Detection of Dangerous Driving Event Using Qualitative Spatial Reasoning from the Video

Kazuko Takahashi[a] and Yurika Yamaguchi
School of Engineering, Kwansei Gakuin University, 1, Gakuen, Uegahara, Sanda, Japan
{*ktaka@kwansei.ac.jp,hfo39210@kwansei.ac.jp*}

Keywords: Event Extraction, Qualitative Spatial Reasoning, Knowledge Representation, Vehicle Action Detection.

Abstract: We describe a method for extracting dangerous driving behavior using data recorded by dashboard cameras. While a machine learning approach is used in most studies, it typically requires a vast amount of data and extensive time for training each event. Furthermore, the semantics of each scene or event are not explicitly provided. To address these drawbacks, we propose a logical approach using qualitative spatial reasoning (QSR). We define a novel QSR tailored to our objectives, represent each scene through relative positional relationships between objects without numerical values, and extract events in the video by logical reasoning. We apply our method to both simulator-generated and real-world video data. As a result, we have found that it is possible to extract events correctly with only a small amount of data.

1 INTRODUCTION

It is important to extract and analyze dangerous driving from existing video data to investigate the causes of traffic accidents and realize safe autonomous driving. Many studies employing deep learning have been undertaken to detect dangerous driving of an ego-vehicle and to create control rules for avoidance. Most of them rely on surveillance cameras (Hacohen et al., 2022; Athanesious et al., 2020), while some use dashboard cameras (Yao et al., 2022; Garefalakis et al., 2024).

The approach based on deep learning is effective but has several shortcomings, such as requiring a huge amount of data and significant learning time for each event. Moreover, the reason for the judgment is usually not provided, although an explanation for what has occurred can be automatically generated. If we use such video data as legal evidence for a traffic accident or an incident of dangerous driving, it is preferable to provide a persuasive explanation that fits human recognition. However, there is currently little research from this viewpoint.

To overcome these shortcomings, a logical approach can provide explanation from a small amount of data. One such approach utilizes qualitative spatial reasoning (QSR) (Chen et al., 2013; Sioutis and Wolter, 2021). In the QSR approach, a scene is represented using relative relations between objects, such as position, direction, distance, size and so on, without using numerical data, and events occurring in the video are detected by logical reasoning.

As vehicles equipped with dashboard camera recorders are increasing in number, and since these devices are relatively cheap and readily available, we discuss dangerous driving events recorded by dashboard cameras. Video data taken by dashboard cameras are projections onto a two-dimensional plane, and the camera itself moves while driving. The data possess several intrinsic features: objects may enter or exit the screen, objects may be occluded, and the size or aspect ratio of a region may be affected by occlusion, the distance to the target object, and its orientation. There have been studies on event detection from video data in driving scenes using QSR (Cohn et al., 2014; Suchan et al., 2019). However, these features are not handled in previous QSR systems. We propose a new QSR capable of handling these features appropriately while remaining easily understood by humans.

In this study, the target kinds of dangerous driving are cut-in, weaving, and pullover. A cut-in is the behavior of a vehicle forcing its way into the path of another vehicle. Weaving is the behavior of a vehicle swaying from side to side on the road instead of moving straight ahead. A pullover is the behavior of a vehicle deliberately closing the distance to the side of a vehicle in the adjacent lane. It is easy to introduce

[a] https://orcid.org/0000-0002-5572-7747

other kinds of dangerous driving, such as tailgating and sudden braking, by adding a logical definition; additional learning phases are unnecessary.

We conduct experiments to check the feasibility of our method for cut-in events using data generated by a simulator and the real data uploaded on YouTube. As a result, we have found that it is possible to extract several basic events and cut-in events correctly. This shows that our QSR-based approach is worth using.

The goal of this study is to investigate a system that provides an explanation aligned with human recognition. More specifically, we aim to clarify how to divide the screen into areas, identify the criteria for defining events, and determine which relative relations of objects primarily influence the judgment. We do not pursue accuracy and efficiency of event detection.

This paper is organized as follows. In Section 2, we propose our representation language in QSR. In Section 3, we show the definition of several events including dangerous ones. In Section 4, we show our experiment and the evaluation of the result. In Section 5, we describe the related works. And finally, in Section 6, we conclude the paper and discuss future work.

