

PC操作性向上のための Webカメラを用いた顔の動き識別手法の提案

渡辺 泰伎^{1,a)} 白石 陽^{2,b)}

概要：

PCの多機能化により、イラストや動画などのUGC(User Generated Contents)の創作や共有が容易になった。それにともない、娯楽やビジネスなどの一環で、一般ユーザがPCを用いてさまざまなUGCを創作している。しかし、一般ユーザが多機能化するPCの機能を十分に活用し、創作作業におけるPCの操作の負担を軽減するためには、PCの操作性を向上させることが重要である。既存のPC操作性向上の方法として、ショートカットキーや音声情報を用いた方法などがある。しかし既存のPCの操作性向上の方法では、「PC操作に対応する入力動作」や「PCの操作内容に対応する入力動作の記憶」に対する負担が十分に考慮されていない。そこで本研究では、これらの負担を軽減するために、顔の動きに着目した。人は普段の何気ないコミュニケーションにおいて、顔の動きをノン・バーバル・コミュニケーションとして用いるため、顔の動きは自然な動作である。自然な動作をPCの入力に用いることで、「PC操作に対応する入力動作」や「PCの操作内容に対応する入力動作の記憶」に対する負担が軽減できると考える。本研究では、PC操作性向上のための、Webカメラを用いた顔の動き識別手法を提案する。本手法では、Webカメラから収集する顔の動き画像から特徴を抽出し、抽出した特徴を用いて識別器を作成する。その後、作成した識別器を利用し、PC操作と関連付けられた顔の動きを識別することでPCを操作する。本稿では、顔の動き画像から顔の動きを識別するために有効な特徴量やアルゴリズムを検討した。実験の結果、顔の特徴点の中心から顔の各特徴点までの距離が、顔の動きを識別する際に有効な特徴量であることが示された。しかし、実環境を想定した実験として、一定時間固定した顔の動きを学習データ、瞬間的な顔の動きを評価データとして用いた場合、F値は0.52であった。今後の課題として、瞬間的な顔の動きと、一定時間固定した顔の動きどちらにも利用可能な特徴の検討を行う必要がある。

1. はじめに

PCの普及や専用のソフトウェアの開発により、UGC(User Generated Contents)が積極的に活用されている。UGCとは、ユーザの手によって創作されたコンテンツの総称である。UGCの例として、SNSに投稿されたデジタルイラストや動画投稿者の動画がある。このように、専門家だけではなく、一般ユーザが娯楽やビジネスなどの一環でさまざまなUGCを創作している。UGCの創作にあたり、専用のソフトウェアをインストールし、そのソフトウェアが提供する多様な機能を利用する必要がある場合が多い。しかし、一般ユーザが多機能化されたPCの機能を

を十分に活用し、創作作業におけるPCの操作の負担を軽減するためには、PCの操作性を向上させることが重要である。

既存のPCの操作性向上の方法として、ショートカットキーを用いた方法や音声情報を用いた方法、専用のデバイスを用いた方法がある。しかし、これらの方法では「PC操作に対応する入力動作」や「PCの操作内容に対応する入力動作の記憶」に対する負担が十分に考慮されていない。ショートカットキーを用いた方法では、ショートカットキーの入力に意識を向ける必要があり、専用のショートカットキーを記憶する必要がある。音声情報を用いた方法では、ユーザは操作する際に、操作内容を具体的に説明する必要がある。専用のデバイスを用いた方法では、操作に対応するボタンやペダルなどの入力を記憶する必要がある。

そこで本研究では、既存のPCの操作性向上手法の負担を考慮するために、NUI(Natural User Interface)に着目した。NUIとは、人間の自然な動作でコンピュータを操作

¹ 公立はこだて未来大学大学院システム情報科学研究科
Graduate School of Systems Information Science, Future University Hakodate

² 公立はこだて未来大学システム情報科学部
School of Systems Information Science, Future University Hakodate

a) g2118047@fun.ac.jp

b) siraisi@fun.ac.jp

する方法の総称である。東京工芸大学の調査 [1] によると、NUI によって IT 機器の操作が誰にとっても簡単になるかどうかの調査項目では、約 7 割の人が肯定的に捉えているという回答結果が得られた。そのため、自然な動作で PC を操作することで、「PC 操作に対応する入力動作」に対する負担が考慮できると考える。また、PC 操作時には入力動作に対応する操作内容を記憶する必要がある、自然な動作の方が不自然な動作よりも、その対応関係の記憶が容易である。そのため、自然な動作を PC の操作に用いることで、「PC の操作内容に対応する入力動作の記憶」に対する負担を考慮できると考える。

本研究では NUI として、顔の動きに着目した。本研究では顔の動きを、表情のように人間の感情が複雑に絡み合った際に表出する顔の形状ではなく、口角を上げることや、顔を右に傾けるなど、面道具や頭部の動きの単純な変化と定義する。人は普段の何気ないコミュニケーションにおいて、顔の動きをノン・バーバル・コミュニケーションとして用いる。そのため、顔の動きは自然な動作であると考えられる。また、PC の操作性向上に利用可能な NUI として、手の動きや体の動きを用いたものがある。しかし、従来の操作方法と併用した利用を想定すると、顔の動きの利用が適していると考えられる。

