

複数の未解釈センサデータの不変性に基づく自己身体の発見

吉川雄一郎 (阪大) 辻義樹 (阪大) 細田耕 (阪大, 阪大 F R C)
浅田稔 (阪大, 阪大 F R C)

Body finding from uninterpreted sensory data with multiple sensory attributes

*Yuichiro Yoshikawa (Osaka Univ.), Yoshiki Tsuji (Osaka Univ.), Koh Hosoda (Osaka Univ., HANDAI FRC), Minoru Asada (Osaka Univ., HANDAI FRC)

Abstract— This paper presents a method of body-nonbody discrimination by complementarily utilizing multiple sensory attributes based on a conjecture about the distribution of the variance of sensations for each observing posture, where it can be approximated by a mixture of two Gaussian distributions corresponding to the body and the nonbody, respectively. By estimating the distribution, the robot can automatically find a discrimination hyperplane to judge whether it observes its body in the current observing posture. Simple experiments show the validity of the proposed method.

Key Words: Body finding, acquisition of body representation, invariance in self-body-observation

1. はじめに

従来のロボティクスでは、ロボットの身体表現は設計者によって与えられることが多かった。しかし、環境やロボット自身の身体の変化に対してロボットを適応させるためには、ロボットが自分自身で身体表現を獲得する能力を備えていることが望ましい。一方近年、脳科学等の生物の知能に関する研究分野において、ヒトの身体表現に関する研究に注目が集まっている。身体表現を獲得するロボットの構築を通じて、いまだ明らかでないヒトの身体表現の獲得過程の理解に対して何らかの示唆が得られる可能性がある。そこで本研究では、ロボット自身による身体表現の獲得の第一歩として、物理的意味が未解釈のセンサデータから、身体を発見するロボットを実現することを目的とする。

ロボットが運動するとき、運動と相関のあるオプティカルフローが生じることを利用して身体を発見する手法が提案されている [1, 2]。しかし、ロボットの視覚センサと他の身体部位の両方が同時に運動しないような工夫、すなわち視覚センサとロボットの自由度の関係についての物理的な解釈が与えられる必要があった。一方、Yoshikawa ら [3] は、身体についての知覚は環境によらずおおむね一定であるという考察から、観測姿勢に対するセンサデータの不変性に基づいて身体を発見する手法を提案している。しかし提案手法では単一のセンサデータの不変性のみが議論され、複数のセンサデータを相補的に利用する方法については明らかでなかった。

本研究では身体を観測している場合のセンサデータの分散は外界を観測するときのそれに比べて小さい [3] という考え方を、複数のセンサデータに適用し、複数のセンサデータの相補的性質を利用する方法を考える。各観測姿勢に対する知覚の分散の分布は、身体の観測に起因するものと環境の観測に起因するものの二つの混合正規分布であると考えられる。従って、複数のセンサデータの分散からなるベクトルの分布を EM アルゴリズムを用いて推定し、現在の観測姿勢における分散

がどちらの正規分布に属するかを識別することで、観測対象が身体であるか否かの判定を相補的に実現可能であることが期待される。

本稿では、複数のセンサデータを用いて身体-非身体の識別を行う手法について説明し、実ロボットを用いて、複数のセンサデータが相補的に利用され、身体-非身体の識別が実現されることを確認する。

2. 複数センサデータを用いた身体発見

複数センサデータの分散の混合正規分布モデル ロボットは、視差や輝度値など、 D 種類のセンサデータによる観測が可能であるとする。観測姿勢 $\theta \in \mathbb{R}^N$ における i 番目のセンサデータを $x_i(\theta) \in \mathbb{R}^{M_i}$, ($i = 1, \dots, D$) とする。 $x_i(\theta)$ の分散を $\sigma_i^2(\theta)$ と表し、

$$z(\theta) = [\tilde{\sigma}_1(\theta)^2, \dots, \tilde{\sigma}_D(\theta)^2]^T \in \mathbb{R}^D, \quad (1)$$

を観測分散ベクトルと呼ぶ。 $\tilde{\sigma}_i^2(\theta)$ は、 $\sigma_i^2(\theta)$ を正規化したものであり、ここでは対数スケールを用いる。

