

# 卒業論文概要書

## Summary of Bachelor's Thesis

Date of submission: 01/30/2024 (MM/DD/YYYY)

学科名 Department	情報通信	氏名 Name	白崎 良侑	指導員 Advisor	渡辺 裕 印
研究指導名 Research guidance	オーディオビジュアル 情報処理研究	学籍番号 Student ID number	1w202173-2		
研究題目 Title	露光の多重化と Bad-NeRF を用いた効率的な自由視点画像生成 Efficient Free-Viewpoint Image Generation Using Multiplexed Exposure and Bad-NeRF				

### 1. まえがき

Neural Radiance Fields (NeRF) [1]は、2021年に発表されて以来、自由視点画像の生成研究分野で注目を集めている。自由視点画像技術は、一般家庭の人々にとって、静止画による記録を3D形式で残し、後でVRゴーグル等で豊かな視覚体験として振り返る、eコマースやオンラインマーケティングにおいて、小規模ビジネスオーナーや個人販売者が商品の3D表示を行い、効果的な商品プレゼンテーションを行う、といった利用価値が考えられる。しかしながら、NeRF技術が一般家庭へ普及する上で、計算コストと時間コストの大きさ、撮影コストの大きさ、動的シーンの取り扱いづらさなどの課題がある。そこで本研究では、安価な撮影デバイスを用いることを前提とし、自由視点画像生成モデルの訓練に用いる学習データの撮影回数を削減する手法を提案する。

### 2. 関連研究

#### 2.1 Programmable Aperture Camera

Programmable Aperture Camera [2]（以後「PA Camera」と表記）は、カメラの絞りを動的に変化させることができるカメラである。センサのフル解像度を活用したライトフィールド撮影が可能でありながら、ライトフィールドカメラと比較して製造コストを大幅に削減できる。先行研究では、カメラのレンズ前に紙製のフィルタを設置し、フィルタ上に小さな穴のパターンを作成することでPA Cameraの機構を実現している。この穴のパターンは3×3のグリッドで構成されており、カメラのレンズよりも小さく作成される。フィルタ上には、一つ、または複数の穴があげられる。こうしてフィルタにあける穴のパターンを変化させることで、異なる露光パターンを作り出すことができる。

#### 2.2 NeRF

Neural Radiance Fields (NeRF) [1]は、3Dシーンを高密度な点群として表現し、それぞれの点での光の放射輝度と色をニューラルネットワークを用いて学習する技術である。複数の視点から物体を捉えた画像をNeRFで学習することで任意の視点からの高品質な画像を合成することが可能となる。この3D表現は物体やシーンの光の散乱や反射の特性を含んでおり、それによって観測されていない視点からでも高精度な画像を生成することができる。

#### 2.3 Bad-NeRF

Bad-NeRF [3]は、NeRFの拡張版であり、モーションブレンダーのかかった画像セットを扱う場合における性能向上に特化した技術である。Bad-NeRFでは、モーションブレンダーのかかった画像を複数の異なる視点からの画像の線形結合と捉え、それらの視点のパラメータを予測するモデル（スプラインモデルと呼ばれる）をNeRFモデルと並行して学習させる。

### 3. 提案手法

少ない撮影回数で高品質な自由視点画像を生成するために、PA CameraとBad-NeRFを組み合わせるアプローチを提案する。撮影にはPA Cameraを使用し、露光を多重化して画像を取得する。さらに、取得画像をBad-NeRFに学習させる。Bad-NeRFではスプラインモデルによりカメラパラメータを推定するが、本研究ではPA Cameraによる撮影時のカメラパラメータが既知であるため、その情報を活用してスプラインモデルの訓練を省略する。提案手法の概要を図1に示す。

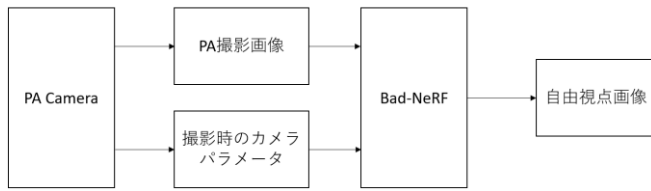


図1 提案手法の概要

#### 4. 実験

提案手法の有効性を示すために、二重化画像4枚を Bad-NeRF にパラメータ付きで学習させる実験、二重化画像4枚を Bad-NeRF にパラメータ無しで学習させる実験、単一視点画像4枚を NeRF に学習させる実験、単一視点画像8枚を NeRF に学習させる実験の4項目それぞれを5種類のデータセットで実施する。データセットは NeRF の訓練済みのモデルを用いて生成する。二重化画像は Python のコードにより二つの画像を線形結合して作成する。生成モデルの評価には PSNR[4]を用いる。実験結果を視覚的に比較するため拡大図を含んだ画像を図2に示す。さらに、実験の結果生成された画像の PSNR を比較し、以下の表1に示す。



図2 実験結果の比較

表1 各生成モデルの PSNR

(単位 : dB)

