Detection of Dangerous Driving Events from Video Streams with Logical Explanations

Kazuko Takahashi[1] [a], Yurika Yamaguchi[1], Daiki Suzuki[2], Duong Dinh Tran[2] [b], Aran Chindaudom[2], Takashi Tomita[2] [c] and Toshiaki Aoki[2] [d]

[1] *School of Engineering, Kwansei Gakuin University, 1, Gakuen, Uegahara, Sanda, 669-1330, JAPAN*

[2] *School of Information Science, Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa 923-1292, JAPAN*

{*ktaka, imd52642*}@*kwansei.ac.jp*, {*s2510093, duongtd, aran.c, tomita, toshiaki*}@*jaist.ac.jp*

Keywords: event extraction, qualitative spatial reasoning, knowledge representation, vehicle action detection, formal software analysis

Abstract: We propose a framework that converts video streams into symbolic representations and performs formal reasoning to derive high-level events. We present the event detection of dangerous driving using this framework. We define several types of dangerous driving events using Qualitative Spatial Reasoning, and represent driving scenes captured by dashboard cameras — specifically focusing on scenarios where the ego-vehicle is subjected to dangerous driving behaviors. These events are then detected through logical reasoning. We conducted experiments using both simulated and real-world data and provided formal explanations for the reasoning process. Furthermore, the event definitions were iteratively revised to encompass a broader range of cases.

1 INTRODUCTION

In recent years, the practical application of autonomous driving technology has advanced significantly. Detecting dangerous driving is crucial for enhancing the safety of autonomous driving systems. Generally, research in this field is conducted using deep learning techniques which have achieved numerous significant results (Athanesious et al., 2020; Hacohen et al., 2022; Yao et al., 2022; Garefalakis et al., 2024). However, this approach has several shortcomings: (i) Data Dependency: A vast amount of training data is required. (ii) Scalability: Long training times are necessary for each event detection model, making it burdensome to introduce new types of dangerous behaviors. (iii) Explainability: These models often fail to provide transparent evidence or logical explanations for their judgments.

To address these limitations, an approach based on Qualitative Spatial Reasoning (QSR) has been proposed (Takahashi and Yamaguchi, 2026). QSR focuses on specific aspects of spatial data to enable logical reasoning (Chen et al., 2013; Sioutis and Wolter, 2021). For example, consider the relative position of a vehicle with respect to the right lane line. If a vehicle located in the right of the line moves to the left, a lane-change event is detected. This demonstrates that a sequence of relative spatial relations between objects can be used to explain events logically. Such information is essential for legal cases or insurance claims following a traffic accident.

Scenario-based verification has become an industry standard in the validation of automated driving systems (JAMA, 2022; ISO, 2021). In the Safety Evaluation Framework proposed by JAMA, road geometry, relative vehicle positions, and vehicle maneuvers are systematically analyzed to derive comprehensive scenario sets. While scenarios such as cut-in and weaving are included in the framework proposed by JAMA, scenarios corresponding to certain dangerous driving behaviors, such as overtake and squeezing, are not explicitly covered.

Takahashi et al. took four categories of dangerous driving, cut-in, overtake, weaving and squeezing, and presented an event detection from video using the QSR approach. However, their experiments were limited to cut-in and overtake scenarios, and the discus-

[a] https://orcid.org/0000-0002-5572-7747
[b] https://orcid.org/0000-0001-7092-2084
[c] https://orcid.org/0000-0003-1249-7862
[d] https://orcid.org/0000-0002-1209-6375

sion on logical reasoning remained insufficient.

In this paper, we evaluate all these four categories of dangerous driving events using a more diverse set of video data. We extract objects from video as regions, transform their relative spatial relations into QSR form, and detect the occurrence of these events. Through experiments on several video streams, we provide formal explanations not only for successfully detected events but also for the reasons behind detection failures. Based on these results, we modified the event definitions to ensure comprehensive coverage of various cases. Furthermore, we propose our approach as a robust framework for video event detection with logical explanations. Our final goal is to provide semantic understanding to video data for dangerous driving detection, rather than to improve raw performance metrics such as precision or accuracy.

