

# Building Safety Road Maps Based on Difference of Judgment of Road Users by their Smartphone

Viet Chau Dang

Department of Computer Science  
National Defense Academy of Japan  
Yokosuka, Kanagawa, Japan

Masao Kubo

Department of Computer Science  
National Defense Academy of Japan  
Yokosuka, Kanagawa, Japan

Hiroshi Sato

Department of Computer Science  
National Defense Academy of Japan  
Yokosuka, Kanagawa, Japan

Akira Namatame

Department of Computer Science  
National Defense Academy of Japan  
Yokosuka, Kanagawa, Japan

**Abstract**—Recently, there has been a growing demand and interest in developing methods for analyzing smartphone logs to extract traffic safety information. Because the log is high time resolution and closely related to user activities but fragmentary and myopic, it is difficult for currently available collision probability based quantitative risk assessment methods to create traffic safety maps automatically from the driving log which require all of concrete information about a collision for example, size of vehicle, speed of pedestrian. This paper proposes a computable risk measurement method for building traffic safety maps with the logs of different users' driving, which does not discuss collision probability. The proposal is designed to compute differences in the recognition of the road environment among road users mathematically. Drivers differ in their recognition, judgment, and handling of a given situation. Suppose that a difference in recognition among users in the same situation is a signal of danger. This signal is easy to calculate by Poisson process. Each user's recognition of road environment and the safety map integrated from the collection of the recognition are generated fully automated. A real-world experiment was carried out, and the results show that the assumption and the proposed method succeeded in generating an accurate and effective traffic safety map.

**Keywords**—Traffic Safety Map; Risk Estimation; Occupancy Grid Map; Driving Model; Smartphone Sensing; Collective Intelligence

## I. INTRODUCTION

This paper proposes a method for generating traffic safety maps based on differences in individual recognition of the road environment by using smartphone data from various users.

Currently, traffic safety maps are plotted annually using accident reports by police bureaus, and parts of these maps are made accessible to road users using a web-based geographic information system (GIS). Traffic safety maps show the locations of occurred accidents and other information, such as the type of collisions, the vehicles involved, and the time the collisions occurred [1],[2],[3].

Safety road map is absolute necessary for safety of road users, however it costs very high to create. That why current safety map is limited reluctantly. Moreover, it is lack of information because only accident records are used without any prediction of new risk locations. Recently, the development of new sensor technologies on smartphones-e.g., global positioning systems (GPS), accelerometers, and gyroscopes which can help to collect massive real-time traffic road data via smartphones, provides us many useful information of road conditions. Mohan et al. developed a road monitoring system named Nericell, which can automatically monitor braking, bumps in the road, honking and stop-and-go traffic via smartphone. It uses a smartphone's accelerometer, microphone, GSM communications and GPS for this purpose. Nericell aggregates sensed data from multiple participating smartphones on a centralized server [4].

Fathi et al. proposed a method for detecting road intersections from GPS traces [5]. Fazeen et al. developed an advanced driver-assistance system on a smartphone. Such a system advises a driver on dangerous situations that emerge from vehicle maneuvers and environmental factors. The aim of these systems is to recognize and classify driving behavior and to map road surface conditions [6]. Zhang et al. presented a method for integrating GPS traces and current road map towards a more accurate, up-to-data and detailed road map [7].

Safety map is a kind of risk analysis result of traffic environment. It is also a traffic road property and closely relates to the relationship among road users but has not been studied enough. There are some works related to risk analysis. For example, Google Live Traffic is a service that utilizes GPS data from Android smartphones to estimate traffic jams based on average GPS speed [8].

Honda initiated a project for traffic safety map in 2013 using data recorded by their Internavi in-vehicle unit, along with police reports and user contributions to their safety map website [9]. The locations where sudden brake occurs as risky spots are plotted automatically in the map.

However, there are gaps between the locations pointed out on these maps and actual hazard locations, for example, a sudden brake can occur for many reasons, and these often depend on the road type, road conditions, weather conditions, road equipment, and any driver distractions. A sudden braking is a consequence of a decision-making; however all of the locations are not actual hazard spots.

It is easy to understand of traffic safety if actual hazard spots are plotted. For example, school zone marks are usually drawn on hand-written safety maps. Currently, there is not any automatic method to detect hazard spots from the logs. In previous works, we have proposed a method for detecting incident locations at which risk situation occurred frequently [10],[11]. The method can be used to detect hazard locations for road user by using smartphone logs.

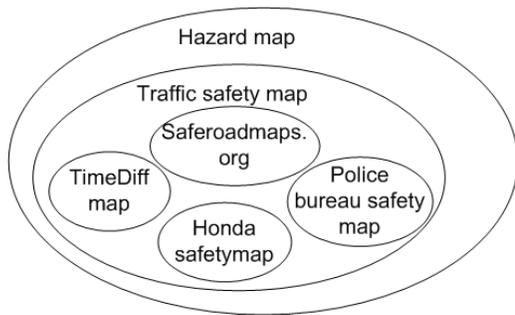


Fig. 1. Hazard map and traffic safety map

This paper proposes a method for integrating a collection of hazard locations based on differences in drivers' recognition of the road environment. The method is inspired by STEP technique [12] of risk analysis in traffic safety field. Suppose that if drivers have the same recognition of a crossroad, the level of danger will depend on one's driving skill. By contrast, where one driver recognizes a crossroad as passable and where other drivers recognize it as impassable, accidents are more likely to occur.

This paper is organized as follows. Section II reviews the related works and the requirements for traffic safety maps. The proposed method, related algorithms, and concept is described in Section III. The experiment and evaluation are discussed in Section IV. Section V provides the conclusion and future work.

