画像運動情報に基づく単眼視覚移動ロボットの行動獲得

○中村 恭之　　浅田 稔
大阪大学工学部

Motion Sketch: Acquisition of Visual Motion Guided Behaviors

○Takayuki NAKAMURA　　Minoru ASADA
Osaka University

1 はじめに

視覚を持つ移動ロボットが，複雑な環境内を自律的に走行するためには，環境に適応して行動を決定することが重要である．そのためには，ロボット自身が環境に関する何らかのモデルを持っている必要がある．

従来からの研究においては，ロボット上の視覚センサからの2次元画像情報を処理して詳細な3次元幾何情報を再構成し，この3次元情報を用いてロボットの取り巻く環境を表現している研究例が多い．この様な手法は，コンピュータビジョンの分野において，Ullman1) によって"shape from motion" の問題が定式化されて以来，多くの研究がなされている．最近の研究では，2次元画像系列から2, 3)，アフィンまたは投影不変量から4, 5, 6, 7)，カメラキャリブレーションから8)，それぞれ3次元情報を求める研究がある．これらの研究では，3次元情報やその不変量をできるだけ正確に求めることを目的としている．しかしながら，どれくらいの再構成の精度が必要であるかを決めることは難しい．3次元情報は一般的な形式をしておりその情報を使用する際に簡単に変換が可能であるとして，3次元情報をもとにした環境の表現が，ロボットのナビゲーション等に有効であると一般的に言われている．しかしながら，ロボットを使用する際には，センサ情報やモーター制御に対して実時間処理しなければならない．さらに，目的とするタスクに依存してロボットを取り巻く環境全体を表現する必要はなく，また見え方の異なる環境を必ずしも区別する必要もない．例えば， たとえば，障害物発見や回避をタスクとした場合，机や椅子が乱雑に配置された屋内環境も，岩などを含む屋外環境も識別する必要がない．これまでの3次元幾何情報の再構成を主眼として来たアプローチでは，これらを同一視することは困難であると考えられる．

そこで，視覚情報を用いて与えられたタスクを達成する自律エージェントにとって，どのような環境表現が適当であるかについて考えなければならない．この問題に対して，ロボット学習の研究者は，ロボットに対して，外界から知覚されたデータに対して行動すること，すなわち，環境状態とロボット自身の行動の最適な関係を学習させようとしてきた9)．この状態と行動の最適な関係が，タスクもしくはロボットの行動に基づく環境表現と考えられる．これによって，タスクを達成するために詳細な3次元情報を再構成する必要がなくなる．しかしながら，これまでのロボット学習の研究においては，特定のタスクの行動学習を行なうために，環境を記述する記述子に対する候補は，知覚された莫大なデータの中から前もって選択されている．またこれらの記述子の候補は，環境シーンの構成要素や特定の状況やタスクに依存して決定されている．

そこで，この報告では，環境シーンの構成要素に独立で，モータコマンドと密接に関連したロバストな記述子である画像運動情報を利用して，実ロボットに，目的のタスクを達成させるための行動に基づく環境表現を獲得させる手法について述べる．ここでは，目的のタスクを達成するための行動に基づく環境表現を**運動スケッチ**と呼んでいる．ロボットのタスクとしては，動的環境内で障害物を回避しながら対象物体を追跡するというタスクを想定している．