学習データ	fern	flower	fortress	horns	trex
二重化画像4枚 (提案手法)	<u>32.12</u>	<u>35.14</u>	<u>37.11</u>	<u>31.44</u>	<u>31.20</u>
二重化画像4枚 (カメラ パラメータ無し)	14.88	15.15	20.27	14.08	15.53
単一視点画像4枚	23.51	22.59	24.83	21.54	22.36
単一視点画像8枚	<b>34.74</b>	<b>36.66</b>	<b>39.18</b>	<b>35.57</b>	<b>32.93</b>

図2より、単一視点4枚を NeRF に学習させた場合と、二重化画像4枚をパラメータ無しで Bad-NeRF に学習させた場合の生成モデルの品質が明らかに低下していることが分かる。また、提案手法による生成モデルは高品質ではあるが、図4.9の椅子の影の部分に注目すると、正解画像や、単一視点画像8枚を NeRF に学習させた場合と比較して、雑音が入っていることが分かる。さらに表1より、8枚の画像で NeRF を学習させた生成画像に迫る品質の画像が、多重化した4回の露光で取得した画像とカメラパラメータで Bad-NeRF を学習させることで生成できることや、Bad-NeRF へのカメラパラメータ入力の効果が高いことがわかった。

#### 5. 結論

本研究では、安価な撮影デバイスを用いることを前提とし、自由視点画像生成モデルの訓練に用いる学習データの撮影回数を削減する手法を提案した。提案手法では、PA Camera により露光を多重化して取得した画像をカメラパラメータと共に Bad-NeRF に学習させる。実験により、露光を多重化することで、8枚の画像で NeRF を学習させた生成画像に迫る品質の画像が、多重化した4回の露光で取得した画像とカメラパラメータで Bad-NeRF を学習させることで生成できることや、Bad-NeRF へのカメラパラメータ入力の高効果がわかった。以上の結果から、提案手法の有効性が確認できたと考える。ただし、品質改善の余地があり、今後の研究を通して品質が劣化する原因を特定し、改善を目指す。

#### 参考文献

- [1] 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, Dec. 2021.
- [2] C. Liang, T. Lin, B. Wong, C. Liu, and H. Chen, "Programmable Aperture Photography: Multiplexed Light Field Acquisition," *In ACM Transaction on Graphics*, Vol. 27, No.3, pp. 1-10 Aug. 2008.
- [3] P. Wang, L. Zhao, R. Ma, and P. Liu, "BAD-NeRF: Bundle Adjusted Deblur Neural Radiance Fields," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023)*, pp. 4170-4179, Jun. 2023.

[4] MathWorks. (n.d.). PSNR. Retrieved January 17, 2024,  
from  
<https://jp.mathworks.com/help/vision/ref/psnr.html>

2023 年度 卒業論文

露光の多重化と Bad-NeRF を用いた  
効率的な自由視点画像生成

Efficient Free-Viewpoint Image Generation Using  
Multiplexed Exposure and Bad-NeRF

指導教員 渡辺 裕 教授

提出日：2024 年 1 月 30 日

早稲田大学 基幹理工学部 情報通信学科

1W202173-2

白崎 良侑

# 目次

第1章	序論.....	3
1.1	研究背景.....	3
1.2	研究目的.....	4
1.3	本論文の構成.....	4
第2章	関連研究.....	5
2.1	まえがき.....	5
2.2	<b>Programmable Aperture Camera</b> .....	5
2.3	NeRF とその応用.....	6
2.3.1	NeRF.....	6
2.3.2	<b>Bad-NeRF</b> .....	7
2.4	むすび.....	7
第3章	提案手法.....	8
3.1	まえがき.....	8
3.2	提案手法.....	8
3.3	提案手法の優位性.....	9
3.4	むすび.....	9
第4章	実験.....	10
4.1	まえがき.....	10
4.2	データセット.....	10
4.3	評価指標.....	12
4.4	実験概要.....	12
4.4.1	概要.....	12
4.4.2	<b>Bad-NeRF</b> による二重化画像4枚の学習.....	12
4.4.3	<b>Bad-NeRF</b> による既知のカメラパラメータ無しでの二重化画像4枚の学習.....	14
4.4.4	NeRF による単一視点画像4枚の学習.....	14
4.4.5	NeRF による単一視点画像8枚の学習.....	15
4.4.6	各実験結果の定性的評価.....	16
4.5	評価実験.....	17
4.6	考察.....	17

4.6.1	Bad-NeRF へのカメラパラメータ入力の有無比較 .....	17
4.6.2	二重露光画像 4 枚と単一視点画像 4 枚の比較 .....	17
4.6.3	二重露光画像 4 枚と単一視点画像 8 枚の比較 .....	18
4.7	むすび .....	18
第 5 章	結論と今後の課題 .....	19
5.1	結論 .....	19
5.2	今後の課題 .....	19
謝辞	.....	20
参考文献	.....	21
図一覧	.....	22
表一覧	.....	23

