PSP-learning Design of Hierarchical Intelligent Controller for Nonholonomic Vehicle

Yaya SURYANA, Seiji YASUNOBU and Noriaki SUETAKE
Institute of Engineering Mechanics and Systems
University of Tsukuba
Ibaraki 305-8573, JAPAN
yasunobu@esys.tsukuba.ac.jp

Abstract - In this paper, the profit sharing plan (PSP) is used to learn targets for achieving final target of nonholonomic vehicle moved in narrow area. The simulation result showed the effectiveness of the proposed method in generating control knowledge that could be time consuming and difficult to acquire by heuristic method.

Keywords— **PSP-learning, hierarchical intelligent control, nonholonomic vehicle, narrow area.**

I. INTRODUCTION

Control system design for nonholonomic vehicle is not an easy task especially in narrow area. Nonholonomic vehicle is not only nonlinear, but also there is kinematics constraint in the system [1]. However, if we observe human driver, it seems not so difficult for him to drive vehicle even if it is, for example, to park a vehicle in parking lot. Therefore, human control strategy is likely to be a suitable candidate for designing control system of nonholonomic vehicle. However, heuristic method is difficult to apply directly without the availability of operator or expert knowledge. Moreover, multilevel controller is needed to control nonholonomic vehicle in narrow area in order to overcome kinematics constraint [2][3]. There are some researches of motion planning of vehicle like searching a feasible path by dividing convex space in a workspace [4], controllability of mobile robot using search based method to find path in some space with obstacles [5], smooth trajectory toward goal using Lyapunov function and linierization techniques [6]. However, those techniques are not readily available for motion planning that need multiple switches for moving forward and backward like parking vehicle in narrow area.

In this paper, the hierarchical intelligent controller for nonholonomic vehicle is designed based on PSP-learning [7]. In this hierarchical intelligent control, PSP-learning is used as path planner. Several targets are made for path planning. Those targets are learned by PSP-learning and called strategy target. Strategy target is attained using several tactical targets. In order to speed up the training, cascade fuzzy controller [8] is used to attain the tactical target. At the end of training, predictive fuzzy controller [9] is applied to make a smooth movement of the designed controller. The validity and effectiveness of the proposed method are verified by computer simulation.

II. THE PROBLEM OF CONTROL SYSTEM FOR NONHOLONOMIC VEHICLE

The basic equation for nonholonomic system, subject to n generalized configuration variables $(q_1, \cdots, q_n)$ and m constraints can be described as follows [10]:

$$M(q)\ddot{q} + V_m(q\dot{q})\dot{q} + F(\dot{q}) + \tau_d = B(q)\tau - A^T(q)\lambda, \quad (1)$$

$$A(q)\dot{q} = 0, \quad (2)$$

where $M(q) \in \Re^{(n\times n)}$ is a symmetric, positive definite inertia matrix. $V_m(q\dot{q}) \in \Re^{(n\times n)}$ is the centripetal and coriolis matrix. $F(\dot{q}) \in \Re^{(n\times 1)}$ denotes the surface friction. τ_d denotes bounded unknown disturbances including unstructured unmodelled dynamics. $B(q) \in \Re^{(n\times p)}$ is the input transportation matrix. $\tau \in \Re^{(p\times 1)}$ is the input vector. $A(q) \in \Re^{(m\times n)}$ is the matrix associated with the constraints. $\lambda \in \Re^{(m\times 1)}$ is the vector of constraints forces. Most nonlinear system can be solved using piece-wise linierization, however for nonholonomic system piece-wise linearization cannot be applied because the system state q is nonintegrable at input vector τ.

The kinematics model of nonholonomic vehicle in two degrees of freedom can be written as follows [5]:

$$\frac{dx}{dt} = v\cos\phi\cos\theta, \frac{dy}{dt} = v\cos\phi\sin\theta, \frac{d\theta}{dt} = \frac{v}{L}\sin\phi, \quad (3)$$

where v is the speed of the vehicle, θ is the direction of the vehicle, ϕ is the steering angle, L is length of the vehicle. However, using only this equation is not guarantee for achieving trajectory planning [6], especially for a difficult situation like parking vehicle in narrow area. Therefore, multilevel controller in the form of hierarchical intelligent controller is proposed.

