Human-like Guidance に適した視線情報データセットの作成

EP19078 難波田雅己　　指導教授：山下隆義

1. はじめに

自動車の次世代ナビゲーションとして，人間と同じ様な誘導を行う Human-like Guidance (HℓG) [1] の研究が取り組まれている．HℓG を実現するには，運転者が注視すべき物体を推定し，その物体に着目した案内をする必要がある．そのため，シーン状況に合わせて運転者がどこを注視するか推定することが重要となる．これまでに，HℓG に特化した視線情報を含むデータセットは存在しない．そこで本研究では，HℓG ナビゲーションに特化した視線情報データセットを作成する．また，作成したデータセットを利用した視線推定の実験を行い，有効性を検証する．

2. HℓG のためのデータセット作成

HℓG に特化した視線データセットの作成を目的とし，運転シーンにおける人の視線を収集する．実世界で様々なシーンのデータを収集することは困難であるため，自動運転開発シミュレータである CARLA を用いる．視線データの収集には，眼球の動きが計測可能な Tobii Pro X3-120 を使用する．これらを用いて作成したデータセットを Driving Gaze Dataset と呼ぶ．図 1 に視線データの収集の流れを示す．

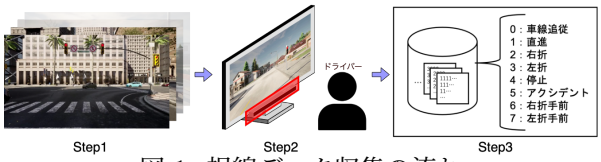

図 1: 視線データ収集の流れ

視線の収集方法は，以下の通りである．

Step1：運転シーンの動画像作成

CARLA の自動運転機能を用いて様々なシーンの動画像を収集する．1 つの動画像は 60fps で約 3 分間とし，20 本作成する．

Step2：視線収集

2 種類の視線データを収集する．1 つ目は通常運転時の視線情報である．既存のカーナビゲーションのように「100m 先を右方向です」といった案内に沿った視線 (自然な視線) の収集を目的とする．2 つ目はナビゲーションに適した視線情報である．事前に動画像を見せ指示文に関連した物体に視線を向けるように指示して,「前方車に追従してください」といった人間のような道案内を行った視線 (恣意的視線) の収集を目的とする．それぞれのデータは人間毎の視線情報の多様性を考慮するため，複数人で収集する．

Step3：ラベルデータの付与

シーン状況を Straight (車線追従/直進)，Turn (右折/左折/右折手前/左折手前)，Stop(停止/アクシデント) と定義し，各フレームに対してラベル付けを行う．また，ラベルの数値で埋められたラベルマップデータを作成し，デコーダに入力する特徴マップと連結する．

3. 評価実験

視覚的顕著性予測の代表的な手法である Dilated Inception Network (DINet)[2] を用いて，Driving Gaze Dataset の有効性を検証する．

3.1. 実験概要

Driving Gaze Dataset には自然な視線，恣意的視線がそれぞれ 6 人分含まれ，学習用は約 100,000 フレーム，評価用は約 3,200 フレームである．これらのデータを用いた 4 つの手法について比較，評価を行う．1 つ目は自然な視線のみでの学習（手法 1），2 つ目は恣意的視線のみでの学習（手法 2）である．3 つ目は手法 2 にラベルマップデータを付与して学習する（手法 3）．4 つ目は手法 3 に用いる動画像データについて，各シーン（Straight/Turn/Stop）の割合を 1:8:1 に調整して学習する（手法 4）．評価指標には CC，sAUC を使用する．

3.2. 実験結果

表 1 に各手法の評価結果，図 2 に各手法の視線推定結果を示す．図 2 から，手法 1 では画面中央付近に視線が集中した．手法 2 では右折時で左前方車両に視線を推定する傾向を確認した．しかし，遠方車両に対しては追従する視線が確認できず，手法 1 と同様に中央付近に視線を推定する結果であった．また，表 1 から手法 1 より低い精度であることを確認した．手法 3 では右折時の左前方車両への視線に加え，前方車両追従時に遠方車両を追従する視線を確認した．それに加え，表 1 から手法 2 より高い精度であることを確認した．手法 4 では，手法 3 と同様の推定結果を獲得しつつ，表 1 から手法 3 よりさらに高い精度であることを確認した．

表 1: 各評価指標での定量的評価

評価指標	手法 1	手法 2	手法 3	手法 4
CC	0.4685	0.4325	0.4306	**0.4802**
sAUC	0.5882	0.5512	0.6002	**0.6446**

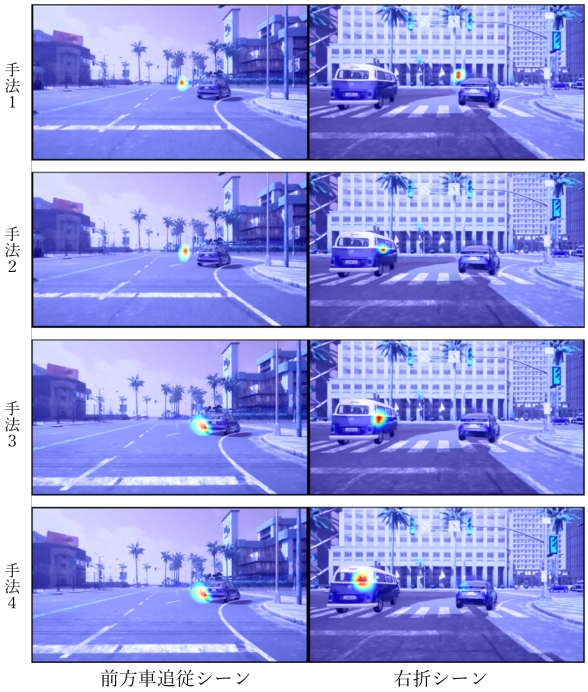

図 2: 各データにおける視線推定結果

3.3. 考察

手法 1 の結果は，人間の視線情報にあるセンターバイアスが原因と考えられる．センターバイアスは，人間が動画像を見る時に周辺視野を用いて情報を得るため，視線が中央に集まる現象である．一方，恣意的視線は視線データ収集時に視線の誘導を行うため，センターバイアスが減少し，手法 2 において左前方車への視線推定が可能となった．手法 3 は，どの状況であるかのシーン情報を追加することで特有の視線特徴を捉えやすくなり，右折時の左前方車両に対する視線に加えて，遠方車両を追従する視線を推定したと考えられる．手法 4 では，センターバイアスの主な原因である直進シーンを削減することでクラス均衡を抑制し，精度が向上したと考えられる．

4. おわりに

本研究では，HℓG に特化した視線情報データセットを作成し，DINet を用いて有効性を調査した．実験結果から，恣意的視線とラベルデータを入力することでナビゲーションにつながる視線を獲得し，各シーンの割合を調整することで精度の向上を確認した．今後は，時系列情報を考慮したモデル等の他モデルでの実験を行い，精度の比較を行う．

参考文献

[1] B.Wang, *et al.*, “Landmarks Based Human-like Guidance for Driving Navigation in an Urban Environment”, ITCS, 2018.

[2] S.Yang, *et al.*, “A Dilated Inception Network for Visual Saliency Prediction”, TMM, 2019.