

공간 미디어 서비스를 위한 실사 다시점 영상 획득에 관한 연구

이봉호, 김준수, 전호성, 윤국진, 정원식, 추현곤
한국전자통신연구원

leebh@etri.re.kr

A Study on the Acquisition of Natural Multiview Video for Spatial Media Services

Bongho Lee, Joonsoo Kim, Hosung Jeon, Kuk Jin Yun, Won-Sik Cheong, Hyun-Gon Choo
Media Research Division of ETRI

요 약

본 논문은 실사 공간 미디어 서비스를 위한 실사 다시점 영상 획득에 관한 것으로, 동종 카메라 및 단일 조건 환경 하에서의 정형화된 획득이 아닌 이종 카메라, 카메라간 비동기 및 비정형 구조를 기반으로 한 획득 시스템 설계/구축 및 이를 기반으로 한 시험 영상을 촬영한 내용을 소개하고자 한다. 정형화된 카메라 셋업의 경우, 일반적으로 획득 공간의 깊이가 클수록 복원력이 떨어지는 단점이 있어 사용자가 줌 효과와 같이 앞으로 이동하여 체험하고자 할 경우 복원 해상력에 한계가 있다. 본 논문에서는 이를 개선하기 위해 주 관심 영역의 복원 해상력을 향상할 수 있는 추가 카메라를 배치하여 획득하한 후 복원해 본 결과를 소개하고자 한다.

I. 서 론

최근 NeRF[1]나 3DGS[2]와 같은 딥 러닝 기반 3 차원 공간 복원 기술의 발전에 힘입어 기존 평면적인 형태의 2 차원 미디어를 넘어 실제와 유사한 3 차원 공간을 복원하고 체험할 수 있는 기술 개발이 활발하게 진행되고 있다. MPEG 에서는 가우시안 스플랫팅 기반 장면 데이터의 압축 및 전송을 위한 Gaussian Splat Coding(GSC) 표준화를 진행하고 있으며[3], 메타버스 표준 포럼의 3D Asset Interoperability WG 에서는 3D 에셋 상호운용성 표준화를 목표로 Gaussian Splat 을 glTF extensions 으로 추가하는 표준화를 진행하고 있다[4].

본 논문은 이와 관련된 연구로 가상 공간이 아닌 실사 공간을 복원하여 사용자가 체험할 수 있도록 하기 위한 실사 다시점 입체 영상 획득에 관한 것이다. 구체적으로는, 기존 정형화된 획득 환경이 아닌 획득 환경의 제약을 완화하고 보다 다양성 및 편이성을 가질 수 있는 실사 다시점 영상 획득 구조를 제안하고, 이를 기반으로 구축한 시스템 및 시험 촬영한 결과를 소개하고자 한다.

또한 복원된 공간의 임의의 관심 영역의 복원 해상력을 향상시키기 위해 별도의 이종 카메라를 사용하여 획득하고 복원한 결과를 소개하고자 한다.

II. 실사 다시점 영상 획득을 위한 카메라 구조 및 획득 시스템

NeRF 나 3DGS 에 사용되는 데이터셋의 영상들은 가급적 동일한 특성을 갖는 카메라를 사용하여 프레임이 동기된 형태로 촬영된다. 이는 GS 복원에 큰 영향을 주는 초기 포인트 클라우드를 잘 찾기 위한 것으로, 획득

후 요구되는 geometry 왜곡, 색 차 및 프레임 비 동기로 인한 오차를 줄일 수 있다.

하지만 실제 촬영 현장을 고려하면, 다양한 카메라와 렌즈, 비 동기 촬영, 비 정형 카메라 배치, 다양한 노출, 카메라 별 다양한 심도 설정 및 무엇보다도 기존 2D 미디어 촬영과의 호환성 내지 활용성을 고려할 필요가 있다. 이를 통해 실사 다시점 영상의 획득 자유도와 편이성 향상을 가져올 수 있을 것으로 사료된다.

