



遠隔全方位マルチ視聴可能な XR メタバースシステム に関する研究

菊地 勇輔¹⁾, 長井 超慧¹⁾, ヤェム ヴィボル²⁾, 池井 寧³⁾

1) 東京都立大学 (〒 191-0065 東京都日野市旭が丘 6-6, kikuchi@vr.sd.tmu.ac.jp, y-nagai@tmu.ac.jp)

2) 筑波大学 (〒 305-8577 茨城県つくば市天王台 1-1-1, yem@iit.tsukuba.ac.jp)

3) 東京大学 (〒 113-8656 東京都文京区本郷 7-3-1, ikei@vr.u-tokyo.ac.jp)

概要: 遠隔地のライブ映像と記録映像の両者を体験でき、かつ多人数で共有可能な XR (拡張現実感) メタバースシステムを提案する。本システムのオペレータは全方位立体視を用いて遠隔地でアバターロボットを移動しながら現地と対話可能である。複数の実際的な通信経路の条件下での遅延評価を示すと共に、操作者の視点を多数の利用者が共有して任意の方向を観察できるための条件を示した。

キーワード: テレプレゼンス, メタバース, 全方位視覚共有, ライブ立体視

1. はじめに

本研究では、3D の仮想空間と連続的に利用することが可能な遠隔の実空間を含む XR (extended reality) 型のメタバースシステムのプロトタイプを構築し、その特性を示すことを目的とする。本システムは、遠隔地に配置可能な全方位ライブ立体視型のカメラシステム (TwinCam)、遠隔の実世界での移動を伴う作業を指向したアバターロボット、身体感覚フィードバックを含むオペレータシステム、多人数がリアルタイムのコミュニケーションに参加可能な配信システムから構成されている。

本構成は遠隔地に配置されたアバターロボットで作業を行う操作者の視点を、多数の参加者が全天周映像として共有可能としている。これは操作者が先導して、実空間での共同作業、グループ教育活動、専門分野が異なる参加者が独自視線の視察で評価する活動や、ガイドの案内で自由な能動的観察が可能なツアーなどに適合する設計である。

2. XR メタバースシステムの構成

2.1 全体構成

本研究で構築した XR メタバースシステムの全体構成を Fig. 1 に示す。システムの主たる操作者 (オペレータ) は、遠隔の実空間または記録された映像空間への入口を配置した Tele-XR ポータルから任意の空間を選択して体験することができる。遠隔の実空間では、アバターロボットの運動を操作し全方位のリアルタイム両眼立体視を映像を見ながら現地を移動することが可能である。記録映像は実空間で撮影された全天周映像であり、再生中に任意方向を見て追体験することができる。これらの体験の際、視点移動時に歩行感覚の身体感覚フィードバックが与えられる。この一人称視点映像は、多数の視点共有者 (非操作者) にもクラウドの映像配信機能から供給される。この映像は多種類のデバイス上で見ることが可能である。

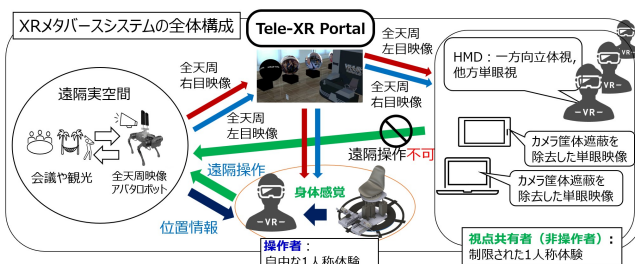


図 1: XR Metaverse System.

