学術論文

ステレオスケッチ： ステレオ視覚を持つ移動ロボットの行動獲得

中 村 恭 之* 浅 田 稔*

Stereo Sketch: Behavior Acquisition for a Stereo Vision-Based Mobile Robot

Takayuki Nakamura* and Minoru Asada*

We have proposed *motion sketch* [1] as a method to represent an interaction between a one-eyed learning agent (a mobile robot) and its environment. In this paper, we extend the basic idea in *motion sketch*, tight coupling of perception and action, to *stereo sketch* by which a stereo-vision based mobile robot learns various kinds of behaviors such as target reaching and obstacle avoidance. Here, we deal with target reaching behavior with which *motion sketch* copes by detecting and avoiding occlusions. First, an input scene is segmented into homogeneous regions by the enhanced ISODATA algorithm with MDL principle in terms of image coordinates and disparity information obtained from the fast stereo matcher based on the coarse-to-fine control method. Then, the segmented regions including the target area and their occlusion status identified during the stereo and motion disparity estimation process construct a state space for a reinforcement learning method to obtain a target reaching behavior. As a result of learning, the robot can avoid obstacles without describing them explicitly. We give the computer simulation results and the real robot implementation to show the validity of *stereo sketch*.

Key Words: Stereo Sketch, Visual Behavior, Motor Behavior, Reinforcement Learning

1. は じ め に

ロボティクスとAIの究極の目標の一つは、目的達成のための動作を選択する内部構造を自己組織化できる自律ロボットを実現することである。つまり、自律ロボットは学習しなければならない。そのようなロボットは、環境と密な相互作用を行うことにより、環境を理解し、行動を洗練させていくと考えられる。したがって、自律ロボット実現のためには、ロボットのセンサ情報と行動の密な結合を生成する必要がある。

このような研究例の一つとして、コンピュータビジョンの分野では、Aloimonos [2] が能動視覚 (Active Vision) を提唱した。この枠組みのなかの、合目的能動視覚 (Purposive Active-Vision) [3]~[5] では、「視覚情報は孤立した形では存在し得ず、外界と相互作用する目的を持ったシステムの一部分として存在する」[3] と考えられている。しかし、実際には、観測者の運動情報が既知であることや観測者の運動を制御可能であることを基に、二次元の画像系列から距離や形状などの三次元情報を復元するために、能動視覚システムを使用している場合が多く、視覚情報とロボットの行動との密な結合を実現する研究例は少ない。また、能動視覚システムとして単眼視覚システムまたは、

ステレオ視覚システムを構築することが多い。ステレオ視覚システムでは、1) 単眼視覚システムに比べ、カメラが一つ増加することにより、それによって得られる情報量が増え、システムのロバスト性が向上する。2) システム自体の運動が、瞬間的に得られる視覚情報に悪影響を及ぼさない (単眼視覚システムでは、視覚情報としてオプティカルフローを利用することが多く、これはシステム自体の運動の影響を大きく受ける) といった利点を持つ。

能動視覚システムを持つ移動ロボットの研究例として、Sandini ら [6] は、ステレオ視覚を持つ移動ロボットを用いて廊下を走行させる実験を行っているが、ロボットの行動はシステムの設計者によって記述された制御則によって規定されており、ロボット自らが、環境との相互作用を通じて、自身の行動を改善していくことはできない。國吉 [7] らは、ステレオ視覚追跡機能として拡張ゼロ視差フィルタ法を提案し、ロバストな目標物追従行動を実現している。しかしながら、このロボットは視覚追跡能力しか持たず、障害物を回避するためには別の視覚処理機能が必要になる。最近、Huber と Kortenkamp [8] は、長時間ステレオ視覚システム [9] を使用して、障害物回避を遂行しながら、移動目標物を追跡した。しかしながら、彼らのステレオ対応アルゴリズムは、ラブラシアン-ガウシアンフィルタによって、抽出されるエッジを基にしているため、以下のような欠点を持っている。視覚システムは、現在の目標物体領域よ

原稿受付 1995 年 11 月 22 日

*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

りもテクスチャの多い領域に簡単に引き寄せられてしまうので、目標物体を見失いやすい。また、目標領域が他の物体によって隠されるような状況に対処することができない。このシステムは三次元空間内で一定距離を保って対象を追跡しようとするので、対象やロボット自身の運動による対象画像（エッジの集合）の大きさの変化に対処できない。

ロボット学習の分野では、ロボッ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[10][11]は少ない。これらの研究では、センサとしてバンパーセンサ、ソナーなどの近接センサのみを使用しているので、行動が局所的かつ、反射的で、簡単なタスクの実験しか行っていない。視覚情報は、大局的な情報を必要とする目標指向型のタスクではより重要であるにもかかわらず、ロボット学習において視覚情報を用いた研究は、視覚情報のセンシングおよび処理コストのために、さらに少ない。

