

リハビリテーション効果推定のための感情識別器の構成と評価

Evaluation of Emotion Detector for Estimating the Effect of Rehabilitation

西田健次^{1*} 山田亨² 藤村友美² 糸山克寿¹ 中臺一博^{1,3}
Kenji Nishida¹, Toru Yamada², Tomomi Fujimura², Katsutoshi Itoyama¹, Kazuhiro Nakadai^{1,3}

¹ 東京工業大学

¹ Tokyo Institute of Technology

² 国立研究開発法人産業技術総合研究所

² National Institute of Advanced Industrial Science and Technology

³ ホンダ・リサーチ・インスティテュート・ジャパン

³ Honda Research Institute Japan

Abstract: 笑顔度検出を、脳卒中後遺症における音楽療法や認知症の進行抑制のための認知活性化療法での介入効果の推定手法に取り入れる提案がなされてきており、その有効性が確認されつつある脳機能障害患者の表情変化は健常者の表情変化よりも乏しいことが多く汎用の笑顔度識別器では「笑顔」と判定されないことが多いため、個人内での表情変化を捉える識別器を構成しなくてはならない。更に介入の効果判定のために個人内での表情変化を正規化し、個人間の比較を行う手法なども提案されてきている。本稿では、笑顔度検出と同様の手法で笑顔（喜び）以外の感情（怒、嫌悪、恐れ、悲しみ、驚き）の識別器を構成し、その有効性の統計的検証を行った。その結果、嫌悪と悲しみの判別など数例に関しては有意な結果が得られたが、笑顔の検出性能が特異的に高いことが示された。

1 はじめに

2018年（平成30年）10月1日現在、我が国の総人口における65歳以上の割合は28.1%に達し、更に75歳以上人口が12.4%を越えるなど本格的な高齢化社会を迎えている[1]。このような高齢化への社会構造の変化にともないアルツハイマー患者数は年々増加し、また患者数は漸減しているものの脳卒中の患者数も100万人を越えており[2]、脳機能障害に対するリハビリテーションの重要性は年々増加している。脳卒中後遺症のリハビリテーションの一つに音楽療法が挙げられる。音楽療法は、「音楽の持つ生理的、心理的、社会的働きを用いて、心身の障害の回復、機能の維持改善、生活の質の向上、行動の変容などに向けて、音楽を意図的、計画的に使用すること」と定義されている[3]。本療法においては従来、患者に対する療法の効果を、病院や音楽療法士が独自に設けた評価基準と介入内容の記録などを通じて質的・量的に評価することが試みられてきたため、客観的で統一的な評価方法を確立することは困難であった。そこで、表情の変化（笑顔度）を検出することによって音楽療法効果の客観的評価手法が提

案され、その有効性が示されている[4]。また、認知症患者に対する心理療法の一つである回想法においても、従来より心理療法士の観察によって効果の評価が行われてきたが[5]、客観的な評価手法の確立が求められており、笑顔度による介入効果の評価への期待が持たれている。

笑顔という表情は「快」あるいは「幸福」の感情を示すものとする、笑顔度は患者に対する療法の正の効果の測るものと考えることができる。逆に、「怒」、「嫌悪」、「悲しみ」、「恐怖」の感情を示す表情は、療法の負の効果の測るものであり、これらの表情を検出することは、療法の不適切な適用を避けるために有効な手段と考えることができる。そこで、本稿では、6種の感情類型、「怒り(anger)」、「嫌悪(disgust)」、「恐怖(fear)」、「幸福(happiness)」、「悲しみ(sadness)」、「驚き(surprise)」に対応する識別器を構成し、感情識別器としての有効性を評価した。表情の本稿の構成は、2章では脳機能障害患者の表情識別に関する課題を述べ、本稿で用いた表情識別器による感情推定手法を説明する。3章では本稿で用いたデータセットについて述べ、表情識別器の訓練手法を説明する。4章において表情識別器による感情推定結果に対し統計的検証によってそれぞれの感情推定結果の有効性を考察する。5

