사물인터넷융복합논문지 Vol. 12, No. 4, pp. 143-150, 2026 https://doi.org/10.20465/KIOTS.2026.12.4.143

로컬 환경에서의 소수 시점 3차원 객체 복원을 위한 특징 기반 이미지 선별과 보간 시점 증강

최원휘[1], 허지욱[2*]
[1]강남대학교 일반대학원 컴퓨터공학과 석사과정, [2]강남대학교 인공지능융합공학부 교수

Feature-Based Image Selection and Interpolated-View Augmentation for Sparse-View 3D Object Reconstruction in a Local Environment

Wonhwi Choi[1], Jee-Uk Heu[2*]
[1]M.S. Student, Department of Computer Science Engineering, Kangnam University
[2]Professor, Division of Applied Artificial Intelligence Engineering, Kangnam University

요 약 최근 3DGS(3D Gaussian Splatting) 기반 3차원 복원은 다수의 입력 이미지와 높은 연산 자원을 요구하며, 소수 이미지 환경에서는 시점 공백으로 인한 부유 잡음(floater)과 품질 저하가 발생한다. 본 논문은 이러한 제약을 해소하기 위해 이미지 선별, 배경 분리, 카메라 포즈 추정, 3DGS 최적화, 보간 시점 합성의 다섯 단계로 구성되며 전 과정이 독립적인 로컬 환경에서 수행되는 소수 시점 3차원 복원 파이프라인을 제안한다. 학습 전 단계에서는 카메라 포즈 정보 없이 이미지 특징 간 거리만으로 시점 차이를 근사하는 최원점 샘플링(Farthest Point Sampling, FPS)으로 시점이 고르게 분포된 샘플 이미지를 선별하고, 돌출 물체 검출의 경계 부정확성을 밝기 정보 기반 재분류로 보완하는 2단계 배경 분리를 통해 객체 영역만을 추출한다. 학습 단계에서는 특징점 기반 SfM으로 카메라 포즈를 추정하고 3DGS 최적화를 수행하여 가우시안을 객체 영역에 집중시킨다. 보강 단계에서는 인접 카메라 포즈를 짝지어 합성한 보간 시점 이미지를 학습 데이터에 추가해 재학습하는 자기 증강(self-augmentation) 기법으로 시점 공백을 보완한다. 실험을 위해 NeRF의 데이터셋을 이용하였으며, 제안 기법은 1차 학습 대비 모든 샘플 설정에서 PSNR +0.44~+0.88, SSIM +0.01~+0.04의 일관된 개선을 보였다. 절대값은 PSNR 10.55~13.20, SSIM 0.38~0.58 범위이며, 배경이 제거된 렌더링을 배경이 포함된 원본과 비교한 조건에서 산출된 상대 비교용 수치이다. 이를 통해 로컬 환경만으로 소수 시점 3차원 복원이 가능함을 보였다.

주제어 : 소수 시점 3차원 복원, 3D 가우시안 스플래팅, 최원점 샘플링(FPS), 보간 시점 증강, 로컬 환경

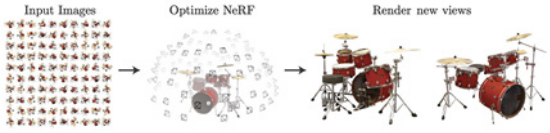
Abstract Recent 3D reconstruction methods based on 3D Gaussian Splatting (3DGS) require many input images and substantial computational resources, and often suffer from floating artifacts caused by viewpoint gaps in sparse-view environments. To address these limitations, this paper proposes a five-stage sparse-view 3D reconstruction pipeline performed entirely within an independent local environment. In the pre-processing stage, sample images with evenly distributed viewpoints are selected by farthest point sampling (FPS), which approximates viewpoint differences from image feature distances without camera pose information, and object regions are extracted by a two-stage background separation that refines salient object detection boundaries using brightness. During the training stage, camera poses are estimated using feature-based structure-from-motion (SfM), followed by 3DGS optimization to concentrate Gaussian primitives on the object regions. In the augmentation stage, images interpolated between neighboring camera poses are added to the training set for self-augmented retraining. Experiments on the NeRF dataset show consistent gains of +0.44 to +0.88 in PSNR and +0.01 to +0.04 in SSIM over the initial training stage, with absolute values of 10.55 to 13.20 dB and 0.38 to 0.58 measured by comparing background-removed renderings against original images with backgrounds. These results demonstrate the feasibility of sparse-view 3D reconstruction in a local environment.

