Acquisition of Fuzzy Target by Reinforcement Learning and Its Application to Four-wheeled Vehicle Control

Tomoya Matsubara
Master's Program in Science and Engineering
University of Tsukuba
Ibaraki 305-8573, JAPAN
E-mail: tomoya_m@edu.esys.tsukuba.ac.jp

Seiji Yasunobu
Department of Intelligent Interaction Technologies
University of Tsukuba
Ibaraki 305-8573, JAPAN
E-mail: yasunobu@iit.tsukuba.ac.jp

***Abstract*— A "fuzzy target" is a control target value defined by "fuzzy set". In this paper, it is acquired by reinforcement learning, which is a machine learning method.**

The control target had been defined as a single target so far. Therefore, when restrictions were newly added and a surrounding situation changed, it was necessary to calculate the target again and set it again. But, by using this fuzzy target, control system is able to correspond to changes in the situation flexibly without target resetting.

This target was applied to the driving control of a four-wheeled vehicle, under the change of the situation. And computer simulations that assumed an actual vehicle were done. Simulation results show the effectiveness of this fuzzy target.

I. Introduction

Fuzzy logic is a method to handle various things including fuzziness by a computer. It can be applied to a complex problem such as natural language processing, decision-making, and more [1].

We described various expert's knowledge in membership functions by using the fuzzy logic, and applied it to a system control for a control target [2].

This control target has been defined as single valued target in the current situation of a controlled system. However, when defining single target in the real world, it is difficult to take into consideration the system dynamics and the change of a surrounding situation.

In this paper, we consider the control target value as a "fuzzy target" [3]. We use fuzzy target to respond to the change of a situation flexibly. This fuzzy target is acquired by the reinforcement learning [4] that is a kind of machine learning method. The membership value of fuzzy target is an evaluation value of each targets obtained by the reinforcement learning. We constructed the control system, which is corresponding to change of a situation flexibly by using this fuzzy target. And it is applied to the driving control of the four-wheeled vehicle under the change of a situation by adding obstacles. Effectiveness of this system is confirmed by the computer simulation that assumed an actual vehicle.

II. Outline of the control system based on fuzzy target

"Fuzzy target" is the control target which was treated by the fuzzy set (figure 1). When the universe of the target is defined as R, fuzzy target $\widetilde{T_n}$ in state c_n is described by the following expressions:

$$\widetilde{T_n} = \int_R \mu_{\widetilde{T_n}}(r_i)/r_i, \qquad r_i \in R.$$

. r_i is an element of the fuzzy target $\widetilde{T_n}$, which is included in R. $\mu_{\widetilde{T_n}}(r_i)$ is a membership value at r_i.

This fuzzy target is beforehand acquired in the situation without restrictions, and foreseeing the future state of a system uses it. This process is shown in figure 2 and figure 3.

First of all, operation instruction candidate $u = Cr_i$ is calculated in current state c_n by using each target element r_i of fuzzy target.

Next, the future state is foreseen by using forward model of the controlled system according to the operation instruction candidate.

Finally, each operation instruction candidate is evaluated by multipurpose fuzzy evaluation, and the operation instruction candidate with the highest value is given to the controlled system as an actual operation instruction

Fig. 1. Fuzzy Target.

The advantage of using the fuzzy target when controlling is "Containing alternatives of a useful control target element". The control target was defined as a single target. Therefore, when restrictions were newly added and a surrounding situation changed, it was necessary to calculate the target again and set it again. However, the fuzzy target contains all elements of the target. Therefore, when the situation changes by the restriction's newly joining, the task can be appropriately achieved by using useful elements in the fuzzy target.

III. Four-wheeld vehicle control based on fuzzy target

The four-wheeled vehicle is known as a system which has the nonholonomic restraint. This restraint is three outputs (position (x, y) and the direction θ) of the vehicle are controlled by only two inputs (operation of the steering wheel and the speed). Therefore, controlling of the vehicle from the current state $c_n = (x_0, y_0, \theta_0)$ to arbitrary state (x, y, θ) is impossible. By such a reason, it is necessary to show appropriate target $T_n = (x_t, y_t, \theta_t)$ to move the vehicle from the current state $c_n = (x_0, y_0, \theta_0)$ to the final target (x_G, y_G, θ_G). This target T_n considers the dynamic characteristic of the vehicle.

