

NeRF의 정확한 3차원 복원을 위한 거리-엔트로피 기반 영상 시점 선택 기술

Distance and Entropy Based Image Viewpoint Selection for Accurate 3D Reconstruction with NeRF

최진원¹ · 서찬호¹ · 최준혁¹ · 최성록[†]

Jinwon Choi¹, Chanho Seo¹, Junhyeok Choi¹, Sunglok Choi[†]

Abstract: This paper proposes a new approach with a distance-based regularization to the entropy applied to the NBV (Next-Best-View) selection with NeRF (Neural Radiance Fields). 3D reconstruction requires images from various viewpoints, and selecting where to capture these images is a highly complex problem. In a recent work, image acquisition was derived using NeRF's ray-based uncertainty. While this work was effective for evaluating candidate viewpoints at fixed distances from a camera to an object, it is limited when dealing with a range of candidate viewpoints at various distances, because it tends to favor selecting viewpoints at closer distances. Acquiring images from nearby viewpoints is beneficial for capturing surface details. However, with the limited number of images, its image selection is less overlapped and less frequently observed, so its reconstructed result is sensitive to noise and contains undesired artifacts. We propose a method that incorporates distance-based regularization into entropy, allowing us to acquire images at distances conducive to capturing both surface details without undesired noise and artifacts. Our experiments with synthetic images demonstrated that NeRF models with the proposed distance and entropy-based criteria achieved around 50 percent fewer reconstruction errors than the recent work.

Keywords: 3D Reconstruction, Next-Best-View, NeRF, Entropy-based Uncertainty

1. 서론

3차원 복원(3D reconstruction)은 비디오나 여러 장의 이미지로부터 물체나 공간의 3차원 형태를 획득하는 것으로 컴퓨터 비전이나 그래픽스를 비롯하여 환경 지도 구축과 같이 로봇 공학에서도 많이 연구되는 분야이다. 3차원 복원을 위해서 다

양한 각도와 위치에서 촬영한 이미지가 필요한데, 이는 물체나 공간의 특징과 카메라 스펙에 대한 이해와 경험이 필요하므로 초보자가 쉽게 고품질의 3차원 복원 결과를 얻기 힘들다.

최근 이러한 과정을 자동화하기 위한 연구가 활발히 진행되고 있다. 특히, 로봇 기술을 활용하여 카메라 이동 및 이미지 획득 과정을 자동화하고, 여러 알고리즘들을 통해 고품질의 3차원 복원을 위해 최적의 이미지를 획득하고자 하는 연구들이 많이 진행되었다.

대표적인 연구 방법 중 하나는 Next-Best-View (NBV) 선택이다. NBV 선택 문제는 로봇이나 드론에 탑재된 카메라와 같은 센서가 주어진 공간이나 물체를 효율적으로 관찰하기 위해 어떤 시점을 선택해야 하는지 결정하는 문제이다. 즉 센서의 시야 내에서 가장 중요하거나 정보가 많은 부분을 획득함으로써, 제한된 관찰 횟수나 시간 내에 정확한 데이터 수집을 가능하게 하는 것이다. NBV 기법들은 새로운 최적의 관측 위치를

Received : Oct. 31. 2023; Revised : Dec. 5. 2023; Accepted : Dec. 18. 2023

※ This research was supported by MSIT/NRF Grant for Bridge Convergence R&D Program (AI-based Localization and Path Planning on 3D Building Surfaces; 2021M3C1C3096810) and CHA/NRICH Grant for a R&D Program (Development of Ultra-High Resolution Gigapixel 3D Data Generation Technology; 2021A02 P02-001)

1. Master's Student, Computer Science and Engineering Department, Seoul National University of Science and Technology (SEOULTECH), Seoul, Korea (randomwons, seochanho0717, dkwngur12@gmail.com)

† Assistant Professor, Corresponding author: Computer Science and Engineering Department, Seoul National University of Science and Technology (SEOULTECH), Seoul, Korea (sunglok@seoultech.ac.kr)

결정하기 위해 다양한 선택 기준을 활용하는데, 정보 이득(information gain), 엔트로피(entropy) 등을 활용한 여러 연구들^[1-3]이 지속적으로 이루어져 왔다.

최근 인공지능망을 활용한 Neural Radiance Fields (NeRF)^[4]는 3차원 영상 표현 및 복원 분야에서 가장 활발하게 연구되고 있는 방법이다. NeRF 또한 다양한 시점에서의 이미지를 입력 받아 3차원 모델을 생성하고, 효율적인 이미지 획득이 필요하기 때문에 NBV 기법을 적용한 다양한 연구들^[5-7]이 진행되었다. Lee et al.^[5]은 NeRF의 ray-sampling에서 가중치 분포를 엔트로피 개념의 불확실성으로 표현하고 이를 NBV의 선택 기준으로 이용하여 새로운 시점을 결정하였다. Jin et al.^[6]은 학습이 필요한 네트워크를 이용하여 시점에서의 불확실성을 이용하여 NBV의 선택 기준으로 이용하는 방법을 제안하였다. Ran et al.^[7]은 렌더링된 물체로부터 RGB 이미지와 깊이 이미지를 이용한 3차원 복원을 수행한 후, 불확실성을 측정하여 NBV의 선택 기준으로 이용한다.