Key Words : Sparse-View 3D Reconstruction, 3D Gaussian Splatting, Farthest Point Sampling (FPS), Interpolated-View Augmentation, Local Environment

*교신저자 : 허지욱(jeeukheu@kangnam.ac.kr)
접수일 2026년 07월 24일 수정일 2026년 08월 10일 심사완료일 2026년 08월 19일

1. 서론

인공지능 기술의 발전과 함께 3차원 이미지 처리 분야는 정밀한 기하 정보 획득에 의존하던 전통적 방식에서 벗어나, 학습 기반의 장면 표현 방식으로 빠르게 전환되고 있다. 이러한 변화는 가상현실, 디지털 트윈, 로보틱스 등 다양한 응용 분야에서 고품질 3차원 콘텐츠에 대한 수요를 촉발하고 있다. 그중 다중 시점에서 찍은 이미지들로 3차원 장면을 만들고, 이를 이용해 원하는 위치에서 본 이미지를 만들어 내는 신경 렌더링(neural rendering)[1]. 기술이 빠르게 발전하고 있다. 이 기술은 가상현실, 증강현실, 자율주행, 문화유산 복원 등 여러 분야에서 쓰이고 있다.



[Fig. 1] Overview of NeRF

[Fig. 1]은 2020년에 발표된 NeRF(Neural Radiance Fields)[2]의 학습 과정과 렌더링 결과를 나타낸다. NeRF는 3차원 장면을 신경망의 가중치로 표현하고, 카메라에서 출발한 광선을 따라 색상과 밀도를 누적하는 방식으로 임의 시점의 이미지를 생성한다.

이를 보완하기 위해 제안된 3DGS(3D Gaussian Splatting)[3]는 장면을 수십만 개의 3차원 가우시안 타원체로 표현하고, 이를 화면에 직접 투영(splatting)하여 이미지들을 렌더링한다. 3DGS는 NeRF 대비 최적화 속도와 렌더링 속도 양면에서 뚜렷한 이점을 보이며, 이에 따라 3차원 복원 연구의 주요 접근법으로 활용되어 왔다.

NeRF와 3DGS는 모두 물체 주위를 촘촘하게 돌면서 찍은 수십에서 수백 장의 이미지가 있어야 한다. 이로 인하여 다수의 입력된 이미지들을 사용할 경우 최적화 과정에서의 연산 비용이 크게 증가하고, 또한 이를 처리하기 위해 막대한 GPU 메모리가 요구되며. 이미지 수가 줄어들면 여러 시점에서 얻을 수 있는 형태 정보가 부족해져 학습이 잘못된 방향으로 진행된다. 그 결과 학습에 사용한 시점에서는 결과가 좋아 보여도, 새로운 시점에서 보면 허공에 뜬 잡음(floater), 배경이 무너지는 현상, 흐릿한 화면 등이 나타난다[4, 5]. 이러한 한계는 문화재 기록[6]과 같이 3차원 복원에 대한 수요가 있으면서도 연산 자원이 제한적인 분야에서 특히 문제가 된다.

이러한 문제를 완화하기 위해 소수의 이미지만으로 3차원 장면을 복원하는 소수 시점(few-shot) 복원 연구가 활발히 진행되어 왔다. 이 중 생성 모델 기반 방법은 사전 학습된 확산(diffusion) 모델로 미관측 시점의 이미지를 합성하여, 기존 정규화 기반 방법 대비 높은 복원 품질을 달성한다.

생성 모델 기반 기법은 대규모 사전 학습 모델에 의존하므로, 모델 확보를 위한 네트워크 연결 환경, 모델 가중치를 보관할 대용량 저장 공간, 그리고 추론에 필요한 충분한 GPU 메모리를 전제로 한다. 이러한 저장 공간 부담을 줄이기 위해 학습된 모델 자체를 압축하는 연구도 제안되었으나, 이는 이미 학습이 완료된 모델의 경량화에 초점을 두어 소수 시점 환경에서의 복원 품질 문제 자체를 해결하지는 못한다[7]. 그러나 외부 네트워크 접근이 차단된 폐쇄망 환경이나 소비자용 장비만 가용한 환경에서는 이러한 조건을 충족하기 어렵다.

본 논문에서는 이러한 환경적 제약을 해소하기 위하여, 사전 학습된 생성 모델과 외부 연산 자원에 의존하지 않고 단일 GPU를 갖춘 로컬 환경에서 독립적으로 동작하는 소수 시점 3차원 복원 방법을 제안한다. 제안 방법은 먼저 특징 공간상의 최원점 샘플링(Farthest Point Sampling, FPS)으로 시점이 고르게 분포된 이미지를 선별한다. 이 과정은 SfM에 선행하므로 카메라 포즈 없이 유사 시점 이미지의 중복 선택을 방지할 수 있다. 선별된 이미지는 배경 분리를 거친 뒤 SfM으로 카메라 포즈를 추정하고 3DGS 최적화를 수행한다. 이후 학습된 3차원 표현으로 인접 카메라 포즈 간 보간 시점 이미지를 렌더링하고, 이를 추가하여 재학습함으로써[8] 시점 공백을 보완한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 살펴보고, 3장에서는 제안하는 방법을 설명한다. 4장에서는 실험 결과를 제시하고, 5장에서 결론을 맺는다.