*連絡先：東京工業大学
152-8552 東京都目黒区大岡山 2-12-1 W8-18
E-mail: nishida@sc.e.titech.ac.jp

章でまとめを述べ、今後の課題についても言及する。

2 表情識別による感情推定手法

表情識別はコンピュータビジョン、および、人工知能技術においても重要な課題であり、これまでも多くの研究がなされてきた。そして、表情は人間の感情を推定する重要な手掛かりでもある。本稿では、脳機能障害患者の感情を推定するために、其々の感情を代表する表情を識別し、その推定値を感情の推定値として用いることを提案し、その妥当性を検証した。

2.1 表情識別に関する関連研究

表情識別は、コンピュータビジョンの分野において重要な課題の一つであり、多くの研究成果があげられてきた。画像による表情識別は特徴点ベースの手法 [6] とアピランス・ベースの手法 [7, 8, 9] の二つに大別される。特徴点ベースの手法は、顔検出後に顔器官（目頭、目尻、口角、鼻など）から特徴点を抽出し、それらの位置関係から表情を検出する。顔の検出位置のずれや顔の向きに対する補正が可能のため、表情の検出性能は高いと言える。その一方で、特徴点の検出性能、特徴点のアノテーションの精度により、表情の検出性能が影響を受けるなどの問題点がある。アピランスベースの手法は、顔検出後に顔器官を検出する必要がないため処置が単純で高速であるが、顔の位置ずれ、顔の向き、個人ごとの顔器官配置、そして、画像の撮影条件（照明、フォーカス）の影響を受けやすい。しかし、訓練サンプルに十分なバリエーションを持たせ、かつ、位置ずれや顔の向きの変化にロバストな特徴を用いることで、これらの問題点を解決する手法が提案されてきた [10]。さらに、大規模データセットと深層学習を組み合わせることで、数多くのアピランス・ベースの表情識別手法が提案されている [11, 12, 13]。また、真顔から笑顔に変化するシーケンスを学習サンプルとすることで、笑顔という表情を検出だけでなく、その強度（笑顔度）を推定する手法も提案されている [14]。

2.2 脳機能障害患者の表情識別

脳機能障害患者は表情が乏しくなることが多く [16, 17]、前節で述べたような汎用的な表情識別器では十分な検出性能を得ることが難しいことは容易に想像できる。しかし、一方で、乏しいながらも表情に変化があるならば、一般的には検出できないものであっても、その変化を捉えることによって、ある個人の表情を検出できる。さらに、無表情からある表情への変化が単調であると仮定するならば、変化の度合いをその表情の強

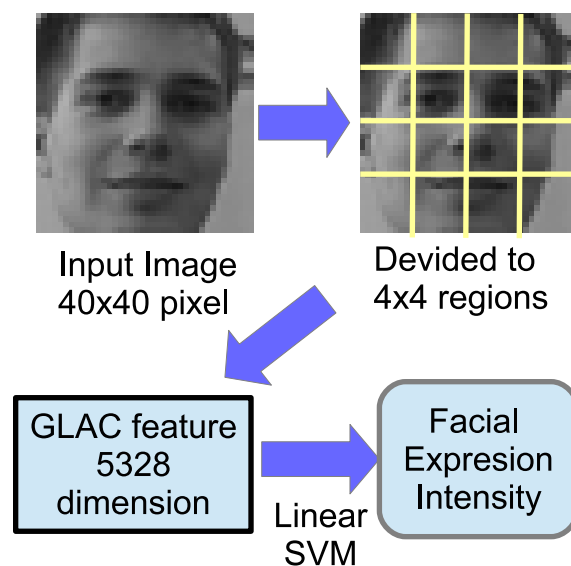


図 1: 表情識別器の構成

度と考えることができる。そして、単調な変化を線形近似すると考えると顔画像の表情の強度は、式 (1) で表すことができる。この表情の強度を、感情推定値として用いることとする。

$$y = \mathbf{w}^T \mathbf{x} - h \quad (1)$$

ここで、 y は表情の強度（スコア）、 \mathbf{x} は顔画像から抽出された特徴量、 \mathbf{w} は係数ベクトル、 h はバイアス値と示す。ある個人の表情の変化は y の変化によって示すことができるが、個人間での表情強度は直接比較することができないため、何らかの方法で正規化する必要がある。この正規化手法については、後述する。

