

Instance Segmentation を利用した屋内の動的環境における 潜在的動物体に基づく SLAM の改善

Improvement of SLAM based on Potential Dynamic Objects
by Using Instance Segmentation in Indoor Dynamic Environments

○学 押久保 雄平 (中央大学) Sarthak Pathak (中央大学)
正 池 勇勳 (JAIST) 正 梅田 和昇 (中央大学)

Yuhei OSHIKUBO, Chuo University, oshikubo@sensor.mech.chuo-u.ac.jp

Sarthak PATHAK, Chuo University, pathak@mech.chuo-u.ac.jp

Yonghoon JI, JAIST, ji-y@jaist.ac.jp

Kazunori UMEDA, Chuo University, umeda@mech.chuo-u.ac.jp

In this paper, we improve a robust SLAM for dynamic environments by using object detection considering potential dynamic objects. Generally, dynamic environments have a negative impact on SLAM because correspondence between frames is wrong and constructed map includes ghosts. Conventional methods resolved this problem by predefining dynamic objects and removing them, however these cannot handle degenerate environments. Therefore, we define potential dynamic objects to increase the number of features as much as possible and to improve the problem of degeneracy that occurs in conventional methods. In this study, we use instance segmentation, which can detect objects more accurately at the pixel level, and verify its effectiveness.

Key Words: SLAM, Object Detection, Scan Matching, Mobile Robot

1 序論

近年、労働人口の減少に伴い空港や工場などの限定された空間では自律移動ロボットが数多く導入されている。これらの自律移動ロボットは一般に事前に作成された高精度な地図を利用していることが多い。

地図構築においてセンサを搭載した移動体が自己位置推定と地図構築を交互に繰り返す Simultaneous Localization and Mapping (SLAM) がある。既に SLAM は実用化されているが、静的環境を前提としていることが多い。動的環境では異なる時刻から取得した情報の誤対応が多くなることで自己位置推定精度の低下や構築した地図に不必要な点群が残ることを招く。また、移動体が完全に存在しない環境を作りだすことは難しい。よって、動的な情報を考慮して SLAM を行う必要がある。これらの問題を解決するため、動的環境に対応した SLAM や動的物体の検出方法に関する研究が数多くなされてきた。Zhong らは物体検出器を用いて移動する可能性がある物体を全て除去しながら SLAM を行う手法を提案した [1]。しかし、画像・形状特徴量に乏しい環境では退化 [2] が発生し、自己位置推定が失敗する恐れがある。退化は廊下やトンネルなどの形状が類似している環境で起こりやすく、位置合わせにおいて局所最適解が多数発生することから自己位置を一意に推定することが難しくなる。また、物体検出器を必要としない差分による方法 [3, 4] も提案されている。状況によっては汎用性が高いが、動的物体の検出に正確なオドメトリ推定が必要とされる。そのため、特徴が少ない環境ではオドメトリの誤差が大きくなり、動的物体の検出が失敗する恐れがある。

我々は潜在的な動物体（準動的物体）と先験的な動物体（動的物体）の関係性に着目し、屋内の特徴が少ない動的な環境においてもロバストな SLAM を提案した [5]。しかし、オフラインでの処理が前提であったことや物体点群の抽出においてノイズを十分に除去できていないことが課題であった。そこで、本論文では物体検出器をリアルタイムでピクセルレベルの検出が可能なインスタンスセグメンテーションモデルに変更し、新たに SLAM システムを提案する。潜在的・先験的物体の定義については、2.2 節で詳しく述べる。また、将来的には自律移動と組み合わせることを想定し、オンラインでの地図構築を想定する。本手法は自律移動における事前地図情報の作成や Active-SLAM [6] に活用することができる。

