1-210　fPSP-強化学習による駐車支援知識の獲得

Fuzzy-Reinforcment-learning Design of Intelligent Controller for Nonholonomic Vehicle

安信誠二 (筑波大学　機能工学系)

Seiji YASUNOBU, Institute of Engineering Mechanics and Systems, University of Tsukuba, Tsukuba-shi, 305-8573 JAPAN

In this paper, a fuzzy profit sharing plan (fPSP) reinforcment-learning is proposed and it used to learn a knowlege of sub-target for achieving final target of nonholonomic vehicle moved in narrow parking area. The simulation results showed the effectiveness of the proposed method in generating control knowledge that could be time consuming and difficult to acquire by heuristic method.

Key Words: Reinforcment-learning, Fuzzy evaluation, Intrlligent control, Vehicle

1. は　じ　め　に

四輪自動車は，ハンドルと速度の 2 つの操作だけで，車体の位置と向きの 3 状態を制御する，ノンホロノミックな特性を持つ熟練と経験を要する運転である[(1)(2)]．特に，駐車場などの限られた領域への駐車を考えた場合，熟練した運転者は，駐車場の状況を把握し，適切な運転を行なえるが，非熟練者は，立ち往生することも多い．しかし，駐車場の誘導員の指示により，駐車目的の達成が可能である．

知的自動車駐車システムは，この熟練運転者や誘導員の車の特性を踏まえた運転知識を組み込み，滑らかに車を自動運転するすると共に，人間が運転する場合にも適切な支援情報を提供するシステムとして開発が進められている[(5)(6)]．しかし，車の駐車状況（空いてる場所には侵入可能など）や車体特性（最小旋回半径など）により駐車知識が異なり，その獲得が課題である．

一方，最適解への勾配情報を用いずに，人間が試行錯誤的に行動を選択し，成功を学習している仕組みを取り入れた強化学習[(7)]の考え方が提案されている．また，この強化学習の手法の一つとして，何段階かの行動の後の成功報酬を各段階にさかのぼって分配する PSP(Profit Sharing Plan)-学習の手法[(8)(9)]が提案され，駐車知識の獲得に適用してきた[(10)(11)]．しかし，連続した位置と方向に移動する自動車を離散的に取り扱った場合，意図しない移動をした場合でも，報酬が高い場合があり，適切な知識の獲得が困難であった．

本論文では，ノンホロノミックな特性をもつ自動車の駐車制御を行う知的自動車駐車システムにおいて，設定に経験と時間を要する駐車知識の設定に，PSP-学習をファジィ評価により拡張した手法（fPSP-学習）を適用し，運転知識の獲得を試みる．

2. 自動車駐車システムの問題

自動車は，図 1 に示すような幾何学的な動きをする．左右前輪の角度 (ϕ_L,ϕ_R) は，四輪全てが旋回中心と直角に旋回半径 R で動くように，アッカーマン・ジャントウの操舵機構により構成されている．

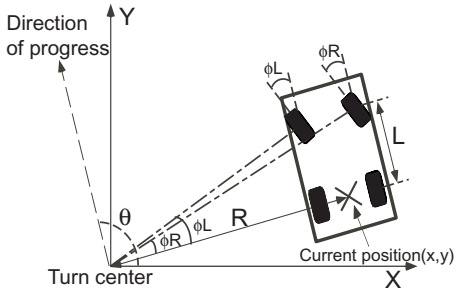

Fig. 1　Kinematics constraint in nonholonomic vehicle

この自動車の動きは，以下の拘束条件式（1）によって記述できる[(2)]．

$$\begin{aligned}\frac{dx}{dt} &= v\cos\phi\cos\theta,\\ \frac{dy}{dt} &= v\cos\phi\sin\theta,\\ \frac{d\theta}{dt} &= \frac{v}{L}\sin\phi\end{aligned} \tag{1}$$

ここで，x,y は自動車の位置，θ は x 軸と自動車の進行方向のなす角，v は車の速度，ϕ は旋回半径 R での動きを指令するハンドルの切れ角，L は車のホイールベース（長さ）である．

この自動車の駐車運転は，最終目標（所定の駐車スロット）に到達することが目的となる．しかし，上記の特性のため，その近くに居ても，車の向きや位置が異なると簡単には移動できず，単純に到達度を求めることができない．そのため，目標状態に到達するような制御指令を決定することが困難である．特に車の最小旋回半径が制御性能に直接関連するような，狭い領域での駐車は難しい．比較的広い駐車場を想定した場合は，追従すべき経路と走行パターンを余裕を持って設定可能であり，それに PID 制御などで追従することも可能であるが，この経路計画だけでも難しい問題である．

