International Journal of Automotive Technology, Vol. 22, No. x, pp. 1−00(2021)
DOI 10.1007/s12239-021-020x-x

pISSN 1229-9138/ eISSN 1976-3832

MODELING MOTORCYCLE MANEUVERING IN URBAN SCENARIOS USING MARKOV DECISION PROCESS WITH A DYNAMICAL-DISCRETIZED REWARD FIELD

Rina Mardiati[1),4)*], Bambang Riyanto Trilaksono[1)], Sony Sulaksono Wibowo[2)], and Dina Shona Laila[3)]

[1)]School of Electrical Engineering and Informatics, Bandung Institute of Technology, Bandung 40116, Indonesia
[2)]Department of Civil Engineering, Bandung Institute of Technology, Bandung 40116, Indonesia
[3)]School of Mechanical, Aerospace and Automotive Engineering, Coventry University, CV1 5FB, UK
[4)]Department of Electrical Engineering, UIN Sunan Gunung Djati, Bandung 40614, Indonesia

(*Received date 27 December 2019; Revised date 25 May 2020; Accepted date 21 July 2020*)

ABSTRACT−This paper proposes a novel MDP framework to deal with the accuracy of the motorcycle driving model. It proposes a weighted and unweighted Dynamical-Discretized Reward Field (DDRF) as a major contribution on modeling motorcycle maneuver in mixed traffic conditions. Other contributions of this work are the integration of a motorcycle trajectory maneuver model in the state transition function, derivation of probability functions, area of awareness (AoA) and its sectorization to perceive vehicles inside the AoA, which is used to determine actions. We conducted some simulations to evaluate the performance of the proposed model by comparing the data from the simulations with real data. In this study, we use 100 simulation data on motorcycle maneuvering, which consisted of two different scenarios, i.e., 50 data of motorcycle maneuvering to avoid other motorcycles and 50 data of motorcycle maneuvering to avoid cars. We adjusted the simulation setting to the real situation and measured the performance of the proposed model using root mean square error (RMSE). In general, the proposed method can properly model the maneuver of motorcycles in heterogeneous traffic with an RMSE value of around 0.74 meters. This model performs twice as good as an existing car-following model. Furthermore, the proposed reward function performs around 4~6% better than the reward function in previous studies.

KEY WORDS : Markov Decision Process, Motorcycle, Reward function, Driving behavior model

NOMENCLATURE

AoA : area of awareness
DDRF : dynamical-discretized reward field
MDP : markov decision process

1. INTRODUCTION

Motorcycles are the most common mode of transportation in developing countries, such as India, Vietnam, Indonesia, etc. In developing countries motorcycles are very dominant because they provide practicality, convenience and low cost. However, motorcycles produce an erratic and unpredictable maneuver pattern since the riders usually do not follow the lane disciplines. Many researchers agree that the study of modeling motorcycle behavior needs to be improved to find the proper model that can represent closely the real situation (Mardiati *et al*., 2014).

In the current literatures, the studies on motorcycle maneuvering have been investigated using a variety of methods for various purposes. For example, Hausser and Saccon investigate the optimal maneuvering trajectory for decreasing lap time in motorcycle racing by using function of frame flexibility, suspension, and tire (Hauser and Saccon, 2006). Lemonakis proposed an optimal maneuvering trajectory model in curve section road using optimal regression on speed and radius of road's curve (Lemonakis *et al*., 2014).

While Hauser and Lemonakis focused their studies on high performance motorcycle, several researchers focused on modeling motorcycle behavior in urban traffic (Minh *et al*., 2012; Lan *et al*., 2010; Babu *et al*.,

Corresponding author. e-mail: r_mardiati@uinsgd.ac.id

2014). Lan *et al*. (2010) studied mixed traffic model which comprises of motorcycle and car, where Cellular Automaton (CA) is used for modeling an erratic behavior of motorcycle. Based on their study, CA offers not only a robust technique to model heterogeneous traffic, especially in describing object movement but also accommodate internal and external decision factors. Although providing this advantage, CA are limited in modeling rider behavior which is very unpredictable.

To deal with this problem, Minh proposed a maneuverability model framework for motorcycle in queues at signalized intersection using basic of lane changing model (Minh *et al*., 2012). The model performance was assessed using data observation which are collected from video estimation software GAUSS in Hanoi city, Vietnam. While, Babu *et al*. studied modeling motorcycle behavior using social force model to describe the lateral movement and intelligent driver model to describe the longitudinal model in mixed traffic condition consisting of many different types of vehicles. In this model, the neighbor vehicles were identified using a perception line logic to be used for determining the direct lead vehicle, front left and front right vehicles (Babu *et al*., 2014). Modeling behavior of driver was also developed by Lee using artificial neural network learning algorithm to determine a collision risk warning system according to the driving characteristic of the driver (Lee *et al.*, 2018). Collision risk warning model based on lane changing recognition has also been studied by Park using Hidden Markov Model to handle the uncertainty data (Park *et al.*, 2019).

Lately, the multi-agent method provides suitable modeling capability for dynamic and complex heterogeneous traffic systems (Wang *et al.*, 2013). Based on some literatures, the multi-agent system offers highly accurate modeling of heterogeneous traffic (Bazghandi and Pouyan, 2011; Mounir et al., 2013). In multi-agent-based traffic modeling, each vehicle acts as an agent with the capability to make decisions according to external and internal conditions.

One popular decision-making process in the agent-based model is the Markov Decision Process (MDP). Several works have been done in modeling vehicle behavior (four-wheel vehicle) using MDP (Wei et al., 2011; Brechtel et al., 2011; Shimosaka et al., 2015). Based on these literatures MDP can describe agent's behavior by modeling it as a state transition, where the executed action in each step triggers the transition of the states. MDP also has a reward for each possible action, which facilitates a selection of an optimal decision (optimal policy). Unfortunately, the previous study only uses MDP for modeling four-wheel vehicle behavior where the traffic condition is homogenous. Since the two-wheel vehicle (motorcycle) is also important to be modeled especially in urban traffic, MDP was potential to be implemented in modeling motorcycle behavior.

Modeling motorcycle maneuver in urban traffic area is challenging, since many factors influence the maneuver, such as different type of vehicles, the presence of neighbor vehicles, and the condition of the surrounding environment. The existing works described above unfortunately did not comprehensively accommodates these factors (Minh *et al*., 2012; Lan *et al*., 2010; Babu *et al*., 2014). The proposed model in this paper tries to fill this gap. This work presents modeling motorcycle maneuver in urban scenarios where the motorcycles move freely regardless the lane disciplines and choose the action based on the presence of neighboring vehicles.

