A Reinforcement-Learning-Based Assisted Power Management With QoR Provisioning for Human–Electric Hybrid Bicycle

Roy Chaoming Hsu, *Member, IEEE*, Cheng-Ting Liu, and Din-Yuen Chan

***Abstract*—In this paper, a reinforcement-learning-based assisted power management (RLAPM) with quality-of-riding (QoR) provisioning is proposed for the human–electric hybrid bicycle or the pedelec, which is a light electric vehicle (LEV) driven mainly by human's pedal force with the assisted force from a battery-powered electric motor. By learning from the changes in the riding environment, the proposed RLAPM adaptively dispatches appropriate assisted motor power to support the rider in meeting the QoR requirement, i.e., safety and comfort, of the pedelec. With the RLAPM, not only pedelec rider's QoR is guaranteed, but energy utilization of battery is improved as well. Simulations of the RLAPM and the RLAPM enhancement (RLAPME) for the pedelec are performed under different road types. Experimental results demonstrate that the achievability of the comfort riding of the RLAPME is improved by 24%, while the energy utilization is improved by 50% by comparing with other existing assisted power methods for the pedelec in urban riding.**

***Index Terms*—Assisted power management, human–electric hybrid bicycle, quality of riding (QoR), reinforcement learning (RL).**

I. Introduction

In recent years, petroleum prices have reached the record high once again. Several reasons cause that phenomenon; one is the expectation of depletion of petroleum, and the other is overdependence on fossil fuel of the existing vehicle industries. According to the latest petrolic consumption statistics conducted by the International Energy Agency (IEA) [1], over 57% petroleum of the world had been consumed by the transportation system. Modern motor vehicles powered by internal-combustion engine give most contribution to these figures. The IEA also forecasts that the figure would show an increase of 3% each year. The usage of internal-combustion engine vehicles was indeed a great evolution for the transportation system in transportation history, which brought many societies a wealth of convenience. However, such closely linked and inseparable relationship of internal-combustion engine vehicles with humans also expedites the depletion of petroleum, where energy crisis is conceivable in the coming years.

Manuscript received April 9, 2010; revised October 22, 2010; accepted January 6, 2011. Date of publication April 7, 2011; date of current version March 30, 2012.

The authors are with the Department of Computer Science and Information Engineering, National Chiayi University, Chiayi 60004, Taiwan (e-mail: rchsu@mail.ncyu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2011.2141092

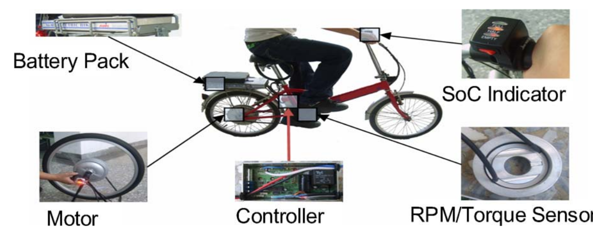


Fig. 1. Prototype of the pedelec and its key components.

Light electric vehicles (LEVs), such as electric bicycle, electric scooter, and electric mobility [2]–[4], are so-called green vehicles for city commuters. LEVs are normally powered by rechargeable battery, and the driving performance of the LEV is constrained by battery capacity, motor power, road types, operation weight, electric control, and particularly the management of assisted power [5]–[8]. In this paper, our focus for the LEV is in electric bicycles. There are two kinds of electric bicycle. One is a pure electric bike [9]–[11], i.e., e-bike, which integrates electric motor into bicycle frame or wheels, and is driven by motor force only using a handlebar throttle. The other is a power-assisted bicycle or called the pedelec [8] hereafter, which is a human–electric hybrid bicycle [12] that supports the rider with electric power only when the rider is pedaling. Fig. 1 shows a prototype of the pedelec, where a battery-powered in-wheel motor in the rear wheel, a battery pack, a torque sensor, a state-of-charge (SoC) indicator, and a controller with assisted power management unit are integrated to the traditional bicycle. When the pedelec rider pedals, the controller will be informed by the torque sensor and request the in-wheel motor to generate the assisted power, which is decided by the assisted power management unit, to support the rider in moving the pedelec, and the battery's SoC will be exhibited on the SoC indicator to show the rider the remaining capacity of the battery.

The motor force determined by the assisted power management unit plays a crucial role in ensuring the comfort and the safety of pedelec riding. Without enough motor-assisted force, the environment forces acting on the pedelec cannot be overcome using the rider's normal pedal force only such that the rider must exercise more energy to keep the pedelec moving under required comfortable speed. However, with too much and abrupt motor-assisted force, the rider might suffer from instantaneous forward propulsion in causing instability of controlling the pedelec, possibly endangering the rider's safety. Thus, the

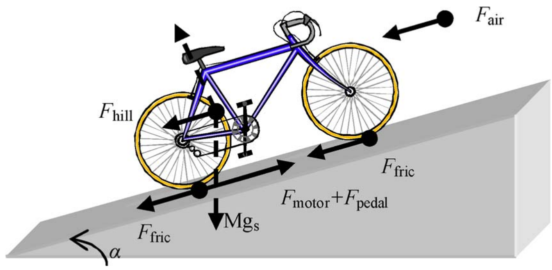


Fig. 2. Resisting environment forces acted on a moving bicycle in a slope.

problem of quality-of-riding (QoR) provisioning in a dynamic pedelec riding environment is how to satisfy the comfort and safety criteria prior to maximize the energy utilization of the battery, i.e., the source of the assisted power, and guarantee QoR constraints [13]. Recently, novel dynamic modelings have been employed in solving the QoR problem for moving vehicles [13], [14]. Algorithms of soft computing were proposed to solve the comfort of ride and other quality-required problems of the vehicle [15]–[17] as well.

The use of reinforcement learning (RL) in providing assisted power with comfort of ride for the LEV is original and is one of our contributions [18]. Unlike supervised learning, whose aim is to learn a mapping from the input to the output with correct values provided by a supervisor, the RL generates a sequence of correct actions to reach the goal. In the situation of pedelec riding, the rider's pedal force varies with different road types and the rider's physical conditions in a nondeterministic manner such that a correct input/output mapping for electric-assisted power is hard to provide by the supervisor. On the contrary, the agent of the RL takes a sequence of action, i.e., different levels of assisted power, in the nondeterministic environment of pedelec riding and receives rewards for its actions in trying to solve the comfort-of-ride problem. After a set of trial-and-error steps, the RL agent learns the best sequence of action that achieves the comfort of ride for rider. In this paper, we further enhance the RL-based assisted power management (RLAPM) to solve the QoR provisioning problem for the pedelec, aiming to satisfy the predetermined QoR constraints and maximize the energy utilization of the battery. The proposed RLAPM and RLAPM enhancement (RLAPME) can adaptively provide proper assisted force at accurate timing such that QoR criteria can be satisfied. By comparing with other existing power-assisted methods for the pedelec, the proposed RLAPM and RLAPME not only perform better in guaranteeing QoR criteria but also achieve better energy utilization.

II. QoR Provisioning in Pedelec Riding

A. QoR Provisioning Criteria for the Pedelec

There are three resisting forces to be overcome, i.e., air drag (F_{air}), friction (F_{fric}), and hill drag (F_{hill}) forces [14], [19]–[21], for a biker to move the vehicle uphill at certain velocity v, as depicted in Fig. 2.