In the no obstacle and enough wide situations, the target is not only one in this situation. And it is possible to think variously like figure 4(a). In this case, each element r_i of the target T_n has the degree of the achievement of the task "Efficient attainment to the final target". This degree contains move costs, such as the move time and amount of operations, and attainment level of the final target. The target that exists in the distance from the current position or the target with many amounts of handle operations has large move costs. On the other hand, it exists near the current position and the target with few amounts of handle operations has small move costs. The target with both small move costs and high achievement to the final target has high value.

In this paper, the evaluation value $\mu_{\widetilde{T_n}}(r_i)$, which is all the elements r_i of the target in the current state c_n is acquired in a large situation without an obstacle. Thus, a "Fuzzy target" $\widetilde{T_n}$ as shown in a figure 4(b) is acquired.

It is very difficult for the car to plan an appropriate route in consideration of a dynamic characteristic and a surrounding situation to make the four-wheeled vehicle reach the final target. Therefore, in this paper, control of the vehicle is

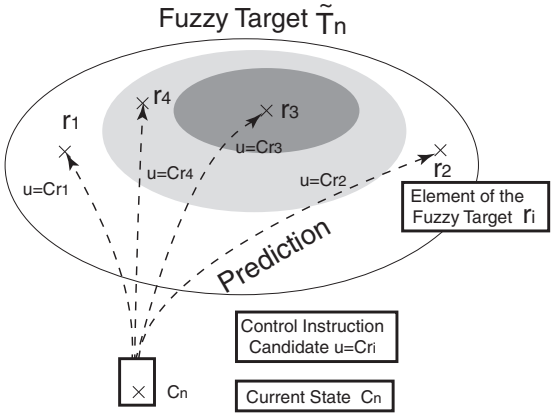


Fig. 2. Conceptual image of the proposed system based on Fuzzy Target.

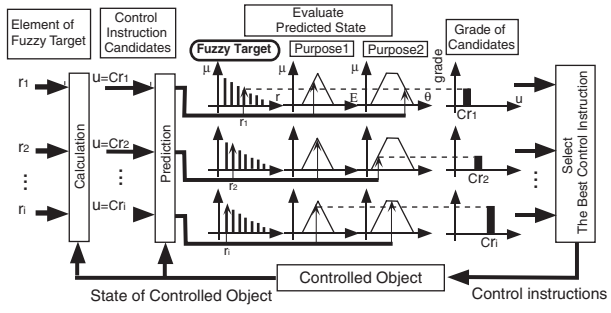
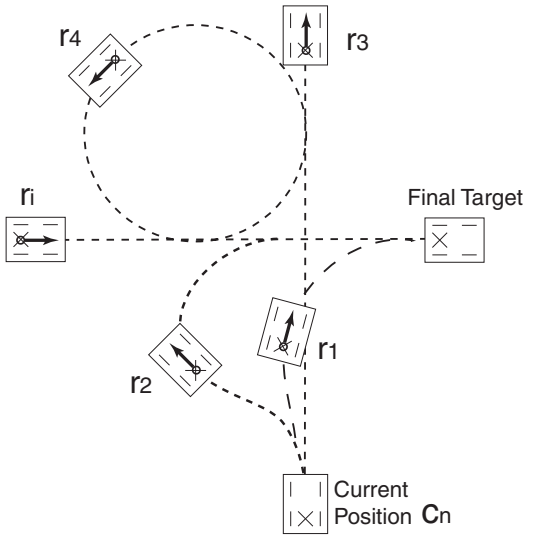
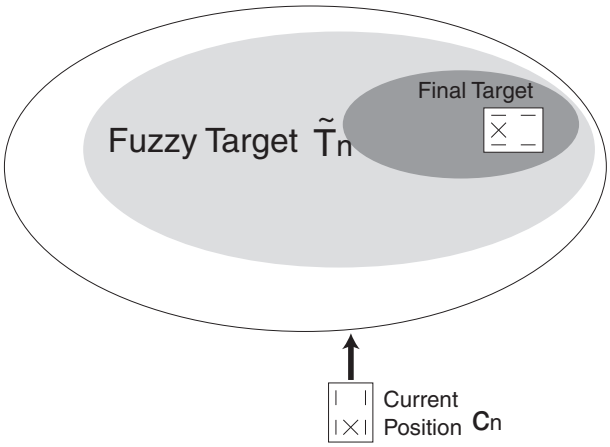


Fig. 3. Outline of the proposed system based on fuzzy target.



