様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　６ 年　 ５ 月 ２７ 日現在

機関番号：３２６２１
研究種目：基盤研究(C)（一般）
研究期間：2021～2023
課題番号：２１Ｋ１１９４３

研究課題名（和文）スナップ写真から全天球画像の生成

研究課題名（英文）omni-directional image generation from snapshot image

研究代表者
山中　高夫（Yamanaka, Takao）

上智大学・理工学部・准教授

研究者番号：２０４３３７９０

交付決定額（研究期間全体）：（直接経費）　　2,500,000円

研究成果の概要（和文）：本研究では，通常のスナップ写真から全天球画像を生成する課題を研究対象とした。全天球画像とは，カメラの全方向を同時に撮影して得られる画像のことで，仮想現実感(VR)や拡張現実感(AR)のコンテンツを作成する際に利用される。全天球画像の撮影には特殊なカメラが必要であり、それが普及の障害になっていると考えられるので、通常の写真から全天球画像を生成する手法の確立を本研究の目的とした。本研究では、生成した全天球画像の多様性を向上する手法、全天球画像生成に適した画像表現方法、事前学習済みVQGANコードブックを利用した高効率で高精度な生成手法を提案し、生成画像の多様性、精細さ、効率性を向上できた。

研究成果の学術的意義や社会的意義
学術的な意義として、画像生成技術の進歩が挙げられる。提案した手法により、通常の写真から生成された全天球画像の多様性、精細さ、効率性を向上できた。同様のアプローチを様々な画像生成タスクに還元できると考えられる。例えば、画像の超解像やInpainting, Outpaintingなどのタスクでほぼ同様の手法を活用できる。また、本研究により全天球画像の応用範囲が広がり、幅広い応用研究につながると考えられる。社会的意義として、仮想現実感や拡張現実感の普及、観光地や文化遺産のデジタル化などに寄与し、特定の場所を訪れていない人にも、その美しさや歴史的価値を伝えることが可能になることが挙げられる。

研究成果の概要（英文）：In this study, we focused on the challenge of generating omni-directional images from regular snapshot photos. Omni-directional images capture the entire surroundings simultaneously, making them useful for creating contents in virtual reality (VR) and augmented reality (AR). Since capturing omni-directional images typically requires a specialized camera, which can be a barrier to widespread adoption, our research aims to establish methods for generating omni-directional images from ordinary photographs. We propose techniques to enhance the diversity of generated omni-directional images, image representations suitable for omni-directional image generation, and an efficient and accurate generation method using the pre-trained VQGAN codebook. These methods improve the diversity, details, and efficiency of the generated images.

研究分野：知覚情報処理

キーワード：コンピュータビジョン　画像生成　全天球画像　360°画像　敵対的生成ネットワーク(GAN)　VQGAN

１．研究開始当初の背景

本研究では、通常のスナップ写真から全天球画像を生成する課題を研究対象とする。全天球画像とは、カメラの全方向を同時に撮影して得られる画像のことで、仮想現実感(VR)や拡張現実感(AR)のコンテンツを作成する際に利用される。全天球画像の活用は徐々に増えつつあるが、通常のカメラで撮影されるスナップ写真に比べると、まだ限られた用途にしか活用されていない。全天球画像の撮影に特殊なデバイス（全天球カメラ）が必要であることが普及の障壁になっていると考えられるので、本研究では手軽に撮影できるスナップ写真から全天球画像を生成する手法を確立することを目的とする。提案する手法の有効性を、様々な応用例を考えたケーススタディで実証する。本研究により、スナップ写真から全天球画像を生成する手法を確立できれば、例えばスマートフォンに搭載されたフロントカメラおよびリアカメラを利用して全天球画像の生成が可能になり、より手軽に全天球画像を撮影できる。このような手法の確立を通して、VR や AR の日常生活への普及に貢献できる。