The method is based on the MDP with improvements in the aspects of probability model, reward function, and weighted rewards function. The probability model is based on states of neighbor vehicles, where the scenario is detailed thoroughly. The reward function is described as a discretized reward field of the road grid based on the state of other vehicles, and finally the weighted reward function is to accommodate a preferable maneuver direction (either left or right). Some weighting functions are investigated, which are Gaussian, Hamming, and Bartlett to represent an aggressive, moderate and safe maneuver behavior, respectively.

The paper is organized as follows. Section 1 gives a general introduction, while Section 2 described the problem formulation. In Section 3 describes the proposed method for modeling motorcycles maneuver using MDP. Moreover, Section 4 presents the process of data acquisition, meanwhile the simulation and model evaluation are described in Section 5. Finally, Section 6 presents the conclusion and future works.

2. PROBLEM FORMULATION

We consider heterogeneous traffic situation with the case study of the traffic in Bandung city, Indonesia. The traffic situation consists of an ego vehicle which is a motorcycle surrounded by neighboring vehicles of various kinds such as motorcycles or cars in one way of two-lane road. The maneuver of ego vehicle and surrounding vehicles are modeled using MDP with ten states and seven actions. The states are defined in terms of position, velocity, and angle, while actions are defined in terms of maneuver and acceleration. More detailed explanation will be provided in Subsection 3.1.

Within the framework of MDP, this work proposes a discretized reward function and its weighted version. The simulations are constructed using the framework of proposed MDP for motorcycle maneuver in mixed traffic and will be carried out and comparison with real world data. The research objective is to investigate how parameter values in the discretized weighted function and its weighted version influence the accuracy of the simulated vehicle maneuver compared with the real data.

3. THE PROPOSED MDP FOR MODELING OF MOTORCYCLE MANEUVER

MDP is a mathematical formulation for decision and control problem when encountering uncertain system behavior (Bellman, 1957). Furthermore, MDP can describe the stochastic behavior of a system, such as maneuver behavior in traffic. Maneuver behavior in MDP is expressed as a state-transition problem where a rider selects certain actions with a goal in mind (Shimosaka *et al.*, 2015). The block diagram of MDP is shown in Figure 1. Based on its definition, MDP has the following five properties (Howard, 1960).

1. S is a finite set of states.
2. A is a finite set of actions (alternatively, A is the finite set of actions available from state s).
3. $P(s', A, s) = P(s'|A, s)$ is the probability that action A in state s at time t will lead to state s' at time $t + 1$.
4. $R_A(s, s')$ is the immediate reward (or expected immediate reward) received after transitioning from state s to state s', due to action A.
5. $\gamma \in (0,1]$ is the discount factor, which represents the difference in importance of future rewards and present rewards.

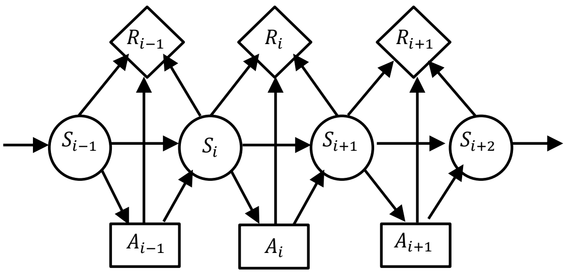


Figure 1. Block diagram system of MDP.

MDPs end goal is to find an optimal policy π^* that will maximize the cumulative function of the random rewards, typically the expected discounted sum over a potentially infinite horizon:

$$\sum_{t=0}^{max} \gamma^t R_{\alpha_t}(s_t, s_{t+1}), \quad (1)$$

where we choose $a_t = \pi(s_t)$.

Suppose $V^\pi(s)$ predicts the cumulative reward when an agent is in the state s with policy π, then $V^\pi(s)$ is computed by

$$V^\pi(i) = E \sum_{k=0}^{\sim} \gamma^k R(s_k, \pi(s_k, s_{k+1})|s_0 = i). \quad (2)$$

3.1. State and Action

The state of the ego vehicle is a function of the position (x, y), velocity v, and steering angle θ which is defined by

$$s = [x, y, v, \theta]^T. \quad (3)$$

An action of motorcycle that will be executed is a function of maneuver $m \in [no\,maneuver,$ $right\,maneuver,$ $left\,maneuver]$ and acceleration $a \in [constant,$ $acceleration,$ $deceleration]$ which is defined as

$$A = [m, a]^T. \quad (4)$$

There are ten states and seven actions to describe the behavior of motorcycles, as mentioned in Table 1 and Table 2.

Table 1. Set of states.

States	Definition
s_0	Vehicle with no speed (stop)
s_1	Vehicle with speed $v_t = v_{t+1}$ and steering straight
s_2	Vehicle with speed $v_t = v_{t+1}$ and steering right
s_3	Vehicle with speed $v_t = v_{t+1}$ and steering left
s_4	Vehicle with speed $v_t > v_{t+1}$ and steering straight
s_5	Vehicle with speed $v_t > v_{t+1}$and steering right
s_6	Vehicle with speed $v_t > v_{t+1}$ and steering left
s_7	Vehicle with speed $v_t < v_{t+1}$ and steering straight
s_8	Vehicle with speed $v_t < v_{t+1}$ and steering right
s_9	Vehicle with speed $v_t < v_{t+1}$ and steering left

Table 2. Set of actions.

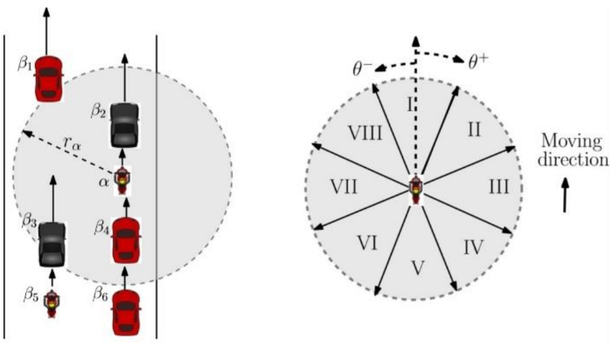
Action	Definition
$A1$	Fixed velocity with no maneuver
$A2$	Increases speed with no maneuver
$A3$	Decreases speed with no maneuver
$A4$	Fixed velocity with right maneuver
$A5$	Increases speed with right maneuver
$A6$	Fixed velocity with left maneuver
$A7$	Increases speed with left maneuver

3.2. Area of Awareness

In this work, we identify a neighbor vehicle using Area of Awareness (AoA). AoA is a virtual area to assess neighbor vehicle which is assumed as obstacles (leading vehicle), room for maneuvers (vehicle in right front or left front), and threats (following vehicle). This simplified model depicts the AoA as a circular area with radius r_α as shown in Figure 2(a). The radius of r_α will depend on the velocity v_α of the ego vehicle as defined by

$$r_\alpha = k_r v_\alpha, \tag{5}$$

where k_r is a scaling factor that indicates the awareness of rider to the surrounding environment.



