

ロボットの行動獲得のための状態空間の自律的構成

浅田 稔^{*1} 野田 彰一^{*2} 細田 耕^{*1}

Action-Based State Space Construction for Robot Learning

Minoru Asada^{*1}, Shoichi Noda^{*2} and Koh Hosoda^{*1}

Robot learning such as reinforcement learning generally needs a well-defined state space in order to converge. However, to build such a state space is one of the main issues of the robot learning because of the inter-dependence between state and action spaces, which resembles to the well known “chicken and egg” problem. This paper proposes a method of action-based state space construction for vision-based mobile robots. Basic ideas to cope with the inter-dependence are that we define a state as a cluster of of input vectors from which the robot can reach the goal state or the state already obtained by a sequence of one kind action primitive regardless of its length, and that this sequence is defined as one action. To realize these ideas, we need many data (experiences) of the robot and cluster the input vectors as hyper ellipsoids so that the whole state space is segmented into a state transition map in terms of action from which the optimal action sequence is obtained. To show the validity of the method, we apply it to a soccer robot which tries to shoot a ball into a goal. The simulation and real experiments are shown.

Key Words: Segmentation, Inter-dependence between state and action spaces, State space construction, Hyper ellipsoid

1. はじめに

動的な実世界でタスクを遂行することを学ぶ自律ロボットを実現することは、ロボティクスとAIの中心課題の一つである。近年、環境とエージェントの相互作用を通して学習する手法として強化学習 [1] や記憶に基づく学習法などのロボット学習 [2] が注目されている。これらのロボット学習法の魅力は、環境やロボット自身の前提知識を必要としないことであるが、(1) 基本的に全探索で、多大な数の試行を必要とし、学習時間が長いこと、(2) ロボットが正しく学習できる状態と行動の空間が必要であること (マルコフ性の仮定) など、拘束や欠点も多い。本来、ロボットがタスクを遂行するために必要十分な情報を含む状態空間の構成は、ロボットの行動能力に依存する。また行動空間もロボット自身の知覚能力に依存し、相互に規定しあう。この問題は「鶏と卵」問題 (Fig.1参照) に類似し、非常に困難な問題である。

この問題に対し、行動空間を先に固定して、状態空間を構成する手法が提案されている。Chapman and Kaelbling [3] は、TVゲームの主人公が敵と戦って目的を達成するタスク設定で、

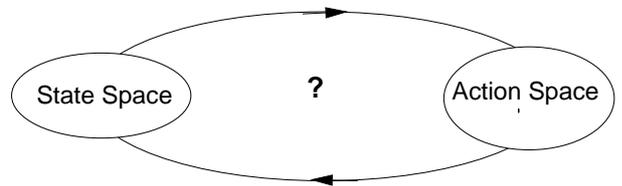


Fig. 1 The inter-dependence between state and action spaces

「敵を撃つ」、「障害物を回避する」などの構造化された行動をもとに、「敵が部屋にいる」、「ドアが開いている」などの既に抽象化された状態の真偽 (オン/オフ) をビット列とする入力ベクトルを、報酬をもとに分割する手法を提案している。しかしながら、もとの状態が既に抽象化されており、一般的なセンサの実数値の連続空間を対象とする問題には適用できない。

連続空間を対象としたものとして、Dubrawski and Reingnier [4] や Kröse and Dam [5] らは、移動ロボットの障害物回避のためのソナー情報の抽象化手法を提案している。また、佐藤ら [6] は、移動ロボットから得られる視覚情報を多変量解析し、障害物回避のための状態構成問題を扱っている。これらの手法では、障害物回避などの反射的なタスクを想定しているので、行動の物理的単位が固定されていても問題は生じにくい。視覚情報などを利用して遠くにある目標物に到達するタスクなどを想定した場合、同じ物理的動作が、画像上で異なる変化を引

原稿受付 1996年6月14日

^{*1}大阪大学工学部

^{*2}大阪大学工学部、現在日立製作所(株)に勤務

^{*1}Osaka University

^{*2}Osaka University, currently with HITACHI, Ltd.

き起こし[†], 正しく学習可能な状態と行動を定義することは難しい. 浅田ら [7] は, この問題を「状態と行動のずれ問題」と呼び, これに対し, 状態空間を先に固定し, 「ずれ」が生じないような行動空間を再構築したが, 最初に設計者が与えた状態空間がロボットにとって最適である保証はない.

そこで, 本論文では, 状態と行動空間を同時に構成するために, まず物理的に定められた単位時間に発生されるモーターコマンドの結果として生じるロボットの動作を「行動要素」と定義し, この行動要素に基づいて, 以下の基本的な考えに従い, センサーから得られる入力ベクトルを再帰的に分割する手法を提案する.

- 目標状態や既に獲得された状態に, 一種類の行動要素の可変時系列で到達できる入力ベクトルの集合を状態と定義する.
- 「状態遷移を引き起こす同一の行動要素の時系列」を一つの行動とする.

