

# 스테레오스코픽 3D(Stereoscopic 3D) 페이스 스왑(Face Swap)을 활용한 3D VR 버추얼 휴먼 제작과 향후 적용 가능성

□ 전우열 / 벤틀브이알

요약

본 기고문은 기존의 2D 기반 페이스 스왑 기술을 확장하여, 스테레오스코픽 3D 페이스 스왑 기법을 활용해 3D VR 버추얼 휴먼을 제작하는 전 과정을 다룬다. 캐논 EOS R5C와 RF5.2mm F2.8 L 듀얼 피셔아이 렌즈를 사용해 좌·우 채널이 분리된 VR 영상을 촬영하고, 이를 AI 딥페이크 기술로 얼굴을 합성함으로써 기존보다 한층 더 입체적이고 실감도 높은 결과물을 도출한다. 특히, 버추얼 휴먼을 '디스플레이 안에서만' 만나던 한계를 넘어, 실제로 VR 환경에서 이들과 직접 상호작용할 수 있다는 점을 강조한다. 이를 통해 엔터테인먼트, 교육, 메타버스 등 다양한 분야에서 활용할 수 있는 새로운 가능성을 제시한다.

## I. 서론

### 1. 페이스 스왑 기술의 확장

페이스 스왑(Face Swap)은 영상을 구성하는 인물의 얼굴을 다른 인물의 얼굴로 교체하는 기술을 가리킨다. 이 기술은 연예·오락 프로그램이나 영화 특수효과 분야에서 널리 활용되면서 대중에게도 많이 알려졌다. 최근에는 생성형 AI 기법이 발달함에 따라, 실제와 구분하기 어려울

만큼 정교한 합성이 가능해졌으며, 각종 SNS나 개인 방송에서도 활발히 사용되고 있다. 그러나 기존의 페이스 스왑은 2D 기반 영상에서만 구현되는 경우가 대다수였다.

### 2. VR 환경의 요구사항

VR(Virtual Reality)은 사용자가 가상 공간에 몰입해 상호작용할 수 있는 환경을 제공한다. 그러나 완전한 몰입을 위해서는 시야각뿐 아니라 깊이감도 충실히 구현되어야



<그림 1>

한다. 단일 시점의 2D 영상에 페이스 스왑을 적용하면, 얼굴 합성 자체는 정교하게 이뤄질 수 있으나, VR 환경에서 보았을 때는 입체감이 부족해 결과물이 어색하게 느껴진다. 이러한 한계를 극복하기 위해서는 좌·우 시점을 모두 고려한 스테레오스코픽(입체) 방식이 필수적이다.

### 3. 연구 목적

본 기고문은 스테레오스코픽 3D 페이스 스왑 기법을 통해 3D VR 버추얼 휴먼을 제작하는 전체적인 프로세스를 소개하고, 이를 통해 얻을 수 있는 기술적·사업적 가치를 논의한다. 특히, 디스플레이 장치 속에서만 볼 수 있던 버추얼 휴먼이 VR 헤드셋 등을 통해 실제 공간에 존재하는 듯한 체험을 선사할 수 있음을 강조한다.

## II. 구현 방법

스테레오스코픽 3D 페이스 스왑 기반의 3D VR 버추얼 휴먼 제작 과정은 크게 (1) VR 무대 영상 촬영, (2) VR 후반 제작 과정, (3) AI 후반 제작 과정, (4) 합성 작업, (5) 마

스터링으로 구분된다. 각 단계에서는 다양한 하드웨어와 소프트웨어 기술이 결합되어 자연스럽게 몰입감 있는 VR 콘텐츠를 완성하게 된다.

