

휴먼 안면인식 기술을 이용한 AR 영상 콘텐츠 개발 연구

A study on development of AR video contents using human facial recognition technology

박은애¹

Eunee Park¹

요약

미디어 시장이 급변하면서 온라인 플랫폼에서 소비 가능한 콘텐츠에 대한 시장의 요구가 커지고 있다. 온라인 플랫폼에 최적화되어 시청자의 흥미를 끌 수 있는 차별화된 영상 콘텐츠 제작이 요구된다. 콘텐츠 제작의 생산성과 효율성을 높이기 위해 AR엔진을 이용한 콘텐츠 제작 사례가 늘고 있다. 본 연구는 유명한 서양의 클래식 음악가인 모차르트, 베토벤, 쇼팽, 리스트의 얼굴을 3D 캐릭터로 제작하여, 실제 사람 얼굴 위에 증강할 수 있도록 안면인식 기술을 적용하였다. 이 기술은 실제 사람의 표정 변화를 분석 및 추적하여 3D 캐릭터의 표정에 실시간으로 적용되어, 3D 캐릭터의 얼굴이 증강된 사람은 유명한 음악가 캐릭터로 변신한다. 증강된 음악가의 얼굴이 씌워진 사람들은 음악 교육 시나리오를 바탕으로 다른 시대에 살았던 음악가가 되어 정보를 전달하고, 현 시대의 시청자들과 소통할 수 있게 된다. 본 연구를 통해 AR 게임엔진을 통한 새로운 영상 제작 시스템을 제시하고, 다변화는 미디어 시장에 맞춰 영상 제작의 새로운 방향성을 제시하고자 한다.

핵심어 : AR, Unity engine, 안면인식, 실시간 추적, 유튜브

Abstract

As the media market changes rapidly, the market demand is increasing for contents that can be consumed on online platforms. It's required to produce differentiated video contents that are optimized for online platforms and can attract viewers' interest. In order to increase the productivity and efficiency of content creation, cases of content production using AR engine are increasing. In this study, the faces of famous Western classic musicians, such as Mozart, Beethoven, Chopin, and List are created as 3D characters, and they augment on human's faces based on facial recognition technology. It Analyzes and traces the change of facial expression of the person, then apply to 3D character's facial expression in real time. Each of persons who augmented musicians' faces can become the musicians who lived in different times, deliver information and communicate with viewers of the present era based on the music educational scripts. This study proposes one of new video production system with AR game engines, and present the new direction of the video production in line with the media market for diversification.

Keyword : AR, Unity engine, Facial recognition, Real-time tracking, Youtube

1 Division of Media, Baekseok Arts University, Seoul, Korea [Professor]

e-mail: eunee@bau.ac.kr

* 본 논문은 백석예술대학교 교수 연구비 지원으로 작성되었음.

* 본 논문은 2021년 (사)차세대컨버전스정보서비스기술학회 춘계공동학술대회에서 “AR 안면인식 기술을 이용한 뉴미디어 영상 콘텐츠 개발 연구”로 발표된 논문을 확장한 것임.

Received(March 12, 2021), Review Result(1st: March 25, 2021), Accepted(April 9, 2021), Published(April 30, 2021)



© 2021 The Authors. Published by NCISS.
This is an open access article licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.
To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

1. 서론

전세계 미디어 시장이 OTT(Over The Top) 시장으로 재편되면서, 모바일 및 인터넷 미디어에서 시청자의 흥미를 끌 수 있는 차별화된 콘텐츠를 제작하기 위한 다양한 형태의 뉴미디어 콘텐츠 제작에 대한 관심이 고조되고 있다 [1]. 최근 유니티나 언리얼같은 게임엔진의 기술 발전은 게임 제작을 넘어 애니메이션, 영화, 교육, 건축, 가상공간 등 다양한 분야로 확장하고 있다 [2]. 가상 캐릭터는 그 동안 주로 게임에만 사용되었지만, 미디어 플랫폼의 다양화 속에서 수아나 슈블같은 버추얼 유튜버의 등장은 주목할 만한 현상이다 [3]. 본 연구는 휴먼 안면인식 기술에 초점을 맞춰 실시간 구현 및 연동이 가능한 증강된 버추얼 교수자를 개발하여, 클래식 음악 교육을 위한 영상 콘텐츠를 개발하였다.