２．研究の目的

近年、全天球カメラの利用例が徐々に増加している。例えば、全天球カメラを利用して不動産の見学がオンラインでできたり、動画サイトが全天球動画に対応していたりする。また、仮想現実感(VR)や拡張現実感(AR)のコンテンツ作成にも活用できる。しかし、通常のカメラと比べると、その活用はまだ限られており、VR や AR も日常生活に普及しているとは言えない。この原因として、全天球画像を撮影するためには、専用のデバイス（全天球カメラ）が必要であることが挙げられる。スマートフォンやパーソナルコンピュータにはカメラが付属することが多くなっており、それらを活用して全天球画像を生成できれば、もっと手軽に全天球画像を撮影でき、VR や AR の普及に役立つと考えられる。

そのような背景のもと、本申請における研究では、単一もしくは複数のスナップ写真から周りの状況を補間して自然な全天球画像を生成する方法の確立を目的とした。本研究では、「自然な全天球画像」を、実際の全天球画像と区別のつかない画像として定義し、深層学習の一手法である Generative Adversarial Networks (GAN)を利用する。また、従来から我々が提案している手法に加え、様々な手法を検討し、高画質な全天球画像を生成できる方法を確立する。検討すべき事項として、全天球画像特有の歪や連続性、解像度、生成速度、モデルの複雑さなどが挙げられる。また、提案する方法の有効性を、様々な応用例を考えたケーススタディで実証する。例えば、様々なカメラ方向で撮影された単一スナップ写真からの生成や、複数画像からの生成などが挙げられる。

３．研究の方法

(1) MLPMixer を利用して多様性向上を目的とした手法

従来、本研究室で提案した全天球画像生成のモデルは、畳み込みニューラルネットワーク(CNN)をベースとしたモデルを利用してきた[1]。CNN では単一の層における受容野が狭い範囲に限られているので、入力の画像を正距円筒図法の中央に埋め込むと、その情報を全天球画像の端まで伝達するのに多くの層を通過させる必要があるという問題点がある。このような構造では、入力の情報が伝わる前に端の画像を生成し始めるため、どのような入力でも端では同じような画像を生成することが多く、生成された画像の多様性が低くなる傾向があった。そこで、本研究では、図 1 に示すように、畳み込み層の代わりに MLPMixer という構造を利用する手法を検討した[2, 3]。この MLPMixer[4]は、Transformer で利用されている Self-Attention の代わりに使うことができる手法として提案されたもので、Self-Attention と同じように、遠く離れた場所でも情報を一度に伝達することができる。この MLPMixer を全天球画像の生成モデルに使用することで、中央に埋め込んだ入力画像の情報を端まで素早く伝達できるようにした。

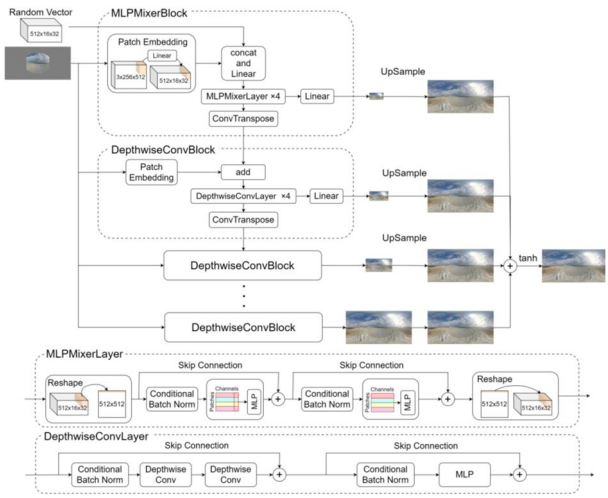

図 1 MLPMixer を利用した全天球画像生成モデル

(2) 全天球画像生成に適した画像表現方法の検討

全天球画像は、通常、地球を世界地図で表現したような正距円筒図法で表現することが多い。全天球画像生成でも、従来、正距円筒図法で生成するモデルが

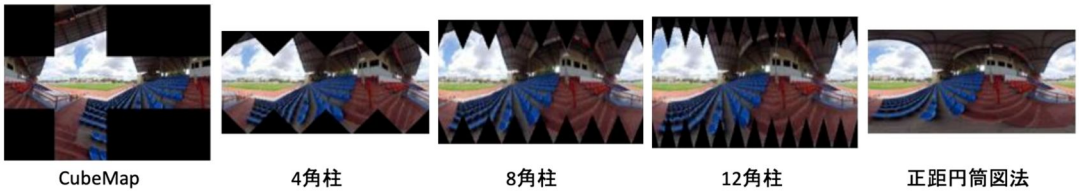

図 2 全天球画像の表現方法[6]