# 第1章 序論

---

## 1.1 研究背景

Neural Radiance Fields (NeRF)[1]は、複数の静止画から自由視点画像を生成する技術である。2021年に発表されて以来、NeRFから派生した研究が盛んに行われ、自由視点画像の生成研究分野で注目を集めている。自由視点画像技術は、一般家庭の人々にとって以下の二つのような利用価値が考えられる。第一に、静止画による記録を3D形式で残すことができる。VRゴーグル等の普及により、3D映像の立体的な視覚体験が手軽に可能になってきている。自由視点画像技術を用いることで、人々が旅行などでの思い出を、後で豊かな視覚体験として振り返ることができるようになる。第二に、小規模ビジネスや個人の商品撮影への利用が考えられる。自由視点画像技術が一般家庭に普及することで、eコマースやオンラインマーケティングにおいて、小規模ビジネスオーナーや個人販売者が商品の3D表示を行い、効果的な商品プレゼンテーションを実現できるようになる。しかしながら、NeRF技術の一般家庭への普及には以下の三つの問題がある。第一に、計算コストと時間コストが大きい点があげられる。高品質な自由視点画像を生成するためには、高価な計算資源と長時間の処理が必要である。第二に、撮影コストの増大が問題となる。NeRFは多くの視点からの画像データに依存しており、対象シーンやオブジェクトの広範な撮影が必要となる。一般的なデバイスでの撮影も可能であるが、高品質な画像を生成するためには多数の撮影が必要である。ライトフィールドカメラのような専用機器も存在するが、解像度と価格の問題で一般家庭へは普及していない。さらに第三に、動的シーンの取り扱いが容易ではないという課題がある。NeRFは静的なシーンでは優れた自由視点効果を示すが、動くオブジェクトや変化する照明条件では生成が容易ではないという課題がある。

本研究では、第二の撮影コストの削減に焦点を当てる。低コストで開発できる撮影デバイスを用いて撮影回数を削減し、軽微な撮影負荷で、自由視点画像を生成できるようにする。これにより、一般家庭への普及が目指せると考える。

## 1.2 研究目的

本研究では、安価な撮影デバイスを用いることを前提とし、自由視点画像生成モデルの訓練に用いる学習データの撮影回数を削減する手法を提案する。そのために、撮影デバイスとして Programmable Aperture Camera[2]（以降「PA Camera」と表記する）を使用する。PA Camera は、センサのフル解像度を活用したライトフィールド撮影が可能でありながら、ライトフィールドカメラと比較して低いコストで製造できる単純な機構を持つ。また、PA Camera では、露光を多重化することで一度の露光で複数視点からの画像の線形結合を取得することができる。この多重露光を活用することで、必要数の半分以下の露光回数で、NeRF に学習させる画像データを効率的に取得することができる。ただし、NeRF では多重露光により取得した画像を直接学習データとして扱うのは効果的ではない。そこで、モーションブラー（動きぼけ）のかかった画像を扱う場合に特化した処理技術である Bad-NeRF[3]を用いる。本研究では、PA Camera による撮影画像を Bad-NeRF に効果的に学習させ、高品質な自由視点画像を生成する。

## 1.3 本論文の構成

本論文の構成を以下に示す。

- 第1章 本章であり、本研究の背景および目的について述べる。
- 第2章 本研究に関連する Programmable Aperture Camera と、NeRF 及びその派生技術について述べる。
- 第3章 安価な撮影デバイスを用いることを前提とし、自由視点画像生成モデルの訓練に用いる学習データの撮影回数を削減する提案手法を紹介する。
- 第4章 提案手法を用いた実験、及びその結果と考察について述べる。
- 第5章 本研究の結論及び今後の課題について述べる。



## 第 2 章 関連研究

---

### 2.1 まえがき

本章では、カメラの絞りを動的に調節できる撮影技術である PA Camera と、自由視点画像生成技術である NeRF、さらにモーションプレーのかかった画像に特化した自由視点画像生成技術である Bad-NeRF について述べる。

### 2.2 Programmable Aperture Camera

PA Camera[2]は、カメラの絞りをプログラム可能に制御できるカメラである。センサのフル解像度を活用したライトフィールド撮影が可能でありながら、ライトフィールドカメラと比較して低いコストで製造できる。Programmable aperture photography: multiplexed light field acquisition[2]では、PA Camera の機構を、カメラのレンズ前に紙製のフィルタを設置し、フィルタ上に小さな穴のパターンを作成することで実現している。この穴のパターンは以下の図 2.1 に示すような 3×3 のグリッドで構成されている。図 2.1 では、各セルを A から I までのラベルで識別している。さらに、このグリッドパターンはカメラのレンズよりも小さく作成される。フィルタ上には、A から I の中から一つ、または複数の穴があけられる。こうして、フィルタにあける穴のパターンを変化させることで、異なる露光パターンを作り出すことができる。以下、複数の穴をあけた露光を「多重化された露光」と記述する。

