

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

Motion Sketch: Acquisition of Visual Motion Guided Behaviors

中村 恭之 浅田 稔

Takayuki Nakamura Minoru Asada

大阪大学工学部電子制御機械工学科

Dept. of Mech. Eng. for Computer-Controlled Machinery, Osaka University

In this paper, we propose *Motion Sketch* as a representation to show the interaction between a vision-based learning agent and its environment. By obtaining this representation, a one-eyed mobile robot can learn several behaviors such as obstacle avoidance and target pursuit. The *Motion Sketch* represents tight coupling between visual and motor behaviors. Visual behavior means optical flow detection and visual tracking by a real-time visual tracking routines. Motor behavior means a set of uninterpreted actions. It is obtained by statistical technique and Q-learning, a most widely used reinforcement learning method based on the visual motion cues. Finally, the experimental results of real robot implementation with real-time visual tracking routines are shown.

1 はじめに

自律エージェントの実現のためにはエージェントと環境の相互作用を記述することが重要であると言われており、そのための計算論的アプローチが、最近、AIの分野において多く研究されている [1]。本研究では、視覚に基づく学習能力を持つ実ロボットを用いた自律エージェントを構築し、そのエージェントに複雑な環境下でいろいろな種類のタスクを実行させることを目標としている。従って、エージェントにいろいろなタスクを実行させるためには、そのエージェントと環境の間の相互作用を明らかにしなければならない。

今までにも、視覚処理能力を持つエージェントと環境の関係性を解析したいろいろな研究例があった。例えば、生物学や生理心理学の分野では、動物実験などから、運動が知覚に本質的に関連していること [2]、自律的な運動無しに知覚が形成されないこと [3] が示されている。このことは、ロボットが自分の知覚を構成する際に、自律的な運動コマンドと得られるセンサ情報の関係を獲得することによって可能であることを示唆している。

ロボット学習の分野では、エージェントと環境との相互作用を通して、エージェントに対して、与えられたタスクを達成するような合目的な行動を学習させようとしてきた。しかしながら、それらの研究の多くは、コンピュータシミュレーションのみの研究であり、実ロボットを用いた研究例 [4, 5] は、数少ない。さらに、強化学習において視覚情報を用いた研究例は、視覚情報のセンシングおよび処理コストのために、非常に少ない。

そこで、本研究では、「運動スケッチ」と呼ぶエージェント(実ロボット)と環境との相互作用を獲得する手法を提案する。実ロボットの持つ視覚センサとモーターシステムとの密な相互作用を実現するために、この枠組にしたがって、視覚情報とモータコマンド群との関係を求め、さらに、エージェントの持つ学習能力により、モータコマンド群をタスク達成のために適した形態に抽象化することで、与えられたタスクを実現する行動を獲得する。実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目標物体を追跡するというタスクを想定している。提案した枠組の有効性を示すために、実ロボットを用いた実験を行っているので、それについても報告する。

2 運動スケッチ

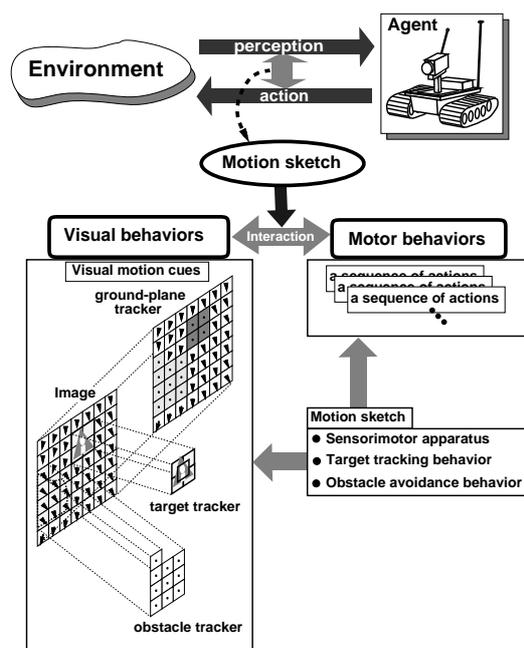


図 1: 運動スケッチ

運動スケッチの基本概念図を図1に示す。運動スケッチは、視覚行動とモーター行動の密接な関係性を表現しており、それを獲得することによりエージェントが目的のタスクを達成できる。複数の視覚追跡ルーチンにより画像運動情報を検出することを、「視覚行動」と呼び、また、動作系列の集合を「モーター行動」と呼んでいる。視覚行動は、個々のタスクによってあらかじめ定義されている。また、モーター行動を構成する各動作は、初期において何の意味づけもなされていないと仮定する。

