Risk map generation system for intelligent vehicles on community roads via data-driven approach

Takuma ITO*, Masatsugu SOYA*, Kyoichi TOHRIYAMA** and Minoru KAMATA*

*The University of Tokyo
5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8589, Japan
E-mail: ito@iog.u-tokyo.ac.jp
**Toyota Motor Corporation
1200 Mishuku, Susono, Shizuoka 410-1193, Japan

Abstract
The aim of this study is to develop a risk map generation system on community roads via a data-driven approach. Because there are no roadside cooperative sensing systems on community roads, intelligent motion controls need the prediction of surrounding traffic participants based on a digital map. In addition, to improve the driver acceptance of the system, adaptive motion control considering the risk level at various intersections is desirable. In general, existing risk map preparation methods need too many human resources to prepare the map data for nationwide community roads. To solve this problem, we design a risk map generation system that can be systemized as much as possible for future automated process. First, our proposed system collects the driving data via non-specialized drivers using relatively usual vehicles at various intersections. Next, the system classifies the driving behavior via feature values with regard to the deceleration operation while approaching non-signalized intersections. Based on the driving data of relatively careful drivers, the system extracts the positions of pseudo intersections that are not registered in the digital map. In addition, the system estimates the risk level of intersections based on the driving data of relatively interactive drivers. To confirm the feasibility of the proposed system, we collect the driving data of 45 elderly drivers. According to the evaluation, the system adequately detects the pseudo intersections on actual roads. The estimation results of the risk level show the substantial agreement with the risk evaluations of driving school instructors regarding basic intersections.

Keywords: Automobile, Digital map, Driving behavior, Data driven analysis, Map deepening

1. Introduction

Traffic accidents caused by elderly drivers on community roads is one of the current major issues in Japan (Cabinet Office, 2017). Collision accidents at non-signalized intersections are one type of typical accidents on community roads (Institute for Traffic Accident Research and Data Analysis, 2016). Intelligent driving systems such as advanced driver assistance systems (ADAS) and autonomous driving are expected to prevent such accidents. Because community roads do not have roadside cooperative sensing systems such as the driving safety support system (DSSS) (Fukushima, 2011), intelligent vehicles should sense surrounding situations by themselves. Additionally, because onboard sensing systems cannot evaluate the situations behind blind areas (e.g., behind walls of roadside houses), proactive motion control systems based on digital map information are necessary (Saito and Raksincharoensak, 2016). If digital maps have sufficient information representing the risk level of a traffic situation, the corresponding proactive systems can conduct adaptive motion control. Further, it may improve the driver acceptance for such systems. Thus, digital maps containing not only the existence information of risk but also information about the risk level are desirable. In addition, to realize the practical use of such proactive systems on nationwide community roads, digital maps generated via an automatized process are necessary. Given these motivations, we aimed to develop a risk map generation system via a data-driven approach. Specifically, we aimed to develop the method extracting the position of pseudo intersections, and the method
estimating the risk levels of intersections. The following parts of this paper are organized as follows. Chapter 2 describes the system design including potential applications, expression of risk levels, and method generating risk map. The experiment of the driving-data collection is described in chapter 3. An evaluation of the generated risk map is presented in chapter 4. Finally, chapter 5 offers a conclusion and an outlook on future research.

2. System design

2.1 Adaptive motion control based on map-based risk information

To discuss the requirements of a risk map, we took the motion control system proposed by Saito et al. as an example for proactive braking intervention systems (Saito and Raksincharoensak, 2016) with a LeanMAP framework, which is a digital map framework proposed in our previous study (Ito et al., 2016). The motion control system calculates time-series reference velocity while approaching the intersection based on the relative position to the blind intersection and the positions and velocities of imaginary pedestrians behind the occluded area. Regarding the relative position to an intersection, localization systems with digital maps and onboard sensors can calculate the time-series variables. However, because intelligent vehicles cannot grasp the position information of pedestrians behind occluded areas, assumption are necessary. On this point, if the intelligent vehicle assumes a constant behavior for pedestrians in occluded areas at blind intersections, the intelligent vehicle conducts exactly the same proactive braking intervention in various driving situations. However, constant braking interventions might decrease the driver acceptance because his or her perceived potential risk varies depending on the driving situation. Thus, although proactive systems cannot perfectly predict the existence and behaviors of pedestrians, assumptions reflecting the risk level of each intersection to a certain degree are desirable.