2 REPRESENTATION OF THE SCENE

2.1 Assumptions

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.

2.2 Regions

When we use an image processing tool to extract an object automatically, its closure is usually extracted in a rectangular form as a bounding box. Therefore, it is natural to regard the rectangle as a unit entity.

We consider three kinds of regions.

1. The outer frame region: It is denoted by *enc* (enclosure), which corresponds to the entire screen.
2. Stable region: As referent regions for representing left and right positions, we use the lane lines drawn beside the driving lane. We set two stable regions: left line (*lline*) and right line (*rline*). 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).

Figure 1: Stable regions in the outer frame.

3. Unstable region: The mobile objects such as the dangerous driving vehicle (*dv*) and a safe driving vehicle in front (*sv*) are treated as unstable regions. They may enter, exit or move within the screen.

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

2.3 Qualitative Spatial Representation

We propose a novel qualitative spatial representation tailored to our objectives. For any given pair of regions, their relative positional relationship is represented by combining mereological and directional relations. Furthermore, we incorporate the temporal change in the size of a region as a key descriptive feature.

2.3.1 Mereological Relation

We define mereological relations between pairs of regions based on RCC-8 (Randell et al., 1992) but in a coarser manner. These relations are divided into four classes: *in*, *shr* and *ec* and *dc*. For a pair of regions X and Y, $in(X,Y)$ indicates X is in the inner part of Y, $shr(X,Y)$ indicates X overlaps with Y (including cases where they are internally connected); $ec(X,Y)$ indicates X is externally connected with Y; and $dc(X,Y)$ indicates X is disconnected from Y.

In cases where a region is invisible in the scene, it may either be absent from the frame or occluded by another object. To distinguish these cases, we introduce the predicates *none* and *cvd*, respectively. $none(X)$ indicates that X does not appear in the scene, and $cvd(X,Y)$ indicates that X is completely occluded by Y.

2.3.2 Directional Relation

To represent directional relations, we partition the screen into multiple areas. While this division is similar to CDC (Skiadopoulos and Koubarakis, 2004), our method for determining direction is specifically adapted for our purposes. In capturing dangerous driving behaviors, it is essential to identify whether a vehicle is deviating toward the left or right side of a lane. Our definition provides a concise representation for this requirement.

Let X and Y be regions. When mereological relation of X and Y is *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. Then the external part of Y is divided into eight areas. 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.

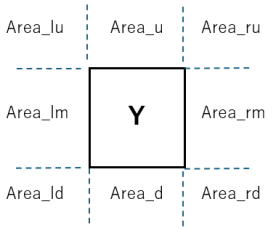


Figure 2: Area division.

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

We first determine the direction if X shares its inner part with the corner areas, which is shown in 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, ✓ 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 $D(X,Y) = lru$, and the fourth line of Table 1(b) shows if X shares its part only with $Area_{rm}$, then $D(X,Y) = rm$.

After all, the directional relations are divided into 15 classes. Considering that X is a rectangular shape, this classification covers all possible directions, ensuring that the relation for any pair of regions is uniquely determined. When focusing solely on horizontal movement, vertical information can be omitted. In this case, *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*.

Table 1: Direction relation.

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

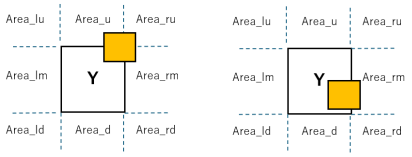
(a) Including corner areas.

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

(b) Not including corner areas.

2.4 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. Figure 3(a) and (b) illustrate examples of the X's location in case of shr_{ru} and shr_{rm}, respevtively.



(a) shr_{ru} (b) shr_{rm}

Figure 3: Examples of relative positional relations.

2.5 Temporal Change

In addition to instantaneous positional relations, we represent the temporal evolution of a region's size.

This is represented by $\mathcal{R}^s(G,t)$ where $\mathcal{R}^s$ is either $larger, same\_size$ or $smaller$, and G is a region or a shared part of regions. The predicates *larger* and *smaller* indicate that the size of G at t has increased and decreased relative to the previous time instant, respectively, while *same_size* indicates that it remains almost unchanged.