ここで、顔の動きを用いて PC を操作する想定シーンの例を 2 つ述べる。まず、利用頻度の高い操作を行う場合である。例えば、Web ブラウジングにおいて、一つ前に戻る機能は多くの人が利用する機能である。しかし、多くのブラウザでは、一つ前に戻る機能がブラウザの上部に存在するボタンを入力したり、バックスペースキーを入力する必要がある。そこで、顔の動きを用いることにより、マウスの操作やバックスペースキー入力のための手の移動がなくなるため、PC の操作性を向上できる。次に、複数のディスプレイを利用する場合である。この場合、もう一方のディスプレイに表示されているアプリケーションを操作する際に、マウスカーソルを移動させる必要がある。そこで、メインディスプレイ上のアプリケーションをマウスなどの従来の方法で操作し、サブディスプレイ上のアプリケーションを顔の動きで操作する。その結果、ディスプレイ間でマウスカーソルを移動させる機会の減少が期待できるため、操作性が向上できる。

顔の動きを識別する研究として、眼鏡型ウェアラブルデバイスや耳掛け型ウェアラブルデバイスなどの専用のウェアラブルデバイスから取得するセンサ値を用いる研究 [5][6][7] と、カメラから取得する画像を用いる研究 [8][9] がある。顔の動きを識別する上で、計算コストの観点では、センサ値を用いる方法が優れる。しかし、対応可能な顔の動きの多様性の観点では、画像を用いる方法が優れる。将来、専用のウェアラブルデバイスが普及する可能性より、PC の処理能力が向上する可能性の方が高いと考える。そ

のため、PC の処理能力が向上した結果、本研究における提案手法をバックグラウンドで動作させた場合でも、PC に対する処理の負荷が少なくなり、従来の PC の操作性を損なわないと考える。よって、対応可能な顔の動きが多く、実現性が高い画像を用いた方が、顔の動きの識別に適していると考えられる。

よって本研究では、「PC 操作に対応する入力動作」や「PC の操作内容に対応する入力動作の記憶」に対する負担を考慮した、PC の操作性向上のために顔の動き識別手法の提案を行う。提案手法を実現するために、本研究では PC 内蔵の Web カメラから取得する画像を用いて顔の動きを識別する。その後、操作に識別した顔の動きを割り当てることで PC の操作性を向上させる。本稿では、顔の動きを識別するために有効な特徴量やアルゴリズムについて検討を行う。

2. 関連研究

本研究では、PC 操作性向上のための顔の動き識別手法を提案する。そこで、IT 機器の操作性向上のためのユーザインタフェースに関する研究 [2][3][4] や、顔の動きを識別するための要素技術として、表情推定に関する研究 [5][6][7][8][9] について述べる。

2.1 IT 機器の操作性向上のためのユーザインタフェースに関する研究

IT 機器の操作性向上のためのユーザインタフェースに関する研究として、腕の動きを用いた研究 [2][3] や、顔の動きを用いた研究 [4] がある。

腕の動きを用いた研究 [2][3] では、IT 機器の操作性向上のために手の形や動きの推定を行っている。池ら [2] は、PC を離れた位置から操作するために、PC に搭載されている Web カメラから取得した画像を用いて手を検出し、手の振りかぶりや手を振るなどの手の軌跡を推定している。久保ら [3] は、手に一組のスピーカーとマイクを装着し、スピーカーから発した音波を、手を通じてマイクで収集する。収集した音波の周波数成分の強度を特徴量とする、SVM により手の形を推定している。しかし、PC の操作時には、両手がふさがっている場合が多く、手をカメラの前に構えていることや、膝の上に置いていることは少ない。そのため、PC の操作への適用を考えると、既存のキーボードやマウスの利用したほうが効率的である。

顔の動きを用いた研究として、顧ら [4] は、タブレットに内蔵されている Web カメラから取得した画像から、顔を認識し、目の両端や口の両端といった特徴点を抽出する。その後、各特徴点のオイラー角を算出し、設定した閾値を超えた際に、自作した電子書籍アプリケーションを操作するシステムの提案を行っている。この研究では、顔の動きを利用しているため、両手がふさがっている状況でも操作性

は向上できる。しかし、あらかじめ設定された顔の動きしか識別できないため、操作に対応する顔の動きが、全てのユーザにとって自然な動作であるとは限らない。また、推定手法として、オイラー角の閾値を用いているが、ユーザごとに顔の形や、動かしやすい表情筋は異なると考える。そのため、ユーザによっては適切な閾値にならず、操作性が損なわれる可能性がある。

2.2 表情推定に関する研究

表情推定に関する研究として、専用のウェアラブルデバイスを用いた研究 [5][6][7] と、カメラを用いた研究 [8][9] がある。

専用のウェアラブルデバイスを用いた研究として、反射光センサを用いた研究 [5][6] と、筋電位センサを用いた研究 [7] がある。

反射光センサを用いた研究では、眼鏡型ウェアラブルデバイスの内側に反射光センサを取り付け、眼鏡と顔の表面との距離を計測する。表情の変化時に眼鏡と顔の皮膚との距離が変化するため、表情変化時のセンサ値を用いた表情の推定を行っている。正井ら [5] は専用の眼鏡型ウェアラブルデバイスから取得したセンサ値を用いて、SVM による表情の識別を行っている。糸数ら [6] は、専用の眼鏡型ウェアラブルデバイスから取得したセンサ値を用いて、事前に設定した閾値による表情の識別を行っている。