III. PSP-LEARNING TO ACQUIRE VEHICLE DRIVER STRATEGY

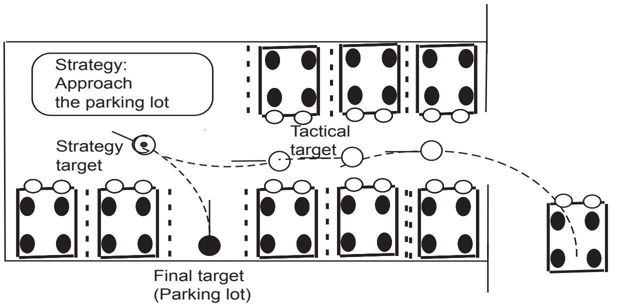

Fig. 1. Human control strategy in parking vehicle

The human control strategy for controlling nonholonomic vehicle can be illustrated by taking example of parking vehicle as in figure 1. First, it can be observed when the driver wants to park a vehicle. He can understands the situation around his vehicle and the current state of the vehicle (*e.g.*: position, direction, etc.). After that, the driver will compare the current state of the vehicle and the parking lot. Finally, the driver will set the target by considering the characteristic of the vehicle when he is driving a vehicle toward the final target (the parking lot).

The vehicle control method as describes above can be illustrated as follows.

1. The shape of the parking site, obstacle, etc. are confirmed.
2. The strategies such as *approach the parking lot* or *enter the parking lot* are mapped out by setting the relative position between the parking lot and the current state.
3. Compare the current states with the strategy target and create the tactical target.
4. Attain tactical target by moving the vehicle based on experiences.

Using this way of strategy, the vehicle parking decomposed the strategy into three targets: the strategy target, the tactical target and the final target. The strategy target is a general situation target. The strategy target is decided from a relative relation to the current state of the vehicle in order to reach the final target. In purpose to control the vehicle toward the strategy target, the tactical targets are created. In other word, to decide the strategy target and to move the vehicle, the tactical targets are determined. The most difficult decision is in strategy target setting, because nonholonomic vehicle has kinematics constraint to attain the final target. PSP-learning is used for this purpose.

The basic concept of PSP-learning is that problem solving is divided into episodes where the reward is given to the active rules. The rules are like in classifier system basically in the form of *IF* $<$ *condition* $c_n>$ *THEN* $<$*action* $a_n>$. The condition part contain patterns that can be match to the detector message which indicate the observe state. On every step, the rules that match the current state make a bid that is depend on the strength of rule (S-value). The rule is active if it is winning bid competition during any step in episode. The reward is given at the end of episode and distributed to the active rules (see figure 2).

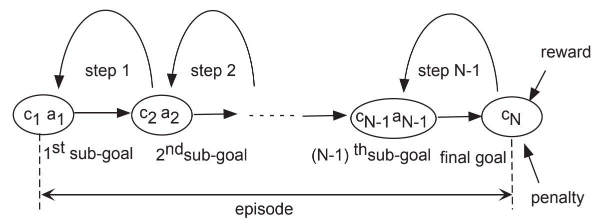

Fig. 2. PSP-learning distributes reward or penalty to the previous fired rules

IV. OUTLINE OF HIERARCHICAL INTELLIGENT CONTROLLER

A. *Control System Outline*

The outline of the hierarchical intelligent control system using PSP-learning for strategy target setting, fuzzy control for tactical target setting and predictive (cascade) fuzzy control for auto driving part can be seen in figure 3.

