

Structure-from-Motion 기법에 따른 MPEG GSC 중심의 Gaussian Splat 품질 변화 연구

이승미 최경호 한건희 정종범

국립군산대학교 컴퓨터정보공학과

{qw easd030622, cgh3578, gunhee9369, jongbeomjeong}@kunsan.ac.kr

Research on MPEG GSC-Based Gaussian Splat Quality Difference by Structure-from-Motion

Seung-Mi Lee GyeongHo Choi GunHee Han Jong-Beom Jeong

Department of Computer Science and Information Engineering,

Kunsan National University (KSNU)

요 약

3D Gaussian splatting (3DGS)은 입력 영상으로부터 추정된 카메라 포즈 (camera pose)와 포인트 클라우드 (point cloud)를 초기 구조로 사용하여 장면을 재구성하는 방식이므로, 선행 structure-from-motion (SfM) 결과의 품질이 최종 렌더링 성능에 영향을 줄 수 있다. 본 논문은 moving picture experts group (MPEG) Gaussian splat coding (GSC) 실험 데이터 중 library 시퀀스를 대상으로 COLMAP, GLOMAP, RealityScan을 적용하고, 각 SfM 결과를 동일한 gsplat 기반 3DGS 학습 파이프라인에 입력하여 재구성 품질을 비교하였다. 실험 결과, SfM 기법에 따라 포인트 클라우드 구성과 최종 렌더링 품질에 차이가 나타났으며, RealityScan 기반 결과가 peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), learned perceptual image patch similarity (LPIPS) 기준에서 가장 우수한 성능을 보였다. 이를 통해 3DGS 기반 3차원 재구성에서 SfM 파이프라인 선택이 초기화와 최종 렌더링 품질에 중요한 영향을 줄 수 있음을 확인하였다.

1. 서론

최근 가상현실, 증강현실 및 실감형 미디어 서비스의 발전에 따라 사용자가 임의의 위치와 방향에서 장면을 관찰할 수 있는 6자유도 (6 degrees of freedom, 6DoF) 기반 몰입형 비디오 기술이 주목받고 있다. MPEG-I Part 12 MPEG Immersive Video는 다수의 텍스처 뷰 (texture view)와 깊이 뷰 (depth view)를 이용하여 제한된 공간 내에서 사용자가 임의의 위치와 방향으로 장면을 관찰할 수 있는 6DoF 몰입형 비디오 표현을 목표로 한다[1]. 이러한 몰입형 콘텐츠를 제공하기 위해서는 다수의 시점에서 취득된 영상으로부터 장면 정보를 표현하고, 사용자의 시점 변화에 따라 새로운 시점 영상을 합성할 수 있어야 한다.

3DGS는 3차원 장면을 다수의 가우시안 프리미티브 (Gaussian primitive)로 표현하고, 이를 렌더링하여 새로운 시점의 영상을 합성하는 방식이다. 3DGS는 SfM 과정에서 생성된 희소 포인트 클라우드 (sparse point clouds)를 초기 3D 가우시안 (Gaussian)으로 사용하고, 학습 과정에서 가우시안의 위치, 공분산, 색상 및 불투명도 등을 최적화한다[2]. 이에 따라 3DGS 기반 장면 표현에서는 입력 영상뿐만 아니라 카메라 파라미터와 희소 포인트 클라우드의 구성이 중요한 요소이다.

3DGS 학습을 위해서는 입력 영상에 대한 카메라 파라미터와 초기 3차원 구조 정보가 필요하며, 이러한 정보는 일반적으로 SfM을 통해 추정된다. SfM은 여러 장의 2차원

이미지로부터 카메라의 위치와 방향, 그리고 희소 3D 구조 (sparse 3D structure)를 추정하는 과정이다[3]. 따라서 SfM 기법에 따라 카메라 정합 결과와 희소 포인트 클라우드의 밀도 및 분포가 달라질 경우, 후속 3DGS 학습 결과에도 차이가 발생할 수 있다.

