3次元画像処理と敵対的生成ネットワークを用いた全方位多視点画像閲覧法

竹内 音[*1] 宍戸 英彦[*1] 亀田 能成[*1] Kim Hansung[*2] 北原 格[*1]

Abstract --- 3次元画像処理と敵対的生成ネットワークを用いた全方位多視点画像閲覧法を提案する．多視点画像の撮影の際，多視点カメラの撮影位置にばらつきがあると，画像（視点）切り替えの際のブレにより視点移動の滑らかさが損なわれ，画像の切り替わりを知覚しやすくなるといった問題が生じる．本稿では，撮影シーンの3次元情報に基づいた自由視点画像生成により，視点切り替え時の見え方のガタつきを軽減し，問題の解決を試みる．3次元形状推定誤差などによって自由視点画像生成時に画質劣化の問題が新たに生じるが，敵対的生成ネットワークを用いて構築した画像生成器を適用することにより，生成画質を改善する．

Keywords: 3次元画像処理，全方位画像，自由視点画像，敵対的生成ネットワーク，画質改善

1 はじめに

多視点で撮影された画像は，様々な方向から撮影空間を観察できるため，単一視点では得られない情報を多面的に獲得することができる．さらに，多視点映像撮影に RICOH THETA[1]をはじめとした，上下左右の空間を一度に記録可能な全方位カメラを用いることにより，Google Street View[2]のように撮影空間に没入した見え方を閲覧者に提供することが可能となる．本稿では，この全方位多視点画像閲覧方式の実現に向けた取り組みについて紹介する．

まず，閲覧者によって異なる関心物体（注視点）に対応可能な全方位多視点画像閲覧方式の実現に取り組んだ．具体的には，全方位多視点画像に3次元情報推定処理（SfM: Structure from Motion）を適用することにより，多視点撮影時に生じる全方位カメラ毎の姿勢（視線方向）の不一致を補正した上で，注目物体の3次元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ことが可能な Bullet-Time 映像生成を実現した[3].

次に，全方位多視点画像の品質を高めるために，カメラの設置位置と間隔について検討する．多視点映像の切替時の見え方のガタつきを考えると，多視点カメラの並び（設置位置の軌跡）は滑らかであることが好ましい．また，切替時に生じる運動視差が大きくなると，画像の切り替わりが知覚されやすくなるため，撮影視点の間隔はなるべく狭い方が好ましい．

これらの条件を十分に考慮した環境で多視点映像を撮影した場合，単に多視点画像を切り替えるだけで高品質映像の閲覧が可能であるが，そのためには三脚等を用いてカメラを均等かつ同一平面上に配置する必要がある．一方で，実際の撮影現場では，構造物の段差や設営時間の制約などによって理想的な状況での撮影が困難である．

そこで本研究では，3次元画像処理によってこれらの課題を解決し，高品質な全方位多視点画像閲覧を実現する手法を提案する．自由視点画像生成技術を活用することで，理想とする視点位置から撮影した全方位画像を合成し，視点切り替え時のガタつきや不連続さを低減する．

2 関連研究

自由視点画像を生成する手法の一つである Depth Image-Based Rendering（DIBR）[4, 5]は，視点依存の奥行き画像と実際に撮影した画像を用いることで，写実的な画像を生成することができる．しかし，生成される自由視点画像は実際に撮影した画像と比較すると，アーチファクトが観測されるなどの画質劣化が確認される．この画質劣化が生じるのは，カメラキャリブレーションの精度や撮影空間の3次元（奥行き）情報の推定精度の誤差によるものである．そのため，完全に3次元空間を復元することは困難である．

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較して重複撮影領域を広く取り，3次元情報の推定に重要な対応点探索の精度の向上が期待できるが，依然として画質劣化は問題となる．デプスカメラ等の奥行き情報が取得可能な装置を用いて，より正確な3次元情報の推定を図る手法[6]もあるが，撮影装置の簡易性が損なわれることが実利用時の課題となる．

