H-025

スマホカメラを用いた隠消現実感アプリケーション
Diminished Reality Application using a Smartphone Camera

小林 海人† 高橋 正信†
Kaito Kobayashi Masanobu Takahashi

1. はじめに

隠消現実感(Diminished Reality 以下 DR)とは，現実空間から物体が削除されたように見せる技術である．拡張現実(Augmented Reality 以下 AR)では現実空間と仮想物体を重ねて表示して，仮想物体が現実空間に存在するように見せるのに対して，DR では現実物体に背景画像を上書きして表示することにより，物体があたかも削除されたかのように見せる．AR アプリケーションを利用する際に，現実物体が邪魔でうまく仮想物体を配置できない場合があるという問題点を DR により障害物を削除することで解決できる．また，景観シミュレーションや壁の透過，過去の景色の再現などにも DR は応用されている[1].

我々は，スマホのカメラを用いた DR アプリケーション[2]を実現したが，背景が単純な床面に限定されるなど原理実証の段階であった．そこで，ネットワーク構成の改善などでリアルタイム性を損なうことなく背景画像の精度を改善し，より実用的な DR アプリケーションを実現したので報告する．

2. 手法

2.1 手法の概要

DR アプリケーションは，スマートフォンと PC を接続した状態で使用する．スマートフォンのカメラから取得した画像に対し，PC 上で DR の処理を行う．DR アプリケーションの処理手順は次のとおりである(図 1).

(1) スマホカメラの撮影画像中で削除対象が配置されている平面を検出し，検出した平面と重なるように仮想空間内に平面を構築する．
(2) 削除対象と重なるように仮想円柱を配置する．配置位置や仮想円柱の大きさはユーザが指定する．仮想円柱内を，DR を行う領域(以下補完領域)とする．
(3) 仮想円柱の 8 点について，その 3 次元座標とカメラパラメータから画面上での各点の座標を算出し，それらの外接長方形を求めて 2 次元的な補完領域とする．
(4) カメラ画像の 2 次元補完領域をデータセットの平均画素値で塗りつぶし，補完ネットワークの入力画像を作成する．
(5) 補完ネットワークを用いて，補完領域を補完する画像(以下補完画像)を生成する．
(6) カメラ画像と補完画像を合成し，DR が行われたカメラ画像を作成する．
(7) 元のカメラ画像の代わりに，DR が行われたカメラ画像を表示する．
(8) (3)~(7)を繰り返す．

なお，平面検出，仮想平面の構築，仮想空間の原点固定やカメラ位置・回転の推定など AR の基本処理は，スマートフォン AR 向けの SDK である ARCore を用いて実現した．

† 芝浦工業大学 Shibaura Institute of Technology

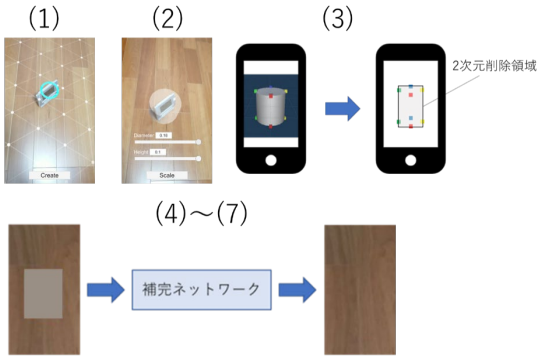

図 1 処理の概要

2.2 画像補完

処理(5)の画像補完には深層学習を用いており，補完ネットワークは GAN に基づいた生成器，局所識別器，大域識別器の 3 つのネットワークから構成される．生成器は補完領域が補完された画像を生成する．局所識別器は生成器の生成した画像の補完領域を中心とする小領域を対象に，大域識別器は画像全体を対象にして，元画像か補完画像かの識別を行う．局所識別器と大域識別器の出力結果は統合され，識別器全体としての識別結果を出力する．

2.3 Channel Attention を採用したネットワーク

ネットワーク構成の改善策の一つとして，Channel Attentionを採用した．Channel Attentionは，どの特徴量チャネルが重要かを重み付けする．これを行うことで，重要な特徴を強調し，そうでないものを抑え，表現の質を向上させる．Channel Attentionは，画像修復の分野でも用いられている[4]．本研究では，従来のネットワーク[3]をベースに Channel Attentionを生成器と2つの識別器にそれぞれ3層ずつ組み込んだ(図2).

