コンピュータビジョンと
イメージメディア 121－12
（2000．3．23）

行動獲得過程における視覚情報の自律的構造化：分節と統合

浅田 稔　内部 英治　細田 耕　鈴木 昭二

asada@ams.eng.osaka-u.ac.jp
大阪大学大学院工学研究科知能・機能創成工学専攻
〒565-0871 大阪府吹田市山田丘 2-1

本稿では，著者らが行ってきたロボットの行動獲得過程における視覚情報の構造化手法について，強化学習の枠組みでの状態（・行動）空間の自律的構成問題の観点から，この問題をとらえ直し，その意義を検討する．強化学習では，センサ出力値（画像データ）やモータへの入力値（運動指令）をそのまま状態・行動空間に射影する場合が多いが，実環境では，そのような状態・行動空間では学習できない．そこで，断片的視覚情報と運動情報をクラスタリングすることで，状態・行動空間を構成した手法，オンラインで視覚情報空間を分節・統合する手法，そして協調などのより複雑なタスクに対応するための状態ベクトル推定法に基づく手法を紹介する．これらの結果が示唆するものは，視覚情報の分節や統合の規範は視覚情報の中にあるのではなく，外部規範（行動基準）に基づいている点である．

Visual Feature Organization through Robot Behavior Acquisition

Minoru Asada, Eiji Uchibe, Koh Hosoda, and Sho'ji Suzuki

Dept. of Adaptive Machine Systems, Graduate School of Engineering
Osaka University, 2-1, Yamadaoka, Suita, Osaka 565-0871, Japan

This paper discuss the issue of visual feature organization in the context of state and action space construction needed in the framework of reinforcement learning, a typical method for robot behavior acquisition. Conventionally, sensor outputs (image data) and motor inputs (motion commands) are directly mapped onto the state and action spaces. However, behavior learning based on such spaces often do not work in real environments. To handle this problem, a method of clustering the instances of visual features and motor commands in the temporal domain, online segmentation of visual feature space, and state vector estimation for cooperative behavior learning have been proposed. In these methods, the principle of the visual feature organization is not inside the vision system but comes from the performance evaluation of learned behaviors.

1 はじめに

視覚情報の分節と統合に関する従来手法は，画像上のデータのある性質の差に起因する規範により処理が行われ，その評価はもっぱら，視覚情報の統計的性質や，設計者などの主観的判断に委ねられる場合が多い．視覚は，本来切り離された存在では無く，動的な環境との相互作用可能な，より大きなシステムの一部として位置づけられるべきであり [1]，そのシステムのパフォーマンスの評価に依存する知覚情報の構造化がなされるべきと考えられる．環境が静的で，システムに与えられたタスク用件が事前に明確にされている場合，専用マシンに特化したシステムとして，視覚情報の構造化は容易かもしれない．しかしながら，環境が動的に変化したり，システム自身の記述が不十分な場合，自身の知覚情報や運動情報をもとに目的を達成する行動を獲得しなければならない．この場合，視覚はシステムから切り離しが困難であり，行動獲得過程の中で，結果として視覚情報が構造化されていくと考えられる．

本稿では，著者らが行ってきたロボットの行動獲得過程における視覚情報の構造化手法について，強化学習の枠組みでの状態 (・行動) 空間の自律的構成問題の観点から，この問題をとらえ直し，その意義を検討する．強化学習では，センサ出力値 (画像データ) やモータへの入力値 (運動指令) をそのまま状態・行動空間に射影する場合が多いが，実環境では，そのような状態・行動空間では学習できない．そこで，断片的視覚情報と運動情報をクラスタリングすることで，状態・行動空間を構成した手法，オンラインで視覚情報空間を分節・統合する手法，そして協調などのより複雑なタスクに対応するための状態ベクトル推定法に基づく手法を紹介する．これらの結果が示唆するものは，視覚情報の分節や統合の規範は視覚情報の中にあるのではなく，外部規範 (行動基準) に基づいている点である．

