

自動車運転者の行動分析のための注視情報抽出法

A New Method of Fixation Identification for Driving Behavior Analysis

針ヶ谷勇星^{1*} 白山晋¹
Yusei Harigaya¹ Susumu Shirayama¹

¹ 東京大学大学院工学系研究科

¹ School of Engineering, The University of Tokyo

Abstract: Determining the starting time, duration, and target of drivers' fixation is useful for driving behavior analysis. However, this information is currently extracted from raw gaze data by hand or by algorithms that are not suitable for mobile eye tracking. In this paper, we firstly propose a novel method for extracting drivers' fixation information that is robust to drivers' head movement utilizing spatial and temporal features of drivers' foveal vision. Then we introduce the gaze data processing system that semi-automatically extracts drivers' fixation information based on our proposed method. It is shown that the proposed method outperforms previous methods in terms of both validity and efficiency. The result indicates that our system is effective in driving behavior analysis.

1 はじめに

日本では高齢運転者による交通事故が増加しており、社会問題となっている [1]。高齢運転者の事故率が高い原因として心身機能の低下が挙げられるが、その程度には大きな個人差があるため、個々の高齢者自身にどのような運転上の問題があるかを把握し、一人ひとりの特性に応じた指導や支援を行うことが必要となる。Palazziら [2] は、このような個人に即した支援にあたっては、運転者が運転中に受容する情報を分析し運転者個人の人認状態を推定することで、より効果的な支援が可能であるとしている。

このような背景から、運転者が受容した情報をもとに彼らの行動分析を行うことで、運転者の行動分析に基づく運転者への助言・指導や運転支援システムの構築に活用できるような知見を得ることが求められている。運転行動は一般的にその 90% が受容した視覚的情報に基づいてなされているため、運転者に対して視線計測を行うことで、彼らが受容した視覚的情報を分析することができると考えられている。藤本ら [3] は、熟練運転者が注目した領域をドライブレコーダーの映像に重ねて表示することで熟練運転者が受容したと思われる情報を可視化し、これを初心運転者に見せることで、初心運転者による映像中の危険因子発見率が有意に向上したと述べており、運転者の注意機構が視線移動に反映されていることが示唆されている。Martin

ら [4] は、運転者がどのような物体および領域を注視したかを分析することで、運転者の状況理解度の把握や次に行う運転行動の予測を行うことが可能であると述べている。このため、運転中の視線計測データを分析する際は運転者の注視情報に着目して分析を行うことが有用であり、様々な形で研究が行われている。

Martinら [4] は、注視情報として以下の3つを挙げている。

- 注視開始時刻
- 注視終了時刻
- 注視対象の属性

すなわち、注視情報の抽出を行うためには、まず各時刻における視線データを注視状態であるか否かの2つに分類し、その後注視対象を識別することが必要である。注視状態の分類手法は広く研究が行われており、代表的な手法に視線移動速度に閾値を設けることによって分類を行う I-VT [5] がある。しかしながら、従来手法は頭部を固定した状態でスクリーンと相対するような特殊な実験室環境下で計測されたデータに対して適用することを前提としており、実環境、とりわけ運転中のような大きな頭部運動を伴う環境において注視状態を安定して検出する手法は未だに確立されていない。

したがって、運転者の注視状態に着目した既存研究は、その多くが研究者が手作業で各視線データに対して注視状態の検出および注視対象の識別を行うものである。手作業での注視情報抽出を行う場合、運転者の注視状態を分析することは原理上可能ではあるが、視線研究の専門家でなければ作業が非常に困難である上

*連絡先：針ヶ谷 勇星，東京大学大学院工学系研究科システム創成学専攻 白山研究室，文京区本郷 7-3-1
E-mail: yharigaya429@gmail.com

に、膨大な視線データ一つひとつを手作業で処理することは非常に高コストであり大規模な分析は非現実的である。したがって、運転者の注視状態の分析を低コストかつ非専門家によって行うことを可能にするためには、実環境下における高精度な注視状態の検出、および注視対象の識別を自動で行う必要がある。

以上を踏まえ本研究では、自動車運転者の注視情報抽出において課題となる

- (i) 自動車運転時のように頭部運動の大きい環境下での注視状態の機械的な検出が困難である
- (ii) このような環境下における視線データから、注視している対象物を機械的に識別することが困難である

の2点を解決する注視情報抽出法を提案する。また、提案した注視情報抽出法に基づく注視イベント抽出システムを構築する。これらによって、運転者の行動分析の精度向上と効率化に寄与する。

