

# GAN による表情のモーフィング作成と高次元表情弁別閾値楕円面の推定 Morphing of Facial Expressions by GAN and Estimation of High-Dimensional Facial Expression Discrimination Threshold Hyperellipsoids

飯野 匠<sup>†</sup>      小林 洋明<sup>†</sup>      趙 晋輝<sup>†</sup>  
Takumi Iino   Hiroaki Kobayashi   Jinhui Chao

## 1 序論

現在、表情認識は言語カテゴリに対してラベルを割り当て分類をして表情を識別している。しかし、表情が常に連続変化するため離散的な認識には限界がある。そこで心理特性を基準に表情を評価して空間上に配置する次元説 [1] が注目された。しかし、心理表情空間と物理刺激との対応が難しい。そこで、炭矢らは物理刺激としての表情画像に対して心理特性として弁別閾値を導入した心理物理空間を提案した [2]。その心理物理空間の構築、すなわち表情弁別閾値楕円面の推定はデータセットの表情推移から表情弁別閾値のデータ点を測定し、その点からその表情弁別閾値楕円面を推定するという手法がとられている。高次元の表情弁別閾値楕円面を測定する際には様々な表情推移で測定をする必要があるため、多数の表情画像が必要である。本研究では、表情画像から深層学習のモデルの一種である SemanticStyleGAN[3] を用いることでモーフィングを生成する新たな手法を提案し、高次元の表情弁別閾値楕円面の測定を可能にした。さらにその表情モーフィング画像を用いた心理物理実験によって得られたデータに対して、最尤推定をすることで高次元の表情空間の構築をおこなった。

## 2 表情弁別閾値楕円面に関する従来研究

### 2.1 心理空間と物理空間

従来の次元説による心理空間は SD 法などの測定を用いて、表情を 2, 3 次元の心理空間上に配置することでその幾何学特性を議論されてきた [1]。しかし、従来の心理空間は物理刺激と直接対応していないため、心理空間の幾何学的特性と実際の実験結果では異なる場合がある。そこで、心理空間に物理刺激を結びつけた心理物理空間が提案されている [2]。

心理物理空間とは画像の画素値の物理刺激と、その刺激における弁別閾による心理計量から構成する空間のことである。本研究で用いる心理物理空間は表情の推移画像の物理刺激と、その表情における弁別閾値によって知覚特性を導入することで表情空間を獲得する。

### 2.2 表情空間の構築

表情空間を構築するためのデータセットは Bosphorus Database[6][7] に含まれる BS002 から本研究の提案手法を用いて、BS002 内の各々の表情間の中間画像を生成し表情画像を増やしている。BS002 は 7 基本表情 (Anger, Disgust, Fear, Surprise, Happiness, Sadness, Neutral) と Facial Action Coding System (FACS) と呼ばれる手法である。その手法で顔の動きをラベル付けした 21 表情がある。これらと増やした表情画像を用いて表情弁別閾値を測定す

ることで得られるデータを主成分分析 (PCA) を用いて次元削減をして表情空間を構築した。

また、表情画像は輪郭より外側は黒く塗りつぶし、輪郭境界はガウシアンフィルタによってぼかしている。

### 2.3 表情の補間手法

各表情画像から表情間の中間画像を生成する手法としてモーフィングが知られている。モーフィングは表情推移に限らず、ある形状から他の形状へ変化させるときに徐々に変化するように補間する技術である。ある画像  $I^S$  から別の画像  $I^T$  への推移を考える。この際のモーフィング画像  $I^M$  は以下のような合成で生成ができる。

$$I^M = \alpha I^S + (1 - \alpha) I^T \quad (0 \leq \alpha \leq 1)$$

これまでの研究では、各画像に特徴点を付与して局所変形することで表情画像補間を生成した。しかし、この手法では画素による変形であるため、表情の筋肉については考慮されていない。また、表情推移中の口の開閉や 2 表情間の各部位の位置の差により、顔の部位が重なって表示されてしまう問題が確認されている。

## 3 提案手法

本研究では各表情画像を学習済みモデルの潜在空間に埋め込み、求めた潜在変数間で 2.3 で述べた手法であるモーフィングをおこない、その潜在変数から SemanticStyleGAN[3] で画像を生成する手法をとる。

### 3.1 SemanticStyleGAN

SemanticStyleGAN[3] とは GAN というモデルの一種で、生成器と識別器という 2 つのネットワークから構成される生成モデルである。ただ、通常の GAN と異なることは潜在変数が定義されている。潜在変数の定義により、従来不可能であった編集を可能にした。また、1 つの潜在空間に埋め込むと各パーツやテキストの間で強く影響を受けてしまい意図しない変化が生じてしまうため、本モデルではパーツやテキストごとに潜在空間を分割することでパーツごとにもつれることを防ぐ。

### 3.2 潜在変数への埋め込み

SemanticStyleGAN などのネットワークモデルは潜在変数から画像を生成する。その逆の操作である GAN inversion という手法をおこなう。適切に潜在空間へ埋め込むための損失関数は以下を用いる。

$$L = \lambda_p L_{\text{percept}}(G(w^*), I) + \lambda_s L_{\text{simple}}(G(w^*), I) + \lambda_m L_{\text{mask}}(w^*, m) + \lambda_r L_{\text{reg}}$$

$I$  は埋め込む画像、 $w^*$  は生成器が出力する潜在変数、 $m$  は SemanticStyleGAN で生成した粗構造マスク、 $G$  は生成器である。また、 $\lambda$  は各々のハイパーパラメータである。 $L_{\text{percept}}(G(w^*), I)$  は知覚的損失 (Perceptual loss) と