본 논문은 이러한 점을 감안하여, 비 정형 카메라 배치 구조, 비동기, 이 기종 카메라 및 렌즈 환경에 기반한 리그 및 획득 시스템을 고려하였다. 비 정형 카메라 배치는 공간의 전경과 주 관심 영역을 별도로 분리 촬영 가능한 구조를 고려하였으며, 비동기의 경우에는 전경을 촬영하는 카메라의 경우 genlock 이 지원되지 않는 IP 기반 동기를 지원하는 카메라를 고려하였다. 마지막으로 이기종 카메라와 렌즈는 전경과 주 관심 영역에 각기 다른 카메라와 렌즈를 적용하는 것을 고려하였다.

획득 테마의 경우 공간 내 빠른 동적인 움직임이 가정하고, 무엇보다도 180 도 내에서 사용자가 이동하며 체험할 수 있도록 실내 스포츠인 당구를 선정하여 획득하였다.

카메라 리그의 경우, 복원 공간 내에서 180 자유시점 이동이 가능하도록 그림 1 과 같이 리그를 설계하고 구축하였으며, 전경과 주 관심 영역을 별도로 획득하기 위한 이종의 카메라를 사용하였다. 전경 획득을 위해 24 대의 Sony RX0 II 카메라를, 당구대 주변을 주 관심 영역으로 획득하기 위해 Blackmagicdesign 사의 Micro Studio 4K 카메라 18 대를 적용하였다. RX0 카메라는 약 83 도의 화각을 갖는 카메라로 24 대를 2x 12(수직 x 수평) 형태로 배치하여 넓은 공간 획득을 목적으로 하며, BMD 카메라는 약 46 도 화각을 갖도록 렌즈를 조정하여 주 관심 영역인 당구대를 중점적으로 획득하도록 구성하였다.

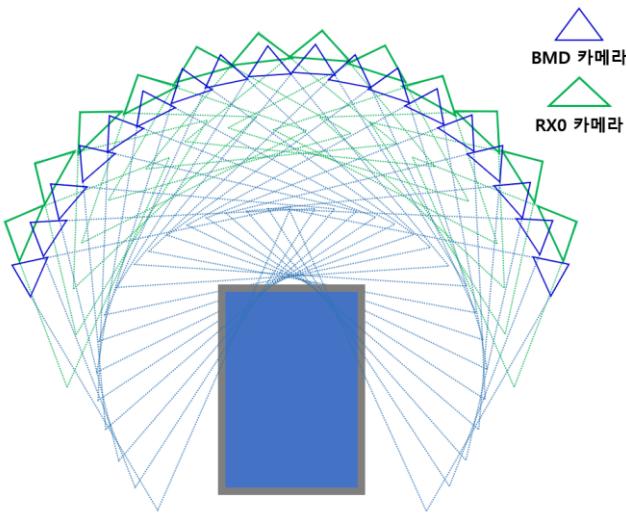


그림 1. 카메라 배치 구조

획득 시스템의 경우, 선행 연구[5]에 사용된 획득 시스템을 기반으로 재 구축하였으며, genlock 프레임 동기를 지원하는 BMD 카메라의 경우, 동기 신호 생성기, BMD smart videohub, ATEM 8K Switcher 및 실시간 4K frame grabber를 장착한 PC로 구성된 획득 시스템을 그림 2와 같이 구축하였다.

전경 영역 및 프레임 비동기 획득을 위한 시스템은 소니 RXO II를 사용하여 그림 3과 같이 구축하였다. 각 카메라는 제어 박스와 연동되며 로컬 LAN 환경을 통해 제어용 웹 UI를 통해 제어된다. RXO의 프레임 동기는 IP 기반 방식으로 최대 반 프레임 이내 동기 오차를 가진다.

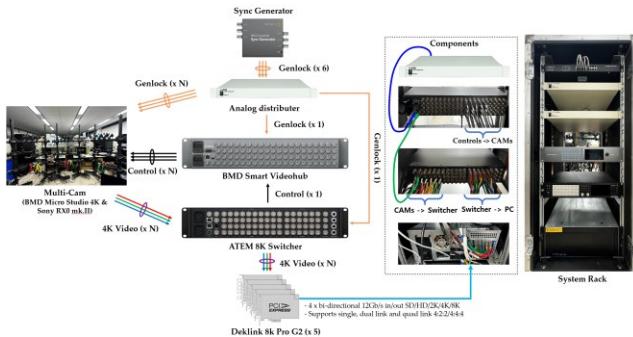


그림 2. 관심 영역 획득 시스템(BMD 카메라)

