JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 30 No. 5, pp. 9-20, May 2025
https://doi.org/10.9708/jksci.2025.30.05.009

Image quality enhancement for arbitrary viewpoint synthesis based on accurate inter-camera point matching

Jong-Su Ha*, Young-Soo Kwon**, Jin-Ah Kim***, Yoo-Joo Choi***

*Student, Dept. of AI S/W Engineering, Seoul Media Institute of Technology, Seoul, Korea
**Director, ZYX Technology, Seoul, Korea
***Professor, Dept. of AI S/W Engineering, Seoul Media Institute of Technology, Seoul, Korea

[Abstract]

Using deep neural networks such as NoPe-NeRF, it has been observed that when generating 2D synthesized images from arbitrary viewpoints—after learning 3D spatial information from 2D input images without prior camera information—the image quality in distant regions tends to degrade significantly. This paper proposes a method to address the issue of image quality degradation in distant regions of synthesized images. The degradation is presumed to result from the use of inaccurately matched point pairs with low correspondence accuracy when computing the point cloud loss function in NoPe-NeRF. To mitigate this problem, we incorporate a Probabilistic Dense Correspondence Network (PDCNet) to improve point matching accuracy between adjacent camera images. These refined correspondences are then used to compute a more reliable point cloud loss. The effectiveness of the proposed method is validated through quantitative evaluation using PSNR, SSIM, and LPIPS, as well as qualitative visual comparisons. Furthermore, to assess improvements in camera pose estimation, ATE and RPE metrics are compared before and after applying our method. The results demonstrate that the proposed approach not only enhances the visual quality of synthesized images from arbitrary viewpoints but also improves the accuracy of camera pose estimation.

▸ **Key words**: Synthesized Image Generation, 3D Spatial Estimation, NeRF, and Probabilistic Dense Correspondence Networks

• First Author: Jong-Su Ha, Corresponding Author: Yoo-Joo Choi
*Jong-Su Ha (ha1153@naver.com), Dept. of AI S/W Engineering, Seoul Media Institute of Technology
**Young-Soo Kwon (dreamhighkwon@gmail.com), ZYX Technology
***Jin-Ah Kim (kkim.jinah00@gmail.com), Dept. of AI S/W Engineering, Seoul Media Institute of Technology
***Yoo-Joo Choi (yjchoi@smit.ac.kr), Dept. of AI S/W Engineering, Seoul Media Institute of Technology
• Received: 2025. 04. 18, Revised: 2025. 04. 30, Accepted: 2025. 05. 09.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

[요 약]

NoPe-NeRF와 같은 딥 뉴럴 네트워크를 이용하여, 사전 카메라 정보 없이 2D 이미지를 입력으로 3D 공간정보를 학습한 후, 임의의 시점(Viewpoint)으로부터 2D 합성 이미지를 생성할 때 원경 영역의 이미지 품질 저하 문제가 관찰되었다. 이에 본 논문은 합성 이미지에서 원경 영역의 화질 저하의 문제점을 해결하기 위한 방안을 제시한다. 해당 문제가 NoPe-NeRF에서 포인트 클라우드 손실 함수 계산 시, 매칭 정확도가 낮은 포인트 쌍을 기반으로 손실을 계산하는 과정에서 비롯된다고 추정하고, 이를 개선하기 위해 확률적 밀집 대응 네트워크(PDCNet)를 이용하여 인접 카메라 이미지 간의 포인트 매칭 정확도를 높이고, 이를 기반으로 보다 신뢰도 높은 포인트 클라우드 손실을 계산하도록 설계하였다. 제안 기법의 효과는 PSNR, SSIM, LPIPS 지표를 활용한 정량적 평가와 합성 이미지에 대한 정성적 비교 분석을 통해 입증되었다. 또한, 카메라 포즈 추정 성능의 개선 여부를 검증하기 위해 ATE 및 RPE 지표를 기준으로 제안 방법 적용 전후의 성능을 비교하였다. 실험 결과, 제안 기법은 임의 시점에서 생성된 합성 이미지의 품질을 향상시켰을 뿐만 아니라, 카메라 포즈 추정의 정확도 역시 함께 개선되었음을 확인하였다.

▸ **주제어**: 합성 영상 생성, 3차원 공간정보 추정, NeRF, 확률적 밀집 대응 네트워크

I. Introduction

최근 영화, 광고 등의 산업 분야에서 영상 촬영의 비용을 줄이고 다이나믹한 장면의 연출을 위하여 몇 장의 2D 이미지 데이터를 바탕으로 3D 장면을 재구성하고, 다양한 시점에서 고품질의 새로운 이미지를 렌더링하는 혁신적인 기술이 활용되고 있다. 이러한 기술을 가능하게 해주는 대표적인 딥-뉴럴 네트워크로 NeRF[1]를 들 수 있다.

NeRF는 서로 다른 시점의 2D 이미지들과 각 2D 이미지에 대한 카메라 포즈 정보를 입력으로 받아 이를 기반으로 3D 공간정보, 즉 3D 공간을 구성하는 각 지점의 색상과 밀도 값을 추정한다. 최종적으로 3D 공간을 구성하는 점들의 색상과 밀도 값을 기반으로 볼륨 렌더링을 통하여 임의의 시점에서의 고해상도의 2D 이미지를 재구성한다. 그러나 3차원 공간정보를 재구성하는 현실에서 2D 이미지를 촬영하는 카메라 포즈, 즉 카메라의 3차원 공간상의 위치(position)와 회전 정보(rotation)를 알아내는 것은 쉽지 않은 일이다. 또한, 노이즈를 가지는 카메라 포즈 정보는 NeRF가 정확한 3D 장면을 복원하는 데 방해 요소가 된다. 이러한 문제를 해결하기 위하여 최근 발표된 딥-뉴럴 네트워크들[2,3,4,5]에서는 입력 이미지에 대한 카메라 포즈를 입력받지 않고, 카메라 포즈도 학습을 통하여 최적화하도록 제안하고 있다. 이들 딥-뉴럴 네트워크 중 우수한 성능을 보이고 있는 NoPe-NeRF[5]는 카메라의 움직임이 큰 환경에서 합성 이미지 화질 저하의 문제를 해결하고자 카메라별 단안 깊이(Monocular depth)값을 단안 깊이 신경망(DPT: Dense Prediction Transformer)[6]을 통하여 추론하고, 스케일과 시프트 매개변수를 통해 왜곡을 바로잡으며 카메라 포즈 정보를 학습 최적화할 수 있도록 하였다.

최근 연구[7]에서 NoPe-NeRF를 통하여 생성된 합성 이미지에서 카메라에서 가까운 전경 영역에 비해 원경 영역에 대한 이미지의 화질이 현저하게 저하되는 현상이 지적되었다. 또한, 실험을 통하여 원경 영역의 화질 저하는 NoPe-NeRF가 학습에 적용하고 있는 손실 중 포인트 클라우드 손실에 의한 영향으로써, 포인트 클라우드 손실을 제외한 경우 원경 영역의 화질은 개선되지만, 전경 영역의 화질은 오히려 저하되는 경향이 나타남을 보고하였다.