This paper is organized as follows. In Section 2, we describe our representation language in QSR. In Section 3, we provide definitions for several events including dangerous driving. In Section 4, we present our experimental results of event detection. In Section 5, we discuss the methodology for generating explanations and summarize the framework. In Section 6, we describe the related works. And finally, in Section 7, we conclude the paper and discuss future work.

2 QSR FOR DANGEROUS DRIVING

2.1 Prerequisites

We currently assume that the video data taken by the dashboard camera satisfy the following conditions: (i) all vehicles (including the ego-vehicle) except for the dangerously driving vehicle are driving safely, (ii) there are one driving lane and one passing lane, both of which are almost straight, (iii) traffic is left-hand and the passing lane is on the right side of the driving lane, and (iv) the ego-vehicle is driving in the driving lane.

We use the following regions in a rectangle form: the enclosure corresponding to the entire screen (*enc*), the lane lines drawn beside the driving lane left line (*lline*) and right line (*rline*), the dangerous driving vehicle (*dv*) and a safe driving vehicle in front (*sv*). The *lline* is extracted as a rectangle defined by the top right and bottom left corners of the lane line in the scene, and the *rline* is extracted as a rectangle defined by the top left and bottom right corners (Figure 1).

Hereafter, in the figures, the regions $lline, rline, dv$ and *sv* are shown as rectangles colored by blue, yellow, orange and purple, respectively.

Figure 1: The regions of lines in the entire screen.

2.2 Qualitative representation

QSR is a method to represent spatial data in a declarative manner, focusing on a certain aspect of an object or relative relation of objects. We treat a rectangle region on a two-dimensional plane, and focus on a mereological relation, directional relation and relative size between a pair of objects. We use the QSR tailored to handle dangerous driving detection proposed in (Takahashi and Yamaguchi, 2026).

2.2.1 Mereological relation

We define mereological relations between pairs of regions based on RCC-8 (Randell et al., 1992) but in a coarser manner. For a pair of regions X and Y, $R(X,Y)$ where R is *in*, *shr*, *ec* and *dc*, indicate that X is in the inner part of Y, sharing a part with Y, externally connected with Y and disconnected from Y, respectively. In cases where a region X is invisible in the scene, $none(X)$ indicates that X does not appear in the scene, and $cvd(X,Y)$ indicates that X is completely occluded by Y.

2.2.2 Directional relation

Let X and Y be regions. When mereological relation of X and Y is either *in*, *cvd* or *none*, it is unnecessary to give a specific direction. Otherwise, we determine the directional relation X with respect to Y.

We set Y in the center and divide the screen by extending its boundaries which are shown by dotted lines in Figure 2. By this division, we got eight areas outside of Y. The four areas $Area_{lu}, Area_{ru}, Area_{ld}$ and $Area_{rd}$ are said to be corner areas, and the other four areas $Area_u$, $Area_d$, $Area_{rm}$ and $Area_{lm}$ are said to be middle areas. The middle areas are defined to include the boundaries adjacent to the corner areas. The corner areas and the middle areas are treated differently.

The directional relation X with respect to Y, denoted as $D(X,Y)$, is determined depending on the areas of X's location.

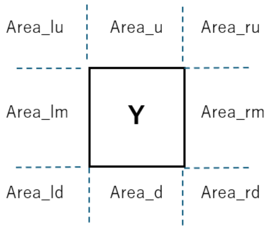


Figure 2: Area division.

We first determine the direction if X shares its inner part with the corner areas (Table 1(a)). In this case, intersection with the middle areas is ignored. If X intersects only with middle areas, the direction is determined according to Table 1(b). In these tables, the symbols ✓ and × show that there exists a shared part and not, respectively. For example, the second line of Table 1(a) shows if X shares its part with $Area_{lu}$ and $Area_{ru}$, and not with $Area_{ld}$ and $Area_{rd}$, then $D(X,Y) = lru$.

Table 1: Directional relation.

corner area $Area_x$				D
lu	ru	ld	rd	
✓	✓	✓	✓	$lrud$
✓	✓	×	×	lru
×	×	✓	✓	lrd
✓	×	✓	×	lud
×	✓	×	✓	rud
✓	×	×	×	lu
×	✓	×	×	ru
×	×	✓	×	ld
×	×	×	✓	rd

(a) Including corner areas.

middle area $Area_x$				D
lm	rm	u	d	
✓	✓	×	×	lrm
×	×	✓	✓	ud
✓	×	×	×	lm
×	✓	×	×	rm
×	×	✓	×	u
×	×	×	✓	d

(b) Not including corner areas.

