

YOLO を用いた人除去処理により動的環境に適応した 屋内 SLAM

秋庭啓吾¹ 鈴木龍紀² 池勇勳³ Sarthak Pathak⁴ 梅田和昇⁵

1 中央大学大学院理工学研究科 〒112-8551 東京都文京区春日 1-13-27

2 中央大学大学院理工学研究科 〒112-8551 東京都文京区春日 1-13-27

3 北陸先端科学技術大学院大学 先端科学技術研究科 〒923-1211 石川県能美市旭台 1-1

4 中央大学理工学部 〒112-8551 東京都文京区春日 1-13-27

5 中央大学理工学部 〒112-8551 東京都文京区春日 1-13-27

概要

本論文では、YOLO (You Only Look Once) を用いた人除去処理を従来の一般的な SLAM (Simultaneous Localization and Mapping) 手法に適用することで、人が存在する屋内の動的環境でもロバストな地図構築が可能な新規 SLAM システムを提案する。一般に、人が多く存在する環境で移動ロボットによる地図構築を行う場合、人によるオクルージョンの影響で高精度な地図構築が困難である。そこで本研究では、深層学習による物体認識技術である YOLO を用いて人を識別し、識別結果に基づいて人のデータを除去することで、人が存在する動的環境下での地図構築の精度を向上させる。

キーワード： 自律移動ロボット 地図構築 SLAM YOLO 人除去

1. 序論

現在、工場や空港など人と共存する屋内環境において、運搬や案内等の目的で自律移動ロボットの導入が進んでいる。近年では、労働人口の減少による人手不足やロボットによる自動化が顕著で、これらの重要性は今後も高まっていくと予想される。また、ロボットが効率よく自律移動するためには、事前に地図を作成しておく必要がある。特に、自律移動ロボットによるナビゲーション[1]では、地図を事前情報として用いる。そのため、これらのロボットの運用では地図の事前作成が必要不可欠であり、その手法としてしばしば SLAM (Simultaneous Localization and Mapping) が用いられる。しかし、人が存在する動的環境における SLAM では、センサ情報に人におけるオクルージョンが生じるだけでなく、人は移動物体であるためセンサ情報同士の位置合わせの際に誤マッチングが生じてしまう問題などがあるため、高精度な地図構築が困難である。そのため、人のデータを除去するなどの対処が求められ、一般的な SLAM は動的環境では用いることができない。ここで藤井らは、測域センサを用いて、スキャンデータの蓄積と遠方点抽出処理により、人混みにおいても人がいない時と同等の壁形状を取得した[2]。しかし、この手法では、壁と障害物の位置関係を利用し、壁より手前の点群を除去することで障害物に対処しているため、人の識別は行えていない。そのため、地図構築時に壁より手前の環境形状が除去されることや、人が壁付近にいる際は人の除去ができず、正確な地図構築ができない問題がある。

そこで本研究では、RGB-D カメラを用いて深層学習を利用した人の識別と人除去処理を SLAM に適用することで、人がいる屋内環境でのロバストな地図構築が可能なシステムを提案する。

2. 提案手法

2.1 概要

提案手法全体の概要を図 1 に示す。本手法では、ロボットに搭載した距離画像センサと LiDAR で取得した RGB-D 画像と 3 次元点群情報を地図構築に用いる。ここで、RGB 画像を用いる場合、画像中における人の

領域情報しか得られないため、RGB-D 画像を用いることで、画像中における人の領域情報と点群の両方が推定可能になる。まず、オドメトリによるロボットの位置予測を行う。次に、LiDAR で取得した点群に対して、画像を用いた人の識別を行い、人に属する点群を除去することで人除去処理を行う[3]。人の識別では、深層学習による高速な物体検出アルゴリズムである YOLO[4]を用いて画像内の人の検出を行う。そして、画像と点群の対応付けによって、人の検出結果を点群にも反映させ、人の点群を検出し除去する。その後、ICP (Iterative Closest Point) [5]アルゴリズムによる点群の位置合わせを行い、オドメトリの誤差を減少させる。ここでは、人除去処理後のフレームの点群と 1 フレーム前の点群に対して ICP による位置合わせを行い、ICP 結果から算出した回転行列 R と並進ベクトル T を用いてロボットの自己位置推定を行うと同時に地図構築を行う。

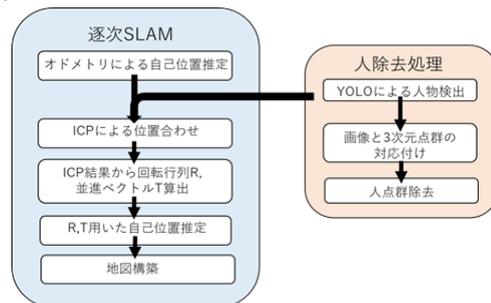


Fig. 1 Outline of proposed method.

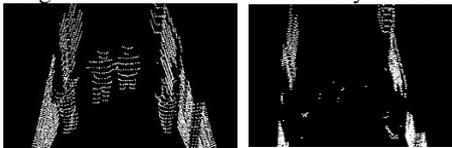
2.1 YOLO を用いた人の 3 次元点群除去処理

本節では、人がいる環境においてロバストな地図構築を行う上で、必要不可欠な人のデータの除去処理について述べる。まず、YOLO による人物検出を行い、図 2 に示すように画像中における人物領域を求める。ここで、この人物領域は人周辺の背景も含む矩形領域である。次に距離画像センサで取得した距離画像から YOLO で求めた人物領域に属する点群のみを抽出する。その点群に対しクラスタリングによる背景除去を行うことで、元の人物領域から背景を除去し、人に属する点群データのみを抽出する。その後、上記プロセスで求めた人物領域の点群と LiDAR で取得した点群対

し、近傍点探索を用いて、距離画像センサで取得した画像の人検出結果を LiDAR で取得した点群に反映させる。この処理によって、人の点群のみを消すことが可能になる。最後に、図 3 に示すように LiDAR で取得した点群の内、対応付けされた人に属する点群を除去することで、人を含まない周辺環境の点群のみを求める。



Fig. 2 Result of human detection by YOLO.



