原著論文

顔画像間の濃度特徴を用いた表情認識†

中村 宗広*1・梶原 祐輔*1・村田 裕章*2・木村 春彦*1

本論文では、顔画像間から濃度特徴を抽出し、5つのパターン識別器を用いて表情を判別する方法を提案する。人間の表情表出シーンを目視確認すると、表情筋の動いた方向に対してしわが現れることがわかる。本手法では、しわが頻繁に現れる箇所に対象領域を設定し、ゼロ平均正規化相関に基づく類似度の特徴量として抽出する。また、最近傍法、ランダムフォレスト法、ロジスティック回帰分析法、サポートベクターマシン、ニューラルネットワークの5つのパターン識別器を用いて表情判別を行う方法を提案する。本手法に対する評価実験では、公開されている複数のデータベースに対し、基本6表情の判別を行った。

キーワード：表情認識、顔画像、表情筋、ゼロ平均正規化相関

1. はじめに

人間同士のコミュニケーションでは、言語コミュニケーションによる情報伝達と比べて、表情や声のトーンなどの非言語コミュニケーションによる情報伝達の割合が93%と大半を占める[1]。そのため、非言語コミュニケーションにおいて人の心を測ることができれば、ヒューマン・コンピュータ・インタラクションに基づくアプリケーションの精度向上が期待できる。その中で、表情には人間の感情表現が含まれることから、画像中の人物の表情を認識するための研究が数多く行われている[2]。

表情は、目や眉毛、鼻、口、表皮付近における表情筋の一時的な動きによって表出し、喜び・嫌悪・悲しみ・怒り・驚き・恐れの基本6表情に大別される[3]。表情をコンピュータにより自動認識する方法として、P. Ekmanらは人種や年齢、性別に依存しない表情筋の動きパターン(AU: Action Unit)を用いるFacial Action Coding System(以下、FACS)[4]を提案している。FACSでは、AUの組み合わせを基に表情を判別する方法により、類似した表情筋の動きに対して高い認識精度を誇っている。しかし、FACSでは67箇所の顔特徴点を専門家が手動で設定する必要があり、人手を要する。このため、顔特徴点を自動抽出する方法が提

案されているが、照明や解像度などの撮影環境に依存して、顔の細かなパーツの認識が難しいケースが少なからずある。FACSを起点として、既存の表情認識システムは、顔の3次元情報を用いる方法[5]、顔領域における特徴点間の距離やオブティカルフロー、輝度変化量などの特徴量を抽出する方法[6][7][8][9]、種々の知的認識手法を用いる方法[10][11][12][13][14]などに大別できる。顔の3次元情報を用いる方法[5]では、2次元の情報を用いた場合と比較して表情認識の精度向上が期待できるが、3次元情報を取得するための前準備が必要であり、また計算コストが2次元情報を用いる場合と比べて膨大になる。顔画像から種々の特徴を抽出する方法[6][9]では、人間が知覚することが困難な輝度値の変化をとらえる等、計算機独自の認識機能を活かすことで表情の認識精度を高めている。また、種々の知的認識システムを応用することにより、顔表情の認識精度は徐々に高くなっている。しかし、人間の表情はあいまいであることが少なからずあることから、未だ人間と同等の表情認識を実現することは難しい。また、認識精度と実行速度とのトレードオフを考慮する必要がある、実用性の高いシステムが求められている。

表情表出シーンを目視確認すると、表情筋が動いた方向にしわが発生することがわかる。無表情時と比べて、表情表出によりしわが発生している領域では、輝度値の分散が高くなることが考えられる。これまで種々のフィルタリングによりしわを検出する方法が提案されてきたが、画像中のノイズや輝度値の急激な変化に対して頑健ではなく、計算コストの増加に伴う実時間性の低下が問題となる。以上を踏まえて、本論文では、表情筋の動きとしてオブティカルフローを抽出し、オブティカルフローの方向に対して一定範囲内の

† Facial Expression Recognition Using Features of Density Distributions between Facial Images
Munehiro NAKAMURA, Yusuke KAJIWARA, Hiroaki MURATA and Haruhiko KIMURA

*1 金沢大学自然科学研究科
Graduate School of Natural Science and Technology, Kanazawa University

*2 金沢大学環日本海域環境研究センター
The Institute of Nature and Environmental Technology, Kanazawa University

対象領域における輝度値の濃度分布を利用する方法を提案する．オプティカルフローとは，時間連続な画像間における物体やカメラの速度をベクトル集合で表記したものである．ベンチマークテストを用いた既存手法の比較実験では，本手法は比較的計算コストが小さく，実時間処理が可能であることを示す．

2. 提案手法

2.1 概要

図1に提案手法の概要図を示す．提案手法では，まず，原画像に8ビット・グレースケール化とヒストグラム伸張化をそれぞれ適用した画像に対して，人物の顔領域および顔特徴点を検出する．次に，無表情時のフレーム（以下，無表情フレーム）と表情表出時のフレーム（以下，表情フレーム）に25箇所の顔特徴点を設定し，対応する特徴間のユークリッド距離としわが発生する領域における濃度分布から特徴を抽出する．そして，抽出したすべての特徴量を種々のパターン識別器に入力し，顔の基本6表情を判別する．

