



映像投影による他者の顔への視覚的変容に向けた 画像生成手法の検討

羽山 大智¹⁾, 渡辺 義浩¹⁾

1) 東京科学大学 工学院 情報通信系 (〒 226-8501 神奈川県横浜市緑区長津田町 4259 G2-31, hayama.d.b7d2@m.isct.ac.jp)

概要: ものまねメイクと呼ばれる手法により、本来の顔の骨格を変えることなく特定の他者への見た目の変容が可能である。同手法は、化粧を用いた肌の色、皺の位置、立体感等の操作により、見た目を他者に近づけている。本稿では、顔への映像投影により同様の効果を得るための方法について検討を行う。具体的には、StyleGAN ベースの画像生成とメッシュ変形を用い、主要な顔パーツの位置を保った状態で変容目標に近づく投影像の生成手法を提案する。

キーワード: プロジェクションマッピング, バーチャルメイク

1. はじめに

自身の顔を目標とする顔に近づけることや置き換えることが可能な技術は、映像制作やコミュニケーションツールなど様々な分野で注目を集めている。そのような中で、画像内における人物を変化させる手法では、主に機械学習により画像を生成する手法が検討されている [1, 2]。特に、画像内の人物の顔を特定の他者の顔へと置き換える編集は Face Swapping と呼ばれている [2]。

同様に、現実世界における見た目を特定の他者へと変容させることで、映像制作やコミュニケーションのさらなる発展に繋がると予想される。しかし、Face Swapping の結果を直接現実再現することは容易ではない。その主因は、画像上での編集と異なり、現実世界において、見た目の操作対象となる顔の目や鼻、口の位置などの骨格を自在に変化させることが難しい点にある。

一方、近年では、ものまねメイクと呼ばれる手法により、本来の骨格を変えることなく、特定の他者へと見た目を近づけることが実現されている。同手法は、化粧を利用し、見かけの肌の色、皺の位置、立体感を操作することで、他者に見た目を近づけている。しかし、ものまねメイクは特殊な技能であり、手軽に利用できる段階には至っていない。

また、実際の化粧をせずに顔の見た目を手軽に変えることが可能な手法として、Dynamic Facial Projection Mapping (DFPM) を用いたメイクが提案されている [3]。同手法は、現実の動的な顔に映像投影を行い、見た目を変化させる。しかし、DFPM による他者への変容は未だ実現されていない。このような投影により、ものまねメイクと同等の効果を達成できれば、実際のメイクよりも容易に他者への変容が可能になると考えられる。

以上を踏まえ、本稿では、投影により顔を特定の他者へと変容させることを目標とする。しかし、現実の顔の変容は骨格を保持する制約があり、同条件の下で特定の他者への変容が可能な投影像は検証されていない。仮に、色補償

を用いたとしても、眉や目など投影が難しい部位は見た目を上書きすることはできない。そのため、投影対象の形状を無視した投影は避けることが望ましい。

そこで、本稿では、StyleGAN ベースの手法を用いて、投影対象の形状を保った他者の要素を持つ顔画像を生成し、3D Morphable Model (3DMM) によるメッシュ変形と組み合わせることで、他者への変容が可能な投影像の生成手法を提案する。このとき、メッシュ変形は僅かな生成ずれの修正に用いる。また、比較のため、メッシュ変形のみを用いて目標とする顔画像を歪め、投影対象の形状を保持した投影像を生成する。本稿では、上記 2 つの手法で生成した像を実際の顔へと投影し、定性的な確認を行った。

2. 関連研究

2.1 生成画像を用いた投影による人体の見た目の変容

Projection Mapping (PM) により、人体の見た目を変容させる操作は、活発に研究されている。特に、顔への投影では、化粧効果の付与や年齢感の操作が実現されている [3, 4]。

年齢感の操作を行う手法は、機械学習により投影対象の年齢を操作した顔画像を変容目標として 1 枚用意し、投影結果の顔画像との差分からフィードバック型の色補償を行っている。また、顔に追従して投影を行うため、3DMM によって推定されたメッシュを用いて、変容目標の 1 枚の顔画像を変形させる。なお、同研究は年齢変容のみに絞った検証を行っており、他者への変容は検証されていない。

また、1 枚の画像を変形させた顔が投影目標となり、変容目標に含まれない表情変化に伴う皺などを表現できない。他者への変容を行う PM で同手法を用いる場合、表情変化に伴う表現を行うには、投影対象とは異なる人物の同じ表情の画像を用意する必要がある。しかし、そのような画像が必ずしも手に入るとは限らない。