Furthermore, 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*. Here, *hor_larger* and *ver_larger* indicate that the ratio of horizontal change is greater than the vertical change (or vice versa) compared to the previous time step, while *same_rate* indicates that the changes are approximately equal.

3 JUDGMENT OF EVENTS

3.1 Basic Events

We define a basic event at a specified time.

$Event \Longleftarrow State$ shows that if $State$ is observed then $Event$ occurs. $State$ is a conjunction of relations defined in Section 2.

In the following, we show definitions of several basic events, where L represents a stable region, whereas X and Y represent unstable regions.

The appearance and disappearance of X are defined using the difference between scenes at the previous and current time instants.

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

$appear\_from\_lt(X,t)$ and $disappear\_to\_lt(X,t)$ are defined similarly.

The approach and leave of X relative to the ego-vehicle can be identified mainly by the temporal changes in the size of X. However, size changes alone are insufficient for a robust judgment. When an object is fully visible, changes in its size may indicate approaching, leaving or a change in orientation. The first two cases can be distinguished from orientation changes by analyzing the aspect ratio (the ratio of changes in the edge lengths). When an object is only partially visible, size changes do not always correspond to these specific actions. Since the actual physical dimensions of the object are unknown, the proportion of its visible area cannot be precisely determined. Furthermore, if unstable 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.

- X approaches
  $approach(X,t)$
  $\Longleftarrow \neg\exists Y; ec_D(X,Y,t) \wedge larger(X,t) \wedge$
  $same\_rate(X,t)$
- X leaves
  $leave(X,t)$
  $\Longleftarrow \neg\exists Y; ec_D(X,Y,t) \wedge smaller(X,t) \wedge$
  $same\_rate(X,t)$
- 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 returns to front facing
  $return\_fwd(X,T)$
  $\Longleftarrow \neg\exists Y; ec_D(X,Y,t) \wedge ver\_larger(X,t)$

We extend these representations of relative relations which hold at an instant, to the ones which hold during a time interval. $\mathcal{R}(X,Y,I)$ indicates that the relative relation $\mathcal{R}$ of X with respect to Y holds during a time interval I.

The expression $I = [I_1,\ldots,I_n]$ denotes that a time interval I is divided into several consecutive smaller time intervals $I_1,\ldots,I_n$, and $t \in I$ denotes that an instant t is included in I.

In the lane changing from right to left, X located in the right side of the region L moves to its left side. This action is usually executed across a time interval. The directional relation of X with respect to L is changed not only from right directly to left, but sometimes passing through the state in which up-relation holds or X is completely contained by L within the moving action (Figure 4).

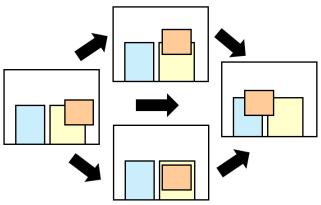


Figure 4: Moving from right to left.

- X moves from the right side of L to its left side
  $move\_lt(X,L,I)$
  $\Longleftarrow (\ I = [I_1,I_2] \wedge$
  $shr_r(X,L,I_1) \wedge shr_l(X,L,I_2)\ ) \vee$
  $(\ I = [I_1,I_2,I_3] \wedge$
  $shr_r(X,L,I_1) \wedge shr_u(X,L,I_2) \wedge shr_l(X,L,I_3)\ ) \vee$
  $(\ I = [I_1,I_2,I_3] \wedge$
  $shr_r(X,L,I_1) \wedge in(X,L,I_2) \wedge shr_l(X,L,I_3)\ )$

  The event $move\_rt(X,L,I)$ is defined similarly.

3.2 Cut-in

Here, we consider two types of cut-in. The one is the case in which dv in front approaches from the adjacent lane by decelerating and moves into the driving lane ahead of the ego-vehicle (called "cut_in_1"). The other is the case in which dv appears on the right of the screen at high speed and then moves to the driving lane, which is caused during an overtaking maneuver (called "cut_in_2"). In the latter case,

$appear\_from\_rt$ occurs first. In both cases, the events of $approach$, $move\_lt$ and $leave$ occur in this order. In addition, ch_orien and $recover\_fwd$ occur within $move\_lt$ action, but the order of the orientation change and the beginning/end of the $move\_lt$ is undetermined.