図1 提案手法の概要図

図2 顔画像から自動抽出された25箇所の特徴点

2.2 前処理

まず，原画像に対して8ビット・グレースケール化とヒストグラム伸張化をそれぞれ適用する．次に，既存手法[9]と同様に，顔特徴点抽出ソフトウェアであるLuxand FaceSDK[15]を用いて顔領域及び顔特徴点を決定する．既存手法[9]は，無表情フレームと表情フレームにおける顔特徴点間について，距離や濃度分布に基づく特徴を抽出し，有用度の高い特徴を選択的に用いる方法である．抽出される特徴は，無表情フレームと表情フレームにおける任意の2点間を結ぶ直線距離に基づく特徴が276個，任意の3点間を結ぶ直線で囲まれた領域から輝度値に基づく特徴が12144個である．図2（左）は抽出された顔特徴点，図2（右）は特徴点の部位番号を表している．設定した特徴点は，図2（右）の15, 16, 25を除いて，既存手法[9]と同様である．特徴点の15, 16は，鼻唇溝のしわを抽出するために設定しているが，既存手法[9]では，これらの特徴点はそれぞれ左右の小鼻に設定されている．また，顎下の特徴点である25は，首のしわを抽出するために設定している．

既存手法[9]では，任意の3つの特徴点を直線で結んだ， ${}_{24}C_3 = 2024$ 箇所の領域内から特徴抽出をしており，その範囲は顔領域の大部分に及ぶ．これに対して，本手法では顔のしわが現れる範囲を予め絞り，有用な特徴となるように設定している．このように，顎下の領域を除いて，既存手法[9]と本手法における特徴抽出の対象となる領域の大部分が一致していることから，既存手法[9]を本手法の比較対象として考える．

2.3 特徴量の抽出

まず，図3に特徴量を抽出するためのパラメータを示す．図3において， op_n は無表情フレームにおける n 番目の特徴点， fp_n は表情フレームにおける n 番目の特徴点， lx_n は fp_n と op_n 間の x 軸方向のユークリッド距離， ly_n は fp_n と op_n 間の y 軸方向のユークリッド距離

図3 特徴抽出のためのパラメータ

ド距離である。また、表情フレームにおける縦 h_n ピクセル、横 w_n ピクセルで囲まれたROI(対象領域)の i 行 j 列の画素値を $I(i, j)$ ($i=1, 2, \dots, h_n, j=1, 2, \dots, w_n$)、無表情フレームにおける縦 h_n ピクセル、横 w_n ピクセルで囲まれたROIの画素値を $T(i, j)$ と設定する。 h_n, w_n の値は、後述する方法により各パーツに合わせて設定する。

次に、表情フレームと無表情フレームの各ROIに対してゼロ平均正規化相関 R_{ZNCC} (ZNCC: Zero-Mean Normalized Cross-Correlation) を以下の式により算出する。

$$R_{ZNCC} = \frac{\sum_{i=1}^{w_n} \sum_{j=1}^{h_n} (I(i, j) - \bar{I})(T(i, j) - \bar{T})}{\sqrt{\sum_{i=1}^{w_n} \sum_{j=1}^{h_n} (I(i, j) - \bar{I})^2 \times \sum_{i=1}^{w_n} \sum_{j=1}^{h_n} (T(i, j) - \bar{T})^2}} \quad (1)$$

ここで、 $\bar{I}$ は表情フレームのROIにおける平均輝度値、 $\bar{T}$ は無表情フレームのROIにおける平均輝度値である。 R_{ZNCC} を用いる理由は、SAD (Sum of Absolute Difference)

などの単純な類似度の算出手法と比較して、表情フレームと無表情フレーム間で輝度の変動があっても安定的に類似度を計算できるからである。

本手法において、各表情フレームの間における類似度ではなく、無表情フレームと表情フレームの間の類似度を利用している理由を述べる。無表情フレームを用いる代わりに、各表情フレームの間における類似度を算出する場合、そのための前処理として、あらかじめ人手を介して同一人物の各表情をシステムに登録しておく必要がある。一方、未知の表情画像が入力される状況としては、不特定人物の表情判別、映像中における表情画像の検索などが挙げられる。本研究では、汎用性を考慮して、未知の表情画像が入力される状況を想定している。なお、同様の理由により、各表情画像の間における特徴点の移動量を利用する場合も、システムに入力される表情画像が既知である必要がある。

ROIを設定する箇所について述べる。図4に基本6表情に対応した表情フレームを示す。図4より、happiness, disgust, anger の表情フレームでは鼻唇

図4 同一人物の基本6表情に対応した表情フレーム

溝にしわが鮮明に現れていることがわかる。同様に、surpriseでは額、featでは首、disgustでは眉間にそれぞれしわが鮮明に現れていることがわかる。これらの表情フレームは、見易くするために筆者の一人が手動で切り取ったものである。なお、図4の四角で囲まれた領域は、本手法により設定されたROIである。ここで、25箇所すべてを用いて多数のROIを設定した場合、しわが現れない、もしくは、しわが殆ど現れない領域から特徴量を抽出することになる。そのため、パターン識別器に不要な特徴が入力され、判別精度が低くなることが予想される。そこで、本研究では、図2の1, 2, 3, 4, 5, 6の両眉毛, 17, 21の鼻唇溝, 25の顎下について、ROIを以下の基準により設定する。