遠隔地のアバターロボットは複数設置され、任意の時点で視点を変更できる。過去に撮影した記録済み全方位映像は、多数の遠隔実空間への没入を提供しており、それらの間の自由な視点移動が可能である。これにより、操作者の一人称体験を多人数で共有し、準一人称の体験を得ることができる。

2.2 Tele-XR Portal

XR メタバースシステムの多種空間の体験への入口として、Fig. 2 の Tele-XR Portal を構築した。これは、3D 空間シミュレーションソフトウェア (Unity) による VR (virtual reality) 空間であり、多人数が接続するメタバースである。操作者または多種の体験を供給するガイド CG アバターは、この Tele-XR Portal の中で、対象となる空間へ視点を移動して利用する。遠隔の実空間と記録映像空間は、映像球で表されている。Portal 中の視点は、次節の TwinCam と等価なオブジェクトを用いて、両眼の映像を生成している。

遠隔の実空間の映像球に入った場合、操作者は Head Mounted Display (HMD) で遠隔地のアバターロボットの視点位置で現場を立体視し、その場にいる作業者と対話することができる。アバターロボットには操作者の 3D 頭部 CG モデルを重畳表示し、観察者の顔を見て双方向に対話することを可能とする構成となっている。

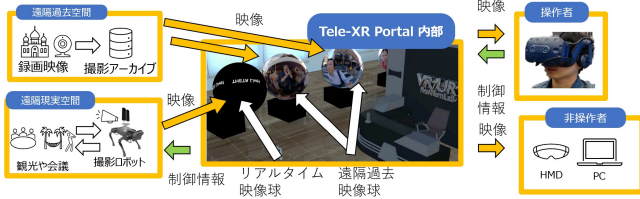


図 2: Configuration of Tele-XR Portal system.

2.3 複数のアバターロボットと全方位ライブ立体視カメラシステム (TwinCam)

アバターロボットは、遠隔の実空間で作業を行うことを想定したテレイグジスタンス [1] 型指向のロボットであり、リアルタイム全方位立体視システム (TwinCam)[2] を搭載しており、操作者および視点共有者への映像配信も行。TwinCam は、全天球カメラ (THETA Z1, Ricoh 社) を 2 台用いて、全方位の正しいステレオ画像を取得し、ステレオ音響と共に遠隔の体験者に臨場感の高いテレプレゼンス体験を提供する視聴覚システムである。

遠隔地に配置されるアバターロボットの構成を Fig. 3 に示す。操作者は、HMD (Quest2) の controller でどちらかのロボットの視点に切り替え、前進/後進、左右旋回を操縦する。操作コマンドは、映像配信のためのクラウドサーバを介して 5G 等の高速回線で接続されたロボットに着信する。

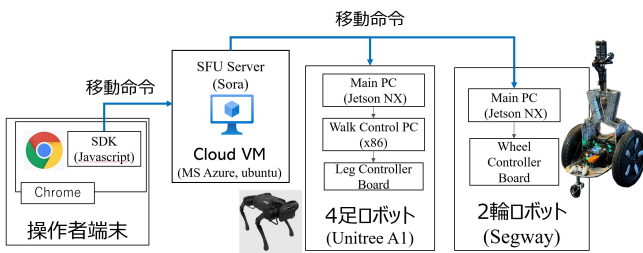


図 3: Control system of two avatar robots.

2.4 視点共有者への映像提示

アバターロボットのカメラからの立体視映像は、操作者が見ている方位の視差を有している。多数の視点共有者は、操作者が注視する前進方向などと同じ方向を見ている場合は、立体視が可能である。操作者と異なる方向を見る場合は視差が異なるため、その場合には単眼視映像に切り替えて提示する。この単眼視映像への切り替えは、Fig. 4 に示すように、視点共有者 (非操作ユーザ) が操作者の視線方位と一定以上の偏差を持つ場合に用いられるが、実際には操作者の視線方位と TwinCam の機構的視線方位と視点共有者の視線方位の 3 者の関係から切り替え時点が決定される。この切り替えを行うことにより、左右どちらかの視野に反対側のカメラの筐体が入り視野を遮蔽することを回避することも可能となる。この機能は、両眼立体視用に 2CH の映像 stream を受信している端末 (HMD または smartphone) における処理として実行される。

