



# NeDDF を用いた単眼 Visual SLAM

高橋 響熙<sup>\*1</sup> 上田 樹<sup>\*2</sup> 宍戸 英彦<sup>\*3</sup> 北原 格<sup>\*3</sup>

Monocular Visual SLAM using Neural Density-Distance Field

Hibiki Takahashi<sup>\*1</sup>, Itsuki Ueda<sup>\*2</sup>, Hidehiko Shishido<sup>\*3</sup>, and Itaru Kitahara<sup>\*3</sup>

**Abstract** – This paper proposes a monocular Visual SLAM by Neural Density-Distance Field (NeDDF), which reconstructs 3D scene geometry using neural fields from a video captured by a monocular RGB-camera. We aim to realize a novel Visual SLAM utilizing the advantages of neural fields such as compactness, resolution, and compatibility. This paper outlines the 3D geometry representation with NeDDF and the camera poses estimation by re-projection error. A proposed method to perform sequential estimation of camera poses and NeDDF model is introduced. The results of preliminary experiments on learning NeDDF models in an indoor simulation environment is shown.

**Keywords** : Visual SLAM, Monocular Camera, NeDDF, Neural Fields, 3D Reconstruction

## 1 はじめに

本稿では、単眼 RGB カメラ映像から再構成したニューラル場三次元シーン表現 (Neural Density-Distance Field (NeDDF)) に基づく単眼 Visual SLAM (Simultaneous Localization and Mapping) について述べる。

ニューラル場による三次元形状表現は、点群 [3][5][7] やボクセル [1][4] と比較し、モデルがコンパクトかつ高分解能であり、また深層学習と親和性が高いことから、近年盛んに研究されている。ニューラル場の有用性を広く知らしめた Neural Radiance Field (NeRF) [11] は、髪などの高周波成分や非ランバート性反射を含むシーンにおいても写実的な画像を合成可能にした。初期の NeRF では、Structure from Motion (SfM) (COLMAP[5] 等) により推定したカメラ姿勢 (カメラ外部パラメータ) に基づいたモデル学習が一般的であったが、近年、未知もしくは真値に十分近い初期値からカメラ姿勢と NeRF モデルを同時に学習する手法に注目が集まっている [13][15]。これらの研究では、レンダリングにより得られる推定画素値と真値画像値の誤差 (Photometric Error) を損失関数とし、6DoF のパラメータで表現されたカメラ姿勢を学習する手法が提案されている。収束性に関する実験から、カメラ姿勢を含めた学習が可能なカメラ配置は、視線方向が平

行に近い画像群に限られることが報告されている [15]。収束性が悪い理由は、物体表面のみに勾配を持つ密度場に対しボリュームレンダリングを実行し算出される損失関数は、凸性を持たず局所解に陥りやすいためだと考えられる。後述する Visual SLAM において、大きなカメラ姿勢の変化にも追従して推定可能であること (追従性) は重要である。NeDDF[16] は、勾配を持つ領域が狭いという“密度場”の問題点に着目し、勾配を広く持てる“距離場”を同時に学習することで、カメラ姿勢の追従性向上を実現した。その結果、従来手法と比べ高い追従性を有しつつ、NeRF と同等の画像合成品質を持つ三次元形状表現が可能となった。

Visual SLAM は、入力映像から、未知環境の三次元形状とカメラ姿勢を逐次推定する手法であり、ロボティクスや自動走行車両、AR、MR などの分野で用いられている。Visual SLAM で再構築される三次元形状は、一般に点群やボクセルによって記述されるが、近年ではニューラル場の有用性から、Visual SLAM のマップ表現にニューラル場を適用する試みが行われている。iMAP[14] は、デプスマップを含めた RGB-D 映像を入力とし、NeRF モデルをリアルタイムに学習するが、RGB 映像入力から NeRF モデルをリアルタイム学習する手法は未だ例を見ない。アクティブステレオ法を用いれば三次元形状をリアルタイムで獲得することができるが、RGB カメラと比べ計測されるデプスマップの解像度が低く、視差推定精度はオクルージョンの影響を受けやすいため、ピクセルレベルでの整合性を想定している NeRF に対して、誤差を十分に抑えたデプスマップの獲得が困難である。我々は、NeRF の性能を十分に活用するためには、RGB 画像