2. 관련 연구

2.1 신경 렌더링과 3D Gaussian Splatting

신경 렌더링은 여러 시점에서 촬영된 이미지로 3차원 장면의 표현을 학습하고 임의 시점의 이미지를 합성하는 기술이다. Mildenhall 등[2]은 3차원 좌표와 시선 방향으로부터 색상과 밀도를 출력하는 신경망으로 장면을 표현하는 NeRF를 제안하였으나, 학습과 렌더링에 오랜 시간이 소요된다는 한계가 있었다. Kerbl 등[3]은 장면을

다수의 3차원 가우시안으로 명시적으로 표현하고 화면에 직접 투영(splatting)하는 3DGS를 제안하여 실시간 렌더링을 가능하게 하였다.

한편, 신경 렌더링의 학습에는 각 이미지의 카메라 포즈가 요구되며, 특징점 기반 SfM으로 카메라 포즈와 희소 점군을 추정하는 COLMAP[9]이 표준 도구로 활용된다. 학습 기반 특징 매칭[10]은 사전 학습 모듈에 대한 의존도가 높아, 본 연구는 재현성과 자원 제약을 고려하여 Nerfstudio[11]의 3DGS 구현과 특징점 기반 SfM을 사용한다.

2.2 소수 시점 3차원 복원

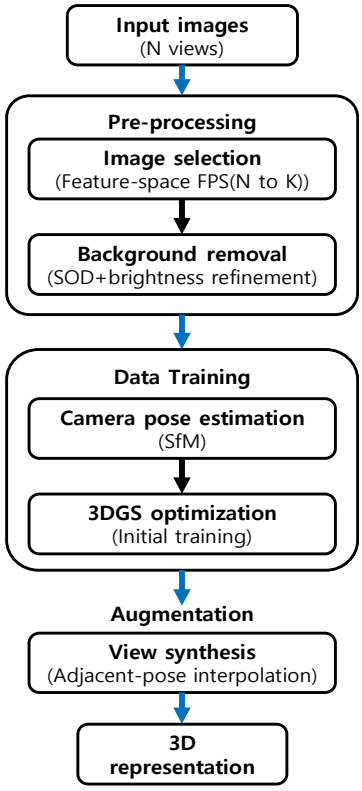
소수 시점 복원 연구는 크게 두 갈래로 진행되어 왔다. 정규화 기반 방법은 주파수 정규화[4], 미관측 시점의 기하·외형 제약[5], 깊이 사전정보 활용[12], 깊이 추정과 부유 잡음(floater) 제거[13], 가우시안의 점진적 증식[14] 등으로 부족한 다중 시점 제약을 보완하지만, 관측되지 않은 영역의 정보를 새로 생성하지는 못한다. 생성 모델 기반 방법[15]은 사전 학습된 확산 모델로 미관측 시점 이미지를 합성하여 이 한계를 넘어서나, 대용량 모델 가중치와 높은 연산 자원을 요구한다.

이와 달리 Re-Nerfing[8]은 외부 모델 없이 학습된 모델이 스스로 렌더링한 이미지를 학습 데이터에 추가하여 재학습하는 자기 증강(self-augmentation) 방식을 취한다. 본 연구도 이 계열에 속하나, 학습 이전 단계에서 특징 기반 이미지 선별과 배경 분리를 수행하고 인접한 카메라 포즈를 짝지어 보간 시점을 생성한다는 점에서 차이가 있다.

2.3 돌출 물체 검출을 활용한 배경 분리

돌출 물체 검출(Salient Object Detection)은 이미지에서 시각적으로 가장 두드러지는 객체를 배경과 분리하는 문제이다. Qin 등[16]이 제안한 U2-Net은 수 MB 수준의 작은 가중치로도 높은 정확도를 보이며, 가중치를 사전에 확보하면 네트워크 연결 없이 사용할 수 있다. 사용자가 위치를 지정하는 대화형 분할 모델[17]은 범용성이 높으나 모델 크기가 크고 사용자 개입이 요구되어 다수 이미지의 일괄 처리에 적합하지 않다. 다만 돌출 물체 검출만으로는 가늘거나 복잡한 구조의 경계가 뭉그러지는 한계가 있어, 본 연구는 밝기 정보 기반 정제 과정을 추가하여 이를 보완한다. 학습 전에 배경을 분리하면 가우시안이 객체 영역에 집중되어 미관측 시점의 부유 잡음(floater)이 억제된다.

3. 제안 방법



[Fig. 2] Overall architecture of the proposed

본 논문에서 제안하는 기법의 목표는 원본 이미지 집합 I = {I_1, I_2, ⋯, I_N}으로부터 K장(K 〈 N)의 부분집합을 선별하여, 전체 집합을 사용한 복원 결과에 근접하는 품질을 유지하는데 있다. 모든 과정은 사전 학습된 대규모 모델이나 외부 연산 자원에 의존하지 않고 로컬 환경에서 수행한다.

제안하는 방법의 전체 구조는 [Fig. 2]와 같이 다섯 개의 모듈로 구성되며, 학습 전 단계, 학습 단계, 보강 단계의 세 단계로 구분된다.