The total required input power P_{req} in moving a bicycle uphill at expected speed V_{exp} is formulated as

$$P_{\text{req}}(V_{\text{exp}}) = (F_{\text{drag}} + F_{\text{fric}} + F_{\text{hill}})V_{\text{exp}}. \tag{1}$$

F_{air}, F_{fric}, and F_{hill} can be obtained by the following:

$$F_{\text{air}} = 1/2(C_d D_a A_f v^2) \tag{2}$$

$$F_{\text{fric}} = M g_s C_r \tag{3}$$

$$F_{\text{hill}} = \sin(\alpha) M g_s. \tag{4}$$

In (2), C_d, D_a, and A_f are the drag coefficient of atmosphere, the atmospheric density, and the frontal area, respectively. In (3), C_r is the rolling coefficient of the bicycle's rubber tire, g_s is the standard gravity, and M is the gross weight of the rider and the vehicle. In (4), α is the slope angle of the hill. To keep the bicycle moving uphill with V_{exp}, the required pedaling power from the rider, i.e., P_{pedal}, must not be less than the total power of the required expected power P_{exp} plus the required power in overcoming the environment change, i.e., ΔP_{env}, as follows:

$$P_{\text{pedal}} \geq P_{\text{exp}} + \Delta P_{\text{env}}. \tag{5}$$

If the rider is riding a pedelec, the left side of (5) will be the power provided by the motor, i.e., P_{motor}, plus the rider's pedal power as follows:

$$P_{\text{pedal}} + P_{\text{motor}} \geq P_{\text{exp}} + \Delta P_{\text{env}}. \tag{6}$$

The power provided by the motor, together with the rider's pedal power, will drive the wheel of the pedelec through the transmission system to cause a vehicle speed V as

$$V = (P_{\text{pedal}} + P_{\text{motor}})/(F_{\text{drag}} + F_{\text{fric}} + F_{\text{hill}} + \Delta F) \tag{7}$$

where V should be as closer to V_{exp} as possible in considering for the comfort of ride and ΔF is the force due to the energy lost in the transmission system and in the electricity transfer, and other influential factors.

In (6), P_{motor} is determined by the assisted power management unit of the pedelec. If the rider's pedal power remains constant, the vehicle velocity of the pedelec will be solely determined by P_{motor}, as in (6). However, the rider's pedal power generally varies in responding to the road conditions and the rider's physical condition, and hence, how to adaptively and autonomously adjust the assisted power for the motor in satisfying the rider's demand becomes a problem to be solved. With appropriate assisted power, constant vehicle velocity can be maintained, and the rider will feel comfortable during the ride. On the contrary, the rider will feel that the pedelec is hard to control and is unsafe if the assisted power abruptly comes out. Hence, the problem of QoR provisioning in a nondeterministic pedelec riding environment is how to manage the dispatched assisted power in satisfying both comfort and safety criteria, i.e., guarantee QoR constraints, prior to maximize the energy utilization of the battery, i.e., the source of the assisted power.

To provide comfort in riding the pedelec, the rider of the pedelec will expect to maintain the vehicle under a preferred velocity, namely, v_{comf}, at a certain road area when the assisted power is provided in collaborating with the rider's pedal power. However, as it is impractical to keep the velocity precisely at a specific speed, the best one could do is to maintain the vehicle within tolerance to the preferred comfortable velocity. Although the riding environment of the pedelec can be very unpredictable, different levels of preferred velocity and tolerance with respect to different road conditions can be defined to satisfy the comfort criteria. Hence, one of the QoR constraints of the assisted power for the pedelec is to define a comfortable zone (CZ) for the vehicle velocity such that, when the vehicle velocity falls within the CZ, the comfort criteria is considered achieved.

In addition, although the velocity can be adaptively maintained within the CZ, it is still desirable to define another QoR constraint concerning the safety criteria. Stability in controlling the pedelec is often considered as the most important factor to the safety of the rider, while the instantaneous acceleration generated by the assisted power directly affect the stability of the pedelec control. With abrupt assisted power, the instantaneous acceleration generated will be too large or too small to cause instability in controlling the pedelec such that the rider will feel unsafe in riding. Similar to the CZ of the velocity in satisfying the comfort criteria, a safety zone (SZ) can be also defined for the safety concern. In order to guarantee the safety criterion in pedelec riding, the instantaneous acceleration generated by the assisted power should be kept within the SZ. Therefore, the QoR provisioning power-assisted management should comply with both the CZ and SZ constraints and maximize the energy utilization of the battery, i.e., the source of the assisted power.

B. Assisted Power Methods for the Pedelec

There are two kinds of conventional assisted power methods for the commercial off-the-shelf pedelec, which are the constant-assisted power (CAP) [8] and proportion-assisted power (PAP) methods [22]. In the CAP method, when the power assistance criteria are met during the riding, the rider is assisted by the motor with the predetermined constant power. The criteria of activating the CAP method are that a predefined paddle torque is sensed, which implies that the rider needs assisted power to move the pedelec forward, and that the vehicle velocity is lower than the velocity regulated by law, i.e., $v_{\max}$, which is 24 or 25 km/h in most countries. The CAP method is simple to implement; however, the same predefined assisted power is provided without considering the riding environment and the rider's physical condition such that the comfort and the safety of the rider might not be guaranteed.

The PAP method is proposed by U.K.'s cyclists' organization and Japan's industry [22] and has been adopted by regulation. The assisted power ratio of the PAP, denoted as P_R, is formulated in the following:

$$P_R = \begin{cases} 1, & v \leq v' \\ 1-\eta(v-v'), & v' < v < v_{\max} \\ 0, & v \geq v_{\max}. \end{cases} \tag{8}$$

While the vehicle velocity v is lower than a predefined velocity v', the assisted power provides the same power as the rider's power; otherwise, P_R gradually decreases according to η, where η is used to attenuate the assisted power ratio and is defined as the inverse of $v_{\max}$ minus v'. Thus, the motor assisted power is P_R multiplied by the rider's power. In some countries, $v_{\max}$, v', and η are set 24 and 15 km/h, and 1/9,

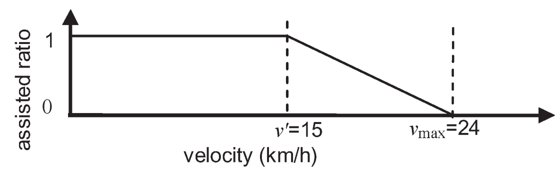


Fig. 3. Diagram of assisted power ratio of the PAP method.

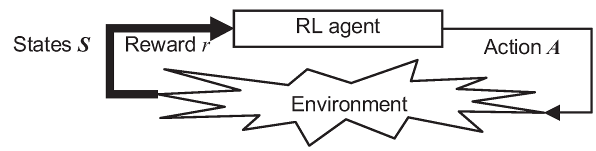


Fig. 4. Diagram of the RL agent's interaction with the environment.

respectively, prior to manufacturing for the PAP method. The PAP assisted ratio versus vehicle velocity is shown in Fig. 3, where the assisted power will be linearly decreased when the vehicle velocity is over 15 km/h to avoid overassistance. The power-assisted rule of the PAP provides proportional power in a piecewise and linearly decreased fashion without considering the rider's output power, which might cause abrupt instantaneous acceleration in case the rider's pedal power is large such that safety concern might occur.