本研究では，この画質劣化問題に対して，自由視点画像と実際に撮影した画像を対応付けて学習し，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．閲覧の際には画像を観察することから，

*1 筑波大学
*2 University of Surrey

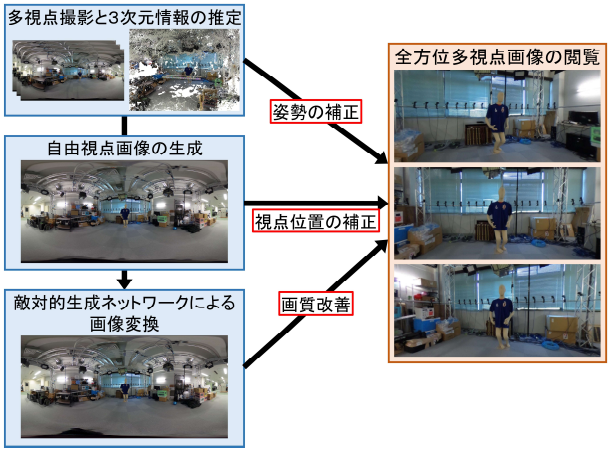

図1　全方位多視点画像閲覧法の概要

Fig.1 Overview of viewing methods for omnidirectional multi-viewpoint images

最終的な提示系である2次元的な見え方の改善を目的とする．画質改善には，近年，画像生成の分野で発展が著しい敵対的生成ネットワーク（GAN: Generative Adversarial Networks）を用いることで，写実的な自由視点画像の生成を行う．

3　全方位多視点画像閲覧法

図1に本論文で提案する全方位多視点画像閲覧法の概要を示す．全方位カメラによって多視点画像を取得し，SfM を使用して全方位多視点画像から各カメラパラメータ，および撮影空間の3次元情報を推定する．各全方位画像に対し，推定されるカメラパラメータを用いて座標変換処理を施すことで，全方位画像の方位を一致させ，撮影時の全方位カメラの回転を補正する．また，VR 環境構築プラットフォームを用いて全方位多視点画像の閲覧システムを構築する．撮影視点の拡張のために，全方位自由視点画像を生成する．撮影空間の3次元情報より各撮影視点における奥行き画像を生成し，新たに設定した視点における全方位画像を DIBR により生成する．全方位自由視点画像の画質改善のために，GAN を用いた深層学習によって，同視点での自由視点画像と撮影画像を対応付けて画像生成器を構築する．この画像生成器を用いて自由視点画像の画質改善を行い，高品質な全方位多視点画像の閲覧を実現する．

4　全方位多視点画像の閲覧

4.1　多視点撮影と3次元情報の取得

全方位カメラを用いて多視点画像を撮影する．三脚等を用いてカメラの高さを揃えることが難しい場合や手持ちでの撮影では，カメラの位置軌跡を滑らかにすることはこんなである．この場合の対応策については，5節以降で述べる．

取得した全方位多視点画像より，各カメラのパラメータ，および撮影空間の3次元情報を取得する．近年，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[7]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[3]．また，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[8]を実行し，密な3次元点群を取得する．

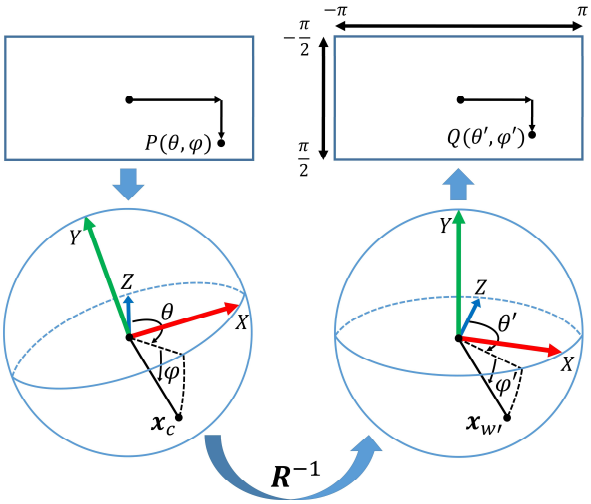

図2　全方位カメラの姿勢情報を用いた全方位画像の回転処理

Fig.2 Rotation processing of omnidirectional images using posture information of omnidirectional cameras