본 논문은 MPEG GSC 실험 데이터 중 library 시퀀스를 대상으로 SfM 기법에 따른 3DGS 품질 변화를 비교한다[6]. COLMAP은 SfM 및 Multi-View Stereo (MVS)를 지원하는 범용 3차원 재구성 파이프라인이며[4], GLOMAP은 글로벌 SfM 기반의 이미지 재구성 파이프라인이다[5]. 본 논문은 COLMAP에 통합된 글로벌 매퍼 (global mapper) 기반 재구성 결과를 GLOMAP으로 표기하여 비교군으로 사용한다. RealityScan은 영상 기반 3차원 재구성 도구이며, 본 실험은 RealityScan에서 생성한 SfM 결과를 COLMAP 텍스트 형식으로 내보낸 뒤 gsplat 입력으로 사용하였다[7].

본 논문의 구성은 다음과 같다. 2절에서는 SfM 및 3DGS의 배경을 설명하고, 3절에서는 실험 조건과 결과를 제시한다. 마지막으로 4절에서 결론을 맺는다.

2. 관련 기술 및 선행 연구

SfM은 다수의 2차원 이미지로부터 카메라 포즈와 장면의 희소 3차원 구조를 추정하는 기법이다[3]. 일반적인 SfM 파이프라인은 특징점 추출, 특징점 매칭, 카메라 포즈 추정, 희소 포인트 클라우드 생성 과정으로 구성된다. COLMAP은 SfM 및 MVS를 지원하는 범용 3차원 재구성 파이프라인이며[4], 본 논문에서는 글로벌 SfM 계열 비교군으로 COLMAP의 글로벌 매퍼 기반 재구성 결과를 GLOMAP으로 구분한다[5].

3DGS는 SfM으로부터 얻은 희소 포인트 클라우드를 초기 3D 가우시안으로 사용하고, 각 가우시안의 위치, 공분산, 색상 및 불투명도 등을 최적화하여 새로운 시점 영상을 렌더링하는 방식이다[2]. 기존 신경망 방사형 필드 (neural radiance fields, NeRF) 계열 방식이 신경망을 이용하여 방사형 필드 (radiance field)를 표현하고 볼륨 렌더링 (volume rendering)을 수행하는 것과 달리, 3DGS는 명시적인 가우시안 프리미티브를 사용하여 장면을 표현한다. 또한 3DGS는 학습 과정에서 가우시안의 밀집화 (densification) 및 가지치기 (pruning)를 수행한다. 따라서 초기 희소 포인트 수뿐만 아니라 카메라 포즈의 안정성, 희소 포인트 클라우드의 공간적 분포, 학습 중 가우시안 수 변화가 최종 렌더링 품질에 영향을 줄 수 있다.

SfM 결과는 3DGS 학습의 초기 입력으로 사용되므로, SfM

단계에서 발생한 차이는 후속 학습 과정에 누적될 수 있다. 예를 들어 카메라 포즈가 불안정하거나 희소 포인트 클라우드가 장면 구조를 충분히 반영하지 못하는 경우, 초기 가우시안의 분포가 장면의 실제 구조와 달라질 수 있다. 반대로 희소 포인트 수가 많더라도 포인트의 공간적 분포나 카메라 포즈의 일관성이 낮다면 최종 렌더링 품질 향상으로 직접 이어진다고 단정하기 어렵다. 따라서 SfM 기법별 결과를 단순히 포인트 수만으로 비교하기보다는, 최종 가우시안 수와 렌더링 품질 지표를 함께 고려할 필요가 있다.

기존 연구에서는 3DGS의 장면 표현 방식과 렌더링 효율[2], SfM 기반 3차원 재구성 파이프라인[3], [4]이 주로 논의되어 왔다. 그러나 서로 다른 SfM 결과를 동일한 3DGS 학습 파이프라인에 입력했을 때의 품질 차이는 추가 검토가 필요하다. 본 논문에서는 library 시퀀스에 대해 COLMAP, GLOMAP, RealityScan 결과를 동일한 gsplat 학습 과정에 입력하여 렌더링 품질 변화를 비교한다.

