Development of Web Content for Music Education Using AR Human Facial Recognition Technology

Eunee Park

Division of Media Arts, Baekseok Arts University, Republic of Korea
E-mail: euneepark@bau.ac.kr

Abstract

As the media market changes rapidly, market demand is increasing for content that can be consumed on web platforms. It's required to produce differentiated web content that can attract viewers' interest. In order to increase the productivity and efficiency of content creation, cases of content production using AR engines are increasing. This study has a development environment in which parametrics and muscle-based model techniques are mixed. The faces of famous Western classical musicians, such as Mozart, Beethoven, Chopin and List are created as 3D characters and augmented on human's face based on facial recognition technology in this study. It analyzes and traces the changed of facial expression of each person, then apply to 3D character's facial expression in real-time. Each person who augmented musicians' faces can become those who lived in different times, deliver information and communicate with viewers of the present era based on the music educational scripts. This study presents a new direction for video production required in the media market.

Keywords: Augmented reality (AR), unity engine, face recognition, facial expression, real-time tracking, YouTube, web content.

1 Introduction

In recent years, the global media market has undergone a significant shift towards a new media market. There has been a growing interest in producing various forms of new media content that can create differentiated content to attract viewers on mobile and internet media [1]. The recent advancements in technology in game engines such as Unity and Unreal have expanded beyond game development into a wide range of fields including animation, movies, education, architecture, and virtual spaces [2]. While virtual characters have been mainly used in games, the emergence of virtual YouTuber's like SUA and Schubl on diverse media platforms is a noteworthy phenomenon [3]. The potential for new media platforms and their ability to reach audiences globally has opened up opportunities for creators to produce unique and engaging content. As the demand for such content continues to rise, new and exciting opportunities for content creators are emerging.

Against this backdrop, this study focuses on the development of an augmented virtual teacher using human face recognition technology. The virtual teacher is designed to be implemented and linked in real-time, enabling the development of video content for classical music education. By utilizing cutting-edge technology, this study seeks to provide a new and innovative approach to music education that combines traditional teaching methods with the latest technological advancements.

Prior to this study, 3D modeling of 2D characters from the TV series animation, "Mind Blowing Breakthroughs", which aired on EBS in April 2020, was created. After licensing HYPRFACE, HYPRESENSE's facial motion capture software [4], and applying it to modeled 3D characters, 3D avatars on real actors and actor's faces joined the cast under the theme of music-related content. It was produced by shooting on the chroma key set and completing the content through the post production. A new type of video production method was implemented that combines game engines by using the technology of synthesizing the background taken through the filming of the chroma key set. Through this, more diverse backgrounds and situations can be realized, and video production can be made quickly.

Using these technologies and techniques, this study will prove that more diverse and realistic video work is possible by presenting new forms of video production methods. This will enable higher levels of the video content production to be done using more advanced technologies in the new media platform.

2 Literature Review

2.1 Production Techniques of Facial Expression Animation

Facial expression animation of 3D characters is used as an essential element in various fields such as games, animations, and movies. Various techniques are used to implement this facial expression animation, such as interpolation technique, parametric technique and muscle-based model technique [5].

The interpolation technique is a keyframe molding technique, which sets the beginning and end points of the facial expression as the keyframe and automatically creates the middle frame to create natural facial movement. This method is one of the most commonly used methods in animation production.

The parametric technique is a method of obtaining various facial variation data by creating parameter controls that can control facial movements. This method is widely used in the field of computer graphics, and various facial expressions can be defined as parameters in advance and adjusted as necessary.

The muscle-based model technique applies actual types of facial muscles and movement data to 3D models, expressing facial expressions more naturally and realistically. This method is implemented by collecting and analyzing data related to facial muscles and applying them to 3D models. Waters [6] classified the types of facial muscles that mainly contribute to expressing expressions as linear muscle, sphincter muscle and sheet muscle according to the direction of movement of each muscle. The linear muscles move obliquely with a series of muscles centered on one point, such as the Zygomaticus major which raises both ends of the mouth. The sphincter muscles are a series of muscles that move horizontally in one direction, such as Depressor anguli oris, Depressor labii inferioris and Occipitalis. The sheet muscles are movements that center on one point and all the muscles around it move toward that point, such as frowning or closing the mouth [6].

The motion capture technique dataizes the flow of real faces and creates facial expressions of 3D characters based on them. This method is that when a real actor or model makes a facial expression, it is captured and applied to the 3D model to realize natural facial movement.

