

실사 공간 영상 캡처링을 위한 6DoF VR 몰입형 비디오 취득 시스템 구현

양이삭¹, 류영일², 김장현², 류은석¹

성균관대학교 실감미디어공학과, 컴퓨터교육과

icarax1491, yeongilryu, amrs73, esryu@skku.edu

6DoF VR Immersive Video Acquisition System for Real-life Spatial Video Capturing

Isaac Yang¹, Yeongil Ryu², Jang Hyun Kim², Eun-Seok Ryu¹

Department of Immersive Media Engineering

Department of Computer Science Education

Sungkyunkwan University

요약

본 논문은 메타버스 기반의 실감형 교육 콘텐츠 제작을 위한 6DoF 몰입형 비디오 취득 시스템을 제안한다. 제안된 시스템은 시그널링을 통해 동작하는 여러 PC들과 이에 연결된 여러 카메라들을 활용하여 취득 대상이 되는 공간의 영상을 취득해 6DoF 몰입형 비디오 제작을 위한 RGB raw 비디오로 저장한다. PC에 연결된 카메라들은 소프트웨어 트리거를 통해 제어되고 PC는 UDP 멀티캐스트 방식으로 동작하며, 제안된 시스템에 따라 취득된 영상들은 최대 48fps까지 균일한 프레임 레이트를 지원할 수 있다. 제안된 시스템을 통해 성균관대학교 VR 첨단 강의실의 다시점 영상을 취득할 수 있었으며, 0.1ms 이하의 표준편차를 보이는 동기화 수준을 확인하였다.

1. 서론

코로나19 이후로 사회적 거리 두기 정책이 시행되면서 비대면 서비스가 산업계는 물론 교육계에도 자리 잡게 되었다. 이후 2022년 앤데미 이후로도 학습효과를 높인 교수법, 의사소통의 효율성, 안목을 넓히는 등시 다현장 교육 방식 등과 같은 장점들로 인해 비대면 교육은 지속적으로 발전되고 있으며, 새로운 기술의 등장과 기존 기술의 발전으로 인해 더욱 주목받게 되었다[1]. 2023년 현재 비대면 교육과 관련된 에듀테크 분야의 가장 큰 두 가지 화두는 AI 기반의 비대면 교육 프로그램과 메타버스 기반의 실감형 교육 콘텐츠라고 할 수 있다. 기존 교수자의 역할을 AI가 대체할 수 있는지에 대한 연구와 함께, 기존 물리적 학습 공간을 연계한 메타버스 상의 가상 교육 공간은 가상공간이기에 가능한 시도들과 함께 실감형 혹은 몰입형 교육 콘텐츠라는 이름으로 연구되고 있다. 이는 Learning Management System(LMS), ZOOM, 유튜브 라이브 등의 플랫폼들이 기술적 장애와 불편한 의사소통 등과 같은 한계를 보이자 이 문제들을 보완하기 위해 현실 공간을 연계한 3차원 공간 정보를 가상공간에 구축하여 활용하고자 하는 필요성이 제시되었고, 특히 3D 몰입형 콘텐츠를 활용한 교육이 강의뿐만 아니라 실습 및 시뮬레이션 등과 같은 자유로운 활동이 가능하다는 점에서 그 필요성이 더욱 강조되고 있다. 이러한 가상공간은 학습자로 하여금 익숙하다고 느낄 수 있는 온-오프라인 학습 환경과 높은 기술적 완성도에 대한 요구 또한 제

시되고 있다[2].

이러한 비대면 교육을 목적으로 한 실감형 콘텐츠를 제작하기 위해서는 3D 그래픽 기반의 가상공간을 제작하거나 현실 스튜디오 혹은 공간에서 취득한 실사 영상을 몰입형 비디오로 제작하여야 하는데, 이 몰입형 비디오는 서비스하는 자유도에 따라 3DoF(Degrees of Freedom)와 6DoF로 구분할 수 있다. 2016년 10월 시작된 MPEG-I(ISO/IEC 23090 Coded Representation of Immersive Media) 프로젝트에서 몰입형 미디어의 몰입도를 사용자 시점의 자유도에 따라서 3DoF와 6DoF로 구분하였다. 3DoF 몰입형 미디어의 경우, 고정된 사용자 위치에서의 X, Y, Z축을 중심으로 한 머리 회전만을 지원하지만, 6DoF의 경우 머리 회전뿐만 아니라 사용자가 임의로 시점 위치를 옮길 수 있으므로 더욱 몰입감 있는 사용자 경험을 제공할 수 있다. MPEG-I는 이러한 6DoF 몰입형 미디어에 대하여 2021년 말부터 사용 시나리오, 요구사항, 부호화 및 전송 기술에 대해 표준화를 진행하고 있다[3].