(a) (b)

Figure 2. (a) AoA area of vehicle under consideration, (b) Illustration of AoA division into eight sectors.

After determining the radius of the AoA, the sectorization stage determines which vehicles are inside the AoA. The AoA sectorization consists of eight sectors as shown in Figure 2(b), which are then categorized as an obstacle, room for maneuvers, and threats.

We can determine the vehicles inside the AoA and categorize these vehicles based on the AoA sectorization by determining the distance and angle between the ego vehicle and the neighboring vehicle using

$$r_{\alpha\beta} = \left(\left(x_{\beta_i} - x_\alpha\right)^2 - \left(y_{\beta_i} - y_\alpha\right)^2\right)^{\frac{1}{2}}, \tag{6}$$

and

$$\theta_i = \tan^{-1}\left(\frac{y_{\beta_i} - y_\alpha}{x_{\beta_i} - x_\alpha}\right), \tag{7}$$

where $r_{\alpha\beta}$ is the distance of motorcycle and other vehicle, (x_α, y_α) is the position of ego vehicle, and $\left(x_{\beta_i}, y_{\beta_i}\right)$ is the position of vehicle β_i. The vehicle β_i is inside AoA if the $r_{\alpha\beta_i} \leq r_\alpha$.

3.3. Probability Function

The probability function is used to calculate the probability value of each action shown in Table 2. This probability function is derived based on the logical thinking of riders when they make a movement that is described in the flowchart diagram as shown in Figure 3. In this figure, there are six parameters that influence the motorcycle movements, namely the absence of leading vehicle, the speed of leading vehicle, the speed of ego vehicle, room or space of maneuver, the safer space or room for maneuver, and threat or the absence of another vehicle behind ego vehicle. These six parameters were represented by probability functions $(P1, P2, \ldots, P6)$, respectively. The derivation of $P1, P2, \ldots, P6$ were done in previous study (Mardiati *et al*., 2018) and rewritten in the following for completeness.

1. The probability of a leading vehicle (obstacle) $(P1)$, which is denoted as

$$P1 = \begin{cases} e^{-k_1(d_{\alpha\beta_1} - d_c)}, & for\ d_{\alpha\beta_1} > d_c \\ 1, & for\ d_{\alpha\beta_1} \leq d_c \end{cases}, \tag{8}$$

where $d_{\alpha\beta_1}$ is the distance of vehicle α (motorcycle) and β_1 (leading vehicle), d_c is the critical distance of vehicle α to make a decision of maneuver or stop, and k_1 is a constant value.

2. The probability of a slower leading vehicle $(P2)$, which is denoted as

$$P2 = \tfrac{1}{2}\left(sign\left(v_\alpha - v_{\beta_1}\right) + 1\right), \tag{9}$$

where, v_α and v_{β_1} are the speed of vehicle α and β_1, respectively. The value of $P2$ are 0 and 1.

3. The probability of motorcycle occupying a comfortable speed v_c $(P3)$, which is denoted as

$$P3 = \tfrac{1}{2}\left(sign(v_\alpha - v_c) + 1\right). \tag{10}$$

The possible values for $P3$ are 0 and 1.

4. The probability of room to maneuver $(P4)$, which is denoted as

$$P4 = \max\left(P_{\theta_L}, P_{\theta_R}\right), \tag{11}$$

where $P_{\theta_L} = \xi e^{k_3 \frac{(d_{\alpha\beta_2} - d_{C_l})}{W} - k_2 \gamma_l}$ is the probability of making a left maneuver, and $P_{\theta_R} = \xi e^{k_3 \frac{(d_{\alpha\beta_3} - d_{C_r})}{W} - k_2 \gamma_r}$ is the probability of making a right maneuver. While, $d_{\alpha\beta_2}$ is the distance of motorcycle α and left side vehicle β_2, $d_{\alpha\beta_3}$ is the distance of motorcycle α and right side vehicle β_3, d_{C_l} is the critical distance of motorcycle α to make a decision of left maneuver, d_{C_r} is the critical distance of motorcycle α to make a decision of right maneuver, $W = d_{\alpha\beta_2} + d_{\alpha\beta_3}$, and k_2, k_3, ξ are the constant.

