



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

太田貴士^{*1}葛岡英明^{*1}

Abstract - 本研究では, 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コミュニケーションシステムを提案する. 同じ実環境を反映したモデル内で遠隔地のユーザとコミュニケーションを行う際には, 従来は事前にモデルを構築するのに膨大な時間が必要とされていたが, 本手法では現実的な時間内でモデルの構築ができるようになり, スムーズに遠隔地の他者にその実環境の情報を提供し, 両ユーザが没入して同一の場での体験を共有できるようになる. 本システムの用途として, 外出のできないユーザと外出のできるユーザが一緒に旅行の体験を共有することなどが想定される.

Keywords : Mixed reality, Neural radiance field, Remote communication

1 はじめに

通信技術の発展により, リモートワーク [1], オンライン教育 [2], 遠隔医療 [3] など, 様々な領域で遠隔コミュニケーションが受け入れられている. 遠隔コミュニケーションには, ユーザが同じ場所へ集まる必要がなく時間的・エネルギー的なコストが削減できることや, 海外などの離れた地域にいる人とコミュニケーションができることなどの利点がある [4]. しかし, ビデオ会議を用いたコミュニケーションにおいては, アイコンタクトやボディランゲージといった非言語的な情報が利用できないためユーザはコミュニケーションに疲れてしまいやすいという課題がある [5, 6].

この課題を解決するための手法の1つとして Virtual Reality (VR) 技術を用いた遠隔同期コミュニケーションが挙げられる. VR を用いた遠隔同期コミュニケーションシステムにおいて, ユーザは同一のバーチャル環境にアバタとして身体化されて他者とコミュニケーションするため, ボディランゲージなどを通してより多くの非言語的な情報を他者に伝達できる [7, 8]. このようにユーザがアバタとして身体化された VR 遠隔同期コミュニケーションシステムでは, 仮想環境で物体を操作したり, 動作を教示したりすることができるため, 遠隔協調作業 [9] や遠隔技能伝達 [10] にも利用される. さらに, 操作対象が実環境に存在する場合の遠隔協調作業のための手法として, Mixed Reality (MR) 技術を用いた遠隔同期コミュニケーションシステムも提案されている. MR 技術を用いた遠隔同期コミュニケーションにおいて, 遠隔地のユーザに実環境の情報を伝達する手法として複数のセンサ情報を用いて三次元モ

デルを再構成する手法 [11] や, 実環境にカメラを配置する手法 [12] などがある. しかし, 前者には三次元モデルの構築に膨大な時間がかかること, 後者には1種類の視点しか提供できないことという課題がそれぞれ存在する.

そこで, 本研究はこれらの課題を解決したシステムとして, 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を提案し, その応用可能性をユーザスタディを通して評価した.

2 関連研究

Neural radiance field (NeRF) とは, 対象物をとらえた複数の2次元画像から, 深層学習を用いてその対象物を含む3次元空間内のシーンの輝度と色を推定する3次元シーンの再構築手法である [13]. NeRF は, シーンを視点依存の放射輝度と体積密度を出力する関数と捉えて, この関数をニューラルネットワーク (NN) で表現する. 学習後の NN には各地点における放射輝度がエンコードされるため, 任意の視点からの対象物のイメージを表現できるという特徴がある. この特徴により, 自然史的な価値のある化石を3次元のデジタルデータとして保存する研究で NeRF の応用が示唆されている [14]. 他にも, 文化遺産を NeRF でエンコーディングすることで対象の文化遺産を複数の視点から観察できるようにすることができる [15]. 任意の視点からの高精細なイメージ表現が可能な NeRF だが, 原著の NeRF [13] では NN の学習に膨大な時間がかかるという課題があった. この課題を解決するために多くの派生手法が提案されているが, そのうち

^{*1} 東京大学大学院情報理工学系研究科

Multiresolution Hash Encoding という手法を導入することにより，原著の NeRF と比べて大幅に学習時間を削減することに成功した．学習対象にもよるが，Instant-NGP は概ね現実的な時間内に学習を収束させることができるため，リアルタイム性の求められる遠隔同期コミュニケーションにも活用できる．

