Annotation Tool for Precise Emotion Ground Truth Label Acquisition while Watching 360° VR Videos

Xue, Tong; Ali, Abdallah El; Ding, Gangyi; Cesar, Pablo

DOI
10.1109/AIVR50618.2020.00076

Publication date
2020

Document Version
Accepted author manuscript

Published in
Proceedings - 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality, AIVR 2020

Annotation Tool for Precise Emotion Ground Truth Label Acquisition while Watching 360° VR Videos

Tong Xue
Beijing Institute of Technology
Centrum Wiskunde & Informatica (CWI)
Beijing, China
xuetong@bit.edu.cn

Abdallah El Ali
Centrum Wiskunde & Informatica (CWI)
Amsterdam, The Netherlands
abdallah.el.ali@cwi.nl

Gangyi Ding
Beijing Institute of Technology
Beijing, China
dgy@bit.edu.cn

Pablo Cesar
Centrum Wiskunde & Informatica (CWI)
Delft University of Technology
Amsterdam, The Netherlands
p.s.cesar@cwi.nl

***Abstract*—We demonstrate an HMD-based annotation tool for collecting precise emotion ground truth labels while users are watching 360° videos in Virtual Reality (VR). Our tool uses an HTC VIVE Pro Eye HMD for displaying 360° videos, a Joy-Con controller for inputting emotion annotations, and an Empatica E4 wristband for capturing physiological signals. Timestamps of these devices are synchronized via an NTP server. Following dimensional emotion models, users can report their emotion in terms of valence and arousal as they watch a video in VR. Annotation feedback is provided through two peripheral visualization techniques: HaloLight and DotSize. Our annotation tool provides a starting point for researchers to design momentary and continuous self-reports in virtual environments to enable fine-grained emotion recognition.**

***Index Terms*—360° video; emotion annotation; continuous; ground truth labels**

I. Introduction

Watching 360° videos using head-mounted displays (HMDs) allows users to interact with the content and feel immersed in such experiences. Previous work [2] has indicated that this setting has the ability to elicit a wide range of emotions in individuals as they interact with the content through head movements. Thus it is important to collect accurate and precise ground truth labels throughout the user's HMD-based experience.

Emotion research are represented in two main ways generally: as categories or as dimensions. Categorical models (e.g. Ekman's six-basic-emotion model [3] and Plutchik wheel model [4]) are easy to understand but they ignore the intensity and continuity of emotions. Russell's Circumplex model [1], as a common used two-dimensional emotion space, combines valence and arousal and allows users to express a wide range of emotions, which is more precise and specific. Based on this, there have been some studies aimed at developing real-time, continuous emotion annotation techniques to collect valence and arousal labels while watching video stimuli, like the joystick-based CASE for desktop [5] and mobile RCEA [6]. But given that virtual environments, users watch videos with an HMD. As Oliveira et al. stated [7], existing self-report methods are carried out either after the VR experience or outside of wearing an HMD. In this work we develop a 360VR HMD-based annotation tool to capture emotions in real-time. The used 360° videos are selected from Li et al.'s database[1].

II. 360VR HMD-based Annotation Tool

Our tool is designed based on the Russell's Circumplex Model, given it is widely used and offer a finer level of granularity for describing emotions [8]. Four colors (HEX values = #eecdac, #7fc087, #879af0, #f4978e for quadrants one to four respectively) are selected based on a simplified version of Itten's color system [9], as shown in Fig.1a. Following the instructions for designing VR HMD-based interactions [10], our 360VR annotation tool contains the following four parts.

A. HTC VIVE Pro Eye HMD

We used the HTC VIVE Pro Eye HMD[2] to present immersive VR environments and display 360° videos. An eye tracker is embedded to collect users' behavioral data, including eye movements and pupil diameter.

B. Joy-Con Controller

We used Joy-Con right controller[3], a primary wireless digital gaming peripheral, to collect users' self-report emotions (valence and arousal). Joy-con has an analog stick (featuring a return spring) and an an array of buttons. Users could report their emotions by moving the joystick head into one of the four quadrants of the Valence-Arousal model space. To increase the emotion intensity, the User can move the joystick head further. Also, we enabled a helper function by clicking the trigger button on the controller, thereby users who forget how to annotate can use it for easy lookup.