The event cut_in_2 is defined as follows.

$cut\_in\_2(dv, I)$
$\Longleftarrow I = [I_1, I_2, I_3, I_4] \wedge t_1, t_2 \in [I_2, I_3, I_4] \wedge t_1 < t_2 \wedge$
$appear\_from\_rt(dv, I_1) \wedge$
$approach(dv, I_2) \wedge move\_lt(dv, rline, I_3) \wedge$
$leave(dv, I_4) \wedge$
$ch\_orien(dv, t_1) \wedge return\_fwd(dv, t_2)$

The event cut_in_1 is defined by deleting the condition $appear\_from\_rt(dv, I_1)$ from this definition.

We consider the scenes s1, s2 and s3 where dangerously driving vehicle is located close to the ego-vehicle, which we define as close states.

$close\_state(t) =$
$shr_{rm}(dv, enc, t) \vee$
$shr_{rm}(dv, rline, t) \vee$
$(shr_{lm}(dv, rline, t) \wedge shr_{rm}(dv, lline, t))$

When a dangerously driving vehicle is located more closely, dv and enc share their lower edges. We define such scenes d1 and d2 as more dangerous states.

$dgr\_state(t) =$
$shr_d(dv, enc, t) \vee shr_{rd}(dv, enc, t)$

Close states and dangerous states are shown in Figure 5.

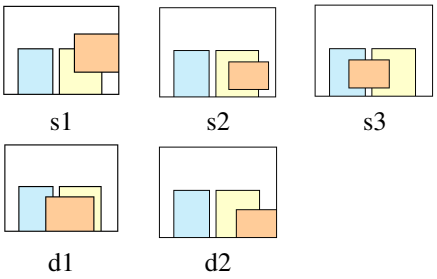


Figure 5: States of close locations (s1,s2,s3) and dangerous locations (d1,d2).

If one of close states occurs within a cut-in event, then it is judged to be a dangerous cut-in. And if one of dangerous states occurs within a cut-in event, then it is judged to be a more dangerous cut-in.

$dgr\_cut\_in(dv, I)$
$\Longleftarrow t \in I \wedge$
$(cut\_in\_1(dv, I) \vee cut\_in\_2(dv, I)) \wedge close\_state(t)$
$more\_dgr\_cut\_in(dv, I)$
$\Longleftarrow t \in I \wedge$
$(cut\_in\_1(dv, I) \vee cut\_in\_2(dv, I)) \wedge dgr\_state(t)$

3.3 Weaving

We define the weaving behavior of dangerously driving vehicle in front, specifically considering a single weave starting from the left side.

Roughly speaking, weaving consists of two consecutive actions of dv: it moves from the left side of $lline$ to its right side, and then returns to the left.

Therefore, at least the following condition should hold.

$weaving\_simple(dv, I)$
$\Longleftarrow I = [I_1, I_2] \wedge$
$move\_rt(dv, lline, I_1) \wedge move\_lt(dv, lline, I_2)$

Although weaving is judged only by this condition, it can be judged more accurately if we know the position relative to another car sv in front of dv. For this purpose, we use the mereological relations ec and dc. It is natural to assume that the size of dv is larger than that of sv, considering the fact that dv is nearer than sv to ego-vehicle. Then, weaving is defined as follows.

$weaving(dv, sv, I)$
$\Longleftarrow I = [I_1, I_2, I_3] \wedge$
$shr_l(dv, lline, I_1) \wedge dc_l(dv, sv, I_1) \wedge$
$shr_r(dv, lline, I_2) \wedge ec_l(dv, sv, I_2) \wedge smaller(sv, I_2)$
$shr_l(dv, lline, I_3) \wedge dc_l(dv, sv, I_3) \wedge larger(sv, I_3)$

dv is located to the left side of sv throughout I. sv is (partly) occluded in I_2 and disconnected from dv in I_1 and I_3 (Figure 6).

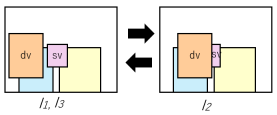


Figure 6: Weaving.