本稿では、アピランス・ベースの表情識別手法を採用し、特徴量として位置ずれや照明条件に頑健な GLAC (Gradient Local Auto-Correlation)[15] 特徴を採用した。検出された顔画像は 40×40 のグレースケール画像に変換され、 4×4 、計 16 個の領域に分割される。各領域について 333 次元の GLAC 特徴が抽出され、顔画像 1 枚につき 5328 次元の特徴量が \mathbf{x} として抽出される (図 1)。係数ベクトル \mathbf{w} 、バイアス h は、特定の表情に対する 2 クラス線形サポート・ベクトル・マシン (SVM) を学習することによって得られる。

3 感情識別器の学習・評価用データセット

感情識別器の学習用のデータセットとして The Face-Grabber Database and Software [18] (図 2) を使用



図 2: FaceGrabber DB 顔画像の例

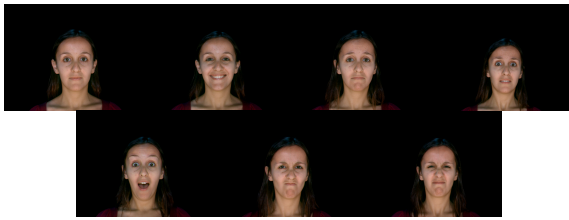


図 3: VISGRAF Faces DB 顔画像の例

した。FaceGrabber データベースは、40 人の怒り、嫌悪、恐怖、喜び（笑顔）、悲しみ、驚きの 6 つの表情とニュートラルとされる表情に分類されており、一つの表情あたり 30 枚（ニュートラルに関しては 90 枚）の計 10800 枚の顔画像が含まれている。左右反転画像まで含めた計 21600 枚の画像を、1 表情分 2400 枚とそれ以外の表情全て（ニュートラル含む）19200 枚の 2 クラスに分け、2 クラス識別器による表情識別器 6 種（怒り、嫌悪、恐怖、喜び、悲しみ、驚き）の訓練を行った。

感情識別器の評価には、AIST 顔表情データベース [19] と VISGRAF Faces DB [20]（図 3）を用いた。AIST 顔表情データベースは、日本人 8 人（女性 4 人、男性 4 人）の 12 種類（怒り（閉口）、怒り（開口）、嫌悪（閉口）、嫌悪（開口）、興奮、恐怖、喜び、ニュートラル、リラックス、悲しみ、眠気、驚き）の表情が含まれているものであり、VISGRAF Faces DB は、主に欧米人 36 人の 7 種（怒り、嫌悪、喜び、ニュートラル、悲しみ、驚き）の表情が含まれている。

4 実験結果

4.1 識別器スコアの正規化

図 4 に、AIST 表情 DB の f01（女性）の感情ごとの識別器スコアのプロットを示す。6 種の識別器は独立に訓練されているため、識別器スコアの値そのものには意味がないため、識別器間での正規化を行う必要がある。本稿では、最大値-最小値での正規化（式（2））を

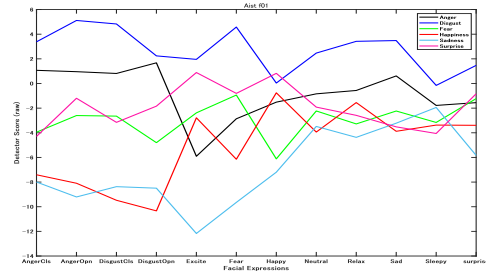


図 4: AIST 表情 DB: f01 の識別器スコア

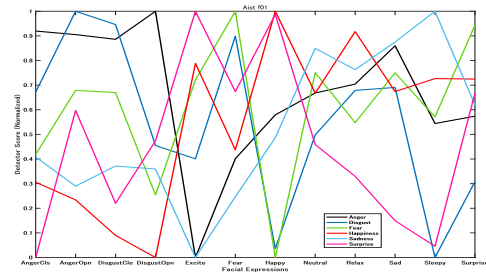


図 5: AIST 表情 DB: f01 正規化スコア

行った。これは、一人のサンプルでの感情識別器の全表情に対するスコアの最大値と最小値により、感情識別器のスコアを正規化するものである。図 5 は最大最小値での正規化の結果である。

$$\tilde{y} = \frac{y - \min(y_e)}{\max(y_e) - \min(y_e)} \quad (2)$$

\tilde{y} は識別器出力 y の正規化値、 y_e は表情 $\{e | e = \text{anger, disgust, fear, happy, neutral, sad, surprise}\}$ での識別器出力を示す。