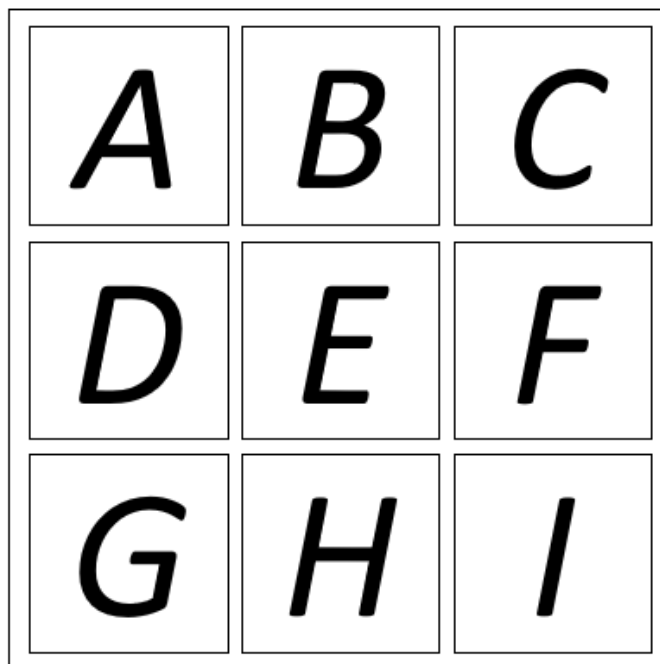


図 2.1 フィルタ上の穴のパターン

また、先行研究では以下の三つの内容が述べられている。第一に、単一の穴をあけた露光では、光量が少ないことから取得画像にノイズが発生する。このノイズは露光を多重化することで減少させることができる。第二に、多重化された露光により得られた画像は、一つずつの穴を通して得られる画像の線形結合になっており、言い換えれば、複数の視点の画像の線形結合になっている。さらに第三に、露光後の計算機処理により、多重化された露光の取得画像を各視点の画像に分解することができる。ただし、シーンを  $N$  個の異なる角度から捉える画像を収集するためには、露光パターンを変えた  $N$  回の露光が必要である。

## 2.3 NeRF とその応用

### 2.3.1 NeRF

Neural Radiance Fields (NeRF) [1]は、3Dシーンを高密度な点群として表現し、それぞれの点での光の放射輝度と色を、ニューラルネットワークを用いて学習する技術である。NeRFは限られた数の画像から、そのシーンの詳細な3D表現を構築する。つまり、複数の視点から物体を捉えた画像をこの手法で学習することで任意の視点からの高品質な画像を合成することが可能となる。この3D表現は物体やシーンの光の散乱や反射の特性を含んでおり、それによって観測されていない視点からでも高精度な画像を生成することができる。

### 2.3.2 Bad-NeRF

Bad-NeRF[3]は、NeRFの拡張版であり、モーショントラッキングのかかった画像セットを扱う場合における性能向上に特化した技術である。Bad-NeRFでは、モーショントラッキングのかかった画像を複数の異なる視点からの画像の線形結合と捉え、それらの視点のパラメータを予測するモデル（スプラインモデルと呼ばれる）をNeRFモデルと並行して学習させる。Bad-NeRFの学習プロセスの概要を図2.2に示す。

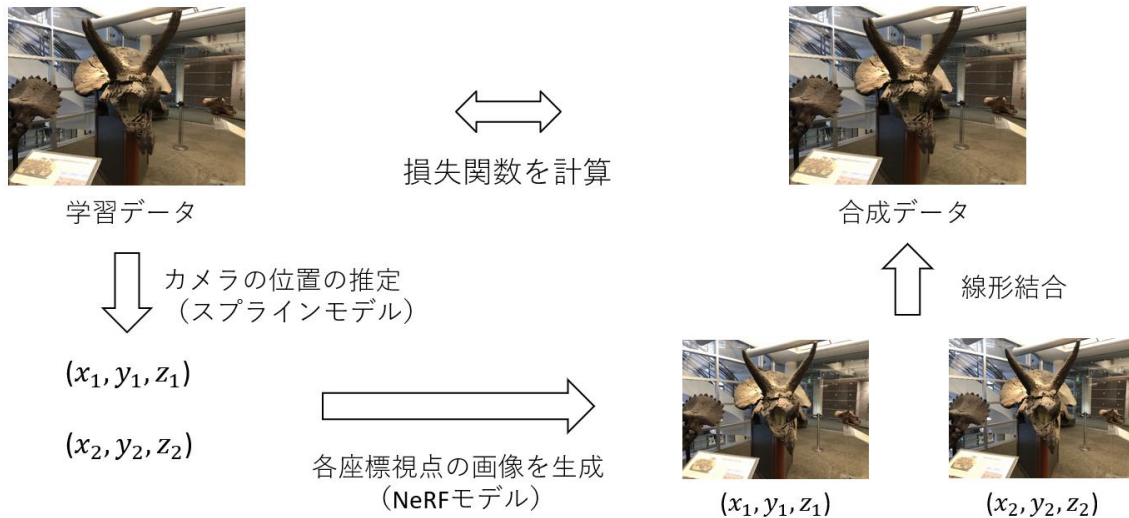


図2.2 Bad-NeRFの学習プロセスの概要

図2.2では、一つの画像に二つのカメラパラメータを対応させた場合の学習プロセスを表している。なお、カメラのパラメータは位置と方向の二つの情報を含んでいるが、簡単のため図2.2では「位置」に焦点を当てている。

### 2.4 むすび

本章では、カメラの絞りを動的に調節できる撮影技術であるPA Cameraと、自由視点画像生成技術であるNeRF、さらにモーショントラッキングのかかった画像に特化した自由視点画像生成技術であるBad-NeRFについて紹介した。