위의 선행 연구들은 모두 NBV의 후보 시점을 미리 생성하고, 생성된 시점에서 물체 복원의 불확실성을 측정하여 다음 최적 시점을 결정한다. 그러나 후보 시점 생성 방식에서 선행 연구 모두 카메라와 물체 사이의 거리를 고려하지 않았다. 같은 물체를 촬영하더라도 촬영 거리에 따라 촬영된 물체의 관찰 영역과 그 디테일에 차이가 있고 이에 따라 복원된 결과가 크게 달라진다. 따라서 본 연구에서는 기존 선행 연구의 문제점을 보다 확장하고 일반화하여 카메라의 시점에 촬영 거리를 고려하여 NBV를 선택하는 방법을 제안한다.

2장에서는 본 연구와 관련된 주요 배경 이론에 대해서 소개한다. 3장에서는 본 논문에서 제안하는 거리-엔트로피 기반의 불확실성 지표에 대해서 설명한다. 4장에서는 제안한 방법이 효과적인지 입증하기 위해 3가지 합성 데이터셋에 대한 실험 결과를 보여준다. 5장에서는 본 연구의 결론을 요약하고 향후 연구 계획에 대해 논의한다.

2. 배경 이론

2.1 NBV (Next-Best-View)

NBV 선택 문제는 카메라 등의 센서가 주어진 환경에서 효율적인 정보 수집을 위해 어떤 시점을 선택하여야 하는지 결정하는 문제이다. 이 문제는 3차원 물체 복원에서 카메라가 어떤 위치와 방향각에서 이미지를 획득하여야 효율적인 3차원 물체 복원이 가능한지 결정하는 과정과 연관^[8]되어 있다.

NBV 선택 기법은 크게 두 가지 측면에서 분석해 볼 수 있다. 첫 번째 측면은 최적 시점을 선정하는 지표이다. 최적의 시

점을 선정하는 기준으로 정보 이득(information gain), 엔트로피 등이 이용되며 이러한 지표들은 특정 시점에서 모델에 대한 정보나 불확실성이 얼마나 되는지를 정량적으로 표현한다. 예를 들어 엔트로피는 해당 시점에서의 불확실성의 정도를 표현한다^[11]. 두 번째 측면은 앞선 지표를 바탕으로 최적의 시점을 선택하는 과정이다. 카메라의 시점, 즉 위치와 방향각은 연속적인 공간에서 표현된다. 그러나 대부분의 연구는 연속적인 공간에서의 최적화 대신, 샘플링된 후보(candidate) 시점을 미리 생성하여, 정해진 후보 시점 중 최적의 시점을 결정하는 조합 최적화의 문제로 NBV 선택의 문제를 해결한다.

2.2 NeRF (Neural Radiance Fields)

NeRF는 인공 신경망을 이용하여 복잡한 3차원 장면을 사실적으로 렌더링하는 방법이다^[4]. NeRF는 여러 각도에서 촬영된 사진들을 학습 데이터로 이용하여 공간 내 각 위치의 관찰의 방향에 따른 색상과 공간의 밀도를 학습한다. NeRF는 인공 신경망이 연속적인 공간을 부드럽게 표현하는 장점을 이용하여 고해상도의 3차원 장면을 복원할 수 있다. 또 NeRF는 점군(point cloud)이나 복셀(voxel)과 같은 기존 3차원 표현 방법들에 비해 더 적은 용량으로 고해상도의 3차원 모델을 표현할 수 있어 최근 큰 관심을 받고 있다.

NeRF는 다층 인공 신경망(multi-layer perceptron; MLP)으로 $(\mathbf{c}, \sigma) = \text{MLP}(\mathbf{X}, \theta)$ 와 같이 공간 내 각 위치 $\mathbf{X} \in \mathbb{R}^3$ 와 관찰 방향 $\theta \in \mathbb{R}^2$ 를 입력으로 하고, 해당 위치에서의 색상 \mathbf{c} 와 밀도 σ 를 출력으로 한다. 한 시점 \mathbf{v} 에서 획득된 이미지의 각 픽셀은 하나의 광선 \mathbf{r} 로 표현된다. 즉 광선 \mathbf{r} 은 시점의 위치 $\mathbf{o} \in \mathbb{R}^3$ 와 해당 픽셀로의 방향 $\mathbf{d} \in \mathbb{R}^3$ 의 합으로 $\mathbf{r} = \mathbf{o} + t\mathbf{d}$ 로 표현된다. NeRF는 직선 \mathbf{r} 위의 여러 개의 3차원 위치 \mathbf{X} 를 샘플링하여 MLP의 입력으로 넣는다.

