한국컴퓨터그래픽스학회
Korea Computer Graphics Society Vol. 31, No. 3, P. 67~77

가우시안 스플래팅을 활용한 실내 공간 복원의 실용적 접근

배종환° 박상훈*

서강대학교 메타버스전문대학원

{bjh0309, mshpark}@sogang.ac.kr

A Practical Approach to Indoor Space Reconstruction Using Gaussian Splatting

Jonghwan Bae° Sanghun Park*

Graduate School of Metaverse, Sogang University

요 약

본 연구는 3차원 가우시안 스플래팅(3D Gaussian Splatting, 3DGS) 기술을 기반으로 대규모 실내 공간의 3D 복원 효율성과 정확성을 개선하는 통합적인 방법을 제안한다. 이 프로세스는 데이터 전처리, 객체 제거, 그리고 포인트 클라우드 병합의 세 단계로 구성된다. 첫째, HyperIQA와 ResNet50 모델을 활용하여 데이터셋에서 품질이 낮거나 중복된 이미지를 효과적으로 제거함으로써 3DGS에 최적화된 데이터셋을 구축한다. 둘째, Grounded-SAM2와 ProPainter를 결합하여 장면에 포함된 불필요한 객체를 탐지 및 제거하고, 인페인팅을 통해 시각적 일관성을 확보한다. 마지막으로, 여러 부분으로 나누어 복원된 공간(PLY 파일)을 병합하기 위해 CloudCompare로 초기 정렬을 수행하고, GaussReg 알고리즘으로 정밀하게 변환 행렬을 적용하여 최종 병합을 완료한다. 이와 같은 2단계 정합 방식은 형태가 복잡하고 특징점이 적은 공간에서도 구조적 일관성을 유지한다. 본 연구는 이미지 전처리부터 객체 제거, 정밀 정합까지의 과정을 통합하여 3DGS 기술의 실질적인 활용 가능성을 높였으며, 디지털 트윈 및 메타버스 분야에서 효율적인 실내 공간 데이터 구축의 기반을 마련했다는 점에서 의의를 갖는다.

Abstract

This study proposes an integrated method to improve the efficiency and accuracy of large-scale indoor space 3D reconstruction based on 3D Gaussian Splatting (3DGS) technology. The process consists of three stages: data preprocessing, object removal, and point cloud merging. First, HyperIQA and ResNet50 models are used to effectively remove low-quality or duplicate images from the dataset, thereby constructing a dataset optimized for 3DGS. Second, Grounded-SAM2 and ProPainter are combined to detect and remove unnecessary objects present in the scene, ensuring visual consistency through inpainting. Finally, to merge the spaces (PLY files) reconstructed in multiple parts, initial alignment is performed using CloudCompare, and the final merging is completed by applying precise transformation matrices with the GaussReg algorithm. This two-step registration approach maintains structural consistency even in spaces with complex shapes and few feature points. This study enhances the practical applicability of 3DGS technology by integrating processes from image preprocessing to object removal and precise registration, and is significant in that it lays the foundation for efficient indoor space data construction in the fields of digital twins and the metaverse.

키워드: 가우시안 스플래팅, 3D 복원, 이미지 필터링, 객체 분할, 인페인팅, 포인트 클라우드 정합

Keywords: Gaussian Splatting, 3D Reconstruction, Image Filtering, Object Segmentation, Inpainting, Point Cloud Registration

*corresponding author: Sanghun Park / Sogang University (mshpark@sogang.ac.kr)

Received : 2025.06.13./ Review completed : 1st 2025.06.30. 2nd 2025.07.10./ Accepted : 2025.07.14.
DOI : 10.15701/kcgs.2025.31.3.67
ISSN : 1975-7883(Print)/2383-529X(Online)

1. 서론

대규모 공간의 3D 복원 기술은 메타버스, 디지털 트윈 등 첨단 산업의 핵심 요소로 자리 잡고 있다. 기존의 3D 모델링은 노동집약적이며 많은 시간과 비용이 소모되는 한계가 있다. 이러한 문제를 해결하기 위해 Nerf [1]와 같은 기술이 등장했지만, 훈련 속도가 느리고 대규모 데이터 처리에 어려움이 있다. 최근 이를 대체하기 위해 3D 가우시안 스플래팅(3DGS) [2]이 빠른 처리 속도와 효율성을 바탕으로 주목받고 있으나 3DGS를 복잡한 대규모 실내 공간에 적용하기 위해서는 몇가지 해결 과제가 존재한다.

- 데이터 과부하: 대규모 공간 촬영 시 이미지 수가 기하급수적으로 증가하며, 불필요한 이미지 프레임과 중복된 이미지는 복원 품질 저하와 처리 비효율을 유발

- 불필요한 객체: 촬영 시 포함되는 사람, 이동형 가구 등 동적 객체는 복원 결과의 일관성을 해치는 노이즈로 작용

- 공간 병합의 어려움: 여러 구역으로 나누어 촬영된 결과물을 하나의 연속된 공간으로 병합하는 과정에서 정합 오류가 발생하거나, 일부 영역이 누락되는 현상이 발생

본 연구는 이러한 문제들을 해결하기 위해 이미지 필터링, 객체 제거, 공간 병합으로 이어지는 체계적인 3단계 파이프라인을 제안한다. 이를 통해 실내 공간 복원의 효율성과 정확성을 극대화하고, 3DGS의 실용적 활용 방안을 제시하고자 한다 [3].

2. 관련 연구

이미지 사전 처리는 3DGS 입력 데이터 품질을 향상시키는 중요한 단계로, HyperIQA와 같은 딥러닝 방식을 사용하여 이미지 품질 평가(IQA)를 수행한다 [4][5][6]. 또한 ResNet50 [7]과 같은 모델을 통해 이미지 특징 벡터를 추출하여 중복 이미지를 효과적으로 제거할 수 있으며, 이러한 기법은 3DGS의 입력 데이터셋을 최적화하여 복원 품질을 높이는 데 기여한다.

객체 제거 및 인페인팅에는 3D 장면 내 불필요한 객체 제거 [8]를 위해 텍스트 프롬프트 기반 객체 탐지가 가능한 Grounded-SAM2 [9]와 동영상 프레임 간의 연속성을 고려하여 자연스러운 복원이 가능한 ProPainter와 같은 인페인팅 모델이 활용된다 [10]. 기존 연구들은 주로 3DGS 복원 이후에 포인트 클라우드를 직접 편집하는 방식에 집중했지만, 본 연구는 입력 이미지 단계에서 객체를 제거하는 전처리 방식의 효율성을 분석한다.

