

# NeDDF を用いた煙霧環境に適用可能な ニューラル場単眼 Visual SLAM

高橋 響熙<sup>\*1</sup> 上田 樹<sup>\*2</sup> 謝 淳<sup>\*3</sup> 北原 格<sup>\*3</sup>

Monocular Visual SLAM Applicable to Smoggy Environments Using NeDDF

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

**Abstract** – We propose a monocular Visual SLAM system based on NeDDF (Neural Density-Distance Field). In contrast to previous methods using point clouds or meshes, our approach has an advantage to reconstruct objects with indistinct boundaries, such as steam, smoke, dust, and similar phenomena. Our method utilizes NeDDF as a 3D shape representation and introduces a photometric error to account for the challenging smoke environment. This approach enables the reconstruction of indoor scenes featuring low-density objects using a monocular RGB camera. To evaluate the effectiveness of our proposed method, we conduct experiments using a simulated RGB dataset of an indoor scene with smoke.

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

## 1 はじめに

本稿は、煙霧が存在する屋内シーンの三次元再構成を目的とした、ニューラル場表現 NeDDF (Neural Density-Distance Field) を用いた単眼 Visual SLAM (Simultaneous Localization and Mapping) を提案する。

メタバースの流行などにより、現実世界の物体や場所をユーザが撮影することで、被写体をリアルな質感で 3D データ化するアプリケーションに注目が集まっている。これには SfM (Structure from Motion) [4] または Visual SLAM とよばれる技術が用いられる。SfM は入力情報である多量の画像から、被写体の三次元形状およびカメラ姿勢をバッチ処理によって推定する技術を指す。対照的に Visual SLAM は時系列な映像データが入力情報であり、オンライン処理によってこれらを推定する技術を指す。ロボットナビゲーションや AR/MR アプリケーションなど、ユーザとのリアルタイム応答が求められるシステムに多く使用されている。本稿では、単眼 RGB 映像を入力情報とする Visual SLAM について述べる。

従来の Visual SLAM[1, 3, 5, 6] はシーンの三次元形状を点群やメッシュで表現するため、適用対象は境

界が明確な物体で構成されるシーンに限られている。そのため、煙や湯気、塵などの境界が曖昧なオブジェクトは視認性を悪化させ姿勢推定の精度が低下することから、センサフュージョンや検出といったデノイズに向けた対応が取られてきた [2, 13]。しかし、煙や湯気のような明確な境界を持たない低密度物体も含め三次元空間中に存在する全ての物体による三次元シーン表現が利用可能であれば、より汎用的な Visual SLAM が実現されるであろう。そこで本研究では、図 1 に示す通り、提案手法は明確な境界を持つ物体は従来通りメッシュデータとして、低密度物体はボリュームデータとして記述することを試みる。こうすることで低密度物体と CG オブジェクトの相互的なインタラクションが実現でき、AR/MR アプリケーションにおける表現の幅が広がると考えられる。例えば、低密度物体が透けて描画されることによる前後関係の表現や、CG オブジェクトの旗が湯気によって揺れる物理現象の表現などが実現できる。

Visual SLAM は、粗なシーン表現手法または密なシーン表現手法に大別される。本研究では再構成結果をメッシュデータとして記録することを想定するため、密なシーン表現手法を用いる必要がある一方で、低密度物体が記述可能な手法でなければならない。これら要件を満たすため、本研究ではシーン表現手法としてニューラル場を導入する。ニューラル場による形状表現は、密度場または距離場が多く用いられる。物体形状を密度場として保持する NeRF は、ボリュームレンダリングを用いることで煙霧などの明確な境界を持

