

敵対的生成ネットワークによる 全方位自由視点画像の画質改善手法

竹内 音^{1,a)} 宍戸 英彦^{2,b)} 亀田 能成^{2,c)} 北原 格^{2,d)}

概要

敵対的生成ネットワークを用いて全方位自由視点画像生成における画質劣化を改善する手法を提案する。多視点で撮影された全方位画像を計算機内部で統合し、撮影空間の3次元情報を推定することで、実際には撮影していない視点からの全方位の見え方の再現が可能であるが、3次元情報の推定誤差によるアーチファクトやオクルージョンによる欠損によって、自由視点画像は画質が劣化するという問題が存在する。敵対的生成ネットワークを用いることによりこれらの問題を解消し、自由視点画像を画質改善する。

1. はじめに

360 カメラと呼ばれる全方位カメラで撮影した画像は周囲を見回すような観察に適しており、HMD(Head-Mounted Display)提示と組み合わせることで没入体験が可能であることから、近年注目を集めている。Google Street View[1]のように、複数視点から撮影された全方位画像を切替えることにより、視点移動を伴った多面的な観察も可能となる。我々は、全方位多視点画像から全方位カメラの位置姿勢と撮影空間の3次元情報を推定することで、観察したい箇所を注視しながら視点を切替える Bullet-Time 映像生成提示方式を提案した[2]。しかし、撮影視点以外での観察には対応していないこと、多視点カメラの撮影間隔が広がると視点移動の滑らかさが損なわれる、という課題が残った。

本研究では、全方位多視点画像に自由視点画像生成技術を適用し、撮影視点以外からの見え方を再現する。全方位カメラを用いることで、一般的なカメラでの撮影と比べて重複撮影領域をより広く取ることができる(つまり、3次元空間中の同一箇所がより多くの視点から観察される)ため、3次元情報推定に重要な対応点探索精度が向上し、カ

メラキャリブレーション精度の向上、自由視点画像の高画質化、効果的な深層学習などの効果が期待できる。

それでもなお、3次元情報の推定誤差によるアーチファクトやオクルージョンによる欠損領域による、自由視点画像の画質劣化の問題は発生する。本稿では、3次元情報に基づいて生成した自由視点画像と実際に撮影した画像(正解画像)に見え方の関係性を深層学習し、その学習結果を用いて自由視点画像の画質を改善する手法を提案する。深層学習には、相反する目的を持つ二つのネットワークから構成される敵対的生成ネットワーク(GAN: Generative Adversarial Network)を用い、撮影画像に近い見え方の自由視点画像を合成する。その際、全方位画像を複数の透視投影画像に分割し、全方位カメラの射影特性による見え方の多様性を低減することにより、深層学習の学習効率を高め、より写実的な画像生成を行う。

2. 全方位自由視点画像の生成と画質改善

図 1 に、全方位多視点画像の撮影から全方位自由視点画像の画質改善までの処理を示す。対象空間を撮影した全方位多視点画像に対して Structure from Motion(SfM)を適用す