<sup>\*1</sup>筑波大学知能機能システム学位プログラム

<sup>\*2</sup>筑波大学エンパワーメント情報学プログラム

<sup>\*3</sup>筑波大学計算科学研究センター

<sup>\*1</sup>Master Program in Intelligent and Mechanical Interaction Systems, University of Tsukuba

<sup>\*2</sup>Doctoral Program in Empowerment Informatics, University of Tsukuba

<sup>\*3</sup>Center for Computational Sciences, University of Tsukuba

のみで三次元形状を再構成することが望ましいと考えている。

そこで本研究では、ニューラル場による三次元シーン形状およびカメラ姿勢の逐次推定を目的とし、NeDDFを用いた単眼 Visual SLAM システムの構築を目指す。

## 2 関連研究

### 2.1 ニューラル場

ニューラル場は、ベクトル場を多層パーセプトロン (MLP) で近似する手法であり、三次元再構成の分野では三次元形状を記述する密度場や距離場の関数近似器として用いられる。三次元形状を記述するベクトル場は様々な形がとられる。例えば、自由視点映像生成の領域で顕著な成果を上げた NeRF[11] では、三次元位置と視線方向を入力とし、密度を出力する密度場と RGB 値を出力する色場を組み合わせた輝度場で三次元形状を表現している。DeepSDF[9] では、三次元位置と latent code を入力とし、物体表面までの符号付き距離を出力する符号付き距離場 (SDF) で表現している。

NeRF は、多量の RGB 画像およびカメラ姿勢に基づいた学習により、三次元形状を表現した輝度場を獲得する。与えられたカメラ姿勢に対し輝度場をボリュームレンダリングすることで、任意のカメラ姿勢 (視点) で観測される画像を生成する。モデルパラメータの学習では、Photometric Error を損失関数とする。実環境で撮影した画像群から NeRF モデルを学習する場合、COLMAP[5] などの SfM 技術を用いてカメラ姿勢を事前に推定する必要がある。しかし、SfM のような画像特徴点を手がかりとする手法はテクスチャ情報が乏しい環境に対して脆弱性を有する。

そのような課題に対し、カメラ姿勢の事前算出を必要とせず、カメラ姿勢と NeRF モデルを同時に学習する手法が提案されている [15]。iNeRF[12] は、NeRF の枠組みにおいて、誤差逆伝播法による最適化スキームがカメラ姿勢推定に適用可能であることを示した。本研究においても、誤差逆伝播法によるカメラ姿勢推定手法を用いる。BARF[13] は、NeDDF と同じく、勾配情報が追従性に与える影響について考察した研究であり、Positional Encoding の高周波成分が勾配に基づく最適化計算を不安定にさせることを指摘し、学習の経過に応じたローパスフィルタを Positional Encoding に適用する手法を提案した。NeRF-[15] は、視線方向が平行に近い画像群ではカメラ姿勢を含めた学習が可能である一方、被写体を 360 度から囲うようなカメラ配置では学習が困難であり、Photometric Error によるカメラ姿勢推定の問題点を実証的に示した。NeDDF[16] は、カメラ姿勢を大域解に収束させることが困難である問

題に対し、密度場の勾配を持つ領域が狭いことが原因であるとし、カメラ姿勢推定に勾配を広く持つ距離場を適用した。また、色場に対し、最近傍表面の色を物体外部に伝播させることで、Reprojection Error を計算可能にした。Reprojection Error は、画像とマップの対応関係から計算される幾何学的な誤差であるため、広域の追従が可能である。一方、Photometric Error は、画像平面に投影される像の形状および色の誤差であるため、細部のカメラ姿勢を推定できる。Reprojection Error および Photometric Error の両者とも用いたカメラ姿勢推定は、従来手法よりも高い収束性を持ち、Visual SLAM のトラッキング手法として有用であると考えられる。