2 提案手法

2.1 概要

本手法全体の大まかな流れを図 1 に示す。RGB-D カメラにおける 1 フレームの入力に対して逐次的に処理を行い、既存の SLAM [7] に利用する。まず、RGB 画像と距離画像を取得する。得られた RGB 画像に対して深層学習による物体検出器を活用してピクセルレベルで画像中の物体の位置を求める。次に、あらかじめ RGB 画像と距離画像の対応が既知であることを利用して検出した各物体の 3 次元点群に変換する。その後、各物体の点群座標と付与した属性の関係性を元に各点群の SLAM への利用方法を決定する。以下、各処理について詳細を述べる。

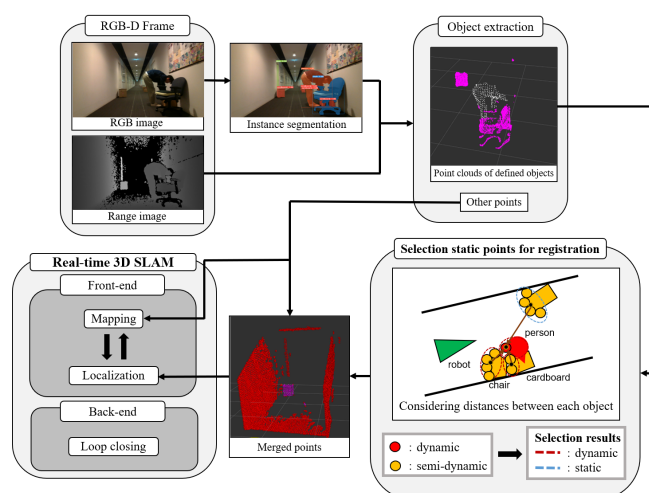


Fig.1 Flow of the proposed method

2.2 物体の定義と前提条件

本研究では、物体を静的物体、潜在的動物体、先験的動物体の 3 種類に分類する [8, 9, 10]。潜在的動物体と先験的動物体は物体自身の移動可能性によって判別する。先験的動物体については移動可否に関わらず SLAM に利用しない。一方、潜在的動物体

については、2.5 節で述べる方法により必要に応じて SLAM に利用する。また、人を先験的動物体、段ボールと椅子を潜在的動物体と定義し、検出器によって検出できなかった領域は静的として扱う。前提として屋内環境を想定し、風などの影響は無いとする。よって、潜在的動物体は先験的動物体のみによって動かされると仮定し、この関係性に着目する。例えば、段ボールや台車などの潜在的動物体が自身で動くことはあり得ず、動くときは人やロボットなどの先験的動物体が作用していると言える。

2.3 物体検出

本手法では、RGB-D カメラで取得した RGB 画像に対して YOLACT[11] を用いてリアルタイムのインスタンスセグメンテーションを行う。本研究で検出したい物体は事前に学習されていなかったため、著者が撮影した 75 枚の画像を拡張して事前学習済みモデルに対してファインチューニングを行い、人、段ボール、ドアをそれぞれ検出できるようにする。出力結果の例を図 2 に示す。次に YOLACT の推論結果から得られた情報を 2.4 節での物体点群の抽出に利用する。

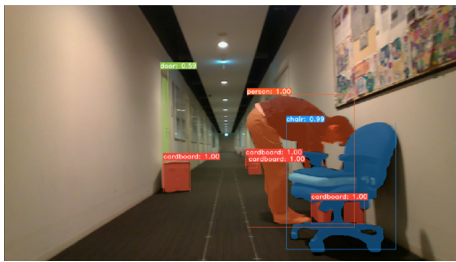


Fig.2 Detection result by YOLACT in customized model

2.4 物体点群の抽出

距離画像はセンサの内部パラメータが既知の場合に 3 次元点群に変換できる。よって、2.3 節で求めた RGB 画像中における各ピクセルで求めた物体情報と距離画像の対応を取ることで物体の 3 次元点群を構築することが可能となる。しかし、一般的に RGB-D カメラで取得した画像の解像度が高いほど点群数は増え、密になることから変換した点群は計算量削減のためにダウンサンプリングを行う。本研究では点群の密度が均一になるようにボクセルグリッドダウンサンプリングを用いる。また、RGB 画像中において物体が検出されなかった領域に対しても点群への変換及びダウンサンプリングを行い、保存する。この点群は静的点群として扱う。