以下では，まず，簡単に強化学習の説明をした後，我々が行ってきた行動獲得手法 [2, 3, 4] を紹介する．具体的な問題設定としては，ロボットによるサッカー競技 (通称ロボカップ [5]) を取り上げ，ドリブル，シュートなどの個々のロボットの行動，パス，センタリングなどの複数ロボット間の協調行動を視覚に基づいた学習によって獲得する．最後に，今後の課題を述べる．

2 強化学習の枠組

強化という言葉は元々行動心理学の用語として用いられていた．アメリカの行動心理学者の代表であるスキナーは，人間をはじめとする動物の行動を説明するための基本原理として「強化」による行動原理を唱え，種々の動物行動実験を行った．代表的な実験として，スキナーボックスによる鼠の行動実験を例に強化学習の基本的枠組みと，その問題点を簡単に説明しよう．

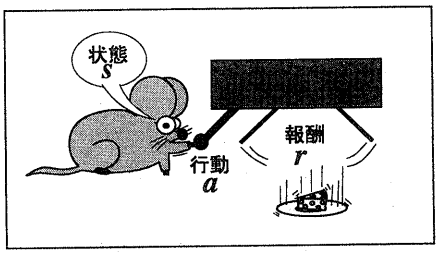


図 1: スキナーの鼠箱

スキナーボックスの実験では，鼠を箱の中に入れ，その中にあるレバーを鼠がたまたま押すと，餌がもらえる実験で，一旦レバー押しを憶えると何回もレバーを押し続ける行動をとる (図 1)．このときレバーを押す行為に正の強化 (餌，報酬，価値などなど) が与えられる．強化学習は，これを確率的動的計画法の枠組で定式化したものである．

鼠は箱の中で，どこにいたり，レバーがどのように見えるかなどの状態 ($s \in \boldsymbol{S}$:状態集合) が分かり，前に進んだり，レバーを押すなどの行動 ($a \in \boldsymbol{A}$:行動集合) をとることができる．このとき，環境は厳密にはマルコフ過程としてモデル化され，現在の状態と鼠がとった行動により確率的に (うまく見えなかったり，脚を滑べらしたりするかもしれないので) 次の状態 ($s' \in \boldsymbol{S}$) 遷移する．その結果報酬 (r : 例えばチーズ) が与えられる．状態遷移が既知であれば通常の動的計画法 (以下，DP と略記) の枠組で最適行動が得られるが，未知のとき環境内で試行錯誤しながら，状態遷移と最適行動を推定しなければならない．これが確率的 DP とか逐次的 DP などと呼ばれる結縁である．最も良く利用される強化学習法として **Q 学習** [6] が有名で，状態 s で行動 a をとる行動価値関数 $Q(s,a)$ は，試行錯誤により，次

式で更新される．

$$Q(s,a) \Leftarrow$$

$$(1-\alpha)Q(s,a) + \alpha(r(s,a) + \gamma \max_{a' \in \boldsymbol{A}} Q(s',a')) \quad (1)$$

ここで，s' は，次状態，α は学習率で 0 と 1 の間の値をとる．γ は，減衰率で，現在の行動が将来に渡ってどれくらい影響を及ぼすかを決めるパラメータで，0 と 1 の間の値をとり，小さい程影響が少ない．行動選択は，学習の収束時間を決める要因の一つで，一旦憶えた成功例を何回も繰り返して上達させるか，別のアプローチを未経験のところから探すかのトレードオフがある．無限の時間を費やして探索することが困難な実ロボットの観点からは前者が有利であるが，準最適解しか発見できない可能性が高くなる．以下では，実ロボットへの適用の観点から二つの問題を取り上げ，それぞれについて説明する．

3 状態・行動空間の構成問題

スキナーボックスでは，鼠は，箱の中のどこにいたり，レバーの見え方などの状態の定義を鼠自身が事前にもっていることを仮定していた．実際の環境では，これらの状態そのものをどのように定義するかが大きな問題である．すなわち，