5. The probability that right maneuvering is safer than left maneuvering $(P5)$, which is denoted as

$$P5 = \arg\max\left(P_{\theta_L}, P_{\theta_R}\right) - 1. \tag{12}$$

6. The probability of a threat $(P6)$, which is denoted as

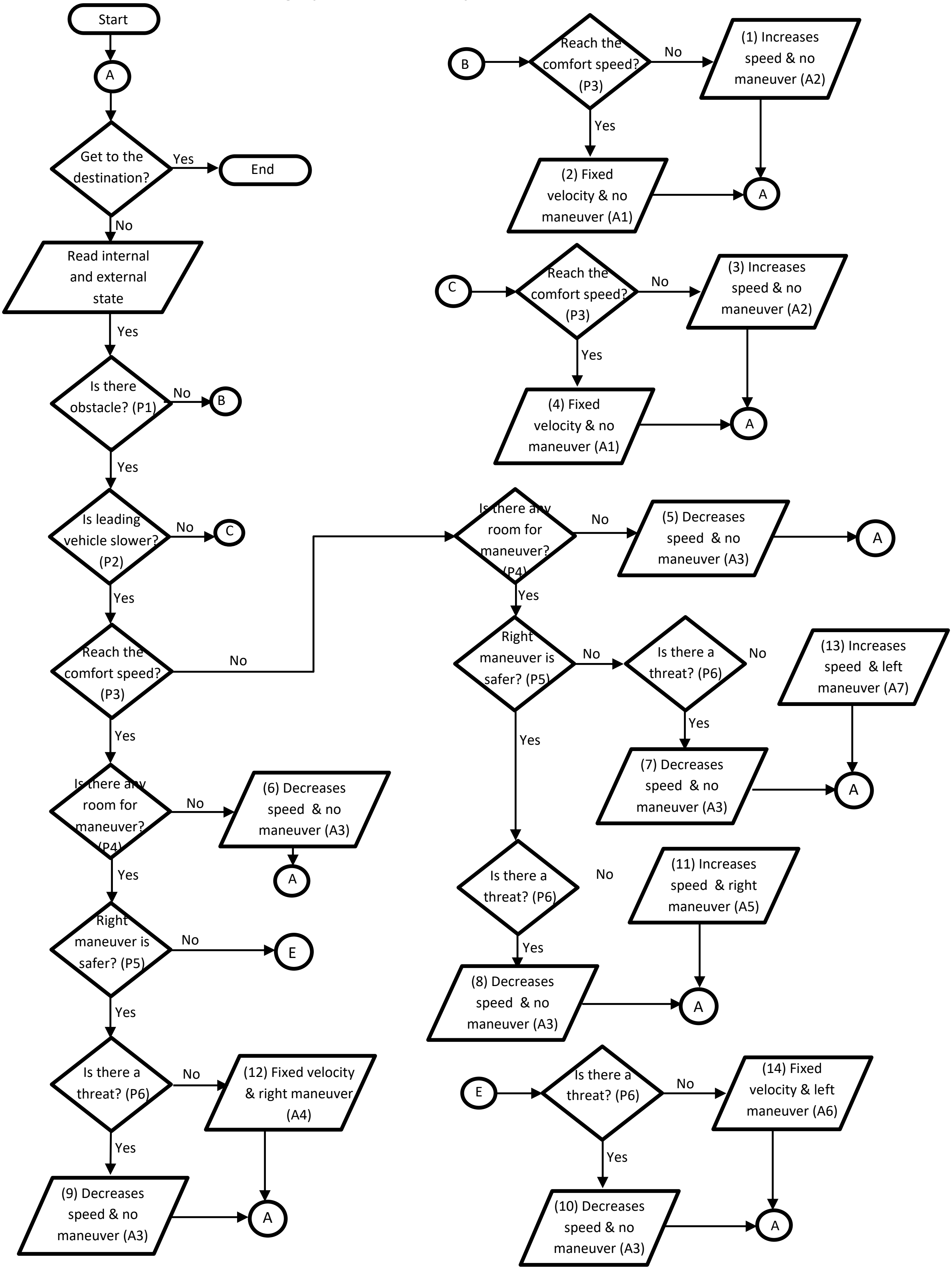


Figure 3. The logic of a rider's decision-making process.

$$P6 = e^{-k_x d_x}, \tag{13}$$

where $d_x = \min(d_{x_1}, d_{x_2}, \cdots, d_{x_n})$, and d_{x_i} is the distance of motorcycle α and vehicle β_i. Note that vehicle β_i is another vehicle behind motorcycle.

After determining the probability function of $P1, P2, \ldots, P6$, the next step is to calculate the probability function of each action. Based on Figure 3, there are fourteen output blocks of actions that refer to actions $A1, A2, \cdots, A7$ which are described in Subsection 3.1. As an example, to find the probability function for action $A1$ (fixed velocity with no maneuver) can be obtained from two scenarios as follows (refer to Figure 3).

1. There is no obstacle or leading vehicle in front, and the rider has reached a comfortable speed (output block number 2).
2. There is an obstacle or leading vehicle in front, the speed of the leading vehicle is faster than the ego vehicle, and the rider has reached a comfortable speed (output block number 4).

In the two scenarios above, each situation represents independent events and produce the same action which is action $A1$. Therefore, we can obtain the probability of action $A1$ for each scenario by multiplying the probabilities of parameters that influenced the action using Equations (8)-(13). Finally, the final probability function for action $A1$ is derived by adding the probability of action $A1$ in both scenarios above. By using the same derivation process, we can find the probability function for all actions as follows.

$$P(A1) = \left(1 - e^{-k_1(d_{\alpha\beta_1} - d_c)}\right)\tfrac{1}{2}(sign\ (v_c - v_a) + 1) + \tfrac{1}{4} e^{-k_1(d_{\alpha\beta_1} - d_c)}\ (1 - sign\ (v_\alpha - v_\beta) + 1)\ (sign\ (v_c - v_a) + 1) \tag{14}$$

$$P(A2) = \left(1 - e^{-k_1(d_{\alpha\beta_1} - d_c)}\right)\left(1 - \tfrac{1}{2}(sign\ (v_c - v_a) + 1)\right) + e^{-k_1(d_{\alpha\beta_1} - d_c)}\left(1 - \tfrac{1}{2}(sign\ (v_\alpha - v_\beta) + 1)\right)\left(1 - \tfrac{1}{2}(sign\ (v_c - v_a)\ + 1)\right) \tag{15}$$

$$P(A3) = e^{-k_1(d_{\alpha\beta_1} - d_c)}\tfrac{1}{2}(sign\ (v_\alpha - v_\beta) + 1)[1 - \max(P_{\theta_L}, P_{\theta_R}) + \max(P_{\theta_L}, P_{\theta_R})]e^{-k_x d_x} \tag{16}$$

$$P(A4) = e^{-k_1(d_{\alpha\beta_1} - d_c)}\tfrac{1}{2}(sign\ (v_\alpha - v_\beta) + 1)\tfrac{1}{2}(sign\ (v_c - v_a)\ + 1)\max(P_{\theta_L}, P_{\theta_R}) \tag{17}$$

$$P(A5) = e^{-k_1(d_{\alpha\beta_1} - d_c)}\tfrac{1}{2}(sign\ (v_\alpha - v_\beta) + 1)\left(1 - \tfrac{1}{2}(sign\ (v_c - v_a) + 1)\right)\max(P_{\theta_L}, P_{\theta_R})\ \left(\arg\max(P_{\theta_L}, P_{\theta_R}) - 1\right)(1 - e^{-k_x d_x}) \tag{18}$$

$$P(A6) = e^{-k_1(d_{\alpha\beta_1} - d_c)}\tfrac{1}{2}(sign\ (v_\alpha - v_\beta) + 1)\left(\tfrac{1}{2}(sign\ (v_c - v_a) + 1)\right)\max(P_{\theta_L}, P_{\theta_R})(1 - (\arg\max(P_{\theta_L}, P_{\theta_R}) - 1))\ (1 - e^{-k_x d_x}) \tag{19}$$

$$P(A7) = e^{-k_1(d_{\alpha\beta_1} - d_c)}\tfrac{1}{2}(sign\ (v_\alpha - v_\beta) + 1)\left(1 - \tfrac{1}{2}(sign\ (v_c - v_a) + 1)\right)\max(P_{\theta_L}, P_{\theta_R})(1 - (\arg\max(P_{\theta_L}, P_{\theta_R}) - 1))\ (1 - e^{-k_x d_x}) \tag{20}$$

3.4. Reward Function

The reward function evaluates the action taken by the vehicle agent. In general, the reward function calculates the value reward of ego vehicle if it performs a certain action. In this research, we propose a reward function which is modeled on the road grid to describe the reward or punishment that is given to the ego vehicle if it selects that grid. The reward model called the dynamical discretized reward field (DDRF). Using DDRF, each grid on the road has its own reward value which depends on the steering angle and distance of the ego vehicle to surrounding grid, and the movement of the vehicle agent, which is denoted as

$$R^{\alpha}_{x_g, y_g} = \frac{\sqrt{\cos\theta_g}}{\sqrt{(x_g - x_\alpha)^2 + (y_g - y_\alpha)^2}}\, sign\ (y_g - y_\alpha), \tag{21}$$

where $R^{\alpha}_{x_g, y_g}$ is the reward value on the grid (x_g, y_g), (x_α, y_α) is the motorcycle's position, and θ_g denotes the angle of motorcycle towards the grid (x_g, y_g).

