



実世界 360 度動画とインタラクティブな VR 空間の融合

A Fusion of Real 360-degree Videos and Interactive Virtual Worlds

武縄 瑞基¹⁾, 杉本 直樹³⁾, Leslie Wöhler²⁾, 池畑 諭⁴⁾, 相澤清晴¹⁾

Mizuki TAKENAWA, Naoki SUGIMOTO, Leslie WÖHLER, Satoshi IKEHATA, and Kiyoharu AIZAWA

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

2) 東京大学・日本学術振興会 外国人特別研究員 (〒 113-8656 東京都文京区本郷 7-3-1 woehler@hal.t.u-tokyo.ac.jp)

3) MMMakerSugi (naoki48916@gmail.com)

4) National Institute of Informatics (〒 101-8430 東京都千代田区一ツ橋 2-1-2 satoshi.ikehata@nii.ac.jp)

概要: 本研究では現実世界を再現したインタラクティブな空間における探索を目的として、360 度動画と 3D の VR 空間を融合により写実的な VR 世界を生成するシステムを提案する。本システムでは道路を歩いて撮影した 360 度動画について、交差点による動画の分割と深層学習モデルを用いて映像以外の情報を付与することで、擬似的な VR 空間を作成する。さらに、360 度動画の描画面を楕円形にすることで、アバターの空間への自然な融合を目指した。

キーワード: 360 度動画, インターフェース, VR

1. はじめに

VR 空間を利用することで、ユーザは他のユーザとオンライン上で交流することができる。さらに実世界に存在する場所を再現した VR 空間では、ユーザ同士の交流だけでなく肉体的に遠く離れた場所の探索を体験することができる。例えば、渋谷を再現した VR 空間で行われたハロウィンのイベントでは 40 万人以上の来場者を集めるなど、現実世界での大規模なイベントをオンライン上に移行することに成功している [1]。しかし、現実世界を再現したような緻密な VR 世界の構築は多くの時間を必要とし、さらに現実感のある 3D 世界をデザインするには専門的な知識が必要である。また、動く生き物や車が存在しないため 3D モデルのみを用いることによる現実世界の再現には限界がある。この課題を解決するために、複雑な 3D モデルを構築する代わりに 360 度動画を使用することを考える。

360 度動画は 3D モデルとは異なり、専用のカメラの普及により容易に作成することができ、撮影した環境の写実的な印象をユーザに与えることができる。360 度動画は VR アプリケーションと相性が良く、多くのバーチャルツアーのサービスは観光地やその周辺風景の 360 度動画をを用いている。Google Street View (GSV) [2] では 360 度画像を使ってユーザに撮影箇所の風景を提供している。しかし、画像は離散的な位置に配置されているだけであり、動く物体が存在しないため撮影地の雰囲気や完全には伝えることはできない。一方で、Movie Map [3] は 360 度画像ではなく 360 度動画を用いて街の連続的な探索を可能にした。このシステムでは、道路を歩いて撮影した 360 度映像から交差点を検出し、ユーザが交差点で再生する動画を切り替えられるインター

フェースを作成している。Movie Map においては動画は自動的に再生されるため、ユーザは交差点で移動する方向を指定するという最小限の動作で、撮影された街を探索することができる。しかし、360 度動画は単なる動画であるためユーザには乏しいインタラクティブしか提供できず、また自由な空間探索ができない。

そこで、3D の VR 環境と 360 度動画を組み合わせることで互いの欠点を補いつつ、写実的でインタラクティブな低コストの VR 世界を生成することを目指す。街の風景のレンダリングには簡単に撮影可能である 360 度動画を用いることで精密な 3D オブジェクトの作成に必要な手間を省き、ユーザが操作可能な 3D アバターとその活動環境を用いることで 360 度動画に不足しているインタラクションを付与する。本論文では、360 度動画中でアバターが活動する際の様子に説得力を持たせるために、動画補完モデルによる撮影者除去と 360 度動画の投影面の変形による違和感の解消を提案する (図 1 (b), (c))。さらに、セマンティックセグメンテーションとその結果の白黒化によってアバターの行動範囲を決定し、擬似的な当たり判定を設定することを提案する (図 1 (d))。これらの手法によってユーザは 360 度動画でレンダリングされた写実的な VR 世界を探索することができる。一連の処理は自動化されているため従来の 3D の VR 世界とは異なり、簡単に写実的な VR 世界を構築可能となっている。

2. 関連研究

人による複雑な 3D 世界の作成を避けるために、360 度動画を用いたシステムが提案されている。Geollery [4] は GSV の画像を用いて 3D 都市モデルの外観をレンダリングし、