2 運動スケッチとタスク

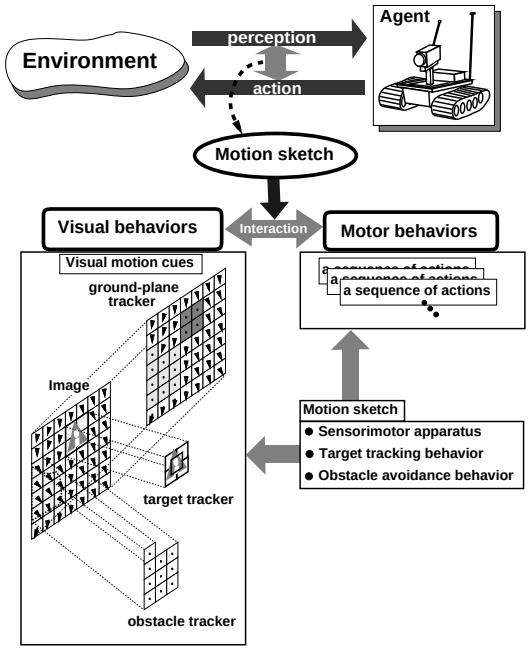


Fig.1 Motion sketch

Fig.1は，運動スケッチの基本的な概念図を表している．運動スケッチは，単眼視覚を持つ移動ロボットが幾つかの行動を学習するために環境を表現する手法である．運動スケッチの基本的な構成要素は，幾つかの視覚追跡ルーチンにより検出された画像運動情報である．視覚ルーチンの行動は，各々のタスクによって決定される．また，画像運動情報は，強化学習の1種であるQ学習により獲得されたモーター行動と密接に関連付けられている．視覚追跡ルーチンにより追跡される画像領域は，各々のタスクに依存して，人間により指定されるか，または自動的に検出される．視覚およびモーター行動は画像内で並列的に動作し，階層構造を構成している．ロボットの運動パターンを検出し続ける行動が一番下の階層である．この階層は，他の階層の行動(障害物検出，回避行動や対象物体追跡行動) によって部分的に包含されている．一番下の階層では，画像上の運動パターンとロボットのモーターコマンド間の関係を獲得する．それを行なうために，視覚追跡ルーチンが画面全体に一様に配置され，ロボットの瞬間的な運動によりオプティカルフローが検出される．この場合，各追跡ルーチンはそれぞれの画像内の位置に固定

されている．障害物検出や回避のタスクにおいては，床面上のオプティカルフローと検出されたオプティカルフローを比較することにより，障害物が存在する候補領域が検出される． そしてこの領域内部が複数の追跡ルーチンにより追跡される．そして，対象物体追跡タスクに対しては，複数の視覚追跡ルーチンにより，対象物体を安定に追跡することができる．モーター行動は，Q学習により獲得されたモーターコマンドの集合である．検出された画像運動情報や与えられたタスクをもとにしてQ学習を行なう．対象物体や検出された障害物の大きさや位置が，Q学習における状態変数として使用される．このように，視覚処理ルーチンとQ学習により獲得されたモーター行動が密接に関連付けられることによって目的のタスクを達成する．そして，視覚処理ルーチンからの画像運動情報とそれらに密接に関連した行動により環境を表現する手法をここでは，**運動スケッチ**と呼んでいる．

ここでは，自律ロボットが障害物を避けながら対象物体を追跡するタスクを想定して，その際に環境がどのように表現されるかを示す．この様なタスクを達成する行動を獲得するために，次のように4段階の過程を踏む．

stage 1 画像運動情報とロボットの行動との間の相関を求めることによりこれらの間の基本的な関係を獲得する．

stage 2 対象物体を追跡する行動を学習する．

stage 3 障害物を検出し，検出された障害物を回避する行動を学習する．

stage 4 対象物追跡行動と障害物回避行動の協調により目的のタスクを達成する．

3 センサアクチュエータ間の関係の獲得

ここでは，センサアクチュエータ間の関係を獲得する手法について述べる．センサ情報とモーターコマンドとの相関をとることによって，それらの関係を獲得する．その獲得手法について述べる前に，想定しているロボットの視覚追跡処理に用いている追跡ルーチン(センサ)と運動機構(アクチュエータ)について説明する．

(A) 視覚追跡ルーチン(センサ)

ロボットの運動による環境状態の変化を検出するために，我々は実時間視覚追跡ルーチンを使用する．それによって，実時間(ビデオレート)で約140個のウィンド(各ウィンドは8×8の大きさ)を，運動推定プロセッサ(MEP)を使用することにより追跡することができる10)．探索領域は16×16の大きさであり，MEPは次式で表される誤差(差の絶対値の総和)を最小にする各ウィンドの位置を出力する．

$$D[i,j]=\sum_{k=0}^{K-1}\sum_{l=0}^{L-1}|R[k,l]-M[i+k,j+l]|$$

$$i,j:0\le i,j\le 15,$$

ここで，$R[x,y]$，$M[x,y]$や$D[x,y]$は，参照テンプレート，マッチング領域，SADの配列をそれぞれ示している．視覚追跡ルーチンは，床面のオプティカルフローを得るためや，人間によって指定された追跡対象を追跡するため，障害物を検出，回避するために使用される． この様な視覚追跡ルーチンにより，時刻$(t=t_i)$における画像と時刻$(t=t_{i+1})$における画像間においてテンプレートマッチングを行なうことによりオプティカルフローを検出することができる．

