JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 27, No. 6, pp. 1477-1486, Jun. 2026

Check for updates

XR 스마트글래스를 이용한 3D Gaussian Splatting 기반 디지털 트윈의 재관측을 통한 모델 정밀화 파이프라인

김 정 현[1] · 이 제 민[1] · 강 형 준[2] · 이 동 희[2] · 류 훈[3] · 김 영 원[3*]
[1]국립금오공과대학교 소프트웨어공학과 석사과정
[2]국립금오공과대학교 컴퓨터공학부 소프트웨어전공 학사과정
[3]국립금오공과대학교 컴퓨터공학부 교수

Re-Observation-Based Model Refinement Pipeline for 3D Gaussian Splatting Digital Twins Using XR Smart Glasses

Jeonghyeon Kim[1] · Jemin Lee[1] · Hyeongjun Kang[2] · Donghee Lee[2] · Hoon Ryu[3] · Youngwon Kim[3*]

[1]Master's Course, Department of Software Engineering, Kumoh National Institute of Technology, Gumi 39177, Korea
[2]Undergraduate Student, Software Major, School of Computer Engineering, Kumoh National Institute of Technology, Gumi 39177, Korea
[3]Professor, School of Computer Engineering, Kumoh National Institute of Technology, Gumi 39177, Korea

[요 약]

디지털 트윈은 물리적 환경의 가상 복제본으로 다양한 산업 분야에서 활용이 확대되고 있으나, 초기 구축 시 제한된 시점으로 인해 특정 방향에서의 렌더링 충실도가 부족한 경우가 빈번하다. 본 연구에서는 XR 글래스로 촬영한 신규 시점 영상을 활용하여 3D Gaussian Splatting(3DGS) 기반 디지털 트윈의 모델 충실도를 개선하는 end-to-end 파이프라인을 제안한다. 제안 파이프라인은 키프레임 선별, 시각 위치 추정, 정밀화 필요 영역 판단, 3DGS 모델 정밀화, 분리 평가의 5단계로 구성된다. 두 개의 실내 환경에서 실험한 결과, 점진적 정밀화를 통해 신규 시점의 PSNR이 19~20 dB에서 30 dB 이상으로 향상되었으며, 기존 시점 품질 저하는 1.36 dB 이내로 억제되었다. 전체 재학습은 신규 시점에서 2.11~2.35 dB 우수한 반면 기존 시점 보존에서는 유사한 수준을 보여, 3DGS의 명시적 표현에서는 학습 전략에 무관하게 catastrophic forgetting이 제한적임을 확인하였다.

[Abstract]

Digital twins are virtual replicas of physical environments with growing industrial applications; however, their initial construction often suffers from limited rendering fidelity in under-observed directions owing to restricted viewpoints. This study proposes an end-to-end pipeline that improves the model fidelity of 3D Gaussian splatting (3DGS)-based digital twins using a newly observed viewpoint video from XR glasses. The pipeline comprises five stages: keyframe selection, visual localization, refinement necessity detection, 3DGS model refinement, and separate evaluation. Experiments on two indoor environments showed that incremental refinement improved the new-viewpoint PSNR from 19–20 dB to over 30 dB while limiting the existing viewpoint degradation to within 1.36 dB. Full retraining achieved 2.11–2.35 dB higher new-viewpoint quality but showed comparable existing-view preservation, suggesting that the explicit representation of 3DGS inherently limits catastrophic forgetting, regardless of the training strategy.

색인어 : 3D 가우시안 스플래팅, 디지털 트윈, 점진적 학습, 모델 정밀화, XR 스마트글래스

Keyword : 3D Gaussian Splatting, Digital Twin, Incremental Learning, Model Refinement, XR Smart Glasses

http://dx.doi.org/10.9728/dcs.2026.27.6.1477

Received 16 April 2026; **Revised** 13 May 2026
Accepted 20 May 2026

***Corresponding Author; Youngwon Kim**

Tel: +82-54-478-7542
E-mail: kim01@kumoh.ac.kr

Ⅰ. 서 론

디지털 트윈(Digital Twin)은 물리적 자산이나 환경의 가상 복제본으로, 건설 현장 모니터링, 시설물 유지보수, 도시 계획 등 다양한 분야에서 핵심 기술로 부상하고 있다[1],[2]. 디지털 트윈의 실용적 가치를 위해서는 모델의 높은 충실도가 확보되어야 하나, 초기 구축 시 제한된 시점에서만 데이터를 수집하게 되어 관측 밀도가 낮은 방향에서의 렌더링 품질이 부족한 경우가 빈번하다. 이를 보완하기 위해서는 동일 장면의 재관측 시점 데이터를 점진적으로 추가하여 디지털 트윈의 모델 충실도를 개선하는 정밀화(refinement) 과정이 필수적이다.

기존의 디지털 트윈 구축 및 유지보수 방법은 주로 LiDAR 스캐닝이나 전문 사진측량 장비에 의존한다. 이러한 방법은 높은 정밀도를 제공하지만, 장비 비용이 고가이며 전문 인력이 필요하고 데이터 수집에 상당한 시간이 소요된다. 특히 건설 현장이나 제조 시설처럼 지속적인 모델 유지보수가 필요한 환경에서는 이러한 비용과 시간이 실질적인 장벽으로 작용하며, 반복적 재관측이 요구되는 환경에서는 비용이 누적되어 더욱 비효율적이다.

최근 Neural Radiance Fields(NeRF)[3]와 3D Gaussian Splatting(3DGS)[4],[5]으로 대표되는 뉴럴 렌더링 기술은 일반 카메라 영상만으로 고품질 3D 장면 재구성을 가능하게 하였다. 특히 3DGS는 명시적 점 기반 표현을 사용하여 NeRF 대비 수십 배 빠른 렌더링 속도를 달성하면서도 동등 이상의 시각적 품질을 제공한다. 그러나 기존 3DGS 연구는 주로 정적 장면의 일회성 재구성에 초점을 맞추고 있으며, 초기 구축 이후 신규 시점의 재관측을 통해 모델을 점진적으로 정밀화하는 연구는 아직 부족한 실정이다.

한편, Meta Ray-Ban Display와 같은 소비자용 XR 글래스의 보급이 확대되면서, 현장 작업자가 별도의 장비 없이 일상적으로 착용한 글래스만으로 데이터를 수집할 수 있는 환경이 조성되고 있다. XR 글래스는 핸즈프리로 자연스러운 시점에서 영상을 촬영할 수 있어, 기존 수동 촬영 방식의 한계를 극복할 잠재력을 가진다.