<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

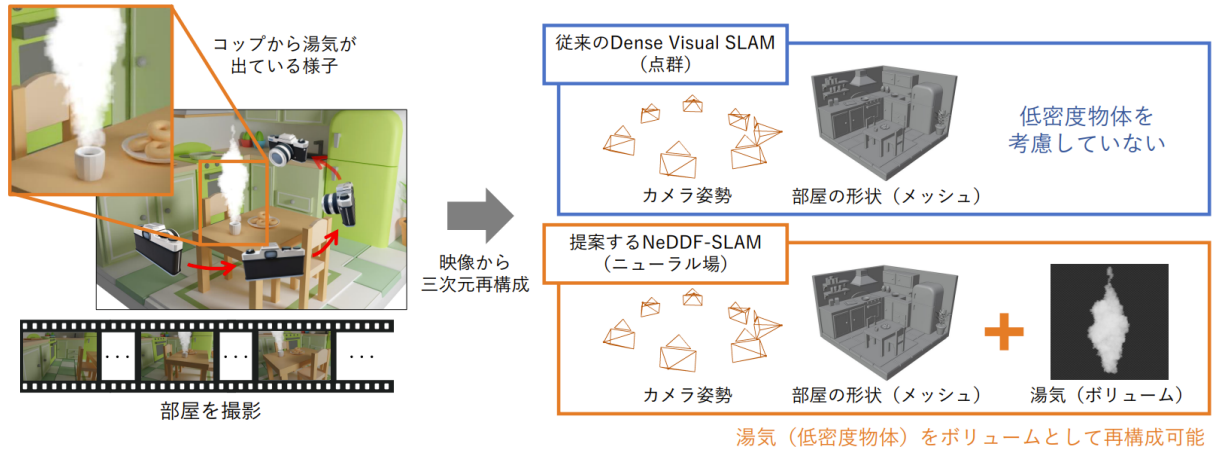


図 1: 従来の Dense Visual SLAM と提案手法 NeDDF-SLAM の比較。ニューラル場 NeDDF および低密度物体を考慮した損失関数を導入することで、湯気のような低密度物体をボリュームデータとして再構成可能。

たない物体のレンダリングが可能である。一方、メッシュへの変換を想定する場合には符号付き距離場（または陰関数表現）が使用される。符号付き距離場のゼロ等高線をマーチンキューブ法によって抜き出すことで、物体表面をメッシュに変換が可能である。しかし、境界が明確でない煙霧などの形状表現は困難である。このような理由から、距離場と密度場の双方を保持可能な NeDDF[11] が上田らにより提案された。NeDDF は距離場の勾配情報を密度に変換するというアイデアのもと、距離場を保持しながら低密度物体の表現を可能にした。以上のことから、低密度物体を含む屋内シーンの三次元再構成を目的とする本研究は、シーン表現手法として NeDDF を用いる。

また、Visual SLAM は目的関数によって非直接系（特徴点系）または直接系に大別される。非直接系手法は画像特徴点の再投影誤差を目的関数とし、直接系手法は三次元点を画像平面に投影した時の測光誤差を目的関数とする。直接系手法は特徴点単位ではなく画素単位での三次元再構成を行うため細かい形状表現が可能である。また、テクスチャレスで輝度変化が小さい環境において頑健である利点を持つ。一方で、非直接系手法は特徴記述子の性能に依存する上、煙霧は視点位置により形や輝度が大きく変化するため本研究が想定する環境では精度が低下すると考えられる。したがって、本研究では直接系手法を採用する。

ニューラル場を用いた Visual SLAM として DIM-SLAM[12] が挙げられる。DIM-SLAM はこれまで直接手法で用いられてきた測光誤差を用いて、NeRF モデルおよびカメラ姿勢を推定する手法を提案した。しかし、低密度物体を含むシーンでは低密度物体が透過して奥の背景が見える現象が発生する。したがって、視点位置に依らず輝度は不変であるという測光誤差の仮定が成立しない。

メッシュや点群によるシーン表現では、物体境界の曖昧さと測光誤差の関係性を議論することが困難であった。一方、密度場に対してボリュームレンダリングを実行することで、分散を持った深度が計算可能である。そこで、深度の分散情報を加味した測光誤差を定義し、検証を行う必要がある。

