



SOMA: Spatial Observation and Movement Archive

HMD の空間構築機能を利用した行動分析ツールの提案

小柳陽光¹⁾

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

概要: 実空間上のユーザ行動を記録する技術は体験のユーザビリティを分析する上で有効である。既存研究では床や壁などの単純な平面への視線のヒートマップなどが主であり、複雑な形状を持つ実空間への分析ツールは提案されていなかった。本研究ではヘッドマウントディスプレイの持つ空間を 3D データ化する機能を用いて、視線を可視化する技術を実装した。これにより、ユーザが実空間のどこに注目していたのか観察することができる。

キーワード: 計測, 可視化, In-Situ

1. 背景

空間内でのユーザの身体動作やオブジェクトに対するインタラクションを可視化することは、ユーザがどのように空間を認識し、身体を動かそうとしているのか重要な洞察を与えることができる。例えば、視線追跡技術はコンテンツに視線情報を埋め込み、ユーザがどこに注目しているのか、あるいは、注目が集まっていないのかを可視化することができる。初期の研究では、Web デザインやゲームにおけるユーザの行動解析のために用いられており、主にデスクトップベースのシナリオの分析が行われてきた。視線追跡技術の進化に伴い、デスクトップベースから現実世界の環境や VR 環境下での視線情報を取得できるようになった。これにより、研究者は美術館の人流やスポーツの動作解析、体験の導線など様々な空間に関する文脈を分析することができる。例えば、建設家は可視化技術により目立たない空間を改善するように設計し直すことができる。そのためには複雑な空間を認識し、得られた情報を可視化するための技術が必要となる。このような現場の空間において、ヒトと空間とのインタラクションを可視化する技術は In-Situ と呼ばれている。近年では、視線追跡技術や身体的位置を追跡することが可能であり、市販の機材一つのみで In-Situ の技術が構築可能となった。これまで In-Situ における主な可視化手法はユーザの身体動作、および、移動軌跡、視線の可視化が提案されてきた。視線の可視化に関して、従来では注視の累積時間による視線のヒートマップ化を用いる手法が提案されてきた。既存研究では、空間上に配置したディスプレイや絵画などの平面的なオブジェクトが視線ヒートマップの主な対象として扱われてきた [1,4]。しかしながら、実際には展示会における彫像など複雑な形状のオブジェクトが配置されており、視線の可視化を立体物に対応して行う必要がある。さらに、空間に対してユーザがどのような動きや姿勢でインタラクションをしたのかを詳細に理解するために、ヒト型の 3D モデルの全身像を介して事前記録した骨格データを再生する方法が提案されている [2]。ただし、

RGB カメラから全身の姿勢推定を行う技術を用いて事前に記録する必要があるため、ユーザを撮影する他者の存在が必要がある。また、記録と再生のプロセスが同一媒体ではないため、単体のアプリケーションでその場で即時可視化することが簡単に行えない。上記の既存研究は全て Microsoft 社の HoloLens を用いたアプリケーションとして実装されている。HoloLens や近年登場した Apple 社の VisionPro などの MRHMD はいずれも約 40 万円以上と一般消費者にとって高額な機材である。一方近年では、約 7 万円の安価なデバイスでありながら MR 機能を搭載している Meta 社の MetaQuest3 が登場している。MetaQuest3 は、空間をメッシュとして構築するだけでなく、機械学習によって全身の身体動作を推定する技術を搭載している。このような廉価製品を用いて可視化技術を実装することで、一般公共施設に来訪したユーザを対象にすることができ、実際に現場を訪れるユーザが体験の際に空間をどのように注目したのかという実践的なデータの記録が容易となると期待できる。そこで本研究では、MetaQuest3 のみで空間を 3D メッシュとして構築する技術を応用することで、複雑な物体へのヒートマップ化を行うほか、全身のアバタ像を介したユーザの身体動作の記録と再生を行う技術 Spatial Observation and Movement Archive(SOMA) を提案する。本稿では、まず In-Situ に関する既存研究について述べ、本研究の新規性を示していく。次に本研究が想定する SOMA のユースケースを最初に述べ、それに基づいた仕様要件を定義する。最後に、SOMA を構成する機能である視線のヒートマップ技術、および、身体動作の記録技術、過去のデータの再生機能について述べていく。

2. 関連研究

民生機器用として販売された MRHMD である Microsoft 社の HoloLens の登場以降、少ない機材で現場の空間におけるユーザの身体動作、および、空間的なインタラクションを記録し、可視化する技術 In-Situ が注目されている。Hololens

