

表情認識のための感情円環モデルを考慮した特徴埋め込み手法の提案 Feature embedding based on emotion wheel for video-based facial expression recognition

○三好遼¹, 秋月秀一¹, 飛谷謙介², 長田典子³, 橋本学¹

1: 中京大学大学院, 2: 長崎県立大学, 3: 関西学院大学

E-mail: miyoshi@isl.sist.chukyo-u.ac.jp

発表内容概要

本研究では、「感情円環モデル」の構造を利用した動画ベース表情認識手法を提案する。表情認識において、類似表情の誤分類が課題となっている。一方、心理学分野における感情に関する研究では、「喜び」と「悲しみ」が対極の関係にあるなど、基本感情が円環状に配列されているというモデルが提唱されている。類似表情において、このモデルでは離れて場所に配置されている場合がある。そこで本研究では、この知見を活用し、動画像データを特徴量化してクラス識別するプロセスにおいて、感情モデルに整合するように特徴量を埋め込むことによって精度向上を図った。表情データセット CK+による性能評価の結果、提案手法は比較手法より 3.05%、従来手法より 0.15%の精度向上を確認した。

1. 感情円環モデルに基づく特徴埋め込み

画像認識分野における表情認識は、表情は顔の時間変化によって表現されることから、時間情報をより直接的に活用した動画ベースの表情認識が活発になっている。これらの研究においては、表情を捉えるための hand-crafted 特徴やアーキテクチャについて提案されているが、類似した表情の誤分類が課題である。

心理学分野においては、「喜び」と「悲しみ」の感情が対極に位置するなど、人の基本感情は円環状に配列されているという感情円環モデルが提唱されている[1]。ここで、類似表情であるが、感情円環モデル上では離れた位置に配列されている表情がある。

そこで本研究では、このような感情と表情に関する心理学的な知見を表情認識アルゴリズムに組み込むことにより、表情認識の精度向上を図る。具体的には、動画像データを特徴量化してクラス識別するという処理プロセスにおいて、この知見に基づく効果的な特徴埋め込みを実現する。

提案手法の概要を図 1 に示す。本手法は、Deep Learning のネットワークアーキテクチャに感情円環モデルを組み込むことによって、モデルに則して特徴を埋め込み、表情を認識する。

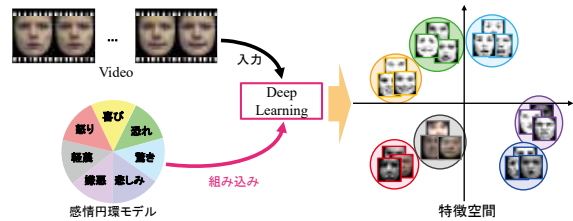


図 1 提案手法の概要

表 1 Baseline および従来手法と提案手法の認識率

Method	Accuracy [%]
C3D-GRU[3]	97.25%
G2-VER[5]	97.40%
Baseline[6]	94.50%
Ours	97.55%

2. 実験

実験では、認識率および特徴埋め込みの観点で比較実験をおこなった。Baseline 手法として、感情の関係を考慮しない一般的な特徴埋め込みである 3 層の全結合を採用した。実験データは、CK+[2]を用いた。評価方法は、10-fold person-independence cross-validation を採用した。

Baseline 手法および RNN や 3D CNN といった時系列ネットワークを用いた従来手法と提案手法の認識率を表 1 に示す。Baseline 手法の認識率は 94.50%であるのに対して、提案手法の認識率は 97.55%であり、3.05%の改善を確認した。また、提案手法は従来手法より 0.15%とわずかであるが高精度であることを確認した。

謝辞

本研究は、JST 研究成果展開事業 COI プログラム「感性とデジタル製造を直結し、生活者の創造性を拡張するファブ地球社会創造拠点」の支援によっておこなわれた

参考文献

- [1] Plutchik R. The emotions, University Press of America, 1991.
- [2] Kanade et al., Comprehensive database for facial expression analysis, FG, 2000.
- [3] Lee et al., Visual scene-aware hybrid neural network architecture for video-based facial expression recognition, FG, 2019.
- [4] T. Albrici et al., G2-VER: Geometry guided model ensemble for video-based facial expression recognition, FG, 2019
- [5] R. Miyoshi et al., Enhanced convolutional LSTM with spatial and temporal skip connections and temporal gates for facial expression recognition from video, Neural Comput & Applic. 2021.