法政大学学術機関リポジトリ
HOSEI UNIVERSITY REPOSITORY

PDF issue: 2025-07-16

動画像における3次元特徴点を用いた日本人の基本感情と表情の相関性について

SHINOHARA, Rikuto / 篠原, 陸人

(出版者 / Publisher)
法政大学大学院理工学研究科
(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 理工学研究科編
(巻 / Volume)
64
(開始ページ / Start Page)
1
(終了ページ / End Page)
6
(発行年 / Year)
2023-03-24
(URL)
https://doi.org/10.15002/00026400

動画像における３次元特徴点を用いた日本人の基本感情と表情の相関性について

USING 3D FEATURE POINTS IN VIDEO IMAGES RELATIONSHIP BETWEEN BASIC EMOTIONS AND FACIAL EXPRESSIONS OF JAPANESE PEPOPLE

篠原陸人
Rikuto SHINOHARA
指導教員　赤松茂

法政大学大学院理工学研究科応用情報工学専攻博士前期課程

With the recent increase in remote meetings and other forms of communication, nonverbal communication that reads the emotions of others through facial expressions and gestures is becoming increasingly important. It is generally accepted that emotions and facial expressions are related, but most of the research has focused on Western human facial expressions. We measure and analyze data based on the assumption that emotions and facial expressions defined by Japanese people cannot be mapped. Test data for facial expressions are recorded as moving images of the subject's specific emotional expressions using MediaPipe, which can capture 3D feature points. The correlation between basic emotions and facial expressions was analyzed by detecting change points based on the recorded 3D feature points.

Key Words *: Face recognition, expressing emotion, change point detection, three dimensions*

１．はじめに

感情認識は幅広い分野で活用され、現在も市場規模を拡大させている. 感情認識において多くの研究が FACS 理論に基づいて顔の表情分析を行っている[1]. FACS とは人間の表情について 40 を超えるアクションユニット(AU)を特定し、そのアクションの組み合わせによって認識を行う技術である. しかし、FACS 理論を含む近年の高精度な感情認識は西洋文化で実施されており東洋人においては西洋人と同様な対応付けができない可能性がある.

そこで本研究では日本人を対象として感情と表情の関係について動画像と 3 次元顔特徴点の側面から分析を行った.

２．研究背景

基本感情と表情は感情認識において相関性があると報告されており、Ekman 氏は FACS 理論に基づいて表情から感情推定を行った.しかし、2019 年度に佐藤氏[2]が東洋人においてはシナリオに基づく感情表現において、怒り、恐怖、悲しみ、嫌悪が従来定義されている感情と表情との関連性がなく異なる表情を生成すると指摘している.また、2021 年度に斉藤氏[3]が 30 代以上の日本人(東洋人)の基本感情と表情には対応づかない傾向があると発表している.しかしこれらの研究は 2 次元の静止画像にて検証を行なっている.そこで本研究は 3 次元の動画像から検証を行う.

３．実験環境

（１）計測ソフトウェア

本研究では顔の 3 次元の動きを計測するために MediaPipe を用いた[4].MediaPipe は、ビデオやオーディオなどの任意の感覚データに対してコンピュータビジョン推論を実行するパイプラインを構築するためのオープンソースフレームワークである.本研究では図 1 に示すように 478 の 3D 顔ランドマークを記録した.

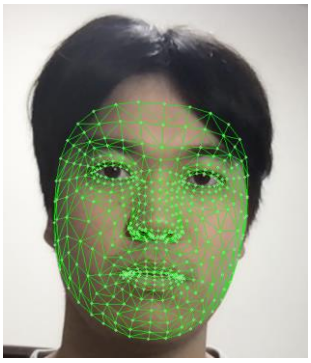

図１　取得できる顔特徴点

（２）計測データ

本研究では基本 6 感情を対象として計測を行った.計測方法としては初めに真顔を 5 秒間撮影し､その後笑顔を 5 秒間撮影、同じ流れで驚き、悲しみ、怒り、恐怖、嫌悪と撮影を行った.すなわち、真顔→笑顔→真顔→驚き→真顔→悲しみ→真顔→怒り→真顔→恐怖→真顔→嫌悪とそれぞれの表情を 5 秒ずつ合計 1 分の動画を 1 サンプルとして計測した.

