Real-time, Continuous Emotion Annotation Technique for Mobile Video Watching

Tianyi Zhang
Centrum Wiskunde & Informatica
Science Park 123,1098 XG
Amsterdam, the Netherlands
Tianyi.Zhang@cwi.nl

Pablo Cesar
Centrum Wiskunde & Informatica
Science Park 123,1098 XG
Amsterdam, the Netherlands
p.s.cesar@cwi.nl

Abdallah El Ali
Centrum Wiskunde & Informatica
Science Park 123,1098 XG
Amsterdam, the Netherlands
Abdallah.El.Ali@cwi.nl

Alan Hanjalic
Delft University of Technology
Van Mourik Broekmanweg 6,
2628 XE Delft, the Netherlands
A.Hanjalic@tudelft.nl

CHI'20,, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.XXXXXXX

Abstract
Collecting accurate and precise emotion ground truth labels for mobile video watching is essential for ensuring meaningful emotion predictions. However, most emotion annotation techniques for mobile video watching rely on post-stimulus and discrete self-reports. To address this limitation, we present a real-time, continuous emotion annotation (RCEA) technique to collect continuous emotion ground truth labels for mobile video watching. Based on *Russell's Circumplex model*, we designed a virtual joystick which allow users to input their valence and arousal continuously when they are watching videos on a mobile phone. Our RCEA provides a starting point for researchers to design momentary self-reports on mobile devices for precise emotion recognition.

Author Keywords
emotion; annotation; mobile; video; real-time; continuous; labels

CCS Concepts
•**Human-centered computing** → **Human computer interaction (HCI); Graphical user interfaces;** User studies;

Introduction
Collecting continuous emotion ground truth labels for video watching poses a challenge, as the annotation procedure demands users to multi-task, which can increase mental

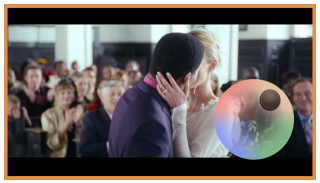

Figure 1: RCEA user interface

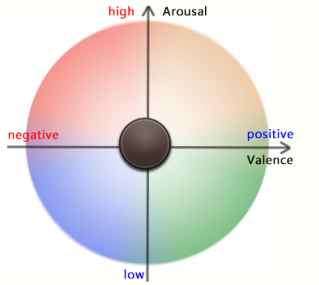

Figure 2: The virtual joystick for V-A annotation

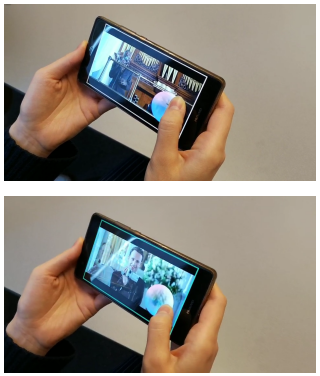

Figure 3: Annotating from positive valence and neutral arousal (up) to positive valence and high arousal (down)

workload while users are watching videos. There are several continuous emotion annotation tools such as *DARMA* [2] or *CASE* [9], which allow users to input their *valence* and *arousal* (V-A) continuously. However, they are limited to static, desktop environments, and require additional auxiliary devices (e.g., physical joystick [9]) to annotate emotions. To address this, we developed a real-time continuous emotion annotation (RCEA) technique [12] suitable for mobile screens. The design of RCEA followed an iterative, user-centric approach and underwent several prototyping rounds. Our technique enables researchers to collect fine-grained, temporal emotion annotations of valence and arousal while users are watching mobile videos. Below, we explain the detail of RCEA.

Design and development of RCEA

As shown in Figure 1, the user interface of RCEA consists of three modules: 1) a video player, 2) a virtual joystick and 3) a color frame. To use RCEA, the users need to put their thumbs on the virtual joystick to rate their V-A. As shown in Figure 3, when the emotion of user changes, the user is supposed to move his or her thumb to indicate this change. The frame around the video player will show the color of the quadrant where user's thumb is touching as peripheral feedback for annotation. After the user watch the whole video, a continuous V-A self-report will be obtained and aligned with the video content. Below we introduce the design details of the virtual joystick and the color frame.

