

顔キーポイントの移動方向コードに基づく 個人差の影響を受けにくい表情認識

非会員 佐々木康輔* 非会員 渡邊 健斗**
正員 橋本 学*^{a)} 正員 長田 典子***

Person-invariant Facial Expression Recognition based on Coded Movement Direction of Keypoints of Facial Parts

Kosuke Sasaki*, Nom-member, Kento Watanabe**, Nom-member,
Manabu Hashimoto*^{a)}, Member, Noriko Nagata***, Member

(2017年3月31日受付, 2017年12月30日再受付)

This paper describes a person-invariant method of classifying subtle facial expressions. The method uses keypoints detected by using a face tracking tool called “Face Tracker”. It describes features such as coded movements of keypoints and uses them for classification. Its classification accuracy was evaluated using the facial images of unlearned people. The results showed the average F-measure was 0.93 for neutral (expressionless) facial images, 0.78 for subtle smile images, and 0.93 for exaggerated smile images.

キーワード: 微妙な表情, 個人依存性, 顔キーポイント, 表情認識

Keywords: subtle facial expression, person invariance, facial parts keypoints, facial expression recognition

1. はじめに

近年, さまざまな場面で, 生活の質の向上に関心が集まっている (Quality of Life)。そのためには, テレビ番組や日用品などに対する楽しい, 面白いといった感情を推定し, それらの情報を制作者にフィードバックする必要がある。

また, 大量にあるライフログデータから幸せそうな笑顔や悔しそうな泣き顔といった特定の写真データを簡易に検索する技術や, ライフログを分析して日常生活を見直すことが求められている。そのためには, 笑顔や泣き顔といった感情や感情の高低を推定する技術が必要になる。感情を推定する研究としては, 接触センサを用いる研究⁽¹⁾と非接触センサを用いる研究⁽²⁾がある。接触センサを用いると, 被験者がセンサを装着することによって違和感を覚えるため, 自然な感情が表れにくいという問題点がある。そこで, 自然な感情を推定するため, 非接触センサによって取得される顔の表情を応用するものが多い。

従来の表情識別手法に用いられる特徴量として, 大きくわけて2つのアプローチがある。1つは, 顔の濃度値を特徴量とするフィルタベースの手法である。もう1つは, 顔の各部位の形状変化を用いる幾何学ベースの手法である。

フィルタベースの手法としては, HoG 特徴量⁽²⁾や Gabor wavelets⁽³⁾, Local Binary Pattern⁽⁴⁾を用いた表情表出時の顔面上の局所的な特徴量を捉える手法がある。これらの手法は, 顔の局所的な位置に注目しているため, 顔パーツの位置ずれによって認識精度が低下し, 識別対象者の顔つきによっ

a) Correspondence to: Manabu Hashimoto. E-mail: mana@isl.sist.chukyo-u.ac.jp

* 中京大学 大学院 情報科学研究科
〒466-8666 愛知県名古屋市中区八事本町 101-2
Graduate School of Computer and Cognitive Sciences, Chukyo University
101-2, Yagoto-Honmachi, Showa-ku, Nagoya, Aichi 466-8666, Japan

** 中京大学 工学部
〒466-8666 愛知県名古屋市中区八事本町 101-2
School of Engineering, Chukyo University
101-2, Yagoto-Honmachi, Showa-ku, Nagoya, Aichi 466-8666, Japan

*** 関西学院大学 大学院 理工学研究科
〒669-1337 兵庫県三田市学園 2-1
Graduate School of Science and Technology, Kwansei Gakuin University
2-1, Gakuen, Sanda, Hyogo 669-1337, Japan

て表情識別が困難になると考えられる。

幾何学ベースの手法では、顔の一部分の動きに注目した手法⁽⁵⁾や、オプティカルフローから顔の皮膚の動きを検出し表情を識別する手法⁽⁶⁾がある。これらの手法は、顔全体の大局的な動きを捉える手法であるため、目の瞬き、口の変形などの局所的な動きである表情の機微を捉えることができない。

このように従来研究は、識別可能な表情が多いという利点がある一方で、実用性の観点からは次の2つの課題が残っている。

第1の課題は、表情の強弱を認識することである。表情が強く表出しているときは感情の高まりが大きく、弱く表出しているときは感情の高まりが小さい。感情の起伏を計測するためには、表情の強弱を計測する必要がある。

第2の課題は、複数人からなる学習データの個人差の影響を受けないようにすることである。実利用する上では、不特定多数の人が表情識別の対象となるため、識別器が学習対象者に影響されないことが要求される。

第1の課題を解決するために、微妙な表情と大仰な表情の識別を目的とした研究としては、松久らの研究⁽⁷⁾と野宮らの研究⁽⁸⁾、画像処理分野において注目を浴びているDeepLearning⁽⁹⁾⁽¹⁰⁾の3つがある。

松久らは、表情の微細な変化を捉えることができるGabor Filterの濃度応答値を特徴量とし、Adaboostを用いて微妙な笑顔と大仰な笑顔の表情を識別する手法⁽⁷⁾を提案した。しかしながら、この手法は、Gabor Filterの濃度応答値が顔の3次元起伏を濃淡情報に反映したものであることに起因して、識別性能が学習者個人に依存しやすいという問題点がある。

野宮らは、顔の主要な各部位から検出されるキーポイントを組み合わせた幾何学特徴量を用いて、表情の強弱を識別する手法⁽⁸⁾を提案した。しかしながら、人ごとに顔の部位の配置が異なることから、松久らの手法と同様に学習者個人に依存しやすいという問題点がある。