본 논문에서는 기존 비대면 강의 및 실습 교육 콘텐츠 플랫폼에서의 기술적 문제 해결에 대한 필요성에 의해 메타버스 기반 몰입형 교육 콘텐츠 제작을 위한 6DoF 몰입형 비디오 취득 시스템을 제안한다. 제안된 시스템은 4대의 PC에 각각 5대의 카메라가 연결되어 총 20대의 카메라로 영상을 취득하는 멀티뷰 카메라 시스템으로, 해당 시스템은 UDP 멀티캐스팅 및 소프트웨어 동기화 기술을 통해 각 카메라에서 취득된 프레임 간 동기화 오차를 최소화한다. 본 논문에서는 제안된 시스템의 구성

과 동작 방식을 설명하고, 멀티뷰 비디오 취득 결과 및 카메라 간의 프레임 취득 동기화 오차 측정을 통하여 제안된 시스템의 성능을 검증한다.

2. 관련 연구

2-1. 기존 메타버스 플랫폼을 활용한 에듀테크 시스템

최근 메타버스 플랫폼을 활용한 교육 콘텐츠 서비스는 콘텐츠 개발자가 시스템 서버에 업로드한 메타버스 플랫폼 기반의 교육 콘텐츠를 강의자와 학습자가 VR HMD(Head Mounted Display)를 착용하고 교육을 진행하는 형태로 제공되고 있다. 동국대학교 연구진이 제안한 교육의 집중도 향상을 위한 몰입형 메타버스 시스템 연구[4]에 따르면, 업로드되는 교육 콘텐츠에는 교육자 아바타, 학습자 아바타, 교재 등이 포함되며, 학습자 아바타에는 학습자가 바라보는 시선과 같은 방향과 움직이면서 교육자 아바타나 교재를 관통하였을 때의 집중도를 측정해 로그 데이터로 남겨 교육자가 확인할 수 있도록 구현되어 있다. 해당 시스템은 학습자의 헤드 트래킹을 위한 HMD 관리 및 추적 모듈이라 명명하였으며, 이렇게 구현된 몰입형 메타버스 교육 시스템은 VR 기반의 가상환경에서 교육자와 학습자가 교육에 몰입할 수 있도록 도움을 줄 것이라고 기대된다. 하지만 현장감과 몰입감의 중요성이 강조되는 실습 중심의 교육 환경에서는 그래픽으로 구현된 가상환경에서의 학습 과정이 실습 교육 현장 환경의 역할을 온전히 재현할 수 없다는 한계점이 존재한다.

에듀테크 기반의 비대면 K-Dance 교육 프로그램에 대한 연구[5]에서는 AI 평가 데이터와 함께 강의용 영상 콘텐츠를 활용하는 실습 교육 플랫폼을 제안하였다. 제안된 실습 교육 플랫폼은 누구나 쉽게 접근할 수 있는 체력운동을 통한 사전 체력 평가 이후, 제시되는 춤 영상 콘텐츠를 시청한 뒤 이를 따라 하는 영상을 촬영하고 업로드하면 AI 평가 점수를 확인하는 춤 사전 평가를 진행한다. 그 후 이론 수업, 춤 실습 강의 그리고 체력 연습 강의 또한 영상 콘텐츠로 제공되는데, 이 영상들은 상, 하, 좌, 우 다양한 각도를 선택하여 시청 및 수강이 가능하다. 이러한 실사 영상 기반의 실습 교육 시스템에서 6DoF 몰입형 콘텐츠를 활용한다면 기존 제한된 각도에서의 사용자 경험에 비하여 수강자가 높은 자유도를 통해 강의자의 동작을 관찰할 수 있는 장점이 있으며, 이는 향상된 몰입 및 학습 효과를 사용자에게 제공하는 것이 가능할 것이다.