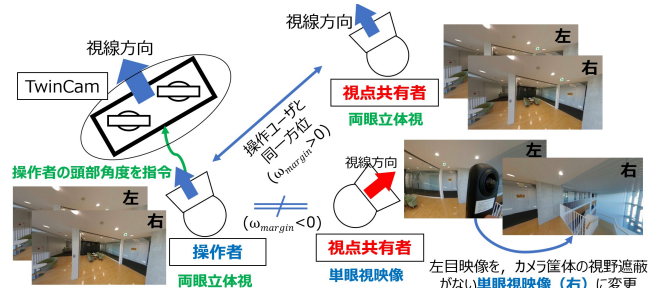


図 4: 3D/2D viewpoint-sharing streaming system.

3. システムの機能の実証

3.1 国際通信時のライブ立体視映像

遠隔立体映像体験の実証として、アバターロボットシステムのカメラ機能を米国に置き、高品質な全方位映像が日本から視認可能であることを実証した。ニューヨークのワールドトレードセンター駅の Oculus に全方位立体視映像送信装置 (Handy TwinCam ユニット) を持ち込んで歩行しながら映像を送信した。通信は 5G モバイルルータを用いて現地の 5G WAN でからインターネットに接続した。遅延を低減するため 2 台のルータを用いた。映像体験は、東京都西部 (青梅市) の参加者宅から光回線で接続して行った。

3.2 実験結果

東京で受信した streaming の視野映像を Fig. 5 左に示す。右側に通信状況のデータを示す。これより、全方位 4K 立体視映像 30 fps が実現されていることが分かる。

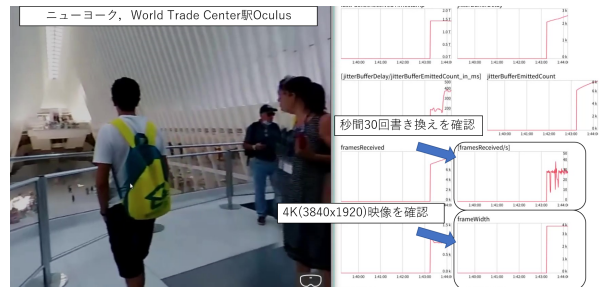


図 5: First-person view of NY Oculus from Tokyo.

3.3 映像遅延時間の検証

3.3.1 実験目的と手法

遠隔地のリアルタイム体験で避けられない通信遅延を評価した。映像を送出する TwinCam カメラは全天周画像データを与えるため、受信側で頭部回転した際の観察方位の変化に対して、暫定的に現画像の中で更新映像を提供可能である。これにより、体験者の映像遅延の知覚確率が大幅に下がるが、その際の立体視条件としては間距離 (IPD) の低下が発生している。また、通常ならば視野外となるカメラ筐体が視野の一部を遮蔽する事象が生起する。

まず実際的な利用条件となるモバイル回線ネットワークから送信時の映像の遅延を計測した。通信経路を Fig. 6 に示す。受信場所は東京都立大学の日野キャンパス 5-201a 室として有線 LAN で接続した測定用 PC で受信した。送信側

場所は日野市西部エリアの公道とした。送信にはモバイル 5G ルータを用い、非混雑時⑤、混雑時⑥、LTE モード④を計測した。さらに比較のため、受信場所と同じ 5-201a 室で有線 LAN に直結した最小遅延送信条件③、wifi で有線 LAN に接続する条件②、学内の Local 5G(sub6) ルータによる送信条件⑦で送信した。①は通信を含まず、カメラ映像を取り込んだ制御器 (ubuntu) 内だけの条件である。

遅延検出のための映像は、送信側のレンズに密着させた LED のパルスの点滅を用いた。受信側のモニタ画面のパルスの点滅をフォトダイオードで検出した。時刻の標準として、両側で GPS パルスで同期した UTC 等価時計を用いた。