Soft computing method can also provide solution in satisfying the QoR constraint for the pedelec. The delta rule or the delta learning rule (DLR) [23]–[26] is widely used for the weighted value updating in a neural network. In the area of power control and management for the pedelec, the rider's pedaling power can be predicted using the delta rule, and the required assisted motor power can be thus calculated from the predicted pedal power under the preferred comfortable speed. The simulation result of the RLAPM with QoR provision for the pedelec will be compared with that of the DLR-based assisted power method as well.

III. Proposed RLAPM for the Pedelec

A. RL

The RL is a heuristic learning method [27], [28], which has been applied in many different areas, such as agent system, robot control [29], [30], power management [31], and image processing [32]. In the RL, there is a decision-making agent, which observes the environment states $\boldsymbol{S}$, takes actions $\boldsymbol{A}$, and receives reward r for its actions in trying to solve a problem in the environment it situates, as shown in Fig. 4.

After certain trial-and-error steps, the RL agent gradually learns the best policy, which is the sequence of actions that maximize the total reward [27]. For the unknown environment, the RL agent will repeatedly count the reward during the early learning phase, and the optimized policy would be obtained when learned. In the beginning of the learning process, the RL agent is unable to make a "for sure" decision, the exploration strategy is hence required for retrieving the sufficient reward information on each state, and the exploitation is then adopted in deciding the action with higher reward and higher probability on certain state after the learning. In such a case, *softmax* function is utilized for the strategy of exploration–exploitation

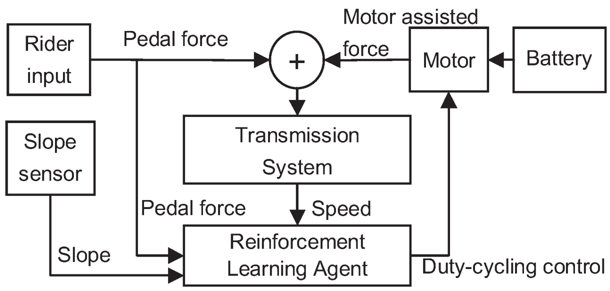


Fig. 5. System architecture of the RLAPM for the pedelec.

```
1 Initialize all Q(s,a) to random number [-1,1]
2 For every round
3  Initialize state vector s_i
4  Repeat pedaling i
5     Choose action a_i from Q (ε-greedy, etc.)
6     Provide assisted power
7     Observe next reward r_{i+1} and next state s_{i+1}
8     Update Q(s_i,a_i):
          Q(s_i,a_i) = (1−η)Q(s_i,a_i) + η(r_{i+1} + γ max_{a_{i+1}} Q(s_{i+1},a_{i+1}))
9     s_i = s_{i+1}
10 Until i = end_of_round
```

Fig. 6. Pseudocode of the RLAPM for the pedelec.

in most applications. In case of a nondeterministic environment or a state transition model, where the transition probability is unknown or unsure, a certain action for the next state cannot be precisely decided by observing the current state. In such a case, the Q-learning algorithm [28] could be utilized in recording the accumulative reward and deciding the best policy. In Q-learning, the accumulative reward $Q(s, a)$ is a function of state s and action a, and the agent iteratively updates the Q values of $Q(s, a)$ with the following updating equation:

$$Q(s_t, a_t) = (1 - \eta)Q(s_t, a_t) + \eta\left[r_{t+1} + \gamma \max_{\forall a_{t+1} \in A} Q(s_{t+1}, a_{t+1})\right] \quad (9)$$

where $Q(s_t, a_t)$ is the accumulative reward at state s, action a is taken at time t, and parameters η and γ are the learning and discount rates, respectively, for the Q-learning with value between 0 and 1. r_{t+1} is a reward value that is obtained by taking action a_t and then making transition from state s_t to state s_{t+1}. Normally, the Q values are stored in a table, namely, the Q-table, as the reference for the agent to determine the next action.

B. System Architecture and the RLAPM Algorithm

The system architecture of the proposed RLAPM for the pedelec is shown in Fig. 5. In the heart of the architecture is an RL agent, which learns the control sequences to provide assisted power in changing the vehicle velocity and improving the comfort of ride by utilizing the rewarding policy and the exploration procedure. The RL agent observes the environment of the pedelec riding, and the observed data is arranged in terms of the state vector, while the necessary action will be then decided by the agent to drive pulsewidth-modulated electric motor [33] for a determined duty cycle. The duty cycle is defined as the percentage of the full motor output power.

In realizing the RLAPM for the pedelec, the state vector, the action, and the reward policy should be first defined. In this paper, each round is defined as 120 paddling cycles, and the state vector consists of three elements, which are the vehicle velocity, the slope grade, and the paddle force. These signals are sensed, quantized into states, and then observed by the RL agent to construct the state vector. The action is defined as different levels of assisted power decided by the RL agent and delivered by the electric motor. According to the designed rewarding policy, the agent can determine the best action from the iterative accumulated reward value. The main purpose of the RLAPM is to satisfy the QoR constraints for the pedelec's rider; hence, a comfortable velocity v_{comf} on riding and the tolerance of deviation μ to the predetermined comfortable velocity should be first defined and arranged into the rewarding policy, which is formulated as follows:

$$r = \begin{cases} \frac{k_1 \mu}{|v - v_{\mathrm{comf}}|}, & \text{if } |v - v_{\mathrm{comf}}| \le \mu \\ -k_2 |v - v_{\mathrm{comf}}|, & \text{otherwise} \end{cases} \quad (10)$$

where v is the pedelec's velocity and k_1 and k_2 are positive weighting factors. In (10), when the pedelec's velocity is within the tolerance range of the comfortable velocity, i.e., within the CZ, a positive reward in proportional to the inverse of the velocity deviation multiplied by k_1 will be given to the agent. Within the CZ, the closer the pedelec's velocity is to the predetermined comfortable velocity, the larger the reward will be given. Outside the CZ, a negative reward, i.e., punishment, in proportional to the deviation multiplied by a positive constant k_2 will be given to the agent. The agent of the RLAPM will use the reward policy and the Q-table to update the reward value for the learning and the action decision as in (9).

The pseudocode for the RLAPM is shown in Fig. 6. The first line of the pseudocode initializes all of the system parameters such as state variables and Q values of the Q-table, which are randomly set between 1 and -1. Each round starts in line 2, and the round's loop goes from lines 3 to 10. The main learning loop starts at line 4 and goes from lines 5 to 10, which represents learning during each paddle iteration or step. In line 5, the agent chooses an appropriate action from its policy of assisted power according to the Q value at the current state, and the assisted power is executed in the corresponding paddle in line 6. The agent then observes the environment's parameters, i.e., the next state vector and reward as in line 7. The iterative update formula of the Q-table is written as in line 8, where parameter η gradually decreases according to the number of the observed state. After the Q values are updated, the next state replaces the current state (line 9). Once the defined steps within a round are completed, as in line 10, another round comes up, and the learning resumes.