本論文では、ステレオ視覚を持つ移動ロボットが、あるタスクを実行するのに十分なだけの視覚情報を検出し、それらを用いてタスクを達成する行動を学習する手法を提案する。この手法をここでは、ステレオスケッチと呼ぶ。また、タスク達成に必要な視覚情報を検出することを「視覚行動」と呼ぶ。ここで用いられた視覚行動は、(1) 実時間ステレオ対応手法として、濃淡画像を用いた局所相関演算[12]を使用し、この手法により、ステレオ・運動視差の推定性能が、エッジベースの手法よりも改善される。(2) 粗密戦略[13][14]に基づくステレオ対応探索から得られる情報を基にした領域分割・領域追跡により、対象やロボットの運動による大きさの変化に対応することができる。(3) 時系列で画像分割を繰り返すことで、ある時刻での画像分割結果の誤りに対処することができる、という特徴を持つ。また、動作系列の集合を「モータ行動」と呼ぶ。後で述べるように、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コマンドを送ることができる。ここでは、単位時間あたり（ここでは、画像処理のサンプリング周期に合わせて1/30秒）に発生されるモータコマンドの実行結果として得られるロボットの運動を「動作」と定義する。

ここで提案する行動学習の枠組みを持つことにより、視覚行動、モータ行動を通じて、ステレオ視覚を持つ移動ロボットが環境と相互作用しながら、自身の移動機構や、自分のおかれた状態とタスクの目的に応じて、合目的な行動を獲得する。ここでは、移動ロボットの一般的なタスクの1例として、目標物体到達タスクを取り扱う。

以下、次章では、ステレオスケッチの基本的な処理の流れについて述べ、次に、視覚行動、行動学習の方法、状態空間の定義について説明する。そして、コンピュータシミュレーションと実機による実験結果を提示し、最後に、考察および今後の課題について述べる。

2. ステレオスケッチ

我々はこれまで、運動スケッチ[1]を提案し、それによって、単眼視覚を持つ学習ロボットに、実時間視覚追跡ルーチンを使用し環境内で障害物を回避し目標物体に到達する行動を獲得させることに成功した。

しかし、運動スケッチでは、以下のような問題があった。

- 障害物を検出するために、現在のシーンとあらかじめ障害物がない環境で得られたオブティカルフローとの差を使用しているが、静止障害物は、小さなフローの差しか検出できないので、障害物を検出できないときがある。

ステレオスケッチでは、この問題を克服するために、ロボットにステレオカメラを搭載し、同じ視覚追跡ルーチンを使用し、実時間ステレオシステムを実現する。本論文では、従来のステレオ視覚システムに多く見られるように、多くの時間を消費して正確な三次元構造を復元するのではなく、望ましい行動を引き出すために必要な情報を実時間で抽出するステレオ視覚システムを構築する。ここでは、目標物体到達タスクを取り扱うため、視覚行動として、ステレオ視差を求める視覚処理と、その処理によって生成される視差マップを基に領域分割・追跡する視覚処理を準備する。それらの視覚行動によって得られる領域分割結果と運動視差情報によって、追跡性能が改善されるだけでなく、オクルージョンやディスオクルージョンといった有益な情報を得ることができる。そのような抽象度の高い情報を基に強化学習のための状態空間を構成すると、状態数を少なくすることが可能になる。モータ行動としては、ロボットが実行可能なモータコマンド群を準備する。ただし、それらのコマンド間の物理的意味やお互いの関係は未知であると仮定する。そして、ステレオ・運動視差の推定によって同定されるオクルージョンの状態等から構成される状態空間を用いて、その視覚行動から得られる環境状態とモータ行動間の関係を求め、目標物体領域に到達する行動を獲得する。学習の結果、ロボットは障害物を陽に記述することなしに障害物を回避し、目標物体領域に到達することが可能になる。Fig. 1 は、ステレオスケッチの基本的な処理の流れである。

ステレオスケッチを基にした行動獲得手法は、以下のような特徴がある。

- [1] 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みが構成できる。
- [2] タスクを達成するためにカメラキャリブレーションや三次元構造の定量的な再構成を一切必要としない。
- [3] 学習の結果得られた行動を観察者の立場である我々が形容する場合に、「観測行動」とみなせる行動が、立案することなく獲得される。

3. 視 覚 行 動

本論文で使用されている視覚行動の処理は、以下のような手順から構成される。

- (1) ステレオ画像から、視差値を粗密戦略に基づくステレオ対応探索によって求める。
- (2) MDL 原理に基づく改良 ISODATA アルゴリズムによって、

Fig. 1 Stereo sketch

特徴パラメータ（画像座標・視差値）を基に入力シーンを領域分割する。そして、領域分割された領域の中から目標物体領域を指定する（この指定は、例えば、人間が行ってもよいし、ロボットに最も近い領域として、自動的に選択してもよい）。

