

360CityGML: 実世界 360 度映像と CityGML を融合した 都市空間可視化システムの構築と評価

360CityGML: Construction and Evaluation of an Urban Visualization
System Integrating 360° Videos and CityGML Model

坂野達郎¹⁾, 武縄瑞基¹⁾, 杉本直樹²⁾, Wöhler Leslie¹⁾, 池畑諭³⁾, 相澤清晴¹⁾

Tatsuro BANNO, Mizuki TAKENAWA, Naoki SUGIMOTO, Wöhler LESLIE, Satoshi IKEHATA, and
Kiyoharu AIZAWA

1) 東京大学 (〒 113-0033 東京都文京区本郷 7-3-1, banno, takenawa, woehler, aizawa@hal.t.u-tokyo.ac.jp)

2) MMMakerSugi (naoki48916@gmail.com)

3) 国立情報学研究所 (〒 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, 3D 都市モデル (CityGML) と街中の 360 度映像を融合した都市空間可視化システム 360CityGML を提案する. これにより, ユーザーはアバターを操作して写実的な都市空間を探索し, 歩行者視点から地理情報を把握することができる. この構築に向け, モデルと映像の位置合わせ手法, およびアバター位置に応じたモデルへの映像の動的なテクスチャ投影を提案し, ユーザー実験でシステムの評価を行なった.

キーワード: 360 度映像, CityGML, 可視化システム

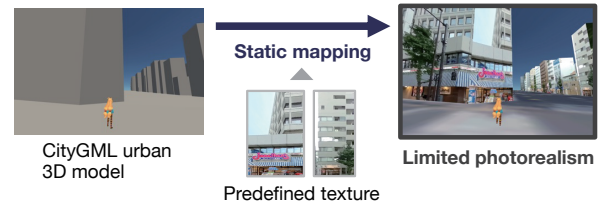
1. はじめに

近年, 都市の 3D モデルにおける国際標準として, CityGML (City Geography Markup Language) が注目を集めている [1]. CityGML は, 建物の 3D 形状に加えて, 各建物の属性情報や洪水リスクなど, 多様な地理空間データを統一された枠組みで扱える点が特徴である. この特性により, CityGML は都市空間の総合的な情報基盤として有用であり, 世界中の都市で導入が進められている [2, 3].

このような地理空間データをユーザーが直感的に理解する手段として, 近年, 歩行者視点から没入的に地理情報を可視化するシステムが数多く提案されている [4]. このような可視化においてユーザーの臨場感を高めるためには, 高い写実性が不可欠である. その実現手段として, 従来は 3D モデルの各面に対応する静的なテクスチャ画像を準備し, それをモデルに貼り付けて表示する手法が検討されてきた [5]. しかし, 公開されている多くの CityGML モデルは, LOD1(Level Of Detail 1) と呼ばれる直方体を組み合わせた単純なジオメトリにとどまっているため [6], この手法ではテクスチャ画像と単純なモデル形状の不整合が顕著となり, 図 1 (上) に示すような写実性に欠ける可視化結果となるという問題がある.

一方, 都市空間を写実的に表現する手法として, 近年, 街中を歩きながら撮影した 360 度映像を活用するアプローチが提案されている [7, 8]. これらのシステムは, 撮影された実世界の映像をそのまま表示するため, 極めて高い写実性

Common approach



360CityGML (ours)

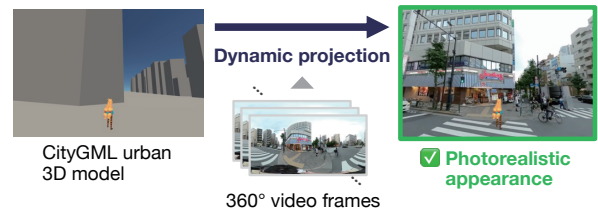


図 1: 従来の各面に対する静的テクスチャマッピング [5] (上) と, 本研究が提案する 360 度映像フレームの動的テクスチャ投影 (下) の比較.

を実現できる. また, 通常の動画とは異なり周囲全体を撮影しているため, ユーザーは映像内で視点を自由に変更することで, 没入感の高い体験を得られる. 一方で, こうした映像に基づくシステムは, シーンの三次元的な幾何情報を持たないため, そのままでは空間的な地理情報と統合した可視化に用いるのは困難である.