본 연구에서는 XR 글래스 영상을 활용한 3DGS 기반 디지털 트윈의 재관측 기반 모델 정밀화 파이프라인을 제안한다. 제안 시스템은 현장 작업자가 XR 글래스로 신규 경로를 촬영한 영상이 모바일 기기를 거쳐 서버로 자동 전송되면, 키프레임 선별부터 3DGS 모델 정밀화 및 품질 평가까지의 전 과정을 자동으로 수행한다. 본 연구의 주요 기여는 다음과 같다. 첫째, XR 글래스 촬영부터 디지털 트윈 모델 정밀화까지 전 과정을 자동화하는 5단계 end-to-end 파이프라인을 설계하고 구현하였다. 둘째, 점진적(incremental) 정밀화 방식과 전체 재학습(full retrain) 방식을 비교 분석하여, 3DGS의 명시적 표현에서는 학습 전략에 무관하게 기존 시점 품질이 잘 보존됨을 실험적으로 확인하고, 두 전략의 렌더링 품질, 지각적 품질, 학습 시간을 정량적으로 비교 분석하였다. 셋째, 키프레임 선별 기준, 학습 반복 횟수, 정밀화 전략 등 주요 설계 요소가 최종 렌더링 품질에 미치는 영향을 정량적으로 분석하였다.

Ⅱ. 관련 연구

2-1 3D Gaussian Splatting

3D Gaussian Splatting[4]은 장면을 수백만 개의 3D 가우시안 원시체(primitive)로 표현하는 명시적 장면 표현 방법이다. 각 가우시안은 위치(mean), 공분산(covariance), 불투명도(opacity), 구면조화함수(spherical harmonics) 계수로 구성된다. NeRF[3]가 암묵적 신경망을 사용하여 볼륨 렌더링을 수행하는 것과 달리, 3DGS는 미분 가능한 타일 기반 래스터라이저를 통해 실시간 렌더링을 달성한다. 우성현과 서정일[5]은 NeRF 계열과 3DGS 계열 모델의 성능을 비교하여, 3DGS가 학습 속도와 재구성 품질 면에서 우수함을 보고하였다. 또한 강희원과 김지윤[6]은 3DGS의 3D 모델링 적용 가능성을 분석하였으며, 김준형과 이동엽[7]은 Unity 기반 디지털 조형물 제작 워크플로우를 제안하였다.

3DGS의 학습 과정은 Structure-from-Motion(SfM)으로 추출한 희소 점군(sparse point cloud)에서 초기화되며, 이후 적응적 밀도 제어(adaptive density control)를 통해 가우시안의 수와 크기를 조절한다. 이러한 명시적 표현의 장점은 점군의 추가·삭제가 용이하여 장면의 점진적 확장 및 수정에 적합하다는 것이다.

2-2 디지털 트윈 구축 및 유지보수

디지털 트윈의 구축과 유지관리는 초기 모델링과 이후의 점진적 품질 개선으로 구분할 수 있다. 전통적인 방법은 LiDAR 스캐닝과 점군 정합을 통해 3D 모델을 구축하고, 추가 스캔으로 모델 품질을 보완한다. 오성택 등[8]은 디지털 트윈 구현을 위한 3D 점군 기반 맵 모델링 프레임워크를 제안하였으며, 윤재석과 김진민[2]은 디지털 트윈 관련 연구 및 표준화 동향을 분석하였다.

최근에는 뉴럴 렌더링을 활용한 디지털 트윈 구축 연구가 등장하고 있다. Block-NeRF[9]는 대규모 도시 환경을 블록 단위로 분할하여 독립적으로 구축하고 확장할 수 있는 구조를 제안하였다. 이순주와 김지윤[10]은 NeRF 상용화가 3D 콘텐츠 제작에 미치는 영향을 분석하였다. 그러나 이러한 연구들은 NeRF 기반으로 렌더링 속도와 학습 시간의 한계가 있으며, 3DGS 기반의 재관측을 통한 점진적 모델 정밀화에 대한 연구는 아직 초기 단계이다.

2-3 시각 위치 추정

시각 위치 추정(Visual Localization)은 영상의 6자유도(6-DoF) 카메라 포즈를 추정하는 문제이다. 계층적 위치 추정(Hierarchical Localization, hloc)[11]은 전역 기술자(global descriptor)를 사용한 영상 검색과 국소 특징(local feature) 매칭을 결합하는 대표적인 파이프라인이다. SuperPoint[12]는 자기 지도 학습으로 훈련된 국소 특징 검출기이며, SuperGlue[13]는 그래프 신경망 기반의 특징 매칭 방법으로, 전통적인 최근접 이웃 매칭 대비 높은 정합 정확도를 보인다. 박승기 등[14]은 SuperPoint를 포함한 특징 추출 기반 이미지 매칭 기술의 동향을 분석하여, 딥러닝 기반 방법이 전통적 방법 대비 우수함을 보고하였다. NetVLAD[15]는 장소 인식을 위한 전역 기술자로, 대규모 영상 데이터베이스에서 유사한 장면을 효율적으로 검색한다.

2-4 XR 글래스 기반 데이터 수집

소비자용 XR 글래스는 카메라, IMU, 디스플레이를 통합한 착용형 장치로, 핸즈프리 데이터 수집을 가능하게 한다. Meta Ray-Ban 글래스는 12 MP 카메라와 1080p 영상 촬영 기능을 제공하며, Bluetooth 및 Wi-Fi를 통해 모바일 기기와 연동된다. 기존 연구에서 XR 글래스를 3D 재구성의 데이터 소스로 활용한 사례는 제한적이며, 글래스 특유의 영상 특성(모션 블러, 제한된 해상도, 자동 노출 변동 등)에 대한 체계적인 분석이 부족하다. 국내에서는 Visual SLAM 기반 모바일 증강현실 시스템 구축 연구[16]와 ORB-SLAM 재지역화 성능 개선 연구[17], 이병권 등[18]의 COLMAP을 활용한 문화재 3D 재구성 연구 등이 수행되었으나, XR 글래스를 뉴럴 렌더링 기반 3D 재구성의 데이터 소스로 활용한 사례는 전무하다.

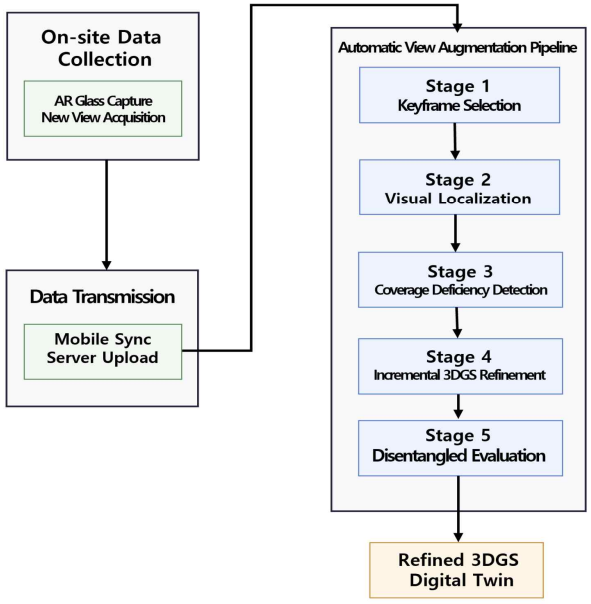

그림 1. 시스템 전체 개요도
Fig. 1. System-wide overview

Ⅲ. 제안 방법

3-1 시스템 개요

제안 시스템은 그림 1과 같이 현장 데이터 수집부, 데이터 전송부, 서버 정밀화 파이프라인의 3개 계층으로 구성된다. 현장 작업자는 XR 글래스를 착용하고 기존과 다른 경로나 각도로 동일 장면을 약 1분간 재관측한다. 촬영된 영상은 모바일 기기를 통해 서버로 자동 전송되며, 서버에서는 5단계의 자동 모델 정밀화 파이프라인이 실행된다.