Instant-NGP を用いた遠隔同期コミュニケーションシステムとして，SharedNeRF が提案されている [17]．このシステムは片方のユーザの手元を RGBD カメラで撮影し，Instant-NGP によって構成された手元のイメージを遠隔地のもう片方のユーザにディスプレイで表示することで，ハードウェアデザインなどの実物を介した協調作業を支援する．他方で，Instant-NGP を VR に応用する研究も行われている．Ke Li らは，Instant-NGP による描画をヘッドマウントディスプレイ（HMD）で表示するためのツールキットである Immersive-NGP を開発し [18]，粒子加速器の環境などの検査などに応用可能な Instant-NGP を用いた VR システムを開発した [19]．

本研究では，Instant-NGP は十分な GPU メモリがあれば部屋全体のように大きな対象物のシーンも構築できるということに着目した．本研究の提案手法では，現地のユーザが撮影した画像データに基づいて Instant-NGP により構成された仮想環境内に遠隔地のユーザが没入する．現地のユーザは AR モードで遠隔地のユーザは VR モードで没入し，両者はアバターとして身体化されることにより，まるで両者が現地でコミュニケーションをしているかのような体験を可能とする．同期的な体験に VR と AR という非対称な提示手法を用いる手法はいくつか提案されているが [20, 21, 22]，これらのシステムでは少なくとも片方のユーザは HMD を装着せずにコミュニケーションするため，両者のユーザが完全に没入して同じ場での体験を他者と共有することは難しい．本研究では，両者のユーザが没入して同じ場での体験を他者と共有することを可能にする手法を提案する．

3 提案手法

3.1 概要

図 1 に提案手法の概念図を示す．本システムは，実環境にいるオンサイトのユーザと遠隔地にいるリモートのユーザの同期コミュニケーションのためのシステムである．2人のユーザはともに HMD を装着するが，オンサイトのユーザは AR モードで没入し，リモートのユーザは VR モードで没入する．オンサイトのユーザのディスプレイには，実環境がビデオスルーで表示され，リモートのユーザのアバターが AR の重畳表示機能によって表示される．一方，リモートのユーザ



図 1: 提案手法の概念図

のディスプレイには，オンサイトのユーザが撮影した画像を元に Instant-NGP が構成した仮想環境と，リモートのユーザのアバターが表示される．これによって，2人のユーザは実環境を反映した環境内でパートナーのアバターとコミュニケーションができる．本システムの用途には，外出のできないユーザと外出のできるユーザと一緒に旅行の体験を共有することや，患者が住み慣れた環境である自宅をリハビリテーション環境として使用する VR 遠隔リハビリテーションなどが考えられる．

3.2 システム構成

図 2 に提案手法のシステム構成図を示す．オンサイトのユーザは専用のアプリがインストールされたスマートフォンと Quest 3 を使用し，リモートのユーザは専用の Windows アプリがインストールされた計算機と Quest 3 を使用する．まず，Instant-NGP による 3D モデル構築の処理をデータの流れに沿って説明する．オンサイトのユーザは，はじめにローカルエリアネットワーク（LAN）内での UDP 通信によって Quest 3 とスマートフォンをペアリングさせる（図中①）．そして，スマホアプリのカメラ機能を使って現地の実環境の画像を十分に撮影すると，撮影された画像は Quest アプリを経由して Google Cloud で構築されたサーバーにアップロードされる（図中②，③）．その後，リモートのユーザが専用の Windows アプリによって，Google Cloud から画像をダウンロードする（図中④）．そして，リモート側の計算機では Windows アプリによって自動的に COLMAP [23] によるカメラポーズ推定や Instant-NGP の学習が行われる．仮想環境のシーン情報は Unity のネイティブプラグインによって Quest アプリに呼び出されて（図中⑤），リモートのユーザは Quest 3 を通して Instant-NGP により構成された仮想環境に没入できる．なお，Unity のネイティブプラグインには Immersive-NGP [18] の実装を一部変更したものが使用された．

