

ブレンドシェイプを用いた個人の表情や個性を反映した 3D 顔モデルのリターゲットイング

足田 善地[†]

[†] 早稲田大学

山口 周悟[†]

[‡] 早稲田大学理工学術院総合研究所

岩本 尚也[†]

森島 繁生[‡]

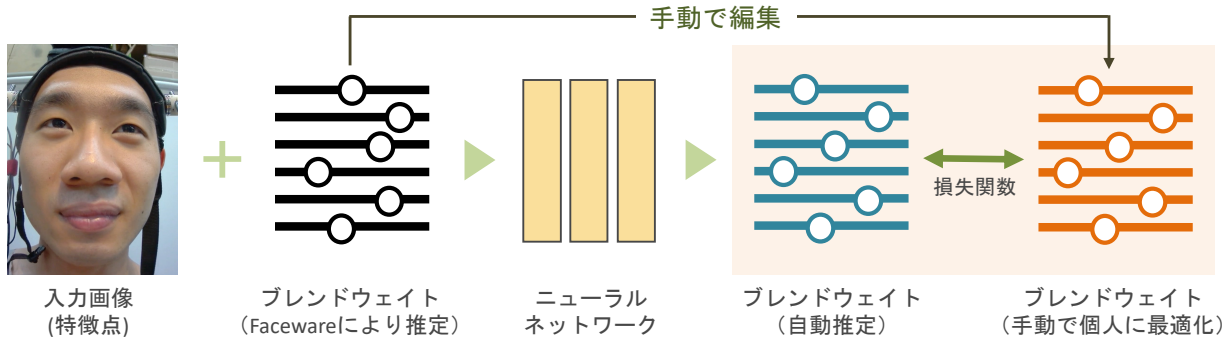


図 1: 提案手法の概要図

1. はじめに

近年、広告業界などを中心として、仮想空間上で現実的な人間の表情を作成することに需要が高まっている。特に個性を反映した高精細な表情アニメーションを安価に作成することが重要である。本研究では、個人の発話動画からアバターの個性を反映した表情コントロールを可能にするため、個性を持った表情のリターゲットを自動化する手法を提案する。個性を反映した三次元顔生成モデルとして、3D モーフアブルモデル [1] を使用したものや、単一スキャンから動的顔モデルを作成する手法 [2] が提案されたが、個性に影響するきめ細かい表情の再現が困難である。そこで本研究では、アバターの作成元となる人物をスキャンすることで得た個人化されたブレンドシェイプ (Personalized ブレンドシェイプ) の独自のデータセットを作成した。また、本データセットを用いたブレンドウェイトの編集結果を学習することにより、個性を転写するネットワークが有効に働くことを示す。

2. 関連研究

人間の顔を撮影し、アニメーション可能なモデルを作成する方法として、3D モーフアブルモデル (3DMM), ブレンドシェイプを用いる手法が利用されている。

2.1 3D モーフアブルモデル (3DMM)

Blanz ら [1] によって提案された 3DMM は、大量の三次元顔モデルに対して主成分分析を適用することにより、顔の個性や表情を主成分の係数を用いてパラメトリックに制御することを可能とした。近年では様々な方法で 3DMM のパラメータ最適化を行うことで、撮影データから被写体の顔を忠実に再現する手法が提案されている。Danecek [3] らは 2 つの表情の違いを計測する指標を導入することで、3DMM によって合成された顔モデルの再投影画像と現実に撮影された単眼画像を比較してパラメータ最適化を行い、単眼画像から顔モデルの

作成を可能にした。しかし、提案された指標では顔のきめ細やかな表情を合成することが困難である。

Wood [4] らは単眼画像から高解像の特徴点を抽出し、特徴点に合うように 3DMM のパラメータを最適化することで、単眼画像から顔モデルを作成する手法を提案した。Wood らの手法は特徴点のみをもとに最適化を行うため計算コストが低く、高速に処理を行うことが可能である。しかし表情合成の精度は特徴点抽出に大きく依存し、予測精度の低い状況下ではモデルへの適合性が失われる。

また 3DMM では主成分分析によって次元削減を行うことでパラメータを作成しているため、平均的な表情が合成されやすく、個性を反映させるようなきめ細やかな表情の表現が難しい。本研究では個人化されたブレンドシェイプを扱うことで、この課題を克服する。