筋電位センサを用いた研究として、Gruebler ら [7] は、専用耳掛け型ウェアラブルデバイスに搭載された筋電位センサから、表情の動きによる筋電位の変化をセンサ値として取得する。その後、取得したセンサ値を入力とした2層のニューラルネットワークにより、表情推定を行っている。これらの専用のウェアラブルデバイスを用いた研究では、識別に利用する特徴のデータ量が少ないため、計算コストの低い表情推定が可能である。しかし、利用するウェアラブルデバイスによって、検出できない顔の動きが存在する。例えば、反射光センサを用いた研究 [5][6] の場合、口を開く際に頬が伸び、センサ値に変化が生じる。しかし、取得したセンサ値の変化が、登録した顔の動きなのか、会話中の顔の動きかを識別することは困難である。

カメラを用いた研究では、カメラから取得した画像内から顔領域画像を切り出し、顔領域画像から抽出した特徴量から表情推定を行っている。Gwen ら [8] は切り出した顔領域画像に、ガウシアンフィルタバンクを適用させた画像情報を特徴量として、SVM による表情推定を行っている。野宮ら [9] は、顔の特徴点を抽出し、3つの顔の特徴点を2つの線分で構成した時のなす角を特徴量として、アンサンブル学習を利用して表情と、その強度を推定している。これらのカメラを用いた研究では、専用のウェアラブルデバイスを利用した研究と比べ、特徴抽出のための画像処理に対する計算コストが高く、顔の表情を推定を行う際の計算コ

ストが高いが、検出できる顔の表情は多い。しかし、これらの研究で用いている特徴量は、顔の傾きの情報を考慮していない。そのため、顔の傾きの識別を考えた際に、対応できない可能性が高い。

3. 提案手法

3.1 研究目的

本研究では、PC の操作性向上のための顔の動き識別手法の提案を目的とする。本手法では、PC 操作の入力に顔の動きを割り当てることで PC の操作を行う。顔の動きを識別するために、顔の動き画像を収集し、画像から抽出した特徴量を用いて識別器を作成する。収集と識別に用いる顔の動き画像は、PC に内蔵されている Web カメラを利用する。また、ユーザによって PC の操作のための表現しやすい顔の動きは異なり、設定された顔の動きではユーザによって、操作性が損なわれる可能性がある。また、ユーザによって顔の形状や表情筋は異なる。そのため、ユーザごとに、記憶や表現が容易な顔の動きを登録することで、入力負担が少ない自然な動作での PC 操作の実現を目指す。

3.2 システム概要

本研究の提案するシステムの概要を図 1 に示す。

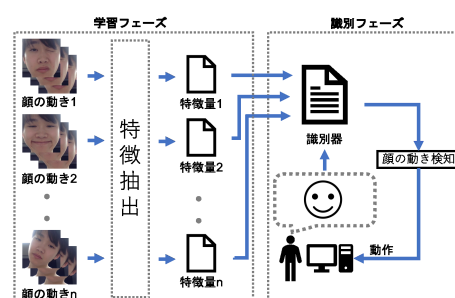


図 1 システム概要

本研究が提案するシステムでは、学習フェーズと識別フェーズに大別される。学習フェーズでは、ユーザが登録したい顔の動きのデータを収集し、識別器を作成する。識別フェーズでは、学習フェーズで作成した識別器を用いて、動きを識別し、実際に操作を行う。本稿では、提案システムの中で、学習フェーズ内の顔の動きを識別するための、顔領域画像中の特徴量やアルゴリズムについて検討する。

3.3 研究課題とアプローチ

顔の動きを識別するための研究課題として以下の3点が考えられる。

- (1) 識別に利用するデータの検討
- (2) 顔の動き画像における特徴抽出のアルゴリズム検討
- (3) 顔の特徴点における特徴量の検討

1 つ目の課題は、識別に利用するデータの検討である。アプローチとして、PC 内蔵の Web カメラから取得可能な画像を用いる。顔の動きを識別可能なデータの種類として、2.2 節の研究で用いられた、専用のウェアラブルデバイスに搭載されたセンサから取得したセンサ値や、カメラから取得した画像が考えられる。センサ値を用いる場合、顔の動きを識別する際の計算コストを軽減できるが、利用するセンサ値の種類によって対応できる顔の動きが限定されるという問題がある。画像を用いる場合、センサ値と比べて情報量が多く、計算コストが高いが、対応可能な顔の動きが多い。将来、専用のウェアラブルデバイスが普及する可能性より、PC の処理能力が向上する可能性の方が高いと考える。そのため、PC の処理能力が向上した結果、PC のバックグラウンドで画像を用いた顔の動き識別を行っても、PC に対する処理の負荷が少なくなり、従来の PC の操作性を損なわないと考える。よって、対応可能な顔の動きが多く、実現性が高い画像を用いる。

2 つ目の課題は、顔の動き画像における特徴抽出のアルゴリズム検討である。アプローチとして、顔の特徴点を用いる。顔の特徴点とは、面道具や顔の輪郭を検出する特徴点のことである。本研究では顔の特徴抽出として、dlib[10][11]という画像処理ライブラリを用いる。dlibによって、図2で示すような、68点の特徴点位置(画像座標系における x, y 座標)が取得できるため、それらの座標を顔の特徴点として利用する。比較対象として、顔領域を切り出した画像データそのものを入力とした、深層学習による識別が考えられる。しかし、識別器を作成する計算コストが高い。また、顔の動き画像は頭部の傾きの違いにより、取得できる画像も異なる。そのため、頭部の傾きに対応するために、さまざまな頭部の傾きにおける顔の動き画像を学習データとして用意しなければならない。顔の特徴点を利用した場合、頭部姿勢推定が可能であるため、推定した頭部姿勢の結果を用いてデータを補正すれば、学習データとしてさまざまな頭部の傾きにおける顔の動き画像を収集する必要がないと考える。そのため本研究では、顔の動きを識別するために顔の特徴点を利用する。