Fig. 1과 Fig. 2는 각각 포인트 클라우드 손실을 적용한 경우와 제외한 경우에 Ignatius 데이터 세트[16]를 학습하고, 임의의 시점에 대한 합성 이미지를 생성한 결과를 보여주고 있다. Fig. 1에서는 원경에 위치한 건물의 이미지 화질이 뚜렷하게 저하됨이 확인되었고, 포인트 클라우드 손실을 적용하지 않은 Fig. 2의 경우에는 원경 영역의 건물 이미지의 화질이 저하되지 않고, 오히려 전경 영역의 화질이 다소 저하되는 결과를 보여주고 있다. 이러한 결과는 Nope-NeRF에서 관찰된 원경 화질의 저하 원인이 포인트 클라우드 손실과 밀접하게 관련됨을 확인할 수 있다.

이에 본 논문에서는 NoPe-NeRF의 한계점인 원경 영역의 급격한 화질 저하의 문제를 해결하기 위하여 포인트 클라우드 손실 계산에 사용되는 새로운 포인트 클라우드 집

합의 구성 방법을 제안한다. 또한, 새로운 포인트 클라우드 집합 구성을 통하여 NoPe-NeRF에서 사용하던 표면 기반 광도 손실 함수를 적용하지 않아도 최종 합성 이미지에서 원경과 전경의 화질 저하 없이 안정적인 품질이 유지됨을 보이고자 한다. 제안 방법의 효과를 검증하기 위해, 방법 적용 전후의 PSNR, SSIM, LPIPS 지표를 활용한 정량적 화질 평가와 정성적 시각 비교 분석을 수행하였다. 아울러, 카메라 포즈 추정 성능의 변화를 확인하기 위해 ATE 및 RPE 지표를 통해 정합 정확도 또한 종합적으로 비교·분석하였다.

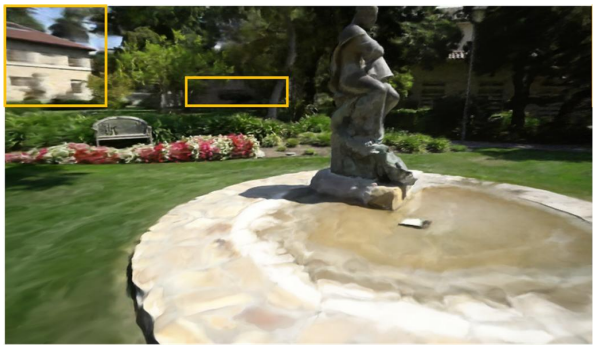

Fig. 1. Synthesized image generated by None-NeRF model with applying the point cloud loss

Fig. 2. Synthesized image generated by the None-NeRF model without applying the point cloud loss

II. Related Work

1. Neural Radiance Field (NeRF)

NeRF(Neural Radiance Field)는 2020년 Google Research와 UC Berkeley의 연구자들이 발표한 딥러닝 기반 3D 장면(scene) 재구성 방법이다. NeRF의 핵심 아이디어는 심층 신경망을 사용하여 3D 공간에서의 각 포인트에 대한 밀도와 색상 정보를 예측하는 것이다. 이때, 주어진 3D 공간 위치 $X=(x,y,z)$와 카메라 시점 방향 벡터 $d=(\theta,\phi)$를 입력으로 하는 다층 퍼셉트론(MLP)을 통해 3D 공간상의 각 포인트의 색상(RGB)(c)과 볼륨 밀도(σ)를 예측한다. 여기서 $d=(\theta,\phi)$는 구면 좌표계(spherical coordinates) 상의 방향(direction)을 나타내는 방위각과 고도 각을 의미한다. 최종적으로 3D 공간상의 각 포인트에 대한 색상(c)과 밀도 값(σ)을 기반으로 볼륨 렌더링을 통해 임의의 시점에서의 2D 이미지를 재구성한다.

볼륨 렌더링은 카메라에서 발사된 광선(rays)을 따라 여러 지점을 샘플링하고, 해당 지점의 RGB와 밀도(σ)를 기반으로 식(1)의 볼륨 렌더링 방정식에 따라 이미지 색상값 $C(r)$을 결정하는 방법이다. 즉, 광선 r이 지나가는 포인트마다 누적되는 색과 불투명도를 계산해 이미지 픽셀 값을 결정한다.

$$C(r)=\int_{t_n}^{t_f} T(t)\cdot\sigma(t)\cdot c(t)\,dt \tag{1}$$

여기서, t_n과t_f는 3D 공간을 정의하는 광선상의 샘플링 시작과 끝 지점을 의미하고, $T(t)$는 누적 투과율을 의미한다.

이전의 3D 공간 재구성 및 렌더링 기법들과 비교했을 때, NeRF는 보다 적은 입력데이터를 이용하여 고품질의 이미지를 생성해 낼 수 있다는 장점이 있다. 다수의 뷰에서 획득한 이미지를 입력으로 사용할 때, NeRF는 복잡한 장면의 세부 사항까지 재현할 수 있다. 반면, 기존의 기법들은 포인트 클라우드, 복셀 그리드와 같은 명시적인 3D 구조를 필요로 하고, 이는 메모리 및 계산 자원의 소모를 심각하게 증가시킨다. 이러한 문제점을 해결하기 위하여 NeRF는 복잡한 장면의 3D 정보를 신경망을 통하여 효율적으로 압축하고, 필요할 때 이를 복원하는 방식을 사용한다. 그러나 NeRF는 3D 장면의 재구성을 위해 정확한 카메라 포즈가 주어진다는 전제하에 동작한다. 정확한 카메라 포즈를 기반으로 카메라로부터 시작하는 광선과 3D 공간상의 샘플링 포인트가 결정된다. 그러나, 실제 환경에서는 모든 이미지 데이터에 대해 완벽한 카메라 포즈 정보를 얻는 것은 쉽지 않은 일이며, 노이즈가 있는 카메라 포즈는 NeRF가 정확한 3D 장면을 복원하는 데 주요한 방해 요소로 작용한다. 이러한 문제를 해결하기 위하여 카메라 포즈 최적화를 NeRF 모델과 결합한 방법들이 제안되었다.

2. NeRF models with camera pose estimation

2.1 NeRF--

기존의 NeRF가 카메라 포즈 정보에 민감한 한계를 극복하기 위해, 노이즈가 섞인 카메라 포즈를 입력으로 받아 이를 NeRF와 동시에 학습하는 방법을 제안했다. NeRF--[2]은 각 뷰(view)에 대한 카메라 포즈를 학습 가능한 변수로 두고 최적화하였고, 카메라 경로의 smoothness를 유지하기 위하여 정규화 항을 추가하였다. 이 방법은 부족한 데이터 세트나 일부 포즈 오차가 포함된 환경에서도 높은 성능을 보이며, 포즈 최적화의 중요성을 보였다. 카메라 포즈 정보를 얻기 위하여 COLMAP 같은 외부 포즈 추정 도구를 사용하지 않으나, 카메라 포즈 학습 파라미터의 초기화가 최적화 성능에 영향을 미치는 문제는 남아 있으므로 랜덤한 방법이 아닌 신중한 카메라 포즈 파라미터 초기화가 필요하다. 또한, 카메라 포즈 간 일관성 보장을 위하여 정규화 항으로 보완 하였지만, 카메라 위치가 크게 변화하는 큰 모션의 경우에는 부정확한 3D 공간추정이 이루어질 수 있다. 이와 더불어, 텍스처가 적거나, 반복 패턴이 있는 장면에서는 포즈 최적화가 어려워 품질 저하가 발생할 수 있다.