## 第3章 提案手法

---

### 3.1 まえがき

本章では、安価な撮影デバイスを用いることを前提とし、自由視点画像生成モデルの訓練に用いる学習データの撮影回数を削減する手法を提案する。撮影には PA Camera を使用し、露光を多重化して画像を取得する。さらに、取得画像を Bad-NeRF に学習させる。通常、Bad-NeRF ではスプラインモデルによりカメラパラメータ推定を行う。しかし、本研究では、PA Camera による撮影時のカメラパラメータが既知であるため、そのパラメータ情報を活用してスプラインモデルの訓練を省略する。

### 3.2 提案手法

本節では、PA Camera と Bad-NeRF の技術を利用して自由視点画像を生成する手法を提案する。PA Camera で取得した画像を、学習データとして直接 Bad-NeRF に学習させ、自由視点画像を生成するモデルを訓練する。PA Camera で撮影された画像のカメラパラメータは既知であり、訓練の際に、学習データと共に Bad-NeRF に入力する。この追加入力により、Bad-NeRF のスプラインモデルのトレーニングを省略することができる。スプラインモデルによる推定カメラパラメータを使用した学習と比較して、カメラパラメータが正確であるため、生成モデルの品質を向上させることができる。この手法の概要を図 3.1 に示す。

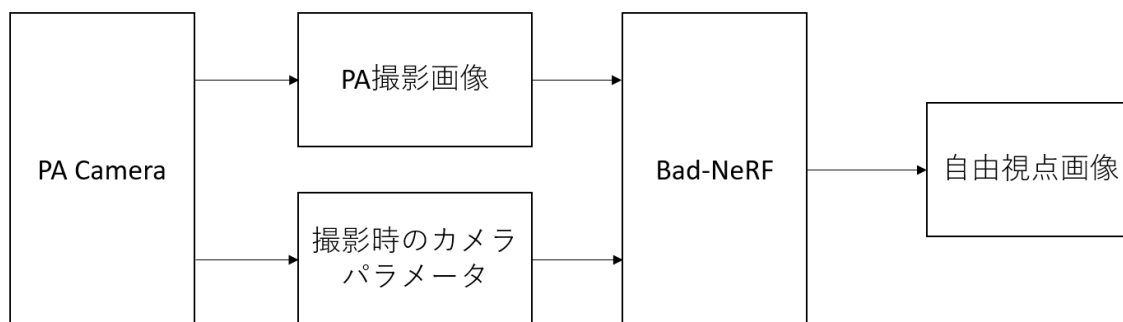


図 3.1 PA Camera と Bad-NeRF を用いた自由視点画像生成の概要

### 3.3 提案手法の優位性

本節では、3.2節で提案した手法の優位性について述べる。2.2.3で述べた先行研究では、撮影画像をもとに単一視点からの画像を抽出するアルゴリズムを用いるために、それぞれの視点が複数回現れるように露光のパターンを設定している。こうすることで、各視点からの画像を正確に分解・再構成することができる。一方で、提案手法では、各視点が全ての露光を通して複数回現れる必要はない。例えば、8枚の異なる視点からの画像を取得したい場合、露光を二重で行うと4回の露光で全ての視点の画像を得ることができる。このように、提案手法では露光回数の削減ができる。

### 3.4 むすび

本章では、安価な撮影デバイスを用いることを前提とし、自由視点画像生成モデルの訓練に用いる学習データの撮影回数を削減する手法を提案した。さらに、その手法の優位性について述べた。

## 第4章 実験

### 4.1 まえがき

本章では、第3章において提案手法として示した手法の有効性を検証する実験について述べる。NeRFの訓練済みのモデルを使用して生成した5種類のデータセットを用いる。提案手法と従来のNeRFの比較及びBad-NeRFにカメラパラメータを入力する効果の検証を行う4項目の実験を行う。定量的な評価方法としてPSNRを用いる。簡単のため本節では5種類のデータセットのうち特にhornsに焦点を当てて実験概要や結果について説明する。実際には全てのデータセットに対して同様の4項目の実験を実施する。

### 4.2 データセット

本研究では、学習データは、シミュレーションを用いて用意する。学習データの生成には「NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis」で提供されている学習データセットであるfern, flower, fortress, horns, trexの5種類を用いる。まず、各データセットをNeRFに学習させたモデルを用いて、16種類の異なるカメラパラメータからの画像を生成する。なお、16種類の異なるカメラパラメータは以下の図4.1のように正方形の軌道上に配置する。また、各カメラの向きは、全てのパラメータにおいて、対象物体付近のある一点を向くように設定する。