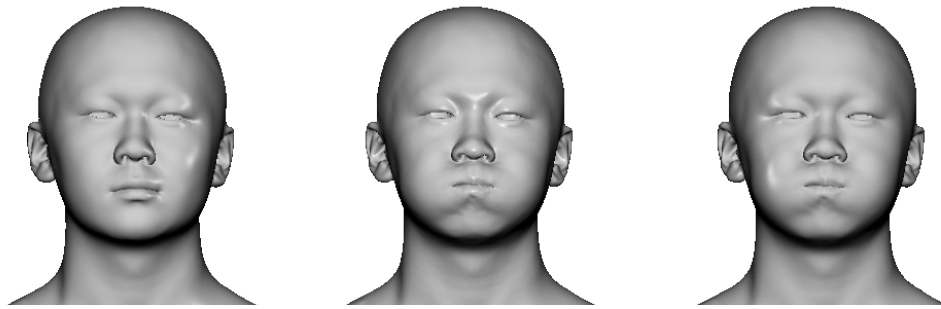
2.2 ブレンドシェイプ

ブレンドシェイプでは、キーシェイプと呼ばれる複数種類の異なる表情によって構成される顔モデルを用いて、これらの線形和で新たな表情を合成する手法が広く用いられる。Li [2] らは一度のスキャンから高品質かつ個人の動きを反映した顔モデル (Personalized ブレンドシェイプ) を合成するフレームワークを提案した。彼らは豊富なデータセットを用いることで、動的な顔モデルを高速かつ頑健に合成した。ただし、精度の高いブレンドシェイプモデルであっても、表情のリターゲットが正確に行われなければ、入力表情に対して忠実性の高い表情の表現はできない。本研究では、表情のリターゲットの精度を向上させることにより、個性を反映した表情の表現を可能とする。

3. 提案手法

本稿では、顔のブレンドシェイプモデルに対し、入力された画像からブレンドウェイトの自動推定をニューラルネットワークを用いて行う手法を提案する。プロダクションにおいては、ブレンドウェイトの推定および編集を行う Faceware [5] というソフトウェアが広く使用されているが、Faceware を用いたブレンドウェイトの推定は正確性に欠けており、アーティストによる編

Modeling Personalized Teaching Methods Using Deep Reinforcement Learning:
Zenchi Hikida[†], Shugo Yamaguchi[†], Naoya Iwamoto[†], and Shigeo Morishima[‡] ([†]Waseda University, [‡]Waseda Research Institute for Science and Engineering)



Faceware自動推定(入力)

手動(正解データ)

提案手法による推定

図 2: 実験結果の比較

集が必須となっている。本研究では、図 1 に示すように、Faceware により取得されたブレンドウェイトおよび画像から取得された特徴点を入力として、ニューラルネットワークを用いて高精度なブレンドウェイト推定を行う。

まず、顔の正面動画を入力として全フレームに対し顔の 4 部位（目、鼻、口、眉毛）の特徴点を取得する。続いて、特徴点と、その人物の Personalized ブレンドシェイプを用いて表情のリターゲットを行う。ブレンドウェイトの自動推定が適用された後、ユーザによる手動編集によってウェイトの修正を行う。本研究では、手動編集結果を正解データとして学習を行うことにより、新たなデータに対する編集作業の自動化を可能とする。

4. 実験

4.1 実験概要

Faceware により取得した編集前のブレンドウェイトおよび画像から取得した特徴点を入力とし、編集済みのブレンドウェイトを正解データとしてニューラルネットワークの学習を行なった。ネットワークは、4 層のマルチレイヤーパーセプトロンで、中間層の活性化層は ReLU、2 層と 3 層に Batch Normalization を適用した。ハードウェアには、NVIDIA A6000 を使用した。損失関数には平均自乗誤差 (MSE) を使用した。図 3 に損失の減衰を示す。

4.2 結果

図 2 に、Faceware の編集前ブレンドウェイト (入力)、編集後のブレンドウェイト (正解)、提案手法による推定結果を示す。それぞれ、ウェイトをブレンドシェイプに適用し、レンダリングしたものである。Faceware による編集前のブレンドウェイトは、口の再現に不正確な点があったが、提案手法による推定結果は、正解に近い形状を示している。

5. おわりに

本研究では、Faceware により取得されたブレンドウェイトおよび画像から取得された特徴点を入力として、ニューラルネットワークを用いて高精度なブレンド

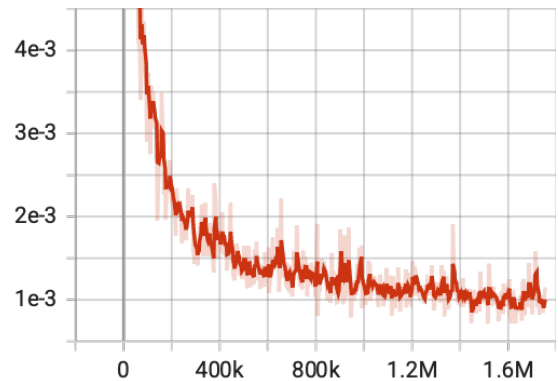


図 3: 実験結果：損失の減衰

ウェイト推定を行うことにより、忠実性の高い表情のリターゲットを可能とした。

一方で、本稿ではフレーム毎の推定を行なっており時系列の連続性が考慮されていないため、今後は LSTM や Transformer など時系列情報に適したネットワークを用いた学習を行うことを考えている。

謝辞

本研究は JSPS 科研費 (19H04137, 21H05054) の補助を受けています。

参考文献

- [1] Blanz *et al.* A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 187–194, 1999.
- [2] Li *et al.* Dynamic facial asset and rig generation from a single scan. *ACM Trans. Graph.*, Vol. 39, No. 6, 2020.
- [3] Danecek *et al.* EMOCA: Emotion driven monocular face capture and animation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20311–20322, 2022.
- [4] Wood *et al.* 3d face reconstruction with dense landmarks. In *European Conference on Computer Vision (ECCV)*.
- [5] Faceware. <https://facewaretech.com/>.