- ①額：両眉毛の上に設定する。具体的には、上方向に縦 $\frac{h}{15}$ ピクセル (h はLuxand FaceSDKにより検出した顔領域の縦ピクセル数), 横幅は図2の2から5までのピクセル幅と設定する。ただし、ROIが両眉毛に重なることを避けるために、両眉毛の6つの特徴点の中で一番上の特徴点からROIを縦 $\frac{h}{30}$ ピクセル離す。
- ②眉間：2の3と4の間に設定する。具体的には、縦 $\frac{h}{20}$ ピクセル, 横幅を図2の3から4までのピクセル幅と設定する。
- ③鼻唇溝：2の15および16に設定する。具体的には、図2の15および16をそれぞれ中心とし、 $\frac{h}{40}$ ピクセル, 横 $\frac{h}{20}$ ピクセルと設定する。
- ④首：顎の下に設定する。具体的には、下方向に縦 $\frac{h}{10}$ ピクセル, 横ピクセル幅を図2の25を中心とした $\frac{w}{3}$ (w はLuxand FaceSDKにより検出した顔領域の横ピクセル数)と設定する。

上記のROIで設定したパラメータの妥当性について述べる。ROIのパラメータは、しわが表れる領域を含む必要がある。しかし、しわの表出位置は人によって異なるため、しわを検出しROIのパラメータを設定するために、エッジ強調などの前処理が必要となり、計算コストが増加する。また、しわの検出においてノイズを含んでしまうことも考えられ、システムとしての頑健性が低下する。そこで本研究では、予備実験により、基本6表情の表出においてしわが頻繁に現れる箇所にROIを設定している。

①～④までの特徴点は自動設定することが望ましい。例えば、肌領域を検出し、ROIが肌領域を含む面積が最大となるように設定することが考えられる。しかし、単にHaar-like特徴などを用いた場合、計算コストが増加し、外乱の影響を受けるため、工夫を要する。このことに関しては、今後の課題とする。

なお、顔表面に現れるしわから特徴を抽出する方法

として、ウェーブレット関数を用いることが考えられる。ウェーブレット関数は、画像中から多次元の特徴を捉えることが可能であり、ゼロ平均正規化相関と比べて細部のテクスチャ特徴を抽出することができる。しかし、本研究では、5つの識別器を用いて基本6表情を判別する方法を提案しており、多次元かつ複数通りの判別となる場合、各識別器の性能が低下することが予想される。

以上より、25箇所の特徴点から合計25個の l_{xn} および合計25個の l_{xn} を、5つの特徴点から合計5個の R_{ZNCC} を特徴量として用いる。

2.4 表情認識

図1に示したように、基本6表情のそれぞれに適したパターン識別器を用いて表情を認識する。学習データにそれぞれのパターン識別器を適用し予測誤差を用いて、テストデータの各表情に対する各パターン識別器の汎化性能を予測することができる。そのため、提案手法では、テストデータの各表情に対するパターン識別器の選定には、学習データの各表情に対して5つのパターン識別器を適用し、各表情に対する表情判別時の予測誤差が最も小さくなったものを用いる。予測誤差 E_r は、以下の式を用いて算出する。

$$E_r = \sum_{i=1}^n p_i \quad (2)$$

ここで、 p_i はクラス所属確率を表している。クラス所属確率は、解らしさを表すパラメータであり、一般には $0 \leq p_i \leq 1$ の範囲をとる実数で示される。クラス所属確率は、ディリクレ分布の確率密度関数を利用する手法[16]を用いて算出する。テストデータの表情判別では、各表情に対して選択された識別器から出力されたクラス所属確率が最も大きくなった表情を選択する。パターン識別器は、最近傍法(Nearest Neighbor)、ランダムフォレスト法[17](Random Forest)、ロジスティック回帰分析(Logistic Regression)[18]、サポートベクターマシン(Support Vector)[19]、ニューラルネットワーク[20](Neural Network)を用いる。なお、パターン識別器は多種多様であることから、各顔表情の認識に最適なパターン識別器を定量的に求めることは困難であり、これは今後の課題とする。

3. 実験

3.1 実験試料

以下の2つの公開データベースに対して提案手法を適用する。

・ The Japanese Female Facial Expression Database

[21]

(以下, JAFFEデータセット)

・ MMI Facial Expression Database[22]

(以下, MMIデータセット)

JAFFEデータセットでは日本人女性10人分の無表情フレームと表情フレームが提供されていることから, 提案手法をそのまま適用することができる. MMIデータセットでは無表情フレームから表情表出を終えて無表情フレームに戻るまでの一連の動画像が22人分収められている. また, 人種は, 白人, アフリカ人, アジア人が含まれ, 男女比については, 女性6人, 男性16人である. MMIデータセットでは, 表情フレー

ムが決まっていない. そのため, 本研究では, 表情フレームを「表情が鮮明に現れてから元の表情に戻る直前のフレーム」とみなし, 著者の一人が目視確認により決定している. 尚, 表1に各顔表情に対するJAFFEデータセットおよびMMIデータセットの内訳を示す.

