2E3-1

能動的な動きにもとづく知覚の恒常性の実現と行動学習

Perceptual constancy based on an active movement and its application to behavior learning

郷古 学[*1] 小林祐一[*2]
Manabu Gouko Yuichi Kobayashi

[*1]日本大学 Nihon University
[*2]東京農工大学，理研 RTC Tokyo University of Agriculture and Technology, RIKEN RTC

This paper presents a state representation using a difference between two distributions of sensor output. The difference written in the form of f-divergence is invariant to invertible transforms. To confirm the effectiveness of the proposed state representation, we conducted experiments using a mobile robot. The result showed that the proposed representation could express the similar state from a converted sensor output.

1. はじめに

一般に，ロボットはセンサ情報にもとづき，外部環境を状態として表現する (state representation)．そして，状態-行動写像により行動を生成しタスクを達成する．近年，様々な環境への適応を目指し，学習による状態-行動写像の自律的獲得に関する研究が盛んに行われている．

外界を状態としてどのように表現すべきかは，タスクに応じて異なる．例として，照明条件が変化する環境と，センサとしてカメラを持つロボットを考える．ロボットが同じ位置・姿勢 (以下，実状態と呼ぶ) で外界を観測したとしても，照明条件の変化により，視覚情報には”差異”が生じる．ロボットのタスクが，照明条件に依存する場合は，差異が識別可能な状態表現を用いる必要がある．一方で，照明条件によらないタスクの場合には，差異の影響が小さくなるような状態表現を用いることで，照明の変化によらず，同一の状態-行動写像によりタスクを達成することができる．本研究では，実状態は同一であるにもかかわらず，センサ情報が変化するような環境の変化を対象とし，このような変化に対しても同一状態と知覚することが可能な知覚システムの実現を目指す．

センサ情報 z_A が環境変化により $z_B(=g(z_A))$ へと変化したとする．ただし，環境変化の前後で実状態は同一とする．前述の例では，照明条件がセンサ情報へ与える影響 (g) を事前知識として用いることで，z_A と z_B を同一状態へと対応させることができる．しかし，事前知識のない新規環境への対応は困難である．また，設計者が環境の変化を想定し，それに対して不変となる量を状態として用いる方法 [伊藤 06] も考えられるが，想定外の変化には対応できない．

人間の知覚には，外界の観測対象から得られる刺激情報が変化しても，対象の特性を同一に保つ性質があることが知られている．たとえば，長方形の板は，視点によって得られる視覚刺激 (網膜像) が様々な台形に変化するが，観測者は板が長方形であることを知覚することができる．このような，刺激 (センサ) 情報は異なるにもかかわらず，同一性を保持する知覚の性質は恒常性と呼ばれている [柿崎 93, 乾 95]．

K. Koffka[Koffka] や J.J. Gibson[Gibson 79, 佐々木 94] は，恒常性はあらかじめ記憶しておいた「様々な視点における長方形の見え方」のような，事前知識によるものではないと主張している．特に Gibson は，恒常性の発現メカニズムとして，知覚主体 (観察者) が自身の能動的な行動により，刺激情報に変化を生じさせ，その変化パターンから観測対象に関する不変的な特徴 (不変項：invariant) をピックアップしていると主張しており，M.T. Turvey が行ったダイナミックタッチの実験結果は [Turvey 96]，この主張を支持している．

以上の考察のもと，本研究ではロボットの能動的な行動を利用することにより，同一の実状態から得られる異なるセンサ情報を，同一状態と知覚可能な恒常性を有する状態表現を提案する．

能動的な動きにもとづく知覚を扱った従来研究として [Nakamura 95, Duchon 98, 寺田 03] がある．これらの研究では，ロボット自身の動きにより生じる視覚情報の変化が，環境変化の影響を比較的受けにくいことに着目し，状態表現に利用している．中でも [Nakamura 95, Duchon 98] では，オプティカルフローを用いて移動ロボットの障害物回避タスクを実現している．しかし，一般に視覚情報の利用は計算コストが高くなるという問題があり，また，これらの研究では，環境変化によるタスク実現への影響については検証されていない．