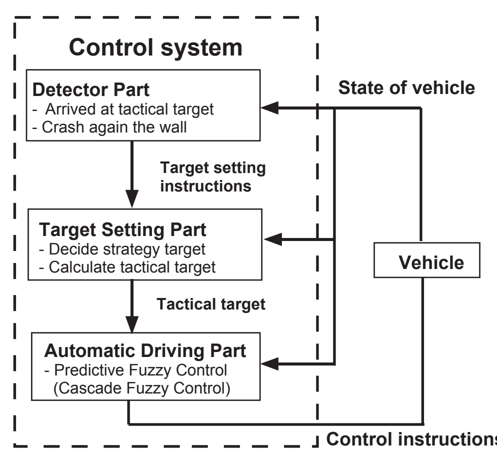

Fig. 3. Outline of the hierarchical intelligent controller

B. *Detector Part*

The detector part detects the current state of the vehicle and the obstacles (see figure 4). A new tactical target is needed when contact to the obstacle is predicted. Moreover, a new tactical target for automatic driving part is needed if the current tactical target was attained. The attainment of the tactical target determined by passing the line. This line is longitudinal with the tactical target. When the condition of detector part is as described above and a new tactical target is necessary, the target set instruction is sent to the target setting part. Finally, when the strategy target and the tactical target reached the final target, the detection procedure is finished.

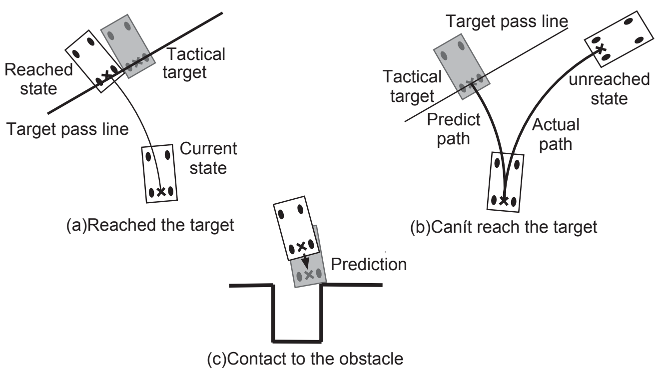

Fig. 4. Detection of the obstacle and target pass line

C. *Target Setting Part*

In the target setting part, current state of the vehicle is compared to the final target when the target setting part received instruction from the detector part. After that, the strategy is selected using PSP-learning method. The strategy can be categorized as follows:

- Change the direction: the direction should be change in order to reach the final target. This change of direction can be to horizontal, make counter direction from current direction or do the same direction with current direction.
- Approach the parking lot: approaching the parking lot means to find the nearest and easiest way to reach the parking lot.
- Enter the parking lot: entering the parking lot is done when the strategy target is almost same as the final target.
- The control is finished: this condition is satisfied when the vehicle reached the final target.

In order to attain the strategy target, the tactical target is calculated. After that, the state of the tactical target is sent to the automatic driving part.

C.1 *Strategy target setting by PSP learning*

The strategy target is decided based on the rule strength table (S-table) learned by PSP-learning.

The basic PSP-learning [7] is modified and employed to overcome the difficulties for parking vehicle in narrow area. The modified PSP-learning is used to learn rule strength table (S-table) for determining strategy target. S-table represents rule strength of (c_n,a_n) pairs which is the center of partitioned state space to reach final target and selected in each step. The selection of strategy target is based on maximum S-value otherwise roulette wheel selection is applied in low S-values. Each path from initial position to final target is one episode.

The reward is distributed to the selected S-values at the end of episode. The reward is the difference between maximum time (200 sec.) allowed to attain final target and time instance needed so far. For example, if the final target attained in 80 seconds then the reward r_i is 120. Penalty is given to overtime or crash to the wall where the value of r_i is -10. Overtime is the time needed to reach final goal exceeding maximum time.

The input states are partitioned state space in three dimension x, y and θ. x and y coordinated partition are labeled in empty space every 2 m from final target position and θ are in -0.75π, -0.5π, -0.25π, 0, 0.25π, 0.5π, 0.75π, π. There are 78 $\times$8 = 624 labeled partitions. The strategy targets $a_n = (x_T, y_T, \theta_T)$ for taking action are also 624. Final target is in (0,0,0.5π).