전통적인 ICP 알고리즘은 널리 사용되나 3DGS 데이터의 고유 속성(밀도, 색상 등)을 완벽하게 보존하기 어렵다. 이를 해결하기 위해 3DGS 데이터 정합에 특화된 GaussReg [11]가 제안되었지만, 특정 파일 구조를 요구하고 겹치는 영역이 적을 경우 실패하는 한계가 있다.

3. 제안 방법

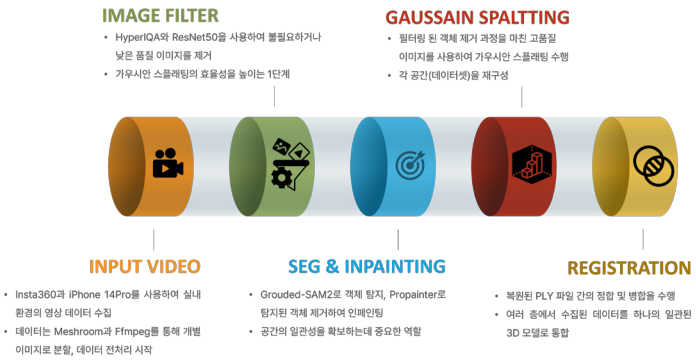


Figure 1: Overall research pipeline

본 연구는 대규모 실내 공간의 효율적인 3D 복원을 위해 그림 1과 같은 통합 파이프라인을 제안하며 각 단계는 다음과 같다.

- HyperIQA 및 ResNet50을 활용한 이미지 필터링

- Grounded-SAM2 및 ProPainter를 활용한 객체 제거와 인페인팅

- CloudCompare 및 GaussReg를 활용한 3DGS 정합

3.1 고품질 이미지 필터링

고품질 이미지 필터링은 2단계로 구성된다. 우선 No-Reference IQA 모델인 HyperIQA를 사용하여 이미지의 품질 점수를 예측한다. 사전 학습된 모델을 통해 각 이미지의 점수를 계산하고, 설정된 임계값(threshold)이하의 저품질 이미지를 제거한다. 이후 ResNet50 모델을 활용하여 이미지 특징 벡터를 추출하고, 유사도 기반으로 중복 이미지를 제거한다. 이 과정에서 이미지 간의 유사도를 측정하기 위해 코사인 유사도(cosine similarity)를 사용하며, 설정된 임계값 이상인 경우 중복으로 판단하여 제거한다.

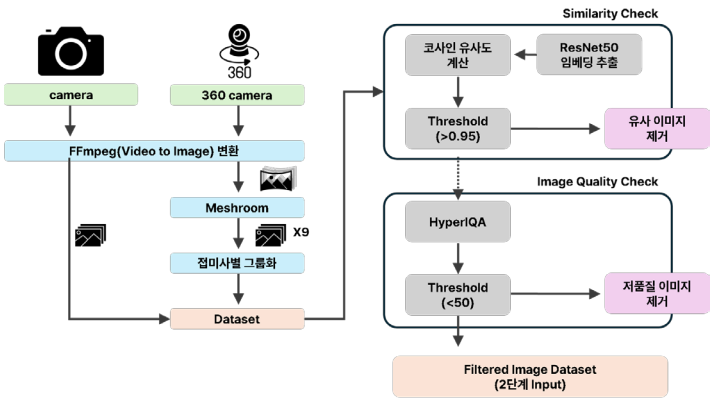
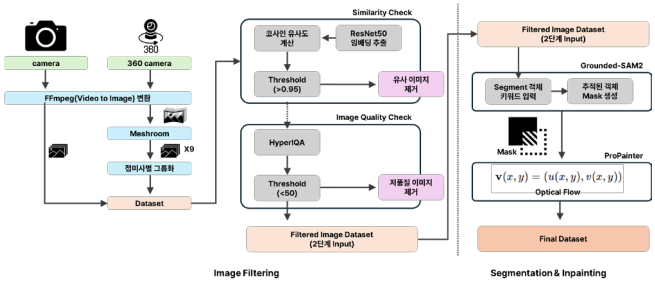
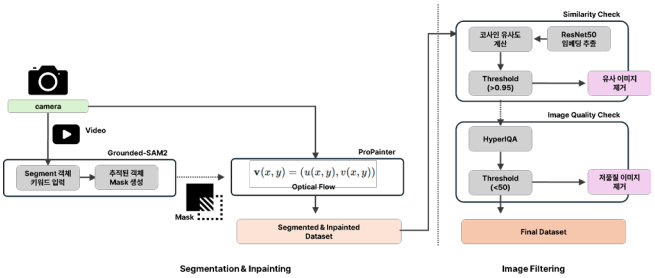


Figure 2: Image filtering pipeline



(a) Object removal and inpainting after image filtering (Experiment 2-1)



(b) Image filtering after object removal and inpainting (Experiment 2-2)

Figure 3: Comparison of two data preprocessing pipelines used in Experiment. (a) performs image filtering first, while (b) applies object removal first.

Table 1: 실험 1: 이미지 필터링 성능 비교

항목	내용
목표	제안하는 필터링 방식(HyperIQA+ResNet50)의 3DGS 복원 효율성 및 품질 기여도 검증
비교 대상	1. 필터링 미적용 원본 데이터셋 2. 제안 방식 (HyperIQA+ResNet50) 3. 전통적 방식 (BRISQUE+CompareHist)
평가 지표	- 처리 시간 (필터링, 3DGS 소요 시간) - 3DGS 결과물 품질 (PSNR, SSIM, LPIPS) - 결합 지표 (품질 및 효율성 종합 평가)

3.2 객체 제거 및 인페인팅

해당 단계에서는 필터링된 이미지 데이터셋에서 원하지 않는 객체를 제거하여 공간의 일관성을 확보한다. 이를 위해 Grounded-SAM2 모델을 사용하여 텍스트 프롬프트 기반으로 객체를 탐지하고, 탐지된 객체에 대해 각 프레임별 세그멘테이션 마스크를 생성한다. 이후 인페인팅 모델인 ProPainter를 사용하여 원본 이미지와 생성된 마스크를 입력받아 객체를 제거하고 인페인팅을 수행한다. ProPainter는 프레임 간 Optical Flow를 활용하여 시각적인 연속성을 유지하는데 강점이 있다.

본 연구에서는 이 과정을 (a)1단계에서 필터링 된 이미지에 적용하는 방식과 (b) 원본 비디오 전체에 적용하는 방식을 모두 실험하여 효율성과 품질을 비교 분석한다.

