

苛立ち表情発見のための自動車運転者の顔画像クラスタリング Clustering Facial Images of a Car Driver for Discovering Irritated Expressions

藤田隆吾[†] 大浦聡一郎[‡] 松川徹[§] 鈴木英之進[¶]
Ryugo Fujita Soichiro Oura Tetsu Matsukawa Einoshin Suzuki

1. 序論

人間の表情はその人の感情に関して有効な情報を多く含み、人同士のコミュニケーションにおいて重要な役割を担っている。[1]によると、Ekman は人の表情はしばしばその人の感情についての情報を与えてくれると述べている。感情はその人の行動に大きな影響を与えるため、表情によってその人がいまだのような感情を抱いているのかを推定することで後に起こしうる行動を予測し、それに対する警告や注意を行うことができるようになる。Ekman らによって提唱された Facial Action Coding System[2]によると、人の基本的な表情は幸福・悲しみ・恐怖・怒り・嫌悪・驚きの6個からなり、この6個の表情は人種・性別・年齢などに関係なく人類に普遍であるとされている。しかし人間の表情はもちろんこれらの6個だけではなく、他の表情が人類に普遍的なものであるかは研究が進められている最中である。

本研究が対象とするのは自動車の運転者である。わが国において自動車保有台数は年々増加の傾向にある。2015年3月においての乗用車の保有台数は[3]によると約6千万台であり、2015年3月のわが国の世帯数が約5460万世帯であること[4]を考えると、1世帯に平均1台以上の乗用車を保有していることになる。それほどまでに自動車は広く普及しており、日常生活においてなくてはならないものになっている。しかし、自動車による悲惨な交通事故が多発していることも見逃すことのできない事実である。国土交通省の調べ[5]によると、死傷交通事故の件数は平成17年以降減少傾向にあるが、それでも年間に60万件近くの死傷事故が起きている。これらの死傷事故の原因はさまざまであるが、運転者の疲労や苛立ちが原因となっている事故も少なからず存在する。よって運転者の表情から運転者が疲労や苛立ちを感じていることを認識することは重要である。

これまで運転者の表情による研究は運転者の疲労や不注意に注目したものが主であった。Jiらは2004年に運転者の疲労をリアルタイムでモニタリングするデバイスについての研究[6]を発表した。これはまぶたの動きや視線、頭部の動きや表情など様々な視覚的情報により疲労の検知を行ったものである。この研究は異なった宗教的背景や、性別、年齢などの個人による相違も考慮して行われた。Dongらは2011年に運転者の不注意をモニタリングすることを目的とした最先端技術

に関する研究[7]を発表した。これは運転中の不注意の原因が心の動揺と疲労の二つに分けられると考え、1) 運転者の主観的な指標、2) 生物学的な指標、3) 身体的な指標、4) 実験時の運転者のパフォーマンスを考慮した指標、5) 以上の4つを組み合わせた指標の5つの指標を考慮し運転者のモニタリングを行った研究である。

本研究では自動車の運転者が運転中に示す表情のクラスタリングを行う。教師つき学習ではなくクラスタリングを用いた理由として、様々な表情が現れるためラベル付けが困難であることや、クラスタリングにすることにより新たな苛立ちの表情を発見できると考えたことが挙げられる。

本論文は全6節で構成されている。第2節では本研究の目的及び困難な点を示し、本研究の評価方法を説明する。第3節では本研究に関連する研究を紹介し、第4節では本実験での提案手法を提示する。第5節で実際に行った実験を5つのステップに分けて説明し、本実験の結果を示しそれに対する考察を述べる。第6節では本研究の結論を述べる。

2. 対象問題

入力には運転者の表情を撮影した動画のフレームから抽出した画像群 I_i ($i = 1, 2, \dots, N$) とし、出力はクラスタリングにより生成されたクラスタ C_i ($i = 1, 2, \dots, M$) とする。ここで N は画像の枚数であり、 M は生成クラスタ数である。

運転者の表情のデータの取得および解析には問題点がある。自動車の運転中は、道路の表面の状態や自動車の発進及び停止などにより車内はほぼ常に振動している状態にあるため、カメラと運転者の顔の距離や撮影された動画内での運転者の顔の位置が一定ではない。よって顔領域抽出によく用いられるOpenCVの顔領域抽出のプログラムでは正確に顔領域を切り取るのは難しいと考えられる。

