

身体性に基づく対象認識

Embodiment based Object Recognition

寺田和憲 中村恭之 武田英明 小笠原司
Kazunori Terada Takayuki Nakamura Hideaki Takeda Tsukasa Ogasawara

奈良先端科学技術大学院大学 情報科学研究科

Graduate School of Information Science, Nara Institute of Science and Technology

In this paper, we propose a new architecture for recognizing objects based on a concept “embodiment” as one of primitive functions for a cognitive robot. We define the term “embodiment” as the extent of the agent’s body, locomotive ability and its sensor. According to embodiment, an object is represented by reaching action paths, which correspond to a set of sequences of movements taken by the agent for reaching the object. Such behavior is acquired by the trial-and-error method based on the visual and tactile information. Visual information is used to obtain sensorimotor mapping which represents the relationship between the change of object’s appearance and the movement of the agent. On the other hands, tactile information is utilized to evaluate the change of physical condition of the object caused by such movement.

1 はじめに

フレーム問題や記号接地問題 [5] など、記号を用いた人工知能に対する批判から、知能の実現のために環境との相互作用が重要であることが指摘され [2][4], 強化学習や遺伝的アルゴリズムを用いた手法が注目されている。これらの手法では、環境との相互作用の中からタスクを遂行するための入出力関係をボトムアップに獲得するため、トップダウンに記号を記述する必要はない。しかし、効率的な事象の記憶や、より抽象度の高いタスクの遂行のために、エージェントが抽象化されたデータとして記号を持つ利点は多い。したがって、環境との相互作用の中からボトムアップに記号表現を獲得する必要がある。本稿では、記号化を行なう上でエージェントの身体的制約が重要な役割を果たすという立場から、身体性を用いた記号表現すなわち身体表象に関する考察を行ない、実現のための手法を提案する。また視覚による対象認識を行なうために、視覚と身体表象の関係を明確にし、具体化するための手法の提案を行なう。以下では順に、身体表象の概説、視覚を用いた対象認識における身体性の役割と、視覚と身体表象の間の関係の獲得方法、身体性を用いた対象の表現方法、視覚を用いたタスク遂行における身体表象の利用方法について説明を行なう。

寺田 和憲 奈良先端科学技術大学院大学 情報科学研究科 〒 630-0101 生駒市高山町 8916-5 電話:0743-72-5265 fax:0743-72-5269 E-mail: kazuno-t@is.aist-nara.ac.jp

2 身体表象

身体表象とは、身体性に基づいた対象の記述のことである。身体性に関しては、文献 [9][11] において議論が行なわれているが、本稿における身体性の定義は、「エージェント固有の身体の広がりや運動能力、そしてそれらを知覚するセンサ」とする。エージェントにおける対象の内部記述方法として、身体性を用いた記述が妥当であるという理由を以下に3つあげる。

ひとつは「相対的記述」の必要性である。対象の持つ機能を実現するためには、対象がエージェントの身体に対して相対的に記述されている必要がある。例えば、椅子は人間にとって座るための機能を有するが、アリにとっては巨大な構造物として、またゾウにとっては単なる障害物として認識され、もはや座るための道具ではない。すなわち、椅子の持つ「座れる」という機能をエージェントが実現するためには、座れるための身体の形状と大きさ、すなわち広がり、そして、座るという行為を実現するための運動能力が必要となる。このように対象の機能はロボットの身体に密接に関係しているため、認知対象は身体性を基準として相対的に記述されている必要がある。

2つ目は、「合目的記述」の必要性である。目的を持たないシステムにおいてはいかなる記述も無意味である。例えば、座る（休息したい）という目的を持たないヒューマノイドロボットにとって、椅子は単なる台でしかない。ところが座るといふ目的のもとでは、適切な高さの台、例えば切り株も椅子として知覚される。したがって目的にしたがって対象を記述する必要

がある。

3つ目は、「視覚に独立した記述」の必要性である。認知心理実験において、プリズムなどを用いて視野を光学的に変換し、網膜に映る像を通常とは逆さにする、さかさめがねを用いた実験がある。さかさめがねをかけると、初期では位置の恒常性が失われ、対象物は全く別のところにある幻影に見える [8]。しかし、ある程度の期間がすぎると知覚的順応が生じ、見えている世界が本物であると感じられるようになる。この現象は、視覚系の変化に対して恒常性を保つ身体を用いた記述、すなわち身体表象が存在し、視覚系と身体表象の間の再マッピングが行なわれていると考えられると理解できる²。