2.2 Existing risk information

To date, various approaches regarding the risk while driving vehicles have been discussed in the automobile research field. For example, various indices such as TTC and THW were proposed for the risk estimation of driving states (Kitajima et al., 2009). The indices have two characteristics: the time-series continuous values and the index calculated via the relative position and velocity between the ego vehicle and preceding vehicle. Regarding risk estimation based on driving environments, Saito et al. analyzed a near-miss incident database and developed a regression formula based on the driving context (Saito et al., 2018). Because their approach focused on various elements of driving environments, it estimated continuous risk values relatively well. For risk estimation based on the driver behavior, Kamon et al. proposed a method for evaluating time-series continuous risk levels based on the eye-movement and pedal operation of the driver (Kamon et al., 2010). Although various approaches have been proposed, the source data for risk estimation vary according to their purposes. In addition, output formats vary (e.g., expressed as continuous time-series values or discretized values). Thus, this study needed to focus on an adequate design from these viewpoints considering the purpose.

2.3 Concept design for risk map

To combine map-based risk information for non-signalized intersections with real-time sensing information (e.g., detection results of surrounding traffic participants), we must avoid information interference to stabilize the motion control. The spatial resolution of map-based risk information as secondary information needs to be roughly discretized because the time resolution of sensor information is relatively fine. Thus, we assumed a risk map that contains one risk level information for each entrance of an intersection. Figure 1 shows a schematic of risk level discretization at various intersections on a community road. Each intersection has risk level information for each entrance. Hence, a normal intersection possesses four risk data, whereas a T-junction has three risk data. The risk information primarily depends on the geometrical shape and visibility of the intersection. For example, the risk level of a blind intersection is basically high, whereas the risk level of an intersection with good visibility is basically low. However, not only geometrical elements but also knowledge-based elements are important to create a useful risk map.
For spatially discretizing the risk level of intersections, the elaborate expression of a risk level, which might be a continuous value with small resolution, is difficult. Thus, we focused not only on a spatially but also on a quantitatively discretized expression of risk levels. Because determining how many levels are adequate for a discretized expression of a map-based risk information was difficult, we simplified the expression as a first step. To be more precise, we assumed three risk levels: high, medium, and low risk levels.

2.4 Approaches for risk map generation

Existing methods for the estimation of map-based risks are roughly classified into three approaches. The first one is a rule-based approach with a precise digital map including various information regarding elements in traffic environments. For example, Armand et al. proposed the ontology-based context-aware system (Armand et al., 2014). Although the drivers can easily understand the risk mechanism in the rule-based approach, the method needs precise digital map information regarding the driving environment, which requires a vast amount of resources for preparing the digital map contents. The second approach is based on expert knowledge. For example, Hasegawa et al. analyzed the driving data of expert drivers in terms of environmental factors and developed a regression model that outputs the reference velocity (Hasegawa et al., 2011). Because the reference velocity represents the risk level of a driving environment to a certain degree, we think that this approach is one of risk map generation methods. Although the driving data of expert drivers is a reliable source for the generation of risk maps, this approach needs many human resources of expert drivers to prepare the risk maps of nationwide community road. The third approach is a data-driven one that collects large amounts of data via non-specialized vehicles. For example, Honda Motor Co. Ltd. developed SAFETY MAP, which collects driving data containing sudden brakes (Honda Motor Co., Ltd., 2013). This approach can reduce the quantity of necessary human resources because the data is collected by their car navigation systems.

In this study, we aimed to develop an as much as possible systemized method for future automated process to reduce the specialized human resources necessary for preparing a risk map for community roads. Thus, a data-driven approach based on non-specialized drivers using usual vehicles was the preferred method. In this context, “specialized human resources” means that, for instance, instructors of driving schools help in the preparation of the risk map. An approach with specialized human resources costs too much to be continuously updated with new information. However, although a data-driven approach requires human resources, focusing on non-specialized drivers and usual vehicles can enable a continuous update of the risk map with reasonable resources. Thus, we selected this approach.

Regarding the data source, driving data featuring sudden brakes contain not only meaningful data but also noise that is not related to the risk of intersections but based on the behaviors of unsafe drivers. Thus, a selection of driving data in terms of the characteristics of driving behavior is necessary. Specifically, we focused on the drivers’ carefulness and interactivity to the driving situation. To evaluate these characteristics, we needed to evaluate the various feature values of the driving behavior regarding deceleration except for those related to sudden brakes on the way toward intersections. Considering these points, we implemented the risk map generation system.