3 つ目の課題は、顔の特徴点における特徴量の検討である。本研究では、顔の特徴点における2種類の特徴量を検討している。1 つ目の特徴量は、顔の各特徴点の x, y 座標の合計 (X, Y) を算出し、各特徴点 (x_i, y_i) を x, y 座標の合計 (X, Y) で除算して正規化した座標の特徴量(136次元)である。2 つ目の特徴量は、顔の各特徴点の x, y 軸の中心 $(\bar{x}, \bar{y})$ を算出し、中心から各特徴点 (x_i, y_i) の距離 d_i を標準化した特徴量(68次元)である。本稿では簡略化のために、1 つ目の特徴量を特徴量1、2 つ目を特徴量2と定義する。

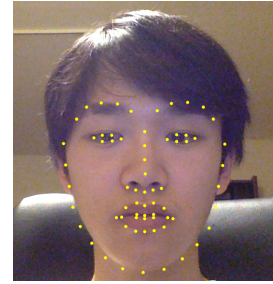


図2 dlib[10]を用いて検出した、68個の顔の特徴点

3.4 特徴抽出の流れ

特徴抽出までの流れについて説明する。まず、dlib[10][11]を用いて、Web カメラから取得した画像内の顔領域画像を切り出し、顔領域画像内の顔の特徴点を抽出する。その後、抽出した顔の特徴点を用いて、3.3 節で述べた2種類の特徴量を算出する。特徴量1, 2の具体的な算出方法は以下の通りである。

特徴量1の算出方法

- (1) 顔の各特徴点座標 (x_i, y_i) の x, y 座標の合計 (X, Y) を算出する
- (2) 顔の各特徴点座標 (x_i, y_i) を (X, Y) で除算した結果の座標 $(\frac{x_i}{X}, \frac{y_i}{Y})$ を特徴量として利用する

特徴量2の算出方法

- (1) 顔の各特徴点 (x_i, y_i) の x, y 座標の合計 (X, Y) を算出する
- (2) (X, Y) を顔の特徴点の数(68)で除算することで、顔の特徴点の中心 $(\bar{x}, \bar{y})$ を算出する
- (3) 顔の各特徴点 (x_i, y_i) と顔の特徴点の中心 $(\bar{x}, \bar{y})$ のユークリッド距離 d_i を算出する
- (4) それぞれのユークリッド距離 d_i を標準化して、特徴量として利用する

4. 実験および考察

4.1 節では、実験環境について述べる。また4.2 節や4.3 節では、顔の動きを識別するために有効な特徴量を検討するための予備実験や、予備実験における考察内容の検討を行うための追加実験についての実験結果や考察を述べる。加えて4.4 節では、予備実験や追加実験を踏まえた、顔の動き識別における精度向上のための今後の課題について述べる。

4.1 実験環境

実験環境の設定を表1に示す。また、実験のパラメータを表2に示す。

本実験は、本研究における初期実験かつ、顔の動きの識別に有効な特徴量検討のための実験である。そのため、顔の動きの識別に有効な特徴量の考察を容易にするために、

表 1 実験の設定

識別器作成システム	
プログラミング言語	Python3.6.2
ライブラリ	dlib19.10.0, OpenCV3.4.0.12
ハードウェア (PC)	
OS	macOS High Sierra
CPU	Intel Core i5 1.6GHz
RAM	8 GB

表 2 実験のパラメータ

項目	項目値
被験者	1 名
識別する顔の動き	平常時の顔 (1 種類) ウィンク (右左 2 種類) 口角をあげる (左右, 両方 3 種類) 顔を前後左右に傾ける (4 種類)
データの収集方法	顔の動きを約 6 秒間固定した動画 (9 種類) 顔の動きを 10 回繰り返した動画 (9 種類)

被験者を 20 代の男子大学生 1 人とした。次に、識別する顔の動きを検討した。平常時の顔や、顔のパーツを用いた単純な顔の動き、顔の傾きを用いた顔の動きによる操作を想定し、実験的に図 3 に示した 10 種類の顔の動きを識別する。また、データの収集として、Quick Time Player を使い、動画を 1280 × 720 のサイズで約 30fps で収集した。



図 3 識別する顔の動き

本手法では、ユーザが利用する顔の動きを登録することを想定する。顔の動きの登録時はユーザの入力負担を考慮し、顔の動きを一定時間固定した動画より、顔の動きの登録を行う。そのため、顔の動き登録時を想定した、顔の動きを約 6 秒間固定した動画を収集した。一方で、ユーザが実際に顔の動きを用いて PC を操作する際は、瞬間的な顔の動きを用いる。そのため、ユーザが実際に顔の動きを利用することを想定した、顔の動きを 10 回繰り返した動画を収集した。また、動画の種類について、簡略化のために、顔の動きを約 6 秒間固定した動画を動画 1、顔の動きを 10 回繰り返した動画を動画 2 と定義する。

動画のデータ量について述べる。動画 1 では、それぞれの顔の動きの動画のフレーム数は約 250 前後である。動画 2 では、平常時の顔のフレーム数が約 1300 フレーム、それ

ぞれの顔の動きのフレーム数が 100 から 200 フレームである。平常時の顔のフレーム数が多い理由としては、収集方法の性質上、顔の動きと顔の動きの間に平常時の顔が動画として入力されるためである。

