

# Estimation of Driver's Insight for Safe Passing based on Pedestrian Attributes

Fumito Shinmura<sup>1</sup>, Yasutomo Kawanishi<sup>2</sup>, Daisuke Deguchi<sup>3</sup>, Takatsugu Hirayama<sup>1</sup>,  
Ichiro Ide<sup>2</sup>, Hiroshi Murase<sup>2</sup> and Hironobu Fujiyoshi<sup>4</sup>

**Abstract**—In order to reduce traffic accidents between a vehicle and a pedestrian, recognition of a pedestrian who has a possibility of collision with a vehicle should be helpful. However, since a pedestrian may suddenly change his/her direction and cross the road, it is difficult to predict his/her behavior directly. Here, we focus on the fact that experienced drivers usually pass by a pedestrian while preparing to step on the brake at any moment when they feel danger. If driver assistant systems can estimate such experienced driver's decisions, they could early detect the pedestrian in danger of collision. Therefore, we classify the driver's decisions into three types by referring to the accelerator operation of drivers, and propose a method to estimate the type of the driver's decision. The drivers are considered to decide their actions focusing on various behaviors and states of a pedestrian, namely pedestrian's attributes. Since the driver's decisions change along the timeline, the use of a temporal context is considered to be effective. Thus, in this paper, we propose an estimation method using a recurrent neural network architecture with the pedestrian's attributes as input. We constructed a dataset collected by experienced drivers in control of the vehicle and evaluated the performance, and then confirmed the effectiveness of the use of pedestrian's attributes.

## I. INTRODUCTION

Pedestrian's safety on the road is one of the most important issues since traffic accidents injure many pedestrians every year. To prevent traffic accidents with pedestrians, it is important to find pedestrians in danger of collision with a vehicle. Research has been widely conducted on pedestrian detection [1] by in-vehicle camera [2] and 3D LiDAR (Light Detection and Ranging) [3]. On the other hand, there are still few studies to predict the behavior of a pedestrian.

For safety in a driving situation, drivers decide their actions focusing on various behaviors and states of a pedestrian. These behaviors and states are important for the prediction of the behavior of a pedestrian. Thus, in this paper, we consider the behaviors and states of a pedestrian as his/her attribute. Actually, many researchers have worked on pedestrian attribute recognition [4], [5]. Some of them focused on the recognition of a pedestrian's body orientation to predict his/her moving direction [6], [7]. Meanwhile, some of them focused on the recognition of risky behaviors by a

pedestrian [10] to estimate the risk of accidents. Some other researchers focused on the early detection of a pedestrian crossing the road [11]. They attempted to predict the road crossing by detecting the beginning of crossing as early as possible.

Including these works, pedestrians crossing the road are usually recognized as dangerous and the others are recognized as not dangerous. However, a pedestrian may suddenly change his/her direction and cross the road. Thus, pedestrians walking on the road side are also not always safe, and it is difficult for the existing works to accurately predict the risk of a pedestrian stepping into the road.

One method to improve driving technique is to imitate the driving of an experienced driver. Likewise, if a system can estimate the driving of an experienced driver, it would be helpful for the assistance of safety driving. When an experienced driver encounters a pedestrian, he/she can take appropriate actions by predicting the pedestrian's behaviors. If passing by a pedestrian cannot be judged as safe, the driver will approach the pedestrian while preparing to step on the brake at any moment. Meanwhile, once passing by the pedestrian is judged as safe, he/she will pass by increasing the speed.

Since an experienced driver decides appropriate driving operations by predicting pedestrian's behaviors, the estimation of the driver's decisions would be helpful for the prediction of pedestrian's behaviors. More specifically, if a system can estimate the driver's decision of stepping on the accelerator pedal or not, it could predict the risk of collision with the pedestrian. Therefore, it is important to estimate such experienced driver's decisions. On the other hand, the estimation of the vehicle speed would not be helpful for the prediction since the vehicle speed does not always represent the driver's decisions. For example, in the case of keeping the vehicle speed, there are two kinds of driver actions; a case that the driver keeps the speed by softly stepping on the accelerator pedal, and a case that he/she keeps the speed by releasing the accelerator pedal while preparing to step on the brake pedal. There is a difference between the two cases regarding the driver's decision.