2.5 潜在的動物体を考慮した静的点群の選択

2.2 節でも述べたように、屋内における潜在的動物体の移動可否については先験的動物体が起因している。すなわち、潜在的動物体の近くに先験的動物体が存在する場合に潜在的動物体の動いている可能性が高いと言える。よって、本手法では潜在的動物体と先験的動物体の物理的な距離を計算することで、各フレームごとに潜在的動物体の状態を判断する。

最初に、各物体点群に対して 3 次元座標の平均をとることでそれぞれの重心を求める。式 (1) の \mathbf{X} は 1 フレーム内で検出した各潜在的動物体の重心の集合、式 (2) の \mathbf{Y} は 1 フレーム内で検出した各先験的動物体の重心の集合である。

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i\} \quad (1)$$

$$\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_j\} \quad (2)$$

i は潜在的動物体の数、 j は先験的動物体の数である。各物体の重心を求めた後、任意の潜在的動物体の重心から先験的動物体の重心までの距離を求める。

$$|\mathbf{x}_n - \mathbf{y}_m| = d \quad (3)$$

それぞれ複数の物体が存在したとき、 n 番目の潜在的動物体の重心 \mathbf{x}_n から先験的動物体の重心 $\mathbf{y}_m (m = 1, 2, \dots, j)$ までの 3 次元ユークリッド距離 d を式 (3) を用いてそれぞれ求める。次にしきい値判定を行う。全ての値がしきい値 λ 以上であれば、重心 \mathbf{x}_n を持つ潜在的動物体の点群を静的として扱う。反対に 1 つで

も求めた距離がしきい値以下である場合は、その潜在的動物体を動的とみなして除去し、位置合わせにも地図構築にも用いない。本研究ではしきい値 λ を経験的に 1 m としている。

図 3(a) は図 2 の場面を上から俯瞰した図である。円がセンサから得られた点群の各点であり、黄色の点が潜在的動物体、赤色の点が先験的動物体に該当する。図 3(b) から重心間距離が λ より小さい d_1 のときは潜在的動物体を動的、 λ より大きい d_2 のときは潜在的動物体を静的と判断していることが分かる。ここで静的と判断された点群は SLAM における点群の位置合わせに基づく自己位置推定に活用する。

また、図 2 で示した RGB 画像と対応する距離画像を元に物体抽出した点群を図 4(a)、本節の処理によって SLAM の自己位置推定に利用すると判断した点群を図 4(b) に示す。先験的動物体である人とその近くにある潜在的動物体の点群が取り除かれ、奥の潜在的動物体である段ボールの点群が残っていることがわかる。これにより、動的点群による SLAM への悪影響を取り除きつつ、できる限り多くの静的点群を利用することが可能となる。また、潜在的動物体を地図の情報に含んでも事前情報として利用する際に移動している可能性がある。よって、提案手法で判断した潜在的動物体の状態によらず、潜在的動物体は地図構築に用いない。