以上の理由から、エージェントの認識において、対象は身体表象を用いて記述されている必要がある。以下の章では、上記の要件を満たす身体表象の具体的な記述方法について述べる。

3 視覚と身体の関係

次に視覚を用いた対象認識における身体性の役割と、視覚と身体との関係の記述方法について考える。人間を含んで実世界で活動するエージェントにおいて視覚は非常に重要な役割を果たす。そのために、人間の認識において視覚が他の感覚系を従えるという視覚優位の感覚系の統合が行なわれているとされてきた [7]。Marr の視覚の計算理論 [3] は、この考えにもとづき視覚を他の感覚モダリティーから切り離し、視覚の目的を2次元の画像からの3次元幾何学的構造を定量的に再構成することとしたものである。近年のコンピュータビジョンの研究において、このような3次元の再構成を行なうのではなく、環境、行動と密に関連した視覚システムの必要性が指摘されている [1]。浅田らは環境との相互作用から強化学習を用いて視覚を直接的に行動に関係付ける方法を提案した [10]。この手法は、タスクを遂行するために、明示的な環境の記述を必要とせず視覚入力から直接的に出力への対応を学習することで、環境の構造や身体性に依存したインプリシットな記述が獲得されるというものである。しかし、身体表象の記述におけるポイントとして先に挙げた「相対的記述」の観点から、運動能力だけでなくエージェントの身体性のもうひとつの属性である「広がり」も考慮する必要がある。そして広がりや直接知覚可能な唯一のセンサである接触センサを考慮した入出力関係の対応づけを考える必要がある。

実世界におけるエージェントのタスクはほとんどの場合接触を伴い、エージェントの身体の大きさ、形状、すなわち広がりや、エージェントの活動に大きな制約を与える。例えば、ロボットサッカーにおけるシュート行動では適切な姿勢からボールにアプローチして、接

²積山は、身体表象を自己身体についての視覚情報と触覚運動情報との関係を貯蔵したものとして考え、視覚との関係を含むという点で我々の言葉の定義とは異なる。

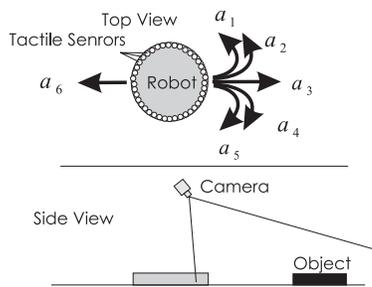


図 1: 前提とする環境とエージェント

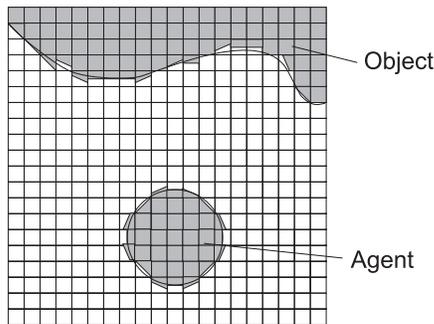


図 2: Input Image.

触を行なう必要がある。また、ロボットアームにおける対象の把持動作においては、手を適切な姿勢から対象に近づけ、適切な形で接触させる必要がある。このように、実世界におけるエージェントの行動においては必ず接触が生じ、特にタスクのゴール状態の判断において接触センサは重要な役割を果たす。翻って、タスク達成における視覚の役割を考えると、視覚はエージェントを適切な接触状態へ導くために非接触状態を評価するためのセンサだと言える。言い換えると、見ることは触ることを予測することである。

3.1 前提とするエージェントと環境

以下の説明では、視覚入力を単純化するために、エージェントとエージェントの存在する環境は2次元平面であるとし、エージェントは2次元平面内を移動するものとする (図 1 参照)。エージェントには、環境との接触を知覚するための接触センサが備えられている。また、身体と環境が接触する状態を観察するために、身体と対象の両方が視野に入る位置にカメラを備えている。カメラは、環境に固定されてもエージェントに固定されてもどちらでもよい。エージェントの行動は、 a_1 から a_6 の6個の行動とし、行動の単位は観測する状態が変化するまでとする。

3.2 視覚と触覚の関係の獲得方法

本節では、「見ることは触ることを予測することである」という観点にもとづいて、視覚と触覚の関係の

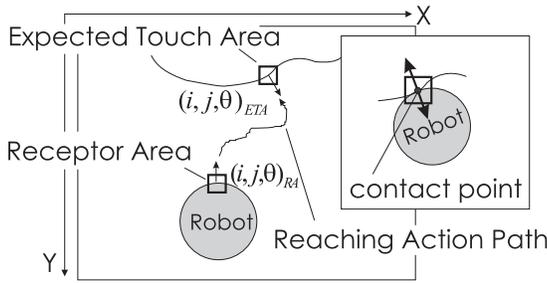


図 3: Expected Touch Area, Receptor Area, and Reaching Action Path.

