Anti-Aliased Grid-Based Neural Radiance Field

2024 winter seminar

Sogang University
Vision & Display Systems Lab, Dept. of Electronic Engineering

Presented By
Jimin Roh

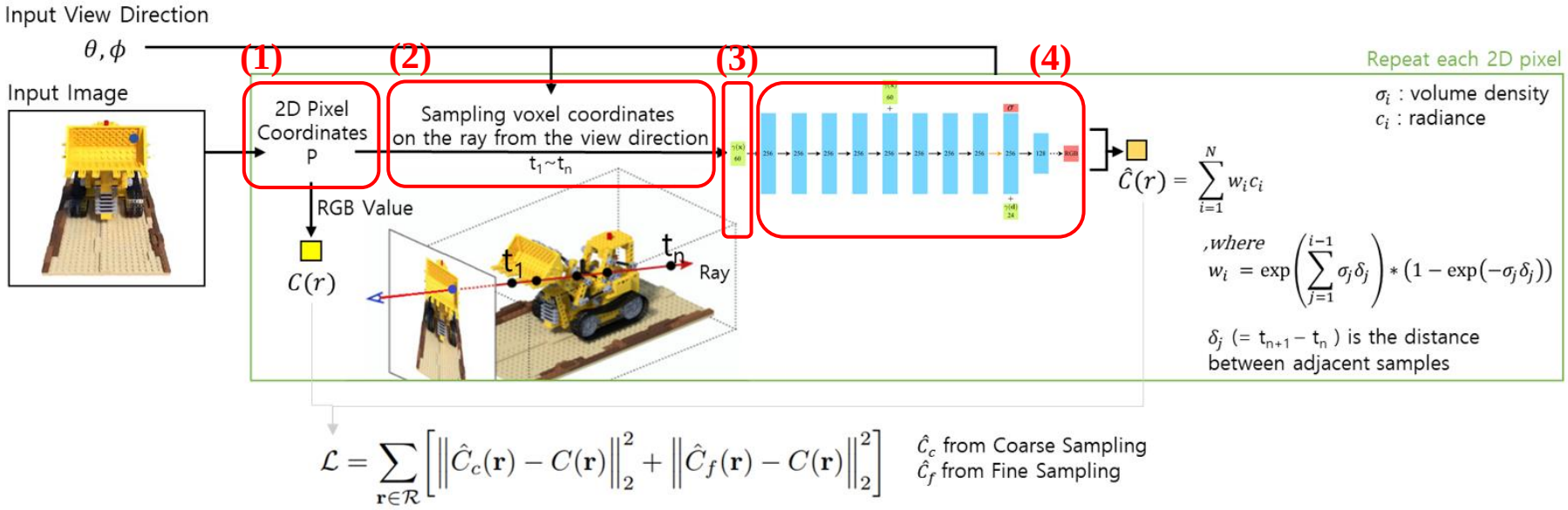
Contents

- Background
  - NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis
  - Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields
  - Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields
  - Instant Neural Graphics Primitives with a Multiresolution Hash Encoding

- Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields (ICCV 2023)
  - Abstarct
  - Methods
  - Results

Background

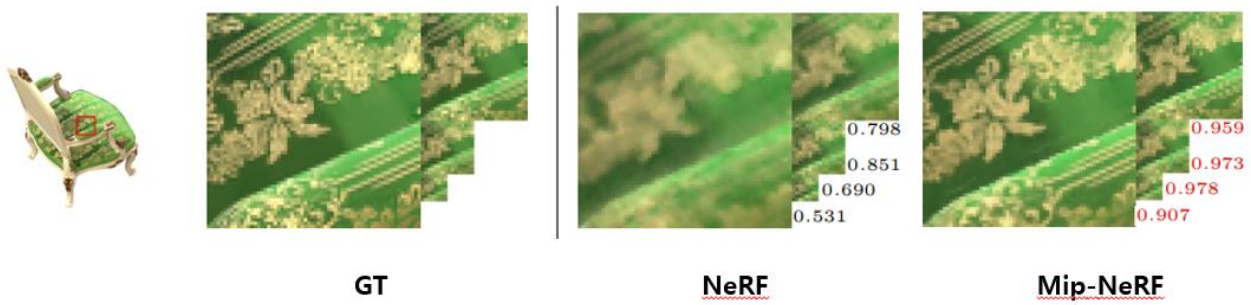
- NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis
  - Overview process
    - (1)Ray tracing, (2)Point sampling, (3)Positional Encoding, (4)MLP layer



<NeRF process 도식화>

Background

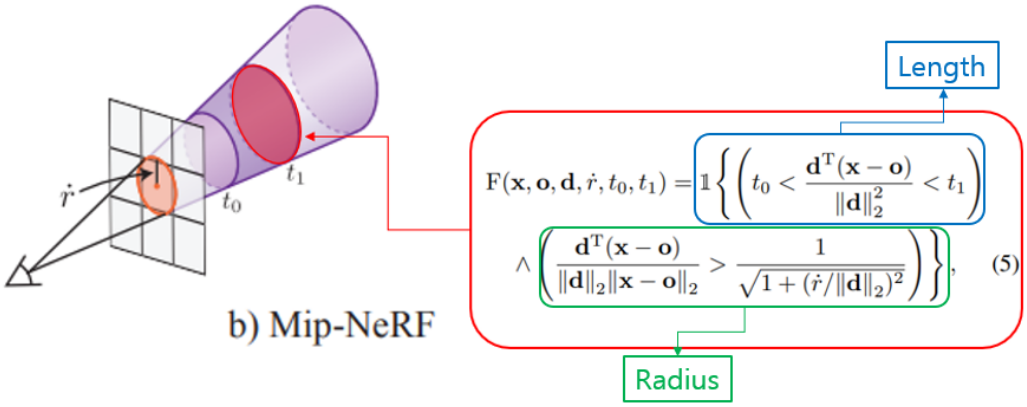
- Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields
  - Abstract
    - 기존 NeRF의 문제
      - 가까운 뷰에서는 지나치게 흐릿하고 먼 뷰에서는 aliasing artifact를 포함함
        - ✓원인 : 모든 train 및 test 이미지가 거의 일정한 거리에서 장면 내용을 관찰한다고 가정하기 때문
        - ✓Aliasing artifact: 카메라로부터 먼 뷰에서 출력이 왜곡되는 현상
    - 해결 방안 제시
      - 연속적인 scale공간에 대해 prefiltering된 radiance를 동시에 나타냄
        - ✓Conical frustum
        - ✓Integrated positional encoding