NeRF는 학습과 새로운 시점의 영상을 렌더링하는데 볼륨 렌더링(volumetric rendering)을 사용한다. 볼륨 렌더링은 영상 내 각 픽셀에 입사된 광선 위의 각 점의 투과율 T 과 밀도 σ , 색상 \mathbf{c} 를 누적하여 해당 픽셀의 RGB 색상 \mathbf{C} 를 결정하는 방법으로 식 (1)과 같이 계산한다.

$$\mathbf{C}(\mathbf{r}) = \int_0^{\infty} T(t)\sigma(\mathbf{r}(t)) \mathbf{c}(\mathbf{r}(t), \mathbf{d}) dt \quad (1)$$

이때, 함수 $T(t)$ 는 t 까지 누적된 광선의 누적투과율로 식 (2)와 같다.

$$T(t) = \exp\left(-\int_0^t \sigma(\mathbf{r}(s)) ds\right) \quad (2)$$

이는 광선이 t 까지 어떤 입자와도 부딪히지 않을 확률과 같다. 이를 이산적으로 표현하면 식 (3)과 같다.

$$T(t) = \exp\left(-\sum_i \sigma_i \delta_i\right) \quad (3)$$

$\delta_i = t_{i+1} - t_i$ 이고, 이는 인접한 두 샘플 사이의 거리이다. 따라서 식 (1)과 같이 연속된 광선 위에 정의된 이미지에 렌더링된 광선의 색상은 이산적으로 식 (4)와 같이 정리된다.

$$\mathbf{C}(\mathbf{r}) = \sum_{i=1}^n T_i (1 - \exp(-\sigma_i \delta_i)) \mathbf{c}_i \quad (4)$$

$T_i = \exp(-\sum_{j=1}^{i-1} \sigma_j \delta_j)$ 로 표현하였다. 또 렌더링된 색상 $\mathbf{C}(\mathbf{r})$ 은 가중치 $w_i = T_i (1 - \exp(-\sigma_i \delta_i))$ 에 대한가중합으로 식 (5)와 같이 간략하게 표현된다.

$$\mathbf{C}(\mathbf{r}) = \sum_{i=1}^n w_i \mathbf{c}_i \quad (5)$$

2.3 Ray-Based Volumetric Uncertainty

Lee et al.^[5]에서는 NeRF 학습을 위한 NBV 선택 지표로 ray-based volumetric uncertainty를 사용한다. 해당 연구에서는 3차원 복원 결과의 불확실성을 표현하기 위해 Shannon 엔트로피를 사용한다. Shannon 엔트로피^[6]는 정보 엔트로피로도 불리며, 정보의 양을 측정하는 방법이다. 이는 랜덤변수 X 에 대해 각 사건 x_i 가 가지는 불확실성의 합을 나타내며 식 (6)과 같이 표현한다.

$$H(X) = -\sum_{i=1}^n P(x_i) \log P(x_i) \quad (6)$$

$P(x_i)$ 는 사건 x_i 가 일어날 확률이다. 식 (6)은 랜덤변수 X 의 이산확률분포의 불확실성을 표현할 수 있다.

식 (2)는 누적투과율로 광선이 t 까지 이동할 때 어떠한 입자에도 부딪히지 않을 확률이므로, Lee et al.^[5]에서는 불투명도 함수 $O(t) = 1 - T(t)$ 로 정의한다. T 는 누적투과율로 단조증가함수이고, $t=0$ 일 때 $O=0$, $t=\infty$ 일 때 $O=1$ 이므로, 불투명도 함수 O 는 누적분포함수로 정의할 수 있다. 따라서, 누적분포함수를 미분한 $\frac{dO(t)}{dt}$ 는 확률밀도함수가 된다. 이는 식 (7)과 같이 정리할 수 있다.

$$\frac{dO(t)}{dt} = \frac{d(1 - T(t))}{dt} = T(t)\sigma(\mathbf{r}(t)) \quad (7)$$

이는 식 (5)에서의 가중치 w 이며, 따라서 광선 \mathbf{r} 의 가중치 분포의 불확실성을 식 (8)과 같이 나타낸다.

$$H(W_r) = -\sum_{i=1}^n P(w_i) \log P(w_i) \quad (8)$$

또한 시점 V 의 N 개의 모든 광선 \mathbf{r}_j 에 대한 평균 엔트로피는 식 (9)와 같이 나타낸다.

$$H(V) = \frac{1}{N} \sum_{j=1}^N H(W_{\mathbf{r}_j}) \quad (9)$$

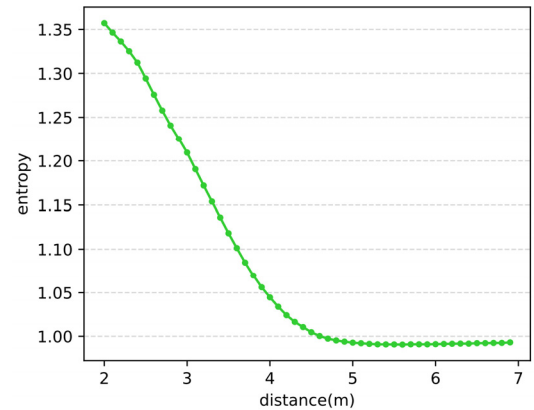
3. 거리기반 엔트로피를 이용한 NBV

본 연구에서는 기존의 NeRF 기반 NBV 선택 기법에서 사용되는 엔트로피 기반의 지표에 물체와 카메라 사이의 촬영 거리를 추가로 고려하도록 하는 방식을 제안한다.

