I-012

360 度カメラを用いた深層学習型自由視点画像生成における最適なカメラ台数の評価
Evaluation of the Number of 360-degree Cameras for Deep Neural Network based Free Viewpoint Image Generation

北條 海斗 † 青木 輝勝 †
Kaito Houjho Terumasa Aoki

1. はじめに

視点を自由に切り替えられる画像の生成技術に自由視点画像と呼ばれる技術がある．自由視点画像は画像生成を行うにあたりカメラの台数を数十～数百台設置する必要がある．そのため，画像の生成が難しく，自由視点画像の普及に至っていない．カメラの設置台数を減らすための一方策としてカメラを 360 度カメラに置き換える手法が考えられる．360 度カメラは 1 台で 360 度全体が取得できるため，自由視点画像の生成に用いることを考えると，少量の台数で自由視点画像の生成が可能になると考えられる．しかし，自由視点画像の生成において具体的にカメラの台数が何台程度必要なのか議論されてはいなかった．本稿では，360 度カメラを用いた自由視点画像の生成について議論し，必要なカメラの台数について考察する．

2. 360 度カメラを用いたニューラル場生成

近年，NeRF(Neural Radiance Fields)[1]の登場により自由視点画像の生成精度は飛躍的に向上した．NeRF はネットワーク内部で光線空間を生成する手法である．入力として光線(Ray)の位置情報(x, y, z) と方向(θ, Φ)の 5 次元情報が与えられたとき，それらから色情報と密度情報(rgb, σ)を予測し，光線空間の生成を行う技術である．しかし，NeRF はネットワークの構造が 360 度パノラマ画像に最適化されていないため，学習が収束しない．オブジェクトの輪郭がぼやけてしまうなどの問題が生じる．

360 度パノラマ画像に対応した手法[2][3]も提案されてはいるものの，これらの手法では入力に深度情報も必要となるため利便性が大幅に低下する．深度カメラは元来 360 度の深度計測に対応していないため，360 度の深度情報を取得するためには，深度カメラを数十から数百台設置する必要があるからである．

3. 天球ドーム画像を用いた前処理

2. で述べた通り，NeRF は 360 度パノラマ画像に対応していない．そのため，360 度パノラマ画像を天球ドームに転写し，天球ドームから画像を切り出す．画像を切り出す流れを図 1 に示す．

360 度パノラマ画像からの画像の切り出しは，パノラマ画像の各画素の幾何変換によって求めることが可能である．まず，360 度パノラマ画像の 2 次元画素インデックスを(θ, Φ)と定義する．ただし(θ, Φ)はそれぞれ 360 度パノラマ画像の横方向と縦方向の画素を表し，$(0 \le \theta < 2\pi, -\pi/2 \le \Phi < \pi/2)$を満たすものとする．360 度パノラマ画像と画素

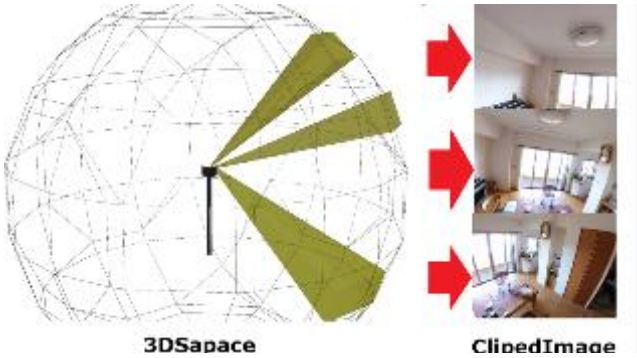

図 1 天球ドームから画像を切り出す流れ

インデックス(θ, Φ)と天球ドーム表面の点(x, y, z)の関係は

$$x = R\cos\phi\cos\theta$$
$$y = R\cos\phi\sin\theta \quad (1)$$
$$z = R\sin\phi$$

