



# 空間拡張現実感を用いた現実物体の仮想化

Exploring the Virtualization of Real-World Objects through Spatial Augmented Reality

川島映輝<sup>1)</sup>, 天野敏之<sup>2)</sup>

Eiki KAWASHIMA, and Toshiyuki AMANO

1) 和歌山大学院 システム工学研究科 (〒 642-8441 和歌山県和歌山市栄谷 930, s266084@wakayama-u.ac.jp)

2) 和歌山大学院 システム工学研究科 (〒 642-8441 和歌山県和歌山市栄谷 930, amano@wakayama-u.ac.jp)

**概要:** 空間拡張現実感の研究として, 現実世界で物体の見目の色彩や質感を変化させる技術が提案されている. そこで, 本研究では物体の外観を操作し, 絵画の世界に同化するというコンセプトで, 現実物体の仮想化を行った. StyTr<sup>2</sup>, Texture Overlay, Average Color の 3 つの手法により仮想化を行い被験者実験により評価した. また, StyTr<sup>2</sup> による投影の定量的評価を行った.

**キーワード:** SAR, 仮想化, StyTr<sup>2</sup>

## 1. はじめに

近年, 仮想空間と現実世界の融合を目指す「Real-World Metaverse」が注目されている. これを実現するには, 現実物体の上に仮想情報を重ねるだけでなく, 現実物体の外観を知覚的に仮想物体へと変換する「現実物体の仮想化」も重要であると思われる. そこで, 本研究では, 現実物体の仮想化として, プロジェクタとカメラから構成される空間拡張現実感 (SAR) に, Transformer ベースのスタイル変換手法を用いることで, 現実物体が絵画の世界に同化する光学操作を試みる.

また, 本研究では仮想化を「物体を認識でき, 物体が仮想空間の色彩・模様に変化すること」と定義する.

## 2. 物体の外観操作

SAR による物体の外観操作の研究として, 劣化した文化財や芸術作品の視覚的復元 [4], アバターへのシワの投影による現実感の向上 [5] などが挙げられる. このような技術は現実物体の外観そのものを意図的に置換・再構成する点に特徴がある. また, プロジェクタカメラ系を用いて現実物体の色彩や質感を動的に変化させる技術として, 同軸光学系とプロジェクタ・カメラシステムを組み合わせたリアルタイム光沢制御手法が提案されている. [1]

これら一連の技術は, 投影を通じた物理対象物の外観を操作する能力を提供し, 現実世界の対象物の仮想化のための強力な基盤を築いている.

## 3. 提案手法

### 3.1 SAR を用いた仮想化

SAR を用いた現実物体の仮想化を実現するためには, 現実物体の物体像を与えられた仮想空間の色彩や質感に変換する目標像生成と, その目標像に光学的に変化させる映像投影技術が必要となる.

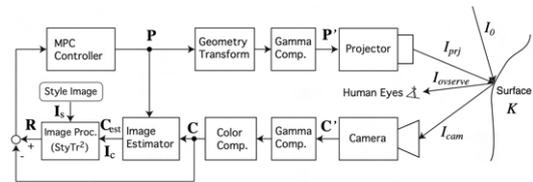


図 1: StyTr<sup>2</sup> を用いた見かけの制御のブロック図

そこで, 本研究では, 絵画の中の世界に没入する仮想化を目標とし, 絵画をスタイル画像として StyleTransformer によって仮想化の目標画像を作成する. 目標画像生成には, Image Style Transfer の一つである, StyTr<sup>2</sup> を, 映像投影技術には, プロジェクタとカメラのフィードバックによって見目を変化させる見かけの制御を用いる.

StyTr<sup>2</sup> (Image Style Transfer with Transformers) [2] は, スケール不変位置符号化と, コンテンツとスタイルの表現分離を統合したスタイル変換手法であり, 入力画像との幾何学的整合性を保つことを可能である. SAR で投影画像を実物体と整合させるためには, 生成された目標画像が実物体と空間的一貫している必要があるため, 本研究では StyTr<sup>2</sup> を採用した.

見かけの制御は, プロジェクタとカメラを用いた光学的なフィードバック系である. 見かけの制御では, 制御系内部で推定された白色照明下での見かけを基準とし, これにユーザの要求に応じた画像処理より生成された画像を制御目標とし, モデル予測制御によって光学的に物体の外観を変化させる.

