



# WEB カメラを用いた VR ゴーグル着用状態での表情認識

Facial expression recognition while wearing a VR goggle using a web camera

吉田北斗<sup>1)</sup>, 影山芳之<sup>2)</sup>

Hokuto YOSHIDA, Yoshiyuki KAGEYAMA

1) 東海大学大学院 医用生体工学専攻 (〒259-1193 神奈川県伊勢原市下糟屋 143, [2ceym006@cc.u-tokai.ac.jp](mailto:2ceym006@cc.u-tokai.ac.jp))

2) 東海大学 (〒259-1193 神奈川県伊勢原市下糟屋 143, [kageyama@tsc.u-tokai.ac.jp](mailto:kageyama@tsc.u-tokai.ac.jp))

**概要:** 仮想空間において、利用者のアバターが利用者と同じ感情を表出できるようになれば、より現実に近いものになると考えられる。そのため本研究では、VR ゴーグルを着用した状態でも web カメラで計測が行える口を用いて、感情を表出する際の口の開閉、口角の移動の時系列データの特徴を 1 DCNN によって解析し、表情認識の手法として有用であるかを検討した。結果から特定の感情に関しては口の結果のみから表情認識ができる可能性が示唆された。

**キーワード:** コミュニケーション, 表情

## 1. はじめに

近年、メタバースの発展により仮想空間でのコミュニケーションが盛んに行われるようになってきた。しかし、仮想空間内で使われるアバターは無機質なものが多く実際のコミュニケーションとは差がある。表情認識の研究が数多く行われている中 [1] [2], メタバース空間内でのコミュニケーションで表情認識が扱われない背景としては、表情認識の研究の多くは顔全体を扱い表情を分類するが、顔の半分が隠れてしまう HMD を装着した表情認識が困難であることが考えられる。

そこで、HMD を装着した状態でも計測が行える口を用いて表情認識を行い、認識した実際の人間と同じ表情をアバターで表出できるようにすることができればより現実に近いコミュニケーションが可能になるのではないかと考え、本研究ではラッセルの円環モデルをもとに、喜、怒、哀、楽、無表情の 5 つの表情の時系列データの特徴を解析し分類することを目的とする。

## 2. 1DCNN

時系列データの解析方法として 1 次元の畳み込みニューラルネットワーク (1DCNN) を用いることとした。1DCNN は 2DCNN に比べ、数値を使うため処理が軽く、リアルタイム計測向きである。本研究では python のライブラリである pytorch を使用しプログラムを作成した。CNN の設定は学習率を 0.001, パッチサイズを 6 とした。また、CNN の層は入力層, カーネルサイズ 3 での畳み込み層×2, Max プーリング層, カーネルサイズ 3 での畳み込み層×2, 全結合層, 出力層を通すことで分類を行う。



図 1 ランドマーク適応

## 3. データセット作成方法

### 3.1 顔検出

python で、顔の部位を 68 個のランドマークを用いて、リアルタイムに検出するプログラムの作成を行った。ランドマーク適応の例を図 1 に示す。ランドマークの位置は測定している顔とカメラとの距離で変化してしまうため、一度顔認識を行い、トリミングした状態で測定することで距離を一定にして扱った。計測は Microsoft 社の web カメラ (LifeCamStudioQ2F-00021) を使い、フレームレートは 20fps で測定を行った。計測部位は先行研究 [3] を参考に口の右端や中心部など距離 6 が所, 開閉度 2 が所の計 8 か所を測定した。

### 3.2 計算方法 (開閉度)

喜と楽の差や怒と哀の差の測定に必要な口の開閉度の変化は先行研究 [3] [4] をもとに式 (1) を用いて部位の動きを測定した。式 (1) を示す。

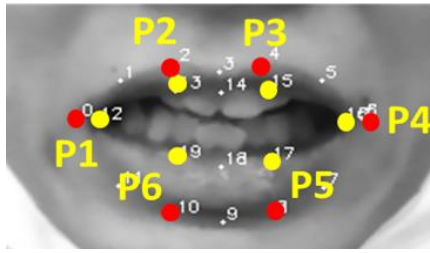


図2 計算用計測点

$$\text{測定値 (開閉度)} = \frac{|p_2 - p_6| + |p_3 - p_5|}{2|p_1 - p_4|} \quad (1)$$

式 (1) に使用した点を図 2 に示す赤い点が唇の外側、黄色い点が唇の内側の測定に使われた点を示す。開閉度は 1 に近づくほど口を大きく開き、0 に近づくほど口を閉じていることを示す。式 (1) を用いて唇の内側、計測点を変更し同様に唇の外側の開閉度の計 2 か所を計測する。