학습 전 단계에서는 전체 입력 이미지로부터 시점이 고르게 분포된 샘플 이미지를 선별하고(모듈 1), 선별된 이미지에서 배경을 분리하여 객체 영역만을 추출한다(모듈 2). 학습 단계에서는 이미지 간에 공통으로 관측되는 특징점의 대응 관계를 이용하는 특징점 기반 SfM으로 카메라 포즈를 추정한다(모듈 3). 이를 바탕으로 3DGS 최적화를 수행하며(모듈 4), 이때 본 논문은 사전 학습된 특징 추출기에 대한 의존을 배제하기 위하여 고전적 특

징점 기반의 SfM 파이프라인을 사용한다. 학습 후 보강 단계에서는 학습된 3차원 표현을 이용하여 인접한 카메라 포즈 사이의 보간 시점 이미지를 합성하고, 이를 학습 데이터에 추가하여 재학습을 수행함으로써 시점 공백에 따른 품질 저하를 보완한다(모듈 5).

3.1 모듈 1: 특징 기반 샘플 이미지 선별 (Image selection)

이미지 선별의 목적은 전체 N장의 입력된 원본 이미지 중에서 시점이 특정 방향에 편중되지 않고 객체를 고르게 둘러싸도록 분산된 K장의 샘플 이미지를 선택하는 데 있다.

우선 원본 이미지로부터 각 이미지 I_i의 특징 벡터 $f(I_i)$를 추출한다. 특징 벡터는 이미지의 시각적 특징을 수치화한 벡터이며, 시점이 유사한 이미지 쌍은 특징 공간에서 가까이, 시점 차이가 큰 이미지 쌍은 멀리 사상된다. 따라서 특징 벡터 간 거리는 시점 간 차이의 근사 척도로 활용될 수 있으며, 본 모듈은 이 거리를 기준으로 기존 선택 집합과 유사한 이미지는 배제하고 가장 상이한 이미지를 반복적으로 채택하기 위하여 특징 공간에서 최원점 샘플링(Farthest Point Sampling, FPS)을 수행한다[18]. FPS는 이미 선택된 이미지 집합과의 최소 거리(min)가 최대(argmax)가 되는 이미지를 반복적으로 추가하는 기법으로 식 (1)로 표현된다.

특징 벡터 $f(I)$는 이미지를 16×16으로 축소한 RGB 화소값(768차원)과 4×4×4 구간 색 히스토그램(64차원)을 연결한 832차원 벡터로 구성하며, 식 (1)의 거리 $d(\cdot,\cdot)$는 유클리드 거리를 사용한다. 두 구성 요소 모두 사전 학습 모델 없이 계산된다.

$$I^{*} = \mathrm{argmax}_{I \in I \setminus S_k} \left[\min_{I_j \in S_k} d(f(I), f(I_j)) \right] \quad (1)$$

3.2 모듈 2: 배경 분리(Background removal)

모듈 2는 샘플 이미지에서 효과적인 객체 학습을 위하여 각 이미지에서 객체 영역을 제외한 배경을 분리하는 작업을 수행한다. FPS로 선별된 K장의 샘플 이미지에는 복원 대상 객체 외에 배경 영역이 포함되어 있어, 시점에 따라 형태가 일관되지 않기 때문에 최적화 과정에서 기하 추정의 모호성을 유발한다. 미관측 시점에서 객체 주변의 허공에 잡음 형태의 가우시안이 떠 있는 floater가 발생하는 주요 원인이 된다.

배경 분리는 두 단계로 진행된다. 우선 돌출 물체 검출이 가벼운 모델인 U2-Net을 이용하여 객체의 초기 마스크를 생성한다. 돌출 물체 검출만으로는 가늘거나 복잡한 구조의 경계가 뭉그러지는 한계가 있으므로, 객체와 배경 간 밝기 차이를 이용해 경계 영역의 화소를 재분류함으로써 마스크의 정밀도를 높인다. 그 후 초기 마스크의 경계를 밝기 정보에 기반하여 이미지에서 객체와 배경을 분리한다. 이렇게 정제된 마스크를 기반으로 배경이 제거된 객체 이미지들은 이후 학습 단계에서 학습을 위한 입력 값으로 전달된다. 배경이 제거된 샘플 이미지들은 특징점 정합이 객체 영역에 집중되어, 배경 영역의 floater로 인한 오류를 줄이고 카메라 포즈 추정값의 정확도를 높이게 된다.

3.3 모듈 3: 카메라 포즈 추정 모듈 (Camera pose estimation)

모듈 2에서 배경이 분리된 K장의 샘플 이미지들은 객체 복원을 위한 카메라 포즈 정보가 없기에 이를 추정하는 작업이 필요하다. 모듈 3은 샘플 이미지로부터 특징점 기반 SfM(Structure from motion)이 적용된 각 이미지의 카메라 포즈와 장면의 희소 점군(point cloud)을 추정한다. 추정된 카메라 포즈값들은 모듈 4의 3DGS 최적화에서 각 이미지 투영 관계를 정의하기 위해 추정된 희소 점군을 가우시안의 초기 위치로 활용한다.

