

Meta Quest 3 から取得される自己位置情報を用いた 3 次元再構成パイプラインの検討 Exploratory Development of a 3D Reconstruction Pipeline with Meta Quest 3 Self-Localization

太田 貴士[†] 葛岡 英明[†]
Takashi Ota Hideaki Kuzuoka

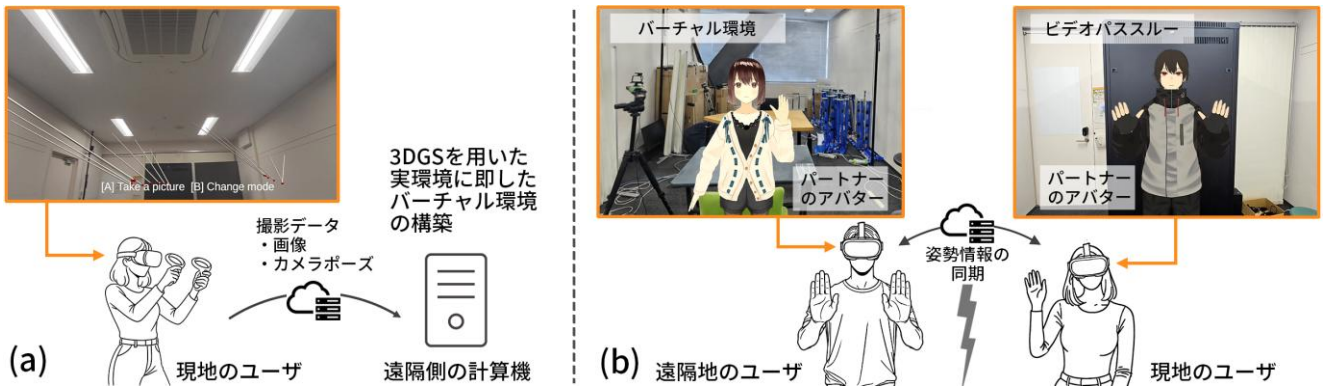


図 1 提案システムの概要. (a) 環境の撮影時および 3D 再構成時. (b) 遠隔コミュニケーション時.

1. はじめに

ビジネス、娯楽、医療など様々な領域において遠隔同期コミュニケーションはなくてはならないものとなった。遠隔同期コミュニケーションは、テキストチャットなどの非同期コミュニケーションと比較して、即時性や伝達情報量の多さで優れているため、ユーザ同士の紐帯の形成において重要であると言われており、様々な研究が行われている。VR および AR の分野においても、アバターによる身体化やヘッドマウントディスプレイ (HMD) による視覚提示を用いた、身体的かつ没入的な遠隔同期コミュニケーションの応用が研究されている。中でも、3 次元再構成により作成されたバーチャル環境を利用した遠隔コミュニケーションは、ユーザに実環境に即した環境ならではの体験を提供する。3 次元再構成が高速で、かつ準備が簡便な手法を利用することで、ユーザは外出先などから遠隔地の他者と実環境に即したバーチャル環境における遠隔コミュニケーションができるようになることが期待される。

近年、CG・CV 分野におけるアルゴリズムの発展とグラフィックスハードウェアの性能向上により、Neural Radiance Field (NeRF) や 3D Gaussian Splatting (3DGS) などの様々な 3D 再構成手法が研究されている[1]。これまでに筆者らは、これらの 3D 再構成を用いた非対称 MR 遠隔同期コミュニケーションシステムを提案してきた[2]。これらのシステムは、学習が高速な 3D 再構成によって比較的短時間で 3D 再構成を行うため、ユーザが外出先などから遠隔地の他者と実環境に即したバーチャル環境における遠隔コミュニケーションができるなどのユースケースが提案されている。