본 연구에 앞서, 2020년 4월 EBS에서 방영된 TV시리즈 애니메이션, ‘허풍선이 음악쇼’에 등장하는 2D 캐릭터를 3D로 모델링하여 생성하였다. HYPRESENSE사의 페이스얼 모션 캡처 소프트웨어인 HYPRFACE를 라이선스하여 모델링된 3D 캐릭터에 적용한 후, 실제 배우와 배우의 얼굴에 증강된 3D 아바타가 음악과 연관된 내용을 주제로 출연자로 참여한다. 크로마키 세트에서 촬영을 진행하고 후반 작업을 통해 콘텐츠를 완성하는 방식으로 제작되었다.

본 연구를 통해 3D 아바타의 얼굴을 HYPRFACE에 적용 및 인식 오차 수정, 3D 아바타의 블렌드쉐입 연동, 크로마키 세트 촬영을 통한 합성 작업을 통해 게임 엔진을 접목한 새로운 형태의 영상 제작 방식을 제시하고자 한다.

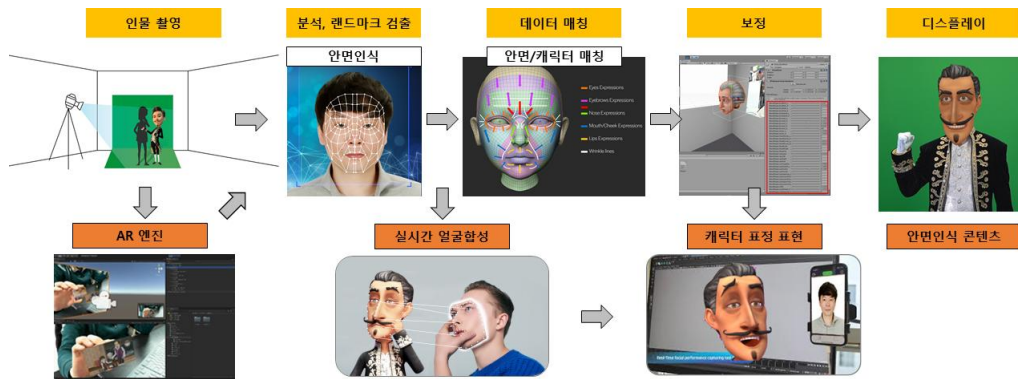
2. 본론

2.1 얼굴 표정 애니메이션의 제작 기법

3D 캐릭터의 얼굴 표정 애니메이션은 인터폴레이션 기법, 파라메트릭 모델 기법, 근육 기반 모델 기법, 모션 캡처 기법 등으로 다양하게 구분할 수 있다. 인터폴레이션 기법은 키프레임 방식의 몰핑 기법이고, 파라메트릭 모델 기법은 얼굴 움직임을 제어할 수 있는 파라미터 컨트롤을 제작하여 다양한 얼굴 변형 데이터를 얻을 수 있다. 근육 기반 모델 기법은 실제 안면 근육의 종류와 움직임 데이터를 3D 모델에 적용하는 방식이다. 모션 캡처 기법은 실제 얼굴의 흐름을 데이터화 하여 이를 기반으로 3D 캐릭터의 얼굴 표정을 제작한다 [4]. 본 연구를 위한 시스템인 HYPRFACE는 파라메트릭과 근육 기반 모델 기법이 혼용된 방식의 개발 환경을 지닌다.

2.2 제작 시스템 설계

시스템은 사용자가 직접 카메라에 얼굴을 대고 어플리케이션에 캐릭터가 인식이 되면 실시간으로 콘텐츠가 생산되도록 설계되었다. 제작 방식은 실제 인물을 촬영하며 AR 엔진이 실제 인물의 얼굴을 분석 및 추적하여 3D 캐릭터의 얼굴과 매칭을 하여 실시간으로 실제 인물의 얼굴에 증강된 캐릭터를 디스플레이 모니터나 모바일을 통해 즉각적으로 확인할 수 있다 [5]. 따라서, 실제 인물이 증강된 캐릭터를 모니터링하면서 촬영을 할 수 있기 때문에 다양한 표정 연출을 실시간으로 확인하고, 오차 범위를 줄일 수 있다. [그림 1]과 같이 실제 인물이 얼굴 표정과 몸동작 연기를 통해 보다 사실적인 연기 구현, 키 애니메이션 시간 단축, 후시 녹음 등의 제작 공정 과정을 줄일 수 있는 장점이 있다. 또한, 크로마키 세트 촬영이 아닌 스튜디오 세트 촬영으로 진행 시에도 별도의 배경 합성이 없어도 실시간으로 콘텐츠 제작 및 송출이 가능하다.