Table 2: 실험 2: 객체 제거 및 인페인팅 효율성 비교

항목	내용
목표	사전 필터링된 이미지셋을 활용한 객체 제거 방식의 시간 효율성 및 품질 평가
비교 대상	1. (A) 필터링된 이미지에 객체 제거 적용 2. (B) 원본 비디오 전체에 객체 제거 적용
평가 지표	- 인페인팅 소요 시간 및 시간 절감율 - 결과물 품질 (PSNR, SSIM, LPIPS) - 시각적 품질 평가 (배경 일관성, 디테일)

3.3 3DGS 복원 및 병합

각기 다른 층(1층, 2층)을 복원한 PLY 파일을 하나의 모델로 정합하기 위해 CloudCompare와 GaussReg의 장점을 결합한 3단계 하이브리드 방식을 제안한다. 이는 각 도구의 한계점인 속성값 손실, 초기 정합 실패를 상호 보완하기 위함이다. 첫 번째 단계는 초기 근사 정렬로, CloudCompare를 사용하여 두 공간에서 공통으로 관찰되는 Marker를 기준으로 직접 소수의 대응점을 지정하여 초기 정합을 수행한다. 이후 동일한 CloudCompare 내 ICP(Iterative Closest Point) 알고리즘을 적용하여 변환 행렬 T를 정교하게 추출한다. 이후 GaussReg를 통해 앞선 정합에서 얻은 변환 행렬 T를 GaussReg 알고리즘에 초기값으로 제공하여 정밀 정합을 수행한다. GaussReg는 3DGS의 고유 속성(`f_rest`, `opacity` 등)을 보존하면서 정합을 수행할 수 있으므로, 3DGS 데이터의 손실 없이 정확한 병합이 가능하다. 이 방식은 초기 정합이 완료된 상태에서 시작하므로 GaussReg가 겹치는 영역 부족으로 인해 실패할 위험을 줄이고자 한다.



Figure 4: Image marker in staircase

Table 3: 실험 3: 3DGS 공간 병합 정확도 비교

항목	내용
목표	제안하는 2단계 하이브리드 병합 방식의 정합 정확도 및 안정성 검증
비교 대상	1. (A) 제안 방식 (CloudCompare + GaussReg) 2. (B) 단일 방식 (GaussReg Only)
평가 지표	- 정합 정확도 (RMSE) - 처리 속도 및 사용 편의성 - 시각적 품질 평가 (정합 일관성)

Table 4: Image Filtering Results(HyperIQA + ResNet50)

Similarity(S)	Location	HyperIQA(H)	Filtered Images	Metric Time(M_t)	3DGS Time(GS_t)	Total Time(T)	Notes
	1F	**-**	**439**	**-**	**69**	**69**	**GT**
	2F	**-**	**905**	**-**	**102**	**102**	**GT**
0.99↑	1F	60↓	213	3	11	14	
		50↓	358	2	20	22	
	2F	60↓	665	8	79	87	
		50↓	843	8	107	115	
0.95↑	1F	60↓	61	1	31	32	
		50↓	88	1	9	10	
	2F	60↓	295	4	35	39	
		50↓	351	5	37	42	
0.90↑	1F	60↓	33	1	Tracking Failed	-	-
		50↓	43	1	10	11	
	2F	60↓	125	3	Tracking Failed	-	-
		50↓	165	3	22	25	

4. 실험 및 결과

4.1 데이터

실험 환경은 NVIDIA RTX 4090 GPU, Ubuntu 22.04 환경에서 실험을 진행했으며, 데이터는 대학교 내 건물 1층과 2층을 아이폰 14Pro와 Insta360 ONE RS 카메라로 촬영하여 4개의 데이터셋을 구축하였다.

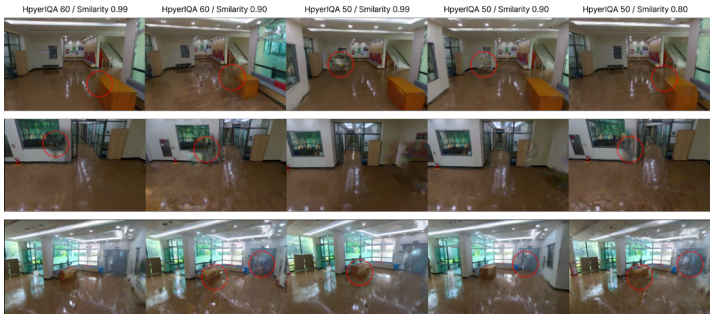


Figure 5: Result images after filtering_1F



Figure 6: Result images after filtering_2F

Table 5: Dataset quality Evaluation

Method	1F			2F		
(S, H)	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
0.99, 60	14.17	0.7790	0.5663	12.62	0.8602	0.3811
0.99, 50	13.90	0.7771	0.5848	13.16	0.8525	0.3528
0.95, 60	14.10	0.7764	0.5961	12.47	0.8549	0.3748
0.95, 50	14.40	0.7860	0.5587	11.16	0.8237	0.4281
0.90, 50	14.35	0.7832	0.5833	13.39	0.8784	0.3559

4.1.1 데이터셋 품질 평가

Figure 5와 6은 필터링 된 이미지를 기반으로 3DGS 복원 결과를 보여준다. 각 실험 조건에 따라 생성된 각 데이터셋은 Unity(게임 엔진)를 활용하여 시각화 하였다. 육안으로 확인하였을 때에는 이미지 장 수가 특정 수준 이하 떨어지면 3DGS 복원 결과가 불안정해지는 것을 확인할 수 있었다.

표 4와 표 5은 이미지 필터링의 기준(HyperIQA 임계값 H, 유사도 임계값 S)에 따라 생성된 3DGS 모델의 최종 렌더링 품질을 정량적으로 평가한 결과로,각 실험 조건별로 1층과 2층 데이터셋에 대한 PSNR, SSIM, LPIPS 값을 보여준다. 특히 S:0.90, H:60 조건에서 이미지의 연속성이 과도하게 훼손되어 SfM 단계에서 특징점 추적에 실패한 것으로 분석된다.

이 평가는 원본 이미지 자체를 Ground Truth로 삼아 절대적인 복원 성능을 측정하기보다, 이미지 데이터를 모두 사용하여 생성된 3DGS 결과물과의 상대적인 품질 차이를 분석하는 데 목적이 있다. 따라서 특정 지표의 절대적인 수치보다는, 각 실험 조건에 따른 값의 변화와 경향성을 비교하여 최적의 필터링 기준을 찾는 것이 더 중요하다.

