Enhancing 3-Dimensional Facial Modeling and Animation

Research performed at the University of Minnesota-Twin Cities

Rochelle Widmer
Gustavus Adolphus College
Distributed Research Experience for Undergraduates
St. Peter, Minnesota
rwidmer@gustavus.edu

Abstract—Despite advances in computer graphics, generating realistic, real-time facial animations with abilities to express emotional intent remains a challenging problem. This is partly because facial animation is still a developing process with a broad array of accepted solutions. In this paper, we bring two significant aspects of facial animation to life. The first aspect includes presenting real-time, photorealistic 3D animated faces in a more interactive way by using head tracking software. The second aspect involves analyzing the essential parameters required in producing genuine smiles on animated faces.

Keywords—facial animation; data-driven research; computer simulations; user study; stereoscopic display

I. Introduction

A. Facial Animation

The field of facial modeling and animation is a flourishing component of computer science that hopes to revolutionize futuristic technology. Some goals for facial modeling include: creating realistic animations, performing in real-time, automating processes, and varying details to cater to actual human faces. Today, there are a wide variety of applications of face animation including: animated films, computer games, communication, education, virtual reality, surgical simulations, and scientific modeling.

Facial modeling is known to be a difficult process because of humans' familiarity with the face. After years of practice, humans are able to recognize faces and observe subtle changes that express emotions [1]. Because of this competency, generating an animated face that has the capability to produce a genuine smile has become a great challenge for computer scientists.

Another challenge of facial modeling involves demonstrating models in an engaging and believable manor. With humans being 3-dimensional beings, viewing computer-generated faces in 3D is a practical solution. One way to display stereoscopic, 3D images is by using an active 3D TV and active shutter glasses. An advantage of stereoscopic display is that the viewer is able to recognize displays as if they were actual objects instead of images being presented on a 2D screen. Another added benefit of stereoscopic viewing, the user is able to perceive depth.

One of our main objectives is to establish a new method for determining parameterization values that facial surgeons would find useful. Since humans will be interpreting the results of the surgical procedures, we decided to have humans evaluate the effectiveness of certain facial expressions. This required a user study where volunteers would be asked about their opinions. We concluded that the Minnesota State Fair was the optimal location for a user study because of its locations, diversity, and popularity.

II. Background and Tools

Today, facial modeling and animation procedures require the use of numerous methods to construct the best results. Some of these techniques include: blend shapes (or shape interpolation), facial action coding, and parameterization.

A. Blend Shapes

Shape interpolation is currently the most popular method for producing animated faces. A blend shape is a way of altering a mesh to generate a specified appearance. Mathematically, a blend shape model is a linear weighted sum of multiple corresponding generic shapes [2]. To create a blend shape, one must use several exaggerated facial expressions that excessively articulated one action such as mouth widening, upper lip movement, etc. and one natural, resting face. To achieve a face between the resting face and a wide mouth face, one needs to factor in half of the resting face and half of the wide mouth face to get a face halfway between the two.

B. Parameterization and FACS

One is capable of displaying any possible expression by using a combination of parameter values independent of each other. By having control over particular facial movements, the user is allowed to produce a wide range of expressions with a limited amount of computation. It is well known that parameterization is a very tedious process of manually adjusting parameter values on a facial mesh.

Another technique is the Facial Action Coding System. FACS depicts facial movements using a derived method based on the anatomy of the face. There are 44 basic action units (AUs) that combine together to create facial expressions [1]. Some FACS elements include: lip tightener, nose wrinkle, cheek raiser, etc. Typically, FACS is used with muscle and muscle-based approaches. Initially, FACS was never intended to be used for computer-based facial animation [1]. Animation involving muscle models provide a wide range of expressions, however, this method contains movements that facial surgeons cannot control. Some of these actions include: jaw dropping and lid opening and closing. In order to make a system that incorporated procedures that surgeons could perform, one would need to determine a new form of parameterization that includes specific data values.

C. *Software/Hardware Used*

Our approach leverages the Unity game engine to perform real-time visualization of the facial animation. Unity is a game development platform that is typically used to create high-quality 3D games. Unity games are multiplatform and are deployable on mobile devices, PC, virtual/augmented reality devices, websites, and various consoles. Developers are able to program in the editor and in either C#, Java, and Boo. Because of Unity's built-in support for blend shapes, we were able to incorporate our pre-made facial model created with blench shapes.