2.5 Implementation of risk map generation system
2.5.1 Base system of digital map framework

We chose the LeanMAP framework as a base platform for the digital map to reduce the quantity of preparation resources (Ito et al., 2018). Because LeanMAP is based on the architecture of the road network defined by the Digital Road Map (DRM), which is used for the contents of car navigation systems, LeanMAP contains the position
information of intersections to a certain degree. However, it does not contain existence information of two kinds of pseudo intersections. The first one is the pseudo intersection in which a pedestrian road is connected to a normal road. The second one is the pseudo intersection in which the exit of a private space, e.g., a car parking lot, is connected to a normal road. Figures 2 and 3 show examples of both cases, respectively. The road network defined by DRM does not contain information regarding these pseudo intersections because the pedestrian road and the parking lot are not roads. Thus, the positions of pseudo intersections must be extracted from collected driving data.

Another challenge for the risk map generation based on the LeanMAP is the estimation of an adequate risk level for each intersection. Because LeanMAP includes only the existence information of intersections and not detailed information about roadside constructions, visibility, and pedestrian traffic, we could not directly estimate the risk level from LeanMAP contents. Thus, we had to estimate the risk level using collected driving data. In these ways, above-mentioned two functions were required for the risk map based on LeanMAP.

2.5.2 Overview of generation flow

Based on the basic concepts discussed in the above sections, we developed a risk map generation system that has the functions of extracting pseudo intersections and estimating risk levels. Figure 4 shows a conceptual flowchart of the risk map generation. The process is roughly divided into seven steps starting from driving data collection to the estimation of risk levels. The details of each step is described in the following subsections. Steps 1-5 are conducted for the driving-data classification. We classified the driver types into three categories: careful, careless, and interactive drivers. The driving data from constantly careful drivers seem effective for the extraction of the positions of pseudo intersections because the drivers pay sufficient attention to the driving situations. On the contrary, because the driving data by constantly careless drivers is not useful for the risk map generation, we want to exclude these data. Additionally, because the remaining drivers are neither constantly careful nor careless, we think that they are interactive drivers to the driving situations. Thus, their driving data is used for the risk level estimation.

To classify the driver types via a data-driven approach, we focused on frequent types of driving behaviors with regard to various intersections. If a driver shows a careful/careless behavior constantly for many intersections, he/she is likely a careful/careless driver. Otherwise, he/she is likely an interactive driver because he/she changes his/her driving behavior depending on the situations. Thus, we focused on his/her frequent type of driving behavior, which can be
determined via the mode type at various intersections. In addition, from a viewpoint of data-driven approach, we assumed that the driving behavior type at a certain intersection can be determined via the mode type of multiple driving behaviors at the intersection. Figure 5 presents the conceptual schematic of the process classifying driver type based on the discussions above. In this process, how to determine the type of driving behavior for a certain intersection is a difficult problem. On this point, we considered two approaches: one based on threshold values regarding various feature values, and one based on an automated clustering method. Because the standard characteristics of driving behavior vary depending on the driving situations and local transportation cultures, determining universal threshold values for each feature value is difficult. Thus, this approach is not appropriate for future automated process. Therefore, we focused on the clustering method considering many drivers. This approach classifies the relative types of driving behavior. Although the relative classification is not as adequate as the absolute classification, increasing the sample number can improve the adequacy of classification results.

2.5.3 Driving data collection

As the source for risk map, the proposed system requires a large quantity of driving data. Because we thought that the use of extraordinarily specialized experimental vehicles with various sensors was impractical, we needed to use relatively normal experimental vehicles and the sensors that are currently common or can be replaced with onboard sensors expected in the near future. Figure 6 shows our experimental prototype vehicle for future applications. The system obtains basic vehicle information from control area network (CAN). We added the following additional sensors for the data collection: cameras for detecting surrounding vehicles and pedestrians, LeanMAP systems with GPS for grasping vehicle positions, and photoelectric sensors for grasping the driver’s foot position. As for the camera for detecting surrounding traffic participants, we used Mobileye 560, which could detect surrounding vehicles and pedestrians. This system might become practical in the near future because of the expected costs of the additional sensors.