提案されていた[1]。しかし、正距円筒図法では画像の上部や下部（極付近）で幾何学的な歪が生じる。そのため、正距円筒図法で全天球画像を生成するためには、その幾何学的な歪を正確に再現して生成する必要がある。また、極付近の歪が少ない全天球画像の表現方法として CubeMap 手法が知られている。これは、球を立方体の各面に投影した図法で、極方向もその方向の接平面に投影するため、極付近の歪が小さいという特徴がある。そこで、この図法で全天球画像を生成する手法を検討した[5]。しかし、この CubeMap 図法は、各面の中央は歪が小さいが、端では歪が大きくなるという問題があった。そこで、本研究では、図 2 に示すように、球を多角柱の各側面に投影する図法を新しく提案して、その表現方法で全天球画像を生成する手法を検討した[6]。この表現方法は、正距円筒図法と CubeMap 図法の間の表現方法となっており、多角柱の側面の数をハイパーパラメータとして、面の数を多くすると正距円筒図法に近づき、面の数を少なくすると CubeMap 図法に近づく。このとき、4 角柱と CubeMap 図法は同等の表現方法である。多角柱の側面の数により、極付近の歪と各面の端付近の歪のバランスをうまく調整することで、全天球画像生成に最適な表現方法を検討した。

(3) 事前学習済み VQGAN コードブックを利用した手法

これまでの全天球画像生成手法では、Generative Adversarial Networks (GAN)を利用してきた[1-3, 5]が、GAN は学習が不安定になりやすかったり、生成画像の多様性が低くなりやすいといった問題点がある。それに対して、最近、拡散モデルや自己回帰モデルを利用した画像生成手法が多く提案されており、高精細な画像生成が可能になっている。例えば、Stable Diffusion では、画像エンコーダを使って画像を潜在空間における特徴ベクトルで表現し、その潜在空間における拡散モデルで画像生成を行っている[7]。また、画像のパッチをベクトル量子化し、それに基づいて画像を生成する VQGAN[8]を活用して、ランダムにマスクしたパッチを周辺の情報から推論する MaskGIT という手法も提案されている[9]。本研究では、従来使用してきた GAN の代わりに、この MaskGIT を利用して全天球画像を生成する手法を検討した[10]。この MaskGIT に利用されている VQGAN では、画像のパッチをベクトル量子化して表現しており、そのコードブックは通常の画像で学習されているので、正距円筒図法の全天球画像に適用するためには、正距円筒図法特有の幾何学的な歪を表現するために、コードブックを再学習する必要がある。本研究では、学習を効率化するために、VQGAN のコードブックを再学習せずにそのまま利用する手法を検討した。そのために、図 3 に示すように 2 段階構造にして、1 段目では全天球画像全体を低解像度で生成し、2 段階目で様々な方向で接平面に投影した NFoV (Normal Field of View)画像をオーバラップさせながら生成することで、1 段目の低解像度画像を Refinement した。1 段目で生成した低解像度画像を 2 段目の条件画像として使用することにより、全天球画像全体で統一したコンセプトの画像を生成できるようにした。1 段目では、正距円筒図法で生成しているので、全天球画像特有の幾何学的な歪を再現することはできない。しかし、2 段目において NFoV 画像として高解像度の画像を生成するため、事前学習済みの VQGAN コードブックでも正確な画像を生成することができる。このように、コードブックを再学習せずに全天球画像特有の幾何学的な歪を正確に再現できる手法を検討した。

