

# 視覚情報を利用した SLAM 地図生成時の人の影響の軽減

## Reduction of Errors Caused by People During making SLAM map Using Visual Object Detection

王 有為<sup>1</sup> 三河 正彦<sup>2</sup> 藤澤 誠<sup>3</sup> 柊 和佑<sup>4</sup>, 石橋 豊之<sup>5</sup>

Youwei Wang<sup>1</sup>, Masahiko Mikawa<sup>2</sup>, Makoto Fujisawa<sup>3</sup>, Wasuke Hiragi<sup>4</sup>, Toyoyuki Ishibashi<sup>5</sup>

<sup>1,2,3</sup> 筑波大学 図書館情報メディア研究科

<sup>1,2,3</sup> Graduate School of Library, Information and Media Studies, University of Tsukuba

<sup>4</sup> 中部大学 人文学部コミュニケーション学科

<sup>4</sup> Department of Communications, College of Humanities, Chubu University

<sup>5</sup> 稚内北星学園大学 情報メディア学部

<sup>5</sup> Faculty of Information Media, Wakkanai Hokusei Gakuen University

**Abstract:** SLAM ( Simultaneous Localization And Mapping) is one of the most important techniques for an autonomous mobile robot. And SLAM using a 2D LRF (two-dimensional Laser Range Finder) is simple and efficient. However, it is not easy to recognize several movable objects based on distance data of the 2D LRF, the objects could be included in the SLAM map as error data. In this paper, a new multiple sensor based strategy for 2D SLAM is proposed to solve this problem. This strategy uses a visual object detection to find out objects that should not included in a SLAM map, such as humans and cars surrounding the mobile robot. After calculating the directions of these objects, the distance from the LRF is calculated and the distance data of the detected objects is deleted. At last, the processed data is used to update the SLAM map. Some experimental results reveal the validity of out proposed method.

## 1 はじめに

移動ロボットの移動制御のためには、ロボットの周辺の地図と、その現在の自己位置が重要となる。地図の生成とロボットの自己位置推定を同時に行う SLAM (Simultaneous Localization And Mapping) が盛んに研究されている。現在 SLAM には、大きく分けて三種類がある。レーザレンジファインダー(LRF) やソナーなどの距離センサを用いる手法、単眼や両眼カメラなどの視覚センサを用いる手法、距離センサや視覚センサなどの複数のセンサを組み合わせ

用いる手法である。

距離センサ、特に LRF を用いる SLAM は様々な研究があり、自動運転車や掃除ロボットなどの商品へ応用もされている。

視覚センサを用いるビジュアル SLAM は画像から特徴点の抽出、マッチングが必要で、計算量が多く、課題も多い。しかし、パソコンの計算能力の向上とディープラーニングの発展とともに、リアルタイムで精度が良いビジュアル SLAM 技術も提案されつつある。例えば Pumarola らが提案する PL SLAM<sup>[1]</sup> は、リアルタイムで動作し、小さい室内環境では精度も良い。しかし大きい部屋などの広い空間やテクスチャがない物体の存在する環境では精度が悪くなる場合もある。

距離センサのみを用いる手法も、視覚センサのみを用いる手法も、人などの移動する可能性のある物体が地図に含まれるなど、環境から悪い影響を受ける場合がある。例えば、距離センサから得られる離散的な距離データだけでは、人と壁を区別することは簡単ではない。視覚センサを用いる手法は、真っ暗の環境や周囲のテクスチャがない時にはうまく動

<sup>1</sup> 筑波大学図書館情報メディア研究科 (〒305-0821 茨城県つくば市春日 1-2) s1865088@s.tsukuba.ac.jp

<sup>2</sup> 筑波大学図書館情報メディア研究科 (〒305-0821 茨城県つくば市春日 1-2) mikawa@slis.tsukuba.ac.jp

