

視覚観測困難な自己の顔の身体像におけるパーツ検出モデルの提案

Invisible facial parts detection in the tactile body image based on the perceptual discontinuity

福家 佐和 (阪大) 正 荻野 正樹 (阪大) 正 浅田 稔 (阪大)

Sawa FUKU, JST Erato Asada Project, Osaka University
Masaki OGINO, JST Erato Asada Project
Minoru ASADA, JST Erato Asada Project, Osaka University

This paper proposes a learning model that enables a robot to acquire a body image for parts of its body that are invisible to itself. A robot estimates the invisible hand positions using the Jacobian between the displacement of the joint angles and the optical flow of the hand. When the hand touches one of the invisible tactile sensor units on the face, the robot associates this sensor unit with the estimated hand position. In addition, we propose a model to discriminate the tactile sensor units which are corresponded to each facial part based on the discontinuity of sensor value. Then finally, the robot becomes able to associate them with the feature points of parts in the image of other's face.

Key Words: Body-image, Multi-modal

1 はじめに

私達人間は、自身の手や道具を使って様々なタスクを行うことが可能である。そのためには、自身の姿勢や運動の結果自身のパーツが外部空間に与える影響といった身体表現を獲得している必要がある。しかしこのような表現がどのように獲得されているのかはまだ明らかにされていない。Ramachandran¹⁾ や Iriki et al.²⁾ の知見より、生物は身体表現は動的に構成していると推測され、視覚だけでなく聴覚や体性感覚等刺激として入ってきた異なるモーダル感覚を、時間的にも空間的にも統合することが重要であると考えられる。

現在多くの場合、ロボットにおいてはこのような身体表現は予め設計者によってプログラムされ与えられている場合が多い。しかし、ハードウェアの条件が変化した時等、人間が最初に埋め込んだ身体表現でタスクを実行するには限界がある。更に、ロボットが人間社会に共存していく上で、様々な種類の道具の適応的な利用や、人間が用いる抽象的な概念の獲得など、自身の身体を利用した高度な学習を行うためには、土台としてこのような身体表現を自律的に獲得する必要があると考えられる。

これまでロボットを用いた身体表現獲得のモデルとしては Nabeshima et al.³⁾ が Iriki et al.²⁾ らの身体表現延長現象のモデルを提案した。また、Yoshikawa et al.⁴⁾ はロボットが自己の身体に触った時の、触覚情報と視覚情報と体性感覚情報の結び付きを Hebb 学習によって構成する手法を提案した。しかしこれらの手法では、身体表現はカメラの視野に依存し、背中や腹部等の死角部分の像を獲得することが困難であると考えられる。Fuke et al.⁵⁾ は、この点に着目しロボットが自律的に死角部位にあたる顔を触った時の、触覚情報と体性感覚情報と推測した視覚情報の結び付きを自律的に獲得する手法を提案した。しかし他者の身体との比較は未実装であり、将来獲得した身体表現を用いた模倣システムなどへ発展させることが困難である。他者の身体の状態と自身の身体の状態を比較するためには、一様な触覚センサの中から”パーツ”という特徴的な部位を検出する必要があると考えられる。よって本研究では、ロボットが死角に存在する身体として顔の触覚センサの空間内での配置を獲得すると同時に、接触運動中の各センサ値の不連続性をもとに顔表面

から特徴的な触覚センサを抽出し、他者の顔画像のパーツに起因する特徴との照合を学習するモデルを提案する。

実際にこのような相手の顔の特徴を認識でき自身の知覚できる特徴との関係を見いだす能力 (Organ Identification) は、新生児模倣の代表的モデルの AIM モデル⁶⁾ において、新生児が持っているべき前提の能力の1つとして位置づけられている。これは人間が各感覚と統合し自身の身体表現を利用している最初の段階であると考えられるが、その獲得課程はまだ知られていない。よって本研究では認知発達ロボティクスの観点から、モデルを提案することで、人間の身体表現の獲得および人間の基本の学習方法である模倣のメカニズムへの理解を助けることも目的としている。