また、表情撮影に関して 3 通りの方法で撮影した.1 つ目は感情表現の単語のみを提示した方法(以下、単語条件と提示)、2 つ目は西洋人が感情表現を成しているサンプルデータを付加した方法(以下､写真条件と提示)にて撮影した.最後の方法は 6 つの刺激感情(幸福、驚き、悲しみ、怒り、恐怖、嫌悪)を与える動画を見てもらいその表情に関して撮影を行った(以下、刺激感情付加条件と提示).すなわち、単語条件と写真条件の方法はこちらが感情表現を指定し意図的に表情を出してもらい、刺激感情付加条件の方法はこちらが指定を行わず自然的に発生した表情を計測する方法となっている.

４．データ処理

（１）正規化

撮影データについて、測定ごとに被験者の顔の位置や向きにばらつきが生じる.そのため顔特徴点座標について座標変換を行った.基準点には両目尻のマーカ点と鼻の頂点のマーカ 1 点を選んだ.

（２）低次元圧縮化

本研究で撮影したデータは 478 個のマーカ点の x,y,z 座標値からなる 1434 次元のベクトルで表される多次元データである.識別実験において多次元データは計算処理の負担が重くなり結果が出ない.そこで主成分分析を行うことにより、多次元データの持つ情報をできるだけ損なわずに低次元空間に情報を縮約する[5].

（３）識別顔パーツ

本研究は撮影したデータについて全体で分析するのではなく顔のパーツや高さで分析を行った.識別に用いた顔のパーツを表 1 に示す.

表１　識別顔パーツ

No.	顔の部位・動作	No.	顔の部位・動作
1	左目	9	右目の虹彩
2	左目の開き具合	10	左眉
3	右目	11	右眉
4	右目の開き具合	12	内側眉
5	唇	13	外側眉
6	唇端点	14	鼻
7	口の開き具合	15	顎
8	左目の虹彩	16	おでこ

パーツについて例えば目に注目した際、MediaPipe では目について 1 点だけではなく 16 点の x,y,z 座標取れる.表 1 の No.1 の左目は x,y,z 座標値を 16 点取り低次元圧縮化してから分析処理を行う.他のパーツにおいても同様の処理を行う.

５．分析手法

（１）ChangeFinder

ChageFinder とは山西氏が NEC 在籍時代に発明した手法である[6][7].従来の外れ値検出手法は時系列データには非対応であったがこれに対し ChangeFinder は時系列データと変化点検出にも対応させた手法である.SDAR モデルを従来手法に導入し、外れ値検出と変化点検出を以下のように 2 段階の「検出」ステップに分離し、同時に処理する.

○第 1 段階学習

①SDAR モデルで、各時点における時系列モデルを構築.

②①で構築したモデルを元に、次の時点の実測値が出てくる尤度を求める.

③この確率の対数損失を求め、外れ値スコアとする.

外れ値スコアは以下のように計算する.

$$Score(x_t) = -\log P_{t-1}(x_t|x^{t-1}) \tag{1}$$

幅 W>0 を設定して、その幅の中で外れ値スコアを平滑化する.平滑化を行うことで、外れ値の状態ではなく、その状態が長く続いているかどうかすなわち変化しているかどうかを判別することができる.t 時点における平滑化されたスコアを y_t とおくと以下のように計算できる.

$$y_t = \frac{1}{W}\sum_{t=t-W+1}^{t} Score(x_i) \tag{2}$$

○第 2 段階学習

①平滑化で得られたスコアをさらに SDAR モデルで学習

②①で構築したモデルを元に、次の時点の実測値が出てくる「確率」(忘却係数による調整あり)を求める

③この確率の対数損失を求め、変化点スコアとする

（２）主観評定実験

これまでの実験方法はコンピュータによる処理による結果であったが、人間による感情推定の結果も必要である.よってある感情表現をしている表情について人間がどの表情かを判断してもらう主観評定実験を行った.評定してもらうデータセットには基本 6 感情(幸福・驚き・悲しみ・怒り・恐怖・嫌悪)を用意し、選択肢は基本 6 感情に加え真顔も選択肢に含めた.真顔を選択肢に含めた理由として、被験者がデータセットの表情を見た際に変化があるのか、ないのか(真顔)判断するためである.感情表現があるのにも関わらず真顔を選択するということは客観的に見て表情だけでは感情表現が推定できないことと同義であると定義できる.