視覚行動として、視覚追跡ルーチン群が画像面全体を一様に覆うように配置され、エージェントの瞬間的な運動によるオプティカルフローの検出するもの (ground-plane tracker)、複数の参照画像を初期に設定し、それら複数の追跡ルーチンによりある1つ

の目標物体を安定して追跡するもの(target tracker),そして、あらかじめ獲得されている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の候補領域を検出し、検出された領域を複数の追跡ルーチン群によって追跡するもの(obstacle tracker)が存在する。また、これらの視覚行動は並列的に動作し、階層構造を構成している。

視覚行動によって検出されるオプティカルフローが統計的手法により、また、検出される目標物体や障害物候補領域の画面上の大きさや位置が強化学習の1種であるQ学習によって、モーター行動と密接に関係付けられることによって「運動スケッチ」が獲得される。

この手法により、環境やロボットに関する事前知識をほとんど必要とせず、視覚情報のみを通してタスクを遂行する枠組が構成でき、環境を表現するために使われている画像運動情報は、環境の構成要素に依存せず特定の状況やタスクに制限されない、さらに、実時間視覚追跡ルーチンを使用することによって、エージェントと環境との相互作用が高速に得られる、といった利点がある。

ここで想定したタスクを達成するために必要な「運動スケッチ」は、まず、画像運動情報とロボットの動作との間の相関を求めることによりセンサアクチュエータ間の関係を獲得し、つぎに目標物体を追跡する行動、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により獲得し、最後に、目標物体追跡行動と障害物回避行動を協調することによって、獲得される。

3 センサアクチュエータ間の関係の獲得

ここでは、最も基本的な「運動スケッチ」として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簡単に述べる。詳しくは、文献[6]を参照して頂きたい。

利用した移動ロボットでは、PWS(Power Wheeled Steering)システムで、左右に独立にモーターコマンドを送ることができる(図2(a)参照)。ここでは、運動コマンドと視覚情報の密な結合を獲得するために、運動コマンドを左右モータに送る単純なボタンスイッチとし、視覚行動(ground-plane tracker)により得られたオプティカルフロー(同図(b)参照)とこれらのスイッチの関係を得る。

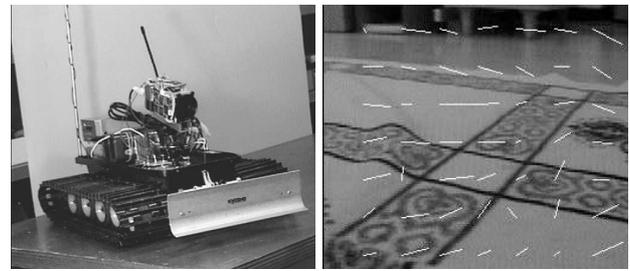
得られるフローパターン(図2(b)参照)とそのとき左右のモーターに送出したモーターコマンドの組合わせを集めて、相関関係を求める(フローパターンの主成分解析)ことにより、ロボットの行動を表現する。ロボットはPWSにより、基本的な運動成分として回転と直進を持つ。解析結果から得られる基本的な運動は図3(a)に示すように、回転と直進のフローパターンに対応している。そして、全ての運動コマンド(この場合25個)が、これら二つの主成分の線形和で表現できる。

2つの主成分パターンによって線形近似した場合の係数によって、もとのフローパターンを表現することによって、全ての運動コマンド(各動作)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。図3(b)は、実ロボットの各動作間の関係を表している。この図内の数字は、動作の番号 i ($i = 1 \sim 25$)を表している。