For head tracking, we made use of the TrackIR infrared-based tracking system. TrackIR is a product made by NaturalPoint; a company dedicated to introducing high-quality tracking technology to the market. The device shown in Fig. 1 is TrackIR camera that we integrated into our project. This device is typically used for gaming because of its ability to track your movements in 3Demensions for optimal game viewing. The product came with a 6 degrees of freedom inferred camera for retrieving signal, inferred reflectors (Fig. 2) to be worn on the user's head, and basic software. With head tracking, the user is able to interact hands-free with the image being presented in front of them.

Fig. 1. NaturalPoint's TrackIR camera.

Fig. 2. NaturalPoint's Track Clip. Attaches to a hat or visor.

III. Real-Time 3D Display of Animated Faces

We started off with a Unity program for adjusting subtle characteristics of a computer-generated face. The user was able to adjust the features of the face by moving slider bars that controlled the face's commissure width, commissure angle, dental top show, and dental bottom show. At this point, the project was only implemented for monoscopic viewing.

A. *Integrating TrackIR and Unity*

Our goal was to add stereoscopic viewing with head tracking abilities to this project. To do this, we used an online Unity plugin that allowed us to get TrackIR data into Unity. The plugin allowed us to access the Cartesian position of the user and Unity would adjust appropriately. For example, when the user moved to the left, Unity would move the image of the face so that the viewer was able to see more of the left side of the face. The same process was used for moving to the right with opposite directions. If the user moved downward, Unity would present an image of the bottom of the face and visa versa for moving upward.

By using a 3D TV and active shutter 3D viewing glasses for users to see stereoscopic images, we incorporated head-tracking software for a 3D viewing program. The program was able to track the user's position using inferred, head-mounted reflectors and a receiving camera. When the user moved, the 3D displayed face would react allowing the viewer to observe all angles of the face using their own motions.

B. *Results*

Fig. 3. Face model in monoscopic view with Unity User Interface on the left.

Initially, the project was in monscopic view and the user was only able to make alterations to the computer-generated face. With TrackIR and stereoscopic viewing incorporated into the project, the user could press the "Init" button and the view would transition to a side-by-side view of the same face where the left image was of the face slightly to the left and the right image was the same but slightly to the right and the head tracking software would begin. We also included a "Shutdown" button to turn off the head tracking functions.

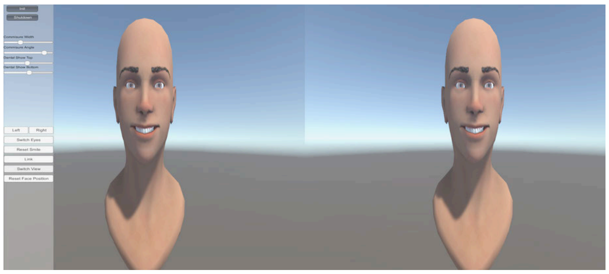

Fig. 4. Face model in stereoscopic view with Unity User Interface.

Fig. 5. Unity User Interface. Includes Init button, Shutdown button, Commisure Width slider, Commisure Angle slider, Dental Show Top Slider, Dental Show Bottom slider, Left and Right buttons, Switch Eyes button, Reset Smile button, Switch View button, and Reset Face Position button.

We were able to get two different images of the face by using two of Unity's camera objects and having one camera be faintly to the left of directly in front of the center of the face and the other camera be faintly to the right. Originally, there was only one camera directly in front of the center of the face. As we were working, we tried to determine the optimal distance between the two Unity camera objects. Since every user has a different distance between their eyes, we implemented buttons that allowed the user to move the cameras together and apart (the right and left buttons). In order to view the two cameras' display next to each other, we included two Unity game objects that rendered the feeds of the cameras. This image was then sent to the 3D TV in the lab via an HDMI cable. With the TV in side-by-side 3D image mode, the images would converge over each other and present itself as a 3Dimensional image to a user wearing active-shutter viewing glasses. Since the face was being rendered in real-time, the image would show as a live feed on the TV as changes were made to alter the face. In order to make the 3D image transition between coming out of the screen and going into the screen, we made a button to switch the locations of the two Unity cameras called "Switch Eyes". With the head tracking activated, the user was able to reset the position of the face to be directly in front of them by pressing the "Reset Face Position" button that sets the Unity cameras back to their original position.