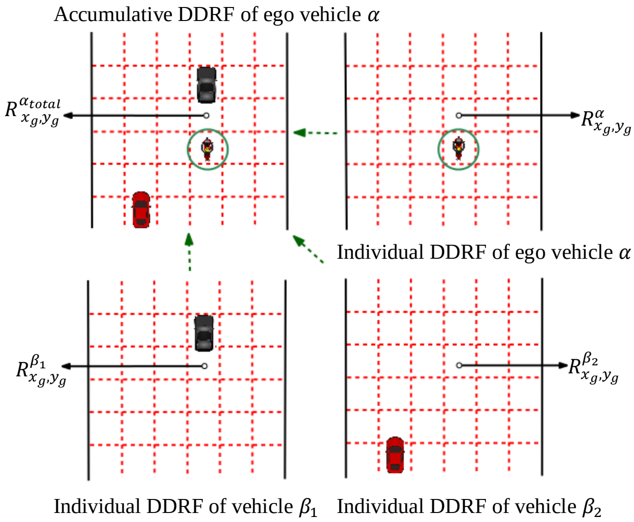


Figure 4. Illustration of individual and accumulative DDRF.

Each vehicle has its own reward field, called as an individual DDRF as shown in Figure 4. In Figure 4, we calculate each grid in the individual DDRF using (21), assuming that the road is empty. Since other vehicles influence the vehicle movement, the final reward field used for each vehicle is produced by accumulating the individual DDRF of neighboring vehicles. In the accumulative DDRF, the value reward of each grid is defined as a superposition function of each individual DDRF, which is expressed by

$$\begin{aligned} R^{\alpha_{total}}_{x_g,y_g} &= R^{\alpha}_{x_g,y_g} - \left(\left|R^{\beta_1}_{x_g,y_g}\right| + \cdots + \left|R^{\beta_n}_{x_g,y_g}\right|\right) \\ &= R^{\alpha}_{x_g,y_g} - \sum_{i=1}^{n}\left|R^{\beta_i}_{x_g,y_g}\right|, \end{aligned} \tag{22}$$

where $R^{\alpha_{total}}_{x_g,y_g}$ denotes the reward value of the grid (x_g, y_g) in the accumulative DDRF and $R^{\beta_i}_{x_g,y_g}$ denotes the reward value of the vehicle β_i on grid (x_g, y_g).

Most countries, including Indonesia, have traffic rules that encourage riders to maneuver in a specific direction, for example maneuvering to the right lane is favorable to maneuvering to the left lane. We can embed this feature by using weighted function. In this paper, we propose three weighted reward functions, i.e., Gaussian, Hamming, and Bartlett as shown in (23), (24), and (25) respectively. These functions have different characteristics in terms of maximum amplitude to the previous amplitude ratio (κ).

$$R^{\alpha}_{x_g,y_g}(f_g) = R^{\alpha}_{x_g,y_g}\frac{1}{\sqrt{2\pi\sigma^2}}exp^{-\frac{(x_g-\mu)^2}{2\sigma^2}} \tag{23}$$

$$R^{\alpha}_{x_g,y_g}(f_h) = R^{\alpha}_{x_g,y_g}\left(0.56 - 0.46\cos\frac{2\pi(n-n_0)}{M}\right) \tag{24}$$

$$R^{\alpha}_{x_g,y_g}(f_s) = \begin{cases} R^{\alpha}_{x_g,y_g}\frac{1}{n_b}(n-n_b), n \le n_b \\ R^{\alpha}_{x_g,y_g}\frac{n}{n_b} - 2, n > n_b \end{cases} \tag{25}$$

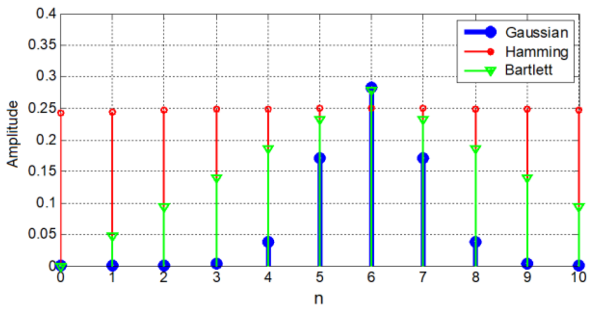


Figure 5. Comparison of three weighted reward functions (Gaussian, Hamming, Bartlett).

In (23), the mean of distribution μ is taken from the actual abscissa of the ego vehicle added by an offset, while the variance σ^2 is taken from the width of the road. In Hamming weighting, M is the width of the road, while n_0 is the abscissa of the ego vehicle. Finally, in Bartlett weighting, n_b is the abscissa of the ego vehicle added by a small offset. Figure 5 illustrates these three weighted reward functions for a road width of 10 meters with the position of motorcycle at $n_0 = 5$ and an offset of 1 meter. As seen in Figure 5, we observe that the Gaussian weighted function has a greater weighted value than Hamming and Bartlett, as shown in $n = 5$ and $n = 6$.