IV. Experimental Results

Experiments of pedelec riding in an urban area utilizing the RLAPM are conducted by simulation on computer, and the results are compared with existing assisted power methods for the pedelec.

TABLE I
SIMULATION SCENARIO OF DIFFERENT ROAD TYPES FOR PEDELEC RIDING WITHIN 15-km DISTANCE

Slope grade / Parameters	0%	1%	2%	3%
Percentage of slope level for urban area	70%	15%	10%	5%
Pedal Force, Gaussian distribution in Watt	(45, 30)	(65, 40)	(95, 50)	(125, 60)

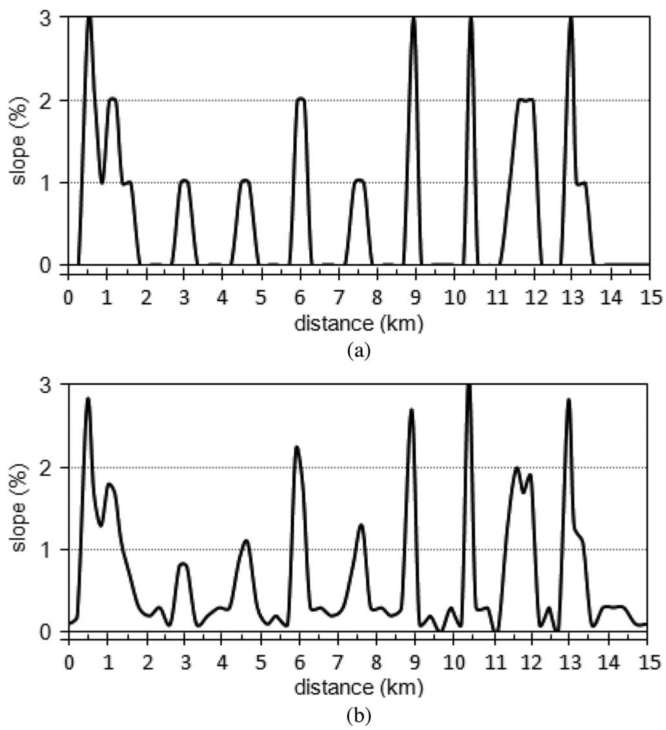


Fig. 7. (a) Designed slope profile for urban riding of 15 km. (b) Designed slope profile of (a) adding with zero-mean 0.2% standard deviation of white Gaussian noise.

A. Experiment Configuration

To conduct the experiment, a simulation scenario of different road types and pedal forces delivered by the pedelec rider are designed first, as shown in Table I. The first row shows the designed slope grades, representing elevation rising per 100-m run in percentage, in modeling the road conditions of pedelec riding [5]. For example, 1% of the slope means that the elevation rises 1 m for every 100-m run. The slope grade is defined among 0%, 1%, 2%, and 3% and added with zero-mean and 0.2% standard deviation white Gaussian noise to simulate the real slope variation in the urban area. The combination of different slope grades for riding simulation in the urban area is shown in the second row of Table I. The rider's pedaling force in responding to the different slope grades is modeled as Gaussian distribution, as shown in the last row of Table I. For example, for the 0% slope grade in the second column of Table I, the mean pedaling force is 45 W with a variance of 30 W.

The slope profile for testing the RLAPM for urban riding within 15 km is shown in Fig. 7. Fig. 7(a) and (b) shows the designed slope profile and the designed slope profile added with zero-mean 0.2% standard deviation of white Gaussian noise, respectively.

TABLE II
PEDELEC'S SPECIFICATIONS AND ENVIRONMENTAL PARAMETERS FOR SIMULATION

Pedelec specification	Driving type: Rear wheel directly-driven with constant gear ratio Motor type: hub brushless DC motor Motor power: 24V/250W Battery pack: Lithium-Ion 24V, 12Ah Wheel type: 16" wheels The total weight of pedelec and the rider: 90kg
Environmental parameters	Drag coefficient: $C_d = 0.5$ Density of air: $D_a = 1.18\text{kg/m}^3$ Rolling coefficient: $C_r = 0.014$ Frontal Area: $A_f = 1\ \text{m}^2$

The pedelec's specifications and environmental parameters are listed in Table II. The actions of the RLAPM agent, i.e., levels of assisted power delivered, are 0%, 25%, 50%, 75%, and 100%, in terms of duty cycle.

B. QoR Criteria

To compare the performance of the RLAPM with existing methods, the vehicle velocity, the instantaneous acceleration, and the accumulative energy consumption are measured and recorded. Qualitative comparison of the achievability of comfortable riding (ACR) and safety index (SI), as criteria for the QoR, which are derived from the vehicle velocity and the instantaneous acceleration, respectively, are evaluated. The ACR is obtained by taking the ratio between the numbers of measured velocity within the CZ and the total number of measured velocity during the riding distance of 15 km. The higher the ACR means the better the performance of comfortable riding provided by the power-assisted method. The instantaneous acceleration generated from the motor-assisted power is a good measurement of stability and safety in controlling a pedelec. Less numbers of the measured instantaneous acceleration falling outside the SZ represents better stability the motor-assisted power provides. Hence, the SI is defined and calculated by the inverse of the number of the measured instantaneous acceleration out of the SZ. The larger the SI, the more stable the power-assisted method is, and the safer the rider will be. How the energy is utilized for the power-assisted methods is evaluated by the accumulative energy consumption of the 15-km pedelec riding in this experiment.

C. Preliminary Experiment of the RLAPM for Urban Riding

For validating the effectiveness of the RLAPM, the experiment of the RLAPM for urban riding is first performed. For urban riding, v_{comf} and μ in the reward function are set to 18 and 2 km/h, respectively, which mean that the CZ is between 16 and 20 km/h. k_1 and k_2, i.e., the weighting factors of the reward function, are both set to 1 for simplicity purposes. The SZ of $[-0.4, 0.4]$ m/s^2 is defined for the instantaneous acceleration. Experiment results of the RLAPM within the first 7.5 km are illustrated in Fig. 8. By examining the vehicle velocity in Fig. 8(b) with respect to the slope profile of Fig. 8(a), one can see that that, after 2.5 km, the RLAPM agent learned all the slope grades and respond with the appropriate actions, i.e., the