近年, 環境との相互作用により, 「知能発現」機構の解明を目指す動きがロボティクスやAIの研究分野で活発になりつつある [8]. 本論文はAIやロボティクスの基本問題の一つである「世界の分節化」問題 [9] に対して, 環境との相互作用により, これを解決する一手法を示すものと位置づけられる. すなわち, 設計者が, 設計者の視点でタスク遂行に必要な状態・行動空間をロボットに与えるのではなく, ロボット自身が自らの経験(環境との相互作用)を通して, 与えられたタスクを遂行するために必要な状態・行動空間を構成する. そして, その構成過程が行動学習に対応し, 結果として目的行動が獲得される. 以下では, まず状態・行動空間の自律的構成法を説明する. 次に, 簡単なシミュレーション結果を示す. 更に, 視覚に基づく移動ロボットに適用した場合のシミュレーション及び実機の実験結果を示して, 本手法の有効性を検証する. 最後に, 本論文で提案する手法の背景にある著者らの考え方を示し, 今後の課題を呈示する.

2. 状態・行動空間の自律的構成

ロボットが識別する状態空間^{††}を, 設計者が適当に分割しても, それがロボットにとって最適な分割になっている保障はない. 本来, 識別されるべき異なる状態が同じ状態としてクラスタリングされていると, 各状態での状態遷移のばらつきが大きく, 目標状態への到達が不確実になる. 逆に, 不必要に分割された状態空間の場合, 状態数が多くなり, 学習に時間がかかる. そこで, ロボットが自らの経験を通して状態空間を構成することにより, この問題の解決を図る.

状態空間は, 与えられたタスクに応じて, 状態遷移が行動とできる限り1対1に対応するように構成すべきである. そのために, Fig.2に示すような, 目標状態と行動に基づいた状態空間を考える. 目標に到達するための行動要素が同一の入力ベクトルをまとめて一つの状態とし, そのような状態で繰り返される同一行動要素の時系列を行動と定義する. 目標状態

に一つの行動で到達できる状態をその行動の種類ごとに, S_1^k , $k = a_1, a_2, a_3, \dots, a_n \in A$ (行動要素集合) とし, それらの集合を S_1 とする. さらに, S_1 に一つの行動で到達できる状態集合を S_2 とする. 同様に, 目標状態に最低 m 個の行動で到達できる状態集合を S_m とすると, 目標に到達可能なものは, いずれかの状態集合に含まれることになる. このような状態空間が構成されていれば, 行動と状態遷移が1対1に対応することになり, パフォーマンスの向上が期待される. 以下に, このような分割を実現するアルゴリズムを示す.

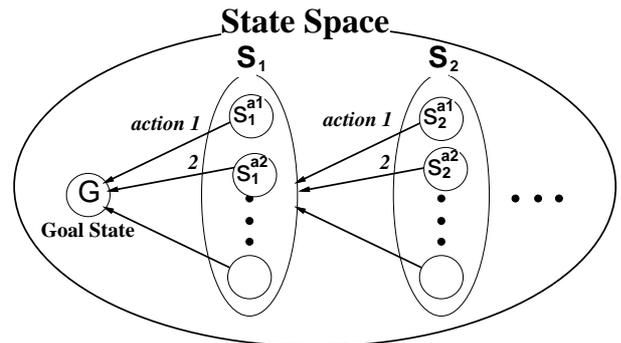


Fig. 2 Action-based state space construction

アルゴリズム

- (1) 目標状態を目標状態とする.
- (2) ランダムに行動する. 行動の定義から, 同一の行動要素を, 目標状態や既に分割された状態に到達するまでか, ある時間を経過するまで実行させ, 目標状態に到達可能な入力ベクトル x を蓄える. ただし, すでに区分された領域内にあるものは蓄えない.
- (3) 蓄えられた入力ベクトルを各行動ごとに状態として領域に区分する. 状態空間が m 次元の時, 状態の分布は m 次元の楕円体内で一様な分布とする. 状態変数ベクトル x の平均ベクトルを μ , 分散共分散行列を Σ とすると, 楕円体の境界は,

$$(x - \mu)^T \Sigma^{-1} (x - \mu) = m + 2$$

で表される [10].

- (4) 各行動で区分された領域の論理和をとった領域を次の目標状態とする. 重なる領域に対しては, 分散で正規化した距離 (マハラノビス距離)

$$\Delta = (x - \mu)^T \Sigma^{-1} (x - \mu)$$

の近い方をとる.

- (5) 目標状態に到達可能な状態変数が無くなれば終了, さもなくば (2) に戻る.