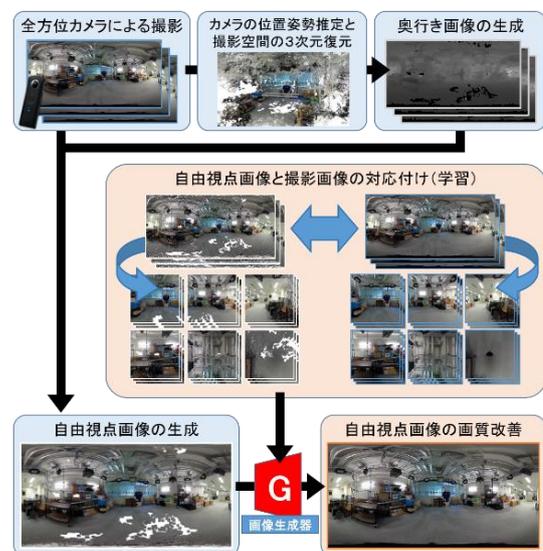


図 1 全方位自由視点画像生成と画質改善

¹ 筑波大学 大学院システム情報工学研究科

² 筑波大学 計算科学研究センター

^{a)} takeuchi.oto@image.iit.tsukuba.ac.jp

^{b)} shishido@ccs.tsukuba.ac.jp

^{c)} kameda@iit.tsukuba.ac.jp

^{d)} kitahara@ccs.tsukuba.ac.jp

ることで、各全方位カメラの位置姿勢と撮影空間の 3 次元点群を推定する[2]。推定したカメラパラメータに基づいて 3 次元点群を各全方位画像面上に投影し、投影点の隙間を補間処理することで各視点における全方位奥行き画像を生成する。この全方位奥行き画像と撮影した全方位画像から任意視点からの見え方(全方位自由視点画像)を生成する。別途学習しておいた深層学習結果(画像生成器)を用いて、生成した自由視点画像の画質を改善する。3 節では、生成処理、4 節では画質改善処理の詳細について述べる。

3. 全方位自由視点画像の生成

3.1 全方位画像の撮影と 3 次元情報推定

全方位カメラを用いて複数の視点から全方位多視点画像を撮影する。多視点画像からの 3 次元情報推定に関する活発な研究開発により、優れた SfM ライブラリ[3]が利用可能であるが、全天球の見え方を 1 枚の画像面に記録する際の射影変換は、汎用的な SfM ライブラリが対象とする透視投影とは異なる。本方式では、全方位画像を複数枚の透視投影画像に分割し、分割画像に対して SfM ライブラリを適用する。各画像に対応したカメラパラメータと疎な 3 次元点群を推定し、推定されるカメラの位置姿勢より全方位カメラの位置姿勢を算出する[2]。推定したカメラパラメータと疎な 3 次元点群を元に Multi-View Stereo 処理[4]を実行し、密な 3 次元点群を取得する。

3.2 全方位奥行き画像の生成

各カメラ視点から 3.1 節で推定した 3 次元点群までの距離を算出し、図 2(a)に示す全方位奥行き画像を生成する。投影する 3 次元点群の色情報と奥行き情報を求める視点で実際に撮影した画像上での観測色の差を求め、差分の絶対値が閾値より小さい 3 次元点群についてのみ、奥行き値を推定する。誤った 3 次元情報が推定された点群では、点群の色情報と観測色が大きく異なることが多いため、この閾値処理によって 3 次元復元誤差の影響を軽減する。

3 次元点群が投影されなかった画素では奥行き値が推定されないため、図 2(a)に示すように隙間のある奥行き画像が生成される。そこで、クロスバイラテラルフィルタ[5]を用いたフィルタリング処理によって隙間を補間する。クロスバイラテラルフィルタでは、同一視点において異なる 2 種類のモダルの画像情報を観測し、そのうち観測ノイズが少ない方の画像を基準として、もう一方の画像のフィルタリングを行う。本方式では、観測ノイズの少ない画像として撮影画像 (RGB 画像) を用いて、観測ノイズの大きい奥行き画像をフィルタリングする。適用するフィルタの式を以下に示す。

$$D_p = \frac{\sum_{r \in N} d(\|p - r\|_2) c(\|I_p - I_r\|_2) D_r}{\sum_{r \in N} d(\|p - r\|_2) c(\|I_p - I_r\|_2)}$$

$$d(\|p - r\|_2) = \exp\left(-\frac{\|p - r\|_2}{2\sigma_1^2}\right),$$

$$c(\|I_p - I_r\|_2) = \exp\left(-\frac{\|I_p - I_r\|_2}{2\sigma_2^2}\right)$$

ここで、 p は注目画素座標、 r は参照画素座標、 D は奥行き値、 I は輝度値、 σ は定数、 N は参照画素座標の集合を表す。注目する欠損画素位置に対し、参照する画素位置との距離と、撮影画像上における色類似度を重みとし、奥行き値を推定する。これにより、図 2(b)に示すように、撮影画像の輪郭を保持しながら奥行き画像を補間することが可能となる。

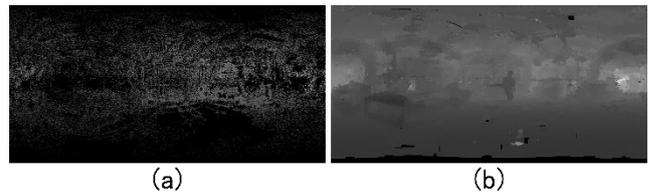


図 2 生成された全方位奥行き画像：補間処理前(a)、後(b)

