

6자유도 로봇암을 이용한 3D 객체 자동 취득 및 Gaussian Splat 생성

최유성, 최경호, 최성식, 이나연, 정종범

국립군산대학교 컴퓨터정보공학과

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

Automated 3D Object Acquisition and Gaussian Splat Generation Using Six Degrees of Freedom Robot Arm

YooSung Choi GyeongHo Choi Seung-Sik Choi Nayeon Lee Jong-Beom Jeong

Department of Computer Science and Information Engineering,

Kunsan National University (KSNU)

요 약

본 논문은 6자유도 (6 degrees of freedom, 6DoF) 로봇암을 이용하여 소형 객체의 다중 시점 영상을 자동으로 취득하고, 이를 기반으로 3D Gaussian splatting (3DGS) 모델을 생성하는 시스템을 제안한다. 객체 중심을 기준으로 객체의 다각도 시점 확보가 가능한 촬영 경로를 자동으로 구성하고, 연속된 시점 간 충분한 중첩을 확보하도록 객체 면을 고려한 촬영 궤적 생성 방식 (face-aware orbit) 을 적용하였다. 또한 표면의 무늬나 특징이 적은 저 텍스처 객체에서 발생하는 재구성 실패 문제를 해결하기 위해 최적화된 COLMAP 설정을 적용하였다. 실험 결과, 최적화된 COLMAP 설정을 기반으로 전체 입력 영상에 대한 안정적인 카메라 위치 추정과 3차원 복원이 가능하였다. 이후 3DGS 학습을 통해 객체의 형상을 효과적으로 복원할 수 있었으며, 이를 통해 로봇암 기반 자동 촬영 방식이 안정적인 3차원 데이터셋 구축에 효과적임을 확인하였다.

1. 서론

현실 객체를 고품질의 3차원 디지털 자산으로 변환하는 기술은 로보틱스, 디지털 트윈, 확장현실, 품질 검사 등 다양한 분야에서 중요성이 증가하고 있다. 특히 사용자가 임의의 시점에서 객체를 자유롭게 관찰할 수 있는 6DoF 기반 콘텐츠는 가상현실과 시뮬레이션 분야에서 높은 활용 가치를 가진다. 기존의 영상 기반 3차원 복원 방식은 사용자가 직접 카메라를 들고 촬영하는 방식에 의존한다. 그러나 이러한 방식은 촬영 거리, 시점 분포, 이동 속도가 일정하지 않아 반복성과 재현성이 낮다. 특히 일부 면이 충분히 관찰되지 않거나 연속된 이미지 간 특징 점 중첩이 부족한 경우 structure-from-motion (SfM) 단계에서 카메라 정합이 실패하고 이후 복원 품질이 크게 저하된다[2]. 3DGS는 빠른 학습 속도와 실시간 렌더링이 가능하여 높은 관심을 받고 있다[1]. 그러나 3DGS 역시 초기 카메라 자세 추정과 입력 이미지 품질에 크게 의존하며, 저텍스처 객체에서는 특징점 부족으로 인해 COLMAP 기반 재구성이 자주 실패한다. 따라서 안정적인 3DGS 복원을 위해서는 균일한 시점 분포와 충분한 특징점 중첩을 확보할 수 있는 자동 촬영 시스템이 필요하다. 본 연구에서는 6DoF 로봇암 기반

자동 객체 취득 시스템을 구현하고, face-aware orbit 촬영과 최적화된 COLMAP 설정을 적용하였다.

본 논문의 구성은 다음과 같다. 2 절에서는 선행 연구를 설명하고, 3 절에서는 전체 파이프라인의 구현 방법을 제시한다. 4 절에서는 실험 결과 및 고찰을 다루며, 마지막으로 5 절에서 결론을 맺는다.

2. 선행 연구

SfM은 여러 시점 이미지에서 특징점을 추출하고 이미지 간 대응 관계를 계산하여 카메라 자세와 3차원 공간상에서 물체의 형태를 표현한 수많은 점들의 집합인 포인트 클라우드를 추정하는 기법이다[2]. 본 연구에서는 COLMAP을 이용하여 키프레임의 카메라 자세와 희소 포인트 클라우드를 생성하였다. 그러나 SfM 기반 복원은 입력 이미지의 특징점 수와 시점 중첩에 크게 의존하므로, 표면 무늬가 적은 객체에서는 특징점 부족으로 카메라 정합이 실패할 수 있다.