3. 実験結果及び考察

3.1 シミュレーション (I)

提案する状態空間の自律的構成法の有効性を示すために, 簡単なシミュレーションを行なった. シミュレーションは, 100×100 の空間の中心に直径 10 の円があり, 前後左右に毎ステップ 1.0

[†] 視覚センサの場合, 同じ動作を行っても, 観測者の近傍の変化は画面上で大きく写るが, 遠方では小さな変化としてしか捉えられない.

^{††} センサーから得られる入力ベクトルの張る空間をここでは想定

だけ進む行動をとることが出来るロボットの中心が円内に入ればゴールとする．初期配置はゴールの円内に入らぬようにランダム，ロボットがフィールドから出ればリセットされる．入力ベクトルはロボットの重心の絶対2次元座標 (x, y) である．

状態空間の分割結果を Fig.3 に示す．分かれた状態数は全部で12個であり，分かれた順が早いほど濃い領域となっている．また，図中の矢印は，その状態がどの行動でラベリングされて分割されたかを示している．1ステップ(1種類の行動要素)でゴールに到達できるのは，楕円領域 $S_1^{up}, S_1^{dn}, S_1^{lt}, S_1^{rt}$ で，それぞれ上下左右の行動に対応し，これらを要素とする集合が S_1 である． S_1 に1ステップで到達する状態は，8個に別れている部分(領域 $S_2^{up}, S_2^{dn}, S_2^{lt}, S_2^{rt}$) で，実際には4種の行動に対応する4つの楕円体の投影である．

この図の解釈は，以下の通りである．例えば，左上の領域 S_2^{dn} は，下向きの行動であるが，そのままに進むと右向き状態 S_2^{rt} に入るように見える．しかしこの領域は，同じ S_2 に属するので， S_1 に属する S_1^{rt} まで下向きに進み，そこで状態遷移を引き起こす[†]．このことは，同じ入力ベクトル(今の場合，座標)でも，自分の行動により異なる状態と見なすことを示している．人間でも，いつも同じ方向から進入する交差点を異なる方向から進入すると同じ交差点とそのとき認識できないことがあり，これに対応すると考えられる．もともと，図の左上の四隅は，下でも右でもよく，たまたま，楕円体中心からの距離で分割されているだけである．

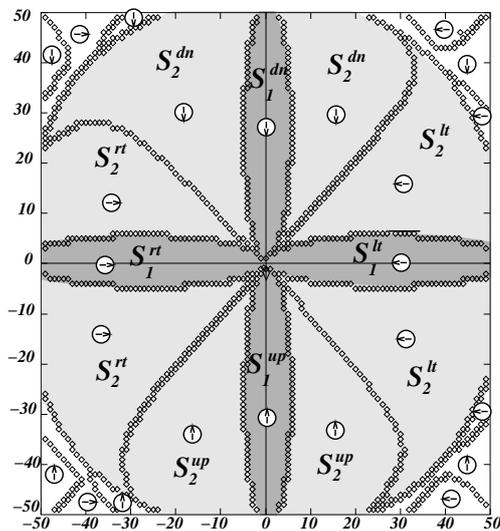


Fig. 3 Result of the variable action-based construction of 2D state space

同じ問題設定に対し，既存手法，例えば [4] [5] の手法の適用を考えると，一つの行動が固定の長さなので，その長さや，行動の速度，フィールドの大きさなどによって，状態数などの結

[†] この図の領域分割を厳密に捉えて，下向きから右向き，右向きから下向きを繰り返して S_2^{dn} と S_2^{rt} 境界線を移動してゴールした場合，状態遷移数は，多いが，実際のゴールへのステップ数は同じである．

果が大きく異なってしまふ．しかし，本手法のように，状態に依存した長さ可変の行動要素の時系列を行動と定義することにより，そのようなパラメータに依存せずに状態空間が構成される．このことは，環境の位相構造が相似していれば，絶対的なサイズにあまり左右されない状態・行動空間が本手法によって構成できることを示している．

3.2 シミュレーション (II)

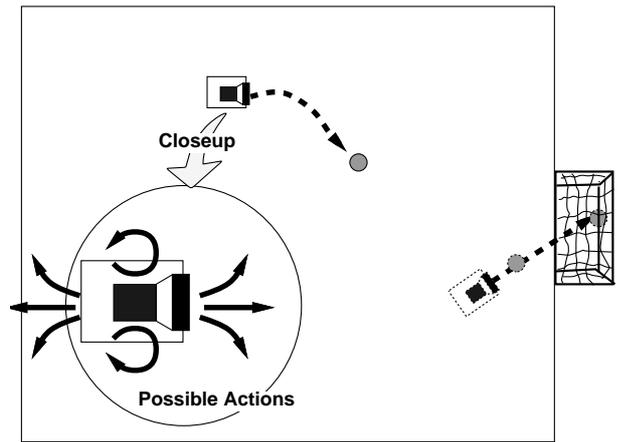


Fig. 4 Task

より複雑なタスクとして，ボールをゴールにシュートするサッカーロボットのコンピュータシミュレーションを行った．サッカーロボットとその環境を表したものを Fig.4 に示す．環境内にはロボットの他に，ボールとゴールしか存在しないものとする．ロボットが得られる情報は，ロボットに搭載されたカメラからのボールとゴールについての画像情報のみである．ボールやゴールの大きさや距離などの三次元情報，カメラパラメータ，ロボット自身の動特性などの先験的な知識は一切与えられていない．