2 関連研究

2.1 注視状態検出手法の研究

先述した、頭部運動に対して頑健な注視状態検出手法が確立されていない課題の解決を目的とするいくつかの研究を紹介する。

Hossain と Miléus [6] は、頭部装着型視線計測装置に搭載されたジャイロセンサ、加速度センサからの出力の積分値から頭部運動を推定し視線位置の補正を行うことでI-VT [5] の適用を試みている。彼らの検証実験ではI-VT の適切な角速度閾値を絞り切れておらず、この原因としてセンサの出力値のノイズによって頭部運動の推定が正しく行えていない可能性が言及されている。

Steil ら [7] は、視野カメラ映像中の隣接するフレームの視線位置周辺の画像を切り出し、これをCNNに入力することで2つの画像の類似度を定量化し、この類似度の値に閾値を設けることで注視状態検出を行う手法を提案している。視線位置の座標値でなく画像特徴を用いることで理論的に頭部運動に頑健な手法となっているが、類似度計算を行った後閾値処理を行うという完全に独立した2段階からなるため、フロー全体での最適化がなされていない。また、彼らが用いたデータセットは非公開となっている。

2.2 運転者の視線分析に関する研究

Sun ら [8] は、運転中の高齢運転者に対して視線計測を行い、視線位置の空間的広がりから視線移動パターンを分類することで、事故に繋がりがやすい視線パター

ンを特定する手法を提案している。彼らは、分析の結果認知負荷が高いとされる特定の状況において特有の視線行動がみられると述べているが、分析には視線位置の座標値のみを用いており、運転者の注視した内容を考慮した分析を行っていないため、議論の妥当性を示すことが困難である。

Tawari ら [9] は車内に設置したカメラで運転者の顔を撮影し、映像に含まれる顔特徴点の情報を入力として、予め分割した車内の領域のいずれに視線が向いているかの多クラス分類を行い、運転者の注視対象を試みている。Tawari らの手法は、運転者の前方に設置したカメラのみを用いる非侵襲的なシステムである点で優れているが、広い領域内のどこが注視されているかが認識できず、また先行車、歩行者など車外環境に存在する対象への注視を検出することは想定されていないため、運転中の注視行動を分析するに足る情報を得ることができているとはいえない。

3 提案手法

関連研究の課題を踏まえ、視線計測データから自動で注視情報を抽出する手法を提案する。運転者の視線計測データとして視野カメラ映像、車載カメラ映像、視線座標データの3つが入力されると、出力として注視開始時刻、注視終了時刻、注視対象からなる注視情報が抽出される。図1に提案手法のフローを示す。

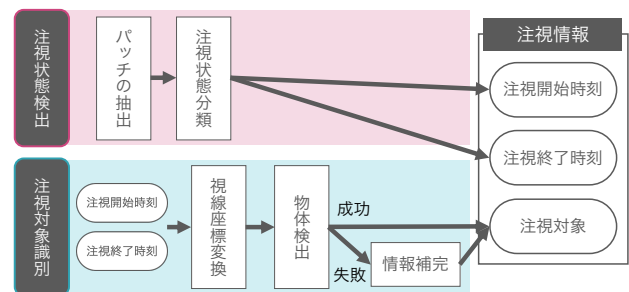


図1: 提案手法のフロー

提案手法は、以下の2つの段階に大別される。以下では、それぞれの詳細を述べる。

3.1 注視状態検出

はじめに、視線計測データから運転者が注視を開始した時刻および終了した時刻の情報を抽出する。

3.1.1 視線座標欠損データの削除

視線計測データには計測装置のデータ計測漏れおよび被験者の瞬目による欠損値が含まれるため、本研究ではこのような異常データを削除して分析を行う。具

体的には、ある視線座標データ $\mathbf{Q} = \{q_1, q_2, \dots, q_N\}$ について、欠損データが $q_k, q_{k+1}, \dots, q_{m-1}, q_m$ であった場合、これら $m - k + 1$ 個のデータを削除し、その前後に存在する $\mathbf{Q}_{\text{before}} = \{q_1, q_2, \dots, q_{k-1}\}$ および $\mathbf{Q}_{\text{after}} = \{q_{m+1}, q_{m+2}, \dots, q_N\}$ をそれぞれ独立した時系列として保持する。

3.1.2 視野カメラ映像からのパッチ抽出

次に、視線計測データの座標情報をもとに、視野カメラ映像の対応するフレームから視線位置周辺のパッチを抽出する。パッチ抽出の概念図を図2に示す。注視状態にあるときには、視線位置周辺の画像特徴は変化が少ないと考えられる。これを利用し、ある時間内の視線位置周辺の画像特徴の変化を入力として学習器により注視状態検出を行うことが本手法の着想である。