具体的には、深度の分散情報はカメラの光学中心から三次元空間中に飛ぶレイに沿って広がっているため、境界点を観測する視点位置によって分散の大きさが変化する。また、その境界点からカメラまでの距離に応じて変化する。本稿では、これらを考慮しながら深度の分散情報を加味した測光誤差の定式化を行う。本手法の有効性を検証するため、煙霧が存在する屋内シーンの RGB 映像データセットを作成し、カメラ姿勢推定および三次元再構成の精度を評価した。

本研究の貢献は以下の三点である。

1. シーン表現手法に NeDDF を用いた単眼 Visual SLAM システムの提案
2. 深度の分散情報を加味した測光誤差の定義
3. 煙霧を含む屋内シーンを再現した単眼 RGB 映像データセットの作成

## 2 関連研究

### 2.1 ニューラル場

ニューラル場は、ベクトル場またはスカラー場をニューラルネットワークで近似する手法の総称であり、三次元再構成の研究領域では三次元形状を記述する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場は密度場または距離場に大別される。

**密度場/距離場** — 密度場は、三次元座標を指定するとその点での体積密度が定まる場である。距離場（もしくは陰関数表現）は、三次元座標を指定すると、そ

の点から最近傍表面までの距離が定まる場である。特に符号なし距離場は物体内部で0、物体外部で正の値を取るような距離場である。

密度場の例として、自由視点映像生成の領域で顕著な成果を上げた NeRF[7] は三次元形状を密度場で表現する。本研究で用いる NeDDF[11] は、密度場と距離場の双方を保持する手法である。NeDDF は距離場を保持しながら低密度物体の表現も可能であり、煙霧を含む環境を想定する本研究において最適なシーン表現手法である。

NeRF ではカメラ姿勢が既知な多量の観測画像を入力情報とし、微分可能レンダリングにより生成した画像と入力画像とのレンダリング誤差を最小化することで輝度場を獲得する。正確なカメラ姿勢情報を必要とせず、カメラ姿勢と NeRF モデルの同時学習（バンドル調整）を行う手法も提案されている [8, 9]。本研究はこれらの手法と同様に誤差逆伝播法に基づくバンドル調整を行う。

## 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オンライン処理により推定する技術である。Visual SLAM は以下の二軸、シーン表現の粗/密および目的関数の非直接/直接によって大別される。

**疎/密** — PTAM[1] などに代表される疎なシーン表現は、画像上のエッジやコーナーなどの特徴点のみを再構成し、テクスチャレスな物体表面の形状情報は含まない表現である。DTAM[3] などに代表される密なシーン表現は、特徴点に限らず全画素を再構成することでオブジェクトの形状全体を記述した表現である。

**非直接/直接** — ORB-SLAM[5] などに代表される非直接系手法は、画像上の特徴点とその特徴マッチングに基づき、三次元点を画像平面に再投影したときの幾何誤差（再投影誤差）を目的関数とする。DSO[6] などに代表される直接系手法は、ある三次元点を観測したときに視点位置に依らず輝度は不変であるという仮定から、複数視点で同一の三次元点を観測したときの測光誤差を目的関数とする。

ニューラル場は密なシーン表現が可能であり、Visual SLAM への適用が試みられている。NeRF を Visual SLAM に適用した iMAP[10] は、RGB-D 映像を入力とし逐次的に NeRF モデルを学習する。この手法は入力デプスマップとレンダリングされたデプスマップの幾何誤差を最小化することで三次元形状を推定する。DIM-SLAM[12] は、RGB 映像のみから逐次的に NeRF モデルを学習する手法として最初期に提案され

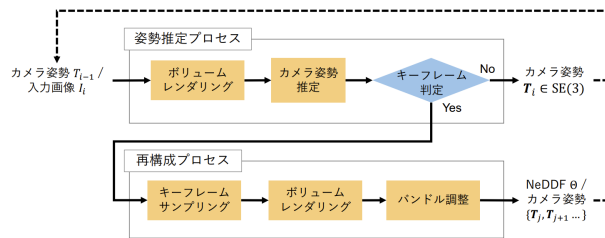


図2: 提案手法の処理の流れ。

た。この手法は、デプス情報が不足する問題に対して測光誤差を導入することで三次元形状の学習を可能にしたが、被写体に明確な境界が存在することを強い仮定としている。本研究では、煙や湯気などの物体境界がないオブジェクトを含めた三次元再構成を目指す。