The S-table is updated by PSP-learning using the algorithms as follows.

1) For $1 \leq i^{th} \leq$ max episode, repeat 2)-8).
2) Initialize state c_1.
3) For step $1 \leq n^{th} \leq$ max step, repeat 4)-7).
4) If current state c_n is final target, then give r_i= (maximum time - time), go to 8).
5) If crash to wall or overtime, r_i=-10, go to 8).
6) Select strategy target a_n by max S or roulette wheel.
7) Use fuzzy control to attain strategy target a_n.
8) Update selected S-values using the equation:

$$S(c_n, a_n) \leftarrow (1-\alpha)\ S(c_n, a_n) + \alpha r_i \gamma^{N-n}, \qquad (4)$$

where α is learning rate and γ is discount factor and N is the last step.

C.2 *Tactical target setting*

The tactical target is needed in order to move the vehicle to reach the strategy target. Firstly, the vehicle is moved approaching the strategy target. Then, the vehicle is moved forward or backward to reach the strategy target. In order to achieve that condition, the tactical target as a stepping-stone is calculated using the equations according to rules applied for each condition. These equations are based on rough characteristics of the vehicle which tracks the trajectory made using circle. In the rules, two variables are evaluated, one is error between the current state and the tactical target. The other is the direction error between the current state and the tactical target. Those linguistic variables are applied to infer the tactical target such as follows:

Rule 1: If *the distance is small* and *the direction is positive small*, then *approach target line in forward direction.*

Rule 2: If *the distance is small* and *the direction is negative small*, then *approach the target line in backward direction.*

$\vdots$

Rule p: If *the distance is very small* and *the direction is very small*, then *tactical target is equal to final target.*

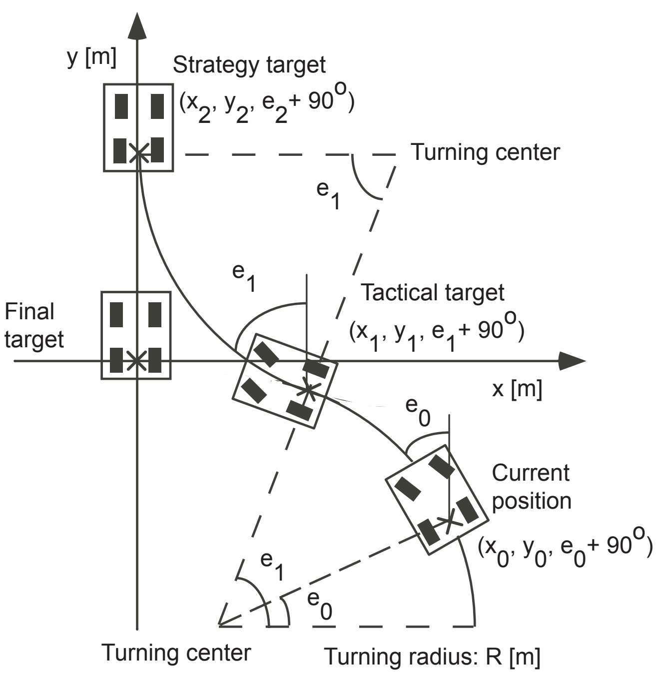

Fig. 5. Tactical targets to attain strategy target

For example, when the strategy target is ((0,0), +90^o) and strategy target rule 1 is selected, the tactical target is calculated as follows:

RULE 1 : IF X_error is *Small* AND θ_error is *Positive Small* THEN approach target line in forward direction (state 1).

The state 1 (see figure 5) is calculated using the equations bellows:

$$
\begin{aligned}
x_1 &= \frac{1}{2}(x_0 + R(1-\cos(e_0))),\\
y_1 &= y_0 + R(\sin(e_1) - \sin(e_0)),\\
e_1 &= \arccos(\frac{1-\cos(e_0)}{2} - \frac{x_0}{2R}),
\end{aligned} \tag{5}
$$

where (x_0, y_0) is current position, (x_1, y_1) is calculated tactical target position, e_0 is direction difference between current state and final target, e_1 is direction difference between tactical target and final target and R is minimum turning radius of the vehicle.