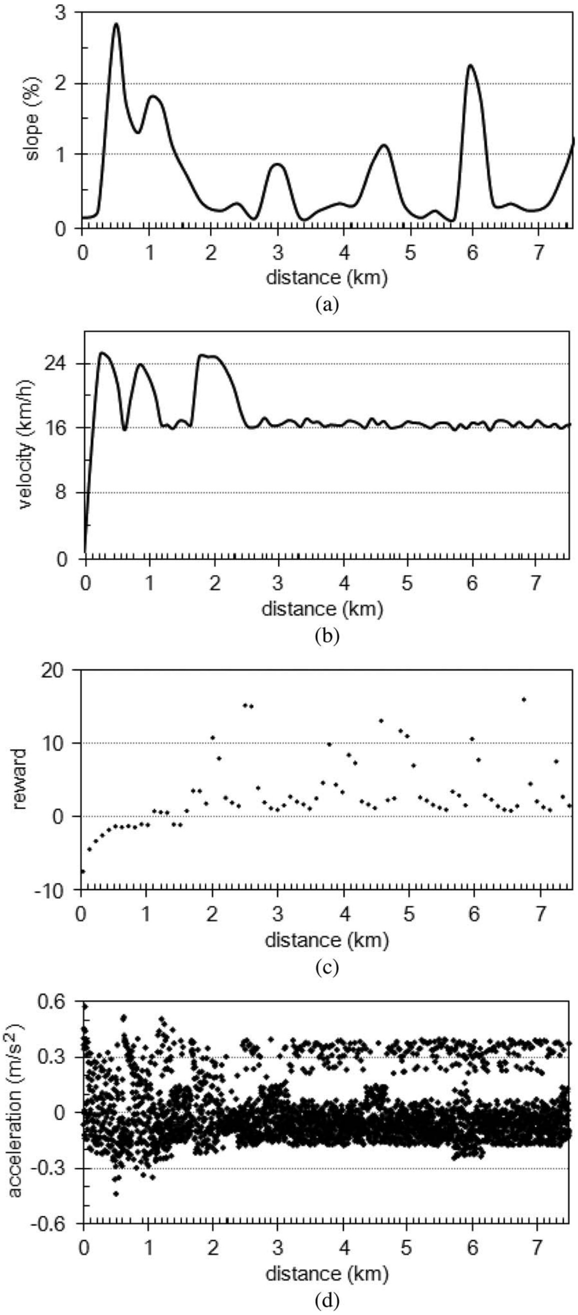


Fig. 8. Experimental results of the RLAPM for urban riding within 7.5 km: (a) slope profile; (b) vehicle velocity; (c) reward values; and (d) instantaneous acceleration.

delivered assisted power, such that most of the vehicle velocities are maintained within the CZ. The trace of the reward value, as in Fig. 8(c), shows that the reward values obtained by the RL agent are most negative before 2 km, indicating that the RLAPM forced the agent to select the appropriate assisted power by punishment in that period. While the reward values obtained after 2 km are most positive, expressing that, by offering a positive reward, the agent interactively and gradually adapted to the environment in selecting the right actions from the learned experiences to maintain the vehicle velocity within the CZ. Trending of instantaneous acceleration is shown in Fig. 8(d), where the instantaneous acceleration is sparsely distributed between -0.4 and 0.4 m/s^2, i.e., within the SZ before 2 km, due to the exploration and the learning phase of the RLAPM; however, after learning, i.e., after 2 km, the instantaneous acceleration is gradually concentrated around the region of $[-0.2, 0.2]$ m/s^2 with only a few samples failing outside 0.4 m/s^2.

D. Experiment of the RLAPME

In the preliminary experiment of the RLAPM for urban riding, after learning, the RLAPM can gradually adapt to varying riding environment and achieve the QoR. However, in the early learning phase before the first 2.5 km, the RLAPM is under the transition state providing with varying assisted power that might cause instability in controlling the pedelec. According to the RLAPM algorithm, the values of the Q-table are initialized with random values within $[-1.0, 1.0]$, which initially means no experience, and information in responding to the decisive reward is provided in the Q-table for the agent such that the agent has to gradually learn by interacting with the environment, i.e., exploration, to encode the learned experience and update the Q-table. Hence, to reach the stable state fast and guarantee the QoR criteria to the pedelec rider, the Q-table of the RLAPM should be initialized with prior learned experience in the beginning of the ride. Herein, a solution for encoding the prior learned experience to the initial Q-table is proposed by executing the RLAPM for certain times within a defined riding distance for a specific riding region, such as the urban area, and the learned experience of the RLAPM, represented as Q values in the Q-table, from the prior rides are accumulated and incorporated into a learned Q-table, i.e., $\boldsymbol{Q}'$, as

$$Q'(\mathbf{S}, \mathbf{A}) = \frac{\sum_{i=1}^{N} Q_i(\mathbf{S}, \mathbf{A})}{N} \tag{11}$$

where N is the total number of test ridings and $\boldsymbol{Q_i}(\boldsymbol{S}, \boldsymbol{A})$ is the resulting Q-table of the ith test riding. The RLAPM initialized with the learned Q-table, i.e., RLAPME. The experiment of the RLAPME for the pedelec riding in the urban area is performed, and the results are compared with the RLAPM, as shown in Fig. 9.

Fig. 9(a) shows the slope profile in the first 4 km. The vehicle velocity of the RLAPM and the RLAPME are shown in Fig. 9(b), which obviously indicates that the RLAPME (in dark black) achieves the convergence fast and enhances the stability in the first 4 km. Fig. 9(c) and (d) exhibits the instantaneous acceleration of the RLAPM and RLAPME, respectively, where the instantaneous acceleration of the RLAPME concentrates in the narrower SZ range of $[0.2, -0.2]$ m/s^2. In the following, the simulation results using the RLAPME will be compared with other existing assisted power methods.

E. Experimental Result Comparison for Urban Riding

The simulation of pedelec riding in the urban area is first performed, where the definition of the slope variation is as in Table I. For comparison purposes, the CAP, PAP, and DLR-based assisted power methods are also conducted for the pedelec riding simulation. The assisted power of the CAP is set as 25% of the maximum power, i.e., 62.5 W, and the threshold

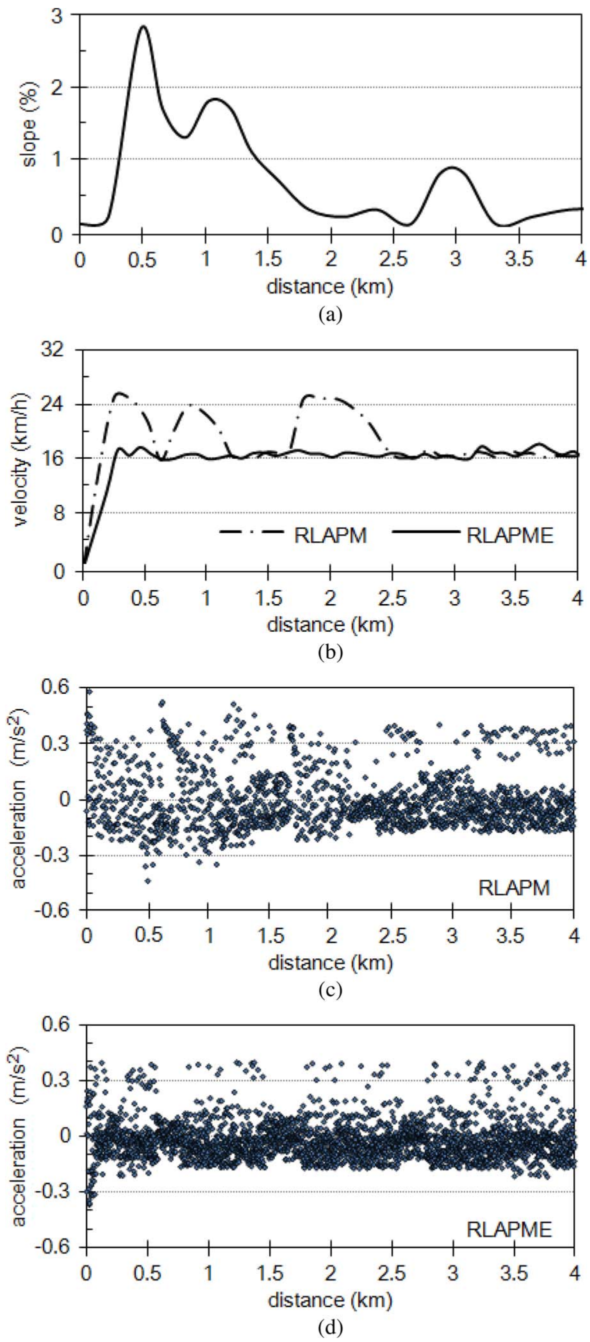


Fig. 9. Performance enhancement of the RLAPME compared with the RLAPM: (a) slope profile; (b) vehicle velocity comparison; and instantaneous acceleration trace of (c) the RLAPM and (d) the RLAPME.