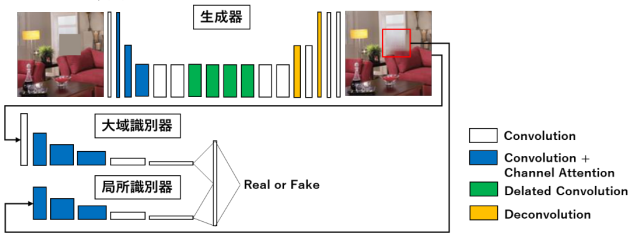

図 2 Channel Attention を採用したネットワーク

2.4 Contextual Attention を採用したネットワーク

また，Channel Attention とは別の改善策として Contextual Attention[5]を採用した．Contextual Attention は，対象領域内の各ピクセルが領域外のどこのピクセルと相関があるかを学習し，画像の再構成に活用する．ただし，Contextual Attention は，補完領域内外で相関性がない場合には有効でない．そのため，補完領域を一度大まかに補完してからもう一度 Contextual Attention を用いてより精度の高い補完を施す構成にした(図 3).

図 3　Contextual Attention を採用したネットワーク

2.5 両方を採用したネットワーク

Channel Attention と Contextual Attention の両方を用いたネットワークも作成した．具体的には，図 3 における 2 つの生成器と 2 つの識別器について，図 2 と同様にそれぞれ 3 層ずつ Channel Attention を追加した．

3. 実験

3.1 ネットワークの学習

以上の 3 つのネットワークと従来のネットワークについて，補完精度と処理速度を比較する実験を行った．データセットは，従来[2]は床面画像を利用したが，本研究では多様な室内画像[6][7][8][9]を利用した．室内画像は画素数を 128×128 画素に変換し，学習用に約 16,000 枚，評価用に約 1,800 枚用意した．

ネットワークの学習では，学習用画像中にランダムな位置に(16～64)×(16～64)のランダムな形，大きさの四角形の補完領域を作成した画像を入力画像とし，補完領域の元画像を正しい結果として学習した．また，最初から GAN での学習を行うと生成器と識別器の均衡が保てず，学習が上手くいかないため，事前に生成器，識別器それぞれ単独である程度学習させてから GAN での学習を行った．

3.2 評価方法

評価用画像を用いて，生成した補完画像の補完領域の精度を評価した．精度は補完領域の大きさに依存するため，評価時の補完領域は正方形として，13 種類の大きさ(16×16，20×20，…，60×60，64×64)のそれぞれで精度を評価した．評価指標には，PSNR(Peak Signal-to-Noise Ratio)と SSIM(Structural Similarity)を用いた．また，リアルタイム性の評価のため，補完画像の生成速度[fps]を求めた．実行環境は，スマートフォン: Google Pixel 4，CPU: AMD Ryzen 9 3900，GPU: NVIDIA GeForce RTX 3080 である．

3.3 結果と考察

図 4　各ネットワークの画像補完結果例

図 4 は，元の画像と，入力画像，および各ネットワークで補完した画像例である．7 種類の補完領域サイズについての評価用画像に対する従来のネットワークの補完精度を表 1 に，従来からの補完精度の改善幅を図 5，図 6 に示す．新たに導入した 3 つの手法の評価値はいずれも従来よりも優れていた．3 つの手法の中では，補完領域が小さいと Contextual Attention を採用したネットワークと両方を採用したネットワークが，補完領域が大きいと Channel Attention を採用したネットワークが優れていた．Contextual Attention は領域が大きいとき，ピクセル間の相関付けがうまく学習できていないことが考えられる．