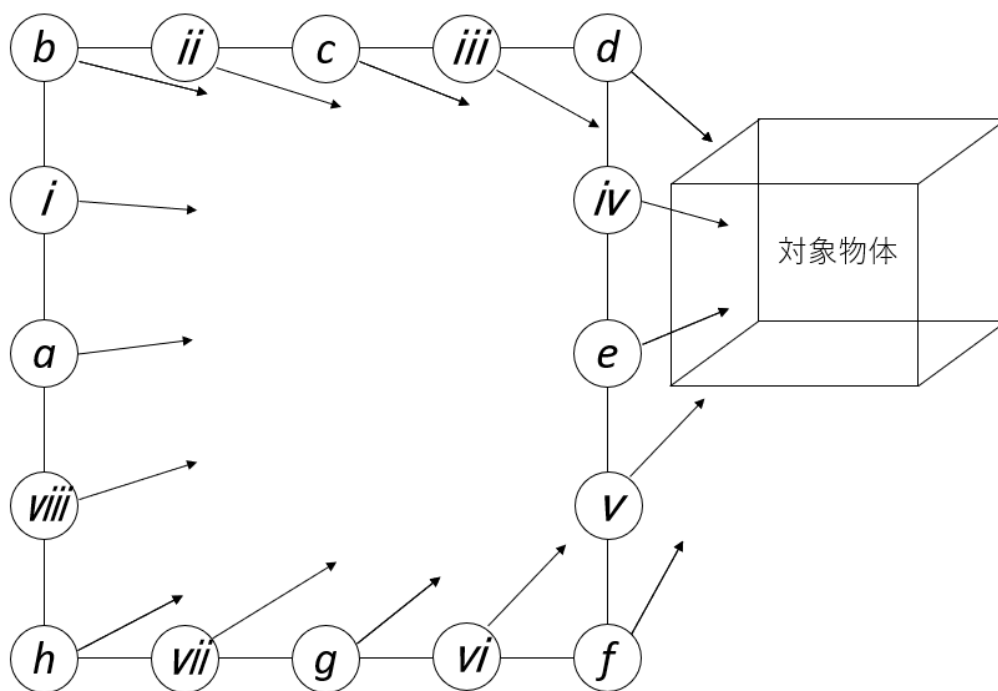


図4.1 生成した画像の16種類のカメラパラメータ

図 4.1 では、正方形の軌道上の丸図形がカメラの位置、矢印が各位置のカメラの向く方向を表している。正方形の軌道上で隣り合うカメラ位置間の距離は全て等しく、図中において、カメラの位置を、 $a \sim h$ ,  $i \sim viii$  で識別している。また、生成した画像の例として、horns データセットを学習して  $b, c, d$  に対応するパラメータで生成した画像を以下の図 4.2 に示す。



図 4.2 horns のカメラパラメータ  $b, c, d$  に対応する生成画像

また、fern, flower, fortress, trex の 4 種類のデータセットについて、 $b$  に対応するパラメータで生成した画像を以下の図 4.3 に示す。



図 4.3 fern, flower, fortress, trex のカメラパラメータ  $b$  に対応する生成画像

以後、各データセットにおいて、生成した 16 枚の画像を、図 4.1 のカメラパラメータの識別文字の大文字で記述する。例えば、生成した画像のうちカメラパラメータ  $a$  に対応するものは  $A$  と表記する。 $A, B, \dots, H$  は各実験の学習データの作成に用い、 $I, II, \dots, VIII$  は、4.7 節で記述する評価実験に使用する。各実験では、必要に応じて画像の線形結合を行い、学習データを作成する。画像の線形結合には Python の Pillow ライブラリの `blend` 関数を用いて、各画像の透過率が等しくなるように設定する。また、以後、記述の際に複数の画像を結合してできた画像を表す関数  $P$  を用いる。例えば  $P(A, B)$  は、 $A$  と  $B$  を線形結合した画像を指す。

### 4.3 評価指標

本実験では、真値 (Ground Truth (GT)) 画像と生成画像の PSNR[4]を計算することでモデルの品質を定量的に評価する。GT 画像には、 $I, II, \dots, VIII$  の 8 枚を用いる。各実験で訓練した生成モデルを用いて、 $i, ii, \dots, viii$  のカメラパラメータで画像を生成する。それぞれの画像について、GT 画像との PSNR を計算し、平均することで最終的な生成モデルの評価値を計算する。なお、PSNR は二つの画像間のピーク信号対雑音比をデシベル単位で計算したものである。計算方法は以下の通りである。まず、MSE (二つの画間のピクセル値の差の二乗の平均) を計算する。計算式を以下の式(4.1)に示す。

$$\text{MSE} = \frac{\sum_{M,N} [I_1(m,n) - I_2(m,n)]^2}{M * N} \quad (4.1)$$

続いて、MSE を使用して PSNR を計算する。PSNR 値を以下の式(4.2)で求める。

$$\text{PSNR} = 10 \log_{10} \left( \frac{R^2}{\text{MSE}} \right) [\text{dB}] \quad (4.2)$$

## 4.4 実験概要

### 4.4.1 概要

提案手法と従来の NeRF の比較及び Bad-NeRF にカメラパラメータを入力する効果の検証を行う 4 項目の実験を行う。以下に各実験の概要と結果を記述する。

### 4.4.2 Bad-NeRF による二重化画像 4 枚の学習

提案手法に基づき、PA Camera で露光を二重化して 4 回の露光を行った場合を想定した実験を行う。学習データには、 $P(A,B), P(C,D), P(E,F), P(G,H)$  の 4 枚を用る。提案手法に基づき、この 4 枚の画像と 1 枚の画像につき 2 つのカメラパラメータを Bad-NeRF に入力し学習させる。例として、horns データセットにおいて学習に用いた 4 枚の画像を以下の図 4.4 に示す。なお、他の 4 種類のデータセット、fern, flower, fortress, trex それぞれについても同様に二重化画像を 4 枚作成している。





図 4.4 horns の二重化画像 4 枚

結果例として、horns の二重化画像 4 枚を学習したモデルを用いて図 4.1 の *b, c, d* のパラメータで生成した画像を以下の図 4.5 に示す。



図 4.5 Bad-NeRF による二重化画像 4 枚の学習結果 (horns)