In this paper, we propose a method to estimate a driver's decision in the case of passing by a pedestrian based on the pedestrian's attributes. We approach this problem by referring to the information on the accelerator operation of experienced drivers, namely,

- Stepping on the accelerator pedal.
- Releasing the accelerator pedal (and preparing to step

<sup>1</sup>Fumito Shinmura and Takatsugu Hirayama are with the Institutes of Innovation for Future Society, Nagoya University, Nagoya, Aichi, Japan [shinmuraf@murase.is.i.nagoya-u.ac.jp](mailto:shinmuraf@murase.is.i.nagoya-u.ac.jp)

<sup>2</sup>Yasutomo Kawanishi, Ichiro Ide and Hiroshi Murase are with the Graduate School of Informatics, Nagoya University, Nagoya, Aichi, Japan

<sup>3</sup>Daisuke Deguchi is with the Information Strategy Office, Nagoya University, Nagoya, Aichi, Japan

<sup>4</sup>Hironobu Fujiyoshi is with the Department of Robotics Science and Technology, Chubu University, Kasugai, Aichi, Japan

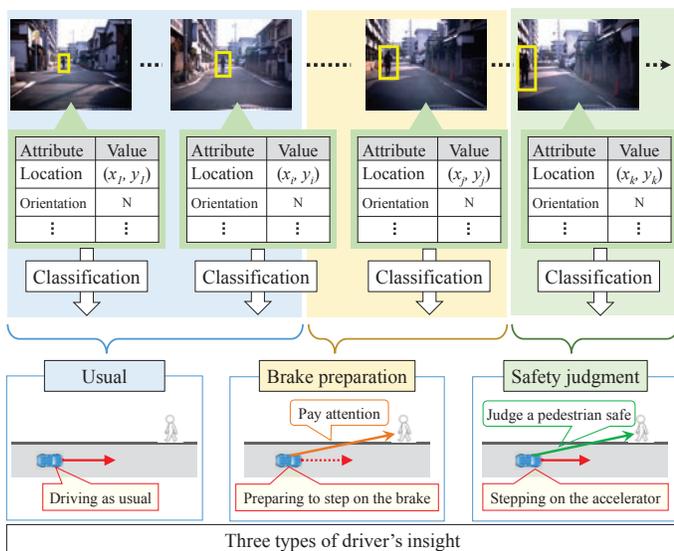


Fig. 1. Outline of estimation of driver's insight.

on the brake pedal).

- Restepping on the accelerator pedal after releasing.

By estimating these three types, we can estimate whether the driver has judged the pedestrian to be dangerous or not. In this paper, these three types of driver's decisions are called as "driver's insight". The outline of this approach is illustrated in Fig. 1.

Our contributions are summarized as follows:

- By extending our previous work [17], we estimate the driver's insight in the case of passing by a pedestrian.
- We propose a method to estimate the driver's insight based on pedestrian's attributes.
- We show the effectiveness of considering the pedestrian's attributes in the estimation of the driver's insight.

## II. RELATED WORK

Many researchers have studied on various recognition methods for pedestrian's attributes from a camera image. The body orientation is one of the most important factors to predict their advancing direction. Accordingly, many researchers have reported on orientation recognition of a pedestrian [7], [8]. A method for estimating orientation by aggregating estimated orientations from various body parts of a pedestrian has been proposed [6]. In order to improve the recognition accuracy, a method utilizing the similarity between the head and the body orientations has been proposed [9]. In addition, the risky behavior on the road is considered as an important factor, such as the use of mobile devices while walking on the road. Since the use of such devices while walking, namely, Texting-while-Walking, is an issue gaining global attention in recent years, a method for recognizing whether a pedestrian is using a smartphone or not has been proposed [10].

There are also researches on the prediction of pedestrian's behaviors. Kohler *et al.* have proposed an early detection

method of road crossing [11]. They combined an edge-based feature and a machine learning method to detect the movements at the beginning of a road crossing. They have reported that their method could detect the movements within 340 ms with an accuracy of 99%. Keller and Gavrila have presented a comparison of four approaches for pedestrian path prediction and action classification that differ in the dynamics modeling and the visual features [12]. They have reported that all approaches yielded similar performances in terms of pedestrian path prediction, but the approaches with non-linear models and motion features predicted the position more accurately in a scene where a pedestrian stopped at a curbstone. Chan *et al.* have proposed a method to anticipate traffic accidents from Dashcam videos [13]. Their method predicts accidents by Recurrent Neural Network with a spatial attention [14] to object regions extracted for each frame. They have reported that their method can predict an accident 1.86 seconds before it occurs with 80% recall and 56% precision.