また、個人間での感情強度の比較を行うためには、一人一人の感情のベースラインと振幅が異なるため、識別器のスコアそのままでは比較することができない。しかし、最大値最小値による正規化を行うことによって、各人のベースラインからの変位を求めることとなるため、個人間での比較も可能になると考えられる。

4.2 AIST 顔データベースに対する感情識別器の正規化スコア

AIST 顔表情データベースに対する識別器のスコアを図 6 から図 11 に示す。

図 9 に示される喜び識別器の結果は、一例を除き、怒り、嫌悪、恐怖、悲しみの負の表情に対するスコアが低く、喜び（笑顔）、リラックスなどに対するスコアが

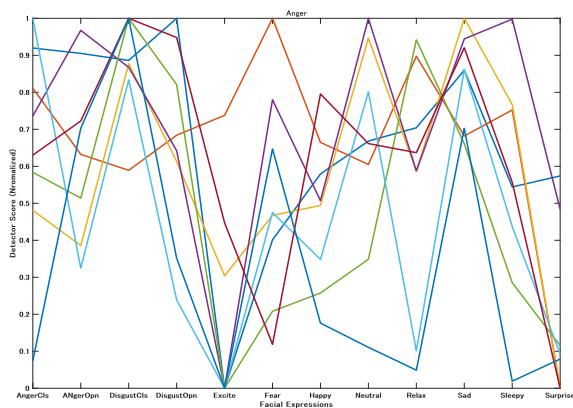


図 6: Anger 識別器の AIST-DB8 人に対する正規化スコア

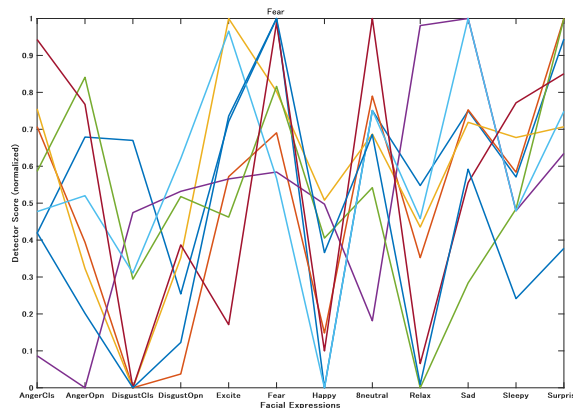


図 8: Fear 識別器の AIST-DB8 人に対する正規化スコア

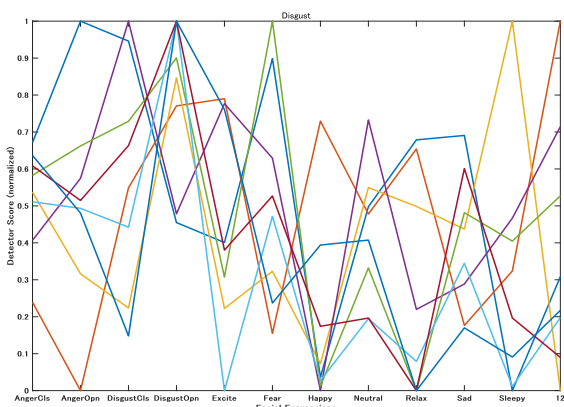


図 7: Disgust 識別器の AIST-DB8 人に対する正規化スコア

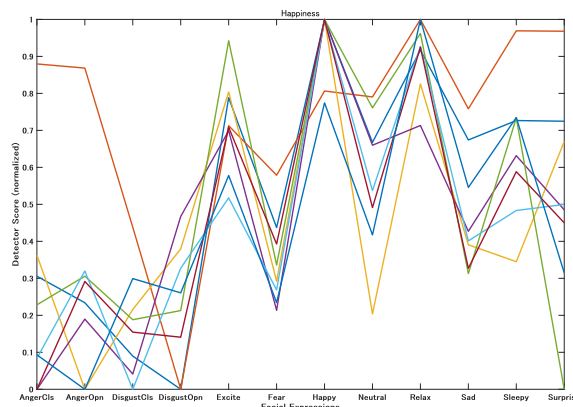


図 9: Happiness 識別器の AIST-DB8 人に対する正規化スコア

高い傾向がみられる。一方、図 6, 図 7, 図 8, 図 10 に示される怒り、嫌悪、恐怖、悲しみに対する識別器の出力は、喜びの表情に対するスコアが低いことは共通するが、対応する表情に対するスコアが必ずしも高いわけではない。

図 11 に示される驚き識別器の結果は、驚きという表情に喜び由来のものと恐怖由来のもの二種類があることを示唆している。

4.3 VSIGRAF DB に対する感情識別器の正規化スコア

VISGRAF-DB に対する識別器のスコアを図 12 から図 17 に示す。

喜び識別器（図 15）の特異性が際立つ結果となっている。一部に、怒り、嫌悪などにも反応している例もあるが、喜びの表情に対しては喜び識別器は高値を出力している。一方、AIST-DB と共通の傾向であるが、怒り、嫌悪、恐怖、悲しみに対する識別器は、喜び表情に対して低値を示すが、負の表情に対しては相互に高値を示す傾向にあり、その識別は難しいと考えられる。