本研究では車のダッシュボードに設置したウェアラブルカメラにより運転者の顔の動画を撮影する。この動画のフレームから顔画像 I_i を抽出し、68個のランドマーク $l_p = (x_p, y_p)$ ($p = 1, 2, \dots, 68$) の座標により得られる特徴を用いて解析を行う。 l_p を抽出する手法については第4節で説明する。BIRCHにより生成されたクラスタ C_i 内の画像を見て苛立ちの表情にどのようなものがあるかを考察する。

提案手法の精度を評価するため、ランドマークの座標を直接入力したときに得られたクラスタと、ランドマークにより得られる特徴を入力したときに得られたクラスタを比較する。出力されたクラスタを評価するため、本研究では正規化相互情報量 (NMI) を用いる。NMIはクラスタリングの結果を評価するために頻りに用いられている指標である。NMIの計算方法および評

[†]九州大学システム情報科学府情報学専攻

[‡]九州大学システム生命科学府システム生命科学専攻

[§]九州大学システム情報科学府情報学部門

(社)電子情報通信学会, IEICE

[¶]九州大学システム情報科学府情報学部門

(社)情報処理学会, IPSJ

価方法は第 3 節で説明する。

3. 関連研究

ここでは本研究の手法と関連する研究を紹介する。

まず本研究でクラスタリング手法として用いた BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) は Zhang らによって提案された [8]。BIRCH はデータ集合 $X = \{x_1, \dots, x_n\}$ を CF (Clustering Feature) ベクトルの集合である $V = \{v_1, \dots, v_m\}$ に変換し、その集合をクラスタリングする手法である。各 CF ベクトル v_i は距離が近い複数個の例を表したものであり、一つの集合のデータを $\{x'_1, \dots, x'_p\}$ ($x'_i \in X$) と表すと、CF ベクトルはデータ数、線形和、および二乗和で表される。CF ベクトル v についての式を以下の式 (1) に表す。

$$v = \left(p, \sum_{i=1}^n x'_i, \sum_{i=1}^p \|x'_i\|^2 \right) \quad (1)$$

あるクラスター $\{C_{\nu(1,k)}\}$ ($k = 1, 2, \dots, N_1$) と別のクラスター $\{C_{\nu(2,l)}\}$ ($l = N_1 + 1, N_2 + 2, \dots, N_1 + N_2$) が与えられた時、その二つの間のユークリッド距離における平均クラスタ間距離 D は次の式 (2) のように表される。

$$D = \left(\frac{\sum_{k=1}^{N_1} \sum_{l=N_1+1}^{N_1+N_2} (C_{\nu(1,k)} - C_{\nu(2,l)})^2}{N_1 N_2} \right)^{\frac{1}{2}} \\ = \sqrt{\frac{N_1 S S_2 + N_2 S S_1 - 2 L S_1 \cdot L S_2}{N_1 N_2}} \quad (2)$$

ここで $\{C_{\nu(1,k)}\}$ とは 1 番目のクラスターの CF ベクトルの k 番目の要素を表す。 LS_1, LS_2 はそれぞれクラスター $\{C_{\nu(1,k)}\}, \{C_{\nu(2,l)}\}$ の線形和を表す。 SS_1, SS_2 はそれぞれクラスター $\{C_{\nu(1,k)}\}, \{C_{\nu(2,l)}\}$ の平均和を表す。 BIRCH において、CF ベクトルにより集合間の距離を求めることで、元の集合を用いずとも近似的にクラスタリングを行うことができる。CF ベクトルは加法性を持っているため、逐次集合が増加しても CF ベクトルとして表されたクラスタ情報を効率よく更新することが可能である。 BIRCH では変換した CF ベクトルを管理するために、高さ平衡木の CF 木を用いている。CF 木は各ノードで CF ベクトルを保持しており、入力に対して逐次 CF 木を構築し、各ノードの CF ベクトルを更新していくことでクラスタリングを行う。 BIRCH の利点の一つとしてクラスタ内の元データの情報をすべて保持する必要がないということがあげられる。元のデータから生成された CF ベクトルからでも平均クラスタ間距離をはじめとする種々の例集合間の距離を計算することが可能であるため、実際に保持すべき情報はクラスタ数分の CF ベクトルだけでよいということになる。さらに CF ベクトルは加法性を持つため、クラスタ内のデータの情報が変更されても容易に CF ベクトルとして表されたクラスタ内情報を更新することができる。