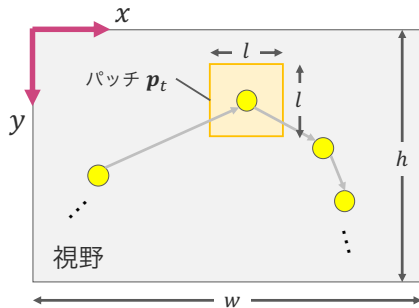


図2: 視野カメラ映像からのパッチ抽出の概念図

抽出するパッチは、視線座標 (x, y) を中心とする一辺 l の正方形領域とする。ここで、 l の値はヒトの中心視野領域に相当するサイズに周辺視野の影響を考慮したマージンを設けたものとする。本研究では中心視野領域の視野角 2° に対して 4° の範囲の画像を抽出した。

パッチ抽出は時系列内のデータのそれぞれについて行い、各時系列内の視線座標データ $\mathbf{Q} = \{q_1, q_2, \dots, q_n\}$ に対してパッチデータ $\mathbf{P} = \{p_1, p_2, \dots, p_n\}$ が作成する。各パッチを注視状態検出部への入力サイズにリサイズした後、 n 個のパッチの各ピクセルの輝度値の平均と分散を求めることにより、各パッチの輝度値の正規化を行う。

3.1.3 各時刻における注視状態分類

抽出したパッチデータを入力として注視状態分類、すなわち注視状態であるか否かの分類を行う。図3に、本研究で提案する注視状態検出部の構造を示す。本研究では Donahue ら [10] の LRCN の構造を参考にした。具体的には、画像特徴を抽出する CNN に、時系列処理を行う LSTM を接続し、最後に全結合層によって注視状態・非注視状態の二値分類を行う学習器を作成する。ここで、LSTM 層としては Bidirectional LSTM [11] (以下 BLSTM) を 2 層に重ねたものを用いる。これによ

り、予測サンプルの未来の情報を入力に加えた予測が可能となる。

はじめに、パッチデータからウィンドウ幅 s のデータ $\mathbf{P}_{\text{input}} = \{p_t, p_{t+1}, \dots, p_{t+s-1}\}$ を取得する。次に、ウィンドウ内の各パッチを CNN を用いて畳み込み、これにより得られた特徴量 $\mathbf{F} = \{f_t, f_{t+1}, \dots, f_{t+s-1}\}$ を BLSTM に入力することで、ウィンドウ内のある時刻における注視状態の二値分類 (注視: 1, 非注視: 0) を得る。

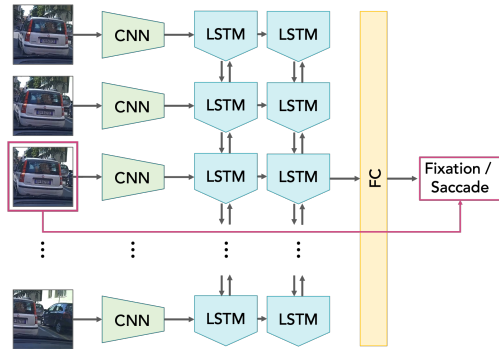


図3: 注視状態検出部の構造

3.1.1 項から 3.1.3 項までの手順によって、全視線計測データは注視状態・非注視状態のいずれかに分類される。このうち、連続して注視状態であると分類されたデータ群それぞれを注視イベントと称する。各注視イベントに属する最初および最後のデータに対応する時刻がそれぞれ注視開始時刻、注視終了時刻となる。

3.2 注視対象識別

次に、上述の手順で検出された各注視イベントにおける運転者の注視対象を識別する。視線位置周辺において物体検出を行うことで注視対象の候補を求め、この中から注視対象を決定する。

3.2.1 車載カメラにおける視線座標の推定

はじめに、注視イベントを構成する各データについて、車載カメラにおける視線座標を推定する。運転者が視野の周縁部を注視している可能性を考慮すると、注視対象の検出は視野カメラよりも画角が広い車載カメラ映像上で行うことが望ましいと考えられる。したがって、視野カメラ座標系における視線座標 (x_s, y_s) から車載カメラ座標系における視線座標 (x_c, y_c) への変換を行う必要がある。

まず、視野カメラ映像、車載カメラ映像それぞれの対応するフレームから AKAZE [12] により画像特徴点を検出する。次に、k 近傍法により双方の画像から検出された特徴点同士のマッチングを行う。一般的に、このマッチングには誤対応が含まれるため、RANSAC [13] を