DeepLearningは、フィルタベースの手法であり、画像を構成する画素値を部分領域として扱い、多段階で画素値を畳み込むことによって、画像自体を1つの高次元特徴量として学習する手法である。この手法は、画像における微細な変化を高次元特徴量として捉えることができるため、表情の強弱を認識することに適した手法である。しかしながら、フィルタベースの手法であるため、フィルタを適用する位置に大きく影響を受ける。表情の微細な変化の出現する位置が個人ごとに異なるため、学習によって生成された識別器は学習データに使用した人の影響を受けやすい。

以上のように、第1の課題に関しては、識別性能が学習対象者の顔形状に依存しやすいという課題が残っている。

第2の課題を解決するために、個人差の影響の軽減を目的とした研究としては、FaisalがCompound Local Binary Pattern特徴量を用いた個人の顔形状に依存しない表情識別手法⁽¹¹⁾を提案した。この手法は、特徴量が顕著に変化する部分を

選択して表情識別に用いるため、微妙な表情を認識できないという問題点がある。

以上のように、第2の課題に関しては、大仰な表情と微妙な表情の識別という課題が残っている。

そこで本研究では、表情の強弱の認識と個人差の影響の軽減という相反する2つの課題を解決する手法として、顔の局所的な位置における動き情報に注目した表情識別手法を提案する。まず、顔の部位の局所的な位置における変化が人に共通であることに注目し、顔の局所的な位置である口角や目頭から検出されたキーポイントの移動ベクトルを用いる。しかし、算出された移動ベクトルの角度と大きさは個人差があるため、2つのアイデアに基づき、局所的な位置における顔キーポイントの移動ベクトルをコードとして表現した移動方向コードを提案する。さらに、表情ごとに不特定多数の人の共通する特徴量のみを選択することによって、個人差の影響を受けずに喜び表情を識別する。本研究では、ニュートラルな感情である無表情と、微妙な笑顔、大仰な笑顔の3種類の喜び表情を対象とする。

2. 基本アイデア

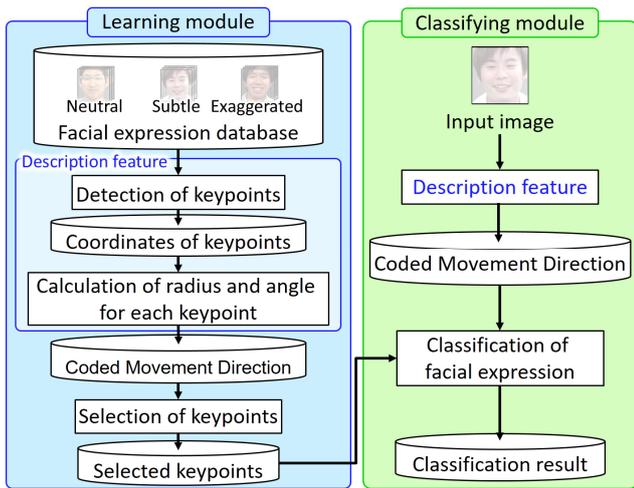
Ekmanらの研究⁽¹²⁾より、主要な感情にはそれぞれ対応する特有の表情があるとされている。これらの感情は、人種、性別、文化に関係なく共通するといわれている。例えば、喜びという感情では、唇の両端が後ろに引かれ、下唇が下に動くといった変化がある。この変化は、人に対して共通性が確認されている。そこで、本研究では、人に対する共通性がある顔の変化に着目する。

顔のパーツ全体に注目すると、個人ごとに筋肉、骨格が異なることから表情表出時の形状の違いがある。そこで、顔画像の目、口といった各主要パーツ内の局所的な位置に注目する。無表情時から表情表出時までの局所的な位置における顔キーポイントの移動ベクトルから特徴量を設計する。算出された移動ベクトルの大きさの値と角度値を用いると、個人差の影響を受け、また微妙な笑顔の識別も困難になる。そこで、算出された移動ベクトルの大きさの値と角度値をそれぞれ量子化し、2つの値をコードによって表現した移動方向コード特徴量を提案する。この特徴量を用いることにより、個人差の影響を受けにくく、微妙な笑顔を認識する。

3. 提案手法

〈3・1〉 概 要 提案手法の流れを Fig. 1 に示す。

提案手法は、学習モジュールと識別モジュールの2つのモジュールからなる。学習モジュールでは、人に汎用的なキーポイントの検出手法⁽¹³⁾を用いて、顔の部位からキーポイントを検出する。キーポイント検出手法⁽¹³⁾については、〈3・2〉節で述べる。また、検出した顔キーポイントを補正することによって、顔キーポイントの移動ベクトルを顔の位置や大きさ、回転の影響を受けにくくする。詳細な補正方法については、〈3・3〉節で述べる。



個人情報保護の観点より加工してある。

Fig. 1. Block diagram of proposed method.

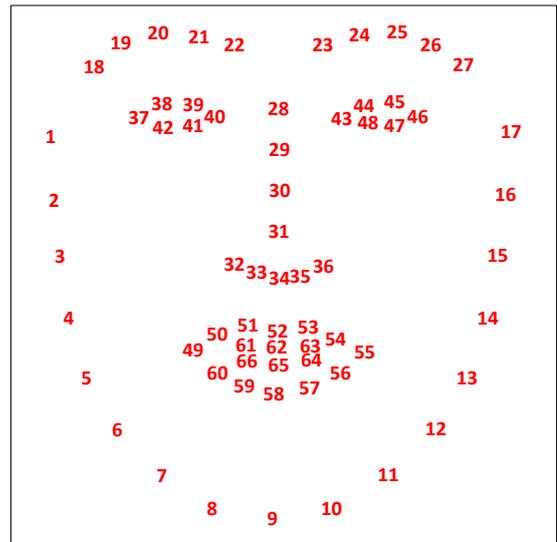


Fig. 2. Keypoint number.

