

自動運転時代におけるドライバモニタリング技術

時系列 Deep Learning によるドライバ状態の推定について

日向 匡史・木下 航一・西行 健太・長谷川 友紀

本稿ではドライバモニタリングシステムにおいて、自動運転時にドライバが起こしうる多様な挙動・状態を推定する技術を提案する。当面の間、自動運転時でもドライバは運転行動に対して責任があり、ドライバの状態を監視する技術が必要であると言われている。そこで我々は運転行動に基づいたドライバの状態を測る指標を定義し、逐次出力する技術を構築した。これにより、ドライバが適切な状態で運転状況を監視し続け、自動運転中の事故リスクを削減することが可能となる。我々はこれを実現する手段としてカメラから得られる画像列から、ドライバの顔の向きや視線といった局所的な情報と、姿勢や動きに関する大局的な情報を抽出した。これらを時系列 Deep Learning により学習されたニューラルネットワークに入力することで、ドライバ状態の推定を高精度に認識することが可能となる。

Driver Status Monitoring System in Autonomous Driving Era

Driver Status Estimation with Time-series Deep Learning

Tadashi Hyuga, Koichi Kinoshita, Kenta Nishiyuki and Yuki Hasegawa

We propose a novel driver's status estimation system based on driver's behavior in autonomous driving scene. In future years, drivers are expected to be responsible for driving and monitoring the surrounding environment even in autonomous driving. To monitor the driver status is very important to ensure that he/she stays proper condition during driving. We define new criteria to measure the adequateness level of the driver during autonomous driving, and realize real-time driver status monitoring based on the criteria. It will contribute to decrease the risk of traffic accidents.

To measure driver's status precisely, we utilize two features: 1st one is driver's facial information, i.e. face direction, gaze direction and eye open-close level, and 2nd one is driver's behavior taken by image sequence. These features are concatenated and fed into a pre-trained recurrent neural network. By using the state-of-the-art neural network technology, the driver's status can be estimated with high accuracy.

1. まえがき

自動運転の実用化へ向けた動きが本格化しており、2020年頃には高速道などの特定の環境下での自動運転が実現すると想定されている¹⁾。しかし、完全な自動運転を実現するには及んでおらず、当面は表1に示す自動運転レベルの中でも、ドライバの責任下での部分的な自動運転をする「自動運転レベル2」が主流になると言われており、ドライバは自動運転中も運転行動が適切であるか監視する必要がある。この段階は当面続くとする調査結果もある²⁾。

そこで我々は、自動運転時におけるドライバの状態に着目し、ドライバが運転に集中している状態かを推定するドライバモニタリング技術を開発した。本稿では、ドライバ

の集中度合いを評価する指標として実際の運転行動に基づいた3指標を新たに定義し、ドライバを撮影した画像の時系列を入力とし3指標の評価結果を逐次出力する、ドライバ運転集中度センシング技術を提案する。

表1 米SAEが提唱する自動運転レベル

レベル0	自動化なし
レベル1	運転支援 (ADAS)
レベル2	部分的な運転自動化
レベル3	条件付き運転自動化
レベル4	高度な運転自動化
レベル5	完全運転自動化

2. 背景

2.1 自動運転を巡る社会動向 自動運転に向けた法改正の議論が現在も続いている。国連の自動操舵に関する国際基準（通称R79）の改正に向けた議論の中で、自動運転車は下記要求を満たすように制度化される見込みである³⁾。

- システムが機能限界に陥る4秒前にはドライバに警告
- ドライバが運転に集中しているか常時監視し、居眠りなどをしている場合には警告
- 警告に応じない場合は自動的に危険を最小化する制御を実行

以上からドライバの状態を監視することは将来的に必須の機能であり、その中でも警告に適切に応じることが可能かを自動運転システムは確認しておく必要がある。

2.2 従来のドライバモニタリングシステム 従来のドライバモニタリングは手動運転中のドライバが正常に運転できている状態かを検知するものであり、基本的に1指標でもって判断がなされるものが多い。たとえば、目の開閉に注目した居眠り検知や顔の向きに注目した脇見検知などがあり、実用化も進んでいる⁴⁾。しかし、これらは手動運転中にドライバが車両前方を注視しているかを判別するのみで、自動運転時に起こりうる多様な動作の検知は困難である。また最近では、耳に装着することで脈波を計測することにより眠気状態を検知するという技術が開発されており⁵⁾、これによりドライバが眠いかを判断することが可能となるが、覚醒時に起こりうる多様な危険に対応することが困難である。また、デバイスを装着することによって運転時のドライバの心理的負担となるため、センサとしては非接触のものが望ましい。

