SICE Annual Conference in Fukui, August 4-6, 2003
Fukui University, Japan

Fuzzy Target Acquired by Reinforcement Learning for Parking Control

Seiji YASUNOBU, Tomoya MATSUBARA

Institute of Engneering Mechanics and Systems
University of Tsukuba
Ibaraki 305-8573, JAPAN
yasunobu@esys.tsukuba.ac.jp

Abstract: In this paper, a flexible action deciding method under various situations like a human being based on the concept of "fuzzy target" is proposed, and it applied to parking control. The "fuzzy target" is acquired by reinforcement learning, and it contains the grade of value of various targets. The simulation results show the effectiveness of the proposed method.

Keywords: Intelligent control, Fuzzy control, Fuzzy target, Reinforcement learning, Parking control

1. Introduction

It is important to achieve a flexible action decision based on knowledge and experience like a human being by the computer to construct a system which is gentle and soft to humans. Humans are using own knowledge and experience to decide various actions[1].

In that case, the people acquire many targets based on knowledge and experience. An action is decided by using these many targets. Therefore, he can flexibly treat the change in various situations.

The parking control is one of the problems that should deal flexibly under various situations. However, it is difficult to obtain a lot of targets independently to deal with the change in a variety of parking lot situations with the computer flexibly, and to use it flexibly.

In this paper, the method of deciding actions as flexibly as a human being based on the concept of "fuzzy target" is proposed. The "fuzzy target" is acquired by reinforcement learning, and it contains the grade of value of various targets. It applied to intelligent parking control system for nonholonomic vehicle. The effectiveness of this proposed method is evaluated by the simulation.

2. Action decision method based on fuzzy target

Figure 1 shows the outline of the proposed action determination method based on fuzzy target.

Reinforcement learning[5] is known as one of the method of acquiring knowledge by the computer. It is a learning method to imitate the technique to which living creatures study the success by the trial and error. As a result, the knowledge based on the experience can be acquired. In this learning method, when a certain target is selected and the action succeeds, the reward is given. On the other hand, the punishment is given when action fails. And worth of the knowledge falls.

As a result, the grade of the value of many targets can be acquired.

Fuzzy set is known as one of the method of showing many objects equipped with grades. The degree of the value of each target can be plainly described by using the fuzzy set.

In this paper, fuzzy set which contains many targets with the grade of the value and based on the experience is "fuzzy target". Because the fuzzy target contains many targets, it is possible to use this target to decide the action even in the situation in which a part of this target cannot be used. Thus, the fuzzy target has the robustness character that is not in the conventional single target.

Predictive fuzzy control[6] is a method of the intelligent control based on human control strategy. That is predicting the future situation by using control instruction candidates.Various values are judged by fuzzy inference. This fuzzy inference uses the expert's knowledge and experience described by fuzzy rules. As a result, a similar control to the predictive control based on expert's experience is achieved.

In this research, flexible action decision according to the situation is achieved as follows: First, the fuzzy target which is the knowledge based on the experience is included in the predictive fuzzy controller as one of the expert's control knowledge. Afterwards, the values of many targets are

