

3D Gaussian Splatting 을 위한 실내 공간 대상 키프레임 추출 최적화

이나연 최경호 최성식 최유성 정종범

국립군산대학교 컴퓨터정보공학과

{nayeon6207, cgh3578, tjdtlr3104, cyscys06, jongbeomjeong}@kunsan.ac.kr

Optimization of Key Frame Extraction for 3D Gaussian Splatting in Indoor Scenes

Nayeon Lee GyeongHo Choi SeungSik Choi YooSung Choi Jong-Beom Jeong

Department of Computer Science and Information Engineering,

Kunsan National University (KSNU)

요 약

본 논문은 실내 환경에서 촬영한 영상을 활용하여 프레임 샘플링 전략이 3D Gaussian splatting (3DGS) 재구성 품질에 미치는 영향을 비교, 분석하는 것을 목적으로 한다. 최근 3DGS는 빠른 학습 속도와 높은 렌더링 품질로 주목받고 있으나, 입력 프레임의 구성 방식에 따라 재구성 성능과 학습 효율이 크게 달라질 수 있다. 이에 동일한 영상 데이터에 대해 시간 균등 샘플링 (uniform sampling), I-frame 기반 샘플링, 그리고 local light field fusion (LLFF) 가이드라인을 참고한 적응형 샘플링 방식을 비교하였다. 실험 결과, 균등 샘플링 방식은 PSNR 25.69, SSIM 0.9408, LPIPS 0.106을 기록하였으며, I-frame 방식은 PSNR 25.87, SSIM 0.9373, LPIPS 0.120을 나타냈다. 반면 적응형 샘플링 방식은 PSNR 27.48, SSIM 0.9521, LPIPS 0.096으로 가장 우수한 성능을 보였으며, 학습 시간 또한 약 336.24초로 가장 짧게 나타났다. 이는 장면 변화량을 고려한 적응형 샘플링이 중복 프레임을 줄이면서도 유효한 시점을 효과적으로 선택하여 재구성 품질과 학습 효율을 동시에 향상시킬 수 있음을 보여준다. 따라서 프레임 샘플링 전략이 실내 영상 기반 3DGS 파이프라인에서 중요한 성능 결정 요소임을 확인하였다.

1. 서론

최근 3차원 장면 재구성과 새로운 시점 합성 분야에서는 neural radiance fields (NeRF) [1]와 함께 3DGS[2]이 주목받고 있다. 3DGS는 다수의 3차원 Gaussian primitive 기반 장면 표현을 통해 빠른 최적화와 고속 렌더링을 가능하게 하며, 실제 응용 환경에서도 높은 실용성을 보이는 방법으로 평가받고 있다. 그러나 3DGS 기반 재구성 성능은 학습 과정뿐만 아니라 입력 프레임의 구성 방식에도 큰 영향을 받는다. 특히 비디오 기반 입력에서는 프레임 선택 방식에 따라 Structure-from-motion (SfM) [3] 기반 포즈 추정 품질이 달라지며, 이는 최종 재구성 품질과 학습 효율에 직접적인 영향을 미친다. 이러한 문제는 반복 구조와

복합적인 깊이 변화가 존재하는 실내 환경에서 더욱 두드러진다. 이에 본 논문은 GoPro로 촬영한 실내 영상을 대상으로 프레임 샘플링 전략에 따른 3DGS 재구성 성능 변화를 비교하였다. 비교 방식은 균등 샘플링 방식, I-frame 방식, 그리고 적응형 샘플링 방식으로 구성하였다. 각 방식에 대해 SfM 기반 포즈 추정을 수행한 뒤, gsplat 기반 3DGS를 Markov chain Monte Carlo (MCMC) [6]모드로 학습하였으며, 정량 평가와 렌더링 결과를 비교, 분석하였다. 이를 통해 프레임 샘플링 전략이 3DGS 재구성 품질과 학습 효율에 미치는 영향을 분석하고, 실내 환경에 적합한 샘플링 방식의 효과를 확인하고자 한다.

