

テクスチャの乏しい環境に対する 線構造光のレンダリングに基づく深度推定

伊賀上 卓也¹, 安 琪², 山下 淳²

¹ 東京大学大学院 工学系研究科 精密工学専攻

² 東京大学大学院 新領域創成科学研究科 人間環境学専攻

本研究では、線構造光を考慮したレンダリングにより、深度推定を行う方法を提案する。屋内廊下やトンネルのようなテクスチャの乏しい環境では、テクスチャを前提とせず構造光に基づき姿勢推定を行うことが有効である。その一方、構造光のみに基づく姿勢推定では、姿勢推定精度を十分に高めることが困難であった。近年の多視点画像のレンダリングに基づく深度推定では、テクスチャの豊富な環境に限定されるものの、高精度な姿勢推定が可能である。以上から本研究では、深層学習に基づくテクスチャを利用した姿勢推定手法に対し構造光を考慮した最適化アルゴリズムを構築することで、深度推定の高精度化を行った。

1. 序論

構造物内部の 3 次元形状はロボットの自動走行や構造物の施工品質管理に活用可能である重要なプロセスである。中でも屋内廊下やトンネルをはじめとした大型構造物においてはテクスチャが少ないことから、限られたテクスチャの中で 3 次元計測を行う手法が要求されている。また、この 3 次元計測においては、労働力不足の観点から、近年ではロボットを使用する需要が高まっている。

テクスチャの僅かな箇所での 3 次元計測を行う手法として、トータルステーション、レーザスキャナ、構造光に基づく三角測量等の手法がこれまで提案されてきた。中でも、構造光に基づく 3 次元計測はロボットに広く搭載されているカメラ以外に追加のセンサを要求せず、UAV や小型ロボットのように積載重量制限が低いロボットに対する計測手法として有用である。構造光に基づく計測手法の中でも光切断法 [1-3] では、ラインレーザを使用することでレーザ光が照射された線上の高密度な点群を取得可能であり、大型構造物への適用において有用である。

3 次元計測装置により、ある地点における局所的な点群が取得可能である一方、図 1 に示すように、長尺構造物である屋内廊下やトンネル全体の 3 次元点群を取得するためには複数地点の計測結果を結合することが必要である。計測地点間の位置姿勢変化を算出する方法として、カメラを用いた計測手法では様々な計測対象において環境中のテクスチャから姿勢を推定する方法が採用されてきた [4]。ところが、屋内廊下やトンネルは単調な色である場合が多く、テクスチャに基づく姿勢推定を達成可能な場面が限定される。これを解決するために、先行研究 [3] では、テクスチャを利用せずに 3 次元計測を行うシステムを構築した。本計測装置では、

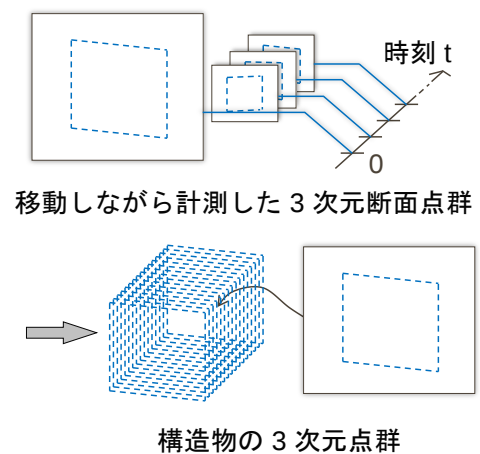


図 1 局所 3 次元計測点群の統合

計測に用いるカメラを 2 台にし、カメラ間の位置姿勢推定を照射されたラインレーザ光に基づき行った。これにより、各カメラを交互に移動させることでテクスチャを利用せずに姿勢推定を可能とした。その一方、このシステムでは UAV へ搭載した場合のドリフト誤差や、レーザ光のみを利用した最適化では姿勢推定精度が十分に高くないという問題があった [5]。特に姿勢推定精度の低さの原因として、環境中のテクスチャを完全に利用しないことに起因し、位置姿勢推定の手掛かりが少なくなっていることが挙げられる。そのため、テクスチャと構造光を組み合わせた位置姿勢の最適化を行うことが望ましい。