である．また， 求めたい切り出し画像の画素インデックスの集合を$\{(i, j)|i = (1, \cdots, I), j = (1, \cdots, J)\}$とする．ただし$I, J$は切り出し画像の縦と横の画素数を表す．続いて，図 1 左に示す天球オブジェクトを天球内部からカメラで撮像するためのパラメータとして，3 次元空間上のカメラスクリーンの縦幅と横幅を(h, w)，カメラからスクリーンまでの距離をcとおく．ここで，天球オブジェクト上の点を撮像するための光線ベクトル集合Sを

$$S = \left\{c, \frac{(2j - J)w}{J}, \frac{(2i - I)h}{I}\right\}, i = (1, \cdots, I), j = (1, \cdots, J) \quad (2)$$

とおく．この光線ベクトル集合 S は 3 次元空間上の原点に配置されているカメラの投影面を表す．今，この光線ベクトル集合 S に対して任意の回転を行い，天球オブジェクト上の点に射影することを考える．ただし，球面の転写するする歪みを極力除去する必要があるため，カメラに任意の回転を施した後，カメラを球面方向に近づける処理を施す．ここで任意のヨーとピッチの角度を$(\hat{\theta}, \hat{\Phi})$とおき，光線ベクトル集合$S$を任意角度$(\hat{\theta}, \hat{\phi})$で回転させた光線ベクトル集合を$\hat{S}$とする．

このとき，単位ベクトル$(1, 0, 0)$をその任意角度で回転させたベクトルを$0 < \hat{r} < 1$倍させたベクトルを$\hat{p}$とし，カメラを$\hat{p}$だけ原点から平行移動させた場合において，天球オブジェクト上にカメラ投影面を射影した際の光線ベクトル集合は以下のように表される．

$$\tilde{S} = \{\hat{p} + \alpha v\} \quad (3)$$
$$v \in \hat{S}, |\hat{p} + \alpha v|_2 = R, \alpha > 0$$

† 東京工科大学 Tokyo University of Technology

この座標集合は(2)より(θ, Φ)に変換することが可能であり，$\hat{S}$上の点(x, y, z)を

$$\theta = \arcsin\left(\frac{y}{R\cos\phi}\right)$$
$$\phi = \arcsin\left(\frac{y}{R}\right) \quad (4)$$

に写像することで 360 度パノラマ画像の画素インデックスと切り出し画像の画素インデックスの対応付を行うことができる．

上記の手法を用いて切り出した画像を入力画像として扱い，SfM(Structure-From-Motion)[5]を用いてカメラの位置情報の取得を行った後，NeRF の学習を行う．

実験

4. 1 実験環境

実験に伴い個別に設定したパラメータや実験条件について述べる．

まず，切り出した画像の生成について説明する．切り出し画像の生成は，3. の通り天球ドーム画像の歪みを最小限にするため，天球ドームの原点と射影するスクリーンの距離を$\tilde{r} = 0.1$に近づける処理を行った．また，360 度パノラマ画像は元画像の解像度が 6720x3260 と非常に大きい．そのため，本実験では，メモリを削減する観点から切り出した画像を 853x481 に縮小し，比較実験を行った．

また，NeRF には入力データとしてカメラの位置情報を入力する必要がある．そのため本実験では，COLMAP[5][6]を用いてカメラの位置情報の推定を行った．深層学習を用いた自由視点画像の生成は InstantNGP[3]を用いて実施し，Iteration 数を 550000 回，ハッシュマップのサイズは 19，最適化関数は Adam を利用した．

本実験では実世界の屋内と屋外のシーンを 360 度カメラにて撮影した画像を用いて実験を行った．入力画像を 3 次元空間上の天球ドームから画像の切り出しはヨーにて 4 枚，ピッチにて 5 枚を等間隔に切り出したものを入力画像とした．

4. 2 カメラの台数と生成精度の違い

カメラの台数と切り出す際の間隔を示したものを表 1，表 2 に示す．撮影画像は自由視点画像生成に NeRF に入力した 360 度カメラの台数を示している．評価は PSNR 値と SSIM 値にて行った．

s

表 1 屋外シーンにおける生成精度

カメラ台数(台)	PSNR (db)	SSIM
3	11. 32	0. 3824
5	12. 37	0. 3838
10	**18. 41**	**0. 8328**
20	12. 55	0. 4291
30	13. 00	0. 4502