そこで本研究では, CityGML モデルと 360 度映像を融

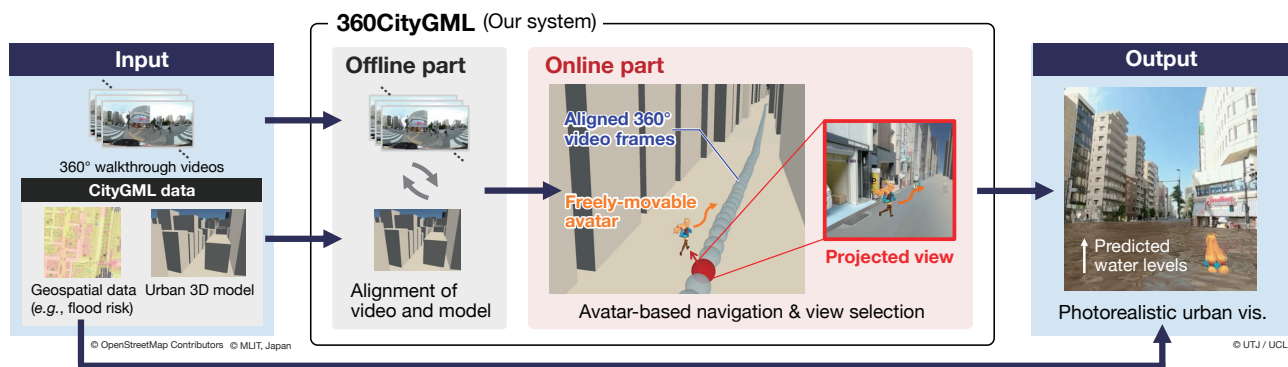


図 2: 提案システムの全体パイプライン。

合した都市空間可視化システムである 360CityGML を提案する。本システムではまず、360 度映像と CityGML モデルに対し、建物の領域情報を用いた位置合わせを行い、両者を空間的に整合する。続いて、空間内にユーザーが操作可能なアバターを導入し、そのアバターの位置に応じて対応する 360 度映像のフレームを建物モデルに動的にテクスチャ投影する。これにより、従来の静的なテクスチャ画像に見られる単純なジオメトリとの不整合が回避され、図 1 (下) に示すような写実的な都市空間が実現される。

こうして構築された都市空間を用いることで、CityGML に含まれる地理空間データを、実際の風景と統合した形で三次元的に可視化することができる。本研究では、日本の秋葉原地区を対象に都市空間を構築し、その中で洪水リスクや日照条件を可視化するシステムを実装した。そして、洪水リスクの可視化に対して、360 度映像による写実的な体験が地理情報の理解に与える影響を評価するユーザースタディを実施し、提案システムの有効性を検証した。

2. 提案手法

提案システムの全体パイプラインを図 2 に示す。本システムは、対象地域の各道路を歩きながら撮影した 360 度映像群と、それぞれの映像に対応する始点・終点の緯度経度情報、および対象地域の CityGML モデルを入力とする。まず、オフライン処理で 360 度映像と CityGML モデルの位置合わせを行い、両者を空間的に整合させる。続くオンライン処理では、ユーザーが操作するアバターの現在位置に応じて最適な映像フレームを動的に選択し、その画像を CityGML モデルにテクスチャ投影する。これにより、映像フレームを用いた写実的な都市空間が構築され、CityGML に含まれる地理空間データの没入的な可視化が可能となる。以下では、都市空間構築における各処理の詳細を述べる。

2.1 360 度映像と CityGML モデルの位置合わせ (オフライン処理)

本処理では、360 度映像の各フレームが撮影された際のカメラの位置 $\mathbf{x} \in \mathbb{R}^3$ および姿勢 $\boldsymbol{\theta} \in \mathbb{R}^3$ を、現実世界の緯度・経度・高度に対応する CityGML モデルのグローバル座標系内で正確に推定する。従来の位置合わせ手法 [9] では各フレームごとに位置と姿勢の最適化を行うが、都市ス