The head tracking software was usually successful. Unity was able to register the user's position by keeping track of the sensors' locations. There appeared to be some bugs with the Unity plugin that would occasionally cause issues with glitchy movements.

We were able to get a few volunteers in the lab to test the program. Most found the system to be enjoyable and stimulating.

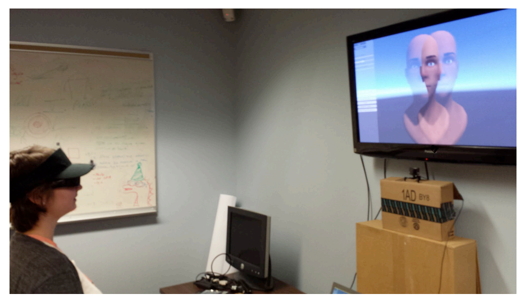

Fig. 6. User viewing 3D face by wearing 3D active shutter glasses and using head tracking device.

IV. Functional Analysis User Study

Building off the experience with the above systems, our next goal was to adopt the system to aide in the study of perceptual aspects of expression emotion using faces. These findings can be especial important in the medical field. For example, there are people every year that suffer from partial facial paralysis due to a wide range of reasons. Often, these people are misjudged and can be difficult to understand. We have combined experts from multiple fields: medicine, psychology, and computer science to tackle this problem. There currently is an abundant amount of research analyzing the psychology of facial and bodily expressions. Unfortunately, there are very limited quantities of data on elements such as how much the mouth moves, angles, and symmetry involved in smiling. With this work, we want to characterize the essential parameters involved in nonverbal facial communication to help assist surgical and therapeutic procedures. We hope that this research will set way for more successful surgical procedures that involve data.

There currently are surgical and medical procedures that can aid those that have undergone facial paralysis but they are not data-driven like our method is. Facial reconstructive surgeons cannot practice procedures on actual humans because of ethical reasons. We were able to create a photo-realistic animated human face using modern computer graphics. By controlling AUPs, Action Unit Parameters, of the animated face, we were able to simulate unique facial expressions. By using this generated facial model, we had access to a much wider variety of expressions than if we had used an actor. We were able to simulate many different examples of facial paralysis using AUPs that we wouldn't have access to otherwise.

A. Study Design

With this study, we wanted to understand what makes a smile look more natural and authentic to a viewer. To do this, we needed to determine parameters of smiling that are crucial. We did this by retrieving data for data-driven research.

B. Approach

Although Unity is typically used with creating games, we used it because of our previous experience with it and because of Unity's compatibility with iPad devices. In order to incorporate the program on an iPad, we made the study in the form of an application.

To start off, a welcome screen is presented to the user to explain our intent. Following the greeting screen, a demographics page asks the users basic information about themselves. The University required some questions be asked of the users to get demographics information (Fig. 7). One questions we decided to ask was about the amount of alcohol consumed by the user to know that validity of the their responses. Before the users are presented with the stimuli, they are directed to a directions screen to briefly explain what they are about to see. The stimuli consisted of videos of our animated face smiling and still images of well-studied faces used by a revolutionary psychologist, Paul Ekman. We wanted to include the Ekman photos in the app so users would have images to reference. Other members of the research group thought it would be best to have the stimuli on one half of the page and a rating system on the other half. First rating screen, shown in Fig. 8, asked about the effectiveness of the expression being displayed. Once the user completed the first rating screen, they were led to the second rating screen as shown in Fig. 9 that asked how much of the 7 basic emotions each of the stimuli expressed. Since the action of smiling takes approximately a second, the videos could be replayed. With the videos, we wanted the users to look at the movements of the smile and not just the still face with a smile. Another aspect of our study was to iterate through several stimuli to get multiple responses from each user. When the user completed all of the necessary components, we sent their information to a University web server by using wi-fi to be later analyzed.

C. Results

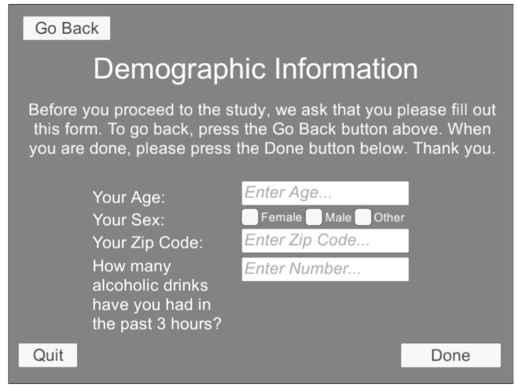

Fig. 7. Demographics page. Includes Go Back button, text boxes and checkboxes for responses, Quit button, and Done button.