加えて、同手法は、顔の 3 次元形状推定誤差により色補償が正しく収束しない問題を解決するため、変容目標と投影結果の顔画像にガウシアンブラーを適用している。しかし、

その結果として、実現される投影像がぼける問題がある。そこで本稿では、ガウシアンブラーを適用しない PM のため、フィードフォワード型の色補償を用いた検証を行う。

一方、投影による手の見た目の操作として、ControlNet を用いた手法がある [5]。同手法は、手の形のマスク画像を基に ControlNet を用いて投影像を生成する。なお、画像生成時には、マスク画像の他にテキストを与えることで、生成の方向性を決定する。同手法により、生成画像を用いた変容を実現しているが、対象は手に限定されている。

2.2 StyleGAN ベースの顔画像生成手法

近年、顔画像を生成する手法として、Diffusion Model と GAN に基づいた 2 つの手法が盛んに研究されている。そのような中で、Diffusion Model に関しては、前述の ControlNet により生成される顔の形状を制御することが可能である。しかし、GAN のように、潜在空間内で明示的に操作を行うことは構造上難しい。潜在空間内で特定の他者を指定することや投影に適した像を求めることなど、生成制御の柔軟性を踏まえると、投影像生成には GAN に基づいた手法を採用することが好ましいと考えられる。

なかでも、GAN の一種に StyleGAN がある [1]。同手法は、GAN の潜在変数を人間がより解釈しやすい変数へと置き換えている。これにより、特定カテゴリ内の高品質な画像生成と操作を実現し、顔画像生成に応用されている。また、同手法を用いて、髪や編集や他者に顔を寄せる編集などが行われている [1]。しかし、スタイルコードのもつれにより、同手法を用いて、顔の形状を保持したまま他者へと変化させることは通常難しい。

その中で Liu らは、Face Swapping の実現のため、StyleGAN ベースの Editing-for-Swapping (E4S) と呼ばれる手法を提案した [2]。同手法は、18 層ネットワークの 1 - 13 層まで顔のセグメンテーションマスクごとに個別の StyleGAN を模倣したネットワークを設け、各セグメンテーションマスクに与えるスタイルコードの制御をすることで、高品質な Face Swapping を達成した。

具体的には、まず上書きされる顔画像 I_t と変容目標の顔画像 I_s を用意し、 I_t に向きと表情を揃えた I_s の顔画像 I_d を生成する。次に、 I_t と I_d のセグメンテーションマスク M_t と M_d を推定した後、 M_t の眉や目、皮膚などの顔の輪郭内の各形状を M_d に含まれる形で置換する。その後、 I_d から抽出したスタイルコードを、置換済みの顔の各マスク領域に与え、顔画像を生成する。なお、髪や装飾品、背景は I_t のものを与える。このように、E4S は形状とスタイルを独立して操作できる。同手法を応用し、セグメンテーションマスクを置換しない場合、上書きされる顔の形状保持と顔の変化の両立が可能であると考えられる。

3. E4S による画像生成と 3DMM を用いたメッシュ変形による投影像生成

3.1 概要

本稿では、E4S とメッシュ変形を組み合わせることによる、投影対象の形状を保持した特定の他者へ変容可能な投

影像の生成を提案する。図 1 に本手法の概要を示す。まず、E4S により、投影対象の形状を保った変容画像を生成する。その後、3DMM を用いたメッシュ変形により、位置補正を行い、投影像を得る。加えて、同ステップにおいて同時に色補償も行う。また、図 2 にシステムの構成を示す。

3.2 E4S による形状を保持した顔画像生成

前述の通り、現実世界における他者への変容では形状の保持が重要となる。そこで本稿では、E4S を用い、投影対象のセグメンテーションマスクを変容目標の形状で置換することなく直接使用し、変容目標から抽出したスタイルコードを与える。これにより、投影対象の形状と変容目標の特徴を両立する画像が得られると考えられる。

そのため、まず白色光を投影し、白色照明下における顔を撮像する。続いて、白色照明下の顔画像に対して E4S を適用し、投影対象の顔の形状を保持した変容目標の要素を持つ顔画像を生成する。なお、本稿では、投影対象の顔の各領域のうち、目・眉・唇・口・鼻・耳・それ以外の顔の皮膚領域のみ変容目標のスタイルコードを与える。