獲得方法の説明を行なう．関係の獲得のための基本方針を以下に挙げる．

1. 画像は微小領域からの集合として構成される．画像から領域を抽出するというは、領域自体がすでに抽象的な意味を持つと仮定するので、ここでは画像を微小領域に分割し個々に意味付けを行ない、認識はそれらの総和として行なわれるとする．
2. 見ることは触ることを予測することであるので、微小領域は行動を通して接触信号に関係付けられる．

次に具体的な手法について説明する．図 2 のような入力画像が得られたとする．入力画像に関しては次の仮定を用いる．

1. 画像は縦横に適当に離散化され微小領域に分割されている．
2. エージェントの能力としてエッジ検出が可能であり、微小領域内に存在するエッジは角度方向に適当に離散化される．
3. エージェントは微小領域内の対象が自分の身体かどうかの判断が可能である．

ここで、対象のエッジが含まれる微小領域を Expected Touch Area(ETA) と呼ぶことにする．また、身体のエッジが含まれる微小領域を Receptor Area(RA) と呼ぶ．ETA と RA は、画像内の位置 i, j と微小領域内のエッジの法線ベクトルの傾き θ を用いて (i, j, θ) として表現する．

ETA を行動を通じて接触信号に関連させるために、検出された ETA に身体を接触させるときにとるべき行動系列を学習することを考える．このような行動系列のことを Reaching Action Path(RAP) と呼ぶ²．RAP によって身体上の点と環境の対象上の点の関係が記述

²グローバルビジョンとローカルビジョンの違い．グローバルビジョンとローカルビジョンを用いた場合では、身体を動かした時にそれぞれ RAs と ETAs が移動するという違いがあるが、ETA と RA の 2 つの領域を接触させるために出力される行動系列は同じである．

表 1: 行動とそれに対する符号

行動	a_1	a_2	a_3	a_4	a_5	a_6
符号	-2	-1	0	1	2	0

される．遷移確率が未知な場合の到達行動の学習方法として強化学習が考えられるが、ここでは、探索しながら学習するのはなく、事前に遷移確率を推定し動的計画法を用いて経路を計算する方法を採用する．

以下に動的計画法について簡単に説明する．マルコフ決定過程において効用関数 U が与えられると、最適な動作はその結果の状態の期待効用が最大となるものとして決められ、状態 i におけるエージェントの政策 f は次のようになる．

$$f(i) = \arg \max_a \sum_j M_{ij}^a U(j) \quad (1)$$

ここで、 M_{ij}^a は状態 i で動作 a がなされたときに状態 j に遷移する確率であり、 $\arg \max_a f(a)$ は $f(a)$ の値が最大となるような a の値を返す．同様に、状態の効用はその全ての可能な次の状態の効用によってあらわされる．

$$U(i) = R(i) + \max_a \sum_j M_{ij}^a U(j) \quad (2)$$

$R(i)$ は状態 i において与えられる報酬である．本手法では $R(i)$ は ETA と RA が接触する状態でのみ与えられるものとする．そのときの接触条件は、

$$(i, j)_{ETA} = (i, j)_{RA}, \wedge \theta_{ETA} = -\theta_{RA} \quad (3)$$

として与える．このようにして RAP は政策 f にしたがって行動したときに生成される行動系列 a_1, a_2, \dots, a_u によってあらわすことができる．したがって、身体表象は身体上の点と環境内の対象の点の間における最適経路、すなわち RAP の集合として記述される．

4 対象認識と判別

次に、RAP を用いて対象の形状を表現することを考える．

4.1 RAPs からの特徴抽出

形状は、対象の物理的境界の変化の様子をあらわすので、RAPs から境界の変化の割合を抽出する．まず、ひとつの RAP における 2 点間の相対角度の表現方法を考える．RAP は行動系列によって表現されているので、行動単位に回転角度に比例するように符号を割当てて RAP を符号の系列として扱う．このような方法はチェイン符号化として知られる．例えば、各行動単位に対して表 1 のような値が割り当てられていたとすると、エージェントが図 4 左のような行動系列を取った場合、その符号は図 4 右のようになる．すな