구체적으로 XR 글래스에서 촬영된 영상은 Bluetooth로 페어링된 스마트폰의 Meta AI 앱을 통해 동기화되어 스마트폰 갤러리에 저장된다. 갤러리에 새 영상 파일이 감지되면 Android 자동화 앱(Tasker)이 Wi-Fi 네트워크를 통해 서버로 자동 업로드하며, 서버 수신 즉시 정밀화 파이프라인이 실행된다.

3DGS 기반 디지털 트윈의 모델 정밀화 시, 베이스라인 구축과 재관측 촬영에 서로 다른 카메라(예: 스마트폰과 XR 글래스)를 사용하면 화각, 색감, 렌즈 왜곡 특성의 차이로 인해 모델 통합 품질이 크게 저하되는 문제가 발생한다. 본 연구의 예비 실험에서도 스마트폰 베이스라인과 XR 글래스 신규 영

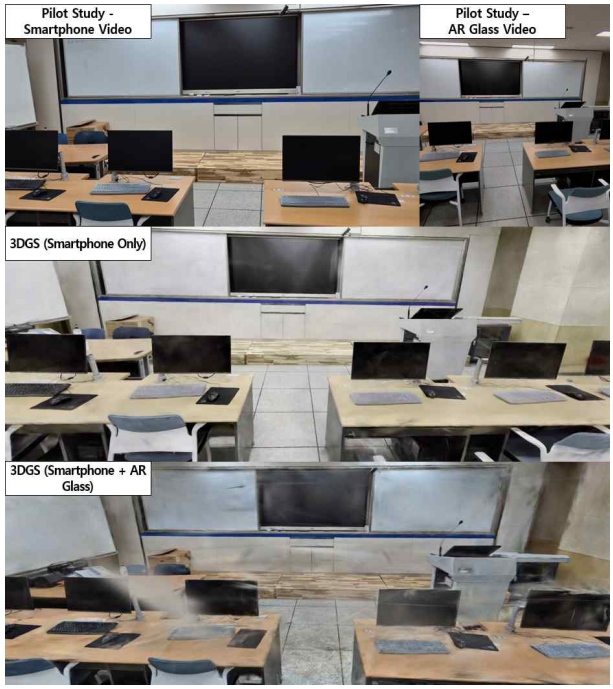

그림 2. 예비실험 Scene과 3DGS 결과
Fig. 2. Pilot experiment scene and 3DGS results

상을 혼합하여 학습한 결과, 신규 시점에서 PSNR 13.99 dB에 불과한 심각한 품질 저하가 관찰되었다(그림 2). 이러한 관찰을 바탕으로, 본 연구에서는 베이스라인과 추가 관측 영상 모두를 동일한 XR 글래스로 촬영하는 단일 카메라 파이프라인을 설계하였다.

3-2 Stage 1: 키프레임 선별

XR 글래스 영상은 착용자의 머리 움직임으로 인한 모션 블러, 연속 프레임 간 중복, 시점 다양성 부족 등의 문제를 포함한다. 고품질 키프레임을 선별하기 위해 3단계 필터링을 수행한다.

블러 필터링 단계에서는 입력 영상을 그레이스케일로 변환한 후 라플라시안 연산자를 적용하고, 그 분산을 블러 점수로 사용한다. 블러 점수가 임계값 이하이면 그 프레임을 제거하며 라플라시안의 분산이 클수록 영상 내 선명한 에지가 많음을 의미한다.

중복 제거 단계에서는 연속 프레임 간 구조적 유사도(SSIM)를 계산하여, SSIM(Structural Similarity Index) 값이 임계값 이상인 프레임을 중복으로 판정하여 제거한다. 이는 정적 장면에서 유사한 시점의 반복 프레임을 효과적으로 제거한다.

다양성 선택 단계에서는 남은 프레임에서 HSV 색상 히스토그램(32x32 bins)을 특징 벡터로 추출하고, 그리디 최대-최소 거리 선택을 수행한다. 첫 프레임을 선택한 후, 이미 선택된 프레임들과의 최소 거리가 최대인 프레임을 반복적으로 추가하여 시점 다양성을 확보한다. 거리 함수로는 카이제곱 거리를 사용한다.

3-3 Stage 2: 시각 위치 추정

새로운 키프레임의 카메라 포즈를 기존 3DGS 모델의 좌표계에서 추정하기 위해, 계층적 시각 위치 추정(hloc)[11]을 적용한다. 이 과정은 다음 5단계로 구성된다. 먼저, 기존 베이스라인 영상에서 SuperPoint[12] 국소 특징을 추출하여 데이터베이스를 구축한다. 다음으로, 신규 키프레임에서 동일하게 SuperPoint 특징을 추출한다. 이어서 NetVLAD[15]을 사용하여 각 질의(query) 영상과 유사한 상위 k개(k=20) 데이터베이스 영상을 검색한다. 그 후 SuperGlue[13]를 사용하여 질의-데이터베이스 쌍 간 국소 특징을 매칭한다.

마지막으로 PnP(Perspective-n-Point)+RANSAC을 통해 각 질의 영상의 6-DoF 카메라 포즈를 추정한다. 인라이어(inlier) 수가 최소 기준인 30 미만인 영상은 위치 추정 실패로 판정하여 제외한다.

추정된 포즈는 기존 베이스라인 SfM 모델과 동일한 좌표계를 공유하므로, 별도의 좌표 정합 없이 직접 통합이 가능하다.

3-4 Stage 3: 정밀화 필요 영역 판단

기존 모델의 충실도가 부족한 영역을 정량적으로 판단하기 위해, 기존 3DGS 모델로 신규 시점을 렌더링한 영상과 실제 촬영 영상을 비교하여 불확실성 맵을 생성한다. 전역 불확실성은 전체 영상 단위로 SSIM과 L1(평균 절대 차이)을 계산한다. 패치 기반 불확실성은 영상을 64×64 크기의 패치로 분할하고, 각 패치에 대해 불확실성 점수를 계산한다.

패치 수준에서 SSIM이 0.80 미만이거나 L1이 0.15를 초과하는 패치를 충실도 부족 패치로 판정한다. 충실도 부족 패치의 비율이 전체의 30%를 초과하면 해당 시점은 기존 모델로 충분히 표현되지 않는 영역을 포함하는 것으로 판단하여 정밀화 대상에 포함한다. 해당 임계값들은 사전 실험을 통해 경험적으로 설정하였으며, 눈에 띄는 렌더링 오류를 안정적으로 검출하면서도 불필요하게 많은 시점을 정밀화 대상으로 포함하지 않는 범위에서 선정하였다. 기준을 더 엄격하게 설정하면 작은 색상 차이나 노출 차이까지 오류로 판단할 수 있고, 기준을 더 완화하면 실제 품질이 낮은 영역이 제외될 수 있어 본 값을 사용하였다.