D. *Automatic Driving Part*

The attainment to the tactical target is done using automatic driving part. However, in training PSP-learning as a machine learning techniques needs much time to learn strategy targets. Therefore, in order to speed up the training process, cascade fuzzy controller [8] is used. In application to make a smooth tracking to attain the tactical target, predictive fuzzy controller [9] [11] is used.

D.1 *Cascade Fuzzy Control*

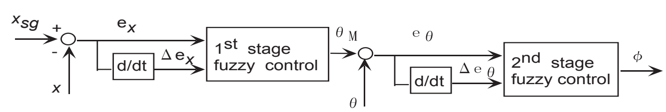

Fig. 6. Blok diagram of cascade fuzzy controller

Cascade fuzzy controller is used to move vehicle toward the tactical target. In cascade fuzzy controller IF-THEN rules are used for inference mechanism in two stages (see figure 6). For example:

First stage fuzzy controller:
Rule (1-1): IF x_e is Z and dx_e is NB THEN θ_M is PM.
Rule (1-2): IF x_e is PM and dx_e is NM, THEN θ_M is NB.

$\vdots$

Rule (1-q): IF x_e is PB and dx_e is NB, THEN θ_M is PM.

Second stage fuzzy controller:
Rule (2-1): IF θ_e is NB and $d\theta_e$ is Z, THEN ϕ is PM.
Rule (2-2): IF θ_e is NB and $d\theta_e$ is NM, THEN ϕ is PB.

$\vdots$

Rule (2-r): IF θ_e is NB and $d\theta_e$ is NB, THEN ϕ is NM, where x_e is the difference between x tactical target and x current state, $dx_e = x_e(t) - x_e(t-1)$, θ_M is output angle of first stage, θ_e is the difference between θ_M and

θ current state, $d\theta_e = \theta_e(t) - \theta_e(t-1)$, ϕ is steering angle. Z(*zero*), NB(*negative big*), NM(*negative medium*), PM(*positive medium*) and PB(*positive big*) are linguistic variables of fuzzy sets.

D.2 *Predictive Fuzzy Control*

Predictive fuzzy control (see figure 7) is used for the application of the designed controller after the knowledge is acquired.

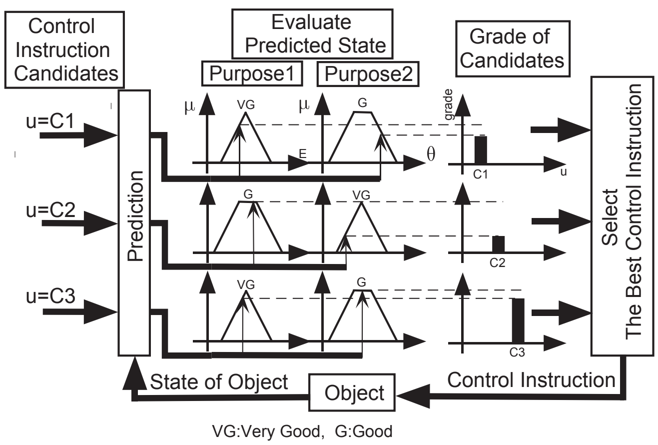

Fig. 7. Outline of predictive fuzzy controller

The control instruction candidates were predicted from the current state prediction by integrating equation (3).

The result of the predicted error to the target and the distance to the obstacle is evaluated. The grade of control instruction candidates is determined by the value of error. The smallest error has the highest grade. The candidate that has highest grade is assumed the best control instruction for the current state. The vehicle is controlled according to these control instructions.

V. SIMULATION RESULTS

The simulation is carried out with the characteristics of the vehicle as follows. The wheelbase is 2.6 m, distance between axis and bumper is 0.4 m, width is 1.7 m, the smallest turning radius is 6 m and the velocities are 0.4 m/s for moving forward, 0.0 m/s for stop and -0.4 m/s for moving backward. The coordinate of the parking site can be seen in figure 8.

