risks, or with low accuracy and user experience. With the increasing popularity of wearables, people tend to equip multiple smart devices, which provides potential opportunity for emotion perception. In this paper, we present a pervasive and portable system called MW-Emotion to recognize common emotional states with multi-modal wearable devices. However, ubiquitous wearable devices perceive shallow information which is not obviously related to human emotions. MW-Emotion excavates intrinsic mapping relationship between emotions and sensing data. Our experiments show that MW-Emotion can recognize different emotion states with a relatively high accuracy of 83.1%.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing**.

KEYWORDS

Wearable devices; Emotion recognition; Multimodal data; Ubiquitous computing

ACM Reference Format:
Dan Wang[†], Haibo Lei[†], Haozhi Dong, Yunshu Wang, Yongpan Zou, Kaishun Wu. 2020. What You Wear Know How You Feel: An Emotion Inference System with Multi-modal Wearable Devices. In *MobiCom 2020 (MobiCom'20), September 21–25, 2020, London, United Kingdom.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3372224.3418161

1 INTRODUCTION

Nowadays, many people suffer from psychological problems such as depression, anxiety, autism and emotional lability due to the increasing life pressure. According to the China national mental health development report (2017-2018), 48% of Chinese believe that people had serious psychological problems.. However, due to the lack of counseling services, people are more inclined to

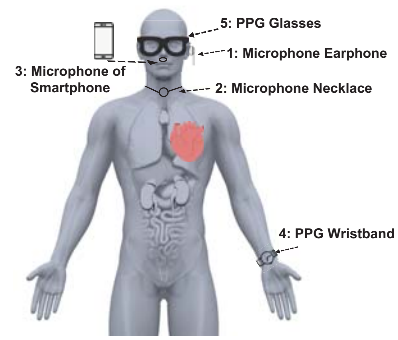


Figure 1: Different wearable devices used in MW-Emotion

relieve psychological pressure through self-regulation. The first step of self-regulation is to obtain good knowledge of one's own emotion states timely. Aiming at this goal, researchers have come up with different techniques to infer human emotions via analyzing facial expressions, walking gait, body postures, linguistic prosodic, smartphone usage, and physiological data (like EMG, ECG, ECG and *etc.*). Nevertheless, these methods posses shortcomings of privacy risks, cumbersome and specialized hardware, low accuracy and user experience.

To overcome the above shortcomings, we propose an automatic and continuous emotion monitoring system called MW-Emotion which infers a user's emotional states using multi-modal commercial wearable devices such as smartwatch, smart glasses and smart earbuds as shown in Fig. 1. This system can infer four common human emotions including happiness, neutrality, sadness and a mixed emotion [1]. Compared with existing methods, it has advantages of preserving privacy, using pervasive devices, achieving accurate recognition and being ease of use in daily life. The rationale of MW-Emotion is the underlying relevance between emotions and physiologic activities such as breathing, heart beating and other kinds of body sounds. However, inferring emotions with these ubiquitous wearable devices has two key challenges. One is that data collected by MW-Emotion demostrates week relevance with human emotions, which brings about great challenges on mining the hidden emotions. The other is that due to the precision of commercial sensors, signals collected by them are usually rather noisy which blurs the relationship between emotion states and sensing data. We handle these challenges by designing appropriate data quality enhancement techniques and deep learning network. We conduct real-world experiments to evaluate MW-Emotion and proves its

[†] indicates equal contribution.

[1] As fear and disgust are very similar, we categorize them into the same group in our work.

MobiCom'20, September 21–25, 2020, London, United Kingdom

ACM ISBN 978-1-4503-7085-1/20/09…$15.00
https://doi.org/10.1145/3372224.3418161

favorable performance. To the best of our knowledge, MW-Emotion is the first work to demonstrate the feasibility to recognize emotion state using low-cost, pervasive and portable wearable devices.

2 SYSTEM DESIGN

MW-Emotion recognizes emotional states only using pervasive wearable devices, which consists of multi-modal sensing, data quality enhancement and emotion recognition.

2.1 Multi-modal sensing signal design

The locations and types of sensing data used in MW-Emotion are as shown in Fig. 1. The device in location 1 is a in-ear earphone with a tiny microphone embedded which forms a resonant chamber inside the ear and amplify sounds of heart beating. The sensors installed in location 2 and location 3 are microphones which are used to collect the tracheal sound and nose sound respectively. Respiratory sound as well as other body sounds can be captured by these two microphones. The sensors installed in location 4 and location 5 are photoelectric volumetric pulse sensors. These physiological signals are closely related to human emotions and possess the potential to reflect them.