(B) 運動機構(アクチュエータ)

本研究で使用されるロボットは，2つの独立したモーターで駆動されるPWS(Power Wheeled Steering)を採用しており，それぞれに個別のコマンド指令を送ることができる．ロボットの並進速度vや回転速度ωは2つのモーターコマンド(より正確には2つのモーターの回転角速度ω_lとω_r)によって表現することができる．(v,ω)と(ω_r,ω_l)の関係は，次式のように表すことができる．

$$\begin{pmatrix} v \\ \omega \end{pmatrix}=\begin{pmatrix} \frac{R_r}{2} & \frac{R_l}{2} \\ \frac{R_r}{T} & -\frac{R_l}{T} \end{pmatrix}\begin{pmatrix} \omega_r \\ \omega_l \end{pmatrix} \qquad (1)$$

ここで，R_r,R_lとTは，左右輪の半径，2つの車輪間の距離をそれぞれ表している．我々の実験においては，$\omega_{l(r)}$を5段階に量子化している．高速正回転(qf)，低速正回転(sf)，停止(s)，低速逆回転(sb)，高速逆回転(qb)の5段階で，これらのコマンドを組み合わせることにより，ロボットは全部で25通りの行動が選択できる．ここで，ロボットはこれらの行動の物理的な意味を知らない．

ここでは，ロボットには自身の持つセンサシステムの構造や，その効果に関する知識をあらかじめ与えられていないものと仮定して，センサアクチュエータ間の関係を求める．この報告では，文献11)の手法を以下のように拡張した．

- ソナー情報(3次元の距離情報)を用いる代わりに，複数の視覚追跡ルーチンにより獲得される床面のオプティカルフロー(2次元の視覚情報)を使用している．
- 環境因子によって，各行動に関するオプティカルフローが変動することを避けるために，あまり障害物の存在しない環境を準備する．オプティカルフローを平均化する際には，ノイズや小さな障害物による外れ値を排除するために，ロバスト推定の手法12)を用いる．

ここでは，ロボットの内部状態空間内に自己運動の情報を含めるために，画像運動情報と自身の行動間の関係を求めている．

3.1 各行動によるオプティカルフロー

画像全体の変化を検出するために49(7×7)個の視覚追跡ルーチンを用いている． 従って，49個のベクトルによって構成されるオプティカルフローが得られる．障害物の存在しない環境中で，ロボットは，ランダムに自身の行動空間の中からある1つの行動$i(\tau_{li},\tau_{ri})$　　$\tau_{li},\tau_{ri}\in\{qf,sf,st,sb,qb\}$を選択し，その時とられた行動に対する平均化されたオプティカルフローp_iが保存される，25個の行動を全てとった後に，ロバスト推定の手法を用いて，外れ値を除去し，平均化されたオプティカルフローを求める．**Fig.2**に，実環境中で，この様な手法により得られた平均化されたオプティカルフローの例を示す．

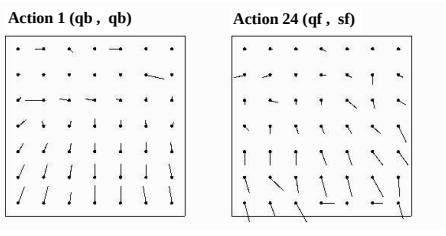


Fig.2 Examples of averaged optical flows in a real environment.