<모델 별 aliasing artifact 비교>

Background

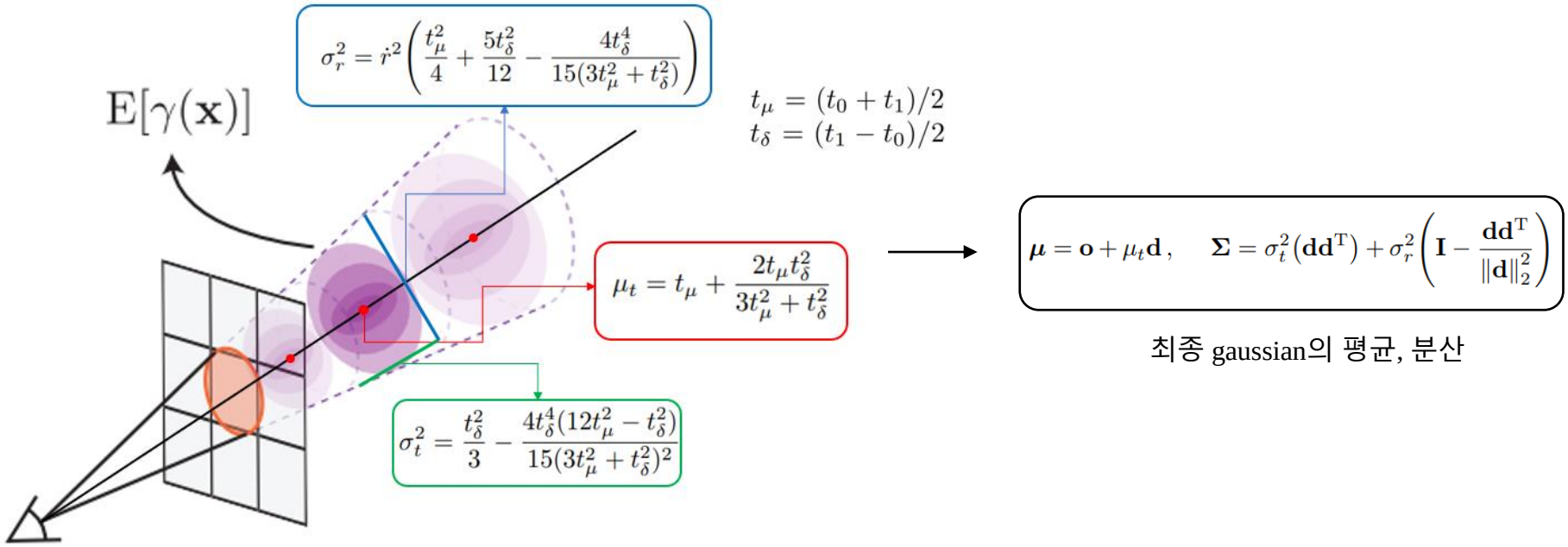
- Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields
  - Cone tracing 및 Conical frustum 구조화
    - Process
      - 기존 NeRF와 동일하게 ray에서 t_i, t_{i+1}를 sampling함
      - Sampled point t_i, t_{i+1}와 카메라 원점 o, view direction d, 반지름 r을 가지고 conical frustum을 구조화
        - ✓이때 반지름 r은 이미지의 한 pixel의 width를 $2/\sqrt{12}$ 만큼 scale한 값을 사용
    - Problem :
      - Conical frustum의 volume을 MLP layer의 input으로 featurized representation을 해야함
    - Solution :
      - Conical frustum을 multivariate gaussian으로 근사화 함



<Conical frustum 정의>

Background

- Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields
  - Multivariate gaussian 근사화
    - μ_t: ray에 대한 distance 평균
    - σ_r^2: ray방향의 수직인 분산, σ_t^2: ray방향에 대한 분산



최종 gaussian의 평균, 분산

<Multivariate gaussian 근사화>

Background

- Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields
  - Integrated positional encoding
    - Multivariate gaussian의 평균과 분산을 푸리에 변환을 통해 feature에 대한 정보를 늘려 최종적인 feature로 사용

기존 Positional Encoding

$$\mathbf{P}=\begin{bmatrix}1&0&0&2&0&0&&2^{L-1}&0&0\\0&1&0&0&2&0&\cdots&0&2^{L-1}&0\\0&0&1&0&0&2&&0&0&2^{L-1}\end{bmatrix}^{\mathrm{T}},\quad \gamma(\mathbf{x})=\begin{bmatrix}\sin(\mathbf{Px})\\\cos(\mathbf{Px})\end{bmatrix}$$

Gaussian Positional Encoding

$$\mathrm{E}_{x\sim\mathcal{N}(\mu,\sigma^2)}[\sin(x)]=\sin(\mu)\exp\left(-(1/2)\sigma^2\right)$$
$$\mathrm{E}_{x\sim\mathcal{N}(\mu,\sigma^2)}[\cos(x)]=\cos(\mu)\exp\left(-(1/2)\sigma^2\right)$$

