敵対的生成ネットワークを用いた 全方位自由視点画像の生成に関する検討

竹内 音[†] 宍戸 英彦[‡] **亀田 能成[‡]** 北原格*

† 筑波大学 大学院システム情報工学研究科 〒305-8573 茨城県つくば市天王台 1-1-1 ‡筑波大学 計算科学研究センター 〒305-8577 茨城県つくば市天王台 1-1-1

E-mail: [†] takeuchi.oto@image.iit.tsukuba.ac.jp, [‡] {shishido, kameda, kitahara}@ccs.tsukuba.ac.jp,

あらまし 全方位自由視点画像の画質向上を実現する手法を提案する. 多視点画像を計算機内部で統合し任意視 点からの見え方を再現する自由視点画像生成処理では、撮影空間の3次元情報復元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が発生し,画質が劣化する問題が知られている.本研究では,全方位画像を用いて自由視 点画像を生成し、敵対的生成ネットワーク(GAN)によって自由視点画像の見え方を再構成することにより、画質 劣化を軽減する、我々は、全方位画像を複数の透視投影画像に分割し、見え方の多様性を低減することで、GANの 効率的な学習を行う.本稿では,全方位多視点画像の撮影から全方位自由視点画像の生成までの処理について述べ, 実証実験を通じてその有効性を示す.

キーワード 自由視点画像,全方位カメラ,全方位奥行き画像,敵対的生成ネットワーク,画像再構成

1. はじめに

上下左右の空間を一度に撮影可能な全方位カメラ で撮影した全方位画像は、周囲を見回す観察が可能で あるため、ヘッドマウントディスプレイのような全方 位提示装置と組み合わせることによって、撮影空間に 没入した視点からの観察を体験することができる. Google Street View [1]のように, 撮影空間の複数箇所に 設置した全方位カメラで全方位多視点画像を撮影し, それらの画像を適宜切り替えることで、視点移動を伴 ったより多面的な撮影空間の観察が可能となる. 我々 は,全方位多視点画像に3次元情報推定処理(SfM: Structure from Motion)を適用することで、全方位カメ ラの位置姿勢と撮影空間の3次元形状を推定し、それ らの情報を用いて観察したい箇所を注視しながら視点 を切り替える Bullet-Time 映像生成提示方式を提案し [2], 撮影空間全体において注目物体の3次元的な理解 を支援することに成功している. 全方位画像を撮影し た視点では全方向の観察が可能であるものの, 撮影し ていない視点からの観察には対応していないため、多 視点カメラの間隔が広くなると、視点移動の滑らかさ が損なわれ,撮影視点以外への視点移動ができないと いう課題が残っている.

任意視点からの見え方を再現することを目的とし た自由視点画像生成に関する研究[3]-[12] が活発に行 われているが、撮影空間の3次元情報復元誤差による アーチファクトやオクルージョンによる見え方の欠損 による画質劣化は未だ問題となっている. RGB-D カメ ラなど奥行き情報の取得が可能な装置を用いることで, 自由視点映像の品質を向上させるアプローチも存在す るが[6][7][12],撮影装置の簡易性が損なわれることが

実利用時の課題である.一方で我々の提案法では,全 方位カメラを用いて撮影を行うため、多視点画像間の 重複撮影領域を広く取ることができる.結果として, 多くの視点から3次元空間中の同一箇所が観察される ため、3次元形状推定にとって重要な対応点探索の精 度が向上し,自由視点映像の高画質化が期待できる.

画像再構成処理によって画質劣化が発生している 領域の見え方を補完する研究が行われている[13].近 年では,深層学習を用いた手法[14][15]も提案されてお り、より自然な画像補完がなされている.しかし、こ れらの手法では,補完対象となる領域の画像中での箇 所が既知であることを前提としているため、撮影状況 によって発生箇所が変化する画質劣化要因の補完への 適用は難しい. また, 深層学習を用いた画像再構成手 法では,学習データの多様性が学習効果に影響を与え るが,全方位カメラでは射影特性上,視点移動に伴う 見え方の変化が大きいため、見え方の効率的な学習が 課題となる.本稿では、これらの問題を解消する全方 位自由視点画像の生成手法を提案する.相反する目的 を持つ二つのネットワークによって構成され、「本物ら しい」画像の生成を得意とする敵対的生成ネットワー ク(GAN: Generative Adversarial Network)によって画像 再構成を実現する.これにより,任意の箇所で発生す る画質劣化に対して実際に撮影された画像と遜色ない 見え方を再現する.また,撮影した全方位画像を複数 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で, GAN の効率的 な学習を行い,生成画像の画質向上を実現する.