（３）評定モデル作成

写真条件、単語条件、主観評定実験の結果から感情推

定モデルを作成する.評定モデルに関しては 3 種類作成した.1 つ目は単語条件の結果から顔パーツの変化度合いによって識別する手法(以下、変化点検知手法と提示する.)である.2 つ目は感情認識の従来手法でもあるアクションユニットを用いた Py-Feat、3 つ目は深層学習 LSTM を用いて単語条件にて撮影した被験者を訓練データとしてモデルを作る手法である.

（４）変化点検知手法

変化点検知手法は顔パーツの変化度合いを組み合わせることで感情推定を行う.本研究では 16 パーツ分析したが、より変化が大きい目の開き具合、口の開き具合、眉の動きに着目した.

（５）Py-Feat

Py-Feat は、画像やビデオから表情(アクションユニット、感情、顔のランドマーク)を簡単に検出し、表情データを視覚化するためのツールとモデルの包括的なセットを提供する[8].本研究では感情認識のために利用した.

（６）LSTM

LSTM(Long short-term memory) は、RNN(Recurrent Neural Network)の拡張として 1995 年に登場した、時系列データに対するモデル、あるいは構造の 1 種である[9].LSTM は RNN の中間層のユニットを LSTM block と呼ばれるメモリと 3 つのゲートを持つブロックに置き換えることで実現できる.

６．実験結果

（１）データセット

単語条件と写真条件は 13 人にそれぞれ 3 回ずつ撮影することで計 39 のデータを取得した.刺激感情条件は 7 人の被験者に 6 個の動画を 1 回ずつ撮影した.主観評定実験は 120 個のデータについて 18 人の被験者が評定を行った.

（２）写真条件

図 2 に写真条件の実験結果を示す.

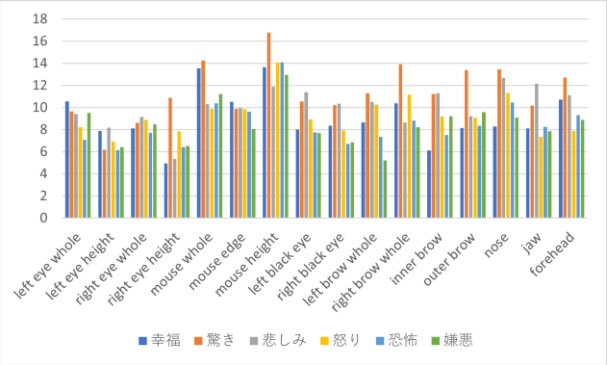

図 2　実験結果（写真条件）

口の開き具合に関しては全ての感情表現で全体的に高いスコアとなった.驚きの感情表現に関してはほとんどの顔パーツにおいて他の感情表現よりもスコアが高くなり、幸福の感情表現は口に関する変化スコアは大きかったが、他の顔パーツに関しては低いスコアとなった.悲しみも全体的にスコアが高く、特に鼻や顎のスコアが高かった.

（３）単語条件

図 3 に単語条件の実験結果を示す.

