

1Y-03

全周画像と Local Light Field Fusion を用いたルームスケール自由視点画像の生成

浦上 詩織 小池 崇文
法政大学 情報科学部

1. はじめに

本研究では、全周画像を用いたルームスケールの自由視点画像の生成手法を提案する。ルームスケールの自由視点画像を生成する際に、スマートフォンや一眼レフカメラなど、通常の画角で撮影を行うカメラを使用すると、膨大な方向・位置で撮影をする必要がある。そこで本研究では、全周画像と、Mildenhall らの提案した Local Light Field Fusion (以下、LLFF) [1] を用いることで、効率よく自由視点画像を生成する方法について検討を行った。全周画像を6枚の平面画像に展開し、LLFFを適用することで、ルームスケールの自由視点画像の生成が可能であることを確認した。

LLFFは、スマートフォンで撮影された、画角が90度以下の画像を用いた自由視点画像の合成法である。まず、入力された画像の位置を推定し、その位置情報をもとに MultiPlane Images (以下、MPI) を生成することで入力画像を深度ごとのライトフィールドである、ローカルライトフィールドに展開する。次に、新しい視点からの画像を生成する際、隣接する4つのローカルライトフィールドを距離の比重に従い合成することで自由視点画像の合成をした。

LLFFでは、視点の向きを大幅に変えることはできない。加えて、生成された自由視点画像の描画範囲は狭いため、描画範囲の広いルームスケールの自由視点画像を生成する際に、処理や撮影にかかる時間が長くなる可能性がある。そこで本研究では、LLFFで自由視点画像を生成する際に、全周画像を用いることで問題の改善を試みる。

2. 準備

本研究で用いる LLFF は、MPI を生成し、入力画像をローカルライトフィールドに展開することで、従来の手法より少ない入力画像で自由視点画像を生成している。MPI は深度ごとに、RGB 画像と RGB 画像の持つ α 値を示す、 α 画像で表現される。MPI を用いることで、単一の画像から、画像の中心を核とした、自由視点画像を生成することができる。

MPI の生成には、Zhou らの手法を使用した [2]。この手法では、膨大な量の画像を用いてデータセットを訓練した。MPI を生成するには、MPI を生成したい画像と、別の角度から同じ場所を写した画像を、データセットにより学習された、ニューラルネットワークに入力する。画像を入力すると、ニューラルネットワークから MPI を構成する背景色と、それを混ぜ合わせる際の重みの比重である α 値が出力として得られる。

MPI を用いて、新たな視点からの画像をレンダリングするには、MPI のレイヤーごとに、処理が必要である。まず、どの程度そのレイヤーの RGB 画像を反映さ

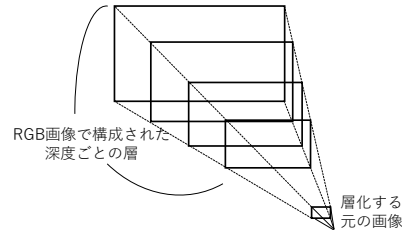


図 1: MPI のイメージ図



(a) RICOH THETA Sで撮影した画像

(b) キューブマップに展開したもの

図 2: 撮影した画像

せるかを定める。これにより、完全に不透明となった領域は、そのレイヤー以降、描写されなくなる。

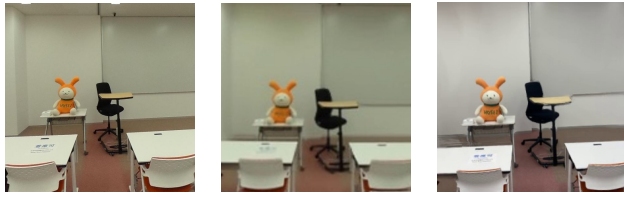
3. 提案手法

まず、全周画像を撮影する。次に、撮影した全周画像をキューブマップに展開する。得られたキューブマップを、一面ごとに LLFF に入力し、MPI を生成する。最後に、新たな視点の位置のパラメータを LLFF に入力すると、新たな視点からの画像を生成される。これを繰り返すことで、自由視点画像を生成する。

3.1. 平面展開

THETA で撮影された画像は図 2 (a) のように正距円筒図法で出力される。この図法ではカメラの位置や向きにより被写体に対するゆがみが大きく変化する。そのため特徴点を検出し、深度推定を行うことは難しい。

そこで本研究では、図 2 (b) のように正距円筒図法をキューブマップに展開する。これにより、画像ごとの被写体のゆがみは軽減し、通常の画角と同様に特徴点の検出及びマッチングを行うことができる。



(a) THETA で撮影し (b) THETA を用いた (c) iPhone を用いた
た画像 生成画像 生成画像

図 3: 入力画像と同地点の生成画像

4. 実験

4.1. 実験条件

三脚にとりつけた RICOH THETA S (以下, THETA) で全周画像を撮影をする。また, スマートフォンで撮影された通常の画角の画像を LLFF に用いた際と, 比較を行うため iPhone 11 Pro (以下, iPhone) を用いて同地点上での撮影を行う。この時, iPhone では全方向の写真を一度に撮ることができないため, 一地点につき, 前後左右にカメラの向きを変えて撮影を行う。