3.4 모듈 4: 3DGS 최적화 모듈(3DGS optimization)

모듈 3에서 추정된 카메라 포즈와 희소 점군을 초기값으로 하여, 샘플 이미지로부터 3차원 객체의 표현을 복원하기 위한 3DGS 최적화를 수행한다. 3DGS에서 장면은 각각 위치, 공분산, 불투명도, 색상 계수를 갖는 다수의 3차원 가우시안의 집합으로 표현되며, 희소 점군의 각 점이 가우시안의 초기 위치로 사용된다. 최적화는 이 가우시안 집합을 각 샘플 이미지의 카메라 포즈로 투영하여 렌더링한 이미지와 해당 입력 이미지 간의 차이를 최소화하는 방향으로 가우시안의 파라미터를 갱신하는 과정이다. 이 과정에서 재구성 오차가 큰 영역의 가우시안을 분할·복제하고 불투명도가 낮은 가우시안을 제거하는 적응적 밀도 제어가 함께 수행된다.

3.5 모듈 5: 시점 합성(View synthesis)

소수의 샘플 이미지로 3차원 객체를 복원하는 경우 학습 시점 사이의 간격이 넓어 시점 공백이 발생하며, 해당 영역에서 렌더링 품질이 저하된다. 이에 본 모듈은 1차

학습된 3차원 표현을 이용하여 학습 시점 사이를 잇는 보간 시점(interpolated view) 이미지를 합성하고, 이를 학습 데이터에 추가하여 재학습을 수행함으로써 시점 공백을 보완한다.

먼저 모듈 3에서 추정된 샘플 이미지의 카메라 포즈를 기준으로, 공간상에서 서로 인접한 카메라 포즈 쌍 (i, j)을 구성하고, 인접 여부는 카메라 위치 벡터 간 유클리드 거리를 기준으로 판정하며, 각 시점에 대해 거리가 가장 가까운 시점을 짝지어 쌍을 구성한다.

보간 범위를 인접한 쌍으로 한정하는 것은, 학습 시점으로부터 멀리 떨어진 가상 시점에서는 렌더링 품질이 보장되지 않아 오류가 누적될 수 있기 때문이다. 각 쌍에 대하여 두 카메라 포즈를 보간한 가상 카메라 포즈의 위치는 식 (2)로 회전은 식(3)으로 각각 계산한다.

$$t_{\text{mid}} = \frac{(t_i + t_j)}{2} \tag{2}$$

$$q_{\text{mid}} = \text{slerp}(q_i, q_j; 0.5) \tag{3}$$

식 (2)의 t_i, t_j는 두 카메라의 위치 벡터, 식 (3)의 q_i, q_j는 회전을 나타내는 단위 사원수이며, slerp(q_i, q_j ; 0.5)는 구면 선형 보간(Spherical linear interpolation)의 중점을 계산한다. 위치는 선형 보간으로, 회전은 구면 선형 보간으로 처리함으로써 두 시점 사이를 자연스럽게 잇는 가상 카메라 포즈 t_{mid}를 얻는다. 생성된 가상 카메라 포즈 t_{mid}로 1차 학습된 3차원 표현을 렌더링하여 보간 시점 이미지가 합성되며, 이 과정을 샘플 이미지로 구성된 모든 인접 쌍에 대하여 반복한다. 합성된 보간 시점 이미지는 대응하는 가상 카메라 포즈와 함께 새롭게 학습 데이터로 등록되어, 기존 K장의 샘플 이미지에 이를 추가한 보강 데이터로 재학습을 수행한다.

보강된 데이터는 학습 시점 사이의 공백 영역에 추가적인 감독 신호, 즉 해당 시점에서 렌더링 결과가 일치해야 할 기준 이미지를 제공한다. 이를 통해 미관측 시점에서 발생하는 부유 잡음(floater), 배경 붕괴, 흐릿한 렌더링 등의 아티팩트가 억제된다.

이 과정이 완료되면, 원본 이미지에 담긴 객체의 시점 정보 중 샘플 선별 과정에서 제외되어 발생한 시점 공백이 보간 시점 이미지로 보완된다. 최종적으로 샘플 이미지와 보강된 보간 시점 이미지를 함께 사용하여 재학습을 수행함으로써, 원본 이미지 전체를 사용한 경우에 근접하는 3차원 객체를 복원한다.

4. 실험 및 평가

4.1 실험 환경 및 데이터

본 논문에서는 소수 시점 3차원 복원을 독립적인 GPU를 갖춘 로컬 환경 구축을 위해서 NVIDIA GeForce RTX 4070 Laptop GPU(VRAM 8GB)에서 실험하였다. 실험은 Nerfstudio 프레임워크 기반으로 수행하였으며, 3DGS 구현으로는 Splatfacto를 사용하고 학습은 30,000 iteration을 기준으로 진행하였다.