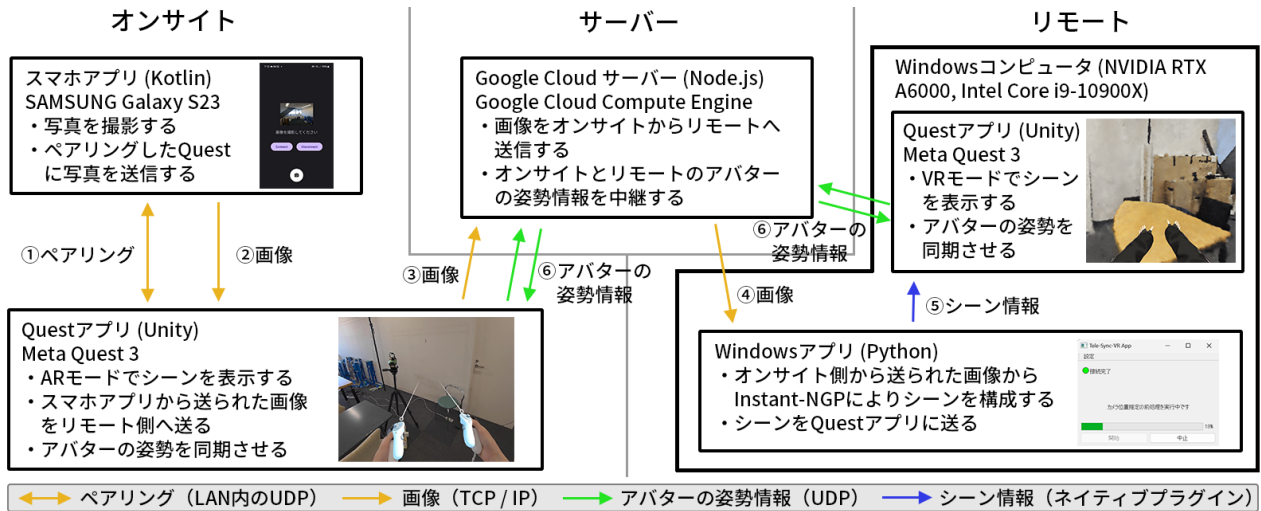


図 2: 提案手法のシステム構成図

続いて、ユーザのアバタの姿勢同期の処理について説明する。リモート側の Quest アプリは HMD およびコントローラから頭部、右手、左手の 3 点の姿勢情報（位置と回転）を取得し、それらの姿勢情報から Final IK¹によってアバターの姿勢を逆運動学的に計算した。オンサイトおよびリモートのユーザの姿勢情報は UDP 通信で Google Cloud サーバーに送信される。Google Cloud サーバーは毎秒 20 回の速度で適宜最新の姿勢情報をパートナーの Quest アプリに配信した。パートナーのアバターの姿勢も配信された 3 点の姿勢情報から Final IK により再構成された。

4 評価実験

本研究では提案手法の有効性の評価および課題の発見を目的として、大学生あるいは大学院生の実験参加者 6 名（男性 6 名，平均年齢: 23.5, 標準偏差: 0.76）を対象にパイロットスタディを実施した。

4.1 実験手法

実験参加者は 2 名のペアで一緒に実験に参加した。実験参加者らは、はじめに VR 酔いについての質問紙である Simulator Sickness Questionnaire (SSQ) [24] に回答した。その後、2 名の実験参加者らには無作為にそれぞれオンサイトのユーザ役とリモートのユーザ役が割り当てられ、別の実験室に分かれて 2 つの課題を遂行した。

課題 1 では、実験参加者らは提案システムのセットアップを行い、5 分間自由に遠隔同期コミュニケーションを体験した。セットアップの際は、両方の実験室は zoom²によって音声会話ができる状態で、実験実施者の指示を受けながら、実験参加者らが遠隔同期コミュ

ニケーションシステムのセットアップを行った。本課題では提案システムのユーザビリティの解析を目的として、セットアップから遠隔コミュニケーションの体験までの実験参加者らと実験実施者の会話が記録された。

