



リアルワールドメタバース実現に向けた クロスバース・プラットフォームの構築 (1) —3次元マップの生成と選択的配信機構—

柳澤祥太¹⁾, 中村文彦²⁾, 木村朝子²⁾, 柴田史久²⁾

1) 立命館大学大学院 情報理工学研究科 (〒 567-8570 大阪府茨木市岩倉町 2-150)

2) 立命館大学 情報理工学部 (〒 567-8570 大阪府茨木市岩倉町 2-150)

概要: 我々は、現実世界とメタバースを融合したリアルワールドメタバースの実現に向け、クロスバース・プラットフォームの構築を進めている。実現に向けた課題の1つとして、実空間をいかにしてデジタルツインとしてVR空間上に再現するかという点がある。近年、デジタルツインの3次元再構成手法として、3D Gaussian Splatting が注目されている。本研究では、大規模な3次元マップの一部分を選択的に配信することで、本手法がクロスバース・プラットフォームに導入可能か検討する。

キーワード: メタバース, デジタルツイン, 3次元再構成, 3D Gaussian Splatting

1. はじめに

近年、複合現実(MR)環境におけるデジタルツインの活用が注目を集めている。特に Niantic 社が提唱する「リアルワールドメタバース」は、現実世界をデジタルツインとしてVR空間に再現し、現実とリモートの双方からのアクセスを可能にする新たな概念である [1]。このアーキテクチャは、物理空間とMR空間のデジタル情報を密接に連携させることで、遠隔地のユーザに高い臨場感を提供し、エンターテインメントからビジネスまで幅広い応用が期待されている。リアルワールドメタバースの研究例として、MRMACでは360度カメラで現実環境をVR空間へリアルタイムにストリーミングしてデジタルツインを構築し、VRユーザとARユーザの非対称な共同作業を実現している [2]。

こうした背景を踏まえ、我々はリアルワールドメタバースを実現するために、クロスバース・プラットフォームの構築を目指している。クロスバース・プラットフォームは、現実世界と仮想空間をシームレスに接続し、両者のユーザが相互にインタラクション可能な環境を提供するものである。クロスバース・プラットフォームでは現実世界のユーザはAR/MR技術により仮想空間のユーザとコミュニケーションをとることができる一方、リモート環境のユーザは、アバターによる仮想空間への没入に加えて、アバターロボットなどを介して現実世界に物理的に働きかけることができる。このような構想の実証例として、大学を舞台とした「MRキャンパス」を構想・提案している(図1)。本構想では、現地に存在するユーザと遠隔にいるユーザが、それぞれMRデバイスおよびVRデバイスを介して同一空間を共有し、円滑にコミュニケーションすることを目的としている。そのためには、現実空間の精緻な3次元再構成と、その配信システムの構築が不可欠である。しかしながら、従来の3次元再構成手法である Structure-from-Motion (SfM) や Neural

Radiance Fields (NeRF) [3] を基盤とした方式では、モデルの品質、レンダリングのフレームレート、3次元マップの編集のしやすさに課題があり、リアルワールドメタバースの実装には適さなかった。

本研究では、これらの課題を解決するために、高速かつ高品質にレンダリング可能な3D Gaussian Splatting [4] を用いた3次元マップ生成と、ユーザの自己位置に応じて3次元マップを選択的に配信することで、クロスバース・プラットフォームのデジタルツインシステムを構築する。具体的には、大規模な範囲でもシステムが動作するように、3次元マップの生成部ではキャンパスのエリアごとに分割し、それぞれのエリアの3次元マップを合成する。そして、配信時には3次元マップを格子状に区切り、チャンクごとにサーバに配置することで、クライアント側は自己位置に基づいて必要な範囲のみを選択的にダウンロードできる設計とする。これにより、描画品質を保ちつつ、高フレームレートかつ低遅延での可視化を目指す。

