Transportation Research Part C 71 (2016) 164–181

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

A probabilistic model of pedestrian crossing behavior at signalized intersections for connected vehicles ☆

Yoriyoshi Hashimoto [a], Yanlei Gu [b,*], Li-Ta Hsu [b], Miho Iryo-Asano [b], Shunsuke Kamijo [b]

[a] Graduate School of Information Science and Technology, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-8654, Japan
[b] Institute of Industrial Science, The University of Tokyo, Komaba, Meguro-ku, Tokyo 153-8505, Japan

ARTICLE INFO

Article history:
Received 30 May 2015
Received in revised form 9 July 2016
Accepted 25 July 2016
Available online 30 July 2016

Keywords:
Pedestrian behavior
Signalized intersection
Active safety system
Connected vehicle
Dynamic Bayesian Network

ABSTRACT

Active safety systems which assess highly dynamic traffic situations including pedestrians are required with growing demands in autonomous driving and Connected Vehicles. In this paper, we focus on one of the most hazardous traffic situations: the possible collision between a pedestrian and a turning vehicle at signalized intersections. This paper presents a probabilistic model of pedestrian behavior to signalized crosswalks. In order to model the behavior of pedestrian, we take not only pedestrian physical states but also contextual information into account. We propose a model based on the Dynamic Bayesian Network which integrates relationships among the intersection context information and the pedestrian behavior in the same way as a human. The particle filter is used to estimate the pedestrian states, including position, crossing decision and motion type. Experimental evaluation using real traffic data shows that this model is able to recognize the pedestrian crossing decision in a few seconds from the traffic signal and pedestrian position information. This information is assumed to be obtained with the development of Connected Vehicle.

1. Introduction

Autonomous driving and connected vehicles are expected to significantly improve traffic safety and convenience by alleviating the burden of a driver. Currently, they are implemented as a form of an advanced driver assistance system (ADAS) to partially aid drivers. It is also expected that fully autonomous and connected vehicles will emerge as the key component of intelligent transportation systems, replacing human drivers in the near future.

Accidents involving pedestrians are one of the leading causes of death and injury around the world. On the other hand, there is no doubt that the reduction of these accidents should be considered in the development of autonomous driving and connected vehicle. Pedestrian detection has been an active research area and significant progress has been reported over the last two decades (Enzweiler and Gavrila, 2009; Dollar et al., 2012; Dalal and Triggs, 2005; Felzenszwalb et al., 2010; Llorca et al., 2012; Alahi et al., 2014). In addition, the classification of the type of road users was proposed by Zangenehpour et al. (2015). However, the detection and classification are not sufficient to direct the operation of vehicle in driving, especially in crowded urban areas. The reason is that many pedestrians appear around a vehicle, and the frequent detection of pedestrians makes the vehicle brake or be at a stop. In order to reduce traffic accidents as well as smooth the driving task, more advanced

☆ This article belongs to the Virtual Special Issue on: Connected Autonomous Vehicles.

* Corresponding author.
E-mail addresses: hashimoto@kmj.iis.u-tokyo.ac.jp (Y. Hashimoto), guyanlei@kmj.iis.u-tokyo.ac.jp (Y. Gu), qmohsu@kmj.iis.u-tokyo.ac.jp (L.-T. Hsu), m-iryo@iis.u-tokyo.ac.jp (M. Iryo-Asano), kamijo@iis.u-tokyo.ac.jp (S. Kamijo).

http://dx.doi.org/10.1016/j.trc.2016.07.011
0968-090X/

collision avoidance systems are required not only to detect pedestrians around vehicles, but also to understand and predict the behavior of pedestrians. For this purpose, many researchers have worked on pedestrian path prediction and motion classification in the last few years.

A popular choice for target state estimation is the Kalman filter (KF). Pedestrian position, velocity, acceleration can be estimated with appropriate dynamical models and measurement models. Generally, the KF is based on the assumption that pedestrian dynamics approximates a linear dynamical system (LDS), which represents that a pedestrian walks at a constant-velocity and it is formulated as a linear operation. The KF can further be used for prediction by propagating the current state with the dynamical model. The KF was proposed to track pedestrian in image space (Binelli et al., 2005), ground plane (Bertozzi et al., 2004) and 3D space (Alonso et al., 2007). Moreover, various derivative versions of KF, such as extended KF (EKF) and unscented KF (UKF) have been applied for pedestrian tracking as well (Meuter et al., 2008; Junli and Reinhard, 2012). Trajectory provides significant information for state prediction, such as position and velocity. However, pedestrians can instantly change their walking direction, abruptly or start/stop walking. It is not sufficient to assume a single dynamical system for the pedestrian movement. Schneider and Gavrila (2013) proposed a more flexible system, which is composed of multiple linear dynamical models. The multiple models were used to distinguish the different motions of pedestrian, such as walking, stopping, bending in and starting.

Since the pedestrian behavior has highly dynamic property, motion changes indicate the crucial information from the view of traffic safety. The motion changes appear on the posture of pedestrian, which can be observed by considering the visual feature in image space. Koehler et al. (2013) proposed to detect pedestrian's initiation of gait using Motion History Image. Keller and Gavrila (2014) applied dense optical flow to two different models to judge whether the pedestrian approaching the curb will cross in front of the ego-vehicle or stop at the curbside. One model is a Gaussian Process Dynamical Models (GPDM) (Wang et al., 2008) trained with walking and stopping motion separately. The other model adopts probabilistic hierarchical trajectory matching, which matches the trajectory of the feature vector with database classified by motion types. While they employed 2-dimensional features which are vulnerable to ego-motion or change of the observing direction, Quintero et al. (2014a,b) used 3D body language to predict path and classify motions. They applied GPDM systems trained with accurate motion capture data to the pose estimated from noisy disparity images.

Besides the pedestrians own intention, the surrounding environment also affects the behavior of pedestrians. Researchers started to consider the contextual information for pedestrian behavior analysis. Kooij et al. (2014a) focused on the surrounding situations and enriched the impact factors, which cause physical motions of pedestrians. They assumed that the pedestrian decision whether to cross a road way or stop before crossing is influenced by three factors: existences of approaching vehicles, the pedestrian awareness of them and the spatial layout of the environment. In addition, the authors employed the Dynamical Bayesian Network (DBN) to model the relations among the behavior and those factors. Kooij et al. (2014b) also proposed a method based on spatio-temporal context, which switches LDSs according to the pedestrian position from a vehicle perspective. In contrast to the above-mentioned approaches, which aimed at short-time path prediction ($\sim$1 s), the method proposed by Bonnin et al. (2014) realized an accurate/early detection for pedestrians' crossing intention at a specific location: zebra-crossing. The authors considered multiple contextual information, such as pedestrian moving direction and distance to the zebra-crossings, to percept the intention of pedestrians. It is valid to take account of contextual information in pedestrian behavior analysis, since pedestrians do not move randomly. Usually, they assess the traffic situations and have intention or planning at the same time. From the view of whole intelligent transportation systems, the contextual information also could be available, which can be provided by V2I or V2V communication.

Besides the context-based model proposed by Kooij et al. (2014a), there are many approaches making use of Bayesian Network (BN) or its application to time-series data: DBN. They are applied to models of semantic traffic situations or human behavior. Gindele et al. (2010) used a DBN to model the vehicle behavior. Integrating drivers' intentions and interactions to achieve the intentions, the proposed DBN estimates their behaviors and predicts their trajectories. Platho and Eggert (2012) proposed to model the traffic situation of intersection using BNs. The complicated intersection scenarios were decomposed into several sets of simple configurations. Each configuration contains an affecting and an affected entity such as a red traffic signal and a stopping vehicle, respectively. Moreover, Patterson et al. (2003) proposed to use DBN model to recognize the traveler's transportation mode: $\{foot, car, bus\}$ from noisy GPS data stream and additional knowledge, such as existences of parking plots and bus stops.

Our work is highly inspired by Kooij et al. (2014a). We focus on a specific yet crucial scenario: the signalized intersections. The pedestrian behavior at the intersections is highly related to the state of traffic signal, instead of distance between pedestrians and vehicles. In this paper, we propose to use DBN to model the pedestrian behavior at signalized intersections. The proposed DBN probabilistically integrates relations among contexts, pedestrian intention, motion type and physical movement in the way pedestrians actually behave. The model estimates the whole state of a pedestrian in the Bayesian filtering framework. In addition, in the development of the model, we incorporated the traffic engineering knowledge (Iryo-Asano et al., 2015). Moreover, we are aiming to estimate the pedestrian intention, because the intention controls pedestrian behavior in a long period, which is the essential source of driving safety.

The remainder of the paper is organized as follows. In Section 2, we discuss pedestrian behavior at intersections as the motivation and problem statement for this study. In Section 3, we present the details for the DBN-based pedestrian behavior model including the filtering method. Experimental evaluation of the proposed system, including training data collection, is described in Section 4. Finally, we conclude this paper and discuss future work in Section 5.

2. Pedestrian behavior at intersections

As we discussed in Section 1, the pedestrian behavior has relationships with the surrounding environment and traffic situations. Especially at signalized intersections, the pedestrian crossing area and the timing for crossing are limited by the road infrastructure and vehicles. Hamaoka et al. (2013) studied on the pedestrian confirmation behavior: head-turning for approaching right/left-turning vehicles when they are crossing intersections. Iryo-Asano et al. (2015) and Zhang et al. (2013) analyzed pedestrians' decision-making and speed at signalized intersections. They proposed to employ probabilistic distributions and considered the intersection layout and conditions surrounding pedestrian as influential factors. Zeng et al. (2014) simulated pedestrian behavior at signalized intersections by considering the influence from turning vehicles, other pedestrians and so on. Li (2013), Marisamynathan and Perumal (2014) and Koh and Wong (2014) analyzed pedestrian behavior violating traffic signals at intersections.

In Japan, pedestrian signals have three phases: pedestrian green (PG), pedestrian flashing green (PFG) and pedestrian red (PR). Pedestrians are allowed to cross during PG phase and are not allowed during PR phase. During PFG phase, pedestrians are not allowed to start crossing. However, a lot of pedestrians start crossing after the onset of PFG in reality. In this case, pedestrians usually accelerate and rush to cross the road. Those behaviors are very dangerous for turning vehicles. Therefore, we choose the intersection scenario for research. In addition, though pedestrians are required crossing to or returning to the nearer sidewalk if they already started crossing at the onset of PFG. However, they rarely return back to the side they start crossing. Since we cannot obtain enough data of the return cases, we do not handle the behavior in this paper.