(a) Various patterns of the target Tn



(b) Fuzzy Target $\tilde{T}_n$

Fig. 4. Image of Fuzzy Target.

performed based on the fuzzy target gained in the past. As a result, it aims to adjust to the changing situation, and to make it reach the final target.

This chapter describes about the characteristic of a four-wheeled vehicle, acquisition method of a fuzzy target, and a control system based on this target.

A. Characteristic of the vehicle

Nonholonomic vehicle has the kinematics constraint shown in figure 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 angle is ϕ, the distance of the front wheel and the rear wheel (wheelbase) is L, the average speed of the front wheel is v. At this time, the equation of motion when turning with the steer mechanism of *Ackerman-Jeantaud* becomes as follows:

$$\frac{dx}{dt} = v\cos\phi\cos\theta,$$
$$\frac{dy}{dt} = v\cos\phi\sin\theta,$$
$$\frac{d\theta}{dt} = \frac{v}{L}\sin\phi.$$

The steer angle ϕ and speed v are kept constant.

For the characteristic like the above-mentioned, the four-wheeled vehicle does not have the guarantee that it is possible to move from the point (x_0, y_0, θ_0) to the arbitrary point (x, y, θ). Therefore, it is necessary to set an appropriate target corresponding to a vehicle state and a surrounding situation, and perform the driving control based on it.

B. Acquisition of fuzzy target

Fuzzy target $\widetilde{T_n}$ at vehicle state c_n is acquired by the "Reinforcement learning"[4], which is a machine learning method. This learning method imitates the mechanism that living creatures study the success by trying and erring. By having chosen operation that is in a certain state, when succeeding, reward is given, and when failing, penalty is given. After repeating it, the knowledge necessary to achieve the task is acquired.

PSP (Profit Sharing Plan) - learning [5] is proposed as one of the techniques of the reinforcement learning. The contingency fee after the action whether the step how many floor goes back at each stage and this study method is distributed in one episode from the state of the first stage to the final target as shown in figure 6. In this paper, fPSP (fuzzy Profit Sharing Plan)-learning [6] is applied to the car drive. This technique evaluates each degree of target achievement by fuzzy evaluation, in order to determine the reward.

The knowledge that acquired by the fPSP-learning is a rule of "IF < condition:c_n > THEN< action:a_n >". This rule expresses the action evaluation of a pair of the state and the action.

It explains the method of obtaining action evaluation $S(c_n, a_n)$ by using fPSP-learning. In each state c_n, action a_n decides by roulette selection, and tries an unknown action. As a result, only when the vehicle can reach the final target, the reward is acquired.

Reward is given by the difference between the limitation time and the time required.

For example, if the limitation time is assumed to be 250 seconds, and it reaches a final target at 80 seconds by selecting action of several steps, the reward acquired by the entire episode becomes 170.

Penalty is given when it cannot reach in the time limit at a final target, or when it cannot be moved with an obstacle etc.

The reward (or penalty) acquired by the entire episode is discounted by the number of steps used to select the action, and it distributes it to each step. And, $S(c_n, a_n)$ is updated based on the distributed reward (or penalty) by following

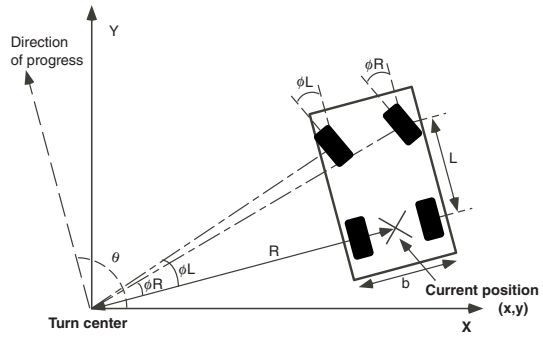


Fig. 5. Kinematics constraint in nonholonomic vehicle.