4.2　全方位多視点画像の回転補正

4.1 節で推定したカメラパラメータを用いて，各全方位画像の方位を一致させるように球面座標系における回転処理を施す．具体的には，ワールド座標系における Z 軸方向（正面方向）とカメラ位置から全方位画像の画像中心に向かう軸を一致させ，全方位画像の縦横をワールド座標系における仰俯角と方位角に対応させるようにする．

図2に処理の流れを示す．全方位画像の正距円筒図法における2次元座標を$\mathrm{P}(\theta,\varphi)$としたとき，単位球面上で表記した点Pをカメラ座標系での3次元座標で表すと，$\boldsymbol{x}_c = (\cos\varphi\sin\theta, -\sin\varphi, \cos\varphi\cos\theta)$となる．これを，ワールド座標系での回転軸に合わせた座標系で表したものを$\boldsymbol{x}_{w'} = (x, y, z)$とすると，カメラの回転行列$\boldsymbol{R}$を用いて，

$$\boldsymbol{x}_{w'} = \boldsymbol{R}^{-1}\,\boldsymbol{x}_c \qquad (1)$$

と表される．正距円筒図法における2次元座標を

Q(θ', φ')とすると，θ', φ'は$\boldsymbol{x}_{w'} = (x, y, z)$を用いて，

$$\theta' = sgn(x)\, cos^{-1} \frac{z}{\sqrt{x^2 + z^2}}$$
$$\varphi' = -\, sin^{-1} y \qquad (2)$$
$$\left(sgn(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}\right)$$

と表される．実際には，点Qに対応する画素を逆変換によって求めることで，全方位多視点画像の方位合わせを行う．

4.3 閲覧環境の構築

VR 環境構築プラットフォームを用いて全方位多視点画像の閲覧システムを構築する．VR 空間内に球体モデルを用意し，その内側に 4.2 節で補正した全方位画像をマッピングする．透視投影モデルのバーチャルカメラを球体モデルの中心に配置し，球面をレンダリングすることで，人の見え方に近い透視投影画像を取得することができる．この球体モデルを 4.1 節で推定した全方位カメラの位置に基づいて VR 空間内に配置する．閲覧者に提示するバーチャルカメラを適宜切り替えることで，多視点全方位画像の閲覧が可能となる．

また，4.1 節で取得した3次元点群を用いることで，バレットタイム映像を生成することができる．閲覧画像中の対象の3次元位置を算出し，これを注視点として設定することで，各バーチャルカメラをその光軸が注視点で交わるようにする．これにより，提示するバーチャルカメラを適宜切り替えることで，視点移動しながら注視点を画面中の同一箇所で常に観察することができる．

5 全方位自由視点画像の生成

5.1 奥行き画像の生成

各カメラ視点から 4.1 節で推定した3次元点群までの距離を算出することで，各視点における奥行き情報を推定する．この際，3次元点群の色情報と実際に撮影した画像上で観測される色情報から色差を算出する．3次元情報の誤推定より，画像上で観測されていない点群がある場合には，この色差が大きくなる．色差が閾値よりも大きい場合は奥行き値を算出しないことで，3次元推定誤差の影響を低減する．色差は，CIBLAB 色空間で記述した色情報のユークリッド距離として算出した．生成した全方位奥行き画像を図3(a)に示す．

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3(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[9]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種類のモー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．

(a)

(b)

図3 生成した全方位奥行き画像 (a)：補間処理前 (b)：補間処理後

Fig.3 Generated omnidirectional depth image (a): Before interpolation processing.