これらを解決するため，到達可能な目標を与える部分と，その目標への自動車を操作する，階層型知的制御系を提案し駐車制御を実現している[(5)(6)]．

3. 人間の駐車知識と fPSP 学習

3.1　人間の駐車知識　　ノンホロノミックな特性を持つ自動車をうまく駐車させている人間の運転方法を図 2 のような駐車場を例として考える．

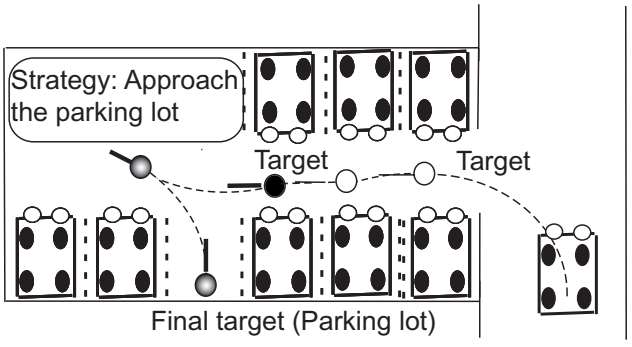

Fig. 2　Human control strategy in parking vehicle

熟練者の駐車手順は，以下のように記述できる．

1.　駐車場の形状や，他のスペースの塞がり状況を把握し，駐車場と現在の車位置から,「車庫に寄せる」,「向きを変え

る」,「切り返し位置へ移動する」「車庫に入れる」「制御を終了する」など，大体実現可能な大局的な目標を考える．
さらに，そこに到達するために，ノンホロノミックな車両特性を考慮し，一回のハンドル操作で到達可能な，局所的な目標を決める．

2. 目標に到達できるよう，ハンドルと速度を操作する．この操作は，基本的には，一定のハンドル切れ角，速度で実現できるものである．しかし，障害物や壁の状況などにより少しの調整や危険回避の行動をとる．
3. 目標に到達したか，到達が困難だと判断した時，現在の目標をリセットし考え直す．

この様な人間の運転方法に基づき，知的自動車駐車システムが構築されている．大局的な戦略目標が根底にあるが，そのとき時に到達すべき，局所的戦術目標が適切に設定されれば，車は移動することができる．本論文では，ファジィ評価を取り入れた fPSP-学習の手法により，この駐車知識の獲得を試みる．

3.2 *fPSP*-学習 PSP-学習[8]は，離散的な状態遷移に対して適用されている場合が多い．ここでは，連続値の状態の制御である車の運転に適用し，各ステップの目標達成度をファジィ評価し報酬の反映に用いる．

この fPSP-学習に関する基本概念を図 3 に示す．行動知識を，”IF <condition:c_n> THEN<action:a_n>” なる形式の状態と行動のペアからなる規則（S-table）で記述する．

ある初期値からの制御開始をエピソードと呼び，各ステップ n において，行動知識に基づいて，現在状態 p_n（連続値）の地点 c_n（離散値）に最も適切な（または，ランダムな）行動 a_n を選択する．選択した行動を実行した結果の状態 p_{n+1} の地点 c_{n+1} において，同様に次の行動 a_{n+1} を選択する．もし，最終目標に到達した時，報酬 R を要したステップ数で γ の割合で減じながら，そのエピソードで使用した N 個の規則へ反映していく．途中で，走行不可能になった場合は，報酬（罰）を負とする．また，第 n ステップでの目標到達の度合いを，図 4 に示すファジィ集合に基づき，式 (2) により評価して報酬を減じる．

$$\mu(p_{n+1}-a_n) = \mu_d x(\Delta x_{n+1}) \wedge \mu_d y(\Delta y_{n+1}) \wedge \mu_d \theta(\Delta\theta_{n+1}) \quad (2)$$