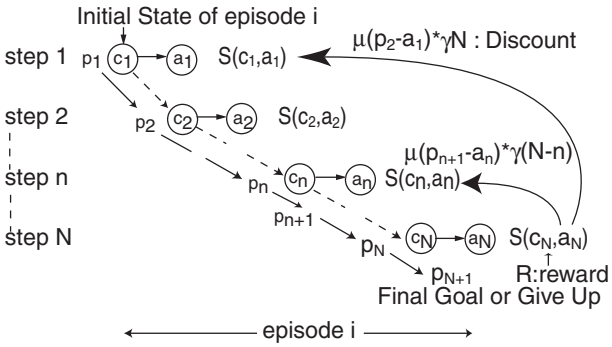


Fig. 6. PSP-learning deitributes reward of penalty to the previous fired rules.

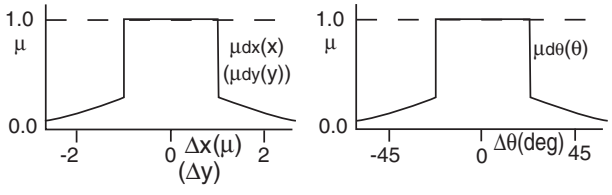


Fig. 7. Fuzzy set of position and angle error.

expression.

$$S(c_n, a_n) = (1-\alpha)S(c_n, a_n) + \alpha\mu(p_{n+1} - a_n)R * \gamma^{(N-n)}.$$

α shows the learning rate, γ shows the discount rate, n shows a number of steps, and N shows a number of maximum steps. "$\mu(p_{n+1} - a_n)$" is a fuzzy evaluation part, and the detail of it is as follows.

$$\mu(p_{n+1} - a_n) = \mu_d x(\Delta x_{n+1}) \wedge \mu_d y(\Delta y_{n+1}) \wedge \mu_d \theta(\Delta\theta_{n+1}).$$

It evaluates an action and that result by using the membership function shown in figure 7.

As a result of acquiring all state-action evaluation $S(c_n, a_n)$ by many trials, a state-action table is made. This table is called "S-table".

By looking up this S-table, the membership value of each target elements $\mu_{\widetilde{T}}(r_i)$ are decided. As a result, each element r_i and membership value $\mu_{\widetilde{T}}(r_i)$ of fuzzy target $\widetilde{T_n}$ in the state c_n are acquired.

C. Outline of the vehicle control system

Figure 8 shows the outline of a hierarchical intelligent driving control system based on the fuzzy target. This system has divided into three parts.

1) The detector part: It is judging of the attainment to the target. And it is detecting contact to the obstacle. When some reasons make difficult to the attainment to the present target, resetting it and the target setting instruction has been output to the fuzzy target setting part.

2) The fuzzy target setting part: It is setting fuzzy target about current position from target setting knowledge (S-table) that is acquired by the reinforcement learning, when target-setting instruction is received.

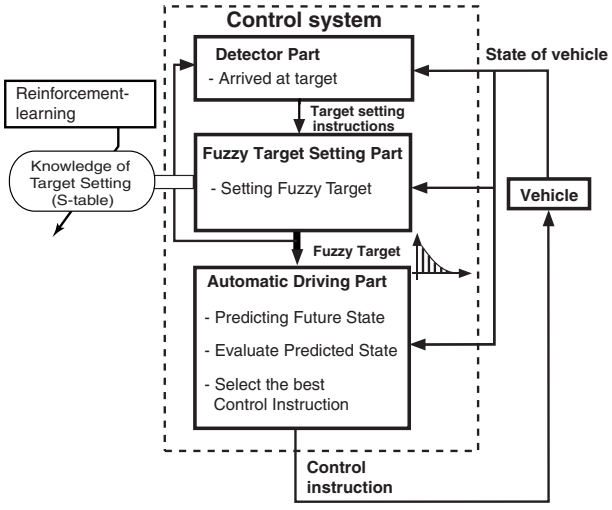


Fig. 8. Outline of the control system.

3) The automatic driving part: it calculates an appropriate control instruction by using fuzzy target $\widetilde{T_n}$, and the vehicle is controlled so as not to come in contact with the obstacle. Figure 9 shows details in the automatic driving part.