3-5 Stage 4: 3DGS 모델 정밀화

기존 3DGS 모델에 신규 시점을 통합하여 모델 충실도를 개선하기 위해, 본 파이프라인은 체크포인트 기반 점진적 학습(incremental training)과 전체 재학습(full retrain)을 모두 수행한다. 점진적 학습은 기존 체크포인트에서 이어서 학습하며, 전체 재학습은 통합된 데이터로 처음부터 학습한다. 데이터 통합 과정에서는 기존 베이스라인의 COLMAP 희소 모델에 새로운 키프레임의 포즈 정보를 추가한다. 이때 기존 영상과 신규 영상이 동일한 SfM 좌표계를 공유하는 것이 핵심이며, Stage 2의 시각 위치 추정을 통해 이를 보장한다.

점진적 학습 과정에서는 베이스라인 모델의 마지막 체크포인트(iteration N)에서 학습을 재개하여 추가 반복을 수행한다. 본 연구에서는 베이스라인 30,000 iterations에서 시작하여 총 60,000 iterations까지 학습한다. 학습률은 베이스라인 대비 스케일링 팩터 0.1을 적용하여, 기존에 학습된 가우시안의 급격한 변화를 방지하고 기존 시점의 품질을 보존한다. 적응적 밀도 제어는 500 iterations 간격으로 수행되며, 불투명도 초기화는 3,000 iterations 간격으로 적용하여 관측 밀도가 보강된 영역에서의 가우시안 재조정을 촉진한다. 전체 재학습의 경우 동일한 총 60,000 iterations를 기본 하이퍼파라미터로 처음부터 학습한다. 두 전략의 렌더링 품질 및 학습 시간 비교는 실험 4에서 상세히 분석한다.

3-6 Stage 5: 분리 평가

정밀화된 모델의 품질을 체계적으로 평가하기 위해, 평가 뷰를 기존 뷰(baseline views)와 신규 뷰(new views)로 분리하여 별도로 측정한다. 기존 뷰 평가는 신규 시점 통합 후

기존 시점의 품질 저하 정도를 측정하며, 신규 뷰 평가는 새로 추가된 시점에서의 렌더링 품질을 측정한다. 평가 메트릭으로는 PSNR(Peak Signal-to-Noise Ratio), SSIM(Structural Similarity Index), LPIPS(Learned Perceptual Image Patch Similarity)[19]를 사용한다. LPIPS는 AlexNet 백본을 사용하며, 입력을 [-1, 1] 범위로 정규화하여 계산한다.

Ⅳ. 실 험

4-1 실험 환경

Meta Ray-Ban Display 글래스를 사용하여 두 개의 서로 다른 실내 환경(장면 A, 장면 B)에서 영상을 촬영하였다. 모든 실험은 동일한 정적 환경에서 재관측한 데이터를 사용하며, 실험 기간 중 장면 내 물체의 추가 및 제거 등 환경 변화는 발생하지 않았다. 각 장면에서 베이스라인 영상과 신규 영상을 약 1분간 촬영하여 2 fps로 프레임을 추출하고, 블러 필터링(임계값 80)을 적용하였다. 장면 A는 베이스라인 111장, 신규 75장(원시 124장), 장면 B는 베이스라인 78장, 신규 72장(원시 76장)으로 구성하였다(표 1, 그림 3 참조). 표 1의 Baseline Frames는 초기 학습용 기존 프레임 수, New Frames는 필터링 후 정밀화에 사용한 신규 프레임 수, Raw Frames는 필터링 전 신규 영상의 전체 프레임 수를 의미한다.

표 1. 실험에 사용된 각 장면별 데이터 구성
Table 1. Data composition for each scene used in the experiments

Scene	Baseline Frames	New Frames	Raw Frames (New Video)
A	111	75	124
B	78	72	76

그림 3. Scene A, B의 원본 사진
Fig. 3. Original picture of scene A, B

서버 학습 환경은 NVIDIA RTX 5000 Ada Generation (VRAM 32 GB), AMD 프로세서를 사용하였다. 3DGS 구현은 공식 gaussian-splatting 저장소[4]를 기반으로 하며, 시각 위치 추정은 hloc[11] 라이브러리를 사용하였다. SfM은 COLMAP[20]을 활용하였다.

4-2 실험 1: 재관측 기반 모델 정밀화 효과

제안 파이프라인의 핵심 성능을 검증하기 위해, 베이스라인 모델(30k iterations)과 정밀화 모델(30k+ 30k iterations)의 렌더링 품질을 비교하였다. 베이스라인 모델은 기존 뷰에서 장면 A 35.67 dB, 장면 B 35.33 dB의 높은 렌더링 품질을 보이나, 신규 시점에서는 장면 A 19.86 dB, 장면 B 19.19 dB로 사실상 렌더링이 불충분한 수준이었다. 두 장면 모두에서 재관측 기반 점진적 정밀화가 효과적으로 동작함을 확인하였다(그림 4, 그림 5 참조). 정밀화 후 신규 뷰 PSNR은 장면 A 31.29 dB(+11.43 dB), 장면 B 30.34 dB(+11.15 dB)로 두 장면 모두에서 11 dB 이상의 대폭적인 품질 향상을 달성하였다. 기존 뷰에서의 PSNR 저하는 장면 A에서 0.70 dB, 장면 B에서 1.36 dB로, 35 dB 이상의 베이스라인 품질 대비 2~4% 수준에 불과하여 catastrophic forgetting이 제한적 수준에서 유지됨을 보인다(그림 6).

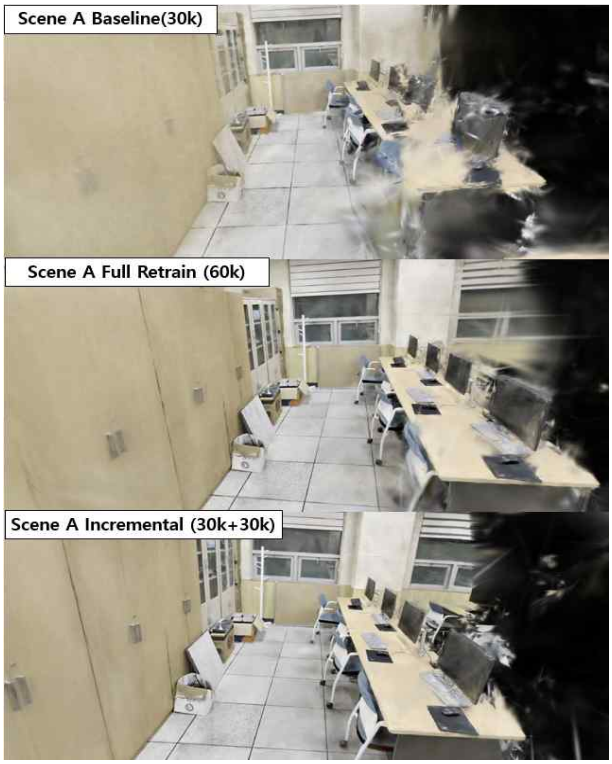

그림 4. Scene A의 실험 결과
Fig. 4. Experimental results of scene A

별도 장면에서 수행한 예비 실험에서, 카메라를 혼합하여 사용한 경우(스마트폰 227장 + XR 글래스 165장) 신규 뷰에서 PSNR 13.99 dB로 사실상 렌더링이 불가능한 수준의 품질을 보인 반면, 동일 카메라를 사용한 경우(XR 글래스 165장 + 122장) 31.73 dB로 17.74 dB의 차이를 보였다. 이는 스마트폰과 XR 글래스 간의 화각, 색감, 렌즈 왜곡 특성의 차이가 3DGS 모델 통합에 치명적인 영향을 미치며, 베이스라인과 재관측 영상을 동일 카메라로 통일하는 것이 필수적임을 실증적으로 보여준다. 이 관찰을 바탕으로 이후 모든 실험은 단일 XR 글래스 카메라로 수행하였다(그림 7).