カメラ 1 台を用いた姿勢推定手法では、前述のように利用場面がテクスチャの豊富な環境に限定される欠点はあるものの、深層学習に基づく最適化を通じ高精度な姿勢推定と深度算出が可能である [6,7]。そして、この深度マップは 3 次元的情

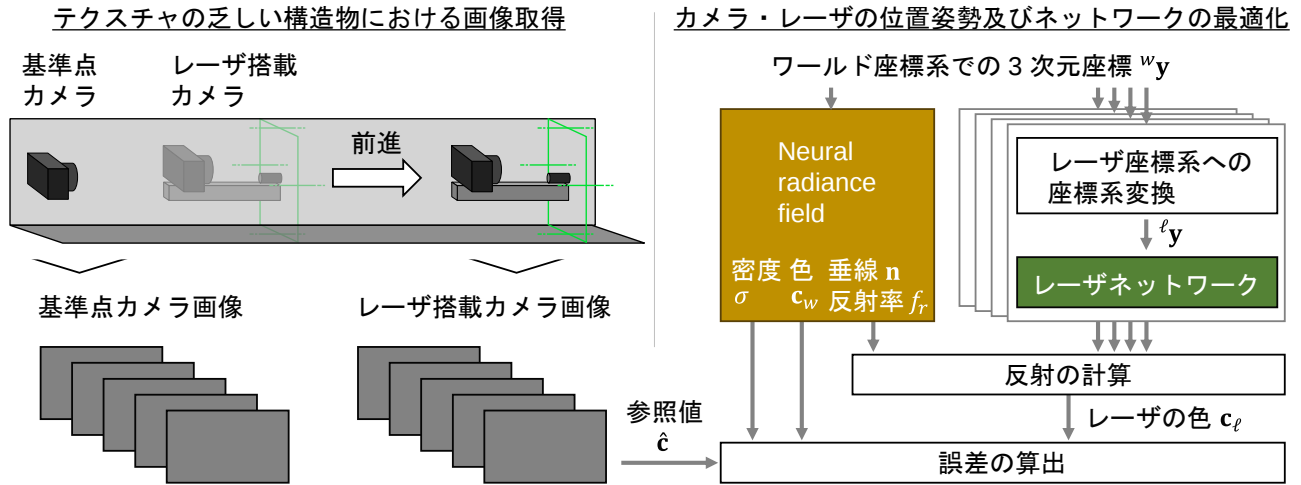


図2 提案手法の概略図

報を含むため、3次元点群と同様に構造物のモデル化や変形の検知等へ利用可能である。以上から、本研究では、テクスチャに基づく手法と構造光に基づく手法を組み合わせた位置姿勢推定と深度推定を研究の目的とした。

2. 関連研究

2-1. 長尺構造物でのロボットの姿勢推定手法

長尺構造物においては、複数地点の3次元計測結果を統合するために、ロボットの位置姿勢推定を行うことが必要である。1つの方法として、車輪の回転量[8]から推定するものがある。ところが、この先行研究で提案された手法では、地上で滑りが発生することや、地上をそもそも走行できないことが問題となり、適用範囲が大きく限定される。また、レーザスキャナのように、ある区間において全方位の3次元点群を取得するシステムに対しては、3次元的な形状をマッチングさせる方法が提案されている[9]。ところが、この手法は環境中に特徴的な形状を要求するため、トンネルのように単調な形状が連続する環境に対しては適用できない場面が多い。カメラシステムに対しては、テクスチャのみを用いた姿勢推定手法[1,2,4]が提案されてきたが、テクスチャの乏しい環境では姿勢推定と同時に取得される3次元形状に大きな誤差を生じる。