본 연구에서는 필터링의 효율성을 종합적으로 평가하기 위해

Table 6: Calculation results of the combined indicator (C_{new})

Method	1F			2F		
(S, H)	Equal Weighting	Quality Priority	Efficiency Priority	Equal Weighting	Quality Priority	Efficiency Priority
0.99, 60	-0.1876	0.0752	-0.3191	-0.2959	0.0856	-0.4866
0.99, 50	-0.5625	-0.3375	-0.6750	-0.3937	0.0914	-0.6362
0.95, 60	-0.3870	-0.2284	-0.4663	-0.0195	0.2405	-0.1495
0.95, 50	**0.1472**	**0.4883**	**-0.0233**	-0.2605	-0.1563	-0.3126
0.90, 50	0.0455	0.2674	-0.0655	**0.1956**	**0.5174**	**0.0348**

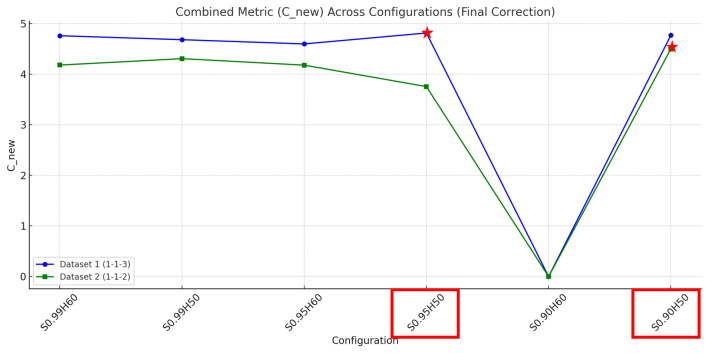


Figure 7: Combined Metric(C_{new})Across Configurations

품질(Q_{avg}), 처리 시간(T), 그리고 입력 이미지 수(I)를 모두 고려한 결합 지표(combined metric) C_{new}를 식 1과 같이 정의했다.

$$C_{new} = w_1 \cdot Q_{avg_norm} - w_2 \cdot \frac{T}{T_{max}} - w_3 \cdot \frac{I}{I_{max}} \quad (1)$$

여기서 각 항목은 다음과 같다.

- Q_{avg}: 정량적 품질 지표(PSNR, SSIM, LPIPS) 가중 평균값
- T / T_{max}: 최대 처리 시간으로 정규화된 처리 시간
- I / I_{max}: 최대 이미지 수로 정규화된 입력 이미지 수
- w_1, w_2, w_3: 각 항목의 중요도를 조절하는 가중치 계수

해당 결합 지표는 3DGS 복원 과정에서 상충관계(trade-off)에 있는 세가지 핵심 요소, 즉 품질(Quality), 처리 속도(Speed), 데이터 효율성(efficiency)을 하나의 점수로 종합하여 직관적으로 최적의 파라미터를 탐색하기 위해 설계하였다. 선형 결합 방식은 각 요소의 기여도를 명확하게 파악할 수 있고, 결과 지표로 널리 사용되어 본 연구에서도 각 평가 항목의 중요도를 직관적으로 조절하고 그 영향을 분석하기에 적합하다고 판단하여 해당 방식을 채택하였다. 제안하는 필터링 기준의 강건함(robustness)을 확인하기 위해, 표 7과 같이 세 가지 다른 가중치 시나리오를 정의하여 민감도 분석을 수행했다. 'Equal Weighting'은 모든 요소를 동등하게 고려하며, 'Quality Priority'는 품질을 가장 중요하게, 'Efficiency Priority'는 처리 속도와 데이터 효율성을 강조한다. 각 시나리오에 따른 결합 지표(C_{new})의 계산 결과는 표 6에 정리하였다.

Table 7: Definition of weight scenarios for calculating the combined indicator (C_{new})

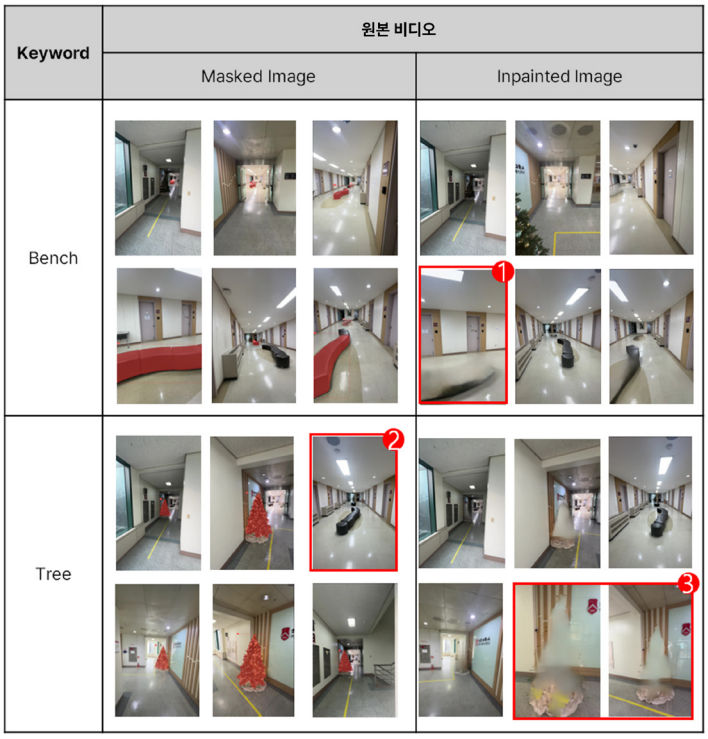
Scenario	w_1 (Q)	w_2 (T)	w_3 (I)
Equal Weighting	1/3	1/3	1/3
Quality Priority	0.6	0.2	0.2
Efficiency Priority	0.2	0.4	0.4

실험 결과, 높은 유사도 (0.99)보다는 적절히 낮은 유사도 (0.95, 0.90)에서 3DGS 결과물이 더 효율적으로 생성되었다는 점은 지나치게 유사한 이미지가 데이터 중복 문제를 야기하여 생성 과정에 방해가 될 수 있음을 시사한다. 이는 3DGS에서 필요한 정보 다양성과 불필요한 중복성 제거가 중요한 요소임을 보여준다. 또한 최적 HyperIQA 값이 50으로 동일하게 유지된 점은 품질이 일정 수준 이상인 이미지가 3DGS결과물 생성에 필수적임을 나타낸다.