Actually, if we observe pedestrian behavior at signalized intersections, we can find tendencies of pedestrian behavior according to signal states. For example, during PFG, pedestrian behaviors are typically categorized into two ways: speeding up to cross and slowing down to stop. We show these tendencies in Fig. 1, which demonstrates the relation between the distance to a crosswalk and the time relative to signal phases when pedestrians change their motion type. The data is generated by observing pedestrian behaviors at a real signalized intersection in Japan. It is also used to evaluate our model and details are described in Section 4. The horizontal axis is the timing relative to the moment of onset of PFG. The negative value means the time before onset of PFG. The vertical axis is the distance to the beginning edge of the crosswalk. The different color of "×" symbol means different type of motion changing, e.g. the blue color is the changing from walking to standing. Therefore, the left most green "×" symbol in Fig. 1 means a pedestrian changed his/her motion from running to walking, when the PFG time is −22 s. At that moment, the distance from him/her to the beginning edge of the crosswalk was about 2 m. Obviously, there are three group of tendencies, which are indicated by different color ellipses. Firstly, the green ellipse shows that when the pedestrian arrives in the 5 m range to the crosswalk and the signal phase is still PG, he/she can cross the crosswalk with sufficient time. Therefore, pedestrians often change running to walking before the crosswalk. Secondly, when the pedestrian signal begins to flash and a pedestrian is approaching the crosswalk, he/she often starts running in order to cross during PFG. This tendency is demonstrated by the pink color ellipse. Thirdly, pedestrians who decide to wait for the next PG stop at 2–6 m in front of the crosswalk edge (blue ellipse) when PFG time passed more than 5 s. The purpose of this paper is to describe these tendencies as a probabilistic model.

There is a practical reason for choosing signalized intersections. A signalized intersection is one of the most hazardous zones in traffic scenes. In fact, nearly half of the accidents between pedestrian and vehicle occurred at intersections in Japan during the year 2015 (Metropolitan Police Department (Japan), 2015). It is because intersections are the main area where

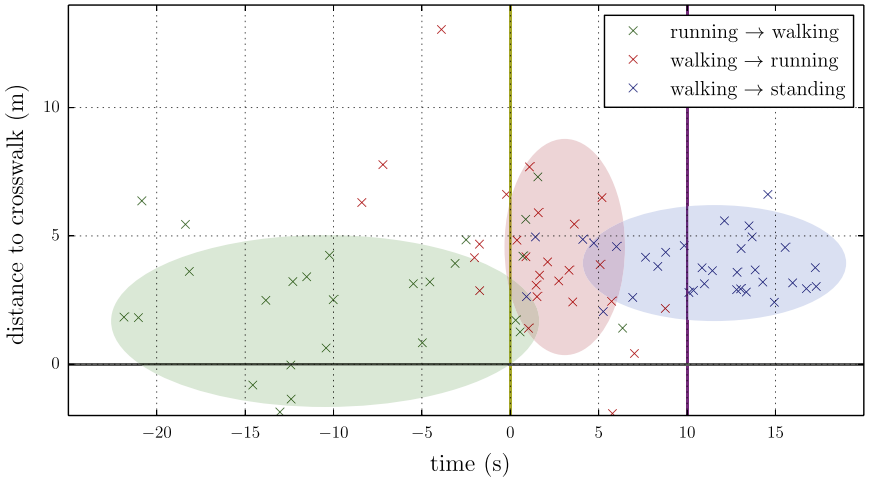


Fig. 1. Illustration of pedestrian crossing behavior at a signalized intersection as a function of time and the distance to the edge of the crosswalk. The origin of the time axis is set to the onset of PFG. Yellow/magenta lines (*time* = 0 s/10 s) represent the onset of PFG/PR, respectively. The distance to the crosswalk is positive/negative value before/after a pedestrian enters the crosswalk, respectively.

pedestrian and vehicle trajectories intersect. Moreover, though traffic signals control traffic flow so that pedestrians and vehicles pass through safely, they frequently hurry up pedestrians as shown by the pink ellipse in Fig. 1. It causes hazardous behavior such as rushing without attention to turning vehicles. In addition, there are wide blind spots on sidewalks for a left-turning driver (in the case of left-hand traffic).

In this paper, we focus on the left-turning situation and near-side pedestrians as shown in Fig. 2. In left-hand traffic context, near-side at an intersection is defined as the sidewalk which locates left side to a left-turning vehicle, and far-side is the other road side. Left and right would be reversed for right-hand traffic. Far-side pedestrians firstly cross halfway before they approach the vehicle and their crossing intention is not easy to be misjudged by a driver or the sensing system. On the contrary, the near-side pedestrians cross the road from the near-side sidewalk to the far-side. Therefore, the collision between a left-turning vehicle and a near-side pedestrian happens in the beginning of pedestrian crossing. This dangerous area is close to near-side sidewalk. In addition, because of a driver blind spot at the left-side, near-side pedestrians are easily caught in hazardous situations. It is necessary for a driver to assess their crossing intentions. The targets of this research are the near-side pedestrians who have intention to cross the crosswalk in front of an ego-vehicle as shown in Fig. 2.

Though we analyze the pedestrian behavior and develop the pedestrian model based on the data obtained in Japan, this research could be significant for other countries or areas such as America (Iryo-Asano and Alhajyaseen, 2014), Singapore (Koh et al., 2014) and Hong Kong (Lee and Lam, 2008), which have the similar traffic situations with PG and PFG, or indicators similar to PFG. Especially in Asian countries, because of the large population, the density of pedestrians in one signal cycle is high. Pedestrians could appear at any time during PG and PFG period like the case in Japan that is shown in this paper. Analyzing and developing the pedestrian behavior model for signalized intersections becomes crucial for Asian countries.

3. Probabilistic model of pedestrian crossing behavior

3.1. DBN overview

Our approach is to construct a pedestrian behavior model with the contexts in the intersection scenario by exploiting the Dynamic Bayesian Network (DBN) (Murphy, 2002). First, the Bayesian Network (BN) is a stochastic model depicted as a directed acyclic graph which consists of nodes and edges. Nodes denote random variables and can take both discrete and continuous values. Edges denote conditional dependencies between nodes as relations of cause and effect. Even non-linear or non-Gaussian can be assigned to them. A DBN is a sequential BN in which nodes can have relations with ones at the adjacent network. This feature makes it possible to model time-series data and a DBN can be regarded as a dynamic system of a Markov process of order one. These probabilistic approaches can take account of sensor noise and estimation confidence. The hidden Markov model can be regarded as a specialization of DBN where unobservable nodes take only discrete values. The Kalman filters can be also regarded as a solution for a specialization of DBN where nodes take continuous values and edges are described with linear operators and Gaussian noise. Because of these flexibilities, DBNs are often used for context-based human behavior models as mentioned in Section 1.

Regarding a DBN as a dynamical system, the Bayesian filtering is employed to estimate its states. The Bayesian filter recursively estimates the posterior probability of the system state $\boldsymbol{x}_t$ conditioned on the all obtained observations $\boldsymbol{o}_{1:t}$. Every

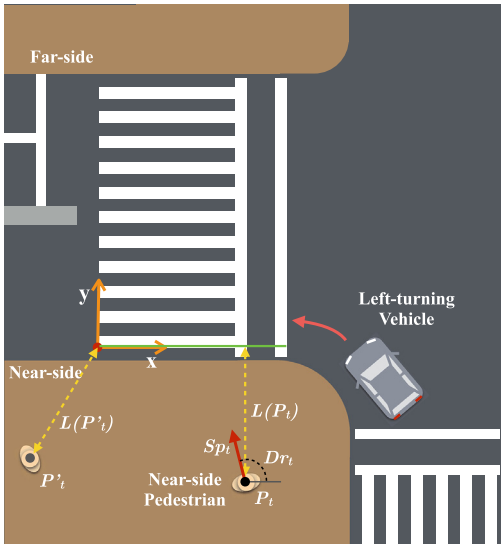


Fig. 2. Definitions of near-/far-side and state variables in our target scenario.

time step, the prior $P(\boldsymbol{x}_t|\boldsymbol{o}_{1:t-1})$ is predicted using the posterior at the previous time step $P(\boldsymbol{x}_{t-1}|\boldsymbol{o}_{1:t-1})$ and the state transition model $P(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ on the assumption of a Markov process of order one. This is achieved by marginalizing out the previous states.

$$P(\boldsymbol{x}_t|\boldsymbol{o}_{1:t-1}) = \int P(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})P(\boldsymbol{x}_{t-1}|\boldsymbol{o}_{1:t-1})d\boldsymbol{x}_{t-1} \tag{1}$$

When a observation $\boldsymbol{o}_t$ becomes available, the posterior $P(\boldsymbol{x}_t|\boldsymbol{o}_{1:t})$ is estimated by correcting the prior with observation likelihood $P(\boldsymbol{o}_t|\boldsymbol{x}_t)$. Based on Bayes' rule, the update formulation is represented as follow:

$$\begin{aligned} P(\boldsymbol{x}_t|\boldsymbol{o}_{1:t}) &= \frac{P(\boldsymbol{o}_t|\boldsymbol{x}_t)P(\boldsymbol{x}_t|\boldsymbol{o}_{1:t-1})}{P(\boldsymbol{o}_t|\boldsymbol{o}_{1:t-1})} \\ &\propto P(\boldsymbol{o}_t|\boldsymbol{x}_t)\int P(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})P(\boldsymbol{x}_{t-1}|\boldsymbol{o}_{1:t-1})d\boldsymbol{x}_{t-1} \end{aligned} \tag{2}$$

In this framework, how to define the state transition model is the key of the filtering performance.

3.2. DBN description

The random variables we consider as state space of the proposed DBN are defined as follows:

S: Traffic signal phases for pedestrians. In Japan, there are three phases for pedestrian:

$$S_t \in \{PG, PFG, PR\} \tag{3}$$

Each value corresponds to pedestrian green/flashing green/red respectively. In this study, we assume that the true signal phase at each time step can be obtained from V2I communication on the assumption of connected vehicles.

D: Pedestrian decisions to cross the crosswalk or wait until next PG.

$$D_t \in \{cross, wait\} \tag{4}$$

We regard this state significant because pedestrian movement follows it. Namely, if a pedestrian decides to cross, he/she will keep moving at least until he/she crosses completely. In this case, a driver or a vehicle system must keep an eye on him/her. Otherwise, he/she stops before he/she enters a roadway and it does not cause any hazardous situations. Knowing this state is helpful to a driver or vehicle system in decisions-makings.

M: Pedestrian motion types.

$$M_t \in \{standing, walking, running\} \tag{5}$$

Running is defined as the moving motion that there are moments when both feet are above the ground, while walking is that one foot is always on the ground. Though running can be divided into detail motion types, jogging, rushing, etc., all these types including walking varies continuously and can be assumed to be a same dynamical system such as constant-velocity model. However, it is important to consider these fundamental motion types separately because they imply pedestrian intentions. For example, running usually accompanies hasty feeling and a running pedestrian definitely has a crossing intention.

Sp: Pedestrian moving speed.

Dr: Pedestrian moving direction.