をはじめとした MRHMD には現実の空間の深度情報を取得する技術を有しており、現実の空間と紐づけて 3D オブジェクトを配置することができる。

Büschel らは、MRHMD を装着した複数のユーザの行動、空間へのインタラクション、イベントデータを可視化できる MR ツール MIRIA を報告している [1]。MIRIA は、記録されたデータを空間中に 2D 表示 (ヒートマップ・散布図) で配置するほか、3D データとしてユーザの移動軌跡を表示することでユーザの空間に対するインタラクションを分析することができる。Reipschläger らは、ユーザの身体の詳細な動き、および、姿勢、移動軌跡を可視化するために、それらの記録データをアバタに反映させる AvatAR を報告している [2]。AvatAR では、ヒト型の 3D モデルに事前記録された骨格データを反映することで特定の時点における人物の姿勢を再現している。さらに、時系列的な移動を表現するためにアバタの上半身に移動軌跡となる曲線を連続して表示している。Luo らは、空間内の特定領域におけるヒトの動作や関心を分析するために関心領域を定義し、その領域内で移動データをフィルタリングすることで、結果のデータを詳細かつ集約された情報として提示する Pearl を報告している [3]。Pathmanathan らは視線情報を空間だけではなく、時間にまたがって視覚化するために、人々の行動データを空間と時間にわたって表示するグラフィック UI を提案し、選択した時点での行動データを視覚化している [4]。

MIRIA、および、Been There は記録されたデータのヒートマップ化することができるが、絵画のような 2D 表示のオブジェクトを対象としている。実際には美術館には彫像などの複雑な立体物が配置されており、ユーザと空間のより詳細なインタラクションを計測するためには視線追跡の可視化を立体物に対応する必要がある。また、AvatAR は事前記録された骨格データを可視化しているため、記録のプロセスと再生のプロセスが異なる媒体で行わなければならない。Pearl は関心領域の定義による可視化のフィルタリングが主な特徴であるが、空間へのインタラクションに対してユーザがどのように身体動作を行ったのかは、頭部と手のモデルのみで表現している。

他方近年では、元々 VRHMD として発売されてきた Meta 社の MetaQuest が前部にあるカメラから外部の環境をモニタに表示する MR 機能を実現している。MetaQuest は上半身の装置の位置から下半身の動きを推定する技術を発表しており、簡便な機材でユーザの身体動作を再現することができる。そこで本研究では、MetaQuest3 を用いて、ユーザの 3D 空間における視線追跡、および、身体動作の記録を行う技術として Spatial Observation and Movement Archive HMD(SOMA) を提案する。

3. 提案手法

本研究では、一般消費者が外出先で MRHMD を装着し、特定の空間に訪れた際に自己の身体動作や視線をその場で保存・再生をするユースケースを想定する。これにより、専

門家が導線改善などの設計に役立てるなどのような従来の使用法だけではなく、消費者であるユーザが過去のほかのユーザの行動を参考にすることができる。近年では、Meta 社の MetaQuest3 にも現実の周囲の映像を取り込みディスプレイに投影するパススルー機能を搭載しており、価格も廉価版の MetaQuest3S では約 4 万円と消費者のエントリーモデルとして採用されることが多い。さらに、MetaQuest3 には上半身の動きや位置から下半身の動きを推定する機能を有しており、単一の HMD のみでユーザのモーションデータを記録することができる。

3.1 仕様要件

SOMA では、ユーザの視線を可視化する機能、および、身体動作を記録する機能、それらを再生する機能を必要とする。ユーザの視線を可視化する機能では、視線が対象物に接触したことへの検知、および、接触箇所を可視化する機能を実装する必要がある。ユーザの身体動作の記録には、ユーザの各身体部位を記録し、再生時にはアバタに反映する機能を実装する必要がある。記録されたデータは、実機で保存されたものを再生可能にするほか、第三者がその空間に訪れた際に再生できるようクラウドストレージにアップロードし、適宜ダウンロードする必要がある。本稿では、QR コードを用いてデータのダウンロードとデータ (再生用のヒト型モデルと事前にスキャンされた空間) の配置を行う。

3.2 環境

本研究では、MRHMD として MetaQuest3 を採用した。MetaQuest3 は前面の RGB カメラがあり、現実空間の風景を映像として内面ディスプレイに投影することで、複合現実型コンテンツを構築することができる。さらに、MetaQuest3 に備えられているセンサとプロジェクタの機構により深度情報を取得することが可能であり、現実空間をメッシュとして生成することが可能である。これにより MR コンテンツにおいて衝突判定や遮蔽関係を表現することができる。ただし、MetaQuest3 には視線追跡技術は搭載されていない。したがって後述の視線ヒートマップでは、実際には追跡された頭部の向きに基づいた代替手法を用いた。具体的には、ユーザの頭部方向にレイ (衝突判定の光線) を飛ばし、現実空間のメッシュの衝突地点をヒートマップとして着色し、可視化している。視線追跡技術を搭載する HMD として MetaQuestPro が存在するが、本研究の実施時点では終売となっていた。今後、視線追跡技術を搭載した新たな HMD が登場する場合には、これを用いた視線ベースの可視化手法へ移行する予定である。アプリケーションの実装にはゲームエンジンである UnityEngine6 を採用し、MetaQuest の機能を使える Unity OpenXR Meta 2.1.1 を使用した。