3.3 3DMM を用いたメッシュ変形と投影像生成

E4S は、18 層ネットワークの 1 - 13 層のみセグメンテーションマスクを使用する。また、セグメンテーションマスク内部の形状は制御できない。これにより、投影対象と生成される顔画像の位置に僅かなずれが生じる。そのため、3DMM を用いたメッシュ変形により位置を修正する [6]。

具体的にはまず、投影対象と E4S 適用後の顔画像から 3DMM を介して得た顔テクスチャを用いて、色補償済みテクスチャを計算する。なお、色補償に関しては次項で詳しく説明を行う。続いて、同軸 IR カメラで顔画像を取得し、再度形状推定を行い、同形状のテクスチャとして前述の色補償済みテクスチャを用いる。その後、同軸 IR カメラで取得した顔画像に対して上記で得た 3D モデルを再投影する。

以上の過程により、メッシュ変形による位置補正を達成する。最後に、再投影を行った補償顔画像に対して、ホモグラフィ変換を適用し、変容目標に近づく投影像を生成する。なお、色補償の反射率計算のため、IR 画像上で推定した形状は白色照明下の形状と同一であると仮定する。

3.4 テクスチャ上における色補償

前項のメッシュ変形の過程に、フィードフォワード型の色補償を組み込むことで、色補償済みテクスチャを取得する。なお、色補償は Fujii らの手法を参考に行う [7]。具体的には、まず図 1 の通り、E4S により生成される顔画像と白色照明下の顔画像から 3DMM によるテクスチャ抽出を行う。次に、これらの 2 枚のテクスチャから、カラーカメラ上の見かけの反射率と色補償済みテクスチャを求める [7]。なお、遮蔽により、テクスチャを抽出できない箇所は、色補償済みテクスチャ上で inpaint を用いて補完を行う。

3.5 メッシュ変形のみを用いた手法

袁らの手法 [4] を参考に、E4S を介さず変容目標から直接テクスチャ抽出を行い、メッシュ変形のみを用いて投影像を求める手法も検証する。同手法も、IR 画像上に再投影

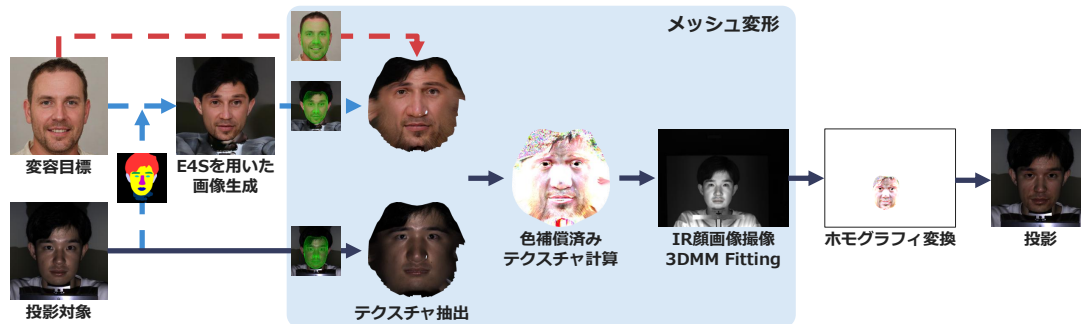


図 1: 提案手法概要。青点線が、E4S+メッシュ変形の手法であり、赤点線はメッシュ変形のための手法を示す。

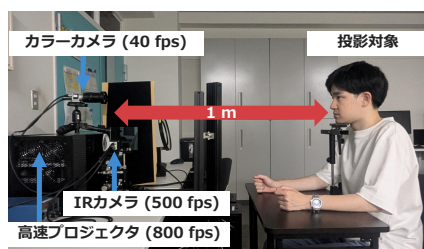


図 2: システム構成と実験環境。

し、投影像を求める過程は 3.3 項と同様である。

以降では、前項までの手法を E4S+メッシュ変形、本項の手法をメッシュ変形のみとする。なお、E4S+メッシュ変形に関しては、E4S の段階で形状が概ね合っているため、各顔パーツのずれが知覚されにくいことや、メッシュ変形のみと比べて歪みが小さくなり、自然な顔が生成されることが予想される。加えて、E4S は 1 枚の変容目標から投影対象と表情を揃えた顔画像が生成されるため、投影時に表情特有の皺や陰影の変化を再現できる可能性がある。