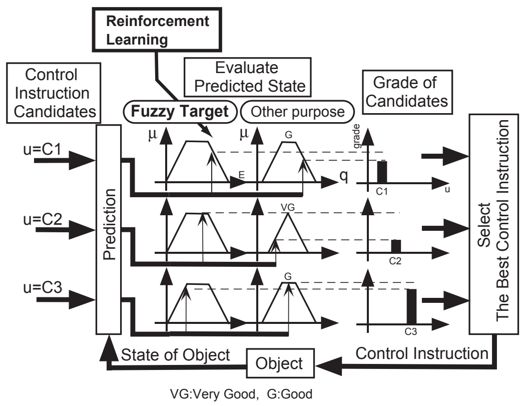

Fig. 1: Outline of the proposed action determing method based on fuzzy target

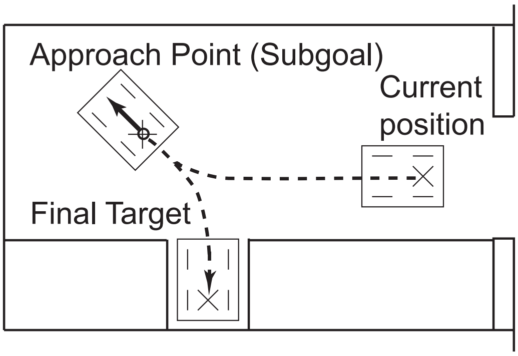

Fig. 2: Human control strategy in parking vehicle

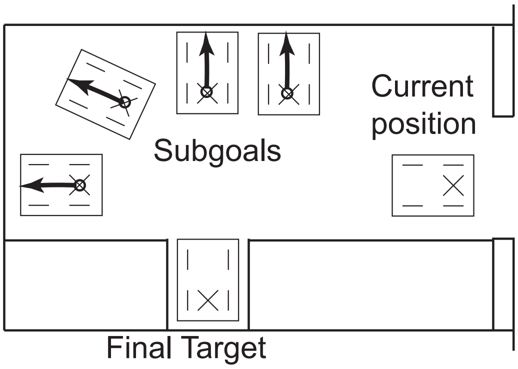

Fig. 3: Various parking strategies

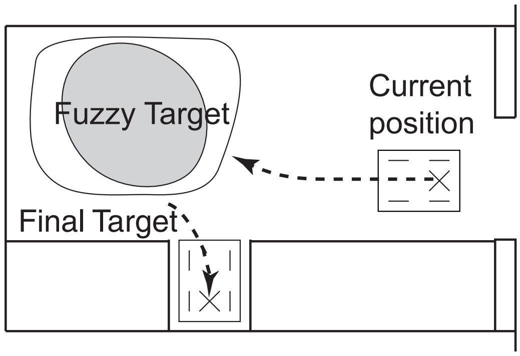

Fig. 4: Outline of the fuzzy target concerning parking control

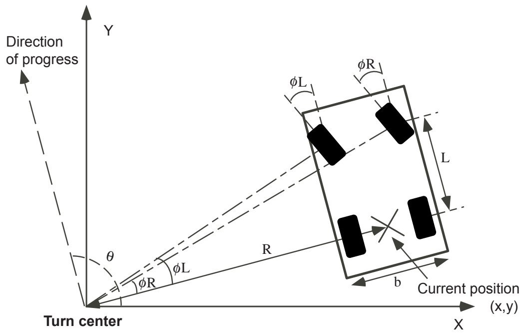

Fig. 5: Kinematics constraint in nonholonomic vehicle

judged under current situation by a fuzzy inference. Finally, the most appropriate target and the best control instruction are selected.

3. Application for parking control based on fuzzy target

The four-wheeled vehicle controls three outputs (position (x, y) and direction) only by two inputs (operation of the steering wheel and speeds), and it is known as nonholonomic character. For such reason, driver's skills and experiences are required for the drive of the four-wheeled vehicle. Especially, parking in the limited area such as parking lots is difficult. To solve this parking problem, a variety of parking systems[2)3)4)] was proposed. The human control strategy to parking nonholonomic vehicle[2)4)] is can be explained (Fig. 2) as follows: First, in order to park a car in the back, the car passes the front of the garage. Next, stop in an approach point. Finally, it puts into a parking lot.

However, many solutions that can be selected exist as shown in Fig.3. Driver understands existing of many targets that can be chosen based on the driving knowledge and experience. Therefore, it is possible to deal flexibly even when the target, which can be selected is limited with the obstacles in the parking lot.

In this research, fuzzy target is acquired by many targets obtained by reinforcement learning. Then, this fuzzy target is included in the predictive fuzzy controller as one of expert's parking control knowledge. Evaluations of many targets are judged by fuzzy reasoning under current situation. And most suitable target is chosen. In such methods, a flexible intelligent parking control system is achieved which is based on many targets like a human being.

Outline of the fuzzy target concerning parking control is shown in Fig. 4.

3.1 Characteristic of nonholonomic vehicle

Nonholonomic vehicle has the kinematics constraint shown in Fig. 5.

