

修士論文  
令和6年度

画像の前処理を用いた Neural Radiance Fields(NeRF)

の精度向上に関する研究

— Research on improving the accuracy  
of Neural Radiance Fields (NeRF) using image preprocessing —

京都産業大学大学院  
先端情報学研究科  
博士前期課程 2年  
386094  
西尾 漱一郎

## 要旨

自由視点画像生成技術である Neural Radiance Fields(NeRF)において、入力画像に前処理を施すことで NeRF の再構成精度を向上させる研究を行なった。また、自身の研究室をコンピュータ上で再現し、その空間とのインタラクションを可能にするシステムの開発を行った。

NeRF とは、2020 年に開発されたニューラルネットワークを用いた 3 次元再構成技術である。現在までの NeRF 研究は、ネットワークを改善することで空間の再構成精度を向上させる研究が主である。一方、精度は入力画像に大きく依存しているにもかかわらず、入力画像の重要性に関する研究は少ない。そこで本研究では、入力画像に重きを置いて、Histogram Stretching や Histogram Equalization などの画像のヒストグラムに対する前処理や、撮影時のカメラの視線方向の工夫、Dynamic Range の変更といった入力画像やその撮影方法に関して検証を行い、再構成精度の向上を目指した。また、NeRF の実用性を示すために再構成したモデルをゲームエンジンに読み込ませ、ヘッドマウントディスプレイで研究室内を見ることができ、またその中を移動するといったインタラクションを行えるゲーム・システムを開発した。

本研究のために、RICOH の THETA を用い 360 度の正距円筒画像を撮影し、Python で前後左右上下の 6 枚の透視投影画像に変換を行った。結果 1 回の撮影で 6 枚の画像を取得することができ、かつ写っていない場所を減らすことができた。次に精度向上に関する研究成果として、Histogram Stretching, Histogram Equalization, Histogram Matching, ガウス分布近似, 黒色成分の調整といった画像のヒストグラムに対する前処理を行った画像, および High Dynamic Range 画像や、撮影時の視線方向に工夫を加えた画像を NeRF に入力し、それらの前処理や撮影方法が再構成精度に寄与することを検証する実験を行った。結果、視線方向の工夫が最も精度の向上に貢献する要因であることが確認できた。また、実験を通して、NeRF は黒色成分の再構成精度が低下する傾向にあることがわかった。さらに NeRF とインタラクション可能なシステムの開発について、ゲームエンジンである Unreal Engine 5 を使うことで、NeRF で再構成した研究室のモデルを読み込み、HMD でそのモデル内に対してインタラクションを行い、歩くことができるシステムを開発した。

以上の結果から、本研究では入力画像に対する前処理を施すことで NeRF の再構成精度を向上させることが可能であり、本研究の結果、視線方向の変更が最も精度向上に起因することを確認した。また、NeRF の前処理において、黒色成分を適切に減らすことで精度上げることができた。さらに、ヘッドマウントディスプレイでインタラクションすることは可能であり、NeRF の実用の一例を示すことができた。

# 目次

<b>第 1 章 序論</b> .....	1
1.1 3次元再構成技術と VR.....	1
1.2 本研究の目的.....	2
1.3 本論文の構成.....	2
<b>第 2 章 3次元再構成技術と NeRF の概要</b> .....	3
2.1 3次元再構成技術の概要.....	3
2.2 Structure from Motion.....	3
2.3 フォトグラメトリ.....	4
2.4 NeRF.....	4
<b>第 3 章 画像の前処理を用いた NeRF の精度向上手法</b> .....	7
3.1 入力画像の取得.....	7
3.1.1 360度カメラを使用した研究室の撮影.....	7
3.1.2 正距円筒画像から透視投影画像への変換.....	8
3.2 画像の前処理.....	12
3.2.1 Histogram Stretching.....	12
3.2.2 Histogram Equalization.....	14
3.2.3 Histogram Matching.....	16
3.2.4 黒色成分の調整.....	18
3.2.5 ハイダイナミックレンジ画像.....	19
3.2.6 撮影時の視線方向の工夫.....	21
<b>第 4 章 実験結果</b> .....	23
4.1 Colmap による SfM の結果.....	23
4.2 再構成結果の評価.....	26
4.3 結果の総評.....	32
4.4 今後の課題.....	33

第 5 章 NeRF 空間とのインタラクション.....	34
5.1 システム開発の背景.....	34
5.2 実装.....	34
5.3 今後の課題.....	35
第 6 章 結論.....	36
謝辞.....	37
参考文献.....	38
業績.....	40
付録.....	41

# 第 1 章 序論

## 1.1 3次元再構成技術と VR

Meta 社の Meta Quest, PICO 社の Pico, Apple 社の Apple Vision, Pimax 社の Pimax Crystal など, さまざまな種類の HMD が販売されており, VR が身近な存在になりつつある. また, PlayStation VR や Steam VR などの VR ゲーム機や VR 環境に対応した VR ゲームが数多く発売されている. さらに, 動画配信サービスである YouTube でも手軽に VR を体験することができる. VR はエンターテインメントでの活用だけでなく, 旅行体験や不動産の内覧などビジネス面でも活用されている.

VR 技術の中に, 3次元再構成技術などを使って実世界を計測し, それを HMD などで体験することを目的とした研究開発がある. この技術では, 実世界の物体や空間の 3次元情報をコンピュータに取り込み, 仮想化することで, HMD などのデバイスを使って, その空間をコンピュータ内で体験することが可能となる. これを実現するためには, 実世界をコンピュータで扱うことができる 3次元データに変換する必要がある.

実世界の物体や空間を 3次元データに変換することを 3D キャプチャや 3次元復元, 3次元再構成と呼ぶ. これらの技術はコンピュータビジョン分野で長く研究されており, Google Earth などで実用化されている. 3次元再構成を大別すると, 深度センサである LiDAR などの 3D キャプチャに特化した撮影デバイスを用いる方法と, 普通のカラーカメラで撮影した大量の写真を利用する方法がある. 基本的には, 3D キャプチャに特化した撮影デバイスを用いる方法がより高精度な結果を得ることができる. 専用の 3D キャプチャデバイスは建築現場での空間把握のためのモデル生成や, 災害に対する都市シミュレーションのための 3D モデル生成に活用されている. この方法は, 専用の装置やシステムが必要となり, 3D キャプチャを手軽に利用することが難しい. またキャプチャ可能な物体の種類も制限される. それに対して普通のカラーカメラを用いる方法は, 手軽に 3D キャプチャを行うことができ, またキャプチャできる物体の種類も多くなるが, 再構成した 3次元データの幾何学的な精度が低くなるという問題点がある.

実世界を撮影した多数の写真画像から 3次元モデルを生成する新しい技術として Neural Radiance Fields (NeRF) や 3D Gaussian Splatting (3DGS) が提案されている. これらの技術は, 3次元モデルのクオリティを大幅に向上させた. これらの技術はコンピュータビジョン分野において盛んに研究が行われており, NeRF のオリジナル論文は, 発表からの 4年間で 8,000 件以上引用された.

NeRF の入力データは写真画像であるため, 3次元データの再構成精度は入力画像の質に依存する. G. Guidi ら[1]や Krzysztof Skabek ら[2]の研究は, フォトグラメトリ技術での研究結果ではあるが, HDR での撮影や Histogram Equalization 等の画像の前処理を行うことで再構成精度が向上することを示している. 本研究では, G. Guidi らや Krzysztof Skabek らの結果を踏まえ, NeRF に入力する画像データに対してこれらの前処理を適用することで,

再構成精度の向上の検討を行う。

NeRF が出力する 3 次元空間の情報は特殊なパラメータで構成されており，一般的なゲーム等に利用できる 3D モデルの形式である OBJ や FBX などとは大きく異なるため，専用のツールを用いなければ生成した 3D モデルを利用することができない。そこで本研究では，NeRF のデータを Unreal Engine 5 などのゲームエンジンで利用することで，HMD でインタラクションが行えるようにした。

## 1.2 本研究の目的

本研究の目的は NeRF の実用性を示すことである。NeRF は革新的な 3 次元再構成技術であり，多くの NeRF の論文は NeRF の深層ネットワーク構造を改善することによる精度向上の研究を行っている。日々，NeRF は進化している一方，入力データは変わらず画像であるため，入力画像の質が再構成精度に大きく影響を与える。また，NeRF が出力する 3D モデルはボリュームデータを表現する特殊なパラメータで構成されているため，その出力結果を Unity などの一般的なツールで扱うことが難しい。結果として，NeRF の応用に関する研究論文は少なく，あまり実用性を示せていない。そのため，この研究では 2 つの目標を掲げた。第一の目標は NeRF を使った 3 次元再構成の精度を向上させることである。第二の目標は，NeRF で取得した 3D データとの，HMD を使ったインタラクションを可能にすることである。

G. Guidi らや Krzysztof Skabek らは HDR での撮影や Histogram Equalization 等の画像の前処理を行うことでフォトグラメトリでの再構成精度が向上すること示している。本研究では NeRF に入力する画像データに対してそれらの前処理を適用することによる NeRF の再構成精度の向上の検証を目的とする。

本論文は，G. Guidi らや Krzysztof Skabek らで検証している前処理方法に加え，それら以外の複数の前処理方法を加えた 9 種の方法での再構成精度の向上の検証の結果を示す。また NeRF で出力される 3D モデルにインタラクションできるようなシステム開発について説明する。加えて，システム開発の結果を示す。

## 1.3 本論文の構成

本論文は，第 2 章で，3 次元再構成技術についての一般的な説明を行う。第 3 章では本研究の実験また提案内容について説明する。第 4 章では実験結果の報告を行う。第 5 章で，NeRF で再構成した空間とのインタラクションに関する成果を報告する。第 6 章で本研究の結論を報告する。

## 第 2 章 3 次元再構成技術と NeRF の概要

2 章では、まず、3 次元再構成技術の概要について説明する。次に従来の 3 次元再構成技術として、SfM とフォトグラメトリについて説明する。最後に、本研究で使用する、近年開発され、現在も研究が盛んに行われている新たな方法である NeRF について説明する。

### 2.1 3 次元再構成技術の概要

3 次元再構成は、立体物や立体空間を撮影した 2 次元情報である画像から、もとの立体形状、表面の反射特性、奥行きなどの 3 次元形状を推定するコンピュータビジョンのタスクの 1 つとして研究が始まった。1960 年代に、2D 画像から 3 次元形状を推論する研究が行われ始めた。これらの研究が 3 次元再構成や 3D スキャンとも呼ばれるものになった。現在では、3 次元再構成の方法は大きく分けて 2 つに分類することができる。一つは LiDAR 等の深度カメラを用い、複数の深度画像を組み合わせることで高品質な物体や環境の 3 次元形状を再構成する方法である。もう一方は、一般的なカラーカメラで撮影した RGB 画像を入力とし、画像処理や Neural Network を使い 3 次元形状を再構成する方法である。本論文の研究は後者の方法に関するものである。

### 2.2 Structure from Motion

Structure from Motion (SfM) とは、多視点から対象物を撮影したカメラ画像を元に、その画像を撮影したカメラの位置姿勢と対象物の 3 次元形状を同時に求める手法のことである。対象物を多視点から撮影した場合、対象物上の同じ位置が複数の画像に映ることがある。この位置を対応点と呼ぶ。その対応点からそれぞれの画像を撮影した時のカメラ位置を推定することができる。推定したカメラの位置関係と画像中の対応点の座標から、その 3 次元位置を計算することで、対象物の 3 次元形状およびカメラの位置姿勢を推定する。

SfM は、原理的には 2 枚の画像でも動作する。3 枚以上の画像を使う場合には、対象物が他の物体の後ろに隠れて見えなくなるオクルージョンの影響を軽減することができる。SfM により出力される 3 次元形状は、特徴点をもとに再構成した形状であるため、疎な 3 次元点群となる。そのため推定された点群をそのまま利用するのではなく Multi View Stereo などを用いることで、密な 3 次元点群を生成する。また、ソフトウェアを使い自由表面形状モデリングを行い、ポリゴンメッシュ化するなどによって、より高精度な 3 次元形状を再構成することができる。

SfM の応用例として、建築現場での環境測量や文化財の調査など幅広い分野で活用されている。また、近年ドローンの普及により、より手軽にかつ大規模な写真測量を行うことができる。さらに自動運転やロボット制御で使用されている自己位置推定と周辺環境地図生成を同時に行う SLAM にも応用されている。

Colmap[3] は、Johannes Schönberger らによって開発された CUI と GUI を備えた SfM お

よび MVS のパイプラインである。オープンソースソフトウェアであり、最近の 3 次元再構成技術において、入力画像の位置関係の推測と撮影画像のカメラパラメータを取得するための前処理として広く利用されている。