Virtual joystick

The virtual joystick is designed based on *Russell's Circumplex model* [8] given it is widely used, and offers a finer level of granularity for describing emotions [10]. The virtual joystick is sampled as 10Hz according to the research on human motor control [5]. As shown in Figure 2, the horizon and vertical axis of the virtual joystick represents for the valence and arousal dimension of users' emotion respectively. Four colors (HEX values = #eecdac, #7fc087, #879af0, #f4978e for quadrants one to four respectively) provided feedback to users on which emotion they are currently annotating. We selected four colors based on a simplified version of Itten's color system [4], which has been shown to be intuitive and easy for users to understand [3]. We also include a gradual transparency from the origin (0% opacity) to the edge (100%) of the wheel to minimize the overlapping area between the video player and the virtual joystick. The transparency is also an indication of the transition of V-A intensity, which means the less transparent the colors are, the stronger the emotions.

The center of the virtual joystick is placed at the bottom right corner of the screen (coordinates: (233 dp, 233 dp)), considering right-hand dominance of users. However, users can choose to use either thumb to annotate. The radius of the virtual joystick is 168 dp. On our testing device (Huawei P9 Plus, 32GB, 5.5 inches, resolution = 1920$\times$1080), it is 23.8mm, which is comfortable for the thumb to move continuously [11]. Touchpoint range radius from the bottom screen edge is 56.8mm, and was determined based on screen size (5.5") and the functional thumb area. Touchpoint size is 7.14mm, as a size of minimum 7mm provides the best touch performance for time-related measures [7], in our case continuous touch.

Color frame

A color frame around the screen display is designed as additional peripheral feedback [6, 1] to the user. We mapped the frame colors to each V-A quadrant. Compared with text labels and emojis, the frame can provide additional feedback in a subtle, peripheral manner, without drawing up screen real estate.

Conclusion

In conclusion, our prototype provides a continuous emotion annotation method that is suitable for collecting more precise emotion ground truth labels for mobile video watching. This demo submission can help capture momentary emotions in mobile environments, which can be used for building more temporally precise emotion recognition and prediction models for mobile video watching.

REFERENCES

[1] Saskia Bakker, Doris Hausen, and Ted Selker. 2016. *Peripheral Interaction: Challenges and Opportunities for HCI in the Periphery of Attention* (1st ed.). Springer Publishing Company, Incorporated.

[2] Jeffrey M Girard and Aidan GC Wright. 2018. DARMA: Software for dual axis rating and media annotation. *Behavior research methods* 50, 3 (2018), 902–909.

[3] Dini Handayani, Abdul Wahab, and Hamwira Yaacob. 2015. Recognition of emotions in video clips: the self-assessment manikin validation. *Telkomnika* 13, 4 (2015), 1343.

[4] Johannes Itten. 1963. *Mein Vorkurs am Bauhaus*. Otto Maier Verlag.

[5] Ian D Loram, Henrik Gollee, Martin Lakie, and Peter J Gawthrop. 2011. Human control of an inverted pendulum: is continuous control necessary? Is intermittent control effective? Is intermittent control physiological? *The Journal of physiology* 589, 2 (2011), 307–324.

[6] Tara Matthews, Anind K. Dey, Jennifer Mankoff, Scott Carter, and Tye Rattenbury. 2004. A Toolkit for Managing User Attention in Peripheral Displays. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04)*. ACM, New York, NY, USA, 247–256.

[7] Yong S Park and Sung H Han. 2010. Touch key design for one-handed thumb interaction with a mobile phone: Effects of touch key size and touch key location. *International journal of industrial ergonomics* 40, 1 (2010), 68–76.

[8] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[9] Karan Sharma, Claudio Castellini, Freek Stulp, and Egon L Van den Broek. 2017. Continuous, real-time emotion annotation: A novel joystick-based analysis framework. *IEEE Transactions on Affective Computing* (2017).

[10] Lin Shu, Jinyan Xie, Mingyue Yang, Ziyi Li, Zhenqi Li, Dan Liao, Xiangmin Xu, and Xinyi Yang. 2018. A review of emotion recognition using physiological signals. *Sensors* 18, 7 (2018), 2074.

[11] Jinghong Xiong and Satoshi Muraki. 2016. Effects of age, thumb length and screen size on thumb movement coverage on smartphone touchscreens. *International Journal of Industrial Ergonomics* 53 (2016), 140–148.

[12] Tianyi Zhang, Abdallah El Ali, Chen Wang, Alan Hanjalic, and Pablo Cesar. 2020. RCEA: Real-time, Continuous Emotion Annotation for Collecting Precise Mobile Video Ground Truth Labels. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, Article Paper #, 15 pages. DOI: http://dx.doi.org/10.1145/3313831.3376808