

オプティカルフローを用いた表情動作の 基本動作単位の抽出

樋口 大悟^{†1} 西原 悠貴^{†1}
伊藤 昭 寺田 和憲

心の状態を顔表情の動きから推測するために、表情動作に用いられる基本動作単位を抽出することを考える。まず様々な表情動作を被験者に指示して、生成される皮膚の動きをオプティカルフローとして取得する。次に得られたフローベクトルを特徴量として主成分分析を行い、その主要成分を表情生成における基本動作単位と捉える。この様にして得られた基本動作単位を用いることで、様々な表情動作を表現できることを示す。

Extraction of The Basic Action Units in Facial Expressions Using Optical Flow

TAIGO HIGHCHI,^{†1} YUUKI NISHIHARA,^{†1} AKIRA ITO
and KAZUNORI TERADA

In order to estimate the state of mind from the skin movement of facial expressions, we attempt to extract the basic action units of facial expressions. First, we instruct subjects to make various expressions and extract the movement of skin as optical flows. Next, we get the basic action units of facial expressions as main component obtained by applying PCA to the vector data of optical flows. In this paper, we show that the basic action units can express the skin movement of facial expressions.

^{†1} 岐阜大学大学院, 工学研究科
Faculty of Engineering, Gifu University

1. はじめに

コミュニケーションにおいて、人は言語情報だけではなく、声の調子、表情、視線や仕草など多くの非言語情報を用いている。相手の発言に対しての驚きの表情や、話を聞いているというサインにもなる顔色、説明をする際のジェスチャーなどがそうである。これらの非言語情報から相手の内部状態を推測することは、コミュニケーションにおいてとても重要なことである。その研究の代表とも言える表情認識は広く研究が行われているが、静止画像を用いて研究を行っているものが多く、連続的な表情の動きを扱っているものは少ない。

そこで我々は、顔面の皮膚の動きに着目し、動画画像を用いて表情の変化から相手の内部状態の変化を推測することを試みている。本研究の最終的な目標は、対話において現れる表情の変化を、画像解析によりとらえることで、対話者の内部状態の変化を推測し、それを用いることで、人と機械とのコミュニケーションを改善することである。

1.1 研究背景

オプティカルフローを用いた表情認識は古くから行われており、目や口などの部位ごとに推定したフローから表情変化時の特徴を捉えたり、また、フローから顔面筋の動きを推定するなど、様々な研究が行われている¹⁾²⁾³⁾。

表情を表現する手法として、P. Ekman の Facial Action Coding System (FACS) の Action Unit (AU) のを用いる方法がよく知られている⁴⁾。FACS は表情を表情動作の最小単位である AU の組み合わせとして記述していくシステムである。しかし、これは人が表情を観察し、その結果を記述するために考案されたシステムであり、どの AU が動いたかの判断は観測者に委ねられている。また、AU を用いた記述は「どの程度」動いたのかという定量的記述法にはなっていない。さらに、AU の動きは非線形なため、AU 同士が相互に影響することもあり、AU を組み合わせた動きは単純に個々の AU を足した動きにはならない。このため、表情変化の AU を用いた記述は、必ずしもコンピュータで利用しやすいものとは言えない。

矢崎ら⁵⁾ は、顔画像に対しオプティカルフローを用いて顔表情の移動量を測定し筋肉パラメータを推定し、それを元に顔表情の合成を行っている。顔面筋肉モデルを用いた手法では、筋肉が皮膚に与える影響を物理的に演算を行うため、動きのリアリティの面で優れている。しかし、筋肉の動きを推定する際、筋肉の移動量の計算や各筋肉の相互の影響も考慮するため、多くの処理が必要となる。

本研究では、表情生成時の皮膚の動きをオプティカルフローによって捉え、推定されたフ

ローに対し主成分分析を行うことで、表情生成時の基本動作単位となるものをこの主成分として抽出する。そして、顔の表情変化をそれらの基本動作単位の組み合わせと考えることにより、表情生成時の皮膚の動きをこの主成分の線型結合として表現する。

1.2 研究の流れ

本研究では、顔動画から得られたフローを特徴量として主成分分析を行い、表情生成時の皮膚の基本動作単位を捉える。また、表情の基本動作単位を捉えた各主成分の線形和によって、表情を表現出来ることを示す。

具体的には、まず初めに「微笑む」「しかめる」「目を見開く」「口を動かす」の4種類の顔表情の動きを被験者にカメラの前で行ってもらい、撮影した顔動画に対しオプティカルフロー推定を行い、推定されたフローに対して主成分分析を行うことで、主成分として表情生成時の顔部品の基本動作単位が捉えられることを示す。また、各表情生成時の主成分得点の変化の特徴によって表情の変化を検出出来る可能性を示す。

次に、顔上部の表情に影響を与えと言われる6つのAUの動きを撮影し、同様に撮影した動画に対してオプティカルフロー推定し、推定されたフローに主成分分析を行い、各AUの動作が抽出されたと思われる主成分を取り出す。その後、「幸福1」「幸福2」「怒り」「悲しみ」「驚き」「恐怖」とした6種類の表情動作を撮影した顔動画に対しオプティカルフロー推定を行う。6種類の表情動作時のフローデータと、AUに対応しているとした主成分から寄与得点という指標を用いて、各表情動作時に用いられている主成分の動作の大きさと、主成分を用いた表情の表現ができる可能性を示唆する。