StyTr<sup>2</sup> を用いた見かけの制御の流れとして, 図 1 に示すように, 見かけの制御の ImageProc. は絵画画像をスタイル画像  $I_s$ , 白色照明下での見かけ  $C_{est}$  を入力コンテンツ画像  $I_c$  として StyTr<sup>2</sup> に入力し, これにより生成された画像を目標画像  $R$  とする. ただし, StyTr<sup>2</sup> によるスタイル変換を

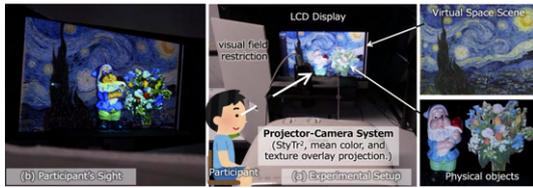


図 2: 実験環境

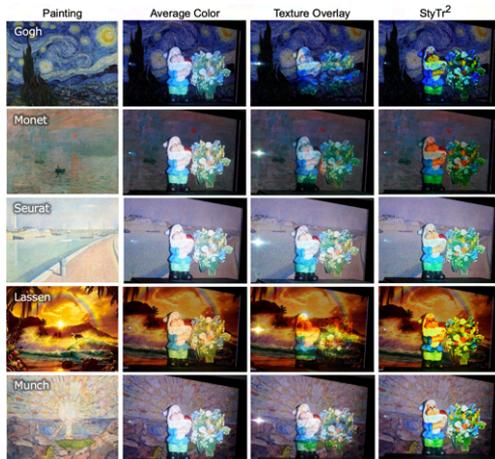


図 3: 投影結果

実時間で行うことは困難であるため、本研究の仮想化では静的なシーンを対象とし、予め生成した  $R$  を用いてフィードバックを行う。

### 3.2 実験環境

仮想化の効果を評価するため、仮想世界は実験環境の後方に配置された 27 インチ LCD モニター (iMac Retina 5K, 27 インチ, 5120 × 2880 解像度) により仮想世界を表す絵画を表示することで仮想空間を擬似的に再現する。物体はディスプレイの前方に配置する。

プロジェクタカメラ系には、フル HD プロジェクタ (EPSON EH-TW6600, 1920 × 1080 解像度, 2500 ルーメン) とカラーカメラ (Point Grey Grasshopper2, 1600 × 1000 解像度 ROI) を採用し、ビームスプリッタによって光学的に同軸の位置に配置した。投影は、対象物体の表面から約 90cm の距離から行われている。参加者の視点は、シーンの前方約 70cm に固定した (図 2 参照)。

本研究では、投影対象として、外観操作が容易な拡散反射を持つ物体とし、物体形状における汎用性を評価するため、立体と平面物体の両方を採用した。

### 3.3 投影結果

多様な擬似的な絵画の世界を作るために、ヴァンセント・ヴァン・ゴッホ、クロード・モネ、ジョルジュ・スーラ、クリスチャン・ラッセン、エドヴァルド・ムンクの 5 つの代表的な芸術作品を用いた。これらの絵画は、印象派や点描画から超現実主義や表現主義まで、多様な画風を表現しており、異なる投影方法が画風とどのように相互作用するかを評価する上で役立つ。図 3 は、各仮想空間内の物体に 3 つの投影方法 (Average Color, Texture Overlay, StyTr<sup>2</sup>) を適

用した結果を示す。

Texture Overlay はスタイル画像  $I_s$  をそのまま投影画像  $P$  とする方式 (すなわち  $P = I_s$ ) である。一方、Average Color は、スタイル画像  $I_s$  の平均色  $\bar{I}_s$  を算出し、それを投影画像  $P$  の全ピクセルに一律に適用する方法である。

### 3.4 評価手順

本研究では投影による物体の仮想化の操作を評価するため、被験者実験を実施した。被験者は仮想化のシーンを観察し、物体が絵画の仮想空間に没入しているかを評価した。サーストンの一対比較法に従い、3 つの手法 (Average Color, Texture Overlay, StyTr<sup>2</sup>) から選択された 2 つの手法を順次投影し、特定の基準をより満たす方を選択してもらった。各ペアに対して 3 つの質問を行った：