Integrated Positional Encoding

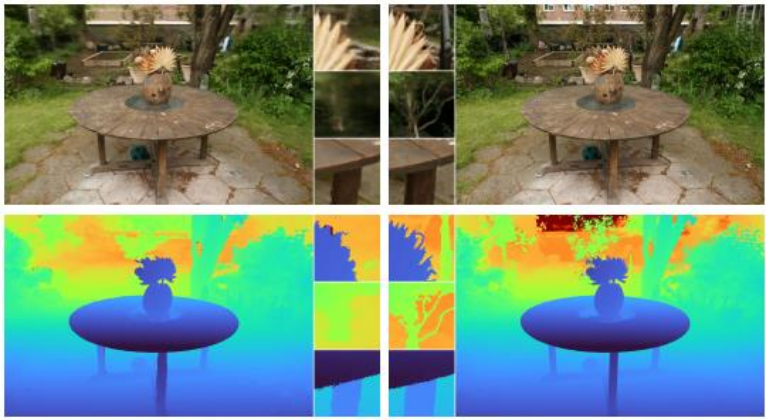
$$\gamma(\boldsymbol{\mu},\boldsymbol{\Sigma})=\mathrm{E}_{\mathbf{x}\sim\mathcal{N}(\boldsymbol{\mu}_\gamma,\boldsymbol{\Sigma}_\gamma)}[\gamma(\mathbf{x})]$$
$$=\begin{bmatrix}\sin(\boldsymbol{\mu}_\gamma)\circ\exp(-(1/2)\operatorname{diag}(\boldsymbol{\Sigma}_\gamma))\\\cos(\boldsymbol{\mu}_\gamma)\circ\exp(-(1/2)\operatorname{diag}(\boldsymbol{\Sigma}_\gamma))\end{bmatrix}$$

$$\operatorname{diag}(\boldsymbol{\Sigma})=\sigma_t^2(\mathbf{d}\circ\mathbf{d})+\sigma_r^2\left(\mathbf{1}-\frac{\mathbf{d}\circ\mathbf{d}}{\|\mathbf{d}\|_2^2}\right)$$

<Integrated positional encoding>

Background

- Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields
  - Abstract
    - 기존 NeRF, Mip-NeRF의 문제점
      - Unbounded scene을 realistic하게 rendering하지 못함
        - ✓Parameterization: Bounded scene 내에서만 3D coordinate에 대해 query할 수 있도록 모델링 됨
        - ✓Efficiency: 크고 detail한 scene을 위해 MLP capacity가 커져야 하는데, 그럼 학습 시에 연산량이 많아짐
        - ✓Ambiguity: Unbounded scene의 물체는 어떤 거리에서든 적은 수의 ray로부터 observed 되기 때문에 어떤 형태로든 ambiguity가 발생함



(a) mip-NeRF [3], SSIM=0.526 (b) Our Model, SSIM=0.804

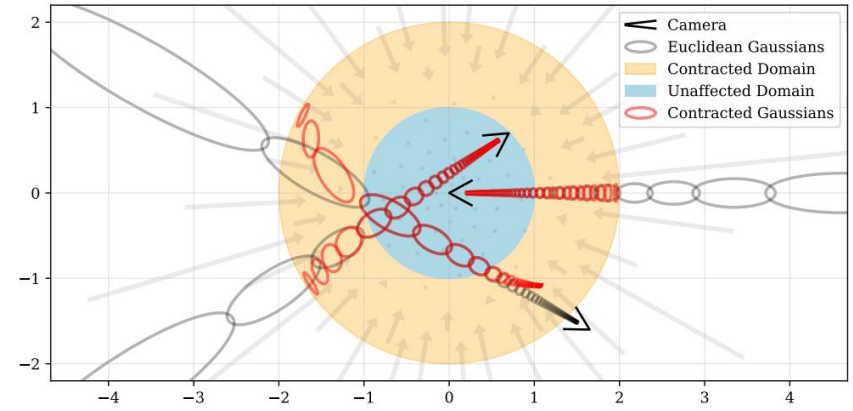
<Unbounded scene에 대한 Mip-NeRF, Mip-NeRF 360 결과 비교>

Background

- Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields
  - Normalized integrated positional encoding
    - Mip-NeRF의 IPE방식을 통해 encoding된 feature의 값을 contract function에 넣어 normalize함
      - 전혀 다른 view direction을 가지는 경우, 타원의 크기가 균등해 지지 않기 때문에 encoding되는 값이 일정하지 않음
      - 따라서 멀리 떨어진 point들은 distance가 아닌 disparity를 기준으로 균등하게 배치할 수 있다는 NDC(Normalized Device Coordinates)를 차용

$$\text{contract}(\mathbf{x}) = \begin{cases} \mathbf{x} & \|\mathbf{x}\| \leq 1 \\ \left(2 - \frac{1}{\|\mathbf{x}\|}\right)\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) & \|\mathbf{x}\| > 1 \end{cases}$$