本研究では、このようなシステムにおける 3D 再構成の待ち時間を短縮するために、Meta Quest3 (以降 Quest3) から取得される自己位置情報を用いた 3 次元再構成パイプラインを検討した。本手法は、実際の撮影時のカメラの姿勢情報を使用することで、従来では COLMAP によるカメラポーズ推定を行っていた前処理が省略できるようになり、3D 再構成の待ち時間が大幅に短縮される。

[†] 東京大学 The University of Tokyo

2. 提案システム

筆者らが提案した 3D 再構成を用いた非対称 MR 遠隔同期コミュニケーションシステム[2]では、現地のユーザと遠隔のユーザがともに、実環境の環境の手がかりを反映したバーチャル環境内でコミュニケーションできる。具体的には、現地のユーザの HMD には、ビデオパススルーによって実環境が表示され、AR の重畳表示によって遠隔のユーザのアバターが表示される。一方で、遠隔のユーザの HMD には、3DGS によって構成されたバーチャル環境と、現地のユーザのアバターが表示される。当システムの主要な特徴は以下の 2 点である：3DGS は学習が比較的高速であるため、3D 再構成の待ち時間が短い (特徴 A)。遠隔のユーザは HMD 以外の特別な機材を使用しないため、外出先での使用も期待される (特徴 B)。

本研究では、当システムの 3D 再構成の待ち時間をさらに短縮するために、Quest3 の位置情報を用いたカメラポーズ推定の機能を追加した。図 1 に本研究の提案システムの概要を示す。使用手順および処理の流れは大まかに以下の通りである：(1) 現地のユーザは実環境を専用の Quest3 アプリで撮影する。(2) 撮影データ (画像および位置情報) は、クラウドを経由して、遠隔のユーザ側に送られる。(3) 遠隔のユーザ側で 3DGS によって、3D 再構成を行う。(4) 遠隔のユーザが完成した 3D シーン的位置合わせを行う。(5) 両ユーザが HMD を装着して、アバターを操作することで前述のようなコミュニケーションができる。

3DGS を用いた 3D 再構成は、COLMAP によるカメラポーズ推定 (前処理) と、3DGS の学習の 2 段階に分けられる。COLMAP によるカメラポーズ推定は特に処理時間が画像の枚数に大きく依存するため、入力画像が増えると前処理段階に時間がかかってしまうという課題があった。本研究の貢献は、Quest3 の自己位置情報を利用した 3D 再構成パイプラインを実装し、その性能を評価したことである。この機能によって、特に入力画像の多い場合に対して、前処理段階の処理時間が大幅に短縮され、3D 再構成の待ち時間が短くなり、特徴 A がより顕著なものとなる。

3. システム実装

COLMAP を用いたカメラポーズ推定では、複数の画像の情報から、各画像のカメラポーズとカメラパラメータを推定する。具体的には各画像について、id, img_name, width (画像の幅), height (画像の高さ), position (ワールド座標におけるカメラの中心位置を表すベクトル), rotation (ワールド座標におけるカメラの回転を表す回転行列), fy (Y 軸方向の焦点距離), fx (X 軸方向の焦点距離) が推定される。よって、本研究の目的を達成するためには、これらのプロパティを Quest3 の自己位置情報で埋めることができれば良い。

本研究では、非対称 MR コミュニケーションシステムで実環境を撮影する際に Quest3 のアプリケーションにおいて自己位置情報を撮影データに付加させる機能を実装した。Quest3 は AR アプリケーションにおいて外部カメラと IMU を用いて自己位置を推定している。その自己位置情報は AR シーンにおける座標と姿勢としてアプリケーション上で取得できる。これらを用いて position, rotation を埋めることができる。続いて、本研究では実装例として Quest3 を使った構成のみを扱うが、外部カメラに劣化などが無いと仮定すれば、fy, fx は固定値として良い。最後に、width, height も既知である。よって、Quest3 の自己位置情報を用いることで、COLMAP によるカメラポーズ推定と同じ形式のカメラポーズ情報を取得することができる。

