IJIBC 20-3-8

Case Study of Short Animation with Facial Capture Technology Using Mobile

Gu Jie, Juwon Hwang, Choi, Chulyoung

*Phd. Candidate, Dept of Media Contents, Dongseo University, Korea,
**Instructor, China-Korea Institute of New Media, China
***Professor, Dept of Media Contents, Dongseo University, Korea
*Gj1223010520@gmail.com, **jamesjuwon@naver.com, ***freechoi@dongseo.ac.kr

Abstract

The Avengers film produced by Marvel Comics shows visual effects that were impossible to produce in the past. Companies that produce film special effects were initially equipped with large personnel and equipment, but technology is gradually evolving to be feasible for smaller companies that do not have high-priced equipment and a large workforce. The development of hardware and software is becoming increasingly available to the general public as well as to experts. Equipment and software which were difficult for individuals to purchase before quickly popularized high-performance computers as the game industry developed. The development of the cloud has been the driving force behind software costs. As augmented reality (AR) performance of mobile devices improves, advanced technologies such as motion tracking and face recognition technology are no longer implemented by expensive equipment. Under these circumstances, after implementing mobile-based facial capture technology in animation projects, we have identified the pros and the cons and suggest better solutions to improve the problem.

Keywords: Facial Capture, Animation, ARkit, Virtual YouTuber, Facial Rig

1. Introduction

The demand for facial capture is based on the results of the unrealistic facial expression of Uncanny Valley. Director 'Robert Zemeckis', who produced the past film 'The Polar Express (2004)', noted that he saw new possibilities in filmmaking through performance capture. It was expected that motion-captured digital actors could be repositioned and re-filmed at the editing stage, which was a production process that was impossible in conventional film production, so motion-capture film production would develop rapidly[1]. However, the movie 'The Polar Express' failed to hit the box office as the characters' facial expressions were not realistic, even though they applied facial capture techniques to the real actor's facial expressions. To overcome the limitations of facial capture, director 'Robert Zemeckis' tried to connect with the Bland Shape

in his next work as the executive producer, 'Monster House (2006)', to overcome the limitations of the interim facial expressions of data collected by facial capture[2][3]. This method seemed to overcome the limitations of existing animation by expressing cartoon characters very effectively. However, there was still an uncanny valley for realistic characters, and the addition of eye-tracking technique 'EOG(Electrooculogram)' to produce natural eye movements in the film 'Beowulf (2007)' resulted in a very natural digital act[4][5].

If the digital act in the movie has tried to establish itself in a realistic form, then in our reality, cartoon characters are making many changes to take their place. The software, known as 'MMD (MikuMikuDance)', allows users to freely create their own music videos using virtual characters to promote games. In order to expose this worldwide, YouTube is making virtual character's music videos as popular as real singers. Game engine software has enabled real-time performances of virtual characters using motion capture or facial capture equipment, which has led to the birth of virtual YouTubers such as 'Kizuna I (A.I. Channel)'[6]. Virtual YouTubers' activities are increasing their IP value by engaging in a variety of activities, including advertising and music videos, as well as YouTube. Recently, the scope of virtual characters' activities has been expanding to virtual spaces. Users are not only doing the characters and activities provided to users in the virtual space called 'VRchat', but also developing and making various digital characters and movements using modeling software and game engines to expand activities in the 'VRchat' space and improve the quality of virtual characters [7][8].

These high-quality virtual characters' activities reflect the fact that the popularization of equipment that delivers real-time user behavior to virtual characters is accelerating. Expensive motion capture equipment has enabled the real-time operation of virtual characters using Microsoft's Kinect, while facial capture linking mobile and game engines is enabling delivery to virtual characters. Apple Inc.'s iPhone is equipped with a camera-based face capture system called 'trueDepth', which allows users to obtain more accurate facial expression data[9]. Non-experts are also sharing and improving information about equipment and software as they engage in YouTube activities using virtual characters. The delivery of information is important in such YouTube activities, so the naturalness or accuracy of facial expressions is not required. This advantage can be attributed to the increased activity of cartoon characters as virtual characters. Under these circumstances, we wanted to apply the mobile-based facial capture to the short animation production process. In the text, we tried to establish an effective production process for the application of animation by analyzing problems that may arise during the implementation process and looking for improvements to them[10].