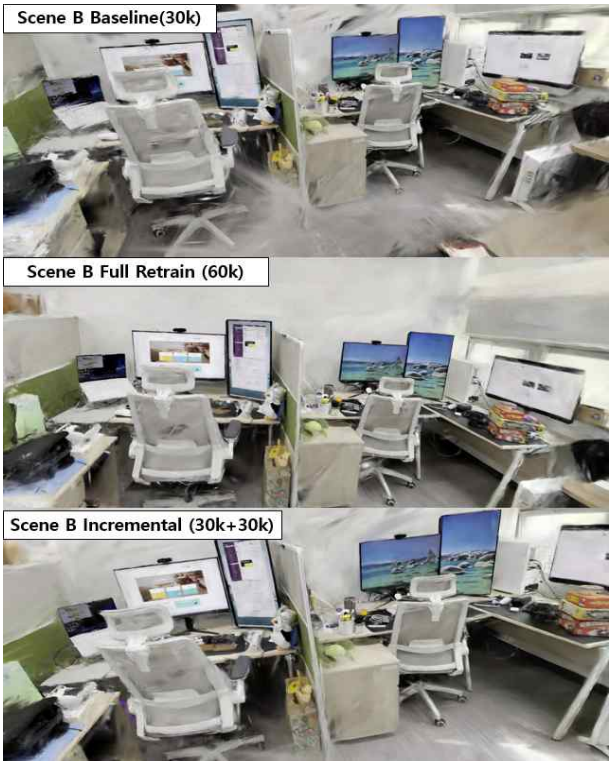

그림 5. Scene B의 실험 결과
Fig. 5. Experimental results of scene B

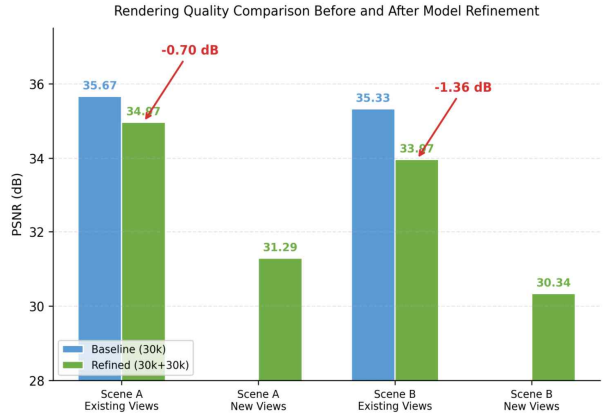

그림 6. 모델 정밀화 전후 렌더링 품질 비교(PSNR)
Fig. 6. Rendering quality comparison before and after model refinement (PSNR)

4-3 실험 2: 블러 임계값

키프레임 선별 단계의 블러 임계값이 정밀화 품질에 미치는 영향을 분석하기 위해, 장면 A에서 세 가지 임계값(60, 80, 100)으로 실험을 수행하였다. 세 임계값 간 신규 뷰 PSNR 차이는 0.14 dB(31.22~31.36 dB) 이내로 크지 않았다. 이는 블러 필터링의 역할이 극단적으로 흐린 프레임을 제거하는 데에 있으며, 임계값의 미세 조정보다는 적용 여부 자체가 중요함을 시사한다. 관대한 임계값(60)은 90장의 키프레임으로 가장 다양한 시점을 확보하여 신규 뷰에서 31.36 dB로 소폭 우수하였으나, 기존 뷰 하락이 0.85 dB로 가장 컸다. 엄격한 임계값(100)은 기존 뷰 보존(−0.55 dB)에서 가장 우수하였으나 키프레임이 62장으로 감소하여 시점 다양성이 부족해졌다. 임계값 80은 기존 뷰 보존과 시점 다양성의 균형점으로 판단된다(그림 8 참조).

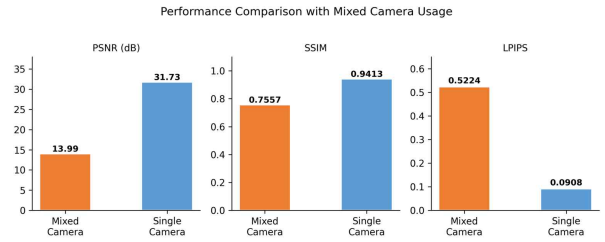

그림 7. 카메라 혼합 사용 시 신규 뷰 렌더링 품질 비교 (예비 실험, 별도 장면)
Fig. 7. New-view rendering quality comparison with mixed camera usage (preliminary experiment, separate scene)

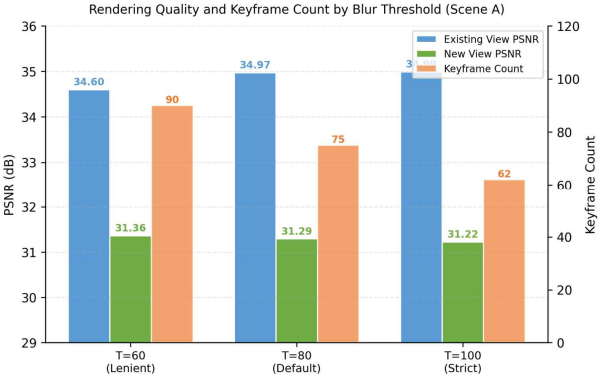

그림 8. 블러 임계값에 따른 렌더링 품질 및 키프레임 수 변화 (장면 A)
Fig. 8. Rendering quality and keyframe count variation according to blur threshold (scene A)

4-4 실험 3: 학습 반복 횟수에 따른 품질-시간 트레이드오프

점진적 정밀화 시 추가 학습 반복 횟수가 렌더링 품질에 미치는 영향을 분석하기 위해, 5,000 iterations 간격으로 중간 체크포인트를 저장하고 각각 평가하였다. 두 장면에서 동일한 경향이 관찰되었다. 신규 뷰 PSNR은 학습 반복 횟수에 따라 단조 증가하며, 35k에서 45k 구간에서 큰 향상(장면 A +1.49 dB, 장면 B +1.18 dB)이 이루어진 반면, 45k에서 60k 구간에서는 증가율이 둔화되었다(장면 A +0.75 dB, 장면 B +0.83 dB). 기존 뷰 PSNR은 두 장면 모두에서 전 구간에 걸쳐 안정적으로 유지되었다. 55k에서 60k 구간의 평균 신규 뷰 향상이 +0.23 dB로 미미한 점을 고려하면, 55k~60k iterations가 품질과 학습 시간 간의 실용적 최적점으로 판단된다(그림 9 참조).