また、人の動作に対してラベルを付与し、画像の時系列からどのラベルに相当するかを認識する技術⁶⁾⁷⁾があり、これをドライバモニタリングに適用することも考えられる。しかし、この手法は所定の動作を認識するものであり、自動運転時に起こりうる多様な動作を網羅するには識別性能の劣化が想定される。以上から、従来手法はドライバの多様な動作に対応できず、自動運転への適用は困難と考える。

2.3 Deep Learningを活用した運転集中度の提案 筆者ら⁸⁾は、自動運転モードから手動運転モードへの走行モード切替時にドライバの運転復帰の可能性を判断するドライバモニタリング技術を提案している。これは走行モード切り替え時にドライバがどのくらいの所要時間で運転復帰可能になるかを判断し、その所要時間に応じて円滑な走行モードの切り替えを実行できるよう車両の制御システムに情報を送信する。しかし、この技術においては想定されたモード切り替えには対応可能であるが、自動運転車の周辺環境の認識が不完全なことによって起こりうる緊急事態への対応を想定していない。このため、本稿で提案する技術では、緊急時の対応および想定される切り替えの双方に対応した新たな指標を提案しこれを実現することを目指した。

3. 自動運転時におけるドライバ状態の評価指標の提案

以上の観点から、我々は自動運転時における運転集中度センシングの指標として新たに3指標、Eyes-on/off、Readiness-high/mid/lowおよびSeated-on/offを定義する。これらは「認知」「判断」「操作」という実際の運転行動と密接に関係するものとしている。図1にその関係を示す。

3.1 Eyes-on/off この指標はドライバが常時走行を監視できているかを確認するためのものである。ドライバが進行方向を確認している状態、もしくは運転上必要となる短時間の確認動作、たとえば計器・ミラーの確認など、を行っている場合はEyes-on、それ以外のドライバの挙動、たとえばスマホや本、カーナビを注意する、目を閉じている、といった状態はEyes-offとなる。

3.2 Readiness-high/mid/low ドライバが運転の準備ができているかを3段階で出力する。覚醒して運転に無関係な動作をしていない場合はReadiness-high、運転に無関係な動作をしているが、システムからの警告を受けて軽い手順で運転に復帰できるような状態をReadiness-mid、寝ているなど運転が困難な状態をReadiness-lowと定義する。

3.3 Seated-on/off ドライバが運転席に着座しているかを指標として、運転行動がとれるかを判断する。ドライバが着座していればSeated-on、離席していればSeated-offとなる。手動運転時には想定できないが、自動運転がより高度になるに従いドライバの監視に対する意識が低くなり、運転行動の準備を怠る可能性があるため、これを定義する。

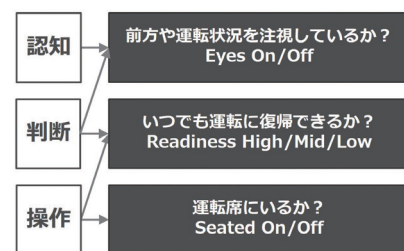


図1 運転集中度の3指標と運転行動との関係

4. 自動運転時における運転集中度センシング

本章では3指標を識別するための提案手法の処理の流れを記載する（図2）。識別器は大きく分けて3段階の構成となっており、近赤外線カメラから入力される画像列に対して、まず顔画像センシング技術を適用することにより顔の局所的な情報を取得する。同時に Convolutional Neural Network（以下CNN）⁹⁾を用いドライバの姿勢に相当する特徴を取得する。そしてこれらの出力を統合し、再帰型ニューラルネットワーク（以下RNN）の1種である、Long Short-Term Memory（以下LSTM）¹⁰⁾を用いて、時々刻々変化するドライバ状態の遷移を認識する。CNNと