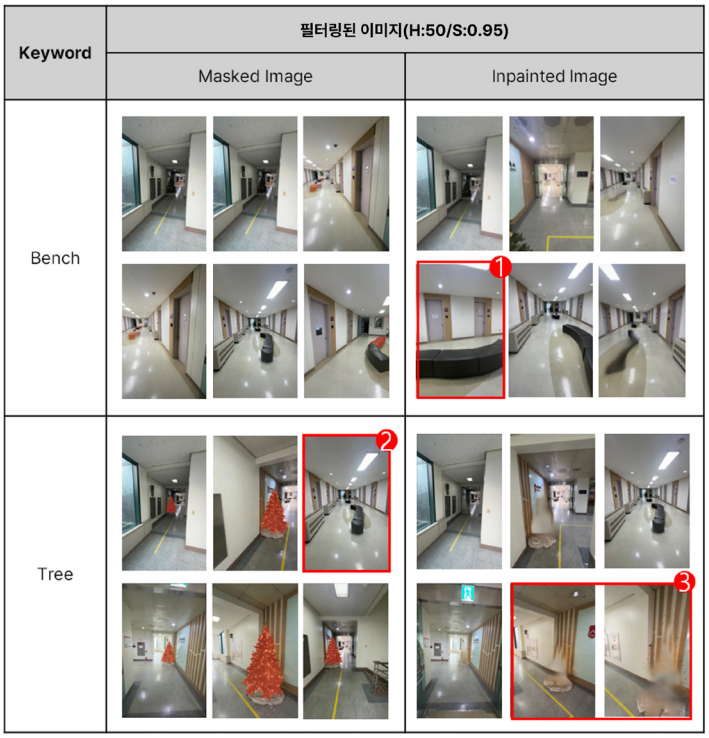
추가적으로 표 1에서 제시한 전통적 방식(BRISQUE + CompareHist)과 비교 실험을 진행했다. 실험 결과, 딥러닝 기반이 아닌 BRISQUE 알고리즘은 동일한 환경(유사도 0.99, IQA 60)에서 이미지 품질 평가 단계(필터링)에서만 대략 60분이 소요되어 새롭게 제안한 딥러닝 기반의 방식보다 월등히 긴 시간이 걸리는 것을 확인하였고, 이외 추가적인 실험을 진행하지 않았다.

4.1.2 객체 탐지 및 인페인팅 성능 비교

본 실험의 목적은 Grounded-SAM2와 ProPainter를 활용한 객체 제거 및 인페인팅의 효율성을 비교하는 것이다. 실험은 두 가지 파이프라인을 비교한다. 첫 번째는 객체 제거 후 이미지 필터링을 수행하는 방식 (그림 8a), 두 번째는 이미지 필터링 후 객체를 제거하는 방식 (그림 8b)이다. 빨간색으로 표시된 3개의 영역(벤치 및 크리스마스 트리)을 대상으로, 인페인팅 결과물의 시각적 품질을 분석한다. 이는 Grounded-SAM2 및 ProPainter 성능을 정성적으로 평가하고, 필터링된 데이터셋과 원본 데이터셋 간 결과를 평가하기 위함이다. 이에, 시각적 품질 평가와 정량적 분석 및 효율성을 평가하고자 한다. 시각적 품질 평가 기준은 다음과 같으며 이에 대한 평가 결과는 표 10에 정리되어 있다.



(a) Image filtering pipeline After Object Removal



(b) Object removal pipeline after image filtering

Figure 8: Comparison between two overall research pipelines

- **디테일 보존**: 인페인팅 영역의 디테일 복원 수준
- **객체 탐지 능력**: Grounded-SAM2의 객체 탐지 능력 수준
- **배경 일관성**: 인페인팅 영역, 주변 배경과 자연스러움 수준

Table 8: Inpainting time comparison

Key Word	Data	Time (min)	Frame Count	Time per Frame (min/frame)	Time Reduction (%)
Bench	O	175.7	302	0.58	-
	F	40.9	259	0.16	76.69
Tree	O	230.8	302	0.76	-
	F	50.9	259	0.20	77.92

Table 9: Inpainted Image Quality Evaluation

키워드	PSNR ↑	SSIM ↑	LPIPS ↓
Bench(Original)	48.97	0.9957	0.0073
Bench(Filtered)	51.26	0.9981	0.0033
Tree(Original)	48.27	0.9871	0.0190
Tree(Filtered)	48.77	0.9914	0.0129

또한 정량적인 분석을 통해 제안 방식의 효율성을 평가하고자, 시간 효율성과 결과물 품질이라는 두 가지 측면을 종합적으로 고려하였다. 먼저, 시간 효율성은 원본 데이터셋 대비 필터링된 데이터셋의 인페인팅 처리 시간 절감율(R_T)로 측정하였으며, 계산식은 아래 식 2과 같다. 표 8는 이 계산 결과를 보여주며, 필터링된 데이터셋이 벤치와 나무 영역에서 각각 76.69%와 77.93%의 높은 시간 절감 효과를 보였음을 확인했다.

$$R_T = \frac{T_{\text{original}} - T_{\text{filtered}}}{T_{\text{original}}} \times 100 \tag{2}$$

여기서 T_{original}은 원본 데이터셋의 인페인팅 소요 시간, T_{filtered}는 필터링된 데이터셋의 인페인팅 소요 시간을 나타낸다.

결과물 품질은 PSNR, SSIM, LPIPS 지표로 평가했다. 각 조건별 상세 품질 지표는 표 9와 같다.

본 연구에서는 이 두 가지 상충 관계(trade-off)에 있는 지표를 종합적으로 평가하기 위해, 식 3과 같이 새로운 복합 지표(C_{comp})를 정의했다. 이때 $Q_{inpainting}$은 각 지표(PSNR, SSIM, LPIPS)를 0과 1 사이로 정규화한 뒤, 세 지표를 동일한 중요도로 간주하여 1/3씩 균등하게 가중평균하여 산출했다.

$$C_{comp} = w_1 \cdot Q_{inpainting} + w_2 \cdot \frac{R_T}{100} \tag{3}$$

표 11은 이 복합 지표 점수를 복합 지표(C_{comp}) 수식 내 w_1과 w_2가 동일한 가중치 기준으로 정리한 것이다. 두 키워드(Bench, Tree) 모두에서 필터링된 데이터셋(Filtered)을 사용한 경우가 원본(Original)보다 월등히 높은 종합 점수를 기록했다. 특히 'Bench' 데이터셋의 경우, 필터링을 통해 종합 점수가 0.413에서 0.883으로 두 배 이상 상승했다. 이는 제안한 방식이 인페인팅 시간을 획기적으로 단축하면서도 결과물의 품질은 유사하거나 더 높게 유지하여, 전반적인 효율성을 크게 향상시킴을 명확히

Table 10: Image filtering results (HyperIQA + ResNet50)

Image Pair	Detail Preservation	Object Detection	Background Consistency	Conclusion
Pair 1	Insufficient	Partially Successful	Inconsistent	The original dataset showed superiority in close-range object removal, but both datasets lacked detail preservation.
Pair 2	-	Failure	-	Both datasets failed to detect objects, and the tree region was not clearly recognized.
Pair 3	Inconsistent	Successful	Partially Successful	The filtered dataset showed better performance in background reconstruction, but still failed to produce a perfect result.

