

顔面動作符号化システム (FACS) を用いる長・短期記憶に基づく異常な顔表情の検知 Detecting Anomalous Facial Expressions Based on Long Short-Term Memory Using Facial Action Coding System

大西 悠貴[†] 松川 徹[‡] 鈴木 英之進[†]
Yuki Onishi Tetsu Matsukawa Einoshin Suzuki

1. 動機

異常検知は、期待されるふるまいと一致しないパターンを識別する問題であり、中でも人の顔画像に異常検知を適用することで心身の異常を検知する研究が近年盛んにおこなわれている。人の顔表情に対する異常検知では、眠気を感じている表情など、時系列な顔画像系列を用いなければ異常かどうか判別が難しい表情が存在する。この問題に対し、時系列的な観点から異常検知をするため、長・短期記憶 (LSTM: Long Short-Term Memory) [1] を用いる手法が存在する [2]。LSTMは時系列なデータの長期依存や短期依存を学習できるネットワークである。

[2] などの顔画像を入力とする研究の多くは、畳み込みニューラルネットワーク (CNN: Convolutional Neural Network) [3] によって特徴を抽出しているが、個人間の顔構造の差異や顔の向き、輝度の変化など、表情とは無関係の情報から影響を受けやすいという欠点がある。これに対し本論文では、顔面動作符号化システム (FACS) [4] を用いて入力顔画像から特徴を抽出し、LSTM に与えることで、時系列的な表情の特徴空間を作成し、異常な顔表情の検知を行う。FACS は表情筋の動作を視覚的にとらえ記述可能にするであり、個人間の顔構造の差異に影響されにくく、より普遍的な特徴抽出が可能になる。

2. 異常顔表情区間の検知

提案手法は、図 1 に示すように、画像特徴量の取得、時系列特徴量の取得、異常検知の手順で構成される。入力された時系列顔画像データ D に対し、一定区間ごとに表情が異常であるかどうか判定し出力する。

画像特徴量の取得では、FACS に基づき画像ごとの特徴量を取得する。FACS は表情の情報を損なわず、構造的な

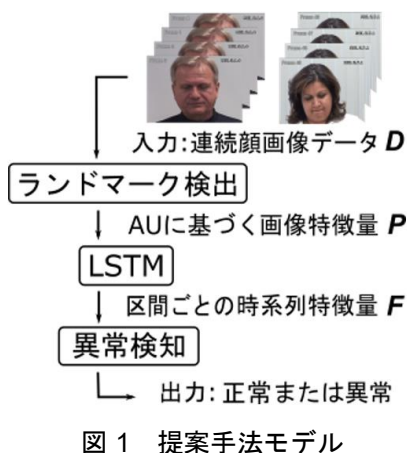


図 1 提案手法モデル

表 1 実験に用いる AU

No.	顔の動作	No.	顔の動作
AU 1	眉の内側を上げる	AU 12	唇両端を上げる
AU 2	眉の外側を上げる	AU 15	唇両端を下げる
AU 4	眉を引き寄せる	AU 17	おとがいを上げる
AU 5	上瞼を上げる	AU 20	唇両端を横に引く
AU 6	頬を持ち上げる	AU 23	唇をすぼめる
AU 7	瞼を緊張させる	AU 26	顎を下げ唇を開く
AU 9	鼻に皺を寄せる	AU 43	瞼を閉じる
AU 10	上唇を上げる		

記述が容易な表情パラメータを取得できる、P. Ekman と W. Friesen による表情理論である。FACS では表情筋の動きを AU: Action Unit と呼ばれる 44 個の動作単位に分解しており、それぞれの AU は 1 つ以上の筋肉に対応している。本研究では、眉を引き寄せる動作や頬を持ち上げる動作など、表 1 に示す主要な 15 個の AU を用いる。各顔画像から顔のランドマークを検出し、顔の向きの影響を除くため、図 2 の左図に示す左右の目尻間の距離 C_w で x 座標を、目尻間の中点と鼻棘の距離 C_h で y 座標を除算し、縦軸と横軸を正規化する。図 2 の右図に示すランドマーク間の距離 Q に関して、それぞれの AU に関するデータセット内の最大値、最小値を Q_{\max} , Q_{\min} としたとき、(1) 式の minmax-scale 関数で正規化し、AU 特徴量を取得する。