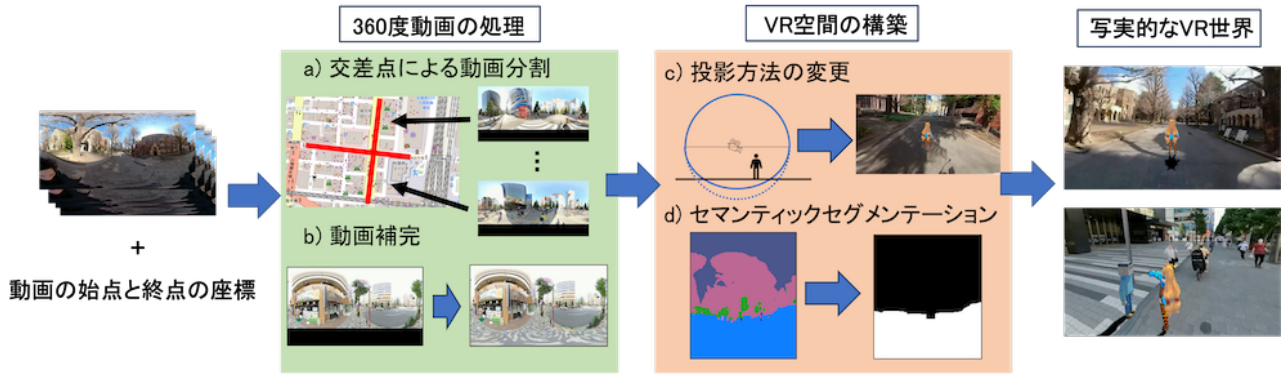


図 1: 提案システムのパイプライン. 道路を歩いて撮影した 360 度動画から, ユーザが自由に探索やインタラクションをすることができる VR 空間を構築する.

さらにソーシャルメディアの情報を付与することで, 写実的な外観を持つ街並みの中で複数ユーザが同時に街を歩き回ることを可能にした. このシステムによって作成された VR 世界では, アバターを用いてユーザ間の豊かなインタラクションが可能ではあるが, 画像は 3D モデルに貼り付けられているだけであるため現実世界を再現したとは言えない. 一方, Tourgether360 [5] では 360 度動画中にカメラの軌道とユーザの目線を表す目のようなアバターを表示している. このインターフェースにより, ユーザは同じ動画を視聴している他ユーザが注目している場所をリアルタイムで知ることができるため, 360 度動画において他ユーザとのインタラクションが容易になっている. しかし, これらのシステムは 360 度動画を再生しているため, 撮影された街の探索はカメラの軌跡上でしかできず, またユーザ側には動画の再生と停止を切り替える程度の能動性しか存在しない.

本論文では, 360 度動画に情報を付与してデジタルアバターによる自由な空間探索を組み合わせることで, 写実的な VR 世界を作成するシステムを提案する. 先行研究とは異なり, ユーザは歩いている人や動いている車を含む現実世界の映像の中でアバターを自由に動かして, 他のユーザと直接インタラクションを行うことができる. さらに, 本システムは 360 度動画とその動画の開始時と終了時のカメラの座標を入力として必要とするだけで, その後は人による処理を介さずに自動的に VR 環境を生成する.

3. 提案システム

図 1 に示すように, 本提案システムでは街の道路を歩いて撮影した 360 度動画と各動画の始点と終点における座標を入力として, 2 段階の処理を行うことで写実的な VR 空間を作成する.

3.1 作成する VR 空間の設定

提案するシステムによって構築される VR 空間は 360 度動画を用いるため, 3D モデルで設計された建物を用いて作られる VR 空間とは異なる. 本稿では, アバターとその歩くための地面のみを 3D オブジェクトとし, その他の建物や道のテクスチャなどを 360 度動画で代用する VR 空間を作成