### 3.3 計算方法 (距離)

表情変化の計測に必要な測定部位の変化は先行研究 [3] [4] をもとに表情変化によって位置が変化しない鼻との距離を利用する。鼻にある変化のない点 ( $q_1, q_2$ ) の値を固定し、6 か所の測定部位 ( $p_x, p_y$ ) との距離をユークリッド距離で求め測定した。測定部位 ( $p_1, p_2$ ) を例に式を以下に示す。

$$\text{測定値 (距離)} = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} \quad (2)$$

距離が大きくなるほど支点である鼻との距離が遠ざかり、小さくなるほど鼻との距離が近づいていることを示す。式 (2) を用いて表情変化の測定に必要な唇の右端、唇の左端、上唇の midpoint (上側) 上唇の midpoint (下側)、下唇の midpoint (上側)、下唇の midpoint (下側) の計 6 か所を計測する。

## 4. 実験方法

### 4.1 実験方法

20 歳から 24 歳の健常な被験者 6 名 (男性 4 名, 女性 2 名) に協力してもらい 3.1 によって web カメラから読み取った口の動きを python のプログラムを扱い変化を測定した。被験者には実際に人がいる想定で、その人物に各感情 (喜, 怒, 哀, 楽, 無表情) が伝わるように表情を表出するよう指示し、画面に表示されるテキストに合わせて 10 秒間無表情, 5 秒間指定の表情を表出という 15 秒のセットを 8 回計 2 分間を各表情を表出させ、テキストの表示をトリガーとしその際の口の動きを測定した。

### 4.2 解析方法

今回の実験の解析方法としては、1DCNN で分類するため、作成したプログラムで epoch 数を 300 と 1000 で解析を行った。式 (1)、式 (2) によって得られる数値を 1 秒当たり 20 回の間隔で 15 秒間分 300 個保存し、8 か所分で 2400 個、1 次元の時系列データを 1 つのデータセットとして扱った。測定した 300 (個) × 8 (か所) × 48 (回分) の時系列データを入力とし、入力データの 80 パーセントを学習データ、20 パーセントをテストデータとして扱うため各

10 回解析を行い、正答率の平均を算出した。

## 5. 実験結果

### 5.1 結果 1 (時系列データ)

図 3, 4 に実際に計測した時系列データの例を示す。図 3 は式 (1) を用いた面積の結果で横軸が時間、縦軸が開閉度を表す。図 4 は式 (2) を用いた距離の結果で横軸が時間、縦軸が距離を表す。まず、図 3 に示した開閉度の変化を確認すると、どの表情も 10 秒からそれぞれ変化している様子が見られる。距離の変化を確認すると喜, 楽は 10 秒から数値が変化している様子が見られるが、怒, 哀は目視ではあまり大きな変化は見られなかった。次に、図 4 に示した開閉度の変化を確認すると、どの表情も 10 秒からそれぞれ変化している様子が見られる。距離の変化を確認すると喜, 楽は 10 秒から数値が変化している様子が確認できるが、怒, 哀は目視ではあまり大きな変化は見られなかった。

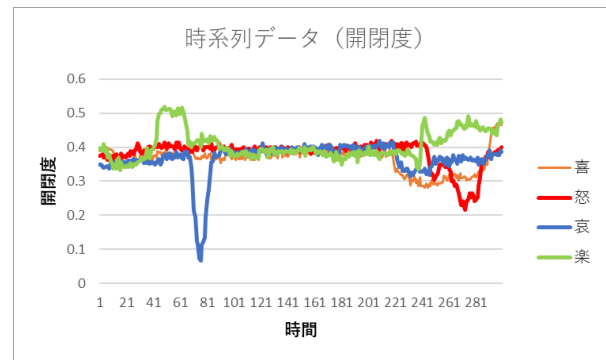


図 3 時系列データ (開閉度)

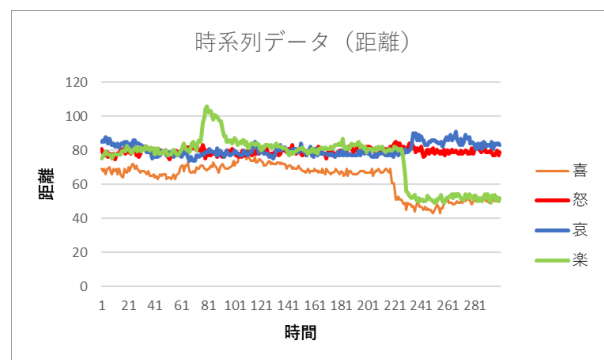


図 4 時系列データ (距離)

