

口頭発表

2024年10月26日(土) 9:00 ~ 10:40 D会場(s-519D 防災研究所大会議室)

[D1] データ取得

座長:永田 彰平(東北大学)

9:00 ~ 9:20

[D1-01]

変化検知に基づく単眼衛星画像を用いた建物フットプリント更新手法の開発

*陳 聖隆¹、小川 芳樹¹、関本 義秀¹ (1. 東京大学空間情報科学研究センター)

9:20 ~ 9:40

[D1-02]

小型2次元レーザスキャナと画像認識技術を用いた時空間解像度の高い歩行軌跡の検出手法：大学病院の集中治療部におけるケーススタディ

*沖 拓弥¹、仙頭 佳起¹、山中 理沙¹、野坂 宜之¹、野口 綾子¹、海塩 渉¹、若林 健二¹ (1. 東京科学大学)

9:40 ~ 10:00

[D1-03]

AIカメラによる歩行者断面交通量データを用いた大規模商業施設来客者のゾーン別滞留時間分布推定

*渡部 隼人¹、大佛 俊泰¹、岸本 まき¹、金子 弘幸²、陳 紹華²、岩崎 庸浩³ (1. 東京科学大学、2. 鹿島建設 技術研究所、3. 鹿島建設 建築設計本部)

10:00 ~ 10:20

[D1-04]

オープンアクセス衛星データを利用した建物検出精度向上のための立地特性分類手法の考察

*荻野 光司¹、大佛 俊泰¹ (1. 東京科学大学)

10:20 ~ 10:40

[D1-05]

書き手はどこを訪れたか？—言語モデルで訪問行動を読み取る—

大友 寛之²、*大内 啓樹^{1,2}、東山 翔平^{4,1}、寺西 裕紀^{3,1}、山本 和太郎⁵、進藤 裕之⁶、渡辺 太郎¹ (1. 奈良先端科学技術大学院大学、2. 株式会社サイバーエージェント、3. 理化学研究所、4. 情報通信研究機構、5. 株式会社NTTドコモ、6. MatBrain株式会社)

変化検知に基づく単眼空中写真を用いた建物フットプリント更新手法の試み

陳聖隆*・小川芳樹**・関本義秀***

Study of building footprint update method based on change detection using monocular aerial images

Shenglong Chen*, Yoshiki Ogawa**, Yoshihide Sekimoto***

Abstract Regular updates to building footprint data are essential for maintaining its accuracy and relevance. Traditional methods rely heavily on bi-temporal aerial imagery to detect building changes. However, this approach can result in discrepancies due to factors such as sensor angle variations, building misalignments, and pseudo-changes from non-building elements. Additionally, existing databases often underutilize the semantic information of buildings. To address these challenges, this study introduces a novel building footprint updating method that integrates up-to-date monocular aerial image with historical building footprint data, enabling the detection of newly constructed and demolished buildings. Validation with the SI-BU and Osaka datasets demonstrate that this method holds strong potential for broad application in the regular updating of building databases.

Keywords: 建物フットプリント更新(Building footprint update), 変化検知(Change detection), 単眼空中写真(Monocular aerial image), セマンティック・セグメンテーション(Semantic segmentation)

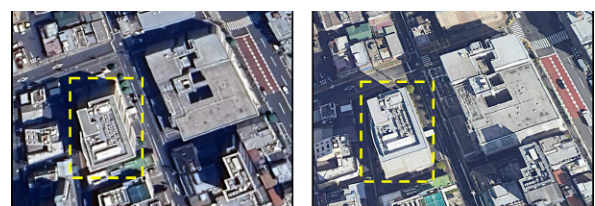
1. はじめに

建物のフットプリントは、都市を分析・管理する際の重要な基盤データである。従来、建物フットプリントの更新は地図会社や自治体が調査により行うか、あるいはボランティアによって（例：OpenStreetMap）行われていた。しかし、リモートセンシング技術の発展により、定期的に撮影される高解像度の空中写真が、効率的かつ高品質なデータソースとして注目されている。一般的な建物フットプリントの更新手法としては、空間参照された二時期の空中写真を比較し、ピクセル単位で変化領域を抽出することにより、解体及び新築された建物を識別する（笹川ほか、2021）。しかし、この方法には、いくつかの課題がある。

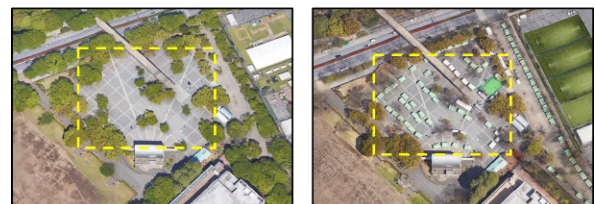
- 1) 二時期の空中写真における建物フットプリントは、撮影角度の違いにより完全には整合していない。この幾何学的なズレは、特に高い建物において予測が難しい問題がある（図 1a）。
- 2) 空中写真において、建物と背景の領域を明確に

区別するのは困難であり、撮影期間や照明条件などの違いによっては、非建物領域に擬似変化が生じる可能性がある。（図 1b）

- 3) 既存データベースに含まれる建物に関する事前情報（例：位置など）が十分に活用されていない。また、既存データベースと完全に一致する過去の空中写真を見つけのは難しい問題がある。



(a) 建物フットプリントのズレ



(b) 非建物領域における擬似変化

図 1. 二時期空中写真を使う変化検知手法の課題

* 正会員 東京大学空間情報科学研究センター (CSIS, University of Tokyo) chen-sl@csis.u-tokyo.ac.jp

** 正会員 東京大学空間情報科学研究センター (CSIS, University of Tokyo) ogawa@csis.u-tokyo.ac.jp

*** 正会員 東京大学空間情報科学研究センター (CSIS, University of Tokyo) sekimoto@csis.u-tokyo.ac.jp

過去の建物フットプリントは、当時の建物分布を記録しており、その変遷を検証するための貴重な参考データとなる。本研究では、従来の二時期の空中写真から建物変化を抽出する方法とは異なり、最新の単眼空中写真と過去の建物フットプリントデータを組み合わせて、建物データの更新手法を提案した。これにより、建物に関するセマンティック情報を最大限に活用し、擬似変化の影響を軽減することを目指している。対象地域として大阪市東住吉区を選定し、本手法の有効性を検証した。

2. 手法

2.1. アーキテクチャ

本提案手法は、最新の空中写真と状況変化を示す建物マスクを組み合わせ、解体および新築された建物を個別に抽出することが可能である。建物マスクは、それぞれ、未変化（グレー）、解体（赤）、新築（青）を注釈している（図 1）。また、学習済みエンコーダ、変形可能畳み込みネットワーク（DCN）ベースの変換モジュール、およびマルチタスクデコーダーの 3 つの部分で構成されている（図 2）。まず、学習済みの ResNet34 エンコーダを用いて、空中写真からマルチスケールの特徴が抽出・融合される。次に、これらの特徴量を過去の建物マスクに基づいて前景と背景の特徴量を分離する。この際、建物マスクと特徴量のオフセットを考慮し、DCN ベースの変

換モジュールで調整が行われる。最後に、マルチタスクデコーダーを用いて、解体と新築された建物を同時に抽出する（Liao et al., 2023）。

2.2. DCN ベースの変換モジュール

過去の建物フットプリントは、空中写真上で正確に位置合わせされていない場合があり、このズレは変化検知において大きなノイズを引き起こす可能性がある。この影響を軽減するため、DCN ベースの変換モジュールが導入され、画像とフットプリントの幾何学的関係を学習し、特徴を適応的に調整する。これにより、入力された融合特徴に基づき、DCN 層が建物（前景）と背景のオフセットを自動的に学習し、空間的位置における特徴とマスクの不整合を修正することが可能となる。

2.3. マルチタスクデコーダー

建物変化を検出するため、補助、解体及び新築建物抽出の三つの抽出タスクを設計した。まず、空間行列の乗算により、前景と背景の特徴量を算出し、それを分割ブロックに入力して建物フットプリントを抽出する。前景と背景の特徴は、それぞれ過去の建造マスクの内部と外部空間の特徴に対応している。解体建物の特徴は、現存する建物の特徴と逆の関係にあるため、空間分割する前にこれらの特徴を反転させる必要がある。さらに、既存建物のセマンティ

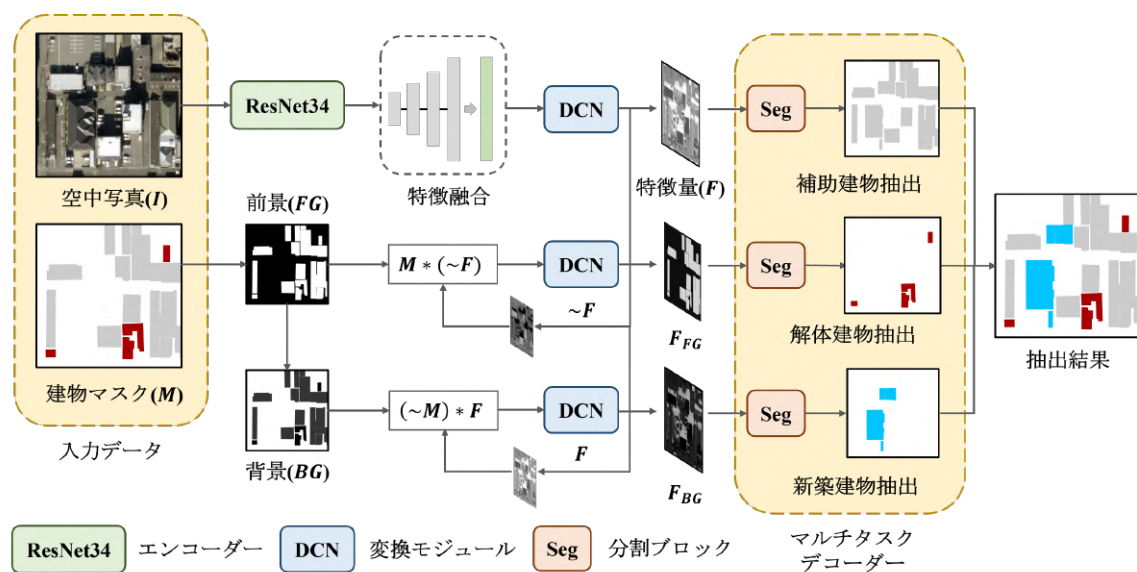


図 2. 変化検知ネットワークのアーキテクチャ (Liao et al., 2023)

ック特徴を十分に活用し、背景から建物を区別するための特徴を学習するために、空中写真から既存建物を抽出する補助抽出ヘッドを設計した。

3. データセット

3.1. SI-BU データセット

SI-BU データセットは、2019 年と 2021 年に撮影された Google Earth の空中写真(解像度 0.5~0.8m)で構成され、中国貴州省の約 172km² の範囲をカバーしている。データセットには合計 4,932 枚のパッチが含まれ、そのうち 3,604 枚が学習用、1,328 枚がテスト用に利用される。建物の変化アノテーションは、背景、未変化の建物、新築の建物、解体の建物の 4 クラスに分類され、それぞれ「0」、「1」、「2」、「3」の値が割り当てられている (Liao et al., 2023)。

3.2. 大阪データセット

大阪データセットは大阪市東住吉区(約 2.3km²)がふくまれおりテストエリアとして選定した(図 3)。国土地理院が令和 6 年 1 月および平成 23 年 1 月に撮影した解像度 0.3m と 0.6m の空中写真(512×512 ピクセル、計 66 枚)と、それに対応する建物アノテーションで構成されている。建物アノテーションは 2022 年版 Plateau3D 都市モデルの建物フットプリントに基づき、10 年間の建物変化を正確に反映するように手動で補正した。