驚き識別器（図 17）の結果も、AIST-DB に対する結果と共通し、喜び、恐怖の表情に対しても高値を示している例がある。

4.4 識別器スコアの統計的解析

前節で述べた識別器の傾向を確認するため、分散分析 (ANOVA)、および、Tukey-Kramer 法による多重比

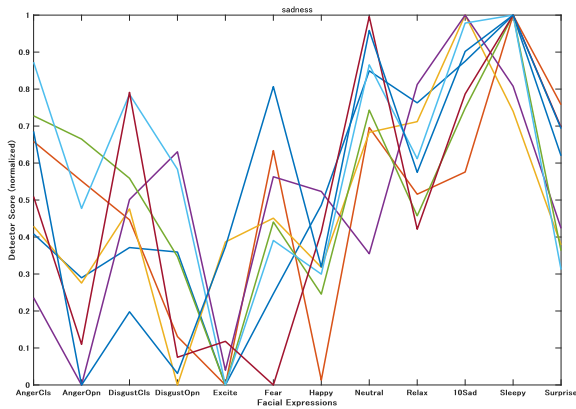


図 10: Sadness 識別器の AIST-DB8 人に対する正規化スコア

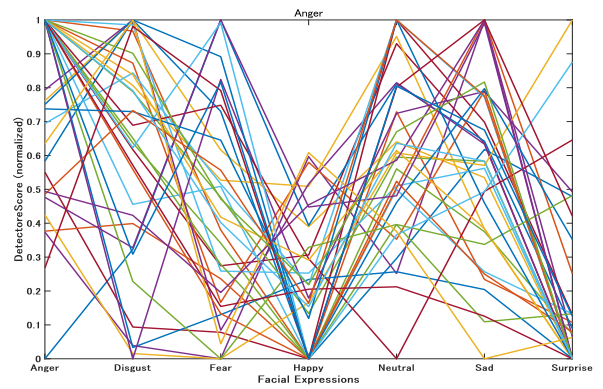


図 12: Anger 識別器の VISGRAF-DB36 人に対する正規化スコア

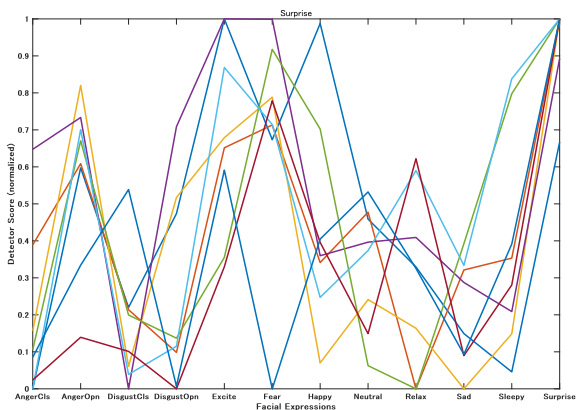


図 11: Surprise 識別器の AIST-DB8 人に対する正規化スコア

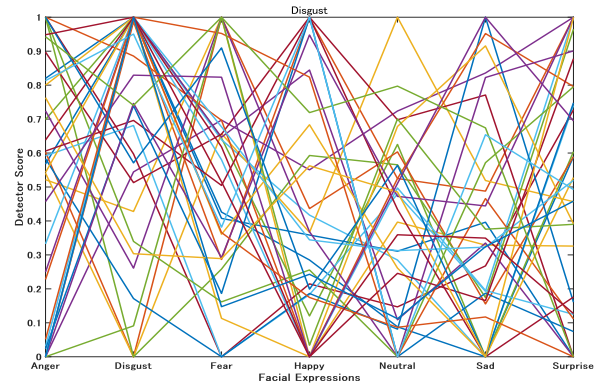


図 13: Disgust 識別器の VISGRAF-DB36 人に対する正規化スコア

較検定 (Post-Hoc) を行った。

表 1 に AIST-DB の結果に対する分析, 表 2 に VISGRAF-DB の結果に対する分析を示す. AIST-DB では, ANOVA において感情識別器の出力と顔表情 (12 種) 間には $p < 0.01$ で交互作用あり, 一連の顔表情に対して感情識別器が異なる出力を示すことが示唆された. これを受けて行った post-hoc 解析では, 恐怖識別器と悲しみ識別器間には $p < 0.01$ で, 怒り識別器と嫌悪識別器間, 怒り識別器と驚き識別器間には $p < 0.05$ で有意差があると判定された. 前節で述べた喜び識別器と他の感情識別器との差異が大きくなっていない. VISGRAF-DB では, ANOVA において感情識別器の結果と顔表情 (8 種) 間には $p < 0.01$ で交互作用があり, やはり一連の顔表情に対して感情識別器が異なる出力を示すことが示唆された. これを受けて行った post-hoc 解析では,