課題 2 では、外出のできるユーザと外出のできないユーザの旅行体験の共有を想定して、実験参加者らは旅行者がよく行う動作である写真撮影を模したタスクを遂行した。オンサイトのユーザ側の実験室にはマネキンが配置されており、実験参加者らはこのマネキンが観光地にある銅像だと想定してこのマネキンと写真撮影を行った。リモートの実験参加者はパートナーのみがマネキンと映った写真と、自撮りの機能を使って 2 人で映った写真を撮影するように指示を受けた。

2 つの課題を完了したら実験参加者らは、VR 酔いについての質問紙である SSQ, 仮想環境に対するプレゼンスについての質問紙である igroup presence questionnaire (IPQ) [25], アバターの身体化に関する質問紙である Virtual Embodiment Questionnaire (VEQ) [26] に回答した。その後、個別に実験実施者による口頭での半構造化インタビューに回答した。半構造化インタビューは以下の質問リストに従って行われた。

- Q1 課題 1 におけるシステムのセットアップは難しいと感じましたか。
- Q2 課題 1 においてパートナーとコミュニケーションをする際に気を付けたことを教えてください。
- Q3 課題 2 におけるカメラを使った写真撮影は難しいと感じましたか。
- Q4 課題 2 においてパートナーと写真撮影をする際に気を付けたことを教えてください。

¹<https://assetstore.unity.com/packages/tools/animation/final-ik-14290?locale=ja-JP>

²<https://zoom.us/>

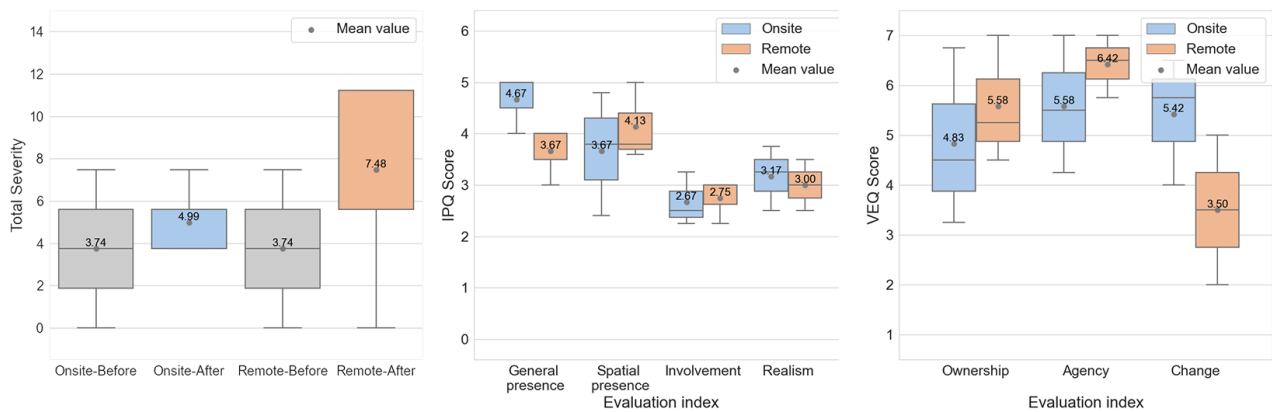


図3: (a) 課題開始前および終了後のSSQのTotal Severity. (b) Onsite条件およびRemote条件におけるIPQの各評価指標のスコア. (c) Onsite条件およびRemote条件におけるVEQの各評価指標のスコア.

Q5 このような遠隔同期コミュニケーションシステムは他には何に應用できると思いますか.

4.2 結果

図3(a)にSSQのスコアを示す. SSQのTotal severityの平均値はオンサイト条件でもリモート条件でも実験前に比べて実験終了後の方が高かった. しかし, SSQのスコアは実験終了後でも一般的な重度のVR酔いの基準値である20を超えている実験参加者はいなかったため, 本システムで重度のVR酔いは発生しなかったと言える [27]. 続いて, 3(b)にIPQのスコアを示す. オンサイト条件でもリモート条件でもGeneral presenceの平均値は3以上であった. Spatial presence, Involvement, Realismの値はオンサイト条件とリモート条件の間に目立つ違いは確認されなかったが, General presenceの平均値はリモート条件に比べてオンサイト条件の方が高かった. 続いて, 3(c)にVEQのスコアを示す. Ownershipの平均値はオンサイト条件に比べてリモート条件の方が高く, Agencyの平均値はオンサイト条件に比べてリモート条件の方が高く, Changeの評価値はリモート条件に比べてオンサイト条件の方が高かった.