## 2.3 フォトグラメトリ

フォトグラメトリは、通常の写真画像や LiDAR などのレーザスキャナーで取得した深度情報をもとに、対象物の 3 次元モデルを生成する技術である。また、その技術を実装したソフトウェアの総称である。前述の SfM はフォトグラメトリを実現する測定原理の一つである。

数々の企業や団体がフォトグラメトリを開発している。Microsoft は 2008 年に Photosynth というフォトグラメトリサービスを提供した[4]。また航空宇宙工学で有名なアメリカの NASA が Ames Stereo Pipeline というオープンソースのフォトグラメトリソフトウェアを開発し、NASA の宇宙船や衛星で取得したステレオ惑星データから、月や火星などの固体表面をもつ天体のデジタル標高モデルやオルソ画像の生成を行った[5]。最新のフォトグラメトリソフトウェアとして Realty Capture が Epic Games 社から教育機関等に無料で提供されている。Realty Capture は、カメラ画像とレーザスキャンの測定結果を同時にインポートすることができ、画像のアライメント、キャリブレーション、ポリゴンメッシュの生成、テクスチャリング、スムージングなど幅広い処理を 1 つのアプリケーションで行うことができる[6]。

フォトグラメトリソフトウェアが生成する 3 次元モデルは、PLY, OBJ, FBX など、一般的な 3DCG で利用されているファイル形式で出力される。これらは物体の表面の有無とその色を 1 つに確定する。そのため、ガラス、水面、金属の表面、鏡といった光の反射の影響で角度によって見え方が変わってしまう物質の再現は難しい。

フォトグラメトリは考古学や、建築、映画など幅広い分野で応用されている。また GoogleEarth もフォトグラメトリを利用し、フォトリアリスティックな地球をコンピュータ上で再現している[7]。

## 2.4 NeRF

Neural Radiance Fields (NeRF) は 2020 年に B.Mildenhall らが開発した新しい 3 次元再構成技術である[8]。従来の 3D モデルの表現手法の多くは「ポリゴンメッシュ」である。「ポリゴンメッシュ」は、三角形や四角形などのポリゴンで 3D オブジェクトの表面を表現する方法である。ポリゴンメッシュは「物の表面か」、「物の内側か」、「空間か」などを 0/1 で表現する。一方、NeRF は「ボリューム」という表現方法を採用している。ボリュームは物の表面かどうかを連続値であるボリューム密度  $\sigma$  で表現する。ボリューム密度  $\sigma$  とは、3 次元空間における点  $p$  について、光線が点  $p$  に衝突した際、それが物体の表面に衝突したとされる確率の高さを表す。NeRF では、3 次元空間上の点  $p$  におけるボリューム密度  $\sigma$  をニ

ニューラルネットワーク  $NN_\sigma$  を用いて、式(2-1)として定義する。

$$\sigma = NN_\sigma(p) \quad (2-1)$$

ここで、 $\sigma \in \mathbb{R}, p \in \mathbb{R}^3$  である。ボリューム密度により、複雑な形状をもつオブジェクトの再構成精度を向上させている。

3D モデルは形だけでなく、色も重要な要素である。NeRF では、3次元空間上にある点  $p$  の色  $c$  はニューラルネットワーク  $NN_c$  を用いて式(2-2)として定義される。

$$c = NN_c(p, d) \quad (2-2)$$

ここで、 $c \in \mathbb{R}^3, p \in \mathbb{R}^3, d \in \mathbb{R}^3$  とし、 $d$  は視線の方向を表すベクトルである。色は見る向きによって影や光の影響により色が違って見える可能性がある。そのため、 $NN_c$  に点  $p$  の座標だけでなく、視線方向を表すベクトル  $d$  を入力することで、鏡やガラス、水面など、見る角度によって色が違うマテリアルを扱うことができ、よりリアリスティックな色の表現を可能にしている。図 2.1 は NeRF の NN を示した図である。まずカメラ光線に沿って、各点の三次元座標  $p(x, y, z)$  と視線方向  $d(\theta, \phi)$  の 5次元情報を  $F_\theta(NN_\sigma, NN_c)$  に入力する。その後、 $F_\theta(NN_\sigma, NN_c)$  から、色情報である  $c(RGB)$  とボリューム密度  $\sigma$  を出力し、レンダリングを行う。

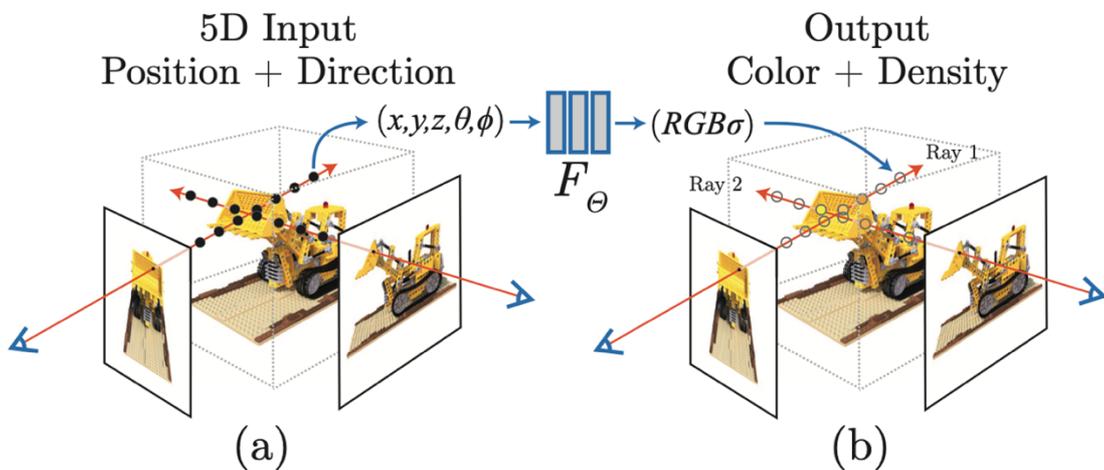


図 2.1 NeRF のネットワークの概要図

- (a) カメラ光線に沿って、三次元座標  $p(x, y, z)$  と視線方向  $d(\theta, \phi)$  の 5次元情報を  $F_\theta(NN_\sigma, NN_c)$  に入力する。
- (b)  $F_\theta(NN_\sigma, NN_c)$  から、色情報である  $c(RGB)$  とボリューム密度  $\sigma$  を出力し、レンダリングを行う。

NeRF はボリュームレンダリングを用いて再構成結果を描画する。カメラ位置から生成する画像の各ピクセルに光線を射出し、その光線上に一定間隔でサンプリング点を取る。そ

のサンプリング点の3次元座標 $p(x, y, z)$ と方向を表すベクトル $d(\theta, \phi)$ をニューラルネットワーク $NN_\sigma(p)$ と $NN_c(p, d)$ に入力し、 $NN_\sigma(p)$ からその点のボリューム密度 $\sigma$ と $NN_c(p, d)$ からその点の色 $c$ を出力する。1つのピクセルには複数のサンプリング点が存在する。ボリューム密度に応じて、各サンプリング点の色を合成し、そのピクセルの色の値を1つに決定する。シーンを変えるたびにNNにサンプリング点のボリューム密度と色を問い合わせる必要があるため、描画コストが非常に高い。

数多くのNeRFが開発されている。NvidiaはNeRFを高速に実行できるオープンソースソフトウェアであるInstant NGPを提供している[9]。Googleの研究チームは、高速なNeRFモデルであるMip-NeRF[10]、リアルタイム再構成可能なSMERF[11]などを開発した。NeRFは一般的には特定のオブジェクトに対しての再構成であるが、空間に対しての応用もされており、Google Mapでは、衛星画像とストリートビューの画像をNeRFに応用し、世界中の風景を3Dで見ることができるimmersive viewを提供している。バークレーAI研究所は2022年からNerf Studioプロジェクトを開始しており、独自のNeRFモデルであるNeRFactoから、Mip-NeRF, Instant NGP, など、2024年12月現在で22種類のNeRFモデルを利用できるAPIを提供している[12]。また、NeRFとは異なる技術になるが、3次元空間を3次元ガウス分布で近似表現することで、リアルな再構成結果を作ることができる3D Gaussian Splatting(3DGS)も開発されている[13]。

## 第 3 章 画像の前処理を用いた NeRF の精度向上手法

本章では, 画像の前処理を用いた NeRF の精度向上のための提案手法について説明する. NeRF に関する研究は, そのニューラルネットワークの改善による精度向上を目指すものが主である. 一方, NeRF の再構成結果のクオリティは, 入力画像の質に大きく依存している. そこで本研究では, NeRF に入力する画像に対して 9 種類の前処理を行い, Colmap によるカメラポーズの推定の結果および NeRF での 3 次元再構成結果の比較を行った. 本研究では, 自身が所属する研究室[図 3.1]の 3 次元再構成を行った.

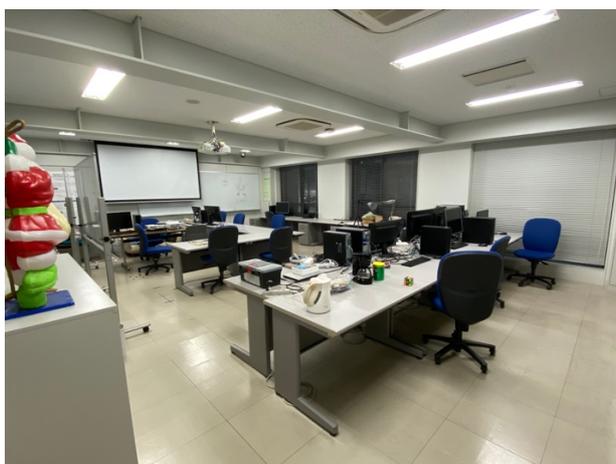


図 3.1 京都産業大学 蚊野研究室 (iPhone での撮影)

### 3.1 入力画像の取得

#### 3.1.1 360 度カメラを使用した研究室の撮影

研究室全体を撮影した RGB 画像で実験を行った. 研究室を撮影するために, 360 度撮影が可能な RICOH の THETA を使用した. 360 度カメラを使用することで, 撮影時間を大幅に短縮することができる. 解像度の違いが再構成結果に影響を与えるかどうかを検証するため, 撮影可能な解像度が異なる RICOH THETA V と RICOH THETA Z1 の 2 つのデバイスでそれぞれ撮影した. THETA V と THETA Z1 の仕様を表 3.1 に示す.

表 3.1 THETA V と THETA Z1 の仕様

	RICOH THETA V	RICOH THETA Z1
静止画解像度	5376×2688	6720×3360
撮像素子	1/2.3 型 CMOS (×2)	1.0 型 CMOS (×2)

研究室全体を撮影した位置を図 3.2 に示す. 図 3.2 は研究室の様子を表したものである. 赤色の×印が撮影位置であり, 24 箇所での撮影を行った. ×印と×印との距離間なるべく等間隔になるようにした. 撮影する際, 三脚に自撮り棒を取り付け, その上に THETA を取り付けて撮影を行った. このようにすることで, 360 度画像に起こりやすいカメラの三脚が

大きく写ってしまう現象を防ぐことができた。図 3.3 は、実際に実験を行った時の撮影状況である。TEATA のレンズ位置は V, Z1 とともに地面から 172cm である。

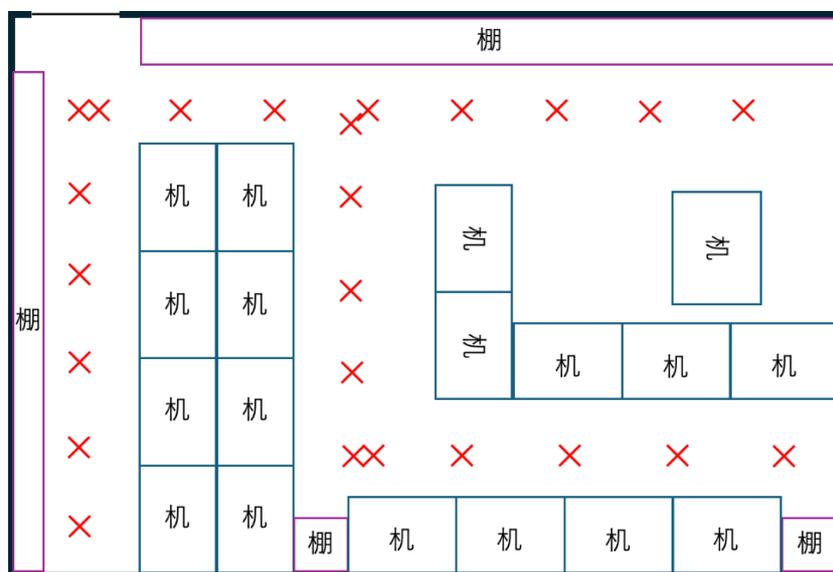


図 3.2 撮影位置を表した図, ×が撮影位置である, 全 24 箇所



図 3.3 三脚と自撮り棒を使った撮影, レンズの高さは地面から 172cm

### 3.1.2 正距円筒画像から透視投影画像への変換

THETA で取得できる画像は正距円筒画像である。正距円筒画像とは縦軸に緯度, 横軸に経度を取り, 2 軸が等間隔かつ直交する正距円筒図法(Equirectangular)で描かれた画像である。正距円筒画像では, 緯度が赤道から縦方向に離れるほど, 像が横方向に引き伸ばされ

る. 図 3.4 は正距円筒図法の例である. 画像中のオレンジの円は球体上では正円であり, 北極・南極に近い円は横方向に引き伸ばされている.