2.2 Facial Expression Animation

Facial expressions play a very important role in how a person expresses emotions. Through the facial expressions, the emotions of the other person

can be understood and empathized with and the feelings of the other person can be understood. Therefore, the facial expressions are considered one the most basic means of communication.

The facial expressions are created through the contraction and relaxation of facial muscles. These muscles are located in the head and neck areas and, when they contract, the tissues connected to the skin move. These movements create wrinkles, lines, and contours on the surface of the face and allow for a variety of facial features to move. Facial muscles can be divided into two major groups. One group consists of expression muscles that are mainly located around the eyes and mouth, and the other group consists of expression muscles located in the head and neck areas. These muscles perform different functions and work together to create each facial expression [7].

The facial expressions also play an important role in animation. In order for animated characters to come alive on the screen, emotions must be expressed through various facial expressions. To this end, animators design various facial expressions to express the emotions of characters and reflect them in animation.

2.3 FACS (Facial Action Coding System) Theory

In the field of psychology, Ekman stated that there is connection between a person's emotions and facial muscles through neural pathways, and that facial expressions are generated as a result of internal emotions, and he argued that, out of the complex and subtle range of human emotions, only six emotions are identifiable through the facial expressions, and they are universally recognized. These six emotions are fear, surprise, anger, disgust, sadness and happiness. According to Ekman, all other emotions are combinations of these basic categories.

The facial action coding system (FACS) is widely used as a method for generating basic facial expressions in facial animation models. FACS, developed by Ekman and Friesen, classifies the movements of muscles involved in facial expressions bellow in Table 1. It represents all visually identifiable movements. FACS classifies muscle movements into 46 combinations of action units (AUs), allowing for the creation of various facial expressions. It should be noted that FACS excludes the changes that the passages of time or muscle movements may have on the overall expression [8].

Table 1   Main action units

AU	Description	Facial Muscle
1	Inner brow raiser	Frontalis, par medialis
2	Outer brow raiser	Frontalis, par medialis
4	Brow lowerer	Corrugator supercilii, Depressor supercilii
5	Upper lid raiser	Levator palpebrae superioris
6	Cheek raiser	Orbicularis oculi, pars orbitalis
7	Lid tightener	Orbicularis oculi, pars palpebralis
9	Nose wrinkler	Levator labii superioris alaquae nasi
10	Upper lip raiser	Levator labii superioris
11	Nasolabial deepener	Zygomaticus minor
12	Lip corner puller	Zygomaticus minor
13	Cheek puffer	Levator anguli oris (a.k.a. Caninus)
14	Dimpler	Buccinator
15	Lip corner depressor	Depressor anguli oris (a.k.a. Triangularis)
16	Low lip depressor	Depressor labii inferioris
17	Chin raiser	Mentalis
18	Lip puckerer	Incisivii labii superioris and Incisivii labii inferioris
20	Lip stretcher	Risorius w/platysma
22	Lip funneler	Orbicularis oris
23	Lip tightener	Orbicularis oris
24	Lip pressor	Orbicularis oris
25	Lips part	Depressor labii inferioris or relaxation of Mentalis, or Orbicularis oris

The action units described in Table 1 show the different movements of facial muscles. Certain combined movements of these facial muscles pertain to a displayed emotion. The following table presents a comprehensive list of AU combinations and the emotions they are associated with. Some examples of action units commonly associated with specific emotions are shown in Table 2.

When AUs are mutually combined, it is possible to define the characteristics of the changes of facial expressions according to the degrees of between eyebrows and trembling of the eyelids, contraction of lips or up and down movement of the lips.

Table 2   List of AU combination and related emotions

Action Units Combination	Emotion
AU6 (cheek raiser) + AU12 (lip corner puller)	Happiness
AU4 (brow lowerer) + AU15 (lip corner depressor)	Sad
AU1 (inner brow raiser) + AU2 (outer brow raiser)	Surprise
AU1 (inner brow raiser) + AU2 (outer brow raiser) + AU25 (lips part)	Fear
AU1 (inner brow raiser) + AU2 (outer brow raiser) + AU5 (upper lid raiser)	Disgust
AU6 (cheek raiser) + AU9 (nose wrinkler) + AU12 (lip corner puller)	Contempt
AU6 (cheek raiser) + AU7 (lid tightener) + AU10 (upper lip raiser) + AU12 (lip corner puller)	Anger
AU4 (brow lowerer) + AU5 (upper lid raiser) + AU7 (lid tightener) + AU23 (lip tightener)	Frustration
AU1 (inner brow raiser) + AU2 (outer brow raiser) + AU5 (upper lid raiser) + AU25 (lips part)	Amusement
AU6 (cheek raiser) + AU12 (lip corner puller) + AU17 (chin raiser)	Relief