1. どのような情報を使えば，タスク遂行に必要かつ十分な状態空間が構成可能であるか，
2. 学習可能性 (マルコフ性) を満足する状態・行動空 間をどのように構成するか．

従来の強化学習の研究では，多くがコンピュータシミュレーションによるもので，実ロボットへの適用可能性を論議しているものは少なく，ロボットの行動により状態が次状態に遷移する理想的な行動及び状態空間を構成している．しかしながら実環境で作動するセンサやアクチュエータの出力が直接，状態や行動に 1:1 に対応するとは限らない．むしろ目的に応じて，センサ空間やモータ空間を抽象化し，状態・行動空間を構成することが望まれる．このとき、センサ情報から状態へ，またモータ出力から行動への抽象化過程は，相互に依存し，鶏と卵問題に類似している (図 2 参照)．ここに，**視覚をはじめとするセンサー情報の抽象化が，行動と切り離して独立に処理できないという問題が浮かび上がる．**

従来の研究では，先に行動空間をプログラマが設計し，それに基づいて状態空間を構成するものがほとんどであった．しかし，相互依存性を考慮すると，必ずしもそのような設計がうまくとは限らない．以下では，センサ出力，モータ入力と状態・行動を厳密に区別しながら，状態・行動空間を構成する手法を実例を交えて説明する．

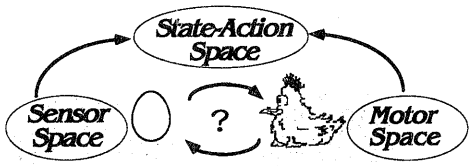


図 2: 「鶏と卵」問題

3.1 状態と行動を同時に構成する手法

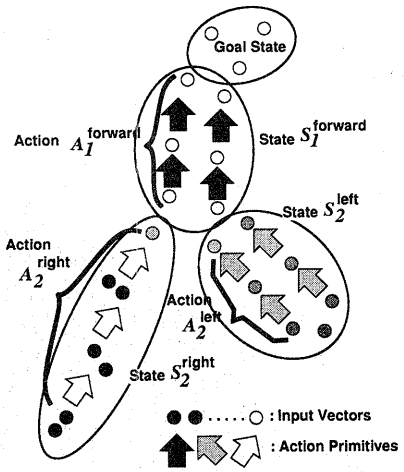
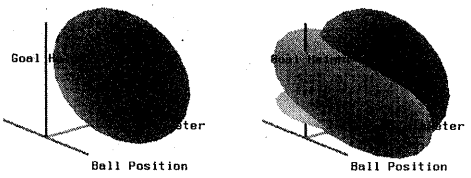


図 3: 状態と行動の相互規定

浅田ら [3] は，状態空間と行動空間を同時に構成する手法を示した．単位時間当たりに実行されるモータコマンドを行動要素，その結果生じる環境の変化を感知するセンサ情報を入力ベクトルとして，同一行動要素の系列でゴール状態 (もしくは既に獲得されている状態) に到達できる入力ベクトルの集合を「状態」，そのときの行動要素の系列を「行動」として定義し，実ロボットがボールをゴールに

シュートするタスクに適用した．図 3 にその基本的な考えを示す．小円が入力ベクトルを示し，濃淡が，知覚の違いを示す．太い矢印が行動要素 (モータコマンドの種類) を示す．異なる入力ベクトルが知覚の違いに関わらず，タスク (ゴールへの状態遷移) に応じて，同じ状態にクラスタリングされている様子がわかる．