### 5.2 結果 2 (正答率)

CNN の分類による無表情, 喜, 怒, 哀, 楽の 5 値分類の正答率をそれぞれ epoch 数 300 の結果を表 1 に、epoch 数 1000 の結果を表 2 に示す。初めに、表 1 の平均正答率を確認するとすべて 20% を超えていることが示された。また、5 回目の解析ではどの表情も正答率が 27% 以上となった。次に、表 2 の平均正答率を確認するとすべて 20% を超えていることが示された。しかし、それぞれの解析結果の中で最低値が一番高い 3 回目の結果を確認すると怒以外は 20% 以上だが、怒が 11% と 20% を下回っている。epoch 数 300 と 1000 を比較すると 1000 のほうが快の表情 (喜, 楽) の正答率が悪くなっている一方、不快の表情 (怒, 哀)

は正答率が高くなっている。

表 1 解析結果 (epoch 数 300)

回数	epoch数300正答率 (%)				
	無表情	喜	怒	哀	楽
1	8	66	50	66	11
2	10	75	0	0	20
3	0	28	75	0	66
4	0	33	50	0	50
5	27	75	40	30	50
6	0	45	0	16	80
7	30	83	0	0	66
8	12	60	30	58	0
9	58	81	0	11	33
10	71	54	12	25	50
平均	21.6	60	25.7	20.6	42.6

表 2 解析結果 (epoch 数 1000)

回数	epoch数1000正答率 (%)				
	無表情	喜	怒	哀	楽
1	8	60	81	10	60
2	20	72	11	21	0
3	20	23	11	88	42
4	36	26	0	40	42
5	83	40	0	25	33
6	87	25	0	16	40
7	0	0	91	0	50
8	0	91	44	16	0
9	25	50	44	0	0
10	60	20	0	22	66
平均	33.9	40.7	28.2	23.8	33.3

5.3 結果 3 (学習の精度)

図 3, 4 に学習過程の損失のグラフを図 5, 6 に精度のグラフを示す。初めに epoch 数が 300 の時の結果だが, 図 3 に示すように epoch 数が増えると損失が 0 に近づき, 150 を超えたところからほとんど 0 になっていることが示された。また, 図 4 に示す epoch 数 300 の精度は 150 を超えたあたりからテストデータは学習データの 50%ほどの精度で安定していることが示された。次に epoch 数が 1000 の時の結果だが, 図 5 に示すように epoch 数が 120 を超えたあたりで 1 度 0 に近づき, 300, 600, 800 のあたりで増加しているが, 最終的には 0 に近づいていることが示された。次に図 6 に示すように精度は epoch 数 120 のところでテストデータが学習データの 40%程度の精度に安定したのちに 300, 600, 800 の点で精度が下がっていったことが示された。

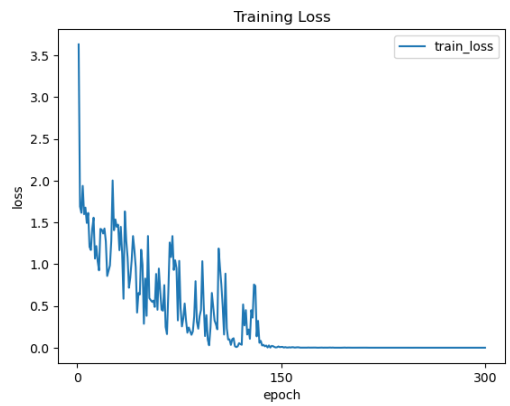


図 5 学習過程の損失 (epoch 数 300)

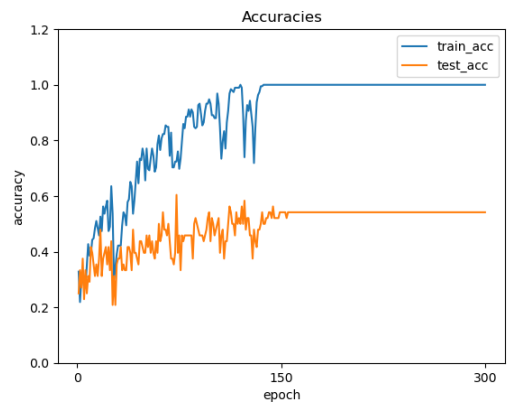


図 6 精度 (epoch 数 300)