실험 데이터는 NeRF[2]에서 제공하는 데이터셋 lego 씬의 원본 이미지 100장(800×800 해상도)을 사용하였으며, 특징 기반 FPS(Farthest Point Sampling)를 통해 이보다 적은 K장(K = 90, 80, 70, 60)을 선별함으로써 few-shot 복원 환경을 구성하였다.

학습에는 배경이 제거된 이미지를 사용하며, [Fig. 3]의 배경이 원 데이터셋과 다르게 보이는 것은 배경 제거 후 재구성된 결과를 렌더링하였기 때문이다. 이때 발생하는 시점 공백은 인접 시점 간 보간 이미지를 합성하고 이를 학습 데이터에 추가하는 자기 증강 방식으로 보완하였으며, 이에 따라 각 K값에 대해 두 가지 조건으로 복원 실험을 수행하였다.

recon1은 선별된 K장의 샘플 이미지만으로 학습한 결과이고, recon2는 부족한 (100 − K)장만큼 보간 시점 이미지를 합성하여 추가함으로써 원본 전체에 근접한 규모로 재학습한 결과이다. 즉 K = 90, 80, 70, 60일 때 추가되는 보간 시점 이미지는 각각 10장, 20장, 30장, 40장이다.

평가 시점은 각 K 설정에서 선별된 K장을 정렬한 뒤 균등 간격으로 약 10%를 분리하여 학습에서 제외하고 평가 전용으로 남겨둔 공통 평가 집합이다. 이 분할은 고정된 규칙을 따라 반복 실행 시에도 동일하게 유지되며, 같은 K 설정의 recon1과 recon2는 동일한 평가 시점에서 비교된다.

4.2 샘플 이미지 수(K)에 따른 품질 비교

평가 지표로는 화소 단위 오차를 측정하는 PSNR(Peak Signal-to-Noise Ratio)과 밝기·대비·구조의 유사성을 측정하는 SSIM(Structural Similarity Index Measure)을 사용하였다. PSNR 지표는 값이 클수록, 복원의 품질이 높음을 의미한다. SSIM 지표 역시 1에 가까울수록 복원 품질이 높음을 의미하며, 화소 오차 기반의 PSNR이 포착하지 못하는 구조적 왜곡을 SSIM이 보완한다. 공정

한 비교를 위해 각 설정에서 기본 복원과 보강 복원은 학습에 사용되지 않은 동일한 공통 평가 시점에 대해 평가하였다. 하지만 일반적으로 이미지 복원 시 잡음도 함께 생성되기 때문에 수치적으로 낮게 측정되는 편이다[19]. 〈Table 1〉은 샘플 이미지 수 K에 따른 recon1과 recon2의 복원 품질을 PSNR과 SSIM으로 비교한 결과이다.

〈Table 1〉 Comparison of reconstruction quality according to the number of sample images (K)

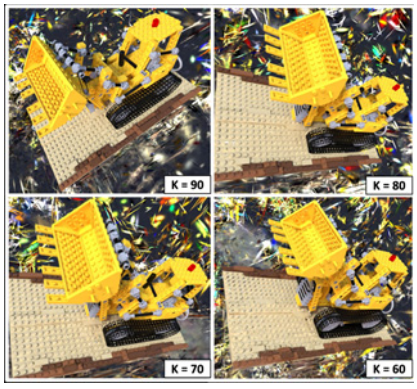
K	90		80		70		60	
Metric	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
recon1	12.32	0.54	11.80	0.52	10.67	0.42	10.55	0.38
recon2	13.20	0.58	12.59	0.55	11.11	0.43	11.13	0.39

모든 K 설정에서 생성 이미지를 추가한 recon2가 샘플 이미지만 사용한 recon1보다 일관되게 우수한 성능을 보였다. PSNR 기준으로 90장 설정에서 +0.88, 80장 설정에서 +0.79, 70장 설정에서 +0.44, 60장 설정에서 +0.58의 향상이 확인되었으며, SSIM 역시 전 구간에서 +0.01~+0.04의 개선을 보였다.

샘플 이미지 수에 따른 경향을 살펴보면, 샘플 수가 90장에서 60장으로 감소함에 따라 recon1의 PSNR은 12.32에서 10.55로, SSIM은 0.54에서 0.38로 단조 감소하였다. 이는 학습 시점(view)의 감소가 재구성 품질에 직접적인 영향을 미침을 보여준다. 주목할 점은 생성 이미지에 의한 개선 폭이 샘플 수가 많은 구간(90장, 80장)에서 상대적으로 크게 나타났다는 것이다. 이는 보간 기반 생성 이미지의 품질이 인접 실측 시점의 밀도에 의존하기 때문으로, 샘플 수가 적을수록 생성 이미지 자체의 품질이 저하되어 개선 효과가 제한되는 것으로 해석된다.