図4に実験参加者のペアごとの各タスクの所要時間を示す. 課題1のセットアップ(COLMAPによるカメラ位置推定やInstant-NGPの学習も含む)にはどのペアも20分以上かかった. 課題1のコミュニケーションは約5分と時間が決まっているため, 全ペアの所要時間は約5分であった. 課題2の写真撮影方法の説明には全ペアが7分から9分程度かかった. 課題2の写真撮影にはどのペアも10分以上かかった.

課題1では, 3組全てがセットアップを完了することができた. 3組のペアは, それぞれ90枚, 71枚, 99枚の実環境の画像を撮影してInstant-NPGによって仮想環境を構築した. Q1に対して, 6名中6名がセッ

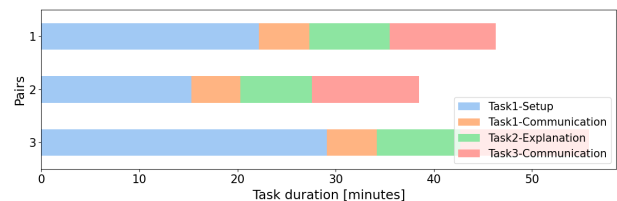


図4: 各タスクの所要時間

トアップが難しいと回答した. 特に, 高精細な仮想環境のモデルを作るために多くの実環境の画像を撮影するのが大変だったというコメントが多かった. Q2に対しては, オンサイト側のユーザが実環境の画像を撮影している間リモートのユーザはやる事が無かったことについて, パートナーにばかり負担が多くて少し申し訳なかったと回答したりリモート側の実験参加者もいた.

図5に課題2で撮影された写真の抜粋を示す. 課題2では, 3組のペアはそれぞれ6枚, 9枚, 12枚の写真撮影した. Q3に対して, 6名中4名はそんなに難しくなかったと回答し, 2名が難しくなかったと回答した. 難しいと回答した2名はリモート側の実験参加者であった. Q4に対しては, Instant-NGPによる実環境の構築が不十分な場所ではリモート側のユーザは環境情報が把握できないことについて, 環境情報が一部分からないため写真をきれいに撮影するのが難しかったというコメントが得られた.

Q5に対しては, 旅行体験の共有と遠隔会議が応用先として挙げられた. 遠隔側の参加者は3名とも構築された実環境のモデルが想像していた以上にきれいだったとコメントした. 景色の良い観光地で本システムを使ってみたいというコメントも得られた.

4.3 考察

SSQの結果より, 本システムの使用により重度のVR酔いを引き起こした実験参加者はいなかった (図



図 5: 課題 2 でリモートの実験参加者が撮影した写真

3 a). また, IPQ の General presence のスコアより, オンサイト条件でもリモート条件でも General presence が 3 以上と十分に高い没入感が生起されたことが確認された (図 3 b). そして, 全ての実験参加者のペアがタスクを完遂することができた. 以上より, 提案システムは遠隔地のユーザに対しても高い没入感を提供しながら, 現地にいるユーザと遠隔地にいるユーザの同期コミュニケーションを可能にすることが示唆された. 特に, Quest3 とスマートフォンという少ない機材のみで高品質な遠隔同期コミュニケーションを実現した本手法は, 現地にいるユーザと遠隔地にいるユーザとの旅行体験の共有といった一般消費者向けの遠隔同期コミュニケーションに適していると考えられる.