There are also researches on the prediction of driving actions. Shimosaka *et al.* have proposed a method to predict the driving actions on a straight road including intersections without a signal [15]. Their method generates a potential map based on the position of the own vehicle, the position of intersections, and the legal speed limit, and then estimates the next driving actions from increasing speed, decreasing speed, and keeping speed. Although their method estimates the driver's decision, it does not consider pedestrians.

## III. DEFINITION OF DRIVER'S INSIGHT

By carefully observing experienced drivers' driving, we have noticed that their decisions can be classified into three types when passing by a pedestrian; When a pedestrian is in a distance, the drivers drive as usual. When the vehicle approaches a pedestrian, the drivers begin to pay attention to the pedestrian from a certain distance. If the drivers perceive danger of collision, they will keep their attention to the pedestrian, and prepare to step on the brake at any moment to stop the vehicle immediately. If the drivers judge that no collision will occur, they will pass by the pedestrian stepping on the accelerator.

Therefore, we define the following three types of driver's decisions:

- 1) *Usual.* A driver drives as usual without focusing on the pedestrian because he/she is in the distance.
- 2) *Brake preparation.* A driver releases the accelerator and prepares to step on the brake at any moment since he/she feels a danger.
- 3) *Safety judgment.* A driver steps on the accelerator under judgment that he/she can safely pass by the pedestrian.

As mentioned previously, we call these three types of driver's decisions as "driver's insight".

### A. Automatic Annotation using Driving Data

We determine the driver's insight frame-by-frame by choosing one of the three types mentioned above based on

the information on the accelerator operation. However, the information on the accelerator operation is unknown when estimating the driver's insight.

The period while releasing the accelerator pedal is annotated as *Brake preparation*. Before *Brake preparation*, if a driver steps on the accelerator pedal, he/she would ignore a pedestrian. Thus, this period is annotated as *Usual*. After *Brake preparation*, the period while stepping on the accelerator pedal from the beginning to the end of passing by a pedestrian, it is annotated as *Safety judgment*. In the case that the driver keeps stepping on the accelerator pedal from the beginning to the end of passing by a pedestrian, it is annotated as *Safety judgment* from the beginning, since safety judgment may occur at a moment when finding a pedestrian.

These types of insight can be discriminated by the accelerator's opening rate. When the accelerator's opening rate is zero, the driver is releasing the accelerator pedal. Thus, *Usual* is the period while the accelerator's opening rate is greater than zero until it becomes zero. *Brake preparation* is the period while the accelerator opening rate is zero, and *Safety judgment* is the period after it becomes greater than zero after *Brake preparation*.

### B. Examples of Driver's Insight

We show an example of transitions of the types of the driver's insight in Fig. 2. The upper images are in-vehicle camera images, and the center graph shows the transitions of the vehicle speed and the amount of acceleration. The bottom bar indicates the driver's insight corresponding to the center graph. This example is a scene where the vehicle passes by a pedestrian on the left side of the road. Here,  $T_1$  represents the timing of releasing the accelerator pedal, which is the turning point from *Usual* to *Brake preparation*.  $T_2$  represents the timing of re-acceleration, which is the turning point from *Brake preparation* to *Safety judgment*.

We show another example in Fig. 3. This example is a scene where the vehicle passes by a pedestrian on the right side. In this example, the driver continues to step on the accelerator pedal from the beginning to the end of the scene. In this scene, the driver may have judged that a collision will not occur at a moment when finding a pedestrian since he/she keeps an enough distance from the course of the vehicle. Hence the driver's insight was *Safety judgment* throughout this scene.