用いて誤対応の除去を行う。最後に、誤対応が除去された対応点を用いて2画像間の透視変換行列を推定し、これを用いて視線座標を車載カメラ座標系に射影する。

3.2.2 各注視イベントにおける注視対象の識別

車載カメラ座標系に射影された視線座標を用いて、その時刻の車載カメラ映像中に存在する物体の検出を行う。提案手法では視野内の画像特徴から注視状態検出を行うため、各注視イベント中の視野はそれぞれ類似した画像特徴を持っていると考えられる。したがって、運転者は各注視イベントにおいて同一の対象を注視していると推察される。そこで本研究では、各注視イベントに対して注視対象を一つ対応付ける。

まず、車載カメラ映像の各フレームを入力として物体検出を行う。出力された物体リストの中から物体領域の範囲内に視線位置座標を含む物体を注視対象候補として保持する。この操作を当該注視イベント中に含まれる全フレームに対して行い、注視対象候補として最も多く挙げられた物体クラスを注視対象とする。ただし、大半のフレームで物体が検出されなかった場合などには、対象の識別結果が妥当でない可能性がある。このような場合に対処するため、データ数 n に対して閾値 n_0 を超える数のデータに同一の物体が対応付けられた場合のみその物体を注視対象とすることで、注視対象として妥当なもののみを記録する。本研究では $n_0 = 0.5n$ とした。なお、物体クラスのラベルとしては、ドライブレコーダー映像を用いた物体検出データセットである BDD100k [14] のラベルに、背景として運転者が注視し得る建物・道路・空・植物の4種および車内空間を示すラベルを加えて用いる。

本研究では、物体検出手法として M2Det [15] の著者が公開している学習済みモデルを用いる。

3.2.3 注視対象情報の補完

物体検出に失敗した場合や、注視対象を表すラベルがデータセットに含まれない場合等は上述の手順では注視対象を識別することができない。このような場合には、例外処理として運転者の注視対象を人間が判断し情報を補完することとする。

3.3 注視イベント検出システム

本研究では、提案手法に基づく運転者の注視イベント抽出システムを、上述の補完作業を行う作業者が操作するアプリケーションとして実装した。実装は、Unityを用いてスタンドアロンな Windows アプリケーションとして行った。図4に本研究で実際に実装したアプリケーションの作業画面を示す。

3.1.1 項から 3.2.2 項に示した処理が機械的に行われる。ある注視イベントに対して自動で注視対象の識別ができなかった場合、作業者にはアプリケーションの



図4: 注視イベント抽出システムの作業画面

画面上で当該イベント時間内の視野カメラ映像、車載カメラ映像および視線位置周辺の映像を拡大した映像の3つが提示される。各映像中には、視線位置の座標が黄丸で示されている。

補完作業は以下の手順により行われる。

1. 注視イベント中の映像の確認。作業者は、まず図4に示した画面左下の領域にある再生ボタンを押下し、1つの注視イベントの時間内の映像を確認する。この映像は、再度再生ボタンを押下することで繰り返しの確認が可能である。
2. 注視対象の決定。映像の確認が終了した後、作業者はその映像中で運転者が注視していると思われる対象を示すラベルを1つ選び、対応するボタンを押下する。

作業者は、以上の手順を 3.2.2 項の処理において物体検出が成功しなかった全ての注視イベントについて繰り返す。この作業により、3.1.3 項の処理において検出された全ての注視イベントに対して注視開始時刻、注視終了時刻、注視対象の3つの情報が視線計測データから抽出される。

3.4 視線計測データの事前処理手法

本研究では、注視状態検出部の精度向上を図るため、入力となる視線計測データに適用する事前処理手法として以下の2つを合わせて提案する。

3.4.1 視線座標データの平滑化

一般に、視線計測装置の計測誤差は 0.5° 程度とされ、視線位置の座標にはデータ点ごとに細かなずれが発生する。このずれによりパッチの画像特徴が変化することで、注視の最中の視線位置周辺の画像特徴が類似することを利用した提案手法の精度を低下させている可能性が考えられる。

この計測誤差の影響を低減するため、視線座標データに対してガウシアンフィルタによる平滑化を行う。本

研究では、 x 座標、 y 座標それぞれのデータに前後 1 近傍の 1 次元ガウシアンフィルタを畳み込む平滑化を行った。

3.4.2 訓練データの選択的利用

人間の視線行動においては、平均 200 ms~300 ms の注視状態のあと 30 ms~80 ms の非注視状態が発生する、ということが繰り返される。注視対象の複雑度によっては注視状態がさらに長く続くことも考慮すると、視線計測データは一般に注視状態のサンプル数が非注視状態のそれより顕著に多い不均衡データであるといえる。

このデータの不均衡を低減するため、多数派クラスである注視状態のデータの削減を行う。具体的には、視線計測データからあるウィンドウサイズで切り出される、注視状態検出部への入力の時系列を以下の 4 種類に分類する。

