

ロボット遠隔操作のための

複数台搭載魚眼カメラを用いた周囲環境の3次元再構成

小松 廉

東京大学大学院工学系研究科精密工学専攻

1. はじめに

ロボット遠隔操作の際にはオペレータは遠隔地にいるため周囲環境を直接見ることはできない。したがって、ロボットに取り付けられた第一人称視点の映像からではロボット周囲環境の状況把握が非常に困難である。そこで、ロボットに取り付けられた複数台の魚眼カメラを用いて、ロボット周囲環境の3次元再構成を行うことで、オペレータが周囲環境を把握することが容易になる。本研究では、屋内環境を対象とした複数の魚眼映像から周囲360度の密な深度推定手法を提案する。特に複数視点映像からの深度推定手法として、学習ベースのPlane-sweeping stereoに注目する。実験においてカメラ姿勢の異なる複数のデータセットを用いて評価することで、従来手法は学習時のカメラ配置と大きく異なると深度推定精度が大きく低下するのに対して、提案手法においてはカメラ配置が大きく変化しても高精度な深度推定が可能であることを示した。また、提案手法は計算効率が非常に高く、GPUを搭載したノートPCを用いて、4枚の魚眼映像から0.42 sで全方位深度推定が可能であった。ソースコードはGithub¹で公開している。

本研究の成果は、2020 IEEE/RSJ International Conference on Intelligent Robots and Systemsに採択されたものである [1]。詳細は[1]を参照することとし、本稿では概要を述べる。

2. 提案手法

深層学習の発展により、通常の透視投影映像にConvolutional neural network (ConvNet) を適用することで、Plane-sweeping stereoを用いて高精度の深度推定が可能になった。しかし、全方位映像や魚眼映像は映像中に大きな歪みが存在するため、普通のConvNetの適用がうまく行かないことがある。例えば、OmniMVS [2]においては魚眼カメラ映像に直接ConvNetを適用したため、学習時のカメラ配置から姿勢変化した際に性能が著しく低下する。したがって、本研究では全方位映像や魚眼映像のために特別に設計された二十面体 (Icosahedron) ベースの新しい表現方法とConvNetsを提案する。この表現方法は折り紙で作った王冠に似ていることからそのConvNetに「CrownConv」と名付けた。CrownConvは、魚眼映像および全方位映像の一般的な表現形式である正距円筒映像へ適用することができ、効果的な特徴量抽出が可能である。また、CrownConvは従来研究であるHexConv [3]の計算効率のボトルネックを解消することで、非常に高速な計算が可能である。

さらに、抽出された特徴量からIcosahedron上にCost volumeを生成するために、IcosahedronベースのPlane sweepingを提案する。Cost volumeは3次元CrownConvを用いて正規化し、最終的なCost volumeから全方位深度推定を行う。従来手法においては、Cost volume作成時のみ

¹ <https://github.com/matsuren/crownconv360depth>

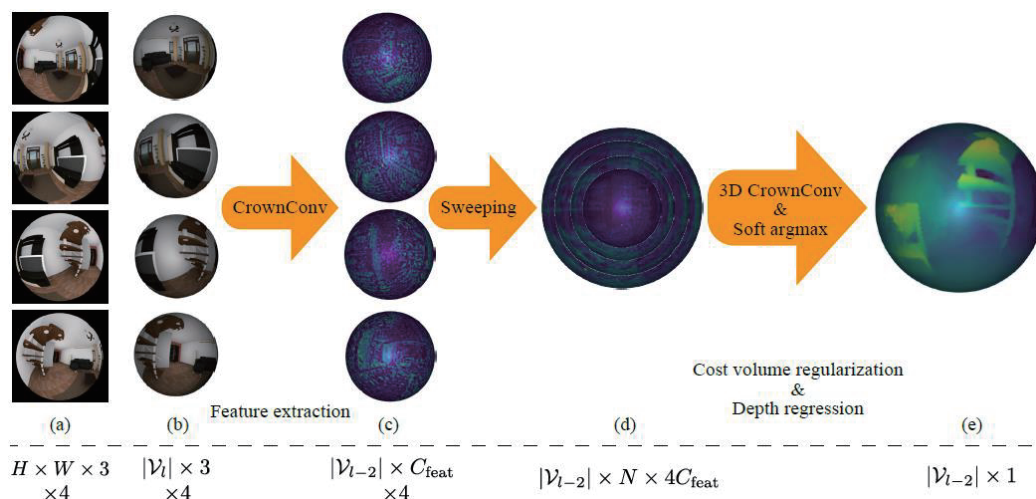


図1 提案手法の流れ.