次に、無表情時と表情表出時のキーポイントからなる移動ベクトルを算出する。移動ベクトルの角度成分と大きさ成分をコード化した移動方向コードを特徴量として定義する。複数人からなる学習画像データセットから、各キーポイントにおいて特徴量を算出し、表情ごとに特徴量の発生頻度のヒストグラムを生成する。全てのキーポイントから、表情識別に有効であり、人に対して共通性が高い特徴量を有するものを選択して識別に用いる。

識別モジュールでは、学習モジュールと同様にキーポイントを検出して、キーポイントの移動方向からなる特徴量を算出する。学習モジュールにて選択されたキーポイントにおける表情ごとの特徴量の発生確率値を用いて、各表情の尤度を求める。この尤度から表情を識別する。

〈3・2〉 顔キーポイントの検出手法 顔画像からキーポイントを検出するために、Jason らが提案した FaceTracker⁽¹³⁾ を用いる。この手法は、Constrained Local Model を用いて、顔の目、鼻、口、輪郭といった主要な器官からキーポイントを検出し、追跡する手法である。また、人に対して汎用的な手法である。Fig. 2 に顔画像から抽出したキーポイントを示す。検出される顔キーポイントの総数は 66 点である。

この手法によって検出された顔キーポイントは顔形状を表すモデルに基づいており、このモデルは顔の首振り方向、傾き方向、カメラに対する光軸方向の 3 次元の回転角を有している。顔形状を表すモデルと顔キーポイントの座標を示す式を(1)式に示す。

$$\mathbf{X} = s\mathbf{R}(\bar{\mathbf{X}} + \Phi\mathbf{q}) + \mathbf{t} \quad \dots\dots\dots(1)$$

\mathbf{X} は顔キーポイントの座標ベクトル、 s はスケール、 \mathbf{R} は顔の 3 次元の回転行列、 $\bar{\mathbf{X}}$ は顔キーポイントの座標の平均ベクトル、 Φ は主成分分析によって求められた基底ベクトル、 \mathbf{q} は顔の変形を表すベクトル、 \mathbf{t} は平行移動を表している。

この式のパラメータから、顔キーポイントを補正する。

〈3・3〉 顔キーポイントの補正 顔の 3 次元の回転角を示している(1)式の回転行列 \mathbf{R} を単位行列に変換し、顔キーポイントの座標を算出すると、正面顔画像から顔キーポイントを検出することと同義になる。これにより、顔の 3 次元回転を補正する。この処理を(2)式に示す。また、画像中の顔の大きさが異なることによって、顔キーポイント間の距離も異なる。そこで、検出した顔キーポイントのうち、左右の目頭の点番号 40 と点番号 43 の距離 l を求める。この処理を(3)式に示す。すべての顔画像において、点番号 40 と点番号 43 の距離 l を同じにすることによって、顔の大きさによる個人差を軽減する。さらに、左右の目頭の 2 点(点番号 40 と点番号 43) と表情表出時に安定して検出できる上唇の 1 点(点番号 52) の計 3 点からなる重心 G を、顔キーポイントの座標軸の原点となるように全ての顔キーポイントを補正した。この処理を(4)式に示す。

$$\mathbf{X} = s\mathbf{R}\mathbf{R}^{-1}(\bar{\mathbf{X}} + \Phi\mathbf{q}) + \mathbf{t} \quad \dots\dots\dots(2)$$

$$l = \sqrt{(x_{43} - x_{40})^2 + (y_{43} - y_{40})^2} \quad \dots\dots\dots(3)$$

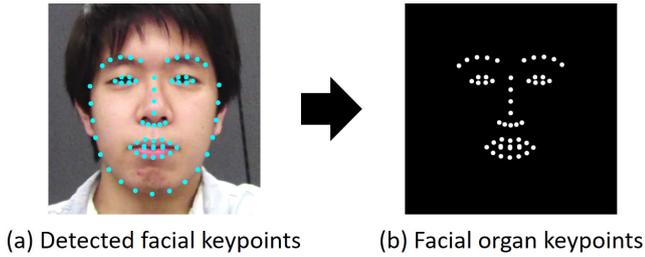
$$g_x = \frac{x_{40} + x_{43} + x_{52}}{3}, \quad g_y = \frac{y_{40} + y_{43} + y_{52}}{3} \quad \dots\dots\dots(4)$$

$$x'_n = x_n - g_x, \quad y'_n = y_n - g_y \quad (n:1,2,\dots,66)$$

x_n は点番号 n の顔キーポイント x 座標であり、 y_n は点番号 n の顔キーポイント y 座標である。

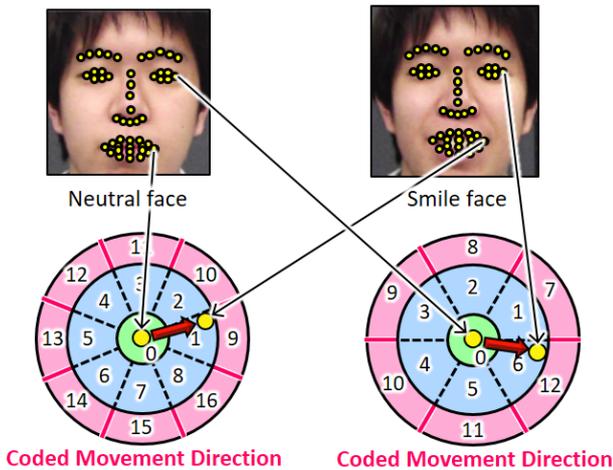
また、顔の輪郭のキーポイントは背景の影響を受けるため、安定して検出することができない。今回は、輪郭のキーポイントを除く 49 点を使用する。Fig. 3 に検出したキーポイントと使用するキーポイントの例を示す。

〈3・4〉 移動方向コード 無表情時と表情表出時のキーポイントからなる移動方向の角度成分と大きさ成分それぞれに注目する。各キーポイントにおいて、それぞれの成分をコード化することによって、個人差を抑えた特徴量で



個人情報保護の観点より加工してある。

Fig. 3. Example of detected keypoints.