3.3 奥行き画像を用いた全方位自由視点画像の生成

図 3 に示すように、3.1 節で撮影した全方位画像と 3.2 節で生成した全方位奥行き画像を用いて、任意視点における全方位自由視点画像を生成する。はじめに、新たに全方位画像を生成する視点位置を定め、その視点から各多視点カメラの位置までの距離を算出する。その中から、距離の近い順に一定数のカメラを選択し、各々の撮影画像の画素値を、全方位奥行き画像を参照して 3 次元空間に逆投影して 3 次元点群を生成する。これらの点群を新規視点位置における全方位画像面に投影することにより、全方位自由視点画像を生成する。自由視点画像の同一画素に異なる点群が投影される場合は、新規視点に近い方の点群を採用する。

同様の処理によって、実際に撮影した視点位置において自由視点画像を生成し、次節で述べる GAN を用いた画質改善処理の学習用データセット (自由視点画像と実際に撮影した画像のペア) を用意する。

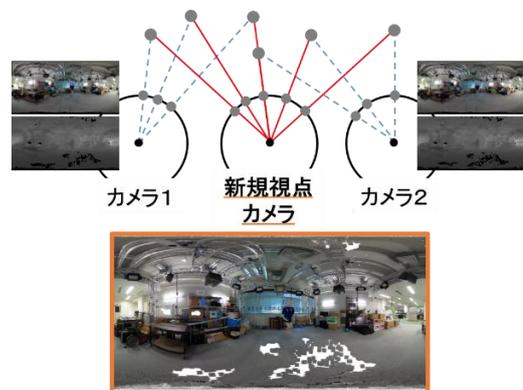


図 3 全方位自由視点画像の生成

4. 全方位自由視点画像の画質改善

3.3 節で生成した自由視点全方位画像には、撮影画像では観測されない 3 次元形状推定誤差によるアーチファクトやオクルージョンによる欠損領域が含まれている。本節では、GAN を用いたそれらの問題の軽減法について述べる。ここでは GAN の実施手段として Pix2Pix[6]を用いる。Pix2Pix は、条件付き GAN の一種であり、線画画像と写真や航空写真と地図など、スタイルの異なる 2 画像間の対応関係を学習し、一方をもう片方のスタイルに変換することができる。本研究では、自由視点画像と撮影画像の画像変換に Pix2Pix を適用し、自由視点画像の画質改善を試みる。

Pix2Pix は具体的に、画像生成器と識別器の二つのネットワークから構成される。学習データとして変換前画像と変換後画像をペアにした画像群を用意し、画像生成器には変換前画像を入力し、識別器には画像生成器が生成した画像、もしくは用意した変換後画像を入力する。識別器はどちらの画像が入力されたのかを判断する。画像生成器は識別器を欺くような画像生成を、識別器は正確な判断ができるよう、互いに競合させながら学習が進められる。学習後、画像生成器を用いることで画像変換を実現する。

今回、学習データには、変換前画像として 3.3 節で生成した実際に撮影した視点における全方位自由視点画像を、変換後画像として 3.1 節で撮影した全方位画像を用意する。これらの学習データを用いて画像生成器を学習させた後、仮想視点における全方位自由視点画像を学習済み画像生成器に入力することで、画質劣化を軽減した写実性の高い画像を生成する。

我々は、学習画像の射影幾何に着目し、学習の効率化を実現する。正距円筒図法に基づく全方位画像は、その射影特性上、視点移動に伴う見え方の変化が大きいため、学習サンプル間での見え方の多様性が大きくなり、効果的な学習が難しくなる。そこで、全方位画像を複数の透視投影画像に分割することで見え方の多様性を軽減し、効率的な GAN の学習を行う。本稿では、全方位画像を 6 面に分割するキューブマッピングを採用し、各面における透視投影画像を用いて画像生成器の構築を行う。