に外部カメラパラメータを用いていたが、提案手法においては特徴量抽出時においても外部パラメータを用いることで、カメラ配置の変化に頑健な深度推定が可能である。したがって、提案手法においてはカメラ配置が変化するたびに学習し直す必要がないという大きな利点がある。

提案手法の流れを図1に示す。図1(a)がロボットに取り付けられた4台の魚眼カメラ映像である。まずは図1(b)に示すように、魚眼カメラ映像の歪みを除去するために魚眼カメラ映像をIcosahedron上に投影する。ここで、提案手法であるCrownConvを用いてIcosahedron上で魚眼カメラ映像の特徴量抽出を行う(図1(c))。得られた複数台の魚眼カメラ映像の特徴量はIcosahedronベースのPlane sweepingを用いてCost volumeを作成する(図1(d))。最後にIcosahedron上のCost volumeを3次元CrownConvを用いて正則化し、全方位の深度推定を行う(図1(e))。本研究は、魚眼カメラ映像を幾何学的に扱うことで魚眼カメラ映像の歪みを補正し、提案手法のCrownConvを用いることで高速に全方位の深度推定を行えるようになった。

3. 実験および結果

本実験において、学習時のカメラ配置から姿勢変化した際の性能を比較した。これは、ロボットに応じてカメラが設置可能である領域が違うことを想定している。従来手法のようにカメラ配置が学習時と異なる場合に性能が低下するという事は、使用するロボットが変わりカメラ配置が変化するたびに学習が必要となってしまう現場適用性が低いことを意味する。

実験結果を図2に示す。図2において、全方位の深度推定結果を正距円筒画像の深度マップで示している。深度マップにおいて明るい色はカメラ群から近い領域であり、暗い色はカメラ群から遠い領域を示している。図2で示すように、15度、30度、45度と学習時から異なるカメラ姿勢を用いると、従来手法のOmniMVS [2]においては部分的に推定結果が悪化していくのに対して、提案手法においては安定した推定結果が可能であることがわかった。

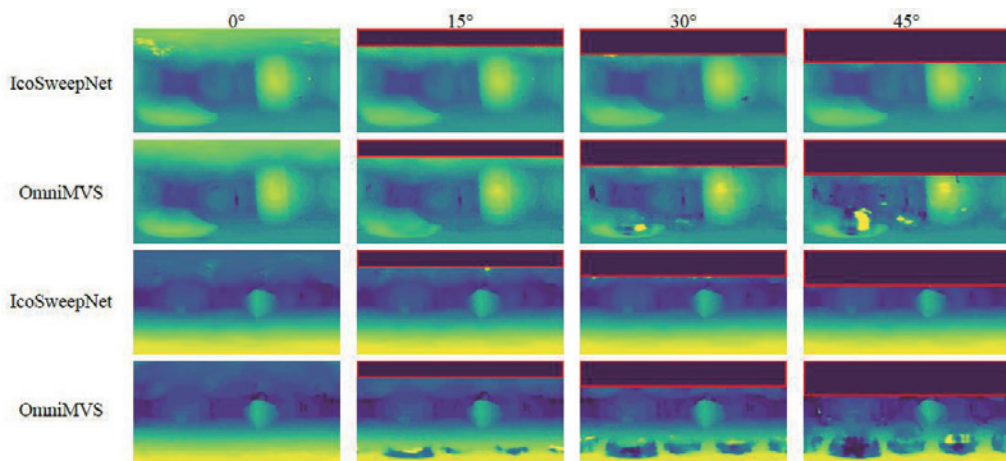


図 2 深度推定比較結果. 提案手法 (IcoSweepNet) と従来手法 (OmniMVS [2]). 上に示した角度は, 学習時のカメラ配置からの姿勢変化を表している.

4. まとめ

本研究では, 屋内環境を対象とした複数の魚眼映像から周囲 360 度の密な深度推定手法を提案した. 魚眼カメラ映像を Icosahedron 上に投影することで魚眼カメラ映像の歪みを除去し, 提案手法である CrownConv を用いることで Icosahedron 上で特徴量抽出および Cost volume の正則化が可能になった. 提案手法により, カメラ配置が学習時から大きく変化しても複数台の魚眼カメラ映像から高精度な深度推定が可能であることを示した.

謝辞

本研究の一部は, 2019 年度後期の若手・女性利用者推薦制度の採択課題として行われ, 東京大学情報基盤センターの Reedbush-H/L を用いた. また東京大学大学院工学系研究科精密工学の浅間一教授および山下淳准教授に多くの指導および助言をいただいた.

参考文献

- [1] Ren Komatsu, Hiromitsu Fujii, Yusuke Tamura, Atsushi Yamashita, and Hajime Asama, “360° Depth Estimation from Multiple Fisheye Images with Origami Crown Representation of Icosahedron”, Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 10092-10099, 2020.
- [2] C. Won, J. Ryu, and J. Lim, “OmniMVS: End-to-End Learning for Omnidirectional Stereo Matching”, Proceedings of the 2019 IEEE International Conference on Computer Vision, pp. 8987-8996, 2019.
- [3] C. Zhang, S. Liwicki, W. Smith, and R. Cipolla, “Orientation-aware semantic segmentation on icosahedron spheres”, Proceedings of the 2019 IEEE International Conference on Computer Vision, pp. 3533-3541, 2019.