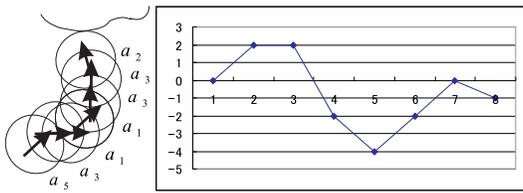


図 4: チェイン符号化の例

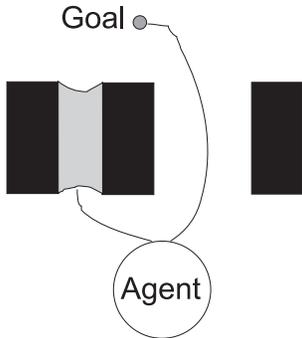


図 5: 通り抜けタスク

わち, a_1, a_2, \dots, a_u のような行動系列が存在した場合, そのコード系列は $c = \{c_1, c_2, \dots, c_u\}$ となり, その積算は,

$$C = \sum_{i=1}^u c_i \quad (4)$$

となる. 積算 C は, RAP における始点と終点の 2 点間の相対角度をあらわしている.

次に, 身体上のある点から対象の境界における全ての点の間で計算される RAPs から形状特徴を抽出することを考える. 形状は, 部分的に見ると対象の境界の変化の割合によってあらわすことができるので, 隣接する RAP から境界の相対角度 ΔC を求める. 全ての RAPs に対して計算される ΔC の系列は, 対象の境界がどのように変化しているかをあらわす. 例えば, 四角形のように直線と直角で構成される図形の場合は, $\Delta C = \{0, 0, 4, 0, 0, 4, 0, 0, 4, 0, 0, 4\}$ のようになる. また円のような場合は $\Delta C = \{1, 1, 1, 1, 1, 1, 1, 1, 1, 1\}$ のように ΔC の値が一定となり, 対象の形状に応じた表現が可能である.

5 タスク遂行時における身体表象の利用

身体表象を用いてタスクを遂行することを考える. エージェントはタスク遂行時には合目的性と相対性にもとづいて入力状態を判断する必要がある. 例えば図 5 のような通路の通り抜きのタスクを考える. 通り抜けるという目的のもとでは, エージェントは通り抜けられるかどうかの状態判断を身体に対して相対的に行なう. 通り抜けられるかどうかは, エージェントの身体の幅と通路の幅の相対的大きさによって決まる. 通り

抜けるためには, 身体の左右が対象に当たらないように行動すればよいので, 身体の左右から対象の通路側の面の間の RAP に注目すればよい.

6 まとめ

本稿では, 実世界で活動するエージェントにおける内部表現として, 身体表象の必要性を議論しその実現手法について概説した. なおタスク遂行のための詳細な手法や実験結果については [6] を参照のこと.

参考文献

- [1] Yiannis Aloimonos. What I have learned. In *CVGIP: Image Understanding*, Vol. 60, pp. 74–85, 1994.
- [2] Rodney A. Brooks. Intelligence without representation. *Artificial Intelligence*, Vol. 47, pp. 139–160, 1991.
- [3] D. Marr. *Vision*. W. H. Freeman, San Francisco, 1982.
- [4] R. Pfeifer. Symbols patterns, and behavior: beyond the information-processing metaphor. *Encyclopedia of Microcomputers*, Vol. 17, pp. 253–275, 1996.
- [5] Harnad S. The symbol grounding problem. *Physica D*, Vol. 42, pp. 335–346, 1990.
- [6] Kazunori Terada, Takayuki Nakamura, Hideaki Takeda, and Tsukasa Ogasawara. Towards cognitive agents: Embodiment based object recognition for vision-based mobile agents. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems 2000 (IROS 2000)*, accepted, 2000.
- [7] 佐々木正人. からだ: 認識の原点. 東京大学出版会, 1987.
- [8] 積山薫. 身体表象と空間認知. ナカニシヤ出版, 1997.
- [9] 浅田稔, 石黒浩, 國吉康夫. 認知ロボティクスの目指すもの. ロボット学会誌, Vol. 17, pp. 2–6, 1999.
- [10] 浅田稔, 野田彰一, 俵積田健, 細田耕. 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, No. 1, pp. 68–74, 1995.
- [11] 國吉康夫, ベルトゥーズリユク. 身体性に基づく相互作用の創発に向けて. 日本ロボット学会誌, Vol. 17, No. 1, pp. 29–33, 1999.