これまでに，著者らは距離センサを持つロボットを用い，ロボットが微少時間動くことで生じるセンサ情報の変化 (差分量) を基礎とする状態表現を提案し，環境変化の前後で，同一の状態-行動写像によりタスクの実現が可能であることを示した [郷古 09]．しかし，この方法は環境変化の影響，すなわち $z_A = g(z_B)$ の g が平行移動の場合にしか対応することができない．

これらの従来研究を踏まえ，本研究では，複数の距離センサを有する自律移動ロボットを用い，より一般的な環境変化 (より一般的な変換 g) に対応可能な状態表現を提案する．提案する状態表現は，センサ出力が確率分布であると仮定し，各センサ出力分布の分布間距離を利用するものである．本稿では，強化学習による行動学習のシミュレーションを通じて，提案する状態表現の有効性を検証する．

2. 提案する状態表現と行動学習

本章では，まず提案する状態表現について概説し，続いて，自律移動ロボットへの適用について説明する．そして最後に，強化学習による状態-行動写像の獲得 (行動学習) について述べる．

連絡先: 郷古 学，日本大学工学部 電気電子工学科，
福島県郡山市田村町徳定字中河原 1 番地，
gouko@ee.ce.nihon-u.ac.jp

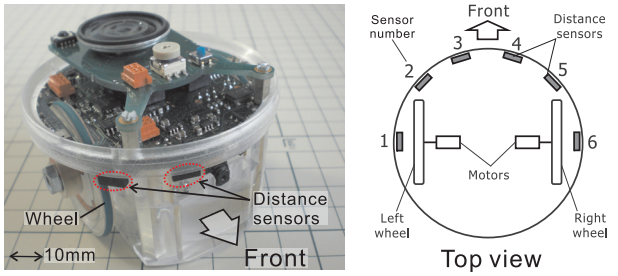

図 1: e-puck.

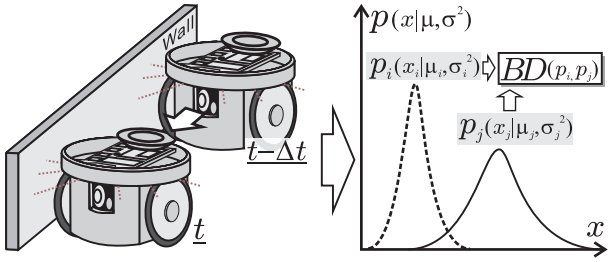

図 2: Proposed state representation using e-puck.

2.1 分布間距離の変換不変性と恒常性の実現

統計学や情報理論において，二つの確率分布の違いを計る量として，f-divergence(以下，f-div.，分布間距離と記す) と呼ばれる尺度が知られている [Csiszar 04]．二つの分布 $p_i(x)$, $p_j(x)$ の f-div. は次式で定義される．

$$f_{div}(p_i(x), p_j(x)) = \int p_j(x) f\left(\frac{p_i(x)}{p_j(x)}\right) dx \qquad (1)$$

ここで，$f(y)$ は凸関数で $y > 0$，$f(1) = 0$ である．Y. Qiao らは，式 (1) の形式で表現される二分布の分布間距離は，各分布に同一な連続かつ可逆な変換を行っても不変となることを示し [Qiao 08]，峯松は，このような変換不変性を音声認識分野に応用し，話者に依存しない音声表象を実現した [峯松 08]．

本研究では，このような f-div. の変換不変性を状態表現に利用する．センサ情報が確率分布であるとするならば，ある環境下で観測される二つのセンサ情報 z_A，$z_{A'}$ の分布間距離 $f_{div}(z_A, z_{A'})$ と，変化した環境における $z_B = g(z_A)$ と $z_{B'} = g(z_{A'})$ との分布間距離 $f_{div}(z_B, z_{B'})$ とは g が連続かつ可逆な変換であるならば，同一となることが期待できる．

著者らは，このような考えにもとづき，変化前後のそれぞれの環境下で，ロボットに異なる二つの姿勢でセンサ情報を取得させ，それらの分布間距離を状態表現として用いる方法を提案してきた [郷古 09]．しかし，この方法ではセンサ情報 (分布) の取得のために各姿勢を一定時間維持する必要があった．この問題に対し，本研究では複数のセンサを持つロボットを用い，ロボットが運動中に観測される各センサ出力同士の分布間距離を用いて状態を表現する．これにより，観測のために同一姿勢を一定時間維持する必要がなく，より円滑な行動生成が期待できる．