2.2 BARF(Bundle-Adjusting NeRF)

BARF[3]는 전통적인 bundle adjustment(BA)와 NeRF를 결합하여, 카메라 포즈를 신경망 학습 과정에서 동시에 최적화하는 방식을 제안하였다. BARF는 초기의 부정확한 카메라 포즈를 NeRF 학습 중에 점진적으로 바로잡음으로써, 포즈가 정확하지 않더라도 NeRF가 제대로 작동할 수 있도록 한다. 이를 위해 카메라 포즈 최적화 문제를 미분이 가능한 형태로 설정하고, NeRF의 MLP 학습과 동시에 카메라 포즈를 최적화하는 공동 최적화 프레임워크를 적용한다. 이러한 방식은 카메라 포즈의 작은 오차가 NeRF의 성능에 미치는 영향을 줄이면서, 보다 정확한 3D 장면 재구성을 가능하게 한다. BARF는 기존 NeRF 구조를 그대로 유지하면서도 학습 과정 중 포즈 보정이 가능한 대표적인 첫 시도로, 초기 포즈가 대략적으로 존재할 때 효과적인 성능을 보인다. 하지만, 카메라 포즈 파라미터의 초기화가 잘 못 주어지면 수렴이 어려운 상황에 빠질 수 있다. 이러한 이유로, BARF는 학습 안정화를 위하여 coarse-to-fine positional encoding 방법을 도입하였다. 이 방법은 높은 주파수의 positional encoding을 점진적으로 도입함으로써 최적화 수렴을 돕고, 학습이 로컬 미니마에 빠지는 위험성을 줄이는데 기여한다.

2.3 SC-NeRF(Self-Calibrating NeRF)

SC-NeRF[4]는 NeRF의 학습 과정에서 카메라 외부 파라미터 (포즈) 뿐만 아니라 내부 파라미터(초점 거리 등)와 비선형 왜곡을 동시에 최적화한다. 이를 위해 투영된 광선과 실제 이미지 간의 기하학적 일관성을 유지하는 'Projected Ray Distance Loss'를 도입하여, 복잡한 카메라 모델에서도 정확한 보정을 가능하게 한다. 이로 인하여, 사전 캘리브레이션 없이도 카메라의 내부 및 외부 파라미터를 자동으로 보정할 수 있고, 비선형 왜곡을 포함한 다양한 카메라 모델을 효과적으로 다룰 수 있다. 기하학적 일관성 유지를 위하여 적용한 'Projected Ray Distance Loss' 적용으로 정확한 3D 공간 재구성 결과를 얻을 수 있다. 하지만, 카메라 내부 파라미터까지 최적화 대상으로 확대함으로써, 학습 시간이 증가되고, 비선형 왜곡 모델링과 관련된 추가적인 구현 복잡성이 증가되었다.

2.4 NoPe-NeRF

NoPe-NeRF[5]는 큰 카메라 움직임에서 전통적인 NeRF 기법이 성능이 저하되는 문제를 해결하고자 제안되었다. DPT[6]를 통해 카메라별 단안 깊이 정보를 획득하고, 단안 깊이 정보를 스케일과 시프트 매개변수를 통해 왜곡을 바로잡는다. 이를 활용하여 연속적인 프레임 간의 상대적 포즈를 제한하였다. 또한, Chamfer Distance를 기반으로 한 포인트 클라우드 손실 함수를 통한 3D-3D matching을 사용함과 동시에 표면 기반 광도(Surfaced-based Photometric) 손실 함수를 이용하여 포인트 클라우드 손실 함수에서 발생하는 부정확한 matching을 줄여줌으로써 3D 장면 복원의 높은 성능과 포즈 추정 정확도를 모두 향상하게 시키며, 복잡한 카메라 경로에서도 우수한 성능을 보여준다. 위와 같은 NeRF와 카메라 포즈 최적화의 결합은 기존의 NeRF가 갖는 실용적 한계를 극복하는 데 중요한 역할을 하고 있으며, 이로 인해 장면 복원 과정에서 더욱 다양한 데이터 세트와 복잡한 환경을 처리할 수 있게 되었다.

NoPe-NeRF는 완전한 pose-free 학습이 가능하며, 단안 깊이 예측 기반 sparse point cloud를 사용하여 움직임이 큰 카메라 포즈 변화에도 3D 공간 재구성이 이루어질 수 있도록 하였다. 카메라 포즈와 NeRF가 통합적으로 학습하는 모델 중에 성능과 안정성 면에서 높은 평가를 받고 있다.

그러나 최근 연구[7]를 통하여 NoPe-NeRF에 의해 생성된 합성 이미지에서 카메라에서 가까운 전경 영역에 비해 원경 영역에 대한 이미지의 화질이 현격히 저하되는 현

상이 지적되었다. 이에 NoPe-NeRF에서 원경 영역의 화질 저하에 대한 원인 분석과 이에 대한 개선 방안의 모색이 요구된다.

3. Limitations of Chamfer Distance

Chamfer Distance(CD)는 포인트 클라우드 비교에 널리 사용되는 지표로, 최근접 거리 기반의 단순성과 계산 효율성으로 인해 딥러닝 기반 3D 재구성 및 정렬 분야에서 자주 활용된다, 하지만 CD는 여러 구조적 한계를 지니며, 이를 극복하고자 하는 다양한 시도가 다양한 연구들에서 이루어지고 있다.

Deng 등[8]은 포인트 클라우드 정렬에서 CD 기반 Iterative Closet Point(ICP)의 한계를 지적하였다. ICP는 두 점군 간의 정합을 위해, 각 반복 단계에서 한 점군의 각 점에 대해 다른 점군에서 최근 접점을 찾고, 그에 따라 전체 점군을 회전 및 평행이동시켜 점점 정렬을 맞춰가는 방식이다. 이 과정은 CD처럼 최근접 거리 기반의 대응 관계에 의존하기 때문에, 노이즈나 부분 겹침이 존재하는 경우 잘못된 매칭이 발생할 수 있고, 이로 인해 정합이 국소 최적해(local minima)에 머물게 되는 취약성이 존재한다. 이를 극복하기 위해 점 대신 임의의 직선을 사용하여 정렬 품질을 측정하는 손실 함수를 제안하였다. 이 접근은 확실한 대응 없이도 안정적인 정합을 가능하게 하며, 비지도 학습 환경에서도 높은 성능을 보였다.

Fan 등[9]은 단일 이미지로부터 3D 포인트 클라우드를 생성하는 초기 연구에서 CD와 Earth Mover's Distance(EMD)를 손실 함수로 비교하였다. CD는 계산 속도가 빠르다는 장점이 있지만, 불확실한 구조를 평균화해 표현하는 경향이 있으며, 그 결과로 세부 구조가 흐려지는 문제가 발생함을 지적하였다. 반면 EMD는 더 정밀한 형태 재현이 가능하지만 계산 비용이 높아 실용성이 떨어지는 것을 설명하였다.

