

조명 불일치를 고려한 3D 가우시안 스플래팅 기반 카메라 자세 정제

강윤석, 이재구*

국민대학교 컴퓨터공학과

*jaekoo@kookmin.ac.kr

요 약

조명·노출 변화는 3D 가우시안 스플래팅(3DGS) 기반 카메라 자세 정제(CPR) 과정에서 외관 불일치 문제를 유발하여 전반적인 자세 추정 성능을 저하시킨다. 본 논문은 GS-CPR 의 전역 색상 보정을 대체하여, 3DGS 렌더링 단계에서 픽셀 단위 외관 분리를 수행하고 채널 혼합을 배제한 대각 제약 변환으로 국소 조명 변화를 모델링한다. 또한 기하학적 정규화를 통해 외관 보정 과정에서의 기하 교란을 완화하고, 추론 단계에서는 제로 임베딩으로 조명 성분의 영향을 억제해 조명 변화 환경에서도 자세 추정의 안정성을 높인다. Cambridge Landmarks 와 Oxford Day-and-Night 에서 OldHospital 의 중위수 이동/회전 오차를 $0.19\text{ m} \rightarrow 0.18\text{ m}$, $0.38^\circ \rightarrow 0.35^\circ$ 로 개선했으며, HB Allen Centre 에서도 이동 오차를 $2.99\text{ m} \rightarrow 2.09\text{ m}$ 로 크게 줄이는 등 조명 변화 환경에서 전반적인 자세 추정 성능을 향상시켰다. 외관 변화량(AS) 분석을 통해 장면의 조명 변화 정도에 따라 제로 임베딩 사용 여부의 효과가 달라짐을 보였다. 추가로 제거 실험으로 기하 정규화와 대각 제약이 성능에 미치는 영향을 확인했다.

1. 서론

시각 기반 자세 추정(Visual Localization, VL)은 단일 이미지로부터 카메라의 6-자유도 자세를 추정하는 문제로, 자율 주행, 로봇 내비게이션, 증강 현실(AR) 등 다양한 응용에서 핵심 기술로 활용된다[1]. 특징 기반 자세 추정[2] 및 학습 기반 접근(APR[3], SCR[4])은 다양한 벤치마크에서 성과를 보여 왔으나, 학습 분포를 벗어나는 환경에서 일반화가 제한되어 정밀도 측면에서 수십 cm 이상의 오차가 발생하는 한계가 보고되어 왔다[5, 6].

이러한 정밀도 한계를 보완하기 위해, 거친 초기 자세(Coarse Pose)를 입력으로 받아 별도의 3 차원 장면 표현과 비교하여 자세를 미세 조정하는 카메라 자세 정제(Camera Pose Refinement, CPR)가 제안되었다[7]. 초기 연구들은 NeRF[8] 기반 장면 표현을 활용하였으나, 반복적인 볼륨 렌더링으로 인한 높은 계산 비용이 실시간 적용의 제약으로 지적되었다[9]. 반면 3D 가우시안 스플래팅(3DGS)[10]은 명시적 3D 표현을 통해 렌더링을 크게 가속하였고, GS-CPR[11]은 재학습 없이도 3DGS 기반 정밀 자세 정제를 가능하게 하여 실용성을 높였다. 그러나 GS-CPR 은 이미지 전체에 동일한 색상 변환을 적용하는 전역적 보정에 의존하므로, 그림자나 국소 조명 변화가 큰 환경에서는 렌더링-입력 간 외관 불일치가 잔존할 수 있다.

한편 외관 분리(Appearance Decoupling)는 기하 구조와 조명·노출 등 외관 요인을 분리하여 모델링함

으로써, 픽셀 단위 외관 변화를 유연하게 표현하는 방향으로 발전해 왔다. NeRF-W[12]와 DAVIGS[13]는 3 차원 좌표와 뷰 임베딩을 이용해 픽셀 단위 색 변환을 예측하는 방식으로 다양한 외관 변화의 표현 가능성을 보였다. 그러나 이러한 외관 분리 모델을 CPR 에 그대로 도입하면, 모델이 정밀한 3D 구조를 맞추기보다 외관 변환으로 픽셀 오차를 감소시키는 방향으로 학습이 진행될 수 있으며, 이는 기하학적 모호성(Geometric Ambiguity)로 이어져 자세 정제 성능을 저하시킬 위험이 있다[12].