2.2 提案する状態表現と自律移動ロボットへの適用

本節では，次章の実験で用いる自律移動ロボットを例に，提案する状態表現の具体的な実現方法について説明する．図 1 に実験で用いる自律移動ロボット e-puck(EPFL 製) を示す．ロボットは計測範囲が 40mm の 8 つの赤外線距離センサを持っている．ただし，後述の実験では図 1 右に示す 6 つのセンサのみ用いる．

図 2 に提案する状態の求め方を示す．まず，ロボットが微少時間 Δt 動く間に，各センサ毎に M 個の出力値を取得する．続いて，得られた各センサ出力分布の分布間距離を求める．なお本研究では，各センサ出力の分布を一次元単一ガウス分布であると仮定し，分布間距離は f-div. の一種であるバタチャリヤ距離 (Bhattacharyya distance 以下 BD と記す) を用いて計算する．二つのセンサ i，j の出力分布 $p_i(x_i|\mu_i, \sigma_i)$，$p_j(x_j|\mu_j, \sigma_j)$ 間のバタチャリヤ距離 $BD(p_i, p_j)$ を次式に示す．

$$BD(p_i, p_j) = \frac{1}{4}\frac{(\mu_i - \mu_j)}{\sigma_i^2 + \sigma_j^2} + \frac{1}{2}\ln\frac{\sigma_i^2 + \sigma_j^2}{2\sigma_i^2\sigma_j^2} \qquad (2)$$

本研究では，時刻 $t - \Delta t$ から t まで動く間に得られたセンサ出力分布をもとに，異なる 2 センサ間の $BD(p_i, p_j)$ を求め，それらを要素とするベクトル $\boldsymbol{v}$ を時刻 t における状態 (ベクトル) と呼ぶ．本稿では，以下のように状態ベクトルを定義した．

$$\boldsymbol{v} = (v_{1,2}, v_{1,3}, v_{1,4}, v_{1,5}, v_{1,6}, v_{2,3}, v_{2,4}, v_{2,5}, v_{2,6}, v_{3,4}, v_{3,5}, v_{3,6}, v_{4,5}, v_{4,6}, v_{5,6}) \qquad (3)$$

ここで，$v_{i,j}$ は $BD(p_i, p_j)$ である．

提案する状態表現では，計測対象がセンサの計測範囲外にある場合など，微少時間のセンサ出力分布の分散が 0 となる場合には，分布間距離を計算することができない．そのため，出力分布の分散が 0 となるセンサに関しては，そのセンサと他のすべてのセンサとの分布間距離は 0 とするとした．

2.3 強化学習による状態-行動写像の獲得

提案する状態表現法を用いた強化学習による状態-行動写像の自律的獲得 (行動学習) について述べる．強化学習は，行動主体であるロボットや環境に関する先験的知識を必要とせず，ロボットが試行錯誤的に行動しながら，タスクを達成する状態-行動写像を自律的に獲得することができる [S.Sutton 98]．

本研究では，強化学習を行うにあたり，自己組織化マップ (Self-Organized Map[Kohonen 95]，以下 SOM) を用いて状態を離散化した．ニューラルネットワークの一つである SOM は，学習により，入力されるベクトルをその類似度に応じて分類することができる．本研究では，強化学習を行う前に SOM に複数の状態ベクトルを入力し学習を行った．学習後の SOM に状態ベクトル $\boldsymbol{v}$ を入力すると，SOM は，その $\boldsymbol{v}$ が分類されるクラスのラベル n $(1, \ldots, n, \ldots, N)$ を出力する．この n を離散化された状態とする．状態数 N は SOM を構成するニューロン数に対応する．

状態-行動写像の獲得には，強化学習の一種である Q-learning[Watkins 92] を用いる．Q-learning では，ある状態 s において行動 a をとることの価値を意味する $Q(s, a)$ 値 (行動価値関数) を導入し，学習により，環境から得られる報酬が最大となるように $Q(s, a)$ 値を調整していく．目的とするタスクに応じて適切な報酬系を設計することにより，タスクの実現が可能な状態-行動写像 ($Q(s, a)$) が獲得される．

時刻 t において，ロボットの状態が $s_t \in S$ であるとする．ロボットが行動 $a_t \in A$ をとり，それにより状態が $s_{t+\Delta t}$ に遷移し，報酬 $r_{t+\Delta t}$ を得たとする．このとき，次式で $Q(s, a)$ 値を更新する．

