

表情強度と単純ベイズ推定を融合した顔表情認識

Facial Expression Recognition by Fusion of Expression Intensities and Naive Bayes Inference

奥田 誠†
Makoto Okuda藤井 真人†
Mahito Fujii佐藤 洋一‡
Yoichi Sato

1. まえがき

人とコンピュータの円滑なコミュニケーションを実現する手段の 1 つとして、画像処理により、顔表情を認識する手法が活発に研究されている。筆者らは、これまで、顔画像について、表情の認識 (Anger, Disgust, Fear, Happiness, Sadness, Surprise への分類) と強度の推定を同時に行う手法を提案した[1]。

本論文では、文献[1]の手法と単純ベイズ推定を融合した新たな顔表情認識手法を提案する。顔表情データベース Extended Cohn-Kanade dataset (以下、CK+と略す) [2]を用いた実験を行い、提案手法により表情認識精度が向上することを確認した。

2. 表情強度を用いた顔表情認識手法

従来研究の多くが、非常に強く表情が出た顔画像のみを表情認識の対象としているのに対し[3]、筆者らは[1]、表情強度が弱い顔画像も表情認識の対象とし、提案した手法の評価を行った。

図 1 に、文献[1]で提案した表情認識手法の構成を示す。

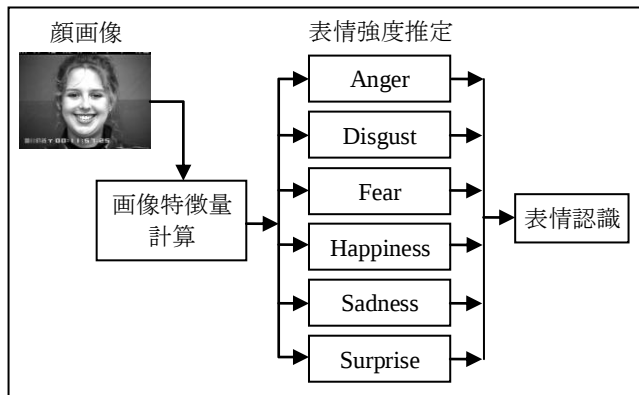


図 1 文献[1]で提案した顔表情認識手法

画像特徴量計算部では、ローカルバイナリパターンを用いた画像特徴量[4]を計算する。

表情強度推定部は、予め、大量のトレーニング用顔画像の特徴量を独立変数、表情強度(0 ~ 100 の実数)を従属変数として表情毎に線形回帰分析し、回帰式を得る。表情毎の回帰分析では、該当の表情の顔画像の表情強度は、人が評価した強度[1]を用いるが、該当の表情以外の表情強度は 0 とする。推定のフェーズでは、入力された顔画像の特徴量と回帰分析により得られた回帰式により、表情毎の強度を計算する。

表情認識部は、表情強度推定部が出力した表情毎の強度

値について、最大値を示した表情を認識結果とする。

3. 単純ベイズ推定との融合

図 1 において、ある表情をした顔画像が入力されたとき、表情強度推定部は、理想的には、その表情以外の表情強度を 0 として出力する。しかし、実際には、Anger: 81, Disgust: 32, Fear: 78, Happiness: 3, Sadness: 18, Surprise: 8 といった数値を出力する。この場合、認識結果は Anger であるが、実際には Fear である可能性もかなり高い。そこで、図 1 で入力される顔画像が、各表情である確率 $p(e|\mathbf{f})$ を、式(1)により定義し、取得する。

$$p(e|\mathbf{f}) = \frac{\hat{I}_e}{\hat{I}_{an} + \hat{I}_{di} + \hat{I}_{fe} + \hat{I}_{ha} + \hat{I}_{sa} + \hat{I}_{su}} \quad (1)$$

ここで、 e は Anger, Disgust, Fear, Happiness, Sadness, Surprise のいずれかであり、 $\mathbf{f}$ は入力顔画像の特徴量である。 $\hat{I}_{an}$, $\hat{I}_{di}$, $\hat{I}_{fe}$, $\hat{I}_{ha}$, $\hat{I}_{sa}$, $\hat{I}_{su}$ は、図 1 の表情強度推定部の各表情の出力強度である。 $\hat{I}_e$ は、表情強度推定部の表情 e の出力強度である。

一方で、単純ベイズ確率モデル[5]による表情の推定も行う。顔画像の各画素のローカルバイナリパターンを $l_1, l_2, \dots, l_n$ とすると、ベイズの定理を用いて、事後確率 $p(e|l_1, l_2, \dots, l_n)$ が式(2)により得られる。

$$p(e|l_1, l_2, \dots, l_n) = \frac{p(l_1, l_2, \dots, l_n|e)p(e)}{p(l_1, l_2, \dots, l_n)} \quad (2)$$

観測された $l_1, l_2, \dots, l_n$ について、条件付き独立性を仮定すると、式(3)が得られる。

$$p(e|l_1, l_2, \dots, l_n) = \frac{p(l_1|e)p(l_2|e)\cdots p(l_n|e)p(e)}{p(l_1, l_2, \dots, l_n)} \quad (3)$$

$p(l_k|e)$ ($k = 1, 2, \dots, n$) は、トレーニング用顔画像を用いて、式(4)により取得する。

$$p(l_k|e) = \frac{t(e, l_k)}{\sum_{l \in L} t(e, l)} \quad (4)$$

式(4)の分子の $t(e, l_k)$ は、表情 e のトレーニング用顔画像において観測されるローカルバイナリパターン l_k の合計数である。一方、 L は顔画像から観測されるローカルバイナリパターンの全種類であり、分母は表情 e のトレーニング用顔画像において観測される全ローカルバイナリパターン