After all, the directional relations are divided into 15 classes. As needed, $lrud$, lru, lrd and lrm are combined to lr; lud, lu, ld and lm are combined to l; and rud, ru, rd and rm are combined to r.

2.3 State representation

A state is represented by relative positional relations and temporal changes which hold at some time instant.

2.3.1 Relative positional relation

The relative positional relation of region X with respect to region Y at a time instant t is represented as $none(X,t)$, $\mathcal{M}^1(X,Y,t)$ where $\mathcal{M}^1$ is either *in* or *cvd*, or $\mathcal{M}^2_D(X,Y,t)$ where $\mathcal{M}^2$ is *shr*, *ec* or *dc*, and D is a directional relation at t (e.g., $shr_{ru}(X,Y,t)$).

2.3.2 Temporal change

Let G be a region or a shared part of regions.

The temporal evolution of a region's size is represented by $\mathcal{R}^s(G,t)$ where $\mathcal{R}^s$ is either *larger*, *smaller* or *same_size*, which indicate that the size of G at t has increased, decreased and almost unchanged, relative to the previous time instant, respectively.

The change of the aspect ratio (the rate of change in the edge lengths of a region) is represented by $\mathcal{R}^r(G,t)$, where $\mathcal{R}^r$ is either *hor_larger*, *ver_larger* or *same_rate*, which indicate that the ratio of horizontal change is greater than, smaller than and approximately equal to that of vertical change, compared to the previous time step, respectively.

3 EVENT DEFINITION

We call a relative positional relation, a temporal change and a temporal constraint as *a primitive fact*. Temporal constraint is either in the form $I = [I_1,\ldots,I_n]$, $t \in I$ or $t_1 < t_2$, where $I, I_1, \ldots, I_n$ are time intervals and t, t_1, t_2 are time instants. $I = [I_1,\ldots,I_n]$ indicates that I is divided into n intervals which may overlap but their beginning times are ordered. $t \in I$ indicates that t is contained in I, and $t_1 < t_2$ indicates that t_1 proceeds to t_2.

An event is defined by a conjunction or a disjunction of the primitive facts. We show the definitions of main events that will be used later [1].

3.1 Basic event definition

3.1.1 The event based on temporal change

The events related to vehicle approaching or orientation change are defined based on the temporal change. We have to carefully treat the change in size. When an object is fully visible, changes in its size may indicate approaching, leaving or a change in orientation. The first two actions can be distinguished from the last one by checking the aspect ratio. When an object is only partially visible, if regions X and Y are in an externally connected relation $ec_D(X,Y)$, where D is any directional relation, it is sometimes difficult to distinguish whether the objects are truly adjacent or one is being partially occluded by the other. Consequently, in the definitions of approaching, leaving, and orientation change, any change in size caused by occlusion must be excluded.

The following shows the representative definitions of the basic events.

- X appears from the right
  $appear\_from\_rt(X,t)$
  $\Longleftarrow none(X,t-1) \wedge shr_r(X,enc,t)$

[1]For simplicity, the following definitions may be represented using intermediate predicates.

- X approaches
  $approach(X,t)$
  $\Longleftarrow \neg\exists Y; ec_D(X,Y,t) \wedge larger(X,t) \wedge$
  $same\_rate(X,t)$
- X turns to the left or right
  $ch\_orien(X,t)$
  $\Longleftarrow \neg\exists Y; ec_D(X,Y,t) \wedge hor\_larger(X,t)$
- X is partially being occluded by Y
  $part\_occl(X,Y,t)$
  $\Longleftarrow (\ dc_D(X,Y,t-1) \vee ec_D(X,Y,t-1)\ ) \wedge$
  $ec_D(X,Y,t) \wedge smaller(X,t)$
- X keeps the same vertical position with Y
  $keep\_ver\_close(X,Y,t)$
  $\Longleftarrow (ec_r(X,Y,t-1) \wedge ec_r(X,Y,t)) \vee$
  $(ec_l(X,Y,t-1) \wedge ec_l(X,Y,t))$

3.1.2 The event with sequence of positional relations

The event of vehicle movement is defined by the sequence of changes of positional relations over a certain period of time.