最後に、自然な対話映像に対しても同様の処理を行う。対話映像で捉えられた表情は「笑顔」のみであったが、今回は自然な対話動画からどのような主成分が得られるか、また、得られた主成分から笑顔を表現出来るかを調査する。

2. オプティカルフロー

オプティカルフローとは、時間的に連続な一連のグレースケール画像の各画素における見かけ上の速度ベクトルである。オプティカルフロー推定法は多数提案されているが、大別すると勾配法とブロックマッチング法の2つがある。

勾配法では、輝度値の勾配（一次微分）についての基本拘束式に第2の拘束条件を付加することでフローを求める。計算コストは低いが、急激な輝度の変化や雑音に弱い。ブロックマッチング法は、画像中にある部分領域をテンプレートとして次のフレーム画像中から誤差値の最小のブロックを移動後のブロックとし、得られたベクトルをオプティカルフローと

する。この手法は誤差が少なく、雑音にも強いが、計算コストが高く、対象の拡大・縮小、回転に対しての精度はテンプレートのサイズ（ボックスサイズ）等に大きく左右される。また、ブロックのフローベクトルを表示する格子点から次のブロックの格子点までの距離をシフトサイズと呼ぶ。本研究では、計算コストは高いがフローに信頼性のあるブロックマッチング法を採用する。ブロックマッチング法の誤差値の算出式を以下に示す。

$$D_i(u, v) = \sum_{x, y} |L_i(x, y) - L_{i+1}(x + u, y + v)| \quad (1)$$

ここで、 $L_i(x, y)$ は、第 i 番目のフレームにおける画素 (x, y) の輝度値であり、和はブロック内の全ピクセルについてなされる。オプティカルフローは $D_i(u, v)$ を最小とする (u, v) 、すなわち $\vec{f}_i(x, y) = (u, v)$ で定義される。なお、本実験で用いるブロックサイズを 20×20 ピクセル、シフトサイズを 20 ピクセルとしてオプティカルフロー推定を行った。

本研究ではブロックマッチング法を用いて皮膚の動きを捉えるが、その際に明らかに皮膚の動きを捉えられていない異常値が算出される場合がある。そのため、本研究では各フローに対し、フローの絶対値が4近傍のフローの大きさの平均値の2倍よりも大きい場合、そのフロー成分を4近傍のフローの平均値と置き換える。次に、各フロー成分を4近傍の平均値とする平滑化を行う。本研究では、以後上記の処理を用いたフローベクトルを用いる。

3. 主成分分析

主成分分析とは、大きな次元を持つ特徴量について、その変化にもっとも大きく寄与している少数の成分因子を抽出し、抽出された少数の成分（主成分と呼ぶ）を用いて元のデータを近似することで、データに含まれる変数間の関係や特徴を容易に把握することを可能とする手法である。具体的には、主成分分析とは、 P 個の変数 $x_P (P=1, 2, \dots, P)$ の持つ情報を、情報の損失を最小限に抑えながら、 x_P の一次結合として与えられる互いに独立な $M (M \leq P)$ 個の主成分（総合指標） z_m

$$z_m = \sum_{p=1}^P w_{pm} x_p (m = 1, 2, \dots, M) \quad (2)$$

により近似的に表現する方法である。なお、 z_m は第 m 主成分と呼ばれ、その結合係数 $w_{pm} (p = 1, 2, \dots, P; m = 1, 2, \dots, M)$ は、以下の条件を満足するように決定される。

【条件】

第 m 主成分 $z_m (m = 1, 2, \dots, M)$ の分散は $z'_m (m' = 1, 2, \dots, m-1)$ の全てと直交する 1 次式の中で最大の分散を持つものである。ただし、

$$\sum_{p=1}^P w_{pm}^2 = 1 (m = 1, 2, \dots, M) \quad (3)$$

とする。

分析データにはフロー成分を、各格子点 i フロー方向毎に規格化したものを用いる。

$$f_{ni} = \frac{f_{ni}^* - \bar{f}_i^*}{\sigma_i}, i = (x_i, y_i, d_i) \quad (4)$$

ここで i はブロックの座標 x_i, y_i , フロー成分 d_i の 3 つの値を持つ。 f_{ni}^* は n 番目の frame における測定値を、 \bar{f}_i^* , σ_i は測定値の平均、標準偏差を表す。基本的には平均値 \bar{f}_i^* は 0 になるが、標準偏差は後に示すように顔の部分によって大きく異なる。規格化を行うことで、顔の少ししか移動しない部分（たとえば頬）の移動も、大きく移動する部分（たとえば口）と同じ重みで寄与することが保証される。

n 番目のデータにおける第 m 主成分の主成分得点を

$$t_n^{(m)} = \sum_i f_{ni} w_i^{(m)} \quad (5)$$

と表す。ここで f_{ni} としては規格化されたものが用いられている。 $t_n^{(m)}$ は、フレーム n において、第 m 主成分の寄与の相対的な大きさを示している。