2. Make face capture using mobile

Recently, as facial expressions such as 'Animoji' have become possible using mobile, videos that implement facial capture have been released online. The iPhone is equipped with a 'trueDepth' camera, which enables more accurate facial expressions compared to Android, so the iPhone is used to implement facial expressions. Furthermore, as shown in Figure 1, we proposed a reasonable manufacturing process to incorporate the iPhone into the pipeline of an animation project. As you can see in Figure 1, we have created a moving structure between software and improved production efficiency so that we can shorten the production time of animation projects and reduce data errors.
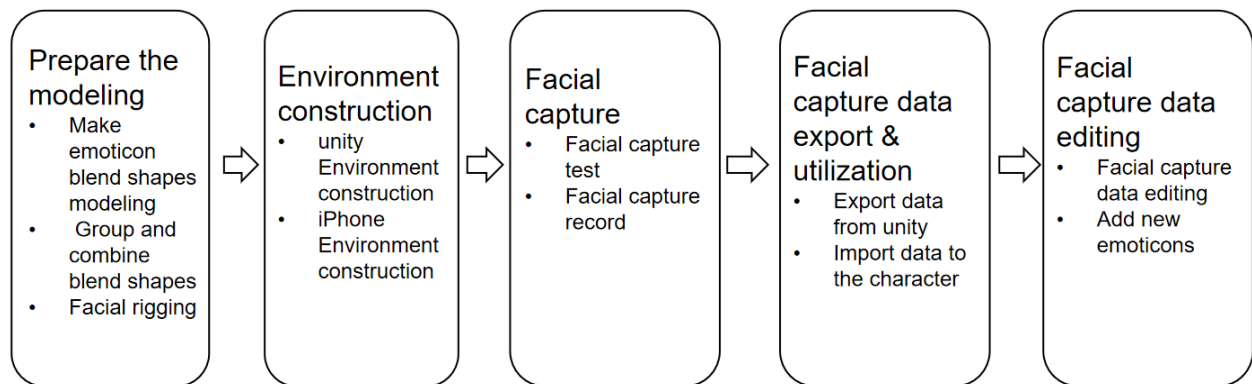
Figure 1. manufacturing process to incorporate the iPhone into the pipeline of an animation

Facial expressions are delivered to the Unity Game Engine via the iPhone. In Unity, which is connected to the iPhone and the network, capture data of facial expressions can be checked in real time through the expressions of CG characters. If the data output is good, it will be forwarded in the 'FBX' file format of Maya software. But in this case, problems arise. Character facial expressions in animation need more than just talking. An animator who works on the need for more exaggerated expressions than is real when applying facial expressions or other simple facial expressions to animated characters. Data made in the 'FBX' format cannot modify facial expressions. Therefore, data must be stored in the Maya format from Unity to modify the expressions produced in the Maya program through 'shape editor'. For this task, the Unity plug-in was created to enable the saving of face data in an 'mb' file format. Copying each facial curve from the curve editor and pasting it into the CG character's 'facial blend shape curve' applies the facial expressions captured in the capture data of the facial expression brought to Maya. The order of copying and pasting in the curve editor depends on the order of the items selected in the list, as shown in Figure 2. The order in which 'blend shape' is created is determined in the curve editor. Therefore, all expression data should be kept in the same order when creating a blend shape. Many characters appear in the project, and the process of applying blend shapes for facial capture must be maintained in the same order of blend shapes in order to work smoothly. If you "bind" in a different order, the data will not be applied correctly to the character.