NeRF는 다중 시점 영상과 카메라 자세를 이용하여 장면을 연속적인 3차원 공간으로 표현하는 대표적인 신경망 기반 복원 기법이다[3]. 새로운 시점의 영상을 생성할 수 있으며 높은 복원 품질을 제공하지만, 반복적인 신경망 연산과 볼륨 렌더링으로 인해 학습 속도가 느리고 계산 비용이 크다는 한계를 가진다. 이를 개선하기 위해 제안된 3DGS는 장면을 Gaussian splat으로 표현하고 이를 직접 최적화하는 방식이다[1]. GPU 병렬 연산을 활용하여 빠른 학습과 실시간 렌더링이 가능하지만, 초기 카메라 자세와 포인트 클라우드 품질에 크게 의존한다. Local light field fusion (LLFF)는 촬영 밀도보다 시점 분포와 중첩의 중요성을 강조한 연구이다[4]. 이는 3DGS 기반 객체 복원에서도 동일하게 적용되며, 특히 작은 객체에서는 균형 있는 시점 확보가 복원 품질에 직접적인 영향을 준다.

기존 연구는 대부분 사용자가 직접 카메라를 들고 촬영하거나 이미 준비된 정적인 이미지 집합을 기반으로 수행되었으며, 실

제 로봇암 기반 자동 촬영 환경에서 발생하는 작업공간 제한, 역기구학 실패, 관절의 급격한 움직임과 같은 물리적 제약을 고려한 연구 역시 필요하다[5]. 또한 저텍스처 소형 객체에 대해 안정적인 COLMAP 초기화를 위한 촬영 경로 설계와 조건 완화 전략 역시 제한적이다. 따라서 반복 가능하고 안정적인 3DGS 데이터셋 구축을 위해서는 로봇암 기반 자동 촬영과 객체 중심의 시점 설계가 함께 고려되어야 한다. 본 연구에서는 이러한 한계를 해결하기 위해 face-aware orbit 기반 자동 촬영 및 최적화된 COLMAP 설정 기반의 객체 취득 시스템을 제안한다. 최적화된 COLMAP 설정은 이미지당 검출 가능한 특징 점 수를 증가시키고, 특징 점 검출 기준을 낮추며, 연속 프레임 간 중첩을 활용하도록 구성하였다. 또한 초기 카메라 자세 계산과 삼각측량 조건을 완화하여 적은 대응점으로도 정합이 시작될 수 있도록 하였다.