3.2 主成分パターンの獲得

各行動群の平均化されたオプティカルフロー群を使用して，ロボットの持つ行動空間を特徴づける主成分パターンを獲得する．これは，ロボットが生成することのできる平均化されたオプティカルフローによって構成される空間を解析することによって得られる．この空間の基本成分，すなわち，求められる成分の線形結合によってロボットの生成できるすべてのオプティカルフローを生成することのできるような代表的なオプティカルフローを，主成分解析(実際は特異値分解(以後，SVDと記す))することによって求める．

ある行動$i(\tau_{li}, \tau_{ri})$に相当する平均化されたオプティカルフローをベクトル表現したp_iを行ベクトルとするような，行列Pを生成する．各行列ベクトルは98個の要素(ある行動により生成されるオプティカルフローは49個のフローベクトルによって生成されている) から構成されており，行列Pは25個の列ベクトルによって構成されている． そして，この行列PのSVDは以下のように表される．

$$P_{m\times n} = U_{m\times n} S_{n\times n} E^T_{n\times n}, \tag{2}$$

ここで，Sは，行列Pの特異値を対角成分とする対角行列であり，E^Tの列ベクトルは，求めようとしている主成分パターンである．ここでは平均化されたオプティカルフローの数は$m = 25$であり，各々の平均化されたオプティカルフローの構成要素数は，$n = 98$である．Uは，列に関する直交行列である．式(2)によって，E^TのK個の列ベクトルの線形結合として，平均化されたオプティカルフローp_iは，以下のように表すことができる．

$$p_i \approx \sum_{k=1}^{K} u_{ik} s_k e_k^T \tag{3}$$

実環境中で得られた25個のオプティカルフローから構成される行列PのSVD を計算することによって，26個の主成分パターンを得た．**Fig.3**は，その時得られた各主成分ベクトルの特異値の値を示している．

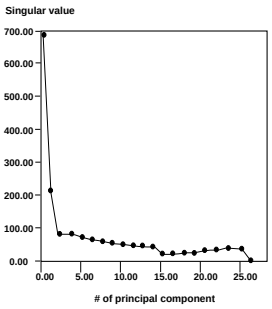


Fig.3 Singular values of principal components.

他に比べて大きな特異値を持つ2つの重要な主成分パターンをそれらの中から選択し，平均化されたオプティカルフローを，これらの2つの主成分パターンによって近似した．こうして，もとのオプティカルフローp_iは，次のように近似される．

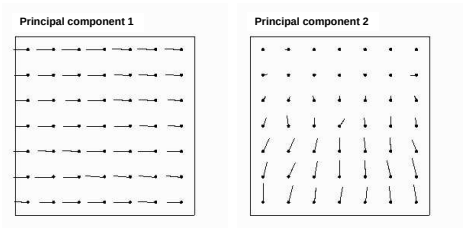
$$p_i \approx u_{i1} s_1 e_1^T + u_{i2} s_2 e_2^T$$

Fig.4は，実環境中で得られた2つの主成分パターンを表している．明らかに，(a) は，純粋な回転運動に相当し，(b) は，純粋な並進運動に相当する．

つぎに，各行動によって生成されるp_iを2つの主成分パターンによって近似した場合の係数$a_k^i = u_{ik} s_k \quad (i = 1, 2)$によって，もとの$p_i$を表現することによって，ロボットの実行可能な行動間の関係を獲得する．これがセンサアクチュエータ間の関係になる．**Fig.5**は，実ロボットの各行動間の関係を表している．この図内の数字は，行動の番号$i \quad (i = 1 \sim 25)$を表している．

4 強化学習による行動獲得

本研究では，対象物追跡行動，障害物回避行動の獲得のために，強化学習の1つであるQ学習を用いている．Q学習についての詳細については，13),14) を参照して頂きたい．ここでは，簡単にQ学習について述べる．Q学習では，ロボットが識別可能なロボットの環境の状態を表す状態空間をS,



(a) (b)

Fig.4 The first two principal components.

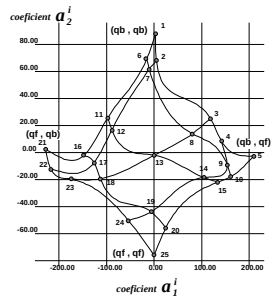


Fig.5 The relation between the possible actions of the real robot.

