



建物外周データと 360 度映像による バーチャル都市空間の構築

川田高輝¹⁾, 坂野達郎¹⁾, Leslie Wöhler¹⁾, 池畑諭²⁾, 相澤清晴¹⁾

1) 東京大学

2) 国立情報学研究所

概要

現実の都市空間を再現したバーチャル空間を構築する方法として、実際の都市を撮影した 360 度映像を、アバターの移動に応じて 3D 都市モデルに動的に投影する手法がある。このような手法は、現実の遠近感や建築物の 3D 構造を活かした高度な拡張現実が可能になる一方で、3D 都市モデルのデータが存在しない地域には対応できない。そこで本研究では、全国的に整備されている建物の 2 次元外周データを活用し、そこから擬似的に作成した 3D モデルと 360 度映像を組み合わせることで、都市のバーチャル空間を構築する新たな手法を提案する。そのために、360 度映像から得られた建物 3D 点群と、建物の 2 次元外周データを平面上で位置合わせすることで、映像と 3D モデルの位置合わせを実現した (図 1)。

1. はじめに

近年、仮想現実技術により都市空間を仮想的に体験することが可能となり、その中で 360 度映像技術は、都市空間のキャプチャと没入型体験において重要な役割を果たしている。閲覧時は撮影された映像内でユーザが自由に視点を動かすことができるため、没入的な体験を実現できる。

また、近年、現実の都市の 3D モデルを構築する動きが広がっており、日本では国土交通省が、全国各都市の 3D モデルを整備する PLATEAU プロジェクト [1] を推進している。CityGML [2] 形式のデータで、ジオメトリに加えて建物の属性情報 (建物利用現況・洪水浸水想定区域など) を統合して扱うことができる。

坂野らは、この PLATEAU モデルに対して 360 度映像を動的にテクスチャ投影することで、360 度映像の写実性と PLATEAU の幾何・属性情報を兼ね備えたバーチャル空間を構築する、360CityGML [3] を提案した。この手法では、アバターの移動に応じて、対応する 360 度映像フレームが都市モデルに動的に投影されることで、ユーザが空間内を自由に探索することができる。ここではアバターが移動した際に視点が最近傍のカメラのフレーム位置に固定されることで、外観情報の整合性を確保している。

一方で、PLATEAU モデルのない都市では、坂野らの実装した 360CityGML でバーチャル都市空間を作成できない。モデル作成に多大な時間と費用を要し、PLATEAU は開始から 4 年で現在約 200 都市のデータが整備されるという拡大速度で、全国のあらゆる都市での整備には程遠い。

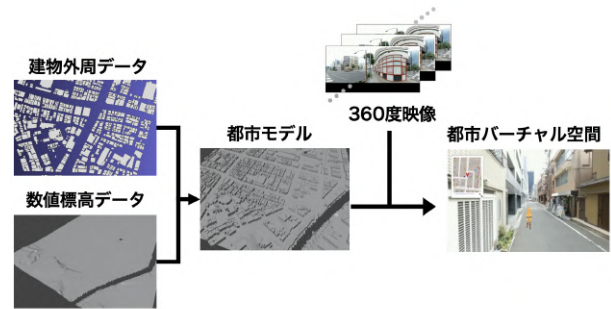


図 1: 本研究での都市空間の作成の流れ。アバターのモデルには Unity-Chan を用いた。©UTJ/UCL

そこで本研究では、都市を撮影した 360 映像と、国土地理院が公開している建物の 2 次元外周データ [4] から、坂野らの手法に準じた都市のバーチャル空間を構築する手法を提案する。このデータは日本全土の約 98% [5] をカバーしている。360 度映像から得られた建物 3D 点群を 2D 平面に射影し、2 次元外周データと平面上で位置合わせすることで映像と 3D モデルの位置合わせを実現できる。本研究の貢献は以下の 2 点である。