3. Proposal for Rigging to Strengthen Facial-captured facial Expression

The reference blend shapes for the facial capture provided by 'ARkit' were very different in form from the blending shapes needed to produce facial expressions in animation. The facial expression acting of an animated character, which is made with only facial capture, is not enough to be applied to the shot as it is, so an additional blend shape should be applied. For example, in the case of 2D animation, the facial expressions of the characters are exaggerated, but the facial expressions do not express the exaggerated feelings as much, so simply applying facial capture to the characters makes them look boring. Two methods can be considered to apply the method of exaggerating facial expressions to animations. One method is to apply the blend shape made for the existing facial capture to the character once again and make additional modifications separately from the facial capture, while the other is to apply a joint to the face to modify the facial expression. The additional process for testing was carried out in the following order:

   a.   How to apply the blend shape provided by 'ARkit' to the character once more to modify the additio

nal facial expressions

b.   How to modify facial expressions by adding joints to the face

In the case of 1, you may consider creating and applying facial expressions for animation with a blend shape, but this method takes too much effort to modify facial expressions. Therefore, we added another blend shape for the existing 'ARkit' as shown in Figure 2 to make it possible to modify the blend shape.
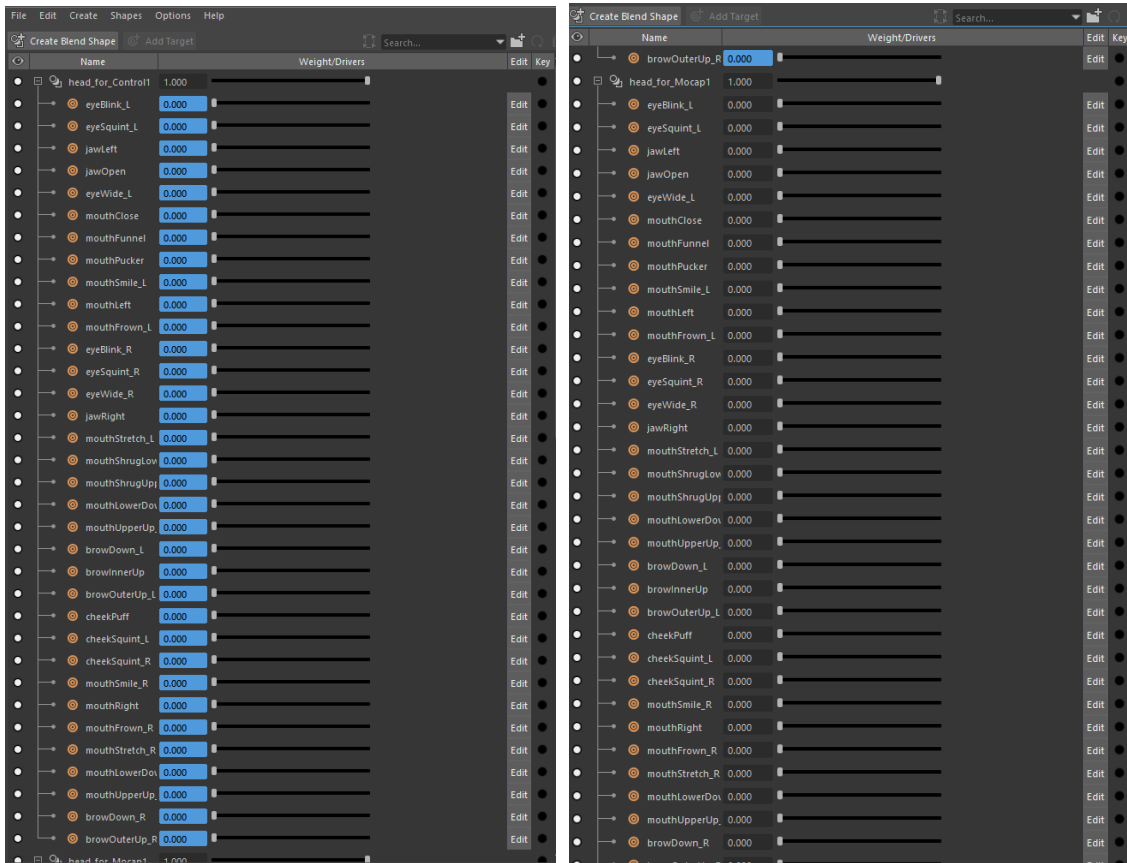

Figure 2. Blend shape for "ARkit" (left), Blend shape for additional modification (right)

Modifying facial expressions by applying additional Bland Shapes is not a success in making facial expressions rich. Blend Sheaf for Facial Capture is not easy to make an exaggerated expression because it differs a lot from the existing blend Sheaf for making facial expressions. For example, when working on a character's mouth, in the case of a 2D animation, the edge of the mouth is tightened and exaggerated in the form of a '口'.