- (3) 目標物体およびその近傍領域を、連続するフレーム間で追跡し、領域分割結果と運動視差情報に基に、目標物体領域の隠蔽状態を判定する。

3.1 ステレオ対応探索

粗密ステレオ対応探索手法 [12] [13] を、左右画像間で SAD (Summation of Absolute Difference) に基づく局所相関演算による実時間視覚追跡ルーチン [14] を用いることによって実現する。第 1 段階の対応探索では、低解像度の画像を 8×5 個の格子状領域（一つの格子領域の大きさは、 16×16 画素）に分け、一つの格子領域に対する対応探索領域を 64×24 画素に設定した。第 2 段階の対応探索では、高解像度の画像を 16×10 個の格子状領域（一つの格子領域の大きさは、 16×16 画素）に分け、一つの格子領域に対する対応探索領域を、低解像度の画像での対応点位置を中心に配置し、その探索領域の大きさを 32×24 画素に設定した。

Fig. 2 は、本手法によって得られたステレオ対応の結果を示している。3 組の左右画像が 3 列になって表示されている。第 1 段階として、 128×120 画素の大きさの画像と、第 2 段階として、 256×240 画素の大きさの画像を使用した。現在、ハードウェア上の制限から、最も粗い画像 (64×60 画素) を用いていない。最終的な対応結果を、図中右下に、視差マップとして示している。

Fig. 2 A coarse-to-fine stereo matching

3.2 領域分割と追跡

目標領域に到達するために、ロボットは、常に追跡する領域を同定する必要がある。その領域は、部分的または完全に他の物体によって隠蔽されることがあるので、このような状況になってもロボットに追跡するために、領域分割・追跡ルーチンを準備する。

3.2.1 領域分割

格子状に分けられた入力シーンを領域分割するために、各格子ごとの特徴が、次のような三次元特徴空間を構成すると考える。

$$\mathbf{x} = (\text{image_location}(x, y), \text{disparity}(d)).$$

ここでは、特徴空間内の距離の定義として、マハラノビス距離を使用する。入力シーン中のすべての格子領域を、ここで提案する修正 ISODATA クラスタリングアルゴリズムによって、いくつかの領域に分割する。ISODATA クラスタリングは、繰り返しアルゴリズムである k 平均クラスタリング手続きと発見的な評価尺度によるクラスタの分割・併合の手続きからなる [15]。ここでは、最小記述長 (MDL) 原理 [16] に基づく記述長の最小値を探索する評価尺度に置き換えることで、特徴ベクトル空間を自己組織化する。MDL 原理の詳細な手順は、文献 [16] を参照。

ここで提案するクラスタリング手法は、次の二つのステップを繰り返すことにより行われる。1) 分割 (併合) により、MDL が減少するクラスタがあれば、そのクラスタを分割 (併合) する。2) k 平均クラスタリングを行う。クラスタリングは、小数のクラスタから出発して MDL を減少させながらすべてのクラスタについて、MDL の減少が飽和するまで領域を分割する (付録 6 参照)。このアルゴリズムにより、自動的に画像分割を行う。

3.2.2 領域追跡

視覚追跡ルーチン群によって得られるオプティカルフロー (運動視差情報) と時刻 t_i における領域分割結果とを用いて領域を追跡する。フローベクトルを、ステレオ対応探索のために設定した 16×10 格子領域の各格子領域で求める。時刻 t_i における特徴パラメータの標本時系列を $\mathbf{x}(t_i) = (x(t_i), y(t_i), d(t_i))$ 、そのときの領域 $j (j = 1 \sim c)$ を記述するパラメータを $\theta_j(t_i) = (\mu_j(t_i), \sum_j(t_i))$ と表す。ここで $\mu_j(t_i), \sum_j(t_i)$ は、それぞれ、時刻 t_i において領域 j を構成している特徴ベクトル $\mathbf{x}(t_i)$ の平均値ベクトル、それらの特徴ベクトル群の共分散行列を表す。次時刻 t_{i+1} における領域記述パラメータ $\mu_j(t_{i+1})$ を $(\bar{x}(t_i) + g_x \bar{u}(t_i), \bar{y}(t_i) + g_y \bar{v}(t_i), \bar{d}(t_i))$ のように推定し、それを次フレームの特徴に関する領域分割における初期クラスタにする。ここで、 $\bar{u}(t_i), \bar{v}(t_i)$ は、オプティカルフロー検出処理によって得られた移動ベクトルを領域ごとに平均したもので、 g_x, g_y は、画像処理による遅れを補償し、次時刻における各領域の画像上での位置を線形予測するための各軸方向に関する係数である。本論文では、 $g_x, g_y = 1.2$ と設定した。

この初期クラスタと時刻 t_{i+1} での特徴パラメータをもとに、MDL 原理を用いた ISODATA アルゴリズムによる再クラスタリングを実施し、時刻 t_{i+1} における領域分割が得られる。もし、目標物体領域が二つのクラスタに分割されたり、他のクラスタに併合されたりすれば、目標物体領域のモデルをこれらの状況を記述するために更新する。また、もし目標物体領域が完全に隠蔽されれば、その位置は線形予測され、その領域の視差値や、この領域の対応探索のための相関演算に用いられる参照画像は更新しない。