2. 선행 연구

영상 기반 3차원 재구성은 다중 시점 기하와 SfM 기술의 발전과 함께 꾸준히 발전해 왔다. 대표적으로 COLMAP[3]은 특징점 추출, 이미지 매칭, 카메라 포즈 추정, 희소 재구성을 통합적으로 제공하는 도구로 널리 활용되고 있으며, 최근에는 NeRF[1] 및 3DGS[2]와 같은 신경 장면 표현 기반 재구성에서도 초기 카메라 포즈 추정을 위한 표준 파이프라인으로 사용되고 있다. NeRF는 연속적인 장면 표현을 통해 높은 품질의 시점 합성을 가능하게 했으나, 높은 계산 비용과 긴 학습 시간이 한계로 지적되었다. 이에 대한 대안으로 제안된 3DGS는 다수의 3차원 Gaussian primitive를 기반으로 장면을 표현하여 빠른 렌더링과 효율적인 최적화를 가능하게 한다. 한편, 시점 합성 품질은 장면 표현 방식뿐만 아니라 입력 시점의 밀도와 분포에도 큰 영향을 받는다. LLFF[4]는 제한된 시점 환경에서 촬영 시 충분한 시차 확보의 중요성을 강조하였으며, 이는 단순 시간 기반 프레임 추출보다 실제 장면 변화량을 고려한 프레임 선택이 중요할 수 있음을 보여준다. 비디오 기반 프레임 샘플링 방식은 일반적으로 균등 샘플링 방식, I-frame 방식, 그리고 장면 변화량이나 광류(optical flow)를 활용한 적응형 샘플링 방식으로 구분된다. 균등 샘플링 방식은 구현이 단순하지만 중복 프레임이 증가할 수 있으며, I-frame 방식은 압축 구조의 영향을 받기 때문에 장면 재구성에 최적화된 시점 분포를 항상 보장하지는 않는다. 반면 적응형 샘플링은 장면 변화량을 반영하여 시점 다양성을 확보할 가능성이 있다. 그러나 기존 연구들은 주로 합성 데이터셋이나 정적인 이미지 집합 기반의 재구성 성능 평가를 중심으로 이루어졌으며, 실제 실내 이동 촬영 환경에서 프레임 샘플링 전략이 3DGS 재구성 성능에 미치는 영향을 비교, 분석하는 연구의 필요성 또한 제기되고 있다. 따라서 본 논문은 GoPro로 촬영한 실내 영상을 대상으로 균등 샘플링 방식, I-frame 방식, 적응형 샘플링 방식을 동일한 파이프라인에서 비교하여 프레임 샘플링 전략이 3DGS 재구성 성능에 미치는 영향을 분석하였다.

3. 본론

원본 실내 영상을 입력으로 사용하고, 세 가지 프레임 샘플링 방식에 따라 이미지 집합을 생성한 후 COLMAP을 통해 카메라 포즈와 희소 포인트 클라우드를 추정하였다. 이후 해당 결과를 기반으로 gsplat의 MCMC[6] 모드를 이용하여 3DGS 학습을 수행하였다. MCMC 모드는

Gaussian의 생성과 제거를 확률적으로 수행함으로써 장면 구조를 보다 안정적으로 최적화할 수 있으며, 복잡한 실내 환경에서도 과도한 Gaussian 증가를 억제하면서 효율적인 재구성을 가능하게 한다. 마지막으로 peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM) [5], learned perceptual image patch similarity (LPIPS)를 이용한 정량 평가와 렌더링 결과 비교를 수행하였다.