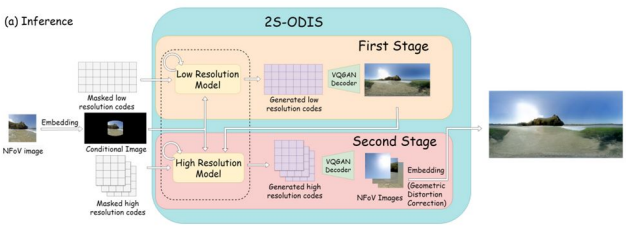

図 3 VQGAN の事前学習コードブックを活用した全天球画像生成モデル[10]

４．研究成果

(1) MLPMixer を利用して多様性向上を目的とした手法

MLPMixer を利用した手法を、CNN を利用した手法と比較した結果を表 1 に示す。評価には 4 つの指標(FID, IS, Accuracy, LPIPS)を使用しており、FID はデータ分布の再現性を、IS は生成された物体のもっともらしさを、Accuracy(認識率)はシーン認識の精度を、LPIPS は生成された画像の多様性を表している。これらの指標は、生成された全天球画像から仰角ごとに NFoV 画像を抽出して算出した。その結果、Accuracy 以外の評価指標では提案手法である MLPMixer の方が優れていることが分かった。特に、LPIPS の比較から MLPMixer により多

様性が向上していることを確認できた。Accuracy が CNN-based の方が高かった理由として、この手法ではシーン情報を条件情報として入力して全天球画像を生成しているので、中央に埋め込んだ入力画像（スナップ写真）に関わらず、同じシーンでは似たような画像を全天球画像の端に生成しているためと考えられる。例えば、図 4 に生成された画像のサンプルを示しているが、同じシーンの 4 枚のスナップ写真に対してそれぞれ生成した結果を並べてみると、MLPMixer(Proposed)に比べて CNN(Baseline)では全天球画像の端で似たようなテクチャーの画像を生成していることが確認できる。このように、MLPMixer を用いた全天球画像生成モデルにより、全天球画像の多様性を向上できることを確認した。この成果は、国際会議 ACPR2023 と国内会議 MIRU2022 において発表した[2, 3]。

(2) 全天球画像生成に適した画像表現方法の検討

全天球画像生成において、多角柱を用いた表現方法と正距円筒図法を比較した結果を表 2 および表 3 に示す。これらの表において、4 角柱は CubeMap に対応する。表 2 において、シーンや物体の情報は水平方向に多く含んでいるので、シーン認識率や物体認識が大きく影響する IS では、水平方向で歪の少ない正距円筒図法で最も良い精度を示した。それに対して、学習データの分布との差を評価している FID では、水平方向と極方向の歪のバランスが取れた 10 角柱で最も良い精度を示した。また、極方向だけで評価すると、表 3 に示すように、極方向で最も歪の少ない 4 角柱で高い精度を示すことが分かった。このようにそれぞれの図法の特徴を反映した生成結果が得られた。これらの成果の一部は、電子情報通信学会の PRMU 研究会で報告した[4]。

表 1 MLPMixer を利用したモデルの定量的評価結果

	Metrics	90°	45°	0°	-45°	-90°	Avarage
MLPMixer-based (Proposed)	FID(↓)	**32.65**	**20.66**	**16.23**	**37.24**	**53.55**	**32.07**
	IS(↑)	**3.92**	**3.50**	**3.77**	**4.28**	**4.61**	**4.02**
	Accuracy(↑)	**26.64**	32.97	49.08	43.23	26.82	35.75
	LPIPS(↑)	**0.648**	**0.605**	**0.622**	**0.657**	**0.724**	**0.651**
CNN-based [1] (Baseline)	FID(↓)	54.84	30.56	21.26	44.58	59.52	42.15
	IS(↑)	3.48	3.07	3.53	3.78	369	3.51
	Accuracy(↑)	24.36	**36.41**	**50.70**	**45.77**	**32.37**	**37.92**
	LPIPS(↑)	0.626	0.591	0.597	0.627	0.674	0.623