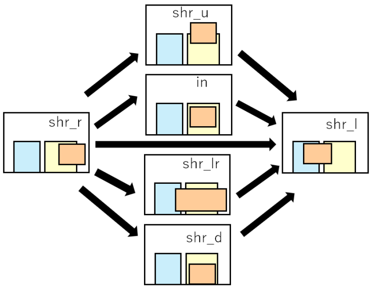


Figure 3: Lane change in the left direction.

Let X and L be the regions of a vehicle and a line, respectively. In the lane changing from right to left, X located in the right side of the region L moves to its left side. The directional relation of X with respect to L does not always change directly from right to left; instead, it may pass through intermediate states in which X is located in the upper side of L, is completely contained in L, exceeds the horizontal size of L or is located at the bottom of L (Figure 3). The third and the fourth cases sometimes appear when the action is done close to the ego-vehicle.

- X moves from the right side of L to its left side
  $move\_lane\_lt(X,L,I)$
  $\Longleftarrow (\ I=[I_1,I_2] \wedge$
  $shr_r(X,L,I_1) \wedge shr_l(X,L,I_2)\ ) \vee$
  $(\ I=[I_1,I_2,I_3] \wedge$
  $shr_r(X,L,I_1) \wedge pass(X,L,I_2) \wedge$
  $shr_l(X,L,I_3)\ )$
- passing state
  $pass(X,L,I)$
  $\Longleftarrow shr_u(X,L,I) \vee in(X,L,I) \vee$
  $shr_{lr}(X,L,I) \vee shr_d(X,L,I)$

On the other hand, X's simple movement in the left direction, not always crossing the line, can be identified considering the relative positional relation with respect to L. Each movement corresponds to a single arrow in Figure 3.

- X moves from the right to left
  $move\_lt(X,L,I)$
  $\Longleftarrow I=[I_1,I_2] \wedge$
  $(\ shr_r(X,L,I_1) \wedge shr_l(X,L,I_2)\ ) \vee$
  $(\ shr_r(X,L,I_1) \wedge pass(X,L,I_2)\ ) \vee$
  $(\ pass(X,L,I_1) \wedge shr_l(X,L,I_2)\ )$

3.2 Dangerous driving event

We define the four dangerous driving events using the above basic events.

3.2.1 Cut-in

Cut-in is a behavior in which dv in front approaches from the adjacent lane by decelerating and moves into the driving lane ahead of the ego-vehicle. The events of $approach$, $move\_lane\_lt$ and $leave$ occur in this order. In addition, ch_orien and $return\_fwd$ which is the event of X's returning to forward facing occur within $move\_lane\_lt$ action but the order of the orientation change and the beginning/end of the $move\_lane\_lt$ is nondeterministic.

- cut-in
  $cut\_in(dv,I)$
  $\Longleftarrow I=[I_1,I_2,I_3] \wedge t_1,t_2 \in [I_2,I_3] \wedge t_1 < t_2 \wedge$
  $approach(dv,I_1) \wedge$
  $move\_lane\_lt(dv,rline,I_2) \wedge$
  $leave(dv,I_3) \wedge$
  $ch\_orien(dv,t_1) \wedge return\_fwd(dv,t_2)$

3.2.2 Overtake

Overtake is a behavior in which dangerously driving vehicle appears on the right of the screen at high speed and then moves to the driving lane [2].

The definition is almost as same as that of cut-in.

- overtake
  $overtake(dv,I)$
  $\Longleftarrow I=[I_1,I_2,I_3,I_4] \wedge t_1,t_2 \in [I_3,I_4] \wedge t_1 < t_2 \wedge$
  $appear\_from\_rt(dv,I_1) \wedge$

[2] In (Takahashi and Yamaguchi, 2026), overtake is considered as a sort of cut-in.

$approach(dv, I_2) \wedge$
$move\_lane\_lt(dv, rline, I_3) \wedge$
$leave(dv, I_4) \wedge$
$ch\_orien(dv, t_1) \wedge return\_fwd(dv, t_2)$

3.2.3 Weaving

Weaving is defined as a repetition of lateral moving. Here, we define the weaving behavior of dangerously driving vehicle in front, specifically considering a single weave starting from the left side.

