講演番号	207

Scenario Extraction from Drive Recorder Footage for Traffic Scene Retrieval

Masafumi Tsuyuki[1)] **Tetsuya Nishida**[2)] **Taminori Tomita**[2)] **Yoshitaka Atarashi**[2)]

1) Hitachi, Ltd. 1-280, Higashi-Koigakubo, Kokubunji-shi, Tokyo 185-8601, Japan (E-mail: masafumi.tsuyuki.wn@hitachi.com)

2) Astemo, Ltd. 37F, Shibuya Sakura Stage SHIBUYA Tower, 1-1 Sakuragaoka-cho, Shibuya-ku, Tokyo, 150-6237 Japan

KEY WORDS: safety, drive recorder/event data recorder, image processing/information processing, scenario extraction (C1)

The continuous development of autonomous driving and advanced driver-assistance systems requires extracting specific traffic scenarios from massive amounts of drive recorder footage. While existing scene retrieval methods using vector search or multimodal Large Language Models (LLMs) are effective for static concepts, they often struggle to identify dynamic events involving complex vehicle interactions, such as lane changes. Furthermore, conventional scenario extraction technologies heavily rely on expensive multi-sensor configurations—including LiDAR, radar, and high-precision IMU—limiting their applicability to general mass-produced vehicles.

To address these challenges, we propose a practical and lightweight approach for extracting traffic scenarios using only monocular camera footage and GNSS location data (Fig.1). Our method first retrieves a localized road network in OpenDRIVE format from a map database based on the ego-vehicle's GNSS coordinates. To mitigate GNSS measurement errors, the ego-vehicle's trajectory is refined using a Kalman smoother. For surrounding vehicles, we apply object detection, multi-object tracking, and monocular depth estimation to the camera frames. These relative positions are then transformed into world coordinates, smoothed, and integrated with the ego-trajectory to generate an OpenSCENARIO XML file.

Experimental evaluation using a lane-change scenario from the nuPlan dataset demonstrates that our smoothing technique effectively suppresses noise from the depth estimation model. The proposed method estimates preceding vehicle trajectories with a root mean square error (RMSE) of 0.36 m in the lateral direction and 2.61 m in the longitudinal direction. This provides sufficient precision to simulate and describe complex traffic events. Moreover, representing the scene as an OpenSCENARIO file (Fig.2) reduces data storage requirements by approximately 97% (from 27 MB to 868 KB) compared to the original video.

Finally, we verified the utility of these extracted scenarios for text-based scene retrieval. When the OpenSCENARIO data was input into an LLM, it generated highly accurate descriptions of spatial-temporal dynamics, correctly identifying a continuous lane-change maneuver (Fig.3). In contrast, feeding raw video frames directly into the LLM caused it to misinterpret relative motions due to the ego-vehicle's movement. While video inputs remain advantageous for capturing visual attributes like vehicle types and scenery, our OpenSCENARIO-based approach provides a highly robust, storage-efficient foundation for understanding and retrieving complex traffic events from everyday drive recorder data.

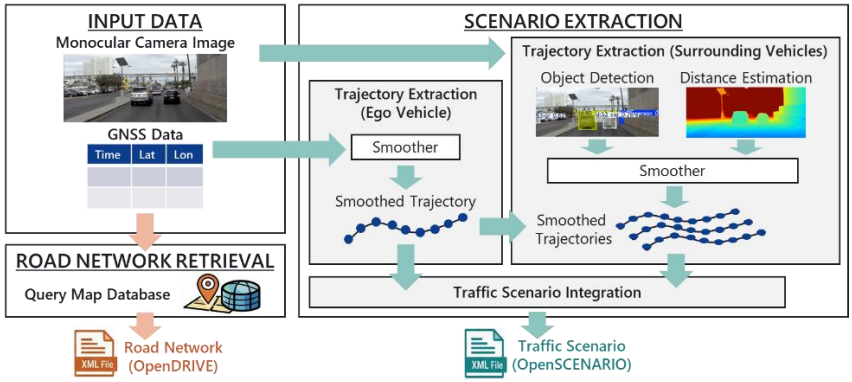
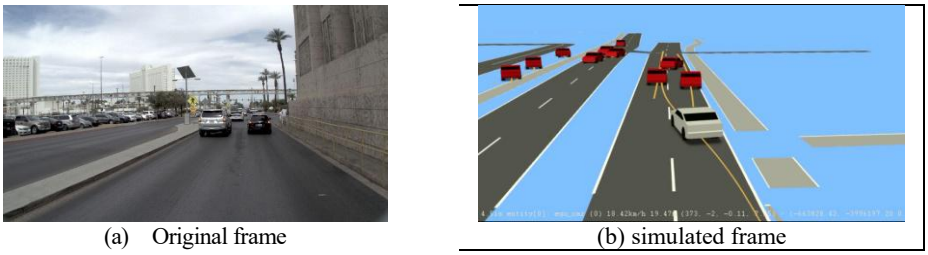


Fig.1 Overview of the proposed method for Extracting Traffic Scenarios from Drive Recorder Footage.



(a) Original frame (b) simulated frame

Fig.2 Comparison between the simulated frame based on the extracted OpenSCENARIO and the original video frame.

自車は片側複数車線の道路を前進しながら，右寄りの車線から左側の車線へ緩やかに移る場面である．自車の少し左前方には中間車線を走る車が先行し，そのさらに左前方には目的の車線を走る別の車がいる．自車はこれら 2 台を追い越すのではなく，前後間隔を保ったまま中間車線を横切って車線変更し，最終的に左前方車の後ろに収まる．周囲にはさらに遠方の並走車もいるが，交差や合流はなく，無理のない連続的な車線変更として自然な交通状況になっている．

Fig.3 Caption Generated from the extracted OpenSCENARIO