

1. はじめに

Structure from Motion(SfM) は 2 枚の画像間の変化からシーンの 3 次元構造, カメラの位置, 姿勢を推定する技術である. 3 次元構造, 姿勢を推定するための手法として, デプスとカメラモーシヨンを推定する Depth and Motion Network(DeMoN)[1] が提案されている. DeMoN はデプスとカメラモーシヨンを出力するため, 自動運転における周辺環境の認識への応用が期待されている. 本研究では, 時系列データを扱うネットワークである Convolutional LSTM を DeMoN に導入し, 高精度化を目指す.

2. Depth and Motion Network

Depth and Motion Network(DeMoN) は, 3 つのネットワークで構成されており, 入力として t および $t+1$ フレームの画像をペア画像として用いる. ペア画像の視差から, 出力としてデプスとカメラモーシヨンが同時に得られる. DeMoN のネットワーク構造を図 1 に示す.

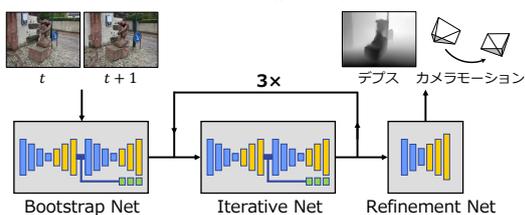


図 1: DeMoN のネットワーク構造

Bootstrap Net 入力したペア画像から, オプティカルフローとその信頼度を算出する. 算出結果とペア画像を用いて, おおまかなデプス, 法線, カメラモーシヨンを推定する.

Iterative Net Bootstrap Net で推定したデプスと法線, カメラモーシヨンを入力として, 新たにデプスと法線, カメラモーシヨンを改善して出力する. Iterative Net は, Bootstrap Net と同様の構造をしているが, 入力にはペア画像とともに以前のネットワークの推定値を用いる.

Refinement Net Iterative Net で推定したデプスマップのサイズを拡大する. そして, 拡大したデプスマップと, ペア画像の 1 枚目を入力することで, 高精細化したデプスマップを出力する.

3. 提案手法

DeMoN の Bootstrap Net に, 時系列データを扱う Convolutional LSTM を導入し, 連続的な変化量を考慮した推定手法を提案する.

3.1. Convolutional LSTM の導入

Convolutional LSTM(ConvLSTM)[2] は画像を入力とし, 全結合層を畳み込み層に変更した Long Short Term Memory(LSTM) である. 画像の特徴を捉える畳み込み処理と, 時系列データを扱う LSTM の機能を併せ持つ. 本手法では Bootstrap Net の畳み込み層の 3 層目と 4 層目の間に ConvLSTM を導入する. ネットワーク構造を図 2 に示す.

3.2. 活性化関数の変更

ConvLSTM の活性化関数を Tanh 関数から Leaky ReLU 関数に変更することで, 勾配消失問題を解決する. しかし Leaky ReLU 関数は不安定であるため, レイヤー正規化を行うことで学習の安定性を向上させる.

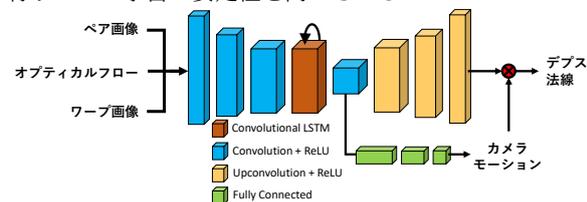


図 2: Convolutional LSTM を導入した Bootstrap Net

4. 評価実験

自動運転シミュレータの CARLA で作成したデータセットを用いて, 提案手法と従来手法の比較実験を行う.

4.1. 実験概要

評価はデプスとカメラモーシヨンの推定結果を比較する. デプスの評価指標には L1-relative と L1-invariant を用いる. L1-relative はデプスの相対誤差, L1-invariant はデプスの逆数の絶対誤差である. カメラモーシヨンの評価には, 回転と並進の推定誤差を用いる.

CARLA で作成したデータセットには, RGB 画像, Raw データから算出したデプス, オイラー角と座標から算出したカメラモーシヨンが含まれる. 学習用には 7810 ペア, 評価用には 2380 ペアのデータを用いる.

4.2. 実験結果

デプスとカメラモーシヨン推定の実験結果を表 1 に示す. 提案手法で推定したデプスは, L1-invariant では推定精度に変化は見られなかったが, L1-relative では誤差を 0.028 削減した. 一方で, カメラモーシヨンは, 回転と並進ともに推定精度に変化は見られなかった.

表 1: 誤差の比較

	デプス		カメラモーシヨン	
	L1-rel	L1-inv	回転	並進
DeMoN	1.121	0.004	0.018	0.337
提案手法	1.093	0.004	0.018	0.337

DeMoN と提案手法のデプスマップの比較結果を図 3 に示す. 提案手法は, DeMoN と同様のデプスマップとなっている. しかし, 電柱の下部を見ると, 従来の DeMoN では電柱を推定できていないが, 提案手法では電柱の一部を推定できていることがわかる. ConvLSTM の時系列を考慮した滑らかな推定により, デプスの推定精度が向上したと考えられる.

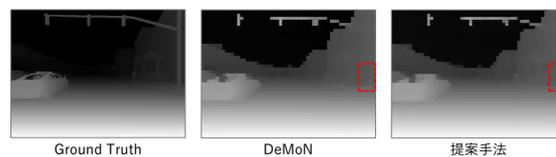


図 3: デプスマップの比較

時系列を考慮した推定が顕著に表れている誤差の変化を, 図 4 に示す. 従来手法は誤差が急に大きくなるフレームがあるが, 提案手法は, 誤差の変化が滑らかなことがわかる.

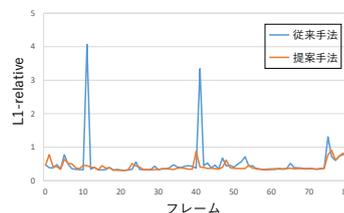


図 4: 推定誤差の変化

5. おわりに

本研究では, ConvLSTM を DeMoN に導入し, 提案手法と従来手法の推定値を比較する評価実験を行った. デプスの推定精度が向上し, 一部の推定できなかった箇所が提案手法では推定できるようになった. 今後は, カメラモーシヨンの推定精度も向上するネットワークの考案を目指す. 参考文献

- [1] B. Ummenhofer, *et al.*, “DeMoN: Depth and Motion Network”, CVPR, 2017.
- [2] S. Xingjian, *et al.*, “Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting”, NIPS, 2015.