In addition, a wider weave indicates a more hazardous situation, which we define as dangerous weaving (Figure 7). In this case, dv moves significantly enough to interact with $rline$ and may occlude sv. Note that the time interval I_{23} during which sv is completely occluded by dv, may not always be present.

$dangerous\_weaving(dv, sv, I)$
$\Longleftarrow I = [I_1, I_2, I_3] \wedge I_2 = [I_{21}, I_{22}, I_{23}, I_{24}, I_{25}] \wedge$
$shr_l(dv, lline, I_1) \wedge dc_l(dv, sv, I_1) \wedge$
$shr_r(dv, lline, I_2) \wedge smaller(sv, I_{21}) \wedge$
$shr_l(dv, lline, I_3) \wedge dc_l(dv, sv, I_3) \wedge larger(sv, I_3) \wedge$
$dc_l(dv, rline, [I_1, I_{21}]) \wedge$
$shr_l(dv, rline, I_{22}) \wedge smaller(sv, I_{22}) \wedge$
$ec_l(dv, sv, I_{22}) \wedge cvd(sv, dv, I_{23}) \wedge ec_l(dv, sv, I_{24}) \wedge$
$shr_l(dv, rline, I_{24}) \wedge larger(sv, I_{25}) \wedge$
$dc_l(dv, rline, [I_{25}, I_3])$

The repetition of this action represents repetitive weavings. Weaving behavior from the right side can be defined similarly.

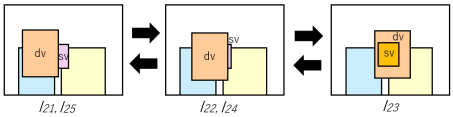


Figure 7: Dangerous weaving.

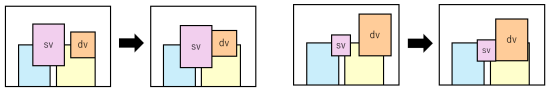
3.4 Pullover

We define the pullover behavior of dangerously driving vehicle, specifically considering a pullover from the right side of sv. (Pullover behavior from the left side can be defined similarly.)

Vertical directional relations are essential for determining whether two vehicles are driving in parallel. The definition accounts for two cases based on the relative sizes of dv and sv.

$$
\begin{aligned}
&pullover(dv, sv, I) \\
&\Longleftarrow I = [I_1, I_2] \wedge \\
&(\ (\ dc_{rm}(dv, sv, I_1) \wedge ec_{rm}(dv, sv, I_2) \ ) \vee \\
&(\ dc_{lm}(sv, dv, I_1) \wedge ec_{lm}(sv, dv, I_2) \ )
\end{aligned}
$$

This definition shows that dv and sv drive in parallel throughout I. dv is disconnected from sv in I_1 and externally connected in I_2 (Figure 8).



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

Figure 8: Pullover.

4 EXPERIMENT

We investigated the feasibility of our method for cut-in events using several videos uploaded on YouTube (e.g., (Japan's dangerous driving reality channel, 2021; Mie Television Broadcasting, 2023; Teleport San-in, 2021)). We also conducted experiments using data generated by a simulation tool provided by Japan Advanced Institute of Science and Technology (Japan Advanced Institute of Science and Technology, 2024), as there is little amount of clear data on dangerous driving in published datasets. These data include 12 cut-in scenes by a car in front (numbered as F#) and 11 by overtaking (numbered as B#).

For each video, we extracted an image every 0.03 seconds as data for each time instant. We detected the unstable regions from the video frames using YOLO (Redmon and Farhadi, 2018) and manually extracted the stable regions. We then transformed the numerical coordinate data of the vertices of these regions into qualitative representations. As a result, we obtained a sequence of states each of which consists of relative positional relations of regions and their temporal changes at each time instant. We then checked the occurrence of events. The transformation process was implemented in Python, and the reasoning for event detection was implemented in Prolog.

Table 2 shows the result of experiment.

Table 2: Result of experiment.