$$\text{minmax-scale}(Q) = \frac{Q - Q_{\min}}{Q_{\max} - Q_{\min}} \quad (1)$$

時系列特徴量の取得では、基本 6 表情 (笑顔, 怒り, 嫌悪, 驚き, 悲しみ, 恐怖) でラベル付けされている時系列顔画像データで訓練した LSTM ネットワークモデルを用いて、AU 特徴量から区間データごとの時系列特徴量 F を取得する。訓練の段階では LSTM の出力を 6 ユニットの全結合層に接続し、時系列顔画像データのラベルを教師信号として訓練を行う。特徴抽出の段階では全結合層を除いた LSTM の出力を時系列特徴量として用いる。LSTM の入力

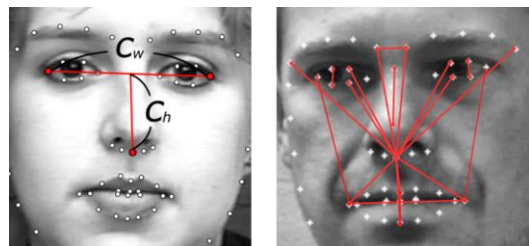


図 2 使用するランドマーク間距離と正規化に用いる距離 C_w , C_h

[†]九州大学大学院 システム情報科学府情報学専攻

[‡]九州大学大学院 システム情報科学研究院情報学部門

に対する活性化関数には \tanh 関数を、再帰計算時に用いる活性化関数には (2) 式に示す hard-sigmoid 関数を用いる。LSTM のユニット数が出力の次元数となるので、ユニット数を u とすると、得られる特徴量 F の次元数も u となる。

$$\text{hard-sigmoid}(x) = \begin{cases} 1 & (2.5 < x) \\ 0.2x + 0.5 & (-2.5 \leq x \leq 2.5) \\ 0 & (x < -2.5) \end{cases} \quad (2)$$

異常検知では作成した各区間データの特徴量 F から、異常なデータ区間を検知する。データに含まれる正常例や異常例は単一の表情ではなく、多数のクラスが生成されるため、異常検知手法には LOF: Local Outlier Factor [5] を用いる。 k 個の近傍点を用いて LOF により異常検知を行う場合を考える。 A の k 個の近傍点集合を $N_k(A)$ とし、 A と B の距離を $d(A, B)$ 、 A と A の k 番目の近傍点の距離を k -distance(A)、 $\max(x, y)$ を x と y の小さくない方と定義する。 A と B の到達可能性距離 $\text{reach-dist}_k(A, B)$ と A の局所到達可能性距離 $\text{lrd}_k(A)$ は (3)、(4) 式と定義する。

$$\text{reach-dist}_k(A, B) = \max(d(A, B), k\text{-distance}(B)) \quad (3)$$

$$\text{lrd}_k(A) = \left(\frac{\sum_{C \in N_k(A)} \text{reach-dist}_k(A, C)}{k} \right)^{-1} \quad (4)$$

このとき、(3) 式と (4) 式を用いると、例 A の異常度を表すスコア $\text{lof}_k(A)$ は (5) 式と定義される。閾値を θ とすると、 $\text{lof}_k(A) > \theta$ のときに限り A は異常であると判断される。

$$\text{lof}_k(A) = \frac{1}{k} \sum_{C \in N_k(A)} \frac{\text{lrd}_k(C)}{\text{lrd}_k(A)} \quad (5)$$

3. 実験

実験には、UNBC-MSPEA [6]、CASME2 [7]、会話中の人物の表情を撮影し作成した会話データの3つのデータセットを用いる。データセットに含まれる異常例を図3に、正常例と異常例の数を表2に示す。UNBC-MSPEA は肩を罹患している患者を対象とし、腕を動かすことのできる範囲を測るテストの様子を撮影した顔画像系列から構成される。実験では50フレームごとに分割し、各フレーム画像にラベル付けされた痛みの程度をもとに、強い苦痛の表情が現れ

