Realtime Facial Expression Representation Method For Virtual Online Meetings System

Yinge Zhu, Bruno Carvacho Yerkovich, Xingjie Zhang, Jong-il Park[1]
Hanyang University
zhuyinge@hanyang.ac.kr, brunocarva@hanyang.ac.kr,
xiaoqie0125@hanyang.ac.kr, jipark@hanyang.ac.kr

Abstract

In a society with Covid-19 as part of our daily lives, we had to adapt ourselves to a new reality to maintain our lifestyles as normal as possible. An example of this is teleworking and online classes. However, several issues appeared on the go as we started the new way of living. One of them is the doubt of knowing if real people are in front of the camera or if someone is paying attention during a lecture. Therefore, we encountered this issue by creating a 3D reconstruction tool to identify human faces and expressions actively. We use a web camera, a lightweight 3D face model, and use the 2D facial landmark to fit expression coefficients to drive the 3D model. With this Model, it is possible to represent our faces with an Avatar and fully control its bones with rotation and translation parameters. Therefore, in order to reconstruct facial expressions during online meetings, we proposed the above methods as our solution to solve the main issue.

Keywords: 3D expression representation, 3D avatar tracking, 3D model fitting.

1. Introduction

3D face reconstruction and avatars for representing human expressions are practical tools widely implemented in diverse fields such as Video games, Movies, Health, and Education. Therefore, the main goal of our research is to find a solution to emulate facial reconstruction on a 3D avatar and use this solution in live online meetings.

During the Covid-19 pandemic, computers as the primary tool for working and studying have changed our lifestyles positively and negatively; We got used to being in front of the computer for online classes or meetings in our houses, letting us save time and increase productivity. Nevertheless, in numerous cases, due to personal reasons or lack of interest, turning off the camera and letting the online video call on hold while doing another activity has become a regular practice among individuals.

Sometimes we are embarrassed about showing our faces or afraid of other people taking pictures of ourselves during a live session. To solve this situation, the use of facial reconstruction in 3D emojis can actively cover our faces but still follow our expressions and movements during the video call, making possible live tracking and expression mapping an accurate system for our daily online meetings.

Our 3D mapping and face expression emulation approach starts with a reliable base model with predefined vertices. On top of that, the extraction of basic animation units from a human face as an input will be the two most fundamental and necessary aspects of the overall process. After the initial procedure, following a data mapping strategy

[1] Corresponding author

in a previously designed 3D Avatar is crucial for emulating human expressions. As the 3D avatar corresponds to a group of vertices that form a 3D mesh, the Model identifies the predominant landmarks of a human face in real-time and establishes specific parameters to fit them in the avatar's vertices. During this step, we will use the Gauss-Newton Algorithm[1] to fit and determine the parameters for animation driving.

2. Facial Expression Representation

In order to accomplish our goal, we decided to work with a pre-established 3D model called Candide3 [2], ERT [3] (Ensemble of Regression Trees) to extract face landmarks and Gauss Newton Fitting algorithm. The reason why we decided to choose these approaches is because the three of them are light and fast to run for the purposes we established in our goal. Finally, for the creation of our 3D model, we chose Blender as the main program as it is open-source and has the necessary functionalities for 3D modeling.

Candide-3 is a parametric 3D face model. It has 113 vertices: 184 triangles. The three-dimensional group of coordinates $(x_i, y_i, z_i)^T$ will represent the vertices $p_i(i \in [0, 112])$.

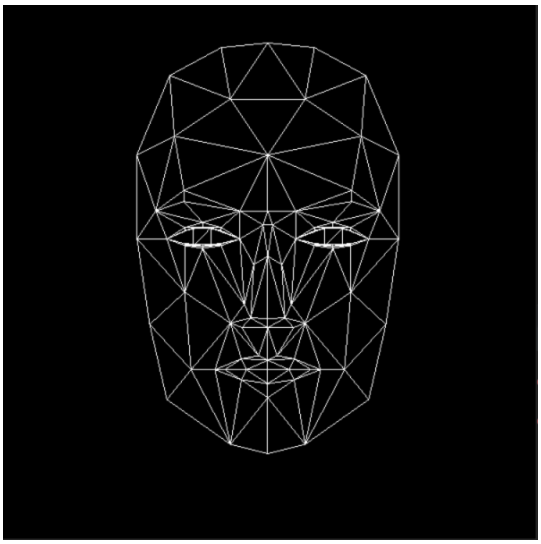

Figure 1. Candide-3

After this, the Candide3 model becomes $C = (p_0, p_1, p_2, \ldots, p_{112})$. It is possible to control the Model by using five parameters: rotation, translation, scale, shape, and animation.

$$C = sR(\bar{C} + A\sigma + S\omega) + t \quad (1)$$

Where $\bar{C}$ is the standard face model, S is the set of shape units representing the face's global shape. It can also be said to be the static shape of Candide3. A corresponds to the animation units; Used to control facial expressions. It can also be said to be the dynamic shape of Candide3. R, s, and t respectively control the rotation, scaling, and translation of the Model. σ , ω are the weights for controlling the dynamic and static control of the face, respectively.