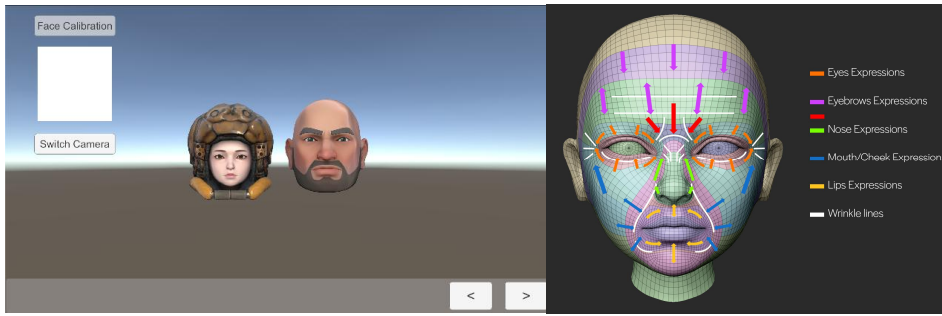
[그림 1] 제작 시스템 개념도

[Fig. 1] Fig. The system Diagram

2.3 HYPRFACE 소프트웨어 분석 및 설계

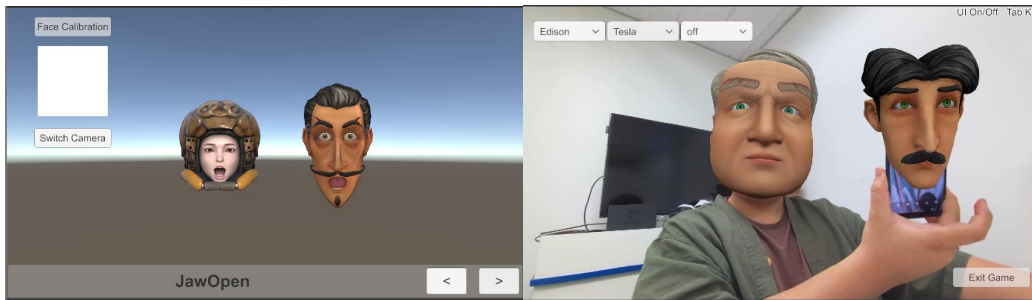
본 연구에서 사용한 HYPRFACE 소프트웨어는 PC, 모바일 환경 모두에 최적화되어 있으며, iOS, 안드로이드, Window OS, Mac OS를 모두 지원한다. [그림 1]은 사용자의 얼굴 특징점을 인식하여, 얼굴 각 부분의 위치 변화를 수치화하여 블렌드 웨입으로 적용한 3D 캐릭터에 적용되는 과정을 보여준다.

[그림 2]는 Hypresense사에서 제공하는 샘플 얼굴로, 코드 작동 방법과 기본적인 인식 UI를 제공하고, Hypreface 블렌드 웨입 가중치 설정을 위한 얼굴 표현 가이드를 제공하였다. 이를 통해 페이스 캡처 정확도 및 민감도를 높일 수 있었다.



[그림 2] 기존 HYPRFACE SDK가 제공하는 샘플 씬 데이터 & 블렌드쉐입 가중치 설정을 위한 각 얼굴 표현 부위 가이드
 [Fig. 2] Fig The sample scene provided by HYPRFACE SDK & Guidance of facial expression areas for setting the blend Shape weight

[그림 3]와 같이 HYPRFACE의 SDK가 제공하는 두 가지 샘플 얼굴 모델링을 통해 코드 작동 방법을 알 수 있으며, 기본적인 인식 UI를 제공한다. Window OS 기반 Kinect 4K 카메라를 사용하여 기본 예제 얼굴과 본 연구를 위해 제작된 3D 캐릭터의 얼굴 모델링에 기능을 연동하여 작동되는지 테스트를 진행하였다.