精度評価の方法として、動画の各フレームから算出した各特徴量を説明変数、識別する顔の動き 10 種類を目的変数とした、SVM を構築する。その後、10 分割交差検定を行い、混同行列を利用することで精度評価を行う。また、識別精度の指標として、再現率、適合率、F 値を用いる。加えて、SVM を構築する段階で、その都度グリッドサーチによるパラメータチューニングを行った。調整したハイパラメータは、SVM のコストパラメータである c 、RBF カーネルのパラメータである γ である。調整したハイパラメータの取りうる値は表 3 に示す。また、SVM のカーネルは RBF カーネルを利用した。今後の精度評価の説明では、利用した特徴量、学習データと評価データの関係のみを述べる。

表 3 SVM のハイパラメータ

ハイパラメータ名	値
c	0.01, 0.1, 1.0
γ	1.0, 5.0, 10.0, 50.0, 100.0

4.2 予備実験

4.2.1 特徴量検討のための予備実験

顔の動きの識別に有効な特徴量の検討のための予備実験を行った。動画 1 から特徴量 1, 2 を算出し、各特徴量をそれぞれ学習データ、評価データとした際の精度評価の結果を表 4, 5 に示す。それぞれの F 値を比較した結果、特徴量 1 より特徴量 2 を用いた際の識別精度が高かった。

表 4 特徴量 1 を用いた 10 分割交差検定の混同行列の結果 ($\gamma = 1, c = 100$)

		計測値										再現率
		0	1	2	3	4	5	6	7	8	9	
実測値	0	180	0	0	0	0	0	0	0	0	0	1.00
	1	0	267	0	0	0	0	0	0	0	0	1.00
	2	0	0	0	0	271	0	0	0	0	0	0.00
	3	0	0	0	0	0	240	0	0	0	0	0.00
	4	0	0	0	0	281	0	0	0	0	0	1.00
	5	0	0	0	0	275	0	0	0	0	0	0.00
	6	0	0	0	0	0	0	307	0	0	0	1.00
	7	0	0	0	0	0	0	0	249	0	0	1.00
	8	0	0	0	0	0	0	0	0	279	0	1.00
	9	0	0	0	0	0	0	0	0	0	292	1.00
適合率		1.00	1.00	0.00	0.00	0.34	0.00	1.00	1.00	1.00	1.00	F 値 =0.66

4.2.2 実環境を想定した予備実験

実環境を想定した予備実験を行った。4.2.1 項の特徴量 2 により作成した識別器を用い、動画 2 について評価した結果を表 6 に示す。表 6 に示すとおり、表 5 の F 値と比べ、識別精度が低くなった。また、右のウィンク (顔の動き 2) の適合率や再現率が低く、右目ウィンクが左目ウィンクに誤識別されたが、左目ウィンクが右目ウィンクの誤識別さ

表 5 特徴量 2 を用いた 10 分割交差検定の混同行列の結果
($\gamma = 1, c = 1$)

		計測値										再現率
		0	1	2	3	4	5	6	7	8	9	
実測値	0	180	0	0	0	0	0	0	0	0	0	1.00
	1	0	267	0	0	0	0	0	0	0	0	1.00
	2	0	8	263	0	0	0	0	0	0	0	0.97
	3	0	0	0	240	0	0	0	0	0	0	1.00
	4	0	0	0	0	281	0	0	0	0	0	1.00
	5	0	0	0	0	0	275	0	0	0	0	1.00
	6	0	0	0	0	0	0	307	0	0	0	1.00
	7	0	0	0	0	0	0	0	249	0	0	1.00
	8	0	0	0	0	0	0	0	0	279	0	1.00
	9	0	0	0	0	0	0	0	0	0	292	1.00
適合率		1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	F 値 =0.99

れなかった。加えて、顔の傾きを利用した顔の動き (顔の動き 6,7) の再現率、適合率がともに低かった。

表 6 動画 1 を学習データ、動画 2 を評価データとした際の 10 分割交差検定の混同行列
(特徴量 2 を利用, $\gamma = 1, c = 1$)

		計測値										再現率
		0	1	2	3	4	5	6	7	8	9	
実測値	0	115	63	128	28	12	2	760	101	95	45	0.08
	1	2	102	0	1	0	0	14	0	0	0	0.86
	2	0	37	56	0	0	0	13	3	0	0	0.51
	3	0	0	1	119	0	0	4	0	0	0	0.96
	4	2	0	0	0	108	0	4	21	0	0	0.80
	5	0	1	29	2	0	81	18	13	0	0	0.56
	6	0	0	8	0	0	0	6	217	0	0	0.02
	7	1	1	200	16	0	0	9	4	0	0	0.02
	8	9	5	2	0	0	0	41	18	88	0	0.51
	9	6	9	12	0	0	0	11	0	1	112	0.74
適合率		0.85	0.47	0.13	0.76	0.90	0.98	0.01	0.01	0.48	0.71	F 値 =0.52

4.2.3 予備実験の考察

4.2.1 項の予備実験の結果、特徴量 1 と比べ、特徴量 2 を用いた際の F 値が高く、顔の動きを識別する際に適した特徴量であると示唆された。しかし、4.2.2 項の予備実験の結果、実環境を想定した際の F 値が 0.52 と低かった。また、右のウィンク (顔の動き 2) の適合率や再現率が低く、右目ウィンクが左目ウィンクに誤識別されたが、左目ウィンクが右目ウィンクの誤識別されなかった。加えて、顔の傾きを利用した顔の動き (顔の動き 6,7) の再現率、適合率がともに低かった。