## IV. ESTIMATION OF THE TYPE OF DRIVER'S INSIGHT

We assume that the type of a driver's insight when passing by a pedestrian is determined according to the combination of the own vehicle's state and the pedestrian's attributes. As mentioned above, we regard pedestrian's behaviors and states as the attributes. Although it is difficult to estimate the driver's insight, a simple approach is to determine from the vehicle speed and the position of the pedestrian. Besides this, the proposed method utilizes the body orientation of the pedestrian as an attribute to estimate the driver's insight. According to our previous work, we have confirmed that the

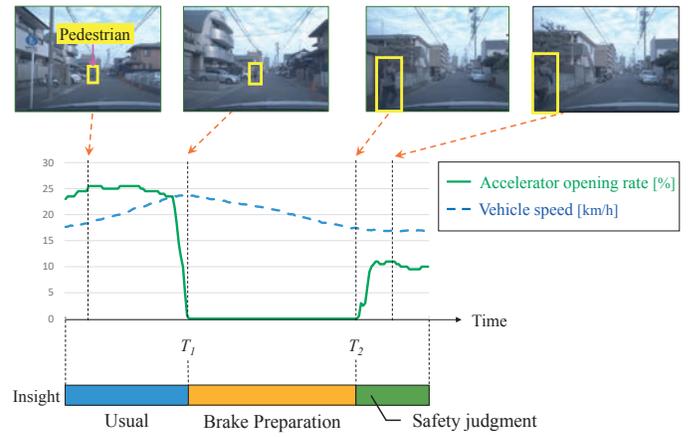


Fig. 2. Example of driving data and driver's insight (1).

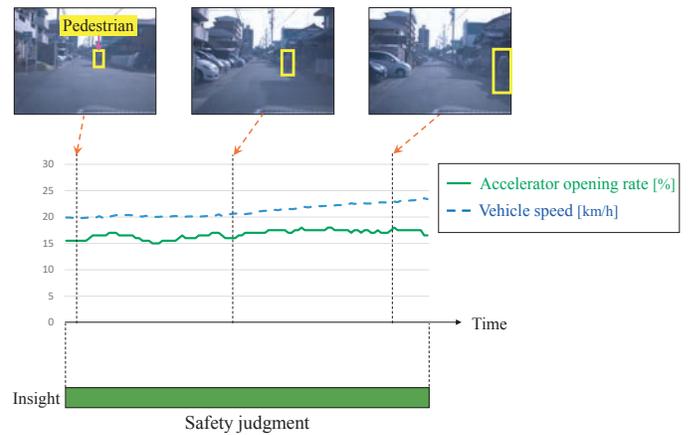


Fig. 3. Example of driving data and driver's insight (2).

body orientation of a pedestrian is particularly important for the estimation of a driver's insight [17].

Therefore, we propose a method to estimate the type of a driver's insight based on the pedestrian's attributes. Our method classifies the driver's insight into three types mentioned in Section III from the attributes. We use the pedestrian's position and body orientation as the attributes. Here, we assume that the pedestrian's position and the body orientation are obtained by existing recognition methods such as those proposed in references [2], [3], and [9] introduced in section II. In this paper, we assume that these recognition methods are highly accurate.

For the estimation, we take a classification approach. Since the driver's insight changes along the time line, we employ a neural network architecture based on Long Short-Term Memory (LSTM) in order to consider temporal features. The network model is shown in Fig. 4. This network architecture consists of two fully-connected layers and an LSTM layer. Leaky Rectified Linear Units (Leaky ReLU) [16] is used as the activation function for the fully-connected layers, and a softmax function is used for the output layer. The network receives the vehicle speed and the pedestrian's attributes, and returns the probability for each type of insight. The details

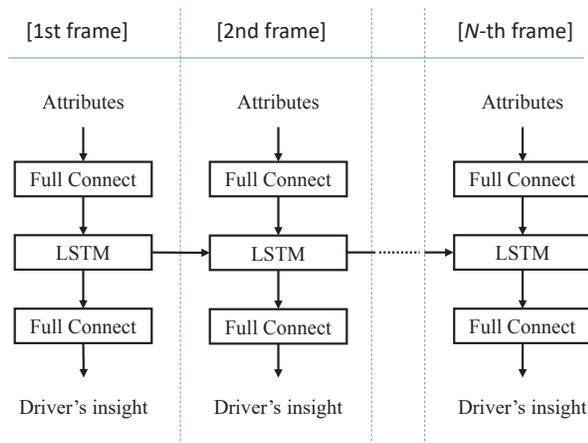


Fig. 4. Network model of the proposed estimation method.

