

# ライトフィールドカメラとニューラルネットワーク学習を用いた カラー単眼風景画像からの深度画像生成実験

三崎 浩杜<sup>†</sup> 高松 真<sup>†</sup> 上原 知菜<sup>†</sup> 長谷川 誠<sup>†</sup>

<sup>†</sup> 東京電機大学工学部情報通信工学科

## 1. はじめに

ライトフィールドカメラ(図1)を用いて被写体とカメラとの距離を推定し、深度画像を生成することが可能である[1]. RGB-D カメラでも深度画像の生成はできるが、長距離の深度画像を得るにはライトフィールドカメラが唯一の製品であるといっても過言ではない. 風景画像を撮影し、深度画像からの立体視が実現できる. しかし、ライトフィールドカメラは高価であり、さらに製品開発が進まず、市場に普及していない状況である. また、ライトフィールドカメラで生成した深度画像における深度の精度は低く、被写体の前後関係が矛盾している結果が得られる場合もある.

筆者らは、ライトフィールドカメラで住宅地における道路風景のカラー画像と深度画像を多量に撮影し、これらを畳み込みニューラルネットワーク(pix2pix)[2]で学習させ、一般的な単眼のデジタルカメラで撮影したカラー画像から深度画像を生成する方法を検討している[3]. ライトフィールドカメラは必要なく、さらにニューラルネットワークの汎化能力による深度精度の向上が期待できる. ここでは、実現可能性を明らかにするための実験について述べる.

## 2. 実験

ここでは住宅地における道路風景をライトフィールドカメラで撮影する. ライトフィールドカメラによって、RGB カラー画像と深度画像の撮影が可能である. 東京都内300地点におけるRGB カラー画像と深度画像を撮影する. RGB カラー画像と深度画像の組 300 セットを pix2pix で学習する. 図2に例を2セット示す. 学習回数(エポック数)は1000回とする.

学習データセットとは別に1地点のRGB カラー画像を撮影し(図3左側)、学習済の pix2pix を用いて深度画像を推定する. 推定結果を図3右側ss側に示す. なお、ライトフィールドカメラによって撮影した同時点の深度画像を図3中央に示す. 学習済の pix2pix さえあればライトフィールドカメラは必要なく、一般的な単眼のデジタルカメラで撮影したカラー画像から深度画像が得られることを示している.

ライトフィールドカメラによる深度画像では、本来存在しない被写体がノイズとして含まれる場合がある. 図3中央図の空領域(赤丸)におけるノイズが例である. 一方、pix2pix による深度画像(本提案方法)ではこのノ



図1. ライトフィールドカメラ



図2. ライトフィールドカメラで撮影した学習データ(300データセットの2セット;右側は深度画像,左側はRGBカラー画像)



図3. 左側はRGBカラー画像,中央はライトフィールドカメラによる深度画像(正解画像),pix2pixによる深度画像(提案方法による推定画像)

イズは発生していない. また、ライトフィールドカメラでは遠距離における被写体がとらえきれない場合があるが、本提案方法の方が再現できている場合もある(図3中央の図,右側の図における中央の赤丸).道路の深度についても本提案方法の精度は高い.このように、ニューラルネットワークの汎化能力による深度精度向上が確認できる.

## 3. まとめ

一般的な単眼のデジタルカメラで撮影したカラー画像から深度画像が得られることを明らかにした. また、本提案方法はライトフィールドカメラの深度画像よりも深度の精度が高い場合があることを示した.

## 参考文献

- [1] Ng, R., Levoy, et al., "Light Field Photography with a Hand-Held Plenoptic Camera", Stanford Tech Report CTSR 2005-02, 2005.
- [2] P. Isola, et al., "Image-to-image translation with conditional adversarial networks," CVPR, 2017.
- [3] 高松他, "ニューラルネットワークによるライトフィールドカメラの深度精度向上に関する一考察," 電子情報通信学会総合大会, 2019.