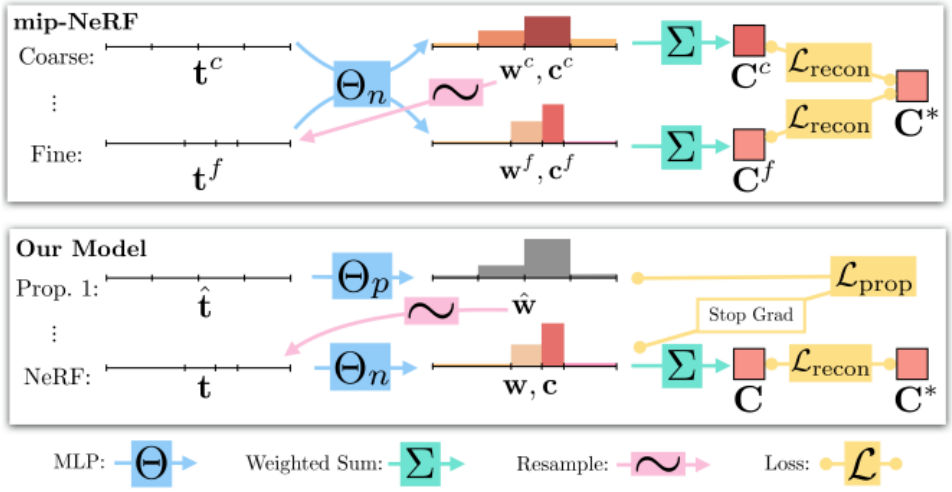
<Contract function>



<Normalize IPE에 의한 gaussian feature를 시각화>

Background

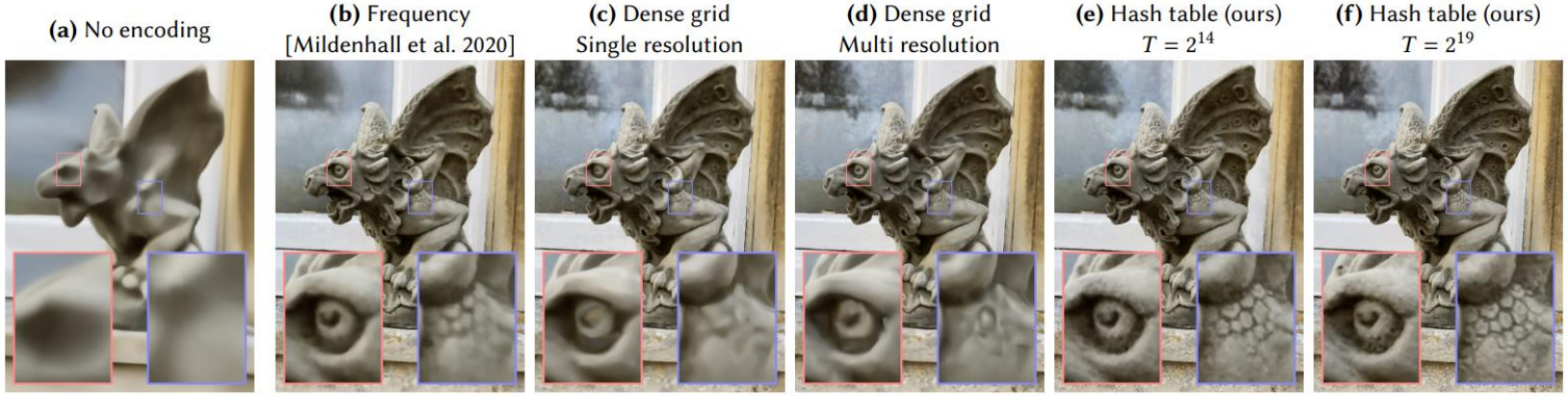
- Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields
  - Online distillation
    - Proposal network
      - Ray위에서 균등하게 sampling을 진행하여 proposal network의 결과로 volume density만을 추출
      - Volume density를 바탕으로 weight를 계산하고, 확률 밀도 함수(PDF)로 변환하여 NeRF network의 input을 추출
      - 학습 시 NeRF network만 backprop되며, proposal network는 NeRF network의 weight를 distillation 받음 (Teacher: NeRF / Student: Proposal)



<MLP network 비교>

Background

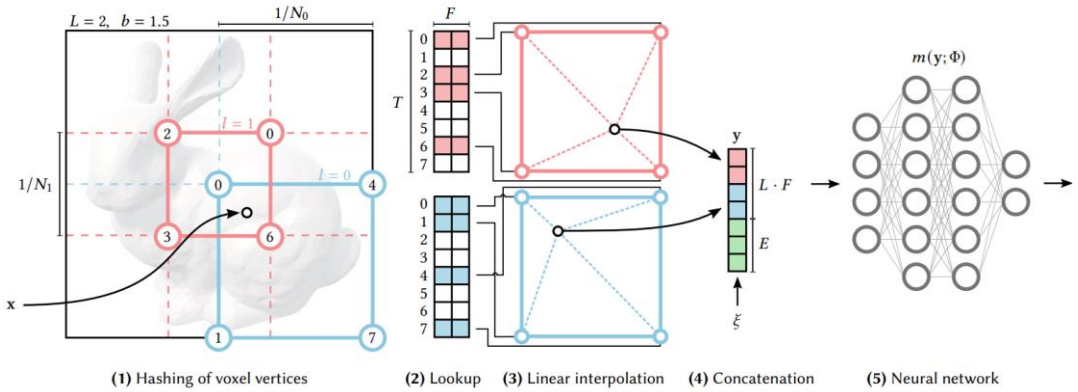
- Instant Neural Graphics Primitives with a Multiresolution Hash Encoding
  - Abstract
    - Task에 국한되지 않고 다양한 task에서 사용할 수 있는 encoding기법을 제안
  - Keyword
    - Multiresolution hash encoding
      - 작은 크기의 MLP network를 사용하여 같은 성능 대비 빠른 속도를 보여줌
      - 학습 시 MLP의 weight와 hash table의 value가 업데이트 됨