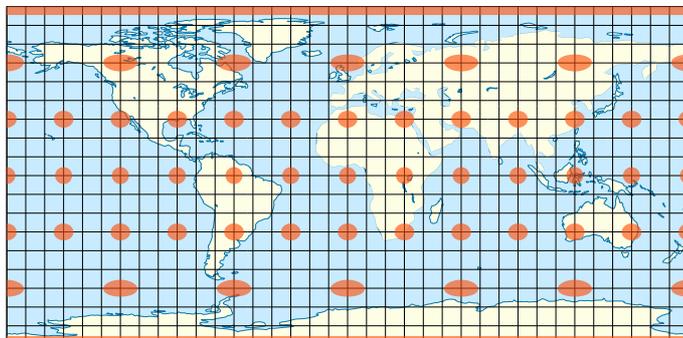


図 3.4 正距円筒図法の例 [14]

図 3.5 に正距円筒画像の座標系を示す. 正距円筒画像は座標を緯度 $\varphi$ と経度 $\theta$ の値で表す. 画像の中央を原点として, 緯度 $\varphi$ の範囲は 180 度である $-\pi/2 \sim \pi/2$ , 経度 $\theta$ の範囲は 360 度である $-\pi \sim \pi$ の範囲をとり, 画素数の縦横の比率は 1:2 になる.

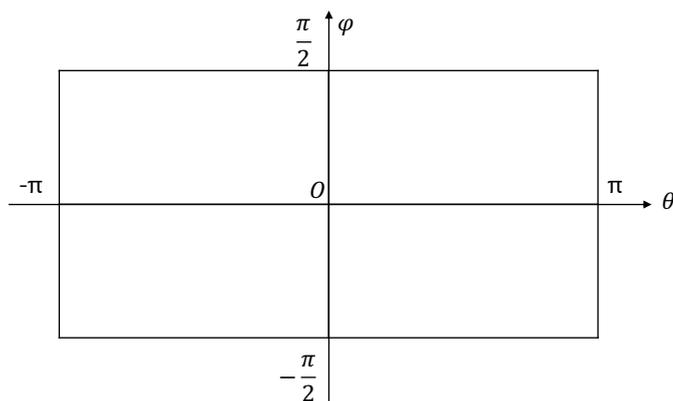


図 3.5 正距円筒画像の座標

正距円筒画像のまま 3次元再構成を行うことは難しい. そのため本研究では, THETA で撮影した正距円筒画像を透視投影画像に変換したものをを用いて NeRF での 3次元再構成を行った.

正距円筒画像は球面上のすべての位置を 1 枚の画像に表現している. 一方, 透視投影画像の視野角は 180 度未満である. したがって正距円筒画像の全体をいくつかの部分に分けて, 部分ごとに透視投影画像に変換する. 今回は図 3.6 に示すように 水平方向を 4 分割した. この時, 水平方向の視野角は 90 度になる. 垂直方向は北極と南極を中心として 90 度分の視野を取る. そして, 図 3.6 に示す領域ごとに緯度経度の値を透視投影画像の座標値に変換した.

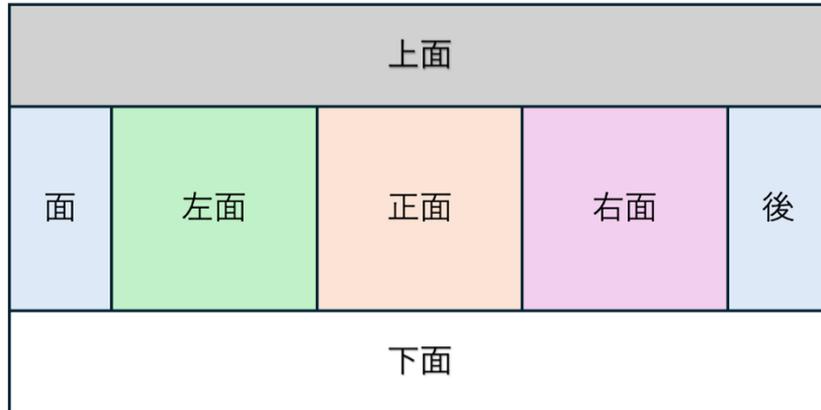


図 3.6 正距円筒画像と透視投影画像の関係

正距円筒画像から透視投影画像への変換は大きく 2 つの手順に分けて行う。まず、正距円筒画像の座標 $(\varphi, \theta)$ を単位球面上の座標 $(X, Y, Z)$ に変換する。地球儀上の 3 次元の世界地図と正距円筒図法で描かれた 2 次元の世界地図が対応しているように、正距円筒画像の緯度・経度を単位球面上の座標に対応付けることが可能である。単位球面は半径 1 の単位球の表面であり、球の表面を仮想的なカメラの撮像面と考える。

正距円筒画像上のある座標 $(\varphi, \theta)$ に対応する単位球面の 3 次元座標 $(X, Y, Z)$ は次の式(3-1,2,3)で求めることができる。

$$X = \cos\varphi\cos\theta \quad (3-1)$$

$$Y = \cos\varphi\sin\theta \quad (3-2)$$

$$Z = \sin\varphi \quad (3-3)$$

よって、単位球面上のある点 $(X', Y', Z')$ に対する正距円筒画像の座標 $(\varphi, \theta)$ は次の式(3-4,5)で求めることができる。

$$\varphi = \arcsin(Z') \quad (3-4)$$

$$\theta = \arcsin\left(\frac{Y'}{\cos(\varphi)}\right) \quad (3-5)$$

次に、単位球の中に 1 辺の長さが 1 の立方体を定義する。立方体の中心点は原点とする。立方体の 6 面は透視投影画像の上下左右前後の 6 枚に対応しており、球面から立方体の各面への写像を考える。立方体の面上のある点 $(x_1, y_1, z_1)$ と原点 $(0,0,0)$ からなる直線と球面との交点 $(x_2, y_2, z_2)$ は以下の式(3-6,7,8)で求めることができる。

$$x_2 = \frac{x_1}{\sqrt{x_1^2 + y_1^2 + z_1^2}} \quad (3-6)$$

$$y_2 = \frac{y_1}{\sqrt{x_1^2 + y_1^2 + z_1^2}} \quad (3-7)$$

$$z_2 = \frac{z_1}{\sqrt{x_1^2 + y_1^2 + z_1^2}} \quad (3-8)$$

これらの式と、先ほどの正距円筒画像と単位球との関係式を合わせると、立方体の面上の点 $(x_1, y_1, z_1)$ に対応する正距円筒画像上の点 $(\varphi, \theta)$ を次の式(3-9,10)で求めることができる。

$$\varphi = \arcsin\left(\frac{z_1}{\sqrt{x_1^2 + y_1^2 + z_1^2}}\right) \quad (3-9)$$

$$\theta = \arcsin\left(\frac{y_1}{\frac{\sqrt{x_1^2 + y_1^2 + z_1^2}}{\cos(\varphi)}}\right) \quad (3-10)$$

透視投影画像の座標は2次元座標である。一方、式(3-9,3-10)は立方体の3次元座標に対する式である。ここで、立方体を図3.7に示すように単位円中に配置する。これは、高さを $z$ 軸でとり、 $x-y$ 平面を切り取った図である。黒色の円が単位球である。赤色の矩形が立方体を表す。円と $x$ 軸 $y$ 軸の交点は単位球であるため $(1, 0), (0, 1), (-1, 0), (0, -1)$ である。立方体と $x$ 軸 $y$ 軸の交点はそれぞれ $(0.5, 0), (0, 0.5), (-0.5, 0), (0, -0.5)$ である。ここで、 $x = 0.5$ の右端の青色の立方体の面を考える。 $y$ 座標は $-0.5 \leq y \leq 0.5$ の値をとる。 $z$ 座標も立方体であるため同様に $-0.5 \leq z \leq 0.5$ の値をとる。 $x$ 座標は常に $0.5$ であるため、青色の面は $y$ と $z$ 座標からなる2次元の平面と捉えることができる。

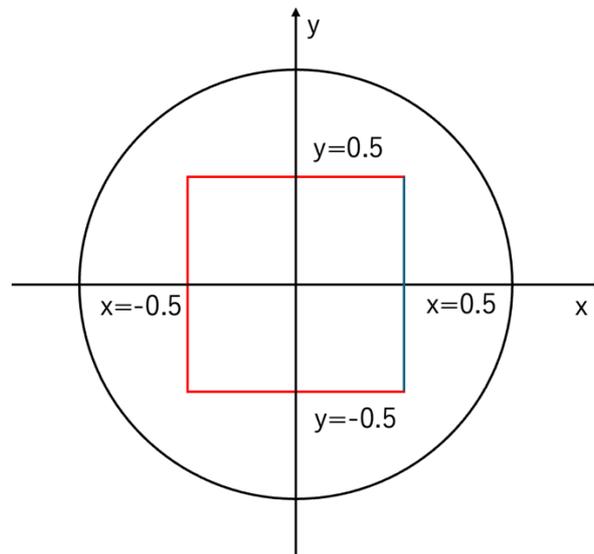


図 3.7 単位円と立方体の関係図

このようにすることで、立方体の各面の3次元の座標軸のうち、1つの座標軸が常に固

定され、立方体の各面を 2 次元座標として定義することができる。以下の表 3.2 は、6 面それぞれの透視投影画像とそれに対応する立方体の座標の関係を示した表である。それぞれ 1 つの座標軸が常に 0.5 か -0.5 に固定される。結果、例えば上面におけるすべての座標の  $z$  軸は 0.5 で固定されているため、立方体の面は 2 次元と捉えることができ、2 次元座標である透視投影画像と 3 次元座標である立方体の各面と対応付けることができる。

表 3.2 透視投影画像とそれに対応する立方体の各面の座標

透視投影画像	立方体の各面の座標ととりうる値
上面	$-0.5 \leq x_1 \leq 0.5, -0.5 \leq y_1 \leq 0.5, z = 0.5$
下面	$-0.5 \leq x_1 \leq 0.5, -0.5 \leq y_1 \leq 0.5, z = -0.5$
正面	$x = 0.5, -0.5 \leq y_1 \leq 0.5, -0.5 \leq z_1 \leq 0.5$
後面	$x = -0.5, -0.5 \leq y_1 \leq 0.5, -0.5 \leq z_1 \leq 0.5$
右面	$-0.5 \leq x_1 \leq 0.5, y = 0.5, -0.5 \leq z_1 \leq 0.5$
左面	$-0.5 \leq x_1 \leq 0.5, y = -0.5, -0.5 \leq z_1 \leq 0.5$

よって、透視投影画像に対応した立方体の各面の座標  $(x_1, y_1, z_1)$  の画素値を対応する座標  $(\varphi, \theta)$  がもつ画素値とすることで、1 枚の正距円筒画像から 6 枚の透視投影画像へ変換することができる。

## 3.2 画像の前処理

この節では、本研究で行った画像の前処理について説明する。NeRF の再構成結果は入力画像の質に依存している。例えば、画像のコントラストが低いと特徴点を抽出しにくい問題がある。また、輝度値が 255 に偏った明るすぎる画像や、0 に偏っている暗すぎる画像も特徴点を抽出できない。そこで、画像に前処理を加え、輝度値のヒストグラムの調整およびコントラストを高め、特徴点を抽出しやすくすることで、再構成結果の精度を上げることができると考えた。

### 3.2.1 Histogram Stretching

Histogram Stretching (HS) とは、元画像の画素値の最小値を 0 に最大値を 255 に拡張することで、画像のコントラストを上げる方法のことである。ヒストグラム拡張と呼ばれる。図 3.8 は Histogram Stretching 処理を行った例である。上の画像が元画像であり、下の画像が処理後の画像である。右のヒストグラムが画素値であり、元画像の画素値は左に偏っており、全体的に暗い画像となっている。一方、処理後の画像のヒストグラムは右側に拡張されており、元画像よりもコントラストが高い画像となっている。