目視では正確に多重化を分解・再構成できていると判断できる。高品質な自由視点画像が生成できたと考えられる。

#### 4.4.3 Bad-NeRF による既知のカメラパラメータ無しでの二重化画像4枚の学習

4.4.2 節同様に、PA Camera で露光を二重化して、4回の露光を行った場合を想定した実験を行う。学習データセットは4.4.1 節で用いたものと同様である。ただし、提案手法とは異なり、Bad-NeRF への既知のカメラパラメータの入力を行わず学習を行う。結果例として、horns の二重化画像4枚をカメラパラメータの入力を行わずに学習したモデルを用いて図4.1の**b,c,d**のカメラパラメータで生成した画像を以下の図4.6に示す。



図4.6 Bad-NeRF による既知のカメラパラメータ無しでの二重化画像4枚の学習結果 (horns)

全ての画像において全体的にぼけが認められ、生成された自由視点画像の品質が低いことが分かる。

#### 4.4.4 NeRF による単一視点画像4枚の学習

続いて、露光の多重化をせずに4回の露光を行った場合を想定した実験を行う。学習セットには**B,D,F,H**の4枚を用いる。これらの画像をNeRFに学習させ、生成モデルを訓練する。結果例として、horns の単視点画像4枚を学習したモデルを用いて、図4.1の**b,c,d**に対応するパラメータで生成した画像を以下の図4.7に示す。



図4.7 NeRF による単一視点画像4枚の学習結果 (horns)

図 4.7 では、視点  $b, d$  からの画像の品質が高い一方で、視点  $c$  からの画像の品質が低いことが分かる。トレーニンググループ中に GT 画像と比較して損失関数を計算していた一部の視点 ( $b, d, h$  の三つの視点) からの画像は高品質に生成できているものの、多くの視点からの画像は正確に生成できなかった。

#### 4.4.5 NeRF による単一視点画像 8 枚の学習

続いて、露光の多重化をせずに 8 回の露光を行った場合を想定した実験を行う。学習セットには、 $A, B, \dots, H$  の 8 枚を用いる。これらの画像を NeRF に学習させ、生成モデルを訓練する。結果例として、horns の単視点画像 8 枚を学習したモデルを用いて、図 4.1 の  $b, c, d$  に対応するパラメータで生成した画像を以下の図 4.8 に示す。



図 4.8 NeRF による単一視点画像 8 枚の学習結果 (horns)

4.4.4 節の結果と比較して学習効果が高く、高品質な自由視点画像が生成できていることがわかる。



#### 4.4.6 各実験結果の定性的評価

4.4.2 節から 4.4.5 節で行った 4 項目の実験結果を定性的に評価する．視覚的に比較するため拡大図を含んだ画像を図 4.9 に示す．



図 4.9 各実験の比較

図 4.9 より，単一視点 4 枚を NeRF に学習させた場合と，二重化画像 4 枚をパラメータ無しで Bad-NeRF に学習させた場合の生成モデルの品質が明らかに低下していることが分かる．また，提案手法による生成モデルは高品質ではあるが，図 4.9 の椅子の影の部分に注目すると，GT 画像や，単一視点画像 8 枚を NeRF に学習させた場合と比較して，雑音が入っていることが分かる．

## 4.5 評価実験

4.4 節で示した5項目の実験の後、4.3 節で述べた評価指標に基づき、GT 画像と、各生成モデルによる生成画像の PSNR を計算することでモデルの品質を定量的に評価する。得られた結果を以下の表 4.1 に示す。

表 4.1 各生成モデルの PSNR

(単位 : dB)

学習データ	fern	flower	fortress	horns	trex
二重化画像 4 枚 (提案手法)	<u>32.12</u>	<u>35.14</u>	<u>37.11</u>	<u>31.44</u>	<u>31.20</u>
二重化画像 4 枚 (カメラパラメータ 無し)	14.88	15.15	20.27	14.08	15.53
単一視点画像 4 枚	23.51	22.59	24.83	21.54	22.36
単一視点画像 8 枚	<b>34.74</b>	<b>36.66</b>	<b>39.18</b>	<b>35.57</b>	<b>32.93</b>

## 4.6 考察

### 4.6.1 Bad-NeRF へのカメラパラメータ入力の有無比較

二重露光による画像 4 枚を Bad-NeRF に学習させる場合、撮影時のカメラパラメータを Bad-NeRF に入力することで 5 種類全てのデータセットにおいて PSNR が向上し、平均で 17.42 dB 向上した。Bad-NeRF において、並行して訓練されるスプラインモデルと NeRF モデルは連携しているため、スプラインモデルによるカメラパラメータの予測精度が NeRF モデルの訓練効果に影響を与える。撮影時のカメラパラメータを入力してスプラインモデルの訓練を省略することで、NeRF モデルの訓練の際に用いられるカメラパラメータの正確さが保証され、より品質の高い自由画像を生成できるようになったと考えられる。