- 都市モデルに依存せずに、360 度歩行映像と公開された建物外周データを用いて、全国の地域の都市バーチャル空間を作成可能にした。
- 詳細なモデルを必要とせず、少ない都市情報量の中から坂野ら [3] と同様に、地理情報とバーチャル空間の融合による拡張現実が可能になった。

2. 関連研究

2.1 都市モデルへの静的なテクスチャマッピング

3D 都市モデルに離散的な歩行者視点での画像群を投影することで実世界に即した 3D 都市空間を可視化でき、地理空間データの理解に役立つ。Geollery[6] や Park ら [7] の研究では都市モデルに対して限られた画像群を直接、あるいは複数視点の画像を混合してテクスチャ投影を行っている。しかし、3D 都市モデルは箱型の単純な幾何構造でモデリングされることが多いため、この複数視点による静的なマッピングでは可視化の際に外観情報での不整合が起こるという問題がある。本研究では、360CityGML と同様な手法で、アバターの位置に応じて対応する 360 度映像フレームを選択

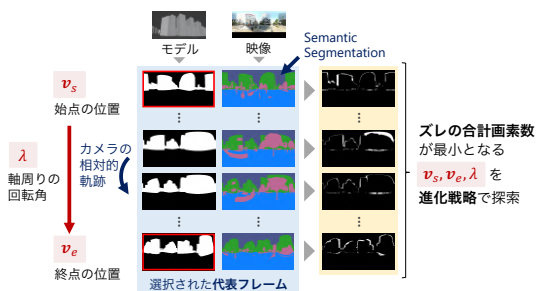


図 2: 360 度映像と 3D 都市モデルの位置合わせ [3]

し、それに応じてカメラ視点を中心として都市モデルと外側に配置した球面に動的に投影することで、不整合の起こらない写実的なバーチャル都市空間を実現した。

2.2 360 度映像群のみを用いた歩行者視点の都市空間

実際の都市を撮影した 360 度映像群のみを用いることで、写実的な都市バーチャル空間を構築することが可能である。Tourgether[8] では、撮影された 360 度映像を用いてユーザーに撮影者と同じ視覚体験を提供できる一方で、撮影されたカメラの軌跡上での視点の変更しかできないという問題がある。一方 360RVW [9] では、3D 空間にアバターを用意し、アバターにカメラが追従する形で、カメラを中心とした天球面にアバターの位置に応じた 360 度映像フレームを投影する。実際に都市を撮影した 360 度歩行映像群とカメラの開始点と終了点の緯度・経度を用いて、360 度映像の各フレームに実世界での位置情報を付与する。しかし、各フレームの位置情報は、OpenVSLAM[10] により推定されたカメラの相対的な軌跡と、手動で指定された始点および終点の位置情報に依存しており、不正確である。また、360 度映像を球面に投影しているため、3 次元的な幾何構造を直接的に表現することができず、結果として実際の都市の構造を反映した拡張現実を提供することが難しいという課題がある。これらの課題を解決するために、現実世界の建物や地形を再現した 3D 都市モデルを導入するアプローチが 360CityGML によって提案された。

2.3 360 度映像と都市モデルの位置合わせ

3D 都市モデルに 360 度映像を動的に投影する際に、実世界の絶対座標系におけるカメラの軌跡の位置情報とカメラの向きを決定する必要がある。360CityGML では、OpenVSLAM によって推定された相対的なカメラの軌跡を都市モデルに合うように最適化する手法を提案している。まず、元データの 360 度映像における建物の領域と、得られた軌跡から実際の 3D 都市モデルを見た時のフレーム上における建物の領域を Semantic Segmentation を用いて抽出し、ずれを計算する (図 2)。そのずれを評価値として、カメラの始点の座標 $\mathbf{v}_s = (x_s, y_s, z_s)$ 、終点の座標 $\mathbf{v}_e = (x_e, y_e, z_e)$ 、始点と終点を結んだ直線を回転軸とした、重力方向に対する軸周りの回転角 λ の 7 個のパラメータを Covariance Matrix Adaptation Evolution Strategy (CMA-ES)[11] を使って評価値が最小となるように最適化を行う。