個人情報保護の観点より加工してある。

Fig. 4. Example of Coded Movement Direction.

ある移動方向コードを算出する。Fig. 4 に各キーポイントにおける移動方向コードの例を示す。

抽出したキーポイントごとに移動ベクトルを求める。移動方向とは、無表情画像のキーポイントから表情表出画像のキーポイントまでのベクトルと、一意に決めた基準ベクトルのなす角度と大きさである。角度に関しては、算出した角度値そのものを用いると、個人間において差が生じる。また、大きさに関しても、個人ごとに表情の強弱に差が生じる。そのため、各キーポイントにおいて、角度値と大きさの値をそれぞれ学習データの分布に基づいてコード化する。このコードのことを移動方向コードと定義し、特徴量とする。表情ごとに、移動方向コードの発生確率値を算出し、キーポイントの選択および表情識別に用いる。

〈3・5〉 コード数の自動決定 顔の部位ごとに筋肉が異なり、キーポイントの移動する方向は異なる。そのため、キーポイントごとに移動方向コードの角度分解能と大きさのしきい値から、コード数を決める必要がある。〈3・5・1〉項で角度分解能の決定方法、〈3・5・2〉項で大きさのしきい値の決定方法について述べる。

〈3・5・1〉 角度分解能の自動決定 キーポイントごとに角度分解能を決めるため、角度分布を利用する。Fig. 5 に角度分布に基づく分解能の自動決定の流れを示す。

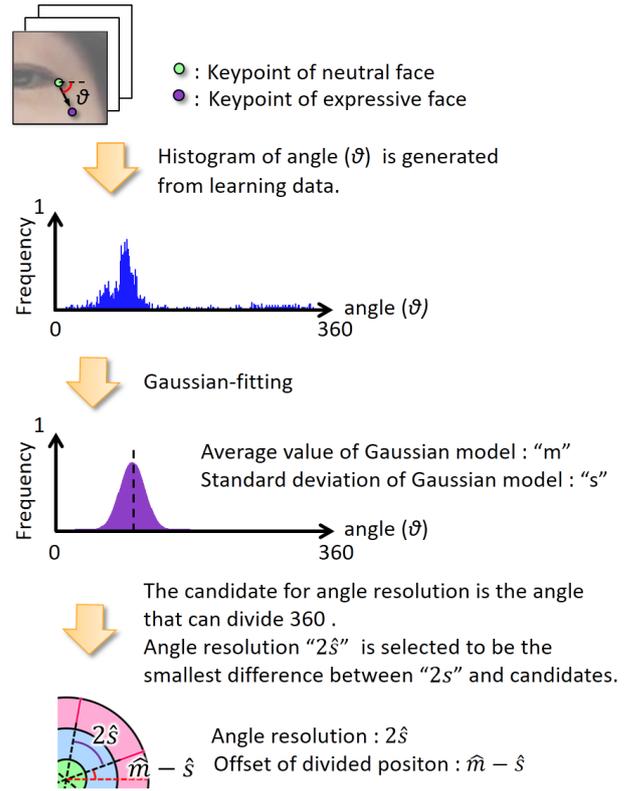


Fig. 5. Automatic determination of angle resolution.

各キーポイントにおいて、無表情画像と表情表出画像の学習画像群から角度を算出し、角度値の頻度を表すヒストグラムを生成する。生成したヒストグラムに平均値 m ($m = 0, 1, \dots, 359$), 標準偏差 s ($s = 0.00, 0.05, \dots, 25.00$) のガウス分布をガウシアンフィッティングする。選択されたガウス分布の信頼区間を 95% と設定し、標準偏差 $\pm s$ の範囲に基づき、角度分解能を決定する。今回は 360 度を割り切れる角度の幅の候補とし、2 分割から 18 分割の間で角度分解能を決定した。

また、角度の分布がコードの分割位置付近に集中することを避けるため、ガウス分布の平均値から標準偏差を引いた値 $m-s$ をオフセットとして付与する。

〈3・5・2〉 大きさのしきい値の自動決定 角度分解能と同様に、キーポイントごとに大きさのしきい値を決めるため、大きさのしきい値ごとに生成した角度分布を利用する。Fig. 6 に大きさのしきい値の自動決定の流れを示す。

各キーポイントにおいて、〈3・5・1〉項と同様に角度値の頻度を表すヒストグラムを生成する。このヒストグラムは、学習画像群の表情ごとに生成したため、他の表情の角度値を含まない理想的なヒストグラムである。また、大きさのしきい値 r (r_1, r_2, \dots, r_n) の値ごとに、角度値の頻度を表すヒストグラムを生成する。今回は、 r_1 を 0.01, r_{100} を 1.0 までの範囲で、大きさのしきい値を求めた。表情ごとに生成したヒストグラムと、大きさのしきい値ごとに生成したヒストグラムの類似度を算出する。類似度の算出には、ハミング距離を用いた。最も類似度の高いヒストグラムは、表

In the case of calculating the threshold \hat{r} between subtle smile and exaggerated smile.