Fig. 3 は、障害物がない場合に目標物体に到達する過程での領域追跡の例を示している。画像内での目標物体の大きさの変化に対処でき、目標物体の追跡に成功したことを示している。

目標物体領域 $\theta_{obj}(t_i) = (\mu_{obj}(t_i), \sum_{obj}(t_i))$ においてオプティカルフローを検出する際に、目標物体領域の一部で非対応

Fig. 3 Region tracking

となり、その近傍で正しく対応がつく状態のときにオクルージョンが発生していると判断する。対応・非対応の判定は、相関演算時に得られる最小誤差値を基に以下のように行う。

$$\begin{cases} D_{\min}[i, j] \geq \text{Thresh} & \rightarrow \text{非対応領域} \\ D_{\min}[i, j] < \text{Thresh} & \rightarrow \text{対応領域} \end{cases}$$

ただし $D_{\min}[i, j]$ と Thresh は、それぞれ最小 SAD 値と閾値を表す。

4. 行動学習

行動獲得のための学習方法として、最も広く使用されている強化学習手法の一つである Q 学習法 [17] を使用する。本論文では、ロボットがオクルージョンを回避しながら目標物体に到達するタスクを想定している。Q 学習をこのタスクに適用するために、上述した視覚行動により得られる目標物体や環境に関する記述から構成される状態空間を定義しなければならない。本章では、まず最初に、強化学習の基本について述べ、次に学習のための状態空間をどのように構成するかについて説明する。

4.1 強化学習の基礎

強化学習ロボットは、環境から受け取る報酬や罰を通して、タスクに対する性能を改善する。それらは、ロボットが間違えた行動をしたときに、その状況に対する正しい反応をエージェントに教える教師がないという点において、教師付き学習ロボットとは異なる。現在のタスクに対する性能を示すロボットのフィードバックは、スカラーの報酬値のみである。Q 学習アルゴリズムは、動的計画法 [18] の枠組から導かれたアルゴリズムである。ここでは、以下の 1 ステップ Q 学習アルゴリズム [17] を用いた。

初期化: $Q \leftarrow$ 行動価値関数の初期値を代入 (通常はゼロ)。

Fig. 4 State space construction

繰り返す：

- (1) $s \leftarrow$ 現在の状態.
- (2) 政策 f に従って行動 a を選択し (任意でもよい), 実行する. 次状態を s' , 即座の報酬を r として,
- (3) $Q(s, a)$ を更新:

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

- (4) f を更新:

$$f(s) \leftarrow a \text{ such that } Q(s, a) = \max_{b \in \mathbf{A}} Q(s, b)$$

4.2 状態空間の構成

状態空間は、与えられたタスクを達成するために必要十分な情報を含んでいる必要がある。一方、ロボットと環境に関する先見的な知識を必要としない Q 学習の学習時間が、状態空間の大きさの指数オーダー [19] になるために、できるだけ状態空間の大きさは小さい方がよい。

そこで、視覚行動によって得られるステレオ視差や運動視差に基づいて識別された目標物体領域やその近傍領域の隠蔽の状況を組み合わせることによって状態空間を構成する。

また、本論文では、環境内に存在する障害物は床面上に存在し目標物体の上側または下側だけが隠蔽されることはない仮定する。障害物の三次元形状に関しても特異な形状 (ロボットの視線方向に大きく凸であるような形状) をしていないものとする。また、目標物体領域の左右両側で隠蔽は発生しないものとする。以上のことを踏まえ、以下のように状態空間を構成した。見え方 (左側隠蔽, 右側隠蔽, 完全隠蔽, 隠蔽なし), 位置 (左, 中, 右), 視差値 (大, 中, 小) を基に状態空間を定義する。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) を状態空間に付加する (Fig. 4 参照)。結局,

- 左・右画像ともに、目標物体が観測される場合。上で定義した、見え方、位置、視差値をすべて考慮しなければならないので、全部で $(3 \times 3 \times 4)^2$ 通り。
- 左画像のみもしくは右画像のみで、目標物体が観測される場合。3 通りある視差値のどれかの状態になった後に、どちらかの画像のみで目標物体が観測される。また、片方の画像のみの情報なので、この状態になった時刻における見え方の情報は得られない。したがって、どの位置にあるかで見失った状態を考慮して、全部で $(5 + 5 + 5) \times 2$ 通り。

Table 1 An example state

image	appearance	position	disparity
left	right side is occluded	center	any
right	completely occluded	center	any

- 左・右画像ともに観測されない場合。1 通り。

の合計 1327 の状態からなる状態空間を構成した。

この状態空間の一つの特徴は、障害物に関する陽な記述を含んでいないという点である。つまり、目標物体領域の隠蔽の状態が、障害物に関する状態を間接的に示している。Table 1 は、障害物が目標物体の右側前方にある場合に対応する状態空間内の状態記述の例を示している。