- (a) ウィンドウに含まれるデータ全てが注視状態
- (b) ウィンドウに含まれるデータ全てが非注視状態
- (c) 両方の状態が含まれ、予測サンプルが注視状態
- (d) 両方の状態が含まれ、予測サンプルが非注視状態

このうち、後述する 2 つのデータセットのいずれも (a) にあたるウィンドウを全体の 6 割以上含んでいた。(a)(b) のように全サンプルが同一ラベルのものは分類容易であると推察されるので、学習困難な例を残すために (a) にあたるウィンドウのうち 75 % をランダムに削除したものを訓練データとして用いる。ただし、この操作により訓練データとテストデータでクラス確率に際が生じ、テストデータに対する予測において非注視状態の事前確率が実際より大きくなると考えられる。したがって、以下の式を用いて実際の学習器による予測確率 p_s を本来の確率である p に補正する。

$$p = \frac{p_s}{p_s + \frac{(1-p_s)}{\beta}} \quad (1)$$

ここで、 β は注視状態のデータのうち、訓練データとして選択される確率である。

4 注視状態検出の評価実験

3.1.1 項から 3.1.3 項に示した注視状態検出手法の評価実験を、以下の記述に従って行った。

4.1 使用したデータセット

4.1.1 GazeCom

GazeCom [16] は予め撮影した日常の映像を被験者に見せた際の視線計測データである。被験者はディスプレ

イに正対しそこに映された映像を見るため、常にディスプレイ方向を向いており頭部運動は発生しない。データセットは 18 種の映像を平均 47 人が見る計 477 試行の視線計測データから構成されており、各試行の長さはそれぞれ約 20 秒である。各サンプルは (Fixation, Saccade, Pursuit, Noise) のいずれかにラベリングされているが、本研究で扱う注視状態の定義より、Fixation と Pursuit を合わせたものを注視状態のラベル、Saccade を非注視状態のラベルとして処理した。また、視線計測データは 250 Hz で記録されているが、一般的な頭部装着型視線計測装置の時間分解能に合わせ 50 Hz にダウンサンプリングして用いた。

4.1.2 DR(eye)VE

DR(eye)VE [17] は運転中の運転者に対して行った視線計測結果のデータセットである。被験者は頭部装着型視線計測装置を着用しており、運転中の左右・後方確認等によって大きな頭部運動が発生している状況を含む。計 74 試行の視線計測データから構成されており、各試行の長さはそれぞれ約 300 秒である。各サンプルは (Fixation, Saccade, Noise) でラベリングされているため、Fixation を注視状態のラベル、Saccade を非注視状態のラベルとして処理した。データは 60 Hz で記録されている。

4.2 実験設定

本実験では、注視状態検出部の CNN として畳み込み層 4 層からなる簡素な構造を使用した。各畳み込み層はバッチ正規化、プーリングの操作を含む。活性化関数には ReLU を用い、プーリングのウィンドウサイズは 2×2 とした。入力となるパッチのサイズは 64×64 ピクセルとし、ウィンドウサイズは 100 ms に相当する長さとした。バッチサイズ 128 のミニバッチ学習を行い、epoch 数は学習状況を考慮し 20 とした。最適化には Adam を用い、学習率は 1.0×10^{-3} より始め、学習が 50 % 経過した段階で 1.0×10^{-4} とした。

提案する注視状態検出部 (CNN-BLSTM) は、ウィンドウ内の最後の時刻のサンプルを予測するもの (CNN-BLSTM_ls)、中央の時刻のサンプルを予測するもの (CNN-BLSTM_md)、視線座標データの平滑化を行うもの (CNN-BLSTM_lm)、訓練データの削減を行うもの (CNN-BLSTM_ds) の 4 つを比較した。前処理を行う 2 つに関しては、ウィンドウの中央の時刻を予測する構造を使用した。比較手法として、注視状態検出で一般的に用いられる手法である I-VT と、頭部運動に頑健な手法として提案されている Steil ら [7] の手法を用いた。

各手法の性能評価はそれぞれのデータセットを 6 分割することで $k = 6$ の交差検証を行い、テストデータに対する分類結果について評価指標を算出することに

よって行った。評価指標としては注視状態および非注視状態それぞれに対する F 値, Cohen の κ 係数, 各手法により検出された注視状態の継続時間分布を用いた。

4.3 実験結果

4.3.1 数値実験 1 : GazeCom

はじめに, 頭部運動の存在しないデータセットである GazeCom について注視状態検出を行った結果を考察する。表 1 に, GazeCom データセットで評価を行った結果を示す。紙面の都合上, 評価指標として F 値のみを示す。