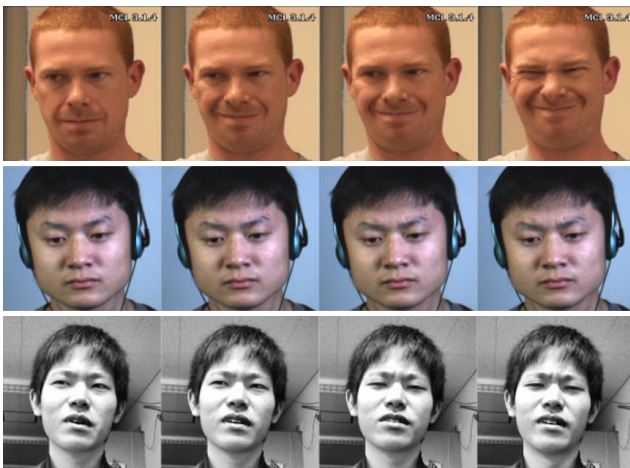


図3 データセットに含まれる異常例

表2 実験に用いたデータの例数

	UNBC-MSPEA	CASME2	会話データ
正常例	877	196	438
異常例	122	40	88

表3 各手法のAUC

	UNBC-MSPEA	CASME2	会話データ
提案手法	0.569	0.520	0.773
比較手法	0.553	0.462	0.734

ているデータ区間を異常とする。CASME2 は、抑圧された感情によって顔のわずかな変化が短時間に表出する微表情のデータセットであり、表情が発生する前から消失するまでの様子を撮影した画像系列で構成される。1つのサンプルデータが1つの表情を含むため分割は行わない。それぞれの画像系列には happiness, disgust, repression, surprise と、これらに分類されない様々な表情のクラス others がラベル付けされており、データセットの中で少数しか現れない表情を含む others クラスのデータを異常とする。会話データは会話中の人物の様子を正面から撮影したデータで構成されており、15フレームごとにデータを分割し、目を閉じたままの区間や強い表情が現れている区間など、少数しか現れていない表情のデータ区間を異常とする。

比較手法として AU を取得する代わりに CNN で画像の特徴量を抽出する手法を用いる。実験には、基本6表情に無表情を加えた、7クラス表情分類データで訓練済みの公開 CNN モデル (https://github.com/oarriaga/face_classification) を使用する。結果の評価には AUC を用いる。

提案手法と比較手法の AUC を表3に示す。それぞれのデータセットにおける AUC は UNBC-MSPEA で 0.016、CASME2 で 0.058、人工データで 0.039、提案手法が比較手法を上回っており、FACS に基づく特徴抽出が異常な顔表情の検知において一定の有効性を持つことを示している。詳しい分析は [8] を参照されたい。

参考文献

- [1] F. A. Gers, N. N. Schraudolph, J. Schmidhuber. Learning Precise Timing with LSTM Recurrent Networks. *Journal of Machine Learning Research*, 3:115-143, 2002.
- [2] J.-M. Guo and H. Markoni. Driver Drowsiness Detection Using Hybrid Convolutional Neural Network, Long Short-Term Memory. *Multimedia Tools and Applications*, 78(20):29059-29087, 2019.
- [3] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE*, 86(11): 2278-2324, 1998.
- [4] P. Ekman, W. V. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, CA, 1978.
- [5] M. M. Breunig, H.-P. Kriegel, R. T. Ng, Jorg Sander. LOF: Identifying Density-Based Local Outliers. In *Proc. ACM SIGMOD International Conference on Management of Data*, pages 93-104, 2000.
- [6] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, I. Matthews. Painful Data: The Unbc-Mcmaster Shoulder Pain Expression Archive Database. In *Proc. FG*, pages 57-64, 2011.
- [7] W.-J. Yan, Q. Wu, Y.-J. Liu, S.-J. Wang, X. Fu. CASME Database: A Dataset of Spontaneous Micro-Expressions Collected from Neutralized Faces. In *Proc. FG*, pages 1-7, 2013.
- [8] 大西悠貴: 顔面動作符号化システム(FACS)を用いる長・短期記憶に基づく異常な顔表情の検知, 九州大学工学部電気情報工学科卒業論文, 2020.