3 Method

3.1 Design of Production System

HYPRFACE, a system for this study, has a development environment with a mixture of parametric and muscle-based model techniques, which enables more natural and realistic 3D facial expression animation. The system is designed to produce content in real-time when the user puts his or her face directly on the camera and the character is recognized in the application. The production method takes a picture of a real person and allows an AR engine to match the face of a 3D character to instantly identify the character augmented to the real person's face in real-time through a display monitor or mobile [9]. Therefore, the real person can shoot while monitoring augmented characters, so it can be checked with various facial expressions in real-time and reduce the range of errors. As shown in Figure 1, recognizing a user's
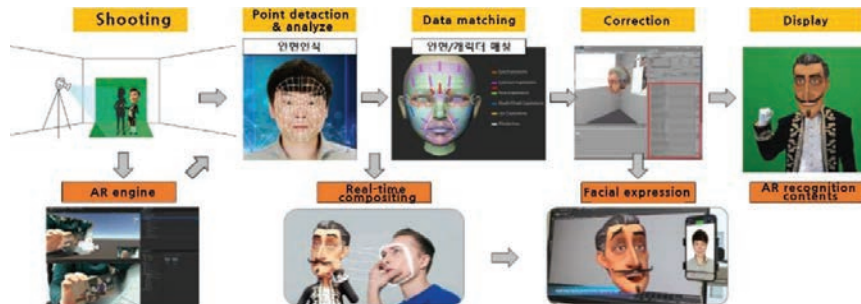

Figure 1   The system diagram.

facial features, quantifies the location change of each part of the face and shows the process of applying a blendShape to the 3D character. There is an advantage that the real person can reduce the production process such using a more realistic performance, reduction of key animation time and post recording through the facial expression and body motion performance. When shooting a set of the studio, not the chroma key, it is possible to produce and send content in real-time without a separate background compositing.

3.2 Analysis and Design of HYPRFACE

The HYPRFACE software used in this study is optimized for both PC and mobile environments and supports IOS, Android, Window OS and Mac OS. Figure 2 shows two sample faces provided by HYPRESENSE, which shows how to operate the code, a basic recognition UI, and facial expression guide for the weight of blendShape of HYPREFACE, as shown in Figure 3. It increased the accuracy and sensitivity of facial capture.

As shown in Figure 4, the sample face modeling provided by HYPRFACE's SDK shows how the code works and it provides a basic

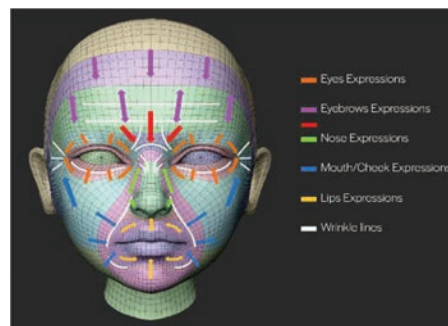
Figure 2   The sample scene provided by HYPREFACE's SDK.

Figure 3   Guidance of facial expression areas for setting the weight of the blendShape.

Figure 4   Guidance of facial expression areas for setting the weight of the blendShape.

recognition UI. It was tested whether it worked in conjunction with the basic example face and the face modeling of a 3D character built for this study using a Window OS based kinect 4k camera. Multi-face recognition was tested, as shown in Figure 5. The faces of musician characters like Mozart, Beethoven, Chopin and List are modeled and mapped. Figure 6 is the modeled musical characters for augmented reality.

After setting up the joint provided by HYPRFACE about 36 blendShape expressions were created, and the blendShape of HYPRFACE created a recipe file corresponding to the naming level and implemented on Unity 3D. The face figures of the produced 3D characters can be easily linked to the blendShape. Unlike the basic character provided in the sample, the face position of the 3D character produced through this study was incorrectly

Figure 5   Multi-face recognition test.