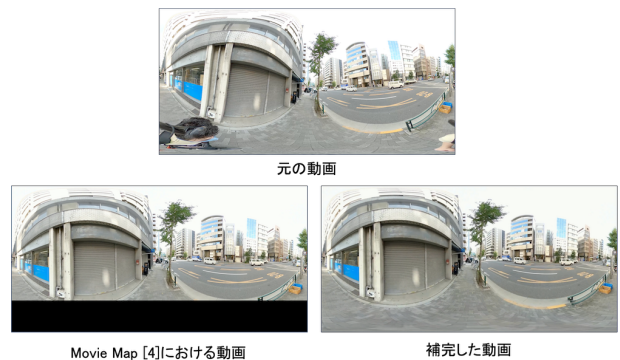


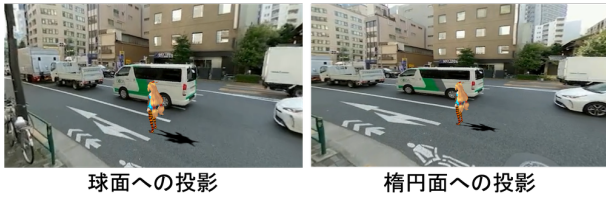
図 2: 補完の有無の比較. Movie Map ではスタビライズした動画を黒塗りしていたが, 本システムでは動画補完モデルによって穴埋めを行う.

することを目的とする. ユーザーが現実世界の街を探検しているように見せるため, アバターの前方に動く速度に合わせて球に投影した 360 度動画を再生する. また, 道路の交差点で映像を切り替えることで右折や左折などユーザの意図した方向への移動を実現し, アバターが自由に空間を探検することを可能にした.

3.2 360 度動画の処理

道路の交差点における映像の切り替えを実現するために, 用意した 360 度動画を交差点によって分割することを考える. この処理は Movie Map [3] に従って行う. まず, 動画をスタビライズしてカメラのブレの影響をなくす. 次に, OpenVSLAM [6] を用いて相対的なカメラの位置と回転を推定し, 入力として与えられた動画の始点と終点の座標に基づいて各フレームにカメラ座標を割り当てる (図 1 (a)). さらに, 割り当てた座標と特徴量を用いてある 2 つの動画で共通する交差点に対応する動画フレームのペアを検出する. 最後に, 全ての交差点について対応するフレームを検出し, その交差点フレームによって動画を分割する. 交差点において分割された動画を切り替えることで, ユーザは動画でレンダリングされた VR 世界を自由に探索することができる.

また, 実世界を撮影した 360 度動画を用いれば写実的な空間を構築可能であるが, 撮影者が必ず写り込んでしまうと



球面への投影

楕円面への投影

図 3: 各投影面におけるシーンの見え方.

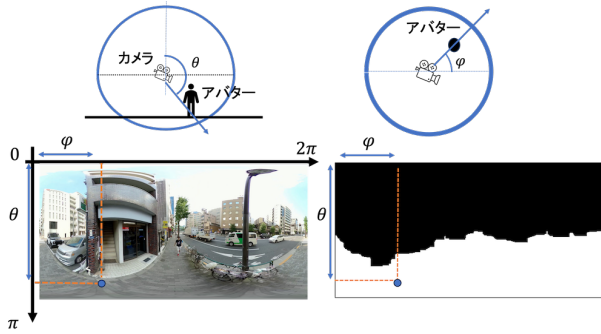


図 4: バイナリ化した画像を用いて道路を判定する方法.

いう問題がある。Movie Map [3] では撮影者を含む地面の一部を黒く塗りつぶすことで映像から除外している。さらに、一人称視点のインターフェースであるため、MovieMapではユーザが黒塗りされた地面の領域を意図して見ることは少ないと考えられる。しかし、3人称視点でアバターによる自由な空間探索を行うことを考慮した場合、黒塗りされた領域がユーザに知覚され、写実的な空間の探索という体験が損なわれる可能性があると考えられる。そこで、動画補完モデルを用いて黒塗り部分を補完して地面のテクスチャを復元する。具体的には、360度動画に動画補完モデル FGVC [7] を特殊な方法で適用する [8]。この補完ステップを追加することで、作成する VR 空間では撮影者が自然に除去された動画を用いることができる (図 2)。

3.3 360度動画を用いた VR 空間の構築

本システムは 360 度動画から VR 空間を生成するために、主に 2 つのステップを踏む。一つ目のステップでは、アバターを動画の風景に自然に溶け込ませるために、360 度動画の投影に通常用いられる球体への投影から、底が浅い楕円への投影に変更する (図 1 (c))。さらに、アバターが歩くための地面に相当する平面を設定する。ここでアバターが動画中に写っている人と同等の大きさに見えるようにするために、設定した地面と VR 空間中のカメラの距離に対するアバターの身長を動画撮影時のカメラの高さに対する人の身長に合わせる。図 3 に示すように、縮尺の調整によって違和感のなくアバターを追加できる上に、楕円面への投影によってアバターが動画中の地面により接地しているように感じることができる。

2 つ目のステップでは、360 度動画に擬似的な当たり判定を設定するためにアバターの移動可能領域をセマンティックセグメンテーションによって決定する。セグメンテーショ



図 5: アバターの移動制限がない場合における物体との重畳.