2 空間知覚モデル

身体表現は視覚・触覚・体性感覚など異なるモーダルの情報が統合され、互いの空間内を関係づける写像が得られることによって獲得されると考えられる。しかし、顔表面に触れた時は手先位置を視覚上で確認することができない。よって、ロボットは視野内での運動中の視覚刺激と体性感覚刺激の結び付きを利用して、直接観測することのできない空間内での位置を知覚する能力が必要である。まずロボットは、視野内で自身の腕を動かした時の視覚空間上での手先位置変化を Δr_m とし、自身の関節角度変化を $\Delta \theta_m$ との関係を

$$\Delta r_m = f(\Delta \theta_m) \quad (1)$$

とすると、この変位と関係する写像する関数 f をニューラルネットを用いて学習する。更に、ロボットは死角に存在する自身の顔を能動的に触れ統合を行うが、この時手先位置は前述の通り直接観測することが不可能である。よって接触運動中の腕の関節角度変位 $\Delta \theta$ と学習した f を用いて位置を推測する。例えば、 t_0 から t_1 までの間、ロボットが自身の顔に触れた場合、 t_1 での手先位置 \hat{r} は

$$\hat{r} = F(\theta_{t_0}) + \int_{t_0}^{t_1} f(\Delta \theta) dt, \quad (2)$$

と計算できる。F は関節角度空間の SOM の cell に結び付いている位置情報への変換である。一般的には、ロボッ

トの姿勢に応じて f は異なる写像となるが、ここでは手先が視野内にあるときに得られた関係をそのまま用いる。

3 視覚空間と触覚空間の統合の学習

ロボットは自分の顔を能動的に触れている間に、各モダルの情報の統合を行う。触覚空間は顔表面上の触覚センサの配列を意味しており、各ユニットにはセンサ ID とともにそのユニットに触れた時の視覚空間上での手先位置情報が結びつけられている。また、体性感覚空間は代表ベクトルが関節角度である自己組織化マップとなっている。触覚空間同様に、各ユニットには視覚空間の手先位置情報が結びつけられている。

接触経験中、2章で推定された仮想的な手先位置 $\hat{\mathbf{r}}$ に対して、視覚空間と触覚空間の統合を自己組織化マップと同様に以下のように学習する。すなわち、触覚の表象空間において、この $\hat{\mathbf{r}}$ に最も近い位置情報を持つ触覚センサ c とする。

$$c = \arg \min_i \|\hat{\mathbf{r}} - \mathbf{r}_i\| \quad (3)$$

このとき、 i 番目のセンサーに対応づけられる仮想的な手先位置 \mathbf{r}_i を以下のように更新する。

$$\Delta \mathbf{r}_i(t) = \alpha(t) \exp(-\|\mathbf{T}_i - \mathbf{T}_c\|/\gamma) (\hat{\mathbf{r}} - \mathbf{r}_i(t-1)) \quad (4)$$

ここで、 $\alpha(t)$ は学習回数とともに減少する学習率であり、 \mathbf{T} は触覚の表象空間上の位置である。

また視覚空間と体性感覚空間の統合も、 $\hat{\mathbf{r}}$ に最も近い位置情報を持つ関節角度空間の SOM のユニットを探し、そのユニットに対しても上記と同様な更新を行う。

4 特徴的な触覚センサの区別

また、ロボットは自身の顔を能動的に触りながら触覚空間と視覚空間の統合を学習すると同時に、顔表面の触覚センサから特徴的なユニットを区別する。ここで実際に生児がどのようにセンサを区別し、パーツという概念を獲得しているかを考えると、まず第一に、先天的にセンサ感度の差に基づいて区別していると仮定することが可能である。しかしその一方で、実際には胎児は母親の胎内でしきりに顔を触ったり指を吸ったりしているという知見から、我々は新生児が接触経験の間、各モダリティのセンサ値の不連続性に基づいてパーツの区別を学習していると仮定する。よってこの仮定に基づき、自己が知覚した各種のセンサ値の不連続性に基づいてロボットが一樣な顔表面の触覚センサから特徴的な部位を区別するモデルを提案する。