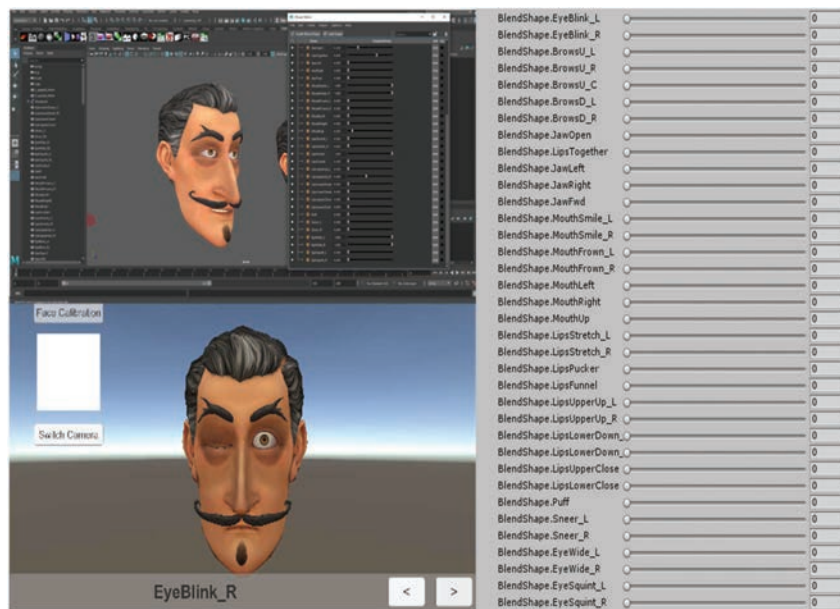**Figure 6**   Modeling and mapping of a classical musician.

Figure 7   BlendShape list of HYPRFACE.

recognized or there was a problem in synchronization according to various angles, so this was solved by initializing the frame values and meshes for each of face modelings. When the ratio of the 3D character's face was adjusted to the human body, the further away from the camera, a position value error occurred, or when the face was turned, the augmented 3D character's face deviated from the user's face. The blendShape list provided by HYPRFACE is shown in Figure 7 and it was tested whether it is normally expressed by applying facial expressions of 3D characters according to each weight of the blendShape.
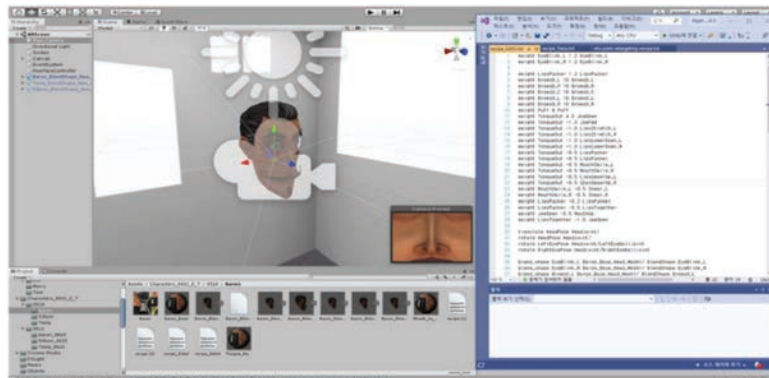
Figure 8　Real-time combination system of character expression.

For this study, fifty basic blendShapes for each character were created. Recipes for individual facial expressions were developed and detailed sensitivity in the facial recognition process was adjusted so that facial expressions can be recognized and combined in real time in HYPRFACE software, as shown in Figure 8.

4 Results and Discussion

This study was to produce classical music education content for YouTube in a chroma key set using virtual teachers. In the production, when actual actors act in front of the camera, the face of the character designated through HYPRFACE software is put on the face of the individual actors and sent to the monitor installed next to it. It allows to compare the actual actor's appearance at the time of shooting with the face of the augmented character through the monitor, as shown in Figure 9.

Both producers and actors can see augmented faces through monitors in real-time, so it is possible to monitor expression tracking errors and

Figure 9　AR characters combining with the real actors.

Figure 10 Compositing in real-time with actual actors and transmitted to the monitor.

Figure 11 Output after compositing and visual effect.

recognition problems in real-time and solve problems. As shown in Figure 10 below, the video footage of AR characters combining with the real actors are sent to the side monitor in real-time.

The green screen taken from the chroma key set was removed in the video, then the background was composited. Some infographics were added to areas where it was necessary to add effects or convey information as necessary. Through the background compositing, it was possible to reduce the sense of difference from the augmented 3D character on the actual actor's face, and to increase the completeness of the content by placing graphic elements on the screen to deliver the information effectively. Figure 11 shows images of the result applying a graphic background, motion graphic, effect and typography in the video footage.

It would have taken about two months to produce a ten-minute length animation, but it contributed to reducing the actual production time as the animation production process was omitted by producing content using AR facial recognition technology. The new production process using this AR facial recognition technology shows its effectiveness as a new media content production tool as well as shortening the production period to shooting, editing and compositing.