また、主成分得点とは別に、本研究では寄与得点という指標を用いる。 n 番目の frame における観測されたフロー f_{ni}^* について、第 m 主成分の寄与得点を、

$$p_n^{(m)} = \sum_i \frac{f_{ni}^* w_i^{(m)}}{\sigma_i} \quad (6)$$

で表す。ここで $\frac{f_{ni}^*}{\sigma_i}$ は規格化されたフローである。我々は、フロー f_{ni}^* として、主成分を求めるのに用いた動画のフローとは異なる動画から求めたフローについても、上式で寄与得点を定義する。これは、任意のフレームのフローについて、第 m 主成分がどの程度含まれているかを表現する。

また、第 m 主成分ベクトルに対して、その格子点、フロー方向の標準偏差を掛けたものを寄与ベクトルと呼び、以下の式で定義する。

$$\tilde{f}_i^{(m)} = w_i^{(m)} \sigma_i \quad (7)$$

これは、主成分ベクトルを各格子点でのフローの大きさの分布として表現したものであり、顔画像上に寄与ベクトルを示すことでその主成分が皮膚のどの様な動きであるかを見やすく表示する。このとき、フローの平均値は 0 と仮定して無視してある。

なお、オプティカルフローから主成分分析を行う際に、各フロー成分の標準偏差に一定の閾値を設け、閾値未満の標準偏差が得られたフロー成分を除外する。これにより、細かい雑音を軽減出来ると考えられる。

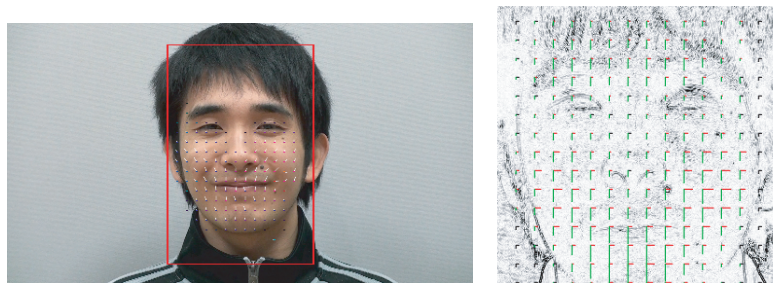
4. 実験 1：表情動作画像からの基本動作単位の抽出

次にこの実験 1 では、被験者に対話によく現れる 4 種類の大まかな顔動作を指示し、その動作をカメラの前で行ってもらった。次に、撮影した動画に対しオプティカルフロー推定を行い、推定されたフローに主成分分析を行った。これにより、表情動作を撮影したデータから求めたオプティカルフローについて、その主成分に基本単位動作となる皮膚の動きが捉えられるかどうかを調べる。

4.1 実験の流れ

皮膚の動きは、顔全体の動きに対して大変微小である。よってオプティカルフロー推定を用いて細かい皮膚の動きを捉えるためには、顔全体の動きを排除する必要がある。そこで本研究では、被験者に頭を動かさないように指示した。撮影したビデオ映像は約 2 分間 (3600frame, 30frame/sec) で「微笑む」「顔をしかめる」「目を見開く」「口を動かす」の 4 種類の動きを、被験者がカメラの前で行ったものである。画像サイズは 960×540 pixel である。図 1(a) に、実際に実験に用いた画像を示す。

撮影したビデオ映像から顔領域を検出し、グレースケールに変換後、その領域に対してオプティカルフロー推定を行い、それらの特微量として主成分分析を行う。本実験ではオプティカルフローのブロックサイズ 20×20 pixel、顔領域のブロック数を 210 個とする。フローの標準偏差に対し 0.05 の閾値を設け細かい雑音の除去を行い、フローの標準偏差から閾値未満の変数を除外してから主成分分析を行った。図 1(b) に、各フロー成分の標準偏差を示す。図中の赤、緑色の線分が、 x, y 成分の標準偏差であり、線分の長さはそれぞれ実際の 100 倍で表示している。図から、頬の部分と額の上下の運動が大きいことが分かる。黒色の線分はその成分の標準偏差が閾値未満であり、除外されたことを表す。本実験では 1 人の被験者のデータと、5 人分の被験者のデータを統合したデータに対し主成分分析を行った。被験者は全て男子学生である。