(a) 初期分割　　(b) 次期分割

図 4: 5 次元から 3 次元に射影された状態空間の分割過程

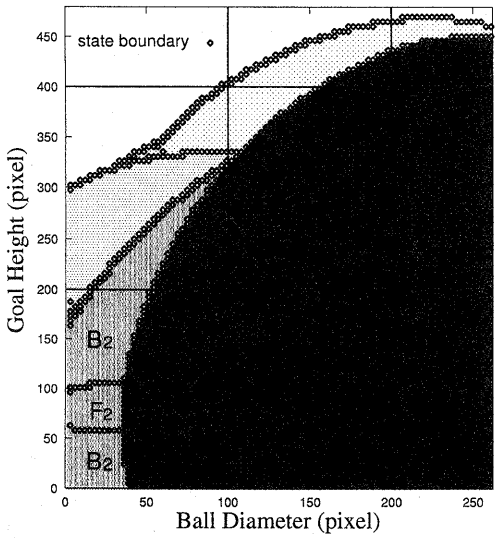


図 5: 状態空間の最終分割結果の 2 次元射影

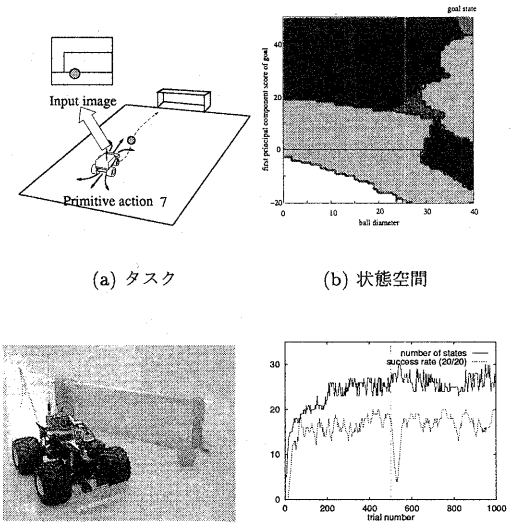
実験では，カラーカメラを搭載した移動ロボットが，青いゴールに赤いボールをシュートするタスクを考える．画面上のボールの位置，大きさ，ゴールの位置，大きさ，向きの 5 次元の状態空間を分割することになる．図 4 に，ゴールの位置，向きがいずれも 0(真正面) での断面をとり，ボール位置，ボールとゴールの大きさの 3 次元で表現したものを示す．(a) では，大きな楕円体 (S_1^F) が一つだけ得られ，直進運動 (**Forward**) に対応している．(b) では，二つの楕円体 (S_2^F と S_2^B) が追加され，それぞれ直進と後退運動 (**Backward**) に対応し，$\mathbf{S_2}$ を構成している．図 5 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．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．

Ishiguro et el.[8] は，全方位に移動可能な 3 自由度移動ロボットの時間的に連続する 2 枚の全方位画像から，ロボットの航行のための状態空間を構成した．但し，探索空間が膨大なので，教示データを基に，特徴ベクトルを直交化し，木構造の状態空間を構成した．結果得られる特徴ベクトルは，ロボットが移動中に視覚情報のどこに注視すべきかの情報も示している．

これら二つの例は、いずれもオフラインの学習である．前者では、ゴール状態からバックトラックすることで，後者では，人がゴール行動の一例を示すことで探索空間を低減し，状態空間を構成するプロセスそのものが学習過程に対応する．タスク遂行に必要な情報の取捨選択は，前者では凸包，後者では特徴空間の直交化により実現されており，センサ情報が入組んだ非線型な特徴空間の分離は困難である．また，後者では行動空間は，モータ空間に対応しており，抽象化は行われていない．また，双方とも分割するのみである．

オンラインの手法として，実ロボットに適用された例ではないが，ゴール状態へ導く初期コントローラ (必ずしも正しく導く必要はない) を想定し，ゴー

ル状態とそれ以外からなる初期の状態空間を再帰的に分割していく PARTI-GAME アルゴリズ [9] が知られている．ゴール状態へ到達できないとき，適当に状態空間を分割し，サブゴールを生成していく．分割法に関する指針が明示的に与えておらず，ゴール状態が厳密に規定されているときは，徒に分割を繰り返す惧れがある．