[1]https://vhil.stanford.edu/360-video-database/
[2]https://enterprise.vive.com/us/product/vive-pro-eye/
[3]https://www.nintendo.com/switch/choose-your-joy-con-color/

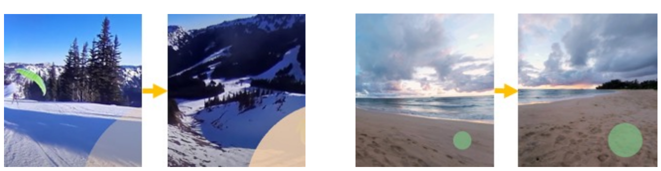
(a) (b) (c)

Fig. 1: 360VR HMD-based annotation tool. (a) Valence-Arousal model space based on Circumplex model [1]. (b) Hardware set-up. (c) One user in our tool, wearing HMD, Empaticia wristband, and annotating with Joy-con controller.

C. Empatica E4 Wristband

We used Empatica E4 Wristband[4], a wireless wearable device to capture users' continuous and real-time physio data, which can be used to analyze and better understand users' physiologucal states in VR. With four embedded sensors, E4 band can measure accelerometer (ACC) data at 64Hz, electrodermal activity (EDA) data in μS at 4Hz, Blood Volume Pulse (BVP) data at 64Hz, and skin temperature (SKT) in Celsius at 4Hz. The Inter-Beat Interval (IBI) and heart rate (HR) data are calculated by BVP data. We stored the physiological data from E4 on a mobile device (Huawei P9 Plus, 32GB, 5.5inches, 1920X1080), and synchronized the timestamps of this mobile and computer via an NTP Server [5].

D. Annotation Interface



(a) HaloLight (b) DotSize

Fig. 2: Two peripheral information visualizations.

We constructed a custom scene in the Unity Engine[6] to display 360° videos at 30 fps. Given that users will be wearing an HMD to view the immersive videos and annotate emotions simultaneously, we provide two peripheral information visualizations (HaloLight and DotSize) to minimize workload and distraction based on our prior work [11]. Whereas HaloLight uses color opacity to indicate intensity (shown in Fig. 2a), DotSize uses the size of the filled circle to indicate intensity (shown in Fig. 2b). Both are fixed at the right bottom corner of the HMD viewport and the color of the component reflects the emotion users are annotating.

[4]https://www.empatica.com/
[5]android.pool.ntp.org
[6]https://unity3d.com/

III. Conclusion

In conclusion, our tool offers a continuous emotion annotation technique that is designed for collecting more precise emotion ground truth labels for 360° VR video content. Furthermore, this demo can be used for building more temporally precise 360° video-based emotion recognition and prediction models.

References

[1] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.
[2] T. Oliveira, P. Noriega, F. Rebelo, and R. Heidrich, "Evaluation of the relationship between virtual environments and emotions," in *International Conference on Applied Human Factors and Ergonomics*. Springer, 2017, pp. 71–82.
[3] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.
[4] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *American scientist*, vol. 89, no. 4, pp. 344–350, 2001.
[5] K. Sharma, C. Castellini, E. L. van den Broek, A. Albu-Schaeffer, and F. Schwenker, "A dataset of continuous affect annotations and physiological signals for emotion analysis," *Scientific data*, vol. 6, no. 1, pp. 1–13, 2019.
[6] T. Zhang, A. El Ali, C. Wang, A. Hanjalic, and P. Cesar, "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*, 2020, pp. 1–15.
[7] T. Oliveira, P. Noriega, J. Carvalhais, F. Rebelo, and V. Lameira, "How deep is a virtual reality experience? virtual environments, emotions and physiological measures," in *International Conference on Applied Human Factors and Ergonomics*. Springer, 2019, pp. 462–471.
[8] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, and X. Yang, "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, p. 2074, 2018.
[9] A. Ståhl, P. Sundström, and K. Höök, "A foundation for emotional expressivity," in *Proc. Designing for User eXperience '05*. AIGA: American Institute of Graphic Arts, 2005, p. 33.
[10] J. Jerald, *The VR book: Human-centered design for virtual reality*. Morgan & Claypool, 2015.
[11] T. Xue, S. Ghosh, G. Ding, A. El Ali, and P. Cesar, "Designing real-time, continuous emotion annotation techniques for 360° vr videos," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–9.