

## 1. はじめに

自動車の案内として、機械による Turn-by-Turn ナビゲーションと人間によるナビゲーションの2つが考えられる。Turn-by-Turn ナビゲーションとは既存のカーナビゲーションシステムのことであり、デジタル地図を用いた道路に基づいた案内文によりナビゲーションをする。そのため、「100m 先を右方向です」といった定型文による案内であり、直感的な理解が難しいことがある。一方、人間によるナビゲーションでは、同乗者が視認した周辺環境の情報を利用して、建物や周辺車等を利用して案内を行う。そのため、「前の黄色い車に付いて行ってください」と言った状況に合わせた案内になり、直感的に理解しやすい。よって、Turn-by-Turn ナビゲーションによる問題を防ぐことが可能であると考えられるが、同乗者がいなければ成り立たない。この問題点を解決するために、本研究では人間のナビゲーションをシステムで再現することを目標とした Human-like Guidance (H<sub>L</sub>G) [1] の実現を目的とする。H<sub>L</sub>G では、状況に合わせたわかりやすい案内文でのナビゲーションがシステムのみで可能となる。本研究では、人間の運転中の視線情報を考慮したシーングラフによる H<sub>L</sub>G の実現を提案する。

## 2. 提案手法

本研究では、時系列を考慮した周囲の環境情報を表すテンポラルシーングラフを提案し、視線情報とテンポラルシーングラフによる H<sub>L</sub>G の実現手法について述べる。提案手法の概略を図3に示す。提案手法ではまず、周囲の状況をシーングラフとして表現する。同時に視線推定モデルを用いて、着目対象を推定し、この視線情報をシーングラフに追加する。次に、視線情報付きシーングラフを任意の時刻分連結した、テンポラルシーングラフを作成する。作成したテンポラルシーングラフを Spatial Temporal Graph Convolutional Network に入力し、テンポラルシーングラフの特徴量を獲得する。最後に、得られた特徴量から、Transformer によって案内文を生成する。

### 2.1 運転者の視線の推定

人間の運転中の視線を再現するために、視線推定モデルである DINet [2] を用いて視線推定を行う。DINet は RGB 画像を入力とし、視線推定結果をヒートマップとして出力する。ヒートマップから注視物体特定するために、閾値以上の領域の外接矩形を求める。学習では、出力したヒートマップと真値の視線座標を基にしたマップとの誤差を求める。ここで、視線情報には、通常運転時の視線情報ではなく、ナビゲーションを行う際に着目しやすい対称に意図的に誘導を行ったデータを用いる。これによって、H<sub>L</sub>G のナビゲーションに適した視線の推定が可能となる。

### 2.2 視線情報を付与したシーングラフ

時刻  $t$  の周囲の状況のシーングラフ  $G$  を定義する。ノードは  $v_i = \{ \text{オブジェクトクラス, 座標, 車両名, 色} \}$  を割り当て、エッジにはノード間の相対距離を用いる。加えて、視線推定により得られた矩形と位置が一致するオブジェクトを着目対象とする。よって、ノード集合を  $vg_i = \{ \text{オブジェクトクラス, 座標, 車両名, 色, 視線情報の有無} \}$  とする。図1に示すように、視線情報を付与したシーングラフを定義する。

### 2.3 ST-GCN による特徴抽出

各オブジェクトの位置関係について時系列情報を考慮するために、ST-GCN [3] を用いる。現在から未来  $T$  時刻分について、各時刻のシーングラフのノード集合を定義する。また、隣接した他時刻における同じオブジェクトを接続するエッジを追加することで、テンポラルシーングラフを定義する。図2に示すように、グラフ畳み込み処理の際に、各時刻で共通するノードを同時に畳み込む。これにより、オブジェクト毎に移動を考慮した特徴量を獲得する。