<서로 다른 Encoding 결과 비교>

Background

- Instant Neural Graphics Primitives with a Multiresolution Hash Encoding
  - Multiresolution Hash Encoding
    - (1) Hashing of voxel vertices : 입력 x가 주어지면, 각 Level의 주변 voxel들을 찾고 hashing기법으로 corner에 indice를 부여함
    - (2) Lookup : 모든 corner의 indice를 hashing table θ를 통해 F-dimensional feature로 일치시킴
    - (3) Linear interpolation : 각각의 i-th voxel에서 x의 상대적인 위치에 따라 linear interpolation을 시킴
    - (4) Concatenation : 각 level별 feature vector들과 auxiliary값(view direction)들을 concat하여 최종 vector로 만듬



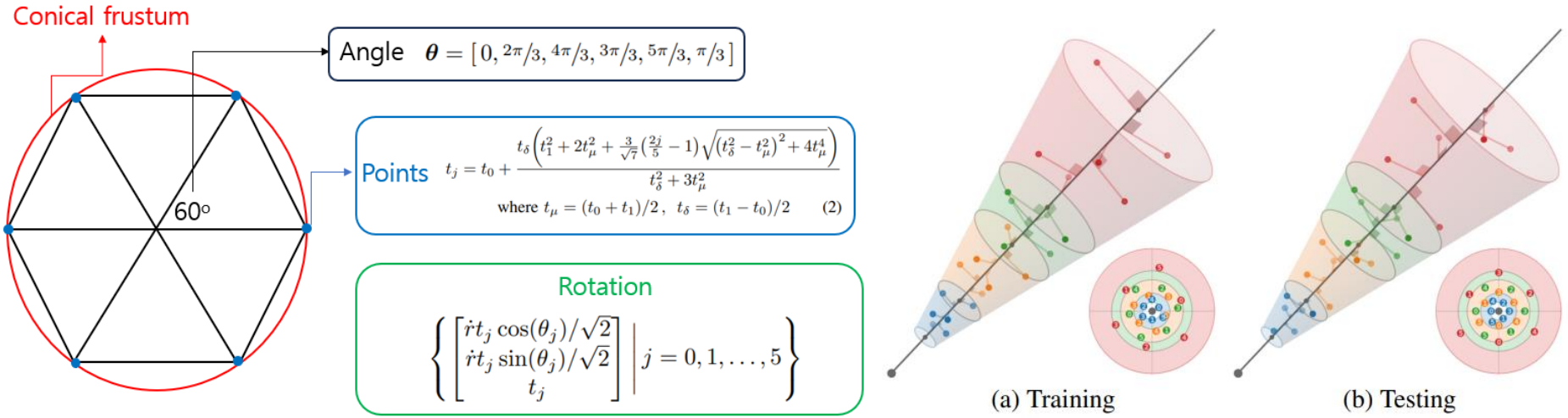
“Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Field”
ICCV 2023

Abstract

- Key idea
  - Instant NGP의 grid-based representation 방식으로 rendering 속도를 올리고 Mip-NeRF 360의 conical frustum방식으로 aliasing 문제를 해결하자
- Problem
  - Encoding 방식의 차이
    - Instant NGP: Ray위의 3D point에 인접한 grid의 3D points를 encoding하여 MLP의 입력으로 사용하고 MLP결과로 나온 값을 interpolation함
    - Mip-NeRF 360: Ray위의 각각의 인접한 3D point 사이를 gaussian 형태로 encoding하고 MLP의 입력으로 사용
  - Z-aliasing
    - 카메라가 z축으로 이동하면, 특정 영역의 density가 사라지는 현상
- Solution
  - Encoding 방식 차이 극복 : Multisampling, Downweighting
  - Z-aliasing 극복 : Anti-aliasing interlevel loss

Methods

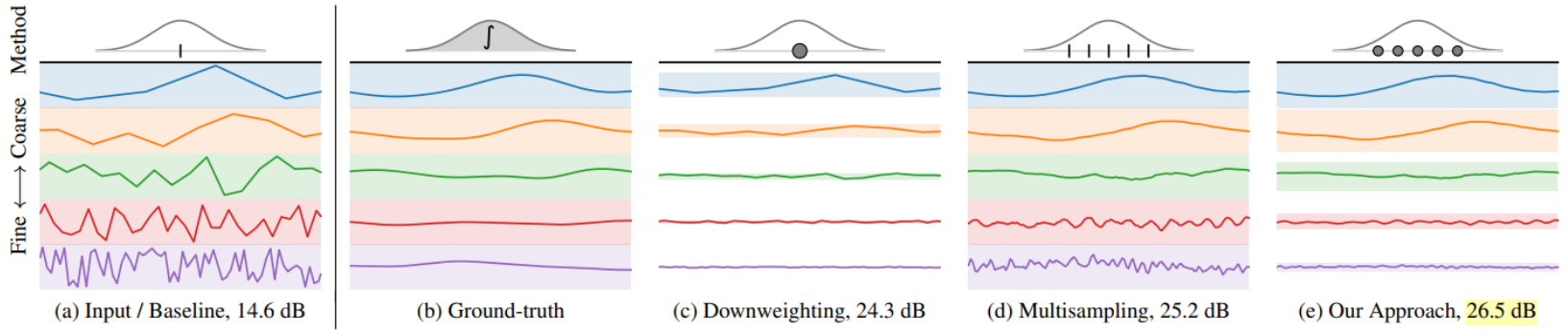
- Multisampling
  - Process
    - Interval $[t_i, t_{i+1}]$에 대해서 6개의 points를 sampling하고, 이때 각 point들 사이의 각도를 60°로 함
    - 이후 각도와 위치를 기반으로 rotation을 진행하여 world coordinate로 변환함
      - 최종적으로 정해진 6개의 multisamples $\{x_j\}$는 isotropic gaussian의 평균과 분산으로 사용