Table 11: Analysis of the Composite Metric(C_{comp}) for the Inpainting Experiment

Experiment	Composite Metirc(C_{comp})
Bench (Original)	0.413
Bench (Filtered)	**0.883**
Tree (Original)	0.000
Tree (Filtered)	**0.677**

보여준다. 정량적 분석 결과, 필터링 된 데이터셋은 인페인팅 처리 시간이 크게 단축되었으며, PSNR 등 품질 평가 지표에서 더 우수한 성능을 나타냈다. 이는 고품질 데이터셋을 선별하여 사용한 결과로 분석되며, 품질 평가를 진행한 주요 이유는 인페인팅 과정에서 발생할 수 있는 이미지상의 노이즈를 정량적으로 확인하기 위함이다. 특히, 품질 지표 점수가 높을수록 노이즈 발생이 적고 인페인팅 결과가 더 우수하다는 판단하에 해당 수치를 활용하였다.

4.1.3 공간 복원 실험 결과

본 실험은 앞서 제안한 정합을 위한 두가지 방법을 정합 정확도, 처리 속도, 속성 보존, 사용 편의성 기준으로 비교하여 포인트 클라우드의 정합 환경을 평가한다. 1차 정합은 CloudCompare를 사용하여 수동으로 수행하며, 2차 정합은 GaussReg를 통해 자동으로 수행한다.

- **1차 정합**: CloudCompare 기반 초기 정합 수행 및 변환 행렬 추출(Point Picking, ICP 알고리즘 활용)
- **2차 정합**: GaussReg를 활용한 3DGS 특성을 보존한 Fine Registration

1차 정합의 경우 아래 그림 4과 그림 9과 같이 1층과 2층에서 모두 확인 가능한 이미지 마커를 기준으로 수동 정합을 진행하였고, 특정 포인트들을 Align(point pairs picking) 및 ICP 알고리즘 사용하여 정합을 진행하였다. 이를 통해 1층과 2층 PLY 정합을 위한 행렬 값 T를 추출하였다. 특히 아래 그림 10에서와 같이 CloudCompare만으로 정합이 가능하지만 단순한 포인트들의 Transformation만 변경하기 때문에 3DGS 정합 정확도에 한계가 있다.

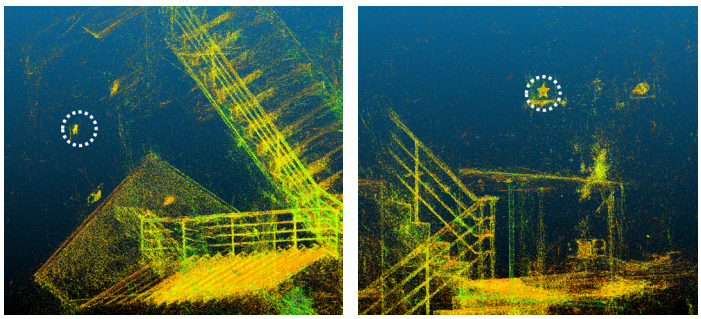


Figure 9: Pointcloud in Staircase

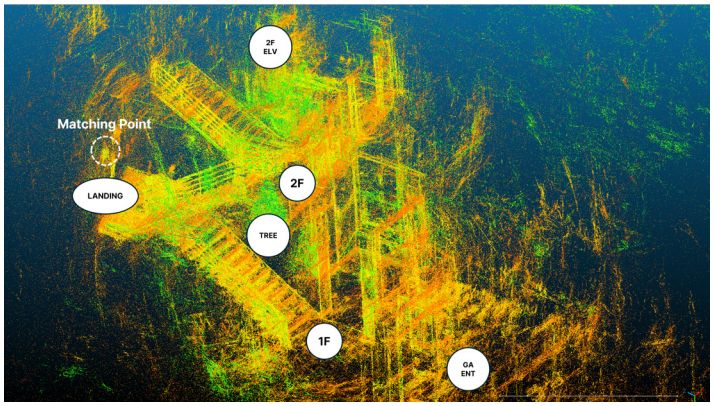


Figure 10: Merge With CloudCompare

$$\mathbf{T} = \begin{bmatrix} -0.977 & -0.060 & 0.130 & -0.604 \\ -0.093 & 0.948 & -0.259 & -6.103 \\ -0.109 & -0.269 & -0.943 & 0.546 \\ 0.000 & 0.000 & 0.000 & 1.000 \end{bmatrix} \tag{4}$$

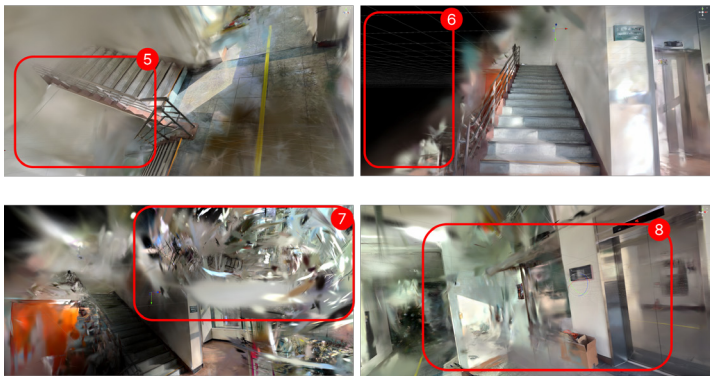
따라서 1차 정합에 사용된 위와 같은 변환 행렬 $\mathbf{T}$를 2층 PLY 데이터 $\mathbf{P}_2$에 적용하면 다음과 같은 수식으로 표현된다:

$$\mathbf{P}_2' = \mathbf{T}\mathbf{P}_2 \tag{5}$$

여기서 P_2'는 변환된 2층 PLY 데이터의 좌표를 나타내며 이 값

을 사용하여 대략적인 정합을 거친 PLY 파일을 GaussReg의 입력으로 사용한다. 2차 정합에서는 GaussReg를 활용하여 3DGS 고유 속성을 반영하여 정합을 수행한다. 다만 아래와 같이 GaussReg를 대략적인 정합 없이 단독으로 사용하는 경우에는 아래 그림 11a과 그림 12a과 같이 전혀 정합이 되지 않는 경우가 발생한다. 이는 특히 계단실과 같이 특징점을 뽑아 낼 수 없는 프레임이 연속되는 공간이거나, 겹치는 영역이 부족할 경우, 좁은 영역에서 정합 성능이 많이 떨어짐을 확인할 수 있었다.