(a) オプティカルフローの図 (b) 各フローの標準偏差

図 1 オプティカルフロー推定

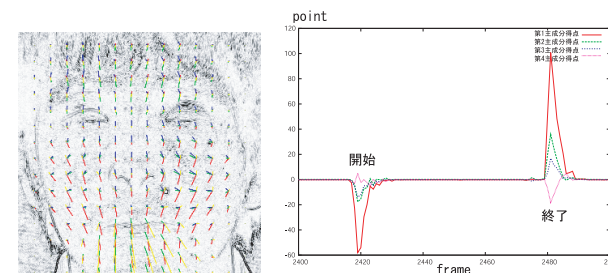
4.2 実験結果

まず、主に一人の被験者から得られたデータについて示す。第 1 から第 4 主成分の寄与ベクトルを、図 2(a) に示す。図中の赤、緑、青、黄色の線分が、各主成分の皮膚の動きを表する第 1、第 2、第 3、第 4 主成分の寄与ベクトルを表しており、ベクトルの長さは実際の 1000 倍で表示している。第 1~4 主成分までのそれぞれの寄与率は 0.26, 0.19, 0.11, 0.07, 累積寄与率は 0.63 である。これらの寄与ベクトルに対し、表 1 に示す AU との対応付けを行った。例として図 3(a) は第 1 主成分、図 3(b) は第 2 主成分の寄与ベクトルに AU を対応させた図を示す。また、表 2 に、各主成分への AU の対応付けの結果を纏めてある。表中の - は、主成分が AU の動きと逆方向であることを示す。正負の符号は寄与ベクトルの方向が AU の動きに沿った方向か、逆方向かを表す。図 3(a) から、第 1 主成分では唇から頬にかけてが斜めに上下し、口元が上下しており、AU6, 12, 26 の動きを表している。図 3(b) から、第 2 主成分では瞼から上の領域が上下し、口元が上下しており、AU1, 2, 5, 26 の動きを表している。この様に対応付けを行っていくことで、第 3 主成分は AU5, 9, 26 の動き、第 4 主成分は AU26 の動きを表しているとし、主成分に表情生成時の基本動作が捉えられていると考えられる。

次に、各主成分の主成分得点の時間変化に着目する。主成分得点の時系列変化の例として、図 2(b) に「微笑み」動作時の時間変化の様子を示す。図 2(b) は「微笑む」動きの得点変化を表している。この図を見ると、第 1 主成分得点が他の得点に比べて大きく上下していることが分かる。これは、第 1 主成分では他の主成分には無い動きである AU6, 12 を捉えているからであると考えられる。この様に、各表情動作時の主成分得点の変化にそれぞれ特徴が現れており、「顔をしかめる」動きでは第 2・第 3 主成分得点が正方向に大きく変化し、

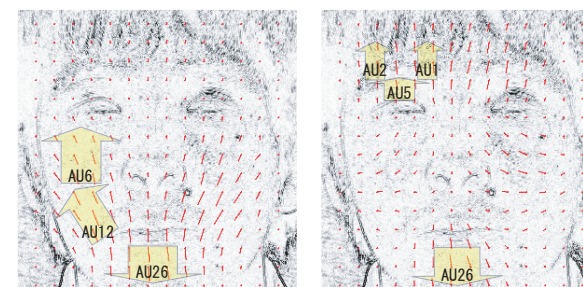
| number | action |
|--------|-------------------------------|
| AU1 | Inner Brow Raiser(眉の内側を持ち上げる) |
| AU2 | Outer Brow Raiser(眉の外側を持ち上げる) |
| AU5 | Upper Lid Raiser(上瞼を上げる) |
| AU6 | Cheek Raiser(頬を持ち上げる) |
| AU9 | Nose Wrinkler(鼻に皺を寄せる) |
| AU10 | Upper Lip Raiser(上唇を上げる) |
| AU12 | Lip Corner Puller(唇の隅を引く) |
| AU26 | Jaw Drop(下顎を落とす) |

表 1 AU 一覧



(a) 主成分の寄与ベクトル (b) 主成分得点の変化例:「微笑み」

図 2 主成分の寄与ベクトル・主成分得点

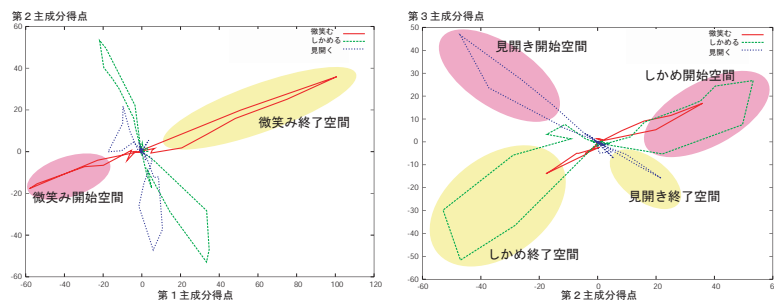


(a) 第 1 主成分の寄与ベクトル (b) 第 2 主成分の寄与ベクトル

図 3 寄与ベクトルに対する AU の対応付け

| 1 人のデータ | 現れた AU | 複数人データ | 現れた AU |
|---------|-------------------------|---------|----------------------|
| 第 1 主成分 | +AU6 -AU12 +AU26 | 第 1 主成分 | -AU6 -AU12 |
| 第 2 主成分 | AU -AU1 -AU2 -AU5 -AU26 | 第 2 主成分 | -AU1 -AU2 -AU5 -AU26 |
| 第 3 主成分 | AU +AU5 +AU9 +AU26 | 第 5 主成分 | -AU9 -AU26 |
| 第 4 主成分 | AU +AU26 | 第 4 主成分 | AU10 -AU26 |