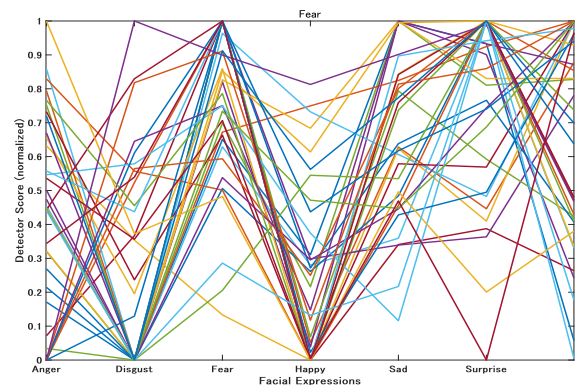


図 14: Fear 識別器の VISGRAF-DB36 人に対する正規化スコア

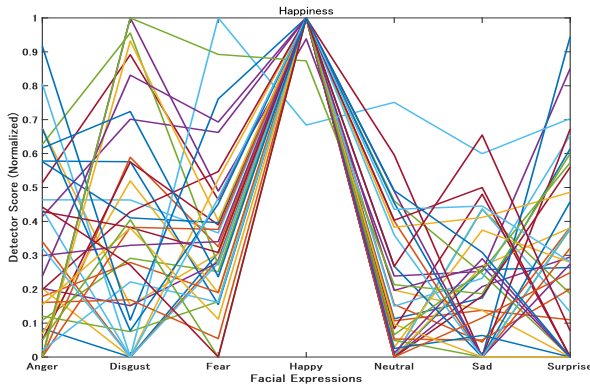


図 15: Happiness 識別器の VISGRAF-DB36 人に対する正規化スコア

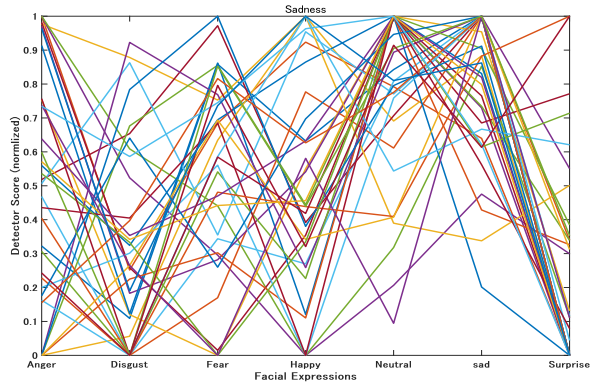


図 16: Sadness 識別器の VISGRAF-DB36 人に対する正規化スコア

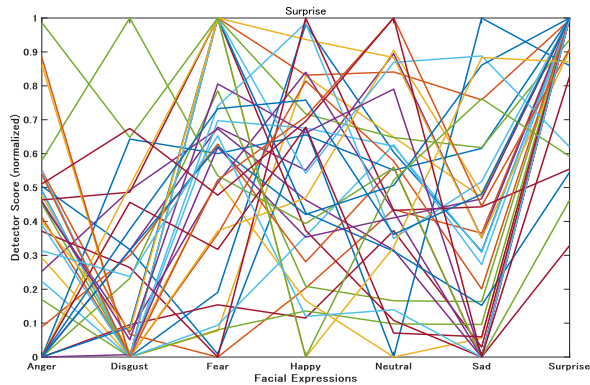


図 17: Surprise 識別器の VISGRAF-DB36 人に対する正規化スコア

表 1: AIST-DB での分散分析結果

ANOVA	Tukey-Kramer	$p <$
$p < 0.01$	anger-disgust	0.01
	anger-surprise	0.01
	fear-surprise	0.05

表 2: VISGRAF-DB での分散分析結果

ANOVA	Tukey-Kramer	$p <$
$p < 0.01$	anger-happiness	0.01
	disgust-happiness	0.01
	fear-happiness	0.01
	happiness-sadness	0.01