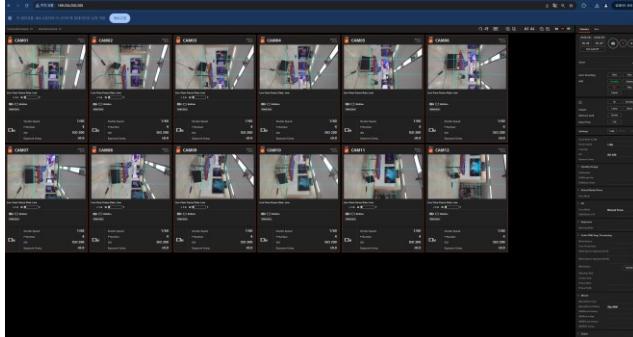


그림 3. 전경 영역 획득 시스템(RXO)

III. 시험 획득 및 공간 복원

II 장에서 서술한 획득 시스템의 기능 검증을 위해 그림 4와 같이 다시점 영상 획득 리그를 구축하고 시험 촬영을 진행하였다.

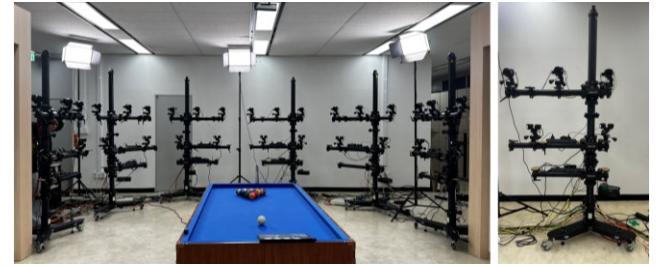


그림 4. 구축된 카메라 리그

3 단으로 구성된 6 개의 리그에 상단에 주 관심 영역 획득을 위한 BMD 카메라 18 대를 거치하고, 중단과 하단에 각각 12 대씩 총 24 대의 RXO 카메라를 거치하였다. BMD 및 RXO 카메라 설정 파라미터는 <표 1>과 같다.

<표 1> 카메라 설정 파라미터

파라미터	BMD	RXO
Baseline(cm)	50	75
Alignment	Toe-in(수평, 수직), Toe-in(수평)/ Parallel(수직)	
FOV (degree)	46	83
Shutter speed	1/50	1/60
White Balance	5600K	5600K
Focus	수동	수동
FPS	30	30
Frame Sync	Genlock	IP

획득 공간 규모를 고려하여 BMD의 경우 베이스라인을 50cm로 조정하였으며, RXO의 경우에는 75cm로 조정하였다.

사전 촬영은 당구 경기 장면에 대해서 각기 카메라 정렬 방식을 달리 하여 촬영을 진행하였다. 수평 1 열로 배치한 BMD 카메라의 경우 toe-in 방식으로 정렬하고, RXO 카메라의 경우 Toe-in(수직, 수평) 및 Toe-in(수평)/Parallel(수직) 방식으로 정렬하였다.



그림 5. 획득 영상 예

그림 5는 전술한 카메라 리그 및 시스템을 통해 획득된 영상에 해당된다. BMD 카메라의 경우 genlock 동기를 통해 프레임 동기가 일치되게 획득이 되었으며 RXO의 경우에는 IP 동기 방식으로 최대 반 프레임 시간차를 가지고 획득됨을 확인하였다. 그림에서 확인 할

수 있는 바와 같이 두 카메라에 적용된 렌즈 및 CCD 특성이 달라 색차가 있음을 확인할 수 있다.

RX0는 83도 화각을 가지고 전경을 충분하게 커버할 수 있도록 촬영을 하였으며, BMD 카메라의 경우 당구대 중앙부터 뒤쪽 부분을 주요 관심영역으로 하여 촬영됨을 확인할 수 있다.

그림 6과 7은 RX0로 획득된 24 이미지 샘플을 이용하여 Inria의 3DGS[6]를 사용하여 모델 데이터를 생성한 후 렌더링하여 복원한 결과 예에 해당한다.