Wu 등[10]은 CD가 점군의 지역 밀도 차이에 민감하지 않고, 이상치의 영향에 취약하다는 점에 주목하였다. 이들은 CD의 계산 방식을 보완한 Density-aware Chamfer Distance(DCD)를 제안하였다. DCD는 포인트 밀도 정보를 반영하여 세밀한 구조를 평가하며, 오차 범위를 제한하여 CD의 단점을 완화하였다. 실험에서는 기존 CD 및 EMD 기반 평가 지표보다 더 일관성 있는 평가를 제공하며, 손실 함수로도 효과적임을 보였다.

Lin 등[11]은 CD의 단점을 극복하기 위해 대조 학습 기반의 InfoCD 손실 함수를 제안하였다. CD가 점 분포의 정렬 능력이 부족하고 이상점에 민감하다는 점을 해결하기 위해, 매칭된 점쌍 간의 정보를 분산시키고, 추가로 곡면 유사도를 도입하여 예측 점군이 실제 구조를 더 균일하게 덮도록 유도하였다. InfoCD는 점군 완성 벤치마크에서 최고의 성능을 달성하며, CD 기반 손실보다 일관된 정밀도를 확보하였다.

이처럼 Chamfer Distance의 구조적 한계를 인식하고 이를 보완하려는 연구들은, 포인트 클라우드 정렬 및 재구성 모델의 정밀도와 안정성을 높이는 데 중요한 역할을 하고 있다.

4. Correspondence-Aware Losses for Camera Pose Optimization

최근 NeRF의 카메라 포즈 최적화 분야에서는, Chamfer Distance 기반의 정렬 손실이 갖는 한계를 극복하기 위해, 정확한 대응 쌍(correspondence)을 활용한 정렬 손실 함수에 대한 연구가 활발히 진행되고 있다. 이러한 접근은 특히 포즈 초기화가 부정확하거나 사전 포즈 정보가 없는 상황에서도 높은 정합 품질을 달성하는 데 효과적임이 입증되었다.

Hong 등[12]은 CoPoNeRF를 제안하며, 2D 대응 쌍 추정, 카메라 포즈 복원, NeRF 학습을 하나의 통합된 최적화 과정으로 결합하였다. 이 프레임 워크는 이미지 간 특징 대응(feature correspondence)을 바탕으로 포즈를 추정하고, 학습된 NeRF로부터 재투영된 픽셀 좌표를 활용하여 대응 쌍을 반복적으로 향상시키는 폐루프 최적화(closed-loop optimization) 구조를 구성한다. 이러한 방식은 대응 기반 손실이 NeRF의 포즈 수렴 안정성과 렌더링 정밀도 향상에 실질적인 기여를 할 수 있음을 실험적으로 보여주었다.

Truong 등[13]은 SPARF를 통해, 소수의 넓은 베이스라인 이미지만으로도 픽셀 간 대응 쌍을 활용한 다중 뷰 기하 정합 손실을 설계하여, NeRF와 카메라 포즈를 공동으로 최적화하였다. 이 방법은 포즈 초기화가 부정확한 경우에도 학습을 수행할 수 있으며, 새로운 시점에서의 렌더링 품질 또한 향상되는 결과를 보였다.

Chen 등[14]은 이미지 매칭을 통해 얻은 2D-3D 대응 쌍을 이용하여 PnP 기반의 단일 단계 포즈 추정(one-step pose estimation)을 수행하는 방식을 제안하였다. 렌더링 된 깊이 정보를 통해 3D 일관성을 검증함으로써 잘못된 대응 쌍을 제거하고, 정밀한 포즈 추정을 가능하게 한다.

Bian 등[15]가 제안한 PoRF는 MLP를 활용하여 카메라 포즈 잔차를 회귀하고, COLMAP으로부터 획득한 대응 쌍

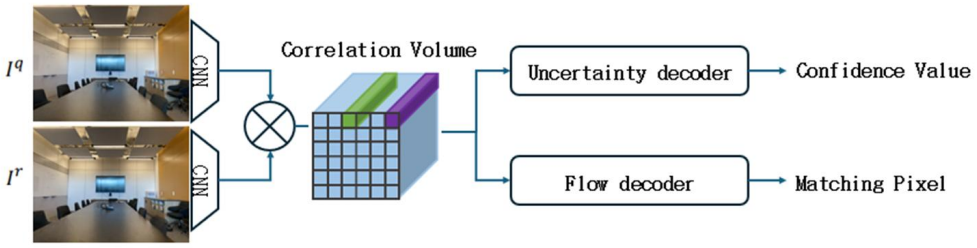

Fig. 3. The Architecture of PDC-Net[17]

을 활용한 에피폴라 기하 손실을 통해 포즈 학습의 감독 신호를 강화하였다. 실제 실험에서는 포즈 및 형상 정밀도의 현저한 향상을 달성하였다.

이와 같은 연구들은 Chamfer Distance 기반 손실 함수의 구조적 한계를 보완하기 위해, 대응 쌍 기반의 정렬 손실이 NeRF 학습 및 포즈 최적화 모두에서 효과적임을 실증적으로 제시한다.

III. The Proposed Scheme

본 절에서는 NoPe-NeRF 모델을 통해 생성된 합성 영상에서 나타나는 원경 영역의 화질 저하 현상이 기존 연구[7]에서 보고된 바와 같이 포인트 클라우드 손실과 밀접하게 연관되어 있다고 가정하고, 이러한 문제를 해결하기 위한 방안으로 포인트 클라우드 손실 계산에 사용되는 새로운 포인트 클라우드 집합의 구성 방법을 제안한다.

제안 방법에서는 새로운 포인트 클라우드 집합을 구성하기 위하여 확률적 밀집 대응 네트워크(PDCNet: Probabilistic Dense Correspondence Network)를 적용하였다. 또한, 기존 NoPe-NeRF에서 적용한 손실 함수 중 표면 기반 광도 손실 함수를 적용하지 않고, 최종 합성 이미지에서 원경과 전경 영역의 화질 변화를 관찰하여 제안 기법의 효과를 분석한다.

1. Probabilistic Dense Correspondence Network (PDC-Net)

밀집 대응(Dense Correspondence) 기술은 서로 다른 관측 시점에서 촬영된 영상 간의 정밀한 대응 관계를 추정함으로써, 이미지 정합이나 3차원 재구성 등의 핵심 전처리 과정에 활용된다. 특히 컴퓨터 비전 및 3D 비주얼 컴퓨팅 분야에서는 정확한 포인트 매칭이 후속 처리 성능에 영향을 미치기 때문에, 그 중요성이 더욱 강조된다.

최근에는 CNN(Convolutional Neural Network) 기반의 밀집 대응 네트워크가 도입되면서 기하학적 정합성 확보 측면에서 상당한 진전을 이루었으나, 조명 변화, 시점 이동, 이미지 노이즈 등 외부 환경 변화에 대한 민감성, 그리고 매칭 결과의 불확실성 정량화 한계와 같은 기술적 제약이 여전히 존재한다. 특히 매칭 신뢰도를 고려하지 못하는 경우, 정밀한 3D 재구성 과정에서 구조적 왜곡을 초래할 수 있다.