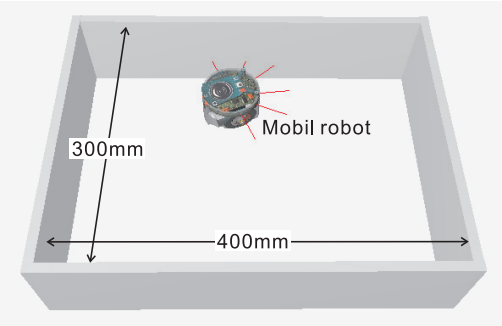

図 3: Experimental environment.

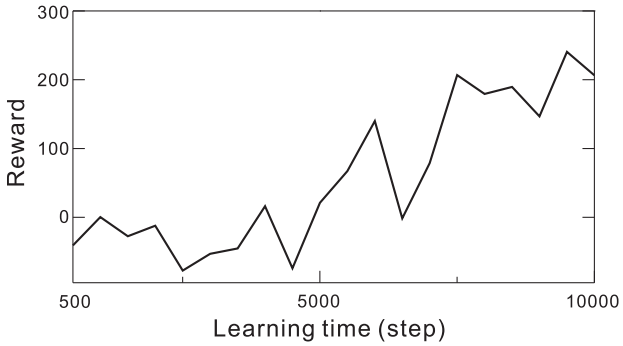

図 4: Reward.

$$Q(s_t, a_t) \leftarrow (1-\alpha)Q(s_t, a_t) + \alpha\left\{r_{t+\Delta t} + \gamma \max_{a' \in A} Q(s_{t+\Delta t}, a')\right\} \quad (4)$$

ここで α，γ はそれぞれ学習率および割引率である．

後述する実験では，状態の離散化の影響により，同じ状態にあるロボットが同一の行動を一定時間 (Δt) 生成したとしても，他の状態へ遷移する場合と同一状態にとどまる場合の二通りの状態遷移が起きる場合がある．このような，行動と状態遷移が 1:1 に対応しない「状態と行動のずれ (state-action deviation)[Asada 96]」に対して，本研究では，ある状態が他の状態へと遷移するまで同一の行動 a を生成し続け，状態が遷移した場合に式 (4) により $Q(s, a)$ 値を更新するとした．

3. 実験及び考察

3.1 実験設定

提案する状態表現の有効性を検証するために，移動ロボットを用いたシミュレーションを行った．実験では，まずロボットに行動学習によりタスクを実現する状態-行動写像を獲得させ，その後，センサ出力を人工的に変化させて，既に獲得した状態-行動写像によりタスクを行った．提案手法を用いることにより，センサ出力が変化しても同程度のパフォーマンスが得られることが期待できる

ロボットの持つ 6 つの距離センサの仕様は同一であり，センサ出力は検出範囲内に物体が存在しない場合は 0 を，検出範囲内の物体との距離が小さく (近くなる) につれて最大で 3600 の整数値をとる．時刻 t におけるセンサ i ($1 \sim 6$) の出力を $o_{t,i}$ とする．ロボットの行動は，前進，左旋回，右旋回の 3 種類で，各行動に対応するモーターコマンドを m_f，m_l，m_r とする．

実験では，はじめに SOM の学習用に，図 3 の学習環境で 3 種類のモーターコマンドをランダムに選択し，Δt 時間行動を生成する作業を繰り返して，複数の状態ベクトルを求めた．学習に用いた状態ベクトル数は 500 であり，SOM のニューロン数は 100 とした．また，$\Delta t = 0.6$ sec，$M = 20$ とした．

次に，学習した SOM を用いて行動学習を行った．タスクは壁沿い移動タスク (wall-following task) とした．時刻 t において，次の条件 1)~3) のすべてを満たす場合にのみ $r_t = 10$ とし，それ以外は $r_t = -1$ とした．

1) $o_{t,2} < 1080$ かつ $o_{t,5} < 1080$
2) $o_{t,1} > 0$ もしくは $o_{t,6} > 0$
3) モータコマンドが m_f

