

ARのための複数自由移動剛体の三次元再構成

小澤 岳大^{*1} 中島 由勝^{*1} 斎藤 英雄^{*1}

Abstract — 拡張現実感において、シーンの三次元再構成及び自己位置姿勢推定 (SLAM) は重要な要素であるが、従来の SLAM の手法の多くは静的なシーンを対象としており、シーン中を自由に移動する複数の物体が存在する場合に対応できないという問題点がある。そこで、本稿では複数の移動剛体が存在するシーンにおける SLAM に基づき、複数の自由移動剛体の三次元再構成を行う手法を提案する。提案手法は、幾何形状に基づいてシーンをオブジェクトに領域分割し、それぞれのオブジェクトごとに三次元再構成及び位置姿勢推定を行う。評価実験では、複数の移動剛体が存在するシーンにおいても、三次元再構成や自己位置姿勢推定が可能であることを確認し、今後の展望について述べる。

Keywords : 三次元再構成, 位置姿勢推定, 拡張現実感

1 はじめに

自己位置姿勢・環境地図の同時推定技術 (Simultaneous Localization and Mapping, SLAM) は、マーカーや三次元モデル等の、事前情報のない状況における拡張現実感 (Augmented Reality, AR) において重要である。特に、RGB-D センサの発達により、環境を密に三次元復元可能である、密な SLAM の研究が盛んに行われている [5, 9]。近年では、スマートフォン上で動く、密な SLAM を用いた AR アプリケーション等も登場している。

しかし、従来の密な SLAM の多くは静的なシーンを対象としており、カメラだけでなく、複数の物体が自由に移動するシーンを対象とした SLAM は技術的課題が多い。

Newcombe らは、Depth センサから得られる非剛体の三次元点群を統合し、動的なシーンにおける三次元再構成を実現している [4]。一方この手法は、動的な領域を一つの物体として三次元再構成するため、物体ごとの 3D モデルの取得や位置姿勢の推定を行うことはできない。

Rünz らは、物体の運動に注目し、背景と異なる運動をするを領域分割し、個別に位置姿勢推定を行うことで、動的シーンにおける自己位置姿勢推定と三次元再構成を実現している [6]。一方この手法では、個別の物体が同一の運動をしている際、それらを同一物体として分割してしまうという問題点がある。

上記手法の問題点に基づき、本研究では以下の利点を持つ SLAM の手法を提案する。

- 複数の物体が自由移動する環境下で、それぞれの物体の三次元再構成及び位置姿勢推定を行う

- 幾何形状に基づいた領域分割を用い、物体の運動に影響を受けない

また、実験によりその有効性を確認する。

なお、Rünz らの Maskfusion[7] は、幾何形状に基づいた領域分割と Mask R-CNN[1] を用いた領域分割の統合による、動的シーンにおける SLAM を実現しているが、本手法は物体認識を出力として考慮しておらず、目的が異なる。

2 提案手法

図 1 に提案するシステムの流れを示す。本手法は Segmentation Part, Labeling Part, SLAM Part により構成される。入力 is 自由移動する RGB-D センサより得られる RGB 画像, Depth 画像であり, 1 フレームずつ図 1 に示した処理を行う。

2.1 Segmentation Part

本手法では、幾何形状に基づいたオブジェクト単位の領域分割を行い、それぞれが個別に三次元マップをもつことで、物体の運動によらない三次元再構成、自己位置姿勢推定を可能にしている。

オブジェクトは凸型である、という仮定のもとで、凹部分や非連続部分をエッジとした領域分割を高速に行うことが可能である。本手法では Tateno らの手法 [8] に基づき、現フレームのデプス画像を、その頂点、法線情報を利用して領域分割を行い、エッジマップを得る。

2.2 Labeling Part

得られたエッジマップに対し、連結部分のラベリング処理を行い、各画素にラベルが振られたラベルマップを得る。前フレームと現フレームのラベルを対応させるため、前フレームで得られたラベルマップ \mathcal{L}_{t-1} と、現フレームで得られたラベルマップ \mathcal{L}_t を比較し

^{*1}慶應義塾大学大学院 理工学研究科

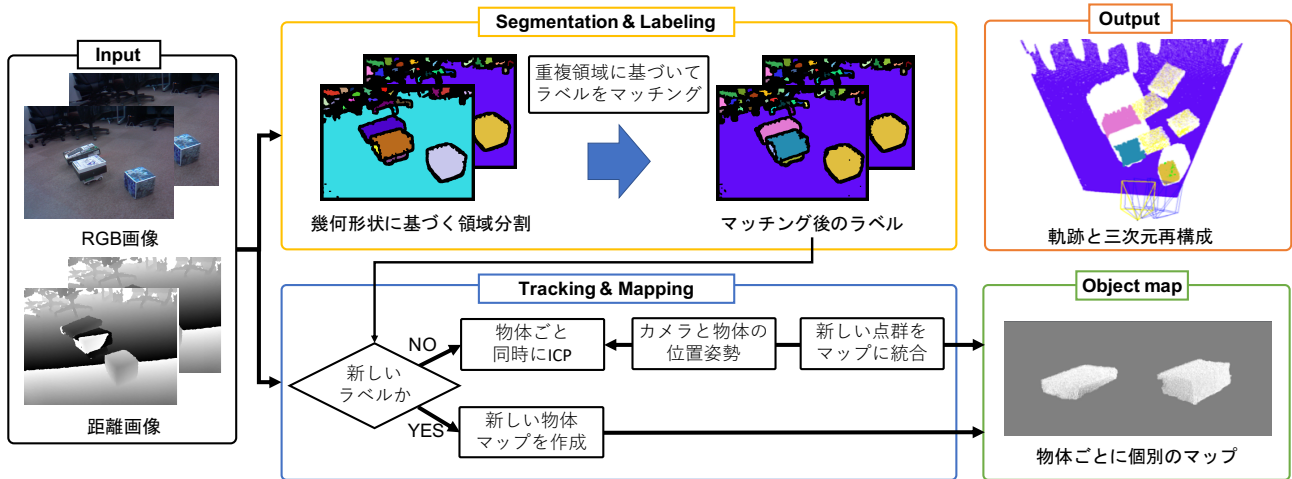


図1 提案手法の流れ
Fig. 1 Flow of our method

て、その対応関係から、現フレームの新しいラベルマップ \mathcal{L}_t^p を作成する。

\mathcal{L}_{t-1} と \mathcal{L}_t を重ねたとき、ラベルが $l_i \in \mathcal{L}_{t-1}$ かつ $l_j \in \mathcal{L}_t$ である画素の数を $S_{\cap}(l_i, l_j)$ 、 $l_i \in \mathcal{L}_{t-1}$ または $l_i \in \mathcal{L}_t$ である画素の数を $S_{\cup}(l_i, l_j)$ とすると、 l_i と l_j がオーバーラップしている割合は、

$$\tilde{S}(l_i, l_j) = \frac{S_{\cap}(l_i, l_j)}{S_{\cup}(l_i, l_j)} \quad (1)$$

である。全ての l_i, l_j の組み合わせについてこの計算を行うと、 l_j と最もオーバーラップしているラベルとその割合 $\tilde{S}_{max}(l_j)$ を算出でき、

$$\tilde{S}_{max}(l_j) = \max_{l_i \in \mathcal{L}_{t-1}} \{\tilde{S}(l_i, l_j)\} \quad (2)$$

と表される。この値に基づいて、 $l \in \mathcal{L}_t^p$ のラベルを以下のように決める。

$$l = \begin{cases} l_i & \tilde{S}_{max}(l_j) > 0.2 \\ l_j & otherwise \end{cases} \quad (3)$$

2.3 SLAM Part

本手法の SLAM Part は Keller らの手法 [2] に基づいている。Keller らの手法は単一のマップを用いて SLAM を行っているが、本手法はオブジェクトごとに持つマップを用いて SLAM を行っている。このため、シーン中に移動する物体が複数存在しても、三次元再構成やセンサの自己位置姿勢推定に影響を及ぼすことはない。

SLAM Part では、まず現フレームにおけるセンサの自己位置姿勢推定、領域分割されたオブジェクトの位置姿勢推定を行い、その推定結果からオブジェクトのマップを更新する。

2.3.1 センサの自己位置姿勢推定

床や壁等の静的なオブジェクトに対し、現フレームの三次元点群と、対応する三次元マップとの間で Point-to-Plane の ICP アルゴリズム [3] を行う。これにより現フレームにおけるセンサの自己位置姿勢 \mathbf{T}_c を推定する。

2.3.2 オブジェクトの位置姿勢推定

オブジェクトのラベルが既知である場合、現フレームのオブジェクトの三次元点群と、対応するオブジェクトマップとの間で Point-to-Plane の ICP アルゴリズムを行う。ここで得られる位置姿勢を $\tilde{\mathbf{T}}_o$ とすると、センサの移動を考慮したオブジェクトの位置姿勢 \mathbf{T}_o は

$$\mathbf{T}_o = \mathbf{T}_c \tilde{\mathbf{T}}_o \quad (4)$$

と計算される。

オブジェクトのラベルが未知である場合、そのオブジェクトは新しいオブジェクトとみなし、新たにオブジェクトマップを作成する。

2.3.3 オブジェクトマップの更新

推定されたオブジェクトの位置姿勢 \mathbf{T}_o に基づき、現フレームのデプス画像から得られる三次元点群を、それぞれのオブジェクトマップを構成する点群に統合することで、オブジェクトマップを更新する。

3 評価実験

本章では、本手法の有効性を示すために行った実験について示す。図2上段のように、複数の剛体が自由移動し、かつ RGB-D センサも自由移動するシーンに対して本手法を適用し、剛体の三次元再構成と位置姿勢、及び RGB-D センサの位置姿勢を推定した。RGB-D センサには Kinect v1 を使用した。

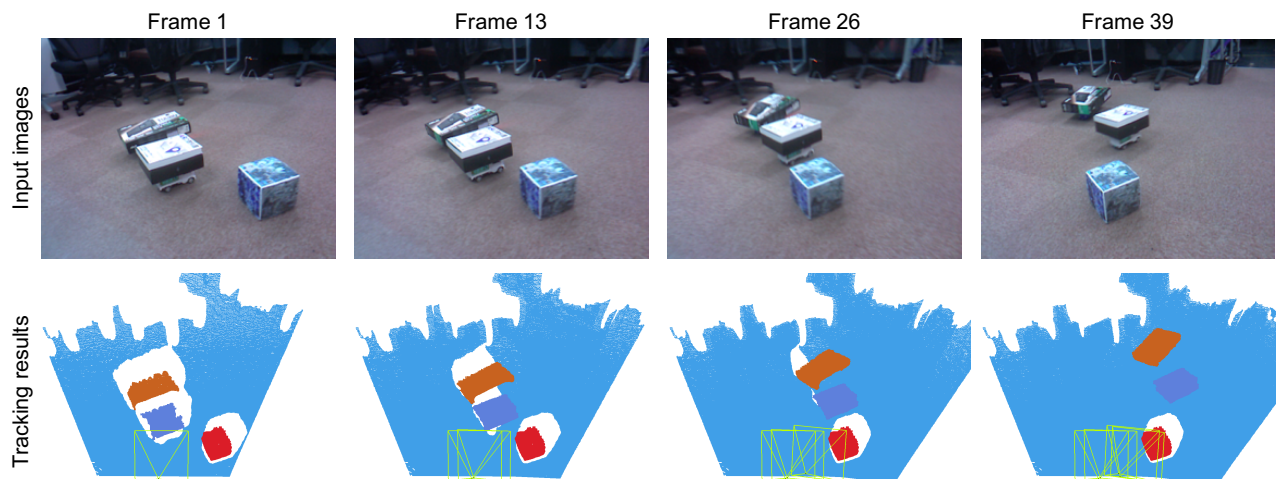


図2 入力画像と位置姿勢推定結果
Fig.2 Input images and tracking results

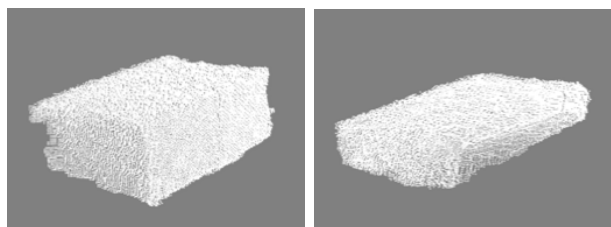


図3 三次元再構成結果
Fig.3 Reconstruction results

図2下段に剛体の位置姿勢及びセンサの自己位置姿勢の推定結果を示した。図2を見ると、自由に移動する二つの剛体の位置姿勢が個別に推定できていることがわかる。また、RGB-D センサの自己位置姿勢は移動剛体からの影響を受けずに推定できている。

また、実験結果から最終的に得られた、オブジェクトの三次元再構成結果を図3に示す。図3から、剛体の形状が個別に再構成されており、物体の三次元再構成が可能であることを示唆している。

4 結論

本稿では、幾何形状に基づいたオブジェクトごとの領域分割を利用した、複数の移動剛体が存在するシーンにおける SLAM の手法を提案した。そして、実際に複数の剛体が移動するシーンにおける実験により、三次元再構成と自己位置姿勢推定が可能であることを示した。

一方で、幾何形状に基づいた領域分割では、1つの物体が複数の分割されてしまい、複雑な物体はオブジェクト単位で正しく三次元再構成、位置姿勢推定ができないことがある。Maskfusion[2]では、その問題点を

Mask R-CNN[1]によるセマンティックなマスクを用いることで補っている。本手法もその問題点を解決する必要がある、今後の課題となっている。

謝辞

本研究は、AIP-PRISM, Japan Science and Technology Agency (JPMJCR18Y2) の支援を受けたものである。

参考文献

- [1] K. He, G. Gkioxari, P. Dollár, and R. Girshick: Mask r-cnn; ICCV, 2017
- [2] M. Keller, D. Lefloch, M. Lambers, S. Izadi, T. Weyrich, and A. Kolb: Real-time 3d reconstruction in dynamic scenes using point-based fusion; 3DV, 2013
- [3] K. Low: Linear least-squares optimization for point-to-plane icp surface registration; Technical report, University of North Carolina, 2004
- [4] R. A. Newcombe, D. Fox, and S. M. Seitz: Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time; CVPR, 2015
- [5] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. W. Fitzgibbon: KinectFusion: Real-time dense surface mapping and tracking; ISMAR, 2011
- [6] M. Rünz and L. Agapito: Co-fusion: Real-time segmentation, tracking and fusion of multiple objects; ICRA, 2017
- [7] M. Rünz and L. Agapito: Maskfusion: Real-time recognition, tracking and reconstruction of multiple moving objects; ISMAR, 2018
- [8] K. Tateno, F. Tombari, and N. Navab: Real-time and scalable incremental segmentation on dense slam; IROS, 2015
- [9] T. Whelan, S. Leutenegger, R. F. Salas-Moreno, B. Glocker, and A. J. Davison: ElasticFusion: Dense SLAM without a pose graph; RSS, 2015