クラスタリングの結果として正解クラスタの数と生成されるクラスタの数と必ずしも一致しない場合がよ

く起こりうる。本研究も正解クラスタと生成されるクラスタの数は一致しない可能性がある。よって双方のクラスタ数が一致しない場合でも評価基準となりうる正規化相互情報量 (Normalized Mutual Information) を評価指標として導入する。NMI の計算方法は文献 [9] にあるものを利用する。生成したクラスタ集合を $X = \{X_1, X_2, \dots, X_n\}$ 、正解クラスタ集合を $Y = \{Y_1, Y_2, \dots, Y_m\}$ とし、 X_i に含まれるデータ数を N_i 、 Y_j に含まれるデータ数を M_j とする。 X_i に含まれるデータの中で、 Y_j にも含まれるデータの数を $K_{i,j}$ とする。すなわち、全データ数を S とすると、

$$\sum_{i=1}^n N_i = \sum_{j=1}^m M_j = \sum_{j=1}^m \sum_{i=1}^n K_{i,j} = S \quad (3)$$

となる。このとき、NMI は以下の式 (4) で表される。

$$NMI = \frac{-2 \sum_{i=1}^n \sum_{j=1}^m K_{i,j} \log\left(\frac{K_{i,j} S}{N_i M_j}\right)}{\sum_{i=1}^n N_i \log\left(\frac{N_i}{S}\right) + \sum_{j=1}^m M_j \log\left(\frac{M_j}{S}\right)} \quad (4)$$

このクラスタリング結果が 1 に近いほど生成されたクラスタと正解クラスタの数が近いということになり、クラスタリングの結果が良いということになる。

4. 提案手法

この節では本研究の手法について説明する。

まず取得した運転者の表情の動画より顔のランドマークの検出を行った。用いたプログラムは文献 [10] に示されているアルゴリズムが実装されている Dlib C++ ライブラリを利用して作成した。このプログラムは機械学習メタアルゴリズムであるブースティングを用いた教師あり学習により顔のランドマークの位置を推定する。このプログラムの出力はランドマークの座標及びランドマークが表された顔画像である。ランドマークが表された顔画像の一例を図 1 に示す。

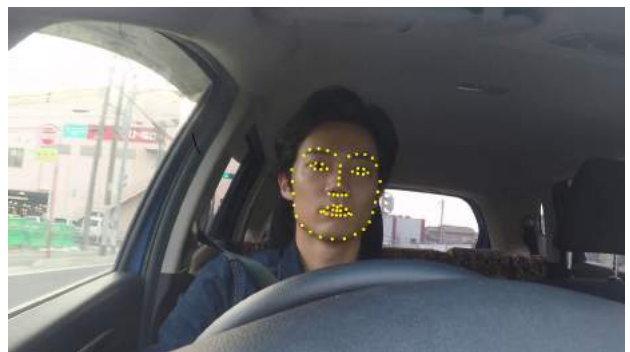


図 1: ランドマークが表された顔画像の一例

次に得られたランドマークの座標により奇立ちに関する特徴を考案した。単純にランドマークの座標をクラスタリングに入力しても運転者の表情をとらえることは難しい。よって運転者が奇立ちを感じている画像と感じていない画像を数枚用意してそれを比較し、ラ

ランドマークの座標を用いて得ることのできる苛立ちの表情の特徴を考えた．下の図 2 に運転者が無表情であるときと苛立ちを感じているときの表情の画像の例を示す．



図 2: 入力した画像群

上の画像群をみて，1) 目の開き具合 F_{eye} ，2) 眉のしかめ具合 $F_{eyebrow}$ ，3) 口の開き具合 F_{mouth} ，4) 顔の 3 次元角度 Θ ，5) 顔とカメラの距離 D_{face} の 5 つが苛立ちの良い特徴であると考えた．よってランドマークの座標によりこれらの特徴を計算し，クラスタリングの入力とした．