Blend shape made for facial capture cannot appear because it mimics the shape of the mouth of a real person's mouth. In general, the shape of the mouth is expressed in exaggerated descriptions in animation, and only a few unique forms can create enough dialogue scenes. In other words, in animation, the form of facial expressions expressing emotions is more important than the movement of the mouth to talk. From this point of view, the reference blend shape provided by Apple has only about five blend shapes for facial expressions that meet the needs of animation production. In general, it is the change in the position of the shoulder and the position of the eyebrow that creates the most easily distinguished form when expressing human emotions. Judging from the above, the first method is an ineffective method to expect great results in order to modify the shape of the mouth and eyebrows for emotional expression. To solve the problems revealed in the first

method, we added a joint and tried the second method of modifying the desired expression as shown in Figure 3. Looking to the left of Figure 5, you can see the control handle for applying the additional blend shape shown in the first method to add an emotional expression. To move the eyebrows, three joints and three movable handles were inserted around the eyebrows of the character model. The eyebrows are added to the results of the facial capture to make an exaggerated expression possible.
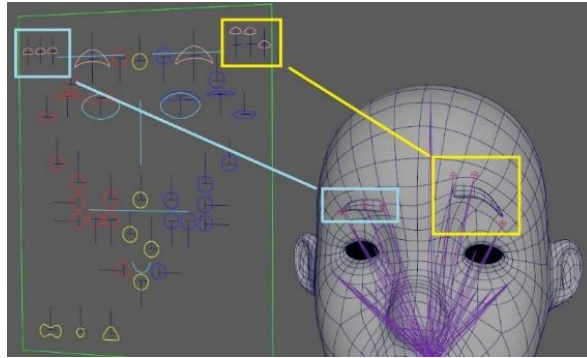

Figure 3. Controllers that can control additional blend shapes and joints

Shape-shifting around the eyes is also important when it comes to emotional expression, as there is no blend shape required by 'ARkit', so you can't make a look of embarrassment or coyness because you can't move up the lower part of the eyes. As shown in the figure, an additional five joints were applied here and three controllers were used to control the movement of the lower part of the eye. In this operation, the x, y, and z axes all had to move, so the controllers were made to be fixed directly to the top of the face joint.
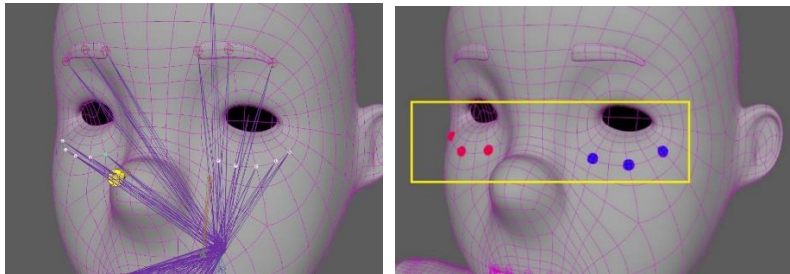
Figure 4. Add joints to move the lower part of the eye