of the pedal power is 20 W, i.e., whenever the rider's pedal power is sensed to exceed 20 W, a 62.5-W power will be provided by the CAP in assisting the rider. The assisted power of the PAP follows (8). The DLR-based method will predict the next motor assisted power, which is the total expected power subtracted by the predicted pedal's power, which is required to achieve the expected vehicle velocity based on the delta (difference) of the previous assisted power and the current estimate. For urban riding, the CZ of the velocity is between 16 and 20 km/h, and the SZ of the instantaneous acceleration is within $[-0.4, 0.4]$ m/s^2. According to the road regulation in most countries, the assisted power for the pedelec will not be

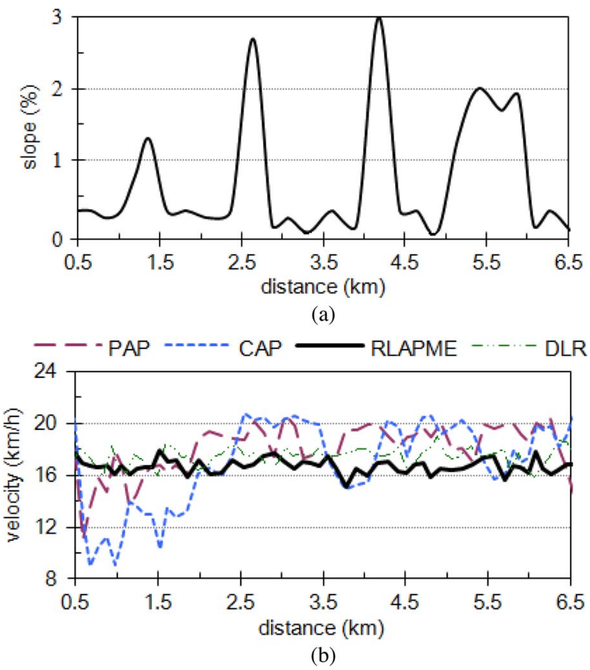


Fig. 10. Simulation of urban riding. (a) Slope profile. (b) Vehicle velocity records.

provided when the vehicle velocity exceeds 25 km/h. Fig. 10 shows the slope profile and the velocity records of the compared methods within the first 6-km range of the 15-km urban riding simulation.

Velocity traces of the RLAPME, DLR, PAP, and CAP methods are exhibited using the black solid, green double-dot-dash, purple long-dash, and blue short-dash lines, respectively. By examining Fig. 10(b), the velocity traces of the CAP and the PAP are fluctuating, and most of the velocity samplings fall out of the range of the CZ in the 6-km range. The velocity variation of the DLR is better than that of the CAP and the PAP; however, the velocity variations are large in some ranges such as 1.8$\sim$1.9, 2.5$\sim$2.7, and 5.6$\sim$5.7 km. The velocity trace of the RLAPME shows that the stable vehicle velocity within the CZ is maintained all the way, starting from the beginning until the end of the ride within the riding range of 15 km. It evidences that the RL agent of the RLAPME is able to exercise the appropriate assisted power for supporting the rider the best, satisfying the criteria of the QoR. Traces of the instantaneous acceleration of the pedelec for the urban riding test using the CAP, PAP, DLR, and RLAPME methods are shown in Fig. 11(a)–(d), respectively, for the 15-km riding distance.

By examining Fig. 11 with regard to the slope variation in Fig. 10(a), one can see that the instantaneous accelerations varied the most whenever the slope changes, which might cause instability and discomfort to the rider of the pedelec. Among the existing power assisted methods, the PAP method in Fig. 11(b) has the worse instantaneous acceleration variation spreading within $[-0.4, 0.4]$ m/s^2, which can be understood since the PAP provides proportional power in a piecewise and linearly decreased fashion without considering the pedal's output power. On the other hand, the assisted power provided by the CAP is constant such that the variation in instantaneous acceleration is smaller than the PAP. The trace of the instantaneous

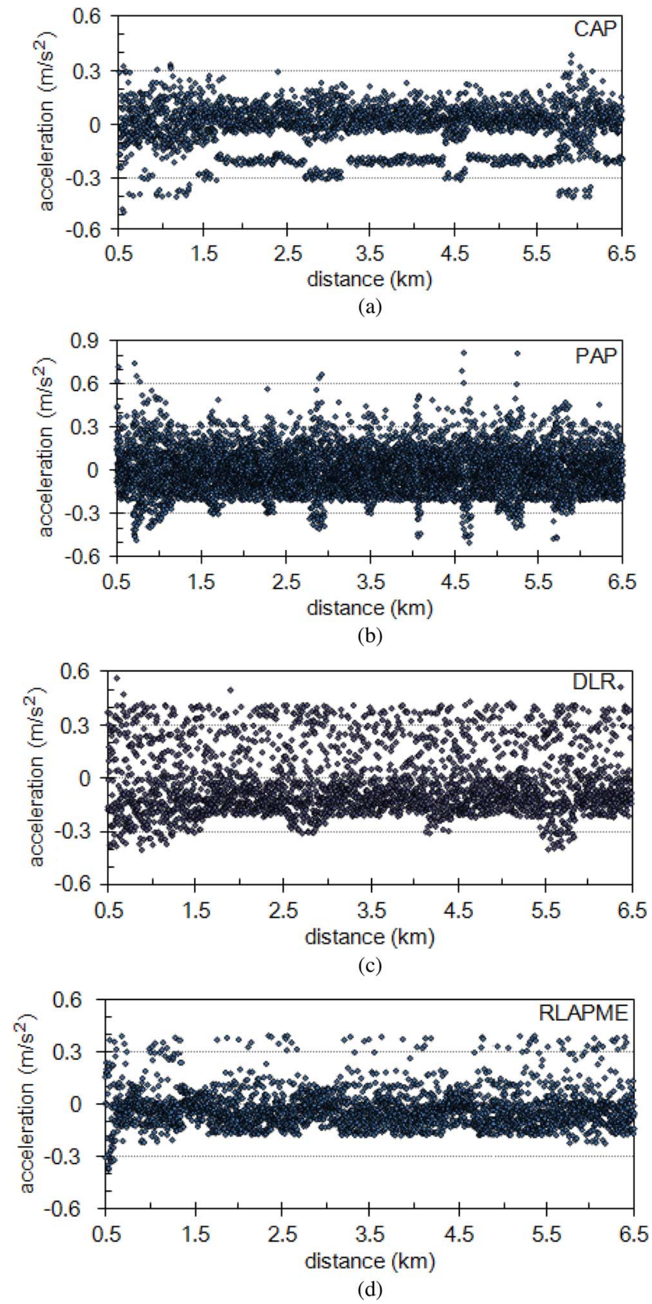


Fig. 11. Trace of the instantaneous acceleration of the pedelec for urban riding using (a) CAP, (b) PAP, (c) DLR-based method, and (d) RLAPME for the first 6-km riding distance.

acceleration of the DLR-based method exhibits similar variation as the PAP method. The trace of the instantaneous acceleration of the pedelec using the RLAPME, as shown in Fig. 11(d), demonstrates that the stable instantaneous acceleration concentrated within the range of $[-0.2, 0.2]$ m/s^2 is achieved all the way, starting from the beginning until the end of the ride, which means that the safety and the stability of the pedelec are maintained by the RLAPME, as compared with other assisted power methods.