<sup>3</sup> 筑波大学図書館情報メディア研究科 (〒305-0821 茨城県つくば市春日 1-2) fujis@slis.tsukuba.ac.jp

<sup>4</sup> 中部大学人文学部 (〒487-8501 愛知県春日井市松本町 1200 番地) wasuke@isc.chubu.ac.jp

<sup>5</sup> 稚内北星学園大学情報メディア学部 (北海道稚内市若葉台 1-2290-28) toyoyuki@wakhok.ac.jp

作しない場合がある。SLAM の精度を上げるため、環境の影響を取り除くことが必要である。そこで本研究では、距離センサと視覚センサの組み合わせることにより、地図の作成時に環境からの影響を受けにくい手法を提案する。

## 2 関連研究と本研究の提案

LRF を用いた SLAM により地図を作成する際、人や車などの移動する可能性のある物体が地図の中に残ることがある。また自己位置推定を行う際、地図の中に残った移動物体が、位置推定に悪影響を与える場合もある。そこで、LRF を用いる SLAM において、環境からの悪影響を取り除く研究が提案されている。

このような移動物体の影響を取り除くため、Tang らは Likelihood Grid Voting (LGV) <sup>[2]</sup> という手法を提案し、特定の閾値を設定することにより、カウンタグリッドの値がこの閾値より小さい場合には移動物体、大きい場合には静的物体として判断することができる。Li らの研究 <sup>[3]</sup> では、バックグラウンド減算法 <sup>[4]</sup> と同じストラテジを用いてグリッドの尤度を更新することにより、動的な物体を計測したグリッドの尤度が、静的な物体を計測したグリッドの尤度より低くなることを利用し、動的な物体と静的な物体を区別することを可能とした。

これらは、LRF を用いる SLAM で得られる地図から、移動物体を取り除くことに対して有効的な手法だが、LRF が計測する範囲内で人や車などが停止している場合には、地図の中に影響が残るという問題がある。Philipp らの研究 <sup>[5]</sup> では、ディープラーニングによって 3D SLAM 地図から大部分の移動可能な物体を消すことができるが、完全に消すことではない。

LRF に基づき作成された SLAM 地図を用い、移動ロボットの自己位置を行う場合、これらの移動する可能性のある物体が地図中に含まれると、事故位置推定の精度に悪い影響を与える恐れがある。

そこで本研究では、SLAM による地図作成時に、視覚センサを用いた物体識別機能を利用することにより、移動や停止状態に関わらず人などの移動する可能性のある物体の影響を取り除く手法を提案する。