環境に対してロボットのとることのできる行動の集合をAとする．現在の状態sにおいて，ロボットの取った行動aにより，ある確率で次状態s'に遷移して，この状態行動対(s, a)に対して，その評価として報酬$r(s, a)$が環境からロボットに与えられる．この報酬の積算$Q(s, a)$(行動価値関数) を評価，更新することによって学習を進める．このアルゴリズムは動的計画法15) の概念から導かれ，アルゴリズムの収束性が証明されている．

4.1 対象物追跡行動の獲得

上述のような状態集合，行動集合や他の関数やパラメータにしたがって，対象物追跡タスクにQ学習を適用した．

我々は，人間によって指定された対象物体を追跡するためや，画像内での対象物体の位置や見え方に関する情報を獲得するために，視覚追跡ルーチンを使用する．この情報は，対象物体追跡行動の獲得のためのQ学習アルゴリズムの中で使用されている．

4.1.1 視覚追跡行動

ここで使用されている視覚追跡ルーチンは以下のような視覚機能を持っている．

1. 複数ウィンドによる追跡:1つの対象物体につき5個の視覚追跡ルーチンから成る追跡器により，対象物を追跡している (図**Fig.6**参照)．これにより，対象物の一部が隠されたり，ロボットの微小振動により入力画像がぶれても追跡し続けることができる．また，解像度の異なる同一画像3枚を1枚の入力画像として使用している．たとえ，対象物体の画面上での大きさが拡大縮小しても，ブロックマッチングのための探索領域を解像度の異なる画像上へ移動させることによって，対象物体の追跡が可能になっている．**Fig.6**は，人間によって指定された初期画像 (a)，複数ウィンドによる追跡のための5つの視覚ルーチンの配置 (b)，拡大縮小に対応する追跡のための解像度の異なる3つの入力画像 (c) をそれぞれ示している．

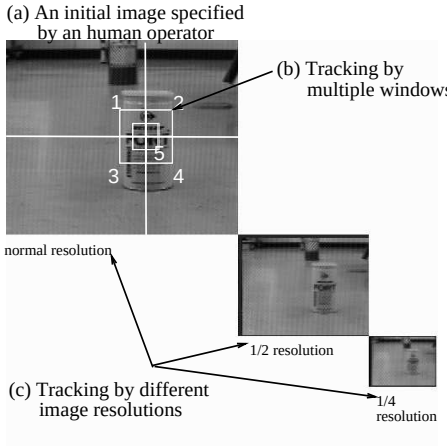


Fig.6 Visual functions of tracking routine.

2. 対象画像の追跡に失敗した場合には，画像内探索ルーチンが呼び出され，再び対象画像が検出される．

我々は，視覚追跡ルーチンによって得られる対象の大きさ(3段階)や対象の位置をもとに，画像内の対象物体の状態を定義している．

4.1.2 Q学習における状態，行動空間

対象物体追跡タスクにQ学習の枠組を適用するために，我々は幾つかの集合とパラメータを定義する．

画像内の対象物体の状態が，位置に関して(左，中，右)の3通りと大きさに関して(大(近く)，中，小(遠く))の3通りの組合せにより，9通りの状態に量子化されている．同様にして，対象物体の位置や大きさの変化を，位置の変化に対する3状態(左へ移動，移動なし，右へ移動)と大きさの変化に対する3状態(大きくなる，変化なし，小さくなる)の組合せで9状態に量子化している．また見失った状態を2状態(左側に見えなくなった，右側に見えなくなった)を状態空間に付加した．さらに，現在の状態を観察する時にとった行動(全部で25状態)を状態空間に付加した．従って，状態集合S内には全部で95×25状態存在する．行動集合A内には，全部で，25個の行動が存在する．エージェントが対象物体に触れた時に1の報酬を与えることにし，それ以外の場合は0を与える．

4.2 障害物回避行動の獲得

4.2.1 オプティカルフローの差による障害物検出と追跡

我々は，障害物の存在しない環境内である行動iをとった時に生じるオプティカルフローp_iをすでに知っている．このp_iは，主成分パターンの線形和として表現されているためにノイズ成分の影響が少ない．これにより，障害物検出が容易になる．障害物の存在候補領域は，オプティカルフローp_iと，障害物の存在する環境内で得られたオプティカルフローp_i^{obs}を比較することによって求められる．