## II. TRAFFIC SAFETY MAPS AND RELATED RESEARCH

In this section, the related research are reviewed. Objective of this research is to propose a method for calculating and providing safety information toward road users. Safety information of smartphone data from different users is extracted and provided to as much road users as possible. However, it is expected that smartphone does not directly senses "safety". Information collected by smartphone only contains the belonging vehicle information, and it cannot be used to measure the road condition and driver's behavior directly. Therefore, it is difficult to use current collision based risk analysis approaches to extract safety information from smartphone. This paper proposed a method based on difference on drivers' recognition of traffic road which inspired by STEP technique (the Sequentially Timed Event

Plotting)[12]. STEP technique is a qualitative risk analysis technique.

First, the requirements for traffic safety maps generated with smartphone data are explained.

### A. Requirements for Traffic Safety Maps

A traffic safety map is a type of hazard map that contains information regarding the road environment of which drivers should be aware(see Fig. 1). The requirements for traffic safety maps generated with smartphone data are as follows:

1) *Accuracy*: Accuracy depends on acknowledging frequent occurrences of unsafe events at particular locations. The locations and the frequency of accidents must be described accurately. This research considers the level of accuracy for alerting drivers of unsafe locations with location errors of 50 m.

2) *Automatic generation*: Generating traffic safety maps requires considerable time and effort. Thus, the map-generation process should proceed automatically. If possible, the process should be conducted on the smartphone, and the results should be forwarded to a GIS server over the Internet.

3) *Privacy considerations*: Private information should not be leaked from the smartphone.

4) *Community participation*: The more users that participate, the more coverage the traffic map has and the more accurate it becomes. Thus, motorcyclists, bicyclists, and pedestrians should also be able to join.

5) *Reflecting individual differences*: Differences in how users recognize the road environment should be reflected on the map.

6) *Reflecting the number of participants*: Busy roads with many commuters should be distinguished from quiet roads with few commuters, and this distinction should be reflected on the map.

7) *Safety map for everyone*: The safety map should be created for everyone uses not for vehicle's drivers only. The safety map can be generated by individual at first. The individual maps are then integrated to be global safety map for public use.

This paper discusses a global traffic safety map that meet above requirements.

### B. Related Research

This section reviews related research for extracting safety information. Traffic safety map is a kind of hazard map which is a result of risk analysis, depicted in Fig. 1. Fig. 2 shows methods for risk analysis [12]. Risk analysis methods are categorized into three types: qualitative technique, quantitative technique, and a hybrid technique. In traffic safety research field, related research on traffic safety extraction are as follows.

1) *Heat-map analysis and plot of past accidents distribution by Hilton et al.[1]*.

2) *Analysis of near miss reports[13]*.

3) *The Sequentially Timed Event Plotting (STEP) technique[12]*.

- 4) Hazard and Operability study (HAZOP) technique[12].
- 5) Estimation of collision probability between a vehicle and road users by Shimizu et al.[14].
- 6) Plot the occurrences of sudden brakes locations by Honda safetymap[9].
- 7) Traffic jams estimation by Google Live Traffic[8].
- 8) SNS based services where riders and pedestrians can contribute their experiences (Honda safetymap, and others)[9],[15].

Summation techniques use different safety elements such as hazard severity and probability is a common way to generate traffic safety map. Namely,  $\sum_{\text{hazard}_i} P_i W_i$ , where  $P_i$ ,  $W_i$  is the occurrence probability and,  $w_i$  is the severity of hazard  $i$ , respectively.

From above techniques, following traffic safety maps and related works have been proposed.

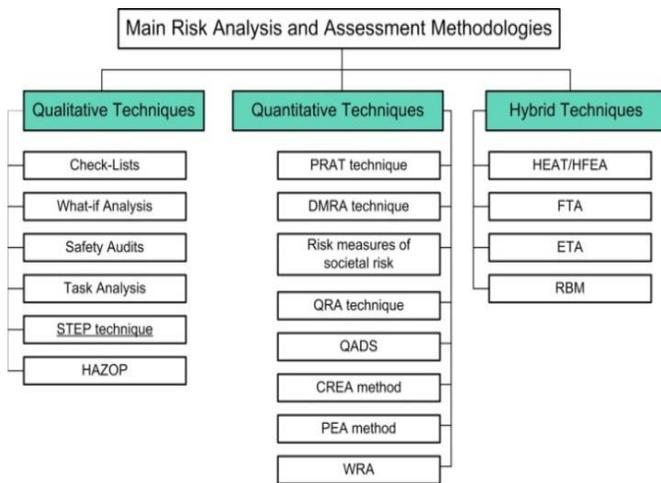


Fig. 2. Main risk analysis and assessment methodologies [12]

Firstly, there are methods to generate maps using accident reports directly. In Japan, each prefecture's police bureau publishes a safety map to improve traffic safety. These safety maps typically plot hot-spot locations with frequently occurring accidents, locations such as intersections with traffic signals. Other non-governmental organizations utilize Google My Map as a platform, relying on community submits hot-spot locations[15].

Next, a strategy to apply risk analysis methods into traffic safety is considered to generate maps of dangerous spots which not being listed in the police reports. Most of the current risk analysis methods usually assume that a hazard is a collision between cars, pedestrians, and others. In this case, a lot of precise information is needed to calculate collision risk probability of a given situation. For example, Shimizu et al.[14] presented a method for calculating collision probability between an ego-vehicle and road users. The model parameters include main road properties, pedestrian's road properties, pedestrian speed and next position probability. Therefore, it is difficult for this approach to predict risk probability of a given location because many collision patterns needed to be considered. It cannot apply this strategy for calculating probability of collision using smartphone data.