[그림 3] HYPRFACE API 샘플 모델링 & 다중 얼굴 인식 작동 테스트
 [Fig. 3] Fig. API sample modeling provided by HYPRFACE & Multi-face recognition test

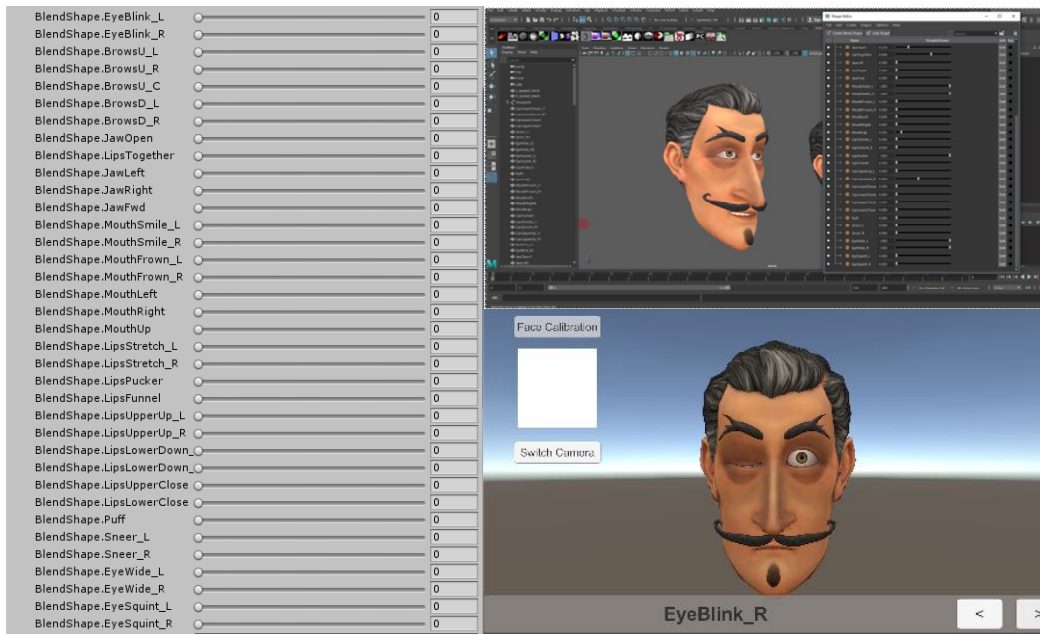
본 연구에서는 음악가 캐릭터인 모차르트, 베토벤, 쇼팽, 리스트의 얼굴을 모델링 및 맵핑하였다. [그림 4]는 증강에 사용된 모델링된 음악가 캐릭터들이다.

HYPRFACE에서 제공하는 joint 설정 후 36개 내외의 블렌드 쉐입 표정을 생성하고, HYPRFACE의 블렌드 쉐입을 네이밍 레벨에 대응시키는 레시피 파일을 생성하여 Unity 3D 상에서 구현했다. 제작한 3D 캐릭터의 얼굴 수치를 블렌드 쉐입과 쉽게 연동할 수 있다. 샘플로 제공된 기본 캐릭터와 달리 본 연구를 통해 제작된 3D 캐릭터의 얼굴 위치를 잘못 인식하거나, 각도에 따른 동기화의 어긋남 문제가 있어 각 얼굴 모델링의 뼈대값과 메쉬를 초기화하여 문제를 해결하였다. 3D 캐릭터의 얼굴을 사람의 몸에 맞춰 비율을 조절하면 카메라에서 멀어질수록 위치값에 오차가 발생하거나 얼굴 돌릴 때 증강된 3D 캐릭터의 얼굴이 사용자의 얼굴에서 벗어나는 현상이 발생하였다.