(a) タスク　(b) 状態空間

(c) 実ロボット　(d) 成功率と状態数

図 6: タスクと実験結果

実ロボットに適用されたオンライン状態・行動空間構成法として，高橋ら [4] は，関数近似を分割指針とし，状態の分割だけでなく融合過程を導入することにより，無駄な再分割を防ぐ手法を提案している．具体的なタスクとして，視覚移動ロボットがボールをゴールにシュートするタスクを考えた．(図 6(a) 参照)．最初に，ロボットの行動時に観測される様々な画像基本特徴量を主成分解析し，支配的と思われる画像特徴を抽出した．それらは，ほぼボールの位置 (x 座標) と大きさ，ゴールの位置，大きさ，向きの 5 次元パラメータに対応した．行動は二つの独立した左右輪への回転指令である．学習の基本的な考え方は，以下である．

1. 5 次元の知覚空間は最初 2 状態 (目標状態とそれ以外) からなる．
2. 行動に関する状態変化を関数近似し，近似による状態変化予測が異なる場合か，ゴール到達に失敗した場合のみ，状態を分割または融合し，あらたな関数近似領域を推定する．これにより無駄な探索を軽減できる．
3. 新たに分割された状態の行動価値のみを初期化し，通常の強化学習を適用する (状態数が少ないので学習時間が短い)．
4. 行動選択にランダムネスを付加し，環境の変化に対応する．

図 6(b,d) に実験結果を示す．図 (d) の実線と破線はそれぞれ状態数と過去 20 試行の成功回数を示している．450 回目にボールの大きさを 2 倍に変更した直後は成功率が下がったが，直ちに持ち直している様子が分かる．図 (b) は分割された 5 次元の状態空間を 2 次元 (ボールとゴールの大きさ) に射影したもので，右上がゴール状態を示している．入り組んだ状態が獲得された様子が分かる．状態変化が生じるまで同じモータコマンドを発生させ，その系列を行動と定義することで，行動の時間的分割を行っている．因みに 1 時間半ほどの実ロボット (図 (c) 参照) の学習時間で目的の行動が達成できた．

3.2 より複雑なタスクへの対応

これまでのタスクでは，一部の隠れ状態を除いて，センサ空間の次元が状態空間の次元と等しい場合を扱ってきた．即ち，ほとんどが静止環境を想定していた．しかし，環境が動的に変化する場合，現在のセンサ情報だけから適切な行動を決定することが困難となる．これまでこの種の問題は，強化学習の分野では部分観測マルコフ問題として定式化されてきた [10] が，他のロボットを含むマルチエーエジェント環境では，問題はより深刻となる．即ち，自身の行動と直接関係なくセンサ情報が変化するので，通常のセンサ情報に直接基づいた状態空間では強化学習を実現できない．

Uchibe et al.[2] は，環境のダイナミクスを学習者自身の運動指令を含めて推定する手法を提案している．部分空間法と呼ばれる次数同定の手法を用いており，過去の知覚情報と運動指令の組を入力として，将来の知覚情報を予測し，これによって状態パ

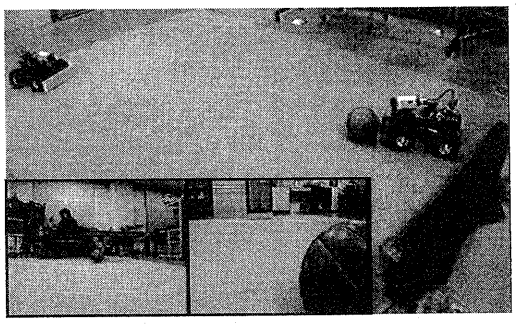


図 7: パッサーとシュータの協調行動

ラメータを推定する．理論的には，観測範囲であらゆる物を無限時間を用いて記述可能であるが，現実には，何らかの規範で同定次数を制限する必要がある．パッサーとシュータが混在する環境でのマルチエージェント学習問題に適用し，実ロボットでの結果を得ている (図 7 参照)．以下に，基本アーキテクチュアと実験結果を示す．