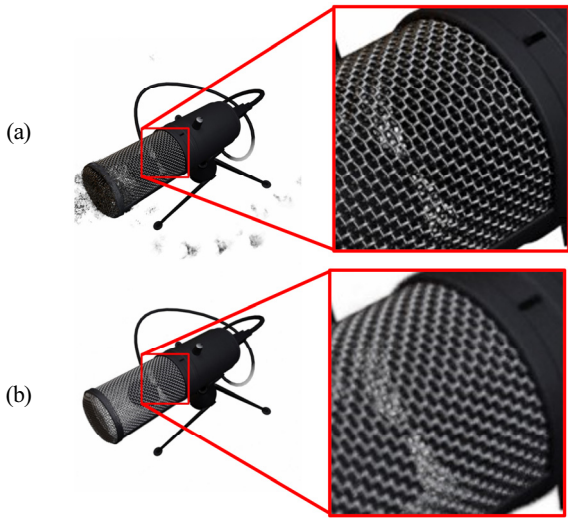
3.1 단순 엔트로피 기반 NBV의 한계

3차원 모델을 고품질로 복원하기 위해서는 다양한 시점에서 이미지를 획득하는 것이 필요하다. 이때 우리는 더욱 정밀한 표면 디테일을 얻기 위해 가까운 거리에서 이미지를 획득하기도 한다.

앞서 설명한 ray-based volumetric uncertainty도 이러한 직관을 그대로 반영한다. 즉 가까운 거리에서 촬영한 이미지의 각 광선은 더욱 명확한 표면 정보를 표현하고, 더욱 높은 엔트로피 값을 가진다. [Fig. 1]은 촬영 거리에 따른 식 (9)로 표현된 엔트로피 값을 실험적으로 나타낸 것이다. [Fig. 1]에서 가까운 촬영 거리에서의 엔트로피가 먼 촬영 거리에서의 엔트로피보



[Fig. 1] Relationship between capturing distance and entropy: As the capturing distance is closer, their corresponding entropy becomes larger

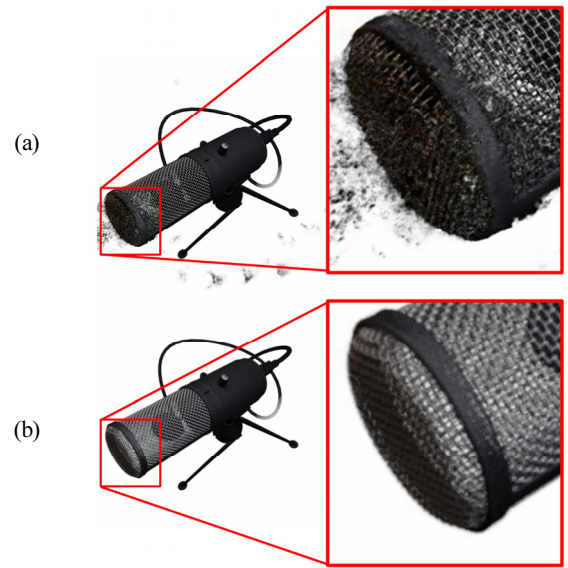


[Fig. 2] Two synthesized images from two NeRF models with (a) images captured in *closer* distance and (b) images captured in *farther* distance: The synthesized image from (a) exhibits better surface reconstruction

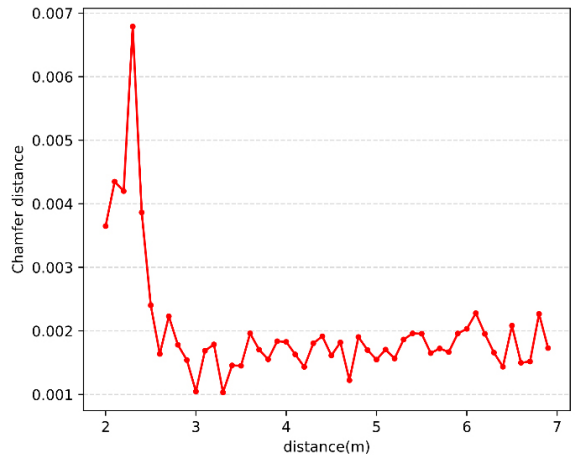
다 높은 것을 관찰할 수 있다. 따라서 Lee et al.^[5]와 같이 엔트로피만을 이용하여 카메라의 시점을 선택하는 경우, 물체 표면의 디테일이 더욱 잘 표현된 모델을 학습시킬 수 있다. [Fig. 2]는 각각 가까운 촬영 거리와 상대적으로 먼 촬영 거리의 이미지를 각각 이용하여 학습한 NeRF 모델로 생성한 렌더링 결과이다. [Fig. 2]에서 가까운 촬영 거리의 이미지를 이용한 NeRF의 렌더링 결과가 더욱 명확하고 자세하게 물체의 표면을 표현함을 관찰할 수 있다.

그러나 가까운 거리에서 촬영한 이미지는 NeRF 학습 결과에 좋은 점만 주는 것은 아니다. 동일한 개수의 이미지를 사용하는 경우, 가까운 거리에서 촬영한 이미지 셋은 먼 거리에서 촬영된 이미지에 비해 관찰된 물체의 영역간 겹침(overlap)이 적다. 관찰 영역의 겹침이 적다는 점은 물체의 한 영역이 관찰되는 횟수가 적다는 점을 의미하고 NeRF 학습에서 해당 영역은 적은 횟수 학습하게 된다. 따라서 이러한 적은 횟수의 학습은 이미지가 갖는 노이즈는 제대로 제거하기 힘들고, 렌더링 후 의도하지 않은 왜곡이나 존재하지 않는 형상(artifact)을 만들어낸다. 이러한 현상은 특히 물체의 경계에서 심하게 관찰된다. [Fig. 3]은 앞서 언급한 두 가지 다른 촬영 거리에 획득된 이미지를 이용하여 학습한 NeRF의 렌더링 결과이다. 물체의 경계에 해당하는 부분에서 많은 노이즈와 인공물(artifact)가 발생한 것을 관찰할 수 있다.