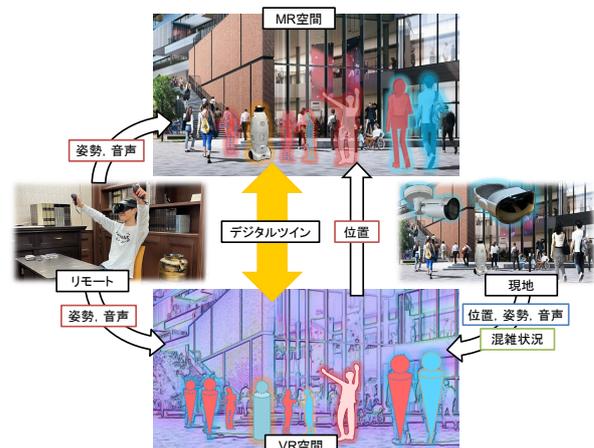


図 1: MR キャンパス概要

2. 3次元マップの生成と選択的配信機構

2.1 システム概要

本研究では、大規模な範囲にも適用することができる3次元マップの生成と選択的配信機構を提案する。システムは、図2に示すように、3つのモジュールから構成される。

1. 動画もしくは連続静止画像から3次元マップを生成
2. 複数エリアの3次元マップの合成と配信
3. 3次元マップのレンダリング

本システムの設計にあたり、3次元再構成の対象が静的な環境に限定され、マップ合成時に使用する変換行列、およびマップの範囲を限定するためのクリップパスがあらかじめ取得済みであることを前提条件とする。

2.2 3次元マップ生成

3次元マップの生成は、対象範囲を複数のエリアに分割し、3D Gaussian Splatting を用いて高精度な3次元マップを生成する。入力データは、対象となる建物内を撮影した動画または連続静止画像である。これらの画像群から SfM を使用して、カメラの位置と姿勢、および初期点群を推定する。本システムでは、SfM を実行するソフトウェアの1つである RealityCapture を使用した [5]。Reality Capture は従来広く利用されている SfM と比較して、大量な画像に対しても高速に処理できる利点を持つため、本システムに採用した。

点群とカメラパラメータが推定された後、3D Gaussian Splatting によって高密度な3次元マップを生成する。3D Gaussian Splatting での再構成に当たっては、Hierarchical 3D Gaussian Representation に基づき、深度マップによる補正と、露出変化への対応を導入する [6]。前者については、Depth Anything V2 を用いて各入力画像に対する高精度な深度推定を行い、これをもとにスプラット配置時の空間精

度を高める [7]。後者については、入力画像間の明暗差を抑えるために、アフィン変換に基づき色調を正規化する。これらの処理により、テクスチャの乏しい領域における精度向上や撮影時の露出差による色調のばらつきから発生するノイズを抑制する。

2.3 3次元マップの合成

3次元マップの合成では、各エリアごとに生成された3次元マップを、グローバル座標系上で一体化する処理を行う。各エリアの3次元マップに対して変換行列を適用し、グローバル座標系へと変換する。その後、他のエリア部分にスプラットが干渉しないよう、各エリアに対応するポリゴンクリップパスによって領域を切り取る。切り取る際には、描画抜けを防止するためにパスをわずかに外側へと拡張し、パスより外側のスプラットは削除する。その後、これらの処理をすべてのエリアに対して適用し、合成されたキャンパス全体の3次元マップを得る。

2.4 選択的配信とレンダリング

統合された3次元マップは、チャンク単位に分割され HTTP ベースで選択的に配信される。チャンクは、エリア内の3次元空間を座標に基づいて格子状に区切ることで定義される。クライアントは、自己位置を基準とした描画距離 $r[m]$ をもとに、必要なチャンクの空間範囲をリクエスト対象領域として動的に決定する。その後、この領域内に含まれるチャンク ID をもとに HTTP リクエストを逐次送信し、対応する3次元マップを非同期に取得する。取得された3次元マップはメモリ上のキャッシュ領域に一時保存され、描画対象となる。一方で、描画距離外に移動した場合には、該当チャンクは削除対象領域に分類され、メモリ上から解放される。描画対象とされた3次元マップのチャンクは、GPU 上で高速にレンダリングされる。