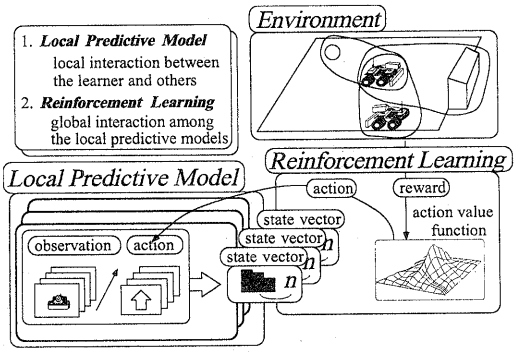


図 8: 提案するアーキテクチュア

3.2.1 アーキテクチュア

図 8 は各ロボットに与えられる行動獲得のためのアーキテクチャである．はじめに，学習者はセンサ情報だけでなく，学習者自身の行動のシーケンスから局所予測モデルを構築する．局所予測モデルは対象の次の運動が予測できるような状態ベクトルを推定する．次に推定された状態ベクトルをもとに，協調行動を獲得のための学習を開始する．

局所予測モデルは，多入力 (行動) 多出力 (観測) の関係を記述する必要がある．状態表現の方法として，システム同定の一つである正準変量解析 (Canonical Variate Analysis : 以下 CVA と略記) [11] を用いて，局所予測モデルを構築する．ここでは簡単な概略だけを述べる．

CVA は離散時間で線形の状態空間モデル

$$\begin{aligned} \boldsymbol{x}(t+1) &= \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t), \\ \boldsymbol{y}(t) &= \boldsymbol{C}\boldsymbol{x}(t) + \boldsymbol{D}\boldsymbol{u}(t), \end{aligned} \tag{2}$$

を用いる．ここで $\boldsymbol{u}(t) \in \Re^m$ と $\boldsymbol{y}(t) \in \Re^q$ はそれぞれロボットの行動ベクトルと観測ベクトルであり，$\boldsymbol{x}(t) \in \Re^n$ は状態ベクトルである．また，$\boldsymbol{A} \in \Re^{n\times n}$, $\boldsymbol{B} \in \Re^{n\times m}$, $\boldsymbol{C} \in \Re^{q\times n}$, $\boldsymbol{D} \in \Re^{q\times m}$ はパラメータ行列である．学習者は観測と行動のシーケンス $\{\boldsymbol{y}, \boldsymbol{u}\}$ から状態ベクトルを次数を含めて推定しなければならない．状態ベクトル $\boldsymbol{x}$ は過去の観測と行動のシーケンスの線形和

$$\boldsymbol{x}(t) = [\boldsymbol{I}_n \ \boldsymbol{0}]\boldsymbol{U}\boldsymbol{p}(t), \tag{3}$$

によって状態を表現する．ここで，

$$\boldsymbol{p}(t) = [\boldsymbol{u}(t-1) \cdots \boldsymbol{u}(t-l)\, \boldsymbol{y}(t-1) \cdots \boldsymbol{y}(t-l)]^T,$$

であり，$\boldsymbol{U} \in \Re^{l(m+q)\times l(m+q)}$ は CVA によって計算される行列であり，l は考慮する履歴長さである．また n は状態ベクトルの次数であり，情報量規準によって決定する．

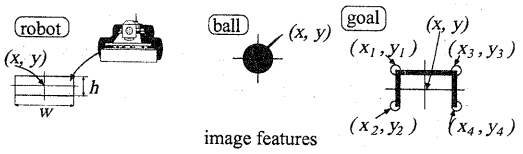
3.2.2 実験

各ロボットは TV カメラを一つ搭載し，そこから得られる画像情報から環境の状況を観測する．モータコマンドとして，各ロボットは 2 自由度を持つ．そこで，ロボットへの制御入力 $\boldsymbol{u}$ は 2 次元ベクトル