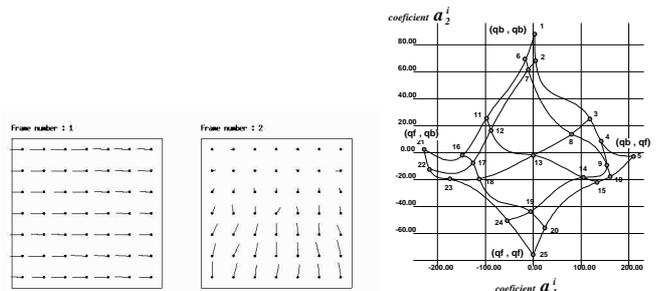
注意しなければいけない事は、ロボット自身は、3次元運動の物理的意味は、理解しておらず、自分の運動とそれによる視覚情報の変化の対応を獲得したに過ぎないが、この関係を利用して障害物の発見・回避などの行動が獲得でき、さらにQ学習における学習時間を短縮することができる。

4 強化学習による行動獲得

本研究では、目標物体追跡行動、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学習の詳細については、[7],[8]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 、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



(a) 移動ロボット (b) 抽出されたフロー
図2: 移動ロボットとフロー抽出の例



(a) 2種類の主成分フロー (b) 実ロボットの各動作間の関係
図3: センサアクチュエータ間の関係の獲得

は動的計画法[9]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 目標物体追跡行動の獲得

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、目標物体追跡タスクにQ学習を適用した。ここでは、視覚行動(target tracker)によって得られる目標の大きさ(3段階)や目標の位置をもとに、画像内の目標物体の状態を定義してQ学習により行動獲得を行った。状態空間や行動空間の構成に関しては詳しくは、文献[6]を参照して頂きたい。

4.1.1 Q学習における学習時間の短縮

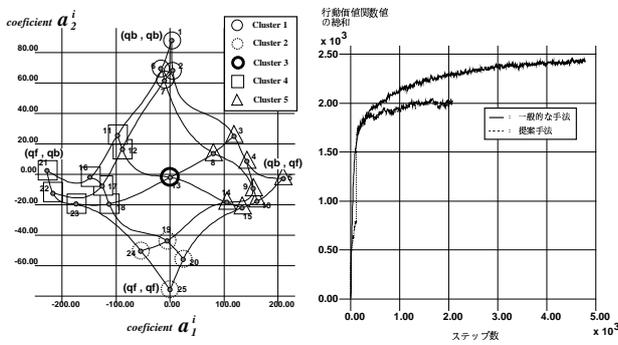
Q学習では、状態空間の数が大きくなるにつれて、学習にかかりの時間が必要になる。これは、実ロボット上で学習を行う際には大きな問題となるため、学習時間を短縮させることが必要になる。ここでは、学習時間の短縮のための手法として、3で求められたロボットの行動空間内の各動作間の関係を用いて、次のような行動選択のスケジューリングを行うことを提案する。1)まず、ロボットの行動空間をクラスタリングすることによって行動空間の次元数を減らして学習し、2)その後、ある状態において政策に沿って、あるクラスターを選択し、その選択されたクラスター内で、ボルツマン分布に従った行動選択を行う。

この手法の有効性を調べるために、目標物体を追跡する行動の獲得の際に、一般的に使用されるボルツマン分布に従った行動選択を行った結果と、ここで提案した手法に従った結果の比較実験を行った。その結果を図4に示す。ロボットの行動空間のクラスタリングは、ISODATA[10]アルゴリズムを用いて行った。図4(a)に示されているように5つのクラスタが生成された。これを用いて学習時間の短縮を試みた結果が、図4(b)である。提案した手法により学習時間が半分以下に短縮されることが分かる。

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

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 は、すでにセンサアクチュエータ間の関係を



(a) 分類結果 (b) 学習時間の比較結果
図 4: 学習時間の短縮

求める際に得られている。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。この領域を追跡すること、すなわち、視覚行動 (obstacle tracker) によって得られる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、目標物体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、障害物追跡行動の関係を用いることによって次のように生成される。1) エージェントの行動空間をクラスタリングすることにより得られた、5つのカテゴリーのうちの停止行動のカテゴリーを除く、4つのカテゴリーを利用する。(図 4(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、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。