There are two approaches to avoid the above problem. The first approach utilizes topographical relations between accidents and set a dangerous value based on accidents distribution. For example, Hilton et al. presented a method to build safety maps using heat-map method, named as Saferoadmaps[1],[3]. This work utilizes data from the Fatality Analysis Reporting System (FARS). FARS's mission is to render vehicle-crash information accessible and useful for the sake of improving traffic safety. Fatality information derived from FARS includes motor-vehicle traffic collisions that resulted in the death of an occupant of a vehicle or of a non-motorist within 30 days of the collision. FARS contains data on all fatal traffic accidents in all 50 states, the District of Columbia, and Puerto Rico. Moreover, Saferoadmaps provides other useful tools, such as heat-map-based map analytics, crash analysis, a real-time safety tracker, and a commuter-stress index.

The second approach tries to estimate the factors caused accidents using vehicle's trace data. This approach considers the locations with high risk probability as unsafe locations, for example Google Live Traffic[8], Honda safetymap[9]. Honda initiated a project for traffic safety map in 2013 using data from their Internavi system, along with police reports and user contributions to their safetymap website. The core data for this safetymap system is sudden-brake data ( $>-0.25G$ , a deceleration of 25 km/h in 3 seconds) from vehicles made by Honda in Japan in 2012. The objective roads are more than 5.5m wide, and they are divided into 100m segments. Then, the system calculates the occurrence of sudden brake in each segment. Based on the occurrence rate, the sudden-brake level is classified into one of three groups: Level 1 is above 2.5%, Level 2 is above 5%, and Level 3 is above 10%. The locations where sudden brake occurs on Honda's safetymap are merely representative points; there is no guarantee that sudden brake occurred in that exact place[9].

These approaches need a technique to correct the errors when mapping the observations and the dangerous spots. There are two types of errors: 1) an observation location is not a dangerous spot, 2) an observation location is dangerous spot with a location error. In Honda safetymap, not all of sudden brake locations are near-miss locations. In addition, there are many accident locations provided by police do not include any sudden brakes provided by Honda Internavi system. This may be one reason that Honda adopts SNS service to allow community users to contribute their experience and upgrade safety road map accuracy.

The proposed method is classified into quantitative method's category and inspired by STEP technique[12] that uses to identify actions that contributed to the accidents: a) the time at which the event started; b) the duration of the event; c) the agent which caused the event; d) the description of the event; and e) the name of the source which offered the information. STEP technique provides a valuable overview of the timing and sequence of events/actions that contributed to the accident, or in other words, a reconstruction of the harm process by plotting the sequence of events that contributed to the accident. In this research, traffic road safety is quantified based on an inherited idea of STEP technique using difference in drivers' behaviors.

Table I shows the works related to integrating traffic maps. These methods use Web 2.0 technology with a server-based architecture. The input data for the mapping system can be classified into three types: smartphone-derived data, data derived from social networking services (SNSs), and data from public agencies. Owing to privacy concerns, data from public agencies is not open to the public. Smartphone- and

SNS-based data usually contains individual differences (i.e., drivers tend to differ in how they recognize, judge, and handle a given situation). To predict traffic jams, mapping systems do not need to consider these individual differences. Google’s Live Traffic service averages data in order to determine whether a traffic jam exists on a particular road[8].

TABLE I. RESEARCH RELATED TO TRAFFIC SAFETY MAPS

Related work	Smartphone	SNS	Public agency data (including road-side camera, probe car,...)	Human factor consideration
Google Live Traffic[8]	O			No
Honda safetymap[9]		O	O	No
Saferoadmaps.org[3]			O	No
Police bureau safety map[2]			O	No
Other SNS traffic safety map[15]		O		No

TABLE II. COMPARISON OF CROWD-SOURCED TRAFFIC SAFETY MAPS AND TRAFFIC SAFETY MAPS BASED ON POLICE REPORT.  
A: HIGH, B: MEDIUM, C: LOW, X: NO

Requirement	Google Live Traffic[8]	Honda Safetymap[9]	Saferoadmaps.org[3]	Police Bureau safety map[2]	Other SNS safety map[15]	Proposed Method
1) Accuracy	C	C	B-	A	C	B
2) Automatic generation	A	B	B	X	X	A
3) Privacy consideration	C	B	B	A	C	B
4) Community participation	A-	B	X	X	C	A
5) Reflecting individual differences	X	C	X	X	C	A
6) Reflecting the number of participants	A	A	X	X	C	A
7) Safety map for everyone	C	A	B	B	B	A

Table II compares the related works and the proposed method. The evaluation marks (A, B, C, X) are based on the requirements for safety road maps, the accuracy of the input data, and the privacy policy of the police bureau.

The police bureau safety map has an A rank for ‘Accuracy’ because it’s based on the report data. Meanwhile, Honda safetymap has a B- rank because their sudden-brake locations just appear at traffic signals only. Google Live Traffic got a C rank for displaying only traffic jams without safety information. Google Live Traffic and the proposal got an A rank for ‘Automatic generation’ because of full automation of generation. Again, the police bureau safety map has an A rank for ‘Privacy consideration’ because it’s based on the report data.

The proposal got an A rank for ‘Community participation’ for supporting both IOS and Android smartphone. Meanwhile, Google Live Traffic only supports Android smartphone. The proposal got an A rank for ‘Reflecting individual difference’ because this is only method to support of reflection of human factors in the model. Honda safetymap, and the proposal got an A rank for ‘Safety map for everyone’. Meanwhile, the police bureau safety maps got a B rank because it needs to make accessible to more road safety information.