表 2 主成分への AU の対応付け



(a) 第 1・第 2 主成分得点空間

(b) 第 2・第 3 主成分得点空間

図 4 各表情動作の主成分得点空間での表現

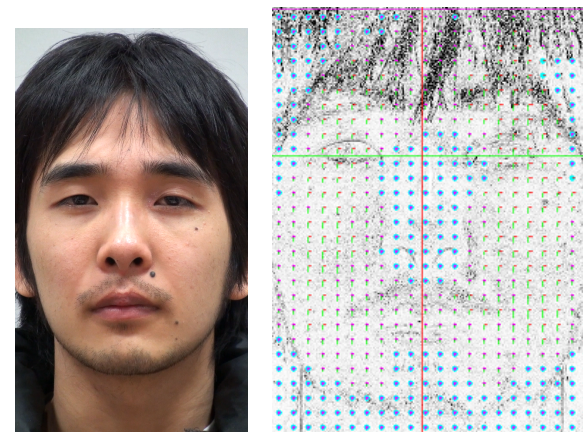
「目を見開く」動きでは第 2 主成分得点は負方向，第 3 主成分得点は正方向に大きく変化，「口を動かす」動きでは，第 4 主成分得点が大きく変化した。

ここで「微笑む」「顔をしかめる」「目を見開く」の 3 つの動きについての得点の時間変化を，第 1，第 2 主成分得点を軸として空間上に表したものを図 4(a) に示す．この図から，「微笑む」動きをした時の得点の軌跡が他の動きの得点の軌跡と大きく異なり，特徴的であることがよくわかる．これより「微笑む」動きの空間を他の動きと分離することができる．同様に第 2，第 3 主成分得点を軸に空間上に表したものを図 4(b) に示す．この図では「顔をしかめる」と「目を見開く」の軌跡の特徴が大きく異なっており，この 2 つの動きも分離することができると思われる．これにより，様々な表情動作を主成分を用いて分類することが出来る．

また，4 種類の動きを撮影した同様のフローデータを 5 人分集め，それらを統合したデータを主成分分析を行った．表 2 に，1 人の被験者データと統合したデータの主成分への AU の対応付けの結果を並べて示した．表から分かる様に，複数人データにおいても 1 人の被験者のデータと同様の AU で対応づけられる主成分があり，また，主成分得点空間で表現した

際も 3 つの動きが分離出来たことから，表情表出における皮膚の動きには複数人で共通性があることがわかり，他の顔動画を用いた際にも同様に表情の分類できると考えられる．

5. 実験 2：基本表情からの基本動作単位の抽出



(a) 撮影した動画

(b) 標準偏差

図 5 実験に用いた顔動画とフローの標準偏差

| | |
|-----|------------|
| AU1 | 「眉の内側を上げる」 |
| AU2 | 「眉の外側を上げる」 |
| AU4 | 「眉を下げる」 |
| AU5 | 「上脛を上げる」 |
| AU6 | 「頬を持ち上げる」 |
| AU7 | 「脛を緊張させる」 |

表 3 実験に用いる AU

| 表情動作 | 使用する AU |
|--------|-----------------|
| 「幸福 1」 | AU6 |
| 「幸福 2」 | AU1 + 2 + 5 + 6 |
| 「怒り」 | AU4 + 5 + 7 |
| 「悲しみ」 | AU1 + 4 |
| 「驚き」 | AU1 + 2 + 5 |
| 「恐怖」 | AU1 + 2 + 4 |

表 4 撮影した表情動作

実験 2 では，FACS の基本的な AU をオプティカルフローとして抽出しそれを基本動作単位として，基本表情に含まれる AU を検出する．

5.1 AU のデータの主成分分析

まず，AU の中でも顔上部の表情に関係する 6 つの AU を個々に撮影した動画を用意す

る．本実験で対象とした AU を表 3 に示す．なお，今回は顔上部の表情に関する動作を対象とし，それ以外の AU は省略する．

この実験において撮影した動画は，撮影するカメラを 90 度傾け，縦に長く撮影を行い，可能な限り大きく顔を映すように撮影したものである．本実験においても，顔全体の動きを排除する必要があるため，顔動画を撮影する際に被験者に顔の位置を固定するよう指示した．画像サイズは 540×960pixel である．図 5(a) に，本実験で用いた顔動画の一つを示す．

オプティカルフローは，ブロックサイズを 20pixel，シフトサイズを 20pixel とし，ブロックの数は縦 29 個，横 22 個の合計 638 個である．顔領域のフローの抽出には，顔の中心線と目の高さを大まかに決め，顔の中心線から左右に 11 個，目から上に 10 個，目から下に 19 個のブロックが展開されるように調整した．