	happiness-surprise	0.01

喜び識別器に対して他の 5 種の感情識別器が $p < 0.01$ で有意差があると判定され、識別器の他の組み合わせに対しては差異が認められなかった。

5 結論

AIST-DB, VISGRAF-DB ともに、喜び表情 (笑顔) による喜びの感情推定の信頼性は高いものと考えられる。一方で、喜び以外の感情推定は、表情によっては判別が難しい組み合わせがあることが示されていると考えられる。驚き識別器が、驚き表情と同時に、喜び表情と恐怖、悲しみなどの表情に対しても高値を示すことから、複数の感情が組み合わせられていると考えられる例もあり、一つの表情が必ずしも一つの感情のみ結び付けられるものではないことが示唆されている。訓練に用いた画像は、主に欧米人の多く含まれるデータセットであったため、AIST-DB の日本人の表情に対しては、推定精度が下がっていた可能性もある。表情に対する文化的背景、性差なども考慮していく必要がある可能性がある。逆に考えると、文化的背景、人種、性差などを越えて、喜び表情 (笑顔) の識別は普遍的なものと考えることができ、興味深い結果となった。

統計的な分析結果は、喜び (笑顔) とそれ以外の表情の判別性が高いため、他の表情間の判別性がマスクされてしまった可能性を示唆している。今後は、喜びの影響を排除した他の表情の識別手法を検討していく必要があると考えられる。