3.2 前実験

2.4節で述べた5つのパターン識別器のパラメータが良好となるように, JAFFEデータセットおよびMMIデータセットを合わせた全データに対して10-fold cross validation[23]を適用した. このとき, 学習データとテストデータに同一人物の対象となる表情画像を含めないように設定した. 各パターン識別器により算出された平均判別率を図5に示す. 図5において, 縦軸は平均判別率, 横軸は各識別器において判別率に最も大きな影響を与えた変数である. その他の判別率への影響の小さなパラメータについては, 事前実験により良好な値となるように設定している. また, 図5の平均判別率(Cr)は以下の式により算出している.

$$Cr = \sum_{i=1}^6 \frac{Num_i \times Cr_i}{Total_i} \quad (3)$$

Num_i は表情 i の事例数, Cr_i は表情 i の判別率, $Total_i$

表1 各データセットに含まれる顔表情の内訳

expression	JAFFE	MMI
happiness	31	40
disgust	29	22
sadness	31	22
anger	30	22
surprise	30	34
fear	32	20

図5 各パターン識別器における良好なパラメータの設定

は表情 i のサンプル数である。また、 Cr_i の計算式を以下に示す。

$$Cr = \sum_{i=1}^6 \frac{Num_i}{Total_i} \quad (4)$$

ここで、 Num_i は表情 i であると正しく判別されたサンプル数である。

以上の結果より、各パターン識別器の良好なパラメータは、Random Forestでは431個の深さ無制限の木、Logistic RegressionではLogistic Model Trees[18]、ブースト回数37~41回、heuristic stopは45、Neural Networkでは学習回数199~213回、活性化関数にシグモイド関数、学習に誤差逆伝播法(学習率0.27、衰退重み0.2)、Support Vector MachineではPolynomialカーネル、 $\gamma=0.2$ 、重み $C=1.0$ であることがわかった。

3.3 実験結果

3.2節の前実験で決定した各パターン識別器のパラメータを用いて、JAFPEデータセットおよびMMIデータセットのそれぞれに対して提案手法を適用し、各パターン識別器により算出された平均判別率を表2に示す。表2において、*Combined* は3.2節で述べたパターン識別器のパラメータを用いて、各表情に適した識別器を用いたことを示している。また、JAFPEデータセットに対する分類表を表3に示し、MMIデータセットに対する分類表を表4に示す。

表5および表6に既存方法[9]と提案手法の再現率(*recall*)と適合率(*precision*)を示す。ここで、既存手法[9]の特徴抽出では、最初のフレームと中央のフレームを用いていることから、同じ条件にするために、提案手法における表情フレームを中央のフレームとして設定している。これらの表において、*recall*と*precision*はそれぞれ以下の式で求めている。

$$recall = \frac{N_{output} \cap N_{category}}{N_{category}} \times 100 \quad (5)$$

$$precision = \frac{N_{output} \cap N_{category}}{N_{output}} \times 100 \quad (6)$$

$N_{category}$ は各顔表情における事例数、 N_{output} は提案手法により判定された事例数である。表5および表6より、既存手法と比べて、再現率・適合率ともに提案手法が優位であることがわかる。また、再現率と適合率に有意差が無いこと、基本6表情の中で喜びと驚きの平均判別率は比較的高く、恐れのリ再現率が最も低いことがわかる。

表2 各データセットに対する各パターン識別器の平均判別率 (%)

tools	JAFPE	MMI
Nearest Neighbor	65.1	52.5
Random Forest	70.8	59.4
Logistic Regression	74.1	66.3
Support Vector	71.2	64.0
Neural Network	74.6	66.9
<i>Combination</i>	78.7	71.9

表3 JAFPEデータセットに対する分類表

	output by the proposed method					
	hap.	dis.	sad.	ang.	sur.	fear
hap.	29	0	0	0	1	1
dis.	0	25	2	1	0	1
sad.	3	3	20	2	3	0
ang.	0	2	4	22	0	2
sur.	1	0	0	0	29	0
fear	2	2	5	2	2	19

表4 MMIデータセットに対する分類表

	output by the proposed method					
	hap.	dis.	sad.	ang.	sur.	fear
hap.	34	1	0	0	3	2
dis.	1	13	3	2	1	2
sad.	3	4	12	2	1	0
ang.	0	1	2	16	0	3
sur.	1	0	0	0	31	2
fear	1	3	3	1	3	9

表5 提案手法と既存手法[9]のJAFPEデータセットに対する再現率*recall*と適合率*precision* (%)

	existing method		proposed method	
	<i>recall</i>	<i>precision</i>	<i>recall</i>	<i>precision</i>
express.				
hap.	83.9	72.2	93.5	82.9
dis.	69.0	66.7	82.8	75.0
sad.	54.8	70.8	69.0	60.6
ang.	73.3	62.9	70.0	80.8
sur.	66.7	64.5	96.7	87.9
fear	56.3	66.7	56.3	81.8
<i>Cr</i>	67.2	67.3	77.8	78.1

表6 提案手法と既存手法[9]の MMI データセットに対する再現率 *recall* と適合率 *precision* (%)

express.	existing method		proposed method	
	<i>recall</i>	<i>precision</i>	<i>recall</i>	<i>precision</i>
hap.	82.5	73.3	85.0	85.0
dis.	59.1	65.0	59.1	56.5
sad.	54.5	80.0	50.0	57.9
ang.	63.6	60.9	72.7	76.2
sur.	61.8	63.6	91.2	77.5
fear	15.0	12.5	40.0	47.1
<i>Cr</i>	60.0	61.7	70.6	69.8