## 3 システム構成

### 3.1 システム概要

本研究は Robot Operating System (ROS) <sup>[6]</sup> を用い

てセンサのデータの取得及びロボットのコントロールを行う。図 1 にシステムのハードウェア構成を、図 2 にソフトウェア構成を示す。

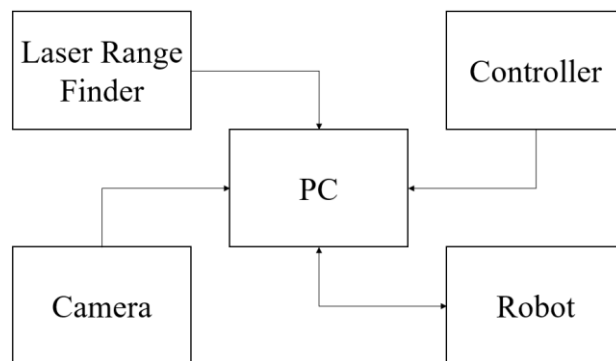


図 1: システムのハードウェア構成

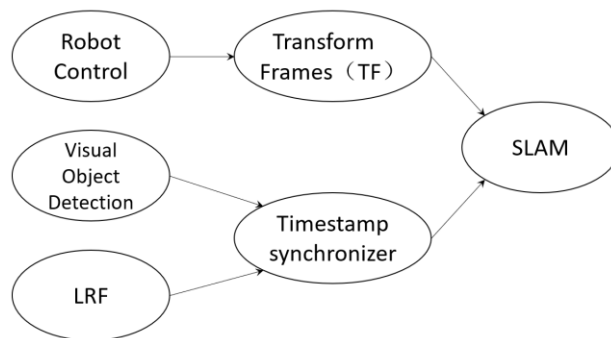


図 2: ソフトウェア構成

本研究の提案手法では、まずカメラを用いて、ロボットの周囲環境の画像を取り込みつつ、画像から周囲の人や車などの物体を検出して、対応した方向情報を算出する。そして、検出した物体の方向データ及びカメラと LRF の位置関係を用いて、LRF の距離データの中から人や車などの物体が含まれる範囲を計算して、人や車などの物体のデータを削除する。最後は、物体の距離データが削除された LRF の距離データに基づき、SLAM のアルゴリズムから地図を作成/更新する。

### 3.2 視覚検出システム

本研究では、機械学習に基づく物体識別手法を利用する。Histogram of Oriented Gradient (HOG) 特徴量に基づき Support Vector Machine (SVM) による学習識別手法 <sup>[7]</sup>、Convolutional Neural Network (CNN) による機械学習手法として LeNet <sup>[8]</sup>、AlexNet <sup>[9]</sup>、GoogleNet <sup>[10]</sup>、VGGNet <sup>[11]</sup>、ResNet <sup>[12]</sup> などのネット構造、Faster R-CNN <sup>[13]</sup>、YOLO <sup>[14]</sup>、SSD <sup>[15]</sup> などの識別アルゴリズムが提案されているが、ここでは計算速

度と識別率のバランスが良い YOLOv3<sup>[16]</sup>を用いることとする。

## 4 提案手法

### 4.1 対象物体の距離データの推定

図3は、LRFから得られる距離データのデータ構造である。この距離データの中から、人や車などの移動物体を取り除くためには、移動物体の両端を知る必要がある。ここでは距離データの中の対象物体の両端の取得について説明する。

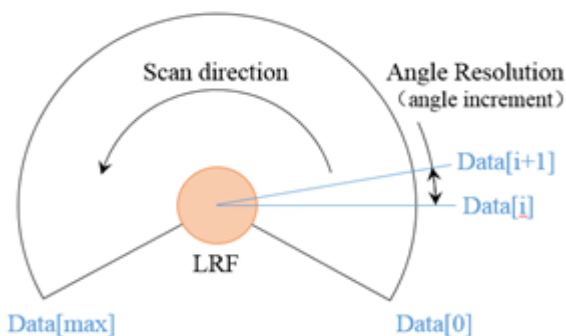


図3: LRF 距離データの構造である

まずは対象物体が占める画像中の角度を求める。YOLOにより物を識別した場合、図4に示すように、検出物体がその大きさに応じてバウンディングボックスが表示される。画面の左下の点を基準として、この領域は  $x_{max}$ ,  $x_{min}$ ,  $y_{max}$ ,  $y_{min}$  で表される。