3. 본론

본 실험은 MPEG GSC 실험 데이터 중 library 시퀀스를 대상으로 수행하였다. Library 시퀀스는 4,000개의 시점 영상으로 구성된 대규모 실외 장면 데이터이며, 원본 해상도는 3840×2160이다. 본 실험은 gsplat 학습 시 data_factor=2를 적용하여 1920×1080 해상도의 입력 영상을 사용하였다. 또한 건물, 수목, 도로, 잔디 영역 등 다양한 구조와 텍스처를 포함한다. 본 논문은 동일한 library 시퀀스에 대해 COLMAP, GLOMAP, RealityScan을 각각 적용하고, 각 SfM 결과의 카메라 포즈와 희소 포인트 클라우드를 동일한 gsplat 학습 파이프라인에 입력하여 최종 렌더링 품질을 비교하였다.

COLMAP은 일반적인 SfM 및 MVS 기반 3차원 재구성 파이프라인으로 사용하였고, GLOMAP은 글로벌 SfM 계열 비교군으로 사용하였다. RealityScan 결과는 COLMAP 텍스트 형식으로 내보내어 gsplat 입력으로 사용하였다. 세 기법 모두 동일한 COLMAP 형식의 SfM 결과를 사용하였으며, library 시퀀스의 이미지 수와 해상도를 고려하여 data_factor=2, 60,000회 반복 조건으로 학습하였다. 성능 평가는 희소 포인트 수, 최종 가우시안 수, PSNR, SSIM, LPIPS 및 학습 시간을 기준으로 수행하였다. 또한 SfM 결과의 카메라 포즈와 포인트 클라우드 시각화 결과, 동일 검증 시점의 3DGS 렌더링 결과를 함께 비교하였다.

표 1. SfM 기법별 gsplat 정량 평가 결과

Sequence	Method	No. of Points	PSNR (dB)	SSIM	LPIPS	Training Time (s)	No. of GS
Library	COLMAP	1,764K	23.23	0.7119	0.2503	3,013.80	2,408K
	Glomap	1,909K	24.96	0.7251	0.2446	3,043.80	2,686K
	RealityScan	1,498K	27.20	0.8443	0.1545	1,540.20	3,623K

표 1은 library 시퀀스에 대해 SfM 기법별 희소 포인트 수, PSNR, SSIM, LPIPS, 학습 시간 및 최종 가우시안 수를 정리한 결과이다. COLMAP은 PSNR 23.23 dB, SSIM 0.7119, LPIPS 0.2503를 보였고, GLOMAP은 PSNR 24.96 dB, SSIM 0.7251, LPIPS 0.2446로 COLMAP보다 개선된 결과를 보였다.

RealityScan은 PSNR 27.20 dB, SSIM 0.8443, LPIPS 0.1545로 가장 우수한 정량 품질을 나타냈다. 특히 RealityScan은 희소 포인트 수가 가장 많은 방법은 아니었음에도 가장 높은 품질을 보여, 최종 품질이 희소 포인트 수에만 비례하지 않음을 확인할 수 있다.

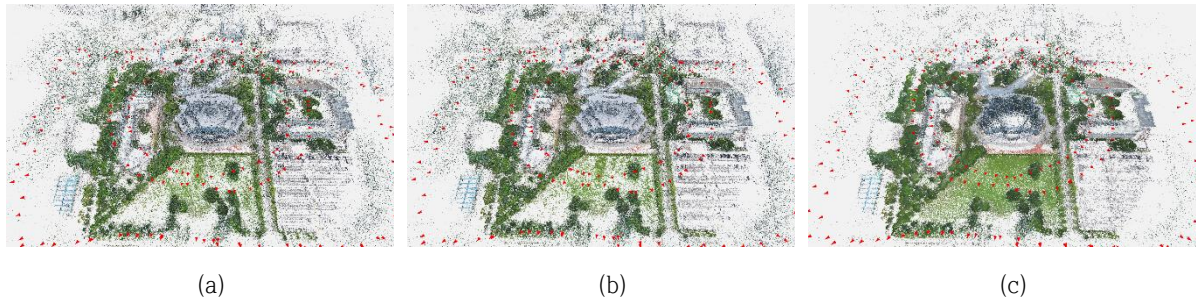


그림 1. 모델 별 카메라 포즈 및 희소 포인트 클라우드 시각화 결과. (a) COLMAP, (b) GLOMAP, (c) RealityScan.