インタビューでは, リモートの実験参加者は全員 Instant-NGP により構成されたモデルの品質の高さを評価していた. 一方で, Q4 への回答で, Instant-NGP による実環境の構築が不十分な場所ではリモート側のユーザは環境情報が把握できないという課題に言及するコメントが得られたように, 本システムは 3D モデルの品質を改善する余地があると考えられる. 図 6 に, ペア 3 が撮影した 99 枚の実環境の画像から構成されたモデルと, 実験担当者が事前に撮影した 672 枚の画像から構成された 3D モデルの比較を示す. 前者では正確に位置が推定されなかったプリミティブが空中に多く残ってしまっているが, 後者ではほとんど残っていない. より高品質なモデルを構築するためには, 良質で大量の画像データが必要である. Instant-NGP の学習には, 様々な座標から様々な角度で撮影された画像が必要とされるため, 例えば高い位置から見下ろす角度や低い位置から見上げる角度での画像の撮影を促

実験参加者が撮影した99枚の画像から構成されたモデル

実験実施者が撮影した672枚の画像から構成されたモデル



両モデルの画像は学習時間が約180秒の時点で撮影. 学習には ASUS ROG Zephyrus (AMD Ryzen 9 6900HS, Geforce RTX 3070ti) を使用.

図 6: ペア 3 が撮影した 99 枚の画像から構成されたモデルと実験担当者が撮影した 672 枚の画像から構成されたモデルの比較.

すようなインストラクションを表示する機能などが必要であると考えられる. また, 本研究では屋内の観光地を想定して実験を行ったが, 屋外においてはより広い実環境を構成する必要があるため, 3D モデルの品質が低下することが想定される. よって, 今後は屋外における本システムの有効性の評価も必要とされる.

また, Q2 への回答ではスマホアプリを使って実環境の画像を大量に撮影するのが大変だというコメントが多く得られた. この課題を解決するために, 現地のユーザはスマホを持ったまま環境を歩き回り動画を撮影するだけで, その動画から Instant-NGP により実環境のモデルが構成されるような実装にしようという方法が考えられる. 他にも HMD の外部カメラから画像を撮影できるようにできれば, 必要な機材が 1 つ減りユーザビリティが向上すると考えられる. しかし, 現状では Quest 3 には外部カメラへはビデオスルー専用の API からしかアクセスできないため, 他の空間コンピューティングデバイスで実装する方が現実的かもしれない.

5 おわりに

本研究では 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を試作し, 本システムの有効性の評価および課題の発見を目的として, パイロットスタディを実施した. 評価実験の結果, 本システムは遠隔地のユーザにも高い没入感を提供しながら, 現地にいるユーザと遠隔地にいるユーザの同期コミュニケーションを可能にすることが示唆された. 一方で, 本研究で提案したシステム構成では, システムの操作性および Instant-NGP により構築される 3D モデルの精度に課題があることが分かった. 今後の研究では,