ロボットは2章で述べたように自身の顔に能動的に触り、顔表面の触覚センサ間の相対的な配置関係を獲得する間に、知覚している種々のモダリティの感覚の不連続性を検出する事で、それらの中から特徴的なユニットを区別する。ロボットの触覚空間の i 番目のユニットには、更にセンサ k からの入力値が不連続になる可能性があることを示す discontinuity level $u_i^k(t)$ という変数が結びつけられている。 $u_i^k(t)$ は、不連続性を検出するたびに以下のように更新する。

$$\Delta u_i^k(t) = \alpha_p \exp(-\|\mathbf{T}_i - \mathbf{T}_{c_p}\|/\gamma_p) \quad (5)$$

c_p は、1式で学習した写像関数 f を用いて推測した不連続性が検出された瞬間の手先位置に最も近い距離情報を持っている触覚空間のユニットである。 k は、 $u_i^{hand}(t)$, $u_i^{R-view}(t)$, $u_i^{L-view}(t)$, $u_i^{hand-move}(t)$ の4種類がある。

$u_i^{hand}(t)$ は、顔表面の凹凸に手先が衝突することによって、手先の接触面に生じる不連続性に基づいて更新される。つまり顔に触れている間に、

$$|s_h(t-1) - s_h(t)| \geq \delta \quad (6)$$

という変化が発生した時に(7)式に従って更新される。ここで $s_h(t)$ とは、運動の各ステップにおいて最も顔表面との距離が小さい手表面のセル (Fig.1における赤いセル) の ID を示す。 $u_i^{R-view}(t)$ および $u_i^{L-view}(t)$ は、ロボットが手先を右および左のカメラの前に持ってきたことにより、視界が暗転し画像入力得られないというイベントが発生した場合に更新されるとする。また $u_i^{hand-move}(t)$ は口の運動と同期して手先の接触面が変化した場合に更新されるとする。

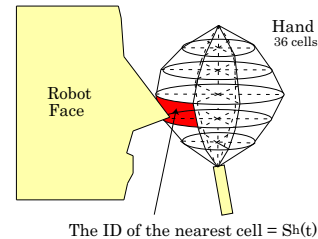


Fig.1 The hand used as a probe.

5 体性感覚空間と視覚空間の統合の学習

$\hat{\mathbf{r}}$ に最も近い位置情報を持つ体性感覚空間の SOM のノードを探し、そのノードに対しても、2章の触覚空間と視覚空間の統合と同様の更新を行う。

6 実験1

本提案手法の有効性をシミュレーションによって示す。Fig.2(a)に示されるように、ロボットは触覚センサが顔表面に格子状に配置され、5自由度の腕を持っている。またロボットの視覚空間の座標系はFig.2(b)のようになっており、原点は右目に設定されている。空間内の (h'_x, h'_y, h'_z) 物体は図中の $z = h_z$ に投影された座標 (h_x, h_y, h_z) で観測される。計算式は以下のように表される。

$$h_y = \frac{h_z \times h'_y}{h'_z}, \quad (7)$$

$$h_x = \frac{h_z \times h'_x}{h'_z}. \quad (8)$$

まず、2章で述べたようにロボットが視野内で手先をラ

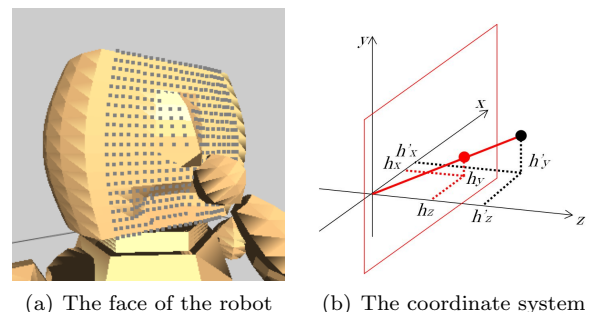


Fig.2 Simulation.