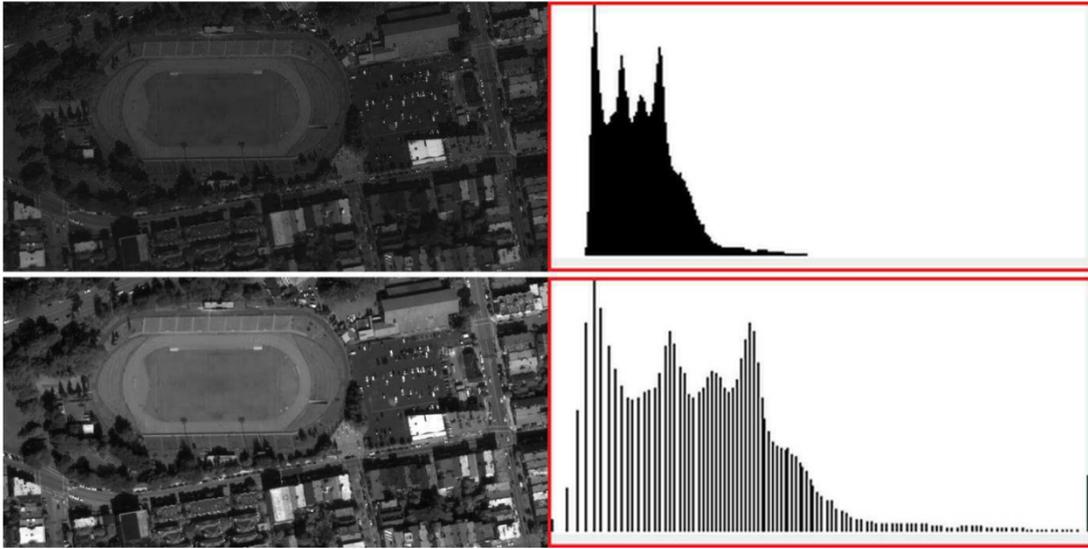


図 3.8 Histogram Stretching の一例, 上図が元画像,下図が処理後の画像 [15]

元画像のある座標の画素値 $v_{before}$ における, Histogram Stretching 処理後の画素値 $v_{after}$ は式(3-11)で求めることができる.

$$v_{after} = \left( \frac{v_{before} - min}{max - min} \right) \times 255 \quad (3-11)$$

ここで,  $max$ は元画像の画素値の最大値であり,  $min$ は元画像の画素値の最小値である. なお, 元画像の画素値に 0 と 255 が含まれている場合,  $v_{after}$ と $v_{before}$ は同じ値となり, 処理後の画像は元画像と同じになる.

本研究では RGB 画像を用いるため, Histogram Stretching での画素値の拡張は輝度値のみに行った. RGB 空間を HSV 空間に変換し, 輝度値である V にのみ前処理を行った. 図 3.9 は実際に研究室の画像に Histogram Stretching を行った例である. 左が元画像であり, 右が Histogram Stretching 処理後の画像である. 左画像に比べると, 右の画像が少しだけコントラストがはっきりしていることがわかる. 実験に用いた画像では, 研究室の照明と机の陰などにより, ほとんどの画像が 0 と 255 の値を持っているため, ヒストグラムに大きな差は発生せず, 画像セットに大きな変化は見受けられなかった.



図 3.9 左:元画像, 右: Histogram Stretching 処理後の画像

### 3.2.2 Histogram Equalization

Histogram Equalization(HE)とは、コントラストを強調するための処理である。コントラストが低い画像は、ヒストグラムに現れる画素値の範囲が狭い。また、一部の画素値の頻度が高い状態である。Histogram Equalization を行うことで、頻度が高い画素値の分布を頻度が低い画素値分布に再分配する。結果、全体的にヒストグラムの頻度を底上げすると同時に、突出した頻度の画素値の分布を減らすことができ、画像のコントラストを上げることができる。Histogram Equalization は X 線画像や衛星画像などの科学的な画像に応用されている。図 3.10 は Histogram Equalization の処理を示した図である。上が元画像とそのヒストグラムであり、下が Histogram Equalization 後の画像とヒストグラムである。ヒストグラム内のグラフは累積分布関数(CDF)を示している。元画像は画素値が 150 付近に集まっており、全体的にぼやけたコントラストが低い画像である。一方処理後の画像は、明暗がはっきりしており、コントラストが高いと言える。またヒストグラムにおいて、元画像では画素値が見られない 0~100, 200~256 の範囲にも画素が存在するようになり、CDF も右斜め上向きの直線となっているため、画素値の頻度の均等化が行われていることがわかる。

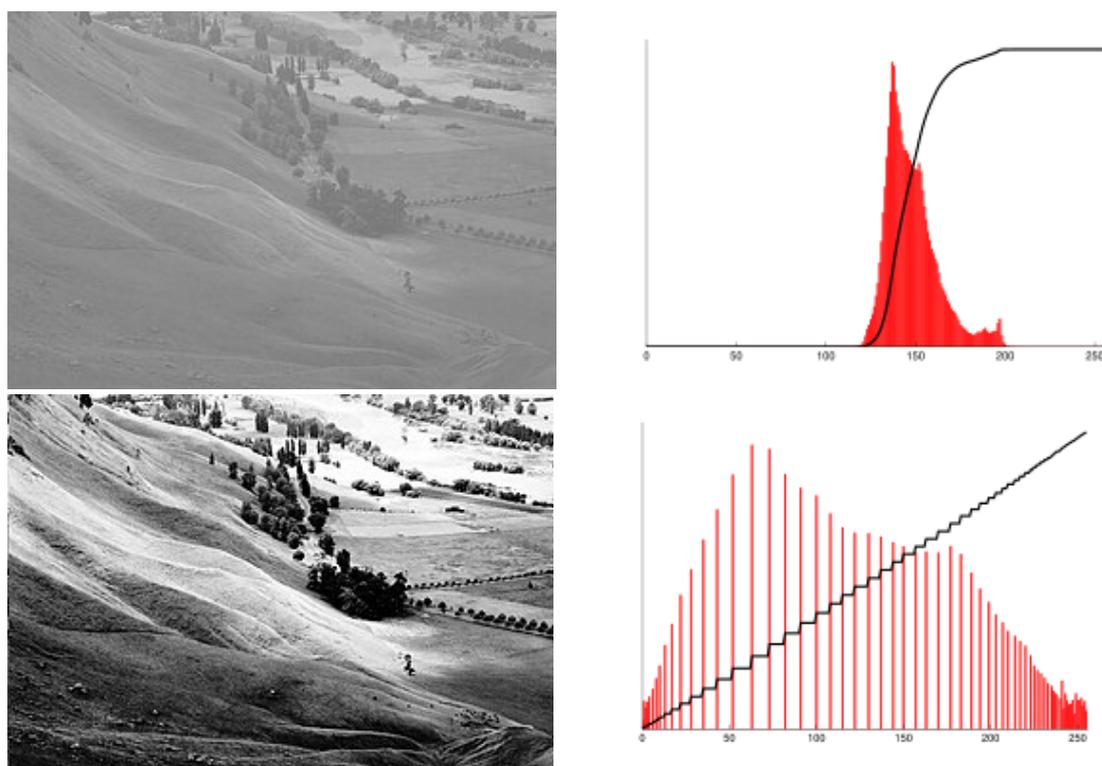


図 3.10 上: 元画像とそのヒストグラム, 下:HE 後の画像とそのヒストグラム[16]

本研究では, Histogram Equalization 処理を OpenCV にある `cv::equalizeHist` 関数を用いて行った. また Histogram Stretching と同様に輝度値にのみ処理を行った.

さらに本研究では, Contrast Limited Adaptive Histogram Equalization(CLAHE)も行った. CLAHE とは, コントラスト制限をかけた適応型 Histogram Equalization である. 前述の HE は画像全体を特に制限なく処理を行うため, 元から明るい場所がさらに明るくなってしまったり, 逆に極端に暗くなるという問題が生じる可能性がある. そこで CLAHE は, 画像を「タイル」と呼ぶ小さなブロックに分割し, そのブロックごとに均等化を行う. ブロック境界で発生する歪みはバイリニア補間により滑らかにしている.

結果, 例えば明るさが大きく違う空と建物との間でのヒストグラム均等化を行うことを防ぎ, 空は空で, 建物は建物で, その領域に適したヒストグラム均等化を行うことができる. またノイズが強調されることを防ぐためにコントラスト制限を設けた. 非常に高い頻度で出現する画素を閾値まで減らし, 閾値を超えた数は他の画素に均等に再分配することで, 突出した頻度の画素を減らす. 本実験では, OpenCV にある `cv::createCLAHE` 関数を用いた. また他と同様に輝度値にのみ処理を行う. 図 3.11 は Histogram Equalization と CLAHE 処理を行った画像である. 左上が元画像, 右上が Histogram Equalization 処理後の画像, 左下が CLAHE 処理後の画像である. どちらも元画像と比べて, 黒と白が強調されており, 天井の柄がはっきりと確認できるほどコントラストが高くなっていることがわかる. また, Histogram Equalization と CLAHE を比べると, CLAHE はホワイトボードの文字が

確認できるほど、コントラストが高くなっている。両処理共にコントラストが強調されている一方、自然画像とは異なる歪な画像となっている。

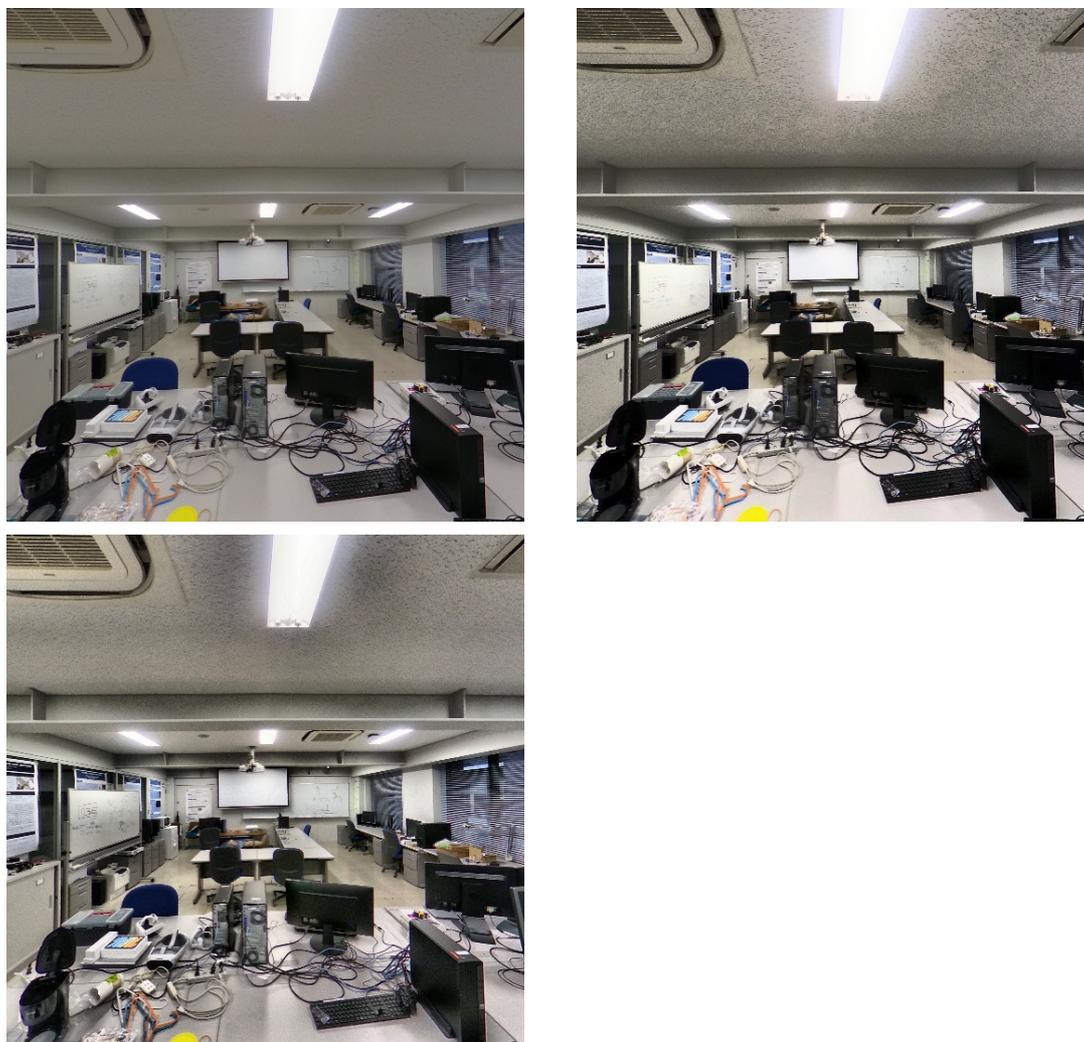


図 3.11 左上:元画像, 右上: HE を適用した画像, 左下, CLAHE を適用した画像

### 3.2.3 Histogram Matching

Histogram Matching とは、画像のヒストグラムをある特定のヒストグラムの形に近似させる、コントラスト調整方法の 1 つである。本研究では、D. Coltuc[17]らによる高精度なヒストグラムマッチングを行うことができる Exact Histogram Matching 手法を用いる。