Fig. 6 Experimental prototype vehicle.
Figure 7 shows the appearance of photoelectric sensors embedded into acceleration and brake pedals. The microcomputer, which is connected to the sensors, considers that the driver’s foot is on a pedal if equal to or more than one sensor embedded into the pedal reacts. Thus, the microcomputer can grasp time-series binary information for each pedal: whether the driver’s foot is on an acceleration pedal or not, and whether it is on a brake pedal or not. Figure 8 shows an example of time-series data obtained from the experimental vehicle while approaching an intersection.

Fig. 7 Photoelectric sensors embedded into acceleration and brake pedals. We embedded two sensors into acceleration pedal and three sensors into brake pedal.

Fig. 8 Example of time-series data obtained from the experimental vehicle. The first row shows the time-series acceleration pedal ratio which was obtained from CAN. The second and third row show the time-series outputs from photoelectric sensors. The fourth row shows the time-series brake pedal switch which was obtained from CAN. If a driver presses the brake pedal even if only slightly, this switch outputs 1. The fifth row shows the time-series velocity which was obtained from CAN. The sixth row shows the time-series offset which was obtained from LeanMAP. The offset is a position expression of the LeanMAP framework, which indicates a distance measured from the origin point along the road.
2.5.4 Exclusion of invalid data

We focused on the deceleration operation of the driver while approaching intersections. In this situation, the existence of other traffic participants such as vehicles and pedestrians are unnecessary factors affecting the driving behavior. Thus, to exclude the effect from other traffic participants, the system excludes data in which other traffic participants coexist. To be more precise, the system excludes data if the experimental vehicle detects other vehicles or pedestrians within 30 m in the target area by the camera. The information from Mobileye camera is used for the exclusion process. The area is defined as segment starting at 30 m in front of the entrance of the intersection and ends 10 m behind the entrance. The position information of intersections is obtained from LeanMAP system. Figure 9 shows the concept of a target area surrounding an intersection.

Fig. 9 Schematic of target area surrounding intersection.

2.5.5 Extraction of feature value

First, to grasp the deceleration operations of the driver while approaching intersections, the system converts the sensor values regarding the brake and acceleration pedals to the information of the “Pedal State” of the driver. Figure 10 shows the time-series example of data conversion. Figure 11 shows the flowchart for the discretization of the pedal state. Then, to evaluate the driving behavior in terms of deceleration, we focused on three feature values. Figure 12 shows example of feature values. The first feature value is the release position of the acceleration pedal. The system extracts the relative position at which the “Pedal State” finally changes from 3 to 2 in the range of 30 m in front of the intersection. Based on the information from CAN and sensors, the system defines the position of “Acceleration pedal off.” If the driver continues to press the acceleration pedal while approaching the intersection, the “Pedal State” keeps 3. In this case, because the system cannot define the position of “Acceleration pedal off,” the system assigns 0 into FV1. Thus, this value is determined between -30 and 0. The second feature value is the average approaching velocity before “Acceleration pedal off.” The system calculates the average approaching velocity in the range of 10 m in front of the point of “Acceleration pedal off” to the point of “Acceleration pedal off.” The third feature value is the minimal velocity in the range of the entrance of the intersection to 10 m behind the entrance. Although various candidate variables exist as feature values, we selected these values based on a preliminary investigation.
Fig. 10 Example of time-series data conversion from sensor data to pedal state. The blue, green and orange arrow marks represent the timing of “Pedal State” change. Regarding the second row, the system binarizes the acceleration pedal state from the acceleration pedal ratio, which was obtained from CAN and shown in the first row, because the system could not obtain the binarized acceleration pedal state directly from CAN. Regarding the sixth row, “Pedal State” indicates the pedal operations of the driver. As for states 1 and 2, “Foot on the pedal” means that the driver’s foot is located on the pedal without pressing it. Based on state information of both pedals and sensor information from both pedals, the system discretizes the pedal state.

Fig. 11 Flowchart for discretization of the “Pedal State.” If all answers of the process in the flowchart are “No,” the system assigns the same state as time-series previous one to the current state. This situation indicates pedal position changes from acceleration/brake pedal to brake/acceleration pedal.
Fig. 12 Example of feature values (FV). To discuss the feature values related to the intersection from the spatial viewpoint, the horizontal axis represents the offset value. The example data shown in Fig. 12, the horizontal axis of which is offset, corresponds to that shown in Fig. 8, the horizontal axis of which is time.