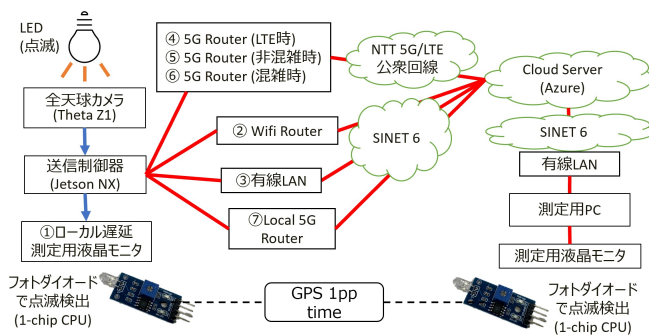


図 6: Image data paths for latency measurement.

3.3.2 実験結果と考察

実験結果を Fig. 7 に示す。THETA カメラの映像を制御器 (NX) で取得して IP 通信無しでモニタに直接描画する際の遅延時間①は 282 ms であり、その大半はカメラ内での映像圧縮と USB 転送である。映像を有線 LAN でクラウドサーバ (azure) に送信し、有線 LAN でダウンロードする③の条件で 368 ms なので、有線ネットワーク経由の遅延部分が計測時で 86 ms であったと考えられる。

遠隔地での条件例として、現地の構内 LAN に wifi でカメラを接続する場合では、②の 459 ms の遅延が該当する。ただし wifi では現地移動の制限は大きい。混雑がない 5G 回線では⑤の 458 ms で wifi とほぼ同等となり、これは近い将来の 5G 回線による屋外使用の目安とみられる。

現在、使用エリアが充実している LTE 回線 (advanced)④では 781 ms の遅延で、混雑時の 5G 回線⑥の 727 ms と同程度である。ただし LTE の方が Jitter が小さいため、WebRTC 通信では画質が優れる可能性がある。Local 5G 条件⑦は、都立大学内の sub6 によるが、安定性能がえられ

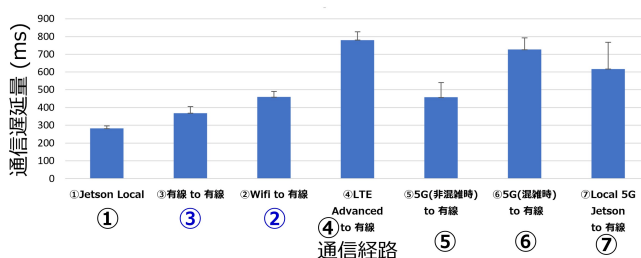


図 7: Latency for seven paths (stdev error bar).

ず jitter が非常に大きいため、映像通信に不利であった。

映像通信に利用した WebRTC[3] では、jitter は映像の停止、かくつきを招来し、映像の解像度を自動的に低下させるため、アバターロボットを利用する際の利用者の利便性を大幅に損なう問題がある。今後は複数の回線を束ねて jitter を小さくする対策などを行う必要があると考えられる。

3.4 全方位立体映像の共有機能の検証

3.4.1 視点共有者の IPD 誤差と視野遮蔽余裕

遠隔地の立体視は、臨場感の向上、現地対話者の存在感とアイコンタクトの感覚の向上、特に近傍の作業対象物の正確な知覚において重要である。遠隔地のアバターロボットを操縦する操作者は、本システムで常に立体視を利用可能であるが、カメラを共有する多数の視点共有者は、操作者と同じ頭部方向で環境映像を見る場合を除いて TwinCam カメラにおける相当眼間距離 (eIPD) が低下する。視点共有者の eIPD 誤差率は、TwinCam カメラ方位との誤差角度 $\delta\varphi$ に対して Fig. 8 で増加し、誤差角度 90 度の時 eIPD=0 となり、視差がない単眼視となる。

また、視点共有者が操作者の頭部 (TwinCam) 方位と異なる方向を向いた場合、カメラ筐体が視野の一部を遮蔽する問題がある。これは、TwinCam カメラの視線方位と観察方位の相違が大きい場合に発生する。また、HMD の視野角が大きいほどカメラ筐体が視野に入る可能性が高まる。視点共有者が頭部回転し視野にカメラ筐体が入るまでの角度 (角度余裕) は、視野角が 104 度 (Meta Quest2 の場合) には、Fig. 9 となる。