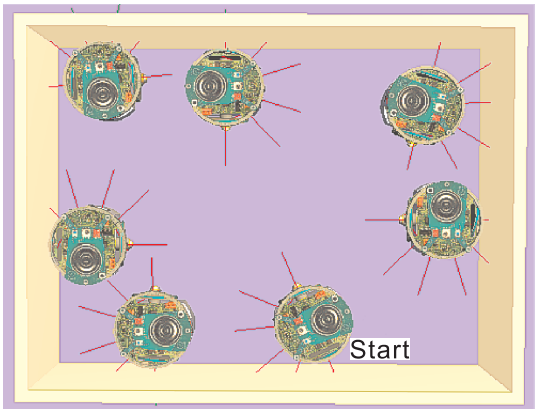

図 5: The behavior of robot.

行動学習の回数は 10000 ステップ (1 ステップは Δt) とし，500 ステップ毎に，ロボットを壁の付近にランダムに配置し直しながら行った．なお，式 (4) の割引率は $\gamma = 0.9$，学習率 α は初期値を 1 とし，学習が進むにつれて 0 に近づくようにスケーリングを行った．

図 4 にロボットが獲得した報酬 (500 ステップ毎の合計) を示す．図は 10000 ステップの学習 5 回分の平均である．また，図 5 は学習後のロボットの挙動 (50 ステップ毎のスナップショット) であり，壁沿い移動を行う様子が確認できる．

3.2 実験結果及び考察

学習後のロボットのセンサ出力を変化させて，同じタスクを行った場合のパフォーマンスを確認した．実験では，次式により変換されたセンサ出力 $o'_{t,i}$ を用いた．

$$o'_{t,i} = a o_{t,i} + b \quad (i = 1 \sim 6) \quad (5)$$

表 1 は，式 (5) の線形変換と平行移動の各成分 (a, b) を $(1, 1000)$，$(1, -100)$，$(0.4, 0)$，$(-1, 0)$，$(0.4, -100)$，$(-1, 1000)$ として，一定時間 (500 ステップ) 行動した際に獲得した報酬を示したものである．ただし，報酬はセンサ出力 o_l を用いて算出した．表 1 で各報酬は，センサ出力が変化していない ($(a, b) = (1, 0)$) 場合に得られる報酬が 1 となるように正規化している．

比較のため，2 つの異なる状態表現 (Type 1，2) を用いて同様の実験を行った．Type 1 は，微少時間のセンサ出力の差分量を用いて状態を表現するもので，状態ベクトルを $\boldsymbol{v}_d = (v_{d,1}, \cdots, v_{d,6})$，$v_{d,i} = o_{t,i} - o_{t-\Delta t,i}$ とする [郷古 09]．また，

表 1: The performance by converted sensor output (normarized reward).

(a, b)	Proposed	Type 1	Type 2
(1,1000)	1.03	1.01	1.04
(1,-100)	1.07	1.02	0.96
(0.4,0)	1.00	0.40	-0.01
(-1,0)	0.99	0.51	-0.17
(0.4,-100)	0.98	0.34	-0.01
(-1,1000)	0.98	0.44	-0.13

Type 2 では，式 (3) の各要素を $v_{i,j} = o_{t,i} - o_{t,j}$ とした状態ベクトルを用いた．学習回数やその他のパラメータはすべて同一とし，学習後はいずれの状態ベクトルを用いた方法でも，壁沿い移動が実現できていることを確認している．

表 1 を見ると，Type 1，2 は，センサ出力の変換が平行移動のみの場合 ($(1, 1000)$，$(1, -100)$) は，獲得報酬は変換前と同程度であるが，線形変換に対してはパフォーマンスが大きく低下してしまうことが分かる．これは，センサ出力の線形変換によって生じる，状態 (差分量) の変化が大きいため，変換前に獲得した状態-行動写像では，適切な行動生成が困難であることを意味している．

一方で，提案する状態表現を用いた場合には，センサ出力の各変換に対して，得られる報酬の変化はほとんど見られない．このことから，提案手法は，線形変換及び平行移動されたセンサ出力を，同一の状態として表現し行動を生成していることが確認できる．

4. まとめと今後の課題

本研究では，人間の知覚の恒常性発現メカニズムをヒントに，分布間距離の変換不変性を用いた状態表現を提案した．移動ロボットを用いたシミュレーションにより，提案する状態表現は，センサ出力の線形変換及び平行移動に対し，それらの影響を受けずに同一の状態-行動写像を使ってタスクが実現できることを示した．

