

類似顔選定による個人差に応じた 3次元顔画像の発話表情生成

小澤 優希[†][†] 法政大学大学院理工学研究科赤松 茂^{††}^{††} 法政大学理工学部

1. はじめに

今まで顔に関する研究が多く行われてきた[1]. 本稿では入力顔に対して複数の類似顔を用いて重みづけしながら表情ベクトルを求め、それを適用することでより本人の発話表情に近いかつ認識率の高い発話表情の生成を試みる.

2. 手法

2.1 特徴量[2]

顔画像の取得にはマイクロソフト社製の Kinect for Windows v2 を用いた. 取得したデータ(顔画像データベース: 男性 32 人, 女性 8 人, 計 40 人)について主成分分析による符号化を行うことで顔画像の特徴量を算出する. 顔画像を多次元ベクトル集合 $\{X_m\}(m = 1, 2, \dots, M)$ とし, その分散共分散行列に主成分分析を行って固有ベクトルを求める. X_m の平均 μ と固有値の大きい上位 K 個に対応する固有ベクトルを用いて任意の顔の特徴量は式(1)に従って $f_m = (f_{m1}, \dots, f_{mk}, \dots, f_{mK})$ のように符号化される.

$$f_{mk} = U_k^T \cdot (X_m - \mu) (k = 1, 2, \dots, K < M) \quad \dots (1)$$

2.2 類似顔の選定

任意の顔と入力顔の特徴量のユークリッド距離を類似度として求める. またデータベース内の真顔と発話表情の差分により表情ベクトル e を求める(式 2).

$$e = [f_{p1} - f_{q1}, f_{p2} - f_{q2}, \dots, f_{pk} - f_{qk}] \quad \dots (2)$$

その際, 式(3)のように類似度の高い上位 n 人のデータを使い, 重み係数を変えながら合成することで表情ベクトル ϵ を得る(式 3). 求めた表情ベクトル $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_K)$ を用いて入力顔の特徴量 f_b , 平均 μ , 固有ベクトル U_k から任意の発話表情 \tilde{X} が得られる(式 4).

$$\epsilon = \sum_{i=1}^n w_i e_i \quad \dots (3)$$

$$\tilde{X} = \mu + \sum_{k=1}^K U_k (f_{bk} + \epsilon_k) \quad \dots (4)$$

3. 結果

2.実験方法に従い, 発話表情を生成した(図 1).



図1. 入力顔(左) 生成した「い」の発話表情(右)

顔形状の異なる 4 人について, 生成した発話表情と実際の発話表情におけるユークリッド距離を比較し, どれだ

け本人に近い発話表情か調べた(表 1). ALL がデータベース内全て, が入力顔と同じ性別のデータ, EUCLID が類似顔を用いた場合である.

表1. 発話表情と実際の表情の比較

| | | ALL | MF | EUCLID | | | ALL | mf | EUCLID |
|----------|---|------|------|--------|----------|---|------|------|--------|
| kumakura | あ | 4.46 | 4.82 | 2.30 | tubouti | あ | 4.46 | 4.82 | 1.79 |
| | い | 3.34 | 3.50 | 3.01 | | い | 2.76 | 3.34 | 2.92 |
| | う | 4.07 | 3.21 | 2.84 | | う | 1.91 | 3.60 | 3.44 |
| | え | 1.78 | 2.07 | 2.00 | | え | 1.85 | 3.48 | 3.17 |
| | お | 3.22 | 3.37 | 2.35 | | お | 3.00 | 3.55 | 3.12 |
| ozawa | あ | 4.46 | 4.82 | 1.83 | sinohara | あ | 4.46 | 4.82 | 2.38 |
| | い | 0.95 | 1.49 | 1.20 | | い | 2.32 | 2.22 | 1.47 |
| | う | 2.60 | 3.17 | 1.78 | | う | 2.00 | 1.98 | 2.00 |
| | え | 1.21 | 1.74 | 1.63 | | え | 4.62 | 4.54 | 3.41 |
| | お | 2.14 | 3.35 | 2.01 | | お | 1.81 | 3.04 | 2.54 |

ほとんどの場合で類似顔を用いた発話表情が実際の発話表情に一番近くなった.

また男女 19 人に発話表情を見て発音しているように見える母音を選択してもらうことで, 生成した発話表情の主観的認識率を調べた(表 2).

表2. 主観的認識率

| | ALL | MF | EUCLID |
|---|------|------|--------|
| あ | 0.94 | 0.86 | 0.91 |
| い | 0.91 | 0.85 | 0.89 |
| う | 0.60 | 0.74 | 0.94 |
| え | 0.40 | 0.46 | 0.54 |
| お | 0.41 | 0.49 | 0.59 |

4. 考察と今後の課題

表 1 より類似顔を使用することで先行研究[3]の手法より本手法の方が本人に近い顔形状を生成することができた. しかし中には先行研究の方が本人に近い場合がある. これはデータベースの中に類似顔が比較的少なかったためだと考えられる. そのためデータを増やすことでより本人に近い発話表情を生成することができる. また今回使った認識率は「う, え, お」に関しては少なからず向上したが高いとは言えない. 認識率を向上させるには類似顔よりも口の形に焦点を当てた方が高い結果を得られるかもしれない.

謝辞

本研究の一部には, 科学研究費補助金(基盤(C)19K12188)の助成を得た.

参考文献

- [1] V. Blanz and T. Vetter, "A Morphable Model for the Synthesis of 3D faces," Proc. of SIGGRAPH '99, pp. 187-194, 1999
- [2] T. Kobayashi, et al., "Discriminant Analysis on Separately Coded Representations of Facial Shape and Texture," Proc. of Int'l Conf. AFGR 2004, pp. 711-716, 2004
- [3] 荒井雄大, 永田俊介, 稲葉善典, 赤松茂, "三次元顔モーフィングモデルにもとづく顔印象の生成 -SVMを用いた印象変換ベクトルによる顔形状の変形操作-", 画像電子学会研究会予稿, pp52-55, 2015-02