실습 교육을 위한 몰입형 교육 콘텐츠 서비스는 간호 현장 실습을 위해 360 몰입형 비디오를 활용하는 형태로 제안되고 있다[6]. Texas A&M University Soft Interaction lab 연구진이 제안한 몰입형 교육 시스템은 대학의 간호학과 학부생들을 위한 평가 수업에 활용하기 위하여 360 몰입형 비디오를 활용하였다. 가상 현실 애플리케이션을 통해 녹화된 360 비디오를 재생하고, 애플리케이션 사용자는 간호사와 의사가 환자 옆에서 있는 것처럼 몰입감 있는 평가를 진행할 수 있다. 또한, 애니메이션, 일러스트레이션, 다이어그램, 나레이션 등을 추가하여 영상으로는 전달하지 못하는 영역의 정보를 추가로 전달하였다. 이 요소들은 카메라 주위의 환경을 360도 동시에 캡처가 가능한 3D 카메라인 Insta360 Evo를 사용하여 현장에서 취득한 영상을 통해 제작되었으며, 현장 촬영 영상을 편집하고 그래픽 및 사운드 효과를 추가하여 만들어진다. 사용자 참여 검사 결과 높은 수준의 참여를 유도하였고, 이는 360 몰입형 비디오 매체를 통한 실습 교육에서의 잠재력을 보였다고 제안하였다.

2-2. 다시점 영상 획득 워크플로우

본 논문에서 제시된 360 VR 비디오 획득을 위한 시스템은 여러 대의 카메라를 통해 획득되는 다시점 영상들을 취득함으로써 6DoF 몰입형 비디오를 얻는 것을 목표로 한다. 이러한 다시점 영상을 획득하는 워크플로우가 선행되어 연구되었으며[7], 해당 논문에서는 (1)카메라를 통한 장면 획득, (2)카메라 캘리브레이션 및 깊이 맵 추정, (3)카메라 파라미터 추정, (4)깊이 맵 생성의 순서로 워크플로우가 구성되어 있다.

장면을 획득을 위한 카메라로 IP 기반의 동기화를 지원하는 Sony RK0mk.2를 사용하였으며, 이 카메라들의 파라미터를 추출하는 작업으로 카메라 캘리브레이션 절차를 수행한다. 이는 영상 품질에 중요한 영향을 미치는 파라미터들을 추출하기 때문에 가장 중요한 절차로 여겨지며, 이때 추정된 파라미터들을 이용해 목표로 하는 획득 영상을 보정하고 별도로 색 보정 파라미터를 추정하여 이들을 이용해 가상시점 합성을 위한 깊이 맵을 생성하는데, 이때 오픈소스 알고리즘인 MPEG DERS(MPEG-I Depth Estimation Reference Software)를 사용한다.

다시점 영상 획득 워크플로우에서는 전술한 절차들을 통해 왜곡 및 색 보정이 된 다시점 영상들을 획득하고 추정된 카메라 파라미터와 이 영상들을 이용해 깊이 맵을 생성하게 된다. 해당 연구는 입체 영상 생성을 위한 워크플로우 구축 및 실험에 큰 의미가 있으며, 깊이 맵 생성을 목표로 한 향후 연구에 도움이 될 것으로 보인다. 하지만 해당 연구에서 진행된 실험은 수렴형 리그를 통해 배치된 3대의 카메라로 실내/외 장면을 취득하였기 때문에, 공간에 대한 전방위 시점을 취득하고자 하는 본 논문에서 제안된 시스템과 유즈 케이스가 맞지 않는다.

3. 6DoF VR 몰입형 비디오 취득 시스템

3.1. 시스템 구축 환경

제안된 시스템 구현을 위하여 본 논문에서 사용한 카메라는 Basler 사의 acA2500-60uc로 그림 1 (a)와 같다. 4.8 μ m x 4.8 μ m 크기의 픽셀 사이즈, 최대 2592px x 2048px 해상도 및 최대 60fps 프레임 레이트 등을 지원한다. 동기적 멀티 카메라 영상 취득을 위해 소프트웨어 트리거 기능을 지원하는 USB 3.0 인터페이스 카메라 모델을 사용하였다. 렌즈는 Edmund Optics 사의 8mm Focal Length, HP Series Fixed focal Length Lens f/1.8 86-569로 그림 1 (b)와 같다. 최대 240.8mm - 83.9°의 수평, 174.7mm - 68°의 수직, 그리고 305.3mm - 94.8°의 사선 화각을 지원한다. 멀티뷰 카메라 취득 SW의 구현은 Basler 사의 Pylon Camera Software C++ API를 기반으로 구현되었으며, 해당 API를 통해 C++ 코드에서 카메라의 프레임 레이트, 픽셀 포맷, 노출 시간, 해상도 및 소프트웨어 트리거 등을 설정할 수 있다. 제안된 시스템 또한 API에서 제공하는 헤더 파일 내 함수를 재정의하여 카메라 취득 설정을 변경하는 기능을 구현하였다.