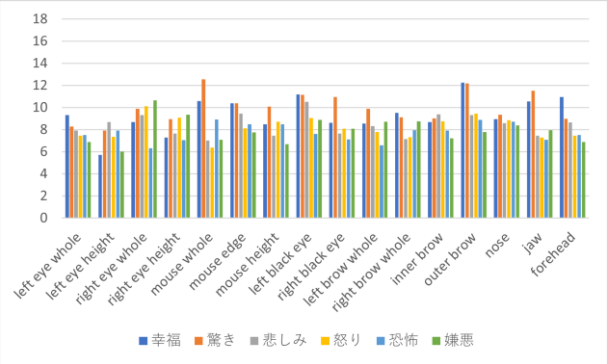

図 3　実験結果（単語条件）

写真条件と比較すると全体的に低いスコアとなり、驚きの感情表現に関しては写真条件よりはスコアは落ちるものの口全体、眉の動きのスコアは高かった.幸福は口端点(口角)に関しては写真条件と変わらないものの口の開き具合はスコアを落とした.悲しみは目の部位に関しては写真条件と比較しても遜色はなかったものの、他の部位のスコアは落ちていた.怒りは目の動きと眉の動きは写真条件と変わらず、口の動きは大幅にスコアを落とした.恐怖はどの部位も大きくスコアを落とした.嫌悪は目の動きは変わらないものの、他の部位はスコアを落とす形となった.

（４）主観評定結果(写真条件)

表 2 に写真条件の主観評定結果を示す.

表 2　主観評定結果(写真条件)単位:%

	真顔	幸福	驚き	悲しみ	怒り	恐怖	嫌悪
幸福	1	95	1	0	2	0	1
驚き	1	1	98	1	0	1	0
悲しみ	14	0	0	65	4	2	14
怒り	1	0	1	17	52	3	27
恐怖	1	0	27	16	9	32	16
嫌悪	0	1	2	4	67	0	26

幸福のデータセットや驚きのデータセットはいずれも 95%以上の一致となり表情と感情の関連性が取れた.悲しみは過半数以上が一致ではあったが、悲しみの感情表情に対して真顔と選ぶ割合は他の感情表現の中で最も高かった.怒りと嫌悪に関してはお互いに誤認識する結果となった.特に嫌悪の表情にも関わらず怒りと選択した割合は 67%と非常に誤認識する割合が高くなってしまった.恐怖は全体的に他の感情表現を選ぶ被験者も多く、驚きを選択した割合は 27%、悲しみを選択した割合は 16%、怒りを選択した割合は 9%、嫌悪を選択した割合は 16%となった.悲しみの感情表情に関しては 14%が真顔と選択していたが、他の感情表現に関しては 1%以下となりほとん

どの写真条件のデータは何かしら表情に変化が起きている(真顔ではない)と感じた被験者が多い結果となった.

（５）主観評定結果(単語条件)

表 3 に単語条件の主観評定結果を示す.

表 3　主観評定結果(単語条件)単位:%

	真顔	幸福	驚き	悲しみ	怒り	恐怖	嫌悪
幸福	27	71	0	2	0	0	1
驚き	4	1	94	1	0	1	0
悲しみ	37	1	1	39	6	5	12
怒り	29	1	1	14	34	1	19
恐怖	53	10	18	8	2	1	8
嫌悪	22	0	2	13	35	3	26

驚きの感情表現に関しては 94%と写真条件の結果と遜色ない結果となった.幸福は正確率においては 71%と高い結果ではあったが、写真条件の 95%よりは低い結果となった.特に幸福の感情表現ではあるが真顔と選択した割合が 27%と高かった.悲しみは真顔、悲しみ、嫌悪、3 つの選択肢を選ぶ被験者が多かった.怒りと嫌悪は写真条件の結果と同様、お互いに誤認識する結果となった.恐怖は主観評定結果の中で最も正確率が悪く、1%となってしまった.恐怖に関しては過半数が真顔を選択しており、そもそも表情に変化が起きていないと感じている被験者が多かった.また、全体的に驚きの感情表現以外は真顔を選択する割合が写真条件と比較した場合多い割合となった.

（６）刺激動画付加条件・変化点検知手法

表 4 に刺激動画付加条件・変化点検知手法の実験結果を示す.行が検出した感情表現を、列が刺激感情動画を示す.