ロボットは，左右の車輪が二つのモータにより独立に駆動される PWS (Power Wheeled Steering) システムを持っている．左右輪がそれぞれ，前進，停止，後進の3段階の速度を出すので，合計9通りの行動要素が選択できる．ただし，この内，両輪とも停止の行動要素は状態の変化をもたらさないで選択せず，残りの8通りの中から行動要素を選択する．

シミュレーション環境は 3.0×3.0 [m] の正方形のフィールドで，上辺の中央に幅 0.9 [m]，高さ 0.23 [m] のゴールがあり，全長 0.45 [m]，幅 0.31 [m] のロボットが直径 0.09 [m] のボールを蹴る．カメラはロボットの中央部についており，画角は 36 度である．ロボットの最大速度は 1.1 [m/s]，最大回転角速度は 4.8 [rad/s] である．ロボットの質量はボールと比べて十分大きいものとし，はねかえり係数は 0.5 とした．また，ボールの転がり速度は床との摩擦を考えて，各ステップ毎に 0.8 を掛けて減衰させている．また，画像処理による 33 [ms] の遅れおよび，モータの立上りの遅れ時間 100 [ms] を考慮している．

サッカーロボットの状態空間は，画面内のボールの位置，大きさ，ゴールの位置，大きさ，向きの5次元の入力ベクトルか

ら構成される．画像のサイズは 512×480 である．入力ベクトルの各要素が物理的にとり得る値の範囲は，

- ボールの大きさ (x_1): 直径に対応し, $1 \sim 271$ (全体像が観察されなくても, 円形により推定)
- ボールの位置 (x_2): $-274 \sim 274$ (ボールが画面の端に映っている場合, その中心位置を求める),
- ゴールの高さ (x_3): 左右のゴールポストの高さの平均. $1 \sim 480$,
- ゴールの位置 (x_4): 左右のゴールポストの x 座標の平均. $-256 \sim 256$,
- ゴールの向き (x_5): 左右のゴールポストの高さの差をゴールの高さ (x_3) で割ったもの. $-1.0 \sim 1.0$ (稀にこの範囲を超えることがある)

である (Fig.5参照) .

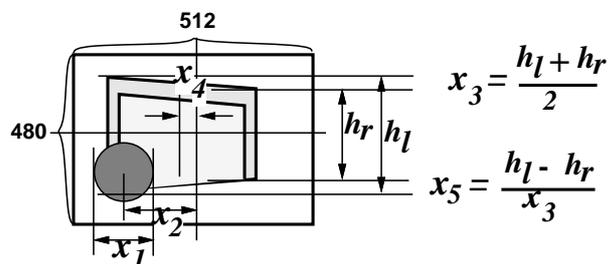
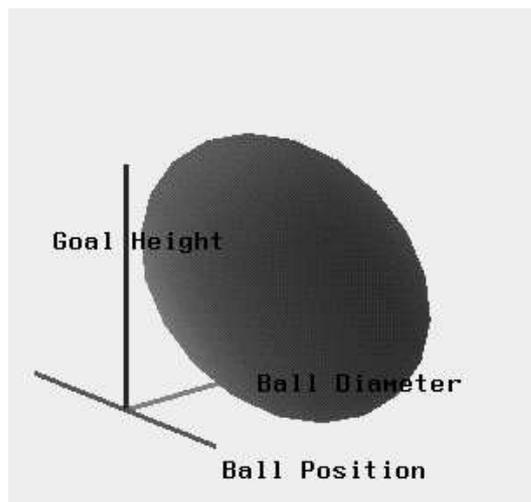


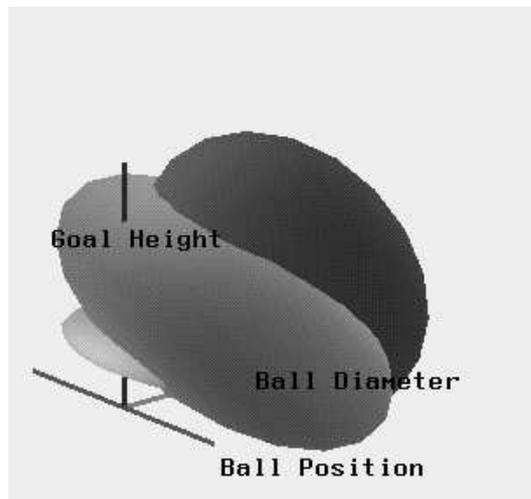
Fig. 5 Input vector consisting of five parameters