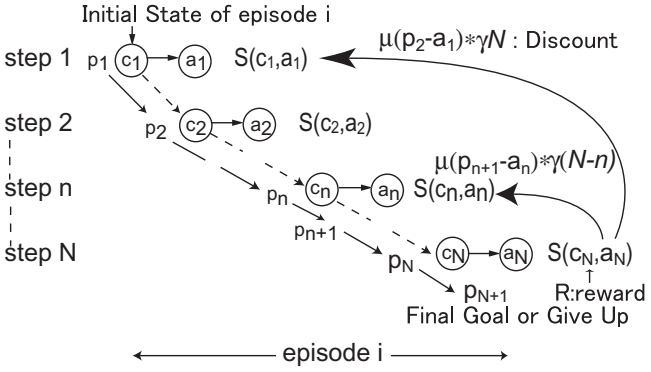

Fig. 3 PSP-learning distributes reward or penalty to the previous fired rules

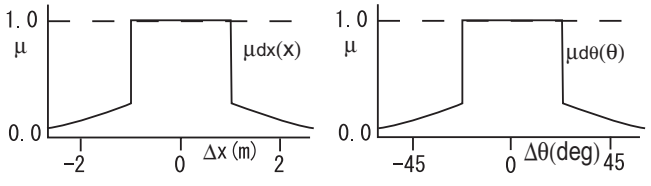

Fig. 4 Fuzzy set of position and angle error

4. 知的自動車駐車システムの概要

4.1 システムの概要 目標への到達を判断する状況監視部，運転知識に基づき目標を設定する目標設定部，目標へ車を操縦する自動運転部からなる知的自動車駐車システム[5][6]に対して，実行結果に基づき目標の駐車知識を獲得するファジィ強化学習（fPSP 学習）部を付加して，今回構築したシステムの概要を図 5 に示す．

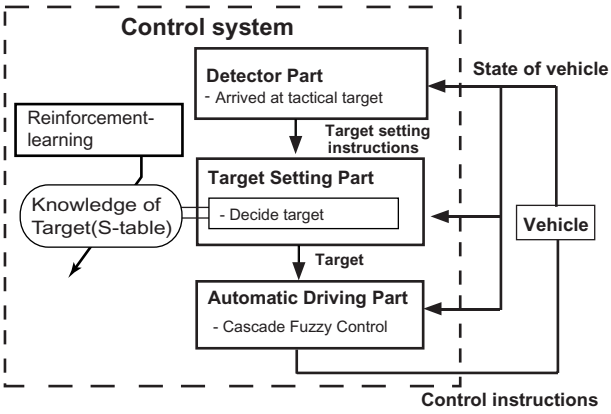

Fig. 5 Outline of the hierarchical intelligent controller

4.2 状況監視部 車は，設定した目標へ到達するよう制御される．現在の目標に到達した場合には，新たな目標が必要である．ここで用いる目標の決定では，車の特性や壁などを考慮していない．現在の目標に到達できなかった場合には，その目標をあきらめる．この状況監視部では，(a) 目標と直角に引いた線より前方に出る，(b) 予定した距離を走行した，(c) 指定した方向へ（障害物に当たり）進めない，場合に，現在の目標をリセットする．

4.3 目標設定部 目標設定部が状況監視部から目標を変えるよう指示を受けたとき，新たに目標を設定する．新しい目標は，fPSP-学習法により獲得した知識（S-table）を使用し選択する．人間の駐車知識[5][6]は，以下のように記述でき，これに対応した知識の獲得を目指す．

- 車庫に寄せる: 所定の駐車位置（車庫）に到達するために，現在位置に最も近くて簡単な目標を設定．
- 向きを変える: 車の向きを変えて車庫に入れやすくするため，一旦，本来進むべき方向とは異なる方向に目標を設定．
- 切り返し位置へ移動する： 次の目標設定で切り返し（逆の進行方向に移動）するため，その開始位置を目標に設定．
- 車庫に入れる: 車庫に入れるため，最終目標を目標に設定．
- 制御を終了する: 車が最終目標に到達し，車庫入れが完了．

4.3.1 fPSP 学習による目標設定 本論文で用いる fPSP-学習の駐車知識 (S-table) は 2m 四方と 45 度刻みに仕切られたセルの中心値をラベルとする (c_n，a_n) の組み合わせである．獲得した運転知識の使用は，現在地点 c_n で最大の S-table 値を持つ目標 a_n が目標として選択される．各初期位置から最終目標までの経路を，1 つのエピソードと呼ぶ．

駐車知識を獲得するため，各地点において目標をルーレットによりランダムに選択し，未知の経験を試みる．実行した結果の報酬は，エピソードの終わりに，選択された S-table 値に分配される．最終目標に達することができた場合，報酬は制限時間 (250 秒) と所要時間の差で与える．例えば，最終的な目標に 80 秒で達したならば，出発地点で用いた知識に対する報酬 r_i は 170 にステップ数で割り引いた値である．さらに，各ステップで，目標への到達をファジィ評価する．制限時間内に到達できない時や，障害物などで動きが取れない状況に陥った場合，報酬（罰）r_i を-10 とする．