4. 実装

4.1 視線の可視化

ユーザが空間においてどこに注目していたのかを可視化するために、視線が空間内の立体物に当たった際に着色を行うヒートマップによる可視化を実現する。MetaQuest は

ランダムドットや網目のパターンの映像を投影するプロジェクタが投影された際の模様の歪みを IR カメラで撮影することで深度情報を計算している。そして、得られた深度情報からメッシュ化する MeshAPI 機能によって、スキャンされた空間をメッシュとして得ることができる。本研究では、MeshAPI によって取得したメッシュの頂点カラーを注視箇所に基づいて白色に塗り、それを基にシェーダによってヒートマップによる可視化を行った。ヒートマップは注視時間の累積で濃淡が濃くなっていき、指定した時間まで注視したときに最大値を示す色を描画する。また、注視時間が事前に定義した閾値を下回る場合、注視箇所の頂点カラーの値が徐々に減少していき、最終的に元の透明色となる(図 1 左)。閾値を超えた場合には、長時間注視していたとして注視箇所に注視時間の UI を空間上に表示する。これにより、ユーザが空間のどこに注目したのかを理解することが期待できる。

視線が空間のメッシュに接触したとき、接触地点からブラシ範囲内にある空間メッシュの各頂点を調べ、該当の頂点の頂点カラーを計算する。このとき空間メッシュの全ての頂点が範囲内にあるか毎フレーム判断するため、メッシュの頂点数に応じてオーダー量が比例して増えていく。そのため、複雑すぎる空間を再構成した場合にはコンテンツのフレームレートが大きく低下する。そこで、Unity Engine にある並列処理機能である DOTS を用いてフレームレートの低下の問題を解決した。DOTS はデータ指向設計に基づくパフォーマンス最適化技術であり、並列化しやすい構造をデータ指向に設計する ECS、および、並列処理を行う JOB、JOB の処理を Single Instruction/Multiple Data(SIMD) 最適化する Burst Compiler から構成される。また、DOTS を用いて変更された頂点カラーを直接反映することはできないため、計算された頂点カラーを CPU で反映させなければならない。この時、頂点カラーの計算と同様に頂点数分の計算が入るため、CPU での頂点カラー反映処理はフレームレートを大きく低下させる。そこで、DOTS の処理の際に該当の頂点のみを配列に格納して、その配列を CPU の処理に渡すことで CPU での計算数を最小限に抑えた。

4.2 モーションキャプチャ

本研究では、ユーザが空間内でどのように行動し、どのような姿勢で立体物を観察していたかを可視化することを目的とする。そのため、ユーザの身体動作を記録し、再生時には全身を持つアバタを用いて、その動作と姿勢を可視化する。ユーザの身体部位のトラッキングには、Meta が提供する Movement SDK を使用する。Movement SDK は、HMD およびコントローラの動きから身体全体の動作を推定し、アバタに反映する機能を備えている。Movement SDK を用いて、アバタの各身体部位の関節角度(回転情報)を毎フレーム JSON 形式で記録する。再生時には、この関節角度の記録をアバタの各関節に対して代入することでユーザの身体動作を再構成する。また、身体動作の記録と同時に、ユーザの音声も録音する。これにより、動作と同期して再

生することで、観察中の発話や意図の理解に活用できる。

アプリケーションの原点および座標軸は、起動時にユーザが最初に立っていた位置に基づいて設定される。そのため、動作と位置の記録のみでは、再生時のアバタの位置が毎回異なってしまう。これを解決するために、本研究では Meta Quest 3 の空間アンカ機能を使用する。空間アンカとは、現実空間の特徴点に基づいて生成される基準点であり、MR 空間内でオブジェクトの位置をセッションをまたいで固定することができる。初回起動時に空間アンカを生成し、その基準に基づいてアバタの再生位置を補正することで、アプリケーションを再起動しても一貫した位置にアバタを再現できるようにしている。さらに、視線の可視化に用いる座標原点は、再生中のアバタの頭部に再配置することで、記録時のユーザがどこを向いていたのかを再現することができる(図 1 右)。