3.4 既存のパターン識別器との比較実験

表情認識における従来法の同種のパターン識別器を用いた場合との比較について述べる。従来法との比較には、動画像の時系列データや3次元情報を対象としたもの、および、従来法の特徴量に特化したものを除いて、Adaboost[24]、Gentle Adaboost[25]、 k 近傍法($k=3$)が挙げられる[26]。また、既存手法[9]の one-versus-one 法を用いたサポートベクターマシンも比較対象とする。one-versus-one法では、表情*i*と表情*j*に対するサポートベクターマシンの判別を、すべての*i*と*j*の組み合わせだけ行い(ただし、 $i \neq j$)、得票数の最も多い表情を決定する。例えば、まず happiness と disgust の判別を行い、happiness であると判別されれば happiness と sadness の判別が行われ、disgust であると判別されれば disgust と sadness を判別する、という動作を繰り返す。表7は、表2に示した提案手法の平均判別率 *Cr* と上記のパターン識別器の *Cr* である。表7において、Adaboost と Gentle Adaboost の組み合わせには、表2の5つのパターン識別器の中で *Cr* が最も高くなった Neural Network を用いている。また、各パターン識別器は、3.2節に述べた方法で良好なパラメータの選択を行っており、Adaboost では反復回数31回、Gentle Adaboost では反

表7 表情認識における従来のパターン識別器を用いた場合の平均判別率 *Cr* (%)

tools	JAFFE	MMI
the proposed method	78.7	71.9
Adaboost	75.1	68.0
Gentle Adaboost	75.4	79.2
one-veresus-one	73.4	65.9
k nearest neighbor ($k=3$)	61.5	48.3

復回数27回、one-versus-one法のサポートベクターマシンでは3.2節で述べた良好なパラメータを用いている。表7より、従来法と比べて、提案手法が優位であることがわかる。

4. 考察

4.1 各顔表情に対する特徴量の有効性

提案した式(1)のゼロ平均正規化相関 R_{ZNCC} に基づく特徴量について述べる。まず、表8に2.3節で述べたROIから抽出した R_{ZNCC} の平均および標準偏差()内の数値)を示す。表8より、各顔表情間の平均値に差が殆ど無いことがわかる。この理由は、表8の各項目における R_{ZNCC} の標準偏差が0.11~0.31の範囲でばらついているからであると考えられる。

そこで、画像一枚ずつの R_{ZNCC} を比較するために、ランダムに選択した2種類の顔画像40枚について、額に設定したROIから抽出された R_{ZNCC} の相関係数を表9に示す。一般には相関係数が-0.4以下では負の相関があり、-0.4~0.4では相関が無く、0.4以上では正の相関があると見なすことができる。つまり、負の相関がある場合や相関が無い場合には、各表情間の R_{ZNCC} が類似していないため表情判別が容易となり、正の相関がある場合には、各表情間の R_{ZNCC} が類似しているため表情判別が困難となる。したがって、表9において、 R_{ZNCC} の相関係数が0.4未満の場合には、各表情間の R_{ZNCC} が類似していないことから、額に設定した R_{ZNCC} が特徴量の一つとして、表情判別に対して

表8 全テストデータにおける ROI から抽出された R_{ZNCC} の平均と標準偏差 (①:額, ②:眉間, ③:鼻唇溝, ④:首)

	①	②	③	④
hap.	0.41 (0.24)	0.46 (0.22)	0.71 (0.23)	0.74 (0.26)
dis.	0.25 (0.22)	0.33 (0.22)	0.67 (0.29)	0.67 (0.29)
sad.	0.37 (0.21)	0.36 (0.20)	0.80 (0.25)	0.84 (0.19)
ang.	0.18 (0.25)	0.33 (0.23)	0.71 (0.31)	0.88 (0.11)
sur.	0.38 (0.24)	0.31 (0.24)	0.77 (0.26)	0.86 (0.16)
fear	0.32 (0.20)	0.30 (0.24)	0.81 (0.21)	0.83 (0.25)

表9 額に設定したROIから抽出された R_{ZNCC} の相関係数

	hap.	dis.	sad.	ang.	sur.	fear
hap.	-	0.11	0.10	0.01	0.43	0.30
dis.	-	-	0.25	0.16	0.27	-0.22
sad.	-	-	-	0.37	-0.17	-0.36
ang.	-	-	-	-	-0.16	-0.33
sur.	-	-	-	-	-	0.18

表10 眉間に設定したROIから抽出された R_{ZNCC} の相関係数

	hap.	dis.	sad.	ang.	sur.	fear
hap.	-	0.54	0.03	0.65	-0.04	0.18
dis.	-	-	0.07	0.32	-0.08	-0.08
sad.	-	-	-	-0.23	0.06	-0.19
ang.	-	-	-	-	-0.46	0.37
sur.	-	-	-	-	-	-0.38

