Demo Abstract: Real-Time Emotion Detection via E-See

Weixi Gu
University of California, Berkeley
guweixi@berkeley.edu

Yue Zhang*
Dep. of Electronic Engineering
Tsinghua University
zyee16@mails.tsinghua.edu.cn

Fei Ma
Tsinghua-Berkeley Shenzhen Institute
Tsinghua University
mf17@mails.tsinghua.edu.cn

Khalid Mosalam
University of California, Berkeley
mosalam@berkeley.edu

Lin Zhang
Tsinghua-Berkeley Shenzhen Institute
Tsinghua University
linzhang@tsinghua.edu.cn

Shiguang Ni
Graduate School at Shenzhen
Tsinghua University
ni.shiguang@sz.tsinghua.edu.cn

ABSTRACT

Real-time emotion detection has being attracted to human attention recently. Recognizing the inner emotion not only assists people to communicate and understand with each other, but also prevents the occurrence of the serious diseases (*e.g.*, autism) and the emergency (*i.e.*, child abuse, sexual invasion). Existing works usually adopt the professional and cumbersome devices to learn the emotions, and therefore limited in the daily usage. In this work, we design a pervasive and wearable device E-See that enables to recognize the emotion in real time. The prototype of the device is deployed in a microcomputer currently, and it can be resized as a small button worn on the collar or extend as a platform to detect the real-time emotion.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**;

KEYWORDS

Real-Time Emotion Recognition

ACM Reference Format:
Weixi Gu, Yue Zhang, Fei Ma, Khalid Mosalam, Lin Zhang, and Shiguang Ni. 2018. Demo Abstract: Real-Time Emotion Detection via E-See. In *The 16th ACM Conference on Embedded Networked Sensor Systems (SenSys '18), November 4–7, 2018, Shenzhen, China.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3274783.3275213

1 INTRODUCTION

Emotion identification is a fundamental part of the affective computing. From the view of communication, recognizing the real-time emotion status or switching helps people to communicate, and thus understand each other better. From the view of safety and health, tracking the emotions could prevent emergency happen such as child abuse or sexual invasion, as well as the diseases including the autism.

How to recognize the emotion in real time has spawned a significant amount of research works and industrial products. To name a few, Fei Tao *et al.*, [11] built a platform to analyze the acoustic signals to infer the human emotions. However, predicting only based on the acoustic signals generally yields an unsatisfied performance, and the complicated workflow limits to a few application scenarios. Robert LiKamWa *et al.*, [10] leveraged the smartphones to sense the user daily operations and then predicted the human mood. Although their operation is ubiquitous, they cannot analyze the user's mood in real time. Therefore, a high-accuracy emotion recognition device with pervasive operation workflow is urgent required

Inspired by the above concerns, we propose E-See , a wearable device to recognize the human emotion in real time. E-See is a micro-computer based machine that automatically monitors the acoustic and visual signals of the users via the embeded microphone and camera. To be specific, it is composed of two modules: the acoustic analytics module and the visual analytics module. The acoustic module is continuously monitoring the user's voice by the microphone. While the potential acoustic signals of emotion occurring, E-See triggers the visual monitoring engine to take pictures of the user's facial expression by the camera. Both the acoustic and visual signals are fed into an inference model on the emotion analytics, which gurantee the recognition performance and and enhance the system resilience. Additionaly, the tiny size and easy-operation workflow makes E-See ubiquious in daily usage. In the current prototype, E-See focuses on four kinds of common emotion analysis: Happy, Neutral, Sad and Angry. It could be extended to more emotion inference in the future.

E-See has been evaluated on a large-scale corpus dataset. This dataset is composed of more than 3000 audio clips per emotion. Around 70% emotion detection accuracy shows the outstanding inference performance on the emotion recognition. The evaluation results show the low energy consumption and CPU utilization, which makes E-See acceptable in daily use.

2 SYSTEM DESIGN AND WORKFLOW

E-See is composed of two parts: the system end is for the real-time analytics and the cloud and the user end is for the long-term analytics.

In the system end, E-See continuously invokes the embedded microphone to monitor the acoustic signals of users with 16KHz

*Weixi Gu and Yue Zhang are co-first authors.

SenSys '18, November 4–7, 2018, Shenzhen, China

ACM ISBN 978-1-4503-5952-8/18/11…$15.00
https://doi.org/10.1145/3274783.3275213

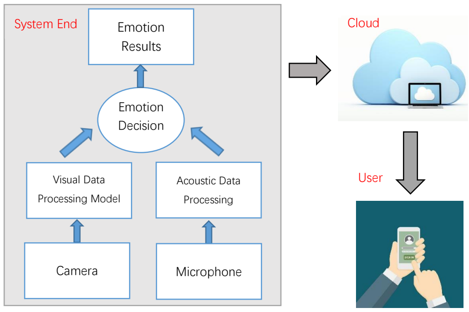

Figure 1: The architecture of E-See .

sampling rate. The raw data are processed via the pyaudioanalysis toolkit [1] to generate 34-dimensional acoustic features. The collected acoustic features are fed into a speech recognition model for the emotion analytics. The speech recognition model is an attention-based LSTM model [6], which has been pre-trained by the corpus datasets involving the four kinds of emotion. Once meeting with the representative acoustic signals of the to-be-detected emotions, E-See further triggers the camera to capture the images. The visual signal sampled by the camera is processed by a CNN-based framework [9], which is also pre-trained by the emotion image datasets. Afterward, the multi-modal (*i.e.*, visual and acoustic) signals are fused together to recognize the emotion, which enhances the system resilience and improves the inference capability. In this way, E-See is able to recognize the emotion in real time.