## 3 NeDDF-SLAM

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

NeDDF は MLP で構成された密度場に変換可能な符号なし距離場および色場  $F_\Theta$  からなる。符号なし距離場は三次元座標を指定すると、物体外部では最近傍表面までの距離を返し、物体内部では0を返す場である。次式を用いることで、符号なし距離場の勾配情報から密度  $\sigma$  に変換が可能である。

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

NeRF と同様に、ボリュームレンダリングは密度場に対して次式を実行する。

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^N w_i c_i, \quad \hat{D}(\mathbf{r}) = \sum_{i=1}^N w_i d_i, \quad (2)$$

$$w_i = T_i (1 - \exp(-\sigma_i \delta_i)) \quad (3)$$

光学中心から画素方向に飛ばしたレイ上を  $\delta_i = d_{i+1} - d_i$  間隔でサンプリングし、透明度  $T_i = \exp(-\sum_{j=1}^{i-1} \sigma_j \delta_j)$  を求める。サンプリングした色および密度を式(3)の重みで加重和を取ることで、画素の輝度およびデプスが計算される。

### 3.2 NeDDFを用いた単眼 Visual SLAM

本研究が提案する NeDDF-SLAM は、シーン表現手法として前述した NeDDF を使用する。また、直接系手法で用いられる測光誤差を煙霧が存在する環境に適用するため、ボリュームレンダリングから算出されるデプスの分散情報を利用した Volumetric Consistency Loss を導入する。従来の測光誤差の問題点および Volumetric Consistency Loss の導出については次節で詳述する。

NeDDF-SLAM のシステム構成を図2に示す。本手法は姿勢推定プロセスと再構成プロセスの並列シス

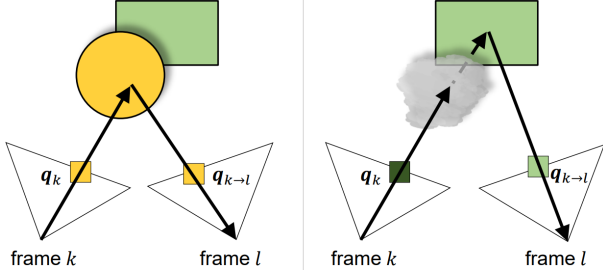


図3: 低密度物体を含むシーンでは同一点の輝度が視点位置によって変化。

テムであり、各プロセスは推定したカメラ軌跡および NeDDF モデルを共有メモリに持つ。各プロセスの詳細な処理フローは5節で述べる。

## 4 目的関数

### 4.1 Photometric Consistency Loss

測光誤差は三次元空間中のある点を複数のカメラで観測したときの輝度の差であり、本稿では Photometric Consistency Loss とよぶこととする。ニューラル場を用いた Visual SLAM においても測光誤差は、三次元形状を学習するための有効的な手段である。Photometric Consistency Loss は次式で定義される。

$$L_{\text{photometric}} = \frac{1}{\mathcal{M}} \sum_{\mathbf{q}_k \in \mathcal{M}} \sum_{\substack{l \in \mathcal{S} \\ l \neq k}} \|\mathbf{I}(\mathbf{q}_k) - \mathbf{I}(\mathbf{q}_{k \rightarrow l})\| \quad (4)$$

$$\mathbf{q}_{k \rightarrow l} = \Pi_c(\mathbf{T}_{k \rightarrow l} \circ \Pi_c^{-1}(\mathbf{q}_k, d(\mathbf{q}_k))) \quad (5)$$

ここで、 $\mathcal{M}$  はサンプリング画素の集合、 $\mathcal{S}$  はキーフレームセット、 $\mathbf{q}_k$  はサンプリング画素の画像座標、 $\Pi_c$  は世界座標系から画像座標系への投影、 $\Pi_c^{-1}$  は画像座標系から世界座標系への逆投影、 $\mathbf{T}_{k \rightarrow l}$  はフレーム  $k$  からフレーム  $l$  への同次座標変換行列、 $d(\mathbf{q}_k)$  は  $\mathbf{q}_k$  のデプス値である。