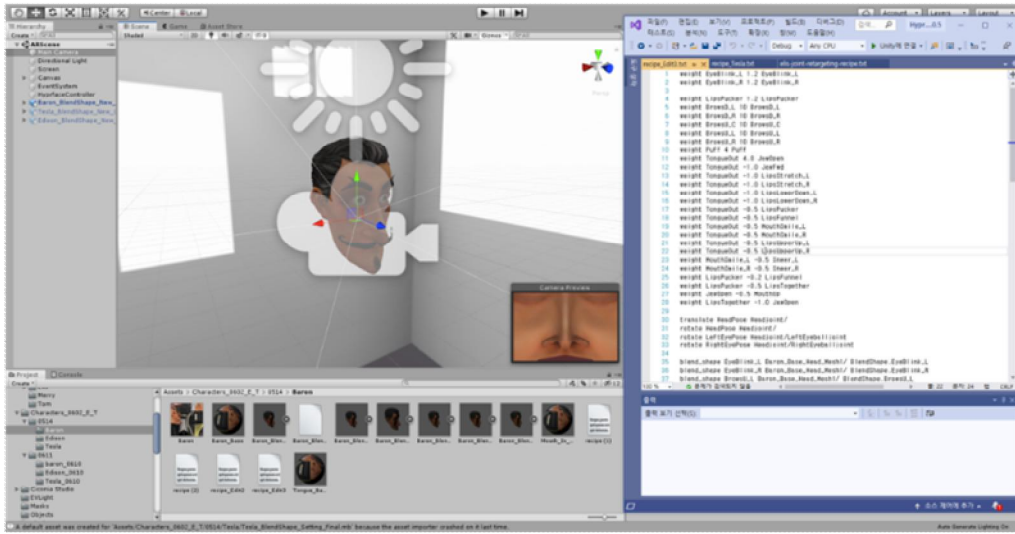
[그림 4] 음악가 캐릭터 모델링 & 맵핑 제작
 [Fig. 4] Fig. Modeling & Mapping of classic musician

HYPRFACE에서 제공하는 블렌드쉐입 목록은 [그림 5]와 같으며, 각 블렌드 쉐입 가중치에 따라 3D 캐릭터의 표정을 적용하여 정상적으로 표현이 되는지 테스트하였다.



[그림 5] HYPRFACE 블렌드 쉐입 목록 & 블렌드 쉐입 적용 테스트
 [Fig. 5] Fig. Blend Shape list of HYPRFACE & Blend Shape test

본 연구를 위해 캐릭터 별로 50개의 기본 블렌드 쉐입을 제작하였고, [그림 6]와 같이 HYPRFACE 소프트웨어에서 실시간으로 표정을 인식 및 조합할 수 있도록 개별 표정 별로 레시피를 개발하고 얼굴 인식 과정에서의 세부 민감도를 조절하였다.



[그림 6] 캐릭터 표정 Blend Shape의 실시간 조합 적용 이미지

[Fig. 6] Fig. Real-time combination of character expression applied the blend Shape

2.4 촬영 시스템

버추얼 교수자를 이용한 OTT용 클래식 음악 교육 콘텐츠를 제작하기 위해 크로마키 세트 촬영을 진행하였다. 촬영은 실제 연극자가 카메라 앞에서 연기를 하면 HYPRFACE 소프트웨어를 통해 지정된 캐릭터의 얼굴이 개별 연기자의 얼굴에 씌어져 옆에 설치된 모니터에 송출 된다. [그림 7]을 통해 촬영 당시 실제 연기자의 모습과 모니터를 통해 증강된 캐릭터의 얼굴 모습을 비교해 볼 수 있다.



[그림 7] 실제 연기자들 촬영 모습 & 증강 캐릭터로 송출된 모습

[Fig. 7] Fig. Real actors shooting footage & AR characters combining to the real actors

촬영자와 연기자 모두 실시간으로 모니터를 통해 증강된 얼굴을 확인할 수 있기 때문에 표정

트래킹 오류나 인식 문제 등을 실시간으로 모니터링하고 문제를 해결할 수 있다. 아래의 [그림 8]을 보면 촬영되는 영상이 실시간으로 옆 모니터로 송출되는 것을 볼 수 있다.



[그림 8] 실제 연기자들 촬영 모습 & 증강 캐릭터로 송출된 모습

[Fig. 8] Fig. Real actors shooting footage & AR characters combining to the real actors

2.5 후반 작업

촬영된 영상은 배경 합성 및 이펙트 추가 등의 후반 작업을 통해 정보 전달이 필요한 부분에 인포그래픽 및 시각 자료를 합성하여 추가하였다. 그래픽 합성 작업을 통해 현실 배경에서 촬영 시 발생할 수 있는 이질감을 줄이고, 효과적으로 정보를 전달할 수 있는 그래픽 요소들을 추가하여 콘텐츠의 완성도로 높일 수 있었다. 아래의 [그림 9]은 촬영된 영상 뒤에 그래픽 배경, 모션 그래픽, 이펙트 등을 적용한 결과물 이미지이다.