Slipping of the tire and the generating of the centrifugal force can be disregarded when the speed is very slow in the four-wheeled vehicle of the front wheel steer. The state of a present vehicle is shown by angle θ of coordinates (x, y) in the middle of the rear wheel and the x axis and the direction of progress. The average of the right and left front wheel steer corner is ϕ, the distance of the front wheel and the rear wheel (wheelbase) is L, the average speed of the front wheel is v, and the rotation speed of the front wheel is ω. At this time, the equation of motion when turning with the steer mechanism of *Ackerman-Jeantaud* becomes the following.

$$
\begin{aligned}
\frac{dx}{dt} &= v\cos(\phi)\cos(\theta) \\
\frac{dy}{dt} &= v\cos(\phi)\sin(\theta) \\
\frac{d\theta}{dt} &= \frac{v}{L}\sin(\phi) \qquad (1) \\
\frac{d\theta}{dt} &= \omega
\end{aligned}
$$

When the steer corner ϕ and speeds v are kept constant, angle θ_0 and present vahicle position (x_0, y_0) are assumed. In that case, after t seconds, car position (x_t, y_t) and angle θ_t are calculated in the following expression.

$$
\begin{aligned}
x_t &= \frac{L}{\tan(\phi)}\cos(kt) + x_0 \\
y_t &= \frac{L}{\tan(\phi)}\sin(kt) + y_0 \\
\theta_t &= kt + \theta_0
\end{aligned}
$$

In this case, k is as follows.

$$
k = \frac{v\cos(\phi)\tan(\phi)}{L}
$$

3.2 Acquisition of parking knowledge by reinforcement learning

In this research, the evaluation value of two or more targets is obtained by the reinforcement learning to make a fuzzy target.

PSP (Profit Sharing Plan)-learning[7)8)] is a reinforcement learning method, which goes back the contingency fee after the action in some stages at each stage and distributes it. PSP-learning is often applied to a discrete state transition. However, fPSP(fuzzy Profit Sharing Plan)-learning[9)], which is the reward to an individual accomplishment of a goal degree is given by a fuzzy evaluation is applied to the drive of the vehicle which is the control of the state of a continuous value.

The knowledge concerning parking acquired by fPSP-learning is consisting of the pair of state and action, like this "*IF* <condition:c_n> *THEN* <action:a_n>". That is called "S-table". The value of S-table is an evaluation value of the selected action.

In this research, the parking lot is assumed to be a state space where three dimensions x, y and θ are partitioned. x and y are divided from the final target into the lattice point based of 2m for each for the labeling. Direction θ of the vahicle at each position is divided into eight azimuths (0, 0.25π, 0.5π, 0.75π, π, 1.25π, 1.5π, and 1.75π). For instance, the parking lot shown by the Fig. 6 is divided into the 84 lattice points and 84×8(azimuth)=672 labels are put under the situation. Therefore, 672 kinds of a_n=(x_T, y_T, θ_T) is divided into the target for the action selection as for state c_n=(x_T, y_T, θ_T) into 672. Final target is (0m, 0m, 0.5π).

The experience of the unknown is tried by selecting target a_n with roulette at random in each state c_n to acquire the action knowledge. The reward of the result of execution is distributed to all the selected S-table values. When it can reach the final target, the reward is given by the difference between the time limit (250 seconds) and the elapsed times. For instance, if it reaches a final target at 80 seconds, the reward to the knowledge used with the place of departure point is a value discounted to 170 by the number of target points passed. When it is not able to finish on the limitation time or movement becomes impossible with an obstacle, the punishment is given. And the evaluation value of the knowledge used in that case is lowered.

After learning, fuzzy target is made from the fuzzy set in which the evaluation value of the obtained target of two or more is a membership value. For purposes of fast acquiring knowledge, cascade fuzzy controller[10)]
(Fig. 7) was used.

3.3 Outline of the system

Figure 8 shows the outline of the proposed hierarchical intelligent parking control system based on fuzzy target. This system is divided into the three parts. The detector part is judging the attainment to the target. The target setting part is deciding target with the fuzzy target acquired by reinforcement learning and expert's driving knowledge. Then, the automatic driving part drives the vehicle toward the target set in the target setting part.

Detector Part: Detector part evaluates contact to the obstacle and the attainment to the target, and judges the necessity of a new target. When some reasons make the attainment to a present target difficult, a present target is reset, and the target setting instruction is output to the target setting part.