4.3 各行動の協調によるタスクの達成

本研究では、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。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各行動を協調することで、タスクを達成する。従って、ここでは学習された行動を統合するために、subsumption architecture [11] 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、障害物追跡行動の行動価値関数の値がある閾値以上になると障害物回避行動が選択されるものとする。

5 実験結果

実際に構築したシステムの詳細については文献 [6] を参照して頂きたい。

5.1 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。図 5-1 ~ 4 は、移動ロボットが目標を追跡している際の画像系列を示している。図 5-1 の上側の画像は、目標の初期位置を示している。図 5-2 ~ 4 の上側の画像は、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、追跡されている目標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。

5.2 障害物の検出と回避行動

図 6-1 ~ 4 に、実ロボットが移動障害物を回避する画像系列を示す。図 6-1 ~ 4 の上側の図は、処理画像を示している。処理画像内の長方形領域は、障害物候補領域を示している。

5.3 各行動の協調によるタスクの達成

図 7 に、実ロボットが障害物を回避しながら目標物体を追跡している画像系列を示す。図 7-1 ~ 8 の上側の図は、処理画像を示している。初期に設定された目標物体 (植木鉢) に向かって実ロボットが移動している時に、人間 (障害物) がその前を通りかかり、ロボットは一端回避する。この時点で、目標物体を見失っているので移動しながら目標物体を探索して、再発見し目標物体に向かって移動をしている様子をあらわしている。

6 まとめ

本研究では、環境やロボットの構造に対する事前知識を前提とせず、視覚情報のみを通してタスクを遂行する枠組として「運動スケッチ」を獲得することを提案した。そのタスク遂行過程において、従来必要とされていたカメラキャリブレーション、3次元情報の再構成を一切必要としないという特徴がある。ここでは、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した。

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 (「知能ロボット」課題番号 07245214) の補助を受けた。

参考文献

- [1] Philip E. Agre. "Computational research on interaction and agency". *Artificial Intelligence*, Vol. 72, pp. 1-52, 1995.
- [2] G. A. Horridge. "The evolution of visual processing and the construction of seeing systems". In *Proc. of Royal Soc. London B230*, pp. 279-292, 1987.
- [3] R. Held and A. Hein. "Movement-produced stimulation in the development of visually guided behaviors". *Journal of Comparative and Physiological Psychology*, Vol. 56:5, pp. 872-876, 1963.
- [4] P. Maes and R. A. Brooks. "Learning to coordinate behaviors". In *Proc. of AAAI-90*, pp. 796-802, 1990.
- [5] J. H. Connel and S. Mahadevan. "Rapid task learning for real robot". In J. H. Connel and S. Mahadevan, editors, *Robot Learning*, chapter 5. Kluwer Academic Publishers, 1993.
- [6] 中村, 浅田. "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". 第5回ロボットシンポジウム予稿集, pp. 133-138, 1995.
- [7] C. J. C. H. Watkins. *Learning from delayed rewards*. PhD thesis, King's College, University of Cambridge, May 1989.
- [8] L. P. Kaelbling. "Learning to achieve goals". In *Proc. of IJCAI-93*, pp. 1094-1098, 1993.
- [9] R. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.
- [10] A. Jain and R. Dubes. *Algorithms for Clustering Data*. Prentice-Hall N. J., 1988.
- [11] R. A. Brooks. "A robust layered control system for a mobile robot". *IEEE J. Robotics and Automation*, Vol. RA-2, pp. 14-23, 1986.