謝辞

表情識別器構成手法に関して有益なご助言をいただいた産業技術総合研究所人間情報研究部門松田圭司氏、

ならびに、笑顔度識別器のプロトタイプを回想法に適用することによってその有用性を示していただいた筑波大学人間系山中克夫准教授に感謝いたします。

参考文献

- [1] 総務省統計局: 人口推計 (2018 年 (平成 30 年) 10 月 1 日現在), <https://www.stat.go.jp/data/jinsui/new.html>
- [2] 厚生労働省: 平成 29 年 (2017) 患者調査の概況, <https://www.mhlw.go.jp/toukei/saikin/hw/kanja/17/d1/05.pdf>
- [3] 関 啓子: 音楽と高次脳機能障害, 音楽医療研究, Vol. 10, No. 1, pp. 14–25 (2017).
- [4] 嶋田敬士, 山田亨, 高橋友香, 野口祥宏, 山崎郁子, 福井和広: SVM による笑顔度推定技術を用いた音楽療法効果の評価, 情報処理学会論文誌, Vol. 55, No. 12, pp. 2569–2581, (2014).
- [5] 中谷淳, 山中克夫: 認知症ケアにおける回想法, 保険の科学, Vol. 48, No. 4, pp. 254–258, (2006).
- [6] Whitehill, J., Littlewort, G., Fasel, I., Bartlett, M. and Movellan, J.: Toward Practical Smile Detection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.31, No.11, pp.2106–2111 (2009).
- [7] Deniz, O., Castrillon, M., Lorenzo, J., Anton, L. and Bueno, G.: Smile Detection for User Interfaces, *ISVC' 08: Proc. 4th International Symposium on Advances in Visual Computing*, Part II, pp.602–611, Springer-Verlag, Berlin, Heidelberg (2008).
- [8] Shan, C., Gong, S. and McOwan, P.W.: Facial expression recognition based on local binary patterns: A comprehensive study, *Image and Vision Computing*, Vol.27, No.6, pp.803–816 (2009).
- [9] Moore, S. and Bowden, R.: Local binary patterns for multi-view facial expression recognition, *Computer Vision and Image Understanding*, Vol.115, No.4, pp.541–558 (2011).
- [10] Shimada, K., Noguchi, Y., Kurita, T.: Fast and Robust Smile Intensity Estimation by Cascaded Support Vector Machines, *Int. natnal. J. of Computer Theory and Engineering*, Vol. 5, No. 1, pp. 24–30, (2013).
- [11] Yu, Z., Cha, Z.: Image Based Static Facial Expression Recognition with Multiple Deep Network Learning, *Proc. 2015 ACM International Conf. on Multimodal Interaction*, pp. 435–442, (2015).
- [12] Zhang, K., Huang, Y., Du, Y., Wang, L.: Facial Expression Recognition Based on Deep Evolutional Spatio-Temporal Networks, *IEEE trans. on Image Processing*, Vol. 29, No. 9, pp. 4193–4203, (2017).
- [13] Zhang, K., Huang, Y., WU, H., Wang, L.: Facial smile detection based on deep learning features, *3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 534–538, (2015).
- [14] Sabri, M., Kurita, T.: Facial expression intensity estimation using Siamese and triplet networks, *Neurocomputing*, Vol. 313, No. 3, pp. 143–154, (2018).
- [15] Kobayashi, T., Otsu, N.: Image Feathre Extraction Using Gradient Local Auto-Correlations, *European Conference on Computer Vision (ECCV)*, pp. 346–356. (2008).
- [16] Borod, J. C., Koff, E., Perlman Loach, M., Nicholas, M., Welkowitz, J.: Emotional and non-emotional facial behaviour in patients with unilateral brain damages, *J. of neurology, Nuerosurgery, and Psyhiatry*, Vo. 51, pp. 826–832, (1988).
- [17] Patel, S., Oishi, K., Wright, A., Sutherland-Foggio, H., Saxena, S., Shppard, S. M., Hillis, A. E.: *Frontiers in Neurology*, Vol. 9, Article 224, pp. 1–7, (2018).
- [18] D. Merget, T. Eckl, M. Schw?rer, P. Tiefenbacher, and G. Rigoll, Capturing Facial Videos with Kinect 2.0: A Multithreaded Open Source Tool and Database, in *Proc. WACV, IEEE*, 2016.
- [19] Fujimura, T., Umemura, H.: Development and validation of a facial expression database based on the dimensional and categorical model of emotions. *Cognition & Emotion*, Vol. 32, pp. 1663–1670, (2018).
- [20] Mena-Chalco J., Marcondes R., Velho, L.: Banco de Dados de Faces 3D: IMPA-FACE3D, *TR 01, IMPA - VISGRAF Laboratory*, (2008).