図3左のような低密度物体が存在しないシーンにおいて、視点に依らず輝度が変わらないという仮定は成立する。しかし、図3右のような背景が透過する低密度物体が存在するシーンでは、この仮定は成立しない。従来手法は低密度物体の有無に差がある画素に対して同じ測光誤差を用いており、低密度物体の性質を考慮しているとは言えない。

### 4.2 Volumetric Consistency Loss

本研究は、低密度物体の有無を考慮した測光誤差を提案する。

第一に、分散を持った三次元点の透視投影について述べる。図3右において、ボリュームレンダリングによって計算されるデプスは低密度物体から緑オブジェクトまでの間で値をとる。また、画像座標  $[u, v]$  にお

けるデプスの分散は次式で計算可能であり、境界の不確か性を表す。

$$\hat{D}_{\text{var}}(u, v)[\text{m}^2] = \sum_{i=1}^N w_i (\hat{D}(u, v) - d_i)^2 \quad (6)$$

図4(a)に示す通り、式(6)は逆投影によって計算される三次元点はレイ方向に分散を持っている。分散を持った三次元点をフレーム  $l$  に透視投影を行ったとき、フレーム  $k, l$  間の視差が鋭角である場合にこの分散は小さく投影される。また、三次元点からフレーム  $l$  の光学中心までの距離が長い場合にも小さく投影される。よって、視点の位置角度に応じて次式でスケールングを行う。

$$\hat{D}_{\text{var}}^l[\text{pixel}^2] = \frac{f}{Z_l} \|\mathbf{d}_k \times \mathbf{d}_l\| \hat{D}_{\text{var}} \quad (7)$$

ただし、 $Z_l$  はフレーム  $l$  に投影した際の三次元点から光学中心までの距離、 $f$  は焦点距離、 $\mathbf{d}_k \times \mathbf{d}_l$  はフレーム  $k, l$  の光学中心から三次元点までのレイ  $\mathbf{d}_k, \mathbf{d}_l$  のクロス積である。式(7)によって、三次元空間中の分散は、図4(b)に示すようにエピポーラ線上の画像座標系における分散に変換される。

第二に、測光誤差から再投影誤差への近似について述べる。前述の通り、測光誤差は画像座標  $\mathbf{q}_k$  および  $\mathbf{q}_{k \rightarrow l}$  での輝度の差  $E_{\text{photo}}(\mathbf{q}_{k \rightarrow l}) = \|\mathbf{I}(\mathbf{q}_k) - \mathbf{I}(\mathbf{q}_{k \rightarrow l})\|$  で表される。空間中の三次元形状と整合する画像座標  $\bar{\mathbf{q}}_{k \rightarrow l}$  において測光誤差  $E(\bar{\mathbf{q}}_{k \rightarrow l})$  は0となる。ここで画像座標系での誤差  $\|\bar{\mathbf{q}}_{k \rightarrow l} - \mathbf{q}_{k \rightarrow l}\|$  は、非直接系手法で用いられる再投影誤差に等しい。再投影誤差は三次元空間中の幾何誤差に比例するという利点から、本研究では測光誤差から再投影誤差への近似を行う。詳細には、次式の通りニュートン法を用いて求める。

$$\bar{\mathbf{q}}_{k \rightarrow l} = \arg \min_{\mathbf{q}'_{k \rightarrow l}} E_{\text{photo}}(\mathbf{q}'_{k \rightarrow l}) \quad (8)$$

$$\simeq \mathbf{q}_{k \rightarrow l} - E_{\text{photo}}(\mathbf{q}_{k \rightarrow l}) \left( \frac{\partial E_{\text{photo}}(\mathbf{q}'_{k \rightarrow l})}{\partial \mathbf{q}'_{k \rightarrow l}} \Big|_{\mathbf{q}_{k \rightarrow l}} \right)^{-1}$$