- weaving identified by the moving direction
  $weaving\_by\_dir(dv, I)$
  $\Longleftarrow I = [I_1, I_2] \wedge$
  $move\_rt(dv, lline, I_1) \wedge move\_lt(dv, lline, I_2)$

Although weaving is judged only by lateral move, it can be judged more accurately if we know the position relative to another vehicle in front of the dangerously driving vehicle. For example, when the dangerously driving vehicle is far from the ego vehicle or affected by the camera angle, the directional relation of dv with respect to $lline$ may be observed to remain u (upper side). In this case, weaving is not detected only by the lateral move. For this purpose, we take a region sv as the reference vehicle and consider the relative positional relations between dv and sv. It is natural to assume that the size of dv is larger than that of sv, considering the fact that dangerously driving vehicle is nearer than the reference vehicle to ego-vehicle. In addition, a wider weave indicates a more hazardous situation. In this case, dv moves significantly enough to interact with the $rline$ and completely occludes sv.

- weaving identified by reference vehicle
  $weaving\_by\_refcar(dv, sv, I)$
  $\Longleftarrow (\ t_1, t_2 \in I \wedge t_1 < t_2 \wedge$
  $part\_occl(sv, dv, t_1) \wedge disclosed(sv, dv, t_2)\ ) \vee$
  $(\ t \in I \wedge cvd(sv, dv, t)\ )$
- weaving
  $weaving(dv, sv, I)$
  $\Longleftarrow weaving\_by\_dir(dv, I) \vee$
  $weaving\_by\_refcar(dv, sv, I)$

The repetition of this action represents repetitive weavings.

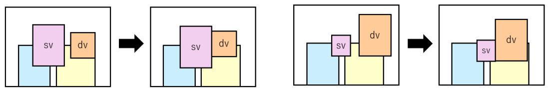
3.2.4 Squeezing

Squeezing is the behavior of a vehicle deliberately closing the distance to the side of a vehicle in the adjacent lane. Here, we define the squeezing behavior of the dangerously driving vehicle, specifically considering a squeezing on the right side of the reference vehicle.

Vertical directional relations are essential for determining whether two vehicles are driving in parallel. We give a definition using a relative directional relation of dv with respect to sv. The directional relation depends on the relative sizes of dv and sv: rm when dv is smaller and rud when dv is larger (Figure 4).

In a squeezing action, dv moves laterally from a distance towards closer to sv and externally connected; there exists a state in which dv is very close right next to sv; and the two vehicles are driving laterally close at least in consecutive time instants.

- $squeezing(dv, sv, I)$
  $\Longleftarrow I = [I_1, I_2] \wedge t_1, t_2 \in I_2 \wedge$
  $(\ (\ dc_r(dv, sv, I_1) \wedge ec_r(dv, sv, I_2)\ ) \wedge$
  $(\ (\ dc_{rm}(dv, sv, t_1) \vee ec_{rud}(dv, sv, t_1)\ ) \wedge$
  $keep\_ver\_close(dv, sv, t_2)$



(a) dv is smaller than sv (b) dv is larger than sv

Figure 4: Squeezing.

4 EXPERIMENTAL RESULTS

4.1 Experiment

We have conducted an experiment to check the validity of our event definitions.

Since there exist little video data in which an ego-vehicle is subjected to dangerous driving maneuvers, we prepare the simulation data generated from the scenario using AWSIM-Labs and Prescan, respectively. AWSIM-Labs was developed to support the development of autonomous driving systems, which provides a high-performance simulation environment for the Autoware (Autoware, 2026). Prescan is a simulation platform based on physics, developed for the development and validation of Advanced Driver-Assistance Systems and automated vehicle functionality (Siemens, 2026). As a result, we got 43 simulation data. We also collected 13 real data uploaded on YouTube (e.g., (Japan's dangerous driving reality channel, 2021)). All the videos contain the scene in which the ego-vehicle is subjected to one of the dangerous driving maneuvers.