3. 본론

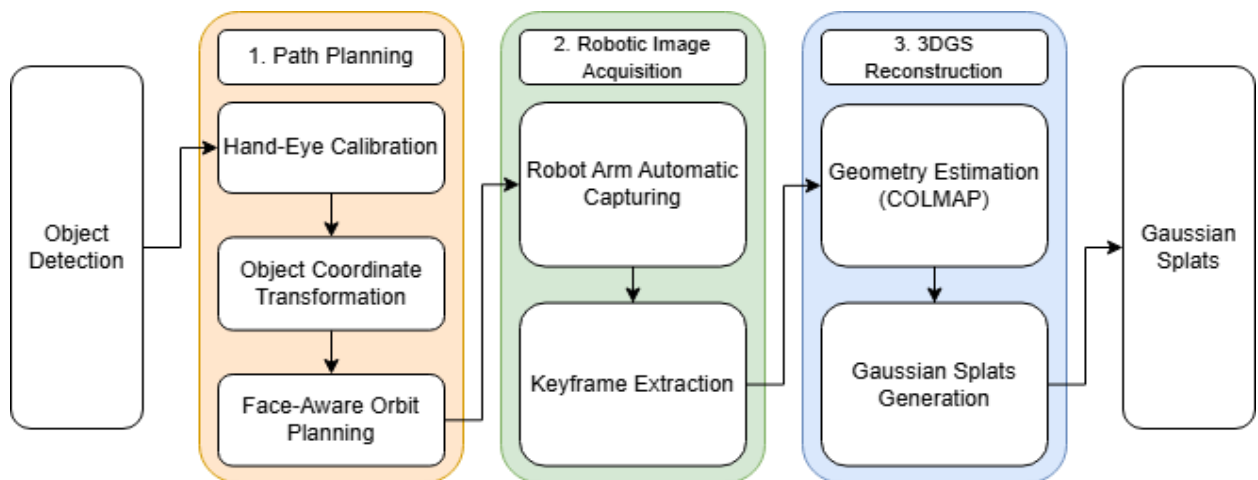


그림 1. 전체 파이프라인 구성 다이어그램

그림 1은 제안하는 전체 파이프라인 구성을 도시한다. 전체 파이프라인은 객체 감지, 카메라와 로봇 좌표계 간 위치 보정, 자동 촬영 경로 생성, 영상 취득, 키프레임 이미지 추출, COLMAP 재구성, 3DGS 학습 단계로 구성된다.

먼저 색상 및 깊이 정보 (RGB-D) 카메라를 이용하여 촬영 대상 객체를 감지하고, 객체 중심 좌표를 계산한다. 이후 카메라와 로봇 좌표계 간 위치 보정을 통해 카메라 좌표계를 로봇 베이스 좌표계로 변환하여 객체의 실제 위치를 추정한다. 변환된 객체 중심을 기준으로 로봇암의 촬영 경로를 생성하고, 다양한 시점에서 객체 영상을 자동으로 취득한다. 촬영된 영상에서는 중복되거나 흐린 프레임을 제거하여 학습에 적합한 이미지만 추출한다. 이후 COLMAP을 이용하여 카메라 자세와 포인트 클라우드를 생성하고, 이를 기반으로 3DGS 학습을 수행하여 최종 3차원 객체 모델을 생성한다.

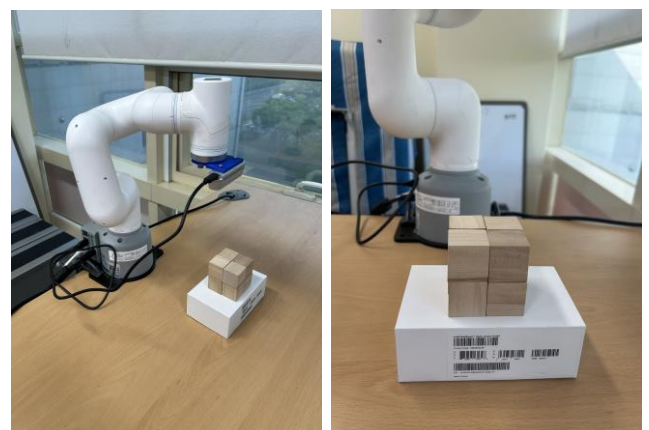


그림 2. 실험에 사용한 장비 및 촬영 대상 사진

그림 2 는 실험에 사용한 로봇암 장비와 촬영 대상 객체를 나타낸다. 실험에는 myCobot 320 Pi 로봇암과 RealSense D435i 카메라를 사용하였으며, COLMAP 과 INRIA 3DGS 구현[1]을 이용해 객체 복원을 수행하였다. 객체 촬영은 플랫폼 위의 6×6×6 cm 크기 정육면체 목재 블록을 대상으로 수행하였다.

객체 감지는 깊이 정보와 색상 정보를 결합하여 수행하였다. 깊이 범위를 제한하고 HSV 색 공간 기반 필터링을 적용하여 배경과 반사 영역을 제거한 뒤, 윤곽선을 이용해 객체 중심을 추정하였다. 선택된 객체의 중심 좌표는 카메라 좌표계 기준의 3 차원 위치로 계산되며, 이를 로봇암이 사용할 수 있도록 로봇 기준 좌표계로 변환해야 한다. 이를 위해 카메라와 로봇 좌표계 간 위치 보정을 수행하였다. 카메라와 로봇 좌표 보정을 위한 기준 패턴인 ChArUco 보드[6]를 기준 좌표로 사용하고, 로봇암이 여러 자세에서 보드를 관찰하도록 하여 보정 데이터를 수집하였다. 다양한 보정 방법을 비교한 결과 Tsai-lenz 보정 방법[7]이 가장 안정적인 결과를 보였으며, 평균 오차 약 5 mm 수준의 정밀도를 확보하였다.