RNNを組み合わせたネットワーク構成はよく研究されているが³⁾、我々は顔画像センシングを組み合わせることによって高精度化と高速化を実現している。以下に詳細を記載する。

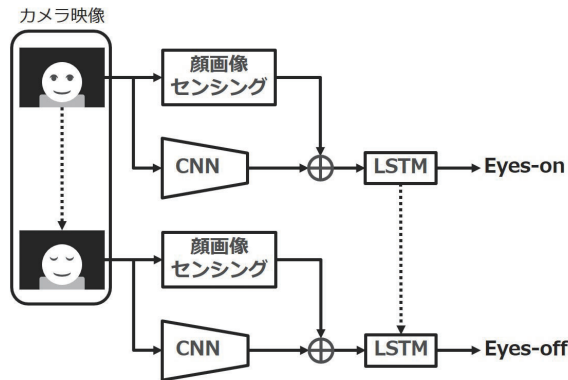


図2 運転集中度センシングのネットワーク概略図

4.1 近赤外線カメラによる画像入力 車両中のドライバの画像を安定して取得するために、我々は近赤外線カメラを採用した。一般的なRGBカメラについては日照条件によっては活用可能ではあるが、昼間は直射日光の入射の影響が大きく、また夜間は照明が必要になる上、昼間の顔とパターンが変化してしまうため望ましくない。今回採用したカメラはカメラユニットに搭載された近赤外線LEDによって不可視領域ではあるが顔領域に常時照明が照射された状態となるため、昼夜問わず安定した顔パターンが取得可能である（図3）。



図3 近赤外線カメラによる人物の画像
(同意書により画像利用許諾確認済)

4.2 顔画像センシング 画像から顔領域を検出し、顔に付随する様々な情報を出力する機能を指す。我々はオムロンが開発したOKAO Visionと呼ばれる技術をベースとして用いており、以下の機能を運転集中度センシングのために活用している。

- 顔検出
- 顔器官点検出
- 顔向き推定
- 目開閉識別
- 視線推定

これら個々の技術は従来RGBカメラでの顔のみに対応しているが、近赤外線カメラで得られた画像へ対応させる

ことで、高速高精度にドライバの特徴を取得することが可能である。

4.3 CNN CNNは従来の全結合型のネットワークと異なり、学習により取得された小領域のフィルタと画像を畳み込み演算を実施する畳み込み層、および畳み込み層で得られた画像を所定のルールで圧縮するプーリング層を幾重にも重ねた構成を持ち、画像の変形に対するロバスト性を向上させたネットワークである。提案自体は古くより行われていたが、近年の汎化性能を向上させる学習方法の提案により、様々な画像認識系のベンチマークテストのState-of-the-Artを更新した、昨今のDeep Learningブームの火付け役である。

4.4 LSTM LSTM（図4）は入力として時系列データを用い、所定の1フレームの認識結果を得るために該当フレームの情報に加えて前フレームの中間出力を入力とする。また、セルと呼ばれる内部記憶を保持し、この値によって出力に対する該当フレームの入力の重みを計算する。この重みは事前学習によって挙動が設定されており、従来のRNNに比べてより長期の記憶が可能になることで知られている。

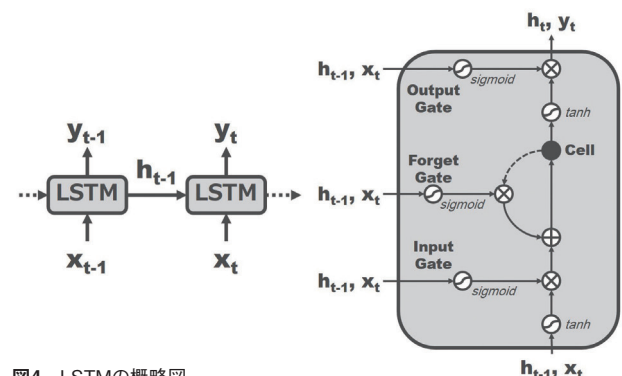


図4 LSTMの概略図

4.5 学習データセットの構築 図2に記載のネットワークに対して、各パラメータを時系列Deep Learningを用いて決定する。我々は自動運転中にドライバが起こしうる動作を洗い出し、その中から代表的なパターンを選定し、学習に利用した。表2、表3および表4にそれぞれの指標に関する動作例を示す。

