

실사 및 CG 환경에서의 다시점 입체영상 획득 기술

정준영, 윤국진, 정원식

한국전자통신연구원

jjy0120@etri.re.kr, kjyun@etri.re.kr, wscheong@etri.re.kr

Acquisition Workflow of Multiview Stereoscopic Video at Real and CG Environment

Jun Young Jeong, Kug Jin Yun, Won-Sik Cheong

Electronics and Telecommunications Research Institute

요 약

고정된 위치를 중심으로 회전운동만 체험할 수 있는 3 자유도(DoF: Degrees of Freedom)를 넘어 위치를 변경하며 운동시차까지 포함된 6 자유도를 지원하는 몰입형 미디어에 대한 연구가 지속해서 진행되고 있다. 특히 부드러운 시점 변경을 제공하기 위해 특정 위치에서 샘플링 된 여러 개의 텍스처(또는 컬러) 및 깊이맵 영상(MVD: Multiview Video plus Depth)으로 구성된 다시점 영상을 통해 실제로 획득되지 않은 위치에서의 영상을 만들어내는 가상시점 합성(virtual view synthesis) 기술이 많이 사용되고 있다. 본 논문에서는 몰입형 미디어의 대표적인 데이터 형식인 다시점 영상을 실사 및 컴퓨터 그래픽스(CG: Computer Graphics) 환경에서 획득하는 방법에 대해 설명한다.

1. 서론

메타버스(Metaverse), VR(Virtual Reality)과 같이 사용자에게 현실에 가까운 현실감 및 몰입감을 제공하는 몰입형 미디어에 대한 관심이 대두되면서, 이의 서비스를 위해 필요한 압축, 전송, 재현 등 다양한 요소 기술에 대한 연구 및 표준화가 현재 활발히 진행되고 있다. 특히, 고정된 위치를 기준으로 회전운동만 가능한 기존의 3자유도를 넘어 위치이동에 따른 병진운동까지 포함된 6자유도를 제공하는 기술에 대한 연구가 지속적으로 이뤄지고 있다. 6자유도가 가능하다는 것은 사용자가 어떤 위치에 있더라도 그 위치의 영상을 볼 수 있다는 의미인데, 현실적으로 이동가능한 모든 위치에 카메라를 설치하는 게 불가능할뿐더러 수 많은 영상을 보관하기 위해선 대용량의 저장 공간이 필요하다. 이를 위해, 몇몇 소수의 위치에서 드문드문 획득된 여러 개의 텍스처 및 깊이맵 영상으로 구성된 다시점 영상을 바탕으로 임의 위치에서의 영상을 만들어 내어 부

드러운 시점 이동을 지원하는 가상시점 합성 기술을 많이 사용하고 있다. 본 논문에서는 6자유도를 지원하는 몰입형 미디어의 대표적인 데이터 형식 중 하나인 다시점 영상을 실사 및 CG 환경에서 획득하는 과정에 대해 설명한다.

2. 실사 기반 다시점 콘텐츠 획득

실사 환경에서 다시점 콘텐츠를 획득하기 위한 첫 번째 단계는 제공하려는 몰입형 서비스의 시나리오를 고려하여 카메라 구조체를 설계하는 것이다. 예를 들어 관광지 체험처럼 특정 위치를 중심으로 주변을 둘러보는 경우에는 그림 1(좌)와 같은 발산형이[1], 스포츠 경기 관람 시에 관중석에서 주로 경기장을 보는 것처럼 특정 관심영역을 시청하는 경우에는 그림 1(우)와 같은 수렴형이[2] 적합할 것이다. 보통은 여러 대 카메라 간의 시간적 동기화, 색 보정 등 작업의 편의를 위해 동일한 사양의 카메라를 사용하며, 카메라 모델은 예산 범위 내에서 지원하고자 하는 서비스의 해



그림 1. 발산형(좌) 및 수렴형(우) 카메라 구조체 예시

상도, 프레임율, 운동시차 범위 등을 고려하여 결정한다. 더불어, 서로 다른 시점에서의 텍스처 영상을 비교하여 대응 영역을 찾는 스테레오 정합(stereo matching) 과정을 통해 거리 정보를 도출하는 깊이맵 추정 단계를 위해 시점 간에는 반드시 일정 범위의 중첩 영역이 존재해야 한다.

두 번째 단계는 카메라 구조체를 기준으로 체커보드(checkerboard)와 같은 패턴 영상을 거리 및 방향을 다양하게 바꾸면서 여러 번 촬영하고, 본 영상을 바탕으로 카메라 보정(camera calibration) 작업을 수행하는 것이다. 카메라 보정은 각 카메라의 내부 변수(초점 거리, 주점 위치, 왜곡 계수) 및 외부 변수(카메라 간 상대적 위치 및 각도) 값을 계산하는 과정을 의미하며, 본 정보는 깊이맵 추정 및 가상시점 합성 단계에서 한 시점에서의 텍스처 정보를 다른 위치로 투영(projection)할 때 반드시 필요하다. MATLAB 등은 안정적인 성능의 체커보드 기반 카메라 보정 툴과 관련 상세 매뉴얼을 제공하고 있어[3], 어렵지 않게 내/외부 변수 정보를 얻을 수 있다. 체커보드를 사용하는 번거로운 작업 없이 촬영 장면에 존재하는 특징 정보(객체 모서리 등)를 사용하는 기술도 여럿 존재하지만, 체커보드만큼 특징 정보가 뚜렷하지 않아 정확한 내/외부 변수 값을 도출하는 데에는 어려움이 있다.

