



生成 AI を用いた手画像の生成による リアルなバーチャルハンドの提示

高橋堅汰¹⁾, 入山太嗣¹⁾, 小室孝¹⁾

1) 埼玉大学 理工学研究科 (〒 338-8570 埼玉県さいたま市桜区下大久保 255)

概要: バーチャルハンドの外見がユーザーの実際の手に近いほど、それを自己の身体の一部として自然に受け入れやすいことが示されている。本研究では、ユーザーの実際の手の外見に近い手画像を簡易に生成し、VR 空間上で提示する手法を提案する。拡散モデルを活用し、ユーザーの手画像と目標の姿勢情報を条件として用いることで、本人の手と見た目が近い任意姿勢の手画像を生成する。

キーワード: 拡張・複合現実, 拡散モデル, 身体所有感

1. はじめに

現在、VR においてハンドトラッキングによるバーチャルハンドを用いて仮想物体を操作する機会が増えている。仮想空間においてバーチャルハンドを用いて仮想物体を操作することは、ユーザーに対してバーチャルハンドが自らの体の一部であるかのような感覚である身体所有感を想起させ、VR における没入感や満足度の向上につながる。

被験者に対してバーチャルハンドを提示し、視覚と他の物理刺激を同期させた場合の身体所有感について研究されている [1][2]。これらの研究では、映像上でバーチャルハンドに物体が触れると同時に被験者の手に振動を加えることや、バーチャルハンドと実際の手動きを同期させることで、現実には存在しないバーチャルハンドに対しても、自分の体の一部と錯覚することが実証された。それ以外にも、バーチャルハンドを実際の長さよりも伸長して提示する身体拡張を被験者に提示したとき、本来の身体とは著しく異なる対称性を持つバーチャルハンドに対しても身体所有感が想起されることが報告されている [3]。

また、いくつかの研究において、バーチャルハンドの形状や見た目が被験者の身体所有感に影響を与えることが報告されている [4][5]。これらの研究では、スケルトンのようなバーチャルハンドや、他の人の手をモデリングして作成したバーチャルハンド、ユーザーの手をモデリングして作成したバーチャルハンドを用いて、被験者が体感する身体所有感について比較実験を実施した。その結果、VR 空間上で操作するバーチャルハンドの見た目が、現実のユーザーの手に近いほど強い身体所有感が引き起こされることが実証された。しかし、現実のユーザーの見た目に近いバーチャルハンドを作成するためには、3D スキャナなど特殊な機器を必要とするため、簡易にリアルなバーチャルハンドを作成するのは困難である。

近年、現実には存在しない画像をリアルに生成する技術として、拡散モデル [6] を使用した画像生成が注目されている。一方で、人間の手のように、部位ごとの相対的な配置や数量

に厳密な制約がある対象については、正確な生成が難しいことが知られている [7]。そこで、手のデプス画像を使用し、手の形状を正しく修正する HandRefiner[8] という研究がある。この研究では、拡散モデルを用いて生成された手の構造が不自然な画像に対して、デプス画像を参照することで手の形状や姿勢を修正する。このように、デプス画像を参照した画像生成をすることで、ネットワークがオクルージョンなどを考慮した手の構造を正しく認識し、手の形状を正確に再現したリアルな画像を生成することが可能となる。

これらの研究を踏まえて、本研究では、生成 AI でユーザーの手に外見に近い画像を生成し、これをバーチャルハンドとして用いることで、強い身体所有感を引き起こす新たな提示手法を提案する。これによって、簡易にリアルなバーチャルハンドを生成することが可能になり、VR におけるユーザーの没入感や満足度向上が期待できる。

2. 生成 AI を用いた手画像生成

2.1 概要

提案手法の概要を図 1 に示す。本手法では、ユーザーの手画像と目標とする手の姿勢情報を入力として、拡散モデルを用いた画像生成を行い、ユーザーの実際の手の外見が近く、入力の姿勢を参照した画像を生成する。手の構造情報を正確に保持するために、手の姿勢情報にはデプス画像を用いる。生成した手画像を VR 空間上でバーチャルハンドとしてユーザーに提示することで、ユーザーの没入感向上を図る。

2.2 手画像生成手法

今回実装した画像生成手法を図 2 に示す。画像生成手法では、任意のデプス画像とマスクを入力できるように変更した HandRefiner[8] を利用する。目標姿勢の手のデプス画像と、ユーザーの手画像、生成範囲を決定するマスクを入力とし、生成モデルをインペインティングとして利用する。インペインティングは、欠損領域を周囲の情報から補うため、ユーザーの手の外見の特徴を保持した生成ができる。またデプス画像を ControlNet に入力することで、生成する画