式(8)の勾配  $\partial E_{\text{photo}}(\mathbf{q}'_{k \rightarrow l}) / \partial \mathbf{q}'_{k \rightarrow l}$  は、ピラミッド画像を用いて再投影点  $\mathbf{q}_{k \rightarrow l}$  の近傍の輝度値から算出される。分散  $\hat{D}_{\text{var}}$  はエピポーラ線上で広がりであるため、画像座標点  $\bar{\mathbf{q}}_{k \rightarrow l}$  のエピポーラ線への射影に次式で変換する。

$$E_{\text{repro}}(\mathbf{q}_{k \rightarrow l}) = \|\mathbf{n}_{\text{epipolar}} \cdot (\bar{\mathbf{q}}_{k \rightarrow l} - \mathbf{q}_{k \rightarrow l})\| \quad (9)$$

最後に、低密度物体を考慮した測光誤差について述べる。前述した分散および再投影誤差から、次式を Volumetric Consistency Loss と定義する。

$$L_{\text{volumetric}} = \frac{1}{\mathcal{M}} \sum_{\mathbf{q}_k \in \mathcal{M}} \sum_{\substack{l \in \mathcal{W} \\ l \neq k}} \frac{1}{\sqrt{\hat{D}_{\text{var}}^l}} E_{\text{repro}}(\mathbf{q}_{k \rightarrow l}) \quad (10)$$

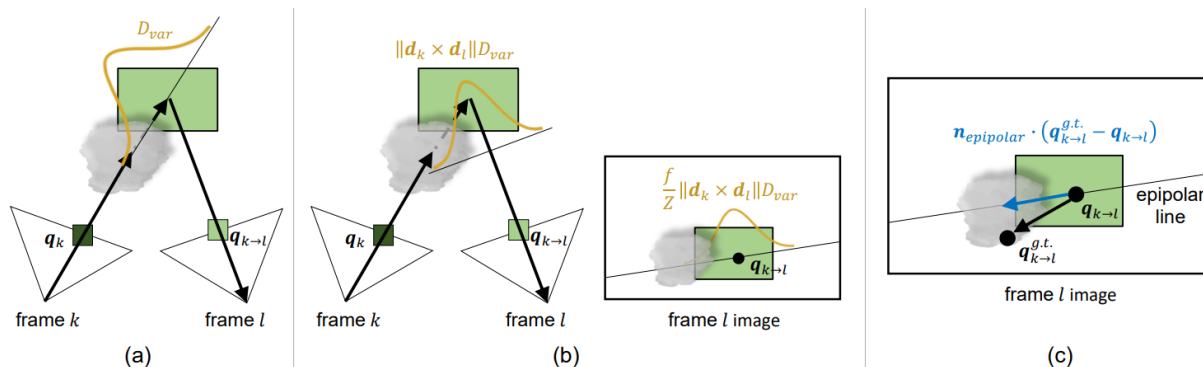


図 4: 低密度物体を考慮した測光誤差

後述する各プロセスでは次式を損失関数とし、カメラ姿勢推定および三次元再構成を行う。  $L_{\text{rendering}}$  はレンダリングされた輝度と入力輝度の誤差、各  $\lambda$  は重みのハイパーパラメータである。

$$\text{Loss} = L_{\text{volumetric}} + \lambda_r L_{\text{rendering}} \quad (11)$$

## 5 システム構成

### 5.1 キーフレームセット

本研究ではキーフレームを、局所キーフレームセット  $\mathcal{L}$  および全域キーフレームセット  $\mathcal{G}$  で管理する。局所キーフレームセット  $\mathcal{L}$  は、現在フレームから  $N_{\text{local}}$  フレーム手前までの集合  $\{I_{i-N_{\text{local}}}, I_{i-N_{\text{local}}+1}, \dots, I_i\}$  とする。全域キーフレームセット  $\mathcal{G}$  は、次項にて後述するキーフレーム判定においてキーフレームと判定されたフレームを逐次追加することで生成する。