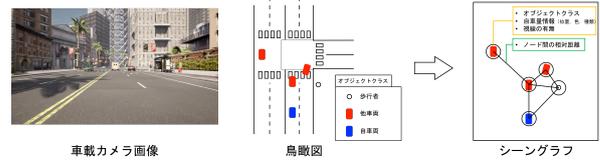


図1：シーングラフの作成

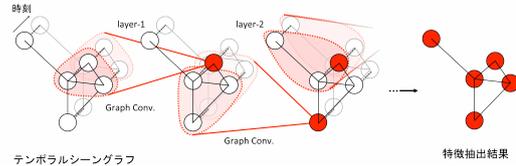


図2：ST-GCN によるテンポラルシーングラフの特徴抽出

## 2.4 Transformer による案内文生成

図4に示すように、ST-GCN で得られたグラフ特徴量から、着目対象を基にした案内文の生成を行う。文章生成のモデルとして Transformer [4] を用いる。Transformer Encoder では、ST-GCN で得られた各ノードの特徴量をトークンとして入力する。Transformer Decoder では、Encoder から獲得したトークンを Decoder の入力として用い、単語を出力する。この時、教師データとして該当シーンにおける正解となる案内文を用いる。正解の案内文は、該当シーンにおける真値の視線情報を用いてルールベースで作成する。これにより、各オブジェクトに対応したナビゲーション文を生成できる。

## 3. 評価実験

本実験では、提案した手法における時系列情報の有無による影響を評価する。

### 3.1 データセット

データセットには、自動運転開発シミュレータである CARLA Simulator で作成した動画画像を用いる。全 14 シーン、合計 1672 フレームの運転シーンの動画画像と、各フレームに存在する全ての各オブジェクトに対する案内文をデータセットとする。ここで、案内文はルールベースでオブジェクトに対する文章を自動生成した。オブジェクトが人の場合、2名以上が近距離に位置している時は集団とし、集団に対する案内文とした。案内文作成のルールは、進行方向、対象オブジェクトの名前、進行方向に対する自車両の行動を含む文とする。対象オブジェクトは、該当シーンにおける視線情報を用いて選択する。自車両の行動に関する指示は、対象オブジェクトの進行方向と自車両の進行方向が同じであれば追従、異なる場合は対象オブジェクトの位置を用いる。

### 3.2 時系列導入による定性的評価

図5に示すように、時系列情報ありの場合は「following kawasaki ninja」と移動を伴う案内文が生成されている。したがって、動的なオブジェクトに着目したわかりやすいナビゲーションの生成できていることが確認できる。一方で、時系列情報なしの場合、「kawasaki ninja is located」と移動を伴わない案内文が生成されている。時系列情報がない場合でも、動的なオブジェクトに着目はできているが、「following」が「located」となっており、動きの情報が欠落していることが確認できる。このことから、時系列情報を導入することは Human-like Guidance において有効であることが確認できた。