3. 動作確認と評価

3.1 評価目的と環境

本研究では、提案した3次元マップ生成および選択的配信システムの有効性を検証するため、3D Gaussian Splatting による3次元マップの再構成精度とエリア間合成の妥当性、レンダリング性能（フレームレート）とスプラット数との関係、選択的配信による通信効率とチャンク単位での取得遅延に着目して評価を行った。

Wang らの研究 [8] によると、VR 空間において視覚応答性能を向上するためには 90fps、シミュレータ酔いを有意に抑制するためには 120fps 以上のフレームレートが必要である。レンダリングでは超解像技術やフレーム生成などの軽量化を行っていないことから、評価では平均 90fps をフレームレートの目標として設定をし、これを満たす条件を検証する。

実験対象としては、立命館大学大阪いばらきキャンパス (OIC) の H 棟 5 階を選定し、図3に示すようにエリアを分割してマップを作成した。撮影には、ミラーレス一眼の Sony $\alpha 7$ IV とジンバルの DJI RS 4 Combo を使用して撮

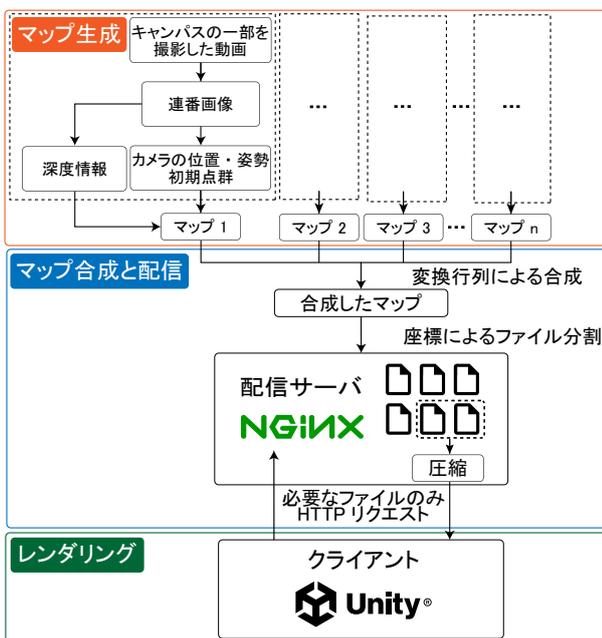


図2: システムの全体構成図

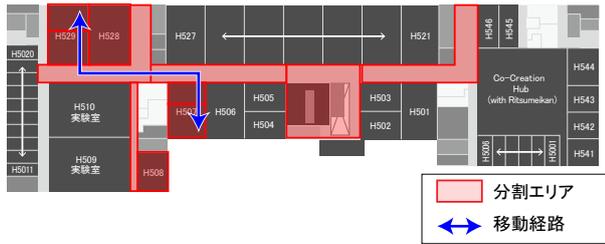


図 3: 3次元マップの範囲と分割場所

表 1: マップ生成とクライアントで使用した計算機

CPU	Intel Core i9-10980XE
GPU	NVIDIA GeForce RTX 3090 NVIDIA RTX A6000
RAM	160GB

影した動画から静止画を抽出した。マップ生成とクライアントに使用した PC 構成を表 1 に示す。マップ生成には高精度な再構成と深度処理のために NVIDIA RTX A6000 を、VR レンダリングには NVIDIA RTX 3090 を使用した。レンダリングを実行するクライアントには Unity 6 を使用した。配信サーバには Amazon EC2 の c7a.large を使用した。また、通信環境については、クライアント PC と配信サーバ間のネットワーク帯域を iperf3 によって評価し、ping が 10ms 程度で推移し、1Gbps 程度の通信速度が確保されていることを確認した。

3.2 3次元マップの再構成と合成

図 4 に、再構成および合成後の 3次元マップの一例として、隣接した 2つのエリアを合成し、レンダリングした結果を示す。各エリアにおいては、壁面、床、天井、机などの構造物が高密度かつ一貫性を持って再構成されており、空間全体として明瞭な視認性を保っている。また、3cm 程度の文字が認識可能な程度の空間解像度も達成された。エリア境界においては、大きなノイズや不連続な描画は確認されず、複数のエリアを接続して閲覧した際にも、隣接エリア間の幾何学的整合性や色情報の不整合は見られず、マップ全体として視覚的な連続性が確保されていた。しかし、薄い壁やガラスでエリアが区切られている場合に限り、合成時の品質低下が見られた。