### 4.6.2 二重露光画像 4 枚と単一視点画像 4 枚の比較

同じ 4 回の露光で得られた画像をもとに生成モデルを訓練する場合、5 種類すべてのデータセットにおいて、露光の多重化を行った場合 (4.4.2 節の実験) の PSNR が露光の多重化を行わない場合 (4.4.4 節の実験) と比較して高い値になり、平均で 10.4 dB 上昇した。露光の多重化を行うことで、4 回の露光でありながら実質的には 8 個の異なる視点からの画像を NeRF に学習させることができる。そのため、多重化画像の分解・再構成という処理の難があるものの、生成モデルの品質が向上したと考えられる。

### 4.6.3 二重露光画像 4 枚と単一視点画像 8 枚の比較

二重露光による画像 4 枚を Bad-NeRF に学習させた場合 (4.4.2 節の実験) と単一視点画像 8 枚を NeRF に学習させた場合 (4.4.5 節の実験) を比較すると, 5 種類全てのデータセットにおいて後者の PSNR がより高い値となったものの, 平均して 2.4 dB の差であり, 品質に大きな差が生まれていないことが分かった.

## 4.7 むすび

本章では, 提案手法の検証に用いた実験の概要, その結果, さらにそれらに対する考察を述べた. 結果として, 二重露光による画像と撮影時のカメラパラメータを Bad-NeRF に学習させる本研究の手法により, 4 回の露光で, 単一視点画像 4 枚を NeRF に学習させた場合と比較して品質向上が達成され, 単一視点画像 8 枚を NeRF に学習させた場合に迫る品質の自由視点画像が生成できた. また, 撮影時のカメラパラメータを Bad-NeRF の学習に活用するアプローチが, 生成画像の品質向上に高い効果があることが確認できた.

## 第5章 結論と今後の課題

---

### 5.1 結論

本研究では、安価な撮影デバイスを用いることを前提とし、自由視点画像生成モデルの訓練に用いる学習データの撮影回数を削減する手法を提案した。提案手法は、PA Camera により露光を多重化して取得した画像をカメラパラメータと共に Bad-NeRF に学習させるという手法である。この手法の有効性を検証するために実験を行い、以下の三つの結果を得た。まず、カメラパラメータを入力することで生成画像の品質が向上する。次に、単一視点画像4枚を NeRF に学習させた場合と比較して、多重化された4回の露光により取得した画像とそのパラメータを Bad-NeRF に入力した場合の生成画像の品質が高くなる。最後に、露光を多重化することで、NeRF に8枚の画像を学習させた生成モデルに迫る品質の生成モデルを、4回の露光で取得した画像により訓練することができる。以上の結果から、提案手法の有効性が確認できたと考える。

### 5.2 今後の課題

露光の多重化と Bad-NeRF の連携により、少ない露光回数で比較的高品質な自由視点画像の生成ができた。しかし、二重化画像4枚を学習した生成モデルは単一視点画像8枚を学習した生成モデルと比較して品質が劣り、改善の余地がある。今後の研究を通して品質が劣化する原因を特定し、さらなる品質改善を目指す。

## 謝辞

本研究において、PA Camera と NeRF 技術を組み合わせるアプローチを提案して下さった東京大学大学院情報理工学系研究科の堀崎遼一准教授, 及び本研究を長期にわたり指導し見守って下さった渡辺裕教授に深く感謝申し上げます。

また, 研究室で貴重なご意見とアドバイスを下さった同期生および先輩方にも心より感謝いたします。

最後に, 私をここまで育ててくださり, 精神面や金銭面で多くの力添えをしていただいた家族に心より感謝いたします。



## 参考文献

- [1] 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, Dec. 2021.
- [2] C. Liang, T. Lin, B. Wong, C. Liu, and H. Chen, "Programmable Aperture Photography: Multiplexed Light Field Acquisition," *In ACM Transaction on Graphics*, Vol. 27, No.3, pp. 1-10 Aug. 2008.
- [3] P. Wang, L. Zhao, R. Ma, and P Liu, "BAD-NeRF: Bundle Adjusted Deblur Neural Radiance Fields," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023)*, pp. 4170-4179, Jun. 2023.
- [4] MathWorks. (n.d.). PSNR. Retrieved January 17, 2024, from <https://jp.mathworks.com/help/vision/ref/psnr.html>

## 図一覧

図 2.1	フィルタ上の穴のパターン.....	5
図 2.2	Bad-NeRF の学習プロセスの概要.....	7
図 3.1	PA Camera と Bad-NeRF を用いた自由視点画像生成の概要.....	8
図 4.1	生成した画像の 16 種類のカメラパラメータ.....	10
図 4.2	horns のカメラパラメータ $b, c, d$ に対応する生成画像.....	11
図 4.3	fern, flower, fortress, trex のカメラパラメータ $b$ に対応する生成画像.....	11
図 4.4	horns の二重化画像 4 枚.....	13
図 4.5	Bad-NeRF による二重化画像 4 枚の学習結果 (horns).....	13
図 4.6	Bad-NeRF による既知のカメラパラメータ無しでの.....	14
図 4.7	NeRF による単一視点画像 4 枚の学習結果 (horns).....	14
図 4.8	NeRF による単一視点画像 8 枚の学習結果 (horns).....	15
図 4.9	各実験の比較.....	16

## 表一覧

表 4.1 各生成モデルの PSNR.....	17
-------------------------	----