그림 1. (a) Basler acA2500-60uc (b) 8mm Focal Length, HP Series Fixed focal Length Lens f/1.8 86-569

3.2. 시스템 구조

제안된 시스템의 구조는 크게 카메라의 영상 취득 전반을 처리하는 1) 비디오 취득 모듈과 멀티뷰 카메라 시스템을 구성하는 각 카메라들 간의 영상 프레임 취득 동기화를 처리하는 2) 카메라 동기화 모듈로 구성되어 있다.

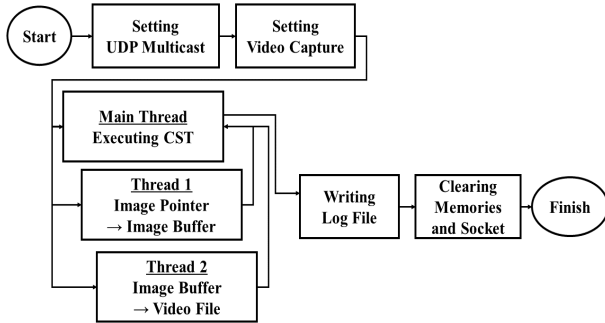


그림 2. 시스템 전체 순서도(CST: Camera Software Trigger)

비디오 취득 모듈의 전체 동작 방식은 그림 2와 같다. 시스템을 시작하면 UDP 소켓과 멀티캐스트 TTL(Time To Live)을 설정하고 소켓 구조체 초기화 및 네트워크 인터페이스 설정 후 멀티캐스트 그룹의 생성 및 가입이 이루어진다. UDP 멀티캐스트 통신 설정이 끝나면 영상 취득과 관련된 설정을 수행한다. 이후 카메라 수만큼 반복하여 각 카메라 객체들에 소프트웨어 트리거를 등록하고, 픽셀 포맷, 프레임 레이트, 노출 시간을 입력된 값으로 설정한다. 모든 설정이 완료되면 API 내 함수를 통해 카메라를 열고 소프트웨어 트리거가 실행되는 즉시 이미지를 취득할 수 있도록 카메라를 대기시킨다.

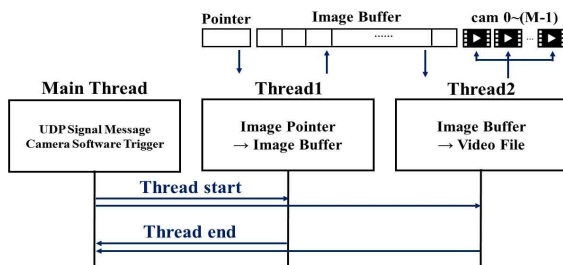


그림 3. 시스템 구성 - 카메라 동기화 모듈

제안된 시스템의 카메라 동기화 모듈은 그림 3과 같다. 모든 설정이 완료된 후 통신이 성공적으로 연결되면 영상 취득을 위한 스레드를 생성하고, 취득이 완료되면 모든 스레드가 종료된다. 메인 스레드는 PC에 연결된 카메라들의 소프트웨어 트리거를 실행한다. Sender의 메인 스레드는 매 프레임마다 소프트웨어 트리거 실행 메시지를 전송하고, Receiver의 메인 스레드는 매 프레임마다 Sender가 전송한 소프트웨어 트리거 실행 메시지를 받은 후 소프트웨어 트리거를 실행한다. 버퍼 저장 스레드(Thread 1)은 메인 스레드에서 소프트웨어 트리거가 실행된 각 카메라들이 취득한 이미지 데이터를 이미지 버퍼에 저장하고, 로그 파일 작성을 위해 카메라의 타임 스탬프를 저장한다. 파일 저장 스레드(Thread 2)는 버퍼 저장 스레드에서 저장된 이미지 버퍼의 데이터를 각 카메라의 RGB raw 비디오 파일에 저장한다.