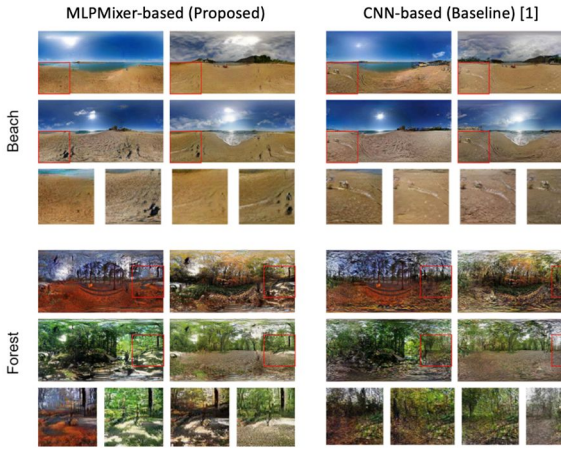

図 4 MLPMixer を利用したモデルで生成したサンプル画像。提案手法では全天球画像の端部分で多様性に富んだ画像を生成できている。

(3) 2S-ODIS: 事前学習済み VQGAN コードブックを利用した手法

事前学習済みの VQGAN コードブックを活用した手法(2S-ODIS)を、従来手法である OmniDreamer [11]、CNN ベースの cGAN による手法[1]、MLPMixer ベースの cGAN による手法[2]、代表的な inpainting 手法である LAMA[12]と比較した結果を表 4 に示す。提案手法は他手法よりも高い評価の全天球画像を生成できることが確認できる。また、生成した全天球画像のサンプルを図 5 に示す。提案手法である 2S-ODIS は、比較対象の OmniDremaer に比べて自然な画像を生成できていることがわかる。2S-ODIS では、1 段目で低解像度の画像を生成し、2 段目でその低解像度画像に基づいて様々な方向で NFoV 画像を生成するので、1 段目の情報から複数の NFoV 画像で連続的な画像を生成することができる。そのため、OmniDremer のように入力画像を埋め込んだ部分と補完している領域で不連続になるような現象が発生しづらいと考えられる。また、提案手法では、学習時の入力画像として、中央に埋め込んだ画像だけを使用するのではなく、様々な部分をマスクした画像を利用しており、1 つのモデルで単一画像からその周辺を補完する outpainting タスクや、全天球画像の特定の領域をマスクしてその部分を周りから補完する inpainting タスクなど、複数

表 2 多角柱表現と正距円筒図法における全天球画像生成結果の比較

形状	認識率(↑)	IS(↑)	FID(↓)
4角柱	21.92%	3.3190	87.0134
8角柱	21.91%	3.3484	84.9514
10角柱	21.74%	3.2816	**84.4998**
12角柱	22.06%	3.3182	89.4978
正距円筒図法	**24.69%**	**3.4298**	88.9866

表 3 極方向(90 °, -90 °)における生成結果の比較

	FID(↓) 90°	-90°
4角柱	**84.621**	**105.373**
8角柱	84.796	107.433
10角柱	90.392	109.412
12角柱	92.847	117.845
正距円筒図法	87.291	117.511

のタスクに対応することができる。例えば、図 6 に様々なタスクに対する生成した画像のサンプルを示す。これらのサンプルでは、モデルの再学習は行っておらず、図 5 と同じ重みパラメータのモデルを使用して生成している。従来手法である OmniDreamer はこのようなタスクに対応することができないが、提案手法では高精細な全天球画像を生成できている。このような結果から、提案手法では、例えば、2 枚の通常の写真から全天球画像を生成したり、撮影した全天球画像から持ち手を消去したり、特定の物体を消去したりすることが可能である。これらの成果は、現在、国際会議に投稿中である。

<引用文献>

[1] Keisuke Okubo and Takao Yamanaka, “Omni-Directional Image Generation from Single Snapshot Image,” SMC2020.

[2] Atsuya Nakata, Ryuto Miyazaki, and Takao Yamanaka, "Increasing diversity of omni-directional images generated from single image using cGAN based on MLPMixer," ACPR2023.