2-2. テクスチャと構造光を考慮した形状推定

Higuchi ら[2]はテクスチャとレーザ光を組み合わせた位置姿勢推定手法を提案した。ところが、この先行研究で提案されたシステムはカメラ1台のみで計測するものであり、姿勢推定はテクスチャのみを利用する。そのため、テクスチャの乏しい環境において利用することができなかった。

テクスチャと構造光を同時に考慮した最適化を

行うためには、光源を含めた深層学習に基づく姿勢推定モデルが有用である。RefNeRF[10]では、光の反射をneural radiance fields (NeRF)[11]に対してモデル化することで、撮像方向によって変化する反射のレンダリングを可能とした。さらにShandilya ら[12]は、RefNeRFを構造光に基づく深度推定に対して拡張した。スポットレーザパターンによる構造光を含めたレンダリングを行っており、構造光を組み合わせることで、テクスチャの乏しい箇所も3次元形状を取得可能としている。ところが、この先行研究はカメラ1台のみからなる手法を提案したため、姿勢推定のために、環境中にテクスチャまたは特徴的な3次元形状があることを要求する。

3. 提案手法

3-1. 概要

先行するNeRFに基づく姿勢推定手法であるBARF[6]は、1台のカメラを使用し、テクスチャを利用した姿勢推定を行った。そのため、この手法はテクスチャの乏しい対象には適用できない。そこで、図2に示すように、提案手法では、レーザ光を手掛かりにカメラ位置姿勢を行う手法を構築した。基準点カメラ C_r を環境中に停止させてレーザ搭載カメラ C_l を移動させることで、基準点カメラ C_r からの相対位置姿勢としてレーザ搭載カメラ C_l の姿勢を推定する。これにより、環境中のテクスチャがない場合も姿勢推定可能とし、精度を向上させるための補助的な情報として、テクスチャを利用する。

また、構造光を用いたレンダリング手法[12]は、スポットパターンを利用していたため、遠方で疎になることから大型構造物内の小物体計測には適用できなかった。そのため、遠方においても照射光が疎になることのない、線構造光ベースの手法に拡張する。レーザ光に基づくレンダリングを行うために、

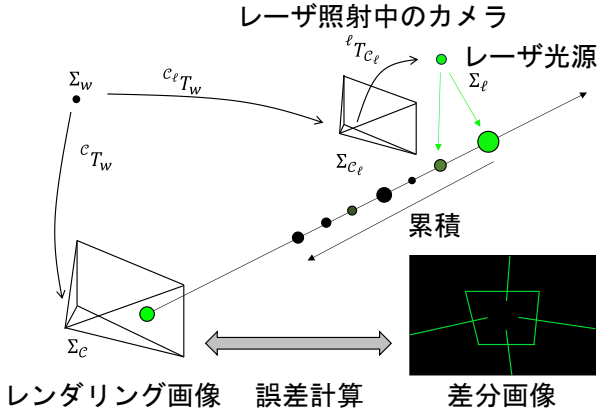


図3 カメラ間レンダリングの概要

レーザーネットワークによって各 3 次元座標における色を変化させ、同時に線構造光の位置姿勢を最適化する。

3-2. カメラ間のレンダリング

提案手法はカメラを 2 台使用し、図 3 に示すように、ある時点においてカメラ $\mathcal{C}$ からサンプリングされた 3 次元座標の色が、レーザー搭載カメラ $\mathcal{C}_\ell$ のレーザーによって変化する。これをモデル化するために、レーザー座標系へと変換を行ったのち、レーザーネットワークによって出力輝度を算出する。この入力ベクトルを算出するためのレーザー座標系における 3 次元座標 ${}^\ell \mathbf{y} \in \mathbb{R}^{3 \times 1}$ は、各カメラ座標系の 3 次元座標 ${}^c \mathbf{y} \in \mathbb{R}^{3 \times 1}$ から、次式で変換する。

$$\begin{bmatrix} {}^\ell \mathbf{y} \\ 1 \end{bmatrix} = {}^\ell \mathbf{T}_{\mathcal{C}_\ell} {}^c \mathbf{T}_{\mathcal{C}} \begin{bmatrix} {}^c \mathbf{y} \\ 1 \end{bmatrix}, \quad (1)$$