つまり，p_i内のフローベクトルと異なったフローベクトルの存在する領域が，障害物の存在する候補領域になる．この領域の画像内での位置や大きさが，障害物追跡行動を獲得するために使用される．

4.2.2 障害物回避行動の学習

障害物回避の行動は，2段階から構成される．まず，対象物追跡行動と同様の方法により，障害物追跡行動が学習される．つぎに，障害物回避行動が，エージェントのとれる行動と，障害物追跡行動の間の関係を用いることによって次のように生成される．

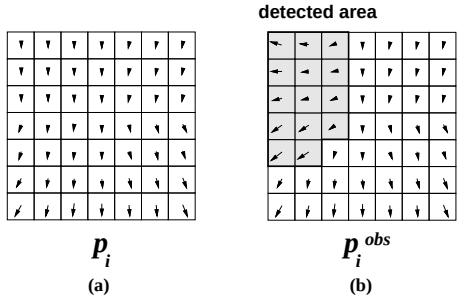
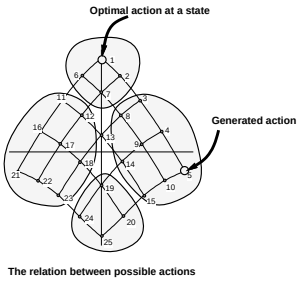


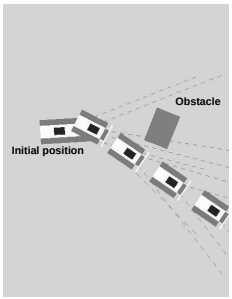
Fig.7 Obstacle Detection.

1. 係数(a_1^i, a_2^i)によって表されるエージェントのとれる行動空間をクラスタリングすることにより，4つのカテゴリーに分類する．(**Fig.8**(a) 参照)
2. 障害物追跡行動をこの行動空間上に写像し，障害物を追跡する行動の属するカテゴリーC^tを見つける．
3. 障害物回避行動は，このカテゴリーC^tを除外するようにして，他のカテゴリー内の行動から選択される．より正確には，カテゴリーC^tを除く他のカテゴリー内に属する行動の中で，障害物追跡行動に関する最小の行動価値関数を持つ行動を，障害物回避行動として選択する．

Fig.8(b)は， ここで述べられた手法によって獲得された障害物回避行動のシミュレーション結果を示している．



(a) Generation of obstacle avoidance behavior



(b) Obstacle avoidance behavior in simulation.

Fig.8 Obstacle avoidance behavior

4.3 各行動の協調によるタスクの達成

ここでは，すでに学習された行動を統合することによってタスクを達成することを考える．ここでは，環境の状態によって，各行動に関する行動価値関数を入れ換えることによってタスクを達成する．従って，我々は学習された行動を統合するために，subsumption architecture 16) と同様の行動選択構造を使用している．行動価値関数の入れ換え条件は，障害物が大きく観測される場合に障害物回避行動が選択されるものとする．

5 実験結果

5.1 実システム

Fig.9は，単眼視覚ロボットのシステム構成を示している．我々はロボットの遠隔操作システムを構築した17)．画像処理装置や，移動ロボット制御システムは，ホストコンピュー

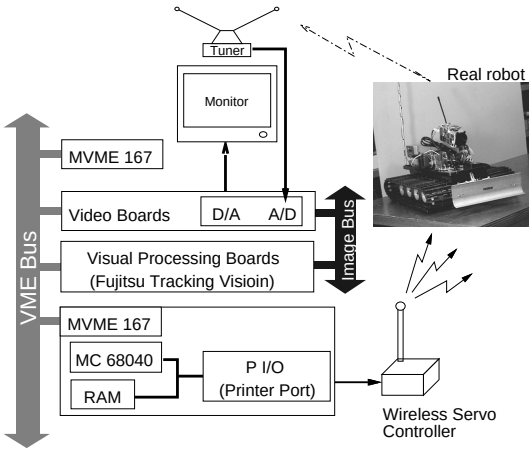


Fig.9 A configuration of the system.