前述した状態空間は、完結しているように見えるが、センサの知覚能力の限界のために、「知覚的見せかけ問題」[20] を含んでいる。この問題は、一般的に、「異なる実際の状況が、一つの観測状態に投影されることにより生じる問題」とであると定義される。複数の状態が 1 状態に写像されることになり、ロボットが最適に行動することを困難にする。したがって、目的行動を獲得するためには、そのような状態 (「隠れ状態」と呼ぶ) を発見しなければならない。そこでここでは、一度学習を行い、その後に、環境状態の遷移に関するモデルとしての状態遷移確率を推定し、それを用いて隠れ状態を発見する。状態遷移確率 $P_{ij}(a)$ の推定は、以下の最尤推定によって求める。

$$P_{ij}(a) = \frac{n_{ij}^a}{\sum_{j \in \mathbf{S}} n_{ij}^a},$$

ここで、 n_{ij}^a は、状態 i で行動 a をとったときに状態 j に遷移した回数を示す。学習中にこれらの履歴を保存し、ある程度学習を行った後に、状態遷移確率 $P_{ij}(a)$ を推定する。そして、各状態 i について状態遷移確率の分布が多峰性になっていないかを調べ、多峰性になっている状態を隠れ状態の候補として識別する。このようにして識別された隠れ状態の候補それぞれに、1 個のメモリを持たせる。そして、一度最適政策に従って行動し、隠れ状態の候補状態へどの状態から遷移したかを記憶する。その際、以前に記憶された状態と、異なる状態からの遷移がある場合、その候補状態を、隠れ状態として識別する [21]。隠れ状態における動作は、識別された隠れ状態を新たに状態空間内に付加し、再び Q 学習を行って、新たに得られた政策に従って決定される。

5. 実験結果

実験は二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。

5.1 シミュレーション結果

5.1.1 行動空間の定義

タスクに対して Q 学習を適用するためには、行動空間を定義する必要がある。ロボットは、環境の現在の状態に対して取るべき、ある一つの行動を選択することができる。ロボットは、二つの独立したモータを持つ PWS (Power Wheeled Steering) システムを用いて移動することが可能である。これら二つの

Fig. 5 An example of hidden states

モータそれぞれに対する制御命令 (ω_l と ω_r) から行動集合を構成する。それぞれのモータ命令を3段階 (前進, 停止, 後退) に量子化して, 全体で9通りの行動が選択できる。

画像情報の特性のために, すなわち, 観測者近傍での小さな変化は画像上では大きな変化として現れ, 観測者から遠方での大きな変化は, 画像上で小さな変化として現れる可能性があり, 一つの動作が常に, 一つの状態遷移に対応するとは限らない。これは「状態・行動空間のずれ問題」と呼ばれている。この問題を回避するために, 以下のように行動空間を構成した。上述で定義したそれぞれの動作を行動要素とする。ロボットは, 現在の状態が変化するまで同じ行動要素を実行する。同一行動要素の系列を一つの行動とみなす [22]。

5.1.2 隠れ状態, ゴール, 報酬

Fig. 5 は, 隠れ状態とその状態における最尤推定された状態遷移確率の例を示している。Fig. 5(a) の上部に, ロボット上のステレオカメラにより撮影された画像を示す。この図 (a) は, 右画像で, 目標物体の左側が隠蔽された状態が観測され, 一方の, 左画像で, 目標物体が観測されないという状態を表している。

視差情報が利用できないので, 目標物体が近傍でも遠方でも存在することが可能になる。これは, 左画像において目標物体が完全に隠蔽されているか, 視野から消えているかのどちらかに相当することになる。Fig. 5(b) での二つのピークは, これら二つの状況に対応している。4.2節で述べた手法により, これらの隠れ状態 (12 個) を発見し, 新たに状態空間に加える。

ゴール状態を Table 2 に示す。ロボットがゴール状態に到達した場合に報酬 1 が与えられ, それ以外の場合は報酬 0 が与えられる。ロボットが障害物に衝突した場合には, マイナスの報酬を与えずに, ロボットを初期状態に戻す (リセットする)。これは, マイナスの報酬が, Q 値の分布に多くの極大点を発生させ, 正しく学習できないことによる。リセットすることや報酬 0 を与えることによって, 学習時間は長くなるが, 「リセットした」, または「報酬 0 を与えた」ときの状態が望ましくない状態であることを間接的に示すことになる。

コンピュータシミュレーションは, 実際の移動ロボットの大きさ, 重さ, 床面の摩擦係数, カメラパラメータ (焦点距離, 傾き角) などを実測し, それらを反映するようにした。学習の

Table 2 A goal state

image	appearance	position	disparity
left	no occlusion	right	near
right	no occlusion	left	near