2.2 Data Quality Enhancement

In the preprocessing stage, we apply denoising and normalization on multi-modal raw signals and reliability evaluation on self-report labels.

2.2.1 ***Multi-modal Signal Preprocessing.*** After collecting multi-modal raw signals, we first subtract the average value of the signals of each channel to remove their DC components. As for the five kinds of signals, we utilize different signal processing methods according to their types and acquisition positions. But all these methods include two same steps, namely, noise filtering and outliers removal. Raw signals are collected with a 1000 Hz sampling rate.

Power line interference: PLI comes from nearby electrical appliances, which are powered by AC current of 50 Hz. Moreover, PLI produces old harmonic components in different frequency. In order to effectively remove this interference, the five channel signals after FFT transform are truncated with the stop bands set to [49, 51] Hz, [149, 152] Hz and [249, 251] Hz, then restore the signal by IFFT.

Preprocessing for different signals: As for the five kinds of signals, we adopt different signal processing methods because of the different signal types and acquisition positions.

- Location 1: Since the main frequency of heartbeat signal is in [1, 6] Hz, we filter this channel using band-pass FIR filter with [0.1, 10] Hz to ensure information redundancy.
- Location 2/3: The throat and nose channels mainly detect breathing information, including the respiratory rate and depth. They also record non-speech body sounds such as laughter and crying.
- Location 4/5: For the PPG signals, we first filter them with band-pass FIR filter with [0.1, 10] Hz. However, PPG signals are affected by contact between sensor and skin and whether sweating or not. Long time experiments lead to signal stable and low-frequency drifts. To eliminate such interference, we directly band stop the components less than 0.5 Hz after FFT transform, and then restore the signal by IFFT.

Normalization: Due to variations in quantization standards for different signals, normalization is necessary. Each channel signal is soft-normalized using the 5th and 95th quantiles.

2.2.2 ***Label Evaluation.*** As there exists errors in self-report labels, for example, a subject may be distracted and mistake the emotion, we only select the sessions with self-report labels in accord with half the people's for each stimulus clip. Furthermore, we only choose sessions with arousal or valence less than 3 or greater than 7 to ensure that it successfully arouse a certain emotion of subjects.

2.3 Emotion Recognition

After the above procedures, we adopt DenseNet as the recognition model. To improve its classification capability, we make every layer of information fully utilized. We transform each channel signal to time-frequency spectrum using short-time Fourier transform (STFT) and then feed them into the model.

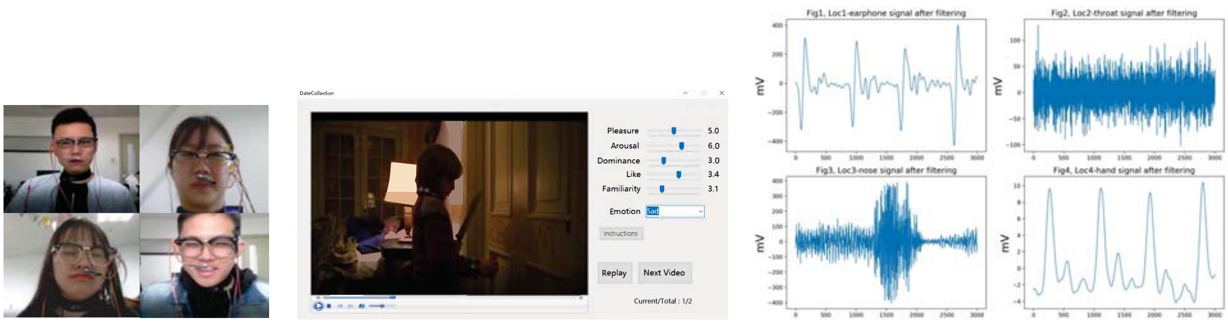
3 EXPERIMENT

3.1 Experiment Setup

We implement a software on a windows platform for collecting experimental data. Fig. 2 shows its screenshots and experiments scenarios. It enables an experimenter report her emotions by choosing emotional labels and entering scores on each scale. It also displays statistics of wearable sensing data. We conduct experiments in an indoor environment with 109 volunteers including 78 males and 31 females aging from 17 to 28. The data are collected by five wearable devices on different locations as shown in Fig. 2(a). In our experiments, we utilize four public emotion stimulation datasets, namely StimFilm[6], EMDB[3], IADS[2] and SEED[4], which contain videos and audios widely used for inducing certain kinds of emotions in psychological experiments. Before showing each stimulus material with the software, experimenters have one minute to calm down. Then they carefully watch videos or listen to audios, and accomplish the self assessment by rating their true feelings in terms of arousal, valence, and dominance with scores from 1 to 9 using the self-assessment manikin(SAM) [1] tool. Moreover, they also need to report the discrete labels of their emotions that are most close to their true feelings at that time.