반면, 아래 그림 11a과 그림 11b 내의 1번과 5번 pair를 보게 되면, 1번의 경우 계단실의 공간이 특정 이하로 깊어지면 화면이 제대로 렌더링 되지 않고 뿌옇게 그려지는 모습을 볼 수 있으나, 1차 정합을 한 후 GaussReg를 적용한 5번의 경우, 계단실의 깊이가 깊어지더라도 화면이 뿌옇게 그려지지 않고, 3DGS 고유 속성을 반영하여 정합이 잘 되는 것을 확인할 수 있다. 이러한 시각적 평가에 더하여, 두 방식의 성능을 객관적으로 비교하기 위해 정합 정확도(RMSE), 처리 시간, 속성 보존, 사용 편의성 측면에서 평가를 진행하였다. 그 결과는 표 12에 정리되어 있다.

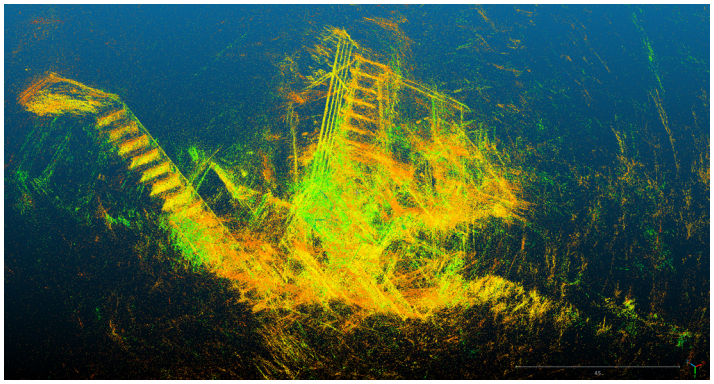


(a) Using only GaussReg

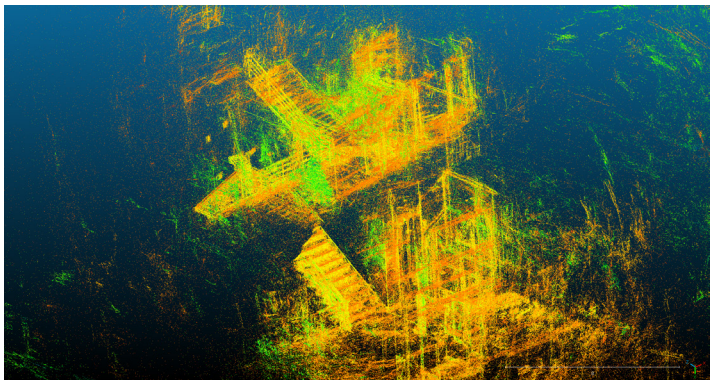


(b) CloudCompare with GaussReg

Figure 11: Comparison Between Two Overall Research Pipelines



(a) Pointcloud using with GaussReg



(b) Pointcloud using hybrid registration approaches

Figure 12: Comparison with GaussReg & hybrid registration approach

Table 12: Quantitative evaluation results of hybrid registration

Evaluation Metric	**GaussReg**	**Combined Method**
RMSE	2.577	2.306
Time	7min	12min
Attribute Preservation	Preserved	Preserved
Ease of Use	Easy	Mostly Difficult

해당 실험을 통해 CloudCompare와 GaussReg를 혼용한 방법은 사용 편의성 측면에서 수동 피킹(Align Point Picking) 과정으로 인해 다소 복잡하고,처리 속도에서도 GaussReg만을 사용한 방법보다 시간이 더 소요되는 결과를 보였다. 이때 측정된 약 5분의 시간 차이는 CloudCompare 내부에서 수행된 ICP 알고리즘의 연산 시간에 기인한다. 사용자가 직접 대응정을 지정하는 수동 피킹 시간은 작업자의 숙련도에 따라 편차가 커 객관적인 비교 지표로 사용하기 어렵다고 판단하여, 본 연구의 정량적인 시간 평가에서는 제외하였다. 그러나 가장 중요한 정합 정확도에서는 두 PLY 파일 간 겹치는 부분이 적은 환경에서도 CloudCompare와 GaussReg를 혼용한 방법이 더 우수한 성능을 나타냈다. 따라서 두 가지 방법을 혼용하는 것이 복잡한 구조나 겹침이 적은 데이

터에서 정확한 정합을 수행하는 데 가장 적절한 방법임을 확인할 수 있었다.

5. 결론 및 한계점

5.1 결론

본 논문은 3D 가우시안 스플래팅(3DGS)을 활용하여 복잡한 실내 공간을 효율적으로 복원하는 실질적인 방법론을 제안했다. 기존 연구들이 주로 객체 또는 실외 공간 복원에 집중하거나, 복원 이후에 포인트 클라우드를 수정하는 방식에 머물렀던 것과 달리, 본 연구는 **입력 이미지 데이터셋을 사전에 정제하는 전처리 과정**의 중요성을 강조했다.

주요 기여는 다음과 같다:

- **이미지 필터링**: HyperIQA와 ResNet을 활용한 사전 필터링을 통해 원본 데이터를 그대로 사용하는 것보다 속도 면에서 월등한 성능을 보이면서도 품질은 큰 차이가 없음을 검증했다.
- **객체 제거**: 필터링된 이미지를 사용함으로써 고사양의 컴퓨팅 자원을 요구하는 인페인팅 시간을 획기적으로 단축했으며, 일부 객체에 대해서는 더 높은 인페인팅 성능을 확인했다.
- **공간 병합**: 특징이 적어 복원이 어려운 계단실 같은 영역에서 CloudCompare와 GaussReg를 혼용하는 하이브리드 방식이 개별적으로 사용하는 것보다 품질 면에서 더 우수함을 보이며, 기존 연구에서 부족했던 층간 정합의 가능성을 제시했다.