表11 鼻唇溝(図2の15)に設定したROIから抽出された R_{ZNCC} の相関係数

	hap.	dis.	sad.	ang.	sur.	fear
hap.	-	0.05	-0.24	0.05	0.24	0.47
dis.	-	-	-0.26	0.12	0.54	-0.07
sad.	-	-	-	0.42	-0.03	-0.15
ang.	-	-	-	-	0.13	-0.03
sur.	-	-	-	-	-	0.42

表12 鼻唇溝(図2の16)に設定したROIから抽出された R_{ZNCC} の相関係数

	hap.	dis.	sad.	ang.	sur.	fear
hap.	-	0.55	-0.19	0.07	0.00	0.48
dis.	-	-	0.01	-0.20	0.14	0.07
sad.	-	-	-	-0.26	0.16	-0.17
ang.	-	-	-	-	-0.43	0.00
sur.	-	-	-	-	-	-0.03

有効性があると考えられる。表9では、幸福と驚きのペアを除いて、 R_{ZNCC} の相関が無いことがわかる。同様に、2.3節で述べた②眉間、③鼻唇溝(図2の15)、④鼻唇溝(図2の16)、④首に設定したROIから抽出された R_{ZNCC} の相関係数をそれぞれ表10、表11、表12、表13に示す。これらの表において相関があることを確

表13 首に設定したROIから抽出された R_{ZNCC} の相関係数

	hap.	dis.	sad.	ang.	sur.	fear
hap.	-	-0.22	-0.10	-0.16	0.02	0.12
dis.	-	-	0.12	0.12	-0.04	0.30
sad.	-	-	-	-0.48	0.12	0.12
ang.	-	-	-	-	-0.02	-0.01
sur.	-	-	-	-	-	0.11

認できるのは、表10の幸福と嫌悪のペアおよび幸福と怒りのペア、表11の嫌悪と驚きのペアおよび幸福と恐れペア、悲しみと怒りのペア、および驚きと恐れペア、表12の幸福と嫌悪のペアおよび幸福と恐れペアである。また、表13では相関のあるペアが存在しないことがわかる。相関があるケースをまとめると、悲しみと怒りのペアの1ケースを除いて、すべてのケースにおいて一方の表情が幸福もしくは驚きであることがわかる。つまり、各部位に設定した R_{ZNCC} は、幸福や驚きを除いて有効であると言える。一方、幸福の表情表出における「口角が上がる」という動作や、驚きの表情表出における「両眉毛が大きく上がる」や「口が大きく開く」という動作は、他の表情では比較的現れないことから、幸福や驚きの判別率は比較的高くなっていると考えられる。

以上より、個々の画像に着目した場合には特徴量に有意差が表れることがわかった。しかし、特徴量のばらつきが大きく、平均的な特徴量の差が乏しいため、安定した特徴量を定義する必要がある、これは今後の課題とする。

4.2 基本6表情の判別結果

表5、表6より、既存手法と提案手法のいずれにおいても、MMIデータセットと比べて、JAFPEデータセットの再現率と適合率はともに高いことがわかる。この理由として、JAFPEデータセットでは日本人の女性のみを対象としているが、MMIデータセットでは複数の人種や男女を対象としており、MMIデータセットのほうが表情の種類が豊富であるため判別が難しいためであると考えられる。また、表3、表4の分類表より、JAFPEデータセットと比べて、誤判別となった表情のばらつきが大きいことがわかる。

表3、表4の分類表より、本手法において誤判別となったケースについて述べる。まず、これらの表より、happinessとsurpriseの判別では、happiness-surprise間で誤判別となっているケースとfearであると誤判別されているケースが大半を占めることがわか

る。次に、表3、表4の分類表より、angerは、fear, sadness, disgustであると誤判別される傾向があることがわかる。表5、表6より、基本6表情の中でfearの判別率は最も低いことがわかる。表3、表4の分類表では、fearの多くがsadness, disgust, surpriseであると誤認識されていることから、fearの顔画像から抽出した特徴量がsadness, disgust, surpriseと類似していることがわかる。表4では、sadnessおよびdisgustの判別率が比較的低い。この理由は、MMIデータセットにおいて、これらの表情画像から抽出した特徴量 l_{xn} と l_{yn} の相関係数が0.45であり、disgustとsadnessの表情が似ているからであると考えられる。以上より、fear, sadness, disgustの3表情の判別に対して、より有効な特徴量を新たに検討する必要がある。

図6に提案手法により正しく判別された画像の一部を手動で切り取った顔画像を示す。これらの画像は、人間にとって比較的判別が容易なものである。次に、しわが顔画像に表出しないため、特徴量が有効的に機能しない事例を図7に示す。図7の(a)、(b)は、それぞれMMIデータセットの画像の一部であるが、首元が衣服で隠れていることがわかる。図7(c)は、JAFPEデータセットの画像中の一枚であり、恐れを表出しているが、しわが表出しておらず、首元も衣服で隠れていることがわかる。このように、提案手法が有効的に

図6 提案手法により正しく判別された顔画像

図7 提案した得著両が有効的に機能しない顔画像

表14 提案手法の処理時間（単位：秒）

processing	proposed method	
	JAFPE	MMI
histogram stretching	—	—
face detection	0.031	0.120
facial feature detection	0.048	0.188
feature extraction	—	—
facial expression recognition	—	—

機能しない画像に対しては、各表情に特化した別の特徴量を提案する必要があると考えられる。例えば、怒りであればにらみを認識すること、悲しみであればうつむき度合いを検出することなどが挙げられる。