図 3: 建物と地形モデル

しかし、この手法は建物の形状を活用して行うため、3D 都市モデルが公開されていない地域に対しては適用できない。そこで本研究は、全国的に存在する 2D の建物外周データから簡易的な都市モデルを作り、映像から取得した建物の 3D 点群を 2D 平面に射影し、2 次元外周データとの位置合わせを行うことで、映像と 3D モデルの位置合わせを実現する。

3. 提案手法

本手法は、対象地域を網羅的に歩いて撮影した 360 度映像群と、各映像の始点と終点の緯度・経度の情報、そして対象地域の建物の 2 次元外周データ、および地形の標高データを入力として、都市バーチャル空間を構築する。

3.1 簡易的な都市モデルの作成

まず、国土地理院 [4] にて公開されている建物の外周線のデータを用いて建物底面のポリゴンを形成する。PLATEAU のような都市モデルはこのような 2 次元地図情報に加え、航空写真や上空から取得した点群を用いて建物上部の複雑な形状や建物の高さを特定し、作成される。PLATEAU の大部分を LOD1 という建物の底面に対して高さ情報を加えた箱型モデルが占める。そのため、今回は外周線に一定の高さを加え LOD1 モデルを作成する。本研究では建物の高さを 20m に設定した。3D モデルは映像の投影先だけでなく衝突判定、建物の壁面クリック機能などに必要となるため、低層部から中層部の幾何構造があれば問題なく機能する。また、都市モデルの外側に球面を配置しているため、作成したモデルの高さによって外観情報の不整合が起きることはない。

次に国土地理院にて公開されている 5 m 間隔の数値標高データ (DEM データ) を用いて地形モデルを作成する。図 3 のように外周データで作成した各建物モデルと地形モデルを組み合わせて、簡易的な 3D 都市モデルが完成する。

3.2 360 度映像からの建物 3D 点群の抽出

まず vSLAM[10] でカメラの軌跡を推定し、映像内で検出された 3D 点群も同時に取得する。作成した建物モデルと位置合わせをするために、vSLAM 時に得られた点群から建物以外の遮蔽物の点群を除去する必要がある。そこで、本研究では、図 4 に示すように、VSLAM を用いて映像内で検出されたスパースな点群を解析し、建物領域に属する点のみを抽出する手法を提案した。建物点群の抽出の手順は以下の通りである。

まず、映像をフレームに分割し、各フレーム画像に Mask2F

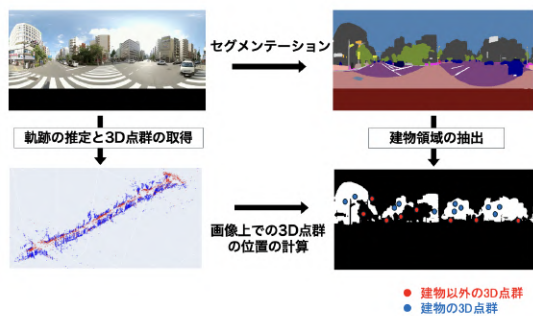


図 4: 建物点群の抽出の流れ

ormer[12] による Semantic Segmentation を用いて、図 4 に示すようなマスク画像を生成する。次に、図 4 に示すように、マスク画像を建物とそれ以外の領域で二値化する。その後、検出された全ての特徴点について、フレーム画像上の 2 次元位置が二値化された建物領域に属しているかをピクセル単位で調べる。この結果、建物領域に含まれる特徴点のみを採用し、含まれない特徴点を除去する。

3.3 点群を用いた 360 度映像とモデルの位置合わせ

建物の 3D 点群を 2D 平面に投影することで、建物の高さに依存せずに建物の 2 次元外周データとの位置合わせを行う手法を提案する。3.2 で得た 3D 点群を 3.1 で作成した都市モデルにフィッティングすることによってカメラの軌跡を最適化する。都市モデルの座標系の中での絶対的軌跡を決定する 7 つのパラメータについて、映像の始点の座標を (x_s, y_s, z_s) 、終点の座標を (x_e, y_e, z_e) 、始点と終点を結んだ軸周りのカメラの回転角を λ とした時、最適化を以下の手順で実行を行う。