$$\boldsymbol{u}^T = [v \ \ \phi], \quad v, \phi \in \{-1, 0, 1\},$$

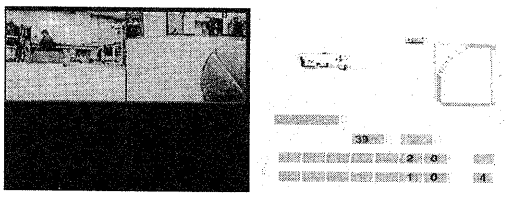
として表現する．ここで，v は台車の移動速度であり，ϕ はステアリングの角度である．また，各ロボットが観測できる画像特徴量 (観測ベクトル) を図 9 に示す．結果として，ボール，ゴール，ロボットに関する観測ベクトルの次数はそれぞれ 4, 11, 5 となる．入力画像を図 10(a) に，処理画像を図 10(b) に示す．

同時学習は困難なので，最初にパッサー (図右) がある方向にボールを蹴り出す学習を実施後，シュー



robot	ball	goal
area	area	area
center position	center position	center position
height	radius	4 corners
width		

図 9: ボール，ゴール，相手ロボットの画像特徴



(a) 入力画像　　(b) 処理画像

図 10: ロボットの入力画像と処理画像

図 11: 獲得された行動

表 1: The estimated dimension

observer	target	l	n	$\log \lvert \boldsymbol{R} \rvert$	AIC
shooter	ball	2	4	0.23	138
	goal	1	2	−0.01	121
	passer	3	6	1.22	210
passer	ball	2	4	0.78	142
	shooter	3	5	0.85	198

ターが転がるボールをシュートする学習を実施する．ともにぶつからないように，障害物回避行動は，事前に学習し，埋め込まれている．現在のセンサ出力の次元を状態空間の次元とした場合の比較として，パスの成功率が約 10%から 50%以上に，シュートの成功率も同じく約 10%から約 80%に改善された．彼らは，環境内の個々のエージェント (ゴール，ボール，敵，味方など) のダイナミクスを同定する過程が状態空間構成にあたり，強化学習がエージェントの相互作用を学習する過程とみなしている．実験結果として，物理的に同一の物体 (例えば，転がるボール) でも，経験 (この場合，タスクの違いによる経験のバイアスが存在) の差異により，推定される状態パラメータが異なったことが挙げられ，ロボットの個性を考える上で興味ある結果と考えられる．今後，協調行動などを実現するとき，このような差異をどのように吸収するかが，課題としてあげられている．

最後に，実環境で収集したデータをもとに推定をやり直し，さらに学習して獲得された行動を図 11 に示す．シミュレーション結果をそのまま適用するよりも，行動は改善された．

ここでの状態表現を視覚情報の構造化の観点から考察してみよう．推定された状態ベクトルは，式 (2,3) から明らかな様に，観測ベクトルと行動ベクトルの重み付き線形和で表現されている．**目的行動獲得に必要な情報 (予測可能な状態ベクトル) 推定過程が，観測行動空間の時空間構造化過程に対応すると**見なすことができる．ロボットが持つ身体的拘束 (知覚，行動，認知能力) が強く反映され，中間表現としての状態は，外部観測者から理解可能なものである保証は無い．たとえば，ボールに対する次元は 4 であり，位置とその時間的変化と直感的に解釈可能であるが，パッサートシューターで入れ替え

ても動作しないことから，タスクと経験にバイアスされて表現となっている．

4 討論

本研究では，観測と行動の経験にもとづく環境表現ならびに行動生成手法を示しており，従来の視覚情報処理のアプローチと異なる．特に，環境表現は，以下の点で異なる．

1. 従来のコンピュータビジョンの研究では，種々の応用が可能と考えられる3次元の定量幾何学的表現の再構成を試みてきた．これらの研究では，視覚情報のみを基にして再構成するため，ノイズに弱い，多大な計算時間を必要とする，などの欠点が挙げられる．これらの手法を，本研究が対象としている問題領域に適応した場合，複数の動物体を移動観測ステーションから観測する問題に対応し，先の欠点が曝け出されるとになり，適切な行動が生成されない．これに対し，本研究では，タスク遂行に必要な情報を経験 (観測と行動) から推定し，実時間で行動を生成している．この意味で，ここで示してきた環境表現は，推定された状態ベクトルと強化学習によって得られる状態遷移図であり，3次元定量的幾何学表現とは異なる．