P: Pedestrian position on the ground plane.

$$P_t = \begin{bmatrix} x_t \\ y_t \end{bmatrix} \tag{6}$$

Note that for Dr and P, the near-side edge line of a crosswalk serves as a reference of the coordinate system as shown in Fig. 2 since we assume pedestrians determine their behavior according to the relative positional relation between them and crosswalks. We denote the distance from a pedestrian to a crosswalk as $L(P)$, which represents the minimum length between the position P and the near-side edge line segment of the crosswalk, and positive/negative value before/after stepping over the edge line. As also shown in Fig. 2, there is usually a lane for cyclists at the side of a crosswalk for pedestrians in Japan. However, we regard it as a part of a crosswalk as pedestrians actually do.

Z: Observation of pedestrian position on the same coordinate system as P.

$$Z_t = \begin{bmatrix} z_{xt} \\ z_{yt} \end{bmatrix} \tag{7}$$

In this state space, we construct a model in accordance with the flow of pedestrian behavior. First, a pedestrian assesses the situation. The situation corresponds to the signal phase and the positional relation to crosswalks in this model. According to the situation, he/she makes decisions of movement. This corresponds to the cross/wait decision and change of motion

types. Then, he/she moves physically based on the decision and motion. The graphical representation of the proposed DBN is shown in Fig. 3. From this representation, we can decompose the state transition model of the DBN as follows:

$$P(D_t, M_t, Sp_t, Dr_t, P_t | D_{t-1}, M_{t-1}, Sp_{t-1}, Dr_{t-1}, P_{t-1}, S_{t-1}, S_t) = P(D_t | D_{t-1}, S_{t-1}, S_t, P_{t-1}) P(M_t | M_{t-1}, S_t, D_t, P_{t-1}) \times P(Sp_t, Dr_t, P_t | Sp_{t-1}, Dr_{t-1}, P_{t-1}, S_t, D_t, M_t) \quad (8)$$

3.3. Filtering models

Each term of the decomposed state transition model (8) represents how each state variable changes probabilistically from the previous time. Using these models, the proposed system filters out improbable states from the huge state space and estimates the true states. We define the subdivided models as follows. In addition to the transition model, we define an observation likelihood model which is also exploited in the filtering process.

3.3.1. Cross-wait decision-making model

For the pedestrian decision-making model, we assume the following rules:

(a) During PG, pedestrians only have *cross* decisions.
(b) At the onset of PFG, pedestrians make decisions to *cross* or *wait*.
(c) During PFG and PR, pedestrians change their decisions at low probabilities.

For the second rule, we follow the model previously proposed (Iryo-Asano et al., 2015). The probability of the wait decision is determined with a logistic function. These rules are formulated as follows:

$$P(D_t = wait | P_{t-1} = \boldsymbol{p}_{t-1}, S_{t-1} = s_{t-1}, S_t = s_t, D_{t-1} = d_{t-1}) = \begin{cases} \text{(a)} & 0 & (s_t = PG) \\ \text{(b)} & \frac{\exp(V_d(\boldsymbol{p}_{t-1}))}{1+\exp(V_d(\boldsymbol{p}_{t-1}))} & (s_{t-1} = PG, \quad s_t = PFG) \\ \text{(c)} & \begin{cases} 1 - q_{w \to c} & (d_{t-1} = wait) \\ q_{c \to w} & (d_{t-1} = cross) \end{cases} & (otherwise) \end{cases} \quad (9)$$

$$P(D_t = cross) = 1 - P(D_t = wait) \quad (10)$$

V_d is a linear function of influential factors. We currently consider only the distance to the crosswalk at the previous time as an explanatory variable:

$$V_d(\boldsymbol{p}_{t-1}) = a_0 + a_1 L(\boldsymbol{p}_{t-1}) \quad (11)$$

This analyzing method called logistic regression is also used for human decision-making models in traffic scenes such as driver/pedestrian gap acceptance model (Rakha et al., 2011; Marisamynathan and Perumal, 2014). Other influential factors such as the crosswalk length considered in the previous model (Iryo-Asano et al., 2015) should be added to the explanatory variables according to the information which can be observed or obtained. However, since we suppose that a pedestrian

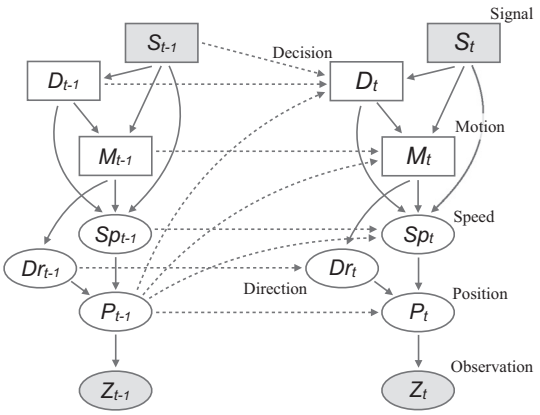


Fig. 3. Two-sliced DBN model of the proposed pedestrian behavior model in a graphical representation. Discrete/continuous/obtained nodes are drawn as rectangular/circular/shaded respectively. Intra-temporal/inter-temporal causal edges are solid/dashed respectively.

determines his/her moving speed according to his/her decision and other contexts as we mention in Section 3.3.3, the speed should not be considered here as the explanatory variable. The parameters $a_0,\ a_1$ are determined by the maximum likelihood estimation.

The parameters $q_{w\to c},\ q_{c\to w}$ are the probabilities that a pedestrian switches the decision from *wait* to *cross*/from *cross* to *wait* at a moment, respectively.

3.3.2. Motion transition model

Pedestrians may change their motion at any time. Therefore, we assume that pedestrians make decisions about a motion change at every time step. The probabilities of the motion changes are defined by the logistic regression in a similar way to the cross/wait decision-making model.

$$P(M_t|M_{t-1}=m_{t-1},S_t=s_t,D_t=d_t,P_{t-1}=\boldsymbol{p}_{t-1})=\frac{\exp(V_m(s_t,d_t,m_{t-1},M_t,\boldsymbol{p}_{t-1})}{1+\exp(V_m(s_t,d_t,m_{t-1},M_t,\boldsymbol{p}_{t-1}))} \quad (12)$$

$$(\text{where}\quad M_t\neq m_{t-1})$$

We also consider only the distance to the crosswalk at the previous time as an explanatory variable. However, this linear function has different constants and coefficients on every combination of the other conditional variables: $S_t,\ D_t,\ M_{t-1}$.

$$V_m(s_t,d_t,m_{t-1},m_t,\boldsymbol{p}_{t-1})=b_{0,s_t,d_t,m_{t-1},m_t}+b_{1,s_t,d_t,m_{t-1},m_t}L(\boldsymbol{p}_{t-1}) \quad (13)$$

For example, $b_{0,PFG,cross,running,walking}$ is a different value from $b_{0,PG,cross,running,walking}$. This means that a pedestrian determines his/her next motion type according to the contexts: the signal phase, decision, present motion type and positional relation to the crosswalk. The parameters are determined by the maximum likelihood estimation.

Eq. (12) defines only the probabilities that a pedestrian switches his/her motion type to another. The probability of staying the same motion type is the remaining one. For example, when a pedestrian is walking at the previous time step, the probability that he/she will keep walking is determined as follows (the other conditional variables $S_t,\ D_t,\ P_{t-1}$ are omitted for simplification):

$$P(M_t=walking|M_{t-1}=walking)=1-P(M_t=standing|M_{t-1}=walking)-P(M_t=running|M_{t-1}=walking) \quad (14)$$

3.3.3. Dynamics model

We decompose the dynamics model into three subsets as follows:

$$P(Sp_t,Dr_t,P_t|Sp_{t-1},Dr_{t-1},P_{t-1},S_t,D_t,M_t)=P(Sp_t|Sp_{t-1},S_t,D_t,M_t,P_{t-1})P(Dr_t|Dr_{t-1},M_t)P(P_t|P_{t-1},Sp_t,Dr_t) \quad (15)$$

We assume that the pedestrian speed is determined in accordance with two factors and further decompose the first term of the right side of (15) into two subsets:

$$P(Sp_t|Sp_{t-1},S_t,D_t,M_t,P_{t-1})\propto P(Sp_t|Sp_{t-1},M_t)P(Sp_t|S_t,D_t,M_t,P_{t-1}) \quad (16)$$

We assume that the moving speed is constant though it is allowed to gradually change with Gaussian process noises v_t (Constant-Speed model). This corresponds to the first terms of the right side of (16).

$$Sp_{t|M_t=m_t}=\begin{cases}Sp_{t-1}+v_t & (m_t=walking,running)\\ 0 & (m_t=standing)\end{cases} \quad (17)$$

$$v_t\sim N\left(0,\sigma^2_{s,m_t}\right)$$

Besides the Constant-Speed model, we also consider a probable speed distribution to each situation and pedestrian behavior. This corresponds to the second term of (16) and we name it the Context-Speed model. For the speed distribution model where $M_t=walking,running$, we employ the gamma distribution with moving parameters as previously proposed (Iryo-Asano et al., 2015).

$$P(Sp_t|S_t=s_t,D_t=d_t,M_t=m_t,P_{t-1}=\boldsymbol{p}_{t-1})\sim Gamma(Sp_t;k(s_t,d_t,m_t,\boldsymbol{p}_{t-1}),\theta(s_t,d_t,m_t,\boldsymbol{p}_{t-1})) \quad (18)$$

The probability density function of gamma distribution is defined using the gamma function $\Gamma(k)$:

$$f(x;k,\theta)=\frac{x^{k-1}\exp(-x/\theta)}{\theta^k\Gamma(k)} \quad (19)$$

The gamma distribution varies its skewness with the shape/scale parameter $k,\ \theta$. In this Context-Speed model, the shape/scale parameters are determined by linear functions of explanatory variables. We consider the distance to the crosswalk at the previous time as the explanatory variables and the linear functions replaces the coefficient and the constant on every combination of the other conditional states: $S_t,\ D_t,\ M_t$ in a similar way to the Motion Transition Model.

$$k(s_t,d_t,m_t,\boldsymbol{p}_{t-1})=\alpha_{0,s_t,d_t,m_t}+\alpha_{1,s_t,d_t,m_t}L(\boldsymbol{p}_{t-1}) \quad (20)$$

$$\theta(s_t,d_t,m_t,\boldsymbol{p}_{t-1})=\beta_{0,s_t,d_t,m_t}+\beta_{1,s_t,d_t,m_t}L(\boldsymbol{p}_{t-1}) \quad (21)$$

These parameters are determined by maximum likelihood estimation. When $M_t = standing$, the speed value is fixed to zero.