ンダムに動かした時に得られる情報から、体性感覚空間上の変位と視覚空間上の変位を関連づける関数 f を、学習係数 0.2 および学習回数 200 回に設定しバックプロパゲーション法により学習させた。 f を用いて死角部位での手先の軌道を推測可能であるという結果が得られたので、ロボットは左手の手先をプローブとして使い、顔表面のランダムな位置をなぞり統合を行った。Fig. 3に触覚センサ

ユニットに結びついている仮想的な視覚空間上の位置と Discontinuity level の学習経過を示す。青は $u_i^{hand-move}$ の、赤は u_i^{hand} の、緑は u_i^{L-view} の、紫色は u_i^{R-view} の Discontinuity level の値を意味し、鮮やかであるほど値が大きい。顔表面の触覚センサユニットの相対関係は、接触機会が少ないため学習が十分に進んでいない端部分には多少のゆがみがあるもの、獲得できていることがわかる。また、同時にロボットは、パーツ付近のユニットを特徴的なものとして区別していることがわかる。

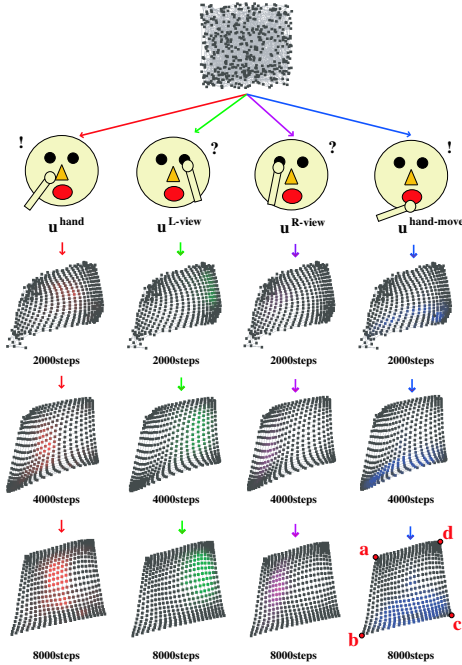


Fig.3 The abstracted tactile sensor units of the parts.

7 Organ Identification モデル

7.1 他者のパーツの特徴抽出

6章で区別した触覚の特徴と、他者の顔のパーツ部分との照合を行うために、視覚情報として取得した他者の顔画像から特徴点を抽出する。まず Fig.4(a) のように肌色検出により、被験者の顔を抽出した後、(b) のようにガウシアンフィルタを用いて画像の平滑化を行った。(b) の画素 (i^V, j^V) の明暗度を $I_{(i^V, j^V)}$ として、(a) で肌色である範囲の (i^V, j^V) について以下のような計算を行う。ただし、本画像処理において、各ピクセルの明暗度は最大値(黒)が 255 で最小値(白)は 0 で設定されているとする。

$$T_{(i^V, j^V)} = \sum_{l=-3}^3 \sum_{k=-3}^3 (I_{((i^V-k), (j^V-l))} - I_{(i^V, j^V)}) \quad (9)$$

$$I'_{(i^V, j^V)} = \begin{cases} 255 & T_{(i^V, j^V)} > \delta \\ 0 & \text{else} \end{cases} \quad (10)$$

計算の結果、 $I'_{(i^V, j^V)}$ に基づいて再描画した画像が Fig.4(c) である。

7.2 パーツの照合モデル

6章で区別した自身が知覚する特徴的な触覚センサユニットのクラスと、7.1章で抽出した他者の顔画像上の

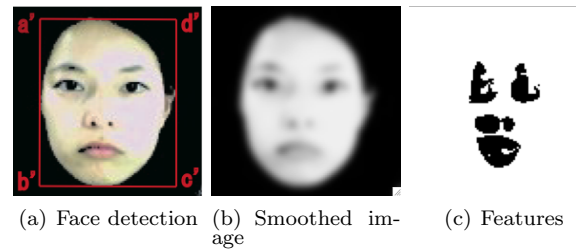


Fig.4 The result of image processing.