Compared with the real-time analytics in the system end, the objective of the cloud and user end is for a long-term emotion analytics. In the first, the inference results output from the system end are transferred and collected in the cloud. By analyzing the percentage of four kinds of emotion (*i.e.*, happy, sad, angry and neutral), we offer the service of weekly and monthly emotion analytics of the users and provide a report. The long-term analytics helps the user for their healthcare.

3 EVALUATION

Dataset. E-See has been evaluated on the corpus dataset composed of 3591 speech clips of each emotion, which involves a total of 212 females and 218 males speech segments. We randomly select 80% of the dataset as the training data and the left is the testing data.

Inference Performance. Fig. 2 illustrates the inference performance of E-See on the four kinds of emotion detection. As shown, all the precision values of the four kinds of emotion are more than 60%, and it performs the best on the "sad" with 67.85%. The recall values of the four emotion ranging from 55.26% ("neutral") to 76.97% ("happy"). In general, E-See achieves with 66% accuracy overall, which is much higher than the accuracy of random guess (25%).

System Overhead. To make sure that E-See is acceptable in daily use, we evaluate its overhead in terms of the energy consumption and the CPU utility. The prototype is installed in the Raspberry Pi 3b platform, with CPU Broadcom BCM2837B0 quad-core A53 (ARMv8) 64-bit with 1.4GHz. We install a power logger and CPU-monitor on the Raspberry Pi 3b. They record the power and CPU dynamics every 30 minutes while E-See is running. Statistically, E-See consumes less 5% power every 1 hour, and it occupies around

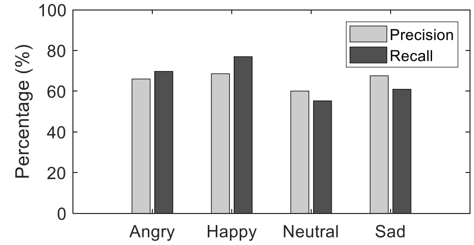

Figure 2: The inference performance of E-See .

40% utility of CPU. The results demonstrate that E-See only consumes a few resources during running, which is acceptable in the daily usage.

4 CONCLUSIONS AND FUTURE WORK

Real-time emotion identification is of great importance for human beings. In this work, we design and implement E-See , a wearable device that recognizes four kinds of daily emotions in real time. It utilizes both acoustic and visual signals to infer the emotion. In the future, we plan to install E-See on the smartphone for the blood glucose monitoring [2, 7], sleep tracking [3, 4] and people behaviors analytics [5, 8].

Acknowledgement. This work is supported by the RD Program of Shenzhen (JCYJ20170307153032483) and the Interdisciplinary Research Project of Graduate School at Shenzhen of Tsinghua University (JC2017005).

REFERENCES

[1] Theodoros Giannakopoulos. 2015. pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis. *PloS one* 10, 12 (2015).
[2] Weixi Gu. 2017. PhD Forum Abstract: Non-intrusive Blood Glucose Monitor by Multi-task Deep Learning. In *Information Processing in Sensor Networks (IPSN), 2017 16th ACM/IEEE International Conference on*. IEEE, 249–250.
[3] Weixi Gu, Longfei Shangguan, Zheng Yang, and Yunhao Liu. 2016. Sleep hunter: Towards fine grained sleep stage tracking with smartphones. *IEEE Transactions on Mobile Computing* 15, 6 (2016), 1514–1527.
[4] Weixi Gu, Zheng Yang, Longfei Shangguan, Wei Sun, Kun Jin, and Yunhao Liu. 2014. Intelligent sleep stage mining service with smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 649–660.
[5] Weixi Gu, Kai Zhang, Zimu Zhou, Ming Jin, Yuxun Zhou, Xi Liu, Costas J Spanos, Zuo-Jun Max Shen, Wei-Hua Lin, and Lin Zhang. 2017. Measuring fine-grained metro interchange time via smartphones. *Transportation research part C: emerging technologies* 81 (2017), 153–171.
[6] Weixi Gu, Yuxun Zhou, Zimu Zhou, Xi Liu, Han Zou, Pei Zhang, Costas J Spanos, and Lin Zhang. 2017. Sugarmate: Non-intrusive blood glucose monitoring with smartphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 54.
[7] Weixi Gu, Zimu Zhou, Yuxun Zhou, Miao He, Han Zou, and Lin Zhang. 2017. Predicting Blood Glucose Dynamics with Multi-time-series Deep Learning. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*. ACM, 55.
[8] Weixi Gu, Zimu Zhou, Yuxun Zhou, Han Zou, Yunxin Liu, Costas J Spanos, and Lin Zhang. 2017. BikeMate: Bike Riding Behavior Monitoring with Smartphones. In *Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, MobiQuitous 2017*. ACM.
[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.
[10] Robert LiKamWa, Yunxin Liu, Nicholas D Lane, and Lin Zhong. 2013. Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 389–402.
[11] Fei Tao, Gang Liu, and Qingen Zhao. 2018. An Ensemble Framework of Voice-Based Emotion Recognition System for Films and TV Programs. *arXiv preprint arXiv:1803.01122* (2018).