<sup>†</sup> 中央大学大学院理工学研究科情報工学専攻  
Dept. of Information and System Engineering,  
Graduate School of Science and Engineering,  
Chuo University

呼ばれ、人間の知覚精度を基準にした損失をあらわす。これは VGG-16[4] により導出される特徴量ごとの平均二乗和 (MSE) となり、 $L_{\text{simple}}$  は画素値同士の MSE である。また、 $L_{\text{mask}}(w^*, m)$  はマスクと潜在変数の MSE であり、マスクから潜在変数がずれないようにしている。 $L_{\text{reg}}$  は L1 正則化項である。

本手法では、この損失関数が最小となるように潜在変数を探索、更新を繰り返す。潜在変数の更新は Adam[5] を用いる。

以上の提案手法により作成した表情推移の画像を図 1 に示す。



図 1: 提案手法で作成した Anger から Disgust の表情推移

#### 4 実験

提案手法により、7 基本表情から他の BS002 の 21 表情間の表情推移の中間画像を作成した。1 つの表情推移における中間画像は 100 枚で総計 19600 枚の表情画像を作成した。その表情画像を用いて表情弁別閾値を測定した。そのデータに PCA を適用し、次元削減されたデータに対して最尤推定をおこなうことで 6 次元の表情弁別閾値楕円の推定、表情空間の構築をおこなった。以下がその結果の一部である。(図 2, 3)

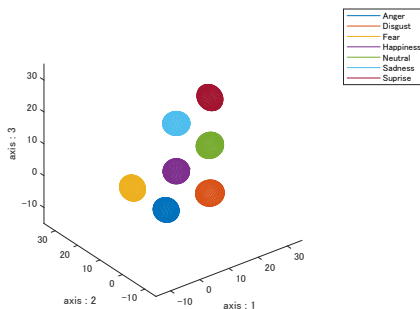


図 2: 第 1, 2, 3 主成分軸に対する楕円面

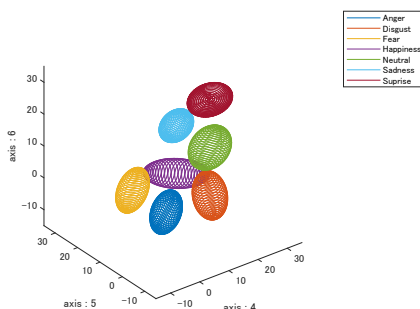


図 3: 第 4, 5, 6 主成分軸に対する楕円面

また提案手法により作成した表情画像と炭矢らが用いた画素値の線形モーフィングで作成した表情画像の、それぞれの測定データから推定した表情弁別閾値楕円の第 1, 2, 3 主軸の方向を図 4 に示す。互いの軸方向が成す角度の最も小さいペアを同じ色で表現している。

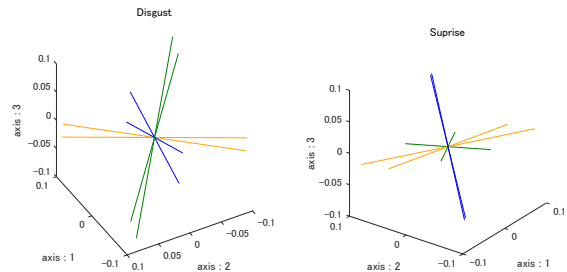


図 4: 異なる観察者の表情弁別楕円の軸方向の比較

#### 5 結論

本研究では SemanticStyleGAN を用いて、精度の高い表情推移の画像を多数に作成することができ、パーツごとに分けた潜在変数をモーフィングすることで中間画像生成に成功した。また、その画像を用いて高次元の表情弁別閾値楕円を推定した。結果は 6 次元のリーマン計量を求めることに成功し、高次元の表情空間を構築した。また、このデータと炭矢らのデータとの楕円軸を比較し、各データごと近い方向の楕円軸のペアが存在することを確認したことから、異なる観察者の間において各表情では共通軸の存在が示唆される。

今後の課題として表情のデータベースの表情は常に表情空間上に均一に布置されているとは限らず、表情が偏ってしまった場合には正常に楕円を推定できない場合がある。したがって、画像が不足しているときの表情推移について検討する予定である。

#### 参考文献

- [1] Russell, J. A, "A circumplex model of affect", *Journal of Personality and Social Psychology*, 39, pp1161-1178, 1980.
- [2] Runa Sumiya, Reiner Lenz, Jinhui Chao "Measurement of JND Thresholds and Riemannian Geometry in Facial Expression Space" 2M. Kurosu (Ed.): HCI 2018, LNCS 10901, pp. 453-464, 2018. Springer.
- [3] Yichun Shi, Xiao Yang, Yangyue Wan, Xiaohui Shen, "SemanticStyleGAN: Learning Compositional Generative Priors for Controllable Image Synthesis and Editing", in *2022 Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, Dec 2021.
- [4] Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", Sep 2014.
- [5] Diederik Kingma; Jimmy Ba (2015). "Adam: A Method for Stochastic Optimization". *Proceedings of the 3rd International Conference for Learning Representations*, Dec 2014.
- [6] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3d face analysis," in *European workshop on biometrics and identity management*. Springer, 2008, pp. 47–56.
- [7] Savran, B. Sankur, M. T. Bilge, "Comparative Evaluation of 3D versus 2D Modality for Automatic Detection of Facial Action Units", *Pattern Recognition*, Vol. 45, Issue 2, p767-782, Feb. 2012.