100人の被験者に対してそれぞれの動作を指示し。その様子を撮影した画像列を用いる。このうち50人分のデータを学習用のデータとし、残りを評価対象とした。評価に際しては1動画中1動作とし、認識に用いたフレーム毎の認識結果を集計し、多数決で決定した運転復帰可否レベルと正解レベルを比較する。カメラの画像サイズは720×480、画角約10度、フレームレートは30 fpsである。このカメラを運転席正面に設置しドライバの動作を一定時間撮影し、計200万フレームのデータを収集した。なお、ニューラルネットワークの学習および評価にはPreferredNetworks社が提供する学習フレームワークであるChainer¹¹⁾を用いた。

表2 Eyes-on/offの動作例

On	運転をする 正面を注視する 窓にもたれる
Off	脇見する スマホを操作する 居眠りをする

表3 Readiness-high/mid/lowの動作例

High	運転をする 正面を注視する 計器を一時的に確認する
Mid	飲食をする スマホを操作する 通話する
Low	居眠りをする 突っ伏す パニックになる

表4 Seated-on/offの動作例

On	上記のドライバ動作
Off	なし（運転席に搭乗しない）

4.6 評価 学習データとほぼ同数のデータを評価データとして用いる。このとき、各指標における正解率は表5のようになった。

表5 各指標の正解率

Eyes	95.4%
Readiness	94.8%
Seated	99.0%

5. 実用化に向けた検討

5.1 現場の課題を網羅した大規模データベースの構築

以上の技術を構築するにあたり、我々は学習および評価に必要なドライバの様々な挙動を撮影したデータベースを構築した。自動運転中のリアルなドライバの挙動データは現状取得が困難なので、①ドライブシミュレータの活用、②実車走行中の助手席での撮影、③手動運転中のドライバの撮影、など様々なシチュエーションでの撮影を実施し、データベースの構築を行った。

表2や表3に示したドライバの不適切な行為や異常な状態を撮影するには、走行中の車両で実際に演技をすると危険が伴うため、ドライブシミュレータにて撮影を行った（図5）。前方のモニタには運転中の風景を表示することで実際の走行中の環境に近づけたうえで、被験者に様々な演技を指示した。また、走行中の助手席に座った搭乗者は、運転状況の監視という点では自動運転中のドライバの挙動に近いが、実際には運転の責任を持たないため、自動運転中のドライバの運転集中度の推移に乖離がある可能性が高い。運転中のドライバは運転の責任があるため、もっとも運転集中度が高い状態を撮影可能であるが、集中度が低下したデータ

の取得が困難という短所がある。我々はこれらのデータを組み合わせて、多様なドライバ状態を網羅したものとみなす。



図5 ドライブシミュレータ
（同意書により画像利用許諾確認済）

5.2 顔画像センシングの耐環境性向上 顔画像センシング技術において、画像中の顔の位置を推定する技術を顔検出と呼ぶが、一般的には顔に現れる固有の明暗パターンを学習し、画像中に現れる同様なパターンを顔として出力する。このとき横顔やマスク装着時の顔（以降マスク顔と呼ぶ）のようにパターンが大きく変化した顔画像の検出は困難となる。そこで我々は上記データベースから横顔やマスク顔のサンプルを抽出し、これらを従来の顔検出の学習サンプルと組み合わせることで、横顔やマスク顔を正面顔と同様に高精度に検出することができる。さらにマスク顔をマスク顔として識別することによって器官点検出などの後段の処理をマスク顔に対応した処理を実施することが可能となる。処理結果の例を図6および図7に示す。