- Q1: 物体輪郭が明瞭に見えるのはどちらか
- Q2: 物体と背景のタッチに差異がないのはどちらか
- Q3: 物体と背景の配色に差異がないのはどちらか

これらの 3 つの質問は、仮想化の定義に基づいた質問である。Q1 は、物体の認識における質問であり、物体輪郭の明瞭さは物体を認識する上で重要な要素であると考えた。Q2 は、絵画の表面テクスチャや筆触が仮想化における模様該当すると考えた。Q3 は、対象物の配色と絵画の配色に差異があるかを評価した。前節で示した 5 つの絵画をスタイル画像として使用した。各試行では、選択された絵画を物体の背後の LCD ディスプレイに表示した。ディスプレイと覗き穴の間に白い遮蔽板を配置し、視認性を制御し残像を排除した。実験手順には以下の通りである：

1. 遮蔽板を配置した状態で選択された絵画を表示。
2. 投影 a を物体に投影。
3. 遮蔽板を外し、参加者にシーンを 15 秒間観察させる。
4. 遮蔽板を再配置してシーンを遮蔽。
5. 投影 b を物体に投影。
6. 再び手順 3 を行う
7. 参加者に Q1~Q3 の回答をしてもらう。

遮蔽方法は、Zhong ら [3] が提案した方法を参考にした。Zhong らの方法では、視覚的残像を軽減するため、観察の間には中立的な輝度平面を使用した。

今回の実験では、白色の遮蔽版で遮蔽することで視覚的残像を軽減した。

## 4. 実験結果

実験には 15 名 (女性 1 名, 男性 14 名; 年齢範囲: 21~25 歳, 平均 = 22.1, 標準偏差 = 1.33) に参加してもらった。

各投影方法がどのように仮想化に寄与するかを評価するため、5 つの絵画と 3 つの評価質問 (Q1: 輪郭の明瞭さ, Q2: 絵画のタッチの統合, Q3: 配色の一致) について、絵画別の各方法のスコアを全体的な性能分布を可視化した。各投影方法の Q1-Q2 と Q1-Q3 の組み合わせを 2 次元散布図に

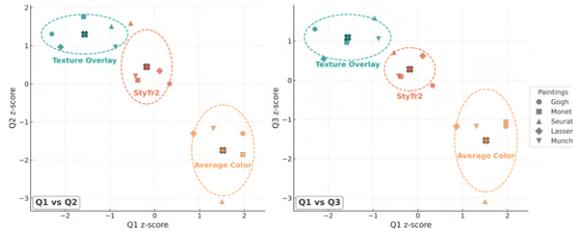


図 4: Q1-Q2, Q1-Q3 の投影別スコアグラフ

プロットし (図 5), プロット内の楕円は, 各絵画のスコア分布の 3  $\sigma$  信頼領域を表す.

図 5 に示すように, Average Color は左上領域に集まり, Q1 での強さを示す一方, Q2 と Q3 では非常に低い性能を示している. 一方, Texture Overlay は右下に位置し, 絵のタッチと配色の一致において優れているが, 輪郭の保持に弱点がある. StyTr<sup>2</sup> は中央に位置し, 最も小さな散布度であり, すべての次元で一貫した中程度の性能を示している.

よって, Texture Overlay と Average Color に比べて, StyTr<sup>2</sup> が 3 つの評価軸すべてにおいて, いずれにおいてもバランスの取れた方法である.

## 5. StyTr<sup>2</sup> の投影に対する定量評価

ここでは, StyTr<sup>2</sup> の投影によって, 物体上に絵画の特徴が付与されているか定量的に評価を行う. 具体的には,

- 目標画像生成に対する評価
- 目標画像とキャプチャ画像の差による評価

の 2 段階に分けて評価を行う.

### 5.1 目標画像生成に対する評価

StyTr<sup>2</sup> では, 画像のスタイル変換の損失関数として, Content loss

$$\mathcal{L}_c = \frac{1}{N_l} \sum_{i=0}^{N_l} \|\phi_i(R) - \phi_i(I_c)\|_2, \quad (1)$$

と Style loss

$$\mathcal{L}_s = \frac{1}{N_l} \sum_{i=0}^{N_l} \|\mu(\phi_i(R)) - \mu(\phi_i(I_s))\|_2 + \|\sigma(\phi_i(R)) - \sigma(\phi_i(I_s))\|_2, \quad (2)$$

が用いられている. ここで,  $\phi_i(\cdot)$  は事前訓練済み VGG19 の  $i$  番目の層から抽出された特徴を表し,  $N_l$  は層の数である. また,  $\mu(\cdot)$  と  $\sigma(\cdot)$  は抽出された特徴の平均と分散を表す.