3.5. State Transition Model

After calculating the probability and reward values, we use (2) to select the optimal action. After generating the action, the state transition model is used to move from the current state to the next state. In the state transition model, the trajectory of the motorcycle's maneuver is obtained from collected data by observation. The transition model for the ego-vehicle is divided into two conditions, i.e., doing a maneuver and no maneuver. The set of states (coordinate velocity and orientation) at time $t + 1$ for doing a maneuver is given by

$$\begin{bmatrix} x' \\ y' \\ v' \\ \theta' \end{bmatrix} = \begin{bmatrix} x \\ y \\ v \\ \theta \end{bmatrix} + \begin{bmatrix} \left(vt + \frac{1}{2}at^2\right)\sqrt{1 - \left(\frac{\Delta B}{vt}\right)^2} \\ \left(vt + \frac{1}{2}at^2\right)\left(\frac{\Delta B}{vt}\right) \\ at \\ \Delta\theta \end{bmatrix}, \tag{26}$$

and no maneuver is given by

$$\begin{bmatrix} x' \\ y' \\ v' \\ \theta' \end{bmatrix} = \begin{bmatrix} x \\ y \\ v \\ \theta \end{bmatrix} + \begin{bmatrix} 0 \\ \left(vt + \frac{1}{2}at^2\right) \\ at \\ 0 \end{bmatrix}, \tag{27}$$

where $[x, y, v, \theta]^T$ denotes the current state, $[x', y', v', \theta']^T$ denotes the next state, a is an acceleration, and $\Delta B = a_1(x_1 - x_0) + a_2(x_1^2 - x_0^2) + a_3(x_1^3 - x_0^3) + a_4(x_1^4 - x_0^4)$ denotes the displacement vector. ΔB was derived based on trajectory model which has been done in (Mardiati *et al.*, 2019).

4. DATA ACQUISITION

4.1. Observation Area

We collected the data using a video recorder set up on the high position (8.2 meters above the ground with a 74^0 elevated angle) near the targeted road as can be seen in Figure 7. The video captured vehicle movements over 73 meters. We conducted the survey from 06:30 a.m. to 08:30 a.m. local time (Bandung city, Indonesia) which is considered as rush hours. An example of the collected data is shown in Figure 6.

(a) (b)

Figure 6. Data collection: (a) motorcycle maneuvering other motorcycle, (b) motorcycle maneuvering a car.

4.2. Data Analysis

We used transformation matrix coordinates to convert video screen coordinates into roadway coordinates. We derived the formula of the transformation matrix coordinates based on the illustration in Figure 7. The derivation of these formula was done in previous study (Mardiati *et al*., 2019).

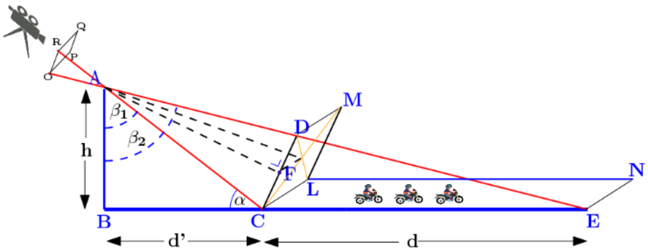


Figure 7. Illustration of the picture in pixels, the picture on the perpendicular side, and the picture on the actual side.

It is necessary to transform the captured video coordinate to real-world coordinate. Using the geometry analysis of Figure 7, the transformation of video coordinate to real-world coordinate is given as (Mardiati *et al*., 2017) :

$$\begin{bmatrix}x"\\ y"\end{bmatrix} = \begin{cases} \begin{bmatrix}\frac{s}{b} & 0\\ 0 & \frac{r}{a}\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & k_1\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}, l_a < \frac{r}{2} \\ \begin{bmatrix}\frac{s}{b} & 0\\ 0 & \frac{r}{a}\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & k_2\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}, l_a \geq \frac{r}{2} \end{cases}, \qquad (28)$$

where (x, y) is pixel coordinate, $(x", y")$ denotes real-world coordinate, r is the length of CD in Figure 7, s is the width of the road, a and b denotes the spatial vertical and horizontal resolution of the video respectively, and the parameters k_1 and k_2 are calculated as:

$$k_1 = \frac{\sin\left[\tan^{-1}\frac{AF}{FH}\right]}{\sin\left[\tan^{-1}\frac{AB}{BC}+\tan^{-1}\frac{AF}{CF}-\tan^{-1}\frac{AF}{FH}\right]} \qquad (29)$$

$$k_1 = \frac{\sin\left[180-\tan^{-1}\frac{AB}{BC}-\tan^{-1}\frac{AF}{CF}\right]}{\sin\left[\tan^{-1}\frac{AB}{BC}+\tan^{-1}\frac{AF}{CF}-\tan^{-1}\frac{AF}{FJ}\right]} \qquad (30)$$

We used observations data from 100 motorcycle maneuvers to avoid other motorcycles or cars to validate our proposed model.

5. SIMULATION AND ANALYSIS

In this section, we present some simulation results to investigate the effect of the proposed model to the accuracy of the trajectory model. In addition, simulations have also been done to compare our MDP model with other models. We evaluate the accuracy using root mean square error (RMSE), denoted as:

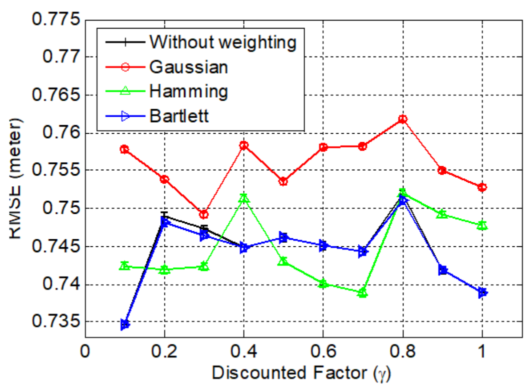
$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x}_i)^2 + (y_i - \bar{y}_i)^2} \qquad (31)$$

where N is the amount of data, (x_i, y_i) is the coordinate of the actual trajectory, and $(\bar{x}_i, \bar{y}_i)$ is the coordinate of the simulation trajectory.

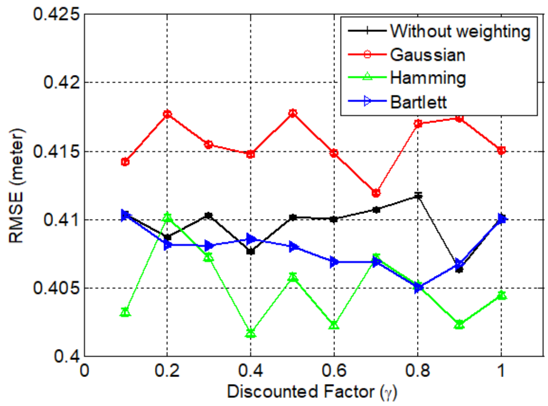
5.1. Simulations to Compare the Weighted and Unweighted Reward Function

In this section, we conduct a series of simulations to determine the effect of the proposed reward function on the accuracy of the trajectory of the maneuvers that our model produced. Simulations were performed for both mixed data (100 observation data of motorcycle maneuvers) and specific data (50 observation data of motorcycle avoid another motorcycle and motorcycle avoid a car) using unweighted and weighted rewards.