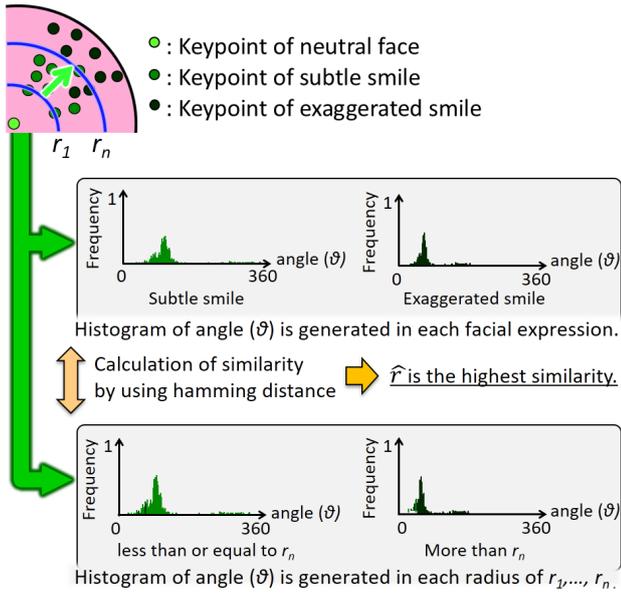


Fig. 6. Automatic determination of radius's threshold.

情ごとの分布を最も再現したヒストグラムである。そこで、最も類似度の高いヒストグラムの大きさのしきい値 r_n を移動方向コードの大きさとする。

〈3・6〉 キーポイントの選択 移動方向コードは、表情ごとに発生確率値を有する。そこで、この発生確率値の分布より、表情識別に使用するキーポイントを選択する。Fig. 7 に選択されるキーポイントが有する移動方向コードの例を示す。

キーポイントを選択する基準として2つが考えられる。

- (1) 移動方向コードが表情間で異なること。
- (2) 移動方向コードが人に対して共通であること。

上記の基準(1)を満たす条件は、各表情において最頻値となる移動方向コードのコード番号が異なることである。コード番号が異なっていれば、表情識別に有効であるといえる。

次に、基準(2)である。移動方向コードの各コードは学習者の移動ベクトルの頻度から成り立っている。最頻値のコードが全ての学習者に対して、一定の頻度を有していれば、人に対して共通であるといえる。そこで、基準(2)を満たす条件は、最頻値となったコードの人に対する分散値がしきい値以下であることとする。

〈3・7〉 表情の発生確率値を用いた識別 学習画像群から得られた移動方向コードの発生確率値と、表情識別に有効かつ人に対して共通性の高い特徴量を有するキーポイントを用いて、属する表情が未知の顔画像を識別する。Fig. 8 に識別の流れを示す。

まず、識別される顔画像に対して、キーポイントを検出し、学習モジュールにて選択されたキーポイントから移動方向コードを求める。移動方向コードの各コードは、各表情の発生確率値を有している。そこで、識別される顔画像

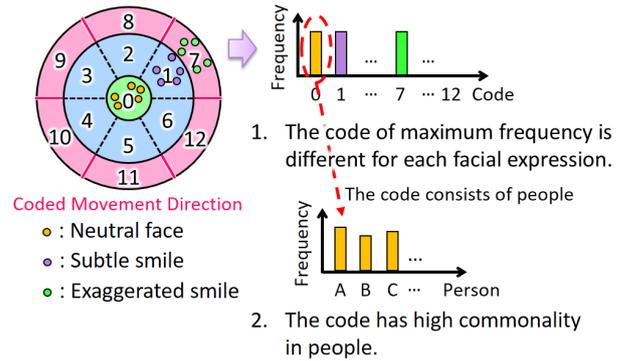


Fig. 7. Characteristic of Coded Movement Direction having selected keypoint.

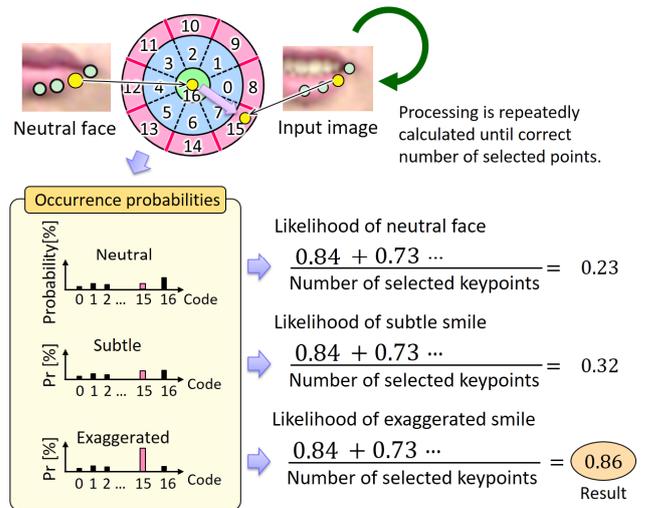


Fig. 8. Flow of facial expression classification.

の各キーポイントの移動方向コードの発生確率値から、表情ごとの尤度を求める。最も高い尤度となった表情を識別結果とする。

4. 実験と考察

〈4・1〉 実験に用いたデータセット 10~20代の男女計28名に対して、自然な笑顔を表出してもらうためにコメディ番組を視聴してもらい、その際の顔を撮影した。撮影した動画をフリーソフトにて30fpsの分解能で画像に分割する。分割した画像を30fpsで自動送りし、表示される画像に対して教師タグを付与してもらった。今回は、撮影の対象者ではない5名の学生が教師タグを付与した。教師タグの付与には、キーボード入力を用いた。表情が無表情に変化した場合は1、微妙な笑顔に変化した場合は2、大仰な笑顔に変化した場合は3をキー入力してもらった。教師タグが付与された画像のうち、5名中4名が同一の教師タグを付与し、顔のキーポイントが顔の各器官の位置で正しく検出された顔画像をデータセットとして使用した。

〈4・2〉 個人に対する識別性能 1名の顔画像を学習に用いて、その学習対象者の学習に使用していない顔画像を無表情、微妙な笑顔、大仰な笑顔に識別した。学習画像に



Fig. 9. F-measure of each facial expression for each person.