ID	appr	apch	lv	mvlt	ch	rtn	cutin	dgrst
F1	×	3	8	8-9	5	9	✓	s2,s3
F2	×	15	18	17-18	15	18	✓	s2,s3
F3	×	14	19	18-20	15	18	✓	s3
F4	×	14	18	18-19	14	18	✓	s2,s3
F5	×	14	19	16-19	14	18	✓	×
F6	×	16	19	18-20	16	19	✓	s3
F7	×	14	19	17-18	15	18	✓	s3
F8	×	14	19	15-19	14	18	✓	s3
F9	×	14	19	19-20	14	19	✓	s2,s3
F10	×	14	18	17-19	15	17	✓	×
F11	×	14	20	18-20	15	20	✓	s3
F12	×	13	17	17-18	13	17	✓	s3
B1	19	20	21	21-23	20	22	✓	s1
B2	14	15	17	18-24	15	18	✓	s1,s2
B3	27	27	28	28-30	28	29	✓	s1
B4	11	12	14	14-21	12	15	✓	s1
B5	17	18	19	20-24	18	20	✓	s1
B6	9	17	10	12-16	10	12	×	s1,s2,s3
B7	14	15	18	14-24	15	22	✓	s1
B8	×	31	25	44-52	×	44	×	×
B9	24	25	29	27-31	26	29	✓	s1
B10	10	11	21	14-16	14	17	✓	s1,s3
B11	×	18	×	21-28	18	24	×	d1

The figures in the table show the following meanings. For the event *appear* (*appr*), the value indicates the first time instant at which the event is observed, while × indicates that it is not observed. For the events *approach* (*apch*) and *ch_orien* (*ch*), the value indicates the first time instant at which the event is observed. For the event *leave* (*lv*), the value indicates the first time instant at which the event is observed after the occurrence of *approach*. For the event *return_fwd* (*rtn*), the value indicates the first time instant at which the event is observed after the occurrence of *ch_orien*. For the event *move_lt* (*mvlt*), the value indicates the time interval during which the event is observed. For the dangerous states (*dgrst*), all observed state labels corresponding to *close_state* and *dangerous_state* are listed.

We compared the experimental results with human observations of the scenes and evaluated how much our method has accomplished.

Basic events and cut-in events were judged almost correctly.

In case of cut-in by a car in front (F#), cut-in events were detected in all data, and the close state s3 was observed in almost all data. For data F5 and F10, cut-in events were detected, but no close or dangerous states were identified, even though the dangerously driving vehicle was actually located very close to the ego-vehicle. This issue arises from the definition of stable regions *lline* and *rline*. Currently, the lower

boundary of the stable region is set to the lower edge of the screen. However, the bonnet of the ego-vehicle is visible in the lower part of the image depending on the position or angle of the dashboard camera. As a result, the detected vehicle (dv) never shares the lower boundary with *enc*. Although the definitions of regions should be consistent across different datasets, the definition of these stable regions should be reconsidered.

In case of cut-in by overtaking (B#), cut-in events were not detected in data B6, B8, and B11, and the close state s1 was observed in all cases except for B8 and B11. For B8, this judgment is reasonable, as the movement is subtle and can be regarded as a simple lane change. For B6, the dangerously driving vehicle overtakes the ego-vehicle from a nearby position and then decelerates and stops very close to it. However, no dangerous state was detected because a certain distance still exists between the stopping position and the ego-vehicle. For B11, the dangerously driving vehicle in front forcibly cuts in from the passing lane between the ego-vehicle and another safely driving vehicle ahead in the driving lane, and then slows down. In this case, the dangerous state d1 was observed. These two cases clearly represent dangerous cut-ins, but they were not judged as cut-in events 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 behavior.

Based on the time instant or interval the events are observed in these results, we can generate semantic explanations as follows. In F2, the dangerously driving vehicle, initially driving far ahead in the passing lane, decelerates, changes its orientation after reaching a position close to the ego-vehicle, and then immediately moves into the front of the ego-vehicle at a sharp angle before leaving. In B1, the dangerously driving vehicle appears at the rear right of the ego-vehicle, approaches it, then changes its orientation immediately after overtaking, moves to a nearby position in front of the ego-vehicle while accelerating, and then leaves.

In summary, the experimental results demonstrate that our method is reasonably effective and worth further use.