実環境を想定した際の F 値が低かったことに関して、学習データの収集方法によって、抽出する特徴量に違いが生じることが原因として考えられる。4.2.2 項の予備実験では、学習データとして動画 1、評価データとして動画 2 を利用した。学習データでは一定時間固定した顔の動きを利用しているが、評価データでは瞬間的な顔の動きを用いている。そのため、一定時間固定した顔の動きと、瞬間的な顔の動きでは、抽出する顔の動きの特徴量が異なるため、識別精度が低下した可能性がある。

右のウィンク (顔の動き 2) の適合率や再現率が低いことに関して、目元の特徴点が、目を閉じる顔の動きの特徴を抽出していないことが原因として考えられる。左右のウィンク (顔の動き 1,2) を行った際に収集した画像データから顔の各特徴点を抽出し、面道具や顔の輪郭で線を結んだ画

像データを、一部抜粋して図 4 に示す。

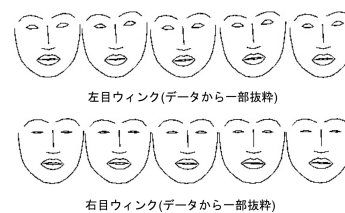


図 4 ウィンクを行った際の顔の各特徴点を線で結んだ画像

図 4 で示す通り、ウィンクしている方の目の周りの顔の特徴点がつり上がっているようには取れるが、ウィンクした方の目を閉じていると判断できていない。そのため、識別する特徴として用いられたのは、つり上がりであると考えられる。結果として、顔の特徴点のつり上がりの度合いが大きい左目ウィンクと比べ、顔の特徴点のつり上がりの度合いが小さい右目ウィンクの再現率、適合率が低くなったと考えられる。また、右目ウィンクが左目ウィンクに誤識別されたが、左目ウィンクが右目ウィンクの誤識別されないことについての考察を述べる。左右のウィンク (顔の動き 1,2) の目周りの顔の特徴点が他の顔の動きと比べ上方に位置しており、類似している。左目ウィンクの顔の特徴点のつり上がりの変化量が高く、識別する際に有効な特徴として用いられたが、右目ウィンクの顔の特徴点のつり上がりの変化量が少ない。そのため、データが与えられた時、右目ウィンクのつり上がりの微小な変化を捉えることができなかった場合に、左右のウィンク (顔の動き 1,2) で類似した特徴である「目周りの顔の特徴点が上方に位置している」という特徴により、右目ウィンクに識別されたことが原因として考えられる。

顔の傾きを利用した顔の動き (顔の動き 6,7) の再現率、適合率が低い原因として、利用した特徴量が顔の傾きの変化に弱い特徴量であることが考えられる。顔の傾きを用いた顔の動きにおける、顔の特徴点と顔の特徴点の中心の関係を図 5 に示す。図 5 で示した通り、顔の特徴点とその中心は鼻の周辺に集まっている。そのため、顔の傾きが変化した場合においても、顔の特徴点中心と顔の各特徴点の距離の変化量は少ない。そのため、顔の傾きを利用した顔の動きである顔の動き 6,7 や、平常時の顔である顔の動き 1 が特徴量について差がなかったため、お互いに誤識別され、顔の傾きを利用した顔の動き (顔の動き 6,7) の再現率、適合率が低下したと考えられる。

4.3 追加実験

4.2.3 項で挙げた、学習データを動画 1、評価データを動画 2 とした際に、精度が全体的に著しく低下した問題が、データの収集方法によって抽出する特徴量に違いが生じることが原因であるという考察内容を検討するために、追加



図 5 顔の傾きの顔の特徴点と顔特徴点の中心の関係

実験を行った。

考察内容を検討するため、動画 2 を学習データ、評価データとして精度評価を行う。予備実験より、動画 1 を学習データ、評価データとした際の識別精度は高いことが示された。動画 2 を学習データ、評価データとした際の識別精度が高い場合、抽出した特徴量は、瞬間的な顔の動きや、一定時間固定した顔の動きごとの識別には有効であると考察できる。加えて、動画 2 を学習データ、動画 1 を評価データとした際の精度評価を行う。予備実験より、動画 1 を学習データ、動画 2 を評価データとした際の精度は低いことが示された。動画 2 を学習データ、動画 1 を評価データとした際の識別精度が低い場合、一定時間固定した顔の動きと瞬間的な顔の動きで抽出する特徴量は、同じ種類の顔の動きの識別には有効であるが、抽出する特徴量に違いが生じると考察できる。以後、簡略化のために、それぞれの精度評価の実験を、追加実験 1 と追加実験 2 と定義する。

4.3.1 追加実験 1

追加実験 1 では、動画 2 から特徴量 1, 2 を算出し、各特徴量をそれぞれ学習データ、評価データとした際の精度評価の結果を表 7, 8 に示す。結果として、予備実験の結果と同様に、特徴量 1 を用いて識別した結果の F 値と比べて、特徴量 2 を用いて識別した結果の F 値が高かった。

表 7 特徴量 1 を用いた 10 分割交差検定の混同行列の結果 (gamma = 1, c = 100)

		計測値										再現率
		0	1	2	3	4	5	6	7	8	9	
実測値	0	1364	0	0	0	0	0	9	11	0	0	0.97
	1	119	0	0	0	0	0	0	0	0	0	0.00
	2	109	0	0	0	0	0	0	0	0	0	0.00
	3	124	0	0	0	0	0	0	0	0	0	0.00
	4	135	0	0	0	0	0	0	0	0	0	0.00
	5	144	0	0	0	0	0	0	0	0	0	0.00
	6	33	0	0	0	0	0	198	0	0	0	0.86
	7	23	0	0	0	0	0	0	206	0	0	0.90
	8	133	0	0	0	0	0	0	0	0	0	0.00
	9	148	0	0	0	0	0	0	0	0	0	0.00
適合率		0.58	0.00	0.00	0.00	0.00	0.00	0.96	0.95	0.00	0.00	F 値 =0.26