$$D_p = \frac{\sum_{\boldsymbol{r} \in N} d(\boldsymbol{p},\boldsymbol{r})\, c(I_{\boldsymbol{p}},I_{\boldsymbol{r}})\, D_{\boldsymbol{r}}}{\sum_{\boldsymbol{r} \in N} d(\boldsymbol{p},\boldsymbol{r})\, c(I_{\boldsymbol{p}},I_{\boldsymbol{r}})},$$
$$d(\boldsymbol{p},\boldsymbol{r}) = exp\left(-\frac{\|\boldsymbol{p}-\boldsymbol{r}\|_2}{2\sigma_1^2}\right), \qquad (3)$$
$$c(I_{\boldsymbol{p}}, I_{\boldsymbol{r}}) = exp\left(-\frac{\|I_{\boldsymbol{p}}-I_{\boldsymbol{r}}\|_2}{2\sigma_2^2}\right),$$

ここで，$\boldsymbol{p}$は注目画素座標，$\boldsymbol{r}$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．$d(\boldsymbol{p},\boldsymbol{r})$と$c(I_{\boldsymbol{p}}, I_{\boldsymbol{r}})$は，それぞれ距離（空間方向）と色差（画素値方向）に関する重みを表す．注目する画素位置と参照する画素位置の間の距離と，撮影画像上での色差で重み付けることにより，奥行き値を計算する．これにより，図3(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の補間を行う．

5.2 全方位自由視点画像の生成

撮影画像と奥行き画像を用いて全方位自由視点画像を生成する．DIBR による生成手法の概要を図4に示す．はじめに，新たに全方位画像を生成する視点を定め，各撮影視点位置までの距離の近い一定数の視点を選択する．選択した各視点における奥行き画像を，新たに定めた視点にワーピングする．このとき，各視点での空間解像度の違いにより生じる小さな穴に対し，メディアンフィルタによるフィルタリング処理を行うことで奥行き値の穴埋めを行う．これらの奥行き値から，元の撮影

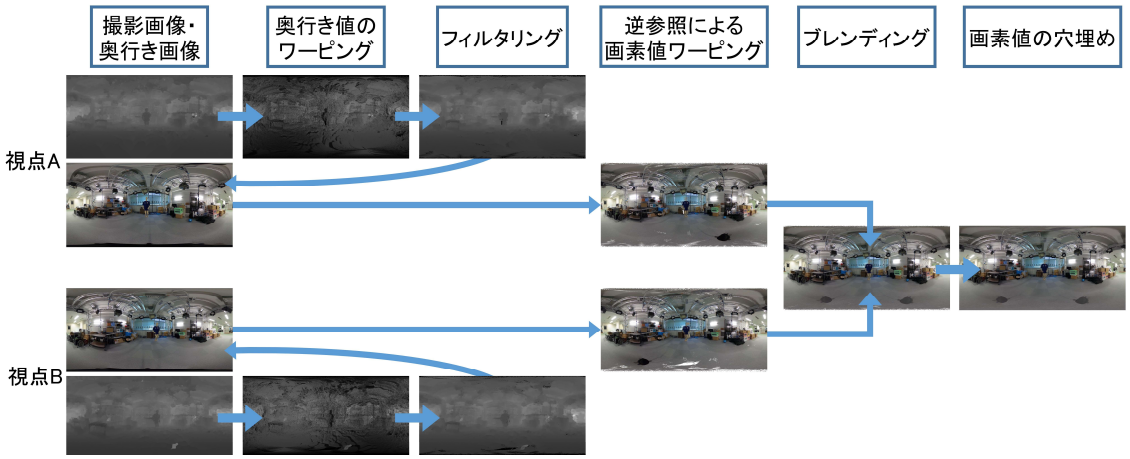

図4　全方位自由視点画像生成の概要

Fig.4 Overview of omnidirectional free-viewpoint image generation

視点における撮影画像を逆参照することで，各視点依存の自由視点画像を取得する．これらの画像を，新規視点までの距離に応じた重み付けによってブレンディングする．最後に，オクルージョン等の影響により画素値が求まらなかった箇所については，周辺の色より穴埋めを行うことで，自由視点画像を生成する．