(a) before removal (b) after removal
Fig. 3 Human removal process.

3. 実験

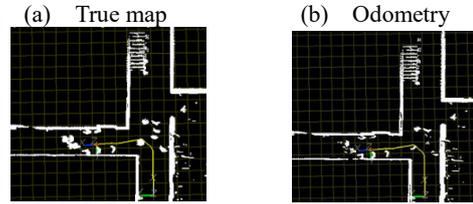
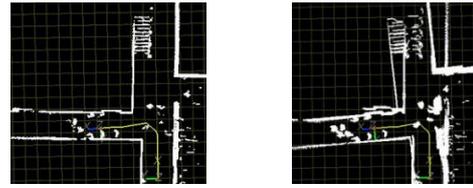
提案手法の検証として、オドメトリによる地図、従来 ICP-SLAM による地図、人除去処理を適用した ICP-SLAM による地図の精度の結果を比較し、評価を行った。なお、地図の精度は、真の地図の各点からそれぞれの手法の地図の各点に対し点間距離を求め、その距離の総和の平均値を誤差とし評価を行った。特に環境地図の精度を評価することが目的のため、人に属する点群部分の計算は行わず、環境の点群のみで計算し誤差を算出した。また、真の地図は各フレームの計測データを手動位置合わせすることで作成した。本実験では、移動ロボットに MobileRobots 製 Pioneer-3DX、距離画像センサに intel 製 RealSense L515、LiDAR に Velodyne 製 VLP-16 を用い、移動ロボットの上に距離画像センサと LiDAR を固定して計測を行った。また、本実験の移動ロボットの軌跡は、図 4 の黄色の線で示す。

本実験は図 2 に示す中央大学後楽園キャンパス 2 号館 7 階廊下にて計 7 フレーム分の計測を行った。実験条件としては、図 2 に示すとおり全てのフレームに人が 4 人存在する環境で計測を行った。また、全てのフレームで 4 人中 1 人が歩行し、残りの 3 人は静止しており、動的環境を再現した。

図 4 と表 1 に示す結果より、人除去処理を行った提案手法を用いた場合に最も誤差が減少しロバストな地図構築が可能であったと分かる。また、図 5 に示す人除去部分の拡大図より、従来手法に比べ、提案手法が環境地図に不要な人の点群がかなり少ないと分かる。つまり、人除去処理により移動ロボットの運用における事前情報として活用しやすい地図の構築が可能であった。

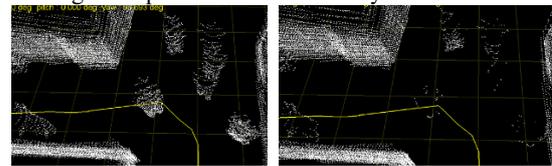
一方で、提案手法の拡大図でも奥にいる人の点群が一部残っていることが確認できる。これは距離画像センサの計測範囲の関係から人の点群の一部が上手く取得できず、除去できなかったためである。また、歩行している人に関しては、距離画像センサと LiDAR で取得した点群に多少ずれが生じてしまうため、その場合は人の点群の一部の対応付けができず除去しきれない

ことがある。



(a) True map (b) Odometry
(c) ICP-SLAM w/o human removal (d) ICP-SLAM w/ human removal

Fig. 4 Map construction results by each method.



(a) ICP-SLAM w/o human removal (b) ICP-SLAM w/ human removal

Fig. 5 Map construction results: enlarged figures of Fig. 4.

Table 1 Error of each map information [m].

Odometry	0.478
ICP-SLAM w/o human removal	0.131
ICP-SLAM w/ human removal	0.097

4. 結論

本研究では、YOLO による物体検出を利用した人の除去処理を ICP-SLAM に応用した手法を提案し、人がいる屋内環境において人のデータを除去することで、精度の高い地図構築を実現した。

一方で、現在は距離画像センサが 1 台のみであるため、人を識別できる計測範囲が狭い問題がある。そのため、今後は複数台のセンサを導入することで計測範囲の問題に対処していく。

参考文献

- [1] 藤井祐介, 大矢晃久, 坪内孝司, “人混み中での移動ロボットの自己位置推定に有効な測域センサデータ処理法の提案”, 第 26 回ロボティクス・メカトロニクス講演会講演概要集 2014, 2014.
- [2] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779-788, 2016.
- [3] A. Dhall, K. Chelani, V. Radhakrishnan and K. M. Krishna, “LiDAR-camera Calibration Using 3D-3D Point Correspondences,” In arXiv preprint arXiv:1705.09785, 2017.
- [4] 伊藤太一, 森岡一幸, “Graph SLAM と局所的な占有格子地図を用いた移動ロボットの自律走行システムの開発,” ロボティクス・メカトロニクス講演会講演概要集 2013, 2013.
- [5] P. J. Besl and N. D. McKay, “A Method for Registration of 3-D Shapes.” In IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, pp. 239-256, 1992.