In the automatic driving part, the operation instruction candidate C_{r_i} is calculated by the cascade fuzzy controller [6], which is shown in figure 10. The future state of the vehicle will be predicted by using operation instruction candidate C_{r_i} and forward model of the vehicle. Then, predicted future state of the vehicle, membership value $\mu_{\widetilde{T_n}}(r_i)$, the angle deflections, and distance to obstacles are evaluated synthetically. This is done to all elements r_i of fuzzy target $\widetilde{T_n}$, operation instruction candidate C_{r_i} with the highest evaluation is selected, and this is given to a real vehicle as a control instruction. That is repeated to all elements r_i of fuzzy target $\widetilde{T_n}$. The operation instruction candidate C_{r_i} with the highest evaluation is selected and it is given to a real vehicle as a control instruction.

IV. Simulation

Effectiveness of the constructed control system is confirmed by the computer simulation which assumes an actual vehicle. The task of moving the vehicle from an arbitrary position to the final target $(20m, 20m, 0.25\pi)$ is assumed in the area of figure 11. In this case, the moveable area of the vehicle is restricted by adding obstacles later.

It aims to correspond to the changed situation flexibly by using the fuzzy target that was acquired in the past. In this paper, an initial position is assumed to be a $(18m, 10m, 0.5\pi)$, and the simulation is executed. Parameters of the vehicle are as follows: width of the car is 1.7m, wheelbase is 2.6m, minimum-turning radius is 6.0m, and car speed is 0.4m/s.

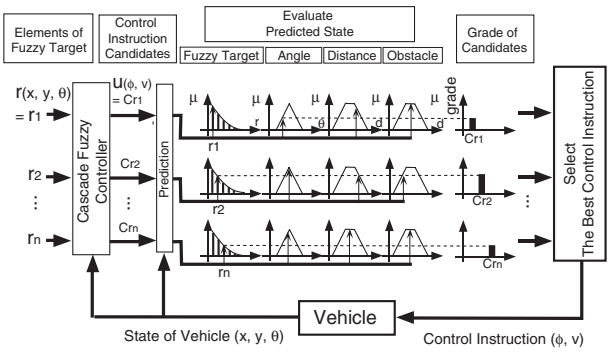


Fig. 9. Control system based on Fuzzy Target.

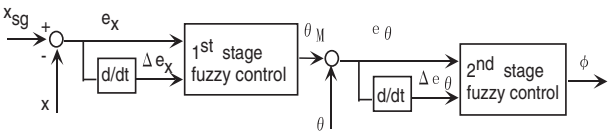


Fig. 10. Cascade fuzzy controller.

Target position	Membersip value
(2m, 6m, 0.25π)	0.033
(2m, 10m, 0.0π)	0.301
(4m, 0m, 0.5π)	0.111
(4m, 18m, 0.0π)	0.043
(6m, 0m, 0.75π)	0.013
(6m, 10m, 0.25π)	0.031
(6m, 14m, 0.0π)	0.058
(8m, 0m, 0.75π)	0.012
(8m, 4m, 0.0π)	0.077
(8m, 10m, 0.25π)	0.026
(8m, 16m, 0.0π)	0.021
(10m, 8m, 1.75π)	0.017
(10m, 16m, 0.25π)	0.015
(10m, 18m, 0.25π)	0.012
(12m, 2m, 0.5π)	0.012
(12m, 4m, 0.75π)	0.043
(12m, 6m, 0.0π)	0.367
(12m, 10m, 0.5π)	0.019
(12m, 12m, 0.5π)	0.041
(12m, 14m, 1.25π)	0.020
(14m, 12m, 0.5π)	0.047
(16m, 2m, 0.0π)	0.011
(16m, 8m, 1.25π)	0.011
(16m, 8m, 1.5π)	0.011
(16m, 10m, 1.75π)	0.012
(16m, 12m, 0.25π)	0.017
(16m, 12m, 1.25π)	0.017
(18m, 0m, 0.75π)	0.101
(18m, 4m, 0.25π)	0.110
(18m, 6m, 0.75π)	0.105
(18m, 8m, 1.0π)	0.030
(18m, 16m, 1.75π)	0.011
(20m, 10m, 0.75π)	0.100
(20m, 12m, 0.75π)	0.013
(20m, 12m, 1.25π)	0.042
(20m, 14m, 1.0π)	0.023
(20m, 20m, 0.25π)	0.999

