MienCap: Performance-based Facial Animation with Live Mood Dynamics

Ye Pan*, Ruisi Zhang, Jingying Wang, Nengfu Chen, Yilin Qiu
Shanghai Jiaotong University,
Shanghai, China

Yu Ding
Netease Fuxi AI Lab,
Hangzhou, China

Kenny Mitchell
Edinburgh Napier University,
Edinburgh, UK
Roblox, CA, US

2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) | 978-1-6654-8402-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/VRW55335.2022.00178

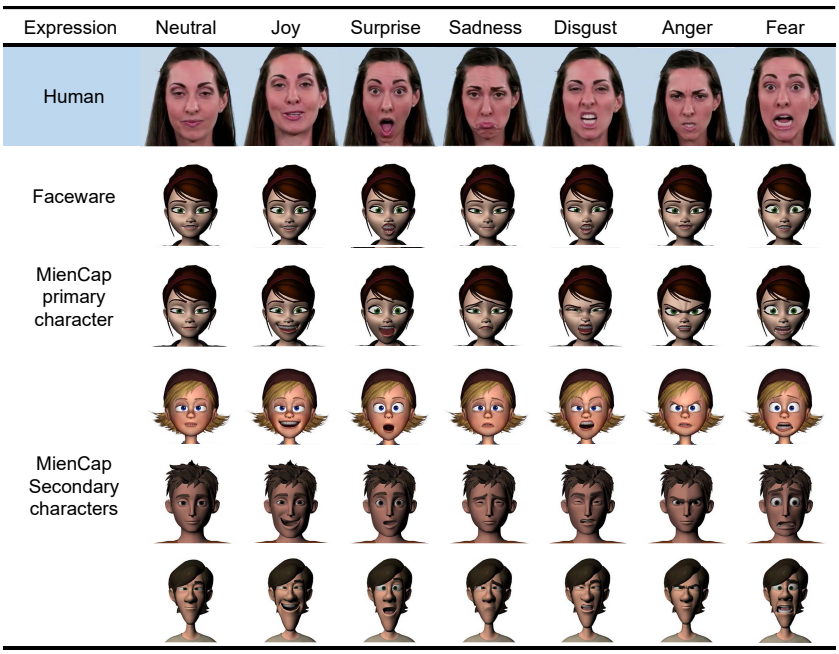

Figure 1: Resulting frames generated from Faceware, MienCap for primary character, and MienCap for secondary characters. Uppermost image in each column is input frame.

Abstract

Our purpose is to improve performance-based animation which can drive believable 3D stylized characters that are truly perceptual. By combining traditional blendshape animation techniques with machine learning models, we present a real time motion capture system, called MienCap, which drive character expressions in a geometrically consistent and perceptually valid way. We demonstrate the effectiveness of our system by comparing to a commercial product Faceware. Results reveal that ratings of the recognition of expressions depicted for animated characters via our systems are statistically higher than Faceware. Our results may be implemented into the VR filmmaking and animation pipeline, and provide animators with a system for creating the expressions they wish to use more quickly and accurately.

Index Terms: Computing methodologies—Computer graphics—Graphics systems and interfaces—Virtual reality; Computing methodologies—Computer graphics—Animation—Motion capture

1 Introduction

"The power of virtual reality cinema is found in its immersive nature. It is transmissive and completely blends the observer and the art". Hollywood is vigorously trying to develop VR movies, and some leading technology companies also have started to build VR theme parks to achieve a more interactive experience via comprehensive motion tracking systems, and video game-like scene design. In particular, creating genuine and accurate animated character facial expressions is critical [3]. Companies such as Disney and Dreamworks continue to use the classic facial animation workflow, which involves an animator creating expressions at discrete keyframes via an intuitive face rig [4, 7]. Character expressions made by animators could be emotive and clear, however, they require significant expertise and a tremendous amount of time.

Motion capture, or "moCap" is an alternative process that provides animators with direct realtime control on 3D characters via performance, providing a more efficient and timely way to generate animation [5, 8]. However, the expressive quality and perceptual validity of craft-based animations are frequently lacking in these moCap systems. This is primarily owing to the restricted two assumptions: First, geometric markers are adequate for expression transfer. Second, the geometric markers and features modeled are for human faces rather than stylized character faces. Despite recent breakthroughs in motion capture, modeling capabilities, and control parameterization, existing geometry-based systems fail to address the essential issue of producing clear expressions that can be correctly identified. Relatively few tools or methods [1, 2, 9] exist that assist animators by including the perception of character expressions while creation, however the most majority are not realtime.