表 1: GazeCom データセットにおける評価指標

	F 値 (注視)	F 値 (非注視)
CNN-BLSTM_ls	0.975 ± 0.002	0.785 ± 0.011
CNN-BLSTM_md	0.977 ± 0.002	<u>0.811 ± 0.010</u>
CNN-BLSTM_sm	0.966 ± 0.006	0.772 ± 0.023
CNN-BLSTM_ds	0.977 ± 0.003	0.808 ± 0.009
Steil et al. [7]	0.966 ± 0.004	0.709 ± 0.038
I-VT	<u>0.977 ± 0.001</u>	0.808 ± 0.006

表 1 より, 提案手法と I-VT はほぼ同等の性能を示し, Steil らの手法はこれらに劣る結果となった。GazeCom は被験者の頭部運動が発生しないデータセットであるため, 頭部運動に対応できない I-VT でも高い精度で注視状態検出が可能であったと考えられる。提案手法は被験者の頭部運動に頑健な注視状態検出を目的とする手法であるが, 頭部運動の発生しない状況においても, そのような状況に特化した I-VT と同等の精度で検出が可能であることが示唆された。これに対して, Steil らの手法も提案手法と同じく頭部運動に頑健な検出を目的とするものであるが, 頭部運動の発生しない状況において I-VT よりも低い精度であることから, 様々な状況で用いることのできる汎用的な手法ではないと考えられる。

また, 提案手法ではウィンドウの中央の時刻の予測を行った場合に, 最後の時刻を予測した場合よりも高い精度を示し, 特に非注視状態の F 値に大きな改善がみられた。中央の時刻の予測を行った場合, 最後の予測の場合と比較して入力に含まれる過去の情報量が半分程度となるが, この影響を踏まえても注視状態検出では予測時点より未来の情報を入力に含めることがより重要であることが示唆された。これは, 特に非注視状態において隣接するサンプル間で画像特徴が大きく異なることから, この変化の前の情報だけでなく変化後の情報を考慮することが精度改善に大きく寄与するからであると考えられる。

次に, 提案した訓練データの前処理手法の効果を考察する。まず, 提案手法において視線座標データの平

滑化を行ったものにおいては, わずかな精度の低下がみられた。この原因としては, 座標の平滑化によって注視状態と非注視状態の境界部分で切り出されるパッチの画像特徴の変化が低減されてしまい, この部分の分類性能が劣化した可能性が考えられる。一方で, 訓練データを間引いて学習を行った場合は, 訓練データの選択を行わない場合とほぼ同等の性能となった。訓練データ量が約半分になっていることを鑑みると, これは提案した訓練データ選択手法によって分類困難なデータを残して学習ができたことを示唆している。

4.3.2 数値実験 2 : DR(eye)VE

次に, 頭部運動の大きい状況を含むデータセットである DR(eye)VE について注視状態検出を行った結果を考察する。表 2 に, DR(eye)VE データセットで評価を行った結果を示す。こちらも同様に評価指標として F 値のみを示す。

表 2: DR(eye)VE データセットにおける評価指標

	F 値 (注視)	F 値 (非注視)
CNN-BLSTM_ls	0.916 ± 0.005	0.419 ± 0.035
CNN-BLSTM_md	<u>0.926 ± 0.007</u>	0.566 ± 0.028
CNN-BLSTM_sm	0.922 ± 0.007	0.521 ± 0.035
CNN-BLSTM_ds	0.917 ± 0.011	<u>0.602 ± 0.042</u>
Steil et al. [7]	0.871 ± 0.016	0.439 ± 0.037
I-VT	0.899 ± 0.017	0.551 ± 0.042

表 1, 表 2 を比較すると, DR(eye)VE では全体として各手法の分類能力が低い。これは, DR(eye)VE が GazeCom よりも分類困難なデータセットであったことが原因であると考えられる。

GazeCom の分類において提案手法と同等以上の性能を示した I-VT は, DR(eye)VE の分類においては提案手法より顕著な精度の低下が見られた。これは, GazeCom には含まれなかった被験者の頭部運動が大きい状況のデータが DR(eye)VE に少なからず含まれることが原因であると考えられ, I-VT に代表される従来手法では頭部運動の大きい状況において適切に注視状態検出を行うことが困難であることが示唆された。また, このような状況においては提案手法が従来手法よりも高精度の検出を行うことができたために, 4.3.1 項の結果と合わせて考えると, 提案手法によって, 頭部運動の有無にかかわらず頑健な注視状態検出が可能であると示唆された。