車が動くのは，三次元（x，y，および θ）の状態空間である．x，y に関しては，最終目標から 2m 毎の格子点を中心とし，各位置における車の向き θ は，0，0.25π，0.5π，0.75π，π，1.25π，1.5π，1.75π，の 8 方位に区分しラベル化している．例えば，5 章の図 7 で示す駐車場の状況では，78 個の格子点が置かれ，78×8=624 個のラベルがある．また，行動選択のための目標 $a_n = (x_T, y_T, \theta_T)$ も 624 である．最終目標は (0m,0m,0.5π) である．

S-table 値（駐車知識）を更新する fPSP-学習のアルゴリズムを以下に示す．

1) 第 i エピソード (i=1,$Imax$) を， 2)-8) で実行．

2) 地点 c_1 にエピソード初期位置 p_1(x_1,y_1,θ_1) をセット．

3) 第 n ステップ (n=1,$Nmax$) を, 4)-7) で実行．

4) もし，第 n ステップが終了し、現在状態 $p_n + 1$ が最終目標のとき，報酬: R_i=$Tmax$-t_n, N=n とし，エピソードを終了 8) へ．

5) もし，動けないまたは $Tmax$ を過ぎた時，報酬 (罰):R_i=-10, N=n とし，エピソードを終了 8) へ．

6) 最大 S-table 値または，ルーレット選択により，目標 a_n を選ぶ．

7) ファジィ制御により，次の目標 a_n へ移動を試み,$p_n + 1$ に到達する．次のステップ実行のため 3）へ．

8) 次式 (3) により，使用した N 個の S-table 値を更新した後，次のエピソードの実行のため 1) へ:

$$S(c_n, a_n) = (1-\alpha)\, S(c_n, a_n) + \alpha \mu_{(p_{n+1}-a_n)}(R_i + t_n)\gamma^{N-n} \quad (3)$$

ただし，R_i はエピソード i での報酬，$Tmax$ は最大時間，t_N は制御終了時刻，α は学習率，γ は割引率，N は使用ステップ個数，$Nmax$ は最大ステップ数，$Imax$ は最大エピソード数である．

4.4 自動運転部　位置と車体方向で与えられる目標 (x_T,y_T,θ_T) に達するように，車を運転するため，カスケードファジィ制御[10]を用いている．ファジィ制御では，目標方向が y 軸となるよう座標系を変換し，目標 X_T との x 軸方向の偏差 x_e から，現時点の目標角度 θ_M をファジィ推論し，さらに車体角度の誤差 θ_e からハンドルの操作角度 ϕ をファジィ推論により求めている（図 6）．

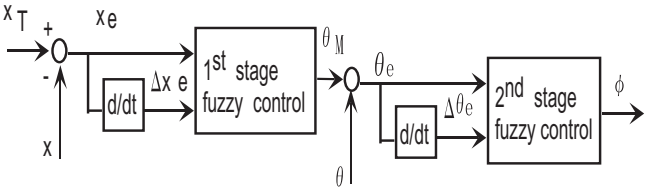

Fig. 6　Blok diagram of cascade fuzzy controller

目標角度推論（第 1 段目）のファジィ制御則は，

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.

...

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

ハンドル操作推論（第 2 段目）のファジィ制御則は，

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.

...

Rule (2-r): IF θ_e is NB and $d\theta_e$ is NB, THEN ϕ is NM,

の形式である．ここで x_e の変化を $dx_e = x_e(t) - x_e(t-1)$, θ_e の変化を $d\theta_e = \theta_e(t) - \theta_e(t-1)$ とし, ϕ はハンドル操作角である．Z(*zero*), NB(*negative big*), NM(*negative medium*), PM(*positive medium*) そして，PB(*positive big*) は，各状態を評価するファジィ集合である．

5. シミュレーション結果

シミュレーションに用いた自動車の諸元は，通常の乗用車を想定し，ホイールベース：2.6m，車幅：1.7m であり，移動速度は，前進，後進共，0.4m/s である．また，駐車場の形状は，5m の道路から，6 台 ×2 列の駐車スペースを設定した（図 7）．