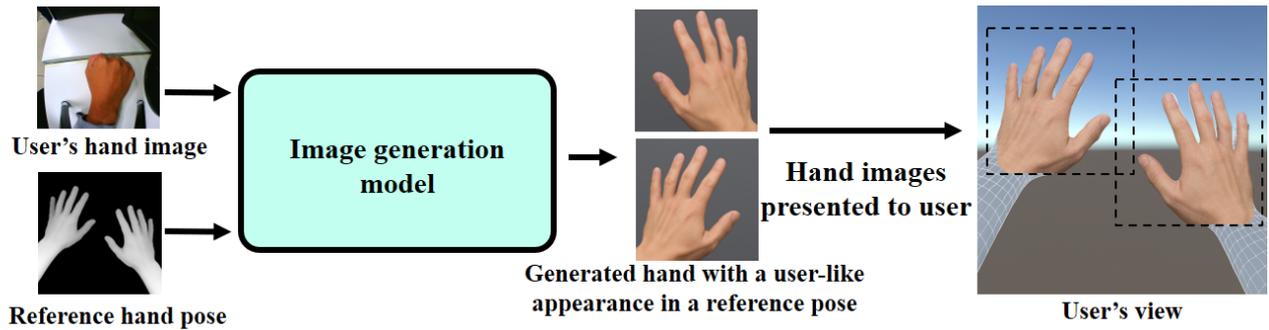


図 1: 提案手法

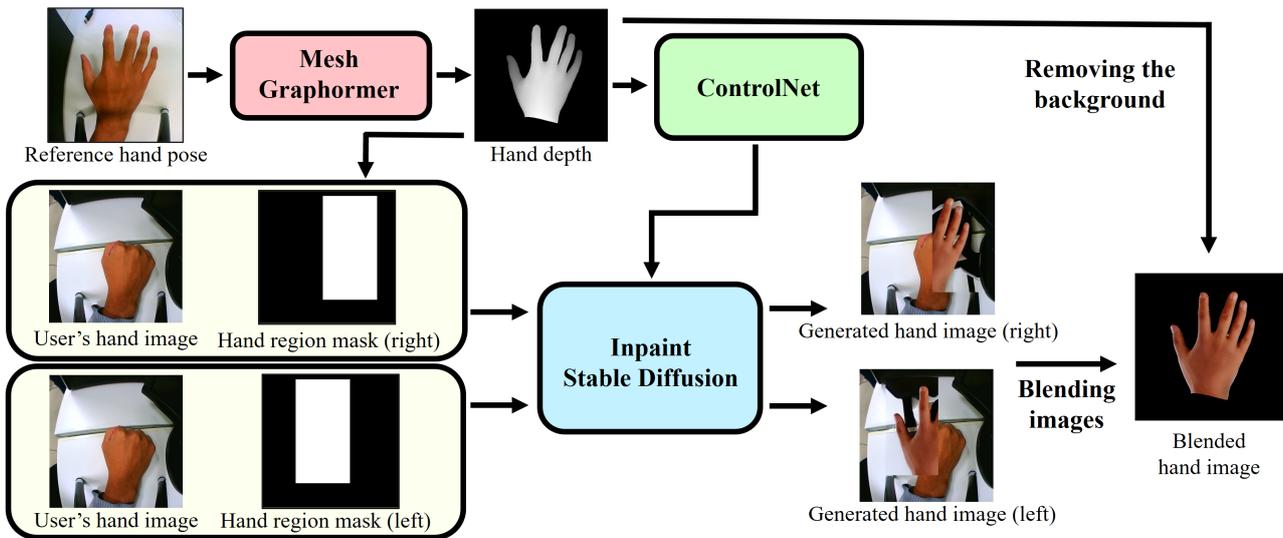


図 2: 実装した画像生成手法

像の手の構造・姿勢情報を指定することができる。これらのアプローチによって、ユーザーの肌の色など外見の特徴を保持した、目標の姿勢の手画像を生成することを実現する。

どの指が手前にあるのかなどより正確な手の構造情報を取得するために、手の姿勢情報としてデプス画像を使用する。手のデプス推定には、HandRefiner と同様に、Mesh Graphormer[9] を使用する。Mesh Graphormer はグラフ畳み込みニューラルネットワークと Transformer を用いることで、単一画像から 3 次元人体姿勢とメッシュを再構成する。作成された 3 次元メッシュをもとにデプスマップを生成することで手の形状を拡散モデルに入力する。広範囲を生成領域に指定するマスクを使用すると、入力であるユーザーの手画像を参照することができず、ユーザーの外見の特徴が失われてしまう場合がある。そこでマスクの範囲を分割し、マスクの形状を変化させることで、ユーザーの手画像の外見の特徴を失わずに画像生成を行うことができる。分割したマスクを利用して生成された画像を合成し、一つの生成画像とする。生成した画像をそれぞれ単純に結合すると、画像ごとに色が若干異なるため境界線が見えてしまう。そのため、重複している部分を線形的に合成することで画像を滑らかに合成する。今回は、生成範囲を左右に分

割したマスクを使用する。画像を生成するとき、生成範囲は矩形のマスクによって決定されるため、デプス画像を参照として生成された手画像以外に、不要な背景が生成されてしまう場合がある。そこで、手のデプス画像を二値化した画像を使用して、生成された画像の背景を除去し、手の部分のみの画像に変換する。

3. 手画像生成実験

提案手法を利用することで、ユーザーの手画像とデプス画像を入力としてユーザーの手の外見に近いリアルな画像を生成できることを検証する。本実験では、手の動作を撮影した動画をもとに、手の姿勢情報となるデプス画像を各フレームごとに作成する。作成したデプス画像とユーザーの手画像を入力することで、ユーザーの手の外見の特徴を保持して目標の姿勢に変更した画像を出力する。

3.1 結果

生成された出力画像の例を図 3 に示す。手のデプス画像を参照とし、指の本数や構造が正しい画像を生成している。また、分割したマスクを使用した場合と使用していない場合の出力画像を比較すると、分割したマスクを使用した出力画像のほうが入力であるユーザーの手画像の肌の色に近