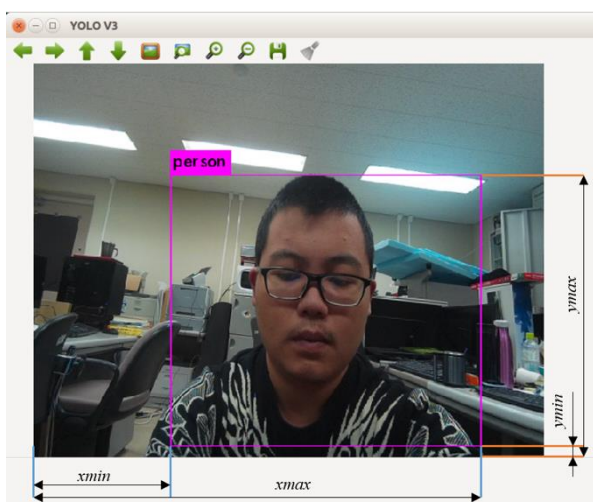


図4: 視覚認識

次に図5に、カメラの画像面に投影される対象物

体との関係を示す。ただし、上面図である。

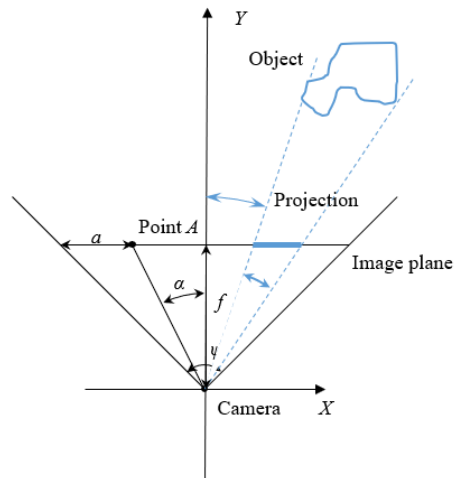


図5: 物体と画像面の投影の関係

式(1)は画面の横の解像度  $x$  とカメラの横の視野の角度  $\psi$  と画像平面からカメラまでの距離  $f$  の関係を表わす。ある点が画像平面上の点  $A$  に投影されたとすると、その  $x$  座標は  $a$  ピクセルとなるものとする。式(2)を用いて  $A$  とカメラの接続線と軸  $y$  の角度  $\alpha$  を算出する。

$$f = x / (2 \times \tan(\psi / 2)) \quad (1)$$

$$\alpha = \tan^{-1}((a - x/2) / f) \quad (2)$$

次に、識別した物体とカメラの位置関係と、カメラと LRF の位置関係を用いて、物体と LRF の位置関係を算出する。図6に、カメラと LRF、識別物体の位置関係を示す。軸  $y$  はカメラの光軸を表し、ここでは単純化のために LRF の原点は光軸上にあるものとする。LRF とカメラの距離を  $d$  とする。 $\alpha$  はカメラの光軸と識別物体の角度、 $\omega$  は LRF の正面と識別物体の角度、 $l$  は識別物体の距離の  $y$  軸への投影を表わす。式(3)に  $l$ ,  $d$ ,  $\alpha$ ,  $\omega$  の関係を表わす。式(4)によって、 $l$  が  $d$  より非常に大きい場合、 $\alpha$  と  $\omega$  の差は小さく、式(4)で近似される。カメラと LRF の位置関係  $d$  を短くすることにより、 $\omega$  を算出する代わりに  $\alpha$  で近似することが可能になる。

$$\tan(\omega) \times l = \tan(\alpha) \times (l + d) \quad (3)$$

$$\frac{\tan \omega}{\tan \alpha} = 1 + \lim_{l \rightarrow 0} \frac{d}{l} \approx 1 \quad (4)$$

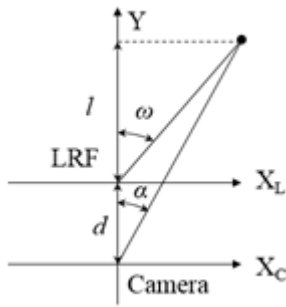


図 6: カメラと LRF と物体の位置関係

最後に、LRF で計測された距離データ中から、識別物体に関するデータ番号を取得する。ハードウェアのレイアウトでカメラの画像の中心方向の LRF の計測データの順番  $n$  を取得する。式(5)により、点 A の計測データ中の番号  $ma$  を計算する。 $r$  は LRF の計測の角度解像度を表わす。

$$ma = n + \alpha/r \quad (5)$$

ただし  $ma$  が整数ではない場合、確実に人などの物体のデータを消すために、端数処理を次のように行う。画像での横の範囲は  $xmin$  と  $xmax$  の物体が距離データの中の順番  $mmax$  と  $mmin$  の計算式は式(6)となる。ただし  $p$  は、LRF からの距離データの計測方向と視覚認識のバウンディングボックスの  $x$  軸の増加の方向で決めるパラメータである。両方の方向が同じ場合、 $p$  の値は 1 となり、両方の方向が違う場合、 $p$  の値は -1 になる。

$$\begin{cases} mmin = n + \left\lfloor \frac{1}{r} \tan^{-1} \left( \frac{xmin - x/2}{p \times f} \right) \right\rfloor \\ mmax = n + \left\lfloor \frac{1}{r} \tan^{-1} \left( \frac{xmax - x/2}{p \times f} \right) \right\rfloor \end{cases} \quad (6)$$