5. 実験

本実験の流れは 1) 運転中の運転者の表情及び車外の状況動画の撮影，2) 動画を切り取った画像からランドマークの座標の抽出，3) ランドマークの座標から考案した苛立ちの特徴の計算，4) 計算した特徴を入力とした BIRCH によるクラスタリング，5) 得られたクラスタの評価の 5 つのステップからなる．

5.1. 運転者の表情及び車外の状況の撮影

まず自動車を運転中の運転者の表情およびそのときの車外の状況を GoPro HERO4 Black を用いて撮影した．撮影日時は 2016 年 2 月 1 日 17 時～18 時頃の約 59 分間で，天候は晴れであった．カメラの解像度は 1920 pixel × 1080 pixel で fps (frame per second) は 30 に設定した．撮影された画像は 89630 枚であった．下の図 3，4 に撮影された画像の一例を示す．



図 3: 運転者の表情の画像 図 4: 車外の状況の画像の一例

5.2. 顔画像からのランドマークの抽出

次に運転者の顔画像からのランドマークの抽出を 4 節で紹介したプログラムにより行った．抽出されたそのままの座標を用いると運転者の顔の角度やカメラとの距離により画像ごとの座標の位置関係にずれが生じるため，その修正を行った．

5.3. 苛立ちの特徴の計算

次に修正したランドマークの座標から 4 節で述べた苛立ちの特徴の計算を行った．具体的な計算方法としては 1) F_{eye} は左右の目の上端と下端のランドマークの距離の平均，2) $F_{eyebrow}$ は両眉の内側のランドマークの距離，3) F_{mouth} は口の上端と下端のランドマークの距離を計算した． Θ は yaw, pitch, roll の三つからなる考え，それぞれを計算した．最後に 5) D_{face} は，本来はカメラとの距離によらず一定である顔の左端と右端のランドマークの距離の比を計算した．

5.4. BIRCH によるクラスタリング

次に計算した苛立ちの特徴を入力として BIRCH によりクラスタリングを行った．撮影により得られた顔画像は 89630 枚であったが，入力には連続する 10 枚につき 1 枚を使用し，計 8963 枚を使用した．画像にはあらかじめ苛立っているかいないかの 2 クラスでラベル付けを行っている．ここで苛立っているとラベルをつけた画像は 141 枚，苛立っていないとラベル付けした画像は 8822 枚である．この入力画像を閾値を 10 通りに変動させてクラスタリングを行い，出力として生成クラスタとそれぞれのクラスタの CF ベクトルを得た．

5.5. 生成されたクラスタの評価

最後に生成されたクラスタの評価を行った．まずすべての入力画像に苛立っているか苛立っていないかという 2 クラスのラベルを付けた際の生成クラスタと NMI の関係を図 5 に表す．

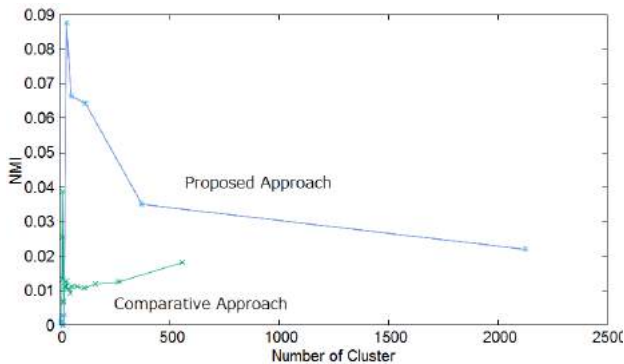


図 5: 生成クラスタ数と NMI の関係

上の図を見ると提案手法である苛立ちの特徴を入力とした時の NMI が、比較手法であるランドマークの座標のみを入力とした時の NMI を上回っていることがわかる。しかし提案手法においても NMI が一番高い時でも約 0.09 と非常に低い値をとっていることが分かる。これは運転者の表情クラスタリングという問題の難しさが原因にあると考えられる。これまで行われてきた人の表情の研究は実験室等の振動がなく顔の照度の変化がほぼない環境において実験が行われることが多かったが、本研究での運転中の車内のような空間においては撮影した画像のノイズとなりうる要因が多く存在するため、非常に難易度の高い問題であると考えられる。さらに、今回用いた苛立ちの 5 つの特徴は画像ごとの変化が微量であるため、ランドマークの抽出が少しずれただけでも大きな影響を受けることも原因の一つであると考えられる。

