

人の意思と時間を錯覚により制御する遠隔操作手法

—第三報：Neural Radiance Fieldsを用いた未知環境への適用可能性検証—

Teleoperation Method by Illusion of Human Intention and Time
-Verification of applicability to unknown environments using Neural Radiance Fields-

○ 青木 惇季 ((株)リコー, 九州大学) 佐々木 史紘 ((株)リコー)
正 山科 亮太 ((株)リコー) 正 倉爪 亮 (九州大学)

Junki AOKI, Ricoh Co., Ltd. and Kyushu University, junki.aoki@jp.ricoh.com
Fumihiko SASAKI, Ricoh Co., Ltd.
Ryota YAMASHINA, Ricoh Co., Ltd.
Ryo KURAZUME, Kyushu University

Illusory control (IC) that is teleoperation method we proposed was based on the premise that the virtual environment is prepared in advance. Therefore, the mobile robot using IC can function only in known environment. This paper presents "Instant illusory control" that no need to prepare the virtual environment in advance. The proposed robot system can instantly create a virtual environment from real images of the robot in motion using Instant NGP as one of methods in Neural Radiance Fields. Additionally, using 360-degree images, the environment around the robot can be comprehensively virtualized without demanding on the camera poses. Furthermore, using the depth estimation result of the images in advance for the optimization of the density of neural radiance fields, the reconstruction accuracy at unknown poses is ensured as well. Using this virtual environment instantly created, teleoperation by transition between real and virtual environments is expected to function in unknown environment. The experimental result shows that the proposed system can achieve a consistent back and forth between virtual and real spaces in mobile robot teleoperation.

Key Words: Human Robot Interaction, Teleoperation, Illusory Control, Neural Radiance Fields

1 はじめに

近年の労働力不足や在宅勤務の増加を背景に、遠隔地からロボットを操縦する遠隔移動ロボットの活用が盛んである。

遠隔地から移動型ロボットを操縦する際、通常は人間の操縦に加えて、ロボットに搭載したセンサなどにより障害物を検出し、ロボットが自律的に回避、あるいは停止するなどの安全システムが備えられている。この人間の操縦とロボットによる自律性を共有しながら移動タスクを効率的に達成する仕組みは Shared control と呼ばれる。しかし、実際にカメラ画像を見ながらロボットを操縦してみると、操縦者が気付かない程度の障害物にロボットが反応し、操縦者は自分が出した指令通りにロボットが移動しない、あるいはロボットが予想外の動きをするなど、強いストレスを感じる場合がある。これはロボットが出す移動指令と人間が出す移動指令が一致していないことにより発生し、操縦者のロボットに対する受容性（システムに対する信頼感）を低下させる要因につながる。

筆者らはこの問題に対処すべく、Illusory control (IC) という新たな遠隔操作手法を提案してきた [1][2][3]。これは、現実空間と仮想空間の切り替えによって移動ロボットの遠隔操作を行うシステムである。操作開始時には、オペレータは現実空間のロボット（現実ロボット）から得られる映像を見ながら現実ロボットを操作する。オペレータの操作によって障害物への衝突が予測される場合には、オペレータへのフィードバックを仮想空間のロボット（仮想ロボット）から得られる映像に切り替え、同時に操作対象を仮想ロボットに切り替える。一方で、現実ロボットはオペレータが操作する仮想ロボットに追従するように、障害物を避けながら安全に自律移動を行う。このような構成を取ることで、オペレータへは意思通りに操作できているかのようにフィードバックしているため、障害物を避けるための煩わしい操作を排除しつつ、安全な走行を保証できる。

しかし、このシステム構成には、仮想空間の事前準備が必要という問題がある。事前に移動対象の場所を訪問し、3D スキャナーを用いて移動環境のセンシングを行い、アピランスの調整などの後処理を行って初めてシステムを動作させることができる。そのため、既知の環境でしか IC の手法を適用することができな

かった。遠隔操作ロボットは、災害時対応をはじめとして、事前にその環境を訪れることが難しいアプリケーションが存在する。そのため、既知環境でしかシステムを動作させることができない問題は、IC の手法の適用範囲を大きく限定してしまう。