We present MienCap, a new method that extends the idea of mocap to the capture of emotional data in real time. By combining traditional blendshape animation techniques with a machine

*e-mail: whitneypanye@sjtu.edu.cn

978-1-6654-8402-2/22/$31.00 ©2022 IEEE
DOI 10.1109/VRW55335.2022.00178

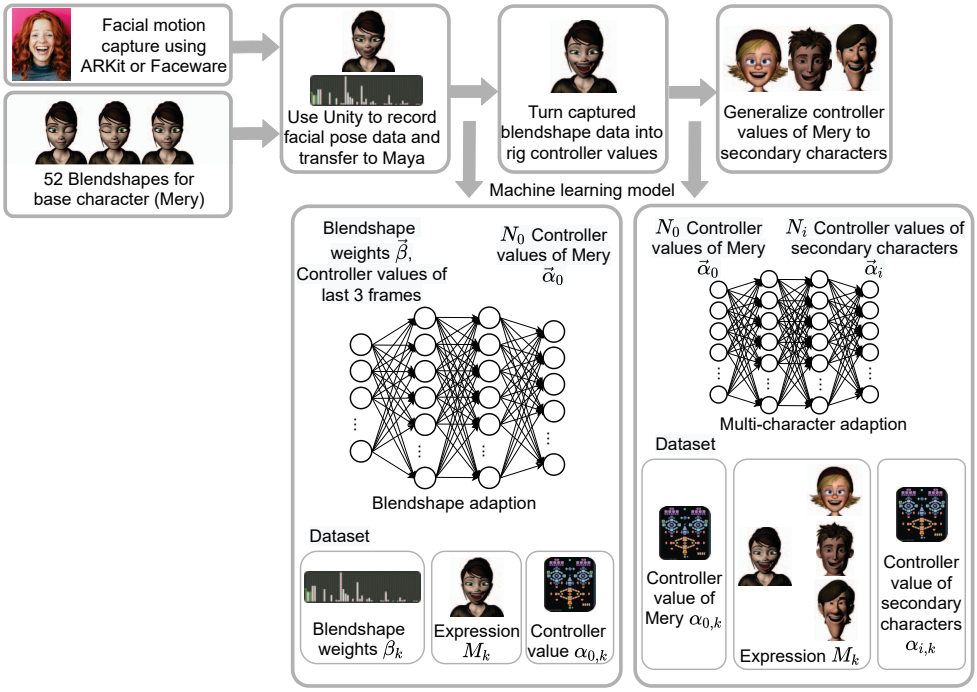

Figure 2: Overview of MienCap system

learning model, we contributed the first realtime system transferring human facial expressions to multiple 3D stylized characters in a geometrically consistent and perceptually correct way. Additionally, blendshape animation methods require a significant amount of manual effort in terms of creating a complete set of blendshapes for each new character. MienCap can flexibly mapping the expressions from the existing character to a new one without the need of creating blendshapes.

2 Method

Figure 2 gives an overview of our pipeline. First, we employed Polywink [6] to generate the fullset of blendshapes of our base character rig (Mery). Note that we generate the blendshapes only for the sake of Faceware comparison. Our system does not necessarily need to create the fullset of blendshapes to drive the character.

Next, we used the award-winning Faceware technology, which is a markless solution for facial motion capture with RGB input. The input image sequence can be either a realtime video stream or a selected video source. The output generated by Faceware Live Client for Unity is a temporal sequence of blendshape weights, which can be directly imported into commercial animation tools.

Then, these realtime blendshape weights with a frame rate of 24 fps are fed into the blendshape adaption model to generate controller values for the base character.

Lastly, the multiple character adaption model uses the controller values for the base character to produce the controller values for other secondary characters. These output controller values are directly assigned to the controllers of the characters in MAYA.

3 Conclusion

In summary, for the first time, we contribute a real time system that captures human facial expressions to drive a stylized character in a perceptually correct and geometrically cohesive fashion. Results (see Figure 1) show that our system creates more perceptually accurate expressions than popular commercially accessible software applications, such as, Faceware.

The ease of use of the MienCap system and the expressiveness of our resulting animations can potentially improve the effectiveness of visual storytelling in areas of online marketing, gaming, animated films, and immersive experiences. Our system can also be used in real-time ("live") animation situations, where facial expression is a useful input modality and amateurs can communicate stories with expressive animation by capturing their own performances. The real-time capability of our method paves the way for many applications in the context of VR filmmaking and teleconferencing.

References

[1] D. Aneja, B. Chaudhuri, A. Colburn, G. Faigin, L. Shapiro, and B. Mones. Learning to generate 3d stylized character expressions from humans. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 160–169. IEEE, 2018.

[2] D. Aneja, A. Colburn, G. Faigin, L. Shapiro, and B. Mones. Modeling stylized character expressions via deep learning. In *Asian Conference on Computer Vision*, pp. 136–153. Springer, 2016.

[3] M. Gonzalez-Franco, E. Ofek, Y. Pan, A. Antley, A. Steed, B. Spanlang, A. Maselli, D. Banakou, N. Pelechano Gómez, S. Orts-Escolano, et al. The rocketbox library and the utility of freely available rigged avatars. *Frontiers in virtual reality*, 1(article 561558):1–23, 2020.

[4] J. Lasseter. Principles of traditional animation applied to 3d computer animation. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pp. 35–44, 1987.

[5] H. Li, J. Yu, Y. Ye, and C. Bregler. Realtime facial animation with on-the-fly correctives. *ACM Trans. Graph.*, 32(4):42–1, 2013.

[6] Polywink. Automatic expressions blendshapes and facial rigs. 2021.

[7] F. Thomas, O. Johnston, and F. Thomas. *The illusion of life: Disney animation*. Hyperion New York, 1995.

[8] T. Weise, S. Bouaziz, H. Li, and M. Pauly. Realtime performance-based facial animation. *ACM transactions on graphics (TOG)*, 30(4):1–10, 2011.

[9] J. Zhang, K. Chen, and J. Zheng. Facial expression retargeting from human to avatar made easy. *IEEE Transactions on Visualization and Computer Graphics*, 2020.