As in (16) the probability density of the proposed speed distribution is proportional to the product of the density of the normal and gamma distribution. Combining the conditional states as $C_t = \{Sp_{t-1},\ S_t,\ D_t,\ M_t,\ P_{t-1}\}$ and defining the probability density function (PDF) of the gamma distribution in Context-Speed model as $G_{cx}(Sp|C_t)$ and the PDF of the normal distribution in Constant-Speed model as $N_{cs}(Sp|C_t)$, (16) is written as follows:

$$P(Sp_t|C_t) = \frac{N_{cs}(Sp_t|C_t)G_{cx}(Sp_t|C_t)}{\int N_{cs}(Sp|C_t)G_{cx}(Sp|C_t)dSp} \tag{22}$$

We also assume the moving direction of the pedestrian movement Dr_t is constant though it is allowed to gradually change with Gaussian process noises ϵ_t (Constant-Direction model).

$$Dr_{t|M_t=m_t} = Dr_{t-1} + \epsilon_t, \quad \epsilon_t \sim N\left(0, \sigma^2_{d,m_t}\right) \tag{23}$$

Finally, the position P_t is uniquely fixed by shifting the position at the previous time P_{t-1} with the speed Sp_t and direction Dr_t.

$$P_t = P_{t-1} + Sp_t \begin{bmatrix} \cos(Dr_t) \\ \sin(Dr_t) \end{bmatrix} \tag{24}$$

3.3.4. Observation model

We assume pedestrian positions are measured with zero-mean Gaussian errors as follow:

$$Z_t = P_t + \boldsymbol{w}_t, \quad \boldsymbol{w}_t \sim N\left(\boldsymbol{0}, \sigma^2_m \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \tag{25}$$

3.4. Inference

To estimate the pedestrian state including the decision and motion type, we apply the Bayesian filtering to the DBN defined above with the state transition model (Section 3.3.1, 3.3.2, 3.3.3) and the observation model (Section 3.3.4). However, since the proposed state transition model includes discrete/continuous state and non-Gaussian conditional probability distributions, the integral part in (2) is intractable. Therefore, we employ a sample-based method, the particle filter (PF) (Doucet et al., 2000).

The PF approximates the posterior probability by numerous weighted samples called particles. Each particle has values in the state space and a weight called importance weight. The general PF algorithm is divided into three steps: sampling, importance sampling and re-sampling steps.

In the sampling step which corresponds to the prediction, each particle moves in its state space according to its previous state and the proposed transition model. Note that if it is hard to get samples from the transition model $P(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$, we can exploit another distribution $q(\boldsymbol{x}_t|\boldsymbol{x}_{0:t-1},\ \boldsymbol{o}_{0:t})$ for the sampling. It is called proposal distribution. In our model, at each time step, the new decision value of each particle is sampled at first with the present signal phase, the previous decision/position and the conditional probability defined in Section 3.3.1. Similarly, its new motion type is determined following Section 3.3.2 after fixing the decision. Then, the speed, direction and position are provided by the models defined in Section 3.3.3. However, since we cannot easily get samples from the rather complicated speed model of (22), we approximate the gamma distribution of Context-Speed model $G_{cx}(Sp|C_t)$ by a normal distribution (PDF: $N_{cx}(Sp|C_t)$) having the same mean and variance values. Using the parameters $k,\ \theta$ of the original gamma distribution, the mean and variance parameters of the normal distribution are set to $k\theta$ and $k\theta^2$ respectively. Since the product of two normal distribution is a normal distribution again (26), we can get samples from it.

$$\begin{aligned} &u \sim N(\mu_u, \sigma_u), \quad v \sim N(\mu_v, \sigma_v) \\ &uv \sim N\left(\frac{\mu_u\sigma_v^2 + \mu_v\sigma_u^2}{\sigma_u^2 + \sigma_v^2}, \sqrt{\frac{\sigma_u^2\sigma_v^2}{\sigma_u^2 + \sigma_v^2}}\right) \end{aligned} \tag{26}$$

In the importance sampling step which corresponds to the correction, the importance weight of the i-th particle $w^{(i)}$ is generally updated as follows by using the symbols in Section 3.1:

$$w_t^{(i)} = w_{t-1}^{(i)} \frac{P(\boldsymbol{o}_t|\boldsymbol{x}_t^{(i)})P(\boldsymbol{x}_t^{(i)}|\boldsymbol{x}_{t-1}^{(i)})}{q(\boldsymbol{x}_t^{(i)}|\boldsymbol{x}_{0:t-1}^{(i)}, \boldsymbol{o}_{0:t})} \tag{27}$$

In our model, since the difference between the transition model and the proposal distribution exists in the speed distribution model, (27) can be written as the following.

$$w_t^{(i)} = w_{t-1}^{(i)} P(Z_t|P_t^{(i)}) \frac{\frac{N_{cs}(Sp_t^{(i)}|C_t^{(i)}) G_{cx}(Sp_t^{(i)}|C_t^{(i)})}{\int N_{cs}(Sp|C_t^{(i)}) G_{cx}(Sp|C_t^{(i)}) dSp}}{\frac{N_{cs}(Sp_t^{(i)}|C_t^{(i)}) N_{cx}(Sp_t^{(i)}|C_t^{(i)})}{\int N_{cs}(Sp|C_t^{(i)}) N_{cx}(Sp|C_t^{(i)}) dSp}} \tag{28}$$

Now, we assume the skewness of G_{cx} is small and the variance of N_{cs} is large. In that case, the lower integral approximates the upper one, and then the equation can be written as follows:

$$w_t^{(i)} \approx w_{t-1}^{(i)} P(Z_t|P_t^{(i)}) \frac{G_{cx}(Sp_t^{(i)}|C_t^{(i)})}{N_{cx}(Sp_t^{(i)}|C_t^{(i)})} \tag{29}$$

In order to avoid the situation that all importance weights except a few ones are close to zero, the re-sampling step is conducted when the variance of the importance weights becomes small. In this step, particles are reproduced/discarded so that the number distribution of particles will be proportional to the importance weight distribution before the procedure. Afterwards, all particles are assigned the same importance weights.

4. Experiments

4.1. Dataset description

In order to train and evaluate the proposed DBN and filtering method, pedestrian data in real traffic scenes were collected from a vehicle. We drove around a certain intersection and passed through a specific crosswalk with left-turning several times. Three monocular cameras were mounted on the car. The camera specifications and experimental setting are shown in Table 1 and Fig. 4. The one on the left side view mirror captured the near-side sidewalk and pedestrians, the other ones were used for obtaining signal phases. At the crosswalk, the duration of PG and PFG was always 35, 10 s, respectively. In this paper, we focus on near-side pedestrians who walk towards the crosswalk which the ego-vehicle is going to pass by left-turning. The dataset contains 289 sequences of uninstructed pedestrians.

Pedestrian grounding points and motion types, the edge line of the crosswalk and traffic signal phases are manually labeled. An example of the labeling is shown in Fig. 5. Since we cannot know the true pedestrian intention at each time step, the decision values are labeled as *wait* only after the onset of PFG for the pedestrian who did not cross according to the assumption mentioned in Section 3.3.1. Pedestrian and crosswalk edge positions are projected to 2D ground based on camera parameters and the assumption of a flat road. By converting the coordinate, pedestrian trajectories relative to the crosswalk edge are obtained. We apply Kalman smoother to each trajectory and regard it as the ground truth trajectory.

We divide the sequences into two scenarios: the cross- and wait-scenarios. A cross-scenario includes a pedestrian who goes through the crosswalk (a cross-pedestrian), and a wait-scenario is corresponding to a pedestrian who stops in front of the crosswalk (a wait-pedestrian). The numbers of cross-pedestrians and wait-pedestrians are shown in Table 2. The numbers of sequences with only walking or running and with motion transitions are also shown in Table 2. As shown in Table 2, while there are 251 cross-pedestrians in our database, only 38 wait-pedestrians were collected. It is because the data were collected from the moving vehicle. When the pedestrian shows the waiting behavior, the signal phase is the PFG or PR, our vehicle cannot be at a stop and has to move. The short observation time causes the small number of wait-pedestrian data. On the contrary, it is easy to collect more data of cross-scenarios, because our vehicle can slowly move or stop, when pedestrians cross the road. All the obtained ground-truth trajectories of cross-scenarios and wait-scenarios are shown separately in Fig. 6. As can be seen in Fig. 6(a), the distribution of trajectories of cross-pedestrians is very diverse. The pedestrians come from different directions, and turn-cross or straightly cross road. This diverse data benefits the generation of a stable model. In addition, though the number of wait-pedestrians is less than the number of cross-pedestrians, the trajectory distribution also shows the different cases.

Although each particle makes a decision to cross or wait at the onset of PFG according to the logistic regression (9) in our system, our dataset contains sequences in which pedestrians enter the visual field after the onset of PFG (18 cross-pedestrians and 31 wait-pedestrians). In this case, each particle is assigned a decision in the same way as mentioned in Section 3.3.1 using the measured position Z_0 instead of the hidden position X_0 at the entering moment, namely the initial frame. Though the decision obviously depends on the time elapsed from the onset of PFG at the entering moment, further analysis

Table 1
Specifications of cameras and experimental setting.

Camera model	Point gray FL2G-13S2C-C
Optical lens	Kowa C-Mount 3.5 mm f/1.4
Focal length	3.5 mm
Resolution	1024 × 768 pixel
Frame rate	13 fps
Attached position	Windshield, left/right side view mirror
Vehicle	Toyota NOAH

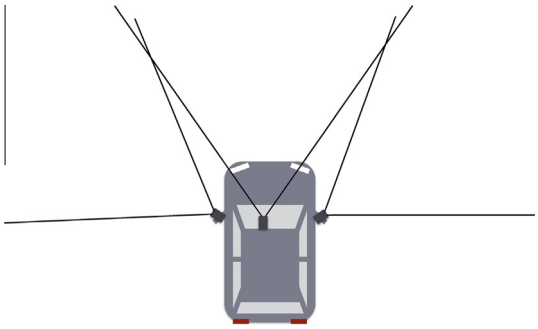


Fig. 4. Positions of attached cameras.

considering time is our future work. The mean and STD value of the distance to the crosswalk of the initial frame are 7.59 and 2.01 m respectively, and those of the last frame are −0.23 and 2.21 m.

4.2. Training & test

We evaluate the performance of the proposed system to estimate the pedestrian states. We use 4-fold cross-validation to divide the dataset into training and test sequences. The parameters in the proposed model are determined by applying the maximum likelihood estimation to the training sub-datasets. As examples, the trained models are visualized in Fig. 7. We can see the pedestrian tendencies are reproduced. For example, pedestrians usually stop around 2–6 m in front of the crosswalk (the cyan line in Fig. 7(b)). As mentioned in Section 2, a pedestrian who decides to cross the road has a slightly faster speed during PFG than the speed during PG (the blue and green lines in Fig. 7(c)). In addition, wait-pedestrians decelerate as they approach the crosswalk (the red and cyan lines in Fig. 7(c)). Note that, only for the parameters $q_{w\to c},\ q_{c\to w}$ in (9), we determine small constants empirically because we cannot know the true moments of decision changing during PFG and assume that it happens at low probability. Though the probabilities might be changed according to the contexts such as the existence of turning vehicles, it is our future work. In this paper, we do not discuss the third rule in Section 3.3 and "decision-making" indicates only the second one.