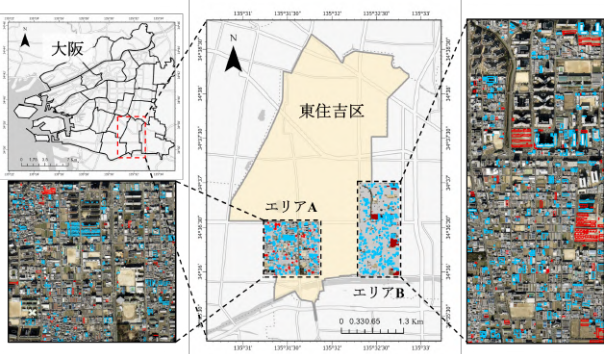


図 3. 大阪データセットの概要

4. 結果・考察

4.1. 評価指標

本研究は、建物単位での変化検知を目的としている。セマンティック・セグメンテーションのフレ

ームワークを用いて建物の変化マスクを抽出するが、結果の定量的評価には、オブジェクト単位のタインスタンス・セグメンテーションメトリクスを使用する。適合率、再現率、及び F 値は、式 1 に示すように定義される。ここで、TP、FP、FN はそれぞれ、正しく検出された、誤って検出された、検出されなかった建物変化を指す。

$$\begin{aligned} \text{適合率} &= \frac{TP}{TP + FP} \\ \text{再現率} &= \frac{TP}{TP + FN} \\ \text{F 値} &= \frac{2 \times \text{再現率} \times \text{適合率}}{\text{再現率} + \text{適合率}} \end{aligned} \tag{1}$$

4.2. 実験結果

提案手法の有効性を検証するため、二時期画像を基ずく変化検知手法である CGNet-CD と比較を行った (Han et al., 2023)。これら 2 つの手法は原理が異なるため、比較の公平性を確保するために、SI-BU データセットで統一して学習させた。実験はすべて PyTorch を用いて実装し、A100 GPU(40GB)を 4 枚搭載した mdx プラットフォーム上で実行した。

表 1 に、SI-BU データセットでの各手法の変化検知精度を示す。提案手法の F 値は 0.57 に達し、CGNet-CD を 0.04 上回っており、提案手法の優位性を示した。これは、過去の建物フットプリントがセマンティック情報を提供し、単なる画像からのピクセルレベルの変化検知よりも、変化領域をより正確に特定できるためである。図 4 には、両手法による変化検知結果を示している。提案手法は新築(青)と解体(赤)の建物を区別できるが、CGNet-CD は全体的な変化領域(黒)しか検出できない。新築(図 4a)と解体建物(図 4b)の結果を示し、提案手法は精度と形状の完全性で優れている一方、CGNet-CD は非建物の領域で擬似変化の影響を受けやすいことを示している。図 4c は、撮影角度の違いによる高層建物のズレを示し、提案手法はこの問題を効果的に回避できることを示している。

表 1. SI-BU データセットでの精度

手法	適合率	再現率	F 値
CGNet-CD	0.49	0.58	0.53
提案手法	0.54	0.60	0.57

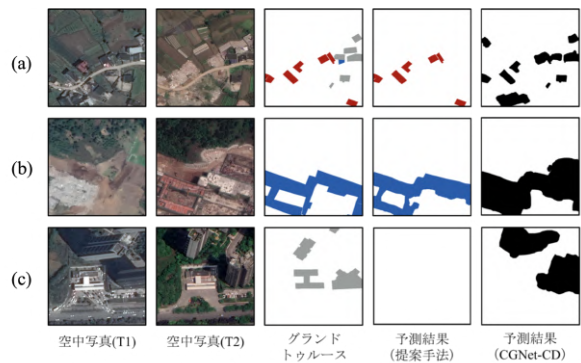


図 4. SI-BU データセットの検出結果

4.3. 考察

提案手法の汎化能力を評価するため、SI-BU データセットで学習したモデルを大阪データセットでテストした。その結果、変化検知の精度が著しく低下している（表 2）。特に新築建物について、再現率の面で大幅に低下している。一方、解体建物の精度は、過去の建物フットプリントが参照できるため比較的良好である。図 5 は、大阪データセットでの提案手法の予測結果を示している。提案手法はセマンティック・セグメンテーションに基づいており、密集した市街地では隣接する建物を正しく分割するのが難しい。また、新築建物の近くにある解体建物の検出も困難である。

表 2. Osaka データセットでの精度

クラス	適合率	再現率	F 値
新築	0.44	0.17	0.25
解体	0.69	0.41	0.51
全体	0.48	0.21	0.29

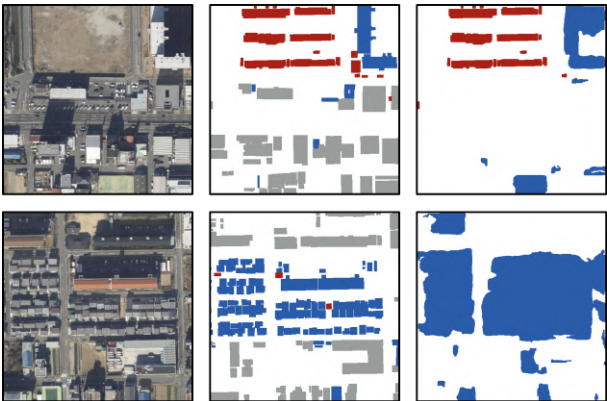


図 4. 大阪データセットの検出結果

5. 結論

本研究では、マルチタスク・フレームワークに基づく新しい建物フットプリント更新手法を提案した。従来の二時期の画像を用いた変化検知手法とは異なり、単眼空中写真と過去の建物フットプリントデータを組み合わせることで、解体と新築された建物を同時に抽出する手法を提案した。この手法により、センサーの角度や照度の変化など、二時期の画像間で生じる擬似変化の影響を効果的に回避できる。また、DCN ベースの変換モジュールは、空中写真と過去の建物フットプリントとの間にある幾何学的不整合を緩和する役割を果たす。同時に、既存の建物データのセマンティック情報を活用し、建築変化領域の特定を支援する。SI-BU および大阪のデータセットを用いて提案手法の性能を評価し、その有効性を検証した結果、提案手法は既存手法より精度が高く、建物データベースの定期的な更新において幅広い応用の可能性が示唆された。しかし、個々の建物、特に新築建物の検出性能には依然として改善の余地があり、さらに、広範囲の領域に対する手法の頑健性の向上も今後の課題である。

謝辞

本研究成果（の一部）は、データ活用社会創成プラットフォーム mdx を利用して得られたものである。

参考文献

笹川啓・田代ゆかり・石塚麻奈・柴田光博（2021）二時期の空中写真と衛星画像による自動変化抽出手法の開発.「国土地理院時報」, 134, 33-42.

Liao, C., Hu, H., Yuan, X., Li, H., Liu, C., Liu, C., ... & Zhu, Q. (2023). BCE-Net: Reliable building footprints change extraction based on historical map and up-to-date images using contrastive learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 201, 138-152.

Han, C., Wu, C., Guo, H., Hu, M., Li, J. & Chen, H. (2023). Change guiding network: Incorporating change prior to guide change detection in remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 99, 1-15.

小型 2 次元レーザスキャナと画像認識技術を用いた時空間解像度の高い歩行軌跡の検出手法：大学病院の集中治療部におけるケーススタディ

沖 拓弥 *1・仙頭 佳起 *2・山中 理沙 *3・野坂 宜之 *2・野口 綾子 *2・海塩 渉 *4・若林 健二 *2

A Method for Detecting Walking Trajectories with High Spatiotemporal Resolution Using Compact 2D Laser Scanners and Image Recognition Technologies: A Case Study in the Intensive Care Unit of a University Hospital

Takuya OKI, Yoshiki SENTO, Risa YAMANAKA, Nobuyuki NOSAKA, Ayako NOGUCHI, Wataru UMISHIO, and Kenji WAKABAYASHI

Abstract: This study proposed a method to detect and track a person using image recognition technologies from 2D point cloud data acquired by compact laser scanners and to mechanically create walking trajectory data with high spatiotemporal resolution. The 2D laser scanner has a small measurement error and excellent privacy protection. A total of 18 scanner units were installed in the intensive care unit (ICU) of the T University Hospital for a three-week trial. Then, the proposed method was applied to the measurement data in a corridor and compared with the number of passersby counted visually. As a result, we demonstrated that walking trajectories can be extracted from 2D point cloud data mechanically, efficiently, and accurately.

Keywords: 動線分析 (traffic flow analysis), レーザスキャナ (laser scanner), 画像認識 (image recognition), 大学病院 (university hospital), 集中治療部 (intensive care unit)

1. はじめに

1.1. 研究の背景

病院の集中治療部 (ICU) は、大手術後や重篤な状況など濃密な医療とケアを要する患者を収容する場所である。多領域にわたる専門性を有する様々な医療スタッフ (医師・歯科医師・看護師・薬剤師・管理栄養士・理学療法士・臨床工学技士など) や機器類が頻度高く出入りするほか、患者のベッドやストレッチャーの往来も多く、人やモノの動線は病院内でも特に複雑である。その結果、至る所で動線の交錯や滞留が発生し、医療スタッフの業務効率性に影響を及ぼしているとされるが、改善につながる定量的なエビデンスは乏しいのが現状である。

また昨今では、人工知能 (AI) をベースとした画像認識技術が普及し、動画から自動抽出した移動軌跡データを用いた研究も盛んに行われている。

しかし病院では、患者のプライバシー保護等の理由により、研究目的で動画を撮影・活用することは容易でない。

1.2. 研究の目的

本研究では、T 大学病院の ICU をケーススタディとして、まず、周囲 360 度をリアルタイムに計測可能な小型 2 次元レーザスキャナ (以下、スキャナ) を用いて、ICU 内における人やモノを 2 次元点群データとして計測する手法を構築する。そして、点群データを集計することで、人やモノの滞留状況の時間推移を可視化する。さらに、筆者らが提案した手法 (沖ほか, 2024) を大きく改善し、AI ベースの画像認識技術を用いて人を検出・追跡することで、時空間解像度の高い移動軌跡データを機械的に作成する手法を構築する。

2. 2 次元点群データの計測

-
- *1 正会員 東京科学大学 環境・社会理工学院 (Institute of Science Tokyo)
〒152-8550 東京都目黒区大岡山 2-12-1 E-mail: oki.t.e60f@m.isct.ac.jp
- *2 非会員 東京科学大学病院 (Institute of Science Tokyo Hospital)
- *3 学生会員 東京科学大学 環境・社会理工学院 (Institute of Science Tokyo)
- *4 非会員 東京科学大学 環境・社会理工学院 (Institute of Science Tokyo)

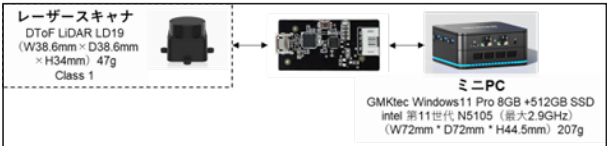


図 1 計測ユニットの構成

2.1. 計測ユニット

計測ユニットの構成を図 1 に示す。使用した 2 スキャナ (LD19, youyeetoo 社製) は、ロボットの自己位置推定やマッピングなどに用いられるもので、センサから発し対象物で反射した光が、センサに戻るまでの時間をもとに、対象物までの距離を計測する dToF (direct Time of Flight) 方式を採用している。計測誤差が極めて小さいだけでなく、プライバシー保護の観点からも優位性を有し、レーザの安全性にも問題はない。周囲 360 度を計測可能で、仕様上の最大計測距離は 12m である。動画では夜間の消灯時における人検出は困難だが、スキャナは時間帯によらず計測できる点も長所である。計測ユニットはスキャナ (幅・奥行・高さがいずれも 4 cm 以下) とミニ PC からなり、小型化に努めることで業務や動線に及ぼす影響を抑制している。