ここで、 $\mathcal{C}$ は任意のカメラ、 $\mathcal{C}_\ell$ は同時刻にレーザーを照射しているカメラ、 ${}^B \mathbf{T}_{\mathcal{A}} \in \mathbb{R}^{3 \times 3}$ は座標系 $\mathcal{A}$ から座標系 $\mathcal{B}$ へと変換する、オイラー角と並進ベクトルから算出された座標系変換行列である。変換後の 3 次元方向ベクトルに基づき、レーザー照射光分の色を加算を行う。レーザーの照射輝度の算出には、ニューラルネットワークを使用することとし、このレーザーネットワーク f_ℓ への入力を単位方向ベクトルとする。続いて、逆二乗の法則と反射強度、入射角に基づき、出力輝度が計算される。各レーザーにおける反射は次式で計算される。

$$\mathbf{c}_\ell = \frac{f_\ell({}^\ell \mathbf{y} / \|{}^\ell \mathbf{y}\|)}{\|{}^w \mathbf{y}_\ell - {}^w \mathbf{y}\|^2} ({}^w \mathbf{n} \cdot {}^w \boldsymbol{\omega}_\ell) f_r({}^w \mathbf{y}, {}^w \boldsymbol{\omega}_o, {}^w \boldsymbol{\omega}_\ell) \quad (2)$$

ここで、 ℓ は全周レーザーのインデックス、 L は全周レーザーの数、 ${}^w \mathbf{n} \in \mathbb{R}^{3 \times 1}$ は ${}^w \mathbf{y}$ における法線ベクトル、 ${}^w \boldsymbol{\omega}_\ell \in \mathbb{R}^{3 \times 1}$ は ${}^w \mathbf{y}$ からレーザーに向かう単位方向ベクトル、 ${}^w \boldsymbol{\omega}_o \in \mathbb{R}^{3 \times 1}$ は ${}^w \mathbf{y}$ からカメラに向かう単位方向ベクトル、 $f_r({}^w \mathbf{y}, {}^w \boldsymbol{\omega}_o, {}^w \boldsymbol{\omega}_\ell) \in [0, 1]$ は ${}^w \mathbf{y}$ における反射率である。ここで、

${}^w \mathbf{n}, f_r({}^w \mathbf{y}, {}^w \boldsymbol{\omega}_o, {}^w \boldsymbol{\omega}_\ell)$ はネットワークから出力される。レーザーからの色 $\mathbf{c}_\ell$ と環境中の色 $\mathbf{c}_w$ を足し合わせ、次式で各 3 次元座標における色を決定する。

$$\mathbf{c} = \mathbf{c}_w + \sum_{\ell=1}^L \mathbf{c}_\ell, \quad (3)$$

ここで、 L は全周レーザーの数である。

各 3 次元座標に対して算出される色 $\mathbf{c}$ を確率密度に基づき累積することで、レンダリングを行う。

3-3. 構造光を考慮した最適化

レンダリングにおいて、カメラ・レーザーの位置姿勢と NeRF ネットワークを同時に最適化する。最適化における誤差関数として、先行研究 [12] で設定された垂線に対する誤差 $\mathcal{L}_p^{\text{normal}}$ と $\mathcal{L}_o^{\text{normal}}$ を利用する。本研究ではカメラ同士の位置姿勢が遠い状態から可視光のレーザーの最適化を行うため、カメラ同士の位置姿勢を RGB 誤差のみからでは行うことが困難である。そのため、環境色を含めた輝度値に対する MSE 誤差 $\mathcal{L}^{\text{RGB}}$ に加え、画像差分によって求めたレーザー光に対する MSE 誤差 $\mathcal{L}_\ell^{\text{RGB}}$ を算出する。この誤差関数により、環境中のテクスチャへの局所的収束を防ぎながら、環境中の色を含めた最適化を可能とする。また、全周レーザーは平面状の形状をしている。そのため、これを誤差関数で表現するため、 xyz 平面からの標準偏差を誤差関数 $\mathcal{L}_f$ として算出した。