4. 実際の顔への投影による他者への変容効果の検証

4.1 概要

投影結果をカラーカメラで撮像し、定性的な確認を行った。なお、投影対象は、顎台を用いて静止させた真顔とした。

実験環境は図 2 の通りであり、プロジェクタから投影面までの距離は約 1 m とした。加えて、事前に標準白色板を用い、カラーカメラとプロジェクタ間における色の対応関係を計算した後、同軸 IR カメラとプロジェクタのホモグラフィ行列を求めた。なお、変容目標を図 3 に示す。同顔画像は StyleGAN2-ADA により生成した [8]。

また、投影には 800 fps の高速カラープロジェクタを用いた [9]。さらに、色補償用カメラとして 1920×1200 px, 40 fps のカラーカメラを使用した。加えて、プロジェクタと同軸なカメラとして、 720×540 px, 500 fps の IR カメラを用いた。なお、抽出した顔画像のテクスチャに対し、ゲインとして 0.55, 0.35 を与えて、画素値を定数倍し、補償目標の明るさを下げ変容を実現しやすくした。

4.2 結果

顔画像の生成結果と投影結果を図 3 に示す。ここで、各ブロックの上段が E4S+メッシュ変形、下段がメッシュ変形

のための手法による結果である。また、各ブロックごとに同じ顔画像を変容目標としている。同顔画像に対して生成されるホモグラフィ変換適用後の投影対象の形状を保った顔画像が、左側から 2 列目の画像であるが、どちらの手法も目標とする他者に近づくような画像を生成可能であることがわかる。特に E4S+メッシュ変形による手法は、眉の形状など、より投影対象の形状に合わせた画像を生成している。

また、上から 1 ブロック目の投影結果は、どちらの手法も変容目標に顔を近づけることができていた。加えて、2 ブロック目は、E4S+メッシュ変形の方が、肌の色や皺の入り方で変容目標に似ているように見える。しかし、1 ブロック目ほど変容目標に近づいては見えない。3 ブロック目は、1, 2 ブロック目と比較すると変容感が弱い。一方、ゲインが 0.35 の際にどちらの手法も少し女性的になっているように見える。加えて、3 ブロック目を見ると E4S+メッシュ変形と比較し、メッシュ変形のための手法は、真顔への投影を行っているにもかかわらず、左目尻の皺や口元の影、皺の深さなど変容目標の笑顔を若干引きずって見えた。

また、すべての結果においてどちらの手法も各顔パーツのずれはあまり知覚されなかった。一方、投影対象のほくら付近の色補償ずれがすべての試行において確認された。

4.3 考察

E4S+メッシュ変形、メッシュ変形のためのどちらの手法も他者への変容を概ね達成可能であると考えられる。一方、変容目標を完全には実現していない。形状を保持した顔画像が変容目標に概ね似ていることから、この原因は、色補償が十分でないことや表面下散乱による投影像のぼけにあると考えられる。また、今回の実験条件下では顔を完全には固定できず、投影までに顔が僅かに動いた結果、反射率推定に誤差が生じたことも同現象の要因であると思われる。

加えて、投影対象と変容目標の組み合わせにより、変容しやすさが異なることも判明した。これは 2 つの顔の類似度や、変容目標における目立つ特徴の有無、色補償の容易さが関与していると考えられる。

一方、E4S+メッシュ変形では 1 枚の画像から投影対象の形状に合わせた投影像を生成するほか、笑顔で強く現れていた皺が薄れるなど投影対象の表情に合わせた生成を確認した。加えて、本手法でも形状がずれている箇所が確認されたが、3 次元形状推定の誤差によるものであると思われる。