2.2. 設置方法

T 大学病院の ICU に計 18 台のユニットを設置した (図 2)。設置箇所の特性に応じて、①カウンターやパーテーション、テーブル等の上面、②作業デスクの裏面、③壁面に取り付けた L 字状のプレート上面、のいずれかの方法でスキャナを設置し、耐震マットで確実に固定した (図 3)。通行者の検知しやすさや障害物による影響を勘案し、設置高さは床面から 100 cm 前後とした。ミニ PC はスキャナと電源コンセントの付近、かつ、ICU 業務の妨げとならない場所に設置した。

2.3. 計測データ

2023 年 12 月 19 日から 2024 年 1 月 10 日までの 23 日間計測を行った。約 0.1 秒単位で取得される 560 点前後の点群情報を、10 分ごとに CSV ファイルとして出力した。点群データには、ファイル作成日時からの経過時間 [s] (1 ms 単位)、方位角 [rad] (1/1000 rad 単位)、スキャナから観測点

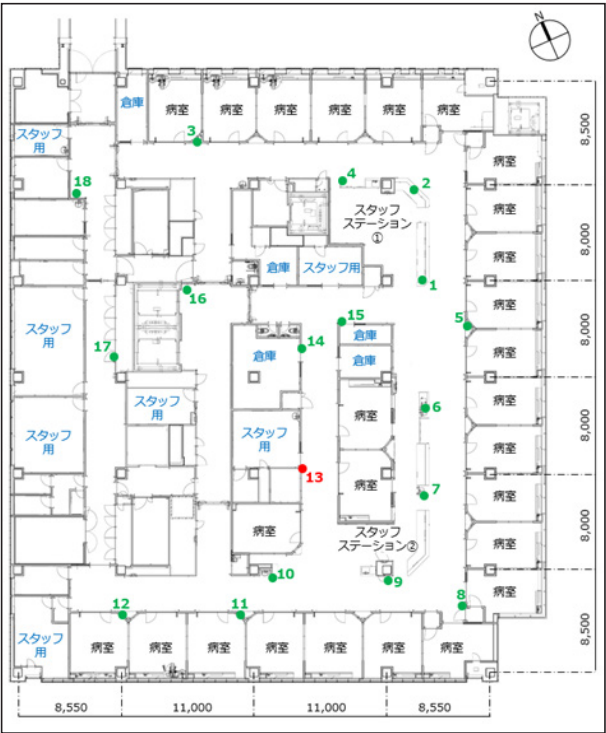


図 2 T 病院 ICU の平面図と計測ユニット (●で示す No. 1 ~ No. 18) の設置位置



図 3 計測ユニットの設置例 (左: カウンター上 [No. 1], 右上: 台の上 [No. 16], 左下: 壁面 [No. 17], 右下: 作業デスク裏面 [No. 3])

表 1 計測データの例

	elapsed	angle	distance	intensity
540	0.129	3.101	0.020	245
541	0.129	3.115	0.012	243
542	0.129	3.129	0.016	243
543	0.129	3.143	0.018	243
544	0.129	3.158	0.019	244
...	...	...	...	...
2556499	599.785	0.571	6.171	211
2556500	599.785	0.585	3.939	220
2556501	599.785	0.599	3.645	220
2556502	599.785	0.612	3.660	227
2556503	599.785	0.626	3.722	221

までの距離 [m] (1 mm 単位), 観測強度の情報が含まれる (表 1)。

1 ファイル (約 10 分間) あたりのファイルサイズは約 56MB で, スキャナ 1 台による 1 日分の出力 CSV ファイルは約 8GB に上った。すなわち, 18 台で 1 日計 144GB, 計測期間 (23 日間) 全体でのファイルサイズは 1 台で計約 184GB, 18 台で約 3.3TB という, 膨大な計測データが得られたことになる。

3. 点群データを用いた滞留・移動の可視化

3.1. 滞留の可視化方法

前章の計測により得られた点群データをもとに, 特定の時間単位で点群密度を集計した上で, 位置合わせや重なり調整を行い, ICU の全体平面図上に描画することにより, いつ, どこで, どの程度の滞留が生じているのかを直感的に把握することができる。

3.2. 滞留の可視化結果

2023 年 12 月 19 日の 1 時間分 (午前 11 時から正午まで) を例に, ICU 全体の点群密度を可視化した結果を図 4 に示す。図面右上などに点群密度が高い箇所が集中している様子が見てとれる。ただし, 人とモノの区別や, 滞留と移動の区別を, この図で直接行うことは難しい。そのため, 後述の移動軌跡を用いた分析が必要となる。

3.3. 移動軌跡の抽出方法

以下の手順で, 約 0.1 秒間隔の点群データから移動軌跡を抽出する。

- ① 点群データから, 特定の時間単位 (例えば 10 分単位) の距離の統計情報に基づき, 「壁」と推定される点を除外する (図 5)。ここでは実際の壁だけではなく, その時間帯に長時間同じ場所にとどまり続けた人・モノも検出されることに留意する必要がある。
- ② 壁以外の点を約 0.1 秒間隔で画像として出力する (図 6 左)。この時間間隔を長くすると, 処理時間は短くなるが, 後述の人物追跡の精度に悪影響を及ぼすため, 時空間解像度の高い分析のためには 0.1 秒程度が適当である。

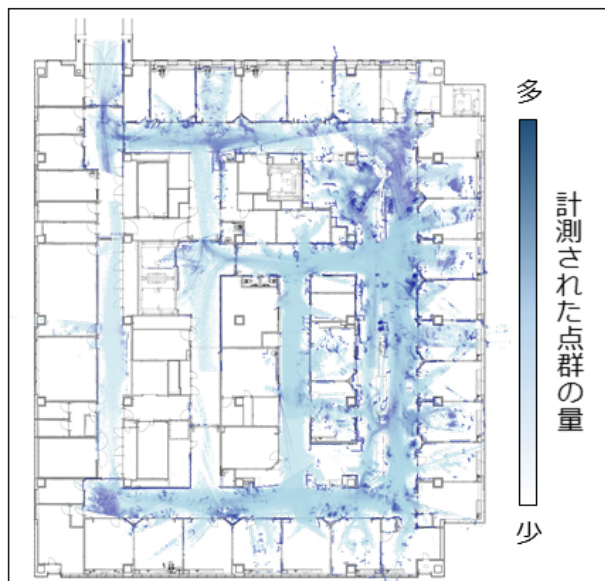


図 4 ICU 全体の点群の集計・可視化例
(2023/12/19 午前 11 時～正午の 1 時間)

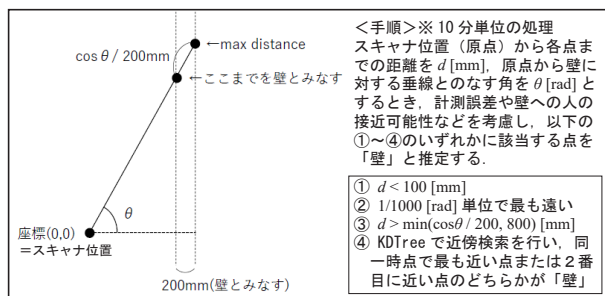


図 5 点群からの「壁」の推定方法

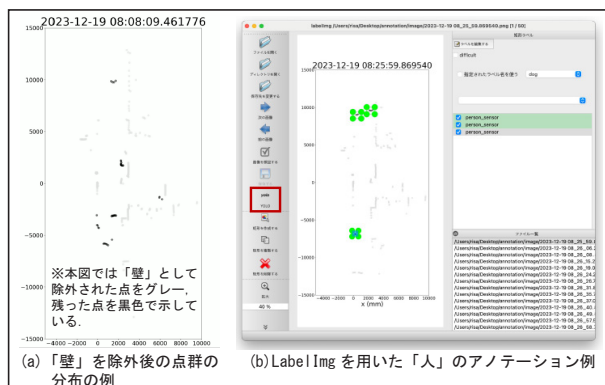


図 6 「壁」を除外後の点群の分布とアノテーションの例

- ③ ②での出力画像から, 人の数や位置の多様性を考慮しながら, 目視でアノテーション用画像をサンプリングする (本稿では 50 枚)。
- ④ 抽出した画像全てに対して, 人のアノテーションデータ (人と思われる位置を矩形で指定したデータ) を手作業で作成する (図 6 右)。本稿では LabelImg (Tzutalin, 2015) を使用してアノテーション作業を行った。クラスラベル数は 1 (「人」のみ) とした。

- ⑤ 作成したアノテーションデータを `train` : `validation = 7:3` に分割し、既存の物体検出モデル YOLO v8 (Jocher et al., 2023) をファインチューニングする。
- ⑥ ファインチューニングしたモデルを用いて、人の検出を行うとともに、測定データからシステムの状態を推定するアルゴリズムの一つであるカルマンフィルタにより人の追跡（トラッキング）を行う。

3. 4. 移動軌跡の抽出結果

2023 年 12 月 19 日の午前 8 時すぎを例に、特定の通路（図 2 のユニット No.13 を設置した通路）における移動軌跡の可視化例を図 7 に示す。横並びでの歩行やすれ違い、回避行動などが抽出できている。

移動軌跡の抽出精度を検証するため、図 7 の縦軸 +5,500[mm] のラインと -6,500[mm] のラインで、図の上下方向それぞれに移動する人数を目視で 30 分間（10 分間×3）カウントし、提案手法で抽出した移動軌跡をもとにカウントした値と比較した。その結果、平均絶対パーセント誤差（MAPE：「誤差の絶対値 / 実際の値」の平均値）は約 14% と良好であった。

軌跡の精度を直接検証することは現時点では難しいが、推定された壁の位置に基づく通路幅の誤差は数 cm 程度であることから、高精度な軌跡抽出が実現できているものと推察される。

4. まとめ

小型 2 次元レーザスキャナとミニ PC からなる計測ユニット 18 台で取得した、T 病院 ICU における点群データに基づき、滞留・移動軌跡を高い時空間解像度で可視化できることを示した。

今後、より多くのスキャナ・時点のデータを用いて、具体的な動線・行動分析を行う予定である。提案手法を用いて医療スタッフの行動分析を行うことで、ICU 業務に関する新たな知見が得られ、それをもとにスタッフの移動や人的・物的資源の配分を最適化することで、ひいては業務効率や患者ケアの質の向上につながることが期待される。