以上の誤差を最小化するように、テクスチャと構造光を同時に考慮した、カメラ・レーザーの位置姿勢、及び NeRF、レーザーネットワークの最適化を Adam [13] により行う。最適化が完了したのち、深度を確率密度関数に基づき計算することができる。この深度を構造物の形状評価へと活用することが可能となる。

4. 計測実験

4-1. 方法

レンダリングに基づく最適化により深度推定を行う提案システムの有効性を検証するために、屋内廊下における深度推定可否の検証実験を行った。

計測装置として、Nikon Z7II を 2 台用いた。各カメラから取得された画像を図 4 に示す。片方のカメラにはアルミフレームで墨出し器を固定した。墨出し器として 4 面の全周レーザーからなるものを使用した。カメラの光線ベクトルに対して垂直なレーザー平面とカメラの距離は、レーザーを視野に収めるために、約 2 m に固定した。使用する画像の解像度は、2064 × 1376 に固定し、カメラの内部パラメータはコーナーの数が 9 × 6 のチェスボードを用

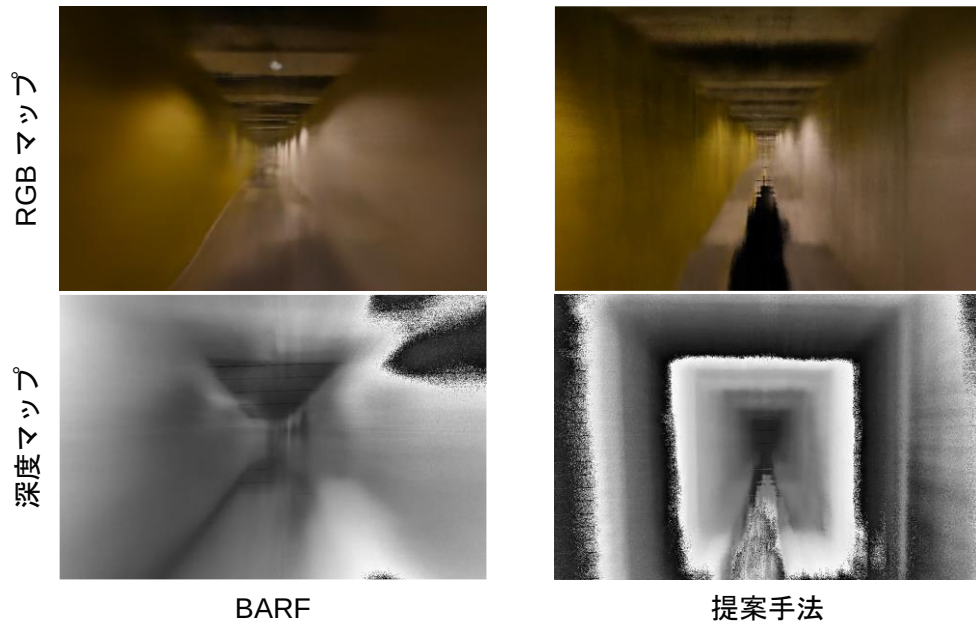


(a) レーザ搭載カメラ画像



(b) 基準カメラ画像

図 4 撮影された画像の一例



BARF

提案手法

図 5 レンダリング結果

い、透視投影モデルに基づき推定した。

全 10 地点（レーザ搭載カメラ 5 地点，基準点カメラ 5 地点）で画像を撮影した。レーザ搭載カメラ，及び基準点カメラ画像上には装置が RGB 値として現れるため，レンダリングに誤差を生じる。そのため，誤差関数を計算する際，装置部分をマスクすることでレンダリングにおいて装置が含まれることを防止した。姿勢推定の初期値として，およそ初期のカメラ同士の距離である 3 m に固定した。