状態空間が構成されると同時に, 目標に到達する状態行動の遷移図が獲得されるので, これを用いれば目的行動が実現できる．実際には, このようにして獲得された状態以外の状態, 例えば, ボールしか観測されない状態, ゴールしか観測されない状態などが存在する．これらの状態から, 獲得された状態への遷移は, 文献 [7] に示すように, 一種の行動要素で到達可能な場合には問題ないが, ボールとゴールが離れていて, その間にロボットが存在し, 大きく回り込まないと, ボールとゴールが同時に観測できる状態にならない場合は, 対応できない．これは回り込んでいる時の状態が識別不能ないわゆる「隠れ状態」を含んでいるため, ここでは扱わない．

Fig.6に, ゴールの位置, 向きがいずれも0(真正面)での断面をとり, ボール位置, ボールとゴールの大きさの3次元で表現したものを示す．(a)では, 大きな楕円体 (S_1^F) が一つだけ得られ, 直進運動 (Forward) に対応している．(b)では, 二つの楕円体 (S_2^F と S_2^B) が追加され, それぞれ直進と後退運動 (Backward) に対応し, S_2 を構成している．Fig.7には, 最終結果をボールの大きさとゴールの高さの2次元空間 (ボールの位置は正面) に投影したものを示す．領域のラベルは, F, Bがそれぞれ前進, 後退運動を, 添字の数字がゴールの到達までの状態遷移数を示す．著者らの以前の研究 [7] では, 設計者が直接状態空間を分割しており, そのときの各状態は, この図で各軸に平行な長方形に対応し, 本手法で得られた形状と大きく異なる．この図で楕円体の集合で覆われない残りの部分は, 「正面にゴールが大きく, ボールが小さく見える」状態を表し, 本来オクルージョン



(a) the first stage ($S_1^F = S_1$)



(b) the second stage ($S_1 + S_2$)

Fig. 6 Construction process of the state space for the shooting robot

によって観測できない．以前の研究ではこのような意味のない状態も含まれていた．

Table 1 Comparison with existing methods

	Number of States	Search Time	Success Rate (%)
Previous work [7]	243	500M*	77.4
Proposed method	33	41M	83.3
cf. Fixed action length	107	222M	71.5

* indicates Q-learning time.

Table 1に既存手法との比較結果を示す．成功率は5000回の試行における結果であり, 状態数はボールとゴールの両方が観測されている時の数である．学習時間は, $33[\text{ms}]$ を単位とする

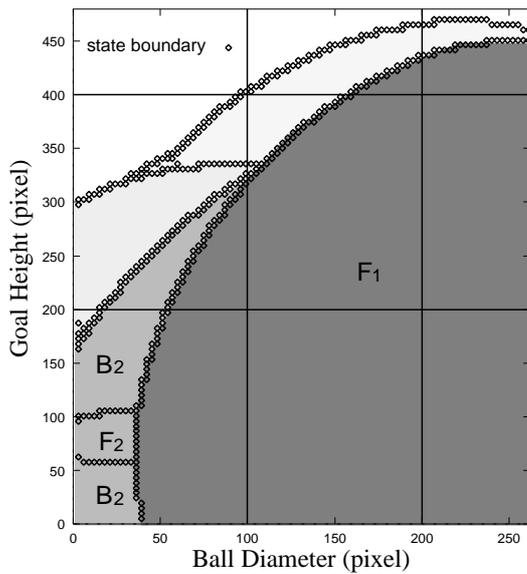


Fig. 7 2-D projection of the result of state space construction

ステップ数 ($M=10^6$) で、ボールとゴールがロボットの前方にランダムに配置された初期配置で行なった時のものである。以前の研究 [7] では、Q 学習の時間を示している。状態数で約 1/8、学習時間を含めた探索時間で約 1/12 と大きく改善されている。参考のために、既存研究などで採用されている固定長の行動系列を採用した場合に対応すると考えられる結果を最下段に示す。長さは、基本単位の 33[ms] を採用した。先の結果 (最上段 [7]) に比べ、状態数、探索時間とも半減しているが、成功率は改善されていない。これは、画像処理の遅れ (33[ms]) や、モーターの上りによる遅れ (約 100[ms]) などの影響を考慮したシミュレーションを実施したので、状態変化により、頻繁に行動要素が変更されるため、それらの影響を大きく受けていることを示す。これに対し、本手法は、同一行動要素を連続して実行することが多いので、時間遅れによる影響が少なく、実際の環境に適していると考えられる。

3.3 実ロボットによる実験

Fig.8に、実験で用いたロボットを示す。システム構成や画像処理手法などは、文献 [7] と同じである。実ロボットを用いた実験では、i) 実ロボットを動かしてその経験をサンプリングする。ii) WS上でサンプルした入力ベクトルのデータに基づいて提案する手法で行動・状態空間を構成する。iii) 得られた行動・状態空間及びその遷移図を用いて実ロボットを動かす。