2. 従来のコンピュータビジョンの研究では，観測対象の環境に対して，観測主体は,「神の眼」的な存在であり，環境に直接働きかけることは無かった．それに対し，マルチエージェント環境では，観測主体と行動主体が同一である場合が多く，観測と行動の主体が環境に埋め込まれ，観測行為そのものが環境に変動を与える．動的に変化し続ける環境で適切な行動を実時間で生成しなければいけない環境では，従来手法における「観測と行動を分離し，環境の中間的な表現 (例えば，3次元定量的幾何学表現) を求める問題設定」そのものの意味が薄れる．

3. 従来のコンピュータビジョンの研究では，環境に対して，共通の普遍的，絶対的表現を求めることを主眼としてきた．本研究では，さきの実験結果に示したように，同じハードウェア仕様の主体でも役割や経験に応じて，異なる環境表現をもつことを示した．これは，主体が主観的環境表現をもちえることを意味し，客観性を前提とする従来の見方とは異なる．主観的な環境表現は，主観的な価値観を持つ可能性を示唆し，身体性による知能発現の一つの過程と考えられる [12].

謝辞

本研究は日本学術振興会未来開拓学術研究推進事業「分散協調視覚による動的3次元状況理解」プロジェクトの一環として行った．

参考文献

[1] Y. Aloimonos. "Introduction: Active vision revisited". In Y. Aloimonos, editor, *Active Perception*, chapter 0. Lawrence Erlbaum Associates, Publishers, 1993.

[2] E. Uchibe, M. Asada, and K. Hosoda. "state space construction for behavior acquisition in multi agent environments with vision and action". In *Proc. of ICCV 98*, pages 870–875, 1998.

[3] 浅田, 野田, and 細田. ロボットの行動獲得のための状態空間の自律的構成. **日本ロボット学会誌**, 15(6):886–892, 1997.

[4] 高橋 and 浅田. 実ロボットによる行動学習のための状態空間の漸次的構成. **日本ロボット学会誌**, 17(1):118–124, 1999.

[5] 北野宏明 and 浅田稔. 「ワールドカップ」ロボットの挑戦. **日経サイエンス**, 28:74–82, 1998.

[6] C. J. C. H. Watkins and P. Dayan. "Technical note: Q-learning". *Machine Learning*, 8:279–292, 1992.

[7] 浅田, 野田, 倹積田, and 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**, 13:1:68–74, 1995.

[8] H. Ishiguro, R. Sato, and T. Ishida96. Robot oriented state space construction. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems 1996 (IROS96)*, pages 1496–1501, 1996.

[9] A. K. Moore and C. G. Atkeson. Parti-game algorithm for variable resolution reinforcement learning in multidimensional state-sapces. *Machine Learning*, 21:199–233, 1995.

[10] M. L. Littman, A. R. Cassandra, and L. P. Kaelbling. Learning policies for partially observable environment: scaling up. In *Proc. of Conf. on Machine Learning-1995*, pages 362–370, 1995.

[11] W. E. Larimore. Canonical variate analysis in identification, filtering, and adaptive control. In *Proc. 29th IEEE Conference on Decision and Control*, pages 596–604, Honolulu, Hawaii, December 1990.

[12] Minoru Asada. An agent and an environment: A view on "having bodies" - a case study on behavior learning for vision-based mobile robot -. In *Proceedings of 1996 IROS Workshop on Towards Real Autonomy*, pages 19–24, 1996.