### 2.2 Visual SLAM

Visual SLAM は、カメラで撮影された映像から、未知環境の三次元形状とカメラ姿勢を推定する技術である。バッチ処理によりこれを実現する SfM とは異なり、Visual SLAM は三次元形状とカメラ姿勢を逐次的かつリアルタイムに推定することが前提となる。本技術は、使用するセンサの種類によって大別され、単眼カメラによって撮影された RGB 画像のみでこれを行うものを単眼 Visual SLAM とよぶ。RGB 画像の他に IMU などの内界センサを用いた Visual-inertial SLAM[8] や、デプスマップを用いた RGB-D SLAM[7][14] がある。その中でも単眼 Visual SLAM は入力される情報が少ないため、他の手法と比較して難しいとされる [6]。

特徴点ベースの Visual SLAM は、キーフレーム間で抽出された特徴点の三次元位置を三角測量により求めるため、主に点群によるマップ表現が用いられる。しかし、点群は物体の幾何的な情報を抽出することが困難な上、非連続性から深層学習との親和性が低い。また、ボクセルは三次元空間を離散化するため、空間計算量と分解能はトレードオフの関係にある。一方、ニューラル場は三次元空間を離散化しないため、一定の空間計算量で連続的な三次元形状が表現可能である。さらに、MLP が持つ微分可能性から、勾配ベースのトラッキングが容易である。

このような利点から、近年では新しいマップ表現としてニューラル場の適用が試みられている。NeRF の枠組みを Visual SLAM に適用した iMAP[14] は、RGB-D 映像を入力とし、リアルタイムに NeRF モデルを学習する。本手法のトラッキング精度は ORB-SLAM2[7] などの従来手法に及ばない一方、オクルージョンにより観測できなかった欠損箇所を尤もらしく推定が可能である。本研究では、RGB 画像のみを入力とし、マップ表現にニューラル場を用いた Visual SLAM の構築を目指す。iMAP は Photometric Error の他に入力デプスマップと推定済みマップ間の幾何学

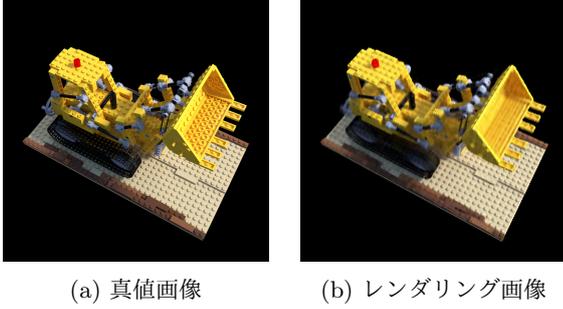


図 1: NeDDF による自由視点画像合成  
Fig.1 Novel view synthesis by NeDDF

的誤差 (Geometric Error) が計算可能であるのに対し, RGB 画像のみを入力とした本研究においては, Geometric Error を代替する別の追従性の高い損失関数を設定する必要がある。

### 3 Neural Density-Distance Field (NeDDF)

NeDDF は, 密度場に変換可能な距離場および輝度場から構成される。NeRF と同様に, 輝度場に対しボリュームレンダリングを実行することで, 表現力の高い自由視点画像が合成可能である。NeRF synthetic Dataset の lego シーンを使用し, 自由視点画像を合成した例を図 1 に示す。NeDDF における距離場の定義およびカメラ姿勢推定手法について以下で述べる。

#### 3.1 NeDDF における距離場の定義

距離場は, 入力される三次元位置に対して, 最近傍表面までの距離を返す場である。前述した通り, 距離場は物体外の広い領域で勾配を持つことが, カメラ姿勢の最適化において有用である。しかし, 距離場は明示的な境界面を前提にしているため, 密度場では表現可能であった半透明なガラスや境界面が曖昧な毛玉は再現が困難である。この問題に対し, NeDDF は境界面を仮定しない距離場を定義するため, 密度場に変換可能な形になるよう距離場の解釈を拡大する。境界面を仮定する符号なし距離場  $D$  の勾配は,  $\mathbf{p}$  を三次元位置とすると物体外部で  $|\partial D/\partial \mathbf{p}| = 1$ , 物体内部で  $|\partial D/\partial \mathbf{p}| = 0$  をとる。そこで, 毛玉表面のような密度値が連続的に変わる位置では  $0 < \partial D/\partial \mathbf{p} < 1$ , ガラスのような低密度な物体の内部では  $D = 1/\sigma_{low}$  ( $\sigma_{low}$  は物体内部の密度) で表現する。距離場  $D$  を, 三次元位置  $\mathbf{p}$  を入力とし, 全視線方向  $\mathbf{v} \in S^2$  での深度値の最小値を返す場と解釈することで, 前述の表現を実現する。すなわち, NeDDF における距離場の定義は, 式 (1) で表される。なお,  $d(\mathbf{p}, \mathbf{v})$  は,  $\mathbf{p}$  を光学中心とし  $\mathbf{v}$  方向にレイを飛ばしたときの深度値である。