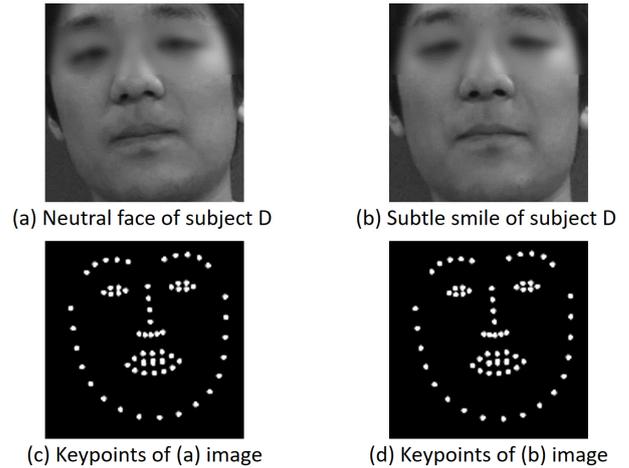
は、それぞれの表情画像 10 枚、識別画像にはそれぞれの表情画像 20 枚を用いた。この実験を 28 名それぞれに対しておこなった。Fig. 9 にそれぞれの人に対する各表情の F 値を示す。

実験結果より、無表情で平均 F 値 0.97、微妙な笑顔で平均 F 値 0.89、大仰な笑顔で平均 F 値 0.98 であった。対象者 D においては、微妙な笑顔を検出できなかった。これは、対象者 D の無表情と微妙な笑顔の顔に違いが少なかったため、キーポイントの移動方向コード特徴量に差が出なかったことが原因である。Fig. 10 に対象者 D の無表情と微妙な笑顔の顔画像とそれぞれの顔キーポイントを示す。

Fig. 10(a)における無表情の顔画像と、Fig. 10(b)における微妙な笑顔の顔画像を比較すると、無表情と微妙な笑顔に変化が少ない。Fig. 10(c),(d)における顔キーポイントの画像は、幾何学補正後の顔キーポイントである。この顔キーポイントを比較すると、顔キーポイントが変化していないとわかる。対象者 D においては、無表情と微妙な笑顔に差異がないため、移動方向コード特徴量が算出できなかったと考えられる。同様に、対象者 U においても表情間で差異がない画像が多くみられた。これは、表情の教師タグを付与するという作業を行った人が、時間的な変化に注目して表情を判断したと考えられる。しかし、提案アルゴリズムは単一フレームから特徴量を算出するため、連続的には確認できる微弱な動きを単一フレームの顔特徴量では表現できずに、微妙な笑顔が認識できなかったと考えられる。これは、単一フレームから算出した特徴量の表現力よりも、人が表情の時間的な変化を見分ける能力が優れていたことを示している。時間的な変化に対応するアルゴリズムの改良が必要であると考えられる。また、単一フレームにおいて表情が表出されている顔画像では、提案アルゴリズムは 5 名中 4 名が同一といった画像を識別できているため、静止画像に対しては人と互角の識別性能を有しているといえる。

〈4-3〉 個人差の影響に関する識別性能の実験 27 名の無表情、微妙な笑顔、大仰な笑顔の顔画像を学習し、学習に使用していない 1 名の顔画像を無表情、微妙な笑顔、大仰な笑顔に識別した。実験に使用したキーポイントの例を Fig. 11 に示す。

口、眉、目のキーポイントが選択されたことによって、提案した特徴量が、口角があがったり、目尻が垂れたりといった、笑顔特有の表情変化を捉えていることを示してい



個人情報保護の観点より一部ぼかしてある。

Fig. 10. Facial image and keypoints of neutral face and subtle smile.

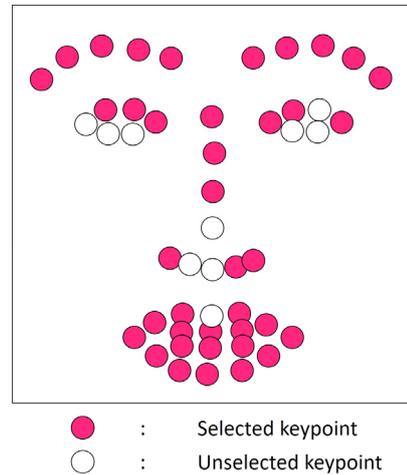


Fig. 11. Example of selected keypoints.

る。また、大きく笑うと鼻孔が広がるため、鼻孔も有効なキーポイントと選択された。これは、微妙な笑顔と大仰な笑顔を識別するために有効なキーポイントとして選択されていた。

次に、比較した手法について述べる。従来研究と性能比較するため、個人差の影響を軽減した表情識別手法である Faisal らの手法⁽¹¹⁾と、微妙な表情の認識を目的とした幾何学特徴量を用いた野宮らの手法⁽⁸⁾、Gabor 特徴量を用いた松久らの手法⁽⁷⁾、DeepLearning 手法の EmotionNet⁽⁹⁾と VGGNet⁽¹⁰⁾の 5 つの従来手法と提案手法を比較した。EmotionNet は Convolutional Neural Network によって識別器を生成する DeepLearning である。学習回数は 15000 回とした。VGGNet も Convolutional Neural Network によって識別器を生成する DeepLearning である。VGGNet については、既に生成された表情識別モデルに対して Fine-tuning という再学習アルゴリズムを適用し、学習回数を 15000 回した。再学習する表情識別モデルは一般に公開されている識別モデルを用いた。