초기 원형 촬영 방식에서는 일부 면이 충분히 관찰되지 않아 재구성이 실패하였다. 이를 해결하기 위해 객체의 각 면을 고려한 face-aware orbit 방식을 적용하였고, 작업공간 제한과 관절 뒤틀림 등의 물리적인 문제를 고려하여 안전한 웨이포인트만 최종 촬영

경로로 사용하였다. 웨이포인트는 로봇암이 안전하게 도달 가능한 촬영 지점을 의미한다. 촬영된 영상에서는 모든 프레임을 그대로 사용하는 대신 학습에 적합한 프레임만 선택해 사용하였다. 흐림이 심한 프레임과 중복도가 높은 프레임은 제거하고, 객체의 다양한 시점이 균형 있게 포함되도록 총 200 장의 이미지를 추출하였다.

기본 COLMAP 설정을 적용한 경우에는 200 장의 입력 이미지 중 3 장만 등록되었으며, 생성된 포인트 클라우드는 67 개에 불과하였다. 이는 표면의 무늬나 특징이 적은 소형 객체의 경우 특징점이 부족하여 카메라 정합이 어려웠기 때문이다. 이를 개선하기 위해 최적화된 COLMAP 설정을 적용하였다. 최적화된 COLMAP 설정에서는 이미지당 특징 점 수를 증가시키고 특징 점 검출 기준을 낮추었다. 또한 연속된 인접 프레임의 중첩 정보를 우선적으로 활용하고, 초기 카메라 자세 계산과 삼각측량 조건을 완화하여 적은 대응점으로도 안정적인 정합이 가능하도록 하였다. 이후 COLMAP 에서 계산된 카메라 자세와 포인트 클라우드를 입력으로 사용하여 3DGS 학습을 수행하였다. 학습은 구면 조화 함수 차수 3, 총 30,000 회 반복 조건으로 수행하였다. 초기 단계에서는 Gaussian splats 의 위치와 색상 정보를 빠르게 최적화하고, 이후 밀집화와 가지치기 과정을 통해 세부 구조를 보완하였다.

4. 실험 결과

표 1. 원본 COLMAP 과 proposed COLMAP 의 비교

Options	Input Images	Registered Images	Sparse Points	Observations	Reprojection Error (px)
Baseline	200	3	67	0	0.4609
Proposed	200	200	3,575	30,207	1.7062

표 2. proposed COLMAP 기반 3DGS 학습 결과

Iterations	L1 Loss	PSNR (dB)	SSIM	LPIPS	No. of GS
7000	0.0132	33.91	0.9689	0.2288	85,587
30000	0.0085	37.40	0.9746	0.2138	109,339

표 1 은 기본 COLMAP 과 최적화된 COLMAP 설정의 재구성 결과를 비교한다. 최적화된 COLMAP 설정을 적용한 결과, 200 장의 입력 이미지 전체가 하나의 재구성으로 안정적으로 등록되었으며 포인트 클라우드 3,575 개와 총 관측 수 30,207 개를 확보하였다. 반면 기본 설정에서는 3 장만 등록되고 포인트 클라우드도 67 개에 불과하여 정상적인 3 차원 복원이 어려웠다. 이는 표면의 무늬나 특징이 적은 소형 객체에서 초기

카메라 정합이 실패하기 때문이며, 특징 점 수 증가와 연속 프레임 간 중첩 활용이 전체 등록 성공에 중요한 역할을 하였다. 표 2 는 proposed COLMAP 기반 3DGS 학습 결과를 나타낸다. 반복 학습이 진행될수록 PSNR 과 SSIM 은 향상되고 L1 Loss 와 LPIPS 는 감소하여 복원 품질이 점진적으로 개선됨을 확인하였다.