세번째 단계는 캘리브레이션 작업이 완료된 카메라 구조체를 통해 여러 시점에 대응하는 텍스처 영상을 촬영하고, 네 번째 단계에서 이전 단계를 통해 획득한 카메라 내/외부 변수 정보와 텍스처 영상을 통해 각 시점의 깊이맵 영상을 추정하는 것이다. 깊이맵 획득에는 라이다, TOF(Time of Flight) 센서와 같이 피사체에 빛을 쬐서 직접적으로 거리를 측정하는 능동형 센서를 활용할 수도 있지만, 가격, 촬영 환경에 대한 제약 등으로 인해 스테레오 정합 기술이 널리 활용되고 있다. 해당 기술은 일반적으로 매칭비용 계산(cost computation), 매칭비용 집적(cost aggregation), 매칭비용 최적화(cost optimization), 깊이맵 정제(depth refinement)를 포함한 네 과정으로 구성되어 있다. 매칭비용 계산은 서로 다른 시점에서의 텍스처 영상의 강도 차이를 픽셀 단위로 여러 깊이 단계에서 계산하고, 각 깊이 단계에서의 색 차이(이하 비용) 정보로 구성된 3D 비용 볼륨(cost volume)을 생성한다. 이때 각 픽셀

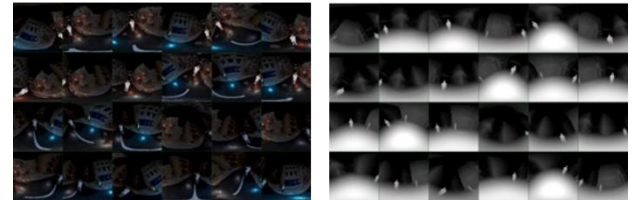


그림 2. CG 모델 내 가상 카메라 배치 형태(상) 및 획득 다 시점 콘텐츠 예시(하)

별로 비용 값을 최소화하는 깊이값을 선택하여 초기 깊이맵을 생성할 수도 있지만, 비용 볼륨의 신뢰도를 높이기 위해 매칭비용 집적을 통해 비용 정보를 지역적으로 누적한다. 예를 들어, 텍스처 정보를 참고하여 색상이 엇비슷한 픽셀의 매칭비용 값을 평균하는 방법이 있다. 하지만, 특징점 부재 영역, 폐색(occlusion) 영역 등에서는 여전히 신뢰도가 떨어진다. 이에, 매칭비용 최적화는 깊이맵은 매끄럽다(smooth)는 가정 하에 인접 영역의 깊이 값까지 함께 고려하는 비용함수(cost function)를 정의하고, 그래프 컷(graph-cut)과 같은 최적화 기법 등을 통해 해당 비용함수를 최소화하는 해를 깊이맵으로 설정한다. 깊이맵 정제는 깊이맵을 보정하기 위해 선택적으로 적용하는 후처리 단계로, 주로 사용하는 방법 중 하나는 각 시점의 깊이맵을 다른 시점의 위치로 투영하여 동일한 위치에 대응하는 픽셀 간의 깊이 값 차이를 통해 신뢰도를 평가하고, 신뢰도가 낮은 영역을 주변 정보로 대체하는 것이다.

3. CG 기반 다시점 콘텐츠 획득

CG 기술을 활용한 기존의 6자유도 체험 서비스는 단말에 3D CG 모델의 다운로드를 완료한 상태에서 사용자의 움직임에 맞춰 영상을 렌더링하는 방식이 주를 이룬다. 하지만, 대용량의 복잡한 3D 모델을 통해 스트리밍 서비스와 같이 압축 및 전송까지 필요한 서비스를 제공하기 위해서는 많은 연구가 필요할 것이다. 반면 2D 형식의 영상은 오래 전부터 TV, 스마트폰 등을 통한 수요가 많아 압축, 전송 등 관련 기술의 성숙도가 매우 높을 뿐만 아니라, 호환성을 이루는 단말 또한 널리 보급된 상태이다. 따라서, 그림 2와 같이 3D 모델 내 특정 공간에 가상의 카메라 구조체를 배치하여 다시점 영상을 얻으면, 비록 운동시차 지원 범위는 줄어들 것이지만, 기존의 서비스 워크플로우를 통해 6자유도 서비스 제공이 가능할 것이다.

실사와 유사한 렌더링 품질로 큰 주목을 받는 유니티(Unity3D), 언리얼(Unreal)과 같은 게임 엔진은 가상 카메라를 통해 텍스처 및 깊이맵 영상 촬영을 지원하는 라이브러리를 제공하고 있다 [4,5]. 실사 촬영 때와는 다르게 자유자재로 원하는 위치에 필요한 사양의 카메라를 배치할 수 있어 별도의 캘리브레이션 작업이 필요 없으며, 이미 공간 정보를 알고 있어 매우 정밀한 실측 깊이맵을 획득할 수 있다.

감사의 글

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(2022-0-00022, 초실감 메타버스 서비스를 위한 실사기반 입체영상 공간컴퓨팅 기술 개발)

참고문헌

- [1] <https://www.blog.google/products/google-ar-vr/experimenting-light-fields/>
- [2] ISO/IEC JTC1/SC29/WG11, Undistorted CBABasketball Test Sequence for MPEG-I Visual, M58500, Dec. 2021
- [3] <https://kr.mathworks.com/help/vision/camera-calibration.html>
- [4] <https://docs.unity3d.com/Packages/com.unity.aov-recorder@0.1/manual/index.html>
- [5] <https://docs.unrealengine.com/4.26/ko/Animating-Objects/Sequencer/Overview/>