$$D(\mathbf{p}) := \min_{\mathbf{v} \in S^2} d(\mathbf{p}, \mathbf{v}) \quad (1)$$

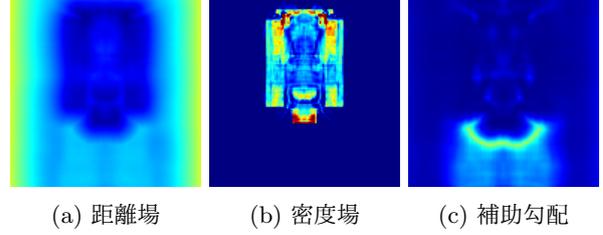


図 2: lego シーンにおける各場  
Fig.2 Fields in the lego scene

採用された視線方向  $\mathbf{v}_n$  は, 距離場の勾配を  $\nabla D(\mathbf{p})$  とすると, 次式で復元できる。

$$\mathbf{v}_n = \frac{-\nabla D(\mathbf{p})}{\|\nabla D(\mathbf{p})\|_2} \quad (2)$$

距離場が密度値の積分計算から求まることに着目し, 距離場に対し微分を実行することで次式が導出される。

$$\sigma(\mathbf{p}) = \frac{1 - \|\nabla D(\mathbf{p})\|_2}{D(\mathbf{p})} \quad (3)$$

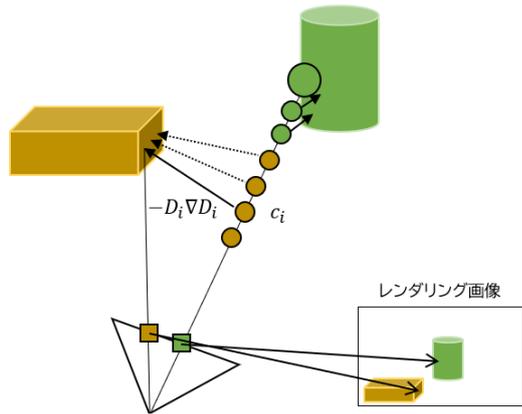
式 (3) を用いることで, 微分可能な変換式で距離場  $D(\mathbf{p})$  から密度場  $\sigma(\mathbf{p})$  を計算できる。

式 (3) による変換において, 本来は密度を持たない位置で誤った密度が発生する問題について述べる。具体例として, 二つの物体表面に挟まれたシーンを仮定する。両物体の中間点では距離場の勾配は逆方向を向き, 距離場が非連続的に変化する尖点が生じる。しかし, MLP で距離場を近似すると尖点は連続的に変化してしまうため, この尖点において式 (3) は誤った密度を持つ。そこで NeDDF では, 距離場に補助勾配  $\partial D/\partial w$  として次元を一つ追加し, 尖点付近では  $\sqrt{(\partial D/\partial t)^2 + (\partial D/\partial w)^2} = \sigma$  となるように勾配成分を分配し, 密度の誤発生を抑制する。