(1) 実ロボットによる経験のサンプリング

まず、ロボットに単一の行動要素をとらせ続け、その時のボールの位置、大きさ、ゴールの位置、大きさ、からなる入力ベクトルをサンプリングした。サンプル数は約 20000 個 (10分強) であった。

(2) 状態空間の自立的構成

構成された状態空間を、ボール位置、ゴールの位置、向きがいずれも 0 (真正面) での断面をとり、ボールとゴールの大きさの 2次元で表現したものを Fig.9に表す。領域のラ

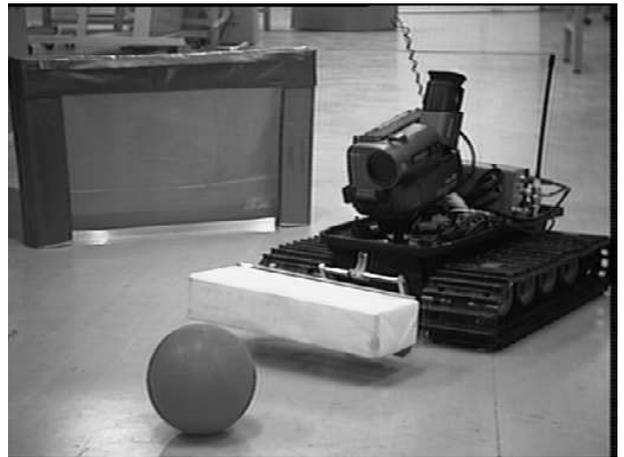


Fig. 8 Our real robot

ベルは、Fig.7と同じで、FLは、左前進を表す。実ロボットでは利用した画像プロセッサの能力の限界から、画像処理の際に面積を 1/16 に縮小するので、画像のサイズは 128×120 画素であり、各軸の値の範囲はシミュレーションの時の 1/4 になっている。状態は 10 個に分割された。状態数がシミュレーションに比べて少ないのは、経験数が少ないためと考えられる。

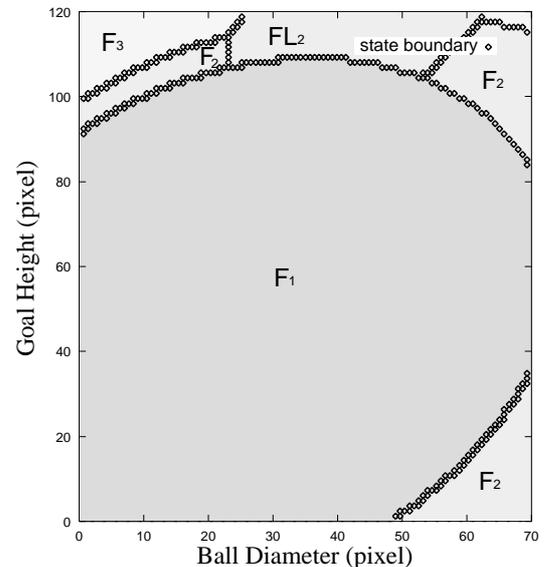


Fig. 9 Result of state space construction for the real robot

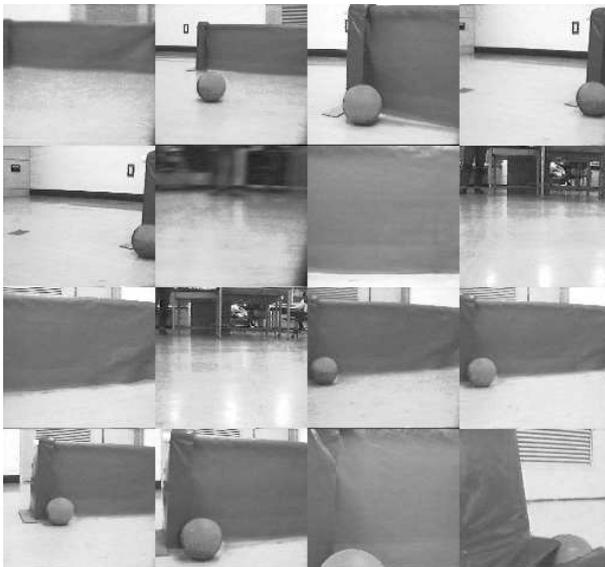
(3) 学習結果の実ロボットによる運用

(2) で得られた学習結果を実機に適用した。成功率はシミュレーションと比べて若干劣る。これは、ボールの重心が偏っているため思いがけない方向に移動したり、床面の滑べりで実際の動きが変化するなど、様々な要因が考えられる。実ロボットがシュートする様子の一例を Fig.10に示す。約 1 秒毎の動きを示している。(a) に、シュートの様子を、(b) にその時ロボットから見た環境の様子を示す。この例では、