#### 4.2 複数カメラを用いる画像物体検出システム

LRF の計測角度の範囲は単眼カメラの視野角より広い。実用化のために、複数のカメラを使って、より広い範囲の映像を撮影する必要がある。ここでは図 7 に示すような複数の広角カメラを使用し、ロボットの全周の画像を取得するシステムを構築した。

#### 4.3 データの同期

LRF のデータの出力の周波数が固定している。しかし視覚認識のデータの出力の周波数は常に変化するので、両方のデータの同期が必要である。なお、データのサンプリング周波数が異なるので、完全一致のタイムスタンプを持つデータを生成することは

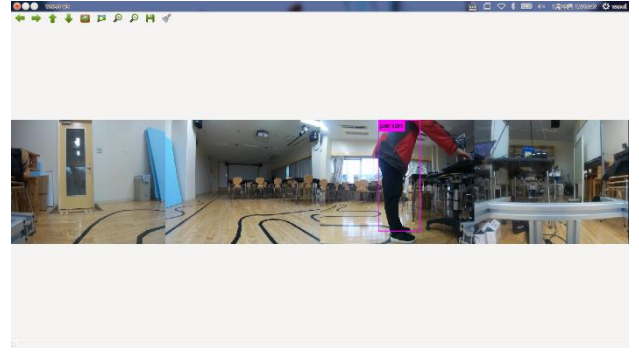


図 7: 複数カメラにロボットの全周画像から物体を検出

困難なので、ここでは、周波数が低い方を基準として、一番近い時間の周波数が高いデータを用いることとする。

## 5 実験結果

### 5.1 実験環境

本研究で実験に使用した移動ロボットシステムを図 8 に示す。移動ロボットの本体には Pioneer 3-DX (Adept Technology, Inc.) を、LRF には UTM-30LX (北陽電機(株)) を、カメラには SJ5000 (SJCAM) を使用した。LRF の計測範囲は 0.1~30 [m]、計測範囲は 270 [deg]、計測点数は 1081 [点]、計測周波数は 40 [Hz] である。LRF の計測角度により、カメラは 3 つで LRF の計測範囲内の画像を取得できる。システムを制御するコンピュータは、OS に Ubuntu16.04LTS、移動ロボット制御ライブラリに ROS-Kinetic を使用し、制御用 PC には、CPU に Intel Core i7-4702MQ、GPU に GTX850M (Compute Capability 5.0) が搭載される。