The simulation was done as follows. Firstly, the initial state is made by driving the vehicle from 78 positions and 8 directions. After that, the intelligent control system start control the vehicle using strategy target and tactical target to reach the final target. The vehicle moved using cascade fuzzy controller to reach tactical targets. In this simulation, the entire procedure to park a vehicle from

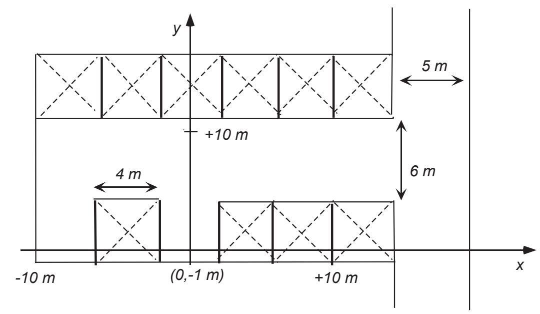

Fig. 8. The dimension of parking lot

initial state to go to the final target must be less then 200 seconds. If the system takes more than 200 seconds to reach the final target, it is defined as failed.

The performance of the controller can be seen in figure 9 for 1000 trials. The strategy target in the S-table is updated based on simulated annealing where at high temperature, roulette wheel selection (random search) is used for 9 trials (it is the lower part of the graph in figure 9) and after that the temperature is freezing. The greedy search is applied to find the maximum value in the subsequent trial (1 trial). It can be seen from figure 9 that the hierarchical intelligent controller design based on PSP-learning can park the vehicle to final target about 160 episodes in one trial.

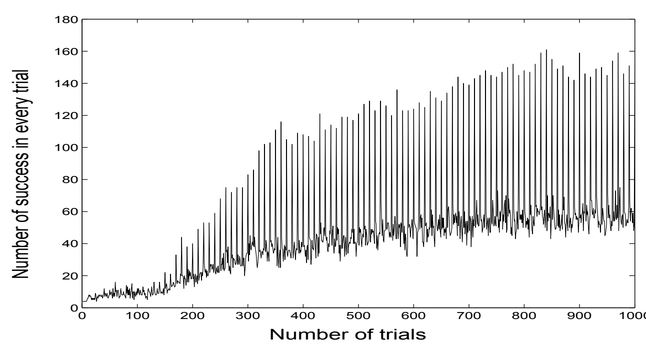

Fig. 9. The performance of PSP-learning

The examples of trajectory can be seen in figures 10 and 11. In figure 10, the first strategy target learned by PSP-learning from initial state (18.0 m, 0, 0.5π) is at position (-6.0 m, 6.0 m, π). In order to reach that strategy target, the tactical target is set to position (12.0 m, 5.9 m, π). After that, the vehicle moved using predictive fuzzy controller to reach the tactical target. The detector part detects "can't reach the target (see figure 4)", therefore the signal sends to the target setting part to set another strategy target. The next strategy target, which is same with tactical target, is at position (-6.0 m, 6.0 m, π). In order to reach that position, the vehicle moves

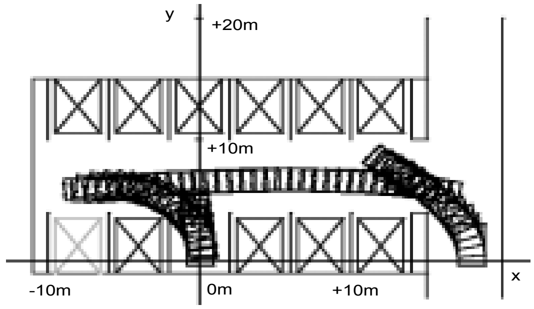

Fig. 10. An example of parking trajectory from (18.0m, 0, 0.5π) to (0, 0, 0.5π)