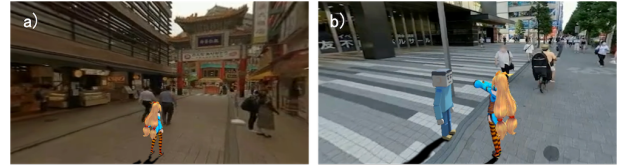


図 6: 本システムによって生成された VR 世界のアプリケーション.

ンモデルによって得られた各ピクセルの属するクラスについて、道路とそれ以外でバイナリ画像化をすることでアバターが移動できる地面領域を取得する (図 1 (d))。作成したバイナリ画像は 360 度画像の Equirectangular projection (ERP) の形式で手に入るため、アバターの 360 度画像上の座標を取得することで移動の可否を判定可能である。具体的には図 4 に示すように、360 度画像が投影されている球の中心からアバターの足元へと向かう半直線と球の交点の緯度と経度を計算することで、ERP 上の座標を求めることができる。その座標におけるバイナリ画像の値を参照することでアバターの行動を制限する。このバイナリ画像による行動範囲の設定を行わない場合、3D モデルで作られた VR 空間では明らかに存在できない場所に移動することができてしまい、図 5 に示すように車や建物と重畳して表示されるなどユーザに違和感を与えてしまうことがある。そのため、バイナリ画像による移動制限は重要である。しかし、厳密に道路とそれ以外で分離した場合は、例えば多くの人が歩いている道などでアバターがスムーズに移動できないことが生じるため、人も道路の一部として移動可能範囲に設定してユーザの快適な操作を優先している。なお、本システムでは Orphans らのセマンティックセグメンテーションモデル [9] を適用している。

4. アプリケーション

本システムによって作られる VR 世界は 360 度動画を用いているため、現実世界を写実的に再現した街を構築することができる。そのため、ユーザは世界に存在する有名な観光地をアバターを用いて能動的に探索することができる (図 6 (a))。さらに、ユーザが操作するアバター以外の 3D オブジェクトを追加することができるため、他のユーザとのインタラクション (図 6 (b)) やユーザが映像中のコンテンツに対して広告や脚注を行うことも可能である。

5. むすび

本稿では写実的な VR 空間の生成を目的として, 360 度動画を拡張してアバターが活動できる VR 世界を作成するシステムを提案した. 本システムは 360 度動画の撮影者除去や投影面の変更, アバターの移動範囲の設定をすることでユーザに違和感のない空間を提供している. ユーザは作成された空間において, 現実世界に存在する街を探検したり他ユーザとインタラクションを楽しむことができる. 今後の展望としては, 360 度動画を VR 空間化する際に用いた手法の効果を被験者実験によって評価することが挙げられる.

謝辞 本研究の一部は, JST-Mirai Program JPMJMI21H1, JSPS KAKENHI 21H03460 の支援をうけた.

参考文献

- [1] Ayumi Sugiyama. Virtual shibuya halloween event ups its game with personal avatars. <https://www.asahi.com/ajw/articles/14460054>, 2021. 最終閲覧日: 2023 年 7 月 21 日.
- [2] Google LLC. Google street view. <https://www.google.com/streetview/>, 2023. 最終閲覧日: 2023 年 7 月 21 日.
- [3] Naoki Sugimoto, Yoshihito Ebine, and Kiyoharu Aizawa. Building movie map - a tool for exploring areas in a city - and its evaluations. In *ACMMM*, 2020.
- [4] Ruofei Du, David Li, and Amitabh Varshney. Geollery: A mixed reality social media platform. In *CHI*, 2019.
- [5] Kartikaeya Kumar, Lev Poretski, Jiannan Li, and Anthony Tang. Tourgether360: Collaborative exploration of 360° videos using pseudo-spatial navigation. *CHI*, 2022.
- [6] Shinya Sumikura, Mikiya Shibuya, and Ken Sakurada. Openslam: A versatile visual slam framework. In *ACMMM*, 2019.
- [7] Chen Gao, Ayush Saraf, Jia-Bin Huang, and Johannes Kopf. Flow-edge guided video completion. In *ECCV*, 2020.
- [8] 相澤清晴 武縄瑞基, 池畑諭. 360° 動画の補完 -入力回転による性能向上. In *MIRU*, 2023.
- [9] Semih Orhan and Yalin Bastanlar. Semantic segmentation of outdoor panoramic images. *Signal, Image and Video Processing*, 2022.