4.3 ファイルシステム

記録されたモーションキャプチャデータおよび音声データは、アプリケーション内の永続データベースに、連番付きのファイルとして保存される。記録の再生は、空間上に配置された専用の UI を通じて行う。各データに対応するボタンを押すことで、対応するモーションおよび音声データを再生システムに読み込む。読み込まれたデータは、Meta Quest 3 のコントローラ操作によって再生が開始され、アバタを介してユーザの身体動作と音声空間内に再現される。さらに、本システムでは記録データの共有・再利用を可能とするため、記録元のユーザがデータをインターネット上にアップロードできる機能を備えている。データのクラウドストレージには AWS S3 を使用し、アップロードされたデータには QR コードが付与される。これを通じて、他のユーザは現地にて QR コードを読み取ることで、該当データを自身の空間内に再構成し、再生することができる。空間のスキャンデータもモーションキャプチャデータに併せて保存可能であり、現実空間の環境情報を .obj 形式の 3D モデルファイルとして記録する。本ファイルは、記録されたモーションキャプチャデータと同様に、ダウンロードおよびアプリケーション内への再配置が可能であり、空間情報と動作情報の統合的な再生を実現している。

5. 結論

本研究では、簡易な機材構成で複雑な空間におけるユーザの視線および身体動作を可視化するツール SOMA を提案した。SOMA では、市販の安価な HMD である MetaQuest3 を用いて、視線の可視化、および、モーションキャプチャを行うシステムを実装している。視線の可視化には、MetaQuest3 の持つ空間を認識し、メッシュ化する機能を用いて、頭部方向のレイとメッシュとの接触点からブラシ範囲内の頂点の頂点カラーの値を変更することでヒートマップを実現した。さらに、体験中のユーザの行動、および、音声を記録する機能を実装した。これにより、空間に対してどのような行動や姿勢をとり、反応を示していたのかを理解するこ

視線の可視化



行動の再生



図 1: SOMA の動作時の画像．視線の可視化 (左) 身体動作の再生 (右)

とが期待できる．記録されたデータは個人の HMD のストレージに保管され、見返すことができるほか、インターネット上のクラウドストレージにアップロードすることができ、QR コードによってダウンロードと同時に QR コードの位置を基準にデータを再生することができる．

6. 将来研究

今回の実装は主に単一ユーザの情報可視化に焦点を当てているが、実際の空間には複数のユーザが同時に存在する．将来的には、同一空間において装置を装着した複数ユーザの行動・視線データを同期的に記録することで、空間全体の時空間的な情報を包括的に保存・再生するシステムの実現を目指す．また、音声や行動はそのままのデータを可視化しているため、第三者が分析をしやすいように意味のあるデータの付与も考えていく．例えば、音声や行動からインタラクション先のオブジェクトに対する感想を抽出し、空間上に配置することで、第三者がそのユーザがオブジェクトに対してどのような感情をいただいていたのかを理解する手助けになると期待できる．また、本実装では QR コードを用いたデータ再生方式により、記録データの呼び出しは可能であるものの、精密な位置合わせを行うことが困難である．この課題を解決する手段として、今後は Visual Positioning System (VPS) の導入が有望である．VPS は空間内の視覚的特徴点をもとに、ユーザの位置・姿勢を高精度で復元する技術であり、これを活用することで再生時には記録当時とほぼ同一の空間座標系でデータを再構成することが可能となる．さらに、VPS と位置情報とを紐づけることで、ユーザの外出時にデータが埋め込まれた位置に行くことでその場でデータの再生が可能となる．これにより、過去にその空間を訪れたユーザの反応から、空間内の体験を楽しむための参考にすることが可能になると期待できる．

参考文献

- [1] R. Dachsel W. Büschel., A. Lehmann. Miria: A mixed reality toolkit for the in-situ visualization and analysis of spatio-temporal interaction data. In the

2021 CHI Conference on Human Factors in Computing Systems, No. 470, pp. 1–15, 2021.

- [2] R. Dachsel. J. Matejka. G. Fitzmaurice. F. Anderson P. Reipschläger., F. Brudy. Avatar: An immersive analysis environment for human motion data combining interactive 3d avatars and trajectories. In the 2022 CHI Conference on Human Factors in Computing Systems, No. 23, pp. 1–15, 2022.
- [3] R. Dachsel. J. Matejka. G. Fitzmaurice. F. Anderson P. Reipschläger., F. Brudy. Pearl: Physical environment based augmented reality lenses for in-situ human movement analysis. In the 2023 CHI Conference on Human Factors in Computing Systems, No. 381, pp. 1–15, 2023.
- [4] M. Becher. M. Sedlmair. D. Weiskopf. K. Kurzhals N. Pathmanathan., S. Öney. Been there, seen that: Visualization of movement and 3d eye tracking data from real-world environments. *Computer Graphics Forum*, Vol. 42, pp. 385–396, 2023.