Fig. 6 Reaching behavior obtained by the learning method

収束時間は, SGI のインディゴ Elan (CPU:R4000) で, 約 12 時間である。

Fig. 6 は, 提案した学習法による目標物体到達行動の例を表している。各図 ((a), (b)) の下部に, 初期位置でのロボットへのステレオ入力画像を右, 左画像をオーバーラップさせて表示している。目標物体と障害物の相対的配置を変えても ((a), (b)), ロボットは, 障害物に衝突することなく目標物体に到達することに成功している。Fig. 6(a) の例では, いったん, 目標物体のよく見える位置まで後退してから, 障害物を避けて目標物体に到達するという行動が学習により獲得された。人間が観測者の立場で見たとき, ロボットが後退せずに, 直角に旋回し, 直進したほうが良いように見える。しかしながら, 我々のロボットは, 現在位置や障害物までの距離などの定量的な三次元幾何情報を陽に用いず, 観測された視覚情報と隠れ状態識別のためにたかだか数ステップのメモリしか持っていないため, このロボットにとっては後退行動が自身にとって最適となっている。更に, このロボットは, ノンホロノミックな移動機構であるため, 直角に旋回するために数ステップ必要とする。また, 目標物体に視線を合わせる行為等を含めると, 更に数ステップの動作が必要になると考えられる。一方, 本論文で学習の結果得られた行動の「後退行動」は, 1 ステップの行為でオクルージョンを回避することが可能で, この意味でも, この行動が最適である。したがって, 本手法に基づいた学習の結果, ロボット自体が積極的に視点を移動させ, オクルージョンの回避ができる視点の位置を設定するという行動が, 立案することなく獲得された。

運動スケッチ [1] では, 視覚行動により得られる情報を移動ロボットの行動を制御するためだけに使用した。視覚を, いわゆる「行動のための視覚」(Vision for Action) として使用していた。これに対して, ここでは, 「視覚のための行動」(Action for Vision), つまり行動することにより, 視覚センサの位置や

Fig. 7 Distribution of Optimal action for reaching behavior

方向をダイナミックに制御し、視点の位置、視線方向を変える事によって観測できる空間を広げ、得られる視覚情報の質を向上させるような行動が学習により獲得された。

Fig. 7 は、Fig. 6 (a) の環境でロボットの初期位置をランダムに変えて行動を開始させたときの、動作系列の履歴を示している。図中の→は、ロボットの動作方向を示している。図で示されているように初期位置を変えても、ロボットが障害物回避に成功することを示している。

5.2 実ロボットによる実験

Fig. 8 に、実ロボットのシステム構成を示す。ステレオカメラが移動ロボット（山彦）上に設置されている。山彦は、RS232Cを通してMVME167上のVxWorks OSによって制御される。ステレオカメラのベース間距離は、約17[cm]で、カメラの光軸はほぼ平行に保たれ、俯角は、約10°である。ステレオカメラの設置高さは、60[cm]で、移動ロボットの移動最大速度は、60[cm/s]である。ロボット上のステレオカメラで取られた画像は、ビデオ送信器でホストの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。1枚の視覚追跡処理ボードは、約140個のターゲットをビデオレートで追跡できる能力を持っている。現在3枚の視覚追跡処理ボードを使用している。しかし、粗密戦略によるステレオ対応探索、MDLに基づく領域分割、追跡処理は、現有のCPUボード（MC68040）にとって負荷が大きすぎ、現在の処理サイクルは160[ms]（約6[Hz]）である。

Fig. 9 に、実ロボットが障害物を回避して目標物体に到達する動作系列の一つの例を示す。この図では、六つの時刻での画像処理結果と環境の様子を示している。Table 3 は、Fig. 9 に示した目標物体到達行動に対応する状態識別結果である。Table 3 では、ロボットが識別したstate step, 目標物体に対する左右画像上での状態（左隠蔽（OL）、右隠蔽（OR）、完全隠蔽（OC）、隠蔽なし（NO）、左存在（L）、中央存在（C）、右存在（R）、視差大（N）、視差中（M）、視差小（F））、左右輪に対する制御コマンド（前進（F）、停止（S）、後退（B））が示されている。表中で“*”印のついている動作系列がFig. 9 に示されている。表中の（ ）内の数字は、処理ステップを示している。ランダムなノイズなどが原因である1ステップでの状態識別を誤る場合

(a) A mobile robot

(b) A system architecture

Fig. 8 Our robot and system

Table 3 State-action data

state step	state		action	
	left	right	L	R
1*(1-11)	(NO,M,C)	(NO,M,C)	F	F
2*(12-14)	(OL,M,C)	(NO,M,C)	S	B
3(15-16)	(OL,M,C)	(NO,M,C)	F	B
4(17)	(OL,M,C)	(NO,M,L)	F	B
5*(18)	(OL,M,L)	(NO,M,L)	F	F
6*(19-21)	(NO,M,L)	(NO,M,L)	S	F
7*(22-28)	(NO,M,C)	(NO,M,L)	F	F
8(29-38)	(NO,N,C)	(NO,N,L)	F	F
9(39)	(NO,N,C)	(NO,N,C)	S	F
10*(40-44)	(NO,N,C)	(NO,N,L)	F	F

はあるが、誤認識の原因が系統的なものである場合を除けば、時系列の情報を利用して連続して画像分割を行い状態識別しているので、続けて状態識別を誤る確率は小さく、ロバストな状態識別が行える。Table 3 のstate step 2 で後退の動作が取られている。これは5.1.2節で述べた「視覚のための行動」が起動されていることを示している。