3.3 レンダリングパフォーマンス

本節では、描画距離の変化がレンダリングに与える影響を評価する。実験では図 3 に示す移動経路を分速 70m で移動するときのシナリオを設定した。レンダリング時の解像度は Meta Quest 3 の使用を想定し、2064 × 2208 のピクセル数を指定した。また、VR での使用時に両眼視差がある場合も想定し、2つのカメラを 10cm 離して配置し、それぞれレンダリングをおこなった。一方、両眼視差がない条件の場合は、1つのカメラでレンダリングするように設定した。

はじめに、スプラット数とレンダリングのフレームレートの関係について評価を行った。目標の平均 90fps が動作可能な読み込みスプラット数は、両眼視差ありの場合は 150

表 2: 描画距離と通信量

描画距離 [m]	2	4	6	8	10
平均通信量 [Mbps]	48.7	75.3	91.1	112.0	114.4

表 3: チャンクサイズと取得時間

チャンクサイズ [m]	0.5	1	2	3	5
平均 [ms]	149	306	809	1444	4361
p95 [ms]	543	1148	2921	3988	15126

万点、両眼視差なしの場合は 320 万点であった。

次に、描画距離と読み込みスプラット数に関して評価を行い、結果を図 5 に示す。描画距離を上げると、読み込みスプラット数は指数関数的に増加した。

3.4 通信パフォーマンス

本節では、描画距離の変化に伴う通信量およびチャンク取得時間への影響を評価する。

図 3 で示した経路に沿って移動する条件下で、描画距離 r を変化した際の平均通信量を表 2 に、チャンクサイズとその取得時間を表 3 に示す。

通信量に関しては、描画距離が増加するにつれて転送されるデータ量も増加し、 $r = 10\text{m}$ の条件で平均 114.4 Mbps を記録した。 $r = 10\text{m}$ より描画距離が小さい場合は、100Mbps 程度の通信に収まっており、一般的な家庭用の 1Gbps 契約の通信環境を前提としたインフラ環境下において、十分に実用的な帯域である。

チャンク取得時間については、描画距離が小さいほど応答時間が短く、0.5m 四方では平均 149ms、p95 (95 パーセントイル) は 543ms であった。チャンクの取得にかかる時間は 1m 四方までは線形的に増加するものの、それよりチャンクが大きい場合は、急激にチャンクの取得にかかる時間が増加し、遅延が顕著になる。

3.5 考察

マップ生成においては、構造物の形状が明瞭に再構成され、エリア間の接合部においても色情報や幾何的な不連続性は見られなかった。3cm 程度の文字認識が可能な程度の空間解像度も達成されており、日常的なキャンパス活動には十分な品質であると考えられる。合成処理について、概ね違和感なく連続的に合成されたが、薄い壁やガラスでエリアが分かれる部分については精度が低下する傾向がみられた。これらの部分については幾何学的な合成だけではなく、追加で機械学習するなどをして、合成精度を高める必要がある。

描画性能に関しては、節 3.3 の結果から、平均 90fps を維持できる読み込みスプラット数は、両眼視差ありで 150 万点、視差なしで 320 万点であった。図 5 を基に、描画距離に換算するとそれぞれ 5m および 8m が上限であり、教室規模の空間であれば十分にカバーすることが可能である。また、チャンクサイズは 1m 四方までであれば、平均取得時



図 4: 隣接するエリアを合成した結果

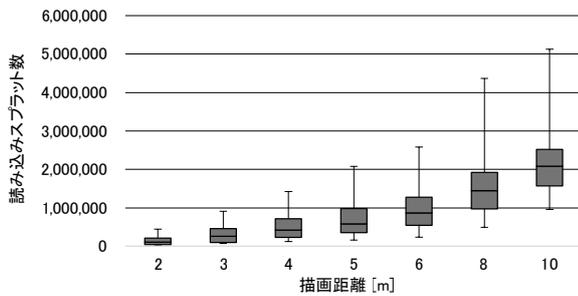


図 5: 描画距離と読み込みスプラット数

間が 500ms, p95 で 1 秒未満に収まり, 体感的に遅延のない取得が可能と判断される. ただし, 遠方を描画する必要のある廊下などのシーンでは, 描画距離は不十分であることが予測される. また, VR 酔いを防ぐためには 120fps 以上が必要であり, 静止に近い状態ではこれを満たすものの, 移動をするシーンではこれを下回ることがあった. さらなる描画性能の向上には, 描画距離に応じて適切な解像度の 3 次元マップを配信する機構が必要がある.