†日本放送協会, Japan Broadcasting Corporation

‡東京大学, The University of Tokyo

の合計数である。式(3)において、 $p(l_1, l_2, \dots, l_n)$ は、表情には依存しないので、式(3)、式(4)より、式(5)が得られる。

$$p(e | l_1, l_2, \dots, l_n) \propto \prod_{k=1}^n \left(\frac{t(e, l_k)}{\sum_{l \in L} t(e, l)} \right) p(e) \quad (5)$$

入力顔画像が各表情である2つの確率 $p(e|f)$ と $p(e|l_1, l_2, \dots, l_n)$ を組み合わせ、次の式(6)により表情 $\hat{e}$ を決定する。

$$\hat{e} = \arg \max_e [\{p(e|f)\}^w q(e|l_1, l_2, \dots, l_n)] \quad (6)$$

ここで、式(5)の右辺を $q(e|l_1, l_2, \dots, l_n)$ とおいた。 w は、 $p(e|f)$ と $q(e|l_1, l_2, \dots, l_n)$ のバランスを取る重み係数である。

4. 実験

CK+を用いて、提案手法の評価を行った。画像特徴量は、文献[1]と同様の手順で計算した。但し、顔領域を検出し、 64×64 のサイズに正規化した後、認識精度を向上させるため、目、鼻、口の領域を図2に示す座標の通り切り出し、それぞれを 4×4 に分割して、各領域でローカルバイナリパターン特徴量を計算し、それらを連結したものを画像特徴量とした。また、文献[1]では、ローカルバイナリパターンの計算は $LPB_{8,3}^{u2}$ オペレータ[4]を用いたが、本論文では、より表情認識率の高かった $LPB_{8,2}^{u2}$ オペレータを用いたときの結果を示す。

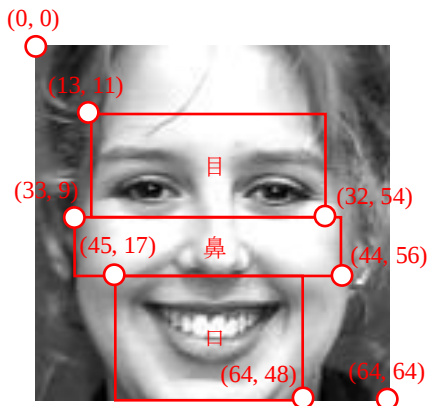


図2 目、鼻、口領域

評価は、5分割交差検定により行った。CK+には、表情がニュートラルからピークに向かって変化する静止画シーケンスが593存在するが、そのうち6表情のいずれかがラベリングされた309シーケンスのみを用いた。被写体の数は111人で、テスト画像とトレーニング画像に、同じ被写体が含まれないようにした。表情がニュートラルの画像は、テストデータから取り除いた。

式(6)における重み係数 w は、さまざまな値を設定して、トレーニングデータ内で5分割交差検定を行い、認識率が最も高かった73にした。

図3は、各手法に対する表情認識率の平均をグラフ化したものである。

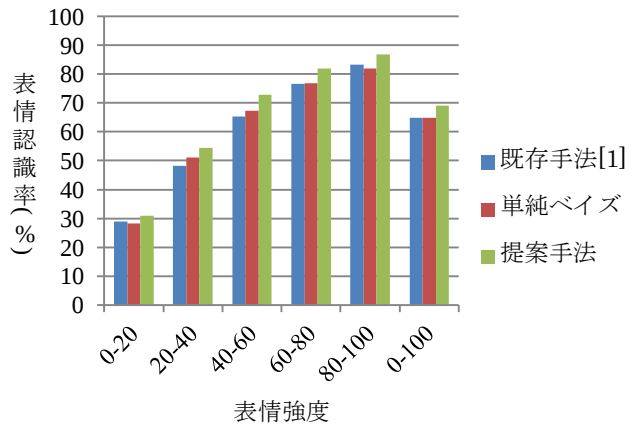


図3 顔表情認識率の比較

既存手法[1]は、図1の手法により表情認識を行ったときの結果である。単純ベイズは、次に示す式(7)により表情認識を行ったときの結果である。

$$\hat{e} = \arg \max_e q(e|f) \quad (7)$$

提案手法は、式(6)により表情認識を行ったときの結果である。単純ベイズも提案手法も事前確率 $p(e)$ は、 $1/6$ とした。

従来手法と単純ベイズの表情認識率は、あまり差がないが、これらを融合することにより、どの表情強度範囲においても、認識率が向上していることが分かる。テストデータ全体の認識率は、既存手法が64.9%、単純ベイズが64.8%であったのに対し、提案手法は69.1%であった。

4. まとめ

表情強度推定と単純ベイズ推定を融合した新たな顔表情認識手法を開発した。個々の手法に比べ、これらを融合した新たな手法の方が、表情認識精度が良いことを、実験により確認した。

今後は、画像特徴量の工夫などにより、さらに精度の向上を図る予定である。

参考文献

- [1] 奥田誠, 藤井真人, 佐藤洋一: "主観的尺度に合致した顔表情の強度推定と認識," 映像学年次大, 8-5, 2012.
- [2] P. Lucey, J. F. Cohn, and T. Kanade: "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," *CVPR4HB*, pp. 94-101, 2010.
- [3] Z. Zhang, C. Fang, and X. Ding: "A Hierarchical Algorithm with Multi-Feature Fusion for Facial Expression Recognition", *Proc. of ICPR*, pp. 2363-2366, 2012.
- [4] G. Zhao et al.: "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions", *Trans. on PAMI*, vol. 29, no. 6, pp. 915-928, 2007.
- [5] P. Domingos and M. Pazzani: "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss," *Machine Learning*, vol. 29, pp. 103-130, 1997.