6. 考察および今後の課題

ステレオ視覚を持つ移動ロボットに目標到達行動を獲得させる手法としてステレオスケッチを提案した。ステレオスケッチは、高度な視覚行動と行動学習過程から構成される。視覚行動から得られる情報の抽象度が高いため、学習過程では、障害物を陽に記述しないコンパクトな状態空間を構成できた。

視覚を持つ移動ロボットには、視覚により環境内から個々のタスクに応じた情報を抽出する能力が必要である。それにかかるコストが、実世界で行動するロボットにとって不利となる場合がある。したがって、タスクやロボットの行動に根ざした環境の記述を最初から作成することが必要であると考えられる。ここでは、目標物体に到達するというタスクに対する重要な情報として、環境内からオクルージョンの状態を抽出して、これを用いて強化学習のための状態空間に構成し、タスク達成のための行動に関する学習を行った。本論文で扱った行動には、明確な位置識別を必要とするタスクのための地図生成行動は、含まれていない。また、当然のことながら、その地図を用いた経

Fig. 9 Reaching behavior by the real robot

路探索行動や経路計画行動は含まれていない。しかし、これらの行動を獲得するには、本論文で獲得された行動を基本にして、目標状態をタスクに応じて動的に決定する機構、局所的な視覚情報から広域的な環境状態を再構成する機構等を新たに設定すればよい。例えば、地図生成行動は、本論文で獲得された行動を基本にして、目標物体領域として地図生成過程における未探索領域を設定する機構、さらに、局所的に得られた視覚情報を統合するような機構を準備することで、獲得できると考えられる。

実ロボットの実験では、シミュレーションで得られた最適政策を実ロボットが実行するのみで、実ロボットによる学習は行っていない。実ロボットでの実験では、視覚処理が状況を誤識別

しない限り、シミュレーションでの行動と一致すると考えられるが、必ずしもシミュレーションで想定した状況だけが出現するとは限らないので、タスクが達成できなくなる可能性がある。これに対しては、シミュレーションで得られた学習結果を初期値として、実ロボットで学習を行えば、実環境中で出現する状況に適した行動を獲得することが可能であると考えられる。

最後に、本論文での仮定や設定について検討する。

- 「目標物体領域の左右両側で隠蔽は発生しない」: この仮定をなくした場合には、左右の隠蔽のどちらか一方を選択するという機構を、視覚行動の中に組み込むことで対処できると考えられる。
- 「移動目標物体、移動障害物の存在する環境を考慮していない」: それらの移動量を記述する状態を状態空間へ組み込むことにより、対処できると考えられる。
- 「視覚行動自体の学習については考慮していない」: 今後は、視覚行動の学習に、自身の行動からのフィードバックを直接的に利用して、生の視覚情報から高次の概念を学習により獲得する問題を扱う予定である。

参 考 文 献

- [1] Takayuki Nakamura and Minoru Asada: "Motion sketch: Acquisition of visual motion guided behaviors," In Proceedings of International Joint Conference on Artificial Intelligence, pp.126-132, 1995.
- [2] Y. Aloimonos, I. Weiss and A. Bandyopadhyay: "Active vision," In Proc. of first ICCV, pp.35-54, 1987.
- [3] Y. Aloimonos: "Reply: What i have learned," CVGIP: Image Understanding, vol.60, no.1, pp.74-85, 1994.
- [4] G. Sandini and E. Grosso: "Reply: Why purposive vision," CVGIP: Image Understanding, vol.60, no.1, pp.109-112, 1994.
- [5] S. Edelman: "Reply: Representatin without reconstruction," CVGIP: Image Understanding, vol.60, no.1, pp.92-94, 1994.
- [6] G. Sandini: "Vision during action," In Y. Aloimonos, editor, Active Perception, chapter 4. Lawrence Erlbaum Associates, Publishers, 1993.
- [7] 國吉: "ステレオトラッキング視覚を搭載した小型移動ロボット", 日本ロボット学会誌, vol.13, no.3, pp.343-346, 1995.
- [8] E. Huber and D. Kortenkamp: Using stereo vision to pursue moving agent with a mobile robot. In Proc. of IEEE Int. Conf. on Robotics and Automation, pp.2340-2346, 1995.
- [9] K. Nishihara: "Practical real-time imaging stereo matcher," Optical Engineering, pp.23-5, 1984.
- [10] P. Maes and R.A. Brooks: "Learning to coordinate behaviors," In Proc. of AAAI-90, pp.796-802, 1990.
- [11] 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.
- [12] D. Marr and T. Poggio: "A computatinal theory of human stereo vision," In Proc. of Royal Soc. London B204, pp.301-338, 1979.
- [13] W.E.L. Grimson: "A computatinal theory of visual surface interpolation," In Proc. of Royal Soc. London B298, pp.395-427, 1982.
- [14] 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.
- [15] G.H. Ball and D.J. Hall: "ISODATA, a novel method of data analysis and pattern classification," Stanford Research Institute, AD-699616, 1965.
- [16] J. Rissanen: Stochastic Complexity in Statistical Inquiry,

World Scientific, 1989.