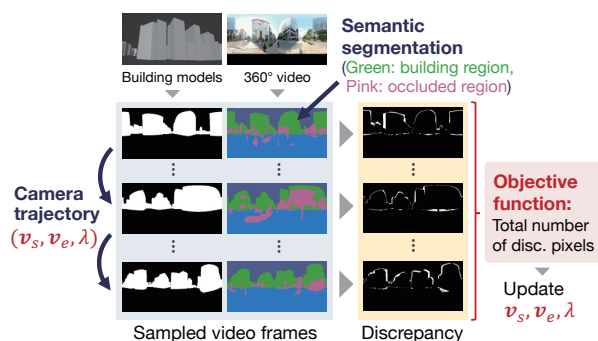


図 3: 360 度映像と CityGML モデルの位置合わせにおけるパラメータ最適化の概要。

ケールの長時間映像では計算コストが高く実用的ではない。

そこで本研究では、各映像単位で得られるカメラの相対的な軌跡を用いた、よりスケーラブルな位置合わせ手法を提案する。具体的には、まず各 360 度映像に対して Visual SLAM [10] を適用し、ローカル座標系における各フレームのカメラ位置 $\mathbf{x}_{\text{local}} \in \mathbb{R}^3$ および姿勢 $\boldsymbol{\theta}_{\text{local}} \in \mathbb{R}^3$ を得る。次に、このローカルなカメラ軌跡を CityGML のグローバル座標系にマッピングする変換行列を推定する。この変換は、カメラ軌跡の始点位置ベクトル $\mathbf{v}_s = (x_s, y_s, z_s)$ 、終点位置ベクトル $\mathbf{v}_e = (x_e, y_e, z_e)$ 、および重力方向まわりの回転角 λ の 7 つのパラメータによって一意に定まる。このため、これらのパラメータ $(\mathbf{v}_s, \mathbf{v}_e, \lambda)$ を軌跡ごとに最適化することで、グローバル座標系における映像の各フレームのカメラ位置 $\mathbf{x} \in \mathbb{R}^3$ および姿勢 $\boldsymbol{\theta} \in \mathbb{R}^3$ を一括して推定することができる。

パラメータ最適化の概要を図 3 に示す。パラメータの初期値として、 x_s, y_s, x_e, y_e には手動で与えた始点・終点の緯度経度を、 z_s, z_e には、CityGML モデルの地面の高さにカメラの高さ (実験では 2m) を加えた値を用いる。 λ の初期値は、SLAM で推定された重力方向が正確であると仮定し 0° に設定する。これらの初期値をもとに、映像と CityGML モデルのずれが最小となるよう、パラメータ $(\mathbf{v}_s, \mathbf{v}_e, \lambda)$ を最適化することで、両者の位置合わせを実現する。

映像とモデルのずれの定量化には、従来手法 [9] で用いら

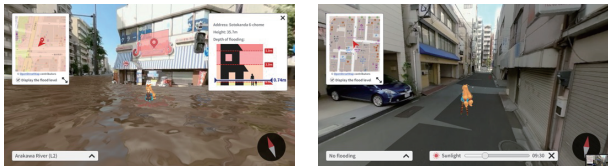


図 4: 本システムの地理情報可視化への応用例. (左) 洪水リスク可視化の例, (右) 日照条件可視化の例.

れた建物の領域情報に基づく目的関数 [9] を映像単位に拡張したものを用いる. まず, 映像から等間隔に 30 フレームをサンプリングし, 各フレームに Semantic Segmentation [11] を適用して建物領域を抽出する. 次に, 現在のパラメータ (v_s, v_e, λ) でグローバル座標系のカメラ軌跡を推定し, 各フレームに仮想的な 360 度カメラを配置して, CityGML の建物モデルを 360 度カメラビューに投影する. そして, 映像フレームと投影結果における建物領域の不一致画素数を算出し, 全サンプルフレームで合計した値を目的関数とする. この目的関数を進化戦略の CMA-ES [12] で最小化することで, 最適なパラメータを決定し, 映像と CityGML モデルの位置合わせを実現する.

2.2 360 度映像の動的なテクスチャマッピング (オンライン処理)

本処理では, 映像テクスチャと幾何形状との位置ずれを最小化するため, カメラ視点を 360 度映像のカメラ軌跡上に制限する. この制約のもとで, ユーザーはアバターを操作して都市空間内を仮想的に歩き回ることができ, システムはそのアバターの現在位置および移動方向に応じて, 表示すべき 360 度映像フレーム (カメラ視点) およびその球面上の視線方向を動的に選択する. 選択された映像フレームは, 対応するカメラ位置から CityGML の建物モデル上へテクスチャ投影され, 写実的な都市空間を実現する.