本稿では、IC における仮想空間の事前準備を不要とする新たな手法を提案する。具体的には Neural Radiance Fields (NeRF)[4] の短時間での学習が期待できる手法である Instant NGP[5] を用いて、仮想空間の即時構築を行う。Instant NGP によって構築した仮想空間と、現実空間の映像や操作対象ロボットの切り替えを行いながら、IC による遠隔操作を行えるシステム Instant illusory control (Instant IC) を提案する。さらに、未知姿勢でのレンダリング精度を上げるための画像の事前深度推定や、現実空間の幾何と仮想空間の幾何の間の一貫性を上げるためのスケールリング処理を追加することで、現実空間と仮想空間の幾何的な一貫性を向上させる手法を実装した。

2 章では提案手法について、3 章では提案手法を適用したロボットシステムの結合動作実験について説明し、4 章では結論を述べる。

2 提案手法

提案手法のデータフローを Fig. 1 に示す。ただし、ここで示すデータフローは、仮想空間の構築とレンダリングに関わる部分のみである。全体ロボットのモジュール構成については [3] を参照されたい。

2.1 IC の基本動作

筆者らがこれまでに構築してきた IC を用いた移動ロボットの遠隔操作フローについて説明する。ここで説明するのは概要であるため、詳細については [3] を参照されたい。

まず、オペレータは、現実ロボットのカメラ映像を見て、現実ロボットを制御対象として操作を開始する。現実ロボットはオペレータからの操作指令を受け、その指令に基づいて将来の軌道を計算する。計算された未来の軌道が障害物に到達するかどうかを判断し、到達する場合は制御対象を現実ロボットから仮想ロボットに切り換える。このとき、システムは現実ロボットの位置情報

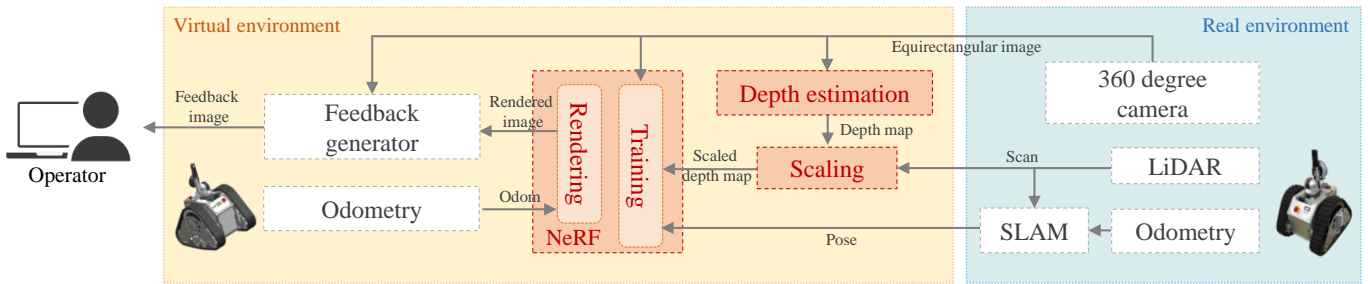


Fig.1 Image dataflow in Instant IC.

をもとに、仮想ロボットを同じ位置に移動させ、オペレータに見せる映像を仮想空間に切り替え、制御対象も仮想ロボットに切り替える。オペレータが仮想ロボットを操作している間、現実ロボットは仮想ロボットの位置・姿勢をサブゴールとして自律移動する。仮想ロボットの位置に現実空間では障害物がある場合は、仮想ロボットの将来の軌道を計算し、軌道上のコストを求め、コストが一定以下となる地点をサブゴールとして設定する。ロボットシステムは、仮想ロボットと現実ロボットそれぞれの位置情報を定期的に取得し、それぞれの姿勢の差が閾値以下になると、オペレータへのフィードバック映像を仮想ロボットから現実ロボットへ切り替え、制御対象を仮想ロボットから現実ロボットへ切り替える。

本稿では、このフローの中で用いる仮想空間の部分についての改善を行った。具体的には、事前準備が必要であった仮想空間を、ロボットの操作中に得られる画像・姿勢情報を用いて即時構築することで、事前準備の要らないシステムを提案する。