To understand how the energy is utilized in the simulations, the accumulative energy consumption is recorded and shown in Fig. 12, where the accumulative energy consumption of the RLAPME, CAP, PAP, and DLR methods are represented in black dot, blue dash, red dot, and green dash lines, respectively. Among the compared methods, the RLAPME consumes the least energy, which is 1627 mAh, as compared with the accumulative energy consumed by the PAP, CAP, and DLR-based methods, i.e., 3289, 4138, and 4141 mAh, respectively.

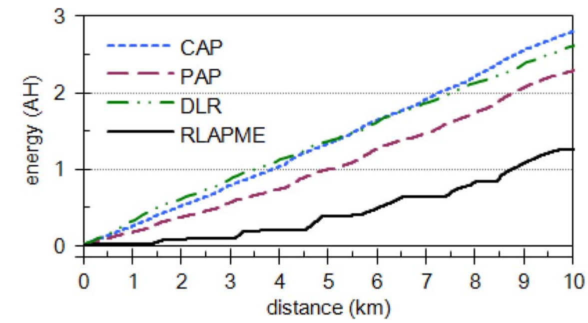


Fig. 12. Accumulative energy consumption during the 15-km test riding in the urban area.

TABLE III
QUALITATIVE COMPARISON OF THE PERFORMANCE FOR URBAN RIDING TEST

Parameters (Unit) / Methods	Achievability of Comfortable Riding (numbers of velocity within CZ/total pedaling numbers)	Safety index (1/numbers of instantaneous acceleration falls out of SZ)	Accumulative energy consumption (mAH)
RLAPME	6163/6482 = 96%	1/6 =0.1667	1627
PAP	4757/6572 = 72%	1/92=0.0108	3289
CAP	4915/6338 = 78%	1/17=0.0588	4138
DLR	5002/5786 = 86%	1/43=0.0232	4141

The qualitative comparison of the QoR, using the ACR and the SI, respectively, and the accumulative energy consumption for the compared methods are listed in Table III.

The ACR comparison (first data column of Table III) shows that the RLAPME has the best ACR of 96%, as compared with the PAP, CAP, and DLR-based methods, which are 72%, 78%, and 86%, respectively, during the 15-km riding distance. The SI values, shown in the second data column of Table III, for the RLAPME, PAP, CAP, and DLR-based methods are 1/6, 1/92, 1/17, and 1/43, respectively. The qualitative comparison of the comfort and the safety of riding, using the ACR and SI, respectively, again confirms that the RLAPME provides better QoR in comfortable riding and in safety for the pedelec rider among other existing power-assisted methods during the 15-km riding distance in the urban area and, at the same time, maximizes the energy utilization of the battery by consuming least energy, as indicated in the last column of Table III. Experiment results on urban riding demonstrate that the ACR of the RLAPME is improved by 24%, while the energy utilization is improved by 50% by comparing with the PAP method for the pedelec.

V. CONCLUSION

In this paper, the RLAPM has been proposed for the pedelec. The riding environment and characteristics of the pedelec has been first modeled and defined as system states, and the Q-learning has been employed for the assisted power management. Experimental results confirmed that the RLAPM and its enhancement, i.e., RLAPME, satisfies the requirement of the QoR and achieves better energy utilization by comparing with other existing conventional and DLR-based assisted power methods. The contributions of the RLAPM can be summarized as follows.

1) It dispatches appropriate assisted power at the right timing for safe and comfortable pedelec riding.
2) It provides better energy utilization of the battery of the pedelec.
3) It provides assisted power without any interference to the rider's operation of the pedelec.

With appropriate modeling of the driving environment and the vehicle system and dynamics, the proposed RLAPM could be possibly extended to the dynamic power management for electrical vehicle (EV) or hybrid EV to satisfy the demand of driver, reach the zero emission of CO_2, and maximize the energy utilization at the same time.

References

[1] T. Onoda and T. Gueret, "Fuel efficient road vehicle non-engine components: Potential saving and policy recommendations," *Int. Energy Agency Inf. Paper*, pp. 1–24, Oct. 2007.

[2] H. Seki, K. Ishihara, and S. Tadakuma, "Novel regenerative braking control of electric power-assisted wheelchair for safety downhill road driving," *IEEE Trans. Ind. Electron.*, vol. 56, no. 5, pp. 1393–1400, May 2009.

[3] C. C. Tsai, H. C. Huang, and S. C. Lin, "Adaptive neural network control of a self-balancing two-wheeled scooter," *IEEE Trans. Ind. Electron.*, vol. 57, no. 4, pp. 1420–1428, Apr. 2010.

[4] A. Emadi, Y. J. Lee, and K. Rajashekara, "Power electronics and motor drives in electric, hybrid electric, and plug-in hybrid electric vehicles," *IEEE Trans. Ind. Electron.*, vol. 55, no. 6, pp. 2237–2245, Jun. 2008.

[5] A. Muetze and Y. C. Tan, "Electric bicycles: A performance evaluation," *IEEE Ind. Appl. Mag.*, vol. 13, no. 4, pp. 12–21, Jul./Aug. 2007.

[6] Y. Tanaka and T. Murakami, "A study on straight-line tracking and posture control in electric bicycle," *IEEE Trans. Ind. Electron.*, vol. 56, no. 1, pp. 159–168, Jan. 2009.

[7] M. Defoort and T. Murakami, "Sliding-mode control scheme for an intelligent bicycle," *IEEE Trans. Ind. Electron.*, vol. 56, no. 9, pp. 3357–3368, Sep. 2009.

[8] E. A. Lomonova, A. J. A. Vandenput, J. Rubacek, B. d'Herripon, and G. Roovers, "Development of an improved electrically assisted bicycle," in *Conf. Rec. IEEE IAS Annu. Meeting*, 2002, vol. 1, pp. 384–389.

[9] P. Fairley, "China's cyclists take charge: Electric bicycles are selling by the millions despite efforts to ban them," *IEEE Spectr.*, vol. 42, no. 6, pp. 54–59, Jun. 2005.

[10] N. Somchaiwong and W. Ponglangka, "Regenerative power control for electric bicycle," in *Proc. IEEE Int. Joint Conf. SICE-ICASE*, 2006, pp. 4362–4365.

[11] M. J. Yang, H. L. Jhou, B. Y. Ma, and K. K. Shyu, "A cost-effective method of electric brake with energy regeneration for electric vehicles," *IEEE Trans. Ind. Electron.*, vol. 56, no. 6, pp. 2203–2212, Jun. 2009.