表 8 特徴量 2 を用いた 10 分割交差検定の混同行列の結果 (gamma = 0.01, c = 100)

		計測値										再現率
		0	1	2	3	4	5	6	7	8	9	
実測値	0	1206	18	14	13	12	24	27	33	12	25	0.87
	1	10	109	0	0	0	0	0	0	0	0	0.92
	2	17	1	90	0	0	0	0	0	1	0	0.89
	3	13	0	0	111	0	0	0	0	0	0	0.96
	4	17	0	0	0	118	0	0	0	0	0	0.87
	5	15	0	0	0	0	129	0	0	0	0	0.90
	6	24	0	0	0	0	0	207	0	0	0	0.90
	7	25	0	0	0	0	0	0	203	0	1	0.89
	8	22	0	0	0	0	0	0	0	111	0	0.83
	9	18	0	0	0	0	0	0	0	0	130	0.88
適合率		0.88	0.84	0.86	0.89	0.91	0.84	0.88	0.86	0.89	0.83	F 値 =0.84

4.3.2 追加実験 2

追加実験 2 では、4.3.1 項の特徴量 2 により作成した識別器を用いて、動画 1 について評価した結果を表 9 に示す。結果として、予備実験の結果と同様に F 値の低下が見られた。加えて予備実験の結果と同様に、右のウィンク (顔の動き 2) の適合率や再現率が低く、右目ウィンクが左目ウィンクに誤識別されたが、左目ウィンクが右目ウィンクの誤識別されなかった。また、顔の傾きを利用した顔の動き (顔の動き 6,7) の再現率、適合率がともに低かった。

表 9 動画 2 を学習データ、動画 1 を評価データとした際の 10 分割交差検定の混同行列 (特徴量 2 を利用, gamma = 1, c = 1)

		計測値										再現率
		0	1	2	3	4	5	6	7	8	9	
実測値	0	284	0	0	0	0	3	0	0	53	4	0.83
	1	228	62	0	0	0	0	0	0	0	0	0.11
	2	280	0	2	0	0	0	0	4	0	0	0.07
	3	10	0	0	266	0	0	0	0	0	0	0.96
	4	2	0	0	0	326	0	0	0	0	0	0.99
	5	2	0	0	0	0	307	0	0	0	0	0.99
	6	340	0	0	0	0	0	0	0	0	0	0.00
	7	246	0	0	0	10	1	2	14	13	0	0.05
	8	165	0	0	0	0	0	0	0	148	0	0.47
	9	4	0	0	0	0	0	0	0	0	307	0.99
適合率		0.18	1.00	1.00	1.00	0.97	0.99	0.00	0.56	0.69	0.99	F 値 =0.62

4.3.3 追加実験の考察

4.3.1 項の追加実験の結果、予備実験と同様に、特徴量 2 を用いた際の識別精度である F 値が 0.84 と高かった。また、4.3.2 項の追加実験の結果、予備実験と同様に、実環境を想定した際の F 値が 0.62 と低かった。よって 4.3 節の考察より、一定時間固定した顔の動きと瞬間的な顔の動きで抽出する特徴に違いが生じることが示唆された。

また、表 6 では、各顔の動きが平常時の顔の動きに多く誤識別されたが、表 9 では、平常時の顔が各顔の動きに多く誤識別された。これは、正解ラベルの付加方法が原因だと考えられる。動画 1 の平常時の顔としてラベル付けされているデータには、平常時の顔しか含まれていない。しかし、動画 2 の平常時の顔としてラベル付けされているデータには、顔の動き変化時の表情も平常時の顔として含まれる。表 6 では、動画 1 を学習データとして用いており、平常時の顔の識別範囲が狭くなったため、平常時の顔が他の顔の動きに誤識別されたと考えられる。表 9 では、動画 2

を学習データとして用いており、平常時の顔の動きの識別範囲が広がったため、各顔の動きが平常時の顔の動きに誤識別されたと考えられる。

加えて、追加実験2において、ウィンクや顔の傾きを利用した顔の動きの再現率や適合率が低かった。そのため、4.2.3項と同様に、利用した特徴量が、目を閉じる顔の動きや顔の傾きの変化に弱い特徴量であることが示唆された。

4.4 今後の課題

4.2.3項や4.3.3項で述べた考察より、顔の動きの識別精度向上のための課題について述べる。今後の課題として、一定時間固定した顔の動きや瞬間的な顔の動きのどちらにも適用可能な特徴量の検討やウィンクや顔の傾きに対応できる特徴量の検討、平常時の顔の識別方法の検討がある。

一定時間固定した顔の動きや瞬間的な顔の動きのどちらにも適用可能な特徴量の検討について、特徴量選択を行うことを検討している。本稿では特徴量として顔の特徴点の中心から、全ての顔の各特徴点の距離を特徴量とした。しかし、識別する顔の動きによって、顔の動きを識別するために有効な特徴量は異なる。そのため、識別する顔の動きごとに適した特徴量を用いることで、一定時間固定した顔の動きや瞬間的な顔の動きの識別のどちらにも適用可能な特徴量の検討を行う。

