

1.はじめに

自動運転システムの安全性評価には、多様なシナリオでの検証が必須である。事故等の危険な状況を実環境で再現することは安全性の面で難しいため、シミュレータの活用が検討されている。しかしながら、人手による多様なシナリオ作成には多大なコストを要する。そこで、本研究では Multimodal Large Language Model (MLLM) を用いたシナリオの自動生成法を提案する。提案手法により、事故につながる危険なシナリオを効率的に作成でき、自動運転システムの安全性評価に貢献することを目指す。

2. OpenSCENARIO XML

OpenSCENARIO XML は、自動化システムと測定システムの国際標準化団体によって策定された、自動運転検証用の標準シナリオ記述フォーマットである。本規格は、車両や歩行者などの交通参加者を定義する Entities や、シナリオ推移を記述する Storyboard などのタグにより階層的に構成される。OpenSCENARIO が動的な挙動を記述するのに対し、OpenDRIVE は静的な道路環境を提供する。シナリオ生成には道路構造を定義する OpenDRIVE との連携が不可欠であり、両者の統合により整合性のとれた交通状況が再現可能となる。

3. 提案手法

本研究では、MLLM として Gemini 3 Pro [1] を用い、実環境の走行動画から OpenSCENARIO XML を自動生成する手法を提案する。本手法の概要を図 1 に示す。入力情報には、実環境のフロントカメラ映像に加え、逆透視投影変換を用いて生成した俯瞰 (Bird's Eye View: BEV) 映像を使用する。BEV 映像を補助的に用いることで、フロントカメラ映像だけでは困難な自他車両間の距離や挙動を捉えられると考える。

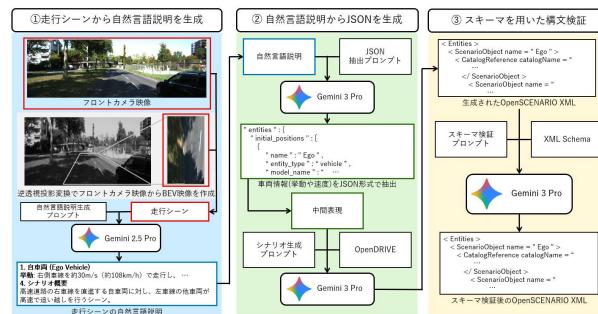


図 1: 提案手法の概要

3.1. 走行シーンから中間表現の生成

フロントカメラ映像と BEV 映像を MLLM に入力し、走行シーンの自然言語説明を生成する。次に、再度 MLLM を用いて自然言語説明から車両情報（車両の挙動や速度）のパラメータを抽出し、JSON 形式に出力する。深い階層構造を持つ XML を MLLM に直接生成すると、タグの不整合や構文エラーが多発する。そこで、JSON 形式にすることで、中間表現としてデータの構造化を行う。MLLM の学習データ特性との親和性を活かし、生成エラーを抑制しながら複雑なシナリオ情報を正確に構造化することを可能にしている。

3.2. 中間表現から OpenSCENARIO XML の生成

生成した JSON 形式の中間表現を MLLM を用いて OpenSCENARIO XML 形式に変換する。MLLM に対し前段の中間表現に加え、OpenSCENARIO XML の記述例と道路情報が記述された OpenDRIVE ファイルを入力する。次に、生成された OpenSCENARIO XML に対し、データ型などが定義された XML Schema を用いた構文チェックを行う。この構文チェックを行うことで、シミュレータでの実行可能率を向上させる。

4. 定量評価

本実験では、OpenSCENARIO XML の生成の成功率とともに、走行軌跡の再現性を定量的に評価する。

4.1. 実験条件

本実験では、提案手法の有効性を検証するため、入力情報の違いによるシナリオ再現精度の比較を行った。比較手法として、図 2(a) のような前方映像のみを入力とする場合と、前方映像に加え図 2(b) のような BEV 映像を補助情報として入力する場合の 2 パターンで比較実験を行う。評価には KITTI Odometry Dataset [2] の直進、右左折を含む計 15 シーン (10 秒間) を用いる。評価指標には、生成した OpenSCENARIO XML をシミュレータ (Esmini) で実行したときの走行軌跡と実走行軌跡 (Ground Truth) との平均位置誤差 (ADE) と最終地点誤差 (FDE) を用いて比較する。



図 2: 入力シーンの例

4.2. 実験結果

表 1 にシーンごとの評価結果を、図 3 に生成したシナリオの軌跡と正解軌跡を示す。表 1 より、動画情報のみを用いた場合は、右左折シーンにおける精度が悪く、動作の再現性が低い結果となった。しかし補助情報 (BEV) を用いた場合は、右左折時の ADE が改善した。また、全体平均においても ADE が半減し、FDE は 6 分の 1 程度となり、誤差の低減が確認できた。さらに、図 3 の通り、補助情報を追加した場合の軌跡は、正解軌跡により近いことが分かる。以上の結果から、OpenSCENARIO XML の生成は、右左折時の軌跡再現精度の向上に効果的である。

表 1: 生成シナリオの軌跡評価と生成成功率

シーン	Video Only			Video + BEV		
	ADE(m)	FDE(m)	成功率 (%)	ADE(m)	FDE(m)	成功率 (%)
右折	12.4	43.1	60.0	6.8	8.7	60.0
直進	2.6	6.3	80.0	2.6	6.2	100.0
左折	12.1	46.7	20.0	4.8	1.2	80.0
平均	9.0	32.0	66.7	4.7	5.4	86.6

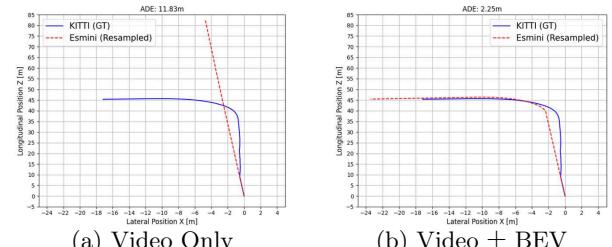


図 3: 軌跡の比較

5. おわりに

本研究では、実環境映像から MLLM を用いて OpenSCENARIO XML を自動生成する手法を提案した。評価実験において、フロントカメラ映像に加えて BEV 映像の補助情報を入力することで、直進のみならず右左折シーンにおいても高精度なシナリオ生成が可能であることを確認した。今後はより長時間のシナリオへの対応や、他車両を含めた動的環境の再現精度の向上が課題である。

参考文献

- [1] Gemini Team, Google, "Gemini 3 Pro Model Card.", Google DeepMind, 2025.
- [2] A. Geiger, et al., "Are we ready for autonomous driving? The KITTI vision benchmark suite", CVPR, 2012.