Twenty-seven of the 52 reference blend shapes provided by 'ARkit' correspond to the blend shape of the mouth. But the 27 mouth-blend shapes are not enough for the work of cartoon-style animation because they are made to be suitable for vocalization. The Japanese animation style evolves in cartoons, and the shape of the mouth is expressed in angular form, as shown in the left image of Figure 6, expressing a very exaggerated and dynamic feeling. As a way to implement this type of mouth, you can think of creating additional Bland Shape and adding a joint. If the form is already being made through facial capture, it is not easy to expect the exact form intended even if additional blend shapes are applied. In such cases, more accurate facial expressions can be made by shape-shifting using a joint. Therefore, we wanted to add 14 joints around the mouth as shown in Figure 5 to express the shape of the mouth as shown in the 2D animation. If you look at the middle image and the right image of Figure 5, you can see that the shape of the mouth with angles like the one seen in the cartoon character is created even if you apply the joint on top of
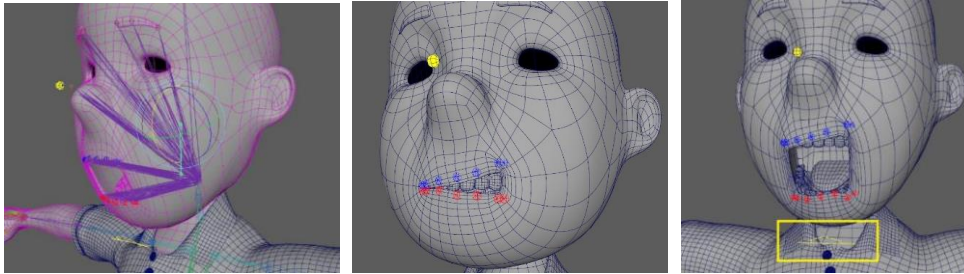
the face-captured mouth.

Figure 5. Additional joints for moving the mouth.

Nevertheless, the effect on the shape of the mouth is limited by adding bones around the mouth. Adjusting with the joints around the mouth after face capture is only a partial complement or improvement. If an exaggerated form of mouth is required, as shown in Figure 6, it is appropriate to add a new type of blend shape as needed. Since additional blend shapes are often applied to specific shots, it is better to create and proceed with additional production schedules after mid-production than in the early stages of production.

Figure 6. Adding a blend shape with an exaggerated form

So far, we have looked for ways to increase production efficiency by establishing a production process for facial capture animation using iPhone and producing short animation. To compensate for the problems encountered in applying facial capture data, we created a controller that can be corrected by applying additional blend shapes for facial capture as shown in Figure 7. As a way to overcome the limitations of facial expressions that still occur, exaggerated facial expressions such as Figure 8 were possible in the final animation results.
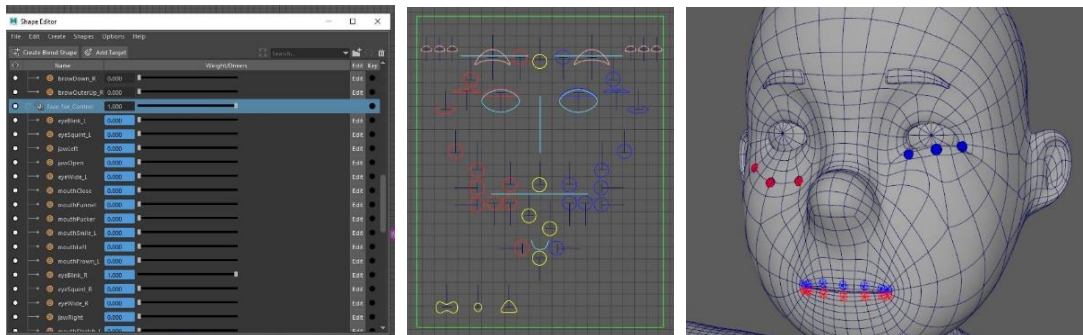
Figure 7. Two Recommendations to complement Facial Capture and related problems

Reviewing the process so far, additional blend shapes to embody exaggerated facial expressions or joint applications to create additional facial expressions should be accompanied in order to make a unique production while keeping the characteristics of animation when planning to produce an animation using mobile facial capture.
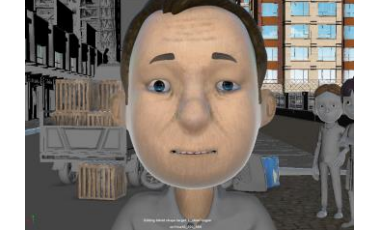
Emotion	Only Facial Capture	Add Blend Shapes	Add Joint Handles
Happy			
Angry			
Sad			
Awkward			
Shock			