まず，AU の動作をカメラの前で繰り返し撮影を行った．6 つの AU を 1 つ 1 つ個別に撮影を行う．各動画の長さは約 90 秒 (2700frame,30frame/sec) とした．次に，その 6 つの動画のフローデータを統合して主成分分析を行い，主成分から AU の特徴を捉えたものを取り出す．本実験では，フローの標準偏差に対しての閾値を 0.2 とし，閾値未満の変数を除外し主成分分析を行った．図 5(b) に，標準偏差を 20 倍した図を示す．赤の線が X 方向の標準偏差，緑の線が Y 方向の標準偏差を表し，実際の値を 20 倍したものを表示している．ピンクの点が表示されているブロックでは X 方向の標準偏差が閾値より小さく，水色の点が表示されているブロックでは Y 方向の動きが閾値より小さいことを表す．図から髪の部分，鼻，顔領域外の小さい動きが取り除かれていることが分かる．

表 5 に，AU の特徴が捉えられた主成分を示す．正負の符号の + は寄与ベクトルの方向が AU の動作方向と一致しているもの，- は動作方向の逆方向を向いていることを表す．図 6(a) に，第 1 主成分の寄与ベクトルを 500 倍したベクトルを示す．AU に対応しているとした第 1 から第 5 主成分の寄与率は，0.42，0.14，0.06，0.05，0.03 となっている．5.2 においてこれらの主成分を用いて各表情動作の表現を行うが，AU5 に対応する動きは，正しく主成分として分離できなかったため，5.2 では表情動作を表現するにあたって AU5 を用いなかった．

5.2 表情動作の主成分による表現

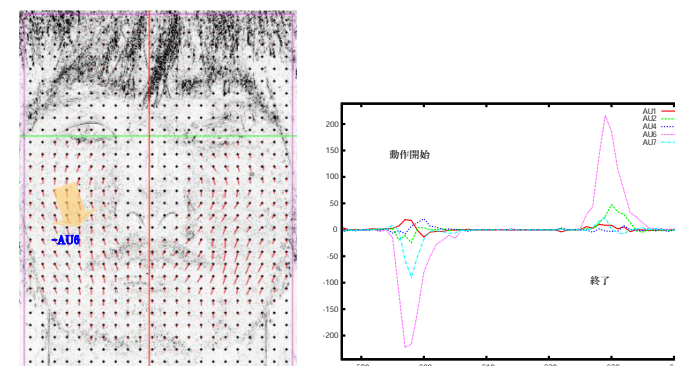
ここでは，表 3 に示した AU を用いて表現出来る表情として，図 4 に示した 6 つの表情を取り上げ，それらの表情から各 AU が正しく抽出出来るかどうかを調べる．各表情時に用いる AU は，他の AU を用いた表情の論文を参考に決定した．⁽⁶⁾⁽⁷⁾⁽⁸⁾⁽⁹⁾ 表 4 に示した 6 種類の表情動作を被験者にカメラの前で各 6 回繰り返し行ってもらった．動画の長さは約 2 分

| 主成分 | AU |
|---------|-------|
| 第 1 主成分 | - AU6 |
| 第 2 主成分 | - AU4 |
| 第 3 主成分 | - AU1 |
| 第 4 主成分 | + AU7 |
| 第 5 主成分 | - AU2 |
| 第 8 主成分 | - AU5 |

表 5 AU の抽出された主成分

| | 使用した AU | 寄与得点が変動する AU |
|--------|-----------|----------------|
| 「幸福 1」 | AU 6 | +AU6 |
| 「幸福 2」 | AU1 2 5 6 | +AU1 -AU4 +AU6 |
| 「怒り」 | AU 4 5 7 | +AU4 +AU6 +AU7 |
| 「悲しみ」 | AU 1 4 | +AU1 +AU4 +AU6 |
| 「驚き」 | AU 1 2 5 | +AU1 -AU4 +AU7 |
| 「恐怖」 | AU 1 2 4 | +AU1 +AU4 +AU6 |

表 6 検出された AU



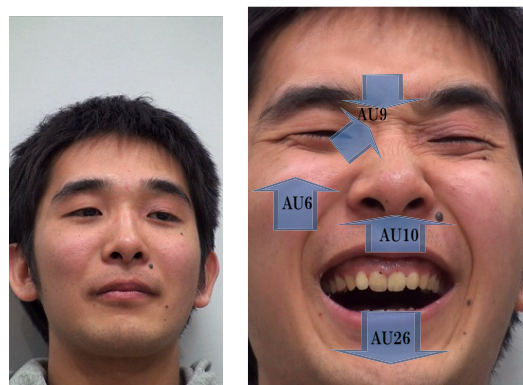
(a) 第 1 主成分 (AU6) (b) 「幸福 1」動作時の主成分得点の変化

図 6 主成分・寄与得点

| | 第 3 (-AU1) | 第 5 (-AU2) | 第 2 (-AU4) | 第 1 (-AU6) | 第 4 (+AU7) |
|------|----------------------|---------------|----------------------|----------------------|---------------------|
| 幸福 1 | 11.40(5.67) | -20.50(10.23) | 3.81(1.90) | -200.33(100) | -44.53(22.23) |
| 幸福 2 | -89.56(38.38) | -24.28(10.41) | 122.08(52.32) | -233.33(100) | -67.20(28.80) |
| 怒り | -25.98(29.15) | -10.39(11.65) | -89.13(100) | -44.25(49.64) | 33.61(37.71) |
| 悲しみ | -66.88(100) | 12.25(18.31) | -45.97(68.73) | -25.46(38.06) | 13.70(20.48) |
| 驚き | -97.27(56.99) | -2.34(1.37) | 170.67(100) | -33.49(19.62) | 69.04(40.45) |
| 恐怖 | -36.55(35.52) | -29.79(28.95) | -40.73(39.58) | -102.90(100) | -3.06(2.98) |