4.3 복원 이미지 분석

[Fig. 3]은 각 설정의 재구성 결과를 Nerfstudio 뷰어에서 렌더링한 것이다. 샘플 수가 감소함에 따라 객체 주변부에 부유 가우시안(floater)에 의한 잡음 아티팩트가 두드러지게 증가하였다. 특히 60~70장 설정에서는 학습 시점이 희소한 영역을 중심으로 배경 전반에 걸쳐 산란형 아티팩트가 관찰되었다. 다만 객체 본체의 기하 구조와 색상은 저샘플 설정에서도 비교적 온전하게 유지되어, 정량 지표의 하락이 주로 배경 및 비관측 영역의 아티팩트에서 기인함을 확인할 수 있었다.



[Fig. 3] Rendering results of the reconstructed 3D object for each sample image count (K = 90, 80, 70, 60)

4.4 분석 및 한계

본 논문의 평가는 배경이 제거된 상태로 학습·렌더링한 결과를 배경이 포함된 원본과 비교하므로, 객체 영역의 복원 품질과 무관하게 배경 화소가 오차로 산입되어 지표의 절대값이 낮게 측정된다. 또한 공통 평가 시점에 학습 시점 분포에서 상대적으로 떨어진 시점이 포함되고, 학습 시점의 감소로 발생한 부유 잡음(floater)이 화면 전체의 지표를 저하시킨다. 한편 본 기법의 이점은 촬영에 요구되는 이미지 수의 경감에 한정되며, 보간 렌더링과 재학습을 수행하는 recon2의 총 연산 비용은 원본 전체를 단일 학습하는 경우보다 크다. 그럼에도 1차 학습 대비 재학습에서 일관된 개선이 관찰되어, 보간 시점 이미지를 통한 데이터 증강이 제한된 촬영 환경에서 소수 시점 3차원 복원 품질을 실질적으로 향상시킬 수 있음을 보여준다. 다만 1차 학습된 표현이 렌더링한 이미지를 다시 학습에 사용하므로 1차 학습의 부유 잡음이 보간 이미지에 전사될 수 있으며, 이는 인접 실측 시점의 밀도가 낮은 저샘플 구간일수록 커진다. 보간 범위를 인접 쌍의 중점으로 한정하여 이를 부분적으로 완화하였으나, 합성 이미지의 신뢰도를 판정해 선별하는 과정은 향후 과제로 남긴다.

5. 결론 및 향후 연구

본 논문에서는 사전 학습된 대규모 생성 모델과 외부 연산 자원에 의존하지 않고, 단일 GPU를 갖춘 로컬 환경에서 독립적으로 동작하는 소수 시점 3차원 복원 방법을 제안하였다. 제안 방법은 특징 기반 샘플 이미지 선

별, 2단계 배경 분리, 카메라 포즈 추정, 3DGS 최적화, 보간 시점 합성을 통한 재학습의 다섯 단계로 구성되며, 전 과정이 로컬 환경에서 수행된다.

본 논문의 기여는 다음과 같다. 카메라 포즈 추정에 선행하는 특징 기반 시점 선별로 포즈 정보 없이도 객체를 고르게 둘러싸는 샘플 집합을 구성하였고, 밝기 정보 기반 재분류를 결합한 2단계 배경 분리로 가우시안을 객체 영역에 집중시켜 부유 잡음(floater)을 억제하였다. 또한 인접 카메라 포즈를 짝지어 생성한 보간 시점 이미지를 대응 포즈와 함께 등록하여 별도의 포즈 추정 없이 재학습에 활용하였다. 실험 결과, 보간 시점 이미지를 추가한 재학습은 1차 학습 대비 모든 설정에서 일관된 품질 향상을 보여, 외부 생성 모델 없이도 소수 시점 환경의 품질 저하를 완화할 수 있음을 확인하였다. 다만 학습·렌더링은 배경이 제거된 상태에서, 평가는 배경이 포함된 원본 기준으로 이루어져 〈Table 1〉의 수치는 동일 파이프라인 내 상대 비교에 한정되며, 검증 범위가 K=60~90에 머무르고 무작위 샘플링과의 대조 실험이 없어 FPS 선별 자체의 기여도는 분리하여 확인하지 못하였다.

향후에는 신뢰도가 낮은 보간 이미지를 걸러내는 선별적 증강과 다중 시점 보간으로 기법을 확장하고, 학습·평가 조건을 일치시킨 프로토콜에서 K 범위 확장 및 무작위 샘플링 대비 실험을 재수행하며, 다양한 객체와 실내외 장면으로 대상을 넓혀 제안 방법의 일반성을 검증할 계획이다.

REFERENCES

[1] A.Tewari, J.Thies, B.Mildenhall, P.Srinivasan, E.Tretschk, Y.Wang, C.Lassner, V.Sitzmann, R.Martin-Brualla, S.Lombardi, T.Simon, C.Theobalt, M.Niessner, J.T.Barron, G.Wetzstein, M.Zollhoefer, and V.Golyanik, "Advances in neural rendering," *Computer Graphics Forum*, Vol. 41, No. 2, pp. 703-735, 2022.

[2] B.Mildenhall, P.P.Srinivasan, M.T., J.T.Barron, R.Ramamoorthi, and R.Ng. "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," in Proceedings of the *European Conference on Computer Vision (ECCV)*, pp. 405-421, 2020

[3] 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 139-1, 2023.