제안된 시스템의 PC 동기화 모듈의 구조는 그림 4와 같다. Sender와 Receiver PC들은 지정된 동일한 멀티캐스트 그룹에 가입하고, UDP 멀티캐스트 통신을 위한 설정 완료 후 Sender가 Receiver들에게 영상

취득 시작을 위한 버퍼 메시지를 전송한다. 이후 Sender가 메인 스레드에서 매 프레임마다 Receiver들에게 카메라 소프트웨어 트리거 실행 메시지를 전송하면 Receiver들은 버퍼 메시지를 받아 Sender와 Receiver들이 각각 연결된 카메라들의 소프트웨어 트리거를 실행하고, 그림 3에 표시된 Thread 1(버퍼 저장 스레드)과 Thread 2(파일 저장 스레드)에서 이미지 취득 및 파일 저장 과정을 실행한다. 제안된 시스템은 Thread 2 종료 시 RGB raw 비디오 파일 쓰기 작업이 종료되고, 취득 시 측정된 매 프레임별 카메라들의 취득 시점 타임 스탬프, 카메라별 프레임 간 타임 스탬프 시간차, 프레임별 카메라들 간의 실행 시간 오차를 계산하여 로그 파일에 저장한다.

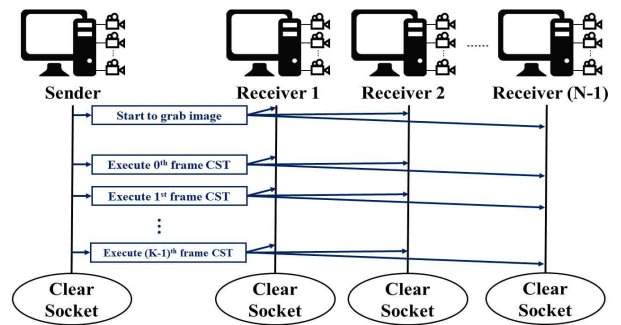


그림 4. 시스템 구성 - PC 동기화 모듈

4. 취득 결과

4.1. 취득 영상

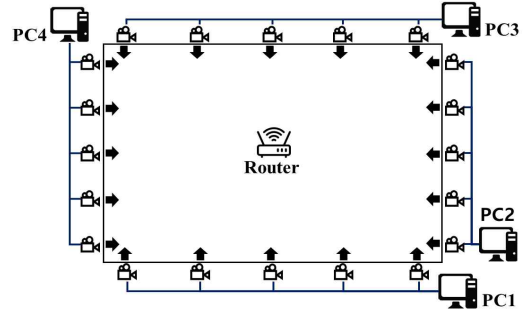


그림 5. 카메라 및 PC 배치 구조

제안된 시스템을 활용하여 성균관대 VR 첨단 강의실의 다시점 영상을 취득하는 테스트를 진행하였으며, 이때의 카메라 및 PC 배치 구조는 그림 5과 같다. 취득 테스트는 표 1의 설정으로 총 3번 진행하였다.

항목	설정값
PC1	Sender
PC2	Receiver
PC3	Receiver
PC4	Receiver
취득 프레임 수	100 프레임
노출 시간	4000
프레임 레이트	30fps

표 1. 취득 테스트 설정

취득된 영상은 그림 6과 같다. 실행 결과 총 20대의 카메라로 취득된 20개의 RGB 8bit 픽셀 포맷의 RGB raw 비디오 파일과 카메라별 타임스탬프가 기록된 로그 데이터를 얻을 수 있었다.