4.3 提案手法の実時間性

表14に提案手法の処理時間を示す。この計測に用いた計算機のCPUはIntel Core i7 3.07 Ghz、メモリは2.0 GB RAMである。並列処理やGPU処理などの特殊な高速化は行っていない。表14において、「—」は処理時間が0.001未満であることを示している。表14より、処理時間を要するのは顔および顔特徴点の抽出であることがわかる。また、JAFPEデータセットとMMIデータセットの処理時間に差があるが、これはJAFPEデータセットの画像サイズが縦256ピクセル、横256ピクセルであるのに対して、MMIデータセットの画像サイズは縦720ピクセル、横576ピクセルと大きいからであると考えられる。尚、既存手法[9]の処理時間は、提案手法における顔および顔特徴点の抽出時間に加えて、MMIデータセットでは2.1秒、JAFPEセットでは0.56秒かかることが報告されている。

5. まとめ

本論文では、表情筋の動きに伴うしわから特徴を抽出し、顔画像から基本6表情を判別する方法を提案した。実験では、喜びと驚きの平均判別率が80%以上となり、既存手法よりも高い精度で表情を認識できることを示した。また、本手法の計算コストは小さく、リアルタイム処理を比較的簡単に実現できることを示した。

提案手法では、同一人物の無表情フレームおよび表情フレームの2枚から表情判別を行うものであり、その前処理として、静止画像の場合には個人の顔認識を、動画画像の場合には無表情の状態の特定を、それぞれ実行する必要がある。一方、初めて見る人物の表情フレーム1枚からでも、人間は表情を判断することができる。このことから、今後の課題として、個人の顔認識および表情フレームから無表情の状態を推定する

ことが求められる。その他、(1)特徴量を抽出するための対象領域の大きさを自動設定すること、(2)最適なパターン識別器の組み合わせを定量的に評価する方法を検討すること、(3)提案した特徴量の分散を小さくすること、(4)各顔表情に特化した特徴量を提案すること、(5)時系列データに対応し、かつリアルタイム処理が可能な手法を提案すること、などが今後の課題として挙げられる。

参考文献

- [1] A. Mehrabian, S. R. Ferris, "Inference of Attitudes from Nonverbal Communication in Two Channels," *Journal of Consulting Psychology*, vol.31, no.3, pp.248-252, Jun 1967.
- [2] B. Fasel and J. Luetttin, "Automatic Facial Expression Analysis: Survey," *Pattern Recognition*, vol.36, no.12, pp.259-275, Jan. 2003.
- [3] P. Ekman, "Darwing and Facial Expression: A Century of Research in Review," Academic Press, New York, June 1973.
- [4] P. Ekman and W. Friesen, "Facial Coding System: A Technique for the Measurement of Facial Movement," Consulting Psychologists Press, Palo Alto, CA, 1978.
- [5] 崔昌石, 原島博, 武部幹, "顔の3次元モデルを用いた顔面表情の分析," 電子情報通信学会論文誌 (D-II), vol.J74-D-II, no.6, pp.766-777, June 1991.
- [6] K. Mase, "Recognition of Facial Expression from Optical Flow," *IEICE Trans.*, vol.E74, no.10, pp.3474-3483, Oct. 1991.
- [7] 太田寛志, 佐治斉, 中谷広正, "顔面筋に基づいた顔構成要素モデルによる表情変化の認識," 電子情報通信学会論文誌 (D-II), vol.J82-D-II, no.7, pp.1129-1139, July 1999.
- [8] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, "Facial Expression Recognition from Video Sequences: Temporal and Static Modeling," *Computer Vision and Image Understanding*, vol.91, no.1-2, pp.160-187, July 2003.
- [9] 野宮浩揮, 宝珍輝尚, "顔特徴量の有用性推定に基づく特徴抽出による表情認識," 日本知能情報ファジィ学会誌, vol.23, no.2, pp.170-185, Apr. 2011.
- [10] 坂口竜己, 森島繁生, "画像の2次元コサイン変換を利用した実時間顔表情認識," 電子情報通信学会論文誌 (D-II), vol.J80-D-II, no.6, pp.1547-1554, June 1997.
- [11] 大谷淳, "遺伝的アルゴリズムを用いた顔画像からの表情と顔の姿勢の推定," 電子情報通信学会論文誌 (D-II), vol.J80-D-II, no.8, pp.2152-2160, Sept. 1997.
- [12] L. Ma and K. Khorasani, "Facial Expression Recognition Using Constructive Feedforward Neural Networks," *IEEE Trans. on Systems, Man and Cybernetics*, vol.34, no.3, pp.1588-1595, June 2004.
- [13] G. Guo and C. R. Dyer, "Learning from Examples in the Small Sample Case: Face Expression Recognition," *IEEE Trans. on Systems, Man and Cybernetics*, vol.35, no.3, pp.477-488, June 2005.
- [14] 熊野史朗, 大塚和弘, 大和淳司, 前田英作, 佐藤洋一, "パーティクルフィルタとこう配法の組合せによる頭部姿勢と表情の同時推定," 電子情報通信学会論文誌 (D-II), vol.J92-D-II, no.8, pp.1349-1362, Aug. 2009.
- [15] Luxand FaceSDK 4.0, <http://www.luxand.com/facesdk/>
- [16] M. Gabel and C. Weihs, "Calibrating Margin-Based Classifier Scores into Polychotomous Probabilities," *Data Analysis, Machine Learning and Applications*, pp.29-36, Jan. 2008.
- [17] L. Breiman, "Random Forests," *Machine Learning*, vol.45, no.1, pp.5-32, Jan. 2001.
- [18] N. Landwehr, M. Hall, and E. Frank, "Logistic Model Trees," *Machine Learning*, vol.59, no.1-2, pp.161-205, 2005.
- [19] C. Cortes and V. Vapnik, "Support-vector Networks," *Machine Learning*, Vol.20, No.3, pp.273-297, Sept. 1995.
- [20] M. F. Kelly, P. A. Parker, and R. N. Scott, "The Application of Neural Networks to Myoelectric signal Analysis: A Preliminary Study," *IEEE ASSP Mag.*, vol.4, no.2, pp.4-22, Mar. 1987.
- [21] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic Classification of Single Facial Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.21, no.12, pp.1357-1362, Sept. 1999.
- [22] M. Pantic, M. F. Valstar, R. Rademaker, and J. Gyoba, "Web-Based Database for Facial Expression Analysis," *Proc. of IEEE International Conference on Multimedia and Expo*, pp.317-321, 2005.
- [23] M. Fosteller, "A k -sample Slippage Test for an Extreme Population," *Annals of Mathematical Statistics*, vol.19, no.1, pp.58-65, Mar. 1948.
- [24] Y. Freund, R. Schapire, "A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, vol.55, no.1, pp.119-139, Aug. 1997.
- [25] A. Torralba, K. P. Murphy, W. T. Freeman, "Sharing Features: Efficient Boosting Procedures for Multiclass Object Detection," *Proc. of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.762-769, 2004.
- [26] Z. Zhuohong, M. Pantic, G. I. Roisman, T. S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.31, no.1, pp.39-58, Jan. 2009.