<Multisampling process>

Methods

- Multisampling
  - Process
    - 6개의 multisamples $\{x_j\}$에 대해 Mip-NeRF 360의 contract function을 진행
    - Contract function으로 normalized된 6개의 multisamples $\{x_j\}$에 대해 각각 trilinear interpolation을 수행
      - 이러한 방식으로 Instant NGP의 grid-based representation에서 Mip-NeRF 360의 prefilter된 feature를 얻을 수 있음 (encoding방식 결합 문제 극복)
  - Problem
    - Multisampling이 aliasing을 완화시켜주지만, high frequency에서는 aliasing이 여전히 발생



<Multisampling, Downweighting 유무에 따른 encoding 시각화>

Methods

- Downweighting
  - 각 feature에 대한 multisample's isotropic gaussian이 얼마나 grid cell에 적합한지를 기준으로 feature를 re-weight함
    - Gaussian의 영역이 interpolation되는 cell보다 더 작아야 하기 때문
  - Process
    - Vertex의 값이 평균 0인 normal distribution을 갖도록 weight decay를 설계
      - Zero-mean 가정 : 각각의 multisample의 gaussian이 grid feature로 근사하게 함
    - 최종적으로 각 multisample의 interpolated feature에 대한 weighted mean값은 conical frustum에 대한 expected feature 근사치를 계산할 수 있게 됨

$$\omega_j \cdot \mathbf{f}_{j,\ell} + (1-\omega_j)\cdot \mathbf{0} = \omega_j \cdot \mathbf{f}_{j,\ell}$$

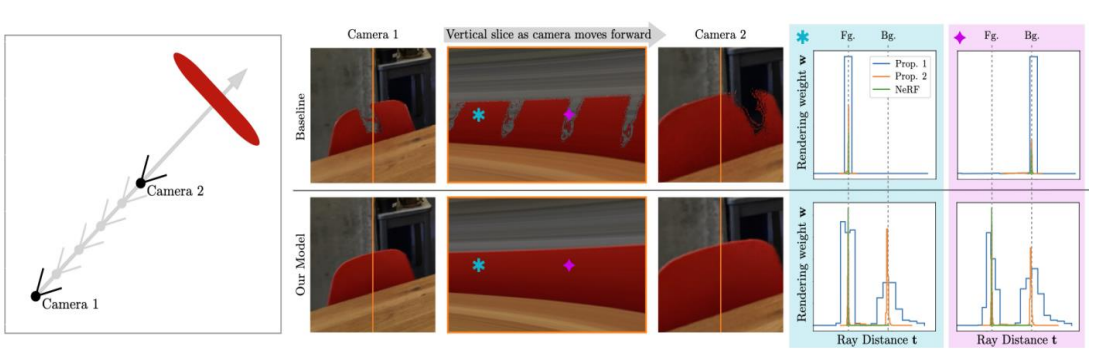
<scale-dependent downweighting factor>

$$\mathbf{f}_\ell = \mathrm{mean}_j(\omega_{j,\ell} \cdot \mathrm{trilerp}(n_\ell \cdot \mathbf{x}_j; V_\ell))$$

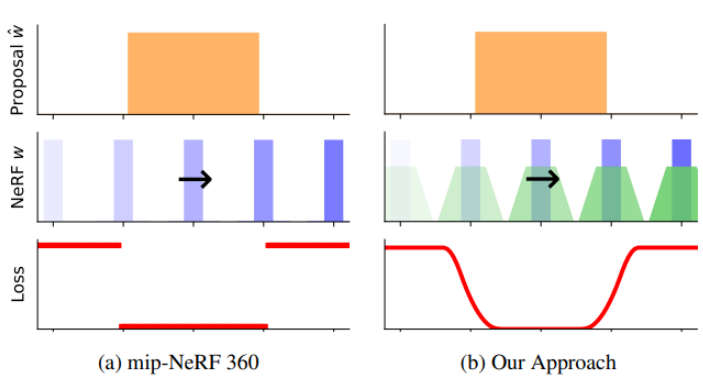
<최종 MLP input features>

Methods

- Problem
  - Z-aliasing
    - Mip-NeRF의 proposal network(teacher)는 NeRF network(student)의 input에 해당하는 sampling interval을 예측함
      - 이때 volume density를 출력하는 MLP의 입력 coordinate에 대해 부드럽지 않게 mapping되는 경향을 보임
  - Solution
    - Blurring a step function
    - Anti-aliasing interlevel loss