Exact Histogram Matching は、入力画像に対して複数のカーネルで畳み込みを行い、各ピクセルごとに畳み込み後の値を複数定義する。それらの値の平均値をとり、その平均値を入力画像の新しいピクセル値とすることで、周囲の明暗を考慮したピクセル値を定義することができ、簡易的な連続値の画像を生成することができる。その入力画像と参照ヒストグラムでヒストグラムマッチングを行うことで、離散的なデジタル画像でも、結果を大きく歪ませることがなく、またヒストグラムの形を正確に近似することができる。

本研究では、参照ヒストグラムは全入力画像の平均ヒストグラムとした。画像セットの平均ヒストグラムを参照することで、時間経過にしたがって空の明るさが変化するなどの、ちょっとした照明条件の違いを均一にすることができる。平均ヒストグラムの計算方法は、入力画像セットの全画像における各画素値[0~255]の頻出回数を取り出し、その頻出回数を画像枚数で割ることで、入力画像セットの平均ヒストグラムを生成する。図 3.12 は左が元画像であり、右が EHM 処理を行った画像である。元画像と比べ大きな違いはないが、全体的に淡い感じの画像となっている。

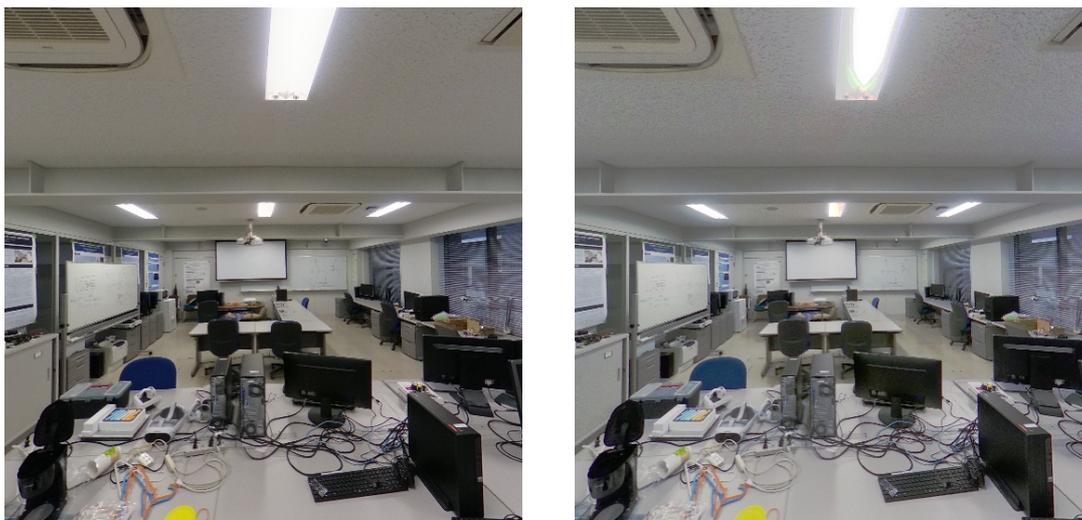


図 3.12 左:元画像, 右: EHM 処理後の画像

ヒストグラムマッチングでヒストグラムを合わせる参照ヒストグラムは画像セットの平均ヒストグラムを用いた。他方、自然風景画の輝度値のヒストグラムは 0 と 255 の近傍に画素値が集まっており、次に 128 付近が高いヒストグラムの画像となる傾向がある。0 と 255 近傍の値を除けば正規分布に近い形をとる。室内画像も同様で、画像ごとのヒストグラムは、一般に正規分布に従わないが、3次元再構成に入力する画像セットのように枚数が膨大であれば、その画像セットの平均ヒストグラムは正規分布に近い形をとる。そこで、EHM を用いて入力画像の輝度値のヒストグラムを正規分布の形に近似させることで、自然風景画に近い画像にする。この時、0 と 255 に近づくほど頻出回数が減っていくため、極端に暗い部分と明るい部分を無くすることができる。SfM において、明るい部分と暗い部分を無くしたことがどのような影響をとるのかという検証を行いたいため、画像の輝度値のヒストグラムを正規分布に近似させた画像セットでの実験も行った。図 3.13 は EHM を用いてヒストグラムを正規分布に従わせた結果の画像である。左上の画像が元画像であり、右上の画像が中央値 128、標準偏差 24 の正規分布に従った画像である。左下の画像が中央値 128、標準偏差 32 の正規分布に従った画像である。元画像と比べて、照明の白飛びや、黒潰れは減っているが、コントラストは極端に減っている。

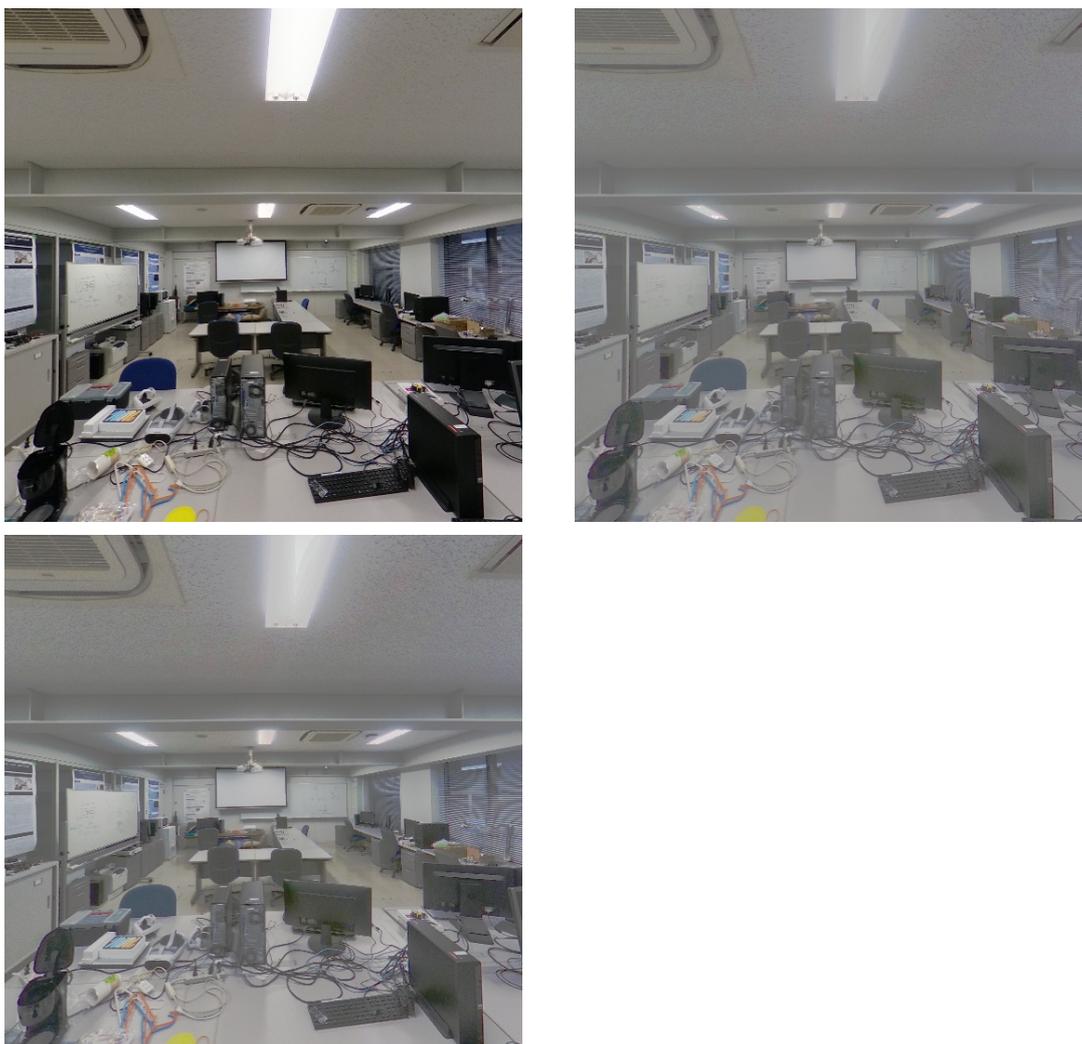


図 3.13 左上:元画像, 右上:標準偏差 24 の画像, 左下:標準偏差 32 の画像

### 3.2.4 黒色成分の調整

3次元再構成において、黒色が特に特徴点を取りづらく、再構成結果の品質に影響を与えていると考えた。そのため、黒色成分を減らすことを行えば、精度向上に寄与することができると考えた。そこで、画像の輝度値の範囲を $[0,255]$ から $[30,255]$ に正規化することで、元の画像のヒストグラムを保ちつつ、黒色成分を調整することができる。図 3.14 は実際に黒色成分を調整した結果である。自然な画像から大きく逸脱させることなく、黒色成分を調整することができる。元画像と比べると、モニターやデスクトップの黒色成分が減少しており、全体的に明るい画像となっている。



図 3.14 左:元画像, 右: 黒色成分減少後の画像

### 3.2.5 ハイダイナミックレンジ画像

ここでのダイナミックレンジとは、センサーデバイスが処理可能な信号の最大値と最小値の比率および範囲のことである。デジタルカメラにおけるダイナミックレンジとは、カメラが捉えることができる限界の明るさと暗さの範囲のことを示す。ダイナミックレンジを超えた強い明るさが複数ある場合、それらは全て同じ強度の光として識別され、写真には白とびしたように写る。同様にダイナミックレンジを超えた暗さの場合は、その部分は黒潰れしたように写る。ダイナミックレンジが広いほど、照明や太陽などの強い光から影や夜空などの弱い光を詳細に描画することができ、よりリアリティが高い写真を撮影することができる。ダイナミックレンジの値はカメラごとに異なっており、値が高いほど、より明暗の範囲が広い写真を撮影することができる。ダイナミックレンジはカメラ選びの指標となる重要な要素の中の1つである。

カメラのダイナミックレンジは年々進化しているものの、自然界の明暗を正確にとらえることは未だできていない。自然界は、太陽光( $10^9 \text{ cd/m}^2$ )という明るい光から夜空の星( $10^{-6} \text{ cd/m}^2$ )という暗い光が存在し、ダイナミックレンジが最大で $10^{15}$ という非常に広い世界である。人の眼球は瞳孔を固定した状態であれば $10^5$ 程度のダイナミックレンジをもつ。また、瞳孔を閉じたり開いたりすることにより最大で $10^{12}$ 程度のダイナミックレンジをもち、自然界の光のほとんどを捉えることができる。一方、カメラのダイナミックレンジは $10^5$ 程度である。カメラは図 3.15 に示すようにレンズの絞り値やシャッター速度を変更することで撮影可能な輝度レベルを調整し、光の強さに応じた最適な撮影を行うことができる。

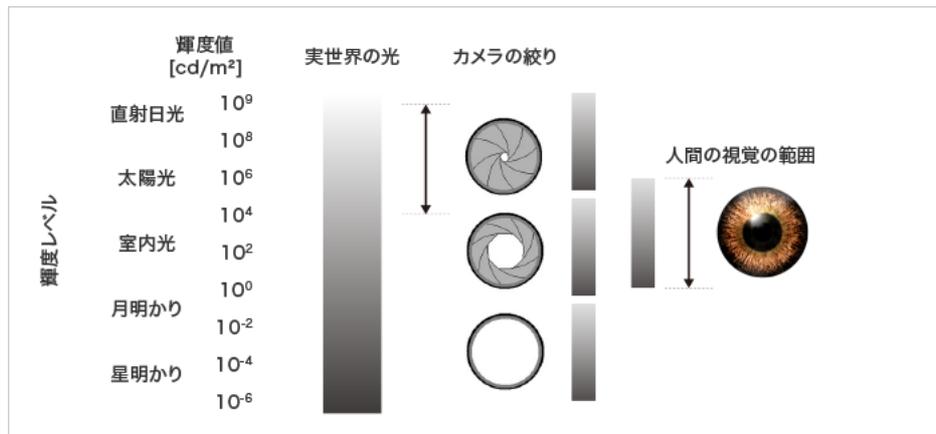


図 3.15 輝度値とカメラの絞りを示した図 [18]