### 2.5.6 Clustering of driving data

The system uses Ward’s method (Ward, 1963) for the clustering and creates a dendrogram using the standardized values of the feature values. In this context, standardization means the scaling process that makes an average value of the measured values 0, and a standard deviation of them 1. Figure 13 shows an example of a dendrogram. Based on the distribution in the dendrogram, the system determines the line classifying the data into three clusters (orange line in Fig. 13). The system conducts this classification process for each intersection. Because the distance of classifying line varies for every intersection, this approach reflects the characteristics of a driver behavior at various intersections.

![Dendrogram via Ward’s method](image)

*Fig. 13 Example of dendrogram via Ward’s method. The horizontal axis indicates the data distribution, whereas the vertical axis indicates the distance of feature values to the contiguous sub-clusters.*

### 2.5.7 Estimation of driver type

Although Ward’s method can classify the data into three clusters, it cannot rank clusters. In other words, the system cannot directly understand which clusters are “careful”/“careless.” Thus, we had to discuss the ranking method. Because the clustering results are not based on the original values of feature values but on their standardized values, the distribution ranges of each feature value do not differ so much from each other. Thus, we focused on a non-weighted summation of average standardized feature values of the clusters. For example, a smaller non-weighted summation value indicates an earlier “Acceleration pedal off” situation, slower approaching velocity, and slower minimal velocity. In short, a smaller value indicates a more careful driving behavior. By contrast, a larger value indicates a later “Acceleration pedal off” situation, faster approaching velocity, and faster minimal velocity, which indicates a more careless driving behavior. Hence, the index reflects characteristics of careful driving to a certain degree. Although a weighted summation has the possibility to realize a better ranking method than a non-weighted one, deciding universal weight values for various intersections is difficult. Thus, the approach using the weighted summation is not appropriate for future automated process. Therefore, we selected the non-weighted approach for the cluster ranking. Based on the ranking results of the summation values of the clusters, the system determines the type of individual driving data at a certain intersection. Then, the system determines the mode type of multiple driving behaviors at an intersection and
repeats this process for all intersections. Figure 14 shows a conceptual schematic of the abovementioned process. Afterward, the system estimates the driver types by determining the mode types of the driving behavior at all intersections. Figure 15 shows the conceptual schematic of determination of the mode types of a certain driver.

2.5.8 Extraction of pseudo intersections

To extract the positions of pseudo intersections, the system focused on the time-series driving data of careful drivers. To be more precise, the system focused on their pedal operations for the preparation of the deceleration in the segment except for the following situations: the areas in which the data of an intersection is registered, area around a curve, slope areas, and area around a stop line. We thought that the preparation of a deceleration process could be used as a clue for the extraction of pseudo intersections. Thus, the system calculates the ratio of drivers who prepare decelerations among the careful drivers, and picks up the area where the ratio is more than a threshold value. Regarding the preparation operation for deceleration, there are two candidates: “Acceleration pedal off” (from pedal state 3 to 2); and “Foot position change” from acceleration pedal to brake pedal (from pedal state 2 to 1). Based on the preliminary
investigation, we found that some drivers conducted “Acceleration pedal off” even at non-intersection parts for controlling the velocity. Thus, if we selected “Acceleration pedal off” as the index, false extractions of pseudo intersections might increase. Therefore, we selected “Foot position change” (from pedal state 2 to 1) as the index to extract pseudo intersections.

2.5.9 Estimation of risk level

To estimate the risk level, we focused on the pedal operation of interactive drivers. In contrast to the extraction of pseudo intersections, the system calculates the ratio of drivers who release the acceleration pedal (from pedal state 3 to 2). Figure 16 shows a conceptual schematic of the percentage of “Acceleration pedal off” and the maximal value at each intersection. This ratio reflects the knowledge-based risk level of the intersection as well as visibility around the intersection: a larger number of interactive drivers who release the acceleration pedal indicates riskier intersections. However, this ratio is a relative index indicating risk level in the town where the data is collected because the frequency of such behavior depends on the local traffic culture and education. Figure 17 shows a conceptual example of maximal percentage of “Acceleration pedal off” at each intersection of a town where the local traffic culture and education are good, whereas Fig. 18 shows another example where those are not good. Because I1 shown in Fig. 17 shows both absolutely and relatively large value, I1 is likely to be a riskier intersection than other intersections. Similarly, because I19 shown in Fig. 18 shows the both absolutely and relatively small value, I19 is likely to be a safer intersection than other intersections. On the other hand, I11 shown in Fig. 18 seems a bit complicated. Although I11 shows the absolutely small value, it shows larger value than other intersection in the same town. As mentioned above, this ratio is a relative index in the same town under the same local traffic culture and education. Thus, we considered that I11 was likely to be a riskier intersection. Similarly, we considered that I9 shown in Fig. 17 was likely to be a safer intersection.