5. 実験

5.1 実験環境

深層学習による全方位自由視点画像の画質改善効果、および、画像分割が学習効率に与える効果に関する実証実験を実施する。図 4 に示すように、屋内環境（筑波大学研究室）の 42 箇所に全方位カメラ（RICOH 社 THETA S）を取付けた三脚を設置し多視点全方位画像を撮影した。CPU：Intel Core i7-7700HQ 2.80GHz、GPU：NVIDIA GeForce GTX 1060、メモリ：16.00GB RAM を装備したノート PC を用いた。SfM は VisualSfM[3]を用いて実行した。

3.3 節で述べた手法により、全撮影視点の全方位自由視点画像 42 枚を生成し、その内 22 視点分の全方位自由視点画像と同視点で実際に撮影した画像を Pix2Pix[6]の学習データとする。

前節で述べた画像分割による GAN の学習効果を検証するため、正距円筒図法による画像をそのまま用いた場合とキューブマッピングによって分割した透視投影画像を用いた場合に対し、画像生成器を学習する。正距円筒図法に基づく全方位自由視点画像は 2018 画素×1024 画素、各透視投影画像は 512 画素×512 画素とし、学習回数は各々 1000 epochs とした。学習に使用していない評価用の 20 視点分の全方位自由視点画像を画像生成器に入力し、生成される画像を観察する。また、画質評価指標の一つであるピーク信号対雑音比(PSNR)を用いて画質の定量的評価を行う。

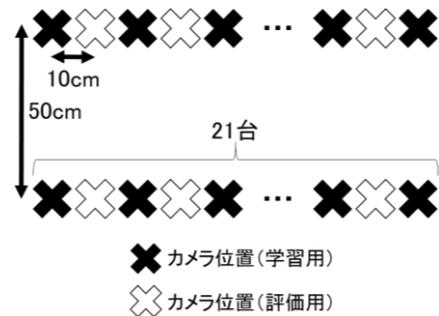


図 4 全方位カメラの配置

5.2 実験結果

比較した画像の一例を図 5 に示す。図 5(a)は、3 節で述べた手法で生成した（画質改善前の）全方位自由視点画像である。図 5(b)(c)は、我々が提案する画質改善を施した全方位自由視点画像であり、(b)は分割した画像を入力とした場合、(c)は正距円筒図法の全方位画像をそのまま入力した場合である。図 5(d)は実際に撮影した画像（正解データ）である。図 5(a)と図 5(b)(c)を比較すると、深層学習によって構築した画像生成器によって画像中の欠損領域が改善されていることが確認できる。正解データの図 5(d)と比較すると、分割した画像を入力とした図 5(b)の方がアーチファクトの少ない画像が生成されており、ボケの少ない鮮明な画像が生成されていることが確認できる。

評価用 20 視点分の全方位自由視点画像で算出した PSNR の平均値を用いて、画質改善の効果と画像分割の有無に関する定量評価を行う。評価結果を表 1 に示す。

表 1 20 視点分の画像の PSNR の平均値

画質改善前	画質改善後	
	画像分割有	画像分割無
12.68[dB]	27.39[dB]	23.01[dB]

表 1 より、深層学習によって構築された画像生成器を用いることで PSNR が向上したことから、画質が改善されていることがわかる。また、画像分割しない手法より画像分