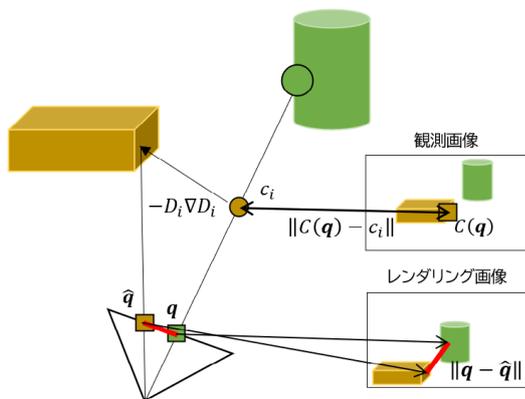
NeRF synthetic Dataset の lego シーンを学習した各場を図 2 に示す。図 2 は,  $z = 0$  でのスライスブルーソーザ上方から見た視点である。

#### 3.2 Reprojection Error によるカメラ姿勢推定

NeRF モデルでのカメラ姿勢推定には, Photometric Error を損失関数とし最適化を行う手法が用いられる [12]。色情報からカメラ姿勢を最適化を行う手法では, 物体表面の色が滑らかに変化している場合に限られる。そこで NeDDF は, レンダリング画像上での二次元点と空間上の三次元点での 2D-3D 対応点を求め, Reprojection Error を計算する手法を提案した。以下で, NeDDF モデルにおける 2D-3D 対応点探索および Reprojection Error の計算方法について概説する。まず, モデルパラメータを学習する際, 物体外部の色場に対して最近傍表面の色を回帰するようにペナルティを与え学習する。レンダリング画像上の二次元点  $\mathbf{q}$  を通るレイ上で物体外部の三次元点をサンプリングする



(a) 最近傍表面の色を物体外部に伝播



(b) 2D-3D 対応点探索, Reprojection Error の計算

図 3: 距離場を用いた Reprojection Error の計算  
 Fig. 3 Calculation of the reprojection error with the distance field

と, サンプル点から最近傍表面までの距離, 方向および最近傍表面の色を取得できる (図 3(a)). これにより, 観測画像 (カメラで撮影した画像) 上での画素値  $C(q)$  と近い画素値を持つサンプル点を探索でき, サンプル点の画素値を  $c_i$  とすると, 画素値の誤差は  $\|C(q) - c_i\|$  で表される. 画素値の誤差およびサンプル点から最近傍表面までの距離が最も小さい三次元点での色  $c_i$  が, 本来レンダリングされる色である. したがって, この点から  $-D_i \nabla D_i$  だけ移動した点が三次元空間上の対応点である. 三次元空間上の対応点を画像平面上に投影した二次元点  $\hat{q}$  および二次元点  $q$  の誤差  $\|q - \hat{q}\|$  が Reprojection Error となる (図 3(b)). Photometric Error および Reprojection Error を用いたカメラ姿勢推定手法は, Photometric Error のみを用いた従来手法と比較し, 高い収束性を持つことが示されている [16].

#### 4 NeDDF を用いた単眼 Visual SLAM

本研究は, マップ表現として NeDDF を用いた単眼 Visual SLAM を提案する. 本システムは, 単眼カメ

ラにより撮影された RGB 映像を入力し, 三次元シーン形状を記述した NeDDF モデルおよびカメラ軌跡を出力する.

##### 4.1 システム構成

提案システムの構成図を図 4 に示す. 一般に, Visual SLAM とよばれる三次元再構成システムは, カメラのトラッキングをリアルタイム (約 10Hz~) で実行可能なものを指す. また, トラッキングで求めたカメラ姿勢を使用し, マップ生成をトラッキングより遅い周期 (約 2Hz~) で実行する. 加えて, トラッキングの累積誤差により生じるドリフトを解消するために全体マップの最適化を行う場合もあるが, 本研究では扱わないものとする. PTAM[3] などの従来手法に倣い, トラッキングスレッドおよびマップ生成スレッドの並列処理によってこれを実現する. Visual SLAM においてトラッキングは, 参照データであるマップ情報 (点群やボクセルからなる三次元シーン形状) に対して, 観測データ (画素値や特徴点位置) の Registration を行うことでカメラ姿勢を算出する処理を指す. 本研究では, 学習された NeDDF モデルがマップ情報, 入力画像の画素値および画素位置が観測データに該当する. マップ生成は, 観測データをマップ座標系 (ワールド座標系) に変換し, 観測データをマップに追加する処理を指す. 両スレッドは, NeDDF モデルに対して並列に読み込み処理もしくは書き込み処理を行うため, メモリの競合に配慮する必要がある. iMAP[14] の実装を参考にし, 各スレッドが開始する時点でモデルパラメータを自スレッドにコピーすることで競合を防ぐ. 表記として, 入力映像のフレーム数を  $N_f$ , 入力映像を  $\{I_i \mid i = 0, 1, \dots, N_f - 1\}$ , 最終的なキーフレーム数を  $N_k$ , 全キーフレームを  $\{k_j \mid j = 0, 1, \dots, N_k - 1\}$  で表す.