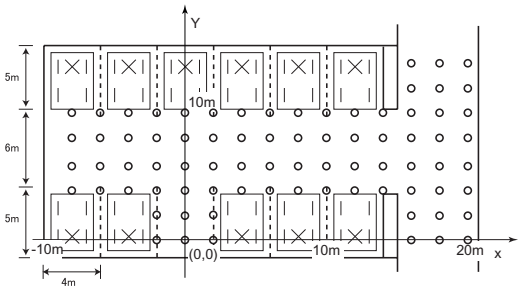

Fig. 7　The dimension of parking lot

ここで，駐車知識を獲得するための初期状態は，図 7 に丸で示す，2m 刻みの 78 個の格子点と 45 度刻みの 8 方位であり，624 エピソードを行う．この全エピソードについて，2000 回のトライアルを行い，駐車に成功したエピソード数の状況を図 8 に示す．各エピソードの初期状態で，車が壁に接触せず，移動可能なほとんど場合の 160 のエピソードで成功している．

このトライアルは，50 回のルーレット（ランダム）選択による冒険（図 8 の下部）と 50 回のグリーディ（最大値）選択（図 8 の上部）による知識の洗練，の繰り返しにより行った．また，学習率 α は 0.5，割引率 γ は 0.8 とした．獲得した運転知識（S-table：状態 c_n に対する目標 a_n で Q 値が最も大きい箇所をプロット）を，図 9 に示す．

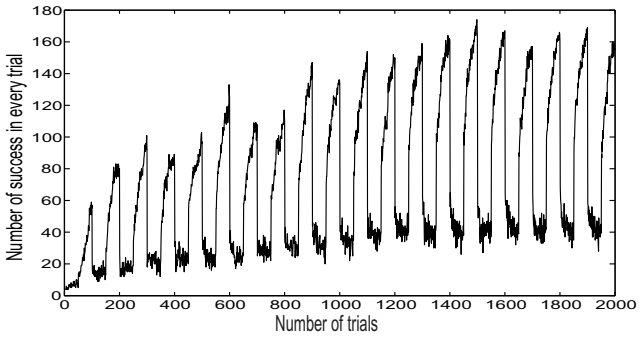

Fig. 8　The performance of PSP-learning

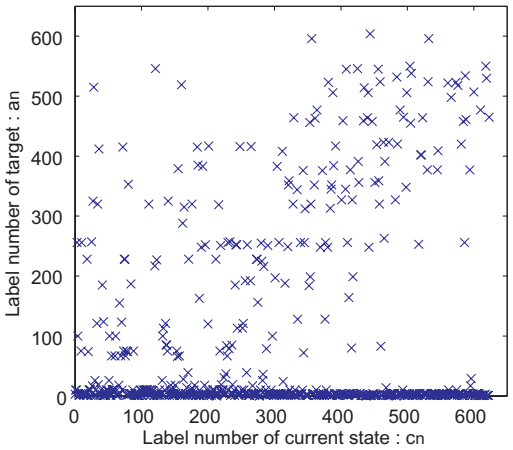

Fig. 9　Obtained drive knowledge (S-Table)

図 10 に，ここで獲得した運転知識を用いて，駐車場の中ほどの位置に 1 つの空きがある場合を想定し，獲得した運転知識を用いた走行（グリーディ選択）の様子を示す．初期状態は，駐車場の外の (18m,0m,0.5π) である．この状態から最初の目標として駐車スペースの奥の (-6m，8m，π) を目標として走行し，最終目標 (0m,0m,0.5π) に後退により進入し終了している．これらの運転知識は,「切り返し位置へ移動する (-6m，8m，π)」,「車庫に入れる (0m,0m,0.5π)」，といった知識に意味付けることができる．

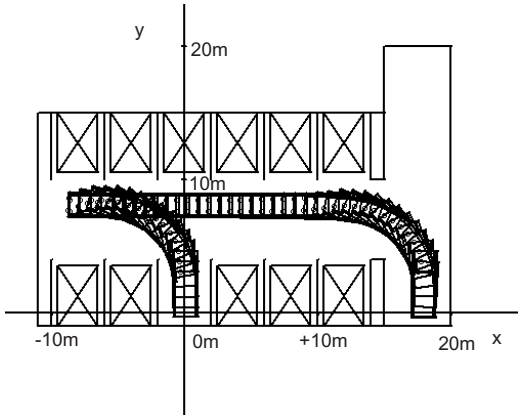

Fig. 10　An example of parking trajectory from (18.0m, 0, 0.5π) to (0, 0, 0.5π)