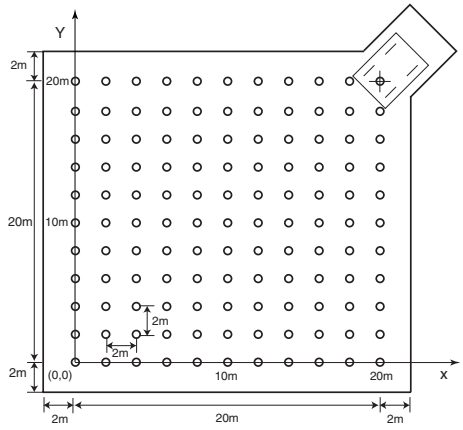


Fig. 11. Demension of the field.

Fuzzy target was acquired by fPSP-learning at the situation which is shown in figure 11. The area was separated discretely in each 2m grid and lattice point was arranged. In centers of each lattice point, direction θ of the vehicle is divided into eight directions $(0, 0.25\pi, 0.5\pi, 0.75\pi, \pi, 1.25\pi, 1.5\pi, and 1.75\pi)$.

In the environment shown in figure 11, there are 121 of lattice points. For such reasons, there are $121 \times 8 = 968$ states $c_n = (x_T, y_T, \theta_T)$ and actions $a_n = (x_T, y_T, \theta_T)$. Therefore, "S-table" where the rule of the state and the action is shown with one table becomes the table of 968×968.

Appearance of the fuzzy target at $(18m, 10m, 0.5\pi)$ is shown in figure 12, and the detail of it is shown in the following table.

The membership value of the element $(20m, 20m, 0.25\pi)$ is 0.999. It is shown that aiming at this element (final target) is the most appropriate for task achievement.

Based on the acquired fuzzy target, the simulation was

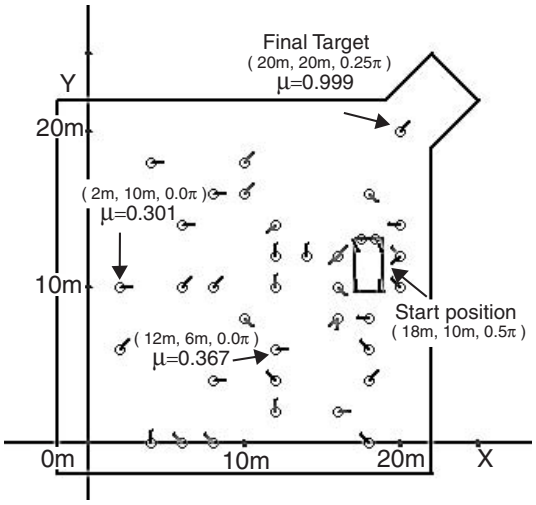


Fig. 12. Fuzzy target at the vehicle state (18m,10m,0.5π).

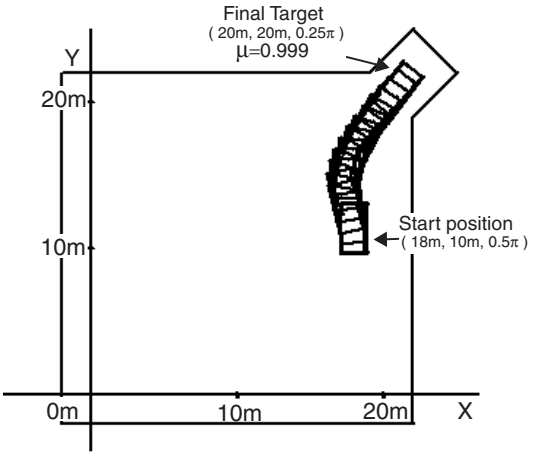


Fig. 13. Trajectory from (18m,10m,0.5π) to (20m,20m,0.25π) without obstacle.

performed about three cases: without obstacle (figure 13), with one obstacle in front of the initial position (figure 14), and with two obstacles (figure 15). Acquired fuzzy target is not arranged for using each case. Even when a situation changed, the same fuzzy target is used.