Target Setting Part: When target setting instruction is received, the target setting part set the fuzzy target from the

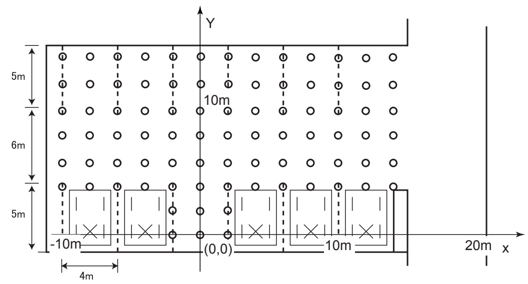

Fig. 6: The dimemsion of parking lot

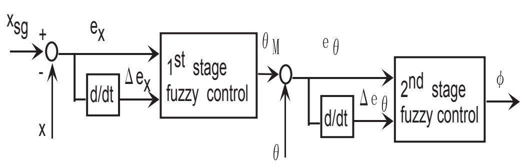

Fig. 7: Block diagram of cascade fuzzyy controller

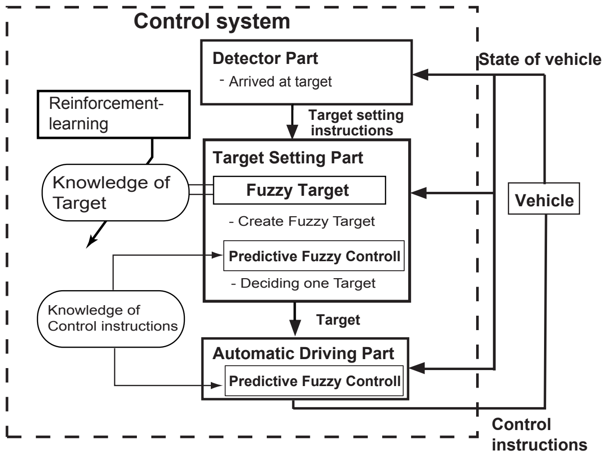

Fig. 8: Outline of the hierarchical intelligent controller

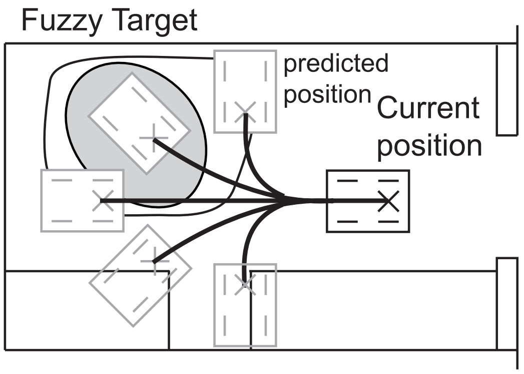

Fig. 9: Use of fuzzy target by predictive fuzzy control

knowledge obtained by reinforced learning. In that case, a surrounding obstacles restricts fuzzy target. The most ideal target position corresponding to the situation is decided from a fuzzy target by the predictive fuzzy control.

The vehicle is moved temporary. And the position is foreseen by using the operation instruction candidate decided beforehand in the future as shown in Fig. 9.

The most ideal target in the fuzzy target is determined by fuzzy evaluation concerning to evaluation value of each target, distance to obstacles and the foreseen future position. As a result, the target position with high possibility that the, parking succeeds without coming in contact with the obstacle is decided and it is output to the automatic driving part.

The fact that should be careful here is not to contain the evaluation of risk to the obstacle in the evaluation value of a fuzzy target. Therefore, in order to know the risk of a target position, predictive fuzzy controller is used for a vehicle car and it is moved temporary based on the operation instruction candidates.

Automatic Driving Part: The operation instruction is decided to reach the target set at the target setting part and outputs it to the vehicle. In that case, the predictive fuzzy controller is used also in this part. It foresees future state based on the same operation instruction candidate sets in the target setting part for the discovery and the evasion of obstacles. Afterwards, the most appropriate operation instruction in the operation instruction candidate sets is selected. As a result, it is possible to cope with surrounding circumstances flexibly.

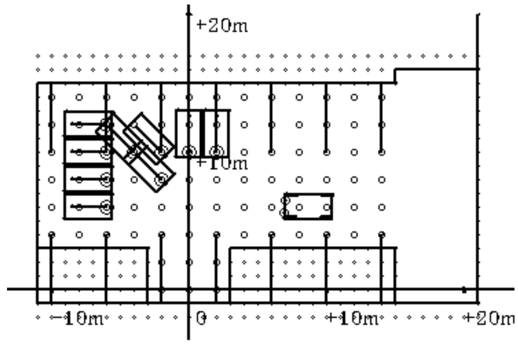

Fig. 10: Fuzzy target at the vehicle state (10m,6m,π) with membersip value more than 0.100

4. Simulation

The effectiveness of this proposed hierarchical intelligent parking control system based on fuzzy target is examined by computer simulation. This simulation is done in the assumed concrete parking lot and the vehicle.