今後は，実機を用いた実験により，提案する状態表現の有効性を確認する他，距離センサ以外のセンサを用いた実験も行う予定である．なお，本稿で扱った問題のように，同一の実状態から得られるセンサ情報が異なる場合，各センサ情報を同一状態と知覚する能力の他，タスクによっては，それらのセンサ情報を異なる状態と知覚する能力も必要となる．様々なタスクに対応可能なロボットを実現するためには，これら二つの能力が相補的に機能する知覚システムを実現する必要があると考える．今後は，この点に関しても検討する予定である．

謝辞

本研究の一部は，栢森情報科学振興財団及び科研費若手研究 (B)(21700219) の助成によるものである．ここに謝意を表す．

参考文献

[Asada 96] Asada, M., Noda, S., Tawaratsumida, S., and Hosoda, K.: Purposive Behavior Acquisition for a Real Robot by Vision-Based Reinforcement Learning., *Machine Learning*, Vol. 23, pp. 279–303 (1996)

[Csiszar 04] Csiszar, I. and Shields, P. C.: *Information Theory And Statistics: A Tutorial*, Now Publishers (2004)

[Duchon 98] Duchon, A. P., Kaelbling, L. P., and Warren, W. H.: Ecological Robotics, *Adaptive Behavior*, Vol. 6, No. 3-4, pp. 473 – 507 (1998)

[Gibson 79] Gibson, J. J.: *The Ecological Approach to Visual Perception*, Houghton Mifflin (1979)

[Kohonen 95] Kohonen, T.: *Self-Organizing Maps*, Springer-Verlag (1995)

[Koffka] Kurt Koffka(著), 鈴木正弥 (監訳)：ゲシュタルト心理学の原理, 1988 (福村出版)

[Nakamura 95] Nakamura, T. and Asada, M.: Motion Sketch: Acquisition of Visual Motion Guided Behaviors, in *Proceedings of International Joint Conference on Artificial Intelligence*, pp. 126–132 (1995)

[Qiao 08] Qiao, Y. and Minematsu, N.: f-divergence is a generalized invariant measure between distributions, in *Proceedings of 10th Annual Conference of the International Speech Communication Association*, pp. 1349–1352 (2008)

[S.Sutton 98] S.Sutton, R. and G.Barto, A.: *Reinforcement Learning: An Introduction*, The MIT Press (1998)

[Turvey 96] Turvey, M. T.: Dynamic Touch, *American Psychologist*, Vol. 51, No. 11, pp. 1134–1152 (1996)

[Watkins 92] Watkins, C. J. C. H. and Dayan, P.: Q-Learning, *Machine Learning*, Vol. 8, pp. 279–292 (1992)

[伊藤 06] 伊藤一之, 福森嘉孝：知覚量に基づく制御系設計-蛇型ロボットの方向の知覚量を用いたフィードバック制御-, 計測自動制御学会論文集, Vol. 42, No. 4, pp. 436–445 (2006)

[柿崎 93] 柿崎祐一：心理学的知覚論序説, 培風館 (1993)

[乾 95] 乾 敏郎, 他：認知心理学 1 知覚と運動, 東京大学出版会 (1995)

[郷古 09] 郷古 学, 伊藤宏司：f-divergence を用いた状態表現の関する基礎的研究, 第 27 回日本ロボット学会学術講演会予稿集 (DVD-ROM) RSJ2009AC2F1-02 (2009)

[郷古 09] 郷古 学, 登美直樹, 長野智晃, 伊藤宏司：状態パターンの変化にもとづく行動生成モデル, 電気学会論文誌, Vol. 129-C, No. 9, pp. 1690–1698 (2009)

[佐々木 94] 佐々木正人：アフォーダンス 新しい認知の理論, 岩波書店 (1994)

[寺田 03] 寺田和憲, 中村恭之, 武田英明, 小笠原司：視覚を有するエージェントのための身体性に基づく内部表現獲得手法, 日本ロボット学会誌, Vol. 21, No. 8, pp. 893–901 (2003)

[峯松 08] 峯松信明：音声言語運用が要求する認知的能力と音声言語工学が構築した計算論的能力, 電子情報通信学会音声研究会，SP2008-84, pp. 31–36 (2008)