IPD 誤差による視差の低下とカメラ筐体遮蔽は、利用者頭部の方位と TwinCam の方位の差により生ずる。この差は、通信の遅延で過去映像を見ていること、および頭部方位指示の通信遅延で TwinCam の回転が遅れるために発生する。

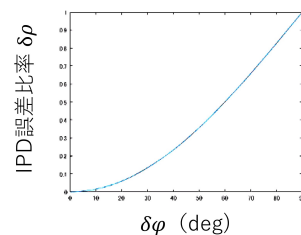


図 8: IPD error ratio.

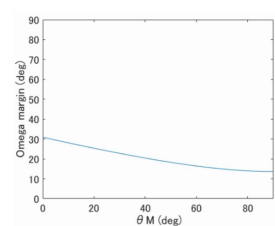


図 9: Margin angle.

3.4.2 TwinCam の角度応答性

TwinCam 機構のカメラ回転の応答性の解析のため、伝達関数を推定した。ここでは近似値をえるため、ステップ応答から PID 制御を含む伝達関数を推定した。モータ制御の PID パラメータは $k_p = 10.523$, $k_d = 0.306$, $k_i = 0.134$ であり、計測のサンプリング周波数は 1 kHz とした。

実機のステップ応答と Matlab システム同定モデルの応答を Fig. 10 に示す。TwinCam の近似伝達関数 T_f は、
$$\frac{-18.79s + 6070}{s^3 + 34.92s^2 + 736s + 6059}$$
 と推定され、実際の応答とのフィッティング率は 97.23% となっている。

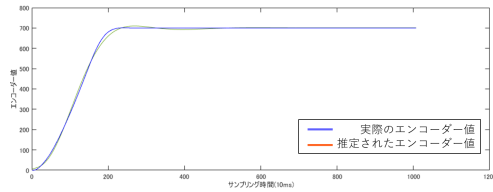


図 10: Step response of TwinCam and the model.

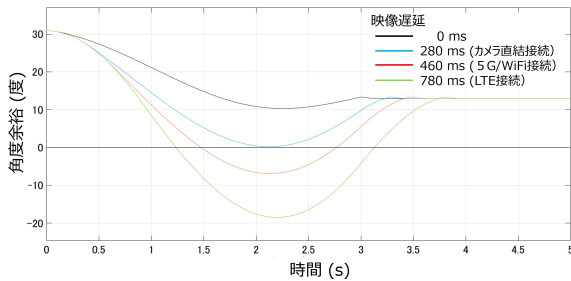


図 11: Margin angle (φ_{margin}).

3.5 視野遮蔽を決定する角度余裕

通信遅延により観察者頭部角度と TwinCam の回転角の間に誤差が生じ、角度余裕が減少し 0 以下となると視野遮蔽が発生する。通信遅延に対する角度余裕を Simlink の計算により求めた。

実際の遅延は、観察者の頭部回転量を TwinCam に伝達する際の通信遅延 (WebRTC の Data Channel), TwinCam 機構の回転遅延, 全天周カメラの映像圧縮遅延, 映像を HMD まで転送し描画する際の遅延の合計である。この時、頭部回転角と筐体映像の角度の差 φ_{margin} が求まる。

DataChannel による遅延は余裕を見て、片道 10 ms とした。TwinCam 機構内部の遅延は推定された伝達関数を利用した。映像系システムの遅延は、カメラに直結①の 280 ms, 5G 接続⑤の 460 ms, LTE 接続④の 780 ms を設定した。頭部回転の例として、ここでは 3 秒で 90 度に達する cosine 波形 ($0.5 * (1 - \cos(\pi/12 * t))$) を用いた。角度余裕 φ_{margin} が正ならばカメラの遮蔽は生じない。