カメラは通常、撮影した写真や動画をディスプレイで表示したり、画像・映像コンテンツで利用するため、Standard Dynamic Range(SDR)規格に合わせ、 $10^3$ 程度のダイナミックレンジでの撮影を行なっている。これは、自然界のダイナミックレンジには遠く及ばず、また人の目のダイナミックレンジよりも狭いため、リアリティが少し欠けてしまう。

High Dynamic Range(HDR)とは、SDRよりも広いダイナミックレンジをもつ、より高い品質の映像体験を実現する技術である。同じシーンを異なる露光レベルで複数回撮影し、異なる明るさで撮影した複数の画像を結合し、カメラの内部もしくはソフトウェアでコントラスト比の調整を行うことで、HDR画像を生成することができる。HDR合成により、ダイナミックレンジを $10^5$ 程度に拡張することができ、人が知覚する自然界のシーンに近い映像を再現することができ、よりディテールが細かい写真を得ることができる。

通常のカメラでHDR画像生成を行うためには、複数枚の写真画像をそれぞれ別の露光レベルで撮影する必要がある。その一方、神納らによれば信号処理を用い1枚の画像から擬似的にHDR画像を生成することもできる[19]。本研究では、RICOHが提供しているHDR撮影機能を使用することが現実的であると考え、RICOH THETAのHDR画像撮影機能を用いて、HDR画像を取得した。RICOH THETAのHDR撮影機能は色を16bitの階調で表現するものの、JPEG等に圧縮するためトーンマッピングを行い、8bitの色の階調で画像を表現している。図3.16は実際に撮影したHDR画像である。左のSDR画像に比べ照明の部分や影がはっきりと写っていることがわかる。



図 3.16 左:元画像, 右: HDR 画像

### 3.2.6 撮影時の視線方向の工夫

3次元再構成において, SfM による画像の位置合わせの精度が高いほど, より正確な再構成結果を得ることができる. SfM の精度を高めるためには, 位置合わせに使われる特徴点の量を増やす必要がある. そこで輝度値のヒストグラムを最適化することで, 抽出する特徴点の量が増えると考え, 信号処理レベルの前処理による精度向上を目指した.

別のアプローチとして, 同一の特徴点が複数の画像に映りやすくなるように, 写真の視線方向の変更を行った. フォトグラメトリにおいて, 出力結果を高品質にするために視線方向を工夫することは一般的であり, 撮影方法のコツについての多数の記事を Web で拝見することができる. NeRF でも視線方向の工夫により精度を高められると考え, 前処理とは別に追加で検証を行った. 本実験で行った視線方向の工夫は, 被写体に対するカメラの撮影アングルの調整である. ここまでの写真撮影では, 壁に向かって垂直に撮影した写真画像を用いていた. そのため, 壁面のみが写っている画像が多数あり, それらから抽出できる特徴点が少なくなっていた. そこで壁に対して垂直に撮影するのではなく, カメラを横に 45 度程度回転させた状態で撮影を行い, 壁のみ写っている画像を無くし, 抽出できる特徴点の数を増やすことを考えた. 図 3.17 は実際にレンズの向きの角度を 45 度回転させ, 撮影によって得た画像である. 左の壁と照明のみが映っている元画像に比べて, 右の画像はデスクトップ PC や椅子, 机といった物体が多く映っており, 特徴点をより多く抽出できると期待した.

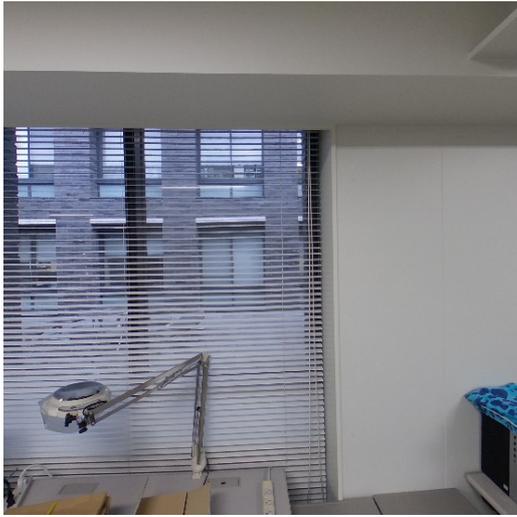


図 3.17 左:元画像, 右:アングル調整後の画像

## 第 4 章 実験結果

本章では, 3 章で示した前処理などを用いた NeRF の精度向上実験の結果を示す. 使用する NeRF モデルとして, NeRFStudio の NeRFacto を用いる. 精度向上を確認するための評価基準として Colmap による SfM の結果と, 9 枚の入力画像について NeRF が生成する画像との PSNR の平均値で評価を行った.

### 4.1 Colmap による SfM の結果

NeRF は入力された画像と, その画像を撮影したカメラの位置・姿勢情報から 3 次元再構成を行う. その際, カメラの位置・姿勢情報が正しくない場合, 再構成結果の品質が低下してしまう. そのため, NeRF において, SfM による画像の位置合わせの精度は非常に重要である. また, 入力画像の枚数も再構成結果に大きく影響する. 一般的には, 入力画像が多いほど, よりディテールが細かな 3 次元モデルを生成することができる. ほとんどの NeRF モデルはカメラの位置・姿勢情報を, オープンソースの SfM ライブラリである Colmap で計算している. Colmap は, 入力画像を撮影したカメラの位置・姿勢情報を特徴点マッチングから計算している. Colmap では, 特徴点が少なくカメラの位置・姿勢情報を計算することができない画像を入力画像セットから削除するため, そのような場合, NeRF への入力画像の枚数が減ってしまう. そのため, Colmap での位置合わせに成功した画像の枚数およびその精度は NeRF の再構成結果の品質に影響を及ぼす. そこで本実験では, Colmap による SfM の結果を NeRF の再構成結果の精度向上の 1 つの指標として評価した.

初めに, 入力画像の解像度の差に関する実験結果を示す. 表 4.1 は入力画像の解像度とカメラの位置・姿勢情報を計算することができた画像枚数との関係を示した表である. 入力画像枚数は共に 96 枚の SDR 画像である. 表 4.1 は, THETA V と Z1 で撮影した画像の比較である. Z1 は V と比べて解像度は 1.25 倍高くなっている. 結果は, V が 86 枚であり, Z1 が 90 枚であった. 解像度が高い方が使用された画像枚数が多いことがわかる. よって, 画像そのものの前提として, 解像度が高いほど, 特徴点を抽出しやすくなり, SfM の精度が向上すると考えることができる.

表 4.1 入力画像の解像度とカメラの位置・姿勢情報を取得できた画像枚数の関係  
(全 96 枚中)

解像度・撮影デバイス	画像枚数
THETA V 解像度: 5376 × 2688	86
THETA Z1 解像度: 6720 × 3360	90

表 4.1 の結果を踏まえ、本実験では、THETA Z1 で撮影した SDR 画像に対して前処理を行った。表 4.2 は前処理の種類と Colmap でカメラの位置・姿勢情報を取得できた画像枚数の関係を示した表である。

表 4.2 前処理の種類とカメラの位置・姿勢情報を取得できた画像枚数の関係  
(全 96 枚中)

前処理の種類	画像枚数
SDR	90
HS	91
HE	86
CLAHE	86
EHM	86
ガウス分布へ近似 標準偏差 24	88
ガウス分布へ近似 標準偏差 32	88
黒色成分の調整	89
ハイダイナミックレンジ	90
視線方向の工夫	93

表 4.2 の結果から、位置・姿勢情報を取得できた画像の枚数は視線方向の工夫が最も多いことがわかった。また HS は処理前の画像である SDR 画像と比べると 1 枚のみ画像枚数が増えている。HE, CLAHE, EHM, ガウス分布, 黒色成分の調整は、処理前の画像である SDR 画像と比べて画像枚数が減ってしまった。

Colmap で計算したカメラの位置・姿勢情報が常に正しいわけではない。その画像に写っている特徴に似ている特徴が他にもあれば、そちらの方に合わせてしまうことがある。そのため、位置・姿勢情報が取得できた画像の枚数には、間違った位置・姿勢情報を持つ画像が含まれている可能性がある。表 4.3 は表 4.2 で示した画像枚数のうち、明らかなハズレ値を除いた画像枚数である。これらの画像は多少の誤差はあるものの、位置合わせに成功した画像の枚数である。

表 4.3 ハズレ値を除いた画像枚数全 96 枚中

前処理の種類	画像枚数
SDR	87
HS	88
HE	83
CLAHE	83
EHM	83
ガウス分布へ近似 標準偏差 24	86
ガウス分布へ近似 標準偏差 32	85
黒色成分の調整	86
ハイダイナミックレンジ	87
視線方向の工夫	92

表 4.3 から、まず視線方向の工夫が、最も多い 92 枚という結果であった。視線方向の工夫により、位置合わせが成功している画像枚数を増やすことができる。そのため、NeRF においても、従来のフォトグラメトリと同様に、撮影時にカメラの視線方向を適切に設定することが必要であることがわかった。次に多いのは HS の 88 枚である。これは、元画像である SDR 画像よりもコントラストがはっきりしたため、SDR 画像よりも多い結果となったと考えた。HE、CLAHE、EHM は自然な状態で撮影した画像と比べると画像のコントラストがいびつであることが SfM に悪い影響を及ぼしたと考えた。ガウス分布への変換は、標準偏差の値に関わらず位置・姿勢情報を計算できた画像の枚数は 88 枚であったが、表 4.3 によると、標準偏差 24 と 32 で、標準偏差 24 の場合に正しい画像が 1 枚多いという結果になった。

次に SfM の評価指標の一つとして用いられる再投影誤差(reprojection error)で評価を行った。再投影誤差とは、画像面に再投影された点と写真撮影された点の間の、画像面上での画素単位の距離を表す幾何学的誤差のことである。再構成した 3D ポイントとそれに対応する画像上の特徴点との距離を定量化するために使用される。この距離は SfM における誤差であり、誤差が小さいほど、推定されたカメラの位置・姿勢と推定された特徴点の位置が正しいことを表している。誤差は小さければ小さいほどよく、一般的には 1 画素以下であれば良い結果であるとされる。表 4.4 は前処理の種類とそれぞれの SfM の結果の平均再投影誤差を示した表である。

表 4.4 前処理と平均再投影誤差

前処理の種類	平均再投影誤差 (画素)
SDR	1.00299
HS	0.99976
HE	1.02202
CLAHE	0.99298
EHM	1.01834
ガウス分布へ近似 標準偏差 24	1.00578
ガウス分布へ近似 標準偏差 32	1.01147
黒色成分の調整	0.99638
ハイダイナミックレンジ	0.96008
視線方向の工夫	0.85033

表 4.4 から、良い結果とされる 1 画素以下の誤差を持つのは、HS, CLAHE, 黒色成分の調整, HDR, 視線方向の工夫であった。中でも最も誤差が小さいのは視線方向の工夫であり、次に小さいのは HDR 画像であった。元画像である SDR 画像の平均再投影誤差よりも小さい値をとるのは、HS, CLAHE, 黒色成分の調整, ハイダイナミックレンジ, 視線方向の工夫であった。

画像枚数と再投影誤差の結果をまとめると、共に最も良い結果であったのは視線方向の工夫であり、次にハイダイナミックレンジと HS 処理となった。CLAHE と黒色成分の調整は、画像枚数は減ってしまったものの、再投影誤差の精度は向上している。SfM において、視線方向の工夫, ハイダイナミックレンジ化, HS 処理を行うことで、精度を向上させることができることを確認できた。

## 4.2 再構成結果の評価

4.1 節の Colmap の結果を用いて、NeRF でそれぞれの画像から 3 次元再構成を行った。再構成結果の評価は、入力画像と、NeRF で再構成した 3 次元空間をそのカメラの位置・姿勢に投影した画像との PSNR で評価した。PSNR は 2 枚の画像の近さ度合いを示す尺度である。画像圧縮における元画像との劣化具合を検証するためなどに用いられる。PSNR の値が 30 程度あれば、その 2 つの画像はかなり近い。画像伝送などで許容できる品質は 20~25 程度とされている。一般的に、PSNR が 25~30 程度あれば、劣化はそれほど問題にならない。

NeRF での PSNR は、NeRF で生成した 3 次元シーンを投影した画像と、その投影位置での入力画像がどのくらいのレベルで同じなのかを示す。NeRF での PSNR が 30 あれば、元画像とそれほど変わらない 3 次元シーンであることを示し、PSNR が 20 以下であれば、元