割を行った手法の方が高い PSNR の値を示しており、画像分割の有効性が確認できる。

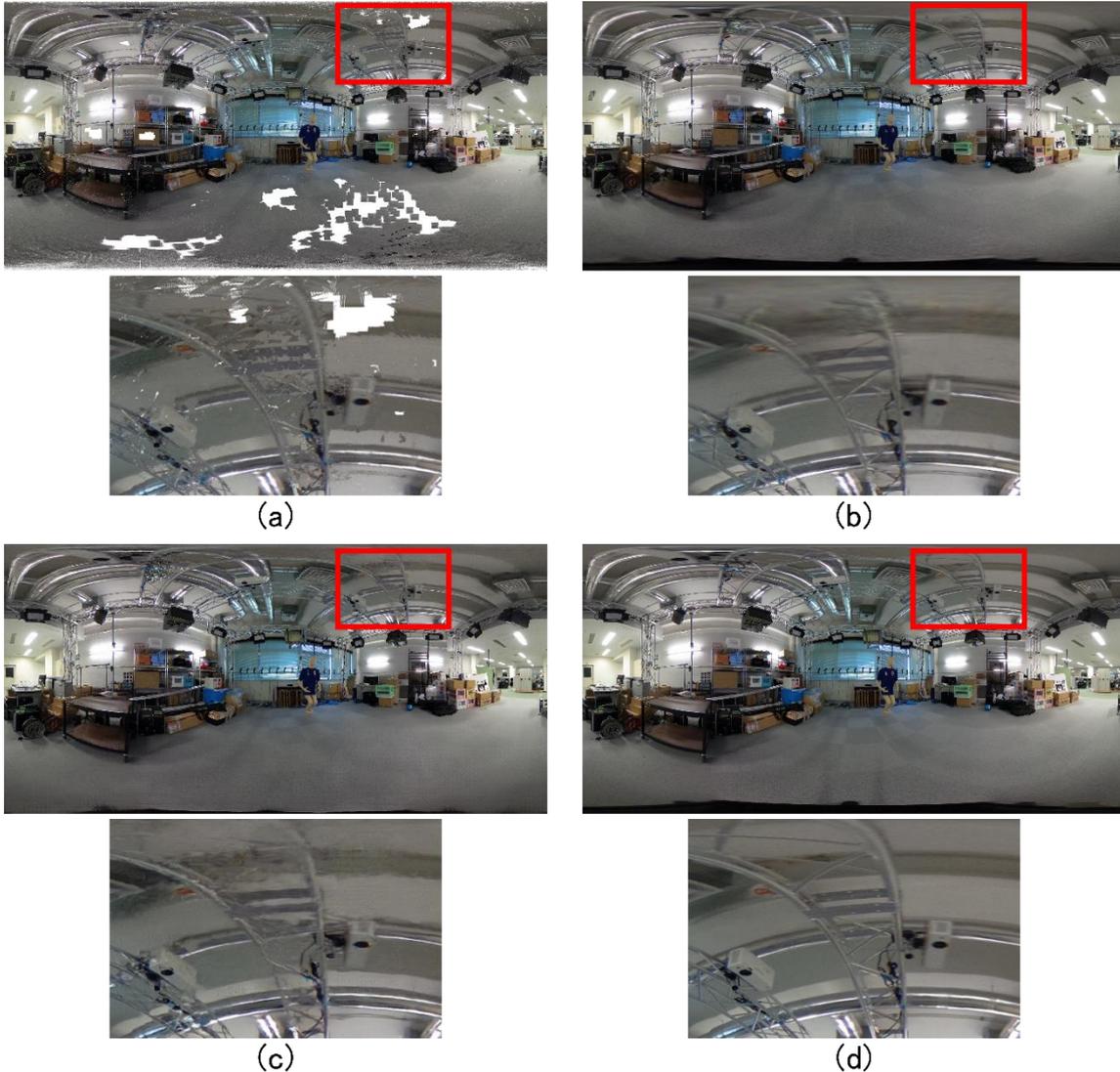


図 5 比較結果（上）とその拡大図（下） (a)：全方位自由視点画像（画質改善なし） (b)：提案手法・画像分割での学習（画質改善あり） (c)：提案手法・全方位画像での学習（画質改善あり） (d)：正解画像（撮影画像）

6. おわりに

本稿では、GAN を用いた全方位自由視点画像の画質改善手法を提案した。全方位多視点画像から撮影空間の 3 次元情報を復元し、画像処理によって奥行き情報の補間を行った後に全方位自由視点画像を生成した。GAN を用いた深層学習を用いることで、従来の自由視点画像に見られたアーチファクトや欠損領域の画質を改善した。学習時の射影幾何に着目し、全方位画像を透視投影画像に変換することで、画質改善性能を向上させた。

References

[1] Google Street View, <https://www.google.com/streetview/>
 [2] O. Takeuchi, H. Shishido, Y. Kameda, and I. Kitahara, Generation Method for Immersive Bullet-Time Video Using

an Omnidirectional Camera in VR Platform, ACM Multimedia Workshop on AVSU, pp.19-26, 2018.
 [3] C. Wu, VisualSFM: A Visual Structure from Motion System, <http://ccwu.me/vsfm>
 [4] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms, CVPR, 2006.
 [5] L. Chen, H. Lin, and S. Li, Depth image enhancement for Kinect using region growing and bilateral filter, ICPR, 2012.
 [6] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, CVPR, 2017.