まず、映像の始点と終点の緯度・経度を用いてカメラの軌跡と建物点群を平面直角座標系 [13] に変換し、都市モデルと同じ空間座標系に配置する。次に点群を 2D 平面に射影し、各点ごとに最近傍の建物の底面外周との最短距離を計算する。その全ての点群に対する合計値を点群数で正規化した値を評価値とする。そして、Optuna[14] の CMA-ES を用いて $x_s, y_s, x_e, y_e, \lambda$ の 5 つのパラメータを変化させて軌跡を変換し、同時にカメラの向きと点群も変化させる。評価値が最小となるようにこれらを繰り返しパラメータ探索を行う。軌跡を決定するパラメーターのうち始点と終点の xy 平面での位置、および 3 次元空間におけるカメラの方向を、 xy 平面における点群と建物外周のフィッティングにより決定する。そして、映像は地面からの高さが 2 m になるように撮影されているので、作成した地形モデルを用いて、カメラの高さが地面から常に一定の高さになるように z 成分を補正することで、都市モデル空間におけるカメラの軌跡を決定する。

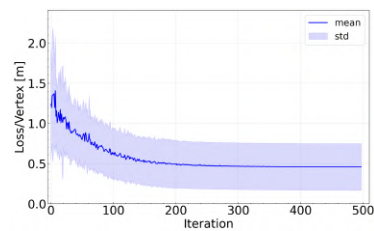
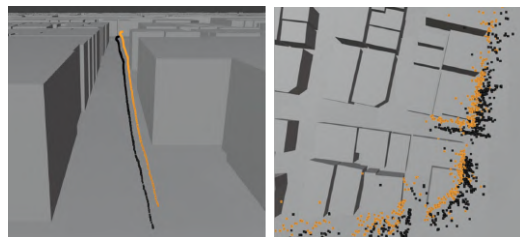


図 5: 最適化でのイテレーション数と評価値。青線は平均値、薄青色の領域は標準偏差を示す。



(a) カメラの軌跡の変化 (b) 建物 3D 点群の変化

図 6: 最適化前後のカメラの軌跡と建物 3D 点群の変化。黒線・黒点が最適化前、橙線・橙点が最適化後を示す。

4. 実験

本実験では、秋葉原の一部の地域を対象として、提案手法に沿って Unity [15] で都市パース空間の構築を行う。計 37 本の映像群を利用し、3.1 で説明したように、都市モデルの代わりに対象地域の建物外周データと数値標高モデルをデータとして使用した。

4.1 点群を用いたカメラの軌跡の最適化

360 度映像と各映像の始点と終点の緯度・経度の情報、作成した都市モデルを用いて、3.3 で提案した 360 度映像と 3D 都市モデルの位置合わせを実行した。ここで、最適化の試行回数を 500 回とし、各パラメータの探索空間は、 $\Delta x_s, \Delta y_s, \Delta x_e, \Delta y_e$ は ± 10 m, $\Delta \lambda$ は $\pm 10^\circ$ とした。イテレーションごとの建物の点群と建物外周との 2 次元上でのずれの評価値について、使用した全映像での平均をとった値の推移を図 5 に示す。最適化の様子を見ると、評価値が小さくなり、300 回目の試行の時点で評価値が収束していることが分かる。さらに、最適化前後におけるカメラの軌跡の変化と、それに伴った 3D 点群の変化を部分的に図 6 に示した。建物の点群がより建物の外周に近づいて配置されるように軌跡のパラメータが変化しているのが分かる。

4.2 3D 都市構造を生かした拡張現実

360 度映像と 3D 都市モデルの位置合わせで得た、360 度映像のカメラの軌跡と姿勢の情報を用いて、3D 都市モデルと外側の球面に対する 360 度映像の動的なテクスチャマッピングを行った。図 7 のように、360CityGML[3] と同様にバーチャル空間内に洪水の水面の CG を配置した。図 7a は 360RVW [9] のようにカメラを中心とした球面に映像を投