表 7 各感情表情動作開始時の AU の寄与得点の最大値の平均

間 (3600frame,30frame/sec) とし，撮影した動画に対して，同様にオプティカルフロー推定を行う．本実験において，表情動作から推定されたフローと，5.1 で求めた AU に対応する主成分から，3 の式 (6) に示した寄与得点という指標を用いる．図 6(b) に「幸福 1」動作時の寄与得点の時間変化を示す．6 回繰り返し替えた各表情動作の開始時における寄与得点の最大値の平均値を表 7 に示す．赤で示された数値は，各表情動作で寄与得点の最大値の平



(a) 実験に用いた対話映像 (b) 「笑い」動作

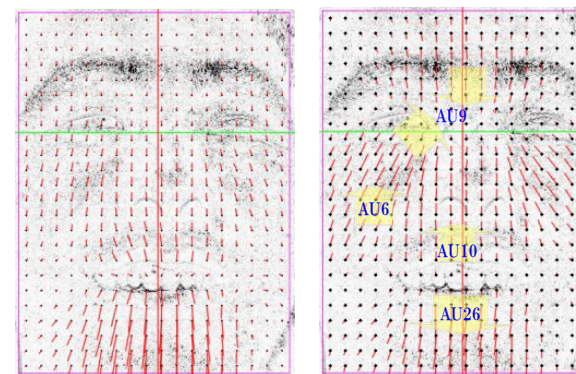
図 7 対話動画と対話中の「笑い」動作

均のが最も高かった主成分であり、括弧内の数値はそ値を 100 とした時の相対的な大きさを表し、以降相対値と記す。相対値の大きさが 30 以上の主成分を各表情動作時に用いられた動作とし、表 6 に示す。表 6 から、検出された主成分の AU と使用した AU が概ね一致していることが分かる。また、「幸福 2」「驚き」の、AU1+2 の眉全体を大きく上げる動きが AU4 の主成分の逆方向の動きとして表現され、主成分の組み合わせが各表情動作の動きとほぼ一致していることが分かった。これにより、表情に用いられる主成分の動きとその大きさが表現でき、主成分により表情を表現出来る可能性がある。

6. 対話映像での主成分による表情表現

これまでの実験において撮影された顔動画は、被験者に動きを指示した人為的な表情を撮影した顔動画を用いた。しかし人の内部状態を表情から推定しようとした際、人の自然な表情を捉えたデータは必要不可欠である。そこで、複数人での対話映像の撮影を行い、対話中に表れる自然な表情を対象に実験を行った。図 7(a) に、実際に用いた対話映像を示す。

対話映像の撮影では、被験者とその知り合いで構成された複数人での会話の撮影を行った。撮影は約 20 分間行い、その内の表情の変化が多い 4200frame(約 2 分 20 秒,30frame/sec) を切り取った動画を実験に用いた。画像サイズは 540×960pixel である。縦 24 個、横 18 個の合計 432 個のブロックを用いて顔領域のオプティカルフロー推定を行った。前述したように、オプティカルフロー推定を用いて細かい皮膚の動きを捉えるためには、顔全体の動きを排除する必要があるため、対話中に被験者には頭を後ろにつけてもらい、できる限り頭



(a) 第 1 主成分 (b) 「笑う」動作の寄与ベクトル表現

図 8 主成分と「笑う」動作の表現

を動かさないように指示した。しかし、自然な対話時には表情変化に伴って顔全体が動くことが多かった。そこで、顔領域の動きをフローから除去するため、顔領域のフローの平均値を各フローから差し引いたデータを用いた。更に、顔領域全体のフローの平均値を引いたことにより起こる雑音を除去するため、フロー成分の大きさに対し閾値 0.6 を設けた。対話映像から様々な表情動作が見られることが期待されたが、「笑う」動作と会話のために口を動かさず動きがほとんどであった。しかし、対話動画中では「笑う」表情が多く見られたため、本実験では、対話映像からどのような主成分が捉えられるかを調査し、またそれらを用いて「笑う」表情の表現を試みた。7(b) に「笑い」動作時の表情を示し、動作時に用いたと思われる AU を示す。この被験者の「笑い」動作には、AU6「頬を持ち上げる」、AU9「鼻に皺を寄せる」、AU10「上唇を持ち上げる」、AU26「顎を下げて唇を開く」の四つの AU が用いられていると思われる。

6.1 実験結果

| frame | 第 1 主成分 | 第 2 | 第 3 | 第 4 | 第 5 |
|-------|--------------------|----------------------|----------------------|---------------------|---------------|
| 545 | 116.30 | 67.75 | -59.56 | 77.14 | 36.84 |
| 1790 | 106.08 | 107.93 | -74.23 | 41.79 | -44.45 |
| 3525 | 121.51 | 137.59 | -41.57 | 26.77 | -33.31 |
| 平均値 | 114.63(100) | 104.42(91.10) | -58.45(50.99) | 48.57(42.37) | -13.64(11.90) |