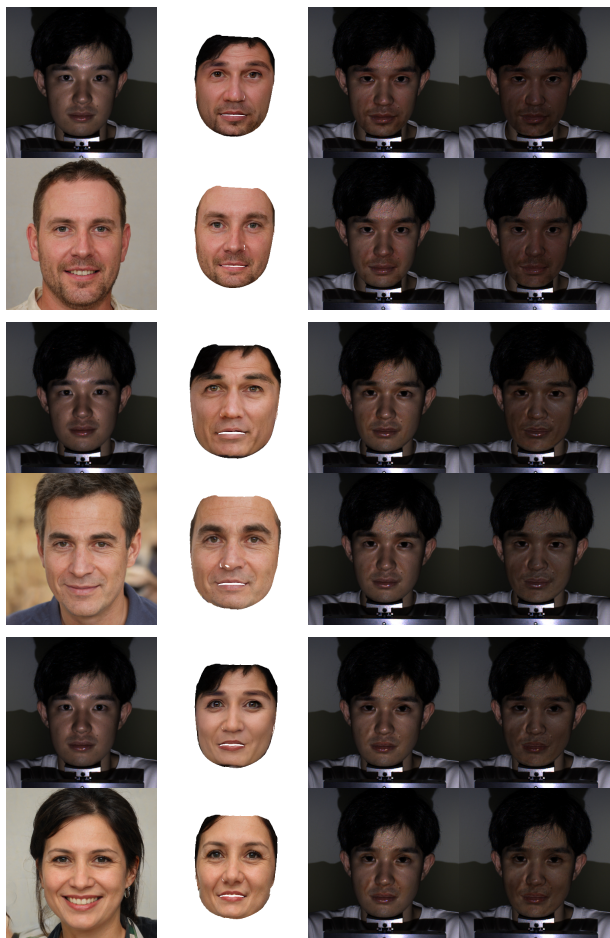


図 3: 顔画像の生成結果と投影結果。左から 1 列目上段が白色照明下の投影対象、下段がそれぞれのブロックにおける変容目標に相当する。また、2 列目が投影対象の形状を保った顔画像であり、3 列目がゲイン 0.55 の投影結果、4 列目がゲイン 0.35 の投影結果である。なお、上段が E4S+メッシュ変形、下段がメッシュ変形のみ結果である。

また、眉などの形状を予め合わせた効果は見られなかった。これは、効果を期待した箇所が眉などの投影が難しい部位であることや色補償が不十分であることが要因と考えられる。

さらに、メッシュ変形のみ手法と比較し、変容目標の髪の分布や色の再現性が劣ることを確認した。同現象は、E4Sの潜在空間で表現の難しい成分があるためだと考えられる。

5. まとめ

本稿では、E4S と 3DMM によるメッシュ変形から投影像を求め、他者へと顔を近づけることが可能であることを定性的に確認した。一方、ものまねメイクと比較し、品質が十分でなく、投影対象と変容目標の組み合わせ次第では、変容感が弱まることを明らかにした。そのため、より適した画像の生成や色補償が必要であると考えられる。

今後の方針として、変容を実現しやすい目標の特定や潜在空間を活用した最適な投影像の生成が挙げられる。投影像の生成に関しては、潜在空間を活用し、変容目標に近づきつつ、投影対象から実現しやすい顔画像を求めることを

考えている。また課題の 1 つに、動的な顔への投影があり、本手法を用いた個人特化 3DMM の構築などが必要である。

参考文献

- [1] Tero Karras, Samuli Laine, and Timo Aila. A Style-Based Generator Architecture for Generative Adversarial Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 43, No. 12, pp. 4217–4228, 2021.
- [2] Zhian Liu, Maomao Li, Yong Zhang, Cairong Wang, Qi Zhang, Jue Wang, and Yongwei Nie. Fine-Grained Face Swapping via Regional GAN Inversion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.
- [3] Nao Tsurumi, Kaoru Ohishi, Ryo Kakimoto, Fumihiko Tsukiyama, Hao-Lun Peng, Yoshihiro Watanabe, and Yuji Masubuchi. Rediscovering Your Own Beauty Through a Highly Realistic 3D Digital Makeup System Based on Projection Mapping Technology. In *International Federation of Societies of Cosmetic Chemists*, 2023.
- [4] 袁璐, 彭浩倫, 林隆介, 渡辺義浩. 色補償を用いたプロジェクションマッピングによる顔の年齢変容に関する検討. 第 28 回日本バーチャルリアリティ学会大会, 2023.
- [5] Yotam Erel, Or Kozlovsky-Mordenfeld, Daisuke Iwai, Kosuke Sato, and Amit H Bermano. Casper DPM: Cascaded Perceptual Dynamic Projection Mapping onto Hands. In *SIGGRAPH Asia 2024 Conference Papers*, 2024.
- [6] Patrik Huber, Guosheng Hu, Rafael Tena, Pouria Mortazavian, Willem P Koppen, William Christmas, Matthias Räscher, and Josef Kittler. A Multiresolution 3D Morphable Face Model and Fitting Framework. In *11th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2016.
- [7] Kensaku Fujii, Michael D. Grossberg, and Shree K. Nayar. A Projector-Camera System with Real-Time Photometric Adaptation for Dynamic Environments. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.
- [8] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training Generative Adversarial Networks with Limited Data. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.
- [9] Yoshihiro Watanabe and Masatoshi Ishikawa. High-Speed and High-Brightness Color Single-Chip DLP Projector Using High-Power LED-Based Light Sources. In *The International Display Workshops*, 2019.