본 연구는 GS-CPR 의 렌더링 단계에 픽셀 단위 외관 분리 모듈을 통합하되, 외관 변환의 자유도를 제한하고 3D 기하 파라미터를 안정화하기 위한 정규화를 도입하여 위 문제를 완화한다. 또한 테스트 단계에서는 외관 임베딩을 최적화하는 대신 영벡터로 고정하는 제로 임베딩(Zero Embedding) 추론 전략을 적용하여, 학습된 조명 편향의 영향을 줄이고 구조 중심의 렌더링을 유도한다. 이를 통해 MAST3R[14] 기반 특징점 매칭과 PnP[15] 정제의 안정성을 향상시키는 것을 목표로 한다.

2. 관련 연구

2.1 시각 기반 자세 추정 카메라 자세 정제

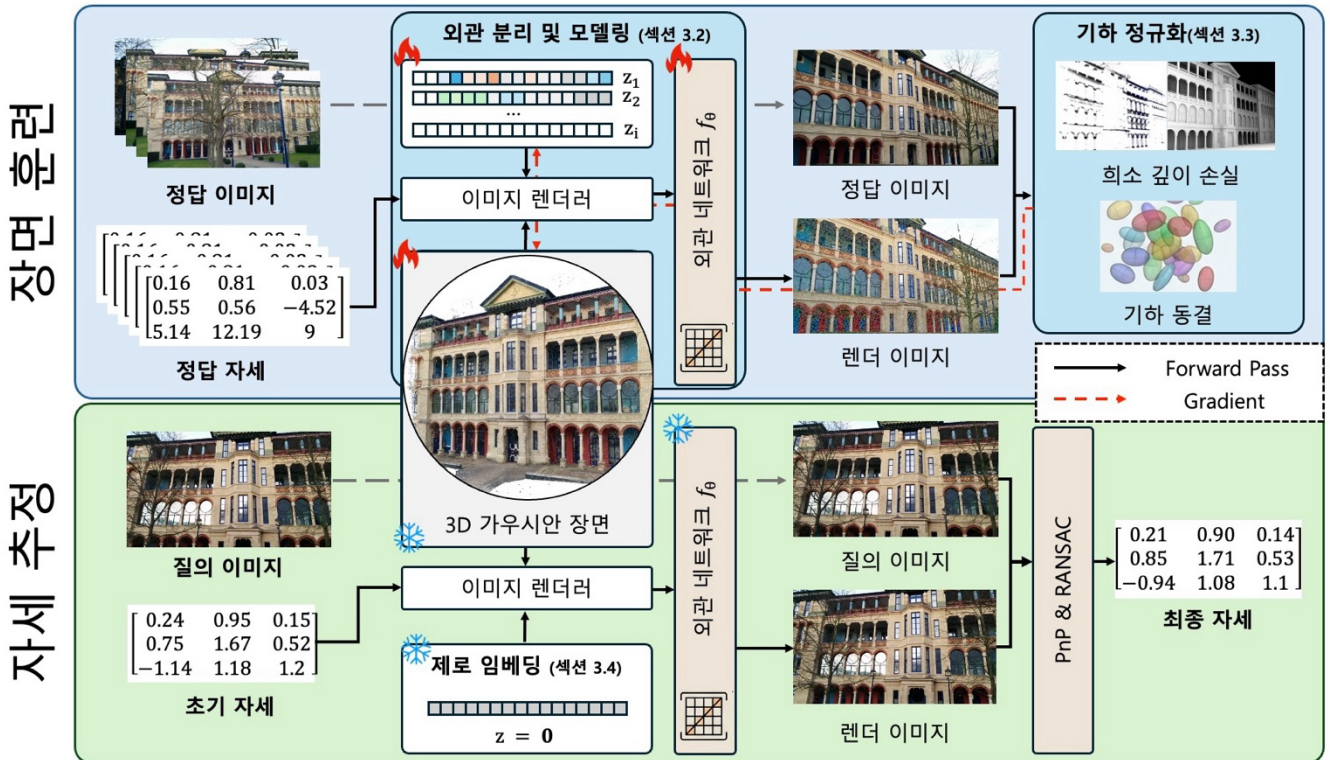


그림 1. 제안 방법의 장면 훈련 및 자세 추정 파이프라인 개요

시각 기반 자세 추정은 매칭 기반 방법과 학습 기반 방법으로 구분된다. 매칭 기반 방법(예: HLoc[2])은 높은 정밀도를 보이지만 특징점 매칭과 RANSAC[16]에 따른 계산 비용이 크며, 학습 기반 방법(APR[3], SCR[4])은 빠른 추론이 가능하나 학습 분포 밖 환경에서 정밀도가 저하되는 한계가 보고되었다[5]. 이러한 정밀도 문제를 보완하기 위해 카메라 자세 정제(CPR)는 초기 자세를 기준으로 렌더링 이미지와 질의 이미지의 불일치를 줄이는 방향으로 자세를 최적화한다. NeRF 기반 CPR[7, 8]은 반복적인 볼륨 렌더링으로 인해 비용이 높았던 반면, GS-CPR[11]은 3DGS[10]의 빠른 렌더링을 활용해 실용적인 정제 파이프라인을 제시하였다. 다만 GS-CPR은 전역 색상 보정(Global ACT)에 의존하므로, 장면 내 국소 조명 변화(Local Illumination Change)가 큰 경우 외관 불일치가 잔존하여 정제 정밀도를 제한할 수 있다[17].

2.2 NeRF/3DGS의 외관 분리와 CPR에서의 한계