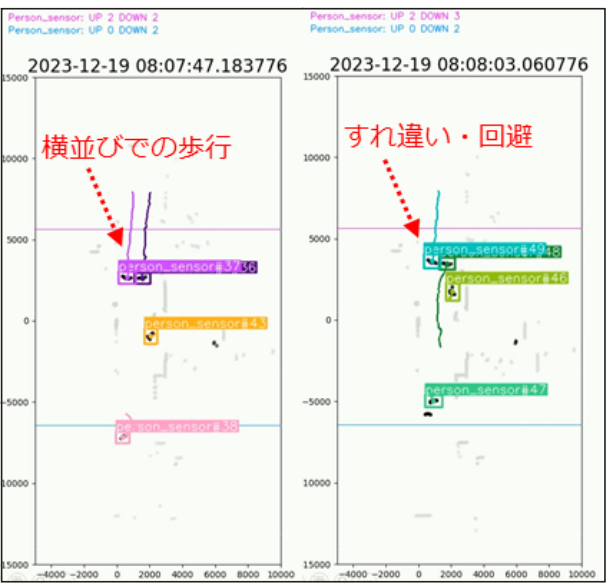


図 7 通路部分（図 2 のユニット No. 13 前）における人検出と移動軌跡可視化の例（2023/12/19 午前 8 時すぎ）

表 2 通行人数カウント結果の比較（目視 / 提案手法）

時間	方法	5500mm ↑	5500mm ↓	-6500mm ↑	-6500mm ↓
07:58:08~	目視（2人）	3	10	15 or 19	16 or 22
	目視平均	3	10	17	19
	提案手法	4	8	21	19
	絶対%誤差	33%	20%	24%	0%
08:05:08~	目視（2人）	5 or 7	13	16 or 17	18
	目視平均	6	13	16.5	18
	提案手法	6	11	17	21
	絶対%誤差	0%	15%	3%	17%
08:15:08~	目視（1人）	9	1	9	8
	提案手法	8	1	9	4
	絶対%誤差	11%	0%	0%	50%

謝辞

本研究の遂行にあたり、東京工業大学・東京医科歯科大学研究マッチングファンドの支援を頂きました。また、ICU での計測に快くご協力頂きました T 大学病院のスタッフの皆さまに、この場を借りて謝意を表します。計測準備や移動軌跡の可視化方法の検討、精度検証にあたっては、東京工業大学沖研究室の学生各位（増喜浩太郎・大塚千聡・グエンホンタイ・山本さつき・LIU Yuan・小泉兵悟・小松なのか・藤田晋太郎・山田拓弥・橋本蒼真・三田村和喜）に協力いただいたことに感謝いたします。

参考文献
沖拓弥, 仙頭佳起, 野坂宜之, 野口綾子, 海塩渉, 若林健二 (2024) 大学病院の集中治療部におけるレーザスキャナを用いた行動分析の試み, 2024 年度人工知能学会全国大会 (第 38 回) .
Tzatalin (2015) Labellmg. <https://github.com/tzatalin/labellmg>
Jocher, G., Chaurasia, A., & Qiu, J. (2023) Ultralytics YOLO (Version 8.0.0) [Computer software]. <https://github.com/ultralytics/ultralytics>

AI カメラによる歩行者断面交通量データを用いた大規模商業施設来客者の
ゾーン別滞留時間分布推定

渡部隼人*・大佛俊泰**・金子弘幸***・福井三穂****

Estimation of Zone-Specific Dwell Time Distribution of Visitors in Large-Scale Commercial
Facilities Using AI Camera-Based Pedestrian Cross-Section Traffic Volume Data

Hayato WATANABE , Toshihiro OSARAGI , Hiroyuki KANEKO and Miho FUKUI

Abstract: Understanding the relationship between the spatial characteristics of commercial facilities and visitor behavior is essential for designing spaces that enhance comfort and encourage purchasing. In this paper, we first collect data on pedestrian traffic in each direction at five-minute intervals using AI cameras placed throughout the facilities, and develop preprocessing methods to ensure data accuracy. We then segment the facilities into zones based on camera locations and construct a statistical model to estimate dwell time distributions for each zone and arrival time. Finally, we apply this model to two large commercial facilities and examine the characteristics of dwell time distributions by arrival times and zones.

Keywords: 大規模商業施設 (large commercial facility), 滞留時間分布 (dwell time distribution), AI カメラ (AI camera), 常時計測 (continuous masurement), 歩行者断面交通量データ (pedestrian cross-section traffic volume data)

1. はじめに

大規模ショッピングモールなどの商業施設における空間特性と来館者の行動特性の関係を把握することは、快適な滞留・購買空間を設計するための基礎情報として重要である。そのため、来館者の行動特性を知るうえで重要な滞留時間分布を把握する調査研究は古くから多く試みられてきた。

本研究の最終的な目標は、滞留時間が長くなる要素（魅力度）は何かを分析し、大規模商業施設の計画案に供する情報を獲得することにある。そのための基礎分析として本稿では、2つの大規模商業施設をそれぞれ複数のゾーンに分割し、ゾーンごとに来館時刻別の滞留時間分布を推定し、その特性について考察する。具体的には、まず、施設内に密に配置されたAIカメラにより5分単位

の方向別通過者数を取得し、観測誤差を調整するための前処理を行う。次に、AIカメラの観測地点に基づき施設を複数ゾーンに分割し、来館時刻別・ゾーン別の滞留時間分布を推定する統計モデルを構築する。最後に、施設ごとに推定した来館時刻別・ゾーン別の滞留時間分布をもとに、それぞれの滞留時間特性について考察する。

2. 歩行者断面交通量データ

2.1 歩行者断面交通量データの概要

歩行者断面交通量データの概要を図1(a)に示してある。施設Aでは計67台、施設Bでは計37台のAIカメラが設置され、各観測地点における方向別通過者数が5分単位で常時計測されている。

* 学生会員 東京科学大学環境・社会理工学院 (Institute of Science Tokyo)
〒152-8550 東京都目黒区大岡山 2-12-1 E-mail: watanabe.h.18ee@m.isct.ac.jp
** 正会員 東京科学大学環境・社会理工学院 (Institute of Science Tokyo)
*** 正会員 鹿島建設 技術研究所 (Kajima Technical Research Institute, Kajima Corporation)
**** 非会員 鹿島建設 技術研究所 (Kajima Technical Research Institute, Kajima Corporation)

2.2 歩行者断面交通量データのの前処理

AI カメラの観測地点に基づき複数のゾーンを設定し（図 1(b)(c)），商業施設内におけるゾーンごとの滞留特性を分析する．各ゾーンは複数の観測地点により他のゾーンと空間的に分離されている．なお，両施設ともにある Left ゾーン，Center ゾーン，Right ゾーンをメインゾーンとし，その他のゾーンを小ゾーンとする．以下では，ゾーン別の滞留時間特性について検討する際，来館者を着目ゾーンへの流入者・流出者と表現する．ゾーン i から j への累計通過者数を X_{ij} とすると，ゾーン k への累計流入者数 $\sum_{i \rightarrow k} X_{ik}$ とゾーン k からの累計流出者数 $\sum_{j \rightarrow k} X_{kj}$ は一致するべきであるが，計測誤差 e_{ij} の影響により一致しない．そこで，図 2(c) に示す制約条件のもと，線形計画法を用いて誤差率（計測誤差 e_{ij} が累計通過者数 X_{ij} に占める割合）の合計 R が最小となる計測誤差 e_{ij} を推定し，計測誤差 e_{ij} を各観測地点の単位時間ごとの観測値の大きさに応じて案分することで，累計通過者数の真値 X^*_{ij} を推定する（図 2(a)(b)）．

2.3 歩行者断面交通量データから得られる基礎情報

各施設における来館者数，退館者数，および，滞留者数を図 3 に示してある．来館者数はいずれの施設においても，多くの店舗が営業を開始する 10 時に最も多くなる（図 3(a)）．退館者数は施設 A は夕食時（18 時頃）と閉店直前（21 時頃）にピークを示す一方で，施設 B は時刻に依存する大きなピークは見られない．（図 3(a)）．また，滞留者数は 14 時から 15 時にピークを示すことがわかる（図 3(a)）．

次に，各観測地点における累計通過者数をみると，累計通過者数は 1 階から 3 階にかけて少なくなっていることがわかる（図 4）．さらに，施設全体で右回りの特徴がみられ，牧野ほか（2021）の研究と同様のことが示されている．この行動特性は来館者が商品や店舗を視認しやすい購買空間の設計に応用できる可能性がある．

3. 滞留時間分布の推定

3.1 滞留時間分布の推定手法

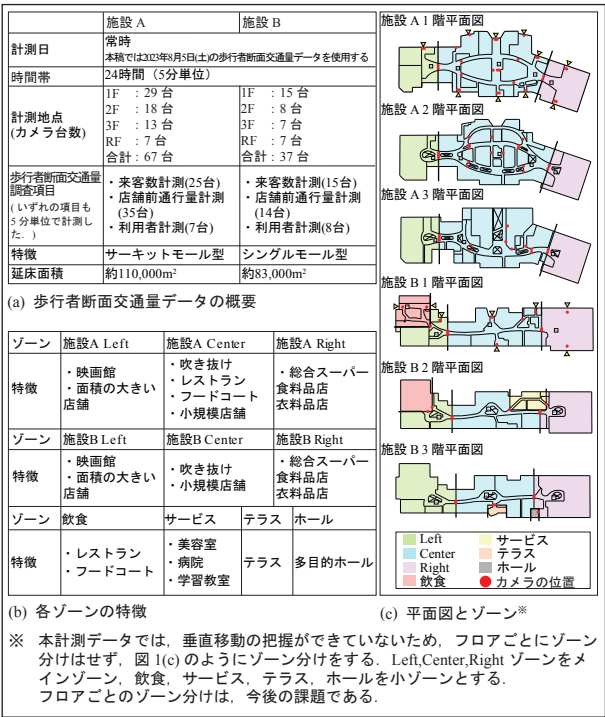


図 1 歩行者断面交通量データ概要

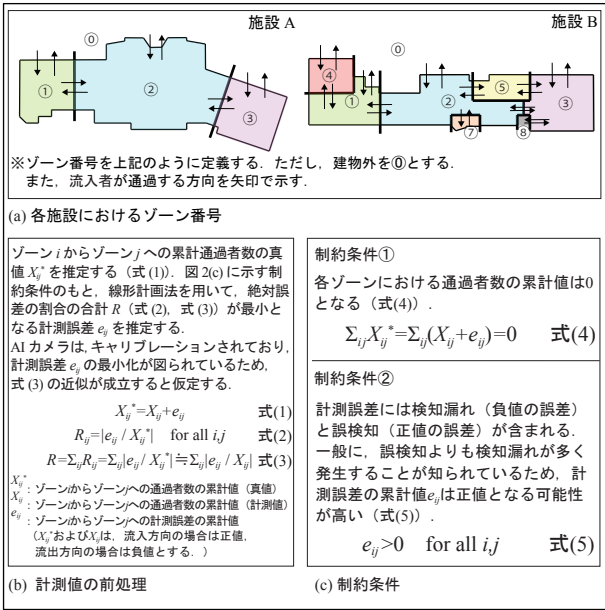


図 2 計測誤差の補正方法

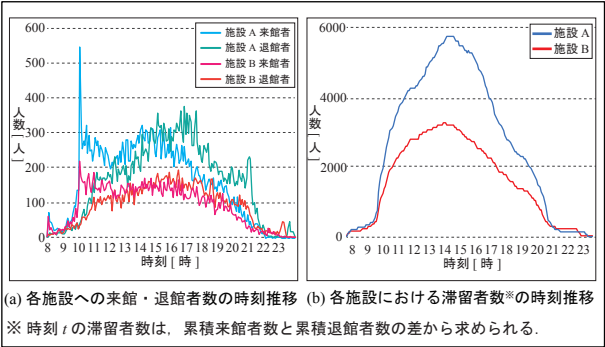


図 3 歩行者断面交通量データから得られる基礎情報

歩行者断面交通量データから、任意の時刻におけるゾーン別滞留者数（どこに何人いるか）は得られるものの、各ゾーンにおける滞留者の滞留時間分布（何分程度そのゾーンに滞留するか）は直接得ることはできない。