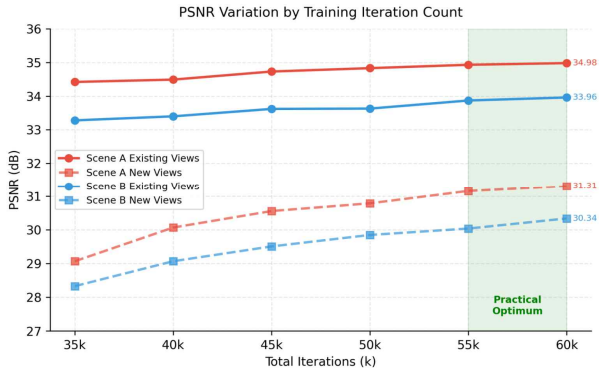

그림 9. 학습 반복 횟수에 따른 PSNR 변화
Fig. 9. PSNR variation according to training iteration count

4-5 실험 4: 전체 재학습 vs. 점진적 정밀화

점진적 정밀화와 전체 재학습의 특성을 비교하기 위해, 동일한 데이터에 대해 처음부터 학습하는 전체 재학습(full retrain, 60k iterations)과 비교하였다. Full retrain은 신규 뷰에서 장면 A 2.11 dB, 장면 B 2.35 dB(평균 2.23 dB) 높은 품질을 달성하였다. 기존 뷰에서는 장면 A의 경우 incremental이 0.36 dB 우수하였으나, 장면 B에서는 full retrain이 0.14 dB 우수하여 두 전략 간 차이가 미미하였다. LPIPS 역시 full retrain이 장면 A 0.076, 장면 B 0.056으로 incremental(장면 A 0.095, 장면 B 0.078) 대비 지각적 품질에서도 우위를 보였다. 주목할 점은 기존 뷰에서의 PSNR 차이가 최대 0.36 dB에 불과한 반면, 신규 뷰 차이는 평균 2.23 dB로 뚜렷하다는 것이다. 이는 3DGS의 명시적 표현 특성상, full retrain으로 전체 데이터를 처음부터 학습하더라도 기존 시점의 구조가 비교적 잘 유지됨을 의미한다. 즉, NeRF에서 우려되는 심각한 catastrophic forgetting이 3DGS에서는 학습 전략에 무관하게 제한적으로 나타남을 확인하였다(그림 10 참조).

학습 시간 측면에서는, 장면 B(150장)에서 incremental이 약 37분으로 full retrain(약 1시간 35분) 대비 약 61% 빠른 반면, 장면 A(186장)에서는 incremental이 약 1시간 46분으로 full retrain(약 1시간 45분)과 유사하였다. 이는 incremental이 이미 densify된 대량의 가우시안에서 시작하여 iteration당 연산 비용이 높아지기 때문으로, 데이터 규모가 커질수록 시간 이점이 감소함을 보인다. 종합하면, 3DGS 기반 디지털 트윈 정밀화에서는 두 전략 모두 기존 뷰를 잘 보존하며, full retrain은 신규 시점 품질에서 우수하고 incremental은 데이터 규모가 작을 때 학습 시간에서 유리하여, 운용 환경의 품질 요구와 시간 제약에 따라 적절한 전략을 선택할 수 있다.

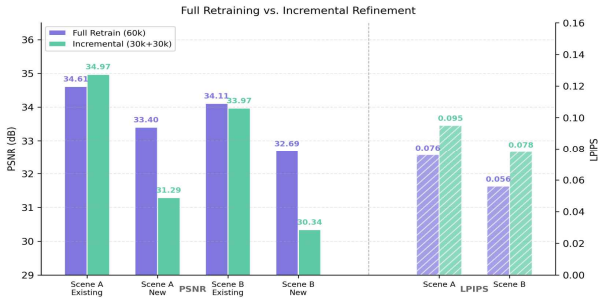

그림 10. 전체 재학습과 점진적 정밀화의 렌더링 품질 비교
Fig. 10. Rendering quality comparison between full retraining and incremental refinement

V. 고 찰

5-1 파이프라인의 실용성

제안 파이프라인은 현장 작업자의 관점에서 최소한의 조작으로 디지털 트윈의 모델 정밀화를 수행할 수 있도록 설계되었다. 작업자는 XR 글래스를 착용하고 기존과 다른 경로로 약 1분간 관심 영역을 둘러보는 것만으로 신규 시점 데이터 수집이 완료되며, 이후 모바일 기기에서 동기화 버튼 한 번으로 서버 전송 및 파이프라인 실행이 자동으로 이루어진다. 서버 측 정밀화 단계는 데이터 규모에 따라 약 37분에서 1시간 46분이 소요되어, 하루 단위의 점진적 충실도 개선에 적합하다.

5-2 카메라 통일의 중요성

본 연구의 예비 실험에서 스마트폰으로 촬영한 베이스라인과 XR 글래스로 촬영한 신규 영상을 혼합하여 학습한 결과, 신규 시점에서 PSNR 13.99 dB에 불과한 심각한 품질 저하가 발생하였다. 이는 두 카메라 간의 화각(FOV), 색 재현 특성, 렌즈 왜곡 모델의 차이가 SfM 좌표계 통합과 3DGS 학습 모두에 부정적 영향을 미쳤기 때문으로 분석된다. 반면, 베이스라인과 재관측 영상 모두를 동일한 XR 글래스로 촬영한 경우 31.73 dB로 17.74 dB의 개선을 달성하였다. 이 결과는 3DGS 기반 모델 정밀화에서 카메라 일관성이 핵심 조건임을 시사하며, 본 파이프라인이 단일 XR 글래스 기반으로 설계된 근거이다.

5-3 기존 뷰 품질 저하 분석

점진적 정밀화 시 기존 뷰의 품질 저하(장면 A 0.70 dB, 장면 B 1.36 dB)는 신규 시점 영상이 학습에 포함되면서 가우시안 분포가 조정되기 때문이다. 이는 NeRF 분야에서 보고된 catastrophic forgetting 현상과 유사하나, 3DGS의 명시적 표현 특성상 그 정도가 제한적이다. 개별 뷰 단위 분석에서도 이를 뒷받침하는 결과를 확인하였다. 장면 A에서는 정밀화 후 기존 뷰의 최솟값이 29.29 dB에서 31.30 dB로 오히려 상승하여, 최악 뷰에서조차 품질이 개선되었다. 장면 B에서도 최솟값 하락이 0.52 dB(29.59에서 29.07 dB)에 그쳤으며, 이는 특정 뷰에서 극단적 품질 하락 없이 전체적으로 균일하게 소폭 하락하는 양상임을 보인다. 전체 재학습과의

비교에서, 기존 뷰 PSNR은 장면 A에서 incremental이 0.36 dB 우수하고 장면 B에서는 full retrain이 0.14 dB 우수하여, 우위 방향이 장면에 따라 달라지면서도 그 차이는 최대 0.36 dB에 불과하였다. 이는 3DGS의 명시적 표현이 학습 전략에 무관하게 기존 시점의 구조를 안정적으로 유지함을 시사한다.