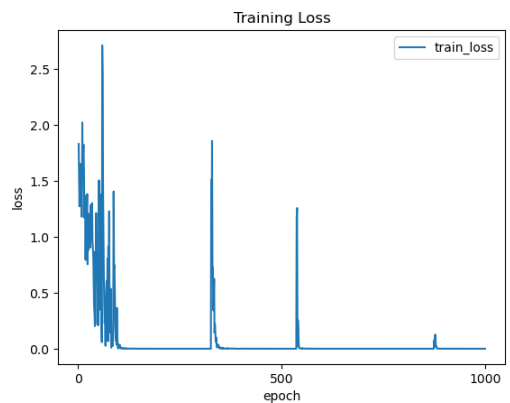


図 7 学習過程の損失 (epoch 数 1000)

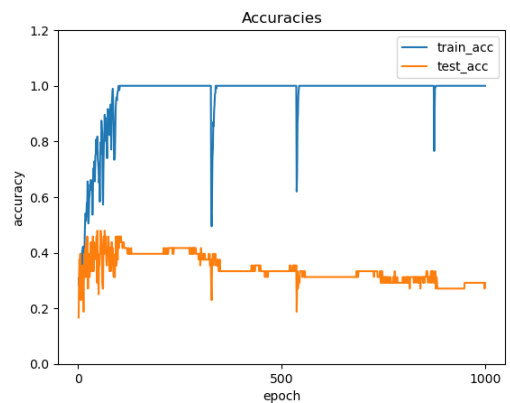


図 8 精度 (epoch 数 1000)

## 6. 考察

今回は口の変化の時系列データから表情分類を行ったが、各解析ごとに正答率が大きく変化し、安定しない結果となった。正答率が大きく上下した理由としては表情という個人差の大きいものの分類を多人数のデータを使い平均化しようとしたが平均化するには人数が足りていなかったこと、他クラス分類をするにはデータセットが少なかったことなどが考えられる。また、解析結果を個別にみていくと、正答率が0%になっている結果があることや、図4のように学習を重ねるごとに正答率が下がっていることから過学習が起きていることが考えられる。さらに、表情ごとの結果を比較すると快の表情（喜、楽）は比較的正答率が高く、不快の表情は比較的正答率が低くなっている。

計測終了後、被験者の一部から不快の表情（怒、哀）の表情の表出が難しかったという意見が出たことから不快の表情（怒、哀）は故意に表出することが難しく、特徴が抽出できなかった可能性も考えられる。実際の時系列データも開閉度ではそれぞれ変化している様子が見られたが、距離は不快の表情では変化が見られなかったことから計測部位も再考の余地がある。

一方で学習の過程で損失が0に近づいていることから各データに何らかの傾向は存在し、学習が行えていると考えられる。また、正答率は高いとは言えないもののすべての表情で平均正答率が5値分類の確率である20%を超えていることから測定部位の変更、データセットの追加で正答率を向上させることができればIDCNNを用いて口の変化の時系列データを解析することで喜怒哀楽の表情分類が行える可能性が示唆された。

## 7. 今後の展望

本研究では表情表出の際の口の動きの時系列データを扱いIDCNNにて5値の表情分類を試みた。結果としては口の動きの時系列データから表情分類が行える可能性は示唆されたが分類器として扱うには正答率が低い。また、測定部位によっては特定の表情で変化が見られていないことから測定部位が今回の5値分類に最適かを再度確認する必要がある。

そのため、今後は被験者の追加によるデータセットの追加、複数人を扱わず個人を扱うことでの個人差の排除、測定部位や分類数の変更などの方法も検討し、正答率の向上を目指す。

### 参考文献

- [1] 後藤みの理, 加納政芳, 加藤昇平, 國立勉, 伊藤英則: 完成ロボットのための感情領域を用いた表情生成, 人工知能学会論文誌, pp. 55-62, 2006.
- [2] 藤村友美, 鈴木直人: 表情の表出過程及び形態学的変化が感情認識に及ぼす影響, 認知心理学研究, 第5巻第1号, pp.53-61, 2007.
- [3] Tereza Soukupova, Jan Cech: Real-Time Eye Blink Detection using Facial Landmarks, Computer Vision Winter Workshop 21<sup>st</sup>, 2016.
- [4] 羽田健一, 中村納: 表情認識のための顔部品からの特徴量の抽出と個人別に認識ルールの生成法に関する検討, 画像電子学会誌, 第34巻第3号, pp.210-219, 2005.
- [5] 池田悠平, 岡田佳子, 堀江亮太, 菅谷みどり: 表情と生体情報を用いた感情の推測方法の検討, マルチメディア, 分散, 協調とモバイルシンポジウム, pp. 149-161, 2016.