知覚は身体の観測と外界の観測の両方によって生じ、身体を観測するときのセンサデータの分散は外界を観測するときのそれに比べ小さいと考えられる。そこで、各観測姿勢に対する知覚の分散の分布は、身体の観測に起因するものと環境の観測に起因するものの二つの混合正規分布であると仮定する (Fig. 1 参照)。すなわち、 z の分布は

$$p(z; \alpha) = w_b \mathcal{N}(z; \mu_b, \Sigma_b) + w_e \mathcal{N}(z; \mu_e, \Sigma_e) \quad (2)$$

で与えられるとする。ここで、 $\mathcal{N}(z; \mu, \Sigma)$ は、平均 μ 、共分散 Σ の z の正規分布を表し、添え字の b と e はそれぞれ、身体と外界を表す。また $\alpha = \{w_b, \mu_b, \Sigma_b, w_e, \mu_e, \Sigma_e\}$ である。正の重み w_b と w_e は $w_b + w_e = 1$ を満たす。

混合正規分布の推定 ロボットの姿勢が q_θ 個に量子化されているとき、ロボットは $z(\theta_i)$, $i = 1, \dots, q_\theta$ を計測することができるが、それぞれの $z(\theta_i)$ がどちらの分布から生じたのかは計測できない。従って、不完全

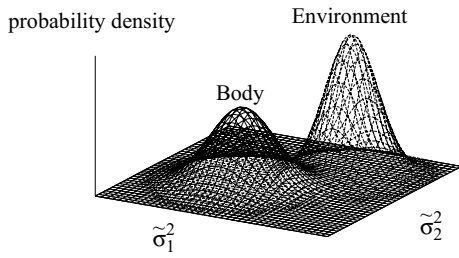


Fig.1 Mixture of Gaussian distribution model of observing variance vector

データからパラメータの最尤推定を行う理論的枠組みである EM アルゴリズム [4] を適用し、混合正規分布の推定を行う。

EM アルゴリズムでは、与えられた初期パラメータに対し、E-step と M-step を収束するまで繰り返し適用し、不完全データ $Z = \{z(\theta_1), \dots, z(\theta_{q_0})\}$ の対数尤度関数、

$$\mathcal{L} = \log p(Z|\alpha), \quad (3)$$

を最大化するパラメータ α を推定する。E-step は、 Z とパラメータの推定値 $\alpha^{(t)}$ が与えられたときの完全データ $\{Z, H\}$ の対数尤度関数の期待値、

$$Q(\alpha|\alpha^{(t)}) = E_Z\{\log p(Z, H|\alpha)|Z, \alpha^{(t)}\} \quad (4)$$

を計算するプロセスである。ここで、 H はどちらの $z(\theta)$ がどちらの分布に属するのを示す隠れパラメータである。M-step では、 $Q(\alpha|\alpha^{(t)})$ を最大化するよう、 α が更新される。上記の E-step と M-step の繰り返しにおいて、対数尤度関数が単調増加することが知られている [4]。

身体-非身体の識別 現在の姿勢で観測しているものが身体か否かを判断するために、現在の観測姿勢に対するセンサデータの分散 $z(\theta)$ が、推定した混合正規分布のどちらのカーネルに属するのかを線形判別法を用いて判定する。識別関数は

$$g(z(\theta)) = \kappa^T z(\theta) + \kappa_0 \quad (5)$$

と表される。ここで、 κ は識別境界 $g(z(\theta)) = 0$ の法線ベクトルであり、 κ_0 はオフセットである。線形判別法を適用すると、

$$\kappa \propto (\hat{\omega}_b \hat{\Sigma}_b + \hat{\omega}_e \hat{\Sigma}_e)^{-1} (\hat{\mu}_b - \hat{\mu}_e) \quad (6)$$

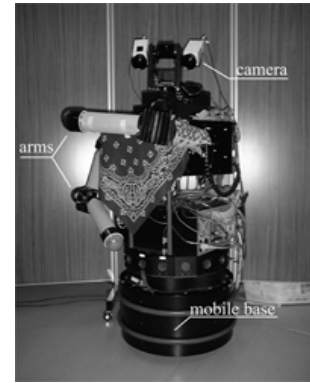
となる。識別境界が二つの正規分布の midpoint を通るように κ_0 を定めると、識別関数は

$$g(z) = (\hat{\mu}_1 - \hat{\mu}_2)^T (\hat{\omega}_b \hat{\Sigma}_b + \hat{\omega}_e \hat{\Sigma}_e)^{-T} (z - \frac{\hat{\mu}_1 + \hat{\mu}_2}{2}), \quad (7)$$