TABLE I  
NETWORK ARCHITECTURE.

Layer	Units	Activation Function
Input	# of attributes	
Fully-Connected 1	# of attributes	Leaky ReLU
LSTM 2	10	
Fully-Connected 3	10	Leaky ReLU
Output	# of classes (= 3)	Softmax

of the network architecture is shown in Table I. Note that the number of units are determined experimentally.

## V. EXPERIMENT AND RESULTS

We conducted an experiment on the estimation of the type of the driver's insight in order to evaluate the effectiveness of the proposed method.

### A. Dataset

We collected driving data of experienced drivers in real traffic scenes when passing by pedestrians. A special vehicle as shown in Fig. 5 was used for the collection. The vehicle was equipped with various sensors (Grasshopper®3 camera, Velodyne LiDAR® HDL-64E, and CAN (Controller Area Network) signals). This vehicle was driven by experienced drivers along a pre-planned path repeatedly for collecting data in various situations. Here, the reason why the experienced drivers were selected is that an inexperienced driver may sometimes misjudge the safety and the estimation of such driver's insight is not helpful for safety driving assistance. The experienced drivers were three instructors of a driving school. They were informed that their driving data will be collected by various sensors, but were not informed about the purpose of the experiment. In this manner we collected in-vehicle camera images, 3D point cloud data obtained from the LiDAR, the vehicle speeds and the amount of acceleration obtained from the CAN signals. The data were collected on sunny days and rainy days, and do not include traffic congestions.

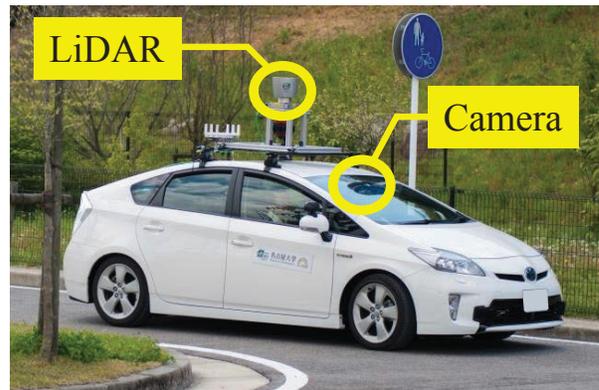


Fig. 5. Vehicle for data collection.

We manually retrieved scenes that the vehicle passes by a pedestrian after collecting the driving data, and then performed two kinds of frame-by-frame annotations; the driver's insight and the pedestrian's attributes. The driver's insight was determined from the amount of acceleration as described in Section III-A. In this experiment, we manually annotated the following attributes.

- Position (Relative position of a pedestrian from the vehicle)
- Body orientation (Eight directions)

The relative position was calculated referring to 3D point cloud data. Meanwhile, the pedestrian's body orientation was visually determined from the in-vehicle camera image.

The constructed dataset included 84 scenes of passing by a pedestrian and consisted of a total of 4,837 frames annotated with the driver's insight and the pedestrian's attributes. Here, the dataset does not include scenes that the vehicle passes by more than two pedestrians. In addition, the dataset includes two types of road; a one-way street with a sidewalk and a street having one lane in each direction without a center line.

### B. Evaluation

We conducted an experiment on the estimation of the driver's insight by using the constructed dataset.

In order to evaluate the effectiveness of adding pedestrian's body orientation as its attribute, we evaluated the following two methods:

- Input is the own vehicle speed and the pedestrian's position.
- Input is the own vehicle speed, the pedestrian's position, and the pedestrian's body orientation

The evaluation was performed in a five-fold cross validation manner and the accuracy rate for the classification was averaged. The accuracy rate for the classification was calculated by comparing between the estimated insight and the annotated insight frame-by-frame. Note that since the number of frames differs for each scene, first the per-frame accuracy rate was averaged for each scene, and then the average of all scenes was calculated as an evaluation criteria.

TABLE II  
ESTIMATION ACCURACY.