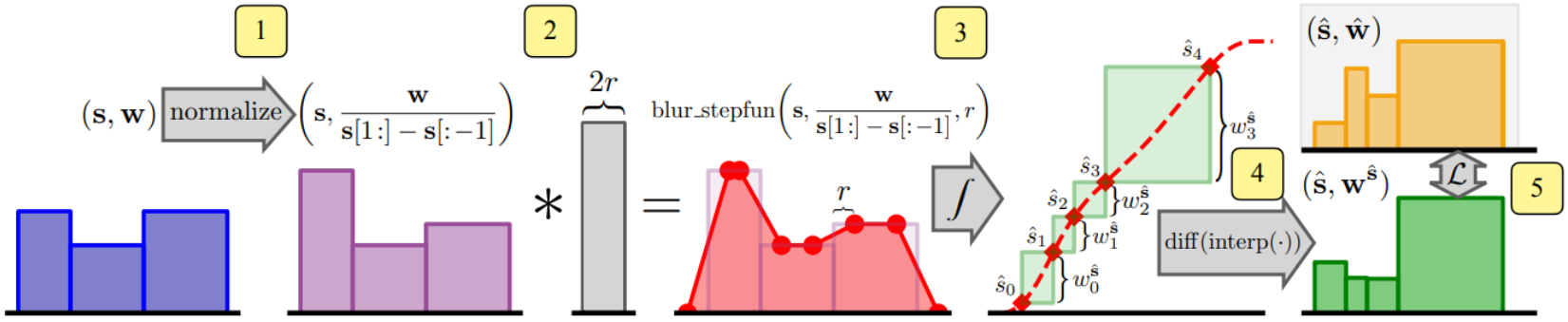
<z-aliasing 문제 시각화>



<Piecewise constant PDF>

Methods

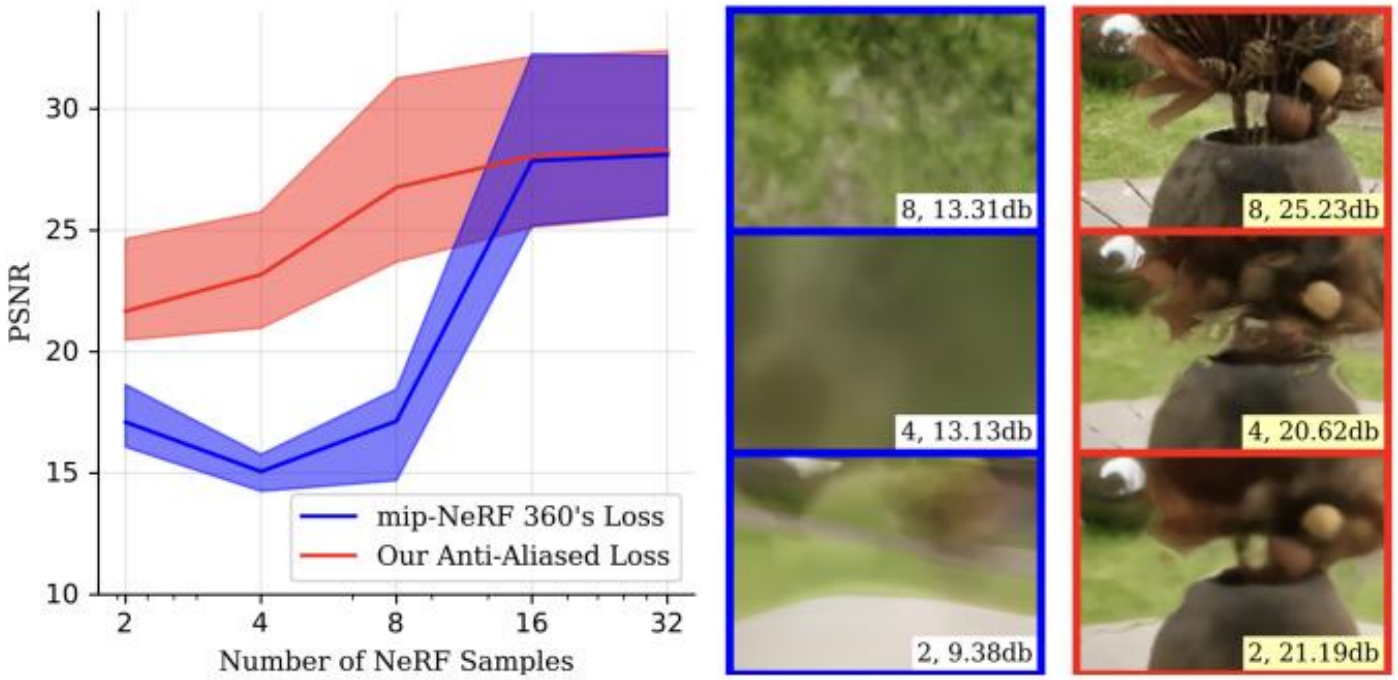
- Blurring a step function & Anti-Aliased Interlevel Loss
  - Process
    - (1) W를 interval로 nomalize 함, 이때 PDF의 적분 값이 1이하가 되도록 함
    - (2) Radius r을 정하고, 2r만큼 convolution 하여 사다리꼴형태의 연속함수로 변환
    - (3,4) PDF를 적분하여 piecewise-quadratic CDF를 생성 후, $\hat{S}$ 마다 piecewise quadratic interpolation으로 변형됨
    - (5) 그렇게 획득한 weight값(초록색)과 proposal network를 통해 계산된 weight값(노란색)간의 loss를 계산하여 proposal network를 업데이트함