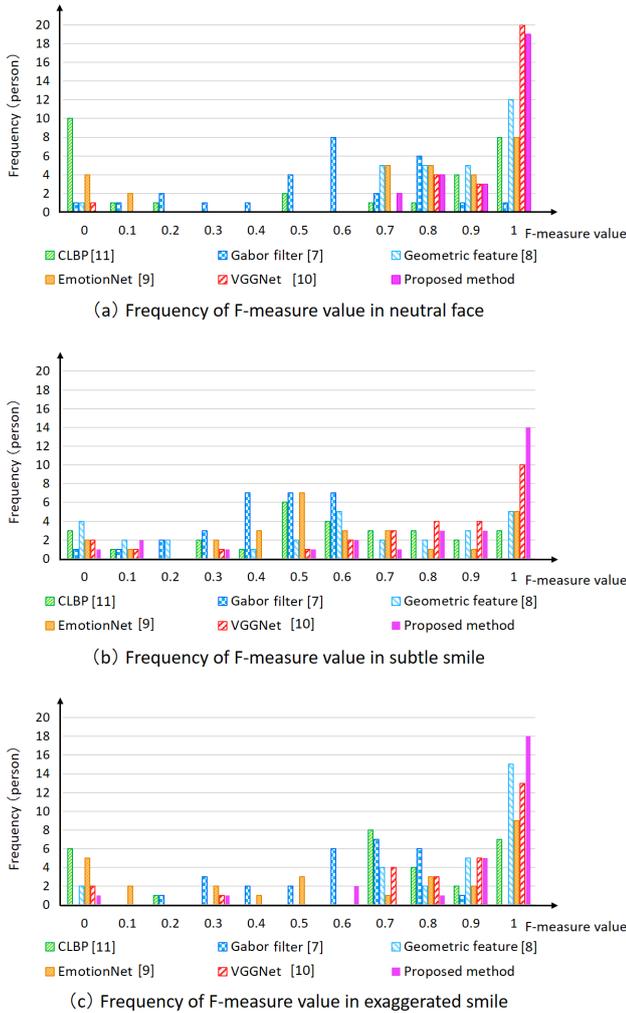


Fig. 12. Frequency of F-measure of unlearned people.

この実験を交差検証にて実施し、各表情の識別結果として人に対する F 値の頻度分布によって個人差の影響を評価した。学習画像および識別画像には、1 名当たりそれぞれの表情画像 30 枚を用いた。Fig. 12 に表情ごとの F 値の頻度を示す。

Fig. 12 のグラフは、分布が右側に集中しているほど、性能が学習データに依存していないことを示す。実験結果より、微妙な笑顔と大仰な笑顔において、提案手法の識別結果が従来手法の識別結果より右側に分布していることを確認した。この結果から、提案手法が従来手法より個人差の影響を受けずに、表情の機微を捉える手法であることを確認した。また、Table 1 に表情ごとの平均 F 値を示す。

実験結果より、提案手法における識別精度は、無表情で平均 F 値 0.93、微妙な笑顔で平均 F 値 0.78、大仰な笑顔で平均 F 値 0.93 であった。個人ごとに学習し、識別した結果と比べて、識別性能があまり低下していないことから、学習データに依存しにくい手法であることが確認された。また、提案手法は従来手法より、無表情において平均 0.22、微妙な笑顔において平均 0.21、大仰な笑顔において平均 0.24 高いことを確認し、高精度に識別できることを確認した。

Table 1. Experimental comparison of facial expression classification performance.

	Neutral	Subtle	Exaggerated
Proposed method	0.93	0.78	0.93
CLBP [11]	0.51	0.56	0.63
Gabor filter [7]	0.58	0.43	0.61
Geometric feature [8]	0.85	0.57	0.85
EmotionNet [9]	0.68	0.57	0.57
VGGNet [10]	0.91	0.74	0.81

5. まとめ

本研究では、個人差の影響を受けにくく、微妙な表情を認識するため、2つのことを提案した。1つめは、個人ごとの骨格、筋肉に依存しない局所的な部位における移動方向に注目した移動方向コード特徴量である。2つめは、表情識別および個人差の影響を考慮した特徴量の選択方法である。提案手法における表情識別の性能は F 値で評価した。個人ごとでは、無表情では平均 F 値 0.97、微妙な笑顔では平均 F 値 0.89、大仰な笑顔では平均 F 値 0.98 を確認した。学習に使用していない人では、無表情では平均 F 値 0.93、微妙な笑顔では平均 F 値 0.78、大仰な笑顔では平均 F 値 0.93 を確認した。また、個人差の影響については、F 値の人に対する頻度分布を用いて評価した。その結果、微妙な笑顔と大仰な笑顔において、提案手法は従来手法より F 値の分布が高い域に集中していた。これにより、提案手法は学習していない人に対しても有効な表情識別手法であることが確認された。

また、今回使用したデータは、5 名中 4 名が同一の表情であると判断した表情のみを用いているため、人の判断能力と互角の識別性能を有しているといえる。

提案手法は、コメディ番組を視聴している人の笑顔度合いを 3 段階で推定できることを確認した。さらに、提案手法は顔画像 1 枚ごとに表情を識別できるため、表情強度に基づく写真データの検索や、顔が連続して写っていない顔動画に対しても利用できると考える。

謝辞

本研究の一部は独立行政法人科学技術振興機構 (JST) の研究成果展開事業「センター・オブイノベーション (COI) プログラム」の支援によっておこなわれた。

文献

- (1) T. Maruyama, H. Sasamoto, N. Arakawa, T. Kawasoe, and M. Nakagawa: "A Study on Improvement of Emotion-Detection Accuracy Using Fractal Characteristics of EEG", *IEICE Trans. Fundamentals (Japanese Edition)*, Vol.J95-A, No.4, pp.343-356 (2012) (in Japanese)
丸山貴司・笹本裕美・荒川尚美・川副智行・中川匡弘:「脳波のフラクタル性を用いた感性推定精度の向上に関する研究」, 信学論 A, Vol.J95-A, No.4, pp.343-356 (2012)