4-2. 深度推定結果

レンダリング結果を図 5 に示す。比較手法の BARF はレーザ光 OFF の画像のみを使用しており，環境中のテクスチャのみを使用した最適化である。BARF は提案手法に比べ，より色をなめらかに出力した。提案手法は色にまばらさを含むものの，廊下の直線的な特徴をより正確に出力した。これらのことから，提案手法は RGB の最適化において課題は残されているものの，より構造的な色変化を表現可能であった。深度マップについては，BARF によって算出した場合，廊下の形状を認識できな

い結果となった。一方，提案手法は屋内廊下の形状に即した深度を推定した。このことから，提案手法は，レーザ光を活用したレンダリングにより，高精度な深度を推定可能であることが示された。

レンダリングに影響を及ぼすカメラ位置姿勢推定の結果を図 6 に示す。図 6 では，原点 (0, 0) に基準点カメラを固定した状態で，レーザ搭載カメラの位置姿勢を推定した様子を示している。テクスチャのみを使用した BARF は限られたテクスチャの中で姿勢を推定するため， xy 軸方向に誤差が大きく推定された。これは，今回の計測対象が奥行き方向に似たパターンが続く屋内廊下であり，ローカルミニマが存在したためであると考えられる。一方，提案手法はレーザ光を手掛かりに姿勢推定を行ったため，廊下内での装置進行方向に沿った位置姿勢を推定した。つまり，提案手法はテクスチャが存在しない場合も正解に近い位置姿勢へと収束するだけでなく，テクスチャの乏しさに起因するローカルミニマへの収束を回避するためにも有効であった。

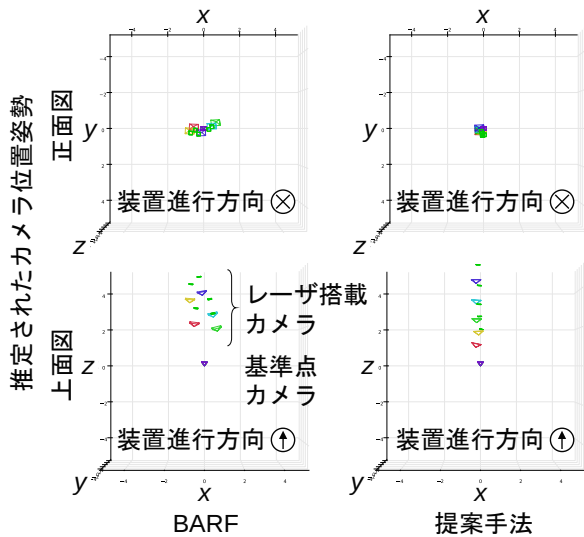


図 6 姿勢推定結果

4-3. 考察

本研究では、全周レーザを線構造光として利用し、テクスチャの乏しい環境で NeRF の最適化を行った。提案手法は線構造光により xyz 軸方向の位置姿勢の推定を可能とすることで、テクスチャを前提とせず姿勢推定可能となった。その結果として、より実際の形状と近い深度マップを算出した。提案手法は僅かなテクスチャを考慮することができるため、[5] で報告されている UAV を使用して計測した場合のドリフト誤差低減にも活用できる可能性がある。

先行手法 [3,5] では全周レーザを 1 面分利用した一方、提案手法は 4 面の全周レーザを使用した。これにより、先行手法で報告されている楕円形断面構造物に対するロール角への不定性は、レーザ光を垂直と水平方向に追加照射することで解消された。今後、楕円形上の物体に対する有効性の検証が期待される。

以上のように潜在的な有用性がある一方、今回のシステムは多くのパラメータを同時に最適化するため、点群マッチングに基づく先行手法 [3] に比べ、計算時間の大きさやハイパーパラメータによる最適化精度の変化が課題となっている。最適化計算を高速化すると同時に多様な環境で機能するパラメータを探索することで、より実用的なシステムの構築が可能となる。