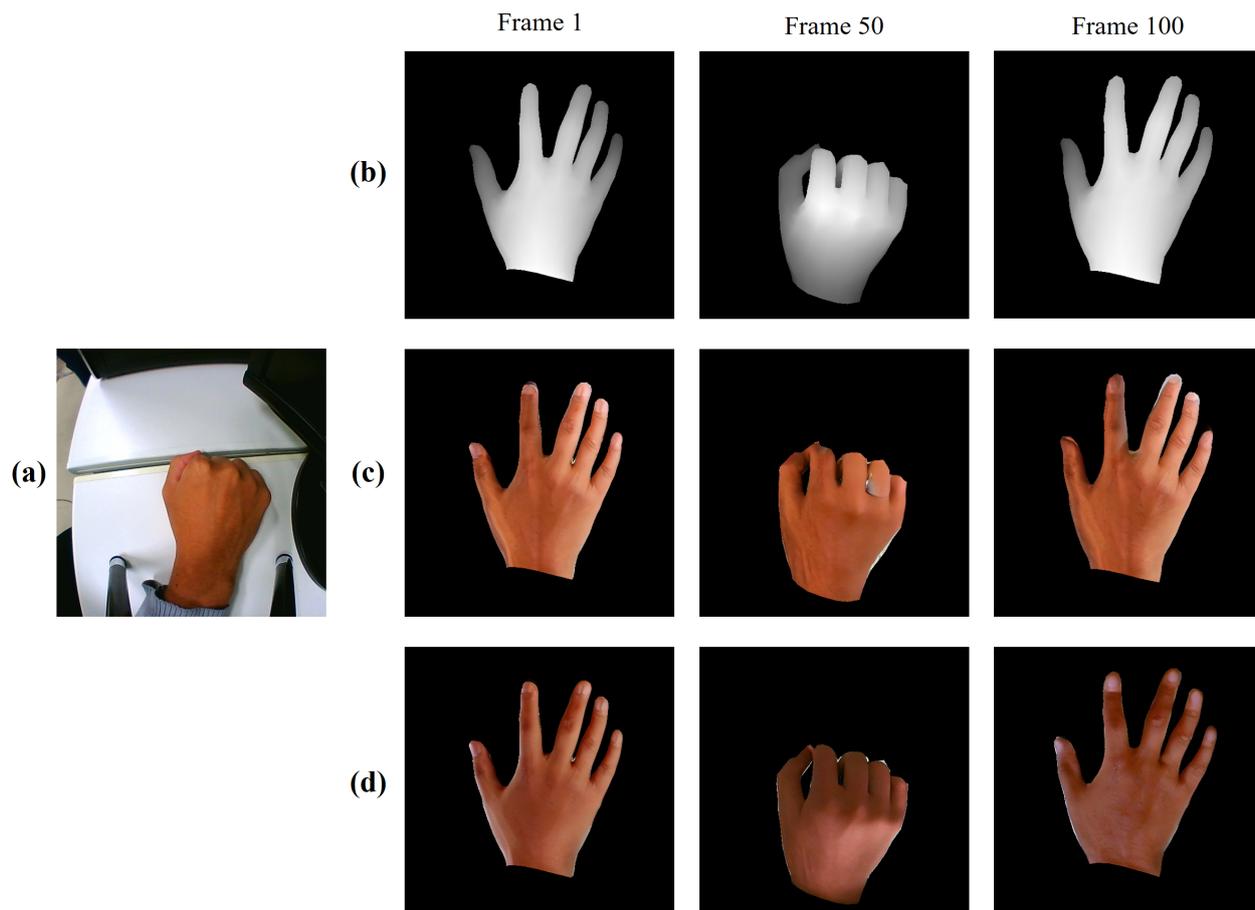


図 3: 連続フレームにおける入力と出力. (a) ユーザーの手画像, (b) デプス画像, (c) 分割したマスクを使用した出力画像, (d) 分割したマスクを使用していない出力画像

い画像を生成している.

3.2 考察

生成された画像は、境界部分に目立った線はみられなかった。これは重複している部分を線形的に合成することによって、色の変化が緩やかになったためであると考えられる。また指の形状が正しく、ユーザーの手画像に近い肌の色の画像を生成していることから、入力画像を参照としたリアルな手画像が生成されていると考えられる。

4. 結論

本研究では、拡散モデルを用いてユーザーの手画像を入力画像として参照することで、ユーザーの手に見た目が近いリアルなバーチャルハンドの画像を生成する手法を提案した。提案手法ではユーザーの手の姿勢情報としてデプス画像を用いることで、肌の色だけでなく手の形状も考慮した手画像の生成を行った。

また、提案手法を用いて手画像の生成を行った結果を検証した。その結果、デプス画像を参照として手の姿勢情報が保持された手画像を生成できることを示した。また、デプス画像に合わせたマスクでは肌の色が参照されないことがあるという問題を解決するため、マスクの形状を変更することを検討し、実験、考察を行った。その結果、重複を含

んで左右に分割したマスクを使用することでデプス画像をもとに正しい指を生成しつつ、ユーザーの肌の色を再現した手画像を生成することを示した。

参考文献

- [1] Mel Slater, Daniel Pérez-Marcos, H. Henrik Ehrsson, and Maria V. Sanchez-Vives. Towards a digital body: The virtual arm illusion. *Frontiers in Human Neuroscience*, Vol. 2, 2008.
- [2] Mel Slater, Daniel Pérez-Marcos, H. Henrik Ehrsson, and Maria V. Sanchez-Vives. Inducing illusory ownership of a virtual body. *Frontiers in Neuroscience*, Vol. 3, pp. 214 – 220, 2009.
- [3] Konstantina Kilteni, Jean-Marie Normand, Maria V Sanchez-Vives, and Mel Slater. Extending body space in immersive virtual reality: a very long arm illusion. *PloS one*, Vol. 7, No. 7, p. e40867, 2012.
- [4] Ferran Argelaguet, Ludovic Hoyet, Michaël Trico, and Anatole Lécuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *Proceedings of 2016 IEEE virtual reality (VR)*, pp. 3–10, 2016.

- [5] Maria Pyasik, Gaetano Tieri, and Lorenzo Pia. Visual appearance of the virtual hand affects embodiment in the virtual hand illusion. *Scientific reports*, Vol. 10, No. 1, p. 5412, 2020.
- [6] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, Vol. 33, pp. 6840–6851, 2020.
- [7] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sd-xl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*, 2023.
- [8] Wenquan Lu, Yufei Xu, Jing Zhang, Chaoyue Wang, and Dacheng Tao. Handrefiner: Refining malformed hands in generated images by diffusion-based conditional inpainting. *arXiv preprint arXiv:2311.17957*, 2023.
- [9] Kevin Lin, Lijuan Wang, and Zicheng Liu. Mesh graphormer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 12939–12948, 2021.