- (2) J. Chen, Z. Chen, Z. Chi, and H. Fu : “Facial Expression Recognition Based on Facial Components Detection and HOG Features”, Scientific Cooperations International Workshops on Electrical and Computer Engineering Subfields, Istanbul, turkey, pp.64-69 (2014)
- (3) E. Owusu, Y. Zhan, and Q. R. Mao : “A neural-AdaBoost based facial expression recognition system”, Expert Systems with Applications, Vol.41, No.7, pp.3383-3390 (2014)
- (4) C. Shan, S. Gong, and P. W. McOwan : “Robust facial expression recognition using local binary patterns”, IEEE International Conference on Image Processing (ICIP), Genoa, Vol.2, pp.370-373 (2005)
- (5) Y. Ogoshi, Y. Mitsuhashi, S. Ogoshi, A. Nakai, S. Matsuura, and C. Araki : “Recognition of Facial Expression Based on Analysis of Resultant Sequential Retinal Outlines over The Course of A Blink”, *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, Vol.23, No.2, pp.218-227 (2011) (in Japanese)
小越康宏・三橋美典・小越咲子・中井昭夫・松浦慎也・荒木睦大 : 「黒眼の形状と瞬きの変化からの表情認識」, 日本知能情報ファジィ学会誌, Vol.23, No.2, pp.218-227 (2011)
- (6) T. Otsuka, J. Ohya, and R. Nakatsu : “Recognizing Facial Expressions Generated by Multiple Persons Based on Hidden Markov Models with Continuous Output Probability Densities”, *The transactions of the Institute of Electronics, Information and Communication Engineers*, Vol.J80-D-II, No.8, pp.2129-2137 (1997) (in Japanese)
大塚尚宏・大谷 淳・中津良平 : 「連続出力確率密度分布を用いたHMMによる動画からの複数人物の表情認識」, 信学論 D, Vol.J80-D-2, No.8, pp.2129-2137 (1997)
- (7) H. Matsuhisa and M. Hashimoto : “Identifying Subtle Facial Expression Changes using Optimized Gabor Features”, *The Journal of The Institute of Image Information and Television Engineers*, Vol.68, No.6, pp.J252-J255 (2014) (in Japanese)
松久ひとみ・橋本 学 : 「Gabor 特徴を用いた顔画像からの微妙な表情変化の推定」, 映像情報メディア学会誌, Vol.68, No.6, pp.J252-J255 (2014)
- (8) H. Nomiya and T. Hochin : “Efficient Emotional Video Scene Detection Based on Ensemble Learning”, *The IEICE Transactions on Information and Systems (Japanese Edition)*, Vol.J95-D, No.2, pp.193-205 (2012) (in Japanese)
野宮浩揮・宝珍輝尚 : 「アンサンブル学習を用いた効率的な映像からの表情表出シーン検出」, 信学論 D, Vol.J95-D, No.2, pp.193-205 (2012)
- (9) B. Kennedy and A. Balint : “EmotionNet”, github: <https://github.com/co60ca/EmotionNet>
- (10) G. Levi and T. Hassner : “Emotion Recognition in the Wild via Convolutional Neural Networks and Mapped Binary Patterns”, Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, Seattle, Washington, USA, pp.503-410 (2015)
- (11) F. Ahmed, H. Bari, and E. Hossain : “Person-Independent Facial Expression Recognition Based on Compound Local Binary Pattern (CLBP)”, IAJIT, Vol.11, No.2, pp.195-203 (2014)
- (12) P. Ekman and W. V. Friesen : “Unmasking the Face”, Malor Books (2003)
- (13) J. M. Saragih, S. Lucey, and J. F. Cohn : “Deformable Model Fitting by Regularized Landmark Mean-Shifts”, *International Journal of Computer Vision*, Vol.91, No.1, pp.200-215 (2011)

佐々木 康 輔



(非会員) 2015 年度, 2016 年度中京大学人工知能高等研究所の準所員を務める。2017 年中京大学大学院情報科学研究科前期課程修了。同年 NEC ソリューションイノベータ (株) 入社。中京大学大学院情報科学研究科前期課程在籍中はヒューマン認識技術の研究に従事。日本顔学会会員。日本認知科学学会会員。電子情報通信学会会員。

渡 邊 健 斗



(非会員) 2014 年 4 月中京大学工学部機械システム工学科に入学し現在に至る。表情認識に興味をもつ。

橋 本 学



(正員) 1987 年大阪大学大学院工学研究科前期課程修了。同年三菱電機 (株) 入社。生産技術研究所, 先端技術総合研究所にて, ヒューマン認識, ロボットビジョン, パターン認識などの研究開発に従事。2008 年中京大学情報理工学部機械情報工学科教授。2013 年より同大学工学部機械システム工学科教授。博士 (工学)。1998 年度日本ロボット学会実用化技術賞, 2012 年度画像センシングシンポジウム優秀学術賞, 2015 年度精密工学会小田原賞, 2017 年度 IWAIT Best Paper Award 等受賞。電子情報通信学会, 日本ロボット学会, 情報処理学会, 精密工学会, IEEE 等各会員。

長 田 典 子



(正員) 1983 年京都大学理学部数学系卒業。同年三菱電機 (株) 入社。産業システム研究所において, 色彩情報処理, 感性情報処理の計測システムへの応用に関する研究に従事。1996 年大阪大学大学院基礎工学研究科博士後期課程修了。2003 年より関西学院大学理工学部情報科学科助教授, 2007 年教授。2009 年米国パデュー大学客員研究員。2013 年感性価値創造研究センター長。2015 年革新的イノベーション創出プログラム (COI) 「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」サテライトリーダー。2013 年文部科学大臣表彰科学技術賞, 2014 年グッドデザイン賞受賞。専門は感性工学, メディア工学等。情報処理学会, 電子情報通信学会, IEEE, ACM 各会員。博士 (工学)。