However, the HYPRFACE software used in the production was optimized for the sample characters provided, making it difficult to track eyelid tremors or fine facial expression changes when applied to other modeled characters. In addition, when the actor turned his or her face from side to side, it was

possible to shoot only within a limited space due to restrictions on the actors' movements when filming due to the departure of the augmented character's face location.

5 Conclusions

This study proposed and verified the effectiveness of an AR image production system based on human facial recognition technology when more advanced technology is required in preparation for the post COVID-19 era. These technologies are useful to reduce the time and cost problems incurred in production as a new content production tool for the new media platforms.

The combination of augmented reality and character animation offers a unique interactive element. Users can actively participate in the animated world by interacting with virtual characters through gestures, voice commands, or even facial expressions. This level of interactivity enhances user engagement and immersion, making the overall experience more captivating and memorable. As this technology continues to evolve, it is possible to expect further advancements in the integration of augmented reality, character animation, and human facial recognition. This will undoubtedly shape the future of animation, pushing the boundaries of creativity and storytelling.

In future research, software specially optimized for modeled characters using APPLE's AR kit will be developed to solve the problem of tracking fine facial expressions and going out of sync according to the angle [10]. By leveraging the capabilities of ARKit, which offers advanced tracking and rendering features, it is expected that the system will be able to accurately capture and replicate even the subtlest facial movements, resulting in more realistic and expressive animated characters. Furthermore, additional research is being conducted to establish a production system that enables real-time transmission of the captured footage. The goal is to create a seamless workflow where the images can be immediately uploaded online while the filming is still in progress. This real-time transmission capability would revolutionize the animation production process, allowing for faster feedback, collaboration, and distribution of content, opening up new avenues for interactive storytelling and entertainment.

The integration of the optimized software and real-time transmission system holds immense potential for the animation industry. It would not only address the existing limitations but also streamline the production process, reducing the time and resources required for post-production editing and rendering. Additionally, it would foster greater connectivity and engagement

between content creators and their audiences, enabling instant sharing and feedback. Additional research will be conducted to establish the production system of real-time transmission that can immediately upload images online at the same time as filming based on the optimized system. As these advancements continue to unfold, it is anticipated that the field of animation will witness significant advancements in terms of realism, interactivity and efficiency.

References

[1] H. J. Kim, "The approaching direction of producing animation contents based on new media", Korea Digital Design Council, Digital Design Studies, vol. 10, no. 3, pp. 185–195, Oct., 2010.

[2] H. Y. Jeon, "Domestic and foreign AR/VR industry status and implications", Hyundai Economic Research Institute, Seoul, Korea, no. 687, 2017.

[3] H. G. Seo, "Domestic virtual characters that go beyond YouTubers and game promotion models", gamemeca.com, https://zrr.kr/yDZE, (accessed February 1, 2021).

[4] Hypresense, hypresense.com, https://www.hyprsense.com (URL), (accessed February 1, 2021).

[5] Y. G. Kim, "A study on marker tracking research for utilization in AU based on facial motion capture -Based on low polygon", The Korean Journal of Animation, no. 10(4), pp. 45–60, 2014.

[6] K. Waters, "Muscle Model for Animating Three Dimensional Facial Expression", Proceeding of SIGGRAPH, vol. 21, no. 4, July, 1987.

[7] J. A. Kim, "A Study on Effective Facial Expression of 3D Character through Variation of Emotions (Model using Facial Anatomy)", Journal of Korea Multimedia Society, vol. 9, no. 7, July, 2006.

[8] P. Ekman, W. V. Friesen, "Facial Action Coding System", Consulting Psychologists Press Inc., 577 College Avenue, Palo Alto, California 94306, 1978.

[9] W. E. Rinn, "The Neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions", Psychological Bulletin, 95(1), pp. 52–77, 1984.

[10] K. W. E. Lin, T. Nakano, M. Goto, (2019). "VocalistMirror: A Singer Support Interface for Avoiding Undesirable Facial Expressions", 16th Sound and Music Computing Conference (SMC2019), National Institute of Advanced Industrial Science and Technology (AIST), pp. 2518–3672, doi: 10.5281/zenodo.3249451.

Biography

Eunee Park received her bachelor's degree in TV, Film and Multimedia from Sungkyunkwan University in 2003, and her master's degree in Computer Art from School of Visual Arts in 2007. She is currently working as an assistant professor at the division of media arts, Baekseok Art University. Her research areas include computer graphics, animation and convergence content design.