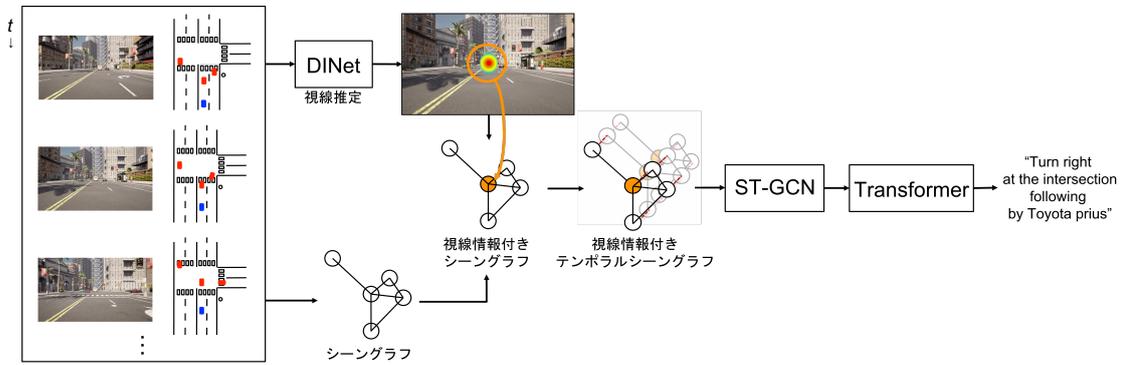


図 3：提案手法の流れ

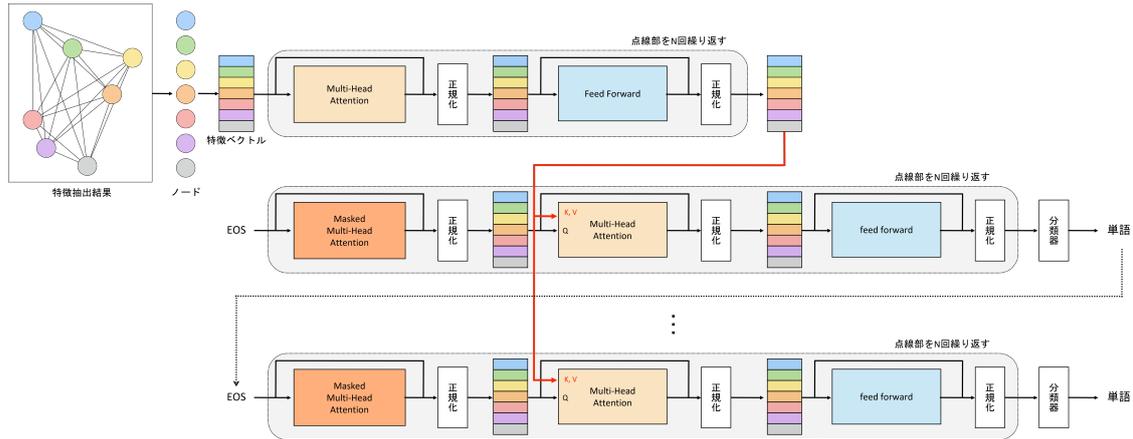


図 4：Transformer による文章生成

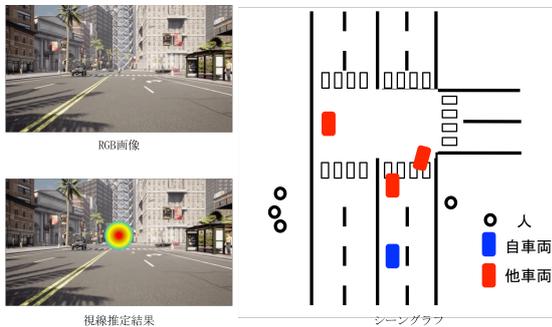


図 5：案内文生成例

### 3.3 定量的評価

自然言語生成の指標における定量的評価を表 1 に示す。表 1 より、時系列情報を導入することによって各評価指標での精度が向上しており、目的である Human-like Guidance に重要となる自然な案内文が生成されている。したがって、時系列情報の導入が有効な手段であることが確認できる。

### 4. おわりに

本研究では、H1G のための、テンポラルシーングラフを用いたナビゲーション生成手法を提案した。視線情報と時系列情報を用いることで、人間のよう状況に合わせたナビゲーションの生成が実現できた。視線推定の評価実験では、ナビゲーションに適した視線が H1G において有効

表 1：時系列の有無による定量的評価比較

	時系列なし	時系列あり
Bleu_4	0.291	<b>0.325</b>
METEOR	0.199	<b>0.223</b>
ROUGE_L	0.450	<b>0.520</b>
CIDEr	2.649	<b>2.945</b>
SkipThoughtCS	0.458	<b>0.487</b>

であると確認できた。時系列情報の有無による評価実験では、時系列情報を用いることが H1G において有効であることが確認できた。今後の課題としては、視線推定モデルの変更や H1G に適切な評価指標の検討などが挙げられる。

### 参考文献

- [1] M. Tsutomu, *et al.*, “Human-Like Local Navigation System Inspired by a Hippocampal Memory Mechanism”, *Brain-Inspired Information Technology*, pp.29 – 32, 2010.
- [2] S. Yang, *et al.*, “A Dilated Inception Network for Visual Saliency Prediction”, *IEEE*, 2019
- [3] S. Yan, *et al.*, “Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition”, *AAAI*, 2018.
- [4] S. Vaswani, *et al.*, “Attention Is All You Need”, *CoRR*, 2017.

### 研究業績

- [1] 宮田昌樹 等, “関係性部分グラフを用いた Graph Convolutional Network による経路予測”, *MIRU*, 2021.
- [2] M. Miyata, *et al.*, “Relational Subgraph for Graph-based Path Prediction”, *MVA*, 2021.