2.2 NeRF による仮想空間構築

NeRF を用いた仮想空間構築の工程は大きく 2 つに分けられる。1 つ目は、操作中のロボットから得られる現実空間の画像データを収集し、NeRF に与えるデータを作成して学習を行う工程である。2 つ目は、学習結果を用いて、シミュレータ上のロボットの姿勢を NeRF に入力して、その姿勢から見える映像をレンダリングする工程である。

ここで、NeRF 技術は Instant NGP[5] を用いた。採用理由は、短時間での学習収束が期待できるため、レンダリング画質を調節すれば 10fps 程度のリアルタイム描画が行えるためである。従来の筆者らの手法では、仮想空間の事前準備が必要であったため、現実空間と仮想空間の作成時点の時間的な差分が大きかった。これに対して、提案手法を用いると、現実空間と、仮想空間の作成時点の時間的差分を数秒～十数秒程度に抑えることができるため、人間などの動的な障害物を除けば、環境の時間的な変化を吸収できる。

まず、操作中のロボットからデータを取得し、NeRF に与えるデータを作成する工程を説明する。操作中のロボットから取得するデータは、画像情報とロボットの姿勢情報の 2 点である。これらの情報は Robot Operating System (ROS) の通信プロトコルを用いて、仮想空間を構築する PC (仮想空間用 PC) に定期的に送信される。仮想空間用 PC では、これらの情報を受け取ると、後述する深度推定と深度スケージングの前処理を行ったうえで、NeRF に与えるデータを作成して蓄積する。

移動ロボットは、リコー製のクローラロボットを用いた。クローラロボットに搭載するセンシングデバイスとして、Ricoh Theta Z1 と Velodyne LiDAR VLP-16 を用いた。NeRF の学習に用いる画像情報は、Theta Z1 から得られる 360 度画像を、水平方向 90 度・鉛直方向 90 度で計 6 枚の Perspective 画像として変換したものである。360 度画像を用いることで、データ取得時のロボットの姿勢によって視野を限定されないように仮想空間を構築できる。NeRF の学習に用いる姿勢情報は、Map フレーム上でのロボットの姿勢である。具体的には、LiDAR とクローラロボットの Wheel odometry を用いて SLAM を行いながら Map フレームを構築し、この Map フレームに対する姿勢情報を用いた。SLAM には Gmapping アルゴリズムを用いた。

次に、レンダリング工程では、Instant NGP の学習と並行してレンダリングを行う。入力には、オペレータが操作するシミュレータ上のクローラロボットから得られる姿勢情報を用いる。オペレータには、仮想ロボットの操作中は、Instant NGP によってレンダリングされた結果がフィードバックとして提示される。

2.3 深度の事前推定

前述した画像情報と姿勢情報を NeRF に与えるのみで学習とレンダリングを行うと、未知姿勢のレンダリング精度が極めて低くなる問題がある。Instant IC のコンセプトは、事前準備が要らないことであるため、ロボットで訪れていない未知姿勢における画像情報は、事前に得られないという制約条件がある。そのため、現在のロボットから得られる画像と姿勢は用いることができるが、未来の想定位置における画像と姿勢を得ることはできない。

この問題に対し、DS-NeRF[6] では、深度に関する事前情報を与えることで、少ない画像からでも品質の良い映像をレンダリングできるアプローチを提案した。筆者らはこの DS-NeRF のアプローチをベースに、深度に関する事前情報を与えることとした。具体的には、360 度画像 (Equirectangular) に対してダイレクトに深度推定を行えるアルゴリズム SliceNet[7] を用いて、深度の事前推定をした画像を NeRF のデータセットとして与えた。SliceNet の選定理由は、360 度画像に対する深度推定が可能であることと、その中で最も短時間での深度推定が可能であることである。SliceNet は教師あり学習技術であるため、事前に深度情報が得られる環境でのデータセット構築と学習が必要である。しかし、本手法では SliceNet の筆者らによって公開されている Pre-trained model のみを用いており、今回の動作実験環境における追加学習は行っていない。