그림 7은 초기 포인트(a), 가우시안 타원체(b), rasterized 된 시점 영상(c) 및 당구대 모서리 주변까지 이동한 시점 영상(d)을 도시하고 있다. 전경을 위주로 획득한 데이터만 사용하여 복원된 공간을 앞으로 깊이 이동한 후 해당 시점을 보고자 할 경우 그림에서 도시된 예(d)와 같이 복원 해상력이 떨어질 수 있다. 획득 공간은 카메라 위치로부터 면 쪽 당구대까지의 거리가 약 4.8m이며 당구대의 길이는 1.8m로, 그림 7 (d)는 당구대 중간까지 이동(약 4m)한 시점에서 복원된 영상이다.



그림 6. 전경 공간 복원 예

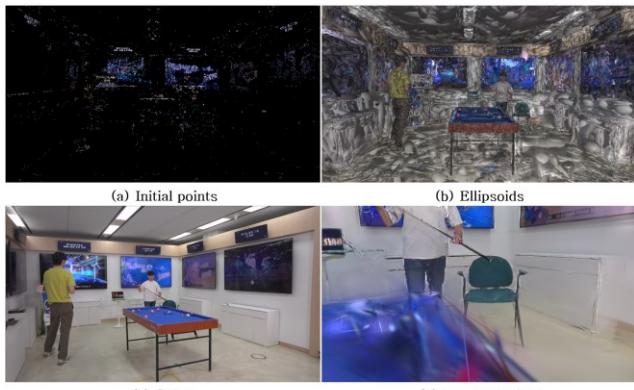


그림 7. 전경 공간 복원 결과

그림 8과 9는 BMD 카메라로 획득된 18 장의 이미지 샘플을 이용하여 Inria의 3DGS[6]로 생성 및 렌더링된 복원 결과이다. 주 관심 영역 공간을 별도로 획득한 영상으로, 그림 9(d)는 전경 공간을 획득한 데이터부터 복원된 공간으로 그림 7(d)에 비해 복원 해상력이 좋은 것을 확인할 수 있다. 이러한 셋업으로 촬영을 할 경우, 스포츠 경기와 같이 전경에 해당하는 공간과 주 관심 공간을 별도로 획득한 후 병합하여 생성 복원하면 사용자 만족도를 개선할 수 있을 것으로 사료된다.



그림 8. 주 관심 영역 복원 예

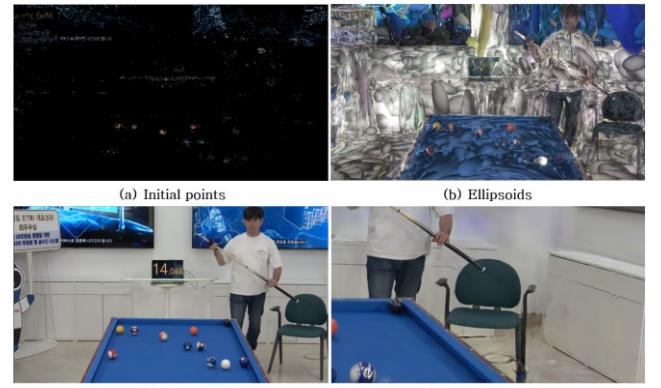


그림 9. 주 관심 영역 복원 결과

아래 그림 10은 RX0와 BMD 카메라로부터 획득된 인접한 프레임의 42 장의 이미지를 이용하여 복원한 예에 해당한다. 이 기종 카메라 특성에 의해 획득한 공간이 중첩되는 부분에 artifacts(그림에서 사각형 창 부분)가 발생함을 확인할 수 있었다.