画像と比べて明らかに3次元シーンが劣化していることを示す。表 4.5 は前処理の種類と PSNR での評価値を示した表である。

表 4.5 PSNR での評価値

前処理の種類	PSNR
SDR	20.0942
HS	20.1335
HE	16.2721
CLAHE	17.8287
EHM	19.5399
ガウス分布へ近似 標準偏差 24	25.6381
ガウス分布へ近似 標準偏差 32	22.4027
黒色成分の調整	21.6221
ハイダイナミックレンジ	20.9791
視線方向の工夫	21.4487

表 4.5 より、元画像である SDR 画像と比べて PSNR の値が高いのは HS, ガウス分布へ近似 標準偏差 24, ガウス分布へ近似 標準偏差 32, 黒色成分の調整, ハイダイナミックレンジ, 視線方向の工夫である。

PSNR の値が最も高かったガウス分布への近似に関して、ガウス分布への近似の画像は明暗がなく、ぼやけた画像である。また画素値が中央値に集まっているという傾向があるため、PSNR の値も高くでやすい。そのため、NeRF においても、PSNR の値が高くなり、明暗が少ないほうが、入力画像の再現度が高くなるということが確認できた。図 4.1 は標準偏差 24 のガウス分布へ近似した画像セットの NeRF 結果と元画像を示した 1 例である。右が元画像であり、左が NeRF の結果である。大まかな形や特徴は捉えているものの、机はぼやけている。色合いがほぼ一様であるため、PSNR の結果が高くなったと考える。同様に、図 4.2 は標準偏差 32 のガウス分布へ近似した画像セットの NeRF 結果と元画像を示した 1 例である。元画像は標準偏差 24 と比べて、自然画像に近い画像である。NeRF 結果は、天井の照明部分が崩れており、また机の再現度も低い。標準偏差 24 と比べ、色の再現度が低くなったため、PSNR の値も 3.2 ほど低い値をとったと考える。

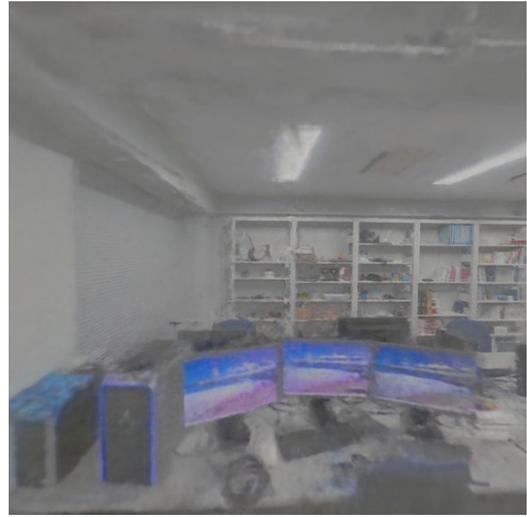
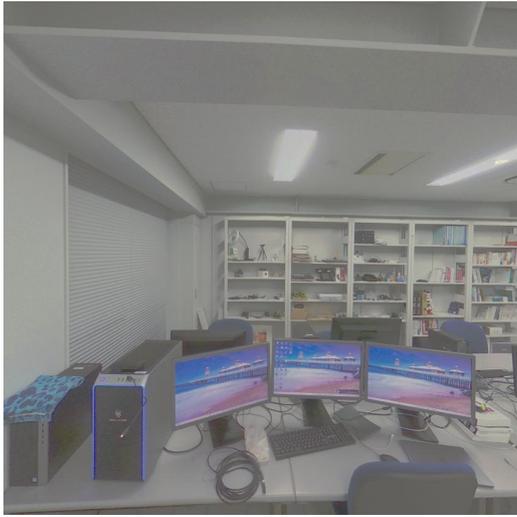


図 4.1 ガウス分布への近似 標準偏差 24 の結果 左図,元画像 右図,NeRF の結果

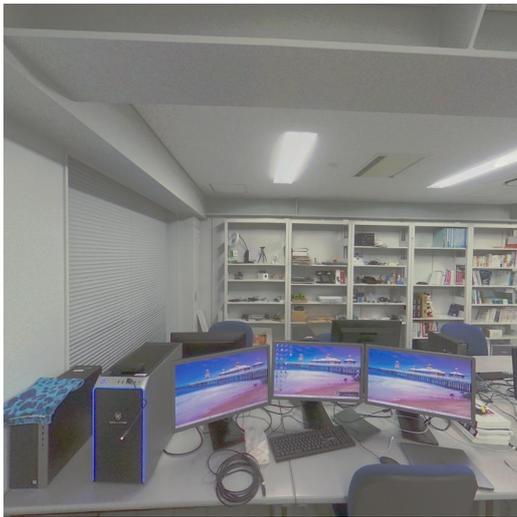


図 4.2 ガウス分布への近似 標準偏差 32 の結果 左図,元画像 右図,NeRF の結果

次に黒色成分の調整が 21.6221 と高い. 図 4.3 は黒色成分を調整した画像を用いた NeRF の結果の一例である. 全体的に特徴を捉える事はできており, デスクトップなどの黒色のぼやけが少なくなっていることがわかる. ガウス分布と同様に, ヒストグラムの取りうる値の範囲が減少しているため, PSNR 評価値の計算では高くでてしまうという点があるものの, 入力画像の再現度は元画像と比較しても高くなった.

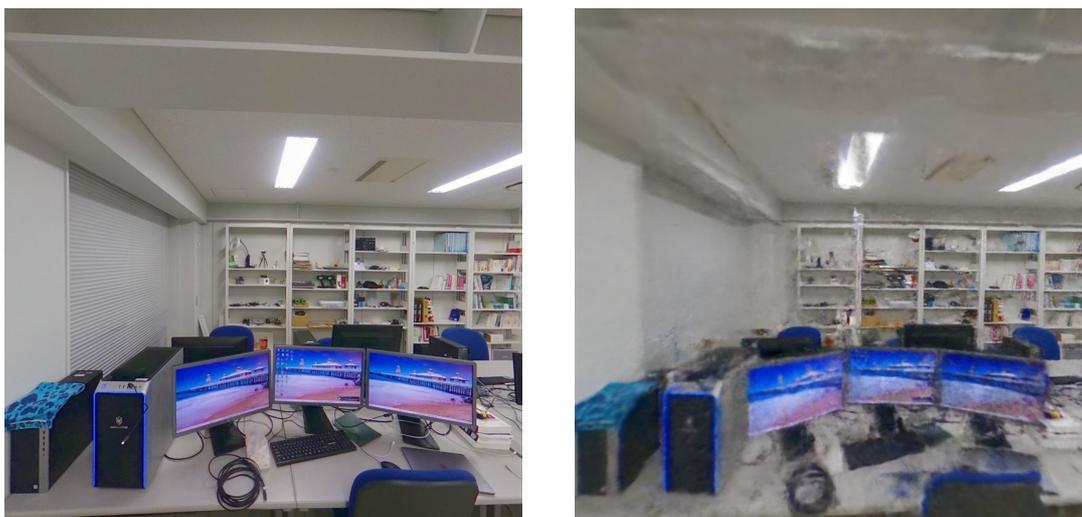


図 4.3 黒色成分を調整した結果 左図,元画像 右図,NeRF の結果

次に、視線方向の工夫が PSNR の値が 21.4487 と高い。図 4.4 は視線方向を工夫した画像セットの NeRF の結果の一例である。他と同様に左図が元画像であり、右図が NeRF の結果である。NeRF の結果を確認すると、天井や壁面はきれいな結果を示している。一方、棚や机はぼやけており、ノイズのようなものが載ってしまっているものの、全体的な再現度は高い。



図 4.4 視線方向の工夫による結果 左図,元画像 右図,NeRF の結果

HDR と HS は SDR に比べて PSNR の値が 1 未満の改善ではあるが、結果の向上を確認することができた。図 4.5 は HDR の結果である。全体的に上手く再現はできているものの、机がぼやけてしまった。HDR 画像は影や照明部分がはっきりと映るため、特徴が捉えやすくなり、精度が向上したと考える。図 4.6 は HS の結果である。全体的に特徴を捉えており、大きな相違点は見当たらない。一方、机の再現度が視線方向の工夫や HDR

に比べ、劣っている。特に影やデバイスの黒色が再現度の劣化に影響を与えていると考える。



図 4.5 HDR の結果 左図,元画像 右図,NeRF の結果

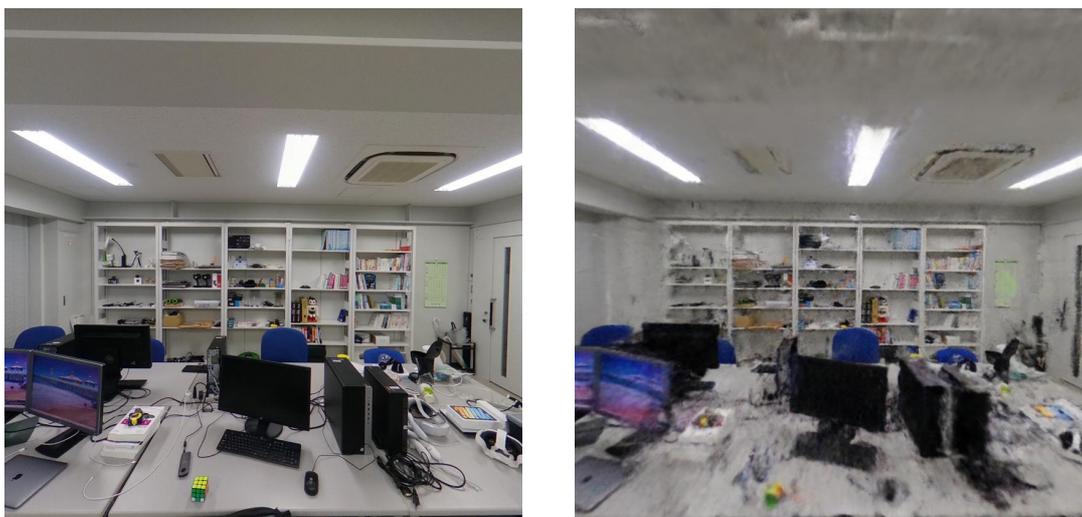


図 4.6 HS 処理の結果 左図,元画像 右図,NeRF の結果

HE や CLAHE, EHM の PSNR の値は元画像である SDR の PSNR の値よりも低くなつてしまい、今回の実験では、NeRF の精度向上に貢献することができなかった。図 4.7 は HE の結果である。左の元画像は白黒の差が激しくなっており、明暗が際立っている。右の NeRF の結果では、全体的に黒色でぼやけている。机や天井の黒色が多い箇所の再現度が著しく低い。図 4.8 は CLAHE の結果である。HE の元画像と CLAHE の元画像を比べると、机のデバイスの影や天井の黒色が薄まっている。NeRF 結果では、HE と同様に全体的に黒

色でぼやけており、机や天井の再現度が低い。HE と比較すると照明部分に黒い霧がかかっておらず、また全体的に黒色が少なく、明るくなっている。図 4.9 は EHM の結果である。この NeRF の結果は、全体的に黒色の霧はないものの、天井部分が大きく崩れており、また机の再現度も低い。HE や CLAHE や EHM も同様に黒色が多い場所の再現度が低くなっていることがわかる。



図 4.7 HE の結果 左図,元画像 右図,NeRF の結果

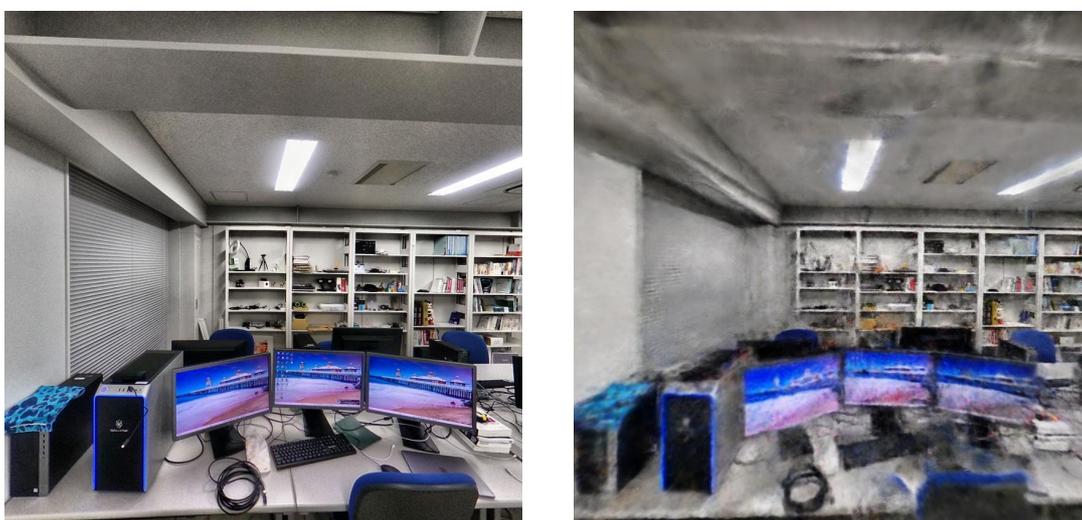


図 4.8 CLAHE 処理の結果 左図,元画像 右図,NeRF の結果



図 4.9 EHM 処理の結果 左図,元画像 右図,NeRF の結果

表 4.6 は解像度と NeRF の PSNR の結果を示した表である。解像度が低いほうが PSNR の値が高くなっている。PSNR は入力画像と 3 次元シーンの同一具合を示している値であるため、解像度が低い方が、詳細な描画が少なく PSNR の値が高くなってしまいう傾向にある。