(a) 球面のみ (b) 都市モデルと球面

図 7: バーチャル空間内での映像の動的なテクスチャマッピング。赤線が水面と都市の外観との境界線を示す。



図 8: 2D 地図と都市バーチャル空間の対応。壁面クリック時に 2D 地図上でその建物の位置が表示される

影した場合、建物の幾何構造が存在せず、外観の不整合が起きている。一方、図 7b では、球面内に建物モデルを配置しており、水面の CG が 3 次元的に整合した形になっている。本研究で作成したモデルと最適化した軌跡に既存のマッピング手法を適用することで、現実世界の構造を反映した外観を再現できる。

また、洪水を可視化する機能に加え、建物のモデル構造を活用した機能を図 8 に示した。建物部分をバーチャル空間上でクリックすることで左上の 2 次元上の地図における位置が表示される。これにより、バーチャル空間における位置情報と実際の 2 次元地図情報が接続され、ユーザーが対応関係を直感的に理解できるようになる。

5. まとめ

本研究¹では、現実の都市空間を再現しつつ、拡張現実への応用ができるバーチャル空間を 2 次元の建物外周情報と 360 度映像から構築する新たな手法を提案した。PLATEAU などの都市モデルがない地域でも対応できるように、映像と建物 2D 外周データから作成を行い、映像から得た建物点群と建物底面の 2D 上でのフィッティングにより位置合わせの最適化を行った。この手法により、少ない都市情報量の中から坂野らと同様の 3D 都市を自由に探索できる空間を再現した。建物の幾何構造を生かした拡張現実が可能となり、新たな応用が期待される。

¹ 本研究の一部は、内閣府 SIP スマート防災および JSPS 25H01164 の支援を受けた。

参考文献

- [1] 国土交通省. PLATEAU. <https://www.mlit.go.jp/plateau/>.
- [2] G. Gröger and L. Plümer. Citygml-interoperable semantic 3d city models. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 71, pp. 12–33, 2012.
- [3] 坂野達郎, 武縄瑞基, L. Woehler, 池畑諭, 相澤清晴. 実世界 360 度映像と CityGML を融合したバーチャル空間の構築. *MIRU*, 2024.
- [4] 国土地理院. 基盤地図情報ダウンロードサービス. <https://fgd.gsi.go.jp/download/menu.php>.
- [5] 国土交通省. 地理空間情報の整備、提供、活用. <https://www.mlit.go.jp/common/001033766.pdf>.
- [6] e. a. Du, Ruofei. Geollery: A mixed reality social media platform. In *proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–13, 2019.
- [7] J. Park, I.-B. Jeon, S.-E. Yoon, and W. Woo. Instant panoramic texture mapping with semantic object matching for large-scale urban scene reproduction. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 27, No. 5, pp. 2746–2756, 2021.
- [8] e. a. Chang, Yung-Ju. Tourgether: Exploring tourists' real-time sharing of experiences as a means of encouraging point-of-interest exploration. In *the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 3, No. 4, pp. 1–25, 2019.
- [9] e. a. Takenawa, Mizuki. 360rvw: Fusing real 360° videos and interactive virtual worlds. In *ACMMM*, 2023.
- [10] e. a. Sumikura, Shinya. Openslam: A versatile visual slam framework. In *the 27th ACM International Conference on Multimedia*, pp. 2292–2295, 2019.
- [11] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, Vol. 9, No. 2, pp. 159–195, 2001.
- [12] e. a. Cheng, Bowen. Masked-attention mask transformer for universal image segmentation. In *CVPR*, 2022.
- [13] 国土地理院. わかりやすい平面直角座標系. <https://www.gsi.go.jp/sokuchikijun/jpc.html>.
- [14] e. a. Akiba, Takuya. Optuna: A next-generation hyperparameter optimization framework. In *the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2623–2631, 2019.
- [15] Unity Technologies. Unity. <https://unity.com>, 2024.