パーツに起因する特徴との照合を行う。画像上の顔部分の領域から、仮想的な視覚空間上の触覚センサの配置の領域への写像は線形写像で行われるとする。変換係数は Fig.3 の a, b, c, d および Fig.4(a) の a', b', c', d' から最小二乗法を利用して求めた。6章で学習した各 Discontinuity level, $u_i^{nose}, u_i^{Reye}, u_i^{Leye}, u_i^{mouth}$ の値がそれぞれ閾値以上の触覚センサを抽出した。そして、Fig.4(c) の特徴点を写像変換し、Fig.5(a) のように同じ空間内に表示した。そして他者の顔画像上の特徴点のうち q 番目の座標を、

$$\mathbf{X}_{(q)}^V = (x_{(q)}^V, y_{(q)}^V). \quad (11)$$

と設定する。また 3.1 章で各 Discontinuity level が閾値以上の触覚センサユニットを取り出す。これらのユニットは活性化したイベントの種類によって既に 4 種類のクラスに分類されていると考えられる。あるクラス k に所属する ID が p のユニットの座標は、

$$\mathbf{X}_{(k,p)}^T = (x_{(k,p)}^T, y_{(k,p)}^T). \quad (12)$$

と表現する。所属するユニットの総数はパーツごとに異なり $num_{(k)}$ と表現する。まず、各クラス k において、所属するユニットの座標を利用してその中心 $center_k$ を計算する。

$$\mathbf{center}_{(k)} = \frac{\sum_{p=0}^{num_{(k)}} \mathbf{X}_{(k,p)}^T}{num_{(k)}}. \quad (13)$$

次に、画像から抽出した特徴からランダムに抽出した点 X_{rand}^V と各クラス k の中心との距離を計算し。この距離が最も小さいクラスを c_k とする。

$$c_k = \arg \min_k \|\mathbf{X}_{(rand)}^V - \mathbf{center}_{(k)}\| \quad (14)$$

更に、クラス c_k の触覚センサユニットの中でも点 X_{rand}^V に最も近い位置情報を持つセンサを求める。

$$c_p = \arg \min_p \|\mathbf{X}_{(rand)}^V - \mathbf{X}_{(c_k,p)}^T\| \quad (15)$$

この時、クラス c_k 中の ID が p のセンサに対応づけられている位置情報 $X_{(c_k,p)}^T$ を以下のように更新する。

$$\Delta X_{(c_k,p)}^T(t) = \alpha \exp(-\|p - c_p\|/\gamma) (\mathbf{X}_{(rand)}^V - \mathbf{X}_{(c_k,p)}^T(t-1)) \quad (16)$$

ロボットは (11) 式から (14) 式を繰り返し、異なるモジュールの知覚上の特徴の照合を学習する。

8 実験 2

7章のモデルの有効性を示すため実験を行った。ロボットが区別した触覚センサユニットと、他者の顔画像から抽出した特徴との照合を学習した結果を Fig.5 に示す。青い点は実験で $u_i^{hand-move}$ が、赤い点は u_i^{hand} が、緑の点は u_i^{L-view} が、黄色の点は u_i^{R-view} が閾値以上であったユニットをプロットしたものである。学習が進むにつれ、各クラスが相当する画像上の特徴に近づき、照合が進んでいることがわかる。

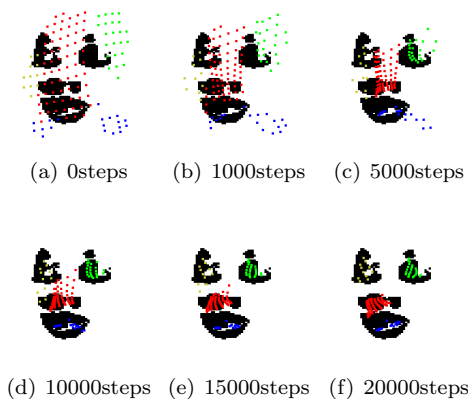


Fig.5 The matching of clusters of tactile sensor units and features of facial parts in image.