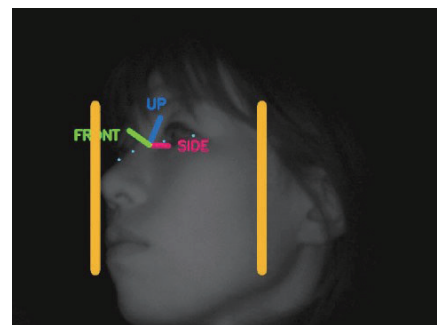


図6 横顔に対する顔検出結果
（同意書により画像利用許諾確認済）

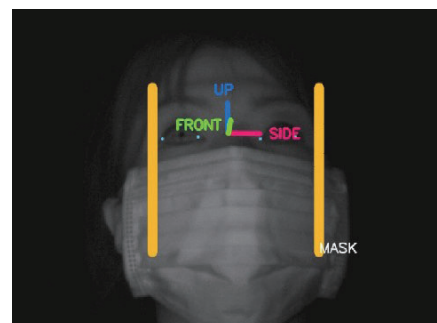


図7 マスク装着時の顔検出結果
（同意書により画像利用許諾確認済）

5.3 ネットワークの共有による出力の効率化 実用化を考える上で、計算コストおよびメモリ消費量を考慮する必要がある。提案した3指標を出力する上で、それぞれ最適化されたニューラルネットワークを構築すると、処理時間として単純に1ネットワークの3倍の時間を要する。そこで、ネットワークの中間層までを共通化し、最終出力の段階で3出力を並列に出力する仕組みを導入する（図8）。ここで問題となるのが3出力を同時に学習することになり、正解情報と損失の計算に困難が生じるため、ラベルの修正が必要となる。具体的には「誰も座っていない画像（Seated-off）はEyes、Readinessでは損失を計算しない」といったルールを追加することになる。

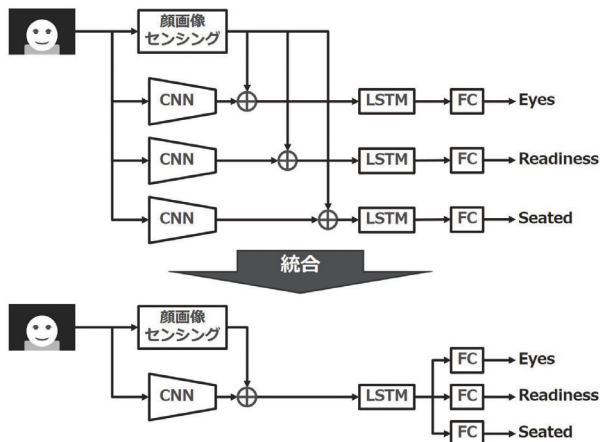


図8 ネットワークの共通化

ラベルが修正された正解情報をもとに学習を実施したところ、その前後で大きな性能変化は生じなかったことから、共通化したモデルを採用した。

5.4 高速化の検討 計算資源が潤沢でないシステムにて提案手法を構築する場合、大規模なNeural Networkをそのまま搭載することは計算コストおよびメモリ消費量双方の観点で困難を生じる。一般的には入力する画像を圧縮することで必要となるネットワークの要素数が大幅に削減できるため、上記課題を克服できる。しかし、単純に画像を圧縮すると写っているドライバの顔の情報が損なわれる。居眠り検知や脇見検知といった技術が提案されたことから、顔をセンシングすることによってドライバの状態推定に必要な情報が獲得できることを示している。

そこで、動き特徴についての解像度のみ下げ、顔画像センシングに利用する画像の解像度は元画像と同じとすることで、顔から推定できるドライバの高次元情報を用いて低解像度画像でも高精度なレベルの推定が可能となる。また、低解像度化によりネットワークのパラメータ数の大幅な削減、処理の高速化が期待できる。筆者らが行った実験⁵⁾では、解像度を下げた状態での学習および評価を実施したところ、720×480から24×18まで解像度を下げても5 pt程度の劣化しか起こらないことが確認されているため、本稿の提案手法においても、解像度を24×18に圧縮としている。