##### 4.2 トラッキング

トラッキングスレッドの処理フローを図 4 上段に示す. まず, 前キーフレームの座標変換行列  $T_{j-1}$  の視点でボリュームレンダリングを行い, レンダリング画像を取得する. レンダリング画像および入力画像  $I_i$  を用いて, 3.2 節で前述した 2D-3D 対応点を求めた後, Reprojection Error を計算する. 次に, Reprojection Error および Photometric Error に対して最適化を行い座標変換行列  $T_i$  すなわちカメラ姿勢を推定する. 最後に, 入力フレームがキーフレームであるか判定する. 全ての入力フレームに対してモデルパラメータおよびカメラ姿勢を最適化することは処理負荷の観点から不可能であるため, 情報利得の大きいフレームをキーフレームとして代表させた疎な画像系列を用いる. 入力フレームが未観測な領域 (モデルパラメータが十分に学習されていない領域) であると判定した場合, 入力フ

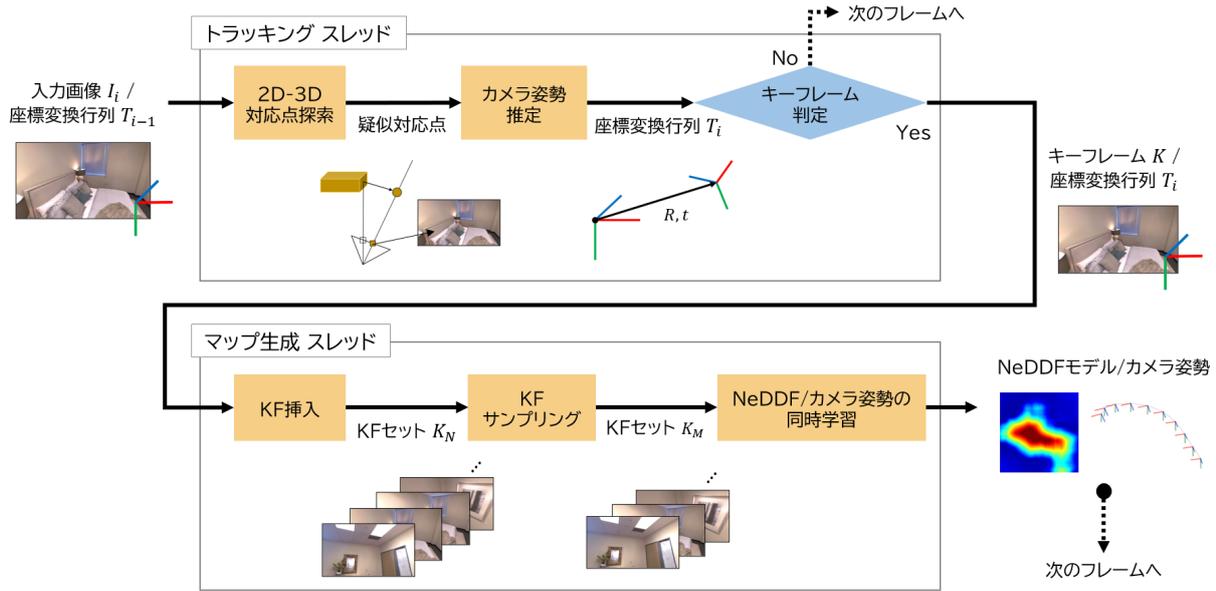


図 4: システム構成  
Fig. 4 Configuration of system

フレーム  $I_i$  および座標変換行列  $T_i$  をマップ生成スレッドに渡す。観測済みか判定する方法は iMAP[14] を参考にし、損失関数に対し閾値処理を行うことで未観測領域か判定する。未観測領域でない、すなわちキーフレームでない判定された場合、自スレッドは終了し、次のフレームのトラッキング処理を始める。