이러한 한계를 극복하기 위해 제안된 확률적 밀집 대응 네트워크(Probabilistic Dense Correspondence Network, PDC-Net)[17]는 각 매칭 결과에 대한 불확실성 추론 모듈을 통합하고, 다중 스케일 기반의 특징 상관 구조를 활용함으로써 기하학적 및 외형적 변형에 강인한 정합 결과를 도출할 수 있도록 설계되었다. 그 결과, Optical flow나 Feature Matching 방식보다 더 정확한 포인트 간 매칭을 가능하게 한다.

PDC-Net의 구성은 Fig. 3과 같다. 두 개의 입력 이미지는 VGG16 등의 CNN 백본을 통해 특징을 추출한 후, 생성된 상관 볼륨(Correlation Volume)은 불확실성 디코더(Uncertainty Decoder)를 통해 각 매칭 포인트의 신뢰도를 추론하며, 플로우 디코더(Flow Decoder)를 거쳐 최종적으로 정밀한 매칭 픽셀 쌍을 예측한다. 본 연구에서는 MegaDepth 데이터 세트[18]를 기반으로 사전 학습된 PDC-Net을 적용하여, 신뢰도 기반의 포인트 클라우드 생성을 수행하였다.

2. Modified NoPe-NeRF based on PDC-Net

Nope-NeRF는 포인트 클라우드의 정확성과 정합 안정성 측면에서 한계를 지닌다. 우선, 깊이 맵의 해상도를 축소한 상태에서 샘플링이 이루어지기 때문에, 공간적 세부 정보가 충분히 반영되지 못하며, 이는 포인트 클라우드의 정밀도를 저하시킨다. 또한, Chamfer Distance는 최근접 포인트 간 거리만을 기준으로 손실을 계산하기 때문에, 점밀도 불균형, 잡음 또는 부분 겹침 등의 상황에서 잘못된 매칭을 유도할 수 있으며, 이로 인해 NeRF의 카메라 포즈 정합 과정에서 오차가 누적될 수 있다[8][10]. 이를 보완하기 위해 적용되는 표면 기반 광도 손실 함수도 조명 변화,

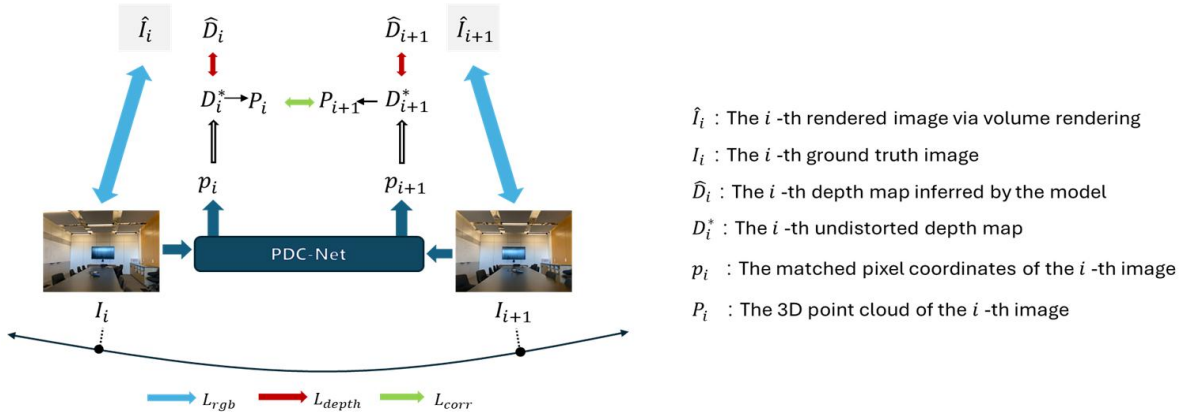

Fig. 4. Learning Process of NoPe-NeRF with a Probabilistic Dense Correspondence Network

질감, 반사 등에 민감하게 반응하기 때문에, 광도 기반 오차가 누적되어 추론 정확도에 부정적 영향을 줄 수 있다.

이러한 구조적 한계는 재구성된 영상의 품질을 저하시킬 수 있다. 이는 해상도가 낮아진 깊이 맵을 기반으로 포인트 클라우드의 깊이 값(Z값)을 결정함에 따라 결과적으로 부정확한 대응 포인트 클라우드 집합을 구축하게 되고, 이는 카메라 포즈 추정 오차를 크게 만들기 때문으로 추정된다. 이에 본 연구에서는 이러한 문제를 해결하기 위해, 확률적 밀집 대응 네트워크(PDC-Net)를 도입하여 두 이미지 간의 신뢰도 높은 밀집 대응 필드를 얻고, 이를 기반으로 정밀한 매칭 쌍 간의 거리 손실을 최소화하는 방식으로 NeRF의 카메라 포즈를 최적화하는 학습 구조를 제안한다.

선행 연구들 [17] [18]에서는 정확한 대응 쌍에 기반한 정렬 손실이 카메라 포즈 정합 성능을 유의미하게 향상시킬 수 있음을 실험적으로 입증한 바 있으며, 이는 본 연구의 접근 방식에 중요한 동기를 제공한다.

특히, PDC-Net은 강한 기하 정합성과 불확실성 기반 매칭 필터링 능력을 통해, Chamfer Distance와 같은 단순 근접 거리 기반 방식보다 더 정확하고 일관된 대응 관계를 유도할 수 있는 강력한 구조적 이점을 제공한다.

본 논문에서는 근접도가 높은 인접 이미지들로 구성된 데이터 세트[16]를 학습에 사용하였으며, PDC-Net 매칭 신뢰도 기준을 99% 이상으로 설정하여 고정밀 매칭 포인트 쌍을 추출하였다.

인접한 두 카메라 이미지 I_i와 I_j 사이에서 매칭된 픽셀은 M개의 2차원 픽셀 좌표로 구성된 p_i와 p_j로 표현된다.

$$p_i = \left\{(x_i^m, y_i^m) \mid m = 0, ..., M-1\right\},$$
$$p_j = \left\{(x_j^m, y_j^m) \mid m = 0, ..., M-1\right\}. \quad (2)$$

여기서 M은 매칭된 픽셀 쌍의 수를 의미하며, 각 픽셀 좌표는 동차좌표(homogeneous coordinates)로 변환된 후, DPT[6]로부터 대응 깊이 값을 추출하고 이를 Z값으로 정의하여 3차원 포인트 클라우드를 생성한다. 포인트 클라우드 P_i는 식(3)과 같이 정의된다.

$$P_i = D_i^*(p_i) \odot p_i^h \quad (3)$$

$D_i^*(p_i)$는 왜곡되지 않은 깊이 맵에서 해당 픽셀 위치 p_i의 깊이 값이며, p_i^h는 p_i의 동차좌표, $\odot$는 원소별 곱을 나타낸다. 생성된 포인트 클라우드 P_i와 P_j는 각각의 프레임에 대응하는 카메라 포즈 행렬 T_i, T_j를 통해 정렬되며, 상대적 정합 손실 함수 L_{corr}는 다음 식 (4)와 같이 정의된다.

$$L_{corr} = \sum_{(i,j)} norm(P_j, T_{ji} P_i) \quad (4)$$

여기서 $T_{ij} = T_j T_i^{-1}$는 P_i를 P_j의 좌표계로 변환하는 행렬이다. 정합 오차는 다음과 같은 식(5)으로 계산된다.

$$norm(P_j, P_i) = \sum_{m=0}^{M-1} \| P_{j,m} - P_{i,m} \|_2 \quad (5)$$

$P_{i,m}$은 P_i내의 m번째 매칭 포인트를 의미한다. 이러한 2D 기반 매칭을 통해 구성된 포인트 클라우드는 기존의 깊이 기반 3D 정합 방식보다 오정합의 가능성이 낮고, 구조적으로 신뢰도가 높은 포인트 셋 생성을 가능하게 한다. 따라서 기존 NoPe-NeRF에서 3D 포인트 매칭의 오류를 보정하기 위해 사용되던 표면 기반 광도 손실 함수를 제거하고도 안정적인 학습이 가능해진다.

