



Style Transferred Reality : 画風変換による絵画世界 VR 没入システムの構築

芹沢信也¹⁾, 脇坂崇平²⁾, 泉原厚史³⁾, 高原慧一⁴⁾, 檜山敦^{1,2)}, 稲見昌彦^{1,2)}

1) 東京大学 情報理工学系研究科 (〒 113-8656 東京都文京区本郷 7-3-1, serizawa@star.rcast.u-tokyo.ac.jp)

2) 東京大学 先端技術科学研究センター (〒 153-8904 東京都目黒区駒場 4-6-1)

3) 東京大学 工学系研究科 (〒 113-8656 東京都文京区本郷 7-3-1)

4) 東京大学 工学部 (〒 113-8656 東京都文京区本郷 7-3-1)

概要: 機械学習アルゴリズム Fast Style Transfer を用い、全周囲カメラで取得するライブ実写映像や CG 等を統一的に特定の画風に変換し没入体験させる VR 装置を構築した。本装置により、例えば現実世界がアニメ化するかの様な体験や、各種映像ソースの区別ができない/難しいために、CG と現実世界が区別なく混ざりあった体験がデザインできる。本研究ではユーザ評価に基づき、基礎的な画像変換技術と演出技法の開発を行った。

キーワード: 視覚, VR 基礎, 代替現実感

1. 背景

近年, VR 技術のコモディティ化に伴い, Oculus Rift をはじめとする一般向けの安価なヘッドマウントディスプレイ (HMD) やバーチャル Youtuber など, 個人でも VR 環境を構築し, 体験することが可能となってきた。その一方で, VR 環境は現実環境とはユーザの主観経験上は独立して (弁別可能な状態で) 存在し, その両者をシームレスにつながる技術開発はそれほどなされていない。数少ない例の一つとして, 代替現実 (SR) システム [1] がある。これは, 現実とバーチャルシーン (記録・編集したシーン) を体験者に気づかれることなくシームレスに切り替え, 体験者がバーチャルシーンを (主観的に) 現実として体験するシステムである。類似したシステムとして, Reality Jockey [2] がある。これは, 音声と, 床及び机から被験者に伝わる振動に特化した SR システムである。これらのシステムでは, 体験者に切り替えに気付かれないようにするため, 切替先と後の構図をある程度以上類似した状態にする必要がある。つまり, 記録映像・音声のような詳細な情報を持った情報源との間で切り替える場合, 体験を行う環境と類似した条件の映像・音声との切り替えに制限される。また, CG モデルなど本来存在しないものを混在させる場合, 現実環境に各種クオリティ (テクスチャ, 運動パターンなど) を十分に合わせないと, バーチャルであるということが被験者に気づかれてしまう。そこで, 本研究では, 現実環境に合わない物体・環境も自然に混在させるため, 現実とそれらを合わせて絵画風に加工することで両者のギャップを埋め, 区別しがたく



図 1: Style Transfer の例。葛飾北斎の富岳三十六景 神奈川沖浪裏 (中央) の画風を学習し, シカゴの街並み (左) に適用した (右)。波のようなテクスチャで海と同じ青である空が加工されているのがわかる。(図は [3] より)

シームレスに混ざった体験を実現することを考えた。

1.1 目的

絵画風世界への没入を実現するにあたり, 画風変換の従来手法として挙げられるのはロトスコープといえよう。ロトスコープは, 実写動画をトレースする形でアニメーションを生成する手法である。しかし, ロトスコープはリアルタイム処理ではない点で映画などの受動的コンテンツには有用だが, 現実環境をリアルタイムに加工することはできず, またインタラクティブな体験コンテンツへの拡張性は低い。そこで, 画風変換を行うことが出来る Fast Style Transfer という技術に着目した。Style Transfer とは, 畳み込みニューラルネットワーク (CNN) を用いた画像変換アルゴリズムの一種で, 特定の画像から画風を抽出し, その画風のように入力画像を変換する技術である (図 1)。特に, Fast Style Transfer [3] はアルゴリズムに改良を加えることで, あらかじめ画風を抽出しているの, リアルタイムに入力画像をその画風で変換することが出来る。

本研究では, Fast Style Transfer を用いて, 現実環境と, それに重畳させた CG モデルを絵画風に加工するシステムを開発し, それらが区別なく混ざりあった体験をデザイン

Shinya SERIZAWA, Sohei WAKISAKA, Atsushi IZUMIHARA, Keiichi TAKAHARA, Atsushi HIYAMA, and Masahiko INAMI