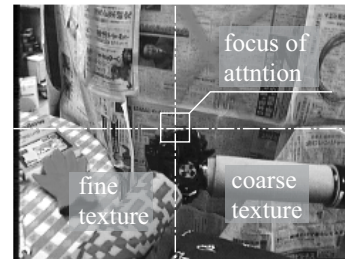
となる。ここで、 A^{-T} は行列 A の逆行列の転置であり、 $\{\hat{\omega}_b, \hat{\mu}_b, \hat{\Sigma}_b, \hat{\omega}_e, \hat{\mu}_e, \hat{\Sigma}_e\}$ は EM アルゴリズムで得られた混合正規分布のパラメータである。最終的に、 $g(\theta) > 0$ であるとき、観測対象がロボットの身体であると判定することができる。

3. 実験

実ロボット (Fig. 2 (a)) を用いた実験により、提案手法の有効性を確認する。ロボットは二つのカメラを有し、センサデータとして、左カメラ中心の視差、輝度値パターン、平均彩度、平均エッジ方向を抽出し、身体-非身体の識別を行う。実験に用いるロボットは二つのカメラが搭載されたカメラヘッド、アーム、移動台車からなり、カメラヘッドはパン・チルト方向に運動することができる。アームは4自由度を有するが、結果の解釈のために、以下の実験では動かさず固定した。提案手法がロボットの身体性に依存しないことを示すため、ロボットの身体表面に異なるテクスチャを貼り付けた。Fig. 2(b) は2種類のテクスチャが貼り付けられたアームを眺めるロボットの視野の一例である。



(a) The robotic test-bed



(b) An egocentric view of the robot

Fig.2 The robotic test-bed: (a) the whole body and (b) an egocentric view

ロボットはランダムにカメラヘッドの姿勢を変えながら、様々な姿勢でのセンサデータを収集し、各観測姿勢におけるセンサデータの平均と分散を学習する。学習の間、ロボットをランダムに移動させるため、ロボットに対する環境は時々刻々変化する。Fig. 3(a) は、ロボットが観測姿勢の各量子に対してセンサデータの平均と分散を保持する様子を図的に示したものである。Fig. 3(a) の右側の図は、例として、保持された輝度値パターンの平均をパン（横軸）とチルト（縦軸）の順に並べて示したものである。ロボットのタスクは姿勢の各量子に収められたセンサデータのうち、Fig.

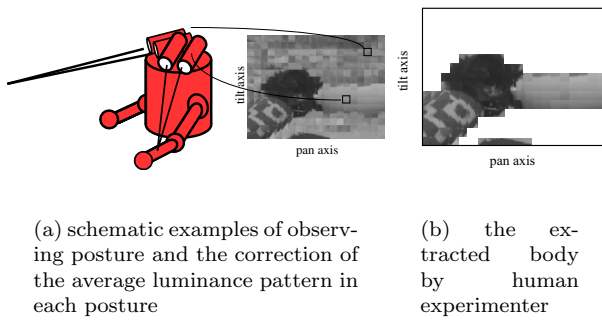


Fig.3 An schematic explanation of (a)the learning process and (b)the desired extraction of the body

3 (b) に示すように、身体を捉えている部分のみを抽出することである。

3.1 単一の視覚センサデータを用いた身体-非身体の識別

センサデータとして、左カメラの中心領域の視差のみ、そして輝度値パターンのみを用いて、身体-非身体の識別実験 ($D = 1$) を行った。Fig. 4 は視差を用いたときの結果を示している。Fig. 4 (a) は、視差の分散の分布 (実線ヒストグラム) と EM アルゴリズムにより推定された混合正規分布 (破線) を示している。分散の低い側にある正規分布のカーネルが身体を観察する姿勢における視差の分散の分布に、高い側が外界を観察する姿勢における視差の分散に対応すると考えられる。Fig. 4(b) は式 (7) の識別境界を用いて、身体を観測していると判定された姿勢における視差の平均値をグレースケールで示している。Fig. 4 (b) を Fig. 3 (b) と比較すると、細かいテクスチャを持つ身体部分については正しく身体であると識別ができていたが、粗いテクスチャを持つ身体部分については、外界であるとみなしてしまっているのがわかる。これは、粗いテクスチャを持つ部分で、視差を計算するためのステレオマッチングの精度が悪くなっているためであると考えられる。