4.1 Simulation condition

Simulations are carried out with the characteristics of the vehicle as follows:

The wheelbase is 2.6m, distance between axis and bumper is 0.4m, width is 1.7m, the smallest turning radius is 6.0m and the velocities are 0.4m/s for moving forward, 0.0m/s for stop and -0.4m/s for moving backward.

The shape of the parking lot is shown in Fig. 6. This parking lot where a fuzzy target was acquired by reinforcement learning is wide without obstacles in front of the final target (0m, 0m, 0.5π).

After the fuzzy target is acquired by the reinforcement learning, the simulation is done. Details of this fuzzy target (elements and membership value) is as follows:

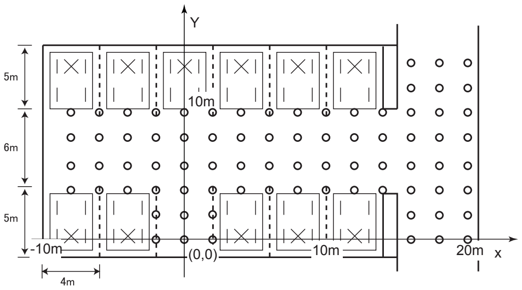

Fig. 11: The dimemsion of parking lot with obstacles

Element number	Target position	Membersip value
0	(0m, 0m, 0.0π)	0.000
1	(0m, 0m, 0.5π)	0.000
2	(0m, 0m, 0.5π)	0.000
⋮		
42	(0m, 10m, 0.5π)	0.131
99	(-2m, 8m, 0.75π)	0.147
107	(-2m, 10m, 0.75π)	0.348
170	(2m, 10m, 0.5π)	0.143
219	(-4m, 10m, 0.75π)	0.103
300	(-6m, 6m, 1.0π)	0.176
308	(-6m, 8m, 1.0π)	0.160
316	(-6m, 10m, 1.0π)	0.304
324	(-6m, 12m, 1.0π)	0.283
⋮		
669	(0m, 0m, 0.0π)	0.000
670	(0m, 0m, 0.0π)	0.000
671	(0m, 0m, 0.0π)	0.000

In this fuzzy target, nine elements with the membersip value more than 0.100 are existed. Each position of the element is shown in Fig. 10.

The parking control simulations are done from the point (10m, 6m, π) to the final target (0m, 0m, 0.5π) by proposed method explained in section 3.3. First, the simulation is done under the situation without obstacles. Then, the simulation with obstacles at the front of the final target that is shown by Fig. 11 is done. The same fuzzy target is used when the simulation is done in two different conditions.

4.2 Simulation results

Figure 12 shows the simulation result in the state without obstacles. As a result of a fuzzy inference from the position (10m, 6m, π), target point was decided to the point (-2m, 10m, 0.75π) that is a part of the fuzzy target. It reached to the point (-2m, 10m, 0.75π) by the operation instruction of controller. Afterwards, the vehicle was reached to the final target. This point (-2m, 10m, 0.75π) is corresponds to the turnabout point on human parking strategy.

Next, Fig. 13 shows the simulation result in the state with obstacles. In this simulation, the motion space is much limited with obstacles in the parking lot. The same fuzzy target as what was used before is used without redoing acquisition of knowledge even if the situation changed. As a result of a fuzzy inference from the position (10m, 6m, π), target point was decided to the point (-6m, 6m, π) that is a part of the fuzzy target without restrictions by obstacles. It reached to the point (-6m, 6m, π) by flexible operation instruction of controller. Afterwards, the vehicle was reached to the final target. This point (-6m, 6m, π) is corresponds to the turnabout point on human parking strategy.