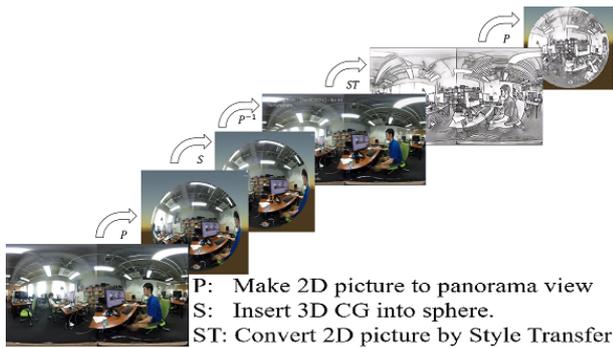


図 2: 各処理ごとの映像の概要. Equirectangular 画像を球に貼り付け, CG モデルを重畳させて撮影することで Cubemap 画像とする. Equirectangular 画像に変換して, Fast Style Transfer を適用後, 再度球に貼り付ける.

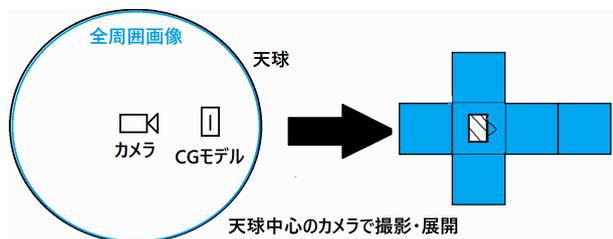


図 3: 全周カメラで取得した画像を天球に貼り付け, 天球内部に CG モデルを配置し, 天球中心のカメラで撮影することで Cubemap 画像を得る.

することを目的とした.

2. 実装

機械学習アルゴリズム Fast Style Transfer を用い, 全周カメラで取得したライブ実写映像とその上に重ねた CG モデルを特定の画風に変換することで, 環境が絵画風世界になりつつ, しかも実際には存在しないものが存在するという没入体験をさせる VR 装置を構築した. これを Style Transferred Reality と呼ぶ.

システム構成を図 2 に示す. 全周カメラによりリアルタイムで全周画像を取得, Style Transfer を適用し, HMD に提示する. 全周カメラとして RICOH Theta S, HMD として Windows MR Acer AH101 を用いた. また, Unity-Python 間の画像の送受信にはアプリケーション間でのグラフィック・バッファ共有を実現する Spout を用いた. PC として, Windows 10, Intel core i9-7900X, メモリ 32GB, GeForce GTX 1080Ti 2つを搭載したものを使用した. 本研究では, モノクロの線画 (図 4) [4] を画風として採用した.

以下に処理の詳細を示す. まず, Theta S から送信されてきた Dual Fisheye 画像を Unity を用いて Equirectangular 画像へと変換し, 球面の内側に貼り付けた. 次に, 球内部に CG モデルを配置した. 三次元上の位置とモデルの大きさ・回転を指定できる. そして, CG モデルを球内部に置いたカメラで撮影し, Cubemap 画像に展開した (図 3). それを再び Equirectangular 画像にし, Fast Style Transfer を適用した. 最後に, 上と同様に球面の内側に貼り付けた (図 2).



図 4: 本研究のシステムで用いた画風のもととなる画像.

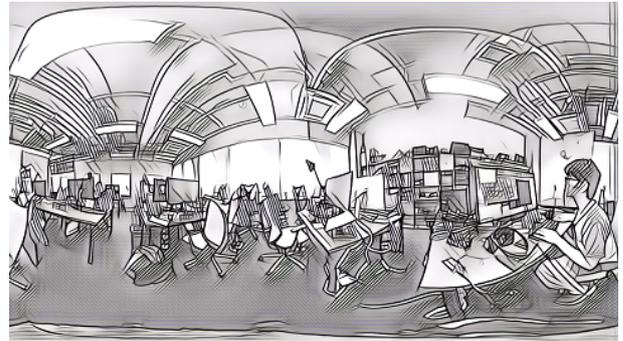


図 5: 実際の Style Transfer 適用例.

実際に適用した結果を図 5,6 に示す. 図 6 中央に置いてあるように見える椅子は実際は CG モデルであるが, 明度や輪郭などの違和感なく環境に溶け込んでいることが確認できる. これらの操作により, 現実環境と CG モデルを絵画風世界に変換したうえで, 少ない違和感で現実と CG モデルを混在させるシステムが構築できたと考えられる.

また, 各フレームごとに独立に処理を行うと, フレームごとのわずかなピクセル値の差の影響により, フレーム間でちらつきが生じてしまう. そこで, 画風を学習する際の損失関数に, 入力にノイズを重畳したものとししないものの出力の差の項を加えることで, ちらつきを抑えた [5].