In this section, the related research has been reviewed. It is confirmed that there are two types of the input data for traffic safety maps. One utilizes accident reports provided by police bureau, and another utilizes probe car- or smartphone- based data. For purpose of generating traffic safety maps, it is necessary to detect hazard locations by some way. In addition, there are two methods of detecting hazards. The first method assumes accident locations reported to be hazard. The second

method uses some sensible information, which sometimes leads accidents. The disadvantage of the first approach is that a near miss situation cannot be reflected on traffic safety maps because an accident has not happened. The disadvantage of the second is deviation from the true hazard locations and the sensed locations. The review’s result pointed out that current difficulty of making safety maps automatically is the difficulty of computation of vehicle’s collision probability. Based on the review’s result, this paper proposes a method for detecting hazard locations based on different behaviors among road users in the next section.

### III. PROPOSED METHOD

This section explains the proposed method.

#### A. Generating the traffic safety map

The process for generating the proposed traffic safety maps is depicted in Fig. 3. The proposed traffic safety map is referred as a TimeDiff map. The TimeDiff map is a kind of hazard map (Fig. 1). The map integrates a collection of traffic-incident maps that are generated from users’ smartphone log data (proposed in [10], [11]). This section reviews the core components of this map-generating process-viz., the IMAC model[16] and the Simple Braking Model (SBM)[11]. Then, the method for constructing a user’s traffic-incident map is explained (the left side of Fig. 3). Finally, the proposed summarization method for integrating these incident maps into a global hazard map is described (the right side of Fig. 3).

#### B. Simple Braking Model(SBM)

This section explains the method for generating a traffic incident map that can be used to interpret how a driver recognizes the road environment.

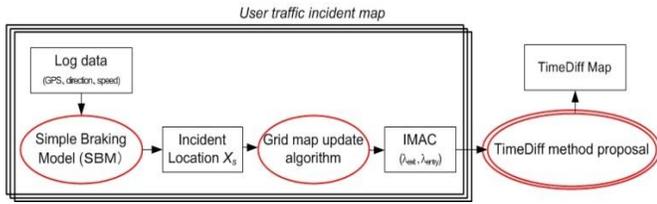


Fig. 3. Process for generating traffic safety map

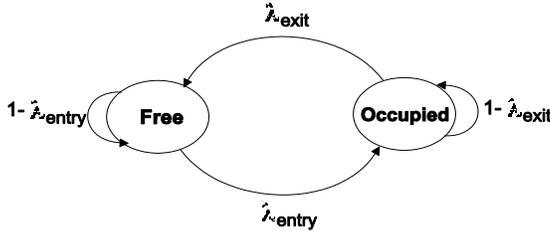


Fig. 4. Grid model – a two state Markov chain[16]

The Simple Braking Model (SBM) is firstly described. The SBM model will be used to measure vehicle deceleration behavior (i.e., upon approaching an incident). The SBM describes the relation between the vehicle’s moving distance, the vehicle’s speed, and the estimated incident location[11].

Assume that whenever a vehicle encounters an incident, it always decreases its speed to avoid accident. The relation between the current speed and the moving distance is described as follow braking equation:

$$\frac{dx}{dt} = v_0 \left(1 - \frac{x}{X_s}\right)^n \quad (1)$$

where  $X_s$  is the initial distance to the incident,  $x$  is current distance to the incident ( $x=0$  when the vehicle begins to slow down),  $n>0$ , and  $v_0$  is the initial speed.

### C. Generating a traffic-incident map with IMAC

This section explains method for estimating the locations of incidents, and discusses individual driver characteristics, and the process for generating a user’s traffic-incident map [10],[11]. Saarinen et al. proposed IMAC, a model for describing dynamic environments with an occupancy grid map[16], and this model is utilized to represent traffic incidents. With the IMAC model, the mapping environment is evenly divided into grids. Each grid is modelled as a two-state Markov chain with two states of being: *free* and *occupied* (see Fig. 4). The IMAC model is suitable for representing dynamic objects, such as traffic signals or traffic incidents, with a grid map. Furthermore, IMAC is used to estimate the transition-probability parameters ( $\lambda_{exit}, \lambda_{entry}$ ) by observing the occurrence of the state *occupied* or *free* and the transitions between them.

Eq. (2) describes the behavior of a grid at each step of the observation.

$$P = \begin{bmatrix} 1 - \lambda_{entry} & \lambda_{entry} \\ \lambda_{exit} & 1 - \lambda_{exit} \end{bmatrix} \quad (2)$$

where  $\lambda_{exit}$  is the probability that the grid state changes from *occupied* to *free*, and  $\lambda_{entry}$  is the probability that the grid state changes from *free* to *occupied*. Suppose that these transitions to follow a Poisson process. A Poisson process

describes the probability of observing a number of events within a certain amount of time. The stationary distribution of the grid state vector  $\pi$  is derived as follows:

$$\pi = \left( \pi_1 = \frac{\lambda_{exit}}{\lambda_{exit} + \lambda_{entry}}, \pi_2 = \frac{\lambda_{entry}}{\lambda_{exit} + \lambda_{entry}} \right) \quad (3)$$

The stationary distribution represents the probability of observing the grid in a particular state given an infinite number of steps. In Eq. (3),  $\pi_1$  is the stationary probability of observing the grid in a free state, and  $\pi_2$  is the stationary probability of observing the grid in an occupied state.