ここでカメラ視点は, 現在表示中の i 番目のフレームであるとし, カメラとアバター間で維持すべき距離を α (実験では 5m), カメラ軌跡の進行方向に沿って測定されるカメラからアバターまでの符号付き距離を d としたとき, $d > \alpha$ の場合には $(i+1)$ 番目のフレームに, $d < -\alpha$ の場合には $(i-1)$ 番目のフレームに切り替えることで決定される. また, 視線方向は, カメラ位置からアバター位置を結ぶ方向ベクトルに合致するよう設定される. ただし, カメラの進行方向とアバターの向きが一致しない場合, 視線に不連続な変動が生じ, 視覚的な揺れとして認識される可能性がある. そこで, 視線方向に対して, 前時刻ステップの視線方向との移動平均フィルタを適用し, 視線ベクトルの時間的な変化を平滑化することで, 急激な方向変化を抑制する.

3. ユーザスタディによる評価

3.1 地理情報可視化システムへの応用

本システムで構築される都市空間を用いることで, さまざまな地理空間データを歩行者視点から没入的に可視化することができる. この有効性を検証するために, 日本の秋葉原地域



図 5: 各視覚化条件における洪水リスク可視化の例. (左) non-texture, (中央) static-texture, (右) dynamic-texture.

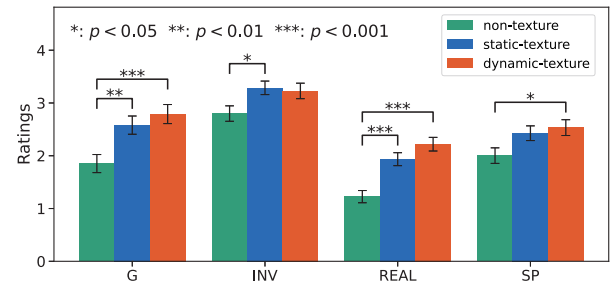


図 6: 各視覚化条件における IPQ [13] の各指標のスコア. 一般的な臨場感 (G), 没入感 (INV), 現実感 (REAL), 空間的な臨場感 (SP) の 4 指標で評価した.

において撮影された 229 本の 360 度映像と, PLATEAU [2] が提供する同地域の CityGML モデルを用いて都市空間を構築し, 洪水リスクと日照条件の 2 種類の地理情報可視化を実装した. これらの実装結果を図 4 に示す. いずれの事例においても, 地理空間データと現実世界の外観とが三次元的に整合された形で統合されており, 全体として没入感の高い可視化表現が実現されていることが確認される.

3.2 ユーザスタディによる地理情報可視化での写実性の有効性の検証

本システムが提供する 360 度映像による高い写実性は, ユーザーのシステムへの臨場感を高め, 結果として地理情報の直感的な理解を促進することが期待される. この仮説を検証するために, 秋葉原における洪水リスク可視化を対象として, 以下に示すユーザスタディを実施した.

3.2.1 刺激条件

本実験では, 秋葉原地域から 9 本の街路を選定し, 各街路に対して 3 種類の視覚化条件を設定した. 1 つ目は, テクスチャなしの 3D 都市モデルをそのまま用いる **Non-texture** 条件, 2 つ目は, 3D 都市モデルの各面に静的な画像テクスチャを貼付して表示する **Static-texture** 条件, 3 つ目は, 本研究の提案手法である 360 度映像の動的テクスチャ投影を用いた **Dynamic-texture** 条件である. 各視覚化条件における洪水リスク可視化の例を図 5 に示した.

3.2.2 実験設計

本実験の実施のために, 27 名 (男性 16 名, 女性 11 名, 平均年齢 28.11 歳, 標準偏差 10.46 歳) の被験者を集めた. 各参加者は対象の 9 本の街路を, あらかじめ割り当てられた視覚化条件で一回ずつ, ランダムな順序で体験した. 実験の際には, 被験者一人に対して 3 種類の視覚化条件が同数現れるように, また被験者間で全ての街路と条件の組合

せが均等に提示されるように調整した。

3.2.3 課題および指標

参加者は各街路において、指定された出発点から目的地までアバターを移動し、その両端における水位をセンチメートル単位で報告するよう指示された。その後、仮想環境における臨場感を評価する iGroup Presence Questionnaire (IPQ) [13] を記入した。IPQ は、一般的な現実感 (G)、没入感 (INV)、現実感 (REAL)、空間的な臨場感 (SP) の 4 指標を 7 段階で評価する。最後に、各視覚化条件についての主観的な意見をインタビュー形式で収集した。