提案手法において, 視線座標データの平滑化を行った場合, 訓練データを間引いて学習を行った場合ともに通常より精度が僅かに低下する結果となった。しかしながら, 訓練データを間引いた場合には少数派クラスである非注視状態の F 値が改善された。したがって, 提案した訓練データの選択手法が, 運転者の視線計測

データの注視状態および非注視状態の不均衡の是正に有効であることが示唆された。他の場合と比較して各指標の標準偏差は増大しているが、これは訓練データの選択により約半分の訓練データ量で学習を行ったことが主因であると考えられる。そのため、データ削減前と同程度の規模の訓練データを用いて学習を行うことでより安定した学習が可能になると推察される。

5 注視イベント抽出の評価実験

5.1 実験概要

本実験では、提案した注視イベント抽出システムの有効性を示すため、従来行われてきた手作業による注視イベント抽出（従来手法）と作業の効率性や妥当性を比較する。抽出作業を行うにあたって、被験者がそれぞれの手法で作業を行う際の UI の差異による影響を可能な限り低減するため、手作業による注視イベント抽出についても提案手法と同様に作業用アプリケーションを実装した。図 5 に手作業用アプリケーションの作業画面を示す。



図 5: 手作業による注視イベント抽出の作業画面

作業者は、作業対象のデータ範囲の映像をコマ送りで確認し、各フレームで示されている視線位置の変化を確認することで対象への注視が開始したフレーム、終了したフレーム、その間の運転者の注視対象を判断し、これらの情報をアプリケーションに入力する。

5.2 実験設定

実験の被験者は、正常な視力を有する男子大学生・大学院生計 6 名とした。6 名全員が普通自動車運転免許を所持し、年に数回以上の頻度で運転を行っていた。被験者は、提案システムによる注視イベント抽出作業および手作業により 3 つの注視情報全てを抽出する作業をそれぞれ一度ずつ行った。実験結果として、各試行における所要時間および抽出された注視情報が記録された。実験データとして、DR(eye)VE から注視状態・

非注視状態の切り替わりのタイミングを多く含む 300 フレーム（5 秒）の範囲のデータを抽出したものを 3 種類用意しこれを用いる。

5.3 実験結果

はじめに、提案手法に基づくシステムの効率性を評価する。実験の結果、各被験者が提案手法および従来手法による作業に要した時間を表 3 に示す。

表 3: 注視イベント抽出作業の所要時間平均 [s]

提案手法	従来手法
221.2 ± 95.2	683.5 ± 242.0

各被験者の提案手法による作業時間と従来の手作業による作業時間の対応をみるため対応あり t 検定を行ったところ、2 つの手法による作業時間の分布に有意差がみられた ($p = 0.006$)。したがって、提案手法による作業で手作業よりも有意に短い時間で作業を完了させることができることが示唆された。また実験後、各被験者が作業前後の疲労度の差を 10 段階で主観評価した結果、提案手法において有意に疲労度が小さくなった ($p = 0.02$)。したがって、提案手法に基づくシステムでは作業者の負担が軽減されることが示唆された。以上の考察より、提案手法による作業は所要時間及び作業者への負担の両面において従来の手作業よりも優れており、より効率的な手法であると考えられる。

次に、システムによる抽出結果の妥当性を評価する。2 つの手法による注視イベント抽出結果から、視線計測データに対して注視状態・非注視状態の分類結果が得られる。これを DR(eye)VE の教師ラベルと比較することで、抽出結果の評価指標 (F 値) を算出することができる。表 4 に、注視情報抽出結果の評価指標を示す。

表 4: 注視イベント抽出結果の評価指標

	F 値 (注視状態)	F 値 (非注視状態)
提案手法	0.925 ± 0.002	0.689 ± 0.048
従来手法	0.798 ± 0.069	0.376 ± 0.189

表 4 に示す通り、提案システムによる分類が全ての指標で従来の手作業による分類を上回り、提案手法による分類結果が従来手法より妥当なものであることが示唆された。従来のような手作業での注視イベント抽出は通常専門家によって行われるため、実際は両手法による結果の妥当性の差は本実験よりも小さくなることが予想される。しかしながら、提案手法を用いることによって非専門家でも従来より妥当に注視イベント抽出を行うことが可能となるため、運転者の注視情報をより幅広い機会を活用できると考えられる。

6 おわりに

本研究では、従来手作業で行われてきた運転者の視線計測データからの注視情報の抽出を自動化することで視線計測データを運転者の行動分析や運転者の支援システムで活用可能にすることを目的として、自動車運転者の注視情報抽出法を提案した。また、提案手法に基づき運転者の視線計測データから半自動で注視情報抽出を行うシステムを提案・実装し、その有効性を検証した。