表 8 「笑い」動作開始時の主成分得点の最大値

対話動画から推定された顔領域のフローに対し主成分分析を行って得られた主成分の例として、第1主成分を、8(a)に示す。第4主成分までの寄与率は第1主成分から順に0.18, 0.12, 0.08, 0.70で、累積寄与率は0.45である。主成分分析時には各フロー成分の標準偏差に閾値0.08を設け、雑音の除去を行った。第1主成分には会話の際の口を動かす動きが現れ、第2主成分には「笑う」動作に類似した動きが捉えられた。他には第5・第6主成分にまばたきの動作が捉えられたが、それら以外に表情に関連がありそうなものはなかった。

次に、主成分得点の変化に着目する。2分間に切り取った対話動画の中から、被験者が大きく「笑う」表情をした3回の表情変化時の各主成分得点のデータを用意し、「笑う」動作時の主成分得点の大きさの最大値の平均を求めた。表8に、3回の「笑い」動作時の第1から第5主成分得点の絶対値の最大値とその平均値を示す。「笑う」動作が確認されたのは、545frame, 1790frame, 3525frame 付近である。平均値に示されている括弧内の数値は、最大値の平均が最も大きい値を100とした時の大きさ、相対値を示す。今回、相対値が30以上の主成分の動きを「笑う」動作時に用いられた動きとした。その結果、第1から第4主成分の相対値が30以上となり、「笑う」動作時に用いられた動きとしてみなした。第1から第4主成分までの寄与ベクトルに相対値の100分の1の値をかけて足したベクトルを、図8(b)に示す。図7(b)と図8(b)を比較すると、「笑う」動作時の皮膚の動きと寄与ベクトルで表現された皮膚の動きは概ね一致している様に見える。図8(b)に示した様に、寄与ベクトルをAUに対応づけを行ったところ、被験者が「笑う」動作の際に用いたと思われるAUと一致した。このことから、自然な対話映像からも主成分に基本動作単位を抽出出来ると考えられる。

7. ま と め

本研究では、様々な顔動画に対してオプティカルフロー推定を行い、フローを特徴量として主成分分析を行うことで、表情生成時の基本動作単位を主成分として取り出し、それらを用いて表情の表現が出来る可能性を示唆した。

まず「微笑み」「しかめ」「目を見開く」「口を動かす」の4つの動作を被験者に行ってもらい、推定されたフローに対し主成分分析を行った。その結果、表情の基本動作単位と考えられる主成分が現れ、主成分得点で各表情動作を分類出来る可能性がある。

次に、6つのAUに対応すると思われる主成分を用いて、6つの表情動作時の皮膚の動きの表現を行った。その結果、主成分の寄与得点から、各表情生成時に用いられた主成分の動きとその大きさを表現することができ、主成分によって表情が表現出来る可能性がある。

最後に、自然な対話映像から、主成分にどのような基本動作単位が現れるかを調査した。その結果、対話中の「笑う」の基本動作単位となる主成分に捉えられ、またそれらを用いて「笑う」動作を表現することが出来た。

今後の課題としては、顔全体の動きに対して強靱な手法の提案と「笑い」以外の自然な表情の顔動画の撮影が挙げられる。

参 考 文 献

- 1) 大塚 尚宏, 大谷 淳, “オプティカルフローを用いた実時間表情認識に関する検討”, 1997年映像情報メディア学会年次大会
- 2) 間瀬 健二: “AN APPLICATION OF OPTICAL FLOW EXTRACTION OF FACIAL EXPRESSION”, MVA'90, Nov.28-30,1990
- 3) 吉川 奈穂美, 桜井 正広, 宮川 道夫: “目・口領域のオプティカルフロー推定による「笑い」認識の試み”, 信学技報 PRMU2003-164 (2003-11)
- 4) Paul Ekman, Wallace V.Friesen, 工藤 力 訳: “表情分析入門”, 誠信書房 (1987)
- 5) 矢崎 和彦, 石川 貴博, 森島 繁生: “オプティカルフローを用いた正面顔画像からの顔面筋パラメータの自動推定”, 信学技報, PRMU97-281(1998-03)
- 6) 大塚 尚宏, 大谷 淳, “ノンバーバル情報を伝達可能な通信環境の検討”, 信学技報, PRMU08-224(1999-02)
- 7) 安藤 泰之, 米倉 達広: “表情を入力に用いた感情対話システムの試み”, 信学技報, MVE96-72(1997-03)
- 8) 番場あやの, 上村 保子: “基本6表情認知における注視部位の基礎的検討”, 昭和女子大学大学院生活機構研究科紀要 Vol.16-2 (2007)
- 9) 柳澤 博昭, 祖川 慎治, 前島 謙宣, 四倉 達夫, 森島 繁生: “モーションキャプチャによる顔表情の定量表現”, 信学技報, HCS2004-50 (2005-3)