3.2.4 結果

各視覚化条件における IPQ の各指標のスコアを Fig. 6 に示す。これによると、一般的な現実感 (G) および現実感 (REAL) の両指標において、Static-texture と Dynamic-texture は Non-texture に比べ有意に高いスコアを示しており、テクスチャが臨場感や現実感といった指標に大きく貢献することを示している。各指標において、Static-texture と Dynamic-texture の間に有意差はなかったが、全体的に Dynamic-texture がよりスコアが高い傾向を示した。

また、ユーザーの水位報告の誤差の平均は 17.9 cm (標準偏差 16.7 cm) であり、条件間に有意差はなかった (Non-texture: 17.6 cm, Static-texture: 16.3 cm, Dynamic-texture: 20.0 cm)。事後アンケートによると、27 名中 24 名がアバターの高さを水深判断の基準にしていたと回答しており、水位の把握においては、視覚化条件に比べて、アバターの存在がより重要な役割を果たしていることが示唆された。一方、洪水の主観的な感覚に関しては、インタビューで Dynamic-texture 条件について「水位の高さがよりリアルに感じられた」(P17)、「洪水リスクを実感しやすかった」(P05) といった肯定的な評価が寄せられており、全体で 9 名の被験者が、Dynamic-texture により直感的な理解が促進されたという旨の供述をした。これらの結果は、360 度動画による写実的な都市空間が、地理空間データの直感的理解に有効であることを示唆している。

4. むすび

本研究では、CityGML モデルと 360 度映像を融合した都市空間可視化システムである 360CityGML を提案した。本システムにより、ユーザーは歩行者視点から、写実的な風景と整合した形で地理情報を把握することができる。これにより、CityGML データの活用領域が大きく拡張され、都市の地理空間データにおける新たな可視化手法と解釈の可能性が拓かれる。

謝辞 本研究の一部は、JSPS KAKENHI 25H01164, SIP3 「スマート防災」の支援を受けた。

参考文献

- [1] Gerhard Gröger and Lutz Plümer. Citygml-interoperable semantic 3d city models. *ISPRS P&RS*, 71:12–33, 2012.
- [2] MLIT, Japan. Plateau, 2020. URL: <https://www.mlit.go.jp/plateau/>.
- [3] Jürgen Döllner, Thomas H Kolbe, Falko Liecke, Takis Sgouros, and Karin Teichmann. The virtual 3d city model of berlin-managing, integrating, and communicating complex urban information. In *UDMS*, 2006.
- [4] Fabio Miranda, Thomas Ortner, Gustavo Moreira, Maryam Hosseini, Milena Vuckovic, Filip Biljecki, Claudio T Silva, Marcos Lage, and Nivan Ferreira. The state of the art in visual analytics for 3d urban data. In *Computer Graphics Forum*, volume 43, page e15112. Wiley Online Library, 2024.
- [5] Saeed Boorboor, Yoonsang Kim, Ping Hu, Josef M. Moses, Brian A. Colle, and Arie E. Kaufman. Submerge: Visualizing storm surge flooding simulations in immersive display ecologies. *IEEE TVCG*, 30(9):6365–6377, 2024.
- [6] 3D geoinformation research group at Delft University of Technology. Cities/regions around the world with open datasets, 2024. URL: <https://3d.bk.tudelft.nl/opendata/opencities/>.
- [7] Naoki Sugimoto, Yoshihito Ebine, and Kiyoharu Aizawa. Building movie map-a tool for exploring areas in a city-and its evaluations. In *ACMMM*, pages 3330–3338, 2020.
- [8] Mizuki Takenawa, Naoki Sugimoto, Leslie Wöhler, Satoshi Ikehata, and Kiyoharu Aizawa. 360rvw: Fusing real 360° videos and interactive virtual worlds. In *ACMMM*, pages 9379–9381, 2023.
- [9] Aparna Taneja, Luca Ballan, and Marc Pollefeys. Registration of spherical panoramic images with cadastral 3d models. In *3DIMPVT*, pages 479–486. IEEE, 2012.
- [10] Shinya Sumikura, Mikiya Shibuya, and Ken Sakurada. Openvslam: A versatile visual slam framework. In *ACMMM*, pages 2292–2295, 2019.
- [11] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *CVPR*, pages 1290–1299, 2022.
- [12] Nikolaus Hansen and Andreas Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evol. Comput.*, 9(2):159–195, 2001.
- [13] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. The experience of presence: Factor analytic insights. *Presence (Camb. Mass.)*, 10(3):266–281, 2001.