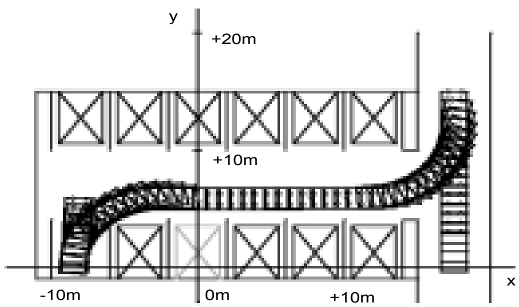

Fig. 11. An example of parking trajectory from (18.0m, 0, 0.5π) to (-8.0m, 0, 0.5π)

backward and after that moves forward. After reaching this position, the PSP-learning set the next strategy target at position (0, 0, 0.5π). The tactical target is at position (0, 0.07 m, 0.5π). However, the obstacles make it not possible to go to directly to this point. The vehicle move forward and then backward but it is not yet reached final target. The next strategy target and tactical target is same with final target. It is hard to reach this point, therefore the strategy target is still at position (0, 0, 0.5π) but tactical target is set to (0, 3.0 m, 0.5π). Finally, the vehicle moves backward to reach position (0, 0, 0.5π) where by that time strategy target, tactical target and final target are at the same position. After reaching this position the control procedure is ended for one episode.

In figure 11, the vehicle move from initial state (18.0 m, 0, 0.5π) to strategy target and tactical target position (18.0 m, 12.0 m, 0.5 π). From this point, the vehicle moves backward to the next strategy target (2.0 m, 6.0 m, 0) via tactical target (12.0 m, 6.0 m, 0). The next strategy target is set to (-8.0 m, 0, 0.5 π) and tactical target (-2.0 m, 5.8 m, 0) then (-8.0 m, -0.2 m, 0.5 π). After that, the vehicle tried to reach final target, but it has to move forward to the tactical target (-8.0 m, 3.0 m, 0.5 π) and then backward to final target.

The simulation results showed that the controller designed using PSP-learning could park nonholonomic vehicle in narrow area without requiring prior knowledge that could be time consuming and difficult to acquire by heuristic method.

VI. CONCLUSION

In this paper, PSP-learning design method for hierarchical intelligent fuzzy controller is proposed and applied to nonholonomic vehicle for parking in narrow area. PSP-learning as path planner learned targets called strategy targets to achieve final target. Strategy target is attained using tactical targets. The simulation results showed the validity and effectiveness of the proposed method in generating control knowledge without prior knowledge available that could be time consuming and difficult to acquire by heuristic method.

References

[1] R.W.Brockett,"Asymptotic stability and feedback stabilization," Differential Geometric Control Theory, Birkhauser, pp.181-191, 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.873-8781, 1996.

[3] R.M. Murray, S.S. Sastry, "Nonholonomic Motion Planning: Steering using Sinusoids," IEEE Transc. on Automatic Control vol.38, no.5, pp. 700-716, May 1993.

[4] H.A. Vasseur, F.G. Pin, J.R.Taylor, "Navigation of a Car-like Mobile Robot Using Decomposition of the Environment in Convex Cells,", Proceedings of the 1991 IEEE International Conference on Robotics and Automation, pp.1496-1502, USA, April 1991.

[5] J. Barraquand, J.C. Latombe,"Nonholonomic Multibody Mobile Robots: Controllability and Motion Planning in the Presence of Obstacle,"Proceedings of the 1991 IEEE Conference in Robotics and Automation, California, pp.2328-2335, April 1991.

[6] J.B. Pomet, B.Thuilot, G. Bastin, G. Campion, "A Hybrid Strategy for The Feedback Stabilization of Nonholonomic Mobile Robots,", Proceedings of The 1992, IEEE International Conference on Robotics and Automation, pp.129-134, France, May 1992.

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

[8] Y.Huang, S.Yasunobu,"Cascade Fuzzy Control Using to Large Scale System," Trans. IEE of Japan, vol.119-C, no.12, pp.1548-1553, Dec., 1999.

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

[10] J.I. Neimark, N.A. Fufaev, "Dynamic of Nonholonomic System," AMS, 1972.

[11] S. Yasunobu, "Fuzzy Engineering", Syokodo, 1991.