최종적으로 본 제안 모델에서 모델 학습을 위하여 적용한 손실 함수는 식(6)과 같다.

$$L = \lambda_{rgb} L_{rgb} + \lambda_{depth} L_{depth} + \lambda_{corr} L_{corr} \quad (6)$$

여기서, L_{rgb}은 식(7)과 같이 모델 학습 결과 추론된 3차원 공간정보를 기반으로 볼륨 렌더링을 통하여 획득된 합성 이미지 $\hat{I}_i$와 Ground Truth 이미지에 I_i 대한 RGB

색상 차이를 나타내는 RGB 복원 손실 (RGB Reconstruction Loss)을 의미한다. 또한, L_{depth} 는 식(8)과 같이 사전 학습된 DPT[6]를 이용하여 획득된 단안 깊이 맵 D_i를 기반으로 연속된 이미지와의 연관성을 고려하여 왜곡이 보정된 깊이 맵 D_i^*와 모델 학습 결과 추론된 깊이 맵 $\hat{D}_i$간의 차이를 보여 주는 깊이 맵 복원 손실을 의미한다. L_{corr}는 식 (4) 에서 설명한 것과 같이 PDC-Net으로 획득한 매칭 3D 포인트 좌표 P_j와 카메라 포즈 행렬 T_i, T_j 를 기반으로 추론한 매칭 3D 포인트 좌표 $T_{ji}P_i$ 간의 차이를 보여주는 포인트 클라우드 손실을 의미한다. λ_{rgb}, λ_{depth}, λ_{corr} 는 각각 손실 항을 제어하기 위한 가중치들이다. 식 (7) 과 식 (8) 에서 N은 학습을 위해 사용된 입력 카메라 이미지 개수를 의미한다.

$$L_{rgb} = \sum_{i}^{N} \| I_i - \hat{I}_i \| \tag{7}$$

$$L_{depth} = \sum_{i}^{N} \| D_i^* - \hat{D}_i \| \tag{8}$$

Fig. 4는 본 논문에서 제안한 전체 학습 구조를 도식화한 것으로, PDC-Net을 중심으로 매칭 픽셀 예측부터 포인트 클라우드 생성, 손실 함수 최적화까지의 전체 흐름을 나타낸다.

IV. Experimental Results

1. Experimental Setup

본 연구의 실험은 Table 1과 같은 하드웨어 및 소프트웨어 환경에서 수행되었다. 학습에는 Tanks and Temples의 4개 장면(Ballroom, Church, Family, Francis) 장면을 활용하였다[16]. 각 장면은 일정한 프레임 간격으로 샘플링하여 학습 프레임과 정합용 프레임으로 분할하였으며, 모든 이미지는 960×540 해상도로 리사이징하였다.

Table 1. Experimental Environment

Item	Value
CPU	Intel Core i9-13900KF
RAM	128GB
GPU	NVIDIA RTX 4090 (24GB)
Framework	PyTorch 1.7.1, CUDA 11.8

모든 데이터 세트에 대해 학습률, 포즈 학습률, 왜곡 학습률은 동일하게 0.0005로 설정하였으며, 학습에는 두 개의 Adam 옵티마이저를 분리하여 사용하였다. 전체 학습은 10,000 epoch 동안 진행되었으며, PSNR(Peak Signal-to-Noise Ratio)이 30 epoch 동안 향상되지 않는 시점으로부터 학습률은 10 step마다 0.9954씩 곱하고, 포즈 학습률, 왜곡 학습률은 100 step마다 0.9씩 곱하는 step decay 스케줄러를 사용하였다. 각 이미지에서 1,024개의 픽셀을 무작위로 샘플링하고, 픽셀당 128개의 지점을 샘플링하여 볼륨 렌더링을 수행하였다.

손실 함수의 초기 가중치는 λ_{rgb} 는 1.0, λ_{depth}는 0.04, λ_{corr}은 1.0으로 설정되었다. 학습 초반에는 고정된 상태로 유지되며, 영상 품질 지표인 PSNR이 30 epoch 동안 향상되지 않을 경우, λ_{depth}와 λ_{corr}는 2000 epoch에 걸쳐 선형적으로 감소시켜 최종적으로 0으로 수렴하도록 하였다. 초기 카메라 단안 정보는 DPT로부터 직접 추출하여 정규화 없이 사용하였고, 카메라 포즈는 초기 외부 추정값으로 시작하여 학습을 통해 최적화되었다.

본 논문은 PDC-Net을 적용하여 정확한 포인트 클라우드 대응 쌍에 기반하여 카메라 포즈 추정 성능을 높임으로써 최종적으로 복원 합성 영상의 화질을 높이는 것에 초점을 맞추고 있다. 이에 실험에서는 추론 합성 영상의 화질 평가 및 카메라 포즈 추정 정확도 평가를 수행한다.

2. Evaluation Metrics

본 연구에서는 합성 영상의 품질을 정량적으로 평가하기 위해 PSNR(Peak Signal-to-Noise Ratio), SSIM(Structural Similarity Index Measure), LPIPS(Learned Perceptual Image Patch Similarity) 세 가지 지표를 사용하였다. PSNR은 원본 이미지와 합성 이미지 간의 픽셀 단위의 정밀도를 측정하는 지표로, 값이 높을수록 화질 왜곡이 적고 원본에 가까움을 의미한다. 그러나 PSNR은 구조적 왜곡이나 시각적 일관성 측면에서는 인간 지각과의 일치도가 낮다는 한계가 있어[19], 일반적으로 SSIM이나 LPIPS와 같은 지표와 함께 병행하여 사용된다. SSIM은 영상의 구조적 유사도를 평가하며, 명암, 대비, 구조 정보를 바탕으로 두 영상 간의 시각적 일관성을 측정한다[20]. LPIPS는 학습 기반 시각 유사도 지표로 인간의 시각적 인지 특성과 유사한 방식으로 이미지 간의 차이를 평가하며, 값이 낮을수록 두 영상 간의 시각적 차이가 적음을 의미한다[21].

본 연구에서는 합성 영상 품질 측정 외에도 카메라 포즈 추정의 정확도를 정량적으로 평가하기 위해 ATE(Absolute Trajectory Error)와 RPE(Relative Pose Error)를 사용하였다. ATE는 예측된 카메라 포즈와

Table 2. Comparison of synthesized image quality metrics across datasets.