表 4.6 解像度と PSNR の値

撮影デバイス, 解像度	PSNR
THETA V 解像度 : 5376 × 2688	21.9994
THETA Z1 解像度 : 6720 × 3360	20.0952

以上の結果より, NeRF の品質は Colmap の結果に依存する一方, 明暗差が少ない方が PSNR の値が高くなる傾向があることを確認した。また, 白色の再現度よりも黒色の再現度が低くなることも確認した。

### 4.3 結果の総評

Colmap での結果と NeRF での結果から, 視線方向の工夫が効果的であることがわかった。これは NeRF においても, 従来のフォトグラメトリと同様に, 高品質な結果を求めるには, 初歩的な撮影ノウハウの工夫や蓄積が必要であることがわかる。また, 通常の SDR 画像よりも HDR 画像や HS 処理, 黒色成分の調整を行うことで, NeRF の精度を向上させることができることを確認した。HE や CLAHE, EHM は輝度値が大きく変化することで自然画像との乖離が起きてしまい, その点で NeRF との相性が悪いため, 結果が振るわなかった。ガウス分布への近似では, Colmap の結果は良くなかったものの, PSNR の計算上では

NeRF の精度は高かった。すべての前処理において、黒色が多かった机の再構成結果の精度が悪かった。これにより、NeRF では明暗が少ない方が再現度は高くなり、また黒色の再構成結果の精度が劣化してしまう傾向にあることがわかった。

以上から、通常撮影で得られる SDR 画像に対して、HS を行う、黒色成分の調整を行う、HDR 画像撮影を行う、視線方向の工夫を行うことで、NeRF の精度を向上させることができることを確認した。

#### 4.4 今後の課題

入力画像のコントラスト調整のために、RGB 画像を HSV 画像に変換し、輝度値である V に対してヒストグラムの前処理を行った。そのため自然画像とは離れた画像となってしまいう前処理が存在し、それらの処理を施すと、精度向上はおろか、下がってしまった。今後の課題としては、RGB 画像に対してヒストグラムの前処理を行うことが考えられる。

今回の実験は、研究室を撮影した画像セットで行なったため、被写体や照明条件の変化により、結果が変わる可能性がある。被写体や照明条件の影響を検証し、さらに、研究室のような空間だけでなく、物体（オブジェクト）に対しても検証を行っていく必要がある。また本実験では、NeRF の評価値として広く使われている PSNR を用いた。しかし PSNR の計算では、ヒストグラムの範囲を狭めるほうが評価値が高くなる性質があり、人の主観的な評価と一致しない可能性がある。SSIM 等の他の評価値も利用して、より正確な評価を行うことが有効である。さらに入力画像セットごとに最適な前処理を行うシステムの開発を行っていく。

## 第 5 章 NeRF 空間とのインタラクション

本章では, NeRF の再構成結果を HMD に映すことで, コンピュータ上に再現した 3 次元空間を移動するなどのインタラクションが可能なシステムの開発について説明する.

### 5.1 システム開発の背景

NeRF は新たな 3 次元再構成技術として注目が集まっており, 2025 年現在も研究が盛んである. 一方, NeRF は 3 次元再構成に NN を使用しており, 出力結果は特殊なパラメータで構成されているため, その結果を利用することが難しく, 応用先が少ない. そこで本研究では, NeRF の実用性を示すため, ゲームエンジンである Unreal Engine 5 と HMD を使い, VR で NeRF の結果を体験できるシステムの開発を行なった.

### 5.2 実装

システムの実装のための開発環境は Unreal Engine 5 を使用した. また VR でインタラクションを行うための HMD は, Meta 社の Meta Quest 3[図 5.1]を用いた.



図 5.1 Meta Quest 3

Unreal Engine 5 は, NeRF の再構成結果そのままでは, エンジン内に描画することができない. Luma AI は NeRF の再構成結果を Unreal Engine 5 上で描画することができるプラグインを提供している. 今回のシステム開発ではこのプラグインを使用し, Unreal Engine 5 上に NeRF の結果を描画した.

描画した NeRF の再構成結果には物理判定が存在しない. そこで, Unreal Engine 5 にデフォルトで用意されている物理演算が可能な球や矩形, 平面を利用する. 描画した NeRF の再構成結果内の床, 天井, 壁, 机, 椅子などの, 現実空間において物理判定が必要な物の形に合うように矩形等のサイズや形, 位置を調整し配置することで, 擬似的に物理判定を付与することができる.

NeRF の再構成結果内の見えているすべてのオブジェクトに物理判定を付与し, VR ゲーム上でその判定に物理演算を行うことで, 違和感なく NeRF の再構成結果を VR 上で体験

することができた。図 5.2 は実装したシステムのゲーム場面の一例である。水色の丸はポインタであり、その位置に移動することができる。また机などの見えているオブジェクトすべてに物理判定を付与しているため、壁の奥や天井の上、床の下など現実的に移動することができない場所には移動することはできず、また机の上にものを置くことができ、壁や床にボールを投げると跳ね返る。



図 5.2 実装したゲームのシーン。ポインタ先に移動することができる。

### 5.3 今後の課題

NeRF の再構成結果を HMD で VR 体験できるシステムの開発を行い、NeRF の実用性を示すことができた。ゲームエンジンでの実装であるため、応用しやすく、多くの分野で活用できると考える。本システムは、NeRF の再構成結果そのものに物理判定を付与したのではなく、内部に同じ形ではあるものの、別の透明のオブジェクトを用意し、そのオブジェクトに物理判定を付与することで、擬似的に物理判定を付与した。そのため、NeRF の再構成結果と物理判定用のオブジェクトのテクスチャが対応していないため、例えば、椅子や机の上の物を動かすことができない。今後の課題として、NeRF の再構成結果そのものに物理判定を付与できるようにすることが考えられる。

## 第6章 結論

本研究では、画像の前処理による NeRF の再構成結果の精度向上の実験を行った。はじめに、実験データの用意として、360度を一度に撮影可能な RICOH THETA を使用して、研究室全体を映した正距円筒画像から透視投影画像へ変換を行なった。その透視投影画像に対してヒストグラムに関する前処理を7種類行った。さらに撮影時に HDR での撮影、視線方向の変更を行った画像を加え、計9種類の前処理を行った。前処理を行った画像で NeRF の再構成結果の評価を行った。前処理の元画像よりも精度が向上したと考えることができたのは、HS, HDR 画像, 黒色成分の調整, 視線方向の工夫であった。精度が向上した理由として、画像を大きく変化させる事なく、特徴点の抽出しやすさの向上させたためであると考える。今後の課題として、RGB 値に対してヒストグラムの前処理を行う。また、照明条件の影響の検証を行い、他の評価値での評価を行っていく。さらに入力画像セットごとに前処理を最適化することを行なっていくことが考えられる。

インタラクションを可能にするシステムの開発において、NeRF の再構成結果を HMD で VR 体験できるシステムの開発を行い、NeRF の実用性を示すことができた。今回の開発はゲームエンジンでの実装であるため、応用しやすく、多くの分野で活用できると考える。今後の課題として、NeRF の再構成結果と物理判定用のオブジェクトが対応していないため、NeRF の再構成結果そのものに物理判定を付与できるような開発を行うことが有効である。

## 謝辞

本研究の遂行にあたり、丁寧に指導して下さった蚊野浩教授に深く感謝いたします。

## 参考文献

- [1] G. Guidi\*, S. Gonizzi, L.L. , “ Image Pre-processing for optimizing automated photogrammetry performances”, ISPRS 2014
- [2] Krzysztof Skabek, Piotr Łabędź, Pawe Ozimek, “Improvement and unification of input images for photogrammetric reconstruction”, Computer Assisted Methods in Engineering and Science, Vol26 2019, p.153-162
- [3] Johannes L. Schonberger, Jan-Michael Frahm “ Structure-From-Motion Revisited ”, CVPR, 2016
- [4] Claire Zimmerman, “Review: Microsoft Photosynth; Stanford Humanities Lab”, jsah, 2010, p.463-466
- [5] Ross A. Beyer, Oleg Alexandrov, Scott McMichael, “The Ames Stereo Pipeline: NASA’S Open Source Software for Deriving and Processing Terrain Data”, Earth and Space Science, Vol 5, p.537-548
- [6] “RealtyCapture – 3D Models from Photos and/or Laser Scans”, [Online] Available: <https://www.capturingreality.com>, [Accessed 25 12 2024]
- [7] Chris Phillips “Google マップがより直感的に”, 8 2 2023, [Online] Available: [https://blog.google/intl/ja-jp/products/explore-get-answers/2023\\_02\\_maps-immersive view/](https://blog.google/intl/ja-jp/products/explore-get-answers/2023_02_maps-immersive-view/), [Accessed 25 12 2024]
- [8] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik et al. “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis”, ECCV 2020
- [9] Thomas Müller, Alex Evans, Christoph Shied, Alexander Keller, “Instant Neural Graphics Primitives with a Multiresolution Hash Encoding”, SIGGRAPH 2022
- [10] Jonathan T.Barron, Ben Mildenhall, Matthew Tancik et al, “Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields”, ICCV 2021
- [11] Christian Reiser, Richard Szeliski, Dor Verbin et al, “MERF: Memory-Efficient Radiance Fields for Real-time View Synthesis in Unbounded Scenes”, SIGGRAPH 2023
- [12] Matthew Tancik, Ethan Weber, Evonne Ng et al, “Nerfstudio: A Modular Framework for Neural Radiance Fields Development” SIGGRAPH 2023
- [13] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, George Drettakis, “3D Gaussian Splatting for Real-Time Radiance Field Rendering” SIGGRAPH 2023
- [14] 西島 啓人, “360 度画像を用いた Structure from Motion” 情報処理学会 第 84 回全国大会 2022
- [15] Miloud Chikr EL-Mezouarm, “Fusion d’images en télédétection satellitaire” , p.45
- [16] “Histogram equalization -Wikipedia”, [Online] Available: [https://en.wikipedia.org/wiki/Histogram\\_equalization](https://en.wikipedia.org/wiki/Histogram_equalization), [Accessed 25 12 2024]

- [17] Dinu Coltuc, Philippe Bolon, J-M Chassery, “Exact histogram specification”,  
IEEE Transactions on Image Processing, 15(5): p.1143-1152 2006:
- [18] “What’s HDR | HDR |映像制作機材|ソニー 製品情報 法人のお客様|ソニー”,  
[Online] Available:  
[https://www.sony.jp/products/Professional/c\\_c/hdr/index01.html?srsltid=AfmBOooOS2DooRKzb0Yo3Oyk4TwK7aZTnchHMOB6OCfk1Lmg\\_dCrMDV0](https://www.sony.jp/products/Professional/c_c/hdr/index01.html?srsltid=AfmBOooOS2DooRKzb0Yo3Oyk4TwK7aZTnchHMOB6OCfk1Lmg_dCrMDV0), [Accessed 25 12 2024]
- [19]神納 貴生, 奥田 正浩 “一枚の画像を用いた擬似高ダイナミックレンジ画像の生成”,  
映像情報メディア学会技術報告 32.54 巻(2008)

## 業績

### 国内学会発表

- (予定)  
西尾 漱一郎, 蚊野 浩, “Neural Radiance Fields(NeRF)の精度向上とインタラククション”, 情報処理学会 第 87 回全国大会 2025

## 付録

本研究で作成した主なプログラムのリストを示す。

プログラム名	プログラム内容
eq2cu.py	正距円筒画像を 6 枚の透視投影画像に変換する Python プログラム。
hs.py	入力画像に Histogram Stretching 処理を行う Python プログラム。
he.py	入力画像に Histogram Equalization 処理を行う Python プログラム。
clahe.py	入力画像に Contrast Limited Adaptive Histogram Equalization 処理を行う Python プログラム。
averagehist.py	入力画像セットの輝度の平均ヒストグラムを計算する Python プログラム。
matching.py	入力画像に Histogram Matching 処理を行う Python プログラム。
blackremove.py	入力画像の輝度値を特定の範囲に正規化する Python プログラム。