As the inputs to the proposed system, we simulated the observations of pedestrian positions $\widehat{Z}$ by adding Gaussian noises $\hat{\boldsymbol{w}}$ to the ground truth positions $\overline{P}$ through the same formulation as (25):

$$\widehat{Z}_t = \overline{P}_t + \hat{\boldsymbol{w}}_t, \quad \hat{\boldsymbol{w}}_t \sim N\left(\mathbf{0}, \sigma_n^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \tag{30}$$

where the magnitude of observation error $|\widehat{Z} - \overline{P}|$ follows Rayleigh distribution (chi distribution with 2 degrees of freedom) and its mean and variance are $\sigma_n\sqrt{\frac{\pi}{2}}, \sigma_n^2\left(2 - \frac{\pi}{2}\right)$, respectively. Supposing practical application of the proposed model to

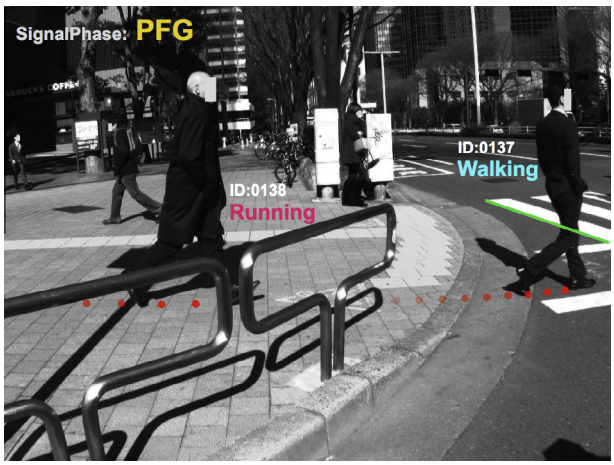


Fig. 5. Image example of the manual labeling. The red points are the labeled grounding points of the pedestrians. The green line is the edge line of the crosswalk. The pedestrian motion type and signal phase (with the other cameras) are also labeled manually. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
The numbers of sequences categorized by scenarios and motion types.

Total	Crossing	Waiting	Walking only	Running only	Walking → standing	Walking → standing	Running → standing
289	251	38	182	29	33	26	26

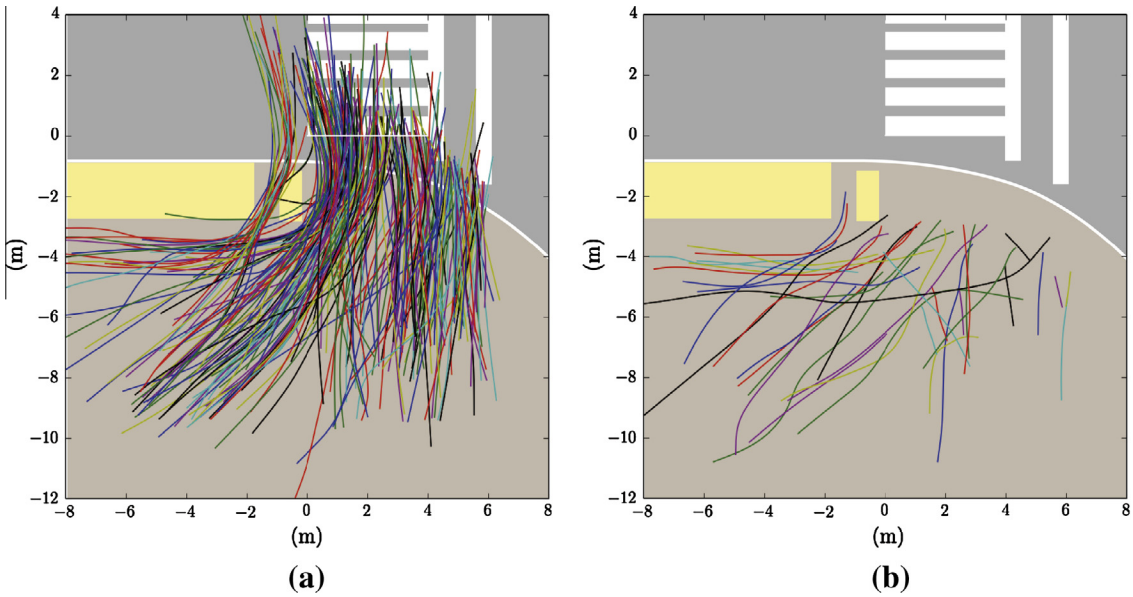


Fig. 6. Ground-truth trajectories of (a) cross-scenarios and (b) wait-scenarios. Yellow rectangles represent the areas where pedestrian cannot walk in due to the design of road structure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

various safety systems, we evaluate noise tolerance of the system based on different levels of the noise: $\sigma_n = 0.1, 0.4, 1.0$ m. Note that the simulated noises are rather large in comparison to the pedestrian dynamics where the average walking distance per frame is 0.1 m in our dataset. We regard σ_n as a known value and σ_m in (25) is set to the same value.

For the particle filtering, we empirically used 2000 particles for each pedestrian. Currently, the algorithm is implemented with a single thread. The computational efficiency was evaluated in a laptop computer with a 2.4-GHz central processing unit. Averagely, the developed algorithm spends 5.1 s to process 1.0 s data of each pedestrian (1.0 s data includes 13 frames). Considering the structure of particle filter algorithm, we could use multiple threads and parallel processing to improve the computational efficiency. We believe that our system is applicable to practical use.

4.3. Case studies

In this section, we show two sequences with the position observations simulated with different variance levels and how the proposed system behaves towards them.

4.3.1. Rushing case

Fig. 8 shows the estimation results for a cross-scenario. Noticing the signal start to flash, a pedestrian decided to cross the crosswalk and began to run. In the left column, the simulated trajectory by each noise variance level is plotted as a green line. As can be seen from Fig. 8(a)(left), the simulated trajectory is very close to the ground truth. In this condition, the system can make the most of the positional information to estimate the other states. On the other hand, in Fig. 8(c)(left), the raw observation trajectory is so complicated that it seems hard to identify the pedestrian states. In this case, the observation likelihood model $P\left(Z_t|P_t^{(i)}\right)$ in (29) becomes a flatter PDF and reduces its influence in weighting particles. It means that the system estimates the pedestrian states relying on the prior knowledge of how a pedestrian behaves in the contexts except positional information.

The left column figures also illustrate the trajectories obtained from weighted averages of particles in our DBN + PF-based system. In addition, comparisons between the observation and filtering errors are shown in the middle column figures. Note that we set the origin of the horizontal axis to the onset of PFG. From these figures, we can see the system smooths the noisy observation trajectories and provides accurate positions.

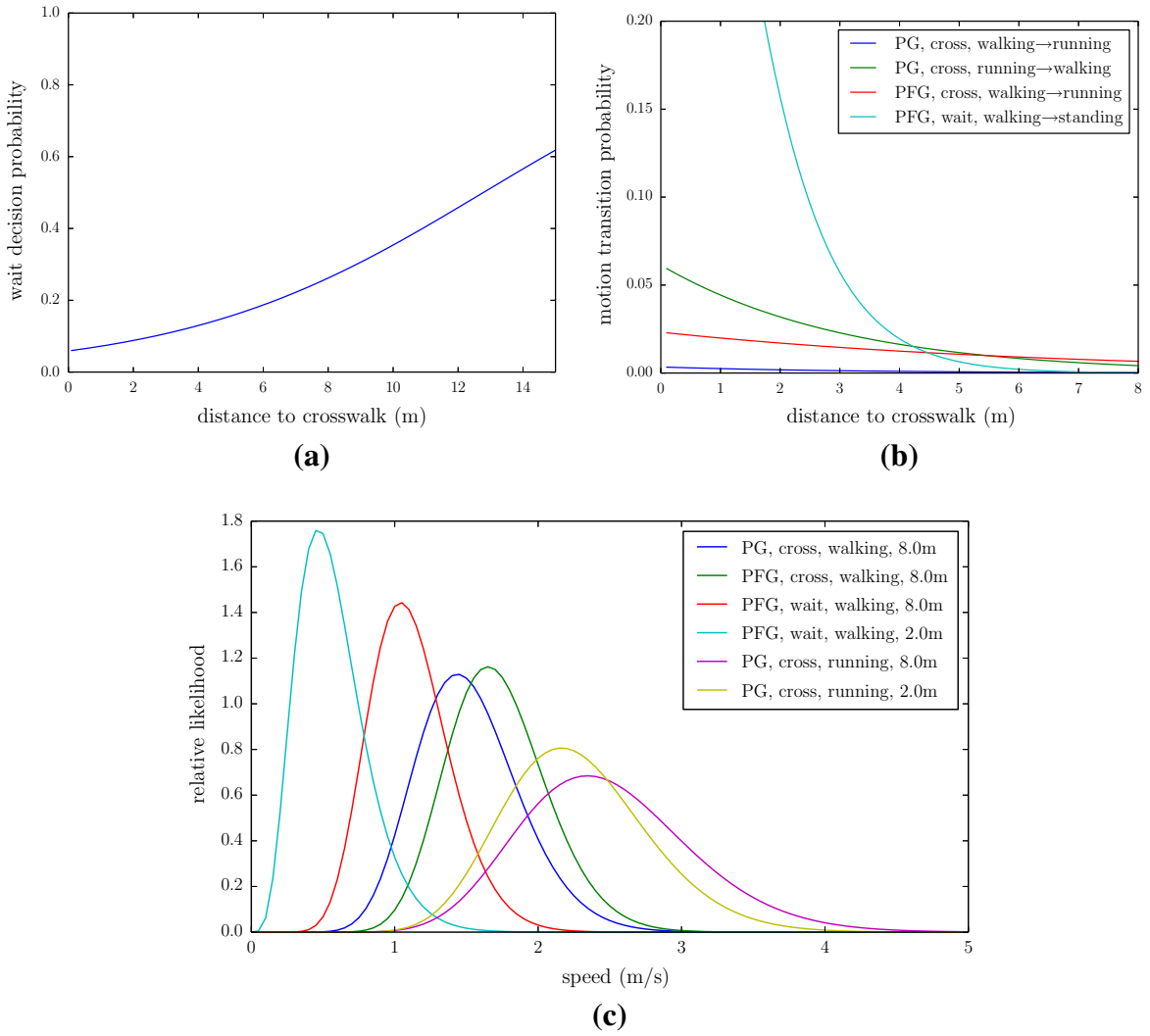


Fig. 7. Visualization of the trained models. (a) Wait decision probability at the onset of PFG in (9). (b) Motion transition probabilities in (12). Legends represent S_t, D_t, M_{t-1} and M_t, respectively. (c) Probability density functions of speed conditioned on the contexts in (18). Legends represent S_t, D_t, M_t and $(\textit{distance to crosswalk})_{t-1}$, respectively.