[4] M.Niemeyer, J.T.Barron, B.Mildenhall, M.S.Sajjadi, A.Geiger, & N.Radwan, "Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs," In Proceedings of the *IEEE/CVF conference on computer vision and pattern recognition*, pp. 5480-5490, 2022.

[5] J. Yang, M. Pavone and Y. Wang, "FreeNeRF: Improving Few-shot Neural Rendering with Free Frequency Regularization," In Proceedings of the *IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR)*, pp. 8254-8263, 2023.

[6] V.Croce, G.Caroti, L.D.Luca, A.Piemonte, and P.Veron, "Neural radiance fields (nerf): Review and potential applications to digital cultural heritage," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 48, pp. 453-460, 2023.

[7] C.L.Deng, and T.Enzo, "Compressing explicit voxel grid representations: fast nerfs become also small." In Proceedings of the *IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1236-1245, 2023.

[8] F.Tristram, S.Gasperini, N.Navab and F.Tombari, "Re-Nerfing: Improving Novel View Synthesis through Novel View Synthesis," arXiv preprint arXiv:2312.02255, 2023.

[9] J.L.Schönberger and J.M.Frahm, "Structure- from-Motion Revisited," In Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pp.4104-4113, 2016.

[10] P.E.Sarlin, D.DeTone, T.Malisiewicz and A.Rabinovich, "SuperGlue: Learning Feature Matching with Graph Neural Networks," In Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pp.4938-4947, 2020.

[11] M.Tancik, E.Weber, E.Ng, R.Li, B.Yi, T.Wang, A.Kristoffersen, J.Austin, K.Salahi, A.Ahuja, D.Mcallister,J.Kerr and A.Kanazawa, "Nerfstudio: A Modular Framework for Neural Radiance Field Development," In Proceedings of the *ACM SIGGRAPH 2023 Conference*, pp. 1-12, 2023.

[12] K.Deng, A.Liu, J.Y. Zhu and D.Ramanan, "Depth-supervised NeRF: Fewer Views and Faster Training for Free," In Proceedings of the *IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR)*, pp.12882-12891, 2022.

[13] J.Li, J.Zhang, X.Bai, J.Zheng, X.Ning, J.Zhou and L.Gu, "Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization," In Proceedings of the *IEEE/CVF conference on computer vision and pattern recognition*, pp. 20775-20785, 2024.

[14] Z.Zhu, Z.Fan, Y.Jiang and Z.Wang, "FSGS: Real-Time Few-shot View Synthesis using Gaussian Splatting," In Proceedings of the *European Conference on Computer Vision (ECCV)*, pp. 145-163, 2024.

[15] R.Wu, B.Mildenhall, P.Henzler, K.Park, R.Gao, D.Watson, P.P.Srinivasan, D.Verbin, J.T.Barron, B.Poole, A.Hołynski, "ReconFusion: 3D Reconstruction with Diffusion Priors," In Proceedings of the *IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR)*,

pp. 21551-21561, 2024.

[16] X.Qin, Z.Zhang, C.Huang, M.Dehghan, O.R.Zaiane, and M.Jagersand, "U2-Net: going deeper with nested U-structure for salient object detection," *Pattern Recognition*, Vol. 106, pp. 107404, 2020.

[17] A.Kirillov, E.Mintun, N.Ravi, H.Mao, C.Rolland, L.Gustafson, T.Xiao, S.Whitehead, A.C. Berg, W.Y.Lo, P.Dollar, R.Girshick, "Segment Anything," in Proceedings of the *IEEE/CVF International Conference on Computer Vision(ICCV)*, pp. 4015-4026, 2023.

[18] C.R.Qi, L.Yi, H.Su and L.J.Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," In Proceedings of the *Advances in Neural Information Processing Systems(NeurIPS)*, pp. 5099-5108, 2017.

[19] J.Gaoxiang, and V.Appia, "Mind the Gap: Standard 3DGS Evaluation Primarily Measures Near-Trajectory Interpolation." arXiv preprint arXiv:2607.01556, 2026.

최 원 휘(Wonhwi Choi) [준회원]

- 2025년 2월 : 강남대학교 ICT융합공학부 소프트웨어전공 (공학사)
- 2025년 3월 ~ 현재 / 강남대학교 일반대학원 컴퓨터공학과 석사과정

〈관심분야〉
컴퓨터 비전, 임베디드 시스템, 온디바이스 AI

허 지 욱(Jee-Uk Heu) [정회원]

- 2016년 2월 : 한양대학교 컴퓨터공학과 (공학박사)
- 2017년 2월 ~ 2017년 12월 : Stanford Center for Professional Development 연구원
- 2021년 3월 ~ 2025년 2월 : 강남대학교 참인재대학 교양교수부 조교수
- 2025년 3월 ~ 현재 : 강남대학교 인공지능융합공학부 조교수

〈관심분야〉
빅데이터 분석, 문서요약, 인공지능, 영상처리, 자연어처리