本稿では、岡田（1975）の著書を参考に、アーラン分布の密度関数を用いて流入時刻ごとのゾーン別滞留時間分布を推定する（図 5(a)）。ここでは、時刻 t における流入者数 A_{in}^t のみ既知であると考ええる。まず、アーラン分布の密度関数 $b(t,s)$ に基づいて（図 5(a)、式 (6)）、流入時刻 s の流入者が x 時間以上滞留する確率を $B(x,s)$ と記述する（図 5(a)、式 (7)）。これをもとに、流入時刻 s の流入者が時刻 t にも滞留している人数を記述し、流入時刻 s が 0 から t まで足し合わせることで、時刻 t における滞留者数 $\hat{A}_{stay}^t$ を推定する（図 5(a)、式 (8)）。

観測データを用いて、時刻 t における滞留者数の計測値 A_{stay}^t と推定値 $\hat{A}_{stay}^t$ の絶対誤差 F^t が最小となるパラメータ $\hat{\theta}$ を勾配法を用いて推定する（図 5(b)）。

3.2 推定結果

時刻別の滞留者数 $\hat{A}_{stay}^t$ 、平均滞留時間 $\hat{\mu}^t$ および、滞留時間分布の推定結果を図 6 に示してある。滞留者数の推定誤差は最大でも 1.3% であり、良好な精度で推定できている（図 6(a) ①）。

まず、施設別の推定結果をみると、両施設とも昼食時（12 時頃）や夕食時（18 時頃）の平均滞留時間は長い傾向が見られる（図 6(b)(c) ①）。また、施設 A の平均滞留時間は施設 B よりも長い。施設 A の方が規模が大きく、サーキットモール型が採用されているため施設内の回遊性が高く、また、施設中央部に休憩スペースとなる大きな吹き抜け空間があるなどの空間特性が影響している可能性がある。

次に、ゾーン別の推定結果をみると、平均滞留時間は流入時刻によってばらつきはあるものの、メインゾーンでは 30 分から 1 時間程度（図 6(b)）、小ゾーンでは 10 分から 40 分程度であることがわかる（図 7(b)）。また、Center ゾーンの平均滞留時間は、施設 A では他のゾーンよりも長く、施設 B では短い傾向がある（図 6(b)）。施

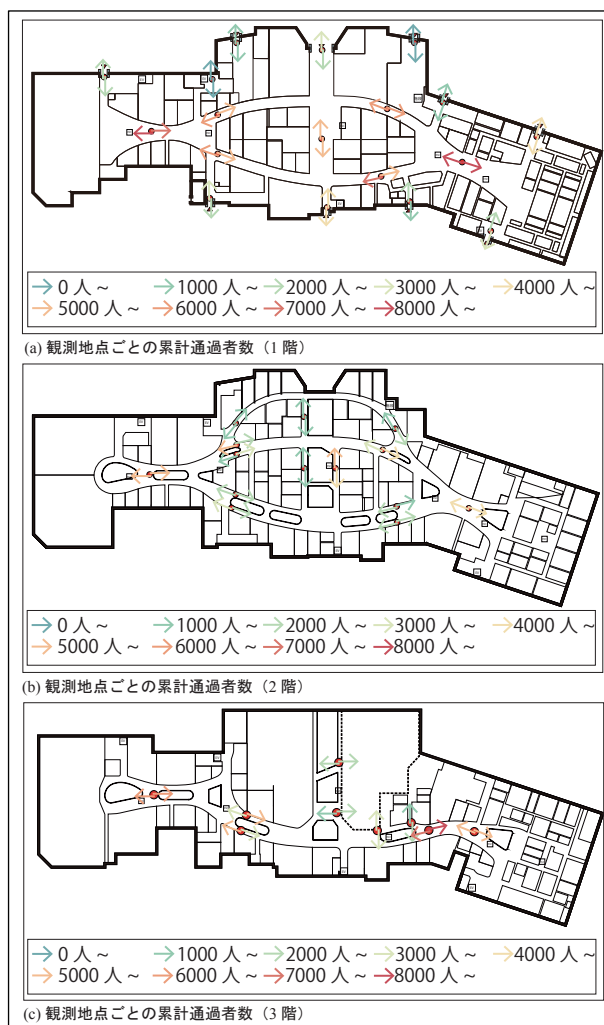


図 4 施設 A の各観測地点における累計通過者数

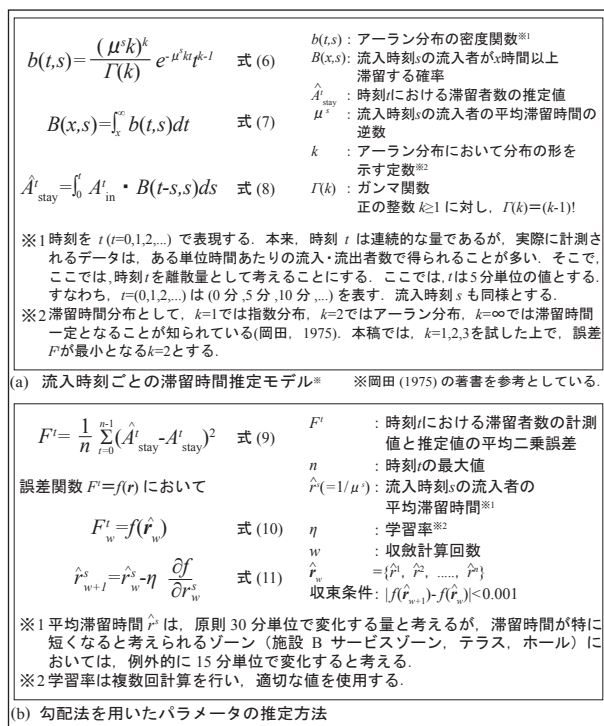


図 5 滞留時間分布の推定方法

設 A の Center ゾーンは相対的に規模が大きく、また、吹き抜け空間のある休憩スペースもあることから平均滞留時間が長くなっていると考えられる。一方、施設 B の Center ゾーンには複数のゾーンが接続しており、Left ゾーンと Right ゾーンの間を移動する来館者が占める割合が高いため、滞留時間は短くなっていると考えられる。また、施設 A の Center ゾーンや施設 B の飲食ゾーンでは、昼食時（12 時頃）に平均滞留時間が特に長くなる傾向が確認できる（図 6(c) ②，図 7(b)(c) ①）。イベント利用が主であるホールでは、流入時刻によって滞留時間が大きく異なり、イベントに合わせた滞留行動の傾向が見られる（図 7(c) ②）。

4. おわりに

大規模商業施設内において常時計測されている流入者数・流出者数に関するデータから来館時刻別・ゾーン別に滞留時間分布を推定する手法を構築し、施設によって異なる来館時刻別・ゾーン別の滞留時間の特性を明らかにした。各ゾーンの特徴を踏まえた、建築空間と滞留時間の関係の分析は今後の課題である。

謝辞

イオンモール株式会社の関係者の皆様にご理解とご協力を賜りました。ここに記して謝意を表します。また、本論文の執筆において、東京科学大学岸本まき助教には多大なご助言と添削をいただき、心より感謝申し上げます。

本研究は、東京科学大学研究倫理審査会の承認を受けて実施したものである（承認番号:第 22484 号）。

参考文献

渡部隼人・大佛俊泰・岸本まき・金子弘幸・陳紹華・岩崎庸浩（2024）常時計測データを用いた大規模商業施設における来館者の滞留時間分布の推定。「日本建築学会大会学術講演梗概集」，6，451-452。
岡田光正（1975）『建築計画学 12 施設規模』，丸善株式会社，121-139。

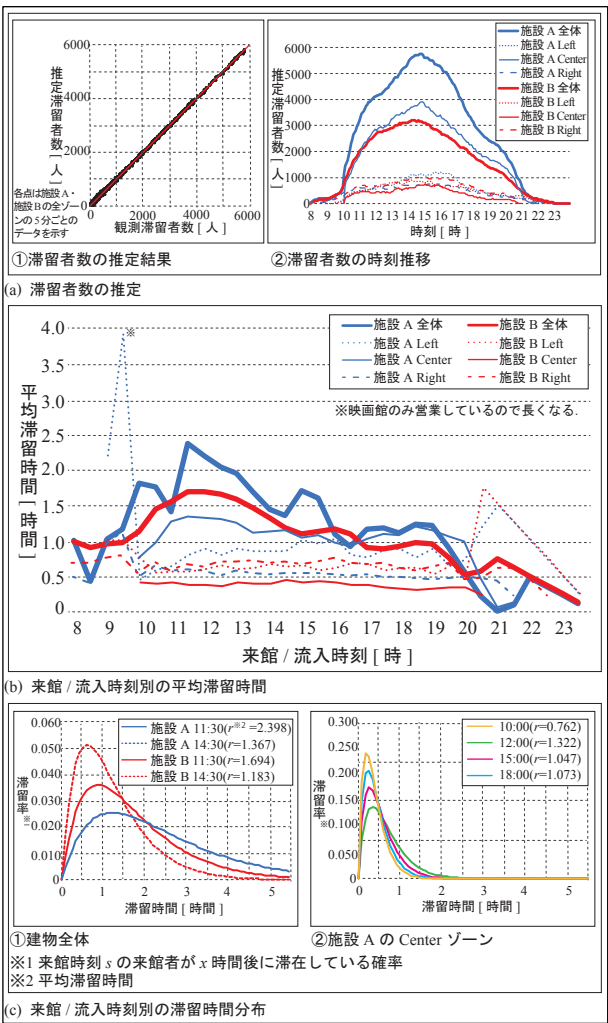


図 6 滞留時間分布の推定結果

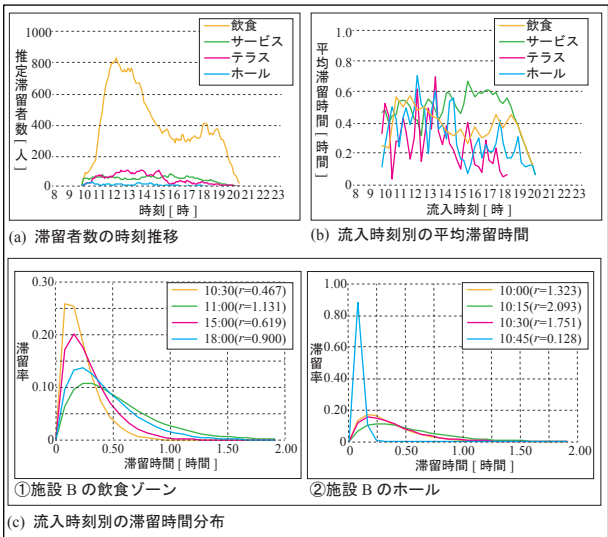


図 7 施設 B の小ゾーンの滞留時間分布の推定結果

藤井健・大口敬（2021）

大規模イベント時における歩行者の通行方法と安全な空間づくり「土木学会論文集 D3」77(3), 174-183

オープンアクセス衛星データを利用した 建物検出精度向上のための立地特性分類手法の考察

荻野 光司*・大佛 俊泰**

Examination of the Building Location Characteristics for Improving the Building Detection Accuracy Using Open-Access Satellite Data

Koji Ogino*, Toshihiro Osaragi**

Abstract Detecting buildings from remote-sensing data usually requires high-resolution satellite images, which are expensive and limited in access. Open-access satellite data providing global coverage every few days are available; however, their low resolution makes accurate building detection difficult. In this paper, we propose to construct building-detection models tailored to the specific characteristics of building locations, such as the area and density of buildings. We believe that this approach can improve detection accuracy compared to a single model that does not consider these characteristics. We classify the area using two methods: one based on population, NDVI, nighttime light data and distance to the coastline, and the other using a Vision Transformer. Next, we construct building-detection models for each building location pattern detected by the two classification models. Finally, we discuss effective approaches for classifying building location characteristics to improve the building-detection accuracy.