We only need the dynamic expression parameters to drive our Model. Therefore, we can simplify equation (1) by deleting the part of the shape:

$$C = sR(\bar{C} + A\sigma) + t \quad (2)$$

We put all the parameters into a list:

$$Parameter = [scale, r_1, r_2, r_3, t, \sigma] \quad (3)$$

Let the difference squared between the point projected by candide3 and the landmarks as our L2 loss function.

$$Parameter = argmax\ cost(x) \quad (4)$$

Finally, we calculate the parameters to drive our designed Model by using the Gauss-Newton algorithm.

3. Experiment

We decided to run the program first in order to prove whether Gauss Newton Algorithm can fit successfully in a 2D face (Figure 2). After that, we checked whether the expression coefficient can drive the 3D model correctly.

Figure 2. Fitting Algorithm

After running the above tests, we proceed to divide the action units in candide-3: AUV6 (eye closed) and AUV5 (Outer brow raiser) into four different new action Units: Left and Right eyes, Left and Right eyebrows.

We implemented the experiment on a computer with CPU AMD Ryzen 9 5900HS with Radeon Graphics, and the FPS reached 12 frames per second. The results of our experiment can be seen in Figure 3.

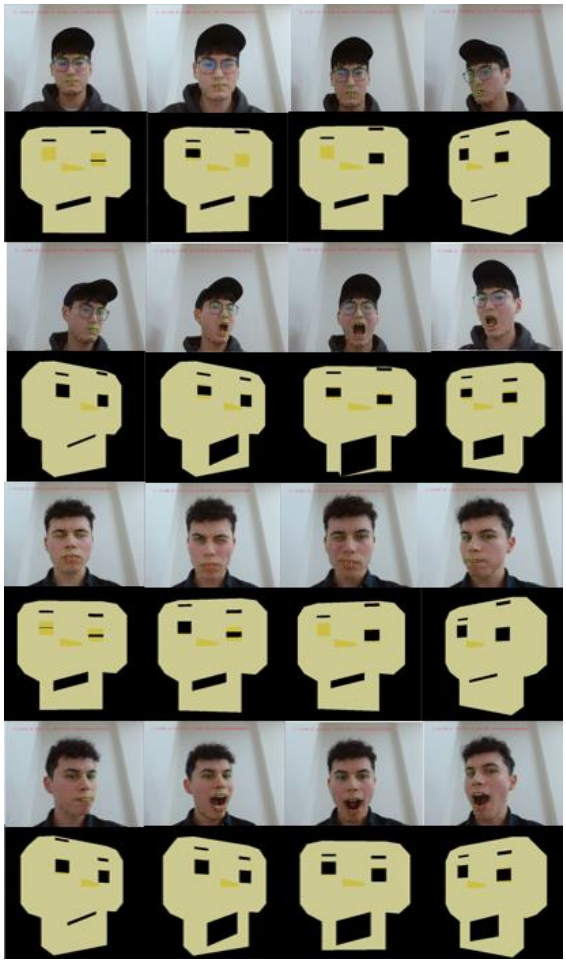

Figure 3. Experiment Results

4. Conclusion

With a reliable and fast algorithm, we can control a 3D model by using a single webcam and complete a 3D reconstruction of a human face. We use this approach to solve common problems that appeared in our new daily life within Covid-19. To accomplish this process, Candide-3, a parametric 3D face model, will control a 3D avatar by using a group of vertices linked to five essential parameters: rotation, translation, scale, shape, and animation. Results showed a successful model that has fully control of a 3D avatar by a webcam at a descent speed. Even though we achieved driving 3D facial expressions actively during the experiments, if we add more expression parameters, the experimental results will be more realistic.

Acknowledgements

"This research was supported by the MISP(Ministry of Science, ICT & Future Planning), Korea, under the National Program for Excellence in SW(2016-0-00023)supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)"(2016-0-00023)

Reference

[1] Gratton, Serge, Amos S. Lawless, and Nancy K. Nichols. "Approximate Gauss-Newton methods for nonlinear least squares problems." SIAM Journal on Optimization 18.1 (2007): 106-132.

[2] Ahlberg, Jörgen. "Candide-3-an updated parameterised face." (2001).

[3] Kazemi, Vahid, and Josephine Sullivan. "One millisecond face alignment with an ensemble of regression trees." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[4] Blanz, Volker, and Thomas Vetter. "A morphable model for the synthesis of 3D faces." Proceedings of the 26th annual conference on Computer graphics and interactive techniques. 1999.

[5] Huang, Jian, Ziling Su, and Ruomei Wang. "3D Face Reconstruction based on Improved CANDIDE-3 model." 2012 Fourth International Conference on Digital Home. IEEE, 2012.

[6] Ren, Shaoqing, et al. "Face alignment at 3000 fps via regressing local binary features." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.