### D. Map estimations with IMAC

To estimate the grid states from the observed events, Saarinen et al.[16] proposed a method for observing two processes in each grid in a dynamic environment with Eqs. (4) and (5):

$$\hat{\lambda}_{exit} = \frac{\alpha_{exit}}{\beta_{exit}} = \frac{\#events:occupied\ to\ free+1}{\#observations\ when\ occupied+1}, \quad (4)$$

$$\hat{\lambda}_{entry} = \frac{\alpha_{entry}}{\beta_{entry}} = \frac{\#events:free\ to\ occupied+1}{\#observations\ when\ free+1}, \quad (5)$$

where  $\alpha_{exit}$  denotes the number of times a grid is observed switching from *occupied* to *free* (#OTF),  $\beta_{exit}$  is the number of observations made in the *occupied* state (#OCC), and  $\alpha_{entry}$  and  $\beta_{entry}$  are the quantities for observing a grid switch from *free* to *occupied* (#FTO) and for observing the grid in the *free* state (#FREE). The additional +1 in Eqs.(4) and (5) follows from the initialization of all the parameters at one. The interpretation of  $\hat{\lambda}_{exit}$  as a Poisson rate parameter is the expected number of state-change events per observation, given that the current state is *occupied*.

### E. Estimations for untraveled roads and locations

Unknown locations are off-road locations upon which vehicles do not travel. In the IMAC model,  $\lambda_{exit}$  and  $\lambda_{entry}$  are probabilistic values ( $0 < \lambda_{exit} < 1$ ,  $0 < \lambda_{entry} < 1$ ). This experiment stipulated  $\lambda_{exit} = \lambda_{entry} = 1$  for unknown locations.

### F. TimeDiff method

This section proposes an integration method for generating a global map from a collection of IMAC users’ grid maps for ( $\lambda_{exit}, \lambda_{entry}$ ). This method is based on the average time that users hold a different recognition of the same road environment, it’s called TimeDiff method. The global map (i.e., the TimeDiff map) shows the total time for road-recognition differences between users in a given area.

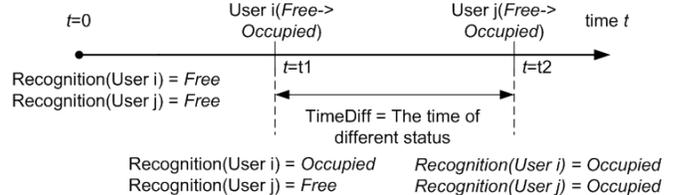


Fig. 5. Time differences in the judgment of a road environment

**TimeDiff concept:** Consider a scenario where two vehicles are driving along a road. At this time, the road upon which the vehicles are traveling is recognized as *Free* (see Fig. 5). As

the two vehicles approach a red traffic signal, they encounter an incident, and the road at this location is recognized as *Occupied*. Consequently, both vehicles come to a stop. At this time, both vehicles must change their recognition of the road environment from *Free* to *Occupied*. TimeDiff method focuses on the time that each vehicle changes its status. It is clear that the occurrence of an accident depends on the recognition of both drivers. If two vehicles change their status at the same time, they both have the same recognition of road environment. Thus, an accident will not occur unless a driver makes a mistake in controlling the vehicle. By contrast, the more difference in time between when the recognition changes - that is, between when one vehicle recognizes the road as *Free* and other recognizes it as *Occupied* - the higher the likelihood of a collision. Thus, the global TimeDiff map is based on the road-recognition level of the users.

Let  $D$  be the set of users. Assume that all users generate their own traffic incident map ( $\lambda_{exit}, \lambda_{entry}$ ) by using the update method in Eqs. (4) and (5).

The hazard level of each grid  $g$  in the global map is defined by Eq. (6). Suppose two users,  $i$  and  $j$  in  $D$ , initially report an *Occupied* status. Suppose further that user  $i$  changes this status to *Free* before user  $j$  does. Alternatively, suppose that  $i$  and  $j$  initially report a *Free* status, and that the status from user  $i$  switches to *Occupied* before that of user  $j$ . The total difference in time between such status changes (whether *Occupied* or *Free*) from all pairs of users  $i$  and  $j$  in  $D$  is calculated as follows:

$$TimeDiff_g = \sum_{i \in D} \sum_{j \neq i, j \in D} E(P_{dif}^f(i, j)) + E(P_{dif}^o(i, j)) \quad (6)$$

where  $E(P_{dif}^f(i, j))$  is the difference in time between switches when each pair of users initially reports a *Free* status, and  $(P_{dif}^o(i, j))$  is the difference in time between switches when each pair of users initially reports an *Occupied* status. Suppose that this follows a Poisson process.

$$E(P_{dif}^f(i, j)) = \int_0^{B_f} t \cdot Prob[X_i = occ \leq t] Prob[X_j = occ > t] dt \quad (7)$$

$$= \int_0^{B_f} t \cdot Prob[X_i = occ \leq t] (1 - Prob[X_i = occ \leq t]) dt \quad (8)$$

$$= \frac{e^{-B_f(\lambda_{i,entry} + \lambda_{j,entry})} (1 - e^{-B_f(\lambda_{i,entry} + \lambda_{j,entry}) + B_f(\lambda_{i,entry} + \lambda_{j,entry})})}{(\lambda_{i,entry} + \lambda_{j,entry})^2} \frac{1 - e^{-B_f \lambda_{j,entry}} (1 + B_f \lambda_{j,entry})}{\lambda_{j,entry}^2} \quad (9)$$

when  $t = 0$ , then  $X_i = X_j = free$ . Moreover,  $B_f$  is a constant. Here,  $\lambda_{i,entry}$  represents the  $\lambda_{entry}$  for user  $i$ ,  $\lambda_{j,entry}$  represents the  $\lambda_{entry}$  for user  $j$ .

$$E(P_{dif}^o(i, j)) = \int_0^{B_o} t \cdot Prob[X_i = free \leq t] Prob[X_j = free > t] dt \quad (10)$$

$$= \int_0^{B_o} t \cdot Prob[X_i = free \leq t] (1 - Prob[X_i = free \leq t]) dt \quad (11)$$

$$= \frac{e^{-B_o(\lambda_{i,exit} + \lambda_{j,exit})} (1 - e^{-B_o(\lambda_{i,exit} + \lambda_{j,exit}) + B_o(\lambda_{i,exit} + \lambda_{j,exit})})}{(\lambda_{i,exit} + \lambda_{j,exit})^2} + \frac{1 - e^{-B_o \lambda_{j,exit}} (1 + B_o \lambda_{j,exit})}{\lambda_{j,exit}^2} \quad (12)$$

when  $t = 0$ , then  $X_i = X_j = Occupied$ . Likewise,  $B_o$  is a constant. Here,  $\lambda_{i,exit}$  represents the  $\lambda_{exit}$  for user  $i$ ,  $\lambda_{j,exit}$  represents the  $\lambda_{exit}$  for user  $j$ .