2. 関連研究

2.1. 全方位多視点画像の提示

Google Street View[1]では、全方位画像を用いること で,上下左右を見回す観察が可能である.複数視点で 撮影された全方位画像を閲覧者が指定した視点移動に 従って切り替えることで,撮影空間を移動しながらよ り詳細に状況を把握することができる.ブレンディン グ処理と画像形状変換を組み合わせることによって, 視点が遷移していることを閲覧者が知覚するよう工夫 が施されている. 我々も, 全方位多視点画像に3次元 情報推定処理を適用することで、全方位カメラの位置 姿勢と撮影空間の3次元形状を推定し、それらの情報 を用いることで、観察したい箇所を注視しながら視点 を切り替える Bullet-Time 映像生成提示方式を提案し [2], 撮影空間全体において注目物体の3次元的な理解 を支援することに成功している.しかし、全方位画像 切り替え方式には,撮影視点以外への視点移動ができ ないという問題が存在する.

2.2. 自由視点映像の生成

自由視点映像の生成提示に関する研究が盛んに行われている.撮影空間の3次元形状を復元した3次元 モデルによって任意視点からの見え方を再現する Model-Based Rendering (MBR) [3]-[7]や,3次元モデル を陽に復元することなく撮影した多視点画像群から直 接的に見え方を合成する Image-Based Rendering (IBR) [8]-[12]が実現されている.

MBR では、生成される自由視点映像の品質は、復元 される3次元モデルの精度に依存するため、3次元復 元誤差が生じやすい複雑な形状の撮影空間を対象とす る場合、誤差の影響による見え方の劣化(アーチファ クト)が発生する.また、オクルージョンによって複 数の多視点カメラから観測されていない箇所は、3次 元情報の推定が困難であり、画像情報が欠損してしま う.

陽に3次元形状を復元しない IBR では,撮影空間の 複雑さに依存せずに自由視点画像の生成が可能である が,画像形状変換時に想定した撮影空間の形状と実際 の撮影物体形状が大きく異なる場合,画像変換によっ て物体の見た目に歪みが生じる.あらゆる視点におい てこの歪みを軽減した良好な見え方を再現するために は,密に配置された多数のカメラでの撮影が必要とさ れる.

2.3. 画像再構成

画像再構成に関する研究も数多くなされている. Barnes ら[13]は、周辺の画像連続性を利用し、対応す る画像情報を見つけることで画像の見え方を補完して いる.しかし、この手法は画像内に含まれていない情 報を復元することはできないという問題点がある.畳



図 1 敵対的生成ネットワーク(GAN)を用いた全方 位自由視点画像の高画質化

み込みニューラルネットワークと GAN を活用し,画 像内に含まれていない情報を復元する手法も提案され ているが,これらは欠損領域が既知であることを前提 としている[14][15].我々は,画像全体を変換する GAN [16]を活用することにより画像再構成を実現し,任意 の箇所で発生する画質劣化に対して実際に撮影された 画像と遜色ない見え方を再現する.

3. 敵対的生成ネットワーク(GAN)を用いた 全方位自由視点画像の高画質化

図1に、GAN を用いた全方位自由視点画像の高画 質化の概要を示す.撮影した多視点全方位画像に対し てSfMを適用することで,各全方位カメラの位置姿勢 (カメラパラメータ)と撮影空間の3次元点群を推定 する.推定したカメラパラメータに基づき,3次元点 群を各全方位画像面上に投影した後,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.任意視点の全方位自由視点画像は,全方 位奥行き画像と実画像(撮影画像)を用いて合成する ことができる.合成画像には3次元情報推定誤差によ るアーチファクトや欠損領域が含まれているため,そ れらの領域の見え方を GAN によって画像変換するこ とにより,見え方を再構成し,全方位自由視点画像の 画質を向上させる.