따라서 실제 물체(ground truth)와 NeRF로 학습된 3차원 모델 간의 유사도를 비교했을 때, 전술한 노이즈와 인공물로 인해 가까운 거리에서 촬영한 이미지로 학습한 NeRF 모델의 경우 실제 물체와 더 큰 차이가 발생하게 된다. [Fig. 4]는 이미지

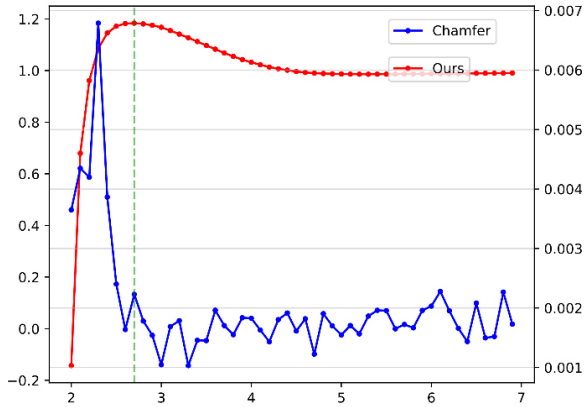


[Fig. 3] Two synthesized images from two NeRF models with (a) images captured in *closer* distance and (b) images captured in *farther* distance: The synthesized image from (a) contains undesired noise and artifacts



[Fig. 4] Relationship between capturing distance and its Chamfer distance: Higher Chamfer distance means more disparity with the ground truth

가 획득된 촬영 거리에서 따라 실제 물체와 NeRF로 학습된 모델 사이의 Chamfer distance (CD)의 변화를 나타낸 것이다. [Fig. 4]에서 촬영 거리가 2.5 m 이내로 근접하여 촬영한 이미지를 사용한 경우 오히려 CD가 극단적으로 커지게 되는 것을 관찰할 수 있다. 약 2.6 m에서 3.3 m 사이에서 가장 낮은 CD 값을 갖고, 우리의 직관과 같이 거리가 멀어질수록 점차 CD 값이 커진다. 따라서 NBV 선택에 있어 단순히 엔트로피만 고려하여 가까운 촬영 거리의 이미지만 획득하지 않고 적절한 촬영 거리의 이미지를 획득하는 것이 중요함을 도출할 수 있다. 또한 이러한 촬영 거리는 이미지 전체에 고정되어야 하는 것이



[Fig. 5] Our proposed NBV selection function (red) with respect to capturing *distance*: Its maximum point (green) and Chamfer distance (blue) are also shown as reference

아니라 관찰된 물체 표면의 복잡도를 고려하여 달라져야 한다.

3.2 거리 제약 조건이 추가된 엔트로피 기반 NBV

본 연구에서는 NBV 선택 기준으로 촬영 거리를 추가적으로 고려한 엔트로피 지표를 제안한다. 엔트로피는 물체 표면의 복잡도를 고려하여 현재 복원된 결과의 불확실성을 표현할 수 있다. 그러나 앞서 관찰한 것과 같이 가까운 거리에서 촬영된 이미지는 물체의 각 영역을 적은 횟수로 관찰하게 되고, 이는 보다 적은 학습의 기회를 의미한다. 따라서 본 논문에서는 식 (9)와 같은 기존 엔트로피 지표^[5]에 촬영 거리 d 에 따른 제약 조건 $p(d)$ 를 추가한 새로운 NBV 지표를 식 (10)과 같이 정의한다.

$$S(V, d) = H(V) - p(d) \quad (10)$$

촬영 거리에 따른 제약 함수 $p(d)$ 는 [Fig. 4]와 유사한 형태로 정의되고 가까운 거리에서 매우 큰 값을 갖고 거리가 멀어짐에 따라 작아지는 지수함수 형태로 식 (11)과 같이 정의한다.

$$p(d) = \alpha d^{-\beta} \quad (11)$$

두 파라미터 α 와 β 는 각각 제약 함수의 가중치와 거리에 따른 변화량을 조절할 수 있는 하이퍼파라미터이다.

본 연구에서는 식 (10)과 같은 새로운 거리-엔트로피 지표를 이용하여 NBV 선택을 수행하였다. 기존 연구들과 마찬가지로 촬영 방향과 거리에 따른 시점의 후보(candidate)를 균일 샘플링을 통해 미리 생성하고, 각 시점의 NeRF 렌더링의 거리-엔트로피 지표가 가장 높은 시점을 NBV로 선정하였다.