3. 評価

3.1 スペック

フレーム全体が更新される頻度は 1280×720 (HD) で 9fps ほどであった. この数値はほぼ Fast Style Transfer の



図 6: CG モデル (図 8) がある時での被験者が見る風景. 画像中央付近に図 5 では存在しない椅子が存在している.

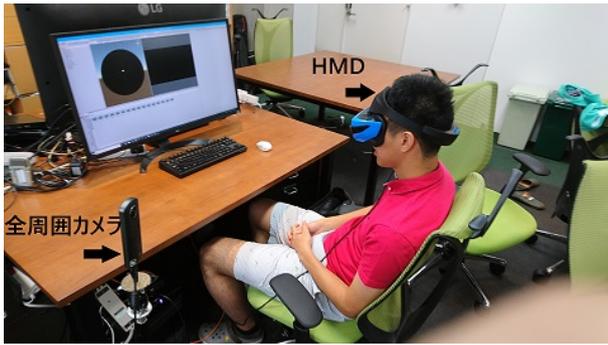


図 7: 実験の様子。被験者を椅子に座らせた上、HMD を装着させ、その左側に三脚付きの全周囲カメラを置いた。



図 8: 使用した椅子の CG モデル。Unity の Asset である Office Chairs chair3(KOBRA GAME STUDIOS) を用いた。

処理に依存するものである。頭部運動に対応して、HMD 上で表示される範囲が更新されるフレームレートは、最大 90fps であった。また、遅延は 300ms 程度で、体験者が自身の体（例えば手）を動かしながら見た場合には、遅れがはっきりと感じられたが、それ以外の場合は、映像と音の位置のずれ次第であり、体験内容次第であった。内訳は、カメラ-Fast Style Transfer への送信（図 2 の一番左の画像が得られるまで）に 150ms 程度、Fast Style Transfer(図 2 の ST) に 110ms 程度、そのほかの処理（図 2 の ST 以外）に 40ms 程度であった。

3.2 実験

Style Transfer Reality 体験時の認知の様相を調べるための基礎的心理実験を、20 代前半の男性 8 人に対して行った。Fast Style Transfer 下で CG モデルを混在させた際に被験者は気付くか、気付くとしたらどのような要因かを調べた。被験者はすべて、周囲に置かれた事物についてある程度事前知識がある状態であった。実験手順は以下の通り：

(1) 被験者を椅子に座らせた上で HMD を装着させ、その横に全周囲カメラを置いた（図 7）ここで、被験者にはまず CG モデル（図 8）なしの状態では画風変換された現実世界（図 5）を体験してもらい、加工されていない現実世界との相違点に慣れてもらった。(2) HMD 呈示映像を暗転させ



図 9: 現実環境と CG の遮蔽関係矛盾。

た。(3)「これから表示する映像には、おかしなところがあります。見つけたら手を上げて教えてください。」と指示した。(4) CG モデル（図 8）を実写映像上に配置した。CG のテクスチャおよび配置は、画風変換を行わない場合には用意に CG であることが被験者に気づかれるものを採用した（図 6）。(5) HMD に映像（図 6）を表示した。(6) 被験者が挙手したら HMD 呈示映像を暗転させた。(7) 被験者が CG モデルの存在（図 8）を報告した場合、気が付いた理由を聞き、映像（図 6）表示から挙手までの時間を記録した。そうでない場合は再度 HMD に映像（図 6）を表示した。(8) CG モデル（図 8）の存在に気が付かないまま 100 秒経った場合、その時点で実験を終了した。

3.3 結果

まず被験者全員が、実験開始後早々に CG が置かれたエリアに注意を払っていた（HMD 装着時の椅子の方向などにより、注意は誘導可能）。正確には頭部および視線方向データと合わせて議論する必要があるが、CG を視野に収めなかったために CG に気づかないでいた、ということはないことは最低限確認されたといえる。参考までに「気付き」を報告したタイミングも付記しておく。被験者 8 人中 5 人は CG モデルの椅子の存在に指摘されるまで気が付かなかった。残りの 3 人はそれぞれ全く異なる理由で CG モデルの存在に気がついた。その理由は以下の通りである。まず被験者 4（開始後 35 秒）は、実験環境に存在する椅子とのデザインの相違により、椅子が実際には存在していない、と判断した（事前知識に基づいた判断）。次に、被験者 5（開始後 45 秒）については、実験中に CG モデルの前を人が偶然通過し、そこをちょうど注視していたために、CG モデルが実写映像に重ね合わされていることに気が付いた。CG モデルは全周囲カメラで取得した画像を貼り付けた天球内部に存在する（図 3）ので、CG モデルに対応する位置を物体が通過した場合、遮蔽関係が矛盾する（遮蔽矛盾。図 9）。