[3] 中田敦也, 山中高夫, MLPMixer を用いた全天球画像生成, MIRU2022.

[4] Ilya Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Andreas Steiner, Daniel Keysers, Jakob Uszkoreit, Mario Lucic, Alexey Dosovitskiy, "MLP-Mixer: An all-MLP Architecture for Vision," NeurIPS2021.

[5] Keisuke Okubo and Takao Yamanaka, "Omni-Directional Image Representation in GAN-based Image Generator," 電子情報通信学会 PRMU 研究会，オンライン，Oct. 2021.

[6] 宮﨑龍斗、多角柱表現に基づいた全天球画像生成、上智大学修士論文、2023.

[7] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," CVPR2022.

[8] Patrick Esser, Robin Rombach, Björn Ommer, Taming Transformers for High-Resolution Image Synthesis, CVPR2021.

[9] Huiwen Chang, Han Zhang, Lu Jiang, Ce Liu, William T. Freeman, "MaskGIT: Masked Generative Image Transformer," CVPR2022.

[10] 中田敦也、2S-ODIS: 幾何学的な歪補正による 2 ステージ全天球画像生成、上智大学修士論文、2023.

[11] Naofumi Akimoto, Yuhi Matsuo, Yoshimitsu Aoki, "Diverse Plausible 360-Degree Image Outpainting for Efficient 3DCG Background Creation," CVPR2022.

[12] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, Victor Lempitsky, "Resolution-robust Large Mask Inpainting with Fourier Convolutions," WACV2022.

表 4 VQGAN の事前学習コードブックを活用した全天球画像生成モデルの評価結果

Method	IS(↑)	FID(↓)	LPIPS(↑)
2S-ODIS (Proposed)	**5.969**	**18.263**	**0.662**
OmniDreamer [11]	4.458	23.101	0.655
CNN-based cGAN [1]	4.684	40.049	0.633
MLPMixer-based cGAN [2]	4.402	47.690	0.634
LaMa [12]	5.784	69.485	0.478

図 5 VQGAN の事前学習コードブックを活用した全天球画像生成モデルの生成画像例

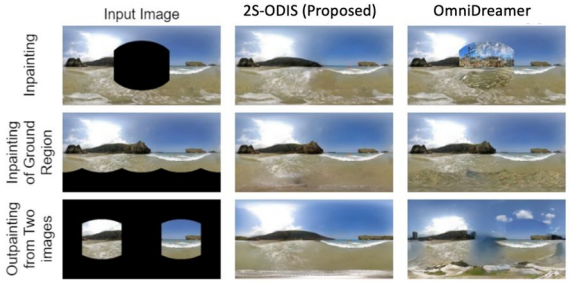

図 6 様々なタスクに適用した例

５．主な発表論文等

〔雑誌論文〕　計0件

〔学会発表〕　計4件（うち招待講演　0件／うち国際学会　1件）

１．発表者名 宮崎龍斗，田畠誠大，山中高夫
２．発表標題 階層型全天球画像生成モデル
３．学会等名 電子情報通信学会PRMU研究会
４．発表年 2022年

１．発表者名 中田敦也，山中高夫
２．発表標題 MLPMixerを用いた全天球画像生成
３．学会等名 第25回画像の認識・理解シンポジウム
４．発表年 2022年

１．発表者名 Keisuke Okubo and Takao Yamanaka
２．発表標題 Omni-Directional Image Representation in GAN-based Image Generator
３．学会等名 電子情報通信学会PRMU研究会
４．発表年 2021年

１．発表者名 Atsuya Nakata, Ryuto Miyazaki, and Takao Yamanaka
２．発表標題 Increasing diversity of omni-directional images generated from single image using cGAN based on MLPMixer
３．学会等名 Asian Conference on Pattern Recognition（国際学会）
４．発表年 2023年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

研究室ウェブページ https://scrapbox.io/islab-sophia/Resarch ソースコード https://github.com/islab-sophia

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関