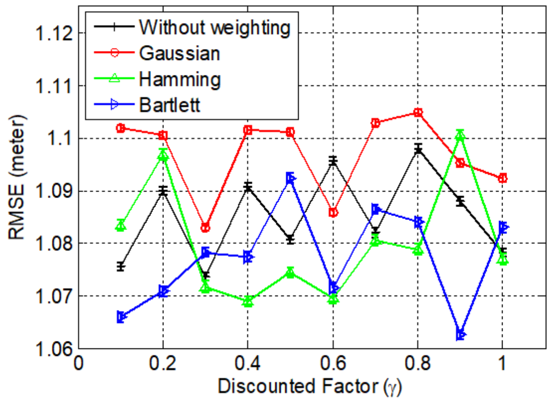
Simulations were conducted to investigate the performance of our proposed model with unweighted and weighted reward. The results are shown in Figure 8. Figure 8 shows that the weighted reward function performs better than the unweighted reward function. Specifically, based on Figure 8, Hamming and Bartlett weighting functions show better performance for three different data cases.



(a)



(b)



(c)

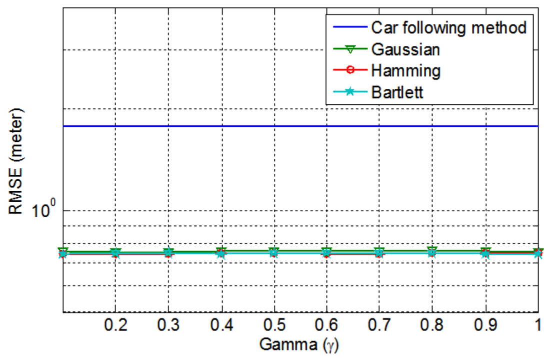
Figure 8. Comparison of unweighted and weighted reward functions for reward depth level 1 and $k_r = 1$ using: (a) mixed data consists of 100 simulation data, (b) specific data consists of 50 simulation data of a motorcycle maneuvering around a car, and (c) specific data consists of 50 simulation data of a motorcycle maneuvering around another motorcycle.

5.2. Simulations to Compare the Proposed Model with Other Model

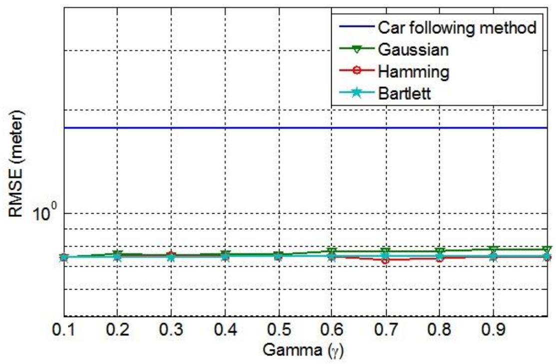
In this section, we compare the proposed MDP model with two other approaches, namely the car-following model that Chang and Chon used (Chang and Chon, 2015) and the reward function model that Brechtel used (Brechtel *et al.*, 2011).

Figure 9 shows the comparison between the car-following model and our proposed model. This comparison shows that our proposed method is two times better than the car-following method. This is due to the simplicity of the car-following model, compared to the MDP method that accommodates stochastic and complex environments. Apart from that, the car-following method also lacks a reward function and produces actions only for that moment without considering the next few events.

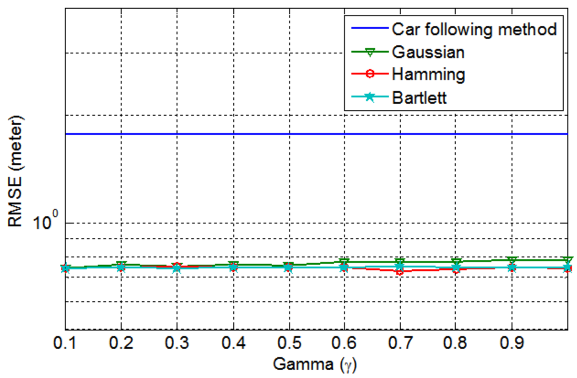
Meanwhile, Figure 10 shows the comparison of the proposed model with the reward function that Brechtel (2011) used in their work. Figure 10 showed the comparison of Brechtel reward function with the weighting reward functions in this study (Gaussian, Hamming, and Bartlett). The simulation result shows that the proposed reward function performs 4-6% better on average than Brechtel's reward function.



(a)

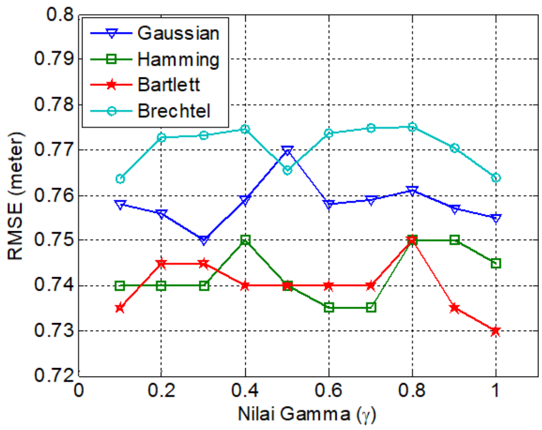


(b)

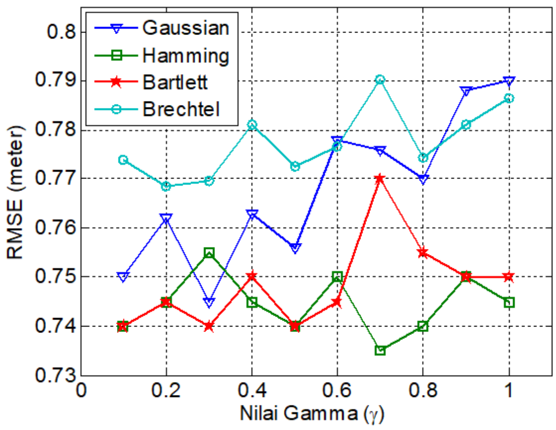


(c)

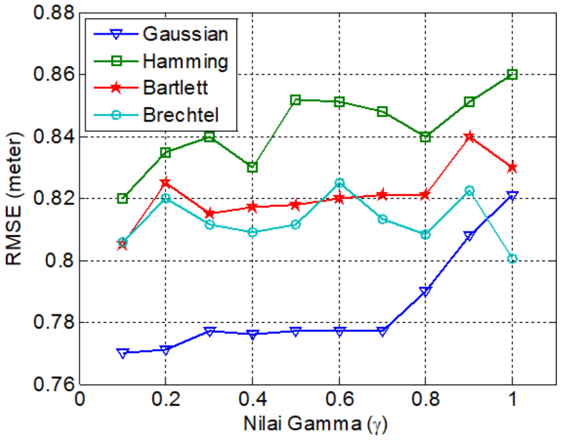
Figure 9. Comparison of the proposed model compared to the basic car-following model with (a) reward depth level 1, (b) reward depth level 2, and (c) reward depth level 3.