For example, when  $\lambda_{i,exit} = 0,01$ ,  $\lambda_{j,exit} = 0.1$ ,  $B_o = 10$ , then  $E(P_{dif}^o(i, j)) = 1.5505$ .

### G. Summary

In summary, this section discussed the proposed map-integration method, TimeDiff, using smartphone data based on differences in the recognition of the road environment. In next section, two kinds of TimeDiff maps are discussed: the high-risk TimeDiff map and the low-risk TimeDiff map. The high-risk TimeDiff map presents locations at which the difference of the time recognition between two users is larger. On the contrary, the low-risk TimeDiff map presents locations at which the difference of the time recognition between two users is small.

## IV. EXPERIMENT

This section evaluates the proposed method with a realistic scenario. The global hazard map (i.e., the proposed TimeDiff map) was quantitatively evaluated in terms of its effectiveness using the F-measure method. The datasets for the experiment comprised a month of data from two drivers with one bike (Honda CBR) and one car (Mazuda 6) on the test route. The smartphone is put on pocket door of vehicle or attached on drivers' arm. The latitude, longitude, speed, and azimuth extracted from GPS data are used for experiment.

*Test course:* The test course covers the Maborikaigan area in Kanagawa Prefecture, Japan, as shown in Fig. 6. The data includes 30 days of smartphone log data. Each day covers 8 km of road along the test course (the red road in Fig. 6). The road between Mabori Elementary School and the Otsu traffic signal is narrow, and traffic jams is common. The No. 16 National Road has four lanes, and it is rarely faced by traffic congestion.

*Police accident data:* The distribution of 260 locations of accident in Maborikaigan area(2010-2014) on this course provided by the Uraga police bureau is shown in Fig. 7. This distribution shows that accidents usually occur either on narrow roads with high-volume traffic and in areas with traffic signals. Fig. 8 shows the sudden-brake distribution of Honda's safety map for the Maborikaigan area. The distribution of sudden-brake locations all occurred near traffic signals (known to be high-risk areas). However, from the accuracy point of view, these results are insufficient because the numbers of the spots in the Honda's map are 18 locations, too less than the number of the accidents provided by the Uraga police bureau.

### A. Traffic safety map with TimeDiff method

The TimeDiff maps ( $B_o = B_f = 10$ ) generated from Honda CBR's dataset and Mazuda 6's dataset are shown in

Figs. 9 and 10. The maps represent hazardous spots in terms of the time differences in the recognition of the road environment.

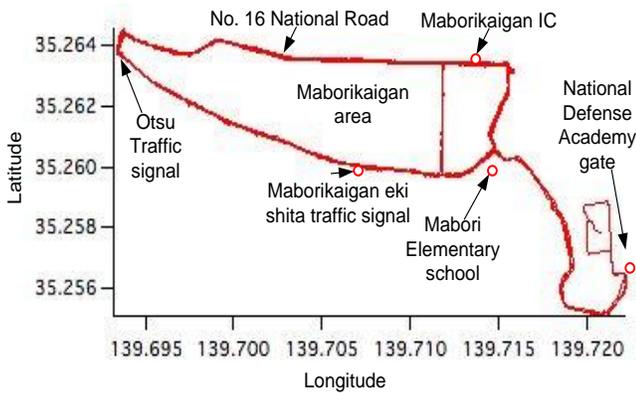


Fig. 6. Test course in Maborikaigan area

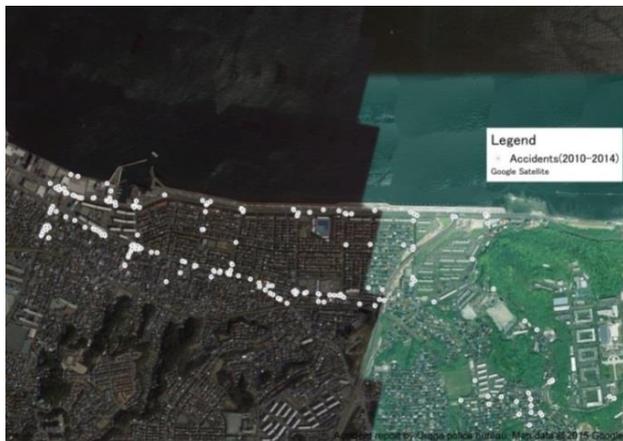


Fig. 7. Accident data in Maborikaigan area provided by the Uraga police bureau

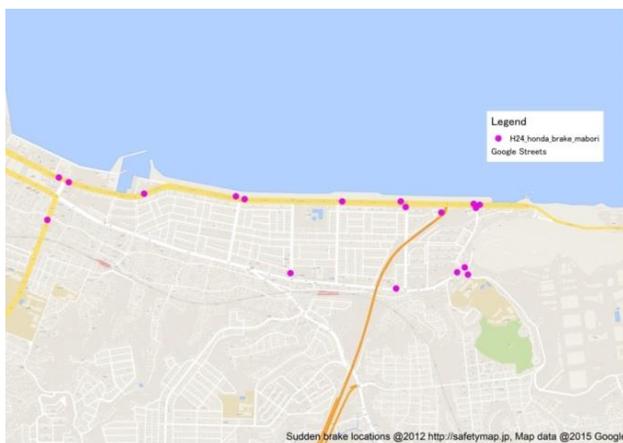


Fig. 8. Honda's sudden brake locations in Maborikaigan area[9]