Input	Accuracy
Basic method: Vehicle speed and pedestrian's position	56.8 %
Proposed method: Vehicle speed, pedestrian's position and body orientation	65.5 %

### C. Results

The estimation accuracy is shown in Table II. The method using the vehicle speed, the pedestrian's position, and the body orientation showed better performance. We consider that our approach based on the pedestrian's attributes is effective for the estimation of the driver's insight. In order to improve the estimation accuracy, the use of pedestrian's attributes related to the driver's insight is very important.

An example of experimental results is shown in Fig. 6. Color bars indicate the types of the driver's insight; the blue, orange, and green bars indicate *Usual*, *Brake preparation*, and *Safety judgment*, respectively. The upper bar (Correct) indicates the correct type of the driver's insight. The middle bar (Baseline) indicates the estimation results by using the vehicle speed and the pedestrian's position, and the bottom bar (Proposed) indicates the result by using the vehicle speed, the pedestrian's position and the body orientation.

Here, the proposed method using three attributes showed a better performance. Focusing on the part of brake preparation, the method estimated the transition of the type of the driver's insight approximately five frames (= 0.25 seconds) earlier than the correct timing. On the other hand, the baseline estimation method without the pedestrian's body orientation showed a worse and unstable performance.

Another example of the experimental results is shown in Fig. 7. Here, both methods could not estimate well. The proposed method estimated the driver's insight to be *Safety judgment* in the early period since the pedestrian was facing the vehicle side. However, the type of the driver's insight actually changed to *Brake preparation* slightly before passing by the pedestrian. Therefore, the driver's insight may have been affected by a factor other than the body orientation of the pedestrian; The driver may have decided to release the accelerator pedal since the vehicle speed became too fast. We need to consider other attributes to improve the estimation accuracy in our future work.

In other cases of the misclassification, the driver's insight was affected by the pedestrian's age. There was little difference between the case of an adult and an elderly pedestrian, but, there was a difference between the case of a child and others. It seemed that the driver decided to decrease the speed more largely since he/she paid more attention to the child. Meanwhile, the driver's insight was also affected by a factor other than pedestrian's attributes. The road structure affected pedestrian's behaviors and the driver's insight. For example, a pedestrian might cross the road near an intersection. The driver releases the accelerator

pedal regardless of the pedestrian's body orientation when passing by such a pedestrian since he/she pays more attention to the crossing.

## VI. CONCLUSION

In this paper, we defined three types of driver's insight, and proposed a method to estimate them using a neural network based on LSTM learned from actual driving data of experienced drivers. Since the driver's insights are strongly affected by pedestrian's attributes, we proposed an estimation method based on them.

The performance of the proposed method is currently not sufficiently good, so we need to improve it. We will need to consider other attributes than those considered in this paper. In addition, the amount of driving data is not sufficient for the training of our network model. There are only a few scenes in the dataset for a certain body orientation, and then the estimation cannot be performed well. We plan to collect a large amount of driving data, and analyze key factors for estimating the driver's insight. Future work also includes estimation in the case of passing by multiple pedestrians. In such a case, we need to first estimate a pedestrian whom a driver is paying attention to.

## ACKNOWLEDGMENT

This research was supported by the Center of Innovation Program (Nagoya-COI) from Japan Science and Technology Agency, JST, and Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology.