本研究では, システムの有用性を確認するために, 部屋の中心から前後左右に 5 点ずつ平行移動して合計 25 地点で高さを変えて撮影する。また, 生成される画像の視点の高さが, 実世界での視点の高さと同じにするために, 地面からのカメラの高さを, 筆者の目の高さと同しくする。したがって, カメラの光学中心の高さを, 床面から 1.55m とし, 部屋の中心から 0.15m ずつ上下前後左右にカメラを平行移動をさせ撮影を行う。この時カメラは手動で動かしているため, 向きや位置にゆらぎが生じる。しかし, このゆらぎはわずかであるため, 隣り合う地点で撮影された画像の被写体が大きく変化することはない。そのため, 処理を行う過程で大きく問題は無い。

4.2. 結果

図 3 (a) は THETA で撮影された画像である。生成された画像は図 4 のように自由に視点を動かすことができた。また, 図 3 (b), (c) はそれぞれ図 3 (a) と同地点から生成した画像である。入力画像と生成された画像を比較して Structural SIMilarity (以下, SSIM) を求めて評価を行う。SSIM の値は, それぞれ 5 地点の入力画像と同地点からの生成画像を比較した値の平均値である。THETA を用いたとき, 入力画像の数は 75 枚, SSIM は 0.9302 となり, 分散は 1.066×10^{-4} となった。これに対して, iPhone を用いたとき, 入力画像の数は 300 枚, SSIM は 0.9162 となり, 分散 4.605×10^{-4} となった。

4.3. 考察

THETA と iPhone の SSIM は最大値 1.00 に対して 0.90 以上という高い結果が得られた。SSIM はピクセル単体ではなく, 周囲のピクセルとの相関を取り込んだ評価法である。そのため, 生成された画像がぼやけていても相対的な類似度は高いという評価が得られた。

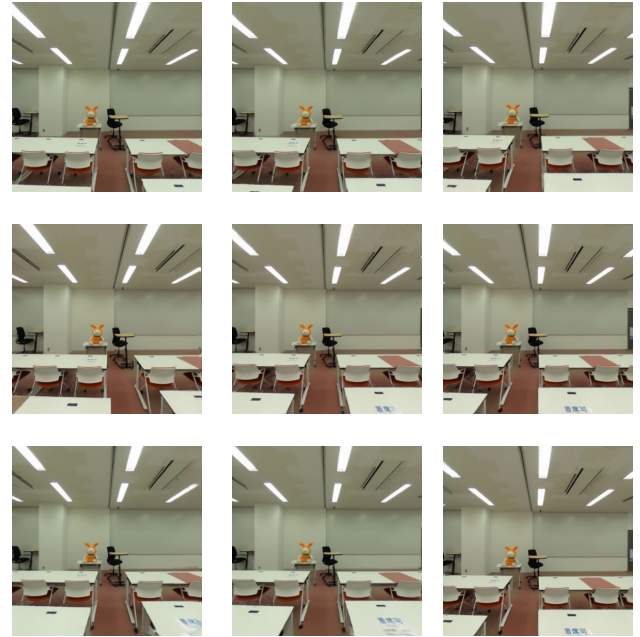


図 4: 生成画像

得られた SSIM の値は高いものの, 図 3 (b) の画質は, 図 3 (a) と比較すると, 高いものといえない。そこで, 画質の改善案として, 画像の撮影間隔があげられる。撮影間隔を狭めると, ローカルライトフィールドごとの距離も近くなる。そのため, 生成したい画像の視点と, 補間に使用するローカルライトフィールドの距離も近くなり, 補間の精度が高くなることが予想できる。また, LLFF では隣接する 4 つのローカルライトフィールドを, 混ぜ合わせて生成画像を出力している。そこで, 混ぜ合わせるローカルライトフィールドの数を増やすと, より高い画質の生成画像を出力できると考えられる。

5. まとめ

本研究では, 全周画像を 6 枚の平面画像に展開し, LLFF を適用することでルームスケールの自由視点画像の生成ができることを確認した。THETA で撮影した全周画像を用いると, iPhone で撮影した画像を用いたときと比較して, 生成に必要な画像の枚数は 75 %削減することができた。これにより, 全周画像に LLFF を適用すると, 効率的にルームスケールの自由視点画像が生成できることがわかった。

今後は, 生成画像の画質の改善と, 定量評価を行い, 発表時にはルームスケールの自由視点映像として生成結果を示す。

参考文献

- [1] Ben Mildenhall, Pratul P. Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. "Local Light Field Fusion: Practical View Synthesis with Prescriptive Sampling Guidelines". *ACM Trans. Graph.*, Vol. 38, No. 4, July 2019.
- [2] Tinghui Zhou, John Flynn, Richard Tucker, Graham Fyffe, and Noah Snavely. "Stereo Magnification: Learning View Synthesis Using Multiplane Images". *ACM Trans. Graph.*, Vol. 37, No. 4, July 2018.