#### 4.3 マップ生成

マップ生成スレッドの処理フローを図 4 下段に示す。まず、トラッキングスレッドにてキーフレームと判定されたフレーム  $I_i$  を、キーフレームセット  $K_N$  に挿入する。次に、キーフレームセット  $K_N$ の中から、NeDDF モデルとカメラ姿勢の同時学習 (図 4 右下) で用いるキーフレームセット  $K_M$  を選択する。マップ生成は 2Hz 程度の動作速度が求められるため、全てのキーフレームに対して同時学習することは困難である。したがって、現在のキーフレーム  $k_j$ 、一つ前のキーフレーム  $k_{j-1}$ 、キーフレームセット  $K_N$  中から選択された  $M-2$  枚のキーフレームを合わせた合計  $M$  フレームに対して、同時学習を行う (iMAP[14] では  $M=5$ )。なお、キーフレームセット  $K_N$  中から選択される  $M-2$  枚のキーフレームは、キーフレームセット  $K_N$  のうち損失が最も大きいフレームが選択される。最後に、 $M$  枚のキーフレームからなる  $K_M$  に対し、NeDDF モデルとカメラ姿勢の同時学習を行う。

#### 4.4 マップ初期化

トラッキングは、これまでに生成されているマップ情報を参照し、観測データの Registration を行う。したがって、4.2 節で前述したトラッキングを実行する前に、マップの初期化を行う必要がある。本研究で

は、特徴点ベース手法である ORB-SLAM[7] のマップ初期化手法を用いる。8 点法もしくは DLT アルゴリズム [2] を用いて座標変換行列を算出した後、Full Bundle Adjustment を行うことで、初期 2 フレームのカメラ姿勢および三次元点群位置を算出する。なお、絶対スケールを取得することは困難であるため、初期 2 フレームでの移動距離を定数値  $l$  で与える。算出されたカメラ姿勢および初期 2 フレームの RGB 画像から NeDDF モデルを学習し、初期状態でのマップとする。

## 5 予備実験

### 5.1 実験構成

NeDDF モデルを使用し、RGB 画像群とそのカメラ姿勢から屋内環境の三次元シーン形状が学習できることを確認する。データセットには、屋内シミュレーション環境である Replica Dataset[10] の room1 シーン (図 5) を使用する。本シーンから取得された 100 枚の RGB 画像群および各カメラの座標変換行列を入力とし、NeDDF のモデルパラメータを学習する。画像 1 枚につき 256 点のサンプリング点をランダムに設定し、損失関数として Photometric Error および 3.1 節で前述した補助勾配に関するペナルティの加重和を計算し、学習を行う。パラメータ数などは、NeDDF[16] に記されたアーキテクチャ構成を用いる。

### 5.2 結果

図 6 に  $z=0$  でスライスした距離場および密度場、補助勾配を示す。図 5 と図 6(a), (b) を比較すると、詳細な形状は失われている一方、長方形である部屋の概形は学習されていることが確認できる。詳細な形状を



図 5: Replica Dataset の room1 シーン  
Fig.5 Room1 scene of Replica Dataset

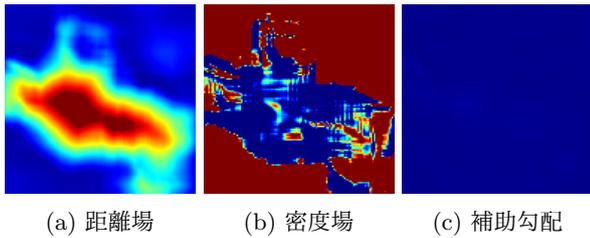


図 6: 学習後の各場  
Fig.6 Fields acquired by learning

学習できていない要因として、入力画像の枚数が少なく詳細な箇所に MLP のパラメータが割り当てられていないことや、壁などの広い領域がテクスチャレスであり深度値が不安定になることが挙げられる。前者の要因に対しては、Visual SLAM は三次元形状の再構成結果をリアルタイムで出力できることから、学習が不十分な箇所をユーザが確認し、不足箇所の画像を追加することで改善すると考えられる。後者の要因に対しては、単眼深度推定による補助などの改善方法が考えられるが、現時点ではテクスチャレスシーンでない別のデータセットを使用した評価を行う予定である。