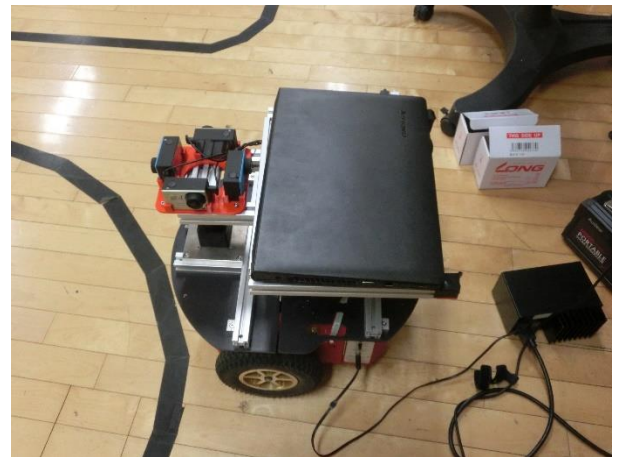


図 8: 移動ロボット



## 5.2 実験結果

図9に、本研究で提案するSLAMが動作している様子を示す。青い点は、視覚データに基づき人と識別された範囲に存在するLRFの距離データを示す。赤い点は、LRFの距離データから人と識別された距離データを削除した距離データを意味する。黄色の矢印は、移動ロボットの位置と姿勢を表す。白い点線の円は、人が存在すると推定した位置を表す。YOLOv3に基づく人検出は15 [fps]程度で実行され、人の影響を受けることなく、SLAMによる地図生成と移動ロボットの自己位置推定が行えることが確認できた。

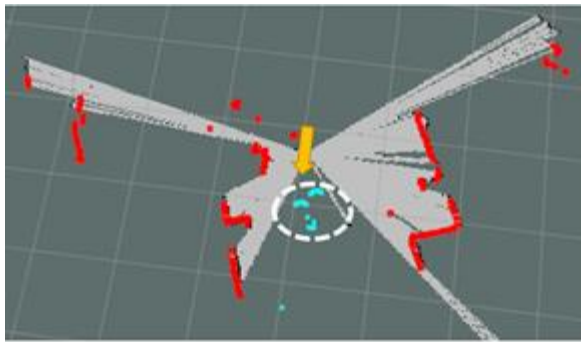


図9: 提案手法

次に、3種類のSLAMアルゴリズムに基づいて提案手法の有効性を確かめた。

図10の左図に示すように、提案手法に基づき画像から人を検出、LRFの距離データから人の関連する距離データを削除したが、環境に影響されやすいicp-slamでは、人のデータが一部残ってしまっている。しかし、図10の右図に示すように、提案手法を用いずicp-slamを行なった場合、人のデータが地図データ中に大きく残り、本手法が有効であることが分かる。本手法を用いず検出された人の計測データに比べて、本手法を用いることにより地図中に残った人の計測データの数は93.6%少ない。

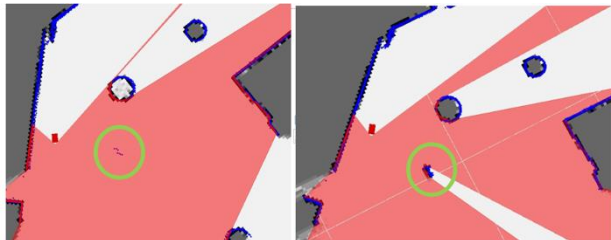


図10: icp-slamで提案手法を用いた場合(左図)と提案手法を用いない場合(右図)

gmapping-slamでは、本手法を用いない場合でも、同じ場所を何度か計測することにより、人などの移動物体を地図中から自動的に削除することが可能であるが、本手法を用いることにより、人の一度だけ通った場合であっても、人を地図中から削除することが可能である。図11左図に、提案手法を用い作成されたSLAM地図を示す。緑の円内に人が立っていたのだが、地図中に残っていないことが分かる。図11右図に、提案手法を用いず作成したSLAM地図を示す。赤い円内に人が立っていたため、地図としてデータが残っていることが分かる。gmapping-slamでは、人が居なくなれば、この場所を何度か移動ロボットが通過すれば、地図中からこの人のデータは自動的に削除されるが、本手法のように、一度通過してだけで正しい地図を作成できる方が良いものと思われる。

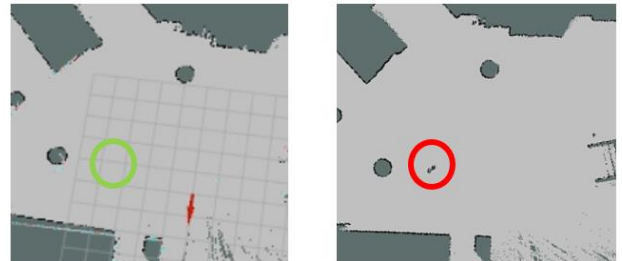


図11: gmapping-slamで提案手法により人の影響を削除できた場合(左図)と提案手法を用いない場合(右図)

rbpf-slamも同様に、提案手法を用いることにより、一度通過するだけの場合でも、地図中に人が出現することは無い。図12右図に本手法に基づき作成したSLAM地図を、図13左図に本手法を用いず作成したSLAM地図を示す。