### 1. VR 무대 영상 촬영

#### 1) 안무가와 협업

- 무대 위에서 발생하는 움직임과 표정을 극적으로 표현하기 위해서는 안무가와 사전 협업이 필수적이다. 특히 VR 환경에서는 360도 전방위로 콘텐츠가 노출되므로, 무대 장치나 출연자 동선을 면밀히 점검하여 카메라 세팅을 최적화한다.
- 안무의 타이밍에 따라 인물의 얼굴이 어느 방향을 향하게 되는지, 얼굴 표정이 또렷이 보이는지 등을 사전에 시뮬레이션한다.

#### 2) 조명 설정

- 캐논 EOS R5C와 RF5.2mm F2.8 L 듀얼 피셔아이 렌즈는 넓은 화각을 제공하지만, 그만큼 조명 세팅이 까다롭다. 인물의 얼굴이 명확히 드러나도록 스포트라이트를 배치하되, 배경이나 주변 인물에게도 균일한

광량이 유지되도록 주의한다.

- 조명은 얼굴에 대한 직접광과 간접광을 균형 있게 구성해야 하며, VR 환경이므로 카메라 주변 360도에서 광원이 노출될 가능성을 고려해야 한다.

### 3) 레이저 조명 반사 주의

- 무대 연출에 레이저 조명을 활용할 경우, 듀얼 피셔아이 렌즈의 광각 범위 안에 불필요한 난반사가 발생하지 않도록 주의한다. 빛 반사가 심한 소재나 금속성 무대 장식품 등을 최소화하거나, 빛의 각도를 조절해 영상 품질 저하를 막는다.

## 2. VR 후반 제작 과정

### 1) 스티칭

- 듀얼 피셔아이 렌즈로 촬영된 좌·우 영상은 기본적으로 원형 왜곡을 가진 어안(Fisheye) 형태다. 이를 VR 환경에서 정상적인 스테레오 영상으로 재현하기 위해서는 전문 스티칭 소프트웨어를 사용해 두 영상을 적절히 연결하고, 왜곡을 보정해야 한다.
- 스티칭 과정에서 가장 중요한 점은 좌안(Left)과 우안(Right) 사이의 정합(Alignment)을 확보하는 것이다. 정확한 스티칭이 이뤄지지 않을 경우, VR 헤드셋으로 시청 시 심각한 어지럼증을 유발할 수 있다.

### 2) 디노이즈 및 색보정

- 라이브 무대 환경에서 촬영된 영상에는 조명 빛 번짐, 고감도 노이즈, 색온도 불일치 등이 발생하기 쉽다. 디노이즈 알고리즘을 적용해 영상 품질을 개선하고, 좌·우 영상 간의 색 차이를 최대한 줄여 시각적 이질감을 감소시킨다.
- 색보정 단계에서는 제작 의도와 무대 연출 분위기에 맞춰 톤을 통일한다. 예를 들어, 따뜻한 분위기를 위해 노란 톤을 강조하거나, 화려한 무대 감성을 위해 채도를 높이는 등의 작업을 진행한다.

### 3) 해상도 변환

- 원본 영상이 8K 이상 고해상도일수록 합성에 필요한 연산 자원이 급증한다. 이를 감안하여 일단 4K로 다운 컨버팅한 뒤, 후속 AI 합성 작업을 진행한다. 이후 최종 마스터링 단계에서 8K로 다시 업스케일링을 실시해 VR 헤드셋에서 높은 해상도를 확보한다.

## 3. AI 후반 제작 과정

### 1) 인물 레퍼런스 제작

- 페이스 스왑의 정확도는 학습 데이터의 품질에 크게 좌우된다. 합성 대상 인물의 얼굴을 다양한 각도와 표정, 조명 조건에서 촬영한 레퍼런스 이미지를 충분히 확보한다.
- 표정 변화가 큰 장면이나 극단적인 앵글에서도 자연스러운 합성이 이뤄지도록, 웃는 얼굴, 찡그린 얼굴, 측면 얼굴 등 최대한 다양하게 데이터셋을 구축한다.