## 6 おわりに

本稿では、単眼カメラで取得される RGB 映像から NeDDF モデルおよびカメラ姿勢の推定を行うための単眼 Visual SLAM を提案した。ニューラル場による三次元形状表現は、モデルがコンパクトかつ高分解能であることから Visual SLAM への適用が期待されている。NeDDF は、NeRF と同等の表現力を持ちながら、カメラ姿勢の追従性を高めた手法である。そこで本研究では、NeDDF を用いた単眼 Visual SLAM について検討し、システムの構築を目指す。予備実験として、屋内シミュレーション環境における NeDDF モデルの学習を行った。今後は、SLAM システムの実装を通じて、カメラ姿勢の推定精度および三次元シーンの再構成精度に関する検証を進める。

## 参考文献

- [1] P. Biber and W. Strasser: The normal distributions transform: a new approach to laser scan matching; IROS, vol. 3, pp. 2743-2748 (2003.10)
- [2] Hartley, R. I. and Zisserman, A.: Multiple View Geometry in Computer Vision Second Edition; Cambridge University Press, Cambridge (2004)
- [3] Georg Klein and David Murray: Parallel Tracking and Mapping for Small AR Workspaces; ISMAR, pp. 1-10 (2007.11)
- [4] R. A. Newcombe et al.: KinectFusion: Real-time dense surface mapping and tracking; ISMAR, pp. 127-136, (2011.10)
- [5] Johannes L. Schonberger and Jan-Michael Frahm: Structure-from-Motion Revisited; CVPR, pp. 4104-4113 (2016.6)
- [6] T. Taketomi, H. Uchiyama, and S. Ikeda: Visual SLAM algorithms: a survey from 2010 to 2016; IPSJ Transactions on Computer Vision and Applications, vol. 9, pp. 1-11 (2017.6)
- [7] Raúl Mur-Artal and Juan D. Tardós: ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras; IEEE Transactions on Robotics, vol. 33, no. 5, pp. 1255-1262 (2017.6)
- [8] Tong Qin, Peiliang Li, and Shaojie Shen: VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator; IEEE Transactions on Robotics, vol. 34, no. 4, pp. 1004-1020 (2018.7)
- [9] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove: DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation; CVPR, pp. 165-174 (2019.6)
- [10] J. Straub, T. Whelan, et al.: The Replica Dataset: A Digital Replica of Indoor Spaces; arXiv preprint, <https://arxiv.org/abs/1906.05797> (2019.6)
- [11] 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; ECCV, pp. 405-421 (2020.8)
- [12] Lin Yen-Chen, Pete Florence, Jonathan T. Barron, Alberto Rodriguez, Phillip Isola, and Tsung-Yi Lin: iNeRF: Inverting Neural Radiance Fields for Pose Estimation; IROS, pp. 1323-1330 (2021.9)
- [13] C. Lin, W. Ma, A. Torralba, and S. Lucey: BARF: Bundle-Adjusting Neural Radiance Fields; ICCV, pp. 5721-5731 (2021.10)
- [14] E. Sucar, S. Liu, J. Ortiz, and A. J. Davison: iMAP: Implicit Mapping and Positioning in Real-Time; ICCV, pp. 6209-6218 (2021.10)
- [15] Z. Wang, S. Wu, W. Xie, M. Chen, and V. A. Prisacariu: NeRF-: Neural Radiance Fields Without Known Camera Parameters; arXiv preprint, <https://arxiv.org/abs/2102.07064> (2022.4)
- [16] Itsuki Ueda, Yoshihiro Fukuhara, Hirokatsu Kataoka, Hiroaki Aizawa, Hidehiko Shishido, Itaru Kitahara: Neural Density-Distance Fields; ECCV, pp. 53-68 (2022.10)