表 2 屋内シーンにおける生成精度

カメラ台数(台)	PSNR (db)	SSIM
3	14. 67	0. 6642
5	15. 03	0. 6832
10	**15. 72**	**0. 7261**
20	14. 99	0. 7249
30	14. 91	0. 7252

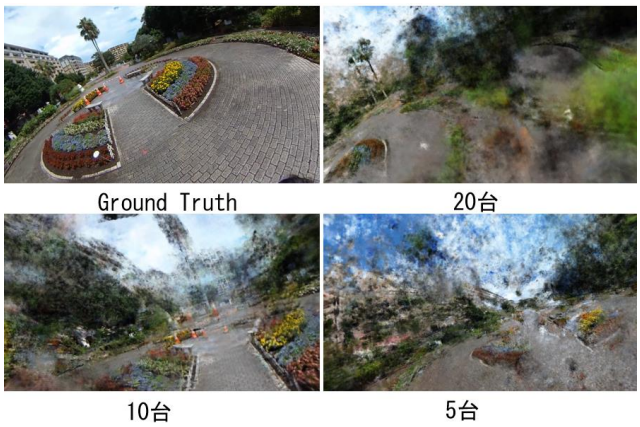

図 2 生成結果

本提案手法を用いて生成した結果を図 2 に示す．上記のように，カメラ 20 台(右上)とカメラ 10 台(左下)を比較すると，カメラ 10 台の画像の方が Ground Truth(左上)と類似している．これらの結果から，屋内，屋外双方において NeRF に入力する 360 度画像の枚数は 10 枚の画像から切り抜く組み合わせが最も精度が高くなることがわかる．また，入力画像が 10 枚を超えてしまうと生成精度が屋内，屋外共に低下してしまうことがわかった．これは，NeRF 内に入力する画像が多すぎるためそれぞれの画像がノイズの要因となってしまい，精度が低下していると考えられる．

また，自由視点画像の生成を更に広いシーンで生成することを考えた場合，必要とする 360 度画像の枚数は生成範囲の広さに応じて増加すると考えられるが，表 1，2 の結果に同様に入力画像の枚数が多すぎると精度が低下してしまうと考えられる．

4. 結論

本稿では 360 度パノラマ画像を NeRF のネットワークに入力する際の最適なカメラ台数と各画像の幾何変換の手法について考察した．また，実画像を用いた自由視点画像の生成を行うことで，自由視点画像の生成精度は画像枚数が多すぎるとノイズの要因となってしまい精度が低下してしまうことを実証的に示した．

本提案手法を用いることで，少量のカメラ台数の環境でも 360 度カメラ特有の広い視野角で光線空間を生成することができ，自由視点画像の生成が可能となることを示した．本提案手法はカメラ台数を減らすのみならず，ドローンやロボティクスにおいても少ない情報量から自由視点画像の生成が可能となる．

文献

[1] Ben Mildenhall and Pratul P. Srinivasan and Matthew Tancik et al, “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis”, ECCV(European Conference on Computer Vision)2020.
[2] Ching-Yu Hsu, Cheng Sun , Hwann-Tzong Chen et al, “Moving in a 360 World: Synthesizing Panoramic Parallaxes from a Single Panorama”, ECCV(European Conference on Computer Vision) 2022.
[3] Shreyas Kulkarni, Peng Yin, and Sebastian Scherer et al , “360FusionNeRF: Panoramic Neural Radiance Fields with Joint Guidance” arXiv:2209. 14265 (arXiv preprint 2022)
[4] Thomas Muller and Alex Evans and Christoph Schied et al “Instant Neural Graphics Primitives with a Multiresolution Hash Encoding”, ACM 2022.
[5] Schonberger, Johannes Lutz and Frahm, Jan-Michael, “Structure-from-Motion Revisited”, CVPR (Conference on Computer Vision and Pattern Recognition)2016.
[6] Schonberger, Johannes Lutz and Zheng, Enliang and Pollefeys, Marc and Frahm, Jan-Michael “Pixelwise View Selection for Unstructured Multi-View Stereo”, ECCV (European Conference on Computer Vision)2016.