Keywords: 建物検出 (building detection), 建物立地特性 (building location characteristics), ロジスティック回帰 (Logistic Regression), Vision Transformer, Sentinel-2

1. はじめに

衛星画像から高精度に建物を検出するためには、高解像度の衛星データが必要であるが、高価なため一般には入手が難しい。代替データとして、全世界を数日毎に観測しているオープンアクセスの衛星データを用いた建物検出に期待が高まっているが、解像度が低いため、精度よく建物を検出することは難しく、超解像等による解像度の向上を行う必要がある (図 1(a))。

本稿では、建物の大きさや密集の程度(立地特性)毎に建物検出モデルを構築することで、精度向上が図れるのではないかと考えている。そこで、まず立地特性(棟数の多少、大規模建築物の有無)を分類する手法として、人口や植生指数、夜間光量、海岸線までの距離による分類手法と深層学習の画像分類手法を検討する (図 1(b))。次に、分類したエリア毎に建物検出モデルを構築し (図 1(c))、オープンアクセスの衛星データから建物を検出する際の効果的な立地特性の分類手法について考察する。

2. 超解像モデル

2.1 使用する衛星データと超解像モデルの概要

欧州連合とヨーロッパ宇宙機関の地球観測プログラムによる地球観測衛星である Sentinel-2 を使用した。Sentinel-2 の回帰は 10 日であるが、2 機体制であるため、5 日に 1 回程度の頻度で同一地点を可視光、近赤外を含む 12 のバンド (最大 10m 解像度) で観測している。10m の解像度で建物検出を行うことは難しいため、超解像モデルによる高解像度化を行った。本稿では、建物検出で高い精度が報告され

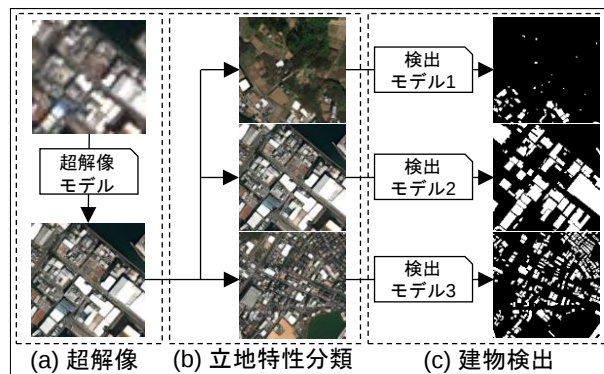


図1 建物検出のフレームワーク

* 学生会員 東京科学大学 環境・社会理工学院 (Institute of Science Tokyo)
〒152-8550 東京都目黒区大岡山 2-12-1 M4-1 ogino.k.275c@m.isct.ac.jp
** 正会員 東京科学大学 環境・社会理工学院 (Institute of Science Tokyo)

ている Real-ESRGAN を採用した (Chen et al., 2023) .

2.2 超解像ファインチューニングの結果

Mapbox から約 1.8m 解像度の都市部・農村部の衛星画像を取得し, Lanczos 法により, 画素数を 4 分の 1 にダウンサンプリングすることにより, 高解像度と低解像度のデータセットを 2,548 個作成した (図 2(1)). このデータセットを用いて, 公開されている Real-ESRGAN の学習済みモデル (Wang et al., 2021) のファインチューニングを行った.

東京都の衛星画像を同じく Mapbox から取得し, 125 個のテストデータによる精度検証を行った. 公開モデルとチューニングモデルによる超解像画像と Mapbox から取得した高解像画像を用いて, 画像類似度を測る MSSIM を計算した結果, チューニングモデルによる画像の方が, 取得した高解像画像に類似していることが確認できる (図 2(2)). 次に, Sentinel-2 の 10m 解像度の衛星画像を 2.5m 解像度に超解像した. 公開モデルによる超解像後の画像 (図 2(3)(b)) では, 小さな建物を視認することはできないが, チューニングモデルによる画像 (図 2(3)(c)) では, 小さな建物まで確認することができる.

3. 建物立地特性の分類モデル

Chen ほか (2023) は, 1 つの建物検出モデルを用いて, 対象地域全域の検出結果を報告した. しかし, 建物検出の精度は, 建物の大きさや密集の程度 (立地特性) に依存している可能性がある. そこで, まず 2 章で構築した超解像モデルを Sentinel-2 に適用し, 解像度 2.5m の衛星画像を得た. 続いて, 建物密度が低い, あるいは大きな建物ほど検出しやすいと考え, 超解像後の衛星画像を目視で観察し, 棟数が少ない, 多い, 大規模建築物を含む, の 3 つに分類したデータセットを作成した (表 1).

3.1 ロジスティック回帰による分類モデルの概要

事例 $\mathbf{x}$ がクラス C_j に分類される条件付き確率を算出し (式 (3.1)), 全訓練データ・全クラスの交差エントロピー誤差 E を算出し (式 (3.2)), 勾配降下法により重み行列 $\mathbf{W}$ を求めた (式 (3.3)). ハイパーパラメータである学習率は, 多数の学習率による勾配降下法を行い, 検証データにおける精度の高い値を

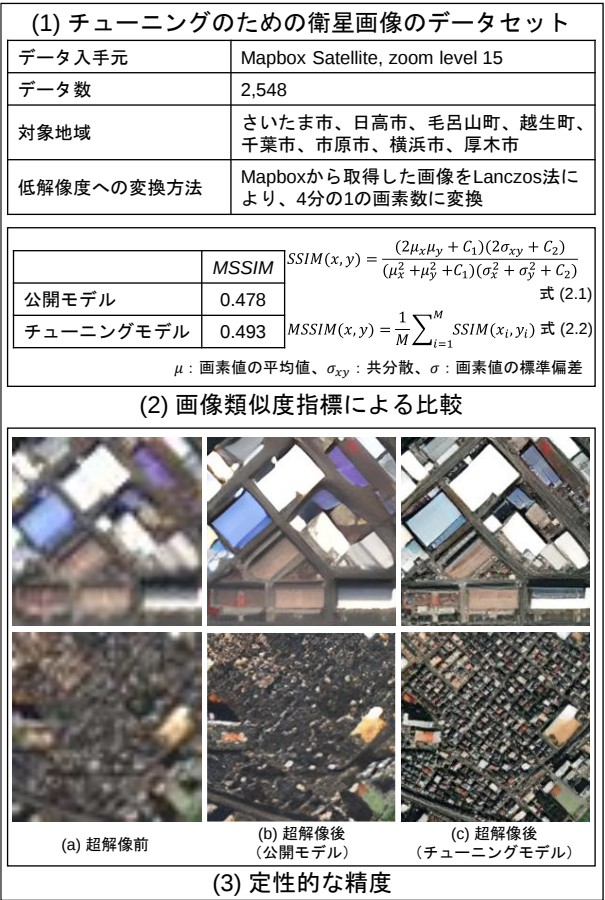


図2 超解像モデルのチューニング結果

表1 立地特性の分類モデルのデータセット数

ラベル名	データセット数			画像サイズ
	学習用	検証用	テスト用	
ラベル0 (棟数少ない)	2,001	521	1,088	200 x 200ピクセル
ラベル1 (大規模建物)	150	46	84	
ラベル2 (棟数多い)	2,472	589	1,305	

$P(\hat{y}_j = C_j | \mathbf{x}) = \frac{\exp(\mathbf{W}_j \mathbf{x})}{\sum_{k=1}^K \exp(\mathbf{W}_k \mathbf{x})}$ 式 (3.1)

$E = - \sum_{i=1}^n \sum_{k=1}^K y_{ik} \ln \hat{y}_{ik}$ 式 (3.2)

$\mathbf{W}^{(t+1)} = \mathbf{W}^t - \eta_t \nabla E(\mathbf{W}^t)$ 式 (3.3)

$NDVI = \frac{IR - R}{IR + R}$ 式 (3.4)

$x_i = \frac{\sum_{j=1}^m w_{ij} x_j}{\sum_{j=1}^m w_{ij}}$ 式 (3.5)

$\mathbf{x}$: 特徴量ベクトル
 $\mathbf{W}$: 重み行列
 K : クラス数
 n : 学習データ数
 y : 正解ラベル
 η : 学習率
 IR : 近赤外域の反射率
 R : 可視域赤の反射率
 m : 全メッシュ数
 $\mathbf{w}$: 空間的隣接行列

変数	説明	データ入手元
人口	2020年時点	WorldPop
NDVI (式 (3.4))	2022年1月、4月、9月、10月の平均値	Sentinel-2A, 2B
夜間光量	2022年の平均値	Earth Observation Group
海岸線までの距離	ダミー変数	国土数値情報 海岸線データ (2006年)

図3 ロジスティクス回帰の定式化・変数

採用した.
建物データが整備されていない地域についても対

応可能とするため、容易に入手可能な変数 $\mathbf{x}$ （人口，NDVI，夜間光量，海岸線までの距離）を用いて，ラベル付けした結果を推定する多クラス分類モデルを構築した．この内，海岸線までの距離以外の変数については，空間的連担性のある工業地帯等を表現するために，式(3.5)に示すメッシュ i と j の空間的隣接関係を表す行列 $\mathbf{w}$ を i と j が隣接する場合は 1，それ以外は 0 とした．隣接行列 $\mathbf{w}_{ij}$ とメッシュ j の変数 x_j を掛け合わせ， $\mathbf{w}_{ij}$ の総数で除すことにより，最大隣接 8 メッシュの属性値の平均値を算出した．

3.2 Vision Transformer による分類モデルの概要

Vision Transformer（以下，ViT とする）は，画像を固定サイズのパッチに分割し，これらパッチを位置情報を付したトークンとして扱うことで画像認識を行うモデルである．画像認識を行う際に一般的であった畳み込み処理を行う必要がないことが従来の手法と異なる特徴である．本稿では，事前学習済みの ViT モデル（vit_base_patch16_224）を用い，表 1 の学習用・検証用データセットを用いて，ファインチューニングを行った．

3.3 分類モデルの推定結果

図 4 にロジスティック回帰及び ViT による立地特性分類の混合行列と F 値を示す．ラベル 0 と 2 は，両分類モデルで高い推定精度が確認できるが，いずれも ViT による精度の方が高い．ラベル 1 は，ロジスティック回帰による F 値は 0.333 と低いが，ViT による F 値は 0.482 であり，大幅に向上していることが確認できる．ロジスティック回帰による分類よりも画像認識による分類の方が，すべてのラベルで高い精度で立地特性を分類できることが分かる．

4. 建物立地特性を考慮した建物検出モデル

4.1 建物検出モデルの概要

表 1 のテスト用データセットを各分類モデルで分類した結果，表 2(1)の内訳のとおりとなった．ラベル 1 のデータ数は著しく少なく，建物検出モデルを学習させるには不十分な数であるため，回転，水平・垂直移動，シア変換，ズーム，左右反転を施し，データ数を増やした（表 2(2)）．なお，表 2(2)のデータの内，8 割を学習用，2 割を検証用とした．また，