다만 LPIPS에서는 full retrain이 두 장면 모두에서 우수하여(장면 A 0.076 vs 0.095, 장면 B 0.056 vs 0.078), PSNR 상의 미미한 차이와 달리 지각적 품질에서는 전체 재학습이 일관된 우위를 보였다. 이는 incremental 방식이 기존 가우시안 위에서 이어 학습하는 과정에서 재구성 과정에서 발생한 미세한 왜곡이 누적될 수 있음을 시사하며, 지각적 품질 면에서는 full retrain이 유리함을 시사한다. 장면 B에서 기존 뷰 하락이 더 큰 것은 베이스라인 대비 신규 영상의 비율이 높아(베이스라인:신규 = 78:72 vs 111:75) 가우시안 분포의 조정 폭이 컸기 때문으로 분석된다. 베이스라인과 신규 데이터의 비율 조절이나 영역별 가중치 적용을 통해 추가 개선이 가능할 것으로 판단된다.

5-4 XR 글래스 영상의 특성과 한계

XR 글래스 영상은 다음과 같은 특성을 보인다. 장면 A의 신규 원시 프레임 124장 중 착용자의 머리 움직임으로 인해 약 40%(49장, 임계값 80 기준)의 프레임에서 블러가 관찰되었으며, 키프레임 선별을 통해 효과적으로 제거하였다. 1080p 해상도는 전문 카메라 대비 낮으나, 3DGS 학습에는 충분한 수준이다. 핸즈프리 촬영으로 인해 사람의 자연스러운 시선 경로를 따르는 시점이 확보되어, 실제 사용 시나리오에 부합하는 렌더링 품질 평가가 가능하다.

5-5 한계점 및 향후 연구

본 연구는 다음과 같은 한계를 가진다. 첫째, 두 개의 실내 환경에서 일관된 결과를 확인하였으나, 실외나 대규모 공간 등 보다 다양한 환경에 대한 일반화 검증이 필요하다. 둘째, 본 연구는 동일한 정적 장면에 대한 신규 시점 확보를 통한 모델 충실도 개선에 초점을 맞추고 있으며, 시간 경과에 따른 환경 변화(물체 추가/제거 등)의 감지 및 반영은 다루지 않는다. 셋째, XR 글래스에서 서버까지의 완전 자동 전송은 Meta AI 앱의 API 제약으로 인해 한 번의 수동 동기화 조작이 필요하다. 넷째, XR 글래스 영상의 품질이 극도로 낮은 경우 파이프라인 적용이 제한된다. 추가로 확인한 사례에서 신규 영상의 블러 점수가 전반적으로 낮아(라플라시안 분산 평균 32.9, 최대 63.4) 임계값 80 기준으로 사용 가능한 키프레임이 0장이 되는 사례가 관찰되었다. 다섯째, 본 실험은 단일 실행 결과를 기반으로 수행되었으며, 반복 실행 시 초기화 조건이나 학습 과정에 따라 성능 변동이 발생할 가능성이 있다.

향후 연구에서는 다양한 환경 및 조건에서의 대규모 실험, 변화 감지 모듈 통합을 통한 환경 변화 대응으로의 확장, 충실도 부족 영역 맵 기반의 선택적 정밀화를 통한 기존 뷰 품질 저하 최소화, 다중 XR 글래스의 동시 데이터 수집을 통한 병렬 정밀화 시나리오를 계획한다.

Ⅵ. 결 론

본 연구에서는 XR 글래스 영상을 활용한 3D Gaussian Splatting 기반 디지털 트윈의 재관측 기반 모델 정밀화 파이프라인을 제안하였다. 제안 파이프라인은 키프레임 선별, 시각 위치 추정, 정밀화 필요 영역 판단, 3DGS 모델 정밀화, 분리 평가의 5단계로 구성되며, 현장 작업자의 최소 조작으로 신규 시점의 재관측을 통한 디지털 트윈의 모델 충실도 향상을 실현한다.

두 개의 서로 다른 실내 환경에서의 실험 결과, 베이스라인 모델이 신규 시점에서 19~20 dB 수준의 불충분한 렌더링 품질을 보인 반면, 점진적 정밀화를 통해 두 장면 모두에서 11 dB 이상의 품질 향상(장면 A 19.86에서 31.29 dB, 장면 B 19.19에서 30.34 dB)을 달성하였다. 기존 시점의 품질 저하는 두 장면 모두 1.36 dB 이내로 억제되었으며, 개별 뷰 분석에서도 특정 뷰의 극단적 하락 없이 균일한 소폭 하락 양상을 확인하였다. 전체 재학습은 신규 시점에서 장면 A 2.11 dB, 장면 B 2.35 dB 우수한 품질을 보이면서도 기존 시점 보존에서는 점진적 정밀화와 유사한 수준(0.25 dB 차이)을 보여, 3DGS의 명시적 표현에서는 학습 전략에 무관하게 catastrophic forgetting이 제한적임을 확인하였다. 다만 LPIPS 기준에서는 전체 재학습이 일관되게 우수하여, 지각적 품질까지 고려할 경우 전체 재학습이 유리함을 확인하였다. 블러 임계값 어블레이션에서는 세 임계값 간 신규 뷰 품질 차이가 0.14 dB 이내로, 극단적 블러 프레임의 제거 여부가 임계값의 미세 조정보다 중요함을 확인하였다. 학습 반복 횟수 분석을 통해 55k~60k iterations가 실용적 최적점임을 확인하였다.

제안 시스템은 LiDAR나 전문 장비 없이 소비자용 XR 글래스만으로 디지털 트윈의 모델 충실도를 개선할 수 있는 실용적인 대안을 제공하며, 건설 현장 순회, 시설물 유지관리 등 지속적인 모델 유지보수가 필요한 다양한 산업 분야에서 활용이 기대된다.

감사의 글

본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단(RS-2026-25478706) 및 한국지능정보사회진흥원의 지원을 받아 수행되었습니다(사업명-2026년 지역특화형 초거대 AI 클라우드 팜 실증 및 AI 확산 환경 조성). 또한 2025년도 교육부 및 경상북도의 재원으로 경북RISE센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)-(아이디어 창업밸리)의 결과입니다(2025-rise-15-105).

참고문헌

[1] M. Grieves and J. Vickers, “Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems,” in *Transdisciplinary Perspectives on Complex Systems*, Cham, Springer, pp. 85-113, 2017. https://doi.org/10.1007/978-3-319-38756-7_4

[2] J. Yun and J. Kim, “Analysis of Digital Twin Research and Standardization Trends,” *Journal of Standards Certification Safety*, Vol. 12, No. 1, pp. 31-47, March 2022. https://doi.org/10.34139/JSCS.2022.12.1.31

[3] 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, 2022. https://doi.org/10.1145/3503250

[4] B. Kerbl, G. Kopanas, T. Leimkuhler, and G. Drettakis, “3D Gaussian Splatting for Real-Time Radiance Field Rendering,” *ACM Transactions on Graphics*, Vol. 42, No. 4, pp. 1-14, 2023. https://doi.org/10.1145/3592433

[5] S. H. Woo and J. Seo, “Performance Comparison of NeRF-Based and 3DGS-Based Models,” *Journal of Broadcast Engineering*, Vol. 30, No. 6, pp. 927-941, November 2025. https://doi.org/10.5909/JBE.2025.30.6.927