For each video, we got an image every 0.03 seconds as data for each time instant. We extracted the region of vehicles from the video frames using YOLO (Redmon and Farhadi, 2018), and manually extracted the line regions, then transformed the numerical coordinate data of the vertices of these regions into qualitative representations. As a result, we obtained a sequence of states represented in QSR. We then checked the occurrence of events by logical reasoning. The transformation process was implemented in Python, and the reasoning for event detection was implemented in Prolog.

4.2 Result

Table 2 shows the number of videos in which the event is detected out of the total number of videos.

Table 2: Result of experiment.

event	simulation	real
cut-in	12/12	0/0
overtake	6/8	8/11
weaving	15/15	1/1
squeezing	8/8	1/1
total	41/43	10/13

The cut-in events were detected in all cases. The overtake events were not detected in five cases. The reasons for the five failure cases of the detection are as follows.

1. The dangerously driving vehicle decelerates and stops very close to the ego-vehicle in two cases. However, overtake was not detected because no leaving event occurred after the approaching event. These cases should instead be regarded as another type of dangerous driving event, namely, sudden braking, rather than cut-in.
2. The judgment is reasonable in one case, as the movement is subtle to be regarded as a dangerous lane change.
3. The orientation change was not detected in two cases. The dangerously driving vehicle very close right next to ego-vehicle had already changed its direction before the whole part appeared.

As for the weaving event, both weaving_by_dir and weaving_by_refcar succeeded in most cases. There is one case in which only the former succeeds, and one case in which only the latter succeeds. It means that both conditions are effective.

Squeezing can be detected in all cases.

5 DISCUSSION

5.1 Giving explanation

We will demonstrate how the explanations are generated using specific examples.

Example of success case. Consider a case where a cut-in event was detected. In this example, the event $cut\_in(dv, [18\text{–}19])$ is successfully identified because all five defining basic events hold: $approach(dv, [15\text{–}17])$, $move\_lane\_lt(dv, rline, [17\text{–}18])$, $leave(dv, [18\text{–}21])$, $ch\_orien(dv, 15)$ and $return\_fwd(dv, 18)$. Consequently, we assign the following semantic interpretation to this case: the dangerously driving vehicle, initially driving far ahead in the passing lane, decelerates, changes its orientation upon reaching a position close to the ego-vehicle, and then immediately moves in front of the ego-vehicle at a sharp angle before leaving. To obtain more granular reasoning — for example, why $move\_lane\_lt(dv, rline, [17\text{–}18])$ succeeded — we examine the definition of $move\_lane\_lt$. There are two cases, and in this example, $move\_lane\_lt(dv, rline, [17\text{–}18])$ is successfully detected since the primitive facts $shr_r(dv, rline, 17)$ and $shr_l(dv, rline, 18)$ are satisfied. This yields the following semantics: the dangerously driving vehicle located on the right side of the right lane line moves to its left side directly.

Conversely, in cases of detection failure, we investigate the underlying causes. We identified four main reasons for faulty judgments: (i) state matching failure, (ii) insufficient conditions, (iii) overly restrictive temporal constraints, and (iv) absence of an actual event. For cases (i) through (iii), we revised the definitions to ensure broader coverage.

Example of failure case (i). Initially, squeezing was defined as a simple state transition from either dc_{rm} to ec_{rm} or from dc_{rud} to ec_{rud}. This indicates that the dangerously driving vehicle moves laterally from a distance toward the ego-vehicle until it is externally connected. However, this definition failed to apply to videos where the state sequence dc_{rd}, ec_{rd} and ec_{rm} appeared. This sequence indicates that the dangerously driving vehicle first approaches laterally at the rear in the adjacent lane and then continues driving along the dividing line. Since this behavior constitutes a form of squeezing, we updated the definition accordingly.

Example of failure case (ii). The initial definition of *pass* was:

$$pass(X,L,I) \Longleftarrow$$
$$shr_u(X,L,I) \vee in(X,L,I)$$

While checking the basic events defining a cut-in, we found that $move\_lane\_lt$ failed since no state in the definition of *pass* was satisfied. In this example, the primitive facts $shr_r(dv, rline, 18)$, $shr_{lr}(dv, rline, 19)$, and $shr_l(dv, rline, 20)$ held. However, $shr_{lr}(dv, rline, 19)$, which indicates the dangerously driving vehicle is straddling both sides of *rline*, was not included in the definition. Such a state may appear depending on the relative speed of the dangerously driving vehicle. To account for this, we added $shr_{lr}(X,L,I)$ to the definition of *pass*. Following similar failure analyses, we also added $shr_d(X,L,I)$ to the definition of *pass*.