次に提案手法において NMI が最も高い値をとった時のクラスタを詳しく見てみた。このとき 141 枚の苛立っている画像は 5 つのクラスタに分かれた。この 5 つのクラスタをそれぞれクラスタ A, B, C, D, E と名前を付け、以下の図 6~10 にそれぞれのクラスタの代表例を示す。



図 6: クラスタ A の苛立ち表情の代表例 (85 枚)



図 7: クラスタ B の苛立ち表情の代表例 (28 枚)



図 8: クラスタ C の苛立ち表情の代表例 (23 枚)

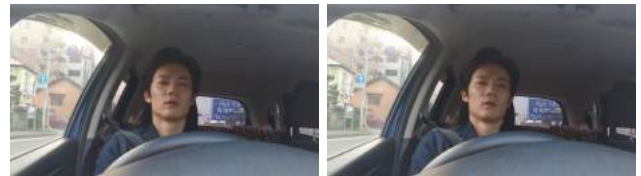


図 9: クラスタ D の苛立ち表情の代表例 (4 枚)



図 10: クラスタ E の苛立ち表情の代表例 (1 枚)

なお図のタイトルの括弧内の枚数はそのクラスタに含まれる苛立っている画像の枚数である。これらのクラスタ内の画像をみると、同じ苛立ちという感情を表している表情が複数種類存在していることが分かる。これらの別れ方は設定した苛立ちの特徴と同じく目の開き具合・眉のしかめ具合・口の開き具合・顔の 3 次元的角度が関係していると考えられる。

6. まとめと今後の展望

本研究では自動車の運転者が運転中に見せる苛立ちの表情の種類を発見することを目的として行った。その結果として苛立ちの表情は 5 つのクラスタに分かれた。本実験では表情撮影の対象者が一人のみであったが、今後対象者を複数人に増やし、さまざまな苛立ち表情を得ることにより運転中の苛立ち検知に貢献することが期待される。

参考文献

- [1] Paul Ekman. Facial Expression and Emotion. *American Psychologist*, 48(4):384-392, 1993.
- [2] Paul Ekman and Wallace V. Friesen. Constants across Cultures in the Face and Emotion. *Journal of Personality and Social Psychology*, 17(2):124-129, 1971.
- [3] 一般財団法人自動車検査登録情報協会. 自動車の保有台数の推移. <https://www.airia.or.jp/>

publish/statistics/ub83el00000000wo-att/
hoyuudaisuusui.pdf. 閲覧日 2016 年 2 月 9 日.

- [4] 総務省. 住民基本台帳に基づく人口, 人口動態及び世帯数のポイント(平成 25 年 3 月 31 日現在). http://www.soumu.go.jp/main_content/000244522.pdf. 閲覧日 2016 年 2 月 9 日.
- [5] 国土交通省. 交通事故死者数・死傷者数・死傷事故件数の状況. <http://www.mlit.go.jp/road/road/traffic/sesaku/genjyo.html>. 閲覧日 2016 年 2 月 9 日.
- [6] Qiang Ji, Zhiwei Zhu, and Peilin Lan. Real-Time Nonintrusive Monitoring and Prediction of Driver Fatigue. *IEEE Transactions on Vehicular Technology*, 53(4):1052–1068, 2004.
- [7] Yanchao Dong, Zhencheng Hu, Keiichi Uchimura, and Nobuki Murayama. Driver Inattention Monitoring System for Intelligent Vehicles: A Review. *IEEE Transactions on Intelligent Transportation Systems*, 12(2):596–614, 2011.
- [8] Tian Zhang, Raghu Ramakrishnan, and Miron Livny. BIRCH: A New Data Clustering Algorithm and Its Applications. *Data Mining and Knowledge Discovery*, 1(2):141–182, 1997.
- [9] Leon Danon, Albert Diaz-Guilera, Jordi Duch, and Alex Arenas. Comparing Community Structure Identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09):P09008, 2005.
- [10] Vahid Kazemi and Josephine Sullivan. One Millisecond Face Alignment with an Ensemble of Regression Trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1867–1874, 2014.