これらは, 絵画の画風に変換する目標画像生成において, Content loss は物体像との類似性, Style loss は絵画のタッチの再現性を評価する指標と考えることができる.

そこで, 物体に対して, 元のコンテンツを維持しているか, スタイル画像の特徴が適切に反映されているかを検証するため, 物体ごとに Content loss と Style loss を用いて評価した.

表 1: Style loss と Content loss の結果

(a) Style loss			(b) Content loss		
No.	平面	立体	No.	平面	立体
Gogh	2.00	1.79	Gogh	1.57	1.87
Monet	0.38	0.53	Monet	1.22	1.16
Seurat	0.82	0.80	Seurat	1.53	1.65
Lassen	1.83	2.46	Lassen	1.45	1.72
Munch	1.10	1.38	Munch	1.50	1.81

表 2: アンケート調査による StyTr<sup>2</sup> のスコア

No.	Q1 のスコア	Q2 のスコア	Q3 のスコア
Gogh	0.33	0	-0.135
Monet	-0.37	0.09	0.09
Seurat	-0.53	1.59	0.70
Lassen	0.10	0.34	0.62
Munch	-0.429	0.20	0.10

表 1(a) より, Style loss の値は Gogh (平面: 2.00、立体: 1.79) および Lassen (平面: 1.83、立体: 2.46) が他作品に比べて顕著に高く, 特に Monet (平面: 0.38、立体: 0.53) や Seurat (平面: 0.82、立体: 0.80) との差は平面・立体ともに 1.0~2.0 程度と大きい. このことから, Gogh と Lassen はスタイルが他の絵画に比べて付与されていないと考えられる.

表 1(b) より, Content loss の値は Monet (平面: 1.22、立体: 1.16) のみが他手法と比べて明確に低いことが確認できる. 一方で, Gogh (1.57 / 1.87), Seurat (1.53 / 1.65), Lassen (1.45 / 1.72), Munch (1.50 / 1.81) はいずれも平面・立体ともに 1.4~1.8 程度の値に集中しており, 内容保持の観点では大きな差は見られない. このことから, Monet のスタイル変換は他の手法に比べて, コンテンツの構造的特徴をより良く保持していると考えられる.

Content loss は物体像との類似性, Style loss は絵画のタッチの再現性を表しているかについて, アンケートのスコアと比べる. 表 2 の Q1 のスコアを見ると, Gogh と Lassen の値が他作品の値より高い数値である. Content loss ではモネだけが他の絵画に比べて低い値を取っていた.

また, 表 2 の Q2 のスコアを見ると, Seurat が 1.59, Lassen と Munch が 0.34 と 0.20 と値が近く, Monet が 0.09, Gogh が 0 となっている. Style loss では, Gogh と Lassen が顕著に高く Seurat と Monet が低くなっており, Munch が中間の値を取っていた.

よって, 定量的評価指標 (loss の評価) と主観的評価 (被験者実験) の間で, 一貫性が見られないことが確認できた. 主観的評価は仮想化の定義に従った評価となっているので,

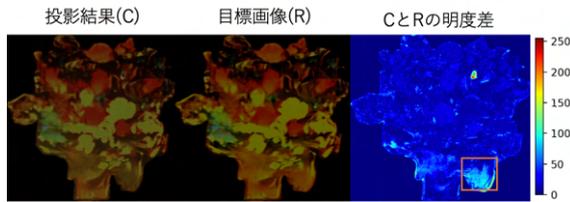


図 5: Lassen 平面物体の目標とキャプチャ画像の明度差

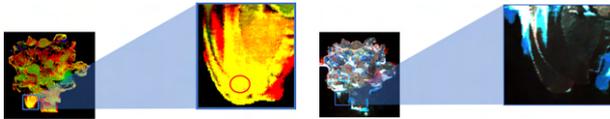


図 6: Lassen の投影画像

一貫性が見られなかったということは loss は仮想化の定義に従っていないと考えられる。

5.2 目標画像とキャプチャ画像の差による評価

それぞれ平面物体と立体物に分けて、目標通りの操作が行われているかどうかを確認するために、目標画像とキャプチャ画像について HSV 変換を行い、色彩と明度について相違を調べた。