The right column figures show the posterior probabilities for each decision/motion type value. From the figures, we can find that the estimated cross probability is almost always higher than 90% after the decision-making. The system is confident that the pedestrian will cross the crosswalk from the onset of PFG. The figures also show that the system reacts to the pedestrian running. However, the reaction is delayed as the observation noises become larger. Although high-speed movement is the main clue for detecting the running motion, the highly scattered observations make it hard for the system to find speeding up.

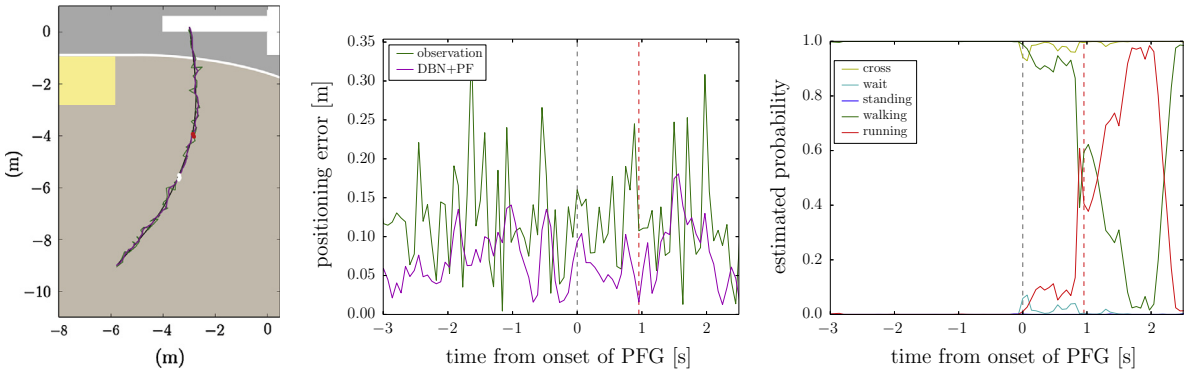
4.3.2. Stopping case

Fig. 9 shows the estimation results for a wait-scenario. Noticing the signal start to flash, a pedestrian decided to wait for next PG and stopped in front of the crosswalk. The matrix of figures is constituted in the same way as (8).

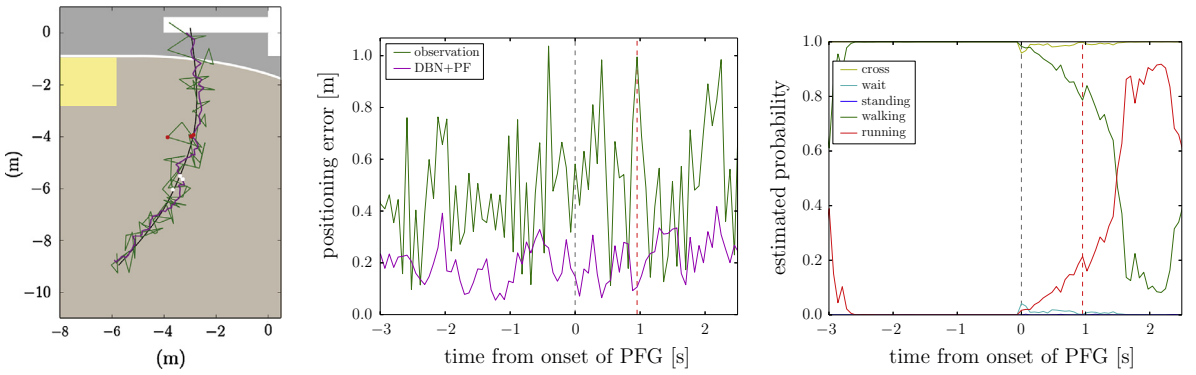
Though the proposed system sometimes outputs worse position than the observation especially around 0 s, it still smooths the noisy position observation in the same way as the previous scenario. The inaccurate positioning results are caused by the pedestrian turning since we assume that pedestrians basically walk straight as defined by (23).

The posterior transitions of the right row figures show that the system quickly recognizes that the pedestrian will not cross. We can also see that the larger the noises are, the later the estimated decision is switched to *wait*. The large noises bring about recognition delay.

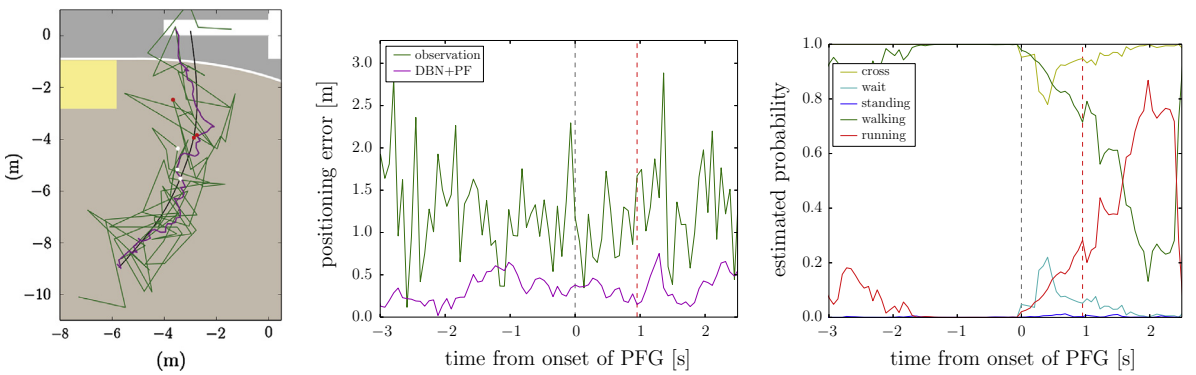
Contrary to the decision estimation, the larger the noises are, the earlier the estimated motion type is switched to *standing*. It seems that the faster detection of pedestrian stop is better. However, *standing* in our model means that a pedestrian is



(a) $\sigma_n = 0.1$



(b) $\sigma_n = 0.4$



(c) $\sigma_n = 1.0$

Fig. 8. Filtering results for the different simulated noise variances σ_n in a cross-scenario. (left) Trajectories of ground truth/simulated observation/estimation result. (middle) Positioning errors of simulated observation/estimated result. (right) Estimated probabilities of each decision and motion type. The white/red points in the left column figure and the gray/red dashed lines in the middle and right column figures represent the onset of PFG/the moment the pedestrian began running, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

not moving. It is different from the *stopping* state which is used in motion classification researches such as (Keller and Gavrila, 2014; Quintero et al., 2014a). The *stopping* represents the decelerating motion of transition from normal walking to stop. Therefore, if *standing* is recognized before the actual stop, it means incorrect. We can say the large noises also deteriorate the estimation of motion type.

4.4. Evaluation on state estimation

In this section, we evaluate the performance of the proposed algorithm based on multiple data. The recognition of decision and motion type are demonstrated by using confusion matrices in Tables 3 and 4, respectively. The decision and motion

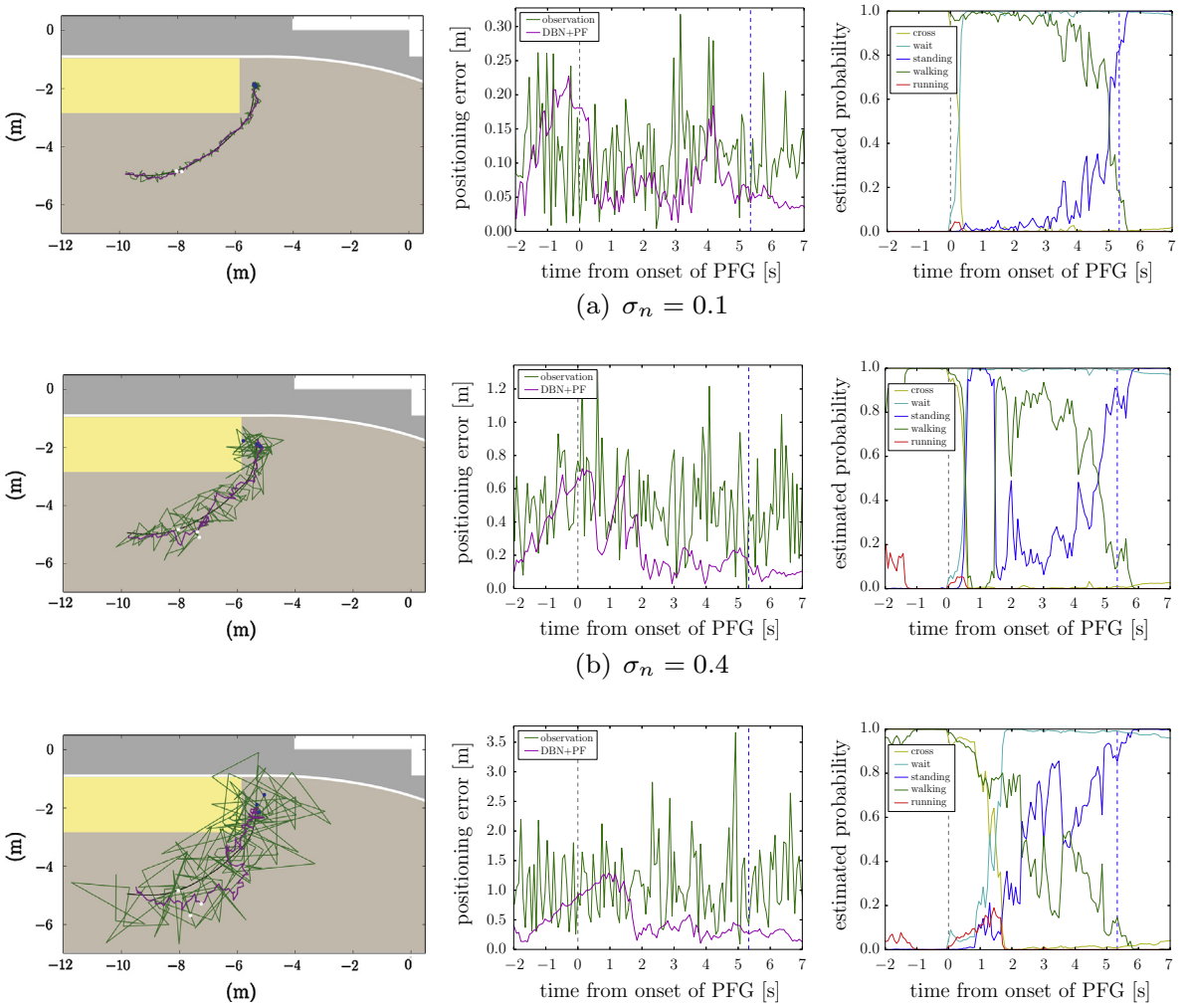


Fig. 9. Filtering results for the different simulated noise variances σ_n in a wait-scenario. (left) Trajectories of ground truth/simulated observation/estimation result. (middle) Positioning errors of simulated observation/estimated result. (right) Estimated probabilities of each decision and motion type. The white/blue points in the left column figure and the gray/blue dashed lines in the middle and right column figures represent the onset of PFG/the moment the pedestrian stopped, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

is determined as the state which has the maximum summation of particle weight. The results in the tables are directly averaged over all frames of all tracks. Precision values in the tables represent the probabilities that decision/motion type recognized as *X* at a time step is actually *X*. In addition, the positioning error is shown in Table 5 as well.