4. 실험 및 평가

4.1 실험 이미지 데이터 생성

제안하는 방법의 효과를 살피기 위해 가상 물체에 대한 합성 이미지 데이터셋을 생성하여 실험에 사용하였다. NeRF를 통해 학습된 3차원 모델의 정확도를 평가를 위해서는 참값(ground truth)인 실제 물체의 3차원 모델이 필요하다. 따라서 일반적인 센서 데이터 획득을 통해 정확한 참값을 얻을 수 없기 때문에 참값이 있는 가상 물체와 Blender를 이용하여 원하는 시점의 영상 데이터를 렌더링하여 사용하였다. 가상 물체 파일은 NeRF^[4]에서 제공하는 데이터 3종(Chair, Hotdog, Mic)을 사용하였다. 각 물체에 대해 대상 물체로부터 2 m에서 6.9 m 사이의 촬영 거리를 0.1 m 간격으로 총 50 세트의 이미지를 생성하였다. 또 각 세트마다 다양한 관찰 방향의 144개의 이미지를 생성하였다. 물체를 중심으로 촬영 거리에 따른 구(sphere)의 윗부분에 해당하는 상반구(upper hemisphere)에서 방위각은 (0도에서 360도까지) 10도 간격으로 고도각은 물체가 놓인 평면에서 위쪽으로 20도 간격(20도, 40도, 60도, 80도)으로 카메라의 관찰 방향을 균일하게 샘플링하여 이미지를 획득하였다. 따라서 각 물체 당 총 7,200개(50개 촬영 거리×144개 촬영 시점)의 합성 이미지를 생성하였다.

4.2 정확도 평가 지표

평가 지표로는 선행 연구^[10, 11]에서 많이 사용된 Chamfer distance (CD)를 사용하였다. CD는 두 점 집합 간의 거리를 측정하는 방식으로 두 3차원 모델 사이의 거리를 정량화 할 때 많이 사용된다. 학습된 NeRF 인공 신경망에서 3차원 메시(mesh) 데이터를 추출하기 위해서는 [12]에서 제시된 방법을 적용하였다. 이때 메시 생성의 해상도는 512로 설정하였고, 임계값(threshold)은 0.5로 설정하여 메시지를 생성하였다.

4.3 실험 설정

제안하는 방법과 기존 기법^[5]과의 비교를 위해 총 7,200개의 후보 시점 중에서 두 NBV 선택 지표를 이용해 정해진 개수의 시점을 선택하고, 이를 이용하여 NeRF 모델을 학습시켰다. 우선 [5]에서 제안된 방식과 같이 초기 NeRF 모델 생성을 위한 coarse training을 진행하였다. 이 과정에서 하나의 특정 거리를 선택하여 방위각 60도에 해당하는 36개의 시점을 사용하여 NeRF의 초기 학습을 수행하였다. 초기 학습 이후, 두 NBV 선택 지표를 이용하여 새로운 시점의 이미지를 하나씩 선택하고

이를 이용해 점진적으로 NeRF 모델을 학습하였다. 3종의 데이터에 대해 모두 $\alpha = 6$ 와 $\beta = 2$ 를 촬영 거리 제약 함수의 파라미터로 사용하였다. [Fig. 5]는 제안하는 방법에 α 와 β 를 적용한 결과와 CD를 비교한 그래프이며, 제안된 방법의 최댓값이 CD의 가장 큰 변화 구간과 유사한 값으로 설정됨을 확인할 수 있다.

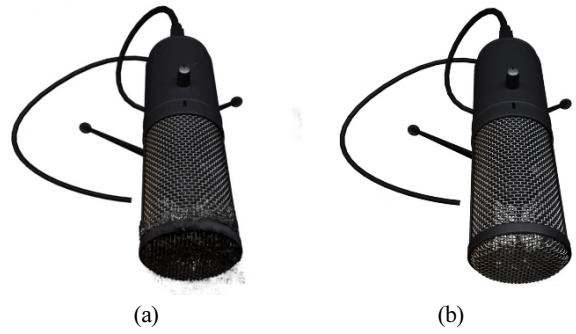
4.4 실험 결과 및 논의

실험을 통해 본 논문에서 제안하는 거리-엔트로피 지표를 사용한 이미지 시점 선택을 통해 보다 정확한 NeRF의 3차원 모델 학습이 가능함을 확인할 수 있었다. [Table 1]은 3종의 물체와 학습 이미지의 개수에 따른 기존 엔트로피 기반의 방법^[5]과 제안하는 거리-엔트로피 지표의 정확도를 Chamfer distance (CD)로 나타낸 것이다. 전체적으로 제안하는 방법이 기존 방법 대비 평균 50% 정도의 작은 CD 값을 갖는 것을 관찰할 수 있었다. 특히 *Chair* 물체의 경우 기존 대비 약 32%로 가장 큰 차이를 보였고, *Hotdog* 물체의 경우 약 76%로 가장 작은 차이를 보였다. 이는 *Hotdog* 물체의 경우 형태가 원통형으로 단순하고 복잡한 경계선을 갖지 않기 때문으로 생각된다. 이미지의 개수에 따라서도 20개로 제한한 경우 거리-엔트로피 지표의 효과가 가장 크게 나타났고, 이미지의 개수가 많아질수록 그 격차가 줄어들었다. 이는 이미지 개수가 적을수록 NBV에 의한 결과가 더욱 중요해지기 때문이다. [Fig. 6]과 [Fig. 7]은 각각 *Mic* 물체와 *Chair* 물체에서 NeRF를 통해 합성한 영상을 비교한 것이다. 합성된 이미지에서도 제안하는 기법을 이용한 결과보다 노이즈가 적고 인공물(artifact)가 적은 결과를 나타냄을 정성적으로 확인할 수 있다.