図 5 に示す Lassen の平面物体の目標画像とキャプチャ画像の明度差に着目すると、右下の部分に差が出ているのが確認できる。

投影画像について、同じ箇所を調べたところ、R チャンネルの画素値が飽和 (8bit 階調画像の 255) しており、投影光量が不足していることが確認された。

図 7 に示す、白色照明下の見かけの推定画像 (図 1 の  $C_{est}$ . 反射率に比例.) を見ると、該当箇所の反射率は低いことがわかる。このような箇所に対して、 $StyTr^2$  で明るい目標画像が生成されると、最大輝度の投影を行っても目標の明度に到達できず、誤差が生じたものと考えられる。

図 8 に示す Gogh の平面物体の目標画像とキャプチャ画像の彩度差に着目すると、平面物体の至る所で彩度差が見られる。

Gogh の投影画像の赤枠で示している領域の RGB の値を調べると、目標画像の彩度が高いにも関わらず、RGB が 0 であった。そこで、図 10 の投影画像が飽和している部分を調べたところ、反射率が高くなっていることが確認された。以上より、元の物体の反射率が高くなっており、投影画像の値が低くなる。それにより、投影画像の画素の値が限りなく 0 に近づくことで目標通りに操作できなかったと考えられる。

6. まとめ

本研究では、SAR を用いた現実物体の仮想化を調査した物体輪郭の明確さ、絵画の絵のタッチの付与、絵画の配色の一致の 3 つの観点から調査を行った。そこで、 $StyTr^2$  の投影と Average Color、Texture Overlay を被験者実験により比較したところ、Average Color は構造を保持するがス

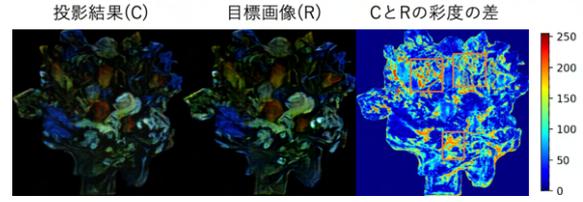


図 8: Gogh 平面物体の目標とキャプチャ画像の彩度差

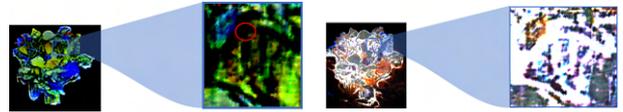


図 9: Gogh の投影画像

図 10: Gogh 白色推定画像

タイルの特徴を反映していない、Texture Overlay はスタイルの特徴を付与できるが輪郭の明瞭さを犠牲にすることがわかった。一方、 $StyTr^2$  はすべての基準において中程度の性能を示し、バランスがとれた方法であることが分かった。

また、 $StyTr^2$  の定量評価により、目標画像とキャプチャ画像の差が出た原因として、元の物体像の反射率に対して光学的に到達不可能な目標が  $StyTr^2$  で生成されることが分かった。Style loss、Content loss の値とアンケート結果のスコアの値を比べると、一貫性が見られないことが確認できた。このことから、loss は仮想化の定義に従っていないと考えられる。

参考文献

- [1] T. Amano: Projection Based Real-time Material Appearance Manipulation, Proc. of CCD 2013 Workshop on CVPR2013 2nd IEEE International Workshop on Computational Camera and Displays in conjunction on CVPR2013 918 - 923
- [2] Yingying Deng, Fan Tang, Weiming Dong, Chongyang Ma, Xingjia Pan, Lei Wang, Changsheng Xu:  $StyTr^2$ : Image Style Transfer with Transformers, In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 14-19 June 2022; pp. 11326-11336.
- [3] Zhong, F., Jindal, A., Yöntem, A.Ö., Hanji, P., Watt, S.J., and Mantiuk, R.K.: Reproducing Reality with a High-Dynamic-Range Multi-Focal Stereo Display. ACM Transactions on Graphics (Proc. SIGGRAPH Asia), 40(6), Article 241, 14 pages(2021)
- [4] Aliaga, D. G., Law, A. J., and Yeung, Y. H.: A virtual restoration stage for realworld objects. ACM Transactions on Graphics 27, 1 (2008)
- [5] Bermano, A., Bruschweiler, P., Grundhöfer, A., Iwai, D., Bickel, B., Gross, M.: Augmenting physical avatars using projector-based illumination. ACM Trans. Graph. 32, 6, 189:1-189:10 (2013)