表 4　刺激動画付加条件・変化点検知手法　単位:%

	幸福	驚き	悲しみ	怒り	恐怖	嫌悪
幸福	17.7	62.6	12.6	4.7	1.6	0.8
驚き	14	72.1	10.7	1.4	0.9	0.9
悲しみ	16.1	66.3	8.2	5.5	1.6	2.4
怒り	17.6	64.7	11.4	4.3	1.2	0.8
恐怖	13.3	71.8	7.9	4.6	0.8	1.7
嫌悪	10.5	69.0	8.9	5.4	3.5	2.7

全ての動画に関して驚きの割合が 6 割を超える形となった.幸福、驚き、怒りの動画に関しては幸福と悲しみが 1 割を超えた.悲しみは幸福が 1 割を超えそれ以外は 1 割未満だった.恐怖の動画は驚きの割合が 7 割を超える形となった.嫌悪の動画は他の動画と比べると嫌悪の割合が最も高くなった.

（７）刺激動画付加条件・Py-Feat

表 5 に刺激動画付加条件・Py-Feat の実験結果を示す.

表 5　刺激動画付加条件・Py-Feat　単位:%

	幸福	驚き	悲しみ	怒り	恐怖	嫌悪
幸福	18.2	61.1	0	20.6	0.1	0
驚き	0	70.3	4.2	22.9	2.2	0.4
悲しみ	0	78.3	1.0	11.8	7.0	1.9
怒り	0.0	30.9	9.1	49.1	10.9	0
恐怖	4.8	84.4	2.5	7.3	1.0	0
嫌悪	0	65.1	3.7	29.4	1.8	0

Py-Feat に関しても怒りを除く全ての動画に関して驚きの割合が高くなった.しかし、怒りと判定する割合も高く特に怒りの動画は半分近く怒りとなった.

（８）刺激動画付加条件・深層学習

表 6 に刺激動画付加条件・深層学習の実験結果を示す.

表 6　刺激動画付加条件・深層学習　単位:%

	幸福	驚き	悲しみ	怒り	恐怖	嫌悪
幸福	32.1	67.9	0	0	0	0
驚き	14.2	61.7	0	1.4	2.1	20.6
悲しみ	12.4	48.8	0	1.6	2.3	34.9
怒り	21.8	54.5	0.6	0	0.6	22.4
恐怖	29.1	69.8	0	0	0.6	0.6
嫌悪	8.1	72.1	0	0	0	19.8

深層学習に関しても驚きの割合が全ての動画で高くなった.他の実験結果と比べると嫌悪の割合が高くなる動画が多い結果となった.

７．考察

（１）写真条件、単語条件、主観評定実験

写真条件、単語条件と主観評定実験の結果から比較を行う. 単語条件のスコア結果から感情表現のイメージ図を記録し西洋人と日本人の感情表出について考察する.

幸福についての表情比較を図 4 に示す.

図 4　幸福イメージ図(左:西洋人、右:日本人)

西洋人に関しては口をはっきりと開けるが日本人に関しては口角だけを上げる被験者が多かった. このことから意図的感情表現に関しては、明らかに幸福を表現できる日本人は少ないと考えられる. ただし、口角を上げるだけも幸福の表現であるため、今後は感情表現の強度も考慮する必要がある.

驚きについての表情比較を図 5 に示す.

図 5　驚きイメージ図（左：西洋人、右：日本人）

驚きに関しては全体的に日本人の方が変化点スコアは低いものの口を開ける動きは多くの被験者が実践しており、基本 6 感情の中では最も表情表出ができる感情だと考えられる.

悲しみについての表情比較を図 6 に示す.

図 6　悲しみイメージ図（左：西洋人、右：日本人）

悲しみは口角が少し下がり、目も開き具合が少し狭くなる傾向であったが眉の変化は少なく真顔に近い結果となった. しかし、写真条件において鼻や顎の変化点スコアが大きかったため細かい部位というよりも全体に動くことで悲しみを表出している可能性もあると考えられる.

怒りについての表情比較を図 7 に示す.

図 7　怒りイメージ図（左：西洋人、右：日本人）

日本人に関しては目を少し細める程度で他のパーツで大きな変化はなかった. すなわち真顔に近い状態となり、怒りについて日本人は表現しにくいと考えられる.