(1) ロジスティック回帰による立地特性分類の精度				
正解 \ 推定	ラベル0	ラベル1	ラベル2	F 値
ラベル0（棟数少ない）	1,012	13	63	0.904
ラベル1（大規模建物）	20	21	43	0.333
ラベル2（棟数多い）	118	8	1,179	0.910

(2) ViTによる立地特性分類の精度				
正解 \ 推定	ラベル0	ラベル1	ラベル2	F 値
ラベル0（棟数少ない）	1,022	4	62	0.942
ラベル1（大規模建物）	11	34	39	0.482
ラベル2（棟数多い）	50	19	1,236	0.936

図4 立地特性分類の精度（混合行列・F値）

表2 建物検出モデルのデータセット数

ラベル名	(1) 元データ数	(2) 画像編集後の数	(3) テスト用
ViT ラベル0	1,027	1,500	56
ViT ラベル1	57	180	50
ViT ラベル2	1,270	1,500	67
ロジスティック ラベル1	1,092	1,500	58
ロジスティック ラベル2	42	180	50
ロジスティック ラベル3	1,220	1,500	65
ラベル分類なし	1,282	1,500	68

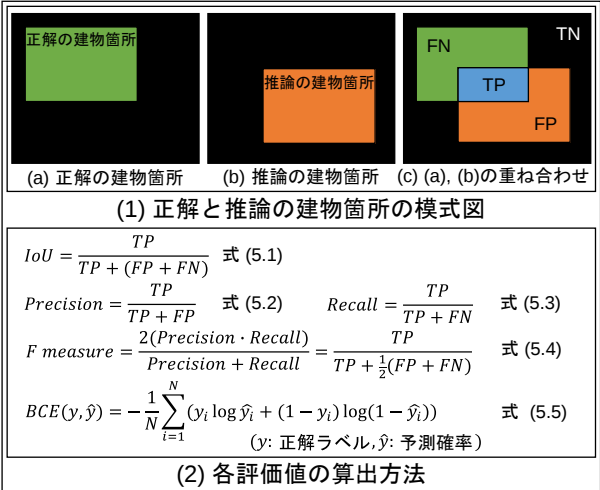


図5 各評価指標の算出方法

分類手法による検出精度への影響を測るために，データ数の少ないラベル 1 を除いて，データ数を 1,500 に統一した．

正解となる建物位置のマスク画像は，国土地理院の基盤地図情報を基に作成した．検出モデルは，建物位置をピクセル単位で推定するセマンティックセグメンテーションモデル U-Net を使用し，バックボーンには ResNet152 を用いた．学習時の損失関数は，Jaccard Loss（以下，JL とする），交差エントロピー誤差（以下，BCE とする），及び JL と BCE の和の計 3 つを設定し（図 5），分類手法の違いに加えて，損失関数の違いによる検出精度の影響も検証した．

4.2 建物検出モデルの推定結果

分類手法別・損失関数別の建物検出の精度（F 値）

を図 6 に示す。ラベル 1 の ViT で分類した結果に基づく検出精度が最も高く、F 値は 0.814 であり、大規模な建物は、良好な精度で検出できることが分かる。

ラベル 0 の検出精度は高くないが、これは棟数が少なくかつ建物も小さいため、式(5.1)の分子 TP (建物箇所を建物と推定した数) が小さいことが起因している。ラベル 0 と 2 では、分類手法の違いによる検出精度は、設定した損失関数により異なっている。

立地特性を分類しないで、1 つの建物検出モデルを構築した場合は、分類結果に基づいたモデルのラベル 2 を下回る程度の精度である。このことから、立地特性毎に建物検出モデルを構築することの優位性を確認できる。

ViT の分類結果に基づいて学習させた建物検出モデルによる検出結果を図 7 に示す。ラベル 1 の大規模な建物は、良好に建物を検出できている。

ラベル 0 と 2 について、損失関数を JL とした場合、小さな建物は複数棟がまとまって検出されてしまう (図 7(1)(c), (3)(c))。これは、IoU は、TN (非建物箇所を非建物と推定した数) を考慮した関数でないことが関係している。一方で、BCE は、建物箇所と非建物箇所をバランスよく評価する関数であるため、建物か否か明確でないピクセルはグレーとなりやすい (図 7(1)(d), (2)(d), (3)(d))。これにより、JL より建物が分割されていることが確認でき、小さな建物を検出する際には、BCE を加味した損失関数の設定が望ましいと考えられる。

5. まとめ

立地特性の分類について、容易に入手可能な変数を用いたロジスティック回帰分類よりも ViT による画像認識の方が、高い精度で分類できることを確認した。特に、大規模建築物を有する地域の分類において、ViT による分類は有効であることを確認した。

建物検出モデルについて、分類手法及び損失関数の違いによる検出精度への影響度合いを検証した。その結果、立地特性毎に建物検出モデルを構築することにより、ラベル 1 で非常に高い精度を確認した。また、対象地域の建物立地特性に応じて、適切な損失関数を設定することの必要性について言及した。

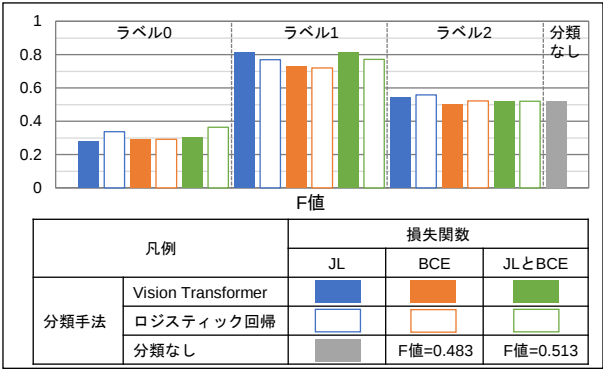


図6 分類手法別・損失関数別の建物検出精度

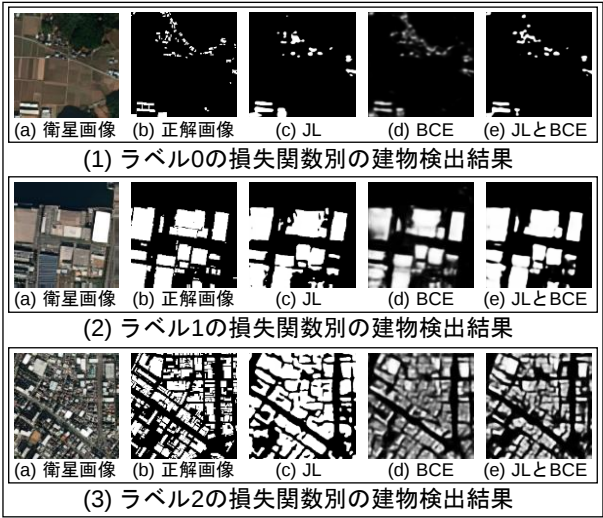


図7 ViT分類に基づく建物検出モデルの推論結果

謝辞

本研究は、日本科学協会の笹川科学研究助成による助成を受けたものである。本研究の遂行にあたり、適宜助言を頂いた東京科学大学 環境・社会理工学院 松岡 昌志教授に深謝致します。

参考文献

Chen, S., Ogawa, Y., Zhao, C. & Sekimoto, Y. (2023) Large-scale individual building extraction from open-source satellite imagery via super-resolution-based instance segmentation approach, *ISPRS Journal of Photogrammetry and Remote Sensing*, **195**, pp. 129-152.

Wang, X., Xie, L., Dong, C. & Shan, Y. (2021) Real-ESRGAN: Training Real-World Blind Super-Resolution With Pure Synthetic Data, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pp. 1905-1914.

書き手はどこを訪れたか？
- 言語モデルで訪問行動を読み取る -

大友寛之*・大内啓樹**・東山翔平***・寺西裕紀****
山本和太郎*****・進藤裕之*****・渡辺太郎*****

Where did the Writer Visit?
- Understanding Human Visit Behavior through Language Models -

Hiroyuki Otomo*, Hiroki Ouchi**, Shohei Higashiyama***, Hiroki Teranishi****,
Aitaro Yamamoto*****, Hiroyuki Shindo*****, Taro Watanabe*****

Abstract: In this work, we address Visit Status Prediction (VSP) by using language models (LMs). The goal of VSP is to identify locations that the writer of each travelogue visited by assigning a visit status for each location mentioned in the travelogue. For example, consider the sentence: “Arrived at Kintetsu Nara Station!” It is possible to judge that the writer visited the station from the description of the real experience. In contrast, consider the sentence: “JR Nara Station is a little far from Kintetsu Nara Station.” This factual statement does not indicate whether the writer visited these stations or not. We investigate how accurately LMs can recognize such visit status. Our experiments demonstrate that LMs achieve around 80% accuracy. Our analysis also reveals that even in cases where the writers did not visit, the LMs tend to mistakenly interpret them as visited, which suggests the points for future improvement.

Keywords: 訪問行動 (Visit Behavior), 旅行行動 (Traveling Behavior), 旅行記 (Travelogue), 言語モデル (Language Models), 自然言語処理 (Natural Language Processing)

1. はじめに

人間がある場所を訪れ、そこで行ったこと、思ったことなど、雑多な経験を綴った文章がウェブ上に蓄積されている。こうした膨大な情報を計算機によって自動で整理・構造化することで、多様な応用のための資源として新たな価値が付加される。

本研究では、場所の経験を綴った記述を自動抽出する問題に取り組む。特に、書き手の訪問行動を読み解き、場所に対する訪問状態（訪れた、訪れていない、訪れる予定、など）を判定するタスクに取り組む。たとえば、「奈良公園にはたくさんの鹿がいました」という記述は、「奈良公園」を書き手が実際に訪れたと解釈できる。一方で、「奈良公園は人気の観光スポットです」という記述は、一般的な事実を表しており、書き手の訪問を必ずしも意味しない。

このような訪問行動に関する文脈上の解釈・判断を、機械学習モデルはどの程度正確に行えるだろうか？この問いを本研究では探求する。具体的には、言語モデルを対象としてその正確性を調査する。昨今の言語モデルは、種々の文章理解タスクにおいて高い正確性を示している。このような背景を鑑み、言語モデルの中でも代表的 Masked Language Model (MLM)と Causal Language Model (CLM)を対象とし、訪問行動認識の正確性を調査する。実験の結果、上記2種類の言語モデルが8割程度の正解率を記録した。また、モデルの予測エラーの分析を通じて、書き手が「訪れていない」場合でも、言語モデルが誤って「訪れた」と解釈してしまうケースも比較的多く見られ、今後の改善の余地について明らかにすることができた。

* 非会員 株式会社サイバーエージェント (CyberAgent, Inc.)
** 正会員 奈良先端科学技術大学院大学・先端科学研究科 (Nara Institute of Science and Technology)
〒630-0101 奈良県生駒市高山町8 9 1 6-5 hiroki.ouchi@is.naist.jp
*** 非会員 情報通信研究機構 (National Institute of Information and Communications Technology)
**** 非会員 理化学研究所・革新知能統合研究センター (RIKEN AIP)
***** 非会員 株式会社 NTT ドコモ (NTT DOCOMO, Inc.)
***** 非会員 MatBrain 株式会社 (MatBrain, Inc.)
***** 非会員 奈良先端科学技術大学院大学・先端科学研究科 (Nara Institute of Science and Technology)

2. 問題設定

2.1. データセット

「地球の歩き方旅行記データセット¹」(株式会社地球の歩き方, 2022; Ouchi et al., 2023)に場所参照表現等のアノテーションを施した ATD-MCL² (Arukikata Travelogue Dataset with Geographic Entity Mention, Coreference, and Link Annotation) (Higashiyama et al., 2024)に対して, 訪問状態ラベルを付与したデータセット(山本他, 2024)を使用する.

