

Faster R-CNN を用いた前方車載カメラ画像による運転手の視線推計

深井 颯† 山崎 啓介† 本村 陽一†

国立研究開発法人産業技術総合研究所†

1. 背景

近年、多発する深刻な交通事故を背景として、車両状況や運転手の状態を観測することによる運転支援システムが数多く開発されている。例えば居眠り運転防止の為に、顔認識によるまばたき増加時のアラート技術などがこれにあたる。運転手の視線を顔画像から推計する研究は、こうした技術と同様に運転手の運転補助を目的としたものが多い。一方で、運転手の視線推計技術を運転支援に活用するには、運転手の視線に正解は存在しないという問題がある。例えば前述の居眠り検知であれば、目を閉じていなければ正解、という明確な評価指標が存在した。これに対し視線では、このように明確な正解や評価指標を事前に設定することができない。これは、状況によって人の注視すべき個所が変わってしまうためである。そこで本研究では、前方カメラ画像のみから運転手が注視する個所を推計するシステムを開発することで、このような運転支援を可能にする。

2. 構築するシステム

本研究では、Faster R-CNN[1]を改変することで、前方カメラ画像のみから運転手の視線を推計するモデルを構築する。このモデルの学習に、仮に十分に熟練したベテラン運転手の視線を使用した場合には、そのような運転手が行うであろう理想的な視線をモデルが出力することになる。この、前方カメラ画像のみから推計した理想的な視線と、実際に運転している運転手の視線を比較することにより、これまで困難であった運転手の視線推計による運転支援を可能にした。システム全体のイメージを図1に示す。

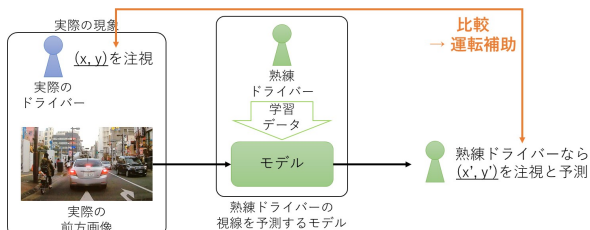


図1. システム全体のイメージ

Estimation of sightlines from images of in-vehicle camera using Faster R-CNN

† So Fukai, Keisuke Yamazaki and Yoichi Motomura
National Institute of Advanced Industrial Science and Technology

3. モデル構成

自動車の運転支援を考えると、重要なのはどの位置を見ているかではなく、どのオブジェクトを見ているかということである。そこで本研究では、どの位置を注視しているかを予測する「位置型視線推計システム」と、どのオブジェクトを注視しているかを予測する「オブジェクト型視線推計システム」の2種類のシステムを構築し、結果を比較した。両システムはそれぞれ、Faster R-CNNにより物体を検知するオブジェクト検知のモデルと、視線推計を行う視線推計モデルの組み合わせによって実現している。

3.1. 位置型視線推計システム

位置型視線推計モデルでは、Faster R-CNNによりオブジェクト検知を行うと同時に、視線推計モデルによって運転手がどの座標を注視するかを予測する。この視線推計モデルの入力は画像情報のみで、オブジェクト情報は含まれていないため、どの点を注視するかという情報のみが出力される。一方で前述のように、運転支援にはどの物体を注視しているか、という事実が重要である。そこで本システムでは、視線推計モデルで出力された注視座標と、Faster R-CNNによる物体検知の結果を比較することにより、アルゴリズム的に注視するオブジェクトを算出する。システム全体の構成を図2に示す。なお、視線推計モデルには ResNet50 を採用している。

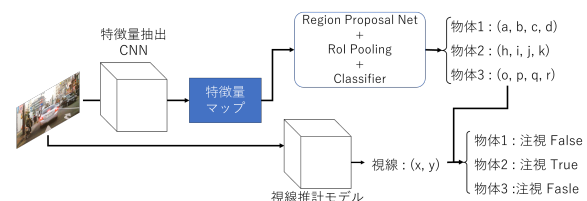


図2. 位置型視線推計システムの概観

3.2. オブジェクト型視線推計システム

位置型視線推計システムでは視線推計モデルの入力は画像情報のみであったのに対し、オブジェクト型視線推計システムの視線推計モデルでは、画像情報の他にオブジェクトの座標も入力としている。例えば Faster R-CNN が画像中に A, B, C という3つのオブジェクトを予測した場合、視線推計モデルにはまず、全体の画像とオブジェクト A の座標が入力され、モデルはそのオブジェクト A が注視されているか否かを予測する。続いてオブジェクト B, C にも同様の手続

きを行うことによって、画像中の全てのオブジェクトに対し、注視されているか否かの出力を得る。システム全体での処理の流れを図3に示す。なお視線推計モデルには、位置型視線推計システムと同様に、ResNeXt50を用いている。