<Anti-aliasing loss process>

Methods

- Blurring a step function & Anti-Aliased Interlevel Loss
  - Result



<Interlevel loss 유무 결과 비교>

Results

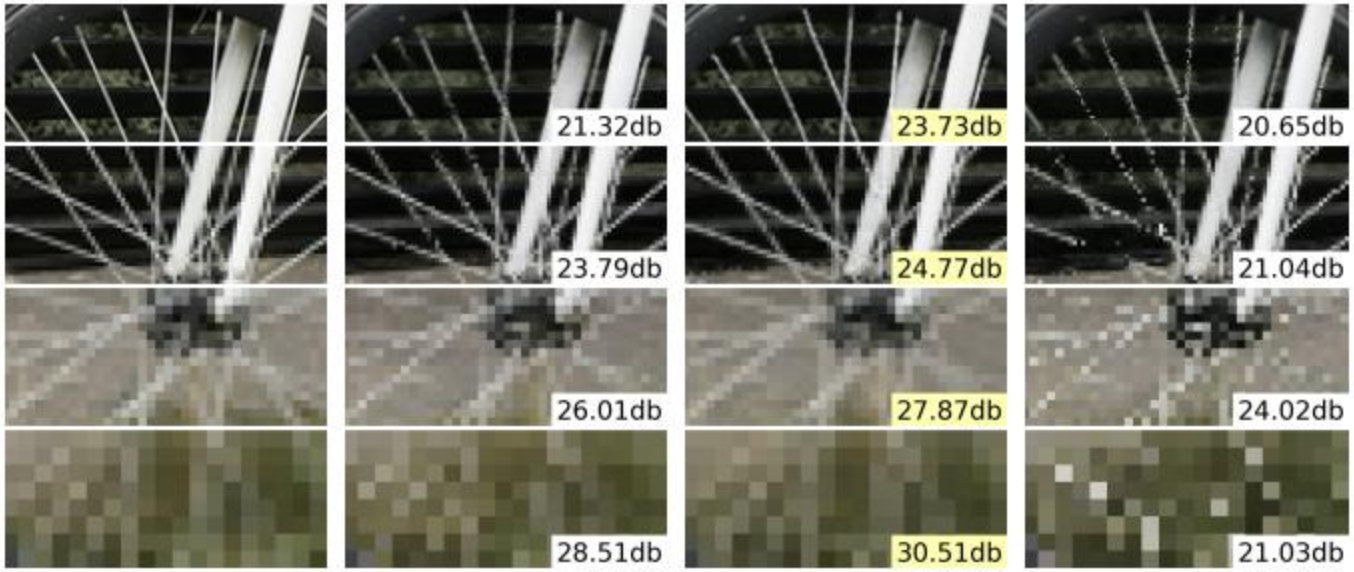
- 정량적 평가
  - 360 dataset에 대한 Mip-NeRF, Mip-NeRF 360, Instant NGP 비교
    - PSNR (Peak Signal-to-noise ratio)
      - MSE로 pixel간의 단순 오차를 계산 후, normalize를 진행하여 평가
    - SSIM (Structural Similarity Index Measure)
      - 두 이미지 간의 상관계수를 휘도, 대비, 구조 총 3가지 측면에서 평가
    - LPIPS (Learned Perceptual Image Patch Similarity)
      - Classification task를 supervised, self-supervised, unsupervised 모델로 학습하고, 비교할 이미지 2개를 각각 학습된 모델을 사용해 feature를 추출하여 feature들의 유사도를 평가

	PSNR ↑	SSIM ↑	LPIPS ↓	Time (hrs)
NeRF [19, 10]	23.85	0.605	0.451	12.65
mip-NeRF [2]	24.04	0.616	0.441	9.64
NeRF++ [36]	25.11	0.676	0.375	28.73
Instant NGP [21, 32]	25.68	0.705	0.302	0.15
mip-NeRF 360 [3, 20]	27.57	0.793	0.234	21.69
mip-NeRF 360 + iNGP	26.43	0.786	0.225	0.30
Our Model	28.54	0.828	0.189	0.89

<Zip-NeRF모델 성능 비교>

Results

- 정성적 평가
  - Unbounded scene(test dataset)에 대해 scale을 달리하여 평가



(a) Ground Truth (b) mip-NeRF 360 (c) Our Model (d) mip360 + iNGP

Results

- Ablations
  - 추가 및 변경 methods에 대해 검증 실험 진행

Scale Factor:	1×			2×			4×			8×			
Error Metric:	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	Time (hrs)
Instant NGP [21, 32]	24.36	0.642	0.366	25.23	0.712	0.251	26.84	0.809	0.142	28.42	0.877	0.092	0.15
mip-NeRF 360 [3, 20]	27.51	0.779	0.254	29.19	0.864	0.136	30.45	0.912	0.077	30.86	0.931	0.058	21.86
mip-NeRF 360 + iNGP	26.46	0.773	0.253	27.92	0.855	0.141	27.67	0.866	0.116	25.58	0.804	0.160	0.31
Our Model	28.25	0.822	0.198	30.00	0.892	0.099	31.57	0.933	0.056	32.52	0.954	0.037	0.90
A) Naive Sampling	27.93	0.797	0.233	29.70	0.880	0.114	29.24	0.887	0.094	26.53	0.820	0.144	0.53
B) Naive Supersampling (6×)	27.48	0.803	0.224	29.03	0.881	0.109	28.42	0.881	0.097	25.97	0.810	0.151	2.54
C) Jittered	27.91	0.797	0.233	29.60	0.879	0.116	29.45	0.893	0.090	27.58	0.855	0.120	0.55
D) Jittered Supersampling (6×)	27.50	0.810	0.212	28.99	0.884	0.105	28.91	0.896	0.086	27.65	0.870	0.109	3.04
E) No Multisampling	28.15	0.817	0.208	29.87	0.886	0.105	31.33	0.927	0.061	32.12	0.948	0.043	0.54
F) No Downweighting	28.22	0.818	0.205	29.94	0.889	0.102	31.25	0.928	0.060	31.67	0.944	0.046	0.88
G) No Appended Scale ω	28.23	0.820	0.200	29.98	0.890	0.101	31.48	0.931	0.057	32.19	0.951	0.041	0.89
H) Random Multisampling	28.09	0.816	0.207	29.75	0.886	0.106	31.18	0.928	0.061	32.03	0.950	0.042	0.95
I) Unscented Multisampling	28.27	0.822	0.198	30.03	0.891	0.100	31.57	0.933	0.056	32.46	0.954	0.038	1.11
J) No New Interlevel Loss	28.12	0.824	0.196	29.82	0.892	0.098	31.31	0.932	0.056	32.23	0.953	0.039	0.86
K) No Weight Decay	27.34	0.814	0.203	28.91	0.881	0.109	30.29	0.921	0.067	31.23	0.941	0.050	0.90
L) Un-Normalized Weight Decay	27.99	0.821	0.196	29.65	0.889	0.100	31.10	0.930	0.058	32.09	0.951	0.040	0.91
M) Small View-Dependent MLP	27.41	0.811	0.207	28.98	0.882	0.109	30.32	0.924	0.065	31.11	0.944	0.047	0.63

<Zip-NeRF모델 성능 비교>

감사합니다