Example of failure case (iii). Initially, the definition of *move_lane_lt* included a strict temporal constraint between the vehicle's movement and its orientation change. We assumed the vehicle would complete its movement after facing forward. However, in practice, the inverse order may occur. Consequently, we removed this restrictive temporal constraint.

5.2 Framework of event detection

We propose our approach as a generalized framework for event detection.

Once the initial event definitions are established, the following iterative process is executed to refine them:

1. Region extraction:
   Extract regions from each video frame.
2. Symbolic transformation:
   Convert numerical data into qualitative representations (logical forms).
3. Event judgment:
   Detect events from these representations using current definitions (logical reasoning).
4. Semantic explanation:
   Generate a semantic explanation for the judgment. If the judgment is incorrect, analyze the cause.
5. Refinement:
   Revise event definitions based on the analysis.

The ability to provide semantic explanations is the most significant feature of this framework. It encompasses not only the reasoning for successful detections but also the investigation of failures, leading to more robust event rules.

Different from deep leaning-based approaches, the proposed method does not automatically construct a system capable of identifying events; instead, it relies on human logical reasoning. This makes it relatively easy to manually design and integrate the system, since QSR provides a representation that are intuitive for human interpretation.

On the other hand, a granularity problem arises as the variety of data increases. For example, we classified directional relation into 15 classes, but it remains unclear whether finer classifications are necessary, or whether additional aspects, such as camera angle or relative speed, should be introduced. Determining the appropriate level of detail is a challenging problem.

6 RELATED WORKS

Several studies based on neural network approaches utilized dashcam footage. These systems often train on normal driving data to detect significant deviations as anomalies (Yao et al., 2022). In contrast, our approach provides explicit logical definitions for specific dangerous behaviors.

QSR-based studies have been found relatively few compared to those using neural networks. Sokeh et al. applied the QSR system CORE9 to the detection of simple actions such as approaching and catching, and demonstrated that classification accuracy can be improved through learning (Sokeh et al., 2013). Weghe et al. proposed Qualitative Trajectory Calculus (QTC) to represent the relative movements between pairs of objects (de Weghe et al., 2005), which was later applied to the traffic domain by Al-Zoubi et al. (AlZoubi and Nam, 2019). In their approach, vehicle interactions are encoded as trajectories of QTC states, and a vehicle behavior recognition system is developed using Deep Convolutional Neural Networks (DCNN). Although their method effectively classifies dangerous driving, it relies on surveillance camera footage. In contrast, our study uses dashboard camera data, which requires handling occlusions caused by other objects and therefore necessitates more complex reasoning.

Suchan et al. emphasized human-centered visual explainability by developing a framework for commonsense-based video understanding using Answer Set Programming (ASP). This work was later extended into a neurosymbolic abduction method that integrates visual and semantic information for driving scenes (Suchan et al., 2019). They also employed learning-based methods to infer occlusions from action sequences (Suchan et al., 2021). Our work is closely related to theirs. However, while their primary objective is to predict potential hazards, our goal is to provide post-hoc semantic explanations that can serve as formal evidence of accidents or dangerous incidents.

Tanaka et al. proposed a method for autonomous driving systems using a description language called BBSL (Tanaka et al., 2023). Control rules for an autonomous vehicle are described using this language. BBSL can be regarded as a form of qualitative representation based on the relative spatial relation between the ego-vehicle and hazardous regions associated with surrounding vehicles. They claimed that BBSL enables more accurate and appropriate control rules compared with simple region-overlap–based approaches. However, their work focuses on vehicle control rather than event extraction. While our objective is to provide semantic interpretations of scenes, our approach can also be applied to vehicle control in future.

7 CONCLUSION

We have proposed a framework that converts video streams into symbolic representations and performs formal reasoning to derive high-level events. We demonstrated the detection of dangerous driving and provided logical explanations using QSR. This is a novel approach totally different from the one using deep learning-based methods. Our experiments validated the effectiveness of the framework.