9 まとめと今後の展望

本研究のモデルの全体図をまとめると、Fig.6 のようになる。顔の触覚センサの仮想的な視覚空間上での配置を自律的に獲得すると同時に、特徴的なセンサを区別するモデルを提案した。更に、その自身の知覚の特徴と、他者の身体画像上のパーツに起因する特徴を照合するモデル (Organ Identification) モデルを提案した。発達学の分野では、在胎期間 24 週から 27 週の頃の胎児は、目を開きしきりに顔を撫でたりする行為が見られるという知見が報告されている⁷⁾。また、出生後間もない新生児は視覚システムが発達していないため平滑化した画像で情報を得ているという考察⁸⁾、母親の横顔より正面顔に好みを示すという知見⁹⁾、また人間の脳内の視覚野には好みを示す部分で囲まれた暗いものに反応するニューロンが存在するという知見⁸⁾などに着目すると、本稿のモデルを人間の初期の Organ Identification 能力の獲得メカニズムのモデルとして提案ができるのではないかと考えられる。

しかし一方で、本研究では視覚においては 2 次元的な認識しか行っておらず、身体表現を 3 次元的に空間表現することはできない。このような身体表現の工学的利用を考えると、今後奥行き知覚も可能であるような高度な視覚システムの構築が重要であると考えられる。またそれに伴い、今後背中や腹部など死角部位に存在する更なる他の身体への適用を目的に、モデル拡張に取り組んでいく予定である。

参考文献

- [1] V.S. Ramachandran, S. Blakeslee: Phantoms in the Brain: Probing the Mysteries of the Human mind. William Mollow, 1998.
- [2] Atsushi Iriki, Michio Tanaka, Shigeru Obayashi and Yoshiaki Iwamura: Self-images in the video monitor coded by monkey intraparietal neurons. Neuroscience Research, 40, 163-173, 2001.
- [3] Cota Nabeshima, Max Lungarella and Yasuo Kuniyoshi: Timing-Based Model of Body Schema adaptation and its Role in Perception and Tool Use: A Robot Case Study, The 4th IEEE International Conference on Development and Learning (ICDL-05), Osaka, Japan, July, 2005.
- [4] Yuichiro Yoshikawa, Hiroyoshi Kawanishi, Minoru Asada, Koh Hosoda: Body Scheme Acquisition by Cross Modal Map Learning among Tactile, Visual and Proprioceptive Spaces. Proceedings of the 2nd International Workshop on Epigenetic Robotics, pp.181-184, 2002.

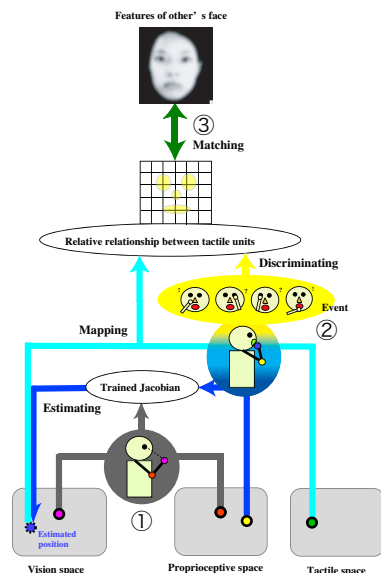


Fig.6 The proposed model to detect the facial parts.

- [5] 福家 佐和, 荻野 正樹, 浅田 稔. 自己受容器を介した視野外の身体像獲得モデルの提案. ロボティクス・メカトロニクス講演会 '06 予稿集, Vol.CD-ROM, 1P1-E13, 2006.
- [6] Andrew N. Meltzoff and M. Keith Moore, Explaining Facial Imitation: A Theoretical Model. Early Development and Parenting, Vol.6, 179-102, 1997.
- [7] S. Campbell: Watch Me Grow, A Unique, 3-Dimensional Week-by-Week Look at Your Baby 's Behavior and Development in the Womb. Carroll Brown Publishers, 2004.
- [8] M. H. Johnson: Subcortical face processing. Nature Review Neuroscience, Vol.6, 766-774, 2005.
- [9] F. Sai and Bushnell I.W. R. The perception of faces in different poses by one-month-olds. British Journal of Developmental Psychology, Vol.6, 35-41, 1988.