[그림 9] 합성 및 효과 처리가 완료된 제작 결과물

[Fig. 9] Fig. Output after compositing and visual effect

실제 10분 길이의 애니메이션을 제작하였다면 두 달 정도의 제작 기간이 소요되었을 것으로 예상되나, 휴먼 안면인식 기술이 적용된 AR 영상 콘텐츠 제작 과정을 통해 촬영, 편집 및 합성까지 8분 기준, 평균 4일 정도로 제작 기간이 단축되었다. 이러한 AR 안면인식 기술을 적용한 새로운 방식의 제작 과정은 촬영, 편집 및 합성까지의 제작 기간 단축은 물론, 새로운 미디어 콘텐츠 제작 도구로서의 효과성을 보여준다.

3. 결론

앞으로 도래할 포스트 코로나 시대에서 본 연구에서 사용된 휴먼 안면인식 기술을 이용한 AR 영상 제작 시스템은 OTT 플랫폼을 위한 새로운 콘텐츠 제작 도구로서의 효과성을 보여주었다. AR로 증강된 캐릭터를 통해 애니메이션 제작 측면에서는 영상 제작 기간을 단축했으며, 라이브 액션과 애니메이션의 장점을 취합하여 새로운 형태의 영상 콘텐츠 스타일을 제시하였다. 물론, 제작에 사용된 HYPRFACE 소프트웨어는 제공된 샘플 캐릭터에 최적화되어서 다른 모델링된 캐릭터에 적용했을 때 눈꺼풀 떨림이나 미세한 얼굴 표정 변화를 추적하는 데에는 어려움이 발생하였고, 연기자가 얼굴을 좌우로 돌릴 경우 증강된 캐릭터의 얼굴 위치가 벗어나는 문제로 촬영 시 연기자들의 움직임에 제약이 있어 제한된 공간 안에서만 촬영이 가능하였다는 한계점이 보였다. 향후 연구로는 APPLE사의 AR Kit를 이용하여 모델링된 캐릭터에 최적화된 소프트웨어를 개발하여 미세한 얼굴 표정 추적과 각도에 따라 싱크가 벗어나는 문제점을 해결하고자 한다 [6]. 최적화된 시스템을 기반으로 촬영과 동시에 영상을 온라인으로 즉시 업로드할 수 있는 실시간 송출 제작 시스템을 구축하는 추가 연구를 진행하고자 한다.

References

- [1] H. J. Kim, "The approaching direction of producing animation contents based on new media", *Journal of Digital Design*, vol. 10, no. 3, October 2010, pp. 185-195, doi: 10.17280/jdd.2010.10.3.019
- [2] H. Y. Jeon, "Domestic and foreign AR/VR industry status and implications", *Hyundai Economic Research Institute*, Seoul, Republic of Korea, no. 687, April 2017. [Online]. Available: [http://hri.co.kr/upload/publication/201743182547\[1\].pdf](http://hri.co.kr/upload/publication/201743182547[1].pdf).
- [3] H. G. Seo, "Domestic virtual characters that go beyond YouTubers and game promotion models", *gamemeca.com*, <https://www.gamemeca.com/view.php?gid=1634502>, (accessed February 1, 2021).
- [4] Y. G. Kim, "A Study on Marker Tracking Research for Utilization in AU based Facial Motion Capture Based on the Low Polygon Character", *The Korean Journal of Animation*, vol. 10, no. 4, December 2014, pp. 45-60.
- [5] Hypresense, "Hypresense", [hypresense.com](https://www.hypresense.com), <https://www.hypresense.com>, (accessed February 1, 2021).
- [6] K. W. E. Lin, T. Nakano, M. Goto, "VocalistMirror: A Singer Support Interface for Avoiding Undesirable Facial Expressions", *16th Sound and Music Computing Conference*, May 28-31, 2019, Málaga, Spain, pp. 2518-3672, doi: 10.5281/zenodo.3249451.