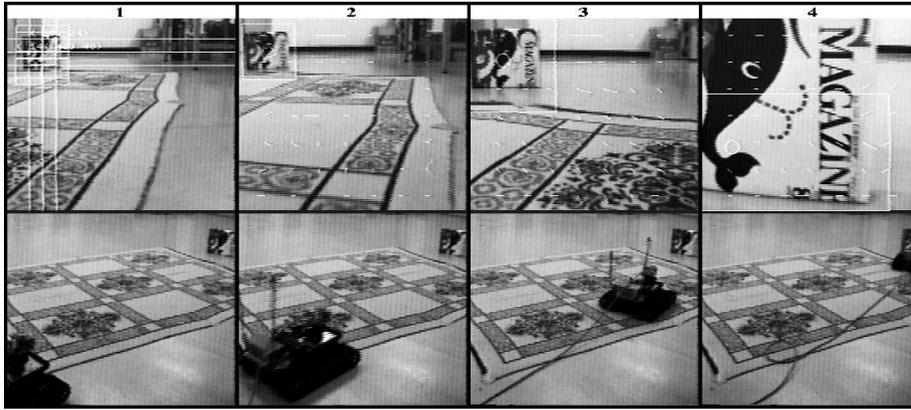


図 5: 目標物体追跡行動

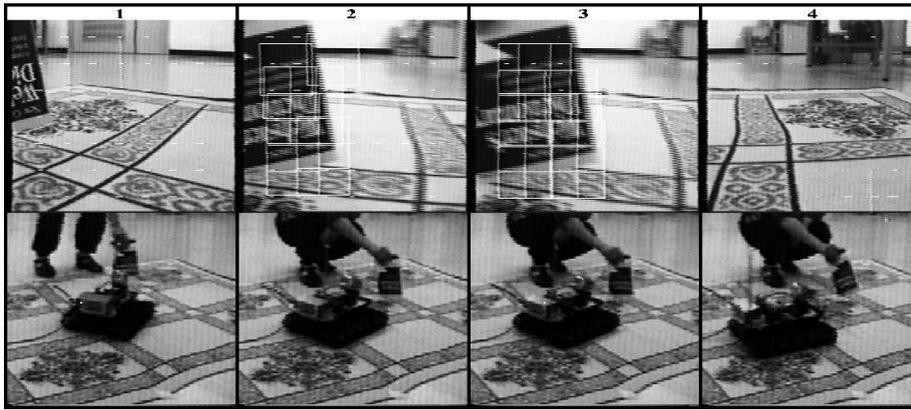


図 6: 障害物回避行動

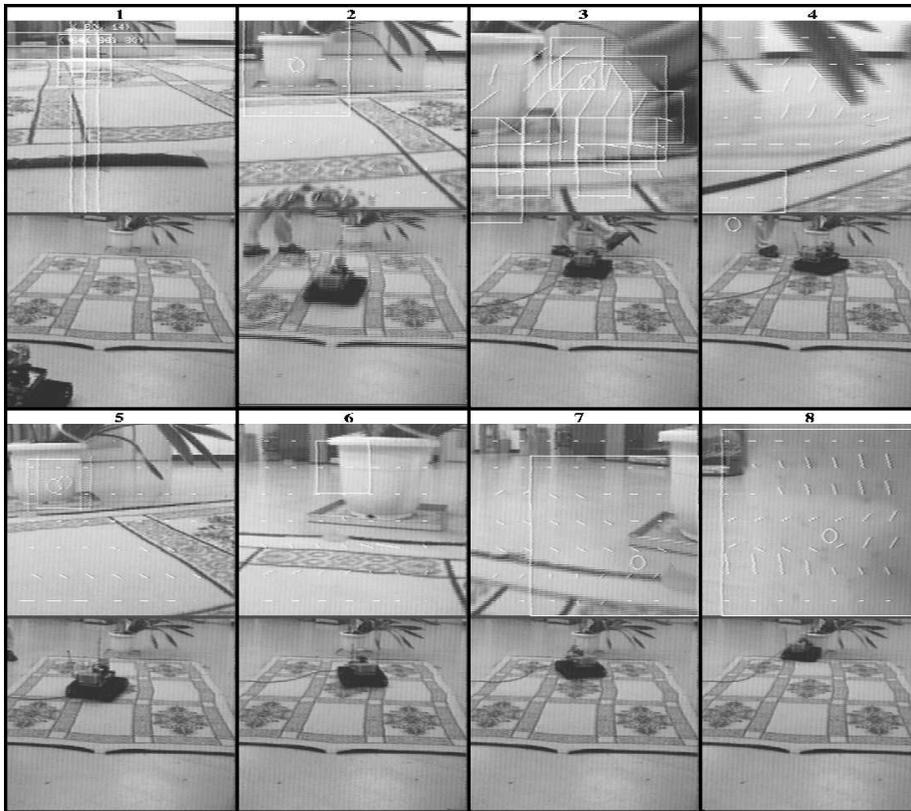


図 7: 障害物を回避し目標物体を追跡する行動