

深層学習を用いた仮想空間提示映像に対する注視物体の判定と評価

Determination and Evaluation of Gazing Objects for Virtual Space Presentation Video Using Deep Learning

和田 健太郎[†]
Kentarō Wada

猿田 和樹[‡]
Kazuki Saruta

陳 国躍[‡]
Guoyue Chen

寺田 裕樹[‡]
Yuki Terata

小林 陽[†]
Hinata Kobayashi

1. はじめに

交通事故の発生要因のうち半数以上がドライバーの認知ミスであり、これまで交通事故減少のために、ドライバーの認知能力向上を目的とした運転者教育システムの研究が行われている。ドライバーが何をどの程度注視するのか判定することができれば、注視物体や見逃しなどを訓練的にフィードバックする効果的な教育が可能となる。

先行研究[1]では仮想空間に運転映像を提示し、同一シーンに対するドライバーの注視物体を視線計測と物体検出モデル YOLOX を用いた物体検出により判定する仕組みを構築したものの、注視物体判定精度は十分ではなかった。

そこで小林ら[2]は、物体検出精度に課題があるのではないかと考え、これまで使用していた CNN-based モデルから Transformer-based モデルに変更することによって精度の向上を図った。その結果、検証用データに対する単純な検出性能の高さだけでなく、システム運用時に生じやすいデータ品質劣化下でも CNN-based モデルよりも検出性能の劣化が生じにくいことを示した。しかし、実際に注視物体判定システムの物体検出の部分に実装し、その評価までは実施されていない。

そこで、本研究では Transformer-based モデルを用いてシステムを再構築し、注視物体判定精度の向上を図り、被験者別、シーン別に注視行動を評価する。

2. 仮想空間を用いた注視物体判定手法

本研究では、視線計測機能付き HMD(FOVE0)を使用して仮想空間上で被験者への走行映像を提示する。映像視聴時の被験者の視線情報を取得した後、注視物体を判定する。この時の提示映像は、360 度カメラを使用して、実際の走行環境で撮影した映像である。

以下では、注視物体判定における、視線計測手法、物体検出手法、注視物体判定手法の順で説明する。

2.1 視線計測手法

本研究の注視物体判定には、被験者の視野角映像と視線座標が必要である。そこで、被験者に FOVE0 を装着させ、走行映像を提示する。その際に出力させた被験者の視野角映像から視線座標を抽出する。

2.2 物体検出手法

視野角映像内の物体を検出するために、小林らの研究において最も高性能かつ有効性を示した Transformer-based モデルである Co-DINO[3] (Backbone を Swim Transformer[4] の Large モデルとした) を使用した。また、モデルの学習には、走行映像のフレーム画像をアンテーションした学習

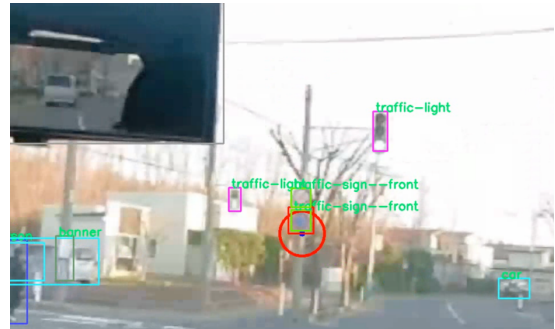


図 1 注視物体判定後の視野角映像

用データ 2045 枚を用いる。物体検出モデルによって検出された物体の座標は視野角映像のフレームごとに保存した。

2.3 注視物体判定条件

本研究では、注視領域が認識した物体の座標領域に 3 フレーム以上停留した場合に注視したと定義する。この時、注視領域は、視線座標を中心とした半径 39pixel の円とする。これは、人間の視機能が最も優れている中心視領域である弁別視野を、映像内に定義したものである。注視物体判定後の視野角映像を図 1 に示す。

3. 注視物体判定実験

運転免許証取得済みの被験者 6 人に対して FOVE0 を装着させ、注視物体の判定実験を実施した。また本実験では、道路環境ごとの注視傾向の違いを分析するために各被験者に 4 本の走行映像を表示する。その後、各視野角映像に対して、提案手法により注視物体を判定する。

また、注視物体のクラス毎の数と、注視物体判定後の視野角映像を目視により確認することによって、被験者がどの走行環境でどういった物体を注視する傾向があるのか考察する。なお実験結果には被験者 A の全ての走行映像に対する注視物体判定の結果を示す。

3.1 被験者概要

本実験の被験者は、全員運転免許証は持っている 20 代前半の学生 6 名である。被験者 A、B、C、D は車を所持しており、普段から運転している。対して被験者 E、F は車を所持しておらず、普段から運転の習慣がない。この 6 名に走行映像を提示し、注視物体を判定する。

3.2 走行映像概要

本研究では 4 本の走行映像を使用した。走行映像の概要を表 1 に示す。また、走行映像 4 に関しては、走行映像 2 と全く同じ走行ルートであるが、天候が異なる。