Fig. 9. Traffic safety map by TimeDiff method( $B_o = B_f = 10$ ): high risk areas

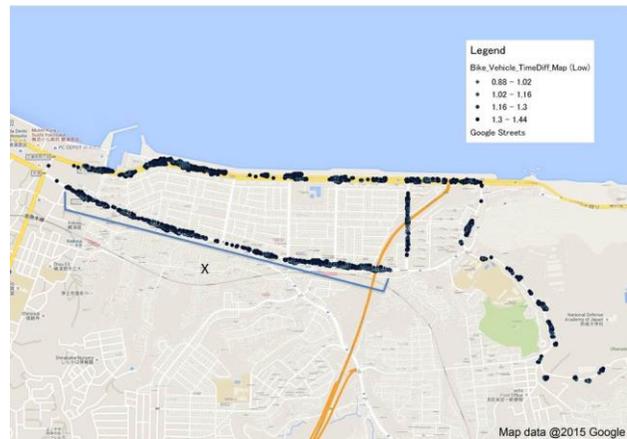


Fig. 10. Traffic safety map by TimeDiff method( $B_o = B_f = 10$ ): low risk areas

There are two TimeDiff maps: the high-risk TimeDiff map (Fig. 9) and the low-risk TimeDiff map (Fig. 10).

The high-risk TimeDiff map is shown in Fig. 9. The white-red graduated circles with the plus sign denotes the level of time difference in recognition by two drivers. The TimeDiff value ranged from 2 to 169 based on the quantile, and this was divided into 5 levels. The high-risk map for the route between the Otsu traffic signal (interval B) to the Maborikaigan eki shita traffic signal through Mabori Elementary School (interval C, D) contains more hot-spots than does the route along the No. 16 National Road between Maborikaigan IC and the Otsu traffic signal (interval A). This result agrees with the accident data provided by the Uraga police bureau.

Moreover, most locations with a high distribution of accidents provided by the police bureau can also be verified with this map. Some representative accident locations which detected by the proposed method are depicted by the arrows with capital letters in Fig. 9. For examples, spots have many vehicles come in and out (the exit entry of packing area - spot H and the entrance to residence area - spot I); spots have many pedestrians (the Maborikaigan eki shita traffic signal - spot J, spot F, spot G).

Figs. 11 and 12 show scene photographs of spot I and K of Fig. 9. The features of spot I are narrow road with high volume traffic, lot of vehicles come in and out, four lanes road decreases to two lanes. The features of spot K are the road is poor visibility with a lot of jump out of the road. The proposed method succeeds in detecting such hazard locations.



Fig. 11. Accident location at the entrance to Otsu Sea Height Mansion (Spot I in Fig. 9)



Fig. 12. Accident location at the entrance to Mabori Elementary School (Spot K in Fig. 9)

Fig. 10 illustrates the low-risk TimeDiff map and the white-blue graduated circles with plus sign denotes the level of time difference in recognition by two drivers. The TimeDiff value ranges from 0.88 to 1.44 based on quantile, and this is divided into 5 levels. The low-risk map shows locations where there is little difference in recognition - that is, where two drivers' recognition of the road was almost same.

The proposed map suggests the traditional assumption that around a hazard spot is dangerous, is not always true and around a safe spot is not always safe. According to the low-risk map, the low-risk locations are distributed all over the test route. This means two drivers' recognition of this road is almost same. For example, two drivers have almost equal recognition of the road from Mabori Elementary School to the Otsu traffic signal (Fig. 10, interval X), meanwhile many hazard locations can be seen for the same area (see Fig. 9, interval B, C, D). That means unsafe locations exist among area looks safe.

From above result, most of hazard locations can be detected using the traffic safety maps based on the recognition difference of road environment. In the subsequent section, the proposed method is evaluated quantitatively by the F-measure method for further discussion.

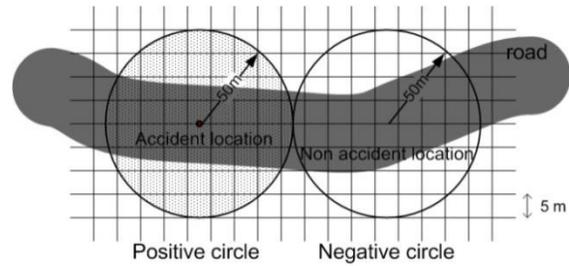


Fig. 13. Evaluation circles

TABLE III. DATASETS FOR EVALUATION

Data	Police data	Honda sudden-brakes data	Smartphone data
Data period	2010-2014	2012	2013,2015
Data region	Maborikaigan area	Maborikaigan area	Maborikaigan area
Data type	Accident report	Sudden-brakes data	Traffic safety map
Data for evaluation (in detail)	Accident distribution	Sudden-brakes distribution	TimeDiff value,

### B. Evaluation result by F-measure method

The F-measure method was adopted to further evaluate the traffic safety maps that were generated. In statistical analysis of binary classification by an information retrieval system, the F-measure is a criteria of a test accuracy. This measure is harmonic mean of precision score and retrieval score. The precision score evaluates the number of the correct items in a search result. The retrieval score is the fraction of found correct items in the total correct items.