データセットに付与されている 6 種類の訪問状態ラベルを表 1 に示す. 最も基本的なラベル Visit は, 書き手が言及する場所(表中の例だと「赤岳山頂」)を訪れたことが読み取れることを意味する. 他のラベルは, 書き手が当該旅行中にその場所を訪れたことが読み取れない場合に付与される. たとえば PlanToVisit は, 当該旅行中に訪問予定であることが言及されている場合に付与される.

データセットの記述統計を表 2 に示す. 表中の Vis は Visit, Plan は PlanToVisit, Past は Visit-Past, Fut は Visit-Future, UN は UnkOrNotVisit にそれぞれ該当する. モデルの性能評価実験のため, データセット全体を学習/開発/評価セットに分割した. セットごと, ラベルごとの事例数を表 2 に示す. これを見ると, ラベルの分布に偏りがあることがわかる. つまり, Visit が最も多く, 全体の約 2/3 を占める. 次に UnkOrNotVisit と PlanToVisit と See が多く, Visit-Past と Visit-Future はほとんど登場しない傾向がある.

2.2. 問題設定

ひとつの場所参照表現(表 1 の例では下線で表示)とその周辺文脈が入力として与えられ, モデルは訪問状態ラベルを出力する. 以下に例を示す.

入力: 穂高神社に向かいます.

出力: PlanToVisit

ここで, MLM では 6 つのラベルから 1 つのラベルを選ぶ問題(多クラス分類)として解く. 一方, CLM では, ラベルに該当する文字列を生成することによって解く.

表 1 訪問状態ラベルの説明

ラベル	定義
1. Visit	参照先の場所を訪問したことに言及している. 例) 無事に <u>赤岳山頂</u> に着きました!
2. PlanToVisit	当該旅行中に, 参照先の場所を訪問する予定であると言及している. 例) <u>穂高神社</u> に向かいます.
3. See	上記 1. と 2. に該当せず, かつ, 参照先の場所を視認したことに言及している. 例) <u>硫黄岳</u> が近くに見えて来ました.
4. Visit-Past	上記 1. から 3. のいずれにも該当せず, かつ, 当該旅行前に参照先の場所を訪問したことに言及している. 例) 数年前に訪れた <u>フランス・ルルドの泉</u> を思い出しました.
5. Visit-Future	上記 1. から 4. のいずれにも該当せず, かつ, 当該旅行後に参照先の場所を訪問する意思があると言及している. 例) 今度は, <u>松江</u> と <u>米子</u> に来てみたいものです.
6. UnkOrNotVisit	上記いずれにも該当しない. 例) <u>糸魚川駅</u> 行きの車両は 1 両です.

表 2 データセットの記述統計

	Vis	Plan	See	Past	Fut	UN
学習	2,577	358	212	10	6	619
開発	332	48	46	1	4	74
評価	748	121	59	10	4	160

¹ <https://www.nii.ac.jp/dsc/idr/arukikata/>

² <https://github.com/naist-nlp/atd-mcl>

3. 実験

3.1. モデルの詳細

3つのモデルを設定し、それらの性能比較を行う。一つめのモデル「最頻ラベル」は、最も頻度の高い訪問順序ラベル Visit を常に出力する。これは一定の基準を示す目的で使用する。二つめのモデルとして、MLM の LUKE (Ri et al, 2023)を採用する。その実装として機械学習ライブラリ Hugging Face Transformers³から提供されている studio-ousia/mluke-large-lite⁴を使用した。事前学習済みモデルを学習データでファインチューニングした。三つめのモデルとして、CLM の Llama (Touvron et al, 2023)を採用する。その実装として elyza/ELYZA-japanese-Llama-2-7b-instruct⁵を使用した。事前学習済みモデルをベースに学習データで LoRA チューニング(Hu et al, 2022)を行った。

3.2. 性能比較結果

表 3 訪問状態予測の性能比較

モデル	正解率
最頻ラベル	.679
LUKE	.789
Llama	.781

表 3 に正解率を示す。MLM の LUKE と CLM の Llama は同等の正解率として 78%を記録した。モデルパラメタ数の観点から考察すると、パラメタ数の少ない LUKE (5.6 億パラメタ) の方が、Llama (70 億パラメタ) よりも効果的にタスクを遂行できていると解釈できる。この理由として、解き方の違いとその困難性が挙げられる。LUKE では多クラス分類問題として訪問状態ラベルを予測しているのに対し、Llama では言語生成問題としてラベルを生成している。一般に多クラス分類よりも言語生成の方が難しく、学習データ量も要すると言える。この解き方の違いが、パラメタ数の開きにも関わらず性能差を生まなかった原因であると推察される。

³ <https://huggingface.co/docs/transformers/index>
⁴ <https://huggingface.co/studio-ousia/mluke-large-lite>

3.3. ラベルごとの性能比較

表 4 訪問状態ラベルごとの正解率

	事例数	LUKE	Llama
Visit	332	.930	.903
PlanToVisit	48	.666	.479
See	46	.760	.826
Visit-Past	1	.000	.000
Visit-Future	4	.000	.000
UnkOrNotVisit	74	.540	.391

開発データを対象として、モデルの解析結果について詳細に分析する。表 4 に、訪問状態ラベルごとの正解率を示す。データセット中で最も多いラベル Visit については 9 割以上を正解できている。See についてもどちらのモデルも 8 割前後を正解できている。一方で、PlanToVisit と UnkOrNotVisit については改善の余地が残る。これらのラベルは頻度も中程度あり、Visit と明確に区別できることが望ましいため、今後さらなる改善が必要となる。

3.4. 予測傾向のモデル間比較

表 5 訪問状態ラベルごとの正解事例数

LUKE Llama	○ ○	○ ×	×	×
Visit	287	22	13	10
PlanToVisit	18	14	5	11
See	34	1	4	7
Visit-Past	0	0	0	1
Visit-Future	0	0	0	4
UnkOrNotVisit	28	12	1	33

LUKE と Llama の正解・不正解の事例数について表 5 に示す。LUKE と Llama の両方のモデルが正解だった事例数を「○|○」の列に示す。同様に、LUKE のみが正解だった事例数を「○|×」、Llama のみが正解だった事例数を「×|○」、両方のモデルで不正解だった事例数を「×|×」の列に示す。最も多いラベル Visit に関しては、どちらのモデルでも正解できている割合が多かった。

⁵ <https://huggingface.co/elyza/ELYZA-japanese-Llama-2-7b-instruct>

ここで UnkOrNotVisit に着目すると、両方のモデルで不正解だった事例数(33)が、両方のモデルで正解だった事例数(28)を上回った。具体的にはどのような事例が両方のモデルにとって難しいかを、UnkOrNotVisit の具体例を観察しながら考察する。

例 1) ホテルの目の前の気になるお店ですが、今回はパスしました。

例 2) 今回は、松江と米子は素通りします。

これらの例で「お店」「松江」「米子」を書き手は訪れていないことが、「～パスしました」「～素通りします」等の語句から読み取れるため、正解のラベルは UnkOrNotVisit となる。しかし,LUKE と Llama はともに Visit ラベルと予測している。その理由として、上記の語句の意味を適切に捉えられなかった可能性が挙げられる。

3.5. 正解ラベルと予測ラベルの混同行列

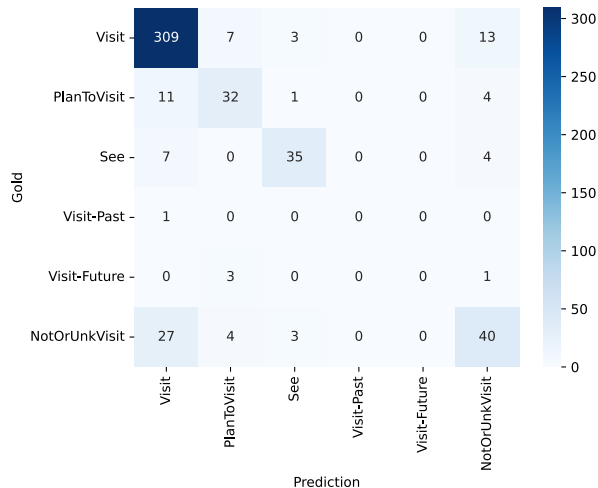


図 1 LUKE の混同行列

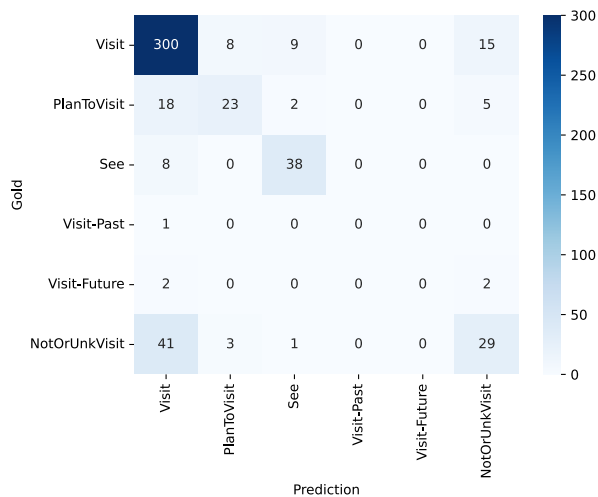


図 2 Llama の混同行列

正解/予測ラベルの混同を図 1 と図 2 に示す。「Gold」は正解ラベルを表し、「Prediction」は予測ラベルを表す。たとえば図 1 では、Gold が Visit で Prediction も Visit であるのが 309 件ある。つまり、対角線が正解事例数となる。一方、Gold が NotOrUnkVisit で Prediction が Visit であるのは 27 件ある。図 2 の Llama も同様のケースが 41 件ある。したがって、書き手が「訪れていない」のに「訪れている」と解釈してしまうケースが比較的多い。今後いかに「訪れていない」ケースで誤りを減らせるかが、モデルの正解率改善の鍵となることが示唆された。

4. おわりに

本研究では、機械学習モデルがどの程度の正確性で文章から訪問行動を読み取れるかを調査した。言語モデルを用いた実験の結果、8 割程度の正解率を記録した。詳しい分析を通じて、書き手が「訪れていない」場合でもモデルが「訪れた」と解釈してしまうケースも見られ、今後の改善点を明らかにした。

謝辞 Acknowledgement

本研究は JSPS 科研費 JP23K24904 の助成を受けたものです。

参考文献 References (in Harvard style)

株式会社地球の歩き方 (2022) 地球の歩き方旅行記データセット.
国立情報学研究所情報学研究データリポジトリ.
<https://doi.org/10.32130/idr.18.1>
山本和太郎他 (2024) 移動軌跡解析: 文章中の人物の地理的な移動を読み取る. 言語処理学会第 30 回年次大会予稿集.
Shohei Higashiyama et al, (2024) Arukikata Travelogue Dataset with Geographic Entity Mention, Coreference, and Link Annotation. In Proceedings of EACL.
Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen (2022) LoRA: Low-Rank Adaptation of Large Language Models. In Proceedings of ICLR.
Ri Ryokan Ri, Ikuya Yamada, Yoshimasa Tsuruoka (2023) mLUKE: The Power of Entity Representations in Multilingual Pretrained Language Models. In Proceedings of ACL.
Hiroki Ouchi et al, (2022) Arukikata Travelogue Dataset. arXiv.
Hugo Touvron et al., (2023) Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv.