

単眼カメラを用いた三次元走行環境の再構成

日大生産工 (院) ○千葉 橋平 日大生産工 風間 恵介
日大生産工 丸茂 喜高

1. 緒言

Simultaneous Localization And Mapping (SLAM) は、三次元の環境地図を作成し、走行環境内で自己位置推定することを目的としている。ロボット制御の一環として提案された手法だが、近年では拡張現実や自動運転などへの応用が進んでいる。LiDAR, Radar など種々のセンサの測定結果が利用されるが、高価であるうえ、色相の認識ができず、人間が想像する三次元環境地図としては不完全である。そのため、より安価なカメラを用いた Visual SLAM の研究が主流になっている。課題は、姿勢の変化によるフレーム間での移動量が大きくなると、特徴量の消失、最適化アルゴリズムの発散、推定誤差の蓄積などにより軌跡の推定に誤差が生じることである。移動速度が速く、姿勢変化の激しい自動車では、実用性に乏しい。

これらを解決する手法として、深層学習の使用が提案されている。既存の研究では、幾何学的に算出された特徴量を学習済みの特徴量に置換し、三次元表現に使用することなどが提案されてきた。他の研究では、SLAM と Visual Odometry から学習することが試みられている。本研究では、従来の手法である ORB-SLAM2¹⁾ と、深層学習と SLAM 技術を組み合わせた DROID-SLAM²⁾ に対象を絞り、姿勢変化の激しい自動車の前方映像を用いて、軌跡推定の精度と、環境地図の再構成について評価を行う。

2. Visual SLAMの概要

従来の Visual SLAM は、自己位置推定と環境地図作成を同時に最適化する。単眼、ステレオ、RGB-D などの様々なカメラ形態に対応しており、一般的に間接法または直接法に分類できる。間接法は、追跡しやすいコーナー点などを特徴量として複数検出し、特徴記述子を付加することで、特徴量が画像間でマッチングされる。投影された三次元点群と画像内での幾何学的な距離を最小化することにより、カメラの姿勢と度値の誤差を最小化する手法をとっている。

直接法の利点は、間接法では使用されない線や長さの変化など、画像に関するより多くの情報を定量化できることである。しかし、画像間で測光誤差の最適化に課題があり、間接法においても、幾何学的歪みに対するロバスト性が低いことが挙げられる。これらの最適化課題においては、さまざまな手法が提案されてきた。このうち、ORB-SLAM2 は画像内の特徴量を利用し自己位置推定と環境地図の作成を行う間接法である。これは CPU 上のみで実行できる手法であるが、低遅延で点群を取得でき、かつカメラ姿勢が正確なことから広く利用されている。Fig. 1 に ORB 特徴量の検出と軌跡推定の様子を示す。この特徴量の検出数が少ない場合、出力された点群も疎なものとなるが、カメラの動きの変化に強く推定軌跡は正確に取得できる利点がある。正確に軌跡を推定することが出来るが、三次元環境地図の描画は認識した ORB 特徴量のみにも留まっている。

DROID-SLAM は、間接法と直接法を融合した手法をとる。カメラ姿勢と深度の推定をニューラルネットワークで実現するのが特徴であり、リカレントニューラルネットワーク (RNN) で、カメラ姿勢と深度を逐次的に更新する。ニュートンラフソン法などの一般的な最適化アルゴリズムのような値の更新を繰り返すことで最適な値に収束する手法ではなく、RNN の繰り返し構造で振る舞いを模倣している。推論が進むに従ってカメラ姿勢と深度を逐次更新し、適当な回数を繰り返した後、正確な値に収束する。これには Visual SLAM のために構築されたデータセットである TartanAir³⁾ を用いて訓練されている。正解データを用いた教師あり学習を行い、Fig. 2 に示すように、Visual SLAM としての能力を獲得したものである。Fig. 1 と比較して、周囲の環境をより鮮明に描画している。

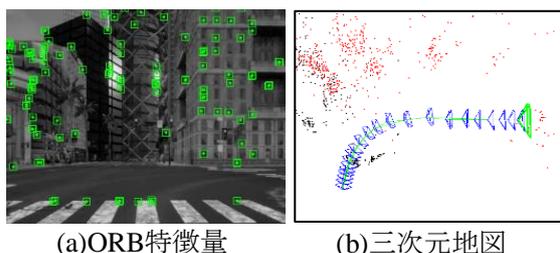


Fig. 1 ORB-SLAM2 の特徴量抽出と地図描画

Reconstruction of 3D Driving Environment Using a Monocular Camera
Kippeï CHIBA, Keisuke KAZAMA and Yoshitaka MARUMO



Fig. 2 DROID-SLAMの地図描画

3. シミュレーション

実際の車両前方映像としてドライブレコーダ映像で軌跡推定を行うと、真値が分からずその性能を検証することができない。そこで、自動運転シミュレータであるCARLA⁴⁾を使用し、疑似的なドライブレコーダ環境で評価する。15fps, 640×480pixelで映像データと、真値となる軌跡データを作成した。緩やかなカーブを走行したときの、ORB-SLAM2とDROID-SLAMでの軌跡推定精度を比較する。このとき、DROID-SLAMの再学習は行っていない。一般的な乗用車で、進行速度は約20km/hとした。単眼のVisual SLAMにはスケール不定性の課題があるため、軌跡推定の倍率が一律に定まらない。そのため、二点パターン間の変換パラメータの最小二乗法による推定⁵⁾を用いた行列変換をした。それぞれ数回軌跡推定を行い平均を取る。評価には、最大誤差 (Max), 平均誤差 (Mean), 二乗平均平方根誤差(RMSE)を算出し比較した。

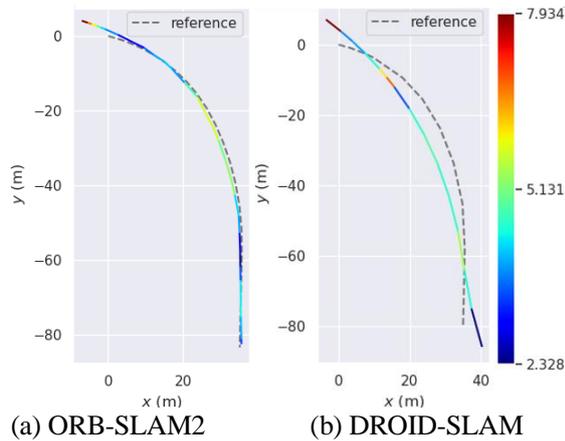


Fig. 3 軌跡推定の比較

Table 1 軌跡推定の比較結果

	ORB-SLAM2	DROID-SLAM
Max[m]	8.133	7.934
Mean[m]	3.275	4.604
RMSE[m]	3.663	5.242

4. 結果

ORB-SLAM2, DROID-SLAMの軌跡をFig.3に示す。また、そのときの各値をTable 1に示す。各数値で比較した結果、ORB-SLAM2の精度が優れていた。DROID-SLAMは描く曲線を真値よりも緩やかに捉えており、姿勢変化にて誤差が生じていた。これは、再学習をすることで精度の向上が見込める。また、ORB-SLAM2は何度か特徴量を見失い、軌跡推定に失敗することがあった。

5. 結語

本研究では、車載カメラから得られる前方画像を用いて自車両の三次元走行環境の再構成を目的とした。ORB-SLAM2と、DROID-SLAMとそれぞれ違う手法から自己位置推定を行い、姿勢変化の激しい自動車での精度を確認した。結果から、パラメータの最適化と一定の条件下では、深層学習による手法以上の精度が確認できた。今後は深層学習に着目し、より学習量を増やすことで自己位置推定の精度向上と、周囲の障害物との位置関係をより鮮明かつ直感的に理解することができる三次元走行環境での地図表現を検討する。

参考文献

- 1) R. Mur-Artal and J. Domingo, "ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras", IEEE Transactions on Robotics, Vol.33, No.5, (2017), pp.1255-1262.
- 2) T. Zachary and J. Deng, "Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras", Advances in neural information processing systems, Vol.34, (2021), pp.16558-16569.
- 3) W. Wenshan, et al., "TartanAir: A dataset to push the limits of visual slam", International Intelligent Robots and Systems (IROS), IEEE, (2020), pp.4909-4916.
- 4) D. Alexey, et al., "CARLA: An open urban driving simulator.", robot learning, PMLR, (2017), pp. 1-16.
- 5) S.Umeyama, "Least-squares estimation of transformation parameters between two point patterns", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.13, No.4, (1991), pp. 376-380.