通信性能について, 描画距離が 10m 以下である場合, 100Mbps 程度の通信に収まっており, 家庭用の 1Gbps の光回線を契約しているユーザ環境であれば, 十分実用可能な帯域消費量である. ただし, 描画距離や視点移動によって最大 500Mbps を超えるピークが生じる場合があり, さらなるキャッシュ制御や優先度付き事前取得や 3 次元マップの圧縮といった配信最適化が必要である.

4. むすび

本研究では, リアルワールドメタバースの実現に向けて, 3D Gaussian Splatting を用いた 3 次元マップの生成と, ユーザの自己位置に応じた選択的配信システムを構築した. 部分的な 3 次元マップ生成とその合成を行うことで, 大規模な範囲の 3 次元再構成が可能であることが示された. 定性的な評価では, 高精度かつ連続性のある 3 次元空間の再構成とエリア間の自然な合成が確認された. また, クライアントの視点に応じて描画距離内のチャンクのみを選択的に配信する仕組みにより, 高い描画フレームレートと通信の効率化を図った.

評価実験では, 描画距離やチャンクサイズがレンダリング性能や通信に与える影響を明らかにした. 両眼視差アリであれば半径 5m, 両眼視差がない条件であれば, 8m 程度であれば平均 90fps 以上で動作可能であることが確認された. また, チャンクサイズが 1m 四方以下であれば低遅延で配信が可能であった. また, 一般的な家庭用インターネット

環境でも実用可能な通信帯域で動作することを確認した.

今後は, 遠距離描画を可能にする多解像度表現の導入, スプラットデータの軽量化, およびキャッシュ制御を含む高度な配信アルゴリズムを導入することにより, さらなる遠距離の描画や軽量化に取り組む.

参考文献

- [1] Niantic, “The Metaverse is a Dystopian Nightmare. Let’s Build a Better Reality.” <https://nianticlabs.com/news/real-world-metaverse?hl=en>, 2021. Accessed: 2025-7-9.
- [2] F. Zaman, C. Anslow, A. Chalmers, and T. Rhee, “MRMAC: Mixed reality multi-user asymmetric collaboration,” in *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 591–600, 2023.
- [3] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “NeRF: representing scenes as neural radiance fields for view synthesis,” *Commun. ACM*, vol. 65, p. 99–106, Dec. 2021.
- [4] B. Kerbl, G. Kopanas, T. Leimkuehler, and G. Drettakis, “3D Gaussian Splatting for Real-Time Radiance Field Rendering,” *ACM Trans. Graph.*, vol. 42, July 2023.
- [5] Epic Games, Inc., “RealityCapture - 3D Models from Photos and/or Laser Scans.” <https://www.capturingreality.com/>. Accessed: 2025-1-5.
- [6] B. Kerbl, A. Meuleman, G. Kopanas, M. Wimmer, A. Lanvin, and G. Drettakis, “A Hierarchical 3D Gaussian Representation for Real-Time Rendering of Very Large Datasets,” *ACM Transactions on Graphics*, vol. 43, July 2024.
- [7] L. Yang, B. Kang, Z. Huang, Z. Zhao, X. Xu, J. Feng, and H. Zhao, “Depth anything v2,” *arXiv:2406.09414*, 2024.
- [8] J. Wang, R. Shi, W. Zheng, W. Xie, D. Kao, and H.-N. Liang, “Effect of Frame Rate on User Experience, Performance, and Simulator Sickness in Virtual Reality,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 5, pp. 2478–2488, 2023.