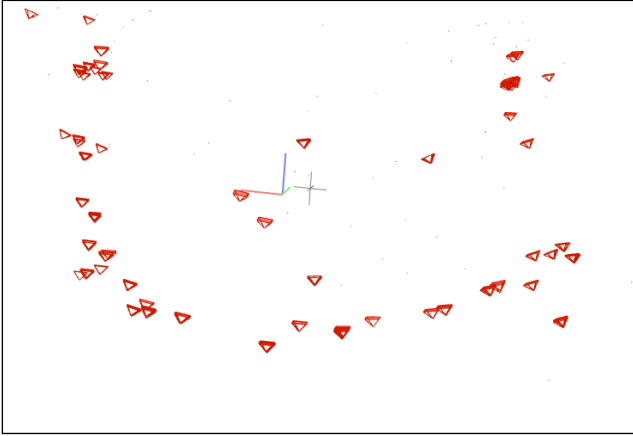


그림 3. 제한한 COLMAP 촬영 경로 시각화

그림 3 은 최적화된 COLMAP 설정을 적용한 촬영 경로와 카메라 위치를 시각화한 결과를 나타낸다. 각 촬영 시점의 카메라 위치와 방향은 객체 중심을 기준으로 원형에 가까운 분포를 형성하고 있다. 이를 통해 객체의 각 면이 고르게 촬영되었고, 연속된 시점 간 충분한 중첩을 확보하여 안정적인 카메라 자세 추정과 포인트 클라우드 생성이 가능함을 확인하였다.



그림 4. 3DGS 학습 결과 시각화

그림 4 는 3DGS 학습으로 생성된 Gaussian splats 를 시각화한 결과를 나타낸다. 복원 결과 객체의 상단부와 정면은 비교적 선명하게 나타났으나, 측면의 모서리와 경계는 다소 흐릿하게 복원되었다. 일부 배경 영역도 함께 포함되어 촬영 시점 분포와 전경 분리의 중요성을 확인할 수 있었다. 또한 로봇암 기반 자동 촬영은 동일한 경로를 반복적으로 재현할 수 있어 수동 촬영보다 높은 반복성과 안정성을 제공하였다.

5. 결론

본 논문은 6 자유도 로봇암과 색상 및 깊이 정보를 제공하는 카메라를 이용한 자동 객체 취득 기반 3차원 Gaussian splats 생성 시스템을 제한한다. 객체의 주요 면이 균형 있게 관찰될 수 있도록 촬영 궤적을 설계하고, 표면의 무늬나 특징이 적은 소형 객체에서도 안정적인 카메라 정합이 가능하도록 최적화된 COLMAP 설정을 적용하였다. 이를 통해 기존 기본 설정에서 어려웠던 안정적인 3차원 복원 및 Gaussian splats 생성을 확인하였다. 또한

로봇암 기반 자동 촬영은 높은 반복성과 안정성을 제공하였다.

향후 연구에서는 분할 마스크 기반의 전경 추출을 적용하여 배경 영역의 영향을 최소화하고, 회전판 기반 객체 회전 방식을 결합하여 더욱 다양한 시점을 안정적으로 확보할 계획이다. 또한 깊이 정보를 초기 카메라 자세 계산에 직접 활용하여 초기 정합의 안정성을 향상시키고, 보다 정밀한 3차원 객체 복원이 가능하도록 연구를 진행할 계획이다.

ACKNOWLEDGEMENT

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

참고 문헌

- [1] B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis, 3D Gaussian Splatting for Real-Time Radiance Field Rendering, ACM Transactions on Graphics, Vol. 42, No. 4, pp. 1-14, 2023.
- [2] J. L. Schönberger, J.-M. Frahm, Structure-from-Motion Revisited, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4104-4113, 2016.
- [3] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, R. Ng, NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, Communications of the ACM, Vol. 65, No. 1, pp. 99-106, 2022.
- [4] B. Mildenhall, P. P. Srinivasan, R. Ortiz-Cayon, N. Kalantari, R. Ramamoorthi, R. Ng, A. Kar, Local Light Field Fusion: Practical View Synthesis with Prescriptive Sampling Guidelines, ACM Transactions on Graphics, Vol. 38, No. 4, pp. 1-14, 2019.
- [5] I. D. Lee, J. H. Seo, and B. Yoo, Autonomous View Planning for 3D Scanning: A Survey, Automation in Construction, Vol. 160, p. 105291, 2024.
- [6] G. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, M. J. Marín-Jiménez, Automatic Generation and Detection of Highly Reliable Fiducial Markers under Occlusion, Pattern Recognition, Vol. 47, No. 6, pp. 2280-2292, 2014.
- [7] R. Y. Tsai, R. K. Lenz, A New Technique for Fully Autonomous and Efficient 3D Robotics Hand/Eye Calibration, IEEE Transactions on Robotics and Automation, Vol. 5, No. 3, pp. 345-358, 1989.
- [8] Elephant Robotics, myCobot 320 Pi Product Specification, Elephant Robotics Official Documentation, 2024.
- [9] Intel Corporation, Intel RealSense Depth Camera D435i Product Specifications, Intel Official Documentation, 2024.