5. 結論

本研究では、テクスチャの乏しい屋内構造物に対し、構造光に基づく深度推定手法を提案した。提案手法は、構造光のレンダリングを通じ実際に近い位置姿勢を推定し、その結果として、テクスチャのみを利用する手法と比べ高精度な深度を算出した。本手法は、カメラのみをセンサとして用いた

め、移動ロボットへと搭載して 3 次元計測を行うことが可能である。

今後は、動画を用いた高密度な光切断法に対する実装と移動ロボットへの搭載を行う。

謝辞

本研究の一部は、JSPS 特別研究員奨励費 24KJ0704、及び JSPS 科研費 22K18826 の援助を受けた。

参考文献

- [1] Atsushi Yamashita, Kenki Matsui, Ryosuke Kawanishi, Toru Kaneko, Taro Murakami, Hayato Omori, Taro Nakamura, Hajime Asama, "Self-localization and 3-D model construction of pipe by earthworm robot equipped with omni-directional rangefinder," in *Proceedings of the 2011 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 1017–1023, 2011.
- [2] Hiroshi Higuchi, Hiromitsu Fujii, Atsushi Taniguchi, Masahiro Watanabe, Atsushi Yamashita, Hajime Asama, "3D measurement of large structure by multiple cameras and a ring laser," *Journal of Robotics and Mechatronics*, vol. 31, no. 2, pp. 251–262, 2019.
- [3] Takuya Igaue, Toko Hayamizu and Hiroshi Higuchi, Mikihiro Ikura, Kenichi Yoshida, Satoshi Yamanaka, Takashi Yamaguchi, Hajime Asama, and Atsushi Yamashita: "Cooperative 3D tunnel measurement based on 2D–3D registration of omnidirectional laser light," *Journal of Field Robotics*, vol. 40, no. 8, pp. 2042–2056, 2023.
- [4] Yadong Xue, Sen Zhang, Mingliang Zhou, Hehua Zhu, "Novel SfM-DLT method for metro tunnel 3D reconstruction and visualization," *Underground Space*, vol. 6, no. 2, pp. 134–141, 2021.
- [5] Takuya Igaue, Toko Hayamizu, Kenichi Yoshida, Satoshi Yamanaka, Hajime Asama and Atsushi Yamashita, "Pose adjustment considering linear cross-section continuity of line-structured light-based 3D tunnel measurements," in *Proceedings of the 2023 IEEE/SICE International Symposium on System Integration (SII)*, pp. 51–57, 2023.
- [6] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, Simon Lucey, "BARF: Bundle-adjusting neural radiance fields," in *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5741–5751, 2021.
- [7] Yu Chen, Gim Hee Lee, "DBARF: Deep bundle-adjusting generalizable neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24–34, 2023.

- [8] Behzad V. Farahani, Francisco Barros, Pedro J. Sousa, Pedro P. Cacciari, Paulo J. Tavares, Marcos M. Futai, Pedro Moreira, "A coupled 3D laser scanning and digital image correlation system for geometry acquisition and deformation monitoring of a railway tunnel," *Tunnelling and Underground Space Technology*, vol. 91, no. 102995, 2019.
- [9] Vassilis Gikas, "Three-dimensional laser scanning for geometry documentation and construction management of highway tunnels during excavation," *Sensors*, vol. 12, no. 8, pp. 11249–11270, 2012.
- [10] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T. Barron, Pratul P. Srinivasan, "Ref-NeRF: Structured view-dependent appearance for neural radiance fields," in *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5481–5490, 2022.
- [11] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
- [12] Aarrushi Shandilya, Benjamin Attal, Christian Richardt, James Tompkin, Matthew O'toole, "Neural fields for structured lighting," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3512–3522, 2023.
- [13] Diederik P. Kingma, Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980*, 2014.