同様の処理によって，実際に撮影した視点位置における自由視点画像を生成し，次節で述べる GAN による画質改善処理の学習用データセットとする．この際，自由視点画像の生成に使用する視点を任意に選択することで，自由視点画像を複数の生成することができる．これにより，学習データの拡張（data augmentation）を行う．

6 全方位自由視点画像の画質改善

5節で生成した全方位自由視点画像には，3次元情報推定誤差の影響によるアーチファクトが観測され，画質の劣化を引き起こす．本節では，GAN を用いた画像変換を施すことによって，画質改善を行う手法について述べる．本手法では，画像のスタイルを変換する Pix2Pix[10]をベースに，超解像処理技術[11, 12]のアイデアも取り入れることで，画質改善を行う画像生成器を構築する．

学習データとして，5.2 節で生成した自由視点画像（入力画像）と実際に撮影した画像（正解画像）を，視点毎にペアにしたものを用意する．GAN には，相反する目的を持つ画像生成器と画像識別器を用意する．画像生成器には，自由視点画像を入力し，出力画像が正解画像に近づくように学習する．画像識別器には，画像生成器の入出力画像のペア，もしくは画像生成器の入力画像と正解画像のペアを入力し，どちらのペア画像が入力されたのかを判断する．画像生成器は画像識別器を欺くような画像生成を，画像識別器は正確な識別ができるように，互いに競合させながら学習が進められる．学習後，生成した自由視点画像を画像生成器に入力することで，実際に撮影した画像と同等の画質の画像を生成することで画質改善を行う．

6.1 ネットワーク構造

本手法でのネットワーク構造を図5に示す．Pix2Pix で画像生成器に採用されている U-Net[13]は，入力層と出力層を接続する大きなスキップ構造を有するため，入力画像の輪郭を保持した変換を可能とする．一方で，本手法での入力画像は輪郭が誤って再現されている場合がある自由視点画像であるため，このような大きなスキップ構造は適切でない．そこで，本手法では残差を用いたネットワークである ResNet[14]を採用する．ResNet は各層ごとにスキップ構造を有しており，誤情報が深い層まで伝搬されることを防ぐことが期待される．また，Pix2Pix の画像生成器のネットワークには batch normalization の層が含まれている．batch normalization には，学習を速く進めることや過学習を抑制する効果がある一方で，アーチファクトをもたらす傾向がある[12]．そのため，本手法では batch normalization を除去し，アーチファクトの影響の低減を図る．

画像識別器には，Pix2Pix でも用いられている Patch GAN を採用する．これは，画像を分割して画像識別器に入力することで，画像全体ではなく，局所的に真偽判定を行うものである．実際には，画像全体を画像識別器に入力し，最終出力をあるサイズの特徴マップとしたとき，各ピクセルで真偽判定を行うことで代替している．

6.2 キューブマッピングによる画像分割

正距円筒図法に基づく全方位画像は，その射影特性上，画像の上下端に近づくほど横に引き延ばされたような歪みが生じる．このような見え方は，視点移動に伴ってさらに変化するため，学習データ（全方位多視点画像）間での見え方の多様性が大きくなり，学習の際の対応付けが困難になる．また，全方位画像は全天球の空