3.2 Evaluation

We have tested different five different learning models including LSTM, DFCNN, MobileNet, ResNet and DenseNet. Fig. 3 shows the 5-fold cross validation overall performance of MW-Emotion with different models. We can see that DenseNet performs the best with precision and recall reaching 83.0% and 83.1% respectively. Fig. 4 shows the confusion matrix achieved by DenseNet. We observe that other emotions are easily confused with neutrality, which may be caused by the actual aroused emotional intensity. In addition, MW-Emotion's recognition accuracy is slightly lower than EQ-Radio [7] (*i.e.*, 87%) which trains and tests the model with data collected from a same subject. But its performance is still better than Memento [5] which recognizes six fitness types with an accuracy of 57%.



(a) The experiment scenario. (b) Self-report screenshot of MW-Emotion. (c) Real-time data screenshot of MW-Emotion.

Figure 2: The experiment scenario and software screenshots of MW-Emotion.

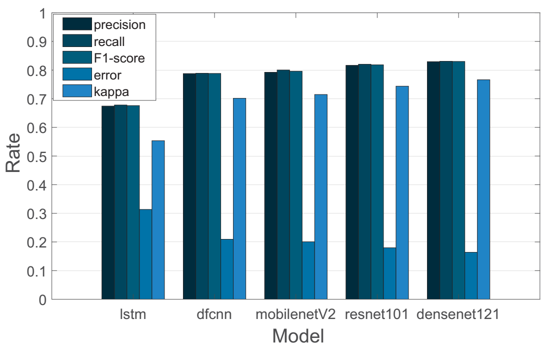


Figure 3: The comparison of recognition performance with different models

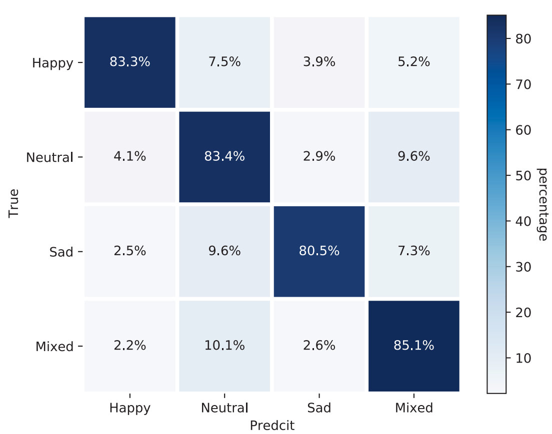


Figure 4: The confusion matrix achieved by the DenseNet

With current setup, MW-Emotion can still achieve a considerable recognition performance.

4 CONCLUSION

We present a noval low-cost, pervasive and portable muti-modal wearable emotion recognition system, called MW-Emotion, that recognizes emotion state with the aim of privacy-preserving, reliable detection and ease of use in daily life. Our experiments show that MW-Emotion can recognize four basic emotion states using multi-modal wearable devices with high accuracy, thereby showing the potential of a future where everyone's emotion can be sensed with low-cost, pervasive and portable wearable devices.

ACKNOWLEDGMENTS

This research was supported in part by the China NSFC Grant (61802264, 61872248), Shenzhen Science and Technology Foundation (No. JCYJ20180305124807337, No. ZDSYS20190902092853047). Yongpan Zou is the corresponding author.

REFERENCES

[1] Margaret M Bradley and Peter J Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25, 1 (1994), 49–59.

[2] Margaret M Bradley and Peter J Lang. 2007. The International Affective Digitized Sounds (; IADS-2): Affective ratings of sounds and instruction manual. *University of Florida, Gainesville, FL, Tech. Rep. B-3* (2007).

[3] Sandra Carvalho, Jorge Leite, Santiago Galdo-Álvarez, and Oscar F Gonçalves. 2012. The emotional movie database (EMDB): A self-report and psychophysiological study. *Applied psychophysiology and biofeedback* 37, 4 (2012), 279–294.

[4] Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. 2013. Differential entropy feature for EEG-based emotion classification. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 81–84.

[5] Shiqi Jiang, Zhenjiang Li, Pengfei Zhou, and Mo Li. 2019. Memento: An emotion-driven lifelogging system with wearables. *ACM TOSN* 15, 1 (2019), 1–23.

[6] Alexandre Schaefer, Frédéric Nils, Xavier Sanchez, and Pierre Philippot. 2010. Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and emotion* 24, 7 (2010), 1153–1172.

[7] Mingmin Zhao, Fadel Adib, and Dina Katabi. 2016. Emotion recognition using wireless signals. In *Proceedings of the ACM Mobicom*. 95–108.