To implement this concept, the standardization and discretization for this ratio are necessary. First, the system calculates the average value and standard deviation of the ratio of the drivers who released the acceleration pedal in the vicinity of the intersection. These values are calculated among the data in the same town. Next, the system discretizes the ratios for all intersections using the average value and standard deviation. If the ratio at a certain intersection is more than the average value plus 0.431 times the standard deviation, the system classifies the risk level of the intersection as high. If the ratio is less than the average value minus 0.431 times the standard deviation, the system classifies the risk level as low. For other intersections, the system determines the risk level as medium. Because the average value plus or minus 0.431 times the standard deviation can divide the normal distribution into three equal segments, we determined this threshold value.

![Fig. 16 Conceptual schematic of the percentage of acceleration pedal off and the maximal value at each intersection.](image-url)

The system calculates the percentage among the valid data of interactive drivers for every offset value, and determines the maximal value in the range of the target area of each intersection, which was discussed in the subsection 2.5.4, as the representative ratio.
3. Experiment for driving data collection

To collect driving data as source for a risk map, we conducted experiments using an actual vehicle on public roads. The following protocol was approved by the institutional review board for human studies of the University of Tokyo. We explained the protocol of the experiment to the participants and obtained their consent.

3.1 Participants

Because we focused on the varieties of driving behavior, participants with various driving characteristics were preferred as experimental participants. On this point, in general, the variety of driving behavior of young drivers is small owing to the short driving experience from the graduation of driving schools. By contrast, the variety of driving behavior of elderly drivers is large owing to the habituation along with the long driving experience (Suzuki, 2011). In addition, declines of various physical functions due to aging are also one of the reasons of the large variety (Akamatsu, 2019). Thus, in this study, we recruited 45 elderly drivers as experimental participants. They lived in Kashiwa city where the experimental courses existed. Their average age was 72.8 years old (SD = 3.7 years). They had driving licenses for 44.5 years on average (SD = 11.3 years). Their average driving frequency was 4.1 days per week. In addition, we recruited three driving school instructors to confirm the reference information regarding the risk level at intersections. Table 1 lists their basic information.

<table>
<thead>
<tr>
<th>Instructor</th>
<th>Age</th>
<th>Experience as an instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor A</td>
<td>41</td>
<td>16 years</td>
</tr>
<tr>
<td>Instructor B</td>
<td>41</td>
<td>16 years</td>
</tr>
<tr>
<td>Instructor C</td>
<td>56</td>
<td>26 years</td>
</tr>
</tbody>
</table>

3.2 Experimental course

We chose three experimental courses to collect driving data. All courses existed in the Kashiwa city. Figures 19, 20, and 21 show the route maps of experimental courses 1, 2, and 3, respectively. These figures are based on the map images published by the Geospatial Information Authority of Japan (Geospatial Information Authority of Japan, 2018). The orange areas indicate the positions of buildings such as houses. As shown in figures, approximately all parts of the experimental courses were community roads. Thus, the velocity was basically limited to 30 km/h. Table 2 lists the basic information of each experimental course. In each course, we set target intersections, which red circles indicate in each figure, where the vehicle drove thorough straightforwardly. In addition, two pseudo intersections were considered: P1
in course 1 and P2 in course 3. The situations shown in Figs. 2 and 3 correspond to P1 and P2, respectively.

![Fig. 19 Route map of course 1.](image1) ![Fig. 20 Route map of course 2.](image2) ![Fig. 21 Route map of course 3.](image3)

Table 2 Basic information of experimental course.