最初シュートを試みたが、ゴールの左隅にボールが留まり、ゴールできていないので、左右に振りながらシュートできる地点まで後退し、再シュートして成功している。



(a) The robot succeeded in finding and shooting a ball into the goal



(b) Images taken by the robot during the task execution

Fig. 10 An example of real robot experiments

3.4 考 察

人間でもそうだが、経験に偏りがあると、変な癖がついて、適切な行動を獲得できないことがある。本手法でも、状態空間を構成する時に経験が偏っていると、楕円体のパラメータが変化し、状態数、探索時間にも影響を及ぼす。今回は、状態空間を構成する物体、すなわちボールやゴールが観測できる初期状

態をランダムに発生させているが、偏りがないという保証はない。なるべく偏りのないデータを獲得するには、一般により多くのデータを必要とし、より多くの時間を必要とする。偏りのないデータを効率よく獲得する手法が望まれるが、これは環境やロボット自身に対する事前知識とのトレードオフと考えられる。

ここでは、集中楕円体モデルを用いて状態空間を分割した。但し、実際の分布が、一様分布である保証はない。モデル中に含まれるべき空間が含まれなかったり、含まれるべきでない空間が含まれることは理想的には避けたいが、実際のロボットには様々な不確定要因が存在するので完全に誤差がなくなることはない。たとえ誤差の少ないモデルであっても、複雑な計算や多くのメモリを必要とするものは望ましくない。種々のタスクに本手法を適用するには、このようなトレードオフを考慮して最適なモデルを選ぶ必要がある。

4. おわりに

本稿では、経験に基づく行動・状態空間構成法を提案し、シミュレーション及び実ロボットを用いた実験でその有効性を示した。「はじめに」で述べたように、筆者らは、行動・状態空間構成問題を「世界の分節化」問題として位置づけている。そして感覚と行動を介して環境と相互作用(経験)できる物理的実体としてのロボットの性質を「身体性」と捉えれば、この「身体性」を有する事が、「世界の分節化」問題を解決できるための条件であると考えている[†]。本論文で示したきたように、ロボットの知覚能力や運動能力に基づいて、獲得された行動・状態空間(例えば、Fig.9は、そのような空間の投影像)は、ロボットがタスクを遂行する上でのロボット主体の環境に対する主観的表現に対応し、非常に限定されたものであるが、ロボットによる世界の見切り方、すなわちロボットの世界観となると考えられる。この世界観は、ロボット固有のものであり、これによりロボットが種々の環境に対して主観的判断を下す事、さらに経験(環境との相互作用)により世界観を発達させることも考えられる。すなわち、「身体性」により、ロボット主体の主観的価値基準が獲得される可能性があり、結果として表れる行為が「自律的」、「知的」と形容されると期待できる。

上で述べて来たように、「身体性」に基づいて、価値基準がロボットの内部に構築されることが重要であるという点は、最近、AIの分野でもこれからの研究の新たな展開のポイントとして指摘されている[11]^{††}。但し、本論文で提案した手法では、まだまだ多くの問題が残されている。特に、以下の三つが重要と考えている。

(1) ロボットに入ってくる画像情報は、色分けされたボールとゴールの画像で、それらの領域特徴を状態空間の軸とした

[†] 「世界の分節化問題」は、コンピュータビジョンの分野では、古くから「画像の分割」問題として取り上げられていたが、分割結果の評価が人間側の主観に委ねられ、その有効性や限界が明らかでなかった。これをロボティクスの視点から考えると、外界から得られるセンサーデータの分割(注意の視点)は、当然のことながら、ロボットの目的(タスク)、能力(感覚・運動能力及び処理能力)及び環境に依存し、その評価は、行動結果でなされるべきである。

^{††} 原文ではこれを「価値の主体内在性」と表現している。

が、一般には、選択すべき特徴を決定することは、より困難な問題である。通常の画像情報を対象とした時、そこからどのような特徴が、タスクを遂行する上で重要か、さらに選択されるべき特徴が状況に依存して変化する場合など、解決しなければいけない問題は多い。逆にいえば、全ての情報を利用することは、不可能なので、与えられた能力に依存して、それらを選択することが重要と思われる。例えば、単眼視における視覚追跡能力を元に行動獲得したり[12]、両眼視差と運動情報の獲得能力を元に行動を学習する手法[13]などが提案されている。結局、ロボットの能力に基づいて、利用可能な画像情報処理能力から何を選択するかを学習する機構の開発が重要と思われる。

