Improving Inclusivity for Emotion Recognition Based on Face Tracking

Mats Ole Ellenberg*
mellenberg@acm.org
Interactive Media Lab Dresden,
TUD Dresden University of Technology
Dresden, Germany

Katja Krug*
katjakrug@acm.org
Interactive Media Lab Dresden,
TUD Dresden University of Technology
Dresden, Germany

Abstract

The limited expressiveness of virtual user representations in Mixed Reality and Virtual Reality can inhibit an integral part of communication: emotional expression. Emotion recognition based on face tracking is often used to compensate for this. However, emotional facial expressions are highly individual, which is why many approaches have difficulties recognizing unique variations of emotional expressions. We propose several strategies to improve face tracking systems for emotion recognition with and without user intervention for the Affective Interaction Workshop at CHI '25 [1].

CCS Concepts

• **Human-centered computing → Mixed / augmented reality**; **Accessibility theory, concepts and paradigms**.

Keywords

Mixed Reality, Emotion Recognition, Neurodiversity, Accessibility

1 Introduction and Background

Emotions are an integral part of interpersonal communication and creating connections between humans. If the mode of communication does not support the transfer of emotions through facial expressions or tone of voice, like in text messages, users often compensate by artificially communicating their emotions, for example, by adding emojis to text messages [3]. In the current landscape of Mixed Reality (MR), and especially Virtual Reality (VR), true emotions are hard to transfer onto virtual avatars, due their fidelity hindering the replicability of subtle or ambiguous facial expressions.

Therefore, many techniques utilize emotion recognition [2, 6], where emotions are characterized and recognized through face tracking, and can then be used to trigger pre-determined facial animations in avatars. This is also especially interesting for accessibility, as emotion recognition is often utilized for systems supporting and understanding the limited emotion recognition skills of neurodiverse people, e.g., with autism spectrum disorder (ASD) [7]. Here, several AR approaches utilize emotion recognition to display a simplified emoji to users [4, 9] or replace faces of others with 3D emojis [11], assisting users in their perception of emotions.

To use these techniques successfully, the underlying emotion recognition must be as accurate as possible. AI-based face tracking is often used for such emotion recognition, as it is widely available and only needs the visual information of the face [8]. However, this is still error-prone, as makeup or facial hair can hinder recognition, and facial emotion expressions differ across cultures and individuals (especially those with ASD, who show emotions less frequently, and their expressions are judged by neurotypical people to be less accurate and of lower quality [10, 12]). Also, artificially over-acted emotions used for training can look different from naturally evoked emotions, especially with regard to intensity. Lastly, many recognition systems use distinct, simple categories for labeling, but emotions do not necessarily fit in these categories, as they can be a more complex combination of multiple emotions [13].

Thus, we want to discuss several directions to increase inclusivity and improve emotion recognition through facial tracking, either directly on the system's side or combined with user input, as the latter has not yet been deeply explored in the literature.

2 Improving Inclusivity

On the system's side, a first step in increasing the quality and inclusivity of emotion recognition is to create more versatile and inclusive emotion datasets to train AI models. Datasets should include acted and naturally evoked emotions, broader scales of emotion intensities, and complex emotion combinations. Additionally, datasets should consist of diverse selections of humans, especially neurodiverse people, as the latter often express emotions differently than neurotypical people, being described as unusual or mechanical-looking expressions [10].

Without expanding the needed sensors, the system can also use available situational and group contexts in emotion recognition. This means that the system could, for example, use the emotional state of other users, audio from voice chats, or textual information as additional input. To make the system more transparent to the user, the results of the emotion recognition should not only be transferred or used as input to trigger functionality but also be communicated

to the user. By mirroring to them how the system would interpret their emotional state, they can identify if that matches their true emotions.

Further including the user, static individual calibration of the emotion recognition can be done at the beginning of the tracking, further refined by dynamic calibration through input-driven reinforcement, where the user can tell the system if it recognized the emotion correctly. Additionally, we propose that the user can scale the intensity of the recognized emotion to counteract low expressivity or even manually trigger certain emotions if the system cannot recognize them independently.

3 Discussion & Conclusion

In the context of emotions, privacy is an especially important consideration. Thus, systems should support user agency by letting them control what personal data can be used to what extent for recognition, and they should be able to turn off the recognition when they do not want to share this information with others.

Even though we believe that expanding emotion training sets is crucial for making facial emotion recognition systems more inclusive, it is unclear how natural emotions, especially negative ones, can be captured ethically. To some degree, emotions could be triggered through external stimuli in a controlled and consensual setting, but the user's knowledge of being recorded can still influence the expression. One approach would be to collect self-labeled artifacts created in natural settings (e.g., videos from family gatherings) from volunteers, but this has a significant overhead and might not result in enough high-quality data.