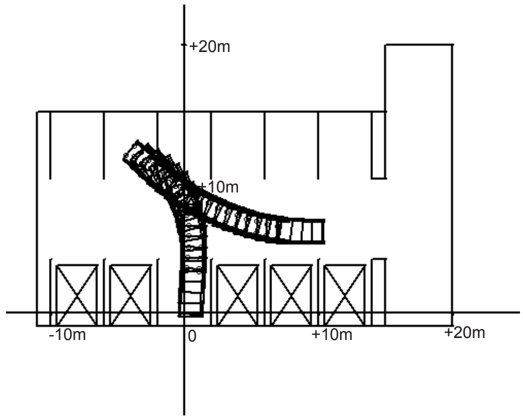

Fig. 12: An example of parking trajectory from (10m,6m,π) to (0m,0m,0.5π) without obstacles

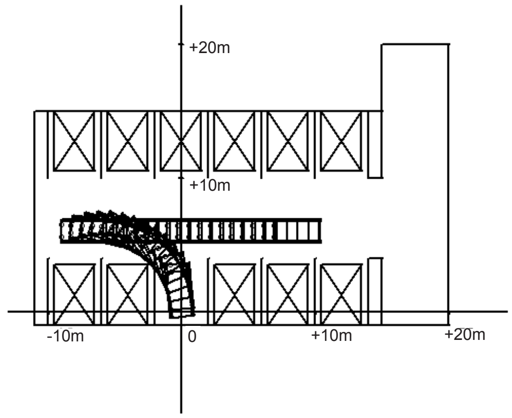

Fig. 13: An example of parking trajectory from (10m,6m,π) to (0m,0m,0.5π) with obstacles

As mentioned above by the proposed system using the fuzzy target, it was able to respond to change of a situation flexibly like man. And the vehicle was able to park at the final target. As a result, the effectiveness of the proposal method was confirmed by these simulations.

5. Conclusion

In this paper, the flexible action deciding method under various situations like a human being based on the concept of fuzzy target is proposed. The fuzzy target is acquired by reinforcement learning, and it contains the grade of many targets. It is included in the predictive fuzzy controller as one of expert's control knowledge. Afterwards, the values of many targets are judged under current situation by a fuzzy inference. And the most appropriate target is selected.

This proposed method is applied to parking control for nonholonomic vehicle. The effectiveness of this method was confirmed by the computer simulation done in the assumed concrete parking lot.

References

[1] Rasmussen. J, Skills Rules, and Knowledge; Signals, Signs, and Symbols, and other Distinctions in Human Performance Models, IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-13, No. 3, pp. 257-266, 1983.

[2] S. Yasunobu and N. Minamiyama, A Proposal of Intelligent Vehicle Control System by Predictive Fuzzy Control with Hierarchical Temporary Target Setting, Proc. of Fifth IEEE International Conference on Fuzzy Systems, New Orleans, pp. 873-878, 1996.

[3] C. Laugier et al., Sensor-Based Control Architecture for a Car-Like Vehicle, IEEE Autonomous Robots, vol. 6, pp. 165-185, 1999.

[4] S. Yasunobu, S. Saitou and Y. Suryana, Intelligent Vehicle Control in Narrow Area based on Human Control Strategy, World Multiconference on Systemics, Cybernetics and Informatics (SICI 2000), Vol. VII, pp. 309-314, 2000.

[5] Richard S. Sutton and Andrew G. Barto, REINFORCEMENT LEARNING: An Introduction, A Bradford Book, 1998.

[6] S. Yasunobu and S. Miyamoto, Automatic train operation by predictive fuzzy control, Industrial Application of Fuzzy Control (M. Sugeno, Ed.), North Holland, pp. 1-18, 1985.

[7] J. J. Grefenstette, Credit Assignment in Rule Discovery System Based on Genetic Algorithms, Machine Learning 3, Kluwer, pp. 225-245, 1988.

[8] T. Horiuchi, A. Fujino, O. Katai and T. Sawaragi, Q-PSP learning: An Exploration-Oriented Q learning and Its Applications, The Society of Instument and Control Enggineers, Vol. 35, No. 5, pp. 645-653, 1999.

[9] S.Yasunobu, Fuzzy-Reinforcment-learning Design of Intelligent Controller for Nonholonomic Vehicle, Proc. of The 12^{th} FAN Intelligent Systems Symposium, pp. 105-108, 2002.

[10] S. Yasunobu and Y. Suryana, Hierarchical Intelligent Control by PSP-learning and Cascade Fuzzy Control for Parking Vehicle in Narrow Area, The Transactions of The Institute of Electrical Engineers of Japan, vol. 122-C, No. 2, pp. 315-316, 2002.