2.4 現実空間に合わせた深度スケージング

上記のように深度の事前推定を行った画像に対して、LiDAR での実測値による深度のスケージングを行った。SliceNet を用いた深度の事前推定は、Pre-trained model を用いているため、実際の動作環境でこれをそのまま用いると、深度のスケールの不一致が生じる。IC は、現実空間と仮想空間の切り替えをシームレスに行う要件があるため、実際の環境のスケールと不一致があると、現実空間と仮想空間で映像が切り替わったときに、姿勢が大きくずれたようにフィードバックされてしまう。そこで、LiDAR による実測値を用いて、事前推定した深度画像をスケージングした。具体的には、LiDAR から得られる点群情報の最大値と深度画像の最大値が一致するようなスケージングを行った。これによって、現実空間と仮想空間で映像が切り替わったときの姿勢ずれを軽減することが見込まれる。

3 実験

Instant IC による動作確認結果を Fig. 2 に示す。左列が Instant IC によるオペレータへのフィードバック映像である。②-④が NeRF によってレンダリングされた仮想空間の映像である。中央列は実際の現実ロボットが動作する様子を俯瞰で撮影した映像である。右列は実際の現実ロボットから得られる映像である。②-④において、現実ロボットから得られる映像はオペレータにはフィードバックされない。Fig. 2 の画像の環境内における廊下の右側の扉部分に顕著に現れるが、データが取得できていない箇

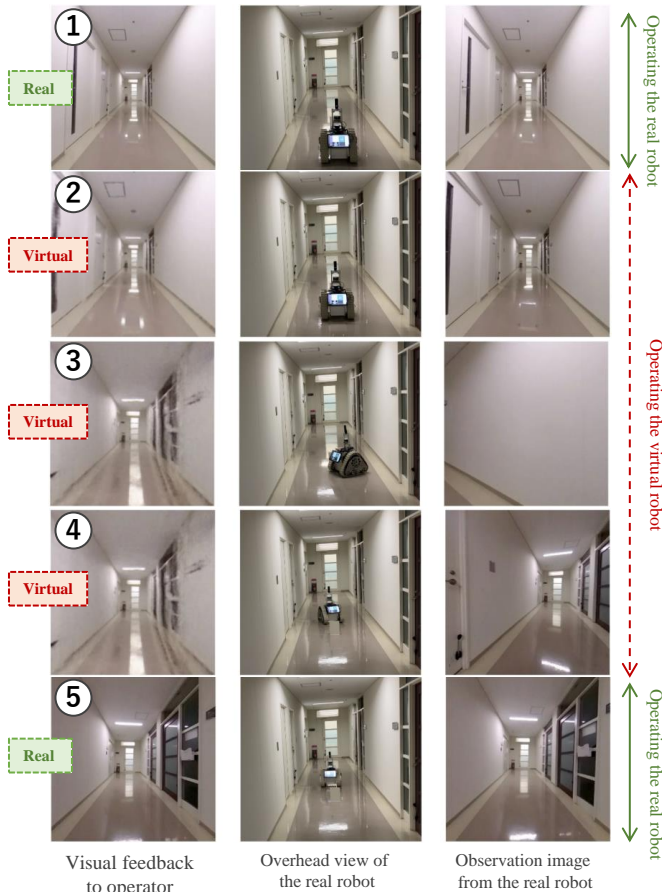


Fig.2 Verification results of the proposed system.

所におけるレンダリング精度は低くなる。このように、データが不足している箇所について、補間する技術については追加で検証の余地がある。

Fig. 3にAblation Studyの定性評価結果を示す。左がGround truth画像であり、右3列分がNeRFレンダリングの結果である。データは、初期姿勢において、1枚の360度画像のみを用いた。Ground truthと同一の姿勢に移動した状態で、NeRFによってレンダリングした結果を示している。Ours (-depth)は、NeRFに対して与えるデータに深度の事前推定の結果を与えなかった場合のレンダリング結果である。Ours (-scale)は、深度の事前推定結果は与えるが、LiDAR実測値によるスケールを行わなかった場合のレンダリング結果である。Oursは深度の事前推定結果とスケールを行ったレンダリング結果である。前述のように、データが不足している箇所についての品質低下は見られるが、Oursは最も現実空間の幾何と一致していることを確認できた。今後は、再構成精度に関する定量的評価と、屋内複数環境での仮想空間構築検証を行う。