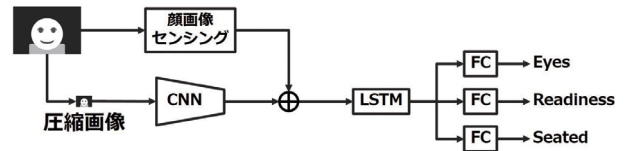


図9 圧縮画像を利用したネットワーク構成

6. 将来の展望

今後の展望としてはカメラから得られる画像と生体情報および周辺の交通情報を組み合わせることでより深くドライバ状態をセンシングすることを目指す。画像センシング技術は非接触での計測が可能のため、ドライバへの負担なくドライバの表情・動作等の表出する現象の認識は可能だが、内的な状態を計測することは困難である。たとえば、ドライバは正面を向いているが、考え事をしており走行環境の監視が不十分である場合でも、今回の運転集中度センシングではEyes-onという出力がなされる。また、右左折時や追い越し時の安全確認といった、画像からだけでは判別できない適切な運転行動を網羅できていない。

生体情報を利用することによってこれら運転上のリスクを大きく削減することが可能になると考える。たとえば脈拍の計測や表情推定を組み合わせることで眠気の予兆を検知することで、ドライバが実際に眠くなる前に警告を促し、運転の引継ぎを安全に実施可能となる。また、体調の変化などを検知し休憩を促すことで、よりドライバの負担を調整することが可能となる。

また近年開発が盛んな外界の周辺監視センサと融合することでドライバの視線の先にある対象を認識することが可能となる。これにより、運転行動が適切であるかの判断の確度を向上させることができるほか、走行中のドライバへの交通情報の提供など、より安全な運転を心がけたフィードバックが可能となる。

7. むすび

本稿では安全かつ円滑な自動運転の実現を目指して、画像列を入力とし、自動運転時におけるドライバの状態に着目し、ドライバが運転に対して責任が負える状態かどうかを推定するドライバモニタリング技術を提案した。提案手法ではCNNとLSTMを組み合わせ、画像のみではなく顔画像センシング結果を使用することで高精度に識別できることが確認できた。今後は一層の性能向上とともに生体情報・周辺監視センサなどの情報を加味した、より深いドライバ状態の推定を開発していく。

参考文献

- 1) 内閣官房IT総合戦略室：官民ITS構想・ロードマップ2017
<https://www.kantei.go.jp/jp/singi/it2/dai71/siryou3-2.pdf>
- 2) 矢野経済研究所：“2030年の自動車技術展望”（2017）

- 3) 国土交通省: “自動運転に係る国際基準の動向” (2016)
- 4) 株式会社デンソー: “ドライバーステータスモニター”, <http://www.denso.co.jp/ja/news/newsreleases/2014/140403-01.html>
- 5) 富士通株式会社: “FEELytm (フィーリズム) ”, <http://www.fmworld.net/biz/uware/ve31/>
- 6) J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko and T. Darrell: “Long-term recurrent convolutional networks for visual recognition and description”, CoRR, abs/1411.4389, (2014) .
- 7) A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar and L. Fei-Fei: “Large-scale video classification with convolutional neural networks”, CVPR (2014) .
- 8) 日向、木下、青位、濱走、山下、藤吉、諏訪、川出: “時系列 Deep Learning を用いたドライバの運転復帰可否レベル推定”, 画像の認識・理解シンポジウム (MIRU) (2017) .
- 9) Y. Lecun, L. Bottou, Y. Bengio and P. Haffner: “Gradient-based learning applied to document recognition”, Proceedings of the IEEE, 86, 11, pp. 2278-2324 (1998) .
- 10) S. Hochreiter and J. Schmidhuber: “Long short-term memory”, Neural Comput., 9, 8, pp. 1735-1780 (1997) .
- 11) Preferred Networks: “Chainer: A flexible framework of neural networks”, <http://chainer.org/>.

執筆者紹介



日向 匡史 Tadashi Hyuga
技術・知財本部 センシング研究開発センタ
専門: 画像処理、パターン認識
所属学会: IEEE



木下 航一 Koichi Kinoshita
技術・知財本部 センシング研究開発センタ
専門: 画像処理
所属学会: 電子情報通信学会、情報処理学会



西行 健太 Kenta Nishiyuki
技術・知財本部 センシング研究開発センタ
専門: 画像処理、パターン認識
所属学会: なし



長谷川 友紀 Yuki Hasegawa
技術・知財本部 センシング研究開発センタ
専門: 画像処理
所属学会: 精密工学会