評価実験の結果、提案した注視状態検出手法は視線計測データにおける頭部運動の有無にかかわらず従来手法以上の精度で注視状態・非注視状態の分類が可能であることを示した。また、本研究で提案した注視情報抽出法による作業は、従来のように完全に人手で情報の抽出作業を行う場合と比較して非常に短時間で完了するものであり、かつ作業への負担も有意に小さいことを確認した。抽出された情報に関しても、従来手法による情報抽出結果と比較してより実際の視線行動に近い結果が得られ、提案手法によって従来の作業より妥当な情報抽出が可能であることが示唆された。

以上の結果から、提案した注視情報抽出法および注視イベント抽出システムを用いることによって、自動車運転者の注視情報を従来の手作業と比較して短時間、低負担、高精度で抽出することが可能であり、本研究は運転者の注視情報を広く活用可能な情報にすることに貢献すると考えられる。

今後の展望としては主に次の2点が挙げられる。まず、より高性能な画像特徴抽出器を用いて注視状態検出部の学習を行うことで、より高精度での分類が可能になると考えられる。次に、運転者の視線がある領域に停留している際に、運転者がそれを注視しているのか、視野を俯瞰しているのかを区別して識別することによって、運転者の認知過程をより妥当に推定することが可能であると考えられる。

参考文献

- [1] 平成 29 年交通安全白書 (概要) - 内閣府. https://www8.cao.go.jp/koutu/taisaku/h29kou_haku/gaiyo/features/feature01.html. (最終閲覧日: 2019/11/12). 2017.
- [2] A. Palazzi, F. Solera, S. Calderara, S. Alletto, and R. Cucchiara. Learning Where to Attend Like a Human Driver. *2017 IEEE Intelligent Vehicles Symposium (IV)*. 2017, pp. 920–925.
- [3] 藤本武司, 砂山渡, 山口智浩, 谷内田正彦. 視線行動の可視化による着眼スキル伝達支援. *人工知能学会論文誌* 19 3 (2004), pp. 174–183.
- [4] S. Martin, S. Vora, K. Yuen, and M. M. Trivedi. Dynamics of Driver's Gaze: Explorations in Behavior Modeling and Maneuver Prediction. *IEEE Transactions on Intelligent Vehicles* 3 2 (2018), pp. 141–150.
- [5] D. Salvucci and J. Goldberg. Identifying Fixations and Saccades in Eye-tracking Protocols. *Proceedings of the Eye Tracking Research and Applications Symposium*. Jan. 2000, pp. 71–78.
- [6] A. Hossain and E. Miléus. “Eye Movement Event Detection for Wearable Eye Trackers”. Master Thesis. Linköping University, 2016.
- [7] J. Steil, M. X. Huang, and A. Bulling. Fixation Detection for Head-mounted Eye Tracking Based on Visual Similarity of Gaze Targets. *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. 23. 2018, pp. 1–9.
- [8] Q. Sun, J. (Xia, T. Falkner, and H. Lee. Investigating the Spatial Pattern of Older Drivers' Eye Fixation Behaviour and Associations with Their Visual Capacity. *Journal of Eye Movement Research* 9 6 (2016), pp. 1–16.
- [9] A. Tawari, K. H. Chen, and M. M. Trivedi. Where is the Driver Looking: Analysis of Head, Eye and Iris for Robust Gaze Zone Estimation. *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. 2014, pp. 988–994.
- [10] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 2625–2634.
- [11] A. Graves and J. Schmidhuber. Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures. *Neural Networks* 18 5 (2005), pp. 602–610.
- [12] P. Fernández Alcantarilla, J. Nuevo, and A. Bartoli. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. *Proceedings of the British Machine Vision Conference*. 2013, pp. 1–12.
- [13] M. A. Fischler and R. C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM* 24 6 (1981), pp. 381–395.
- [14] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell. BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling. *arXiv preprint* 1805.04687 (2018).
- [15] Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen, L. Cai, and H. Ling. M2Det: A Single-Shot Object Detector Based on Multi-Level Feature Pyramid Network. *Proceedings of the AAAI Conference on Artificial Intelligence* 33 (2019), pp. 9259–9266.
- [16] M. Dorr, T. Martinetz, K. Gegenfurtner, and E. Barth. Variability of Eye Movements When Viewing Dynamic Natural Scenes. *Journal of Vision* 10 10 (2010), pp. 1–17.
- [17] S. Alletto, A. Palazzi, F. Solera, S. Calderara, and R. Cucchiara. DR(eye)VE: A Dataset for Attention-Based Tasks with Applications to Autonomous and Assisted Driving. *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2016, pp. 54–60.