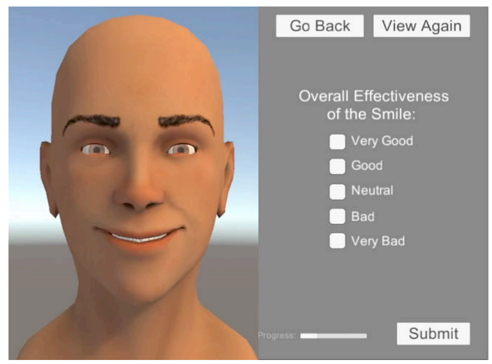

Fig. 8. First rating screen display. Includes Go Back button, View Again button, checkboxes, Progress bar, and Submit button.

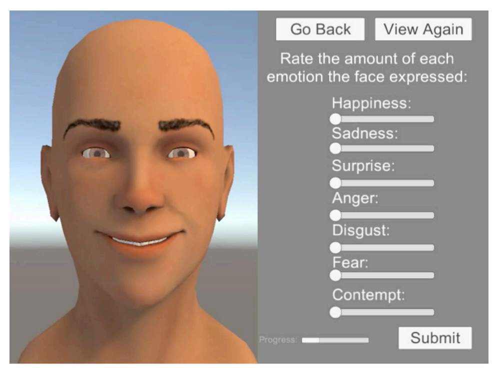

Fig. 9. Second rating screen display. Includes Go Back button, View Again button, slider bars, Progress bar, and Submit button.

V. CONCLUSION AND FUTURE WORK

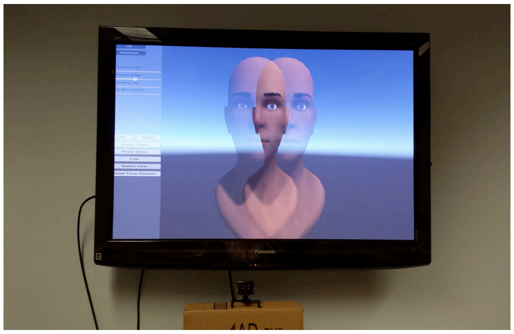

Fig. 10. Image being displayed in 3D mode on 3D TV with TrackIR camera below.

In the future, we could test the user's response to this project and get their feedback on stereo vs. monoscopic viewing. After testing subjects, we could implement their suggestions into the program. Before we would perform such a study, we would first need to find a way around the tracking technology's instability and noise by possibly using filtering, smoothing, and better calibration.

Fig. 11. Univeristy of Minnesota's Driven to Discover Building located at the Minnesota State Fair.

With the data from the State Fair's pilot study, we plan to make a system to measures socio-emotional results to assist with the reanimation process. This assessment tool will judge the facial actions from a socio-emotional perspective. This will be a step up from the current system: a surgeon-evaluated rating system. Determining the quality and success of facial reanimation would have an improved systematic procedure. From there, we would be able to apply this method to individuals with varying facial abilities and assess current approached to rehabilitation.

One major goal we have for this project is to construct a virtual mirror displaying the rehabilitated face of the patient. This revolutionary technology will allow the user to fully interact with an animated image of themselves. This research could also be used for those that have difficulties recognizing facial emotions. Those with autism and socio-emotional processing disorders would benefit from this research. Using insight gained from the user study could help us learn more about abnormalities evident in patients with psychological disorders.

VI. Acknowledgements

We would like to thanks the Distributed Research Experience for Undergraduates program for funding the student researcher. We would also like to thank Nick Sohre, Stephen J. Guy, Moses Adeagbo, Sofia Lyford-Pike, Nathaniel Helwig, members of the UMN graphics department, and State Fair User Study volunteers for their assistance with the project. Lastly, we thank the University of Minnesota for housing the research project.

V. References

[1] Parke, Frederic I., and Keith Waters. *Computer Facial Animation*. Wellesley, MA: K Peters, 1996. N. pag. Print.

[2] Deng, Zhigang, and Ulrich Neumann. "Chapter 1: Computer Facial Animation: A Survey." *Data-driven 3D Facial Animation*. London: Springer, 2008. N. pag. Print.