タの MVME167(MC68040 CPU) 上の VxWorks OS により管理されている．MVME167 コンピュータは，イーサネットを介して Sun ワークステーションに接続されており，プログラム開発が容易に行なえる．ロボット上の CCD カメラ (Sony 社製 EVI-310) でとられた画像は，ロボット上のビデオ送信器から，ホスト側の UHF 受信器に送られ，スキャンコンバータ (Sony 社製) によりサブサンプリングされる．そして，ビデオ信号は高速相関演算処理装置トラッキングビジョン (富士通社製) に送られる．トラッキングビジョンは，幾つかの指定されたテンプレート画像と高速にブロック相関をとる機能を持っており，実時間で画像上の運動情報を獲得することができる．

トラッキングビジョンは画像上の各領域におけるフローベクトルをホスト CPU へ送り，ホスト CPU 上で，平均化運動ベクトル場を計算し，それらを保存しておく．保存された平均化運動ベクトル場は，ワークステーション上で特異値分解処理される．**Fig.10**は，我々の製作した移動ロボットと検出されたオプティカルフローを示している．

(a) our robot (b) detected image

Fig.10 A picture of our robot and motion detection.

5.2 障害物が存在しない環境内での追跡行動

実験は2つの段階から構成される．まず最初に，コンピュータシミュレーションにより，最適な状態行動対を学習により獲得する，そしてそれを実際の環境に対して適用する．コンピュータシミュレーションの効用は，アルゴリズムの正当性を確認するだけでなく，学習過程において実ロボットを使用するコストを節約することにある．**Fig.11**は，移動ロボットが対象を追跡している際の画像系列を示している．**Fig.11** (a) 内の左画像は，対象の初期位置を示している．**Fig.11** (b), (c), (d) 内の左画像は，処理画像を示している．これらの画像内での白い枠の長方形は，追跡されている対象の位置を表している．また，画像内の白線群は，オプティカルフローを示している．

5.3 オプティカルフローの差による障害物検出

Fig.12は，実環境内での障害物検出の結果を示している．ここで (a) は，その実環境を示し，(b) に，検出された候補領域が示されている．画像内の白い円が，障害物候補領域を示している．

6 考察および今後の課題

我々は，エージェントが外界に対して行動するための環境因子に依存しない一般的な手がかりの1つとして，視覚的な運動情報を利用することを提案した．視覚的な運動情報は，環境構造に依存せず，エージェント自身の持つモーターコマンドと密接な関係を持っている．手がかり自体は環境因子に依存しないものであるが，視覚運動情報を検出したり，対象物体を追跡するために使用されている追跡ルーチンが，単純な相関演算を基礎に処理しているために，エージェントが移動したことにともなう照明条件の変化の影響を受けてしまう．画像の輝度値を正規化する処理によってそれらの影響を受けなくなると考えれるが，その処理に多くの時間が必要となると考えられる

つぎに，障害物回避行動獲得に関する問題点について述べる．障害物回避行動の獲得の際に用いられている状態変数として障害物の大きさを考慮しているが，ここでは，障害物が大きく見えることと障害物が近くにあるという事を等価であるかのように考えられている．従って，実際の大きさが小さい物体が障害物として存在する場合，回避できない．例えば，細長い棒状の物がロボットの前方に存在すると，回避できない．また，ロボットの運動量が小さい場合や障害物の高さが低い場合に，フローベクトルによるセグメンテーションに失敗して，障害物を検出できないという問題がある．