Dataset	Ours			Nope-NeRF		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Ballroom	**21.17**	**0.63**	**0.36**	21.12	0.61	0.41
Church	**23.62**	**0.71**	**0.38**	23.18	0.67	0.43
Family	**23.03**	**0.70**	**0.39**	22.41	0.65	0.49
Francis	23.79	0.68	**0.39**	**24.06**	**0.69**	0.43

GT(Ground Truth) 카메라 포즈 간의 위치 오차를 계산하며, 전체 궤적에서의 누적적인 위치 차이를 측정한다[22]. 값이 작을수록 실제 경로와의 정합 정도가 높음을 의미한다. RPE는 프레임 간 상대적인 이동 변화를 기반으로 예측 포즈의 기하학적 일관성을 평가한다[22]. 즉, 프레임 i와 i+1 사이의 상대적인 위치(RPE_t)/회전 변화(RPE_r)가 GT와 얼마나 유사한지를 측정하며, 일반적으로 위치 오차와 회전 오차로 나뉘어 계산된다. ATE와 RPE는 둘 다 낮을수록 더 우수한 정합 품질을 의미한다.

3. Experimental Results

Table 2는 Tanks and Temples 데이터 세트의 4개 장면(Ballroom, Church, Family, Francis)에 대해, 제안 기법과 Nope-NeRF의 합성 영상 품질을 PSNR, SSIM, LPIPS 지표를 기준으로 비교한 결과를 보여준다.

Ballroom, Church, Family 세 장면에서는 제안 기법이 PSNR과 SSIM에서 Nope-NeRF보다 높은 값을 기록하며, 픽셀 기반 정밀도 및 구조적 일관성 측면에서 개선된 성능을 보였다. 특히 LPIPS에서도 Ballroom(0.41→0.36), Church(0.43→0.38), Family(0.49→0.39) 모두에서 낮은 값을 기록, 사람의 시각적 인지 기준에서도 품질 향상을 확인하였다.

한편, Francis 장면의 경우 PSNR과 SSIM에서 Nope-NeRF가 다소 높은 수치를 기록하였으나, 두 지표 간의 차이는 상대적으로 미미한 수준에 그쳤다. 반면, LPIPS 지표에서는 제안 기법이 Nope-NeRF보다 낮은 값을 기록함으로써, 시각적 일관성과 지각 기반 품질 측면에서 오히려 더 우수한 성능을 보였다. 이는 Francis 장면에서도 제안 기법이 구조적 왜곡을 효과적으로 억제하며, 전반적인 영상 품질을 안정적으로 유지했음을 시사한다.

이러한 결과는 다양한 장면에서 제안 기법이 보다 선명하고 구조적으로 일관된 이미지 생성을 가능하게 함을 보여주며, 특히 LPIPS의 일관된 개선을 통해 시각적 품질 향상 효과를 정량적으로 확인할 수 있다.

Table 3은 제안 기법과 Nope-NeRF의 카메라 포즈 추정 성능을 ATE, RPE_r, RPE_t 지표를 기준으로 비교한 결과를 제시한다. 모든 장면에서 제안 기법은 ATE, RPE_r, RPE_t에서 Nope-NeRF보다 일관되게 낮은 값을 기록하며, 전반적인 카메라 포즈 정합의 정확도와 안정성 측면에서 우수한 성능을 입증하였다.

특히 Church 장면에서는 ATE가 0.178에서 0.013으로, RPE_t는 0.196에서 0.043으로 크게 감소하였으며, 이는 경로 기반 정합뿐만 아니라 단일 프레임 간 이동 정합의 정밀도가 크게 향상되었음을 시사한다. 또한 Family 장면의 RPE_r은 Nope-NeRF의 1.014에서 제안 기법의 0.009로 급감하여, 짧은 구간 간 회전 정합 정확도의 극적인 개선을 보여준다. Ballroom과 Francis 장면에서도 RPE_r, RPE_t 모두에서 일관된 오차 감소가 관찰되었으며, 제안 기법이 다양한 장면에서도 견고한 포즈 최적화 성능을 안정적으로 유지함을 확인할 수 있다. 이와 같은 결과는, Chamfer Distance 기반의 불확실한 3D 정합 대신, 제안한 확률적 밀집 대응 기반 매칭 방식을 활용함으로써 더욱 정확하고 신뢰성 있는 카메라 포즈 추정이 가능함을 보여준다.

정량적 결과와의 일관성을 확인하기 위해 GT, 제안 기법, 그리고 Nope-NeRF의 합성 결과를 비교하였으며, Fig. 5는 Church와 Family 장면에서 각 기법 간의 시각적 차이를 보여준다.

Church 장면에서는 전경과 원경 모두에서 의자 배열의 표현 차이가 두드러진다. Nope-NeRF는 의자 간 경계가

Table 3. Comparison of camera pose estimation performance across datasets.

Dataset	Ours			Nope-NeRF		
	ATE ↓	RPE_r ↓	RPE_t ↓	ATE ↓	RPE_r ↓	RPE_t ↓
Ballroom	**0.003**	**0.023**	**0.048**	0.003	0.069	0.061
Church	**0.013**	**0.009**	**0.043**	0.178	0.057	0.196
Family	**0.001**	**0.009**	**0.032**	0.006	1.014	0.048
Francis	**0.003**	**0.008**	**0.036**	0.008	0.346	0.083

Fig. 5. Visual comparison of synthesized results for Church and Family scenes.

흐릿하게 표현되어 배열 간격의 일관성이 무너지는 경향을 보이는 반면, 제안 기법은 전경과 원경에서 모두 의자의 배열 간격과 형태를 정밀하게 복원하였다.

Family 장면에서는 전경의 동상과 배경의 나무 경계 표현에서 품질 차이가 확인된다. Nope-NeRF는 동상 표면의 윤곽 디테일이 손실되고, 배경의 나무와 건물 외곽이 번지거나 흐릿하게 표현되는 경향이 나타난다. 이에 반해, 제안 기법은 동상의 형태와 표면을 더 정확하게 보존하고 있으며, 배경의 나무와 건물 경계 또한 선명하게 복원되어 시각적 일관성이 향상되었음을 확인할 수 있다.

이러한 비교는 앞선 정략적인 지표에서의 개선 결과와도 일치하며, 제안 기법이 Nope-NeRF의 주요 한계인 원경 영역에서의 화질 저하 문제를 효과적으로 완화하였을 뿐만 아니라, 전경과 원경 전반에 걸쳐 시각적 품질을 향상시켰음을 입증한다.

V. Conclusions

본 연구는 NeRF 기반 모델에서 발생할 수 있는 부정확한 포인트 정합 문제를 개선하기 위해, 확률적 밀집 대응 네트워크를 활용한 새로운 정합 구조를 제안하였다. 제안된 방법은 두 이미지 간의 신뢰도 높은 매칭 픽셀을 추론하고, 이를 단안 깊이 정보와 결합하여 정밀한 포인트 클라우드를 생성함으로써, 기존의 Chamfer Distance 기반 매칭 방식에서 발생할 수 있는 오차를 효과적으로 줄였다. 또한, 정확한 매칭 포인트 클라우드 셋을 사용함에 따라 불필요한 손실 항목인 광도 기반 손실을 제거함으로써 학습 구조를 단순화하면서도, 포즈 정합의 정확도는 오히려 향상되도록 하였다.