4. 全方位自由視点画像の生成

4.1. 全方位画像の撮影と3次元点群の取得

複数の視点に設置した全方位カメラを用いて全方 位多視点画像を撮影する.全方位画像は,全天球を1





ф

 $\pi/2$

枚の画像面に記録するため,汎用的な SfM ライブラリ [18][19][20]が対象としている透視投影幾何とは異な る射影幾何で表現されている.一方で,SfM は目覚ま しい開発が進む分野であり、高精度な3次元情報を推 定するためには、最新のアルゴリズムを適用すること が望ましい. そこで本方式では、全方位画像を複数枚 の透視投影画像に分割し、分割結果を SfM ライブラリ に適用することで,分割画像に対応した透視投影カメ ラの位置姿勢と疎な3次元点群を推定する.透視投影 カメラの位置姿勢から全方位カメラの位置姿勢を求め, 全方位画像の縦横が SfM 座標系の仰俯角と方位角に対 応するように座標変換処理を施すことで、すべての全 方位画像の方位を一致させる. 撮影して得られた画像 (座標変換処理前)を図 2(a)に,座標変換後の方位を 一致させた画像を図 2(b)に示す. 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 を行うことで,撮影空間の密な3次元点群を生成する.

4.2. 全方位奥行き画像の生成

4.1 節で生成した3次元点群を,各全方位画像面上 に投影する.各カメラ位置から3次元点群までの距離 を計算し,図 3(a)に示す全方位奥行き画像を生成する. 投影された3次元点群の色と投影点で実際に観測され ている色の差(色差)を求め,色差が閾値よりも大き い場合,その3次元点群の投影は行わず処理を進める. これにより,3次元復元誤差の影響を軽減することが 可能となる.

上述した処理では、3次元点群が投影されなかった



図 3 生成された全方位奥行き画像 (a):補間処理前 (b):補間処理後

画素についての奥行き値は得られないため,図 3(a)に 示すような隙間の多い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[17]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モーダ ルの画像情報を2枚観測し、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 像として実画像(RGB 画像)を用いて、観測ノイズの 大きい奥行き画像のフィルタリング処理を行う.適用 するフィルタの式を以下に示す.

$$D_{p} = \frac{\sum_{r \in N} d(\|\boldsymbol{p} - \boldsymbol{r}\|_{2}) c(\|\boldsymbol{I}_{p} - \boldsymbol{I}_{r}\|_{2}) D_{r}}{\sum_{r \in N} d(\|\boldsymbol{p} - \boldsymbol{r}\|_{2}) c(\|\boldsymbol{I}_{p} - \boldsymbol{I}_{r}\|_{2})}$$
$$d(\|\boldsymbol{p} - \boldsymbol{r}\|_{2}) = \exp\left(-\frac{\|\boldsymbol{p} - \boldsymbol{r}\|_{2}}{2\sigma_{1}^{2}}\right),$$
$$c(\|\boldsymbol{I}_{p} - \boldsymbol{I}_{r}\|_{2}) = \exp\left(-\frac{\|\boldsymbol{I}_{p} - \boldsymbol{I}_{r}\|_{2}}{2\sigma_{2}^{2}}\right)$$

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、**o**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、実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 **3**(b)に示 すように、実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.誤った奥行き値が補間さ れることを防ぐために、フィルタサイズは制限する(す べての画素に対しての補間処理は行わない).

4.3. 全方位奥行き画像を用いた

全方位自由視点画像の生成

4.2 節で生成した全方位奥行き画像と実画像を用い て,任意視点における全方位自由視点画像を生成する. 図 4 に生成の概要を示す.はじめに,新たに全方位画 像を生成する視点位置を決定し,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, 距離の近い順に規定台数のカメラを選択し,各々の実



図 4 全方位自由視点画像生成の概略図

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3次元点群を生成する.これらの点群 を新たに生成する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が生成される. 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,自由視点から近い方の(手前の)点群を採用す る.

同時に、5節で述べる GAN を用いた画像変換の学 習用データセットとして、実際に撮影した視点位置に おける自由視点画像も同様の処理によって生成する. その際、同視点で生成された全方位奥行き画像は用い ずに自由視点映像を生成することによって、任意視点 における全方位自由視点画像と同等の情報量を有する 学習データを獲得する.