障害物回避行動と各行動の協調行動の実現については，現在インプリメント中である．

参考文献

1) S. Ullman. *The Interpretation of Visual Motion.* MIT Press, Cambridge, MA, 1979.
2) C. Tomasi and T. Kanade. “Shape and motion from image streams under orthography: a factorization method”. *IJCV*, Vol. 9:2, pp. 137–154, 1992.
3) Carlo Tomasi. “Pictures and Tracils: a New Framework for the Computation of Shape and Motion from Perspective Image Sequence”. In *Proc. IEEE Int’l Conf. on CVPR*, pp. 913–918, 1994.
4) A. Shashua. “Projective depth: A geometric invariant for 3D reconstruction from two perspective/orthographic views and for visual recognition”. In *Proc. IEEE Int’l Conf. on Computer Vision*, pp. 583–590, 1993.
5) Harpreet S. Sawhney. “3D GEOMETRY FROM PLANAR PARALLAX”. In *Proc. IEEE Int’l Conf. on CVPR*, pp. 929–934, 1994.
6) D. A. Forsyth, J. L. Mundy, A. Zisserman, and C. A. Rothwell. “Using Global Consistency to Recognise Euclidean Objects with an Uncalibrated Camera”. In *Proc. IEEE Int’l Conf. on CVPR*, pp. 502–507, 1994.
7) Q. T. Luong and T. Vieville. “Canonic Representations for the Geometries of Multiple Projective Views”. In *Proc. of Third European Conference on Computer Vision*, pp. 589–597, 1994.

8) Richard I. Hartley. "An Algorithm for Self Calibration from Several Views". In *Proc. IEEE Int'l Conf. on CVPR*, pp. 908–912, 1994.

9) J. H. Connel and S. Mahadevan, editors. *Robot Learning*. Kluwer Academic Publishers, 1993.

10) H. Inoue, T. Tachikawa, and M. Inaba. "Robot vision system with a correlation chip for real-time tracking, optical flow and depth map generation". In *Proc. IEEE Int'l Conf. on Robotics and Automation*, pp. 1621–1626, 1992.

11) D. Pierce and B. Kuipers. "Learning to Explore and Build Maps'. In *Proc. of AAAI'94*, pp. 1264–1271, 1994.

12) P. Meer, D. Mintz, A. Rosenfeld, and D. Y. Kim. "Robust Regression Methods for Computer Vision: A Review". *IJCV*, Vol. 6:1, pp. 59–70, 1990.

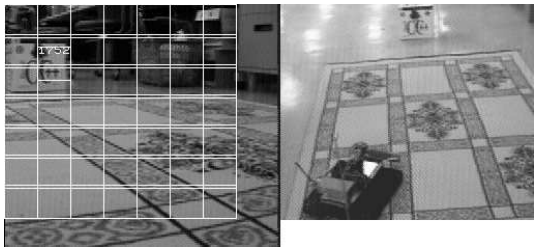
13) C. J. C. H. Watkins. *Learning from delayed rewards"*. PhD thesis, King's College, University of Cambridge, May 1989.

14) L. P. Kaelbling. "Learning to achieve goals". In *Proc. of IJCAI-93*, pp. 1094–1098, 1993.

15) R. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.

16) R. A. Brooks. "A robust layered control system for a mobile robot". *IEEE J. Robotics and Automation*, Vol. RA-2, pp. 14–23, 1986.

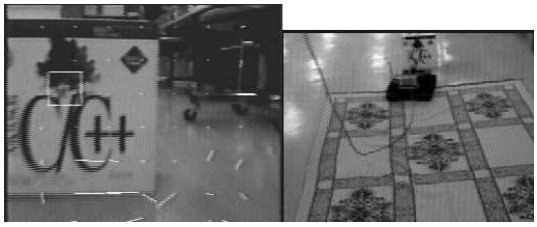
17) M. Asada, S. Noda, S. Tawaratsumida, and K. Hosoda. "Vision-Based Behavior Acquisition For A Shooting Robot By Using A Reinforcement Learning". In *Proc. of IAPR / IEEE Workshop on Visual Behaviors-1994*, pp. 112–118, 1994.



(a)

(b)

(c)



(d)

Fig.11 The robot succeeded in pursuing a moving target.

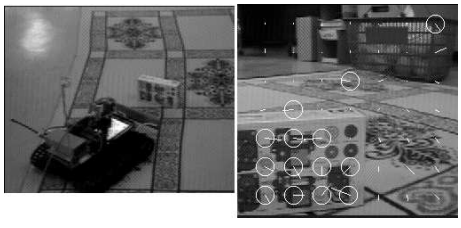


Fig.12 A picture of the environment with an obstacle and obstacle detection.