恐怖についての表情比較を図 8 に示す.

図 8　恐怖イメージ図（左：西洋人、右：日本人）

恐怖はすべての変化点スコアが低い結果となり、真顔とほぼ差異がなかった. すなわち、恐怖は基本 6 感情の中で日本人が最も表現できない感情であると考えられる.

嫌悪についての表情比較を図 9 に示す.

図 9　嫌悪イメージ図（左：西洋人、右：日本人）

嫌悪は眉の動きが怒りより高かったが、他のパーツは変化点スコアが似ており怒りと嫌悪は表情においては同じものと認識している日本人が多いと考えられる.

（２）刺激動画付加条件

全ての手法で共通していることは驚きを多く検出しているという点である. すべての動画に関して驚きの感情表現があったとは考えにくく驚きと誤認識した可能性が高い. これは特に変化点検出法と深層学習法に当てはまるが驚き以外は口を閉じているという定義をしている. すなわち口を開けている表情は全て驚きと認識する可能性が高い. 単語条件において、驚き以外は口を開ける動作をする被験者は少なかったが、刺激動画付加条件に関しては驚き以外の感情でも口を開ける動作がある可能性があると考えられる. Py-Feat による感情推定は怒りを検出するケースが多かった. しかし、幸福の動画で怒りを検出するのは誤認識の可能性が高い. Py-Feat に関しては日本人を対象とした感情認識ではないため日本人の真顔を怒りだと認識したのではないかと考えられる. 西洋人と日本人では真顔でも顔の構造が違うためその差も考慮することが重要である.

８．今後の展望

本研究では、刺激動画を付加した場合の表情の撮影を行ったが、被験者からカメラや近くに人がいるため緊張して自然に見るのが難しいという意見も上がった. よって、動画を見せて反応を見るより普段の日常生活や会話をカメラが存在すると意識させないで撮影することができればよりよい分析が行えると考えられる. また、表情のみではなく声などのマルチモーダルな情報を含めて感情推定を行うことが今後の課題である[10].

９．結論

本研究では、日本人における表情と感情の相関性について連続した 3 次元特徴点情報から分析を行った. 結果、驚き以外は西洋人のサンプルデータとは異なる点が多いことが分かった. また、刺激感情を付加した動画を被験者に見てもらい自然的な表情を撮影したが、こちらは驚きを検出する割合が高かった. また、従来手法では怒りを検出する割合が高い形となった.

謝辞：本研究の一部には、科学研究費補助金（基盤(C)19K12188)の助成を得た.

参考文献

1) P.Ekman:Facial action coding system (FACS),A human face, 2002.

2) W. Sato, S. Hyniewska, K. Minemoto, and S. Yoshikawa: Facial Expressions of Basic Emotions in Japanese Laypeople, Frontiers in Psychol, vol. 10, pp.259, 2019.

3) 斉藤功樹,中川靖士:日本人における基本感情と表情の関係分析,日本認知科学会,vol38,pp2-25,2021

4) Grishchenko,Ivan,et al:Attention Mesh:High-fidelity Face Mesh Prediction in Real-time.,arXiv preprint arXiv:2006,10692,2020

5) H.Hotelling:Analysis of a complex of statistical variables into principal components,Journal of Educational Psychology,Vol.24,pp.417-441,1933

6) 山西健司:データマイニングによる異常検知,共立出版株式会社,2016

7) 平山孝輔,佐伯幸郎,中村匡秀: 変化点検知を活用した顔特徴量時系列データからの注目点の抽出,電子情報通信学会,HCS2020-78,2021

8) Jin Hyun Cheong,Eshin Jolly,Tiankang Xie,Sophie Byrne, Matthew Kenney, Luke J.Chang:Py-Feat:Python Facial Expressions Analyze Toolbox

9) Sepp Hochreiter, Jiirgen Schmidhuber: LONG SHORT-TERM MEMORY,Neural computation 9,1997

10) K. Crawford:Time to regulate AI that interprets human emotions,Nature,vol.592,no.7853,pp.167,2021