5. GAN を用いた全方位自由視点画像の再構成

4節で生成した自由視点全方位画像に見られる,実 画像には含まれていない欠損領域やアーチファクトを, GANを用いた画像変換によって再構成する.今回,こ の画像変換には Pix2Pix[16]と呼ばれる手法を用いる. Pix2Pixとは,条件付き GANの一種であり,スタイル の異なる2画像間の対応関係を学習し,変換するシス テムである.これまでに,線画画像から写真への変換 や航空写真から地図への変換などに用いられており, 様々な画像変換タスクに適用可能である.我々は,画 像変換タスクとして画像再構成による画質劣化の軽減 を実現する.

具体的には、Pix2Pix は生成器と識別器の二つのネットワークから構成されている.学習データには、変換前画像と変換後画像をペアにした画像群を用意し、変換前画像を生成器に入力する.識別器には、生成された画像もしくは用意した変換後画像を入力し、それが 生成器によって生成された画像かどうかを判断する. 生成器は識別器を欺くような画像生成を、識別器は正



図 5 キューブマッピングによる画像分割

確な判断ができるように互いに競合させながら学習が 進められる.

学習データには、変換前画像として 4.3 節で生成し た撮影視点における全方位自由視点画像を用意し、変 換後画像として撮影視点における実画像を用意する. これらの学習データを用いて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生成 器に入力することで画像再構成を行い、画質劣化を軽 減した写実性の高い画像を生成する.

一方,正距円筒図法によって記録された全方位画像 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,GANの学習に影響を与えると考えられる. 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,正距円筒図射影によって生じる見え方の多様 性を軽減し,効率的な GAN の学習を試みる.本稿で は,図 5に示すように全方位画像を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 像を用いて生成器の構築を行う.画像分割による学習 効果は6節にて考察する.

6. 実験

6.1. 実験環境

以上の手法を実装し,全方位自由視点画像の生成, および画像分割による学習の効果についての実験を行った.撮影は筑波大学内で行い,図 6 に示すように, 42 視点から全方位カメラを用いて撮影を行った.各視 点の高さは三脚を用いて一定にしている.全方位カメ ラには THETA S[21]を使用し,カメラキャリブレーシ ョンと撮影空間の3次元復元を行う SfM には VisualSFM[18]を使用した.Pix2Pix[16]を用いて,全方 位自由視点画像の画質劣化を軽減する生成器を構築す る.生成器を構築する際の学習データには,22 視点分 の実画像と全方位自由視点画像を用意した.画像分割 による GAN の学習の効果を検証するため,生成器の 入力に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された透視投影画 像を用いた場合を比較する. 正距円筒図法による全方 位自由視点画像の解像度は 2018×1024 画素,各透視投 影画像の解像度は 512×512 画素とし,学習回数はそれ ぞれ 1000 epochs とした. 学習後,学習に用いていな い 20 視点分の自由視点画像を生成器に入力し,画像再 構成の結果を観察する. また,本実験の処理には, 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.

6.2. 実験結果

画像分割による学習効果を図 7 に示す. 図 7(a)は4 節で述べた手法で生成した自由視点画像であり、生成 器への入力とした画像である.図 7(b)は分割した画像 を入力とした際の生成画像,図 7(c)は正距円筒図法の 形でそのまま入力した際の生成画像,図 7(d)は実画像 である.図 7(b)(c)より,どちらも画像の欠損領域が GAN によって再構成されていることが確認できる.ま た, それぞれ図 7(d)と比較すると, 分割した画像を入 力とした図 7(b)のほうがアーチファクトの少ない画 像変換がなされていることが確認できる. このことか ら,画像分割によって効率的な学習が行われているこ とが示される.図8に透視投影分割した画像を入力し た際の画像再構成の結果(キューブマッピングにおけ る6面分)を示す.図 8(a) は生成器の入力とした画 像,図 8(b) は学習済み生成器の出力画像,図 8(c) は 真値となる実画像である.図8より、補正前の自由視 点画像でみられるアーチファクトや欠損領域が、画像 変換によって6面分それぞれが補正されていることが 確認できる.