Figure 8. Images according to additional facial expression change function

4. Conclusion

Animation can be said to be a way of expressing images. We may stick to the traditional way of making, but new technologies are showing us a new way of creating. With some technical support from programmers, the threshold for video presentation due to the development of mobile technology is lowered to the point where video producers can produce complex video content using facial capture technology. YouTube

contents using virtual characters produced by ordinary people who isn't professional are being uploaded every day. However, it is considered that the study of facial capture and additional necessary facial production process should be preceded by the character concept since the exaggerated and individual facial expressions are crucial factors due to its characteristics. Recently, the level of demand for the quality of virtual YouTubers' gestures and facial expressions using CG characters has been increasing. Most of these characters have the form of the Japanese cartoon character, and in 'VRchat', users of virtual society participate in the game using these cartoon characters. Textures and modeling technologies are being unveiled that make the form of 3D modeling applied look like 2D animation characters in these cases. At present, facial capture is applied directly to 2D characters, but the timing and form of application should be approached differently from real-life people. As a follow-up study, we are going to conduct a study that allows access to 'Cell animation' using the mouth and facial form of 2D virtual characters using facial capture.

References

[1] IAN FAILES, From Speed to Spidey: 20 years of VFX and animation.
https://www.fxguide.com/fxfeatured/from-speed-to-spidey-20-years-of-vfx-and-animation/?highlight=The%20Polar%20Express.

[2] Prnewswire, Digital Production Studio Gives Moviegoers A Multidimensional Experience. *http://www.prnewswire.com/*

[3] Sony Pictures Imageworks. *http://www.sonypictures.com/imageworks/index.php*

[4] Andrew Edgecliffe-Johnson, COMPANIES & MARKETS: Beowulf to lead audiences into new 3D realm.
http://search.ft.com/ftArticle?queryText=Beowulf&y=8&aje=true&x=19&id=071112000028&ct=0

[5] Bob Tourtellotte and Eric Walsh, Epic "Beowulf' gets high-tech, Hollywood makeover.
http://news.yahoo.com/s/nm/20071115/film_nm/beowulf_dc_3

[6] MIKE SEYMOUR, ehind the Scenes at SIGGRAPH Asia's REAL TIME LIVE.
https://www.fxguide.com/fxfeatured/behind-the-scenes-at-siggraph-asias-real-time-live/?highlight=The%20Polar%20Express.

[7] Kaylee Fagan, A large number of people have come out saying VRChat has saved their lives here's what it's like to experience the online meeting place of the 21st century.
https://www.google.com/amp/s/www.businessinsider.com/vrchat-explained-2018-2%3Famp

[8] Developer.Vive, VRChat Partners With HTC And Makers Fund To Close $10m Series C. *https://developer.vive.com/resources/2019/09/17/vrchat-partners-with-htc-and-makers-fund-to-close-10m-series-c/*

[9] MIKE SEYMOUR, BebyFace in Bebylon. *https://www.fxguide.com/fxfeatured/bebyface-in-bebylon/?highlight=ARKit.*

[10] MIKE SEYMOUR, ARKit and UE4. *https://www.fxguide.com/fxfeatured/arkit-and-ue4/?highlight=Animoji*

[11] Yang Pan , Chul-Young Choi , Zilu Meng. Research on the Development of One IP(intellectual property) Animation & Game in Chinese Marke. The Institute of Internet, Broadcasting and Communication(IJIBC), Vol.11 No.2. May 2019.
DOI: *http://dx.doi.org/10.7236/IJIBC.2019.11.2.67*

[12] Ji-Woo Han. A Study on Effects of AR and VR Assisted Lessons on Immersion in Learning and Academic Stress. The Institute of Internet, Broadcasting and Communication(IJIBC), Vol.10 No.2. June 2018.
DOI: *http://dx.doi.org/10.7236/IJIBC.2018.10.2.4*