본 논문은 세 가지 프레임 샘플링 전략을 비교하였다. 첫 번째 방식은 균등 샘플링 방식으로, 본 실험의 baseline으로 사용하였다. 원본 영상에서 1초마다 1장의 프레임을 추출하였으며, 약 10분 길이의 영상으로부터 총 605장의 이미지가 생성되었다. 이 방식은 구현이 단순하고 재현성이 높다는 장점이 있으나, 장면 변화량과 무관하게 동일한 시간 간격으로 프레임이 선택되므로 정적인 구간에서는 중복 시점이 증가할 수 있으며, 빠른 이동 구간에서는 모션 블러(motion blur)가 포함될 가능성이 존재한다. 두 번째 방식은 압축된 비디오 내 I-frame만을 선택하는 방식이다. 이 방법은 독립적으로 인코딩된 프레임만을 사용하므로 상대적으로 완전한 시각 정보를 포함할 가능성이 있다. 실험에서는 총 627장의 이미지가 생성되었으나, 이 방식은 장면 변화량보다 GOP 구조의 영향을 크게 받기 때문에, 장면 재구성에 적합한 시점 분포를 항상 보장하지는 않는다. 세 번째 방식은 적응형 샘플링 방식이다. 해당 방식에서는 우선 더 조밀한 시점 확보를 위해 원본 영상으로부터 1차 프레임 집합을 생성한 뒤, ffmpeg의 mpdecimate 필터를 이용하여 중복도가 높은 프레임을 제거하였다. mpdecimate는 인접 프레임 간 픽셀 변화량을 기준으로 유사 프레임을 제거하여 중복 시점을 감소시키는 방식으로 총 1209장의 이미지가 생성되었다. 각 샘플링 방식에 대해 COLMAP의 OPENCV 카메라 모델을 사용하여 특징점 추출, 이미지 매칭 및 희소 재구성을 수행하였다. 이후 동일한 조건에서 gsplat 기반 3DGS 학습을 수행하였으며, 모든 실험에서 Gaussian 개수는 1,000,000으로 동일하게 유지하였다. 이는 샘플링 전략에 따른 재구성 성능 차이를 보다 명확하게 비교하기 위한 설정이다.

표 1. 프레임 샘플링 방식별 정량 평가 결과(MCMC 기준)

Method	No. of Images	PSNR (dB)	SSIM	LPIPS	Training Time (s)	Number of GS
Uniform 방식	605	25.69	0.9408	0.106	854.49s	1,000,000
I-frame	627	25.87	0.9373	0.120	862.05s	1,000,000
적응형 샘플링 방식	1209	27.48	0.9521	0.096	336.24s	1,000,000

표 1은 세 가지 프레임 샘플링 방식에 대한 정량 평가 결과를 정리한 것이다. 적응형 샘플링 방식은 PSNR 27.48, SSIM 0.9521, LPIPS 0.096으로 가장 우수한 성능을 보였으며, 학습 시간 또한 약 336.24초로 가장 짧게 나타났다. 균등 샘플링 방식은 안정적인 baseline 성능을 보였으나 일부 정적 구간에서 중복 시점이 포함되는 한계가 존재하였다. I-frame 방식은 균등 샘플링 방식과 유사한 수준을 유지했지만, SSIM과 LPIPS에서는 상대적으로 낮은 성능을 나타냈다. 이는 I-frame 방식이

장면 변화량보다 GOP 구조의 영향을 크게 받기 때문으로 해석할 수 있다. 특히 적응형 샘플링 방식은 가장 높은 정량 성능과 가장 짧은 학습 시간을 동시에 보였다. 이는 중복 프레임 감소와 시점 다양성 확보가 SfM 및 Gaussian 최적화 과정에 유리하게 작용했기 때문으로 판단된다.