The main advantages are threefold:

1. Low Data Requirement: It requires a minimal amount of data compared to deep learning.
2. Scalability: It is easy to introduce and define new types of events.
3. Explainability: It provides explicit logical evidence for every judgment.

Our final goal is providing semantic understanding of video data not optimizing performance metrics. This framework is applicable to fields where logical evidence and transparent reasoning are critical, such as medical imaging and robotics.

ACKNOWLEDGEMENTS

This work is supported by KAKENHI Grant Number JP24K15096 and by JST, CREST Grant Number JP-MJCR23M1.

REFERENCES

AlZoubi, A. and Nam, D. (2019). Vehicle activity recognition using mapped QTC trajectories. In *VISIGRAPP2019*.

Athanesious, J., Srinivasan, V., Vijayakumar, V., Christobel, S., and Sethuraman, S. C. (2020). Detecting abnormal events in traffic video surveillance using superorientation optical flow feature. *IET Image Process.*, 14:1881–1891.

Autoware (2026) AWSIM-Lab. https://autowarefoundation.github.io/AWSIM-Labs/main/, accessed on 2026-04-02.

Chen, J., Cohn, A. G., Liu, D., Wang, S., Ouyang, J., and Yu, Q. (2013). A survey of qualitative spatial representations. *The Knowl. Eng. Rev.*, 30:106–136.

de Weghe, N. V., Kuijpers, B., Bogaert, P., and Maeyer, P. D. (2005). A Qualitative Trajectory Calculus and the composition of its relations. In *GeoS 2005*, pages 60–76.

Garefalakis, T., Michelaraki, E., Roussou, S., Katrakazas, C., Brijs, T., and Yannis, G. (2024). Predicting risky driving behavior with classification algorithms: results from a large-scale field-trial and simulator experiment. *Eur. Transp. Res. Rev.*

Hacohen, S., Medina, O., and Shoval, S. (2022). Autonomous driving: A survey of technological gaps using google scholar and web of science trend analysis. *IEEE Trans. Intell. Transp. Syst.*, 23(11):21241–21258.

ISO (2021). ISO 34502 Road vehicles – Scenario-based safety evaluation framework for Automated Driving Systems.

Japan Automobile Manufacturers Association, Inc. (2022). Safety Evaluation Framework Ver 3.0. Tech. rep. Version 3.0.

Japan's dangerous driving reality channel (2021). Dangerous overtaking. https://www.youtube.com/watch?v=Vil9UgppyDo, accessed on 2025-11-01.

Randell, D. A., Cui, Z., and Cohn., A. G. (1992). A spatial logic based on regions and connection. In *KR92*, pages 165–176.

Redmon, J. and Farhadi, A. (2018). YOLOv3: An incremental improvement. Technical report, arXiv.

Siemens (2026). Prescan. https://plm.sw.siemens.com/ja-JP/simcenter/autonomous-vehicle-solutions/prescan, accessed on 2026-02-11.

Sioutis, M. and Wolter, D. (2021). Qualitative spatial and temporal reasoning: current status and future challenges. In *IJCAI2021*, pages 4594–4601.

Sokeh, H. S., Gould, S., and Renz, J. (2013). Efficient extraction and representation of spatial information from video data. In *IJCAI13*, pages 1076–1082.

Suchan, J., Bhatt, M., and Varadarajan, S. (2019). Out of sight but not out of mind: An answer set programming based online abduction framework for visual sensemaking in autonomous driving. In *IJCAI 2019*, pages 1879–1885.

Suchan, J., Bhatt, M., and Varadarajan, S. (2021). Commonsense visual sensemaking for autonomous driving – On generalised neurosymbolic online abduction integrating vision and semantics. *Artif. Intell.*, 299:103522.

Takahashi, K. and Yamaguchi, Y. (2026). Detection of dangerous driving event using qualitative spatial reasoning from the video. In *ICAART2026*, pages 3205–3212.

Tanaka, K., Aoki, T., Kawai, T., and Tomita, T. (2023). Specification based testing of object detection for automated driving systems via BBSL. In *ENASE2023*, pages 250–261.

Yao, Y., Wang, X., Xu, M., Pu, Z., Wang, Y., Atkins, E., and Crandall, D. (2022). DoTA: unsupervised detection of traffic anomaly in driving videos. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(1):4544–459.