最後に、被験者 7（開始後 19 秒）は、頭部方向を動かした際に CG モデルが背景からわずかに浮いて見えることから、CG モデルの存在に気が付いた（奥行き矛盾）。被験者 4, 5, 7 の結果はそれぞれ、本システムの基本的な特徴をよく表しているといえる。より具体的な考察については後述する。

Style Transfer による画風変換の利点を利用した演出技



図 10: 異なる部屋を 2 つつなげ合わせた画像。



図 11: 図 10 の境界に観葉植物 (CG) を配置した上で画風変換を適用したものの。

法をもう一つ紹介しておく。図 10 は、異なる部屋を 2 つつなげ合わせたもので、図 11 は、境界に観葉植物 (CG) を配置した上で画風変換を適用したものである。Style Transfer の性質上、オリジナルの実写画像に含まれる情報が適度に低減されつつ境界部分が円滑化されるため、二つの環境が混ざっていることに気づきづらい、という効果がある。より詳細なユーザー評価については本発表で説明する予定である。

4. 考察・展望

本研究では、Style Transfer により CG と現実環境を区別なく混ぜ合わせる手法の提案と基礎的評価実験を行った。しかし実験で示されたように、現状のシステム設計では「天球内部に CG モデルを配置し、それをソフトウェア上で仮想的に撮影して二次元画像にしたうえで加工する (図 3)」という手順を踏んでいるために、2 つの技術的課題が存在している。ひとつは、現実上のオブジェクトが CG が配置された場所を通過した際の遮蔽関係の矛盾 (遮蔽矛盾: 図 9) で、もう一つは天球内部のカメラから見たときの CG の見え方の違いによるわずかな動きである (奥行き矛盾)。これらは、体験コンテンツの内容を調整することによりある程度回避可能である (明らかな遮蔽矛盾と奥行き矛盾が生じない CG 配置およびコンテンツとする、等)。根本的な解決には、環境における物体の 3 次元的位置情報を取得した

うえで、3 次元的に Style Transfer を適用できるようにする必要がある ([6] など) が、これは Style Transfer にかかわらず、拡張現実全般の課題でもある。我々はむしろ、過半数の被験者が CG を体験中多くのタイミングで視野中心に捉えつつも、そこに存在する奥行き矛盾および CG の存在に自発的に気づけなかった点に注目したい。画風変換が、適度に奥行きを低減もしくは補完している可能性がある。この点については、今後被験者を増やして統計的検証を進める予定である。その際、Style の差異の影響も注意深く考慮する必要がある。なぜなら、Style によってオリジナル情報の縮減・変容度合いが大きく変化し、その結果として認知・知覚上の帰結が変化することは明らかであるからである。例えば Style によって、よりコンテンツの矛盾にシビアなもの (被験者が気づきやすいもの)、あるいはその逆のものなどがあると考えられる。最後に、長時間日常的に現実と CG の区別がない Style Transferred Reality を体験し続けた場合に、認知・知覚だけでなく行動にどのような変容が起こるかというのも興味深いテーマであると指摘しておきたい。これには、全周囲カメラを据え置きではなく体験者の頭部に搭載し、ある程度の空間移動を可能とするといった装置改良が必須となる。これもまた今後の課題とする。

謝辞

本研究は JSPS 科研費 (15K16014) および JST CREST (JPMJCR16E1) の支援によるものである。

参考文献

- [1] Keisuke Suzuki, Sohei Wakisaka and Naotaka Fujii : Substitutional Reality System: A Novel Experimental Platform for Experiencing Alternative Reality Scientific Reports, Vol. 2, No. 1, pp. 459, 2012.
- [2] Kevin Fan, Hideyuki Izumi, Yuta Sugiura, Kouta Minamizawa, Sohei Wakisaka, Masahiko Inami : Reality Jockey: Lifting the Barrier between Alternate Realities through Audio and Haptic Feedback Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2557-2566, 2013.
- [3] Justin Johnson, Alexandre Alahi and Li Fei-Fei : Perceptual Losses for Real-Time Style Transfer and Super-Resolution European Conference on Computer Vision, 2016.
- [4] Rui Shu : fast-style-transfer, <https://github.com/RuiShu/fast-style-transfer>.
- [5] Jeffrey Rainy : Stabilizing neural style-transfer for video, <https://medium.com/element-ai-research-lab/stabilizing-neural-style-transfer-for-video-62675e203e42>.
- [6] Hiroharu Kato, Yoshitaka Ushiku, and Tatsuya Harada: Neural 3D Mesh Renderer The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.