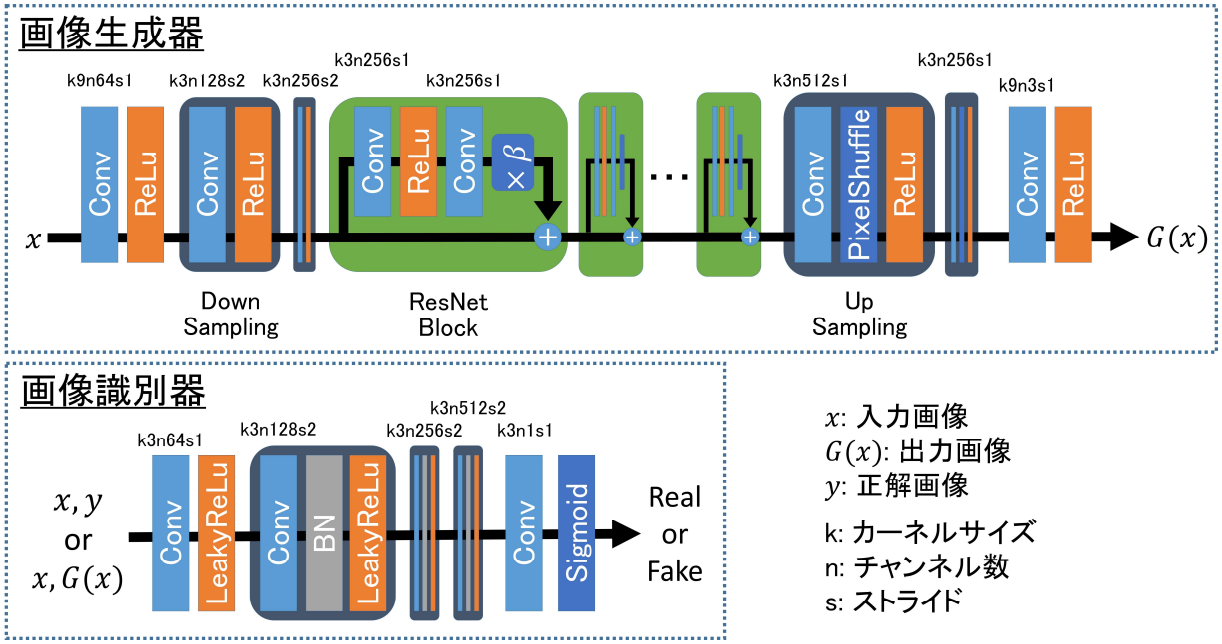

図5　ネットワーク構造

Fig.5 Network architecture

間を記録しているため，画像サイズが大きく，深層学習に適した形ではない．

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，画像サイズを調整することで効率的な 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に対して画像生成器の構築を行う．

7 実験

7.1 実験環境

全方位多視点画像閲覧のための全方位自由視点画像の生成，および敵対的生成ネットワークによる画質改善の効果に関する実証実験を行った．図6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[1]）を取り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 GTX1060，メモリ：16.00GB RAM を搭載したノート PC を用いた．SfM には VisualSFM[7]を使用し，閲覧のための VR 環境構築プラットフォームには Unity[15]を使用した．

5.2 節で述べた手法により，全撮影視点における全方位自由視点画像を生成し，その内 21 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本手法における学習データとする．1視点あたり，4枚の全方位自由視点画像を生成することで，学習データの拡張を行った．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 400 epochs とした．学習

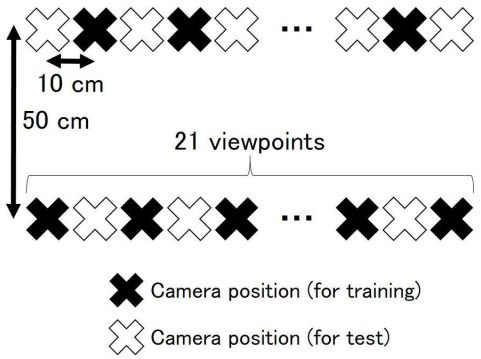

図6　撮影実験での全方位カメラの配置（上から見た図）

Fig.6 Arrangement of omnidirectional cameras in capturing experiments (viewed from above)

後，学習に使用していない評価用の 21 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: Peak signal-to-noise ratio）を用いて画質改善の定量的評価を行う．