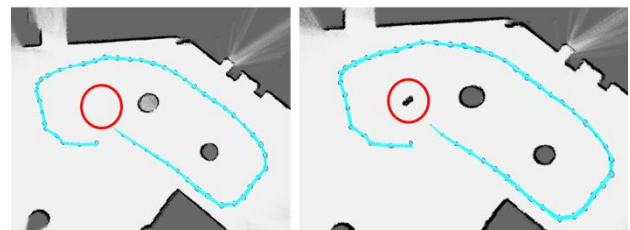


図12: rbpf-slamで提案手法を用いた場合(左図)と用いない場合(右図)

gmapping-slamとrbpf-slamでは、移動ロボットが同じ場所を何度か通過することにより、人が居なくなれば地図中から人の影響を削除する事ができる。しかし、新たに人が出現した場合など、図13に示す

ように、生成した地図中に新たに人が表示される場合もある。これに対して、提案手法は、生成済みの SLAM 地図に人が表示されることは無い。

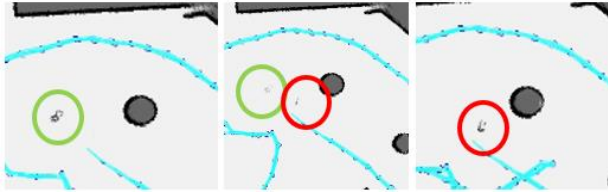


図 13: 地図の更新により、人が新たに出現すると地図中にも人の位置データが出現

## 6 むすび

人や車などの移動する可能性のある物体が、LRFに基づく SLAM で生成した地図の中に残るという問題を解決するため、本研究は視覚に基づく物体検出を用いて人などの物体の方向データを取得し、LRFで計測される距離データから検出物体の距離データを削除し SLAM を行う手法を提案した。3種類の SLAM の実験により、本手法の有効性を確認した。

## 謝辞

本研究の一部は科研費(課題番号 JP15K00358, JP18K04041), 筑波大学人工知能科学センターにおける共同研究の助成を受けたものである。

## 参考文献

- [1] Pumarola, Albert, et al. "PL-SLAM: Real-time monocular visual SLAM with points and lines." IEEE International Conference on Robotics and Automation IEEE, pp.4503-4508, (2017).
- [2] Tang, Jian, et al. "An approach of dynamic object removing for indoor mapping based on UGV SLAM." Sensors & Transducers Vol.190.No.7, pp.40-46, (2015).
- [3] Li, Li, et al. "Laser-based SLAM with efficient occupancy likelihood map learning for dynamic indoor scenes." ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol.3, No.4, pp.119-126, (2016).
- [4] Yao, Jian, and Jean-Marc Odobez. "Multi-layer background subtraction based on color and texture." Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, pp.1-8, (2007)
- [5] Ruchti P, Burgard W. Mapping with Dynamic-Object Probabilities Calculated from Single 3D Range Scans[C]//2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp.6331-6336, (2018).
- [6] Quigley, Morgan, et al. "ROS: an open-source Robot Operating System." ICRA workshop on open source software. Vol. 3. No. 3.2.p.5. (2009).
- [7] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1, pp.886-893, (2005).
- [8] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE Vol.86.No.11, pp. 2278-2324 (1998).
- [9] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. pp.1097-1105, (2012).
- [10] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. pp.1-9, (2015).
- [11] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556, (2014).
- [12] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. pp.770-778, (2016).
- [13] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. pp.91-99, (2015).
- [14] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. pp.779-78, (2016)..
- [15] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, pp.21-37, (2016).
- [16] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).