図 7 画像分割による学習効果 (a):入力画像 (b):画像分割した際の生成画像 (c):全方位画像をそのまま 適用した際の生成画像 (d):正解画像(実画像)



図 8 GAN による画像再構成の結果 (a):入力画像 (b):生成画像 (c):正解画像(実画像)

7.おわりに

本研究では、全方位自由視点画像の生成、および画 像再構成の手法について提案した.デプスカメラ等の 特別な機材を使用することなく、全方位カメラのみを 用いた簡易的な撮影によって、全方位自由視点画像の 生成を行った.画像再構成にはGANを用いることで、 アーチファクトや欠損領域の改善が見られた.本研究 は科研費(17H01772)の助成を受けたものである.

文 献

- [1] "Google Street View," https://www.google.com/streetview/
- [2] O. Takeuchi, H. Shishido, Y. Kameda, H. Kim, and I. Kitahara, "Generation Method for Immersive Bullet-Time Video Using an Omnidirectional Camera in VR Platform," Proc. of the 2018 Workshop on Audio-Visual Scene Understanding for Immersive Multimedia, pp.19-26, Seoul, Republic of Korea, Oct.2018.
- [3] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz and R. Szeliski, "Building Rome in a Day," International Conference on Computer Vision, Kyoto, Japan, Sep.2009.
- [4] I. Kitahara, Y. Ohta, "Scalable 3D Representation for 3D Video in a Large-Scale Space," Presence: Teleoperators and Virtual Environments, vol.13, no.2, pp.164-177, Apr.2004.
- [5] T. Kanade, P. Rander, and P. J. Narayanan, "Virtualized reality: Constructing virtual worlds from real scenes," IEEE MultiMedia, vol.4, no.1, pp.34-47, Jan.1997.
- [6] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon, "KinectFusion: Real-time dense surface mapping and tracking," IEEE International Symposium Mixed and Augmented Reality, Oct.2011.
- [7] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, V. Tankovich, C. Loop, Q. Cai, P. Chou, S. Mennicken, J. Valentin, V. Pradeep, S. Wang, S. B. Kang, P. Kohli, Y. Lutchyn, C. Keskin, and S. Izadi, "Holoportation: Virtual 3D Teleportation in Real-time," Proc. of the 29th Annual Symposium on User Interface Software and Technology, pp.741-754, Tokyo, Japan, Oct.2016.
- [8] S. M. Seitz and C. R. Dyer, "View Morphing," Proc. of SIGGRAPH, pp.21-30, 1996.
- [9] M. Levoy, and F. Hanrahan, "Light Field Rendering," Proc. of SIGGRAPH, pp. 31-42, 1996.
- [10] M. Tanimoto, "FTV: Free-viewpoint television," Signal Processing: Image Communication, vol.27, no.6, pp.555–570, Jul.2012.
- [11] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan, "Image-Based Visual Hulls," Proc. of SIGGRAPH, pp.369-374, 2000.
- [12] P. Hedman, T. Ritschel, G. Drettakis, and G. Brostow, "Scalable Inside-out Image-based Rendering," ACM Transactions on Graphics, vol.35, no.6, pp.231:1–231:11, 2016.
- [13] C. Barnes, E. Shechtman, A. Finkelstein, and D. B.

Goldman, "PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing," ACM Transactions on Graphics, vol.28, no.3, pp.24:1–24:11, Aug.2009.

- [14] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context Encoders: Feature Learning by Inpainting," IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, Jun.2016.
- [15] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and Locally Consistent Image Completion," Proc. SIGGRAPH, vol.36, no.4, pp.21-30, 2017.
- [16] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [17] L. Chen, H. Lin, and S. Li, "Depth image enhancement for Kinect using region growing and bilateral filter," Proc. of the 21st International Conference on Pattern Recognition, Tsukuba, Japan, Nov.2012.
- [18] C. Wu, "VisualSFM: A Visual Structure from Motion System," http://ccwu.me/vsfm
- [19] J. L. Schönberger, J. Frahm, "Structure-from-Motion revisited," IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [20] C. Sweeney, T. H Höllerer, M. Turk "Theia: A Fast and Scalable Structure-from-Motion Library," ACM International Conference on Multimedia, Oct.2015.
- [21] "RICOH THETA S," https://theta360.com/ja/about/theta/s.html