그림 1은 SfM 기법별 카메라 포즈와 희소 포인트 클라우드를 시각화한 결과이다. 세 기법 모두 library 시퀀스의 전체적인 카메라 이동 경로와 장면 구조를 복원하였지만, 희소 포인트 클라우드의 분포에는 차이가 나타났다. COLMAP과 GLOMAP은 전체 장면을 넓게 복원하였으나 일부 영역에서

포인트 분포가 상대적으로 퍼져 있는 형태를 보였고, RealityScan은 주요 구조 주변에서 포인트 클라우드가 비교적 안정적으로 형성된 것으로 관찰되었다. 이는 표 1의 정량 평가 결과와도 일관된다.

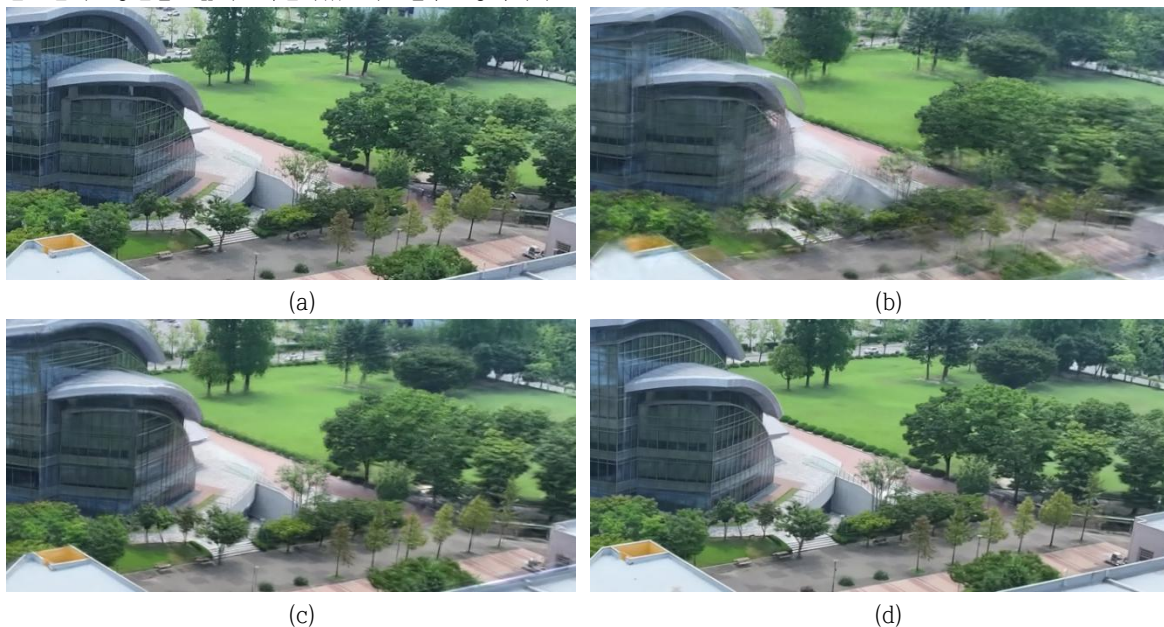


그림 2. 모델 별 렌더링 결과 비교. (a) 정답 이미지, (b) COLMAP 기반 3DGS, (c) GLOMAP 기반 3DGS, (d) RealityScan 기반 3DGS.

그림 2는 library 시퀀스의 동일 검증 시점에서 정답 이미지와 각 SfM 기법 기반 3DGS 렌더링 결과를 비교한 것이다. COLMAP과 GLOMAP 결과는 전체 장면 구조를 유지하였지만, 건물 외곽선과 수목 영역에서 흐림 현상과 세부 구조 손실이 나타났다. 특히 건물 경계와 창문 영역은 정답 이미지에 비해 선명도가 낮고, 일부 구조가 부드럽게 번지는 경향을 보였다. 반면 RealityScan 기반 결과는 건물 외곽선, 잔디와 수목의 경계, 전체적인 장면 구조가 상대적으로 선명하게 유지되어 정답 이미지와 가장 유사한 시각적 품질을 보였다.