실험 결과, 일부 장면에서 PSNR은 소폭 감소하였으나, LPIPS는 대부분 개선되었으며, 모든 장면에서 ATE와 RPE 지표가 하락하는 양상을 보여 제안된 구조의 실효성을 확인할 수 있었다.

특히 본 구조는 기존 NeRF 프레임워크에 직접적인 아키텍처 변경 없이 통합 가능하다는 점에서 실용적인 확장성이 높다. 향후, 카메라 포즈과 생성 깊이맵의 지역적 일관에 대한 분석을 추가적으로 수행하고, 이를 기반으로 다양한 단안 깊이 추정 모델이나 비선형 카메라 모델과의 결합을 통해 더 일반화된 정합 구조로 발전시킬 수 있을 것으로 기대된다. 이러한 결과는 NeRF 기반 모델의 정합 정확도와 학습 효율성 개선에 기여하며, 후속 연구의 유의미한 기반이 될 수 있을 것이다.

REFERENCES

[1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, R. (n.d.) Ng, “Representing Scenes as Neural Radiance Fields for View Synthesis,” ECCV 2020. DOI:

10.1007/978-3-030-58452-8_24

[2] Zirui WANG, S. Wu, W. Xie, M. Chen, V. A. Prisacariu, " NeRF--: Neural radiance fields without known camera parameters," arXiv preprint arXiv:2102.07064, 2021.

[3] C. LIN, W. Ma, A. Torralba, S. Lucey, "Barf: Bundle-adjusting neural radiance fields," Proceedings of the IEEE international conference on computer vision(ICCV), pp. 5741-5751, 2021. DOI: 10.1109/ICCV48922.2021.00569

[4] Y. Jeong, S. Ahn, C. Choy, A. Anandkumar, M. Cho, J. Park, "Self-calibrating neural radiance fields," Proceedings of the IEEE/CVF International Conference on Computer Vision(ICCV). pp. 5846-5854, 2021. DOI: 10.1109/ICCV48922.2021.00579

[5] W. Bian, Z. Wang, K. Li, J. Bian, V. A. Prisacariu, "NoPe-nerf: Optimising neural radiance field with no pose prior," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). pp. 4160-4169, 2023. DOI: 10.1109/CVPR52729.2023.00405

[6] R. Ranftl, A. Bochkovskiv, V. Koltun, "Vision Transformers for Dense Prediction," Proceeding of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), pp. 12179-12188, 2021. DOI: 10.1109/ICCV48922.2021.01196

[7] Y. Kwon "Improving Neural Radiance Field performance through point cloud loss adjustment," Master's thesis, Seoul Media Institute of Technology, 2024.

[8] Z. Deng, Y. Dou, J. Wang, Q.-Y. Zhou, Y. Yuan, and H. Huang, "A robust loss for point cloud registration," Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), p. 6138-6147. 2021. DOI: 10.1109/ICCV48922.2021.00608

[9] H. Fan, H. Su, and L. Guibas, "A point set generation network for 3D object reconstruction from a single image," Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 605–613, 2017. DOI: 10.1109/CVPR.2017.264

[10] T. Wu, Y. Xia, M. Zhang, L. Liu, J. Yu, and X. Chen, "Density-aware chamfer distance as a comprehensive metric for point cloud completion," arXiv preprint arXiv:2111.12702, 2021.

[11] F. Lin, Q. Liu, X. Song, J. Yao, and Z. Deng, "InfoCD: A contrastive Chamfer distance loss for point cloud completion," Advances in Neural Information Processing Systems, 36: 76960-76973. 2023. DOI: 10.48550/arXiv.2306.00908

[12] S. Hong, D. Wu, Z. Shen, Y. Zhong, and J. Yu, "Unifying correspondence, pose, and NeRF for generalized pose-free novel view synthesis," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 20196-20206, 2024. DOI: 10.1109/CVPR52733.2024.01909

[13] P. Truong, M. Danelljan, R. Timofte, and L. Van Gool, "SPARF: Neural radiance fields from sparse and noisy poses," IEEE. In: CVF Conference on Computer Vision and Pattern Recognition, CVPR, p. 6, 2023. DOI: 10.1109/CVPR52729.2023.00408

[14] J. Chen, D. Yu, S. Yang, and J. Wang, "Marrying NeRF with feature matching for one-step pose estimation," Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), p. 7302-7309, 2024. DOI: 10.1109/ICRA57147.2024.10610766

[15] W. Bian, X. Xu, K. Li, J.-W. Bian, and V. A. Prisacariu, "PoRF: Pose residual field for accurate neural surface reconstruction," arXiv preprint arXiv:2310.07449, 2023.

[16] Tanks and Temples Benchmark Dataset, *https://www.tanksandtemples.org/download/*.

[17] P.Truong, M. Danelljan, L. Van, R. Timofte, "Learning accurate dense correspondences and when to trust them," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). p. 5714-5724, 2021. DOI: 10.1109/CVPR46437.2021.00566

[18] Z. Li, N. Snavely, "MegaDepth: Learning Single-View Depth Prediction from Internet Photos," Proceedings of the IEEE conference on computer vision and pattern recognition. p. 2041-2050, 2018. DOI: 10.1109/CVPR.2018.00218

[19] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," IEEE Signal Processing Magazine, vol. 26, no. 1, pp. 98-117, Jan. 2009. DOI: 10.1109/MSP.2008.930649

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004. DOI: 10.1109/TIP.2003.819861

[21] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 586-595, Salt Lake City, USA, Jun. 2018. DOI: 10.1109/CVPR.2018.00068

[22] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 573-580, Vilamoura, Portugal, Oct. 2012. DOI: 10.1109/IROS.2012.6385773

Authors

Jong-Su Ha received his B.S. degree in Applied Mathematics from Hanyang University(ERICA), Korea, and his M.S. degree in AI Software Engineering from Seoul Media Institute of Technology(SMIT), Korea. His research interests include Computer Vision, Deep Learning, Vision AI and Large Language Models (LLMs).

Young-Soo Kwon received his M.S. degree in AI Software Engineering from Seoul Media Institute of Technology(SMIT), Korea. He is currently a research worker in CAD lab. He is interested in computer graphics, NeRF, deep learning, and LLM applications using RAG, LangChain, and LLM API.

Jin-Ah Kim received her B.S., M.S., and Ph.D. degrees in Computer Engineering from Hoseo University, Korea. She is currently an Assistant Professor in the Department of AI Software Engineering at Seoul Media Institute of Technology (SMIT). Her research interests include Big Data Processing & Analysis, Time Series Analysis, Deep Learning, and Multimodal Learning.

Yoo-Joo Choi received the B.S., M.S. and Ph.D. degrees in Computer Science and Engineering from Ewha Womans University, Korea, in 1989, 1991 and 2005, respectively. She is currently a professor in the Department of AI Software Engineering at Seoul Media Institute of Technology. She is interested in computer vision, computer graphics, augmented reality, generative AI and human-computer interaction.