We believe that systems should not encourage or force users to change their way of expressing emotions, as there is no correct way to do so, and especially in people with ASD, camouflaging and masking (changing behavior to hide ASD related traits) is linked to exhaustion and threats to self-perception [5]. Instead, systems should adapt their recognition to the user and provide them with the tools to help the system understand them, without changing who they are.

Acknowledgments

References

[1] Naseem Ahmadpour, Danielle Lottridge, Jonas Fritsch, Corina Sas, Marta Cecchinato, Daniel Harrison, Kristina Höök, Pin Foong, Kiran Ijaz, Phillip Gough, Yidan Cao, Xuefei Li, Shaimaa Lazem, and Thida Sachathep. 2025. Affective Interaction and Affective Computing -Past, Present and Future. In *Xtended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25) (CHI EA '25)*. ACM, New York, NY, USA. doi:10.1145/3706599.3706743
[2] Darren Bellenger, Minsi Chen, and Zhijie Xu. 2024. Facial Emotion Recognition with a Reduced Feature Set for Video Game and Metaverse Avatars. *Computer Animation and Virtual Worlds* 35, 2 (2024), e2230. doi:10.1002/cav.2230
[3] Anthony Cherbonnier and Nicolas Michinov. 2022. The Recognition of Emotions Conveyed by Emoticons and Emojis: A Systematic Literature Review. *Technology, Mind, and Behavior* 3, 2: Summer 2022 (April 2022). doi:10.1037/tmb0000067
[4] Nick Haber, Catalin Voss, and Dennis Wall. 2020. Making Emotions Transparent: Google Glass Helps Autistic Kids Understand Facial Expressions through Augmented-Reaiity Therapy. *IEEE Spectrum* 57, 4 (April 2020), 46–52. doi:10.1109/MSPEC.2020.9055973
[5] Laura Hull, K. V. Petrides, Carrie Allison, Paula Smith, Simon Baron-Cohen, Meng-Chuan Lai, and William Mandy. 2017. "Putting on My Best Normal": Social Camouflaging in Adults with Autism Spectrum Conditions. *Journal of Autism and Developmental Disorders* 47, 8 (Aug. 2017), 2519–2534. doi:10.1007/s10803-017-3166-5
[6] Mounika Kakarla and G.Ram Mohana Reddy. 2014. A Real Time Facial Emotion Recognition Using Depth Sensor and Interfacing with Second Life Based Virtual 3D Avatar. In *International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014)*. 1–7. doi:10.1109/ICRAIE.2014.6909153
[7] Connor Tom Keating and Jennifer Louise Cook. 2020. Facial Expression Production and Recognition in Autism Spectrum Disorders: A Shifting Landscape. *Child and Adolescent Psychiatric Clinics of North America* 29, 3 (July 2020), 557–571. doi:10.1002/aur.2037
[8] Byoung Chul Ko. 2018. A Brief Review of Facial Emotion Recognition Based on Visual Information. *Sensors* 18, 2 (Feb. 2018), 401. doi:10.3390/s18020401
[9] James Ren Hou Lee and Alexander Wong. 2020. AEGIS: A Real-Time Multimodal Augmented Reality Computer Vision Based System to Assist Facial Expression Recognition for Individuals with Autism Spectrum Disorder. doi:10.48550/arXiv.2010.11884 arXiv:2010.11884 [cs]
[10] Katherine A. Loveland, Belgin Tunali-Kotoski, Deborah A. Pearson, Kristin A. Brelsford, Juliana Ortegon, and Richard Chen. 1994. Imitation and expression of facial affect in autism. *Development and Psychopathology* 6, 3 (July 1994), 433–444. doi:10.1017/S0954579400006039
[11] Ran Sun, Harald Haraldsson, Yuhang Zhao, and Serge Belongie. 2019. AnonEmoji: An Optical See-Through Augmented Reality System for Children with Autism Spectrum Disorders to promote Understanding of Facial Expressions and Emotions. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 448–450. doi:10.1109/ISMAR-Adjunct.2019.00052
[12] Dominic A. Trevisan, Maureen Hoskyn, and Elina Birmingham. 2018. Facial Expression Production in Autism: A Meta-Analysis. *Autism Research* 11, 12 (Dec. 2018), 1586–1601. doi:10.1002/aur.2037
[13] Martin Wöllmer, Florian Eyben, Stephan Reiter, Björn Schuller, Cate Cox, Ellen Douglas-Cowie, and Roddy Cowie. 2008. Abandoning emotion classes - towards continuous emotion recognition with modelling of long-range dependencies. In *Interspeech 2008*. ISCA, 597–600. doi:10.21437/Interspeech.2008-192