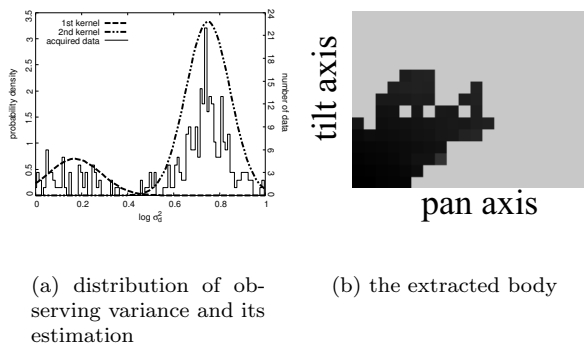


Fig.4 Body-nonbody discrimination with disparity

Fig. 5 は左カメラの中心領域 (8×8 [pixel]) の輝度値パターンを用いた場合の結果を示している。Fig. 5 (a) は、輝度値パターンの分散の分布 (実線ヒストグラ

ム) と EM アルゴリズムにより推定された混合正規分布 (破線) を示している。Fig. 4 (a) と同様に、分散の低い側に身体を、高い側に外界を観察する姿勢における輝度値パターンの分散の分布に対応する正規分布が推定されていると考えられる。Fig. 5(b) は身体を観測していると判定された姿勢における輝度値の平均パターンを示している。Fig. 5 (b) と Fig. 3 (b) を比較すると、センサデータとして視差を用いた場合の結果とは逆に、粗いテクスチャを持つ身体部分については正しく身体であると識別ができていたが、細かいテクスチャを持つ身体部分については、外界であるとみなしてしまっているのがわかる。これは、細かいテクスチャの輝度値パターンが、観測姿勢の違いに対して敏感に変化するためであると考えられる。

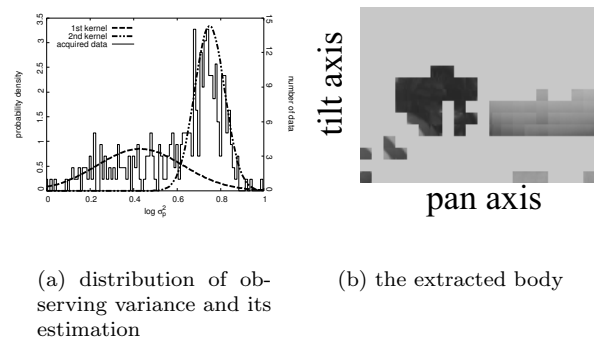


Fig.5 Body-nonbody discrimination with luminance pattern

3.2 複数の視覚センサデータを用いた身体-非身体の識別

単一のセンサデータを用いた身体-非身体の識別実験では、それぞれのセンサデータの性質に応じて、正しく身体-非身体の識別が可能である部分が異なることがわかった。そこで次に、これらのセンサデータを組み合わせることにより、相補的な正しく身体-非身体の識別が可能であることを示す実験を行った。

Fig. 6 はセンサデータとして、左カメラ中心領域の視差と 8×8 [pixel] の輝度値パターンの両方を用いた実験 ($D = 2$) の結果を示している。Fig. 6(a) は視差と輝度値パターンの分散の分布であり、Fig. 6(b) は EM アルゴリズムにより推定された混合正規分布である。単一のセンサデータを用いた場合と同じく、分散の低い側に身体を、高い側に外界を観察する姿勢における輝度値パターンの分散の分布に対応する正規分布が推定されていると考えられる。Fig. 6(c) は身体を観測していると判定された姿勢における輝度値の平均パターンを示している。Fig. 6(c) は単一のセンサデータを用いた場合の結果 (Fig. 4 (b) および Fig. 5 (b)) に比べ、Fig. 3 (b) に近く、それぞれ単一のセンサデータを用いた場合では、正しく識別できていなかった身体部分に関して相補的に働いていることがわかる。

さらに、左カメラ中心領域の平均彩度および平均エッジ方向を加えて、実験を行った。Fig. 7 (a) は左カメラ中心領域の平均彩度のみを用いたとき ($D = 1$) に、