## REFERENCES

- [1] P. Dollar, C. Wojek, B. Schiele and P. Perona, Pedestrian detection: An evaluation of the state of the art, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, 2011, pp. 743–761.
- [2] J. Li, X. Liang, S. Shen, T. Xu, J. Feng, and S. Yan, Scale-aware Fast R-CNN for pedestrian detection, *IEEE Trans. Multimed.*, vol. 20, no. 4, 2018, pp. 958–996.
- [3] Y. Tatebe, D. Deguchi, Y. Kawanishi, I. Ide, H. Murase, and U. Sakai, Pedestrian detection from sparse point-cloud using 3DCNN, In *Proc. 2018 Int. Workshop on Advanced Image Tech. (IWAIT)*, 2018, C4-4.
- [4] H. Fukui, T. Yamashita, Y. Yamauchi, H. Fujiyoshi, and H. Murase, Robust pedestrian attribute recognition for an unbalanced dataset using mini-batch training with rarity rate, In *Proc. 2016 IEEE Intell. Veh. Symp. (IV)*, 2016, pp. 322–327.
- [5] D. Li, X. Chen, and K. Huang, Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios, In *Proc. 3rd IAPR Asian Conf. on Pattern Recogn. (ACPR)*, 2015, pp. 111–115.
- [6] J. Tao and R. Klette, Part-based RDF for direction classification of pedestrians, and a benchmark, *Computer Vision —ACCV 2014 Workshops*, Lecture Notes in Computer Science, vol. 9009, 2014, pp. 418–432.
- [7] Y. Kawanishi, D. Deguchi, I. Ide, H. Murase, and H. Fujiyoshi, Misclassification tolerable learning for robust pedestrian orientation classification, In *Proc. 23rd IAPR Int. Conf. on Pattern Recogn. (ICPR)*, 2016, pp. 481–486.
- [8] D. Baltieri, R. Vezzani, and R. Cucchiara, People orientation recognition by mixtures of wrapped distributions on random trees, *Computer Vision —ECCV 2012*, Lecture Notes in Computer Science, vol. 7576, 2012, pp. 207–283.
- [9] F. Flohr, M. Dumitru-Guzu, J.F.P. Kooij, and D.M. Gavrilu, Joint probabilistic pedestrian head and body orientation estimation, In *Proc. 2014 IEEE Intell. Veh. Symp. (IV)*, 2014, pp. 617–622.

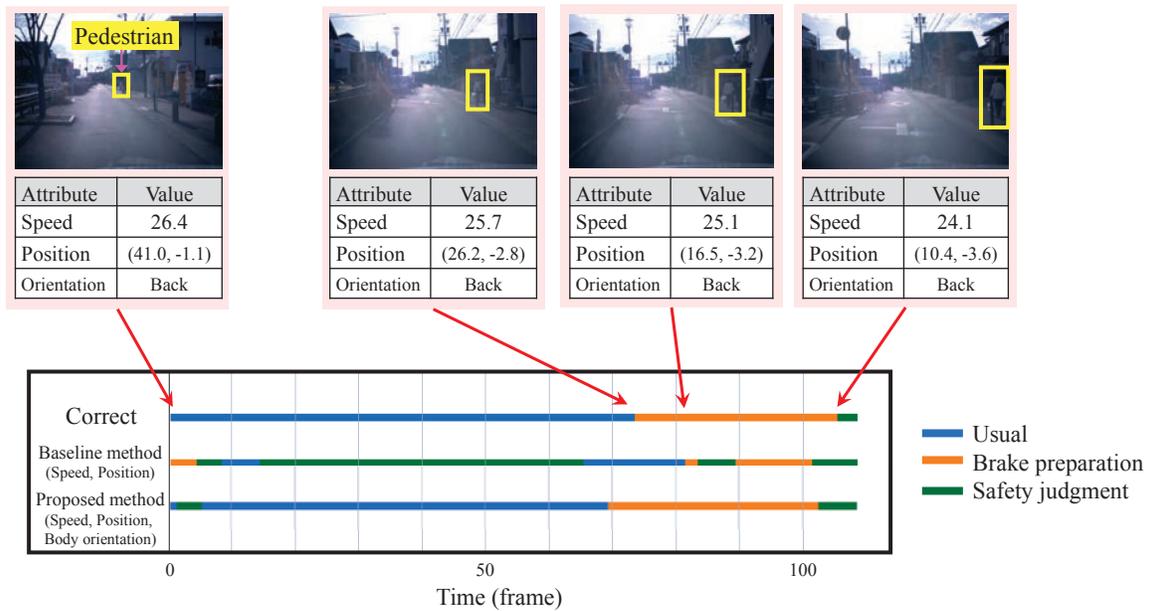


Fig. 6. Example of the experimental results (1).