그림 6. 3rd Test 취득 영상

4.2. 로그 데이터 분석

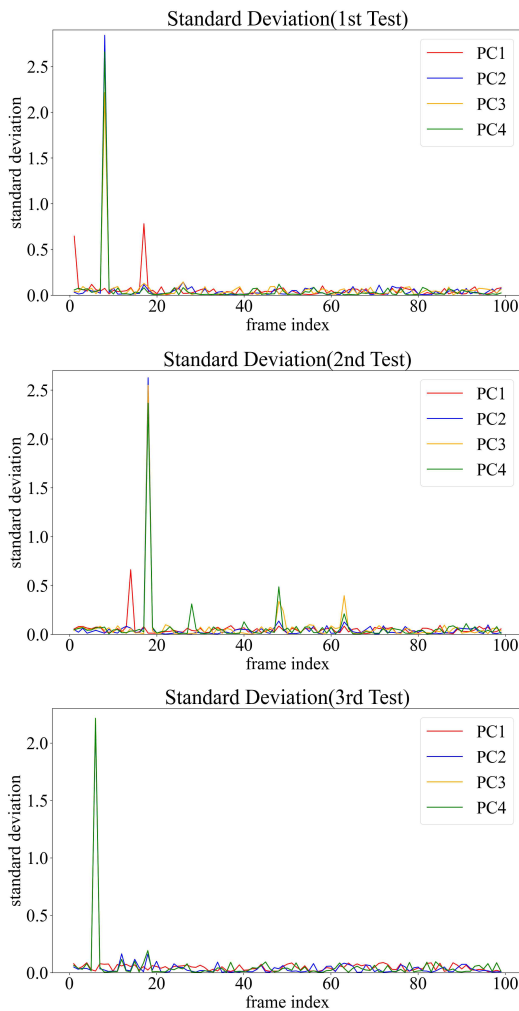


그림 7. 카메라의 프레임별 취득 시간 차 표준편차 그래프

그림 7은 취득 테스트 결과 PC 별로 기록된 카메라의 프레임별 취득 시간 차의 표준편차 그래프이다. 세 번의 테스트 결과 Receiver PC에서 지연이 발생하여 표준편차가 크게 기록되는 프레임이 발생하였고, 이는 Receiver PC가 메시지를 받을 때 멀티캐스트 네트워크에서의 문제가 발생하였기 때문이라고 추정하였다. 반면 각 테스트 별로 취득된 총 400개의 프레임 중 표준편차가 0.1ms 이하로 발생하는 프레임의 수가 각각 383개, 382개, 386개로, 각 테스트 별로 전체 프레임 중 95.75%, 95.5%, 96.5%의 프레임에서 동기적으로 영상이 취득되었다.

5. 결론

엔데믹 이후로도 시공간적인 이유로 인하여 비대면 교육은 여전히 진행되고 있고, 시공간적인 한계를 극복할 수 있게 해주는 메타버스의 몰입형 콘텐츠는 비대면 교육의 가능성을 더욱 높이고 있다. 하지만 몰입형 콘텐츠의 기술적 특성상 실사 영상을 활용한 몰입형 콘텐츠 제작의 난이도가 높고 고가의 장비와 높은 기술력이 요구되기 때문에 상용화 과정에서 어려움을 겪고 있다. 본 논문은 이러한 문제 해결을 위해 6DoF VR 몰입형 비디오 취득 시스템을 소개하였고, 다수의 Basler 카메라를 활용하여 직접 구현한 멀티뷰 카메라 시스템과 이를 활용한 멀티뷰 비디오 취득 결과를 보였다.

본 논문은 6DoF VR 몰입형 비디오 취득을 위한 카메라 시스템을 구축하고, 카메라 SDK에 맞는 API를 호출하여 동기적으로 영상을 캡처하는 취득 시스템을 구현하였다. 향후 연구에서는 네트워크를 통해 시그널을 받는 Receiver PC에서 지연이 발생할 수 있는 점을 보완하여 더욱 정확한 네트워크 시그널링 제어가 가능하도록 하고, 추가로 깊이 이미지 추정을 통해 MIV 표준 encoder 실행이 가능하도록 시스템을 보완하고자 한다.

Acknowledgement

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.RS-2022-00167169. 이동형 로봇 기반 실사 메타버스 실감형 비디오의 획득 및 처리 기술 개발)

7. 참고 문헌

- [1] 장대련, “[2021 대한민국 교육브랜드 대상] 대면·비대면 장점 살린 혼합형 교육 불가피”, 중앙일보, 2021.06.22.
- [2] 이성균, “메타버스기반 가상공간에서의 교육 생태계 성장 가능성”, 대한공간정보학회 학술대회, 2022-05, pp.241-244.
- [3] 최기호, 방건, “MPEG-I 비디오 표준화 동향”, 한국방송미디어공학회 방송과 미디어 제23권 4호 (360VR 및 몰입형 미디어), 2018-10, pp.353-369
- [4] 이재규, 김의창, “교육의 집중도 향상을 위한 몰입형 메타버스 시스템 연구”, e-비즈니스 연구, 2021-12, pp.3-14
- [5] 김보미, 최진영, 이규하, 김미경, “에듀테크 기반 비대면 K-Dance 수업 프로그램 개발 연구”, Asian Journal of Physical Education of Sport Science(AJPSS), 2023-11, pp.71-86
- [6] Jinsil Hwaryoung Seo, Caleb Kicklighter, Brittany Garcia, Soo Wan Chun, Elizabeth Wells-Beede, “Work-in-Progress-Design and Evaluation of 360 VR Immersive Interaction in Nursing Education”, 2021 7th International Conference of the Immersive Learning Research Network (iLRN), 2021-05, pp.1-3
- [7] 이봉호, 김준수, 정준영, 윤국진, 정원식, “실사 환경에서의 다시점 영상 획득 워크플로우”, 한국방송미디어공학회 학술발표대회 논문집, 2022-11, pp.154-156