計算結果の角度余裕 φ_{margin} を Fig.11 に示す。

3.6 視野遮蔽と回避制御

上記の結果から、遅い頭部回転 (30 度/秒) でも、5G 接続の場合 7 度程度の視野遮蔽が発生する。ただし、それは約 1.25 秒で解消するので、視野の端に少し見えてすぐ消える。また、これは 90 度の頭部回転をした場合に、45 度を過ぎてから見えるという事象である。先行研究によると周囲を見る状況で最大の頭部回転速度は 2.27 rad/s (130 deg/s) [4] とされており、大きな角度の回転を行うと遮蔽の発生可能性は高い。他方、Fig. 9 の結果から、頭部回転が正面から 30 度以下であれば、頭部回転が最大速度でも筐体の遮蔽は発生しない。

視野の遮蔽を防止するためには、筐体の裏側の映像を利用することが有効である。これを行うために、遮蔽物となる側のカメラから得られた映像を利用することが考えられる。すなわち、観察者の頭部角度が Fig. 9 で得られた角度余裕

に近づいた際に、筐体が遮蔽する側のカメラの視野 (右目視野または左目視野) の映像を、反対側のカメラからの映像に切り替えて描画すれば良い。

この制御においては、カメラからの映像を配信する際に、TwinCam の回転角を描画端末に送信することで、配信された映像を見る端末側 (HMD, Smartphone, Flat display) で解決することが可能である。このアルゴリズムを実装し、筐体による遮蔽を回避できることを確認した。

4. おわりに

本研究では、現実空間のライブ映像と過去の空間の記録映像を、多人数かつ多種のデバイスに配信できるメタバースシステムを構築し、その機能を検証した。本メタバースシステムの Tele-XR Portal は、現実空間のリアルタイム全方位両眼立体視映像、過去に撮影を行った視差固定立体映像、さらに 3 次元仮想空間の体験を自由に選択して行き来することを可能とする空間である。システムの主たる操作者 (オペレータ) はこれらの空間で遠隔のアバターロボットを操縦する際、移動に伴う身体感覚を得ると同時に、遠隔環境と相手の顔を立体視しながらライブ対話することができる。操作者の視点の全天周映像は、多数の視点共有者 (参加者) に配信され、各受信者は任意の方位を見ることが可能である。操作者と同じ方位の立体視をしない場合は、方位が異なるほど IPD が減少し、連成するカメラの筐体を回避する単眼視に移行する制御を開発し、その特性をシミュレーションで確認した。

システム機能の検証として、1) 米国 New York と東京間で、5G および LTE 通信による実証実験を行い、全方位の 4K 立体視映像、30 fps を実現できることを確認した。2) 東京において GPS の時計を用いて、経路の異なる配信の遅延を計測し、現在の 5G 無線通信を用いた映像配信の遅延時間が、合計で約 450 ms で、無線 LAN 接続とほぼ同じ程度となることを実証した。

今後の課題としては、多数の参加者が相互にコミュニケーションを行う手法の導入が考えられる。

謝辞

本研究は、総務省 SCOPE #191603003, 19189681, JSPS 科研費 18H04118, 21K19785 および東京都立大学 local-5G プロジェクトの支援により実施された。ここに謝意を表す。

参考文献

- [1] S. Tachi, Y. Inoue, and F. Kato. TELESAR VI: Telexistence surrogate anthropomorphic robot VI. *Int. J. Humanoid Robotics*, Vol. 17, No. 5, 2020.
- [2] Y. Ikei et al. Live stereoscopic 3d image with constant capture direction of 360 cameras for high-quality visual telepresence. In *2019 IEEE VR*, 2019.
- [3] Webrtc 1.0: Real-time communication between browsers. <https://www.w3.org/TR/webrtc/>.
- [4] JR. Tian et al. Dynamic visual acuity during transient and sinusoidal yaw rotation in normal and unilaterally vestibulopathic humans. Vol. 137, No. 1, pp. 12–25, March 2001.