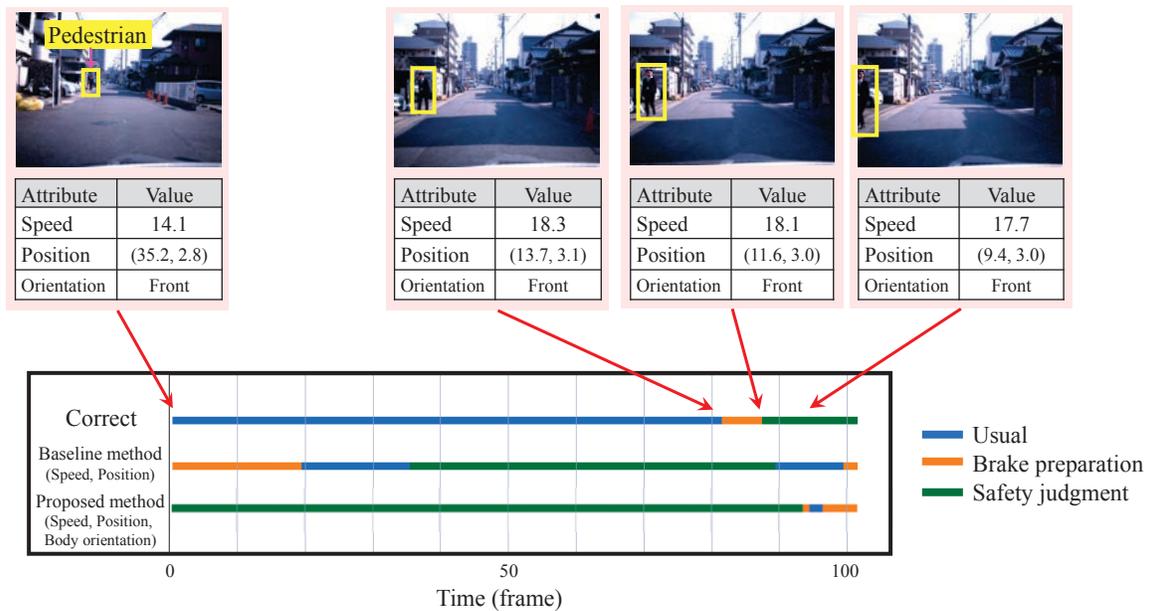


Fig. 7. Example of the experimental results (2).

- [10] F. Shinmura, Y. Kawanishi, D. Deguchi, I. Ide, H. Murase, and H. Fujiyoshi, Recognition of texting-while-walking by joint features based on arm and head poses, *Computer Vision —ACCV 2016 Workshops, Lecture Notes in Computer Science*, vol. 10116, 2016, pp. 452–462.
- [11] S. Köhler, M. Goldhammer, S. Bauer, K. Doll, U. Brunsmann, and K. Dietmayer, Early detection of the pedestrian's intention to cross the street, In *Proc. 15th IEEE Intell. Transp. Syst. Conf. (ITSC)*, 2012, pp. 1759–1764.
- [12] C.G. Keller and D.M. Gavrila, Will the pedestrian cross? A study on pedestrian path prediction, *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, 2014, pp. 494–506.
- [13] F.-H. Chan, Y.-T. Chen, Y. Xiang and M. Sun, Anticipating accidents in dashcam videos, *Computer Vision —ACCV 2016, Lecture Notes in Computer Science*, vol. 10114, 2016, pp. 136–153.
- [14] K. Xu, J. Ba, R. Kiros, A. Courville, R. Salakhutdinov, R. Zemel, and Y. Bengio, Show, attend and tell: neural image caption generation with visual attention, *J. Mach. Learn. Res.*, vol. 37, 2015, pp. 2048–2057.
- [15] M. Shimosaka, K. Nishi, J. Sato, and H. Kataoka, Predicting driving behavior using inverse reinforcement learning with multiple reward functions toward environmental diversity, In *Proc. 2015 IEEE Intell. Veh. Symp. (IV)*, 2015, pp. 567–572.
- [16] A.L. Maas, A.Y. Hannun, and A.Y. Ng, Rectifier nonlinearities improve neural network acoustic models, 30th Int. Conf. on Mach. Learn. (ICML) Workshop on Deep Learning for Audio, Speech, and Language Processing, 2013.
- [17] F. Shinmura, Y. Kawanishi, D. Deguchi, I. Ide, H. Murase, and H. Fujiyoshi, Driver's decision analysis in terms of pedestrian attributes —A case study in passing by a pedestrian—, In *Proc. Workshop on Human Factors in Intelligent Vehicles in the 20th IEEE Int. Conf. on Intell. Transp. Syst. (ITSC)*, 2017, pp. 32–36.