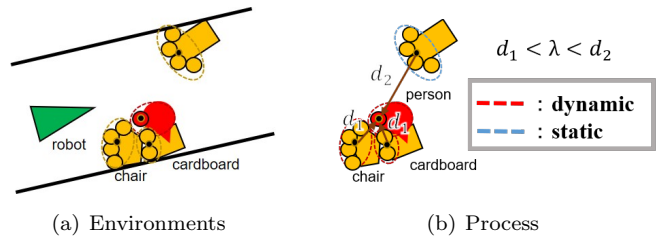


Fig.3 Selection of static point clouds for registration

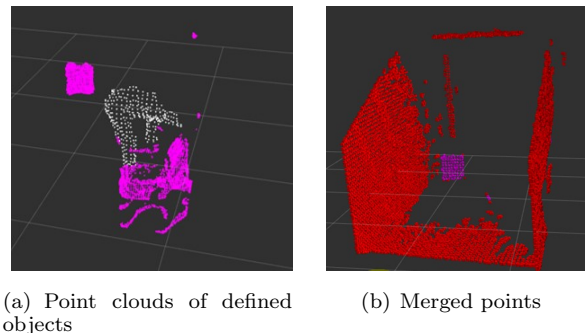


Fig.4 Processing results of point clouds in the proposed method

3 実験

特徴が少ない動的環境において提案手法のロバスト性を検証するために実験を行った。

3.1 実験条件

シナリオとして図 5(a) に示す直線的な廊下において人が段ボールを運んだ。実験環境には事前に定義した潜在的動物体である椅子や段ボールが設置されており、実験中に先験的動物体である人によって動かされる。本研究では、ロボットとしてヴィストン株式会社のメガローパー Ver.3.0 を用い、図 5(b) に示す RGB-D カメラである Intel 製 RealSense LiDAR Camera L515 を固定した。制御はロボットに搭載した GPU 付きノートパソコンで著者が行い、事前に計測して貼り付けた目印に沿って計 10 m 前進させた。評価は定量的に自己位置推定誤差で、定性的に構築した地図の不必要な点群の有無で行った。自己位置推定誤差は通常の SLAM、動く可能性がある物体を全て除去する関連手法、提案手法を組み込んだ SLAM で比較を行った。地図構築結果は通常の SLAM と提案手法を組み込んだ SLAM を比較した。

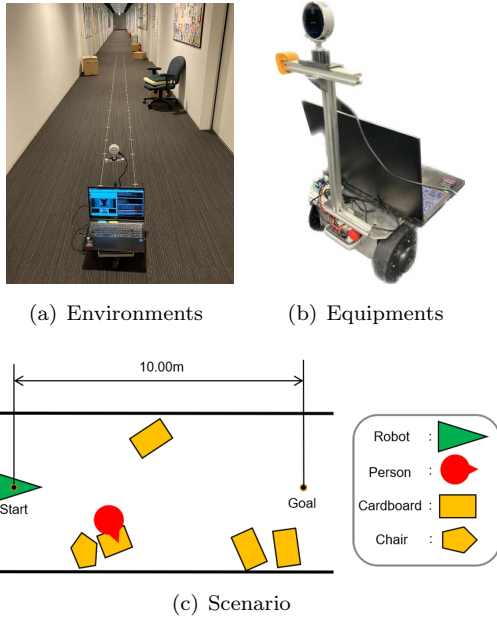


Fig.5 Experimental conditions

3.2 実験結果

まず、自己位置推定誤差の結果を表 1 に示す。自己位置推定の軌跡から横方向の誤差はほぼなかったため、進行方向の誤差に着目し、退化の影響を調べた。自己位置推定誤差の結果から、2つの事柄について述べる。1つめは提案手法では関連手法に対して自己位置推定の精度が高くなったことがわかる。これは、動く可能性があるものを全て除去する関連手法では退化が発生してしまったからだと考えられる。2つめは提案手法において動物体を考慮していない従来手法よりも精度が低下したことである。これは、本シナリオにおいてセンサから取得した情報に占める動的点群の割合が少なく除去せずとも自己位置推定を十分にいたしたことや提案手法で検出漏れが発生し挙動が不安定になってしまったことが原因だと考えられる。

次に、地図構築結果を比較した。人除去処理をしていない場合と提案手法における地図構築結果を図 6 に示す。図 6(a) は構築した点群に人・段ボール・椅子が含まれているのに対して図 6(b) ではこれらの物体がほぼ除去できていることがわかる。このことから、提案手法では動的環境でもロボストに地図構築ができると言える。