조명·날씨 변화에 대응하기 위해 기하(Geometry)와 외관(Appearance)을 분리하는 연구가 진행되어 왔다. NeRF-W[12]는 외관 임베딩을 도입해 강건한 뷰 합성을 보였고, 3DGS에서도 3D 좌표와 뷰 임베딩을 입력으로 픽셀 단위 색 변환을 예측하는 방식이 제안되었다[13]. 그러나 이러한 외관 분리 모델을 CPR에 그대로 적용하면, 외관 네트워크가 픽셀 오차를 줄이는 방향으로 과도하게 작동하면서 그림

자·반사 등의 외관 변화가 기하로 흡수되는 기하학적 모호성(Geometric Ambiguity)이 발생할 수 있으며 [12], 이는 자세 정제의 안정성을 저해할 위험이 있다. 따라서 CPR에서는 외관 분리의 표현력을 활용하되, 기하 일관성을 유지하도록 자유도를 제어하는 설계가 요구된다.

3. 방법론

3.1 전체 시스템 개요

본 연구는 3D 가우시안 스피클래팅(3DGS)[10] 기반 장면 표현을 활용하여, 주어진 초기 자세를 정밀 자세로 보정하는 카메라 자세 정제(Camera Pose Refinement, CPR) 파이프라인을 제안한다. 그림 1은 제안 방법의 전체 흐름을 보여주며, 파이프라인은 (1) 초기 자세 추정, (2) 3DGS 기반 장면 렌더링 및 외관 처리, (3) 특징점 매칭, (4) PnP 기반 자세 정제의 네 단계로 구성된다.

먼저 질의 이미지에 대해 ACE[18]와 같은 전역 자세 회귀 모델로 초기 자세를 얻는다. 이후 해당 자세에서 사전 학습된 3DGS 장면 모델을 렌더링하고, 렌더링 단계에서 외관 처리를 수행하여 조명·노출 차이로 인한 외관 불일치를 완화한다. 본 연구의 외관 처리는 (i) 픽셀 단위 외관 분리 모듈(3.2 절), (ii) 기하 안정화를 위한 정규화 전략(3.3 절), (iii) 추론 시 외관 임베딩을 영벡터로 고정하는 제로 임베딩 전략(3.4 절)으로 구성된다. 이를 통해 매칭과 PnP가

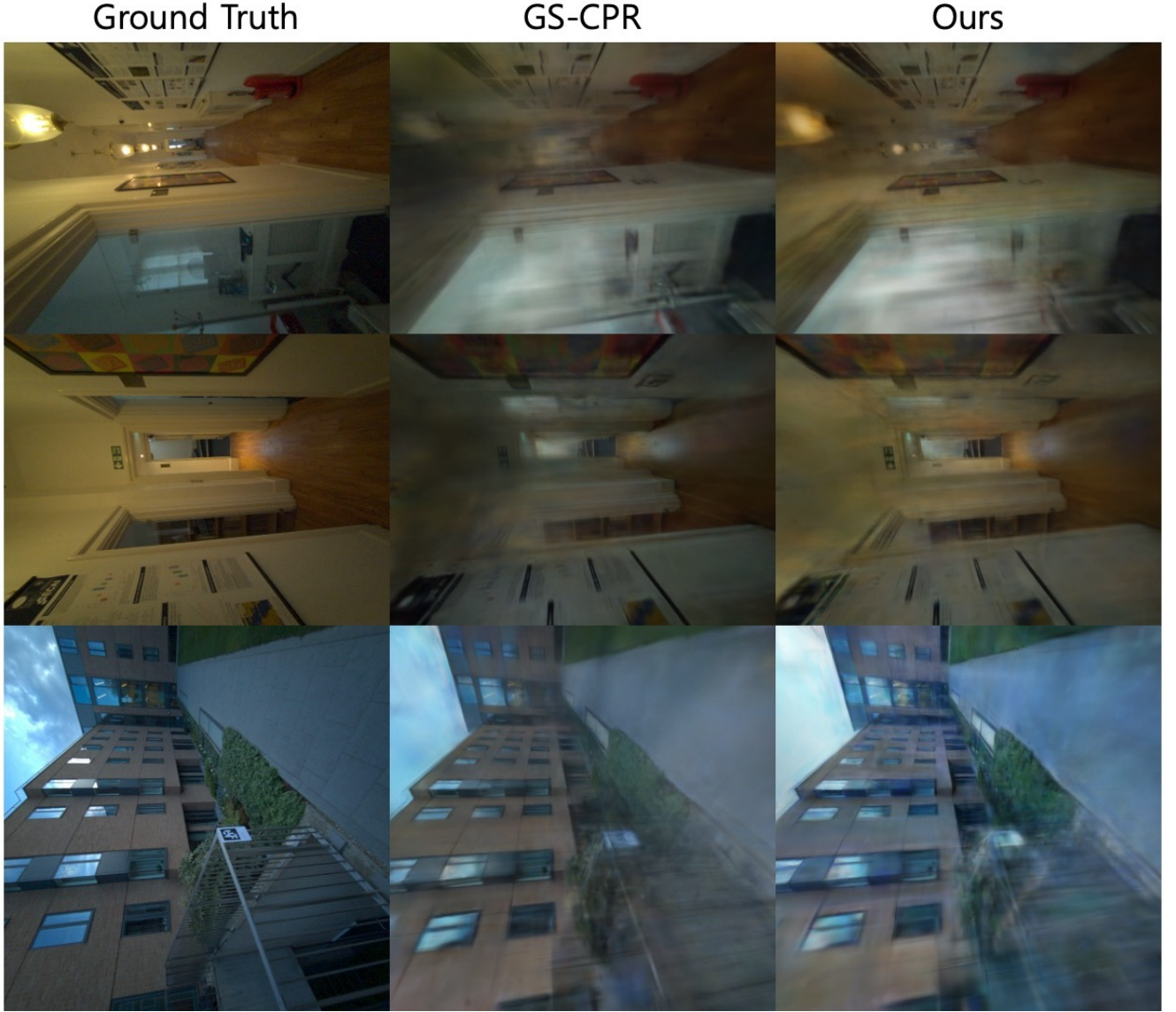


그림 2. GS-CPR 과 Ours 의 렌더링 품질 비교

색감·밝기 변화보다 구조적 단서(윤곽, 에지 등)에 기반해 안정적으로 동작하도록 유도한다.

외관 처리가 적용된 렌더링 결과와 질의 이미지는 MAST3R 기반 매칭 네트워크에 입력되어 2D-2D 대응점을 계산하며, 렌더링 시 얻은 깊이로 2D-3D 대응을 구성한 뒤 RANSAC 기반 PnP[15]로 최종 자세를 정제한다. 이후 절에서는 외관 분리 모듈(3.2 절), 기하 정규화 전략(3.3 절), 제로 임베딩 추론(3.4 절)을 순서대로 설명한다.

3.2 외관 분리 및 모델링

본 절에서는 렌더링 이미지의 국소 조명·노출 변화를 보정하기 위해, 픽셀 단위 외관 분리 (Appearance Decoupling) 기반의 외관 모델을 정의한다. 렌더링 이미지의 각 픽셀 색상은 동일한 기하·카메라 조건에서의 기하 기반 성분과 환경 변화에 의해 달라지는 외관 성분의 결합으로 보고, 후자를

채널별 스케일과 오프셋으로 모델링한다. 구체적으로, 3DGS로부터 초기 자세에서 렌더링한 이미지 I_{render} 에 대해 픽셀 좌표 $u = (x, y)$ 에서의 최종 출력 색상 $I_{\text{out}}(u)$ 은 RGB 채널 간 혼합을 허용하지 않는 대각 아핀 변환으로 다음과 같이 정의한다.

$$I_{\text{out}}(u) = s(u) \odot I_{\text{render}}(u) + t(u) \quad (1)$$

여기서 $I_{\text{render}}(u) \in \mathbb{R}^3$ 는 렌더링된 RGB 값이며, $I_{\text{out}}(u) \in \mathbb{R}^3$ 는 외관 보정 이후의 RGB 값을 나타낸다. $s(u) \in \mathbb{R}^3$, $t(u) \in \mathbb{R}^3$ 는 각각 채널별 스케일과 채널별 오프셋을 나타내고, $\odot$ 는 원소별 곱을 의미한다.

$s(u)$ 와 $t(u)$ 는 픽셀 단위 외관 네트워크 f_θ 가 예측하며, 네트워크 입력은 픽셀 u 에서 추출한 렌더링 기반 특징 $\phi(u) \in \mathbb{R}^d$ 와 뷰(또는 조명) 정보를 압축한 외관 임베딩 $z \in \mathbb{R}^D$ 의 결합으로 구성된다.

$$[s(u), t(u)] = f_\theta(\phi(u), z) \quad (2)$$

외관 임베딩 z 는 학습 단계에서 장면의 조명·노출 스타일을 요약하도록 최적화되며, 추론 단계에서는 3.4 절에서 설명하는 제로 임베딩 전략에 따라 $z = 0$ 으로 고정된다. 생성된 I_{out} 은 이후 특징점 매칭 및 PnP 기반 자세 정제 단계의 입력으로 사용된다.

3.3 기하학적 정규화 및 학습 안정화

외관 네트워크의 표현력이 커질수록, 모델이 3D 가우시안의 기하(위치/형상)를 맞추기보다 색 변환으로 손실을 줄이는 방향으로 학습되어 기하학적 모호성이 발생할 수 있다. 이를 완화하기 위해 본 연구는 학습 과정에서 외관 변환을 정규화하고, 후반 학습 단계에서는 가우시안의 기하 파라미터(예: 위치, Opacity)를 동결(Freeze)한 채 외관 모듈 및 색 관련 파라미터만 갱신하도록 제한한다. 추가로 SfM 희소 포인트 클라우드를 활용하여, 렌더링 깊이와 실제 3D 포인트 깊이의 차이를 줄이는 희소 깊이 손실을 적용한다. 이를 통해 외관 보정이 오차를 줄이는 과정에서 기하 불일치를 외관 변화로 흡수하는 경향을 억제하고, 3D 기하가 관측과 일관되게 유지되도록 유도한다.

3.4 제로 임베딩 추론 전략

3.2 절의 외관 임베딩 z 는 학습 과정에서 장면의 조명·노출 변화를 설명하도록 최적화되지만, 테스트 단계에서는 장면·조건 차이로 인해 학습 편향이 매칭 안정성을 저해할 수 있다. 따라서 본 연구는 추론 시 $z = 0$ 으로 고정(제로 임베딩)하여, 뷰에 종속된 조명 성분을 억제하고 구조 중심의 외관이 렌더링되도록 한다.