5 RELATED WORKS

Several studies have explored event detection from video data using QSR. Sokeh et al. applied the QSR system CORE9 to detect simple actions, such as approaching and catching, demonstrating that classification accuracy could be improved through learning (Sokeh et al., 2013). Weghe et al. proposed Qualitative Trajectory Calculus (QTC) to represent the relative movements of object pairs (de Weghe et al., 2005), which was subsequently applied to the transportation domain by Al-Zoubi et al. (AlZoubi and Nam, 2019). In their approach, vehicle interactions are encoded as trajectories of QTC states, and a system for vehicle activity recognition is developed using Deep Convolutional Neural Networks (DCNN). While their method effectively classifies dangerous driving, it relies on surveillance camera footage (bird's eye view). In contrast, our study utilizes dashboard camera data, which introduces specific challenges such as ego-motion and complex occlusion.

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). While they utilize learning methods to infer occlusion from movement sequences, their primary objective is to predict potential dangers (Suchan et al., 2021). Our goal, conversely, is to provide a post-event semantic explanation that can serve as formal evidence for accidents or dangerous incidents.

Tanaka et al. presented a method for autonomous driving systems using a description language called BBSL (Tanaka et al., 2023). BBSL can be considered as a kind of qualitative representation based on the relative positional relation of dangerous zones of the ego-vehicle and the dangerously driving vehicle. Control rules for an autonomous vehicle are described using this language. They argued that BBSL allows for more precise and adequate control rules compared to those based on simple overlap percentages of the regions. However, their work focuses on vehicle control rather than event extraction. Our research shifts the focus toward providing a semantic interpretation of the scene, filling a gap that autonomous control rules do not address.

6 CONCLUSION

We have presented a QSR-based method for detecting dangerous driving events using dashboard camera footage. We proposed a novel QSR specifically designed for driving event extraction, which enables the generation of semantic explanations for the event detection. The proposed QSR framework is tailored for

handling driving images and utilizes a unique qualitative classification scheme that distinguishes it from existing QSR systems.

Furthermore, this logical approach serves as a promising alternative to address the limitations of deep learning-based methods, particularly in terms of data requirements and interpretability. Unlike deep learning approaches, our system does not require additional training phases to adapt to different traffic conditions, road configurations, or camera angles. For example, the constraint of straight lanes could be relaxed to handle curved roads by adapting the definitions of lane line regions based on their varying horizontal dimensions.

Our experiments, using both simulator-generated and real-world data, demonstrated that the system can detect event occurrences correctly. While the feasibility of detecting weaving and pullover behaviors remains to be validated, the results obtained for cut-in events show the viability of our approach.

Our objective is not to improve the accuracy and efficiency of event extraction, but the result shows that reasonable judgments can be achieved without requiring vast amounts of data or extensive training time. Notably, when an event is reported, our system provides a detailed context and rationale for the judgment by tracing the reasoning process backward.

In the future, we intend to validate the method against a broader range of dangerous driving events, incorporate speed factors into the reasoning process, and enhance the system to handle multiple vehicles and probabilistic movements.

ACKNOWLEDGEMENTS

This work is supported by KAKENHI Grant Number JP24K15096.

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.

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.

Cohn, A. G., Li, S., Liu, W., and Renz, J. (2014). Reasoning about topological and cardinal direction relations between 2-dimensional spatial objects. *J. Artif. Intell. Res.*, 51(1):493–532.

de Weghe, N. V., Kuijpers, B., Bogaert, P., and Maeyer, P. D. (2005). Vehicle activity recognition using mapped QTC trajectories. In *GeoS 2005*.

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.

Japan Advanced Institute of Science and Technology (2024). AW-runtime-verification. https://github.com/duongtd23/AW-Runtime-Verification, accessed on 2025-11-01.

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

Mie Television Broadcasting (2023). 67-year-old man arrested for reckless driving. https://www.youtube.com/watch?v=XeOOoyfhmFs, 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.

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

Skiadopoulos, S. and Koubarakis, M. (2004). Composing cardinal direction relations. *Artif. Intell.*, 152(2):143–171.

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.

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.

Teleport San-in (2021). Nearly a major accident. https://www.youtube.com/watch?v=a574Smv-13Q, accessed on 2025-11-01.

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.