Suppose the accident distribution provided by the police bureau is the correct answer of the hazard locations for the F-measure method. The Honda's sudden brakes distribution is used for the baseline of evaluation. The detailed dataset for evaluation is shown in Table III.

“Evaluation circles” is used for calculating of F-measure (Fig. 13). Every road is divided into multiple circles with a radius of 50 m, including “positive circles” or “negative circles”. Positive circles are created with the location of an accident at its center, and negative circles are continually created in other locations where accidents do not occur.

The F-measure method is defined as in Eqs. (15), (16), (17).  $AD_R$  is a set of the positive circles (the answer set).  $HL_R$  is a set of circles which detected as “hazards” by a given map method (the predict set).

$$F - measure = 2 \cdot \frac{Recall \cdot Precision}{Recall + Precision} \quad (15)$$

$$Precision = \frac{|AD_R \cap HL_R|}{|HL_R|} \quad (16)$$

$$Recall = \frac{|AD_R \cap HL_R|}{|AD_R|} \quad (17)$$

The result is closer to 1, the better performance they are.

Table IV shows the results from evaluating the proposed method. The proposed method's F-measure is the higher (=0.6542923), meaning that the TimeDiff map agrees with the police bureau's accident distribution by approximately 65.42%. Meanwhile, the Honda's sudden brakes distribution agrees with the police bureau's accident distribution of about 44.5%. From the value of precision and recall, the traffic safety maps by the proposed method better covers more hazard locations in the test route.

In summary, these results show that the traffic safety maps generated by the proposal provide better result than Honda's safetymap when compared to the police bureau's accident distribution. Moreover, the difference of recognition of road users can be used to estimate hazard locations.

TABLE IV. F-MEASURE EVALUATION RESULT

Method	Precision	Recall	F-measure
Honda sudden-brake	0.9746835	0.2883895	0.4450867
TimeDiff	0.8597561	0.5280899	0.6542923

## V. CONCLUSION

This paper proposed a method that uses smartphone data to automatically generate traffic safety maps based on the differences in how individuals recognize the road environment. The generating traffic safety map is referred as a TimeDiff map. The TimeDiff maps achieved better results when compared to Honda's safetymap, which is based on the occurrence of sudden brakes. From the experiment result, the difference of recognition of road users can be used to estimate hazard locations.

The power consumption is a limitation of the proposed system. It must be minimized so that will not affect a user's smartphone usage. The power consumption could be decreased by developing computationally efficient algorithms and minimizing the GPS usage.

## ACKNOWLEDGEMENT

We would like to thank the Uraga police bureau for providing the accident data of Maborikaigan area.

## REFERENCES

- [1] B. Hilton, T. Horan, R. Burkhard, and B. Schooley, SafeRoadMaps: Communication of location and density of traffic fatalities through spatial visualization and heat map analysis information visualization. *Information Visualization* 10.1(2011): 82-96.
- [2] Traffic Safety Map & Near-miss Map. Kanagawa Pref. Police Department. 06 August 2015 Accessed.
- [3] Saferoadmaps project. University of Minnesota. 2012 Published. 06 August 2015 Accessed.
- [4] P. Mohan, N. Venkata, and R. Ramjee, Nericell rich monitoring of road and traffic conditions using mobile smartphones. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*. pp. 323-336, 2008.
- [5] A. Fathi and J. Krumm, Detecting road intersections from GPS traces, *Geographic Information Science, Lecture Notes in Computer Science*. Volume 6292, 2010, pp. 56-69.
- [6] M. Fazeen, B. Gozick, R. Dantu, M. Bhukhiya, and M. C. Gonzalez, Safe driving using mobile phones, *IEEE Transactions on Intelligent Transportation Systems*, 13.3: 1462-1468, 2012.
- [7] L. Zhang, F. Thiemann, and M. Sester, Integration of GPS traces with road map, *Proceedings of IWCTS '10, Proceedings of the Second International Workshop on Computational Transportation Science*, pp. 17-22, 2010.
- [8] Google Live Traffic. Wikipedia. 01 June 2015 Last Modified. 06 August 2015 Accessed.
- [9] Safety Map Project. Honda Company. 2013 Published. 06 August 2015 Accessed.
- [10] V. C. Dang, M. Kubo, H. Sato, T. Shirakawa, and A. Namatame, Occupancy grid map of semi-static objects by mobile observer. *Artificial Life and Robotics*. 20.1: 7-12, 2014.
- [11] V. C. Dang, M. Kubo, H. Sato, A. Yamaguchi, and A. Namatame, A simple braking model for detecting incidents locations by smartphones. In *Computational Intelligence for Security and Defense Applications (CISDA), 2014 Seventh IEEE Symposium on*, pp. 1-5, 2014.
- [12] P. K. Marhavilas, D. Koulouriotis, and V. Gemeni. Risk analysis and assessment methodologies in the work sites: On a review, classification and comparative study of the scientific literature of the period 2000-2009. *Journal of Loss Prevention in the Process Industries* 24.5 (2011): 477-523.
- [13] How to prevent accidents. Safety Transportation. Japan Ministry of Land, Infrastructure, and Transportation. 06 August 2015 Accessed.
- [14] T. Shimizu, Y. Ohama, S. Nagata, J. Sakugawa, A computational framework for estimating collision risk against pedestrians. *R&D Review of Toyota CRDL, Vol.43. No.1 (2012) 22-42*.
- [15] Traffic Accident Hotspots Map. General Insurance Association of Japan. 6 August 2015 Accessed.
- [16] J. Saarinen, A. Henrik, and J. L. Achim, Independent markov chain occupancy grid maps for representation of dynamic environment. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3489-3495, 2012.