정량 결과와 시각적 결과를 종합하면, library 시퀀스에서는 RealityScan 기반 SfM 결과가 gsplat 학습에 가장 유리한 초기 입력을 제공한 것으로 해석된다. 이는 동일한 학습 조건에서도 SfM 단계에서 생성된 카메라 포즈와 희소 포인트 클라우드의 차이가 후속 가우시안 최적화 결과에 영향을 줄 수 있음을 보여준다. 따라서 3DGS 기반 재구성에서 SfM 기법의 선택은 초기 희소 포인트 클라우드의 구성뿐만 아니라 최종 가우시안 최적화와 최종 렌더링 품질에도 영향을 줄 수 있음을 확인하였다.

4. 결론

본 논문은 MPEG GSC 실험 데이터 중 library 시퀀스를 대상으로 SfM 기법에 따른 3DGS 재구성 품질 변화를 비교하였다. 이를 위해 COLMAP, GLOMAP, RealityScan으로부터 생성된 카메라 포즈와 희소 포인트 클라우드를 동일한 gsplat 학습 파이프라인에 입력하고, PSNR, SSIM, LPIPS, 희소 포인트 수, 최종 가우시안 수 및 학습 시간을 기준으로 성능을 평가하였다.

실험 결과, SfM 기법에 따라 희소 포인트 클라우드의 구성과 최종 렌더링 품질에 차이가 나타났다. RealityScan 기반 결과는 가장 높은 PSNR과 SSIM, 가장 낮은 LPIPS를 보여 세 기법 중 가장 우수한 정량 품질을 보였다. 또한 희소 포인트 수가 가장 많은 결과가 반드시 가장 높은 품질로 이어지는 않았으며, 이는 카메라 포즈의 안정성 및 포인트 클라우드의 공간적 분포가 3DGS 재구성 품질에 중요한 영향을 줄 수 있음을 보여준다. 향후 연구는 다양한 장면 구조와 촬영 조건을 갖는 시퀀스로 실험을 확장하여 SfM 기법별 특성과 3DGS 재구성 품질 사이의 관계를 추가적으로 분석할 계획이다.

ACKNOWLEDGEMENT

본 연구는 2026년도 과학기술정보통신부 및 정보통신기획평가원의 “SW중심대학사업” 지원을 받아 수행되었음 (2023-0-00065).

참고 문헌

- [1] ISO/IEC JTC1 SC29/WG4, “ISO/IEC 23090-12:2023, Information technology — Coded representation of immersive media — Part 12: MPEG immersive video,” 2023.
- [2] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, “3D Gaussian Splatting for Real-Time Radiance Field Rendering,” *ACM Transactions on Graphics*, Vol. 42, No. 4, pp. 1-14, 2023.
- [3] J. L. Schönberger and J.-M. Frahm, “Structure-from-Motion Revisited,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4104-4113, 2016.
- [4] COLMAP, “COLMAP: Structure-from-Motion and Multi-View Stereo,” *Official Documentation*. [Online]. Available: <https://COLMAP.github.io/>
- [5] L. Pan, D. Baráth, M. Pollefeys, and J. L. Schönberger, “Global Structure-from-Motion Revisited,” *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 58-77, 2024.
- [6] R. Koo, Y.-G. Kim, I. Yang, S. H. Ahn, and E.-S. Ryu, “Library: A Large-Scale Outdoor Gaussian Splat Reconstruction Dataset,” *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops*, pp. 153-156, 2026.
- [7] V. Ye, R. Li, J. Kerr, M. Turkulainen, B. Yi, Z. Pan, O. Seiskari, J. Ye, J. Hu, M. Tancik, and A. Kanazawa, “gsplat: An Open-Source Library for Gaussian Splatting,” *arXiv preprint, arXiv:2409.06765*, 2024.