### 2) AI 딥페이크 적용

- 다운 컨버팅된 4K 영상을 일정 간격의 시퀀스로 나눈 뒤, AI 딥페이크 모델(예: GAN 기반)을 적용해 얼굴을 교체한다. 이때, 좌안과 우안 영상에 동시에 동일한 얼굴 합성이 이뤄져야 하므로, 채널 간 동기화(Synchronization) 처리가 중요하다.
- 합성 결과를 검사하면서, 얼굴 경계선이나 특정 표정 변화에서 생길 수 있는 ‘이음새’ 문제를 확인한다. 필요하다면 후술하는 합성 작업 단계에서 추가 보정(compositing)을 진행한다.

## 4. 합성 작업

### 1) 트래킹

- 3D 랜드마크 트래킹 알고리즘을 통해, 합성된 얼굴이 원본 영상의 움직임과 정확히 일치하도록 조정한다. 좌·우 영상 모두에서 얼굴 위치, 각도, 표정 변화

를 동시에 추적해야 하므로, 계산량이 크고 정밀도가 요구된다.

- 얼굴 표정뿐만 아니라, 목이나 어깨 라인 등 주변 영역도 자연스럽게 연결되도록 마스킹 작업을 수행한다.



<그림 2>

## 2) 업스케일링

- AI 합성 완료 후, 4K 결과물을 업스케일링 기법(딥러닝 슈퍼 레졸루션 등)을 통해 8K로 확대한다. 스테레오스코픽 VR 환경에서는 해상도가 높을수록 몰입감이 향상되므로, 업스케일링 시 세부 디테일을 최대한 살리는 것이 핵심 과제다.
- 업스케일링 과정에서 한쪽 채널만 과도하게 선명해지지 않도록, 좌·우 채널을 동일한 알고리즘과 파라미터로 처리한다.

## 3) 최종 합성

- 얼굴 합성 및 주변 배경, 조명 효과 등 모든 시각 요소를 종합하여 최종 영상을 완성한다. 이때, 인물과 배경 간 조명·색보정 일관성을 재점검하고, VR 환경에서 입체감이 어색하게 들뜨지 않도록 주의한다.

## 5. 마스터링

### 1) 편집 및 색보정

- 편집 단계에서는 전체 시퀀스를 다시 한번 확인하며, 장면 전환이나 음악과의 싱크 등을 조율한다. 특히 VR 영상에서는 시선 유도 요소가 중요하므로, 관찰자가 어느 방향을 바라보게 될지를 고려해 컷 전환이나 효과를 배치한다.
- 색보정은 콘텐츠의 컨셉에 따라 전체 톤을 결정하는 단계로, 제작 의도에 부합하는 분위기를 연출할 수 있다. 필요하다면 HDR(High Dynamic Range) 영상을 적용해 한층 풍부한 명암비를 구현한다.

### 2) 최종 출력

- 마스터링이 완료된 후 8K 해상도의 VR 영상을 렌더링한다. 최종적으로 VR 헤드셋을 착용해 영상을 시청하면서, 입체감이 어긋나거나 특정 구간에서 일시적인 어지럼증이 발생하지 않는지 철저히 검수한다.
- 완성된 결과물은 이후 스트리밍 플랫폼이나 오프라인 이벤트에서 3D VR 체험 형태로 제공될 수 있으며, 필요에 따라 2D 버전으로 재가공될 수도 있다.

## III. 향후 적용 가능성

스테레오스코픽 3D 페이스 스왑은 기존의 버추얼 휴먼이 단순히 2D 디스플레이 속에서만 존재하는 것과 달리, 가상 세계에서 실제와 비슷한 상호작용이 가능하다는 점에서 의미가 크다. 이러한 특징은 엔터테인먼트, 교육, 메타버스 등 다양한 산업 분야로 확장될 가능성을 갖는다.