[6] H. W. Kang and J. Y. Kim, “Performance Analysis and Applicability of 3D Modeling Using 3D Gaussian Splatting in the 3D Model Production Process,” *Contents Plus*, Vol. 22, No. 6, pp. 23-39, 2024. https://doi.org/10.14728/KCP.2024.22.06.023

[7] J. H. Kim and D. Y. Lee, “A Study on Digital Sculpture Production Using 3D Gaussian Splatting Technique,” *Journal of Basic Design and Art*, Vol. 26, No. 6, pp. 55-67, December 2025. https://doi.org/10.47294/KSBDA.26.6.5

[8] S. Oh, J. Lee, and S. Park, “3D Point Cloud-Based Map Modeling for Digital Twin Environment Implementation,” *Journal of Korean Institute of Intelligent Systems*, Vol. 35, No. 5, pp. 483-488, October 2025. http://dx.doi.org/10.5391/JKIIS.2025.35.5.483

[9] M. Tancik, V. Casser, X. Yan, S. Pradhan, B. P. Mildenhall, P. Srinivasan, ... and H. Kretzschmar, “Block-NeRF: Scalable Large Scene Neural View Synthesis,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans: LA, pp. 8238-8248, June 2022. https://doi.org/10.1109/CVPR52688.2022.00807

[10] S. Lee and J. Kim, “A Study on Changes in Production Paradigm That NeRF Commercialization Will Bring to the 3D Animation Industry,” *The Korean Journal of Animation*, Vol. 20, No. 3, pp. 147-172, September 2024. https://doi.org/10.51467/ASKO.2024.09.20.3.147

[11] P.-E. Sarlin, C. Cadena, R. Siegwart, and M. Dymczyk, “From Coarse to Fine: Robust Hierarchical Localization at Large Scale,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach: CA, pp. 12716-12725, June 2019. https://doi.org/10.1109/CVPR.2019.01300

[12] D. DeTone, T. Malisiewicz, and A. Rabinowitz, “SuperPoint: Self-Supervised Interest Point Detection and Description,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City: UT, pp. 337-349, June 2018. https://doi.org/10.1109/CVPRW.2018.00060

[13] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinowitz, “SuperGlue: Learning Feature Matching with Graph Neural Networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle: WA, pp. 4938-4947, June 2020. https://doi.org/10.1109/CVPR42600.2020.00499

[14] S. Park, M. J. Chae, and S. I. Cho, “Recent Trend Analysis of Image Matching Technology Using Feature Extraction,” *Journal of Broadcast Engineering*, Vol. 28, No. 4, pp. 429-438, July 2023. https://doi.org/10.5909/JBE.2023.28.4.429

[15] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, “NetVLAD: CNN Architecture for Weakly Supervised Place Recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas: NV, pp. 5297-5307, June 2016. https://doi.org/10.1109/CVPR.2016.572

[16] J. E. Song and J. Kook, “Building a Mobile Augmented Reality System Based on Visual SLAM,” *Journal of the Semiconductor and Display Technology*, Vol. 20, No. 4, pp. 96-101, December 2021.

[17] H. Cho, Y. G. Seo, and S. Lee, “Utilizing Specialized Visual Vocabulary for Improving Relocalization Performance of ORB-SLAM,” *The Journal of Digital Contents Society*, Vol. 22, No. 11, pp. 1877-1883, 2021. https://doi.org/10.9728/dcs.2021.22.11.1877

[18] B.-K. Lee, B.-J. Kim, W.-J. Yoo, M. Ahn, and S.-J. Han, “A Study on COLMAP 3D Reconstruction Application for Cultural Heritage Restoration,” *Journal of the Korea Computer Information Society*, Vol. 28, No. 8, pp. 95-101, August 2023. https://doi.org/10.9708/jksci.2023.28.08.095

[19] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric,” in *Proceedings of the IEEE/CVF*

Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City: UT, pp. 586-595, June 2018. https://doi.org/10.1109/CVPR.2018.00068

[20] J. L. Schonberger and J.-M. Frahm, "Structure-from-Motion Revisited," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, pp. 4104-4113, June 2016. https://doi.org/10.1109/CVPR.2016.445

김정현(Jeonghyeon Kim)

2025년 : 국립금오공과대학교 컴퓨터공학부 (학사)
2026년 : 국립금오공과대학교 소프트웨어공학과 (석사 과정)

2020년～2025년: 국립금오공과대학교 컴퓨터공학부 학사
2025년～현 재: 국립금오공과대학교 소프트웨어공학과 석사과정
※관심분야 : AX(AI Transformation), 확장현실(Extended Reality), 인간컴퓨터상호작용(Human Computer Interaction) 등

이제민(Jemin Lee)

2025년 : 국립금오공과대학교 컴퓨터공학부 (학사)
2026년 : 국립금오공과대학교 소프트웨어공학과 (석사 과정)

2020년～2025년: 국립금오공과대학교 컴퓨터공학부 학사
2025년～현 재: 국립금오공과대학교 소프트웨어공학과 석사과정
※관심분야 : AX(AI Transformation), 확장현실(Extended Reality), 인간컴퓨터상호작용(Human Computer Interaction) 등

강형준(Hyeongjun Kang)

2023년 : 국립금오공과대학교 컴퓨터공학부 재학중

2023년～현 재: 국립금오공과대학교 컴퓨터공학부 학사과정
※관심분야 : AX(AI Transformation), 확장현실(Extended Reality), 인간컴퓨터상호작용(Human Computer Interaction) 등

이동희(Donghee Lee)

2023년 : 국립금오공과대학교 컴퓨터공학부 재학중

2023년～현 재: 국립금오공과대학교 컴퓨터공학부 학사과정
※관심분야 : AX(AI Transformation), 확장현실(Extended Reality), 인간컴퓨터상호작용(Human Computer Interaction) 등

류훈(Hoon Ryu)

2011년 : Purdue 대학교 전기컴퓨터공학부 (공학박사)
2005년 : Stanford 대학교 전기공학부 (공학석사)
2003년 : 서울대학교 전기공학부 (공학사)

2005년～2011년: 삼성전자 반도체총괄 System LSI 사업부
2011년～2024년: 한국과학기술정보연구원 국가슈퍼컴퓨팅본부 선임/책임연구원/양자정보응용연구단장
2024년～현 재: 국립금오공과대학교 컴퓨터공학부 교수
※관심분야 : 반도체 소재 전자구조 계산(Electronic Structure Calculation for Semiconductor Materials), 거대수치해석 병렬처리(Parallel Processing of Numerical Analysis), 양자 알고리즘 활용기술(Applications of Quantum Algorithm) 등

김영원(Youngwon Kim)

2012년 : 고려대학교 (공학사)
2018년 : 고려대학교 대학원 (공학박사)

2018년～2019년: 고려대학교
2019년～2023년: 한국전자기술연구원(KETI)
2024년～현 재: 국립금오공과대학교 컴퓨터소프트웨어공학과 교수
※관심분야 : AX(AI Transformation), 확장현실(Extended Reality), 인간컴퓨터상호작용(Human Computer Interaction) 등