4. 実験

本実験では、提案手法である Quest3 の自己位置情報を利用した 3D 再構成パイプラインと従来の COLMAP によるカメラポーズ推定を利用した 3D 再構成パイプラインの処理時間を比較した。

4.1 実験手法

研究従事者が、専用の Quest3 のアプリを用いて、大学の一室の撮影を行った。実際に撮影された画像は全部で 480 枚であった。画像の大きさは 1280×960 であった。撮影された画像から Quest3 の自己位置情報を利用したパイプライン（提案手法条件）と COLMAP によるカメラポーズ推定を利用したパイプライン（従来手法条件）のそれぞれで 3D 再構成を実行した。提案手法条件において、fy, fx の固定値はそれぞれ 860, 860 とした。なお、本実験は以下の環境で実行された。CPU: Intel Core i9-10900X, GPU: NVIDIA RTX A6000, メモリ: 128GB, CUDA: 11.8

4.2 結果と考察

表 1 に各条件におけるパイプライン処理の所要時間を示す。まず、提案手法条件は従来手法条件と比較して、前処理の所要時間および合計所要時間が短い。これより、提案手法は、COLMAP によるカメラポーズ推定を省略することで、システム全体における 3D 再構成の待ち時間を短縮させるのに有効であることが示唆された。なお、提案手法条件では従来手法条件と比較して、3DGS の学習の所要時間が長い。これは、COLMAP では冗長な画像やカメラポーズが推定できない画像が省かれる一方で、提案手法ではそれらが省かれないためではないかと考えられる。

図 2 に提案手法条件により構築された 3D シーンのビューの例を示す。これより提案手法条件は COLMAP を使用していないが十分に高精細な 3D 再構成ができていること

表 1 パイプライン処理の所要時間の比較.

	提案手法条件	従来手法条件
前処理の所要時間	10 秒	15 分 22 秒
3DGS の学習の所要時間	14 分 31 秒	12 分 7 秒
合計所要時間	14 分 41 秒	27 分 29 秒



図 2 提案手法で構築された 3D シーンの例.

が確認できた。これは、Quest3 から取得される自己位置情報が COLMAP によるカメラポーズ推定を代替する妥当な情報として使用できることを示唆している。

5. おわりに

本研究では、Quest3 から取得される自己位置情報を用いた 3D 再構成手法を検討した。実験を通して、提案手法は従来手法と比較して、3D 再構成の待ち時間が大幅に短縮されることが確認された。

本研究には主要な制約が 2 つある。1 つ目は外部カメラの劣化などが無いものと仮定して fy, fx の値を固定値としたことである。実際には外部カメラが劣化することは珍しくない。よって、今後は本手法の機体の個体差に対する頑健性を調査したり、機体ごとに適切な値を設定する手法を検討したりする。

2 つ目は検証実験がカバーしているケースが限定的であるということである。本研究では 1 つのデータに対してのみ検証したが、本システムの有効性を明らかにするためには、さらなる検証が必要である。特に、COLMAP は処理時間が入力画像の枚数に大きく依存するため、入力画像数の異なるケースに対して検証することは有益である。

また、本研究により 3D 再構成の待ち時間が大幅に削減されることが示唆されたため、今後は、本パイプラインを著者らが提案する非対称 MR 遠隔コミュニケーションシステム[2]に導入して、ユーザスタディを行う予定である。

謝辞

本研究は立石財団研究助成の支援を受けたものです。

参考文献

- [1] Ben Fei, Jingyi Xu, Rui Zhang, Qingyuan Zhou, Weidong Yang, Ying He, "3D Gaussian Splatting as New Era: A Survey", IEEE Transaction on Visualization and Computer Graphics, doi: 10.1109/TVCG.2024.3397828 (2024).
- [2] 太田 貴士, 葛岡 英明, "Instant-NGP による実環境のモデリングおよびアバターによる身体化を用いた没入型遠隔同期コミュニケーションシステムの試作", 第 73 回復合現実感研究会, (2024).