- (2) 本論文では「隠れ状態」を扱わないと断ったが、「隠れ状態」問題を解決することも重要である。(1)が入力ベクトルの空間的な構造の複雑さへの対応を求めるの対し、これは時間軸での構造の複雑さへの対応を求める。ある変数に対して、何回までの時間微分が、そのタスクに有効かを検出する機構である。本質的問題は、時間軸も含めた入力ベクトルの選定問題と考えられる。
- (3) 提案した手法では、行動要素を単位として、その可変時系列で行動を定義し、その長さをパラメータ化することで、状態を分割すると同時に、行動を時間軸方向に抽象化したと考えられる。これは、移動ロボットの場合、行動自由度が2と少なく、空間的な抽象化を行う必要がないからである。一般には、空間的な抽象化(人間などは、物理的に多自由度でも、実際はかなり自由度を減らして制御しているので、そのような意味での抽象化)が、より重要と考えられ、今後、多自由度のロボットに対する運動空間を抽象化するための基本単位が何かを研究する必要があると考えられる。

謝 辞

本研究は、平成7年度文部省科学研究費重点領域研究(「創発システム」課題番号07243214及び「知能ロボット」課題番号07245214)の補助を受けた。

参 考 文 献

- [1] C. J. C. H. Watkins and P. Dayan. "Technical note: Q-learning". *Machine Learning*, Vol. 8, pp. 279-292, 1992.
- [2] J. H. Connel and S. Mahadevan, editors. *Robot Learning*. Kluwer Academic Publishers, 1993.
- [3] D. Chapman and L. P. Kaelbling. "Input Generalization in Delayed Reinforcement Learning: An Algorithm and Performance Comparisons". In *Proc. of IJCAI-91*, pp. 726-731, 1991.
- [4] A. Dubrawski and P. Reingnier. Learning to Categorize Perceptual Space of a Mobile Robot Using Fuzzy-ART Neural Network. In *Proc. of IEEE/RSJ/GI International Conference on Intelligent Robots and Systems 1994 (IROS '94)*, pp. 1272-1277, 1994.
- [5] B.J.A. Kröse and J.W.M. Dam. Adaptive state space quantisation for reinforcement learning of collision-free navigation. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems 1992 (IROS '92)*, pp. 1327-1332, 1992.
- [6] R. Sato, H. Ishiguro, and T. Ishida. State-space construction considering robot properties. 第13回日本ロボット学会学術講演会予稿集, pp. 453-454, 1995.

- [7] 浅田, 野田, 俵積田, 細田. "視覚に基づく強化学習によるロボットの行動獲得". 日本ロボット学会誌, Vol. 13:1, pp. 68-74, 1995.
- [8] 井上. 感覚と行動の統合による機械知能発現機構の研究. 第12回日本ロボット学会学術講演会予稿集, pp. 151-152, 1994.
- [9] 國吉. 「実世界エージェントにおける注意と視点-情報の分節・統合・共有-」. 人工知能学会誌, Vol. 10, No. 4, pp. 507-514, 1995.
- [10] H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, NJ, 1951.
- [11] 中島, 有馬, 佐藤, 諏訪, 橋田, 浅田. 「新しいAI研究を目指して」. 人工知能学会誌, Vol. 11, No. 4, pp. ??-??, 1996.
- [12] T. Nakamura and M. Asada. Motion Sketch: Acquisition of Visual Motion Guided Behaviors. In *Proc. of IJCAI-95*, pp. 126-132, 1995.
- [13] T. Nakamura and M. Asada. Stereo Sketch: Stereo Vision-Based Target Reaching Behavior Acquisition with Occlusion Detection and Avoidance. In *Proc. of IEEE Int. Conf. on Robotics and Automation*, pp. 1314-1319, 1996.

浅田 稔

年, 大阪大学基礎工学部助手。1989年大阪大学工学部助教授。1995年同教授。1997年大阪大学大学院工学研究科知能・機能創成工学専攻教授となり現在に至る。この間, 1986年から1年間米国メリーランド大学客員研究員。知能ロボットの研究に従事。1989年, 情報処理学会研究賞, 1992年, IEEE/RSJ IROS'92 Best Paper Award受賞。1996年日本ロボット学会論文賞受賞。博士(工学)。電子情報通信学会, 情報処理学会, 人工知能学会, 日本機械学会, 計測自動制御学会, システム制御情報学会, IEEE R&A, CS, SMC societiesなどの会員
(日本ロボット学会正会員)

野田 彰一

1971年1月19日生。1993年大阪大学工学部機械工学科卒業。1995年同大学院工学研究科博士前期課程修了。現在, 日立製作所勤務。在学中は知能ロボットの学習に関する研究に従事。

細田 耕 (Koh Hosoda)

1965年11月9日生。1993年京都大学大学院工学研究科機械工学専攻博士後期課程修了。同年大阪大学工学部助手となり, 現在に至る。1995年日本ロボット学会研究奨励賞受賞。視覚サーボ系, 知能ロボットの研究に従事。博士(工学)。IEEE, 計測自動制御学会の会員。
(日本ロボット学会正会員)