7.2 実験結果

実験結果の一例を図7に示す．図7(a)は5節で生成した全方位自由視点画像，図7(b)は6節で提案した画質改善手法を適用した結果，図7(c)は実際に撮影した全方位画像（撮影画像）である．図からも確認できるように，自由視点画像に生じていたアーチファクトが改善され，撮影画像に近い写実的な画像が生成されていることから，本手法で生成した全方位自由視点画像を用いることの有効性が期待できる．

評価用 21 視点分の全方位自由視点画像で算出した

(a) (b) (c)

図7 実験結果の一例(上)とその拡大図(下) (a):全方位自由視点画像(画質改善手法適用前) (b):全方位自由視点画像(画質改善手法適用後) (c):撮影画像(正解画像)

Fig.7 An example of the experimental results (top) and an enlarged view (bottom) (a): Omnidirectional free-viewpoint image (before applying image-quality improvement method) (b): Omnidirectional free-viewpoint image (after applying image-quality improvement method) (c): Captured image (correct image)

PSNR の平均値を用いて，画質改善の効果に関する定量評価を行う．入力画像群の PSNR の平均値(±標準偏差)は，27.87(±1.03)[dB]，出力画像群の PSNR の平均値(±標準偏差)は，30.25(±1.74)[dB]であった．これにより，PSNR において画質改善が達成されたことが確認できる．

8 おわりに

本稿では，3次元画像処理と敵対的生成ネットワークを用いた高品質な全方位多視点画像の閲覧手法を提案した．全方位多視点画像から SfM によりカメラパラメータと撮影空間の3次元情報を推定し，DIBR により自由視点画像を生成することで，視点位置の補正を行った．また，敵対的生成ネットワークによって自由視点画像の画質改善を行い，高品質な画像生成を実現した．

本研究は JSPS 科研費 17H01772 と JST CREST Grant Number JPMJCR14E2 の助成を受けたものである．

参考文献

[1] RICOH THETA; https://theta360.com/ja/ (access: 2019.12.16)
[2] Google Street View; https:/www.google.com/streetview/ (access: 2019.12.16)
[3] 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 AVSU, pp.19-26 (2018)
[4] S. Zinger, L. Do, and P.H.N. de With: Free-viewpoint depth image based rendering; J. Vis. Commun. Image Represent., 21(5-6), pp.533-541 (2010)
[5] W. Sun, L. Xu, O. C. Au, S. H. Chui, and C. W. Kwok: An overview of free viewpoint depth-image-based rendering (DIBR); Proc. of the Second APSIPA ASC, pp.1023-1030 (2010)
[6] P. Hedman, T. Ritschel, G. Drettakis, and G. Brostow: Scalable Inside-out Image-based Rendering; ACM Trans. Graph, 35(6), pp.231:1–231:11 (2016)
[7] C. Wu: VisualSFM: A Visual Structure from Motion System; http://ccwu.me/vsfm (access: 2019.12.16)
[8] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski: A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms; Conf. on CVPR (2006)
[9] L. Chen, H. Lin, and S. Li: Depth image enhancement for Kinect using region growing and bilateral filter; Conf. on ICPR (2012)
[10] P. Isola, J. Zhu, T. Zhou, and A. A. Efros: Image-to-Image Translation with Conditional Adversarial Networks; CVPR (2017)
[11] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network; CVPR (2017)
[12] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao, and X. Tang: ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks; ECCV Workshops (2018)
[13] O. Ronneberger, P. Fischer, and T. Brox: U-net: Convolutional networks for biomedical image segmentation; In MICCAI, vol.9351 pp.234–241 (2015)
[14] K. He, X. Zhang, S. Ren, and J. Sun: Deep residual learning for image recognition; In Conf. on CVPR, pp.770–778 (2016)
[15] Unity; https://unity.com/ja (access: 2019.12.16)