[12] D. Schneider, "Easy rider: Convert your bicycle to a human–electric hybrid," *IEEE Spectr.*, vol. 46, no. 9, pp. 26–27, Sep. 2009.

[13] W. Du, D. Zhang, and X. Zhao, "Research on battery to ride comfort of electric bicycle based on multi-body dynamics theory," in *Proc. IEEE Int. Conf. Autom. Logistics*, Aug. 2009, pp. 1722–1726.

[14] F. Yu, D. F. Li, and D. A. Crolla, "Integrated vehicle dynamics control—State-of-the art review," in *Proc. IEEE Int. Conf. Veh. Power Propulsion*, Sep. 2008, pp. 1–6.

[15] J. Cao, H. Liu, P. Li, and D. Brown, "State of the art in vehicle active suspension adaptive control systems based on intelligent methodologies," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 392–405, Sep. 2008.

[16] D. Yin, S. Oh, and Y. Hori, "A novel traction control for EV based on maximum transmissible torque estimation," *IEEE Trans. Ind. Electron.*, vol. 56, no. 6, pp. 2086–2094, Jun. 2009.

[17] J. Park, Z. Chen, L. Kiliaris, M. L. Kuang, M. A. Masrur, A. M. Phillips, and Y. L. Murphey, "Intelligent vehicle power control based on machine learning of optimal control parameters and prediction of road type and traffic congestion," *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 4741–4756, Nov. 2009.

[18] R. C. Hsu, W. M. Lee, C. T. Liu, and C. H. Chen, "A reinforcement learning based power assisted mechanism with comfort of riding for light electric vehicle," in *Proc. IEEE Int. Conf. Veh. Technol.*, 2010, pp. 1–5.

[19] A. Haddoun, M. E. H. Benbouzid, D. Diallo, R. Abdessemed, J. Ghouili, and K. Srairi, "Modeling, analysis, and neural network control of an EV electrical differential," *IEEE Trans. Ind. Electron.*, vol. 55, no. 6, pp. 2286–2294, Jun. 2008.

[20] A. Chaibet, C. Larouci, and E. Grunn, "An electric simulator of a vehicle transmission chain coupled to a vehicle dynamic model," in *Proc. IEEE IECON*, 2008, pp. 1578–1583.

[21] C. Larouci, A. Chaibet, and M. Boukhnifer, "A vehicle transmission simulator applied to the automated driving," in *Proc. IEEE Int. Conf. Veh. Technol.*, 2009, pp. 1–5.

[22] K. Masatoku, "Regulations on electric bicycles in Japan," in *Proc. Jpn. Conf. Safety Popularization Elect. Assist Bicycles*, 2001, pp. 1–5.

[23] L. Cai, A. B. Rad, and W. L. Chan, "An intelligent longitudinal controller for application in semiautonomous vehicles," *IEEE Trans. Ind. Electron.*, vol. 57, no. 4, pp. 1487–1497, Apr. 2010.

[24] T. Li and Y. Fujimoto, "Control system with high-speed and real-time communication links," *IEEE Trans. Ind. Electron.*, vol. 55, no. 4, pp. 1548–1557, Apr. 2008.

[25] S. Cong and Y. Liang, "PID-like neural network nonlinear adaptive control for uncertain multivariable motion control system," *IEEE Trans. Ind. Electron.*, vol. 56, no. 10, pp. 3872–3879, Oct. 2009.

[26] F. Moreno, J. Alarcon, R. Salvador, and T. Riesgo, "Reconfigurable hardware architecture of a shape recognition system based on specialized tiny neural networks with online training," *IEEE Trans. Ind. Electron.*, vol. 56, no. 8, pp. 3253–3263, Aug. 2009.

[27] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, pp. 237–285, 1996.

[28] L. Hu, C. Zhou, and Z. Sun, "Estimating biped gait using spline-based probability distribution function with Q-learning," *IEEE Trans. Ind. Electron.*, vol. 55, no. 3, pp. 1444–1452, Mar. 2008.

[29] K. S. Hwang, Y. J. Chen, and C. H. Lee, "Reinforcement learning in strategy selection for a coordinated multirobot system," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 6, pp. 1151–1157, Nov. 2007.

[30] C. F. Juang and C. H. Hsu, "Reinforcement ant optimized fuzzy controller for mobile-robot wall-following control," *IEEE Trans. Ind. Electron.*, vol. 56, no. 10, pp. 3931–3940, Oct. 2009.

[31] C. T. Liu and R. C. Hsu, "Adaptive power management based on reinforcement learning for embedded system," in *Proc. Int. Conf. IEA/AIE*, Jun. 2008, vol. 5027, pp. 513–522.

[32] P. Y. Yin, "Maximum entropy-based optimal threshold selection using deterministic reinforcement learning with controlled randomization," *Signal Process.*, vol. 82, no. 7, pp. 993–1006, Jul. 2002.

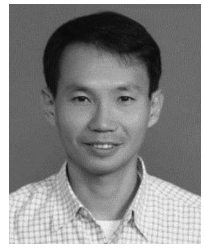
[33] V. Trifa, C. Marginean, and E. Trifu, "Contributions regarding the development of a light urban transportation vehicle—Motor and PWM inverter design," in *Proc. IEEE Int. Conf. Optim. Elect. Electron. Equipment*, 2008, pp. 307–312.

Roy Chaoming Hsu (S'93–M'01) received the M.S. and Ph.D. degrees in electrical engineering from The Pennsylvania State University, University Park, in 1991 and 1995, respectively.

From November 1995 to December 2000, he was with the National Space Organization, Hsinchu, Taiwan, working on satellite flight software. From December 2000 to July 2001, he was with Loop Telecom, working on telecommunication–Internet management software. From July 2001 to January 2005, he was an Associate Professor with the National Formosa University, Taiwan. He is currently an Associate Professor with the Department of Computer Science and Information Engineering, National Chiayi University, Chiayi, Taiwan. His expertise is in machine learning, dynamic power management for embedded system, image processing, and pattern recognition.

Cheng-Ting Liu received the B.S. degree in aeronautical engineering in 2005 from the National Formosa University, Yunlin, Taiwan, and the M.S. degree in computer science and information engineering in 2008 from the National Chiayi University, Chiayi City, Taiwan, where he is currently working toward the Ph.D. degree in the Department of Computer Science and Information Engineering.

His research interests include artificial intelligence, embedded systems, and image processing.



Din-Yuen Chan received the M.S. degree in medical engineering and the Ph.D. degree in electrical engineering from the National Cheng Kung University, Tainan, Taiwan, in 1992 and 1996, respectively.

He was a Senior Engineer with Multi-Media Inc., Minneapolis, MN, developing the videophone products. From 1997 to 2003, he was an Associate Professor with the Department of Information Engineering, I-Shou University, Kaohsiung, Taiwan. Currently, he is a Professor with the Departments of Electrical Engineering and of Computer Science and Information Engineering, National Chiayi University, Chiayi, Taiwan, where he was the Founder and has been the Chairperson of the Department of Electrical Engineering since 2008. His research interests are in the areas of data compression, visual communication, and image/video/audio signal processing.