図 11 は，逆向きの初期状態 (18m,14m,1.5π) から駐車場へ進入し，駐車スペースの近くに寄せるため (-6m,8m,π) に目標を設定し，その後最終目標 (-8m,0m,0.5π) に移動させている．

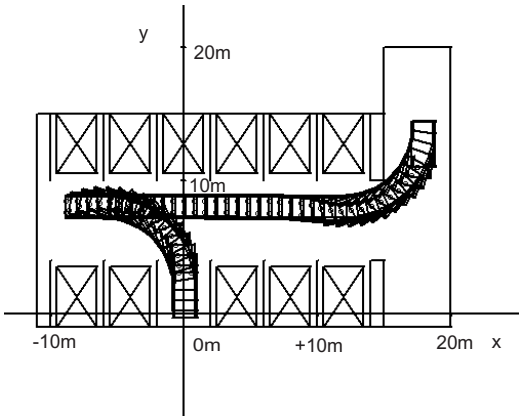

Fig. 11　An example of parking trajectory from (18.0m, 14.0m, 1.5π) to (0, 0, 0.5π)

このシミュレーションによる知識獲得では，報酬として出発してから駐車までの所要時間を用い，ランダム選択と最大値選択（グリーディ選択）により，局所最適解からの脱出を試みているが，さらに，図 9 に示した知識のグループ化によるまとめなどの必要がある．また，各ステップの評価において，設定した目標への到達度をファジィ評価している．これにより，目標に到達できなくても偶然よい場所に行ったため，高い報酬を得て，知識として生き残る可能性が少なくなっている．

図 4 に示したファジィ集合による評価では，評価値の低い部分を広げている．これは，実数値を扱う場合には重要であり，特に，初期状態から知識獲得が動き始めるのに重要な働きをしている．今回のシミュレーションでは，全く先見的知識が無いとしたため，知識獲得に時間を要している．しかし，少しの共通的知識を用いることにより，早く良い解を得ることが可能である．

6. お　わ　り　に

本論文では，ノンホロノミックな特性を持ち運転が難しい自動車の駐車制御に対して，実行結果から得た報酬を，途中の行動知識にファジィ評価を行いながら反映する fPSP-学習法を適用し，知的自動車制御システムの運転知識の獲得を試みた．具体的駐車場を想定し，シミュレーションを行った結果，適切な運転知識を獲得できることを確認した．

ここでハンドル，速度操作に用いたファジィ制御は，人間と同様に地点と方向で与えられた目標に車を移動させる．この機能は，人間の操作を代替するものであり，ここで用いた駐車システムは，人間を運転部分に組み込んで構成することが可能であり，現在，適用実験を進めている．ここで獲得した知識を用いることにより，駐車場の誘導員が行っているように，状況に柔軟に対応し，未熟な運転者などに適切な支援をすることが可能である．

本研究の一部は，日本学術振興科学研究費補助金基盤研究 (C)「福祉車両操作の知的運転支援システムの研究」(課題番号 12650407) の支援によるものである．

参 考 文 献

(1) R.M. Murray, S.S. Sastry: Nonholonomic Motion Planning: Steering using Sinusoids, IEEE Transc. on Automatic Control, vol.38, no.5, 700/716 (1993).

(2) J. Barraquand, and 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, 2328/2335 (1991).

(3) 安信誠二: ファジィ工学, 1/177, 昭晃堂 (1991).

(4) S. Yasunobu and S. Miyamoto: Automatic train operation by predictive fuzzy control, Industrial Application of Fuzzy Control (M. Sugeno, Ed.), North Holland, 1/18 (1985).

(5) 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, 873/8781 (1996).

(6) 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, 309/314 (2000).

(7) Richard S. Sutton and Andrew G. Barto (三上貞義，皆川雅章 訳): REINFORCEMENT LEARNING: An Introduction（強化学習）, 森北出版，1/351 (2000).

(8) J.J. Grefenstette: Credit Assignment in Rule Discovery System Based on Genetic Algorithms, Machine Learning 3, Kluwer, 225/245 (1988).

(9) 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, 645/653 (1999).

(10) Yaya Suryana and Seiji Yasunobu: Hierarchical Intelligent Control by PSP-learning and Cascade Fuzzy Control for Parking Vehicle in Narrow Area, T.IEE Japan, Vol.122-C, No.2, 315/316 (2001).

(11) 安信誠二，ヤヤ・スラヤナ，末竹規哲: 知的自動車駐車システムの強化学習による運転知識獲得, 第 29 回知能システムシンポジウム, pp.99/104 (2002).