In the case of no obstacle (figure 13), it was able to reach at the final target directly. The elapsed time until reaching the final target was 27 seconds.

On the other hand, in the case with one obstacle (figure 14), it is not able to reach at the final target directly because of the obstacle. Moreover, because of the nonhoronomic characteristic of four-wheeled vehicle, the obstacle cannot be evaded by moving in parallel.

Under such a condition, the control instructions were calculated by using fuzzy target at the initial point $(18m, 10m, 0.5\pi)$, and selected the control instructions which move to the suitable element $(12m, 6m, 0.0\pi)$ in the current situation. Fuzzy target was set again after it approached the vicinity of $(12m, 6m, 0.0\pi)$. And the control instruction which move to the suitable element $(6m, 2m, 0.0\pi)(\mu = 0.548)$ in that by the retreat was selected. Afterwards, the obstacle was avoided and it changed the state which could be moved directly to final target $(20m, 20m, 0.25\pi)$. Finally, the car was reached to the final target $(20m, 20m, 0.25\pi)$. The elapsed time until reaching the final target was 100 seconds.

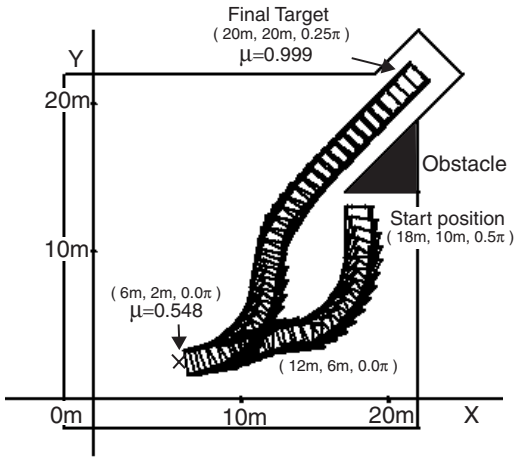


Fig. 14. Trajectry from (18m,10m,0.5π) to (20m,20m,0.25π) with one obstacle.

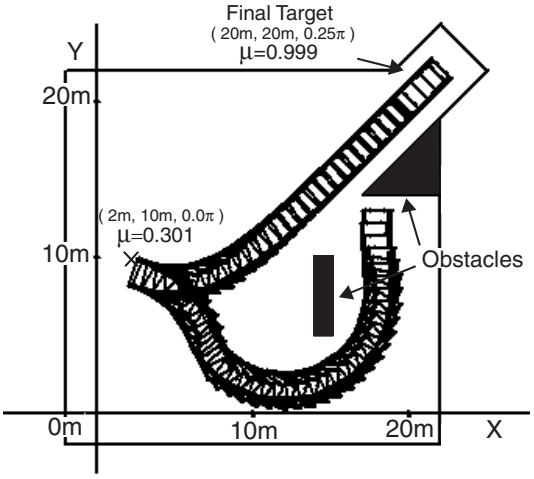


Fig. 15. Trajectry from (18m,10m,0.5π) to (20m,20m,0.25π) with two obstacles.

In the case with two obstacles (figure 15), the obstacle is added to the case of figure 14, and the moveable area was limited more. Even in such a complex situation, the obstacle was able to be evaded by using fuzzy target, and achieve the task. The elapsed time until reaching the final target was 121 seconds.

Thus, even when a situation changed, it was reached to the final target by using the fuzzy target which is acquired in the past. As a result, the effectiveness of the fuzzy target was confirmed by the simulation.

V. Conclusions

"Fuzzy target" was acquired by "Reinforcement learning", which is a machine learning method. It applied to the driving control system of the four-wheeled vehicle, which has nonhoronomic restraint. As a result, even when a situation changed, the vehicle was moved to the final target by using the fuzzy target that acquired in the past. In conclusion, the effectiveness of the fuzzy target that acquired by the reinforcement learning was confirmed by the simulation.

References

[1] L. A. Zadeh: "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-3, No. 1, pp. 28-44, 1973.

[2] 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.

[3] S. Yasunobu and T. Matsubara: "Fuzzy Target Acquired by Reinforcement Learning for Parking Control,"*Proc. of SICE Annual Conference 2003 in Fukui (SICE2003)*, pp.1303-1308, 2003.

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

[5] 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.

[6] 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.