Table 1 Translation error for each method

Method	進行方向誤差 [m]
Conventional	0.049
Related	0.196
Proposed	0.079

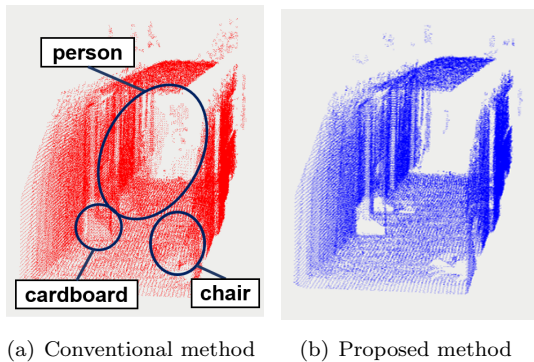


Fig.6 Maps built by each method

4 結論

本研究では、屋内の動的環境においてインスタンスセグメンテーションを用いて検出した物体の属性と距離の関係から動的点群を除去する方法を提案し、SLAM に用いることの有用性を示した。物体の関係性に着目し、インスタンスセグメンテーションを利用して SLAM に組み込んだことが本論文の貢献である。

一方、物体検出器の検出漏れに対応できていないことが問題点として挙げられる。以上より、今後の展望として2つ述べる。1つめは、各フレームで連続して物体を追跡することである。これによって検出漏れに対応できる。2つめは、用いるセンサを LiDAR に変更することである。検出範囲が広がることでより多くの特徴を得ることができると考えられる。自己位置推定の精度が向上するのではないかと考えられる。2.5 節で述べた手法は点群処理であるため物体検出方法は別途検討する必要があるが、LiDAR に対しても利用することが可能である。更に、様々なシナリオでの実験を重ねて本手法の有効性を検証していく。

参考文献

- [1] Fanguwei Zhong, Sheng Wang, Ziqi Zhang, China Chen, Yizhou Wang, "Detect-slam: Making object detection and slam mutually beneficial," *IEEE Winter Conference on Applications of Computer Vision*, pp.1001-1010, 2018.
- [2] 友納 正裕, "SLAM 入門", 株式会社オーム社, p.20, 2018.
- [3] Lukas Schmid, Olov Andersson, Aurelio Sulser, Patrick Pfrendschuh, Roland Siegwart, "Dynablox: Real-time detection of diverse dynamic objects in complex environments," *IEEE Robotics and Automation Letters*, vol. 8, no. 10, pp.6259-6266, 2023.
- [4] David Yoon, Tim Tang, Timothy Barfoot, "Mapless online detection of dynamic objects in 3d lidar," *Conference on Computer and Robot Vision*, pp.113-120, 2019.
- [5] Yuhei Oshikubo, Keigo Akiba, Sarthak Pathak, Yonghoon Ji, Kazunori Umeda, "Improvement of sequential slam by considering semi-dynamic objects in indoor dynamic environment," *Joint Conference on 14th France-Japan and 12th Europe-Asia Congress on Mechatronics & The 9th Asia International Symposium on Mechatronics*, 2023.
- [6] Julio A. Placed, Jared Strader, Henry Carrillo, Nikolay Atanasov, Vadim Indelman, Luca Carlone, José A. Castellanos, "A survey on active simultaneous localization and mapping: State of the art and new frontiers," *IEEE Transactions on Robotics*, vol. 39, no. 3, pp.1686-1705, 2023.
- [7] Kenji Koide, Jun Miura, and Emanuele Menegatti, "A portable 3d lidar-based system for long-term and wide-area people behavior measurement," *International Journal of Advanced Robotic Systems*, vol. 16, no. 2, 2019.
- [8] Xiangcheng Li, Huaiyu Wu, Zhihuan Chen, "Dynamic objects recognizing and masking for rgb-d slam," *International Conference on Intelligent Autonomous Systems*, pp.169-174, 2021.
- [9] Hongjun Zhou, Shigeyuki Sakane, "Localizing objects during robot slam in semi-dynamic environments," *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pp.595-601, 2008.
- [10] Joydeep Biswas, Manuela Veloso, "Episodic non-markov localization: Reasoning about short-term and long-term features," *IEEE International Conference on Robotics and Automation*, pp.3969-3974, 2014.
- [11] Daniel Bolya, Chong Zhou, Fanyi Xiao, Yong Jae Lee, "Yolact: Real-time instance segmentation," *IEEE/CVF International Conference on Computer Vision*, pp.9157-9166, 2019.