4. 실험 결과 및 분석

4.1 실험 설정

본 연구는 Cambridge Landmarks[3], Oxford Day-and-Night [19]에서 제안 방법의 유효성을 평가하였다. 비교 방법은 3DGS, GS-CPR, Ours의 세 가지이며, 성능은 중위수 이동 오차(Med. Trans., m)와 중위수 회전 오차(Med. Rot., °)로 측정하였다. Cambridge Landmarks는 GS-CPR에서 제공하는 COLMAP 재구성 결과 설정을 그대로 사용하여 동일 조건에서 비교하였다. Oxford Day-and-Night은 GS-CPR 공개 구현에 장면 구성 절차가 포함되어 있지 않아, GS-CPR과 동일한 절차를 적용해 장면을 재구성하였다. 또한 일부 장면은 GPU 메모리 제약으로 학습 단계에서 OOM이 발생하여 제외하였으며, 평가 대상은 지표 산출 이전에 OOM 미발생 여부만을 기준으로 결정하였다.

표 1. Cambridge Landmarks 데이터셋에서의 자세 추정 성능 비교

Scene	Method	Med. Trans. (↓)	Med. Rot. (↓)
OldHospital	3DGS	0.31	0.65
	GS-CPR	0.19	0.38
	Ours	0.18	0.35
KingsCollege	3DGS	0.28	0.46
	GS-CPR	0.16	0.26
	Ours	0.21	0.28
ShopFacade	3DGS	0.06	0.33
	GS-CPR	0.06	0.24
	Ours	0.05	0.25

표 2. Oxford 데이터셋에서의 자세 추정 성능 비교

Scene	Method	Med. Trans. (↓)	Med. Rot. (↓)
HB Allen Centre	GS-CPR	2.99	102.3
	Ours	2.09	96.7
Robotics Institute	GS-CPR	4.56	133.9
	Ours	4.67	126.5

4.2 정량적 평가

표 1은 Cambridge Landmarks에서의 결과이다. Ours는 OldHospital에서 GS-CPR 대비 이동 오차가 소폭 감소하는 경향을 보였고(0.19m→0.18m), ShopFacade에서는 두 방법이 유사한 성능을 나타냈다. 반면 KingsCollege에서는 Ours의 오차가 GS-CPR 대비 증가하는 경향이 관찰되었으며, 이는 4.4 절에서 논의하듯 외관 변화량(AS)에 따라 제로 임베딩 적용 여부의 적합성이 달라질 수 있음을 시사한다.

표 2는 Oxford Day-and-Night에서의 결과이다. HB Allen Centre에서는 GS-CPR 대비 이동/회전 오차가 감소하였다(2.99m→2.09m, 102.3°→96.7°). 반면 Robotics Institute에서는 두 방법 모두 큰 오차를 보여, 강한 외관 변화 조건에서 렌더링 기반 자세 정제의 성능 한계가 남아 있음을 확인하였다.

4.3 정성적 평가

그림 2는 Oxford HB Allen Centre에서의 렌더링 결과를 정답 영상과 함께 비교한 것이다. GS-CPR은 조명 분포가 단일 전역 보정으로 충분히 반영되지 않아 천장 조명 및 벽면 밝기 변화가 상대적으로 뭉개지는 반면, Ours는 픽셀 단위 보정을 통해 조명 분포가 더 선명하게 나타나 천장 라인과 구조 윤곽이 비교적 명확히 드러난다. 이는 동일 자세에서 생성된 렌더링임에도, 외관 처리 방식에 따라 매칭에

표 3. 장면별 외관 변화량에 따른
제로 임베딩 적용 성능 비교

Scene	AS	Med. Trans. (↓)	
		w/ Zero Emb	w/o Zero Emb
HB Allen Centre	0.5	2.09	2.21
OldHospital	0.17	0.18	0.20
ShopFacade	0.05	0.05	0.05
KingsCollege	0.04	0.21	0.16

유리한 구조 단서의 가시성이 달라질 수 있음을 보여준다.

4.4 제로 임베딩 효과 분석

본 절에서는 외관 변화 수준에 따라 제로 임베딩의 효과가 달라지는지를 분석한다. 이를 위해 외관 변화량 AS(Appearance Shift)를 정의하고, 장면별 AS와 제로 임베딩 적용 여부에 따른 이동 오차를 비교하였다. AS는 훈련/평가 이미지 집합을 HSV로 변환한 뒤 H-S 2차원 히스토그램(각 30 bins)을 계산하고, 정규화된 평균 히스토그램 간 코사인 거리로 산출한다(0에 가까울수록 유사, 1에 가까울수록 상이).