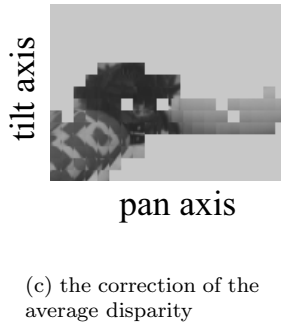
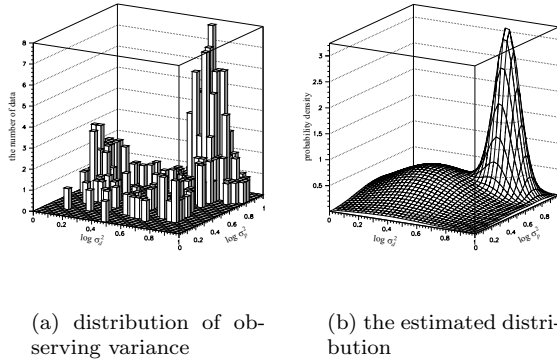


Fig.6 Body-nonbody discrimination with both disparity and luminance pattern

身体を観測していると判定された姿勢における彩度の平均値を示したものであり、Fig. 7 (b) は平均エッジ方向を用いたとき ($D = 1$) のものである。単一のセンサ属性として、視差や輝度値パターンを用いた識別結果 (Fig. 4(b) および Fig. 5(b)) に比べ、身体の広い部分を身体として識別できていることがわかるが、逆に外界の一部も身体として識別してしまっていることが分かる。Fig. 8 は、センサデータとして、左カメラ中心の視差、輝度値パターン、彩度、平均エッジ方向を用いたとき ($D = 4$) に、身体を観測していると判定された姿勢における輝度値パターンを示したものである。Fig. 6 に比べ、Fig. 3 (b) により近く、より精度の高い、身体-非身体の識別が実現されていることがわかる。

従って、単一で用いると外界の一部を身体と誤ってみなしてしまうようなセンサデータを、複数センサデータによる身体-非身体の識別に加えた場合でも、正しく身体を身体として識別している部分の情報のみを最終的な身体-非身体の識別結果に貢献させることができるといえる。

4. 結言

本稿では、観測姿勢に対する観測センサデータの分散の分布が二つの混合正規分布で表されるという仮定に基づき、複数のセンサデータの相補的性質を利用して身体-非身体の識別を学習する手法を提案した。複数のテクスチャを持つロボットを用いた実験により、ロボットの身体表面の特性によらず、複数のセンサデー

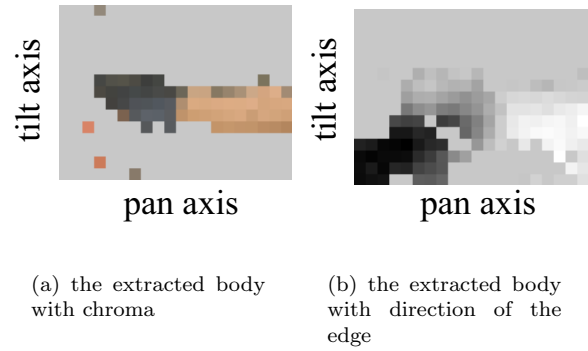


Fig.7 Body-nonbody discrimination with one sensory attribute

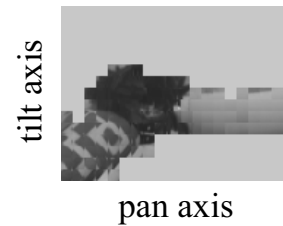


Fig.8 Body-nonbody discrimination with multiple sensory attributes, disparity, luminance pattern, chroma, and direction of the edge

タを相補的に利用した身体-非身体の識別が可能であることを確認した。

謝辞

本研究は、21 世紀型革新的先端ライフサイエンス技術開発プロジェクト、萌芽・融合開発プログラム「動的インタラクションによるコミュニケーション創発機構の構成と解明 (タイプ A)」の援助を受けた。

参考文献

- [1] M. Asada, E. Uchibe, and K. Hosoda. Cooperative behavior acquisition for mobile robots in dynamically changing real worlds via vision-based reinforcement learning and development. *Artificial Intelligence*, Vol. 110, pp. 275–292, 1999.
- [2] P. Fitzpatrick and G. Metta. Toward manipulation-driven vision. In *Proc. of the Intl. Conf. on Intelligent Robot and Systems*, pp. 43–48, 2002.
- [3] Y. Yoshikawa, K. Hosoda, and M. Asada. Does the invariance in multi-modalities represent the body scheme? - a case study with vision and proprioception -. In *Proc. of the 2nd Intl. Symposium on Adaptive Motion of Animals and Machines*, pp. SaP-II-1, 2003.
- [4] 上田修功. ベイズ学習 [1]-統計的学習の基礎-. 電子情報通信学会誌, Vol. 85, No. 4, pp. 265–271, 2002.