그림 10. 전경과 주 관심 공간 병합 복원 예

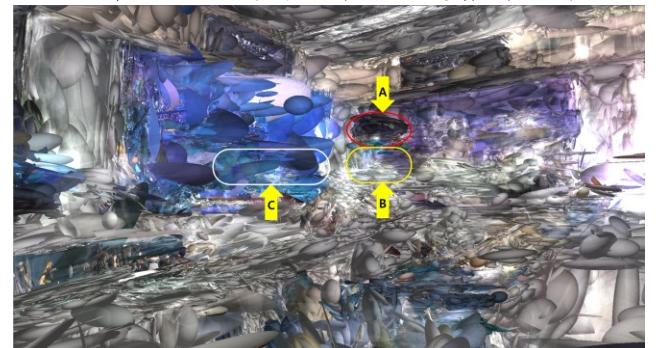


그림 11. 인물 주변 Ellipsoids

그림 11 은 그림 10 에 해당하는 복원 공간내 인물 주변 타원체들로, 화살표 A 는 인물의 머리에 해당하며, B 는 얼굴, C 는 뒤 TV 스크린 영역에 해당한다. 분석 결과 인물의 얼굴에 해당하는 타원체가 구성되지 않은 것(B)을 확인하였다. 이는 두 카메라 특성에 기인한 것으로 geometry 왜곡 및 색 차로 인한 것으로 관련 보정을 통해 극복이 될 수 있을 것으로 사료된다. 또한 프레임 동기를 만족하는 genlock 기반의 BMD 카메라와 IP 기반의 반 프레임 시차를 가지는 RXO 카메라간 시차에 의한 현상으로도 보이며 추후 이러한 보정을 거친 후 복원할 경우 복원력이 좋아질 것으로 사료된다.

IV. 결론

본 논문에서는 실사 공간 미디어 서비스를 위한 실사 다시점 영상을 획득하기 위해 이종 카메라, 카메라간 비동기 및 비정형 구조를 기반으로 한 실사 다시점 영상 획득 시스템을 설계/구축하고, 이를 기반으로 한 시험 영상을 촬영한 결과를 소개하였다.

카메라 센서 및 렌즈 특성이 다른 이종 카메라를 이용하여 전경과 주 관심 영역을 별도로 획득하였으며, vanilla 3DGS[6]를 사용하여 가우시안 프리미티브를 생성하고 복원해 보았다. 복원 결과 전경 영역에 해당하는 획득 데이터만을 이용하여 복원한 결과, 전반적인 전경에 대한 복원력은 GT 이미지 대비 양호한 것으로 확인 되었으나 복원 공간내에서 주 관심 영역 근처까지 이동하여 확인 한 복원 해상력은 다수의 artifacts 가 생성됨을 확인하였다. 주 관심 영역에 대한 획득 데이터를 이용한 복원 결과, 주 관심 영역 주변의 복원 해상력은 전경 영역에 비해 양호함을 확인하였다.

이러한 현상을 종합해 볼 때, 본 논문에서 시도한 전경과 주 관심 영역을 별도로 획득하여 복원할 경우 사용자의 이동에 따른 주 관심 영역의 체험 만족도를 개선할 수 있을 것으로 사료된다.

마지막으로, 두 영역을 병합하여 복원한 영상의 경우 이종 카메라 및 렌즈의 특성으로 인해 geometry 및 색상 왜곡이 발생함을 확인하였다. 이를 개선하기 위해서는 각 시점 별 geometry 왜곡 보정 및 색 보정이 적용되어야 할 것으로 사료된다.

향후 연구 방향으로는, 왜곡 보정, 색 보정, 시차 보정을 기반으로 기능 검증용 GS 모델을 포함하는 워크플로우를 개발할 예정이며, 촬영 현장에서 빠르게 복원 품질을 확인할 수 있는 고속 프리뷰 GS 모델을 연구할 예정이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신 기획평가원의 지원을 받아 수행된 연구임 (2022-0-00022, RS-2022-II22002, 초실감 메타버스 서비스를 위한 실사기반 입체영상 공간컴퓨팅 기술 개발)

참 고 문 헌

[1] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng., “Nerf: Representing scenes as neural radiance fields for view synthesis.” arXiv preprint arXiv:2003.08934, 2020.

[2] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis, “3d gaussian splatting for real-time radiance field rendering,” ACM Transactions on Graphics (ToG), 42(4):1– 14, 2023.

[3] 이광순, 정준영, 추현곤, “MPEG에서의 Gaussian Splat Coding(GSC) 표준화 현황,” 방송과 미디어, Vol. 30, No. 3.

[4] Neil Trevett, Patrick Cozzi, “Town Hall_Gaussian Splats – Ready for Standardization,” SIGGRAPH Frontiers ’ 25, August 10– 14, 2025, Vancouver, BC, Canada.

[5] 이봉호, 김준수, 유상현, 윤국진, 이광순, 정원식, 추현곤, “다시점 실사 입체영상 획득 및 MIV 인코딩,” 2023년도 한국통신학회 추계종합학술발표회.

[6] <https://github.com/graphdeco-inria/gaussian-splatting>. “3D Gaussian Splatting for Real-Time Radiance Field Rendering”.