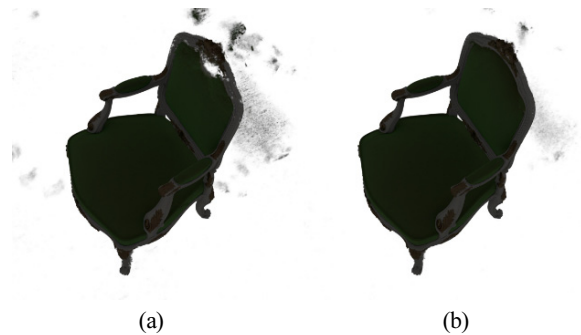
실험을 통해 본 논문에서 제안하는 거리-엔트로피 지표를 통해 기존 방법^[5]보다 조금 더 멀고 다양한 거리에서 NBV가 선택되는 것을 확인할 수 있었다. [Fig. 8]은 선택된 NBV와 물체 사이의 촬영 거리를 히스토그램으로 나타낸 것이다. 기존 방법은 2.5 m 이내의 거리에서 시점 선택이 집중되지만, 제안

[Table 1] Chamfer distance (Note: lower (↓) is better.) of the original entropy-based method^[5] and our proposed entropy-based methods with distance regularization on three different object datasets

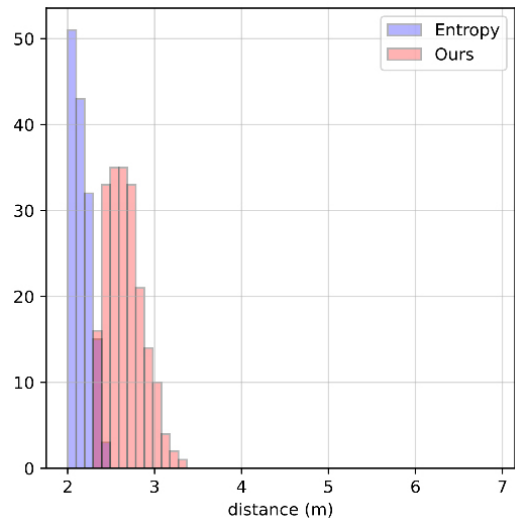
Objects	The Number of NBV Images					
	20		40		60	
	Entropy ^[5]	Ours	Entropy ^[5]	Ours	Entropy ^[5]	Ours
Chair	0.1538	0.0561	0.0892	0.0239	0.0536	0.0173
Hotdog	0.1026	0.0950	0.0746	0.0426	0.0327	0.0258
Mic	0.2521	0.0108	0.0134	0.0080	0.0065	0.0041



[Fig. 6] Two synthesized images of the Mic dataset by (a) the original entropy-based method^[5] and (b) the proposed entropy-based method with distance regularization



[Fig. 7] Two synthesized images of the Chair dataset by (a) the original entropy-based method^[5] and (b) the proposed entropy-based method with distance regularization



[Fig. 8] Histogram of selected NBVs with respect to capturing distance: The proposed method (red) selected NBVs more farther and more distributed than the original entropy-based method (blue)

하는 방법은 그보다 먼 거리에서 그리고 더 넓은 범위에서 시점을 선택하는 것을 확인할 수 있다. 엔트로피는 복원된 물체 표면의 불확실성을 표현하고, 추가된 거리 제한 함수를 통해

단순히 최대한 가까이가 아닌 표면의 복잡도를 고려하여 가까운 거리의 시점 선택이 가능한 것이다. 즉 물체 표면이 복잡한 경우 가까운 촬영 거리의 시점이 선택되고, 물체 표면이 단순한 경우 물체의 더욱 넓은 영역을 관찰할 수 있도록 보다 먼 거리의 시점이 선택되는 것이다. 3종의 물체는 크기가 비슷하여 동일한 값의 파라미터 α 와 β 를 사용하였다. 만약 물체가 탁상 위의 물체의 크기가 아니라 방 크기 또는 건물 크기와 같이 달라지는 경우 또는 NBV를 통해 선택하는 이미지의 개수가 크게 적어지는 경우 다른 값의 α 와 β 를 사용하여야 될 것으로 생각된다.

5. 결 론

본 논문에서는 보다 정확한 NeRF의 3차원 모델 생성을 위해 학습 이미지의 시점 선택에 있어 촬영 거리를 고려한 새로운 지표를 제안하였다. 기존의 엔트로피 지표^[5]를 사용하는 경우 가까운 촬영 거리의 시점만 선택하여 관찰 영역이 좁아지거나 관찰 횟수가 적어지는 문제를 갖는데, 제안하는 거리-엔트로피 지표는 촬영 거리에 따른 제약 함수를 도입하여 물체 표면의 복잡도에 따라 적절한 촬영 거리를 선택할 수 있었다. 참값(ground truth) 물체 모델을 갖는 합성 이미지 데이터셋 3종에서 이를 실험하여 기존 방법 대비 약 50% 적은 에러(CD)를 갖는 것을 확인할 수 있었다.