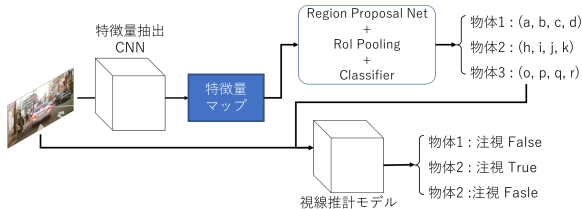


図3. オブジェクト型視線推計システムの概観

4. データセット

モデルの学習の為に、車載カメラで撮影された前方カメラ動画（5分×6本=30分）と、その動画を観たときの普通自動車免許を持つ1名の被験者の視線を計測し、そのうち7712フレーム分の画像と視線データのセットを活用した。被験者を1名に限定したのは、前述のように本研究ではある特定のベテラン運転手の理想的な目線を推計することを目指しているためである。本研究ではこのデータのうち、6168セットを学習、1544セットをテストに活用した。

5. 学習結果

前述のデータセットを用いて、両システムに30エポックの学習を行った。結果を図4に示す。なお、物体検知を行うFaster R-CNN部分は学習済みモデルを使用している。

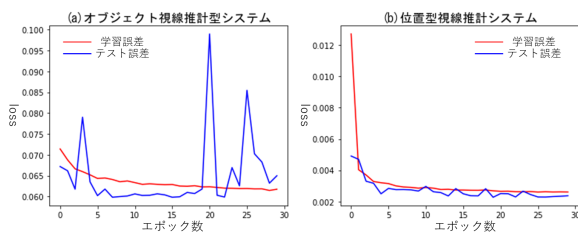


図4. 全体画像を活用した学習結果

まず学習データでのlossの推移に注目すると、両システムではエポックを重ねるごとに順当にlossが低下していることが分かる。一方でテストデータでのlossの推移に注目すると、位置型視線推計システムでは学習データと連動してlossが減少しているのに対し、オブジェクト型視線推計モデルでは、後半において学習データでの結果に反してlossが上昇していることがわかる。これは、lossの取り方と視線の特性が原因だと考えられる。運転手は通常、自らの正面を中心に同心円的に注視する確率が低くなると考えられる。この特性により、仮に位置型視線推計システムで過学習を起こしても、予測は中心近くである確率が高く、大きなlossの上昇は

起こりづらい。これに対しオブジェクト型視線推計システムでは、対象のオブジェクトを注視しているか否かの2値分類で予測を行う。このため過学習により実際と異なった注視傾向を学習した場合、1枚の画像の中で異なった分類を数多く重ね、lossが跳ね上がってしまう。こうした傾向が図4には表れたものと考えられる。

6. システムの検証

前述のように、lossの比較では評価の仕方が異なるため、両システムの優劣は判定できない。そこで両システムに対し別途用意した検証用データを適用し、その結果を同じ評価方法で比較する。用いた評価方法は、画像中のどのオブジェクトを注視しているか（もしくはどれも注視していないか）を予測し、その正答率を比較するというものである。結果を表1に示す。

表1から、位置型視線推計システムでは高い精

表1. 各モデルの検証結果

	位置型	オブジェクト型
精度	547/586 (93.3%)	362/586 (61.8%)

度が得られた一方で、オブジェクト型視線推計システムでは精度が低かったことが分かる。これは、評価方法において注視オブジェクトを、実際の人間と同様に高々1つとしたためだと考えられる。前者では推計するのは座標であるため、注視オブジェクトは必ず高々1つに絞られる。一方で後者では各オブジェクトの注視確率を独立に算出するため、複数オブジェクトの注視確率が同時に上昇し、誤予測を起こす可能性がある。これにより、オブジェクト型視線推計モデルでは本検証での精度が低かったものと考えられる。

7. まとめ

本研究では、特定の運転手の目線を予測するシステムを構築することによる運転支援の可能性を議論した。運転支援では注視する位置ではなく注視するオブジェクトの予測が必要だが、この予測においてオブジェクト情報を与えたオブジェクト型視線推計システムよりも、与えていない位置型視線推計システムの方が良い精度であった。これは前述のように、後者がより人間の、高々1つのオブジェクトを注視するという性質を反映しているためだと考えられる。今後の研究では、オブジェクト型視線推計システムでもこのような制約を設定し、オブジェクト情報を十分に活かしたモデルの構築が必要となる。

8. 参考文献

- [1]Faster R-CNN, Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, NIPS 2015
- [2]It's written All Over Your Face, X Zhang, Y Sugano M Frits and A Bulling, CVPRW 2017