ウィンクや顔の傾きに対応できる特徴量の検討について、それぞれの顔の特徴点の系列性や位置関係に着目した特徴量を用いることを検討している。本稿で検討した特徴量では、それぞれの顔の特徴点の位置については考慮していたが、変化量などの系列性や、お互いの位置の関係性について考慮できていなかった。そのため、系列性やお互いの位置の関係性を特徴量とすることで、ウィンクや顔の傾きへ有効な特徴量の検討を行う。

平常時の識別方法の検討について、OC-SVM(One Class Support Vector Machine)を用いることを検討している。本稿では、SVMを識別に用いた。そのため、正解ラベルの付加方法により、平常時の顔が各顔の動きに識別されやすいか、各顔の動きが平常時の顔に識別されやすいかという違いがあった。OC-SVMでは、顔の動きごとに識別器が作成されるため、平常時の顔について考慮する必要がない。そのため、OC-SVMを用いることにより、上記の違いを考慮することが可能になると考える。また、平常時の表情を学習データとして収集する必要がなくなる。しかし、OC-SVMを用いると、登録した顔の動きの数の識別器を用いるため、評価するデータが1つの場合でも、顔の動きの識別結果が複数になる可能性がある。そのため、顔の動きの識別結果が複数になった場合も考慮した手法の検討が必要になる。

5. おわりに

本研究の目的は、PCの操作性向上のための顔の動き識別手法の提案である。本手法では、PC内蔵のWebカメラの利用を想定し、Webカメラから取得する画像を用いて、顔の動きを識別する。その後、PC操作の入力に顔の動きを割り当てることでPCの操作を行う。また、ユーザの違いに対応するために、PCの操作に利用する顔の動きは、ユーザ自身で登録を行う。

本稿では、顔の特徴点を用いて顔の動きを識別するための、特徴量やアルゴリズムの検討を行った。予備実験では顔の動きを識別する際に、有効な特徴量の検討を行った。まず、顔の特徴点の座標をx,y座標の合計で除算することで正規化した特徴量と、顔の特徴点の中心から顔の各特徴点の距離を標準化した特徴量を算出した。その後、算出した特徴量を説明変数とし、目的変数を識別する顔の動きとしたSVMによる精度評価を行った。その結果、顔の特徴点の中心から顔の各特徴点の距離を標準化した特徴量の方が、顔の動きを識別するために有効な特徴量であると示唆された。

追加実験では、予備実験の結果から、顔の動きの収集方法によって、抽出される特徴量に違いが生じると考察し、検討するための実験を行った。追加実験の結果、顔の動きの収集方法によって、抽出される特徴量に違いが生じることが示唆された。

予備実験と追加実験の結果から、本稿で用いた特徴量がウィンクや顔の傾きの特徴の抽出が困難であることが示された。また、平常時の顔の識別方法も課題であった。

今後は、一定時間固定した顔の動きと瞬間的な顔の動きのどちらにも適用可能な特徴量の検討や、ウィンクや顔の傾きに有効な特徴量の検討、および平常時の顔の識別方法を検討する。

参考文献

- [1] 東京工芸大学：東京工芸大学，ナチュラユザインターフェースに関する調査，(入手先<<https://www.t-kougei.ac.jp/static/file/nui.pdf>>)(2018.06.12)。
- [2] 池司，中洲俊信，岡田隆三：自然な手振りによるハンドジェスチャー ユーザーインタフェース，東芝レビュー，Vol.67，No.6，pp.36-39(2012)。
- [3] 久保勇貴，安藤宗孝，志築文太郎，高橋伸：能動的音響計測に基づくマイクロハンドジェスチャー認識，情報処理学会研究報告ヒューマンコンピュータインタラクション(HCI)，Vol.2018-NaN-177，No.32 pp.1-8(2018)。
- [4] 顧全，服部哲：顔によるタブレット端末の操作方法の提案，情報処理学会研究報告グループウェアとネットワークサービス(GN)，Vol.2017-NaN-101，No.6，pp.1-8(2017)。
- [5] 正井克俊，杉浦裕太，尾形正泰，クンツェカイ，稲見昌彦，杉本麻樹：AffectiveWear: 装着者の日常的な表情を認識する眼鏡型装置，日本バーチャルリアリティ学会論文誌，Vol.21，No.2，pp.385-394(2016)。
- [6] 糸数泰成，馮晨，吉田怜司，高橋裕也，菅谷みどり：ハン

ズフリーメガネ型端末による情報伝達, 情報処理学会研究報告ユビキタスコンピューティングシステム (UBI), Vol.2018-NaN-58, No.12, pp.1-9(2018).

- [7] A. Gruebler, K. Suzuki: Design of a Wearable Device for Reading Positive Expressions from Facial EMG Signals, IEEE Transactions on Affective Computing, Vol. 5, No. 3, pp. 227-237(2014).
- [8] L. Gwen, F. Ian, Bartlett, B. S. Marian, M. R. Javier: Fully Automatic Coding of Basic Expressions from Video, University of California, San Diego, San Diego, CA(2002).
- [9] 野宮浩揮, 阪上翔太, 宝珍輝尚: 決定木アンサンブルを用いた表情認識と表情強度の推定, 第8回データ工学と情報マネジメントに関するフォーラム, E5-1(2016).
- [10] D.E. King: Dlib-ml: A Machine Learning Toolkit, J. Machine Learning Research, Vol.10, pp.1755-1758(2009).
- [11] Dlib: Dlib C++ Library, , (入手先<<http://dlib.net/>>)(2018.07.12).