### 5.2 姿勢推定プロセス

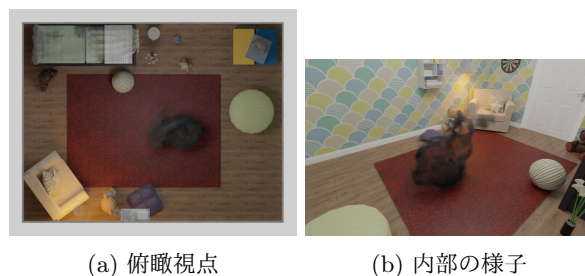
図 2 上段の姿勢推定プロセスでは、現在フレーム  $I_i$  のカメラ姿勢を推定する。以下、詳細について述べる。

まず、前フレームのカメラ姿勢  $T_{i-1}$  の視点でボリュームレンダリングを行い、レンダリング誤差を求める。また、局所キーフレームセット  $\mathcal{L}$  および現在フレームの間で、Volumetric Consistency Loss を求める。次に、前述した損失関数を誤差逆伝播法を用いて最小化することで、カメラ姿勢  $T_i$  を推定する。最後に、キーフレーム判定を行う。キーフレーム判定では、直近の全域キーフレームから現在フレームまでのオプティカルフローの平均が閾値以上であるとき、キーフレームと判定し全域キーフレームセット  $\mathcal{G}$  に追加する。

### 5.3 再構成プロセス

図 2 下段の再構成プロセスでは、バンドル調整を行う。以下、詳細について述べる。

まず、全域キーフレームセットからバンドル調整に用いるキーフレームをサンプリングする。具体的には、現在フレームのカメラ視点方向  $R_i \in \text{SO}(3)$  および全域キーフレームセットのカメラ視点方向  $\{R_j | j \in \mathcal{G}\}$



(a) 俯瞰視点

(b) 内部の様子

図 5: 煙霧あり屋内シーンデータセット。部屋の中央に兎の形をした煙を配置。

の間で相対回転  $R_i R_j^{-1}$  を求め、回転角の小さい順から  $N_{\text{global}}$  個のキーフレームを選択する。次に、サンプリングしたキーフレームおよび局所キーフレームセットでボリュームレンダリングを行い、レンダリング誤差および Volumetric Consistency Loss を求める。最後に、前述した損失関数を誤差逆伝播法を用いて最小化することで、バンドル調整を行う。

## 6 実験

### 6.1 仮説

煙霧を含む屋内シーンにおいて、従来手法である Photometric Consistency Loss を使用した場合と比較し、カメラ姿勢推定および三次元再構成の精度が向上すると考えられる。本節では、これらについて検証を行う。

### 6.2 データセット

検証のため、煙霧が存在する屋内シーンを再現したデータセットを作成した。部屋の構造を図 5 に示す。煙霧は OpenVDB が配布するサンプルモデルである。

### 6.3 構成

DIM-SLAM[12] と同様に、入力映像の冒頭 15 フレームの RGB 画像とカメラ姿勢からマップの初期化を行った後、カメラ姿勢推定およびバンドル調整を逐次的に実行する。図 5 のシーン中を撮影した計 500 フレーム (約 17 秒) の RGB 映像を入力とする。シーン表現手

表 1: カメラ姿勢の推定誤差

目的関数	ATE RMSE[m]
$Loss_{render} + Loss_{photo}$	1.503
$Loss_{render} + Loss_{volume}$	1.293

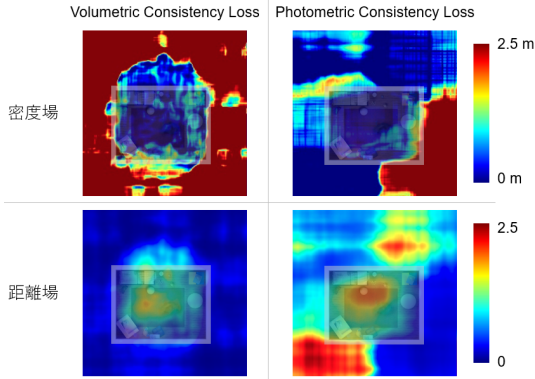


図 6: 推定距離場および密度場のスライス。部屋の俯瞰画像を重畳し誤差を可視化。

法としてニューラル場を使用するが本手法は拡散反射のみを想定するため、輝度値推定に関わるネットワークには視線方向を入力しない。実験では、従来手法である Photometric Consistency Loss を用いる手法および提案手法について比較する。