표 3에서 AS가 큰 장면(HB Allen Centre, OldHospital)에서는 제로 임베딩 적용 시 더 낮은 이동 오차가 관찰되었다. 이는 추론 시 외관 임베딩을 $z=0$ 으로 고정함으로써 테스트 이미지의 조명·색상 편향이 렌더링 외관에 과도하게 반영되는 것을 제한하고, 매칭 및 PnP 단계에서 구조 단서가 상대적으로 우선되도록 하기 때문으로 해석할 수 있다. 반대로 AS가 작은 장면(KingsCollege, ShopFacade)에서는 제로 임베딩을 적용하지 않았을 때 동등하거나 더 우수한 결과가 나타났으며, 외관 정보가 유효한 조건에서는 해당 제약이 정보 손실로 작동할 수 있음을 시사한다.

4.5 구성요소 제거(ablation) 분석: 기하학적 정규화 및 변환 제약의 기여

표 4는 HB Allen Centre에서 Ours의 핵심 구성요소 기여도를 분석한 결과이다. Ours(Full)의 중위수 이동 오차는 2.09m이며, 기하학적 정규화를 제거할 경우 2.64m로 증가하였다(Δ vs Ours +26.3%). 이는

표 4. HB Allen Centre 장면에서의 구성요소
제거(ablation) 실험

구성	Med. Trans. (↓)	Δ vs Ours (%)
Ours(Full)	2.09	-
w/o Geometric Regularization	2.64	+26.3%
w/ Full Affine (w/o Diagonal Constraint)	2.43	+16.3%

3.3 절의 정규화/안정화 전략이 외관 보정 과정에서의 기하 왜곡을 억제하는 데 기여함을 보여준다. 또한 채널 혼합을 허용하는 변환(Full Affine)으로 제약을 완화하면 오차가 2.43m로 증가하였다(Δ vs Ours +16.3%). 이는 조명 변화가 큰 조건에서 채널 혼합이 학습/최적화 안정성을 저해하거나 매칭에 필요한 구조 단서를 약화시킬 수 있음을 시사하며, 본 연구의 채널 독립(대각) 제약이 안정화에 유효함을 뒷받침한다.

4.6 고찰: 외관 변화(AS) 환경에서의 유효 조건과 한계

표 1-2와 그림 2는 Ours가 일부 장면에서 오차 감소 및 구조 단서 가시성 개선을 보일 수 있음을 보여준다. 표 3의 AS 분석은 제로 임베딩이 외관 변화가 큰 조건에서 유리하며, 변화가 작은 조건에서는 미적용이 더 적합할 수 있음을 시사한다. 표 4의 제거 실험은 기하학적 정규화와 채널 독립(대각) 제약이 성능 안정화에 핵심임을 확인한다. 다만 Oxford의 일부 장면에서는 두 방법 모두 큰 오차가 관찰되어, 강한 외관 변화 조건에서 렌더링 기반 자체 정제의 성능 한계가 남아 있음을 확인하였다.

5. 결론

본 논문에서는 3D 가우시안 스플래팅 기반의 시각 기반 자세 추정에서 외관 모델링이 야기할 수 있는 기하학적 붕괴 문제를 분석하고, 이를 해결하기 위한 제로 임베딩 추론 전략과 학습 안정화 기법을 제안하였다. 실험 결과, 제안 방법은 조명 변화가 큰 장면에서 기존 GS-CPR 대비 자세 정제 오차를 감소시키는 경향을 보였으며, AS 분석과 제거 실험을 통해 제로 임베딩 및 기하 정규화, 대각 제약의 기여를 확인하였다.

향후 연구로서는 앞서 논의된 외관 분리 및 정규화 과정에서의 정보 손실 문제를 극복하기 위해, 저조도 향상(Low-Light Enhancement) 기술을 도입하거나 생성형 모델을 활용하여 도메인 간의 스타일 차이를 능동적으로 해소하는 방향을 모색할 예정이다. 이를 통해 3DGS 기반 자세 추정 기술의 실용성을 향상시킬 수 있을 것으로 기대된다.