결론적으로, 본 연구는 3DGS 기반 실내 공간 복원 시 사전 이미지 처리의 효율성과 중요성을 입증했으며, XR 및 디지털 트윈 분야에 기여할 수 있는 가능성을 보였다.

5.2 연구의 한계

본 연구는 다음과 같은 한계점을 가진다.

- **일반화의 한계**: 실험이 특정 건물과 두 종류의 카메라로 제한되어 결과를 모든 실내 공간에 일반화하기 어렵다.
- **실험 범위의 제약**: 객체 제거 실험이 특정 공간의 두 가지 객체로 한정되었고, 인페인팅 성능이 완벽하지 않았다.
- **기술적 제약**: 계단실과 같이 수직적 변위가 큰 공간은 SfM [12] 특징점 추출이 어렵고, 3DGS의 표현력에도 한계가 있었다. 또한, 정합에 대한 새로운 이론을 제시하지는 못했다. 또한 제안한 하이브리드 정합 방식이 정확도는 개선했지만, CloudCompare를 활용한 수동 정합 과정이 필요하여 완전 자동화에는 한계가 있으며, 사용자의 숙련도에 따라 결과의 편차가 발생할 수 있다.

5.3 향후 연구 방향

연구 과정에서 발견된 한계점을 보완하기 위해 다음과 같은 후속 연구가 고려되어야 한다.

- **촬영 효율성 증대**: 촬영 시 누락된 영역을 실시간으로 시각화하여 피드백을 주는 AR 기반 시스템을 개발하여 데이터 수집의 완전성을 높인다.
- **인페인팅 최적화**: 최신 객체 탐지 및 인페인팅 알고리즘들을 다양하게 조합하고 테스트하여, 더 나은 성능을 내는 최적의 방법론을 도출한다.
- **공간 정합 문제 해결**: 좁은 계단이나 경사로처럼 복잡한 수직 구조물을 효과적으로 복원하고, 높이 차이를 고려하여 정밀하게 정합할 수 있는 새로운 알고리즘을 탐구한다.

본 논문은 3DGS를 활용한 대규모 실내 공간 복원을 위해 이미지 필터링, 객체 제거, 공간 병합 과정을 체계적으로 통합한 실용적인 파이프라인을 제안하였다. 효율적인 데이터 전처리를 위해 HyperIQA와 ResNet을 활용한 이미지 필터링이 3DGS 복원 속도와 효율성을 크게 향상시킴을 입증하였으며, 사전 필터링된 이미지를 인페인팅에 활용함으로써 고사양의 컴퓨팅 자원을 요구하는 처리 시간을 단축시켰다. 또한 CloudCompare와 GaussReg를 결합한 하이브리드 정합 방식을 제안하여, 기존 연구에서 어려움을 겪었던 복층 계단실과 같은 복잡한 구조의 정합 정확도를 개선하고 그 가능성을 제시하였다. 본 연구는 3DGS 기술을 활용한 실내 공간의 복원 과정에서 사전 이미지 처리의 중요성을 강조하며, 제안된 방법이 XR, 디지털 트윈 등 다양한 분야의 실내 공간 콘텐츠 제작에 효과적으로 기여할 수 있음을 보여주었다. 향후 연구로는 촬영 가이드 시스템 개발, 인페인팅 기술 고도화, 복잡한 구조의 정합 알고리즘 개선 등이 필요할 것으로 보인다.

감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이고(RS-2023-00251681), 정보통신기획평가원의 대학ICT연구센터사업(RS-2023-00259099)과 메타버스융합대학원(RS-2022-00156318)의 지원으로 수행되었음.

References

[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," *European Conference on Computer Vision (ECCV)*, pp. 405–421, 2020.

[2] G. K. Bernhard Kerbl, "3d gaussian splatting for real-time radiance field rendering,," *ACM Transactions on Graphics*, pp. vol. 42, no. 4, pp. 1–13, 2023.

[3] 배종환, “가우시안 스플래팅을 활용한 실내공간 복원의 실질적 접근,” 공학 석사학위논문, 서강대학교 메타버스전문대학원, 서울, 12 2024.

[4] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[5] S. Su, Q. Yan, Y. Tai, C. Shen, X. Shen, X. Tao, Y.-W. Tai, and C.-K. Tang, “Blindly assess image quality in the wild guided by a self-adaptive hyper network,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 3667–3676.

[6] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, “Deep neural networks for no-reference and full-reference image quality assessment,” *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206–219, 2018.

[7] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[8] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, P. Dollar, and R. Girshick, “Segment anything,” *arXiv preprint*, 2023, arXiv:2304.02643.

[9] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, H. Su, and J. Zhu, “Grounding dino: Marrying dino with grounded pre-training for open-set object detection,” *arXiv preprint*, 2023, arXiv:2203.17200.

[10] Y. Li *et al.*, “Improving propagation and transformer for video inpainting,” https://arxiv.org/abs/2309.03897, 2023, arXiv preprint arXiv:2309.03897.

[11] J. Chang, Y. Xu, Y. Li, Y. Chen, W. Feng, and X. Han, “Gaussreg: Fast 3d registration with gaussian splatting,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 23 454–23 463.

[12] J. L. Schonberger and J.-M. Frahm, “Structure-from-motion revisited,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4104–4113.

[13] A. Myronenko and X. Song, “Point set registration: Coherent point drift,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.

〈저자소개〉



배종환

- 2016 경희대학교 건축학과 건축학사
- 2025 서강대학교 메타버스전문대학원 메타버스테크놀로지 전공 공학석사
- 관심분야: 인공지능, 메타버스, 확장현실, 건축/건설, 디지털트윈
- https://orcid.org/0009-0007-7520-2319



박상훈

- 1993 서강대학교 수학과 학사
- 1995 서강대학교 컴퓨터학과 석사
- 2000 서강대학교 컴퓨터학과 박사
- 2022 ~ 2005 대구가톨릭대학교컴퓨터정보통신공학부조교수
- 2001 University of California, Davis 방문연구원
- 2005 ~ 2023 동국대학교 멀티미디어학과 교수
- 2023 ~ 현재 서강대학교 메타버스전문대학원 교수
- 관심분야 : 실시간 렌더링, 사실적 렌더링, 과학적 가시화, 고성능 컴퓨팅 등
- https://orcid.org/0000-0001-5383-7005