- [17] C.J.C.H. Watkins: Learning from delayed rewards, Ph.D. thesis, King's College, University of Cambridge, May 1989.
- [18] R. Bellman: Dynamic Programming, Princeton University Press, Princeton, NJ, 1957.
- [19] S.D. Whitehead: "A complexity analysis of cooperative mechanisms in reinforcement learning," In Proc. AAAI-91, pp.607-613, 1991.
- [20] S.D. Whitehead and D.H. Ballard: "Active perception and reinforcement learning," In Proc. of Workshop on Machine Learning-1990, pp.179-188, 1990.
- [21] R.A. McCallum: "Instance-based utile distinctions for reinforcement learning with hidden state," In Proc. of the 12th Int. Conf. on Machine Learning, pp.387-395, 1995.
- [22] M. Asada, S. Noda, S. Tawaratsumida, and K. Hosoda: Vision-based reinforcement learning for purposive behavior acquisition. In Proc. of IEEE Int. Conf. on Robotics and Automation, pp.146-153, 1995.
- [23] R. Duda and P. Hart: Pattern Classification and Scene Analysis, John Wiley and Sons, Inc, 1973.

付録 A. 修正 ISODATA クラスタリング アルゴリズム

ここで提案するクラスタリング手法は、次に二つのステップを繰り返すことにより行われる。1) 分割 (併合) により、MDL が減少するクラスタがあれば、そのクラスタを分割 (併合) する。2) k 平均クラスタリングを行う。クラスタリングは、小数のクラスタから出発して MDL を減少させながらすべてのクラスタについて、MDL の減少が飽和するまで領域を分割する。

概略を述べると、全記述長は、モデル式 (1) の記述長と、そのモデルに当てはめたときの当てはめ誤差の記述長の和である。ここでは、領域分割された領域内の画素の分布が、混合密度分布モデル (多変量正規分布) [23] によって表現できると仮定する。そのモデルは、以下のように記述できる。

$$p_i \left(\mathbf{x} | \boldsymbol{\mu}_i, \sum_i \right) = \frac{1}{(2\pi)^{d/2} |\sum_i|^{1/2}} \times \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^t \sum_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right\}, \quad (1)$$

ここで、 $\mathbf{x}$, $\boldsymbol{\mu}$, $\sum$ は、それぞれ列ベクトル、平均値ベクトル、共分散行列である。 i と d は、それぞれ、クラス番号、

Fig. 10 Region segmentation

ベクトルの大きさを表している。

モデルの記述長は、モデル自体と平均値ベクトル $\boldsymbol{\mu}$ を記述するために必要なビット数になる。当てはめ誤差の記述長は、共分散行列 $\sum$ に関するビット数になる。もし多くのクラスタがあれば、前者の記述長が後者の記述長より大きくなる。MDL 原理は、情報理論にもとづいて記述長を最小化することによって、モデル自身の記述長とモデルへの当てはめ誤差の記述長とのトレードオフを取って、与えられたデータのある確率モデルに当てはめる際に最適な基準を与える。本手法では、一般的な多次元正規分布式 (1) を仮定しているため、MDL の評価に次式を用いる。

$$l_R^{(i)}(\mathbf{x}^n) = -\log_K p_i \left(\mathbf{x} | \boldsymbol{\mu}_i, \sum_i \right) + \frac{d}{2} \log_K n + \log_K M, \\ l_{MDL}(\mathbf{x}^n) = \min_{1 \leq i \leq M} l_R^{(i)}(\mathbf{x}^n),$$

ここで、 n と M は、それぞれ標本データ数、情報源のモデルの数であり、ここでは $M = 1$ or 2 であり、それぞれ、あるクラスを一つの多次元正規分布で表現する場合、あるクラスを二つの多次元正規分布で表現することに相当する。Fig. 10 は、提案した手法によって、Fig. 2 に示された画像に対する領域分割結果を示している。画像中央の目標物体を一つの領域として分割した場合の MDL 基準を評価すると、その記述長は、683.2 [bits] であり、二つの領域に分割した場合の記述長は 451.1 [bits] となる。これは、目標物体がカメラに対して平行でない配置にあったので、目標物体領域が一つの視差値で記述することができないためである。

中村恭之 (Takayuki Nakamura)

1968 年 1 月 17 日生。1996 年大阪大学大学院工学研究科後期課程修了。同年、日本学術振興会特別研究員。現在、米国ブラウン大学客員研究員。コンピュータビジョン、人工知能、ロボティクスの研究に従事。博士 (工学)。人工知能学会の会員。(日本ロボット学会正会員)

浅田 稔 (Minoru Asada)

1953 年 10 月 1 日生。1982 年大阪大学大学院基礎工学研究科後期課程修了。同年、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。1995 年、IEEE Int. Conf. on Robotics and Automation Best Paper Award の最終候補に選択される。第 10 回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、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)