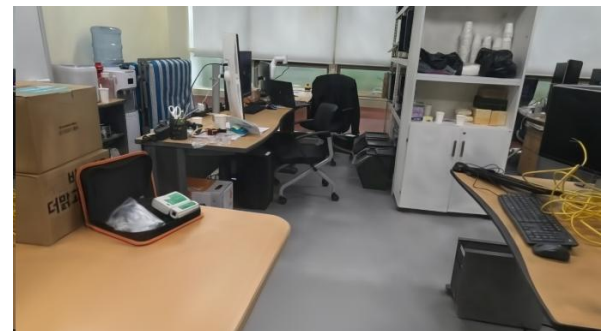
(a) Ground Truth



(b) 균등 샘플링



(c) I-frame



(d) 적응형 샘플링

그림 1. 프레임 샘플링 방식별 3DGS 렌더링 결과 비교

그림 1 은 각 프레임 샘플링 방식의 대표 렌더링 결과를 비교한 것이다. 균등 샘플링 방식은 전반적인 장면 구조를 안정적으로 재현하였으나, 상자 모서리와 책상 경계 영역에서 다소 흐릿한 표현이 관찰되었다. 특히 상자 전면의 인쇄 문자와 모서리 부근에서는 경계가 번지는 경향이 나타났다. I-frame 방식은 균등 샘플링 방식 대비 상자 윤곽은 비교적 유지되었으나, 상자 상단 모서리와 측면 경계 영역에서 세부 표현이 균일하지 못하고 일부 영역에서는 경계가 흐려지는 현상이 관찰되었다. 반면 적응형 샘플링 방식은 상자 전면과 측면의 직선 경계가 가장 선명하게 유지되었으며, 상자 상단 모서리와 책상 가장자리의 구분 또한 가장 명확하게 나타났다. 특히 인접한 물체 간 경계 표현과 구조적 일관성이 상대적으로 우수하게 유지되었으며, 이러한 경향은 정량 평가 결과와도 유사하게 나타났다. 또한 실험 과정에서는 COLMAP 의 OPENCV 카메라 모델을 사용할 경우 재구성 품질이 보다 안정적으로 유지되는 경향을 확인하였다. 이는 카메라 모델 설정 역시 재구성 품질에 영향을 줄 수 있음을 의미한다.

4. 결론

본 논문은 GoPro로 촬영한 실내 연구실 영상을 대상으로 프레임 샘플링 전략에 따른 COLMAP 기반 3DGS 재구성 성능을 비교하였다. 이를 위해 균등 샘플링 방식, I-frame 방식, 그리고 적응형 샘플링 방식을 구성하고, OPENCV 카메라 모델 기반 COLMAP 정합 후 gsplat의 MCMC 모드를 이용하여 학습을 수행하였다.

실험 결과, 적응형 샘플링 방식은 세 가지 방법 중 가장 우수한 재구성 성능을 보였으며, 학습 효율 측면에서도 가장 안정적인 결과를 나타냈다. 균등 샘플링 방식은 안정적인 baseline 성능을 제공하였으며, I-frame 방식은 비교적 높은 정합률을 보였으나 전체적인 재구성 품질 측면에서는 한계를 나타냈다. 이러한 결과는 실내 영상 기반 3DGS에서 프레임 샘플링 전략이 단순한 전처리 과정이 아니라 재구성 품질과 학습 효율에 직접적인 영향을 미치는 중요한 변수임을 확인하였다. 또한 동일한 학습 조건에서 샘플링 방식만 변경하여 입력 데이터 구성의 중요성을 비교하였다. 특히 적응형 샘플링 방식은 중복 프레임을 감소시키면서도 시점 다양성을 효과적으로 유지함으로써, SfM 기반 포즈 추정과 Gaussian 최적화 과정 모두에 유리하게 작용하는 경향을 보였다.

향후 연구에서는 광류 (optical flow), blur score, feature overlap과 같은 지표를 활용하여 장면 변화량을 보다 정량적으로 반영하는 적응형 샘플링 전략으로 확장할 필요가 있다. 또한 다양한 실내 및 실외 환경에 대한 추가 실험을 통해 프레임

샘플링 전략의 일반화 가능성을 검증하고자 한다.

ACKNOWLEDGEMENT

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

참고 문헌

- [1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis,” *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 405–421, 2020.
- [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 (CVPR)*, pp. 4104–4113, 2016.
- [4] B. Mildenhall, P. P. Srinivasan, R. Ortiz-Cayon, N. Kalantari, and R. Ramamoorthi, “Local Light Field Fusion for Practical View Synthesis,” *ACM Transactions on Graphics*, Vol. 38, No. 4, pp. 1–14, 2019.
- [5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp. 600–612, 2004.
- [6] B. Kheradmand, B. Kerbl, G. Kopanas, and G. Drettakis, “3D Gaussian Splatting as Markov Chain Monte Carlo,” *arXiv preprint arXiv:2404.09591*, 2024.