As shown in Table 3, the recognition rates for both *cross* and *wait* decisions are higher than 86% in all the levels of the measurement noise. The experimental results also indicate that the True Positive Rate of *cross* decision is higher than the

Table 3
Normalized confusion matrices of decision recognition for the different simulated noise levels.

		$\sigma_n(m)$					
		0.1 Estimated		0.4 Estimated		1.0 Estimated	
		Cross	*Wait*	*Cross*	*Wait*	*Cross*	*Wait*
Actual	*Cross*	0.98	0.02	0.98	0.02	0.97	0.03
	Wait	0.11	0.89	0.14	0.86	0.14	0.86
Precision		0.98	0.92	0.97	0.91	0.97	0.87

Table 4
Normalized confusion matrices of motion recognition for the different simulated noise levels.

		$\sigma_n(m)$								
		0.1 Estimated			0.4 Estimated			1.0 Estimated		
		Standing	*Walking*	*Running*	*Standing*	*Walking*	*Running*	*Standing*	*Walking*	*Running*
Actual	*Standing*	0.94	0.06	0.00	0.88	0.12	0.00	0.81	0.18	0.00
	Walking	0.01	0.90	0.09	0.02	0.90	0.08	0.02	0.91	0.07
	Running	0.00	0.44	0.56	0.00	0.51	0.48	0.01	0.65	0.35
Precision		0.84	0.93	0.48	0.77	0.91	0.48	0.76	0.89	0.43

one of *wait* decision, which are 98% and 89% respectively in the case of $\sigma_n = 0.1$ m. The reason is the proposed system has a unique decision in the PG time, and recognizes the intention as *cross*. The recognition results are the same as the ground truth of the pedestrian decision in our database during the PG time. It is important to note that pedestrians always have a *cross* intention in the PG time, which is a natural assumption based on the experience in the real world. Further evaluation is conducted by analyzing the recognition rates in the period of PFG, which is discussed in Section 4.5.

The motion recognition performance of the proposed system is shown in Table 4. A recognition rate of 90% is maintained for *walking* motion in different levels of noises. The recognition rate for *standing* motion decreases with the increase of the noise level. The large observation noises raise the uncertainty of the observed movement, which makes it difficult to judge whether the pedestrian is moving or not. Table 4 also shows that *running* motion is difficult to be distinguished from the *walking* motion. In our proposed model, the pedestrian speed is the main clue for estimating the motion type. However, the speed distributions of walking and running overlap each other as we can see in Fig. 7(c), and the input information is not enough to classify them. In the view of the practical system, the running behavior is the most dangerous for the driver and ADASs. There is room for improvement in the proposed algorithm. In addition to speed measurements, the walking and running motions show the difference on the appearance of pedestrian as well. The difference of the two motions has been discussed based on temporal feature in images (Ismail and Tahir, 2013; Arunnehru and Geetha, 2013). Integration of these information with our proposed system could improve the recognition performance. Moreover, Table 4 indicates that large measurement errors decrease the capability of the proposed system to detect deceleration/acceleration in stopping/running progresses, thus, degrades the motion recognition performance.

Table 5 provides the statistic evaluation of the positioning results of our DBN + PF-based system. The mean and standard deviation prove that the proposed method can reduce the positioning errors coming from the measurements, especially for the largest noise case ($\sigma_n = 1.0$ m). At the same time, we can observe that the positioning errors in wait-scenarios are smaller than the one in cross-scenarios. The reason can be explained by referring to the trajectories of the pedestrians in the two scenarios. As shown in Fig. 6(a), cross-scenarios contain many curving trajectories. As we discussed in Section 4.3.2, a pedestrian suddenly turns, which brings difficulty in position estimation due to the assumption of Constant-Direction model in (23). In addition, wait-scenarios contain a standing period. When a pedestrian becomes standing, the detection of standing will constrain the movement of the pedestrian, which reduces the impact of larger positioning error compared to the cross-scenarios.

4.5. Evaluation on crossing decision recognition

In this section, we focus on the evaluation of the decision recognition during the PFG period. The reason we especially evaluate the PFG period is that pedestrians usually make decisions or change behavior during the PFG, and these information need to be recognized.

The pedestrian sequences are aligned along the time axis, whose origin is defined as the decision-making moment in our system. Here, we define the time elapsed from the decision-making moment as time-from-decision (TFD). As mentioned in Section 4.1, there are two kinds of sequences in our dataset. The first one includes the pedestrians who had been in the camera before the onset of PFG. In this case, TFD represents the time from the onset of PFG. The second one corresponds to pedestrians who entered the view of the camera after the onset of PFG, where the TFD represents the time from the entering moments.

Table 5
Mean/STD value (m) of positioning errors for the different simulated noise levels.

	$\sigma_n(m)$		
	0.1	0.4	1.0
Cross	0.08/ 0.07	0.25/ 0.18	0.49/ 0.32
Wait	0.07/ 0.06	0.19/ 0.13	0.39/ 0.23
All	0.08/ 0.07	0.24/ 0.17	0.47/ 0.31

Table 6 shows confusion matrices according to different TFD moments and noise levels of the simulated observations. It consists of 15 3-row × 2-column sub-matrices, and each of them represents a normalized confusion matrix under the joint condition of one TFD moment and one noise level. The experiment results demonstrate that the proposed system achieves more than 85% accuracy to recognize *cross* decision after 2 s observation. On the other hand, the recognition accuracy for *wait* decision is almost always worse than that for *cross* decision, and it takes 3–4 s to obtain an accuracy of 85%. Based on the analysis for Table 6, three tendencies can be concluded:

1. It takes longer time to perceive *wait* decision than *cross* decision.
2. The longer TFD the system has, the more accurately it recognizes *cross* or *wait* decisions.
3. The larger the errors the observations contain, the lower the accuracy the system obtains in the recognition of the decisions.

The conceptual reasoning for these three conclusions can be explained as follows.

The reason for the first tendency is that when a pedestrian makes a decision to wait, he or she does not always start to decelerate just after the decision-making. Since the low speed indicated from the deceleration to stop is the main clue for distinguishing *wait* decision from the cross one, the system maintains the possibility of *cross* decision until the deceleration occurs. In addition, this result also implies that some pedestrians make or change the decisions after the onset of PFG, and then they shift their behavior for stop. In that case, the system cannot provide *wait* decision before the decision-making or changing.

The same reason also can be used to explain the second tendency. In addition to that, observation for a longer period of time could provide more information of the pedestrian state. Therefore, it enables the filtering method to extract the accurate state, and enhance the performance of the proposed system. The reason of the third tendency is that large errors make it difficult to estimate the correct speed of pedestrians. The system utilizes the pedestrian speed to distinguish *cross* and *wait* decisions. However, large errors confuse the system and lead to the decrease of the recognition accuracy.

In order to demonstrate the effectiveness of our proposed method, we additionally conducted the comparison experiment. In the comparison, the baseline method is model previously proposed by Iryo-Asano et al. (2015), because this work also focuses on signalized intersections. This model employs logistic regression to represent the probability of the crossing decisions in the same way as (9)(b) and (11), and this model has two explanatory variables, one is the distance from a pedestrian to the crosswalk, and the other is the traveling speed at the onset of PFG besides. Though the work also considers the crosswalk length as one of explanatory variables, we omitted it because the experiment was conducted at a single crosswalk. We applied this model to the sequences in which the pedestrians are in the visual field at the onset of PFG and regenerated the parameters using our experiment data. The value of the state variables at the onset of PFG were input to the model. Table 7 shows the confusion matrix of the decision recognition results by this previous model on our data at the onset of PFG.

The previous model classifies the crossing decision from the pedestrian state at the onset of PFG. Therefore, we firstly compare the correct recognition rate in Table 6 (diagonal elements in confusion matrix) with the case of TFD = 0.0 s and Table 7. Our proposed method is worse than the baseline method at the moment of the onset of PFG. However, our proposed system considers the different cases (moments after the onset of PFG), and continuously performs state estimation. As shown in Table 6, the performance of our system was improved with the increase of time. We can conclude that, after 2.0 s of TFD, the correctness of the estimation from our proposed model is higher than the one from the previous model, especially for the recognition of waiting behaviors.

These results were caused by two kinds of pedestrians: one is that the pedestrians begin to hurry when they notice the signal starting to flash, the other is that the pedestrians give up crossing evaluating the traffic conditions after the onset of PFG. Since highly dynamic changing of pedestrian behaviors at intersection, it is difficult to distinguish the crossing decisions

Table 6
Normalized confusion matrices of decision recognition during PFG period for the different simulated noise levels.

σ_n (m)			*TFD* (s)									
			0.0 Estimated		1.0 Estimated		2.0 Estimated		3.0 Estimated		4.0 Estimated	
			Cross	*Wait*	*Cross*	*Wait*	*Cross*	*Wait*	*Cross*	*Wait*	*Cross*	*Wait*
0.1	Actual	*Cross*	0.75	0.25	0.91	0.09	0.89	0.11	0.85	0.15	0.96	0.04
		Wait	0.39	0.61	0.34	0.66	0.18	0.82	0.00	1.00	0.03	0.97
	Precision		0.77	0.59	0.79	0.83	0.85	0.86	1.00	0.88	0.96	0.97
0.4	Actual	*Cross*	0.77	0.23	0.80	0.20	0.93	0.07	0.97	0.03	1.00	0.00
		Wait	0.39	0.61	0.29	0.71	0.29	0.71	0.21	0.79	0.03	0.97
	Precision		0.77	0.61	0.80	0.71	0.80	0.90	0.80	0.97	0.96	1.00
1.0	Actual	*Cross*	0.78	0.22	0.65	0.35	0.87	0.13	0.91	0.09	0.88	0.12
		Wait	0.39	0.61	0.26	0.74	0.26	0.74	0.13	0.87	0.05	0.95
	Precision		0.77	0.62	0.78	0.60	0.80	0.82	0.86	0.92	0.91	0.92

Table 7
Normalized confusion matrices of decision recognition at the onset of PFG by Iryo's model.

		Estimated	
		Cross	*Wait*
Actual	*Cross*	0.78	0.22
	Wait	0.29	0.71
Precision		0.94	0.36

using the pedestrian states only at the onset of PFG in the previous model. Our proposed system continuously observes the pedestrian behaviors and overcomes the difficulty caused by highly dynamic changing. In addition, there are some pedestrians who entered the view of the on-board camera after the onset of PFG as mentioned above. This is also a challenge for the previous method, our proposed system could solve this difficulty as well.

5. Conclusion & future work

In this paper, we presented a probabilistic model of pedestrian behavior at signalized intersections. The model is constructed using the Dynamical Bayesian Network by imitating the way pedestrians assess the situations and decide their behavior in real traffic scenes. It describes the stochastic connections among external contexts, pedestrian behavior and physical movement. In the inference step, the model estimates their states jointly by exploiting the particle filter.