[†] 秋田県立大学大学院 Graduate School of Akita Prefectural University

[‡] 秋田県立大学 Akita Prefectural University

表1 走行映像の概要

	車線	道路幅	運転タスク	天候	交通量
1	片道一車線	狭い	緩やかなカーブ	晴れ	普通
2	片道一車線	狭い	交差点右折	晴れ	普通
3	片道二車線	広い	交差点右折	晴れ	比較的 多い
4	片道一車線	狭い	交差点右折	雪	普通

4. 実験結果

4.1 走行映像毎の注視物体判定結果

被験者 A の走行映像 1~4 に対する提案手法を用いた注視物体判定数を図 2 に示す。図 2 を見ると雪の日の走行映像 4 では、走行映像 2 と比較して、車の注視回数が増えていることが分かる。また、信号機のような小物体に対しても注視回数が増えていることが分かる。

また、交通量が比較的多く、交差点を右折するタスクもある走行映像 3 では、車の注視回数が多いだけでなく、右ミラーと左ミラーの注視回数が他の走行映像と比較して多くなる傾向がある。

全ての走行映像に対して、標識（traffic-sign-front）が出現するにも関わらず、ほとんど注視していなかった。

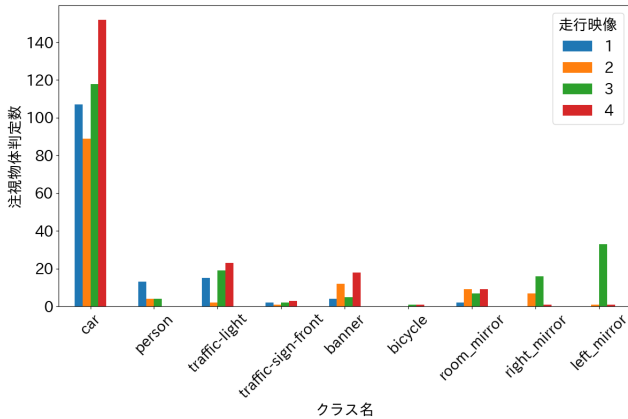


図 2 被験者 A の注視物体判定数

4.2 被験者毎の注視物体判定結果

表 2 に、走行コース 1 における被験者毎の注視物体判定結果を示す。表から被験者によって車やルームミラーの注視回数に違いが生じていた。

4.3 目視による注視物体判定との比較

走行映像 1、2 に関しては、歩行中の人に対しては、注意深くその人を注視している傾向があった。また、全ての走行映像において基本的に前の車両を注視しているが、狭い道路の時ほど、すれ違う対向車に視線を送る傾向が見られた。注視物体の判定性能に関しては、従来の CNN-based 手法と比較し、信号機や遠方の物体などのような小物体に対して判定性能が向上したことが確認できた。

表 2 被験者毎の注視物体判定数

クラス名	被験者					
	A	B	C	D	E	F
car	107	115	136	107	97	164
person	13	18	11	7	10	13
traffic-light	15	36	17	32	34	21
traffic-sign-front	2	5	8	11	3	6
banner	4	6	4	7	4	8
bicycle	0	0	0	0	0	0
room_mirror	2	2	6	7	7	0
right_mirror	0	0	4	1	2	0
left_mirror	0	0	1	1	0	0

5. 注視物体判定に関する考察

雪の日の走行映像に対し、車や信号機の注視回数が向上した要因として、雪によって道路環境が悪化し、さらに視界が悪くなることが挙げられる。これによって、遠くから接近する車や前方の車に対する注意力が高まっていると考えられる。

被験者毎の注視物体判定結果では、本研究の手法によって被験者間で注視傾向に違いがあることが確認できた。

全ての走行映像に対して、標識がほとんど注視されなかったことに関しては、被験者にとって慣れた道路環境であるため、標識を見る必要がないことが要因と考えられる。

6. おわりに

本研究では、6名の被験者を対象に注視物体判定実験を実施した。まず、先行研究と比較して注視物体の判定精度が向上したことによって、従来よりも高精度に被験者の注視物体を判定し評価できることが分かった。

また、本研究では被験者全員が見慣れた走行環境を提示したので、被験者が見慣れていない走行映像を提示することで注視物体に違いがあるのかについても検証したい。

注視物体判定数だけでなく、シーン別の注視物体を判定し、出力する機能をシステムに実装することによって、さらに被験者間の注視傾向の評価を詳細に実施していきたい。また、本実験では、同年代の被験者を対象にしていたが、高齢ドライバーに対する実験も実施したい。

参考文献

- [1] 和田 健太郎ほか, “360度カメラ映像を用いた注視物体の判定と評価”, 日本交通科学学会誌, Vol.23, No.G6-3 (2023).
- [2] 小林 陽ほか, “Transformer-based モデルによる運転者視界に対する物体検出性能の検討”, 第 24 回情報科学技術フォーラム, H-051 (2024).
- [3] Z. Zong, et al., “DETRs with Collaborative Hybrid”, Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.6748-6758 (2023).
- [4] Z. Liu, et al., “Swim Transformer: Hierarchical Vision Transformer using Shifted Windows”, In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.10012-10022 (2021).