システムの実装を改善し、大規模な評価実験を行う予定である。

謝辞

本研究は、JSPS 科研費 24KJ0926 の助成を受けたものである。

参考文献

- [1] José María Barrero, Nicholas Bloom, and Steven J. Davis. The evolution of work from home. *Journal of Economic Perspectives*, 37(4):1–28, 2023.
- [2] Florence Martin, Ting Sun, and Carl D Westine. A systematic review of research on online teaching and learning from 2009 to 2018. *Computers and education*, 159:104009, 2020.
- [3] Sathyanarayanan Doraiswamy, Amit Abraham, and Ravinder et al. Mamtani. Use of telehealth during the covid-19 pandemic: scoping review. *Journal of medical Internet research*, 22(12):e24087, 2020.
- [4] Joana Fialho. *Benefits and Challenges of Remote Work*, pages 1–17. Springer International Publishing, Cham, 2022.
- [5] Sarah Morrison-Smith and Jaime Ruiz. Challenges and barriers in virtual teams: a literature review. *SN Applied Sciences*, 2(6):1–33, 2020.
- [6] Dani Paul Hove and Benjamin Watson. The Shortcomings of Video Conferencing Technology, Methods for Revealing Them, and Emerging XR Solutions. *PRESENCE: Virtual and Augmented Reality*, 31:283–305, 12 2022.
- [7] Elena Dzardanova, Vlasios Kasapakis, and Damianos et al. Gavalas. Virtual reality as a communication medium: a comparative study of forced compliance in virtual reality versus physical world. *Virtual Reality*, 26(2):737–757, 2022.
- [8] Ahsan Abdullah, Jan Kolkmeier, and Vivian et al. Lo. Videoconference and embodied vr: Communication patterns across task and medium. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2), oct 2021.
- [9] Zhenyi He, Ruofei Du, and Ken Perlin. Collabovr: A reconfigurable framework for creative collaboration in virtual reality. In *2020 IEEE ISMAR*, pages 542–554, 2020.
- [10] Hawkar Oagaz, Breawn Schoun, and Min-Hyung Choi. Performance improvement and skill transfer in table tennis through training in virtual reality. *IEEE TVCG*, 28(12):4332–4343, 2022.
- [11] Huidong Bai, Prasanth Sasikumar, and Jing et al. Yang. A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. In *Proceedings of the 2020 ACM CHI*, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.
- [12] Peng Wang, Xiaoliang Bai, and Mark et al. Billingham. 3dgam: using 3d gesture and cad models for training on mixed reality remote collaboration. *Multimedia Tools and Applications*, 80:31059–31084, 2021.
- [13] Ben Mildenhall, Pratul P. Srinivasan, and Matthew et al. Tancik. Nerf: representing scenes as neural radiance fields for view synthesis. *Commun. ACM*, 65(1):99–106, dec 2021.
- [14] Claudio Silva, Yurii Piadyk, and João et al. Rulff. Paleoscan: Low-cost easy-to-use high-volume fossil scanning. In *Proceedings of the 2024 ACM CHI*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [15] V Croce, G Caroti, L De Luca, A Piemonte, and P Véron. Neural radiance fields (nerf): Review and potential applications to digital cultural heritage. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48:453–460, 2023.
- [16] Thomas Müller, Alex Evans, and Christoph Schied et al. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):102:1–102:15, July 2022.
- [17] Mose Sakashita, Balasarayanan Thoravi Kumaravel, and Nicolai et al. Marquardt. Sharednerf: Leveraging photorealistic and view-dependent rendering for real-time and remote collaboration. In *Proceedings of the ACM CHI*, pages 1–14, 2024.
- [18] Ke Li, Tim Rolff, and Susanne et al. Schmidt. Bringing instant neural graphics primitives to immersive virtual reality. In *2023 IEEE Conference on VRW*, pages 739–740, 2023.
- [19] Ke Li, Tim Rolff, and Susanne et al. Schmidt. Interacting with neural radiance fields in immersive virtual reality. In *Extended Abstracts of the 2023 ACM CHI*, CHI EA '23, New York, NY, USA, 2023. Association for Computing Machinery.
- [20] Balasarayanan Thoravi Kumaravel, Cuong Nguyen, and Stephen et al. DiVerdi. Tranceivr: Bridging asymmetrical communication between vr users and external collaborators. In *Proceedings of the 33rd ACM UIST*, UIST '20, page 182–195, New York, NY, USA, 2020. Association for Computing Machinery.
- [21] Kohta Seki and Shinichi Fukushige. Teleghost: Asymmetric telepresence system using ar and vr avatars in a shared real space. In *2023 IEEE SMC*, pages 4550–4555, 2023.
- [22] Botao Hu, Yuchen Zhang, and Sizheng et al. Hao. Mofa: Exploring asymmetric mixed reality design strategy for co-located multiplayer between handheld and head-mounted augmented reality. In *Extended Abstracts of the 2023 ACM CHI*, CHI EA '23, New York, NY, USA, 2023. Association for Computing Machinery.
- [23] Johannes L. Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *2016 IEEE CVPR*, pages 4104–4113, 2016.
- [24] Robert S Kennedy, Norman E Lane, and Kevin S et al. Berbaum. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.
- [25] igroup.org. igroup presence questionnaire (ipq), 2001. <https://www.igroup.org/pq/ipq/index.php>.
- [26] Daniel Roth and Marc Erich Latoschik. Construction of the virtual embodiment questionnaire (veq). *IEEE TVCG*, 26(12):3546–3556, 2020.
- [27] Kay M Stanney, Robert S Kennedy, and Julie M Drexler. Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, volume 41, pages 1138–1142. SAGE Publications Sage CA: Los Angeles, CA, 1997.