(a)



(b)



(c)

Figure 10. Comparison of the proposed reward model to the Bretchel reward function in: (a) reward depth level 1, (b) reward depth level 2, and (c) reward depth level 3.

6. CONCLUSION

This paper presented a novel maneuvering model of motorcycle for an urban road scenario using MDP with Dynamical-Discretized Reward Function. Compared with prior models that used a basic reward function model, our proposed reward model performs better in describing motorcycle behavior in mixed traffic conditions.

In general, our proposed method can properly model the behavior of motorcycles in heterogeneous traffic with a root mean square error (RMSE) value of around 0.74 meters, based on a comparison of the simulation results with actual data. This result is twice as good as the car-following model. Furthermore, the reward function (DDRF) proposed in this study performed around 4-6% better than the reward function in previous studies.

The simulation results lead to a number of specific conclusions. Firstly, the RMSE of a motorcycle maneuvering around another motorcycle is greater than that of a motorcycle maneuvering around a car, which shows that modeling motorcycles maneuvering around another motorcycle is more difficult than modeling a motorcycle maneuvering around a car. This is due to the motorcycles' highly dynamic movements. Moreover, generally, the proposed method performs better at the reward depth levels 1 and 2 than at reward depth level 3, which shows that motorists tend not to think about the possibility of moving far ahead. Furthermore, the effect of (AoA) on the proposed method shows that motorists have moderate AoA coverage while riding. In addition, the discount factors performed well at small range values; and adding a weighting function to the reward model led to better performance, especially for the Hamming and Bartlett weighting function.

Furthermore, the results of this study provide further research opportunities to be implemented in the behavior of four-wheeled riders, as well as opportunities to develop the functions of AoA, DDRF, probability to improve model performance. This proposed model also could be extended to Partially Observable MDPs (POMDP) which able to cope with uncertain and incomplete perception.

REFERENCES

Babu, F.A.A., Vortisch, P. and Mathew, T.V. (2014). Modelling of motorcycle movements in mixed traffic conditions. *In: 94th Annual Meeting Transportation Research Board.*

Bazghandi, A. and Pouyan, A.A. (2011). An agent-based simulation model for urban traffic system. *Computer and Information Science*, **4, 4,** 72−82.

Bellman, R. (1957). A Markovian Decision Process. *Journal of Mathematics and Mechanics*, 679−684.

Brechtel, S., Gindele, T. and Dillmann, R. Probabilistic MDP-behavior planning for cars. *In: Intelligent Transportation System (ITSC), 2011 14th International IEEE Conference*, 1537−1542.

Chang, K. and CHON, A. (2005). A car-following model applied reaction times distribution and perceptual threshold. *Journal of the Eastern Asia Society for Transportation Studies,* **6,** 1888−1903.

Hauser, J. and Saccon, A. (2006). Motorcycle modeling for high-performance maneuvering. *IEEE Control System Magazine* **26**, **5**, 89−105.

Howard, R. A. (1960). Dynamic Programming and Markov Process. *Willey for The Massachusetts Institute of Technology*.

Lan, L.W., Chiou, Y.-C., Lin, Z.-S. and Hsu, C.-C. (2010). Cellular automaton simulations for mixed traffic with erratic motorcycle behaviors. *Physic A: Statistical Mechanics and Its Applications,* **389, 10,** 2077−2089.

Lee, S. H., Lee, S. and Kim, M. H. (2018). Development of a driving behavior-based collision warning system using a neural network. *International journal of automotive technology*, **19 (5)**, 837-844.

Lemonakis, P., Eliou, N., Karakasidis, T. and Botzoris, G. (2014). A new methodology for approaching motorcycle rider's behavior at curved road sections. *European Transport Research Review* **6, 3,** 303–314.

Mardiati, R., Ismail, N. and Faroqi, A. (2014). Review of microscopic model for traffic flow. *ARPN Journal of Engineering and Applied Sciences,* **9, 10,** 1794−1800.

Mardiati, R., Mulyana, E., Maryono, I. and Usman, K. (2019). The derivation of matrix transformation from pixel coordinates to real-world coordinates for vehicle trajectory tracking. *In: 2019 5th International Conference on Wireless and Telematics (ICWT)* (pp.1-5). IEEE.

Mardiati, R., Trilaksono, B.R., Gondokaryono, Y.S. and Wibowo, S. S. (2017). Motorcycles trajectory tracking model based on polynomial least-squares approximation. *Advanced Science Letters,* **23, 5,** 4537−4541.

Mardiati, R., Trilaksono, B.R., Gondokaryono, Y.S. and Wibowo, S. S. (2018). Motorcycle movement model based on Markov chain process in mixed traffic. *International Journal of Electrical and Computer Engineering (IJECE),* **8, 5,** 3149−3157.

Minh, C.C., Sano, K. and Matsumoto, S. (2012). Maneuvers of motorcycle in queues at signalized intersections. *Journal of Advanced Transportation,* **46, 1,** 39–53.

Mounir, G., Rais, N. and Idrissi, M.A. (2013). The system-agent system and stochastic process in the road traffic. *International Journal of Computer Application,* **62, 12,** 20−26.

Park, S., Lim, W., & Sunwoo, M. (2019). Robust Lane-Change Recognition Based on An Adaptive Hidden Markov Model Using Measurement Uncertainty. *International Journal of Automotive Technology*, ***20*** **(2)**, 255-263.

Shimosaka, M., Nishi, K., Sato, J. and Kataoka, H. (2015). Predicting driving behavior using inverse reinforcement learning with multiple reward functions towards environmental diversity. *In: Intelligent Vehicles Symposium (IV), IEEE,* 567−572.

Wang, S., Dai, Z. and He, J. (2013). A multi-agent simulation of traffic system under mixed traffic condition. *In: 2013 Third International Conference on Instrumentation, Measurement, Computer, Communication and Control,* 1646−1649.

Wei, J., Dolan, J. M., Snider, J. M. and Litkouhi, B. (2011, May). A point-based MDP for robust single-lane autonomous driving behavior under uncertainties. In *2011 IEEE International Conference on Robotics and Automation* (pp. 2586-2592). IEEE.