<table>
<thead>
<tr>
<th>Course</th>
<th>Length</th>
<th>Number of intersections</th>
<th>Number of pseudo intersections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course 1</td>
<td>Approximately 1.7 km</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Course 2</td>
<td>Approximately 0.9 km</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Course 3</td>
<td>Approximately 1.8 km</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

3.3 Experimental tasks

Participants drove the experimental vehicle three times in each course according to the following instructions:

- Obey the speed limit.
- Slow down if there are other traffic participants (e.g., pedestrians and cyclists) near the ego vehicle.

Further, the instructors were asked to express subjective risk levels in terms of high, medium, or low after all trials. The evaluation results were collected as reference and not as data source.

4. Experimental results

4.1 Estimation of driver type

Figures 22 and 23 show the examples of distributions of standardized feature values at intersection I2 in course 1. Because we measured three trials for 45 participants at each intersection, we collected 135 trials of driving data. Regarding intersection I2, 114 driving data were analyzed because the remaining 21 driving data satisfied the exclusion conditions. Because the non-weighted summation of average standardized feature values, which meant a summation of average standardized FV1 + average standardized FV2 + average standardized FV3, of cluster 3 was minimal among the clusters, the driving data of cluster 3 were classified as careful driving behavior. Similarly, the driving data in clusters 1 and 2 were classified as careless and interactive driving behaviors, respectively. Based on these results, the system calculated the mode type for each participant at this intersection. However, for few participants, the system could not determine the mode type because driving data of three trials were classified equally as different types. In addition, for participants whose driving data were excluded, the system could not select a mode type. For both cases, the system determined the last valid driving data as representative data for the participant at the intersection because we thought that it reflected the characteristics of the driving behavior most owing to the habituation.
Similarly, the system determined every mode type for the driving behavior at each intersection. Afterward, the system determined a representative type of driving behavior for each participant by picking up the mode type at all intersections. As a result, 20, 16, and 9 participants were classified as constantly careless, interactive, and constantly careful drivers, respectively.

### 4.2 Extraction of pseudo intersections

Figure 24 shows the percentage of “Foot on brake pedal” among careful drivers in course 1. The system extracted the segments in which the ratio exceeds the threshold value. In this study, we determined the threshold value as 20% through the trial-and-error approach for reducing false extractions at non-intersection parts due to the incidental velocity control, and missed extractions at actual pseudo intersections. The gray areas indicate non-target segments in which the participants needed to decelerate owing to a stop line or to turn around a corner. The green areas indicate segments in which the ratio exceeds the threshold value near the non-target segments and already known intersections. The orange area indicates the remaining segment in which the ratio exceeds the threshold value. These areas correspond to the position of pseudo intersection P1. Thus, the system could extract the pseudo intersection adequately via a data-driven approach. Further, the system could detect the position of P2 in course 3 and another pseudo intersection P3, which we did not anticipate before the experiment, in course 2. Figure 25 shows the appearance of P3. The detected position of P3 corresponds to the exit of a park, which was a potentially risky situation.
Fig. 24 Percentage of “Foot on brake pedal” in course 1 among careful drivers. The horizontal axis indicates the offset, whereas the vertical axis indicates the ratio.

Fig. 25 Appearance of the newly extracted pseudo intersection P3. There is an exit of a park, as indicated in the orange dotted line.

4.3 Estimation of risk levels of intersections

Figure 26 shows the percentage of “Acceleration pedal off” among interactive drivers in course 1. Based on the map information, the system set the segments to focus on the local maximal for the ratio of “Acceleration pedal off.” In these results, we found that results regarding I4 and I11 might be increased by the experimental conditions of the routes. Although environmental factors such as visibility and shape of intersection I4 were approximately equivalent to those of I3, the ratio of I4 is larger than that of I3. Similar tendencies can be confirmed for I10 and I11. Because participants turned at the next intersections behind I4 and I11, some participants might have prepared the deceleration. If the experimental routes were different, the preliminary deceleration might not have occurred. To correct this effect, we determined a corrective coefficient to decrease the ratios for I4 and I11; the ratios for I4 and I11 were decreased by approximately 10% according to the determined balances at the corresponding intersections. In addition, we applied the same corrective coefficient to the ratios for I2, I5, I7, I13, I17, I19, I21, I26, and I27, behind which participants turned at the next intersections. Because the corrective coefficient was a tentative one, we need to discuss a more universal approach in the future.