본 연구는 식 (10)과 식 (11)의 형태로 촬영 거리에 따른 제약을 엔트로피 지표에 부여하였는데, 해당 방법에 대한 보다 최적의 제약 함수와 물체의 크기와 촬영 조건에 따른 파라미터 α 와 β 에 대해 추가적인 고찰과 연구가 필요하다. 또 합성 이미지 데이터 외에 실제 로봇이나 드론, 또는 포터블한 3차원 스캐닝 장치에 이를 적용하여 실제 센서 데이터에 대한 결과를 확보하여야 한다.

향후 본 연구의 결과와 후속 연구를 통해 물체나 공간의 3차원 복원에 있어서 자동으로 카메라의 시점과 촬영 거리를 선택할 수 있을 것으로 기대한다.

References

- [1] S. Isler, R. Sabzevari, J. Delmerico, and D. Scaramuzza, "An Information Gain Formulation for Active Volumetric 3D Reconstruction," *2016 IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden, pp. 3477-3484, May., 2016, DOI: 10.1109/ICRA.2016.7487527.
- [2] L. hou, X. Chen, K. Lan, R. RasMussen, and J. Roberts, "Volumetric Next Best View by 3D Occupancy Mapping Using Markov Chain Gibbs Sampler for Precise Manufacturing," *IEEE Access*, vol. 7, pp. 121949-121960, Aug., 2019, DOI: 10.1109/ACCESS.2019.2935547.
- [3] D. Peralta, J. Casimiro, A. M. Nilles, J. A. Aguilar, R. Atienza, and R. Cajote, "Next-Best View Policy for 3D Reconstruction," *arXiv:2008.12664*, pp. 558-573, Jan., 2020, DOI: 10.48550/arXiv.2008.12664.
- [4] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," *European Conference on Computer Vision*, vol. 123456, pp. 405-421, Nov., 2020, DOI: 10.1007/978-3-030-58452-8_24.
- [5] S. Lee, L. Chen, J. Wang, A. Liniger, S. Kumar, and F. Yu, "Uncertainty Guided Policy for Active Robotic 3D Reconstruction Using Neural Radiance Fields," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 12070-12077, Oct., 2022, DOI: 10.1109/LRA.2022.3212668.
- [6] L. Jin, X. Chen, J. Rückin, and M. Popovic, "NeU-NBV: Next Best View Planning Using Uncertainty Estimation in Image-Based Neural Rendering," *arXiv:2303.01284*, 2023, DOI: 10.48550/arXiv.2303.01284.
- [7] Y. Ran, J. Zeng, S. He, J. Chen, L. Li, Y. Chen, G. Lee, and Q. Ye, "NeurAR: Neural Uncertainty for Autonomous 3D Reconstruction With Implicit Neural Representations," *IEEE Robotics and Automation Letters*, vol. 8, no. 2, pp. 1125-1132, Feb., 2023, DOI: 10.1109/LRA.2023.3235686.
- [8] L. M. Wong, C. Dumont, and M. A. Abidi, "Next-best-view algorithm for object reconstruction," *Sensor Fusion and Decentralized Control in Robotic Systems*, vol. 3523, pp. 191-200, Oct., 1998, DOI: 10.1117/12.327001.
- [9] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379-423, Jul., 1948, DOI: 10.1002/j.1538-7305.1948.tb01338.x.
- [10] J. Tang, H. Zhou, X. Chen, T. Hu, E. Ding, J. Wang, and G. Zeng, "Delicate Textured Mesh Recovery from NeRF via Adaptive Surface Refinement," *arXiv:2303.02091*, 2023, DOI: 10.48550/arXiv.2303.02091.
- [11] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu, and Y.-G. Jiang, "Pixel2mesh: Generating 3d mesh models from single rgb images," *European conference on computer vision*, vol. 11215, pp. 52-67, Oct., 2018, DOI: 10.1007/978-3-030-01252-6_4.
- [12] T. Muller, A. Evans, C. Schied, and A. Keller, "Instant Neural Graphics Primitives with a Multiresolution Hash Encoding," *ACM Transactions on Graphics*, vol. 41, no. 4, pp. 1-15, Jul., 2022, DOI: 10.1145/3528223.3530127.



최진원

2023 서울과학기술대학교 기계시스템디자인
공학(학사)
2023~현재 서울과학기술대학 컴퓨터공학과
(석사과정)

관심분야: Robot Vision, 3D Computer Vision



최준혁

2022 단국대학교 기계공학과(학사)
2022~현재 서울과학기술대학 컴퓨터공학과
(석사)

관심분야: Visual SLAM, Robotics



서찬호

2023 서울과학기술대학교 전기정보공학과
(학사)
2023~현재 서울과학기술대학 컴퓨터공학과
(석사과정)

관심분야: Camera Calibration, 3D Object Detection



최성록

2006 서울대학교 기계항공공학부/전기공학부
(학사)
2008 KAIST 로봇공학학제전공(석사)
2019 KAIST 로봇공학학제전공(박사)
2008~2020 ETRI 지능로봇틱스연구본부
(선임연구원)
2021~현재 서울과학기술대학교 컴퓨터공학과
(조교수)

관심분야: Robot Navigation, 3D Computer Vision