カメラ姿勢推定の精度は、推定されたカメラ軌跡の誤差である Absolute Trajectory Error により定量評価を行う。三次元再構成の精度は、床から 1m 地点での密度場および距離場のスライスを可視化することで定性評価を行う。

#### 6.4 結果

表 1 にカメラ姿勢の推定誤差を示す。結果から目的関数に Volumetric Consistency Loss を用いた手法は従来手法よりも高い推定精度であった。しかし、双方ともに映像の途中でトラッキングできなくなり、真値カメラ姿勢から大きく外れた結果となった。これは RGB 画像の約 1/4 を煙霧を占めるような視点ではカメラ姿勢を制約することが困難であったためだと考えられる。

図 6 に推定した距離場および密度場の床から 1m 地点でのスライスを示す。Volumetric Consistency Loss を用いた手法は部屋の概形を獲得できているのに対し、従来手法は壁の一部が欠損した結果となった。前景である煙霧を撮影しながら壁などの背景を同時に推定するため、従来の明確な物体境界を仮定した従来手法では部屋の形状を獲得することができないと考えられる。しかし、本研究の目的である煙霧形状を推定することは、双方の手法において実現できなかったため、今後の課題したい。

## 7 おわりに

本稿では、低密度物体を考慮した測光誤差を目的関数とする NeDDF を用いた単眼 Visual SLAM を提案した。単眼映像から低密度オブジェクトを含む室内シーンの三次元再構成を目的とする。ニューラル場を用いて境界が不明瞭なオブジェクトに対して定義可能な重み付き測光誤差を定義した。提案手法が有用であるか検証するため、煙を含む室内シーン RGB 映像データセットを使用しカメラ姿勢および三次元形状の推定精度に関する実験を行った。

## 参考文献

- [1] Georg Klein and David Murray: Parallel Tracking and Mapping for Small AR Workspaces; ISMAR, pp. 1–10 (2007.11)
- [2] B. Christopher, P. Thierry, V. Teresa: Combining multiple sensor modalities for a localisation robust to smoke; IROS, pp. 2489–2496 (2011.9)
- [3] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison: DTAM: Dense tracking and mapping in real-time; ICCV, pp. 2320–2327, (2011.11)
- [4] Johannes L. Schonberger and Jan-Michael Frahm: Structure-from-Motion Revisited; CVPR, pp. 4104–4113 (2016.6)
- [5] 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)
- [6] J. Engel, V. Koltun, and D. Cremers: Direct Sparse Odometry; TPAMI, vol. 40, no. 3, pp. 611–625, (2018.3)
- [7] 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)
- [8] 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)
- [9] C. Lin, W. Ma, A. Torralba, and S. Lucey: BARF: Bundle-Adjusting Neural Radiance Fields; ICCV, pp. 5721–5731 (2021.10)
- [10] E. Sucar, S. Liu, J. Ortiz, and A. J. Davison: iMAP: Implicit Mapping and Positioning in Real-Time; ICCV, pp. 6209–6218 (2021.10)
- [11] Itsuki Ueda, Yoshihiro Fukuhara, Hirokatsu Kataoka, Hiroaki Aizawa, Hidehiko Shishido, Itaru Kitahara: Neural Density-Distance Fields; ECCV, pp. 53–68 (2022.10)
- [12] Heng Li, Xiaodong Gu, Weihao Yuan, luwei yang, Zilong Dong, and Ping Tan: Dense RGB SLAM With Neural Implicit Maps; ICLR (2023.5)
- [13] Z. Jun, X. Renxiang, L. Heshan, L. Yiyao, S. Xudong, H. Chaoyu, L. Zhongxu, W. Danwei: 4DRT-SLAM: Robust SLAM in Smoke Environments using 4D Radar and Thermal Camera based on Dense Deep Learnt Features; CIS-RAM (2023.6)