Fig. 26 Percentage of “Acceleration pedal off” in course 1 among interactive drivers. The horizontal axis indicates the offset, whereas the vertical axis indicates the ratio.

Figure 27 shows the maximal percentages of “Acceleration pedal off” for all intersections. Based on the results, the system calculated average values and standard deviations of the ratios. The system discretized the risk level at each intersection based on the dotted line. Table 3 shows the risk estimation results. In the results, we found that estimation results of pseudo intersections with regard to pedestrian roads (I1, P1, and P2) obtained relatively lower risk estimation than the evaluations by instructors. Figure 28 shows intersection I1. Similar to the cases of P1 and P2 (Figs. 2 and 3), there was a pedestrian road, which the participants hardly noticed. Because the effect of these barely visible pseudo
intersections was only slightly reflected in the driving behavior, we thought that the relatively low risk estimations were confirmed. Thus, a correction method for the estimated risk level regarding barely visible pseudo intersections must be discussed in the future.

Fig. 27 Maximal percentage of “Acceleration pedal off” for all intersections. The dotted line indicates the average line, whereas the dashed lines indicate the average value plus or minus 0.431 times the standard deviation.

Table 3 Comparison of risk evaluation between proposed system and instructors. L, M, and H in the table indicate a low, medium, and high risk, respectively. In addition, the lower table row presents the risk evaluation results of the instructors. The mode was chosen based on the evaluation results. “-” in the gray cell indicates that we could not choose a mode because the evaluation results of the three instructors were different from each other.

| Number of intersections | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | P1 | P2 |
| Risk estimation by the system | L | H | L | M | M | L | H | L | H | L | L | L | M | L | M | L | H | L | H | M | H | H | H | M | M | L | L |
| Risk evaluation by instructors | H | H | L | L | - | M | H | L | L | L | M | H | L | L | L | - | M | H | H | H | L | M | M | L | M | M |

Fig. 28 Appearance at intersection I1. The barely visible pedestrian road is connected to the normal road at left side, as indicated in the orange dotted line.

To evaluate the total performance of the proposed system, we calculated Cohen’s Kappa value (Fleiss et al., 2003), which evaluates the degree of the observers’ agreement regarding categorical data, between the risk estimations of the proposed system and the risk evaluations by the instructors. Because the estimation results were not on a nominal scale but on an ordinal scale, we calculated the modified Kappa value with linear weights. For the 22 intersections that include I1, P1, and P2, \( \kappa = 0.5489 \) (\( p = 0.0054 \)), which indicates a moderate agreement (Landis and Koch, 1977). For the 19 intersections that exclude them, \( \kappa = 0.6724 \) (\( p = 0.0008 \)), which indicates a substantial agreement. Therefore, as for the basic intersections, our proposed system achieved the risk estimation that had substantial agreement with risk evaluation of the instructors via a data-driven approach. However, further challenges exist regarding barely visible pseudo intersections.

5. Conclusion

We developed a risk map generation system for non-signalized intersections on community roads via a data-driven approach. By using an automatized clustering method based on the deceleration behavior in the vicinity of intersections, the proposed system picked up the relatively careful and interactive drivers, respectively. Based on the driving data of careful drivers, the system extracted the positions of pseudo intersections, the information of which was
not included in the digital map based on the road network defined by DRM. Regarding the driving data of interactive drivers, the proposed system estimated the risk level of intersections on community roads. Through the experiments on actual community roads, we confirmed a substantial agreement of the risk estimation results for basic intersections with the reference risk evaluations by the instructors of the driving school.

However, future work is still needed. Although the proposed system could estimate the risk level of basic intersections, the system outputt lower risk estimations at pseudo intersections with barely visible pedestrian roads. Hence, a correction method with a hybrid approach based on driving data of both relatively interactive and careful drivers must be discussed in the future. In addition, we need to confirm the usability of the generated risk map in combination with motion control systems. A confirmation through user tests is one of our future tasks.

Acknowledgement

This research has been conducted as a part of the research project “Autonomous Driving System to Enhance Safe and Secured Traffic Society for Elderly Drivers” granted by Japan Science and Technology Agency (JST), S-Innovation (Strategic Promotion of Innovative Research and Development). The authors would like to thank the agency for providing financial support.

References

Saito, Y. and Raksincharoensak, P., Shared Control in Risk Predictive Braking Maneuver for Preventing Collisions