(2011年12月14日 受付)

(2012年 4月30日 採録)

[問い合わせ先]

〒920-1192 石川県金沢市角間町

金沢大学 自然科学研究科

中村 宗広

TEL: 090-9864-6156

FAX: 06-6359-0158

E-mail: m-nakamura@blitz.ec.t.kanazawa-u.ac.jp

著者紹介

なかむら むねひろ
中村 宗広 [非会員]

2008年近畿大学理工学部情報学科卒業。2010年同大学大学院総合理工学研究科博士前期課程修了。2011年金沢大学大学院自然科学研究科博士後期課程修了。工博。現在、同大学理工研究域電子情報学系助教。パターン認識、医用画像処理に関する研究に従事。情報処理学会、電子情報通信学会、IEEE、映像情報メディア学会各会員。

かじわら ゆうすけ
梶原 祐輔 [非会員]

2009年東京電機大学理工学部情報システム工学科卒業。2011年金沢大学大学院自然科学研究科博士前期課程修了。現在、同大学大学院自然科学研究科博士後期課程に在籍。

むらた ひろあき
村田 裕章 [非会員]

2010年9月金沢大学大学院自然科学研究科博士後期課程修了。博士(工学)。同年10月、同大学大学院研究協力員。同年12月、金沢大学理工研究域博士研究員。2012年2月同大学環日本海域環境研究センター博士研究員。現在に至る。電気学会、電子情報通信学会、情報処理学会、人工知能学会各会員。

きむら はるひこ
木村 春彦 [非会員]

1979年東北大学工学部研究科博士課程修了。同年富士通(株)勤務。1980年金沢女子短期大学講師。1984年金沢大学経済学部助教授。1992年同大学工学部電気・情報工学科助教授。1994年同学科教授。現在、同大学理工研究域電子情報学系教授。工博。ソフトコンピューティングの応用や独居老人の介護支援に関する研究に従事。情報処理学会、電気学会、人工知能学会各会員。

Facial Expression Recognition Using Features of Density Distributions between Facial Images

by

Munehiro NAKAMURA, Yusuke KAJIWARA, Hiroaki MURATA and Haruhiko KIMURA

Abstract :

This paper presents a method of discriminating human facial expressions using 5 pattern classifiers, where features based on density distributions are extracted from target regions set in facial feature points. Observing human facial expressions in videos, we could find that wrinkles appear in regions of correlated mimic muscles. The proposed method extracts the degree of similarity based on Zero-Mean Normalized Cross-Correlation as features from target areas where wrinkles often appear. And, 5 pattern classifiers, namely, Nearest Neighbor, Random Forests, Logistic Regression, Support Vector Machine, and Neural Network, are applied to discrimination of 6 basic facial expressions.

The efficiency of the proposed method has been confirmed using public facial expression databases.

Keywords : Facial Expression Recognition, Facial Image, Mimic Muscle, Zero-mean Normalized Cross-Correlation

Contact Address : **Munehiro NAKAMURA**

*Graduate School of Natural Science and Technology, Kanazawa University
Kakuma, Kanazawa, Ishikawa, 920-1192, JAPAN*

TEL : 090-9864-6156

FAX : 06-6359-0158

E-mail : m-nakamura@blitz.ec.t.kanazawa-u.ac.jp