감사의 글

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 생성 AI 선도인재양성사업(IITP-2024-RS-2024-00397085)과 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(No.RS-2023-00212484, 복잡한 실제 주행환경에서 설명 가능한 움직임 예측)을 받아 수행된 연구임.

참고문헌

- [1] J. Miao, K. Jiang, T. Wen, Y. Wang, P. Jia, B. Wijaya, X. Zhao, Q. Cheng, Z. Xiao, *et al.*, “A survey on monocular re-localization: From the perspective of scene map representation,” *IEEE Transactions on Intelligent Vehicles*, 2024.
- [2] P.-E. Sarlin, C. Cadena, R. Siegwart, and M. Dymczyk, “From coarse to fine: Robust hierarchical localization at large scale,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12716–12725, 2019.
- [3] A. Kendall, M. Grimes, and R. Cipolla, “Posenet: A convolutional network for real-time 6-dof camera relocalization,” in *IEEE International Conference on Computer Vision (ICCV)*, pp. 2938–2946, 2015.
- [4] E. Brachmann and C. Rother, “Visual camera relocalization from rgb and rgb-d images using dsac,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 44, no. 9, pp. 5847–5865, 2021.
- [5] T. Sattler, Q. Zhou, M. Pollefeys, and L. Leal-Taixe, “Understanding the limitations of cnn-based absolute camera pose regression,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3302–3312, 2019.
- [6] T. Sattler, W. Maddern, C. Toft, A. Torii, L. Hammarstrand, *et al.*, “Benchmarking 6dof outdoor visual localization in changing conditions,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8601–8610, 2018.
- [7] S. Chen, Y. Bhalgat, X. Li, J.-W. Bian, K. Li, Z. Wang, and V. A. Prisacariu, “Neural refinement for absolute pose regression with feature synthesis,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20987–20996, 2024.
- [8] 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,” *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
- [9] L. Yen-Chen, P. Florence, J. T. Barron, A. Rodriguez, P. Isola, and T.-Y. Lin, “inerf: Inverting neural radiance fields for pose estimation,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1323–1330, IEEE, 2021.
- [10] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, “3d gaussian splatting for real-time radiance field rendering,” *ACM Transactions on Graphics (SIGGRAPH)*, vol. 42, no. 4, pp. 139:1–139:9, 2023.
- [11] C. Liu, S. Chen, Y. Bhalgat, S. Hu, M. Cheng, Z. Wang, *et al.*, “Gs-cpr: Efficient camera pose refinement via 3d gaussian splatting,” *arXiv preprint arXiv:2408.11085*, 2024.
- [12] R. Martin-Brualla, N. Radwan, M. S. Sajjadi, J. T. Barron, A. Dosovitskiy, and D. Duckworth, “Nerf in the wild: Neural radiance fields for unconstrained photo collections,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7210–7219, 2021.
- [13] J. Lin, Z. Li, B. Huang, X. Tang, J. Liu, S. Liu, X. Wu, *et al.*, “Decoupling appearance variations with 3d consistent features in gaussian splatting,” vol. 39, pp. 5236–5244, 2025.
- [14] V. Leroy, Y. Cabon, and J. Revaud, “Grounding image matching in 3d with mast3r,” in *European Conference on Computer Vision (ECCV)*, pp. 71–91, Springer, 2024.
- [15] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng, “Complete solution classification for the perspective-three-point problem,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 25, no. 8, pp. 930–943, 2003.
- [16] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [17] M. Tancik, V. Casser, X. Yan, S. Pradhan, B. Mildenhall, P. P. Srinivasan, *et al.*, “Block-nerf: Scalable large scene neural view synthesis,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8248–8258, 2022.
- [18] E. Brachmann, T. Cavallari, and V. A. Prisacariu, “Accelerated coordinate encoding: Learning to relocalize in minutes using rgb and poses,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5044–5053, 2023.
- [19] Z. Wang, W. Bian, X. Li, Y. Tao, J. Wang, M. Fallon, and V. A. Prisacariu, “Seeing in the dark: Benchmarking egocentric 3d vision with the oxford day-and-night dataset,” *arXiv preprint arXiv:2506.04224*, 2025.