We conducted experiment using pedestrian data in real traffic scenes and evaluated the system performance. In the evaluation, the system utilized signal phases and pedestrian positions with various simulated measurement noises on the assumption of connected vehicles. Though the large errors cause lower accuracy and reaction delay, the proposed system correctly recognizes the decision to cross the crosswalk with 2 s of observation and the decision to wait for the next PF period with 3–4 s of observation in every noise case. At the same time, we demonstrated the capability of the system to classify the pedestrian motion and estimate the true pedestrian position from the noisy measurements.

The results showed that the contextual information which is not easy for a stand-alone on-board system to obtain is crucial for advanced vehicle safety systems. The proposed model can find rushing pedestrians who might be in a turning driver's blind spots. This capability can also help drivers to reduce unnecessary waiting for pedestrians who might have already given up crossing. Unlike discriminative models which classify binarily, this stochastic model provides probabilities of the system states as confidence. At the same time, it can take noises into account. These properties make it possible for a vehicle system to behave variously, not limited only to go or stop, according to the probabilities. Especially in the context of autonomous driving, it is helpful to have vehicles behave naturally like a human and prevent undesirable excessive braking.

In this paper, we assumed a strong statement that pedestrians make decisions to cross or wait at the onset of PFG. Unfortunately, it is not always true. Pedestrians sometimes change their decisions and behavior according to the traffic situation such as the existence and positions of turning vehicles. For more exact modeling, an analysis that takes additional contexts into account should be conducted.

We regarded the crossing decision as critical information because it determines pedestrian behavior over a long period while motion classification by the previous works (e.g. walking or stopping) can give only short time prediction. Therefore, we emphasized the evaluation on decision recognition. However, in order to show the advantage of recognizing the decision over previous works, path prediction is required. The reason why we avoided predicting paths is that our model can predict only straight trajectories while our dataset contains curved trajectories. However, as we can see in Fig. 6, the turn mainly occurred at a specific area ($x = (-2\mathrm{m}, 2\mathrm{m}),\ y = (-6\mathrm{m}, -2\mathrm{m})$). Based on this tendency, we are planning to learn the turning motion as a spatial feature and achieve path prediction.

Acknowledgments

The authors gratefully acknowledge the financial support of Semiconductor Technology Academic Research Center (STARC). The pedestrian face images are masked for the purpose of privacy in this paper. This research is permitted by the Compliance Committee of the University.

References

Alahi, A., Bierlaire, M., Vandergheynst, P., 2014. Robust real-time pedestrians detection in urban environments with low-resolution cameras. Transport. Res. Part C: Emerg. Technol. 39, 113–128.

Alonso, I., Llorca, D., Sotelo, M., Bergasa, L., Revenga de Toro, P., Nuevo, J., Ocana, M., Garrido, M., 2007. Combination of feature extraction methods for SVM pedestrian detection. IEEE Trans. Intell. Transport. Syst. 8 (2), 292–307.

Arunnehru, J., Geetha, M., 2013. Behavior recognition in surveillance video using temporal features. In: Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), pp. 1–5.

Bertozzi, M., Broggi, A., Fascioli, A., Tibaldi, A., Chapuis, R., Chausse, F., 2004. Pedestrian localization and tracking system with Kalman filtering. In: IEEE Intelligent Vehicles Symposium (IV), pp. 584–589.

Binelli, E., Broggi, A., Fascioli, A., Ghidoni, S., Grisleri, P., Graf, T., Meinecke, M., 2005. A modular tracking system for far infrared pedestrian recognition. In: IEEE Intelligent Vehicles Symposium (IV), pp. 759–764.
Bonnin, S., Weisswange, T., Kummert, F., Schmuedderich, J., 2014. Pedestrian crossing prediction using multiple context-based models. In: 17th IEEE International Conference on Intelligent Transportation Systems (ITSC), pp. 378–385.
Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 886–893.
Dollar, P., Wojek, C., Schiele, B., Perona, P., 2012. Pedestrian detection: an evaluation of the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 34 (4), 743–761.
Doucet, A., Godsill, S., Andrieu, C., 2000. On sequential monte carlo sampling methods for bayesian filtering. Stat. Comput. 10 (3), 197–208.
Enzweiler, M., Gavrila, D., 2009. Monocular pedestrian detection: survey and experiments. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 31 (12), 2179–2195.
Felzenszwalb, P., Girshick, R., McAllester, D., Ramanan, D., 2010. Object detection with discriminatively trained part-based models. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 32 (9), 1627–1645.
Gindele, T., Brechtel, S., Dillmann, R., 2010. A probabilistic model for estimating driver behaviors and vehicle trajectories in traffic environments. In: 13th IEEE International Conference on Intelligent Transportation Systems (ITSC), pp. 1625–1631.
Hamaoka, H., Hagiwara, T., Tada, M., Munehiro, K., 2013. A study on the behavior of pedestrians when confirming approach of right/left-turning vehicle while crossing a crosswalk. In: IEEE Intelligent Vehicles Symposium (IV), pp. 106–110.
Iryo-Asano, M., Alhajyaseen, W., Nakamura, H., 2015. Analysis and modeling of pedestrian crossing behavior during the pedestrian flashing green interval. IEEE Trans. Intell. Transport. Syst. 16 (2), 958–969.
Iryo-Asano, M., Alhajyaseen, W.K., 2014. Analysis of pedestrian clearance time at signalized crosswalks in Japan. Proc. Comput. Sci. 32, 301–308, The 5th International Conference on Ambient Systems, Networks and Technologies (ANT-2014).
Ismail, A., Tahir, N., 2013. Analysis of walking and running based on markerless model. In: Fifth International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN), pp. 212–216.
Junli, T., Reinhard, K., 2012. Tracking of 2d or 3d irregular movement by a family of unscented Kalman filters. J. Inform. Commun. Converg. Eng. 10 (3), 307–314.
Keller, C., Gavrila, D., 2014. Will the pedestrian cross? A study on pedestrian path prediction. IEEE Trans. Intell. Transport. Syst. 15 (2), 494–506.
Koehler, S., Goldhammer, M., Bauer, S., Zecha, S., Doll, K., Brunsmann, U., Dietmayer, K., 2013. Stationary detection of the pedestrian's intention at intersections. IEEE Intell. Transport. Syst. Mag. 5 (4), 87–99.
Koh, P., Wong, Y., 2014. Gap acceptance of violators at signalised pedestrian crossings. Acc. Anal. Prevent. 62, 178–185.
Koh, P., Wong, Y., Chandrasekar, P., 2014. Safety evaluation of pedestrian behaviour and violations at signalised pedestrian crossings. Safety Sci. 70, 143–152.
Kooij, J., Schneider, N., Flohr, F., Gavrila, D., 2014a. Context-based pedestrian path prediction. In: European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science, vol. 8694. Springer International Publishing, pp. 618–633.
Kooij, J., Schneider, N., Gavrila, D., 2014b. Analysis of pedestrian dynamics from a vehicle perspective. In: IEEE Intelligent Vehicles Symposium (IV), pp. 1445–1450.
Lee, J.Y., Lam, W.H., 2008. Simulating pedestrian movements at signalized crosswalks in Hong Kong. Transport. Res. Part A: Policy Pract. 42 (10), 1314–1325.
Li, B., 2013. A model of pedestrians intended waiting times for street crossings at signalized intersections. Transport. Res. Part B: Methodol. 51, 17–28.
Llorca, D., Sotelo, M., Hellín, A., Orellana, A., Gavilán, M., Daza, I., Lorente, A., 2012. Stereo regions-of-interest selection for pedestrian protection: a survey. Transport. Res. Part C: Emerg. Technol. 25, 226–237.
Marisamynathan, Perumal, V., 2014. Study on pedestrian crossing behavior at signalized intersections. J. Traffic Transport. Eng. 1 (2), 103–110, English Edition.
Metropolitan Police Department (Japan), 2015. Traffic Accident Situations of Pedestrians in the First Half of the Year 2015 (in Japanese). <http://www.keishicho.metro.tokyo.jp/toukei/jiko/data/jiko_hoko.pdf>.
Meuter, M., Iurgel, U., Park, S.-B., Kummert, A., 2008. The unscented Kalman filter for pedestrian tracking from a moving host. In: IEEE Intelligent Vehicles Symposium (IV), pp. 37–42.
Murphy, K.P., 2002. Dynamic Bayesian Networks: Representation, Inference and Learning. Ph.D. thesis, Univ. of California, Berkeley.
Patterson, D., Liao, L., Fox, D., Kautz, H., 2003. Inferring high-level behavior from low-level sensors. In: Ubiquitous Computing (UbiComp). Lecture Notes in Computer Science, vol. 2864. Springer, Berlin, Heidelberg, pp. 73–89.
Platho, M., Eggert, J., 2012. Deciding what to inspect first: incremental situation assessment based on information gain. In: 15th IEEE International Conference on Intelligent Transportation Systems (ITSC), pp. 888–893.
Quintero, R., Almeida, J., Llorca, D., Sotelo, M., 2014a. Pedestrian path prediction using body language traits. In: IEEE Intelligent Vehicles Symposium (IV), pp. 317–323.
Quintero, R., Parra, I., Llorca, D., Sotelo, M., 2014b. Pedestrian path prediction based on body language and action classification. In: 17th IEEE International Conference on Intelligent Transportation Systems (ITSC), pp. 679–684.
Rakha, H., Sadek, S., Zohdy, I., 2011. Modeling differences in driver left-turn gap acceptance behavior using bayesian and bootstrap approaches. Proc. – Soc. Behav. Sci. 16, 739–750, 6th International Symposium on Highway Capacity and Quality of Service.
Schneider, N., Gavrila, D., 2013. Pedestrian path prediction with recursive bayesian filters: a comparative study. In: Pattern Recognition. Lecture Notes in Computer Science, vol. 8142. Springer, Berlin, Heidelberg, pp. 174–183.
Wang, J.M., Fleet, D.J., Hertzmann, A., 2008. Gaussian process dynamical models for human motion. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 30. Springer, pp. 283–298.
Zangenehpour, S., Miranda-Moreno, L.F., Saunier, N., 2015. Automated classification based on video data at intersections with heavy pedestrian and bicycle traffic: methodology and application. Transport. Res. Part C: Emerg. Technol. 56, 161–176.
Zeng, W., Chen, P., Nakamura, H., Iryo-Asano, M., 2014. Application of social force model to pedestrian behavior analysis at signalized crosswalk. Transport. Res. Part C: Emerg. Technol. 40, 143–159.
Zhang, X., Chen, P., Nakamura, H., Asano, M., 2013. Modeling pedestrian walking speed at signalized crosswalks considering crosswalk length and signal timing. Proceedings of the Eastern Asia Society for Transportation Studies, vol. 9.