4 結論

本稿では、ICにおける仮想空間の事前準備を不要とする新たな手法Instant ICを提案した。具体的にはNeRFのリアルタイムな学習が期待できる手法であるInstant NGPを用いて仮想空間の即時構築を行った。また、画像に対しての事前深度推定を行うことで、未知姿勢におけるレンダリング精度を向上させる手法と、LiDARによる実測値を用いて現実空間に対する深度スケールを行うことで、実際の幾何との一致度を上げる手法を提案した。これらの提案手法を用いて、Instant NGPによって構築した仮想空間と、現実空間の映像の切り替えを行いながら遠隔操

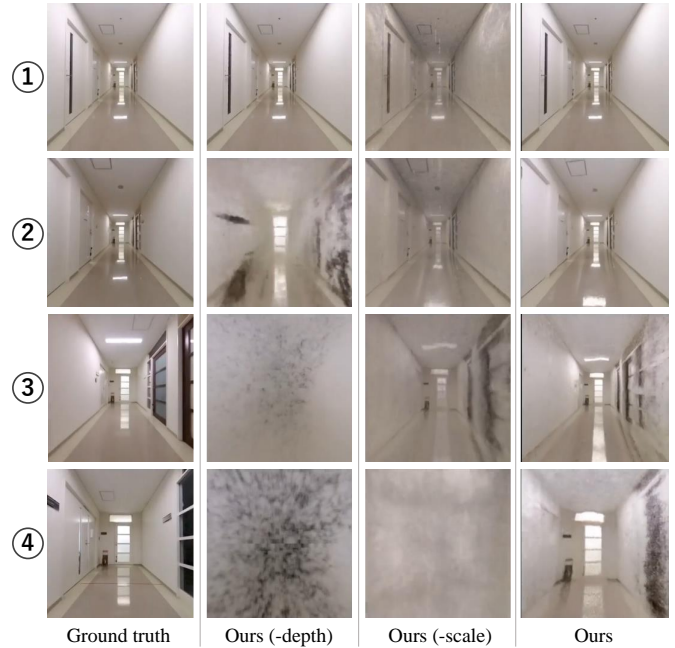


Fig.3 Verification results of the ablation study.

作を行うInstant ICシステムを構築し、クローラロボットで動作検証を行った。今後は、事前準備が必要な従来手法との再構成精度の比較や、深度の事前推定、深度スケールの詳細な効果検証を行う。

謝辞

本研究の一部は、JSPS 科研費 JP21K18701 の助成を受けたものである。

参考文献

- [1] 青木 惇季, 山科 亮太, 倉爪 亮, Illusory Control: 人の意思と時間を錯覚により制御する遠隔操作手法, ロボティクス・メカトロニクス講演会講演概要集, vol. 2021, 2A1-M01, 2021.
- [2] 青木 惇季, 山科 亮太, 倉爪 亮, 人の意思と時間を錯覚により制御する遠隔操作手法 - 第二報: 移動ロボットの実機検証 -, ロボティクス・メカトロニクス講演会講演概要集, vol. 2022, 2P2-G06, 2022.
- [3] Junki Aoki, Fumihiro Sasaki, Ryota Yamashina, Ryo Kurazume, Teleoperation by Seamless Transitions in Real and Virtual World Environments, Robotics and Autonomous Systems, 2023. DOI: 10.1016/j.robot.2023.104405.
- [4] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng, NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, The European Conference on Computer Vision (ECCV), 2020.
- [5] Müller, Thomas, Evans Alex, Schied Christoph, Keller Alexander, Instant neural graphics primitives with a multiresolution hash encoding, ACM Transactions on Graphics, vol. 41, no. 4, 2022.
- [6] Kangle Deng, Andrew Liu, Jun-Yan Zhu, Deva Ramanan, Depth-supervised NeRF: Fewer Views and Faster Training for Free, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- [7] Giovanni Pintore and Marco Agus and Eva Almansa and Jens Schneider and Enrico Gobbetti, SliceNet: deep dense depth estimation from a single indoor panorama using a slice-based representation, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11536-11545, 2021.