### 1. 엔터테인먼트 및 게임

- 버추얼 콘서트나 뮤직비디오에서 아티스트의 얼굴을

버추얼 휴먼으로 교체하거나, 게임 캐릭터에 실제 배우의 표정과 움직임을 투영할 수 있다. 이렇게 제작된 3D VR 콘텐츠는 기존 2D 화면과 비교할 수 없을 정도로 높은 몰입감을 제공한다.

상품·서비스에 대한 주목도를 높일 수 있다. VR 환경에서 인터랙티브하게 브랜드를 체험할 수 있어, 고객 만족도를 높이는 데에도 효과적이다.

## 2. 교육 및 훈련

- 역사적 인물이나 저명 강사의 얼굴을 입체적으로 재현해, VR 환경에서 학습자와 직접 대화하는 듯한 교육 콘텐츠를 제작할 수 있다. 의료나 군사 훈련에도 활용될 수 있으며, 실제 상황을 안전하게 시뮬레이션하면서도 사실감을 극대화한다.

## 3. 소셜 플랫폼 및 메타버스

- 메타버스 플랫폼에서 아바타를 단순 캐릭터가 아니라, 실제 인물의 얼굴을 가진 버추얼 휴먼으로 표현할 수 있다. 이를 통해 사용자들 간의 소통이 한층 자연스러워지고, 아바타에 감정 표현이나 미세한 표정을 부여해 ‘존재감’을 높일 수 있다.

## 4. 브랜드 마케팅 및 광고

- 유명 모델이나 배우의 이미지를 실제로 ‘착용’한 버추얼 휴먼을 통해 광고나 홍보 콘텐츠를 제작함으로써,

## IV. 결론

본 기고문에서는 스테레오스코픽 3D 페이스 스왑 기술을 활용해 3D VR 버추얼 휴먼을 제작하는 과정을 상세히 살펴보았다. 캐논 EOS R5C와 RF5.2mm F2.8 L 듀얼 피서아이 렌즈를 통한 입체 촬영부터, AI 딥페이크 합성, 최종 합성, 마스터링에 이르는 전체 단계를 통해 디스플레이를 넘어선 몰입형 VR 콘텐츠가 완성된다. 이는 기존 2D 페이스 스왑의 한계를 극복하고, 가상 세계에서 사람과 버추얼 휴먼이 더욱 자연스럽게 교류할 수 있는 길을 열어준다.

앞으로 VR 헤드셋과 같은 하드웨어가 발전하고, AI 알고리즘이 더욱 정교해지면, 버추얼 휴먼의 활용 범위는 더욱 넓어질 것으로 보인다. 게임·엔터테인먼트부터 교육·훈련, 소셜 플랫폼까지 무궁무진한 가능성이 열려 있으며, 사용자 개인이 직접 VR 공간에서 버추얼 휴먼을 만들어 내고 조작하는 시대가 도래할 것으로 예측된다. 이러한 기술은 인간의 창의성을 극대화하면서도 새로운 형태의 미디어 소비 및 협업 구조를 만들어낼 것으로 기대된다.

## 참 고 문 헌

- [1] R.Y. Tsai, "A Versatile Camera Calibration Technique for High Accuracy 3-D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses," IEEE Journal of Robotics & Automation 3 (1987), pp. 323-344.
- [2] 캐논 EOS R5C 공식 사이트: <https://www.usa.canon.com/cameras/eos-r5-c>
- [3] J. K. An, "Deep Learning-based Face Swapping for 3D Virtual Reality," Journal of Advanced Media Technology, vol. 12, no. 2, pp. 45-56, 2023.

저 자 소 개



**전우열**

- 2015년 ~ 현재 : 벤틀브이알 대표
- 2020년 ~ 현재 : 3D VR 다중교육솔루션 VENTA CMS 운영 총괄
- 2022년 ~ 현재 : 8K 3D VR 기술을 활용한 K-POP XR콘텐츠 서비스 (VENTA X) 운영 총괄