表 1　従来のネットワークの補完精度

	16×16	24×24	32×32	40×40	48×48	56×56	64×64
PSNR[dB]	25.215	23.422	22.072	20.797	20.079	19.159	18.569
SSIM	0.5820	0.5532	0.5265	0.5028	0.4902	0.4748	0.4613

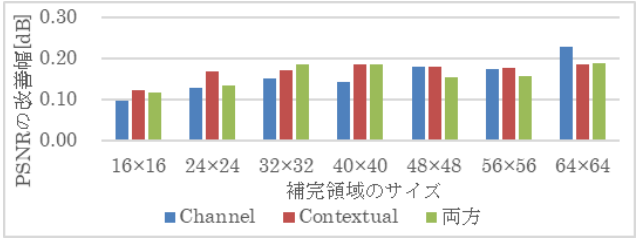

図 5　補完精度の改善幅(PSNR)

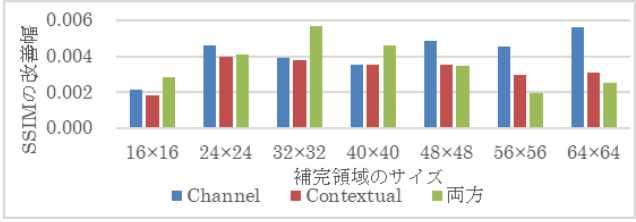

図 6　補完精度の改善幅(SSIM)

補完画像の生成速度は従来が 22.8[fps]，Channel Attention を採用した場合が 23.1[fps]，Contextual Attention を採用した場合が 19.4[fps]，両方を採用した場合が 19.5[fps]であった．生成速度の面では Channel Attention を採用した場合が従来と同等以上に速かった．Contextual Attention を採用した場合は生成器を 2 段階で用いるため速度が少し遅くなったと考える．従って，総合的には Channel Attention を採用したネットワークが DR アプリケーションに適している．

4. おわりに

ネットワークと学習画像を改善することで，背景の制約を緩和するとともに，リアルタイム性を損なうことなく生成される背景画像の精度を改善し，より実用的な DR アプリケーションを実現した．ネットワークとしては，補完精度，補完速度の面から Channel Attention を採用したネットワークが DR アプリケーションに適していると考える．

今後は，補完機能の改善と DR アプリケーションの利便性の向上に取り組みたい．

参考文献

[1] J. Valentin, et al., “Depth from Motion for Smartphone AR”, ACM Transactions on Graphics, Vol. 37, No. 6, Article 193, 2018.
[2] 澤田悠暉，他，“敵対的生成ネットワークを用いた隠消現実感”, 2019 年度電子情報通信学会東京支部学生会研究発表会，60, 2020.
[3] S. Iizuka, et al., “Globally and Locally Consistent Image Completion”, ACM Transactions on Graphics, Vol. 36, No. 4, pp.1-14, 2017.
[4] Y. Chen, et al., “The improved image inpainting algorithm via encoder and similarity constraint”, International Journal of Computer Graphics, The Visual Computer, Vol.37, issue.7, pp.1691-1705, 2021.
[5] J. Yu, Z. Lin, et al., “Generative image inpainting with contextual attention” arXiv:1801.07892, 2018.
[6] A. Quattoni, et al., “Recognizing Indoor Scenes” IEEE CVPR 2009, pp.413-420, 2009.
[7] “Pixabay”, <https://pixabay.com/ja/>, 2022 年 6 月 22 日アクセス.
[8] “Unsplash”, <https://unsplash.com>, 2022 年 6 月 22 日アクセス.
[9] “GAHAG”, <https://gahag.net>, 2022 年 6 月 22 日アクセス.