視覚的注意に基づく手先の感覚統合

Integration of bimodal sensation using visual attention based on saliency

○ 疋田 麻衣 (阪大)　　福家 佐和 (阪大)
正 荻野 正樹 (JST ERATO, 阪大)　　正 浅田 稔 (JST ERATO, 阪大)

Mai HIKITA, Osaka University
Sawa FUKE, Osaka University
Masaki OGINO, JST Erato Asada Project
Minoru ASADA, JST Erato Asada Project, Osaka University

Body representation is one of the most fundamental issues for physical agents (humans, primates, and robots) to perform various kinds of tasks. This paper proposes a method that constructs cross-modal body representation from vision, touch, and proprioception. Tactile sensation when the robot touches something triggers the construction process of the visual receptive field for body parts that can be found by visual attention based on saliency map. Simultaneously, proprioceptive information is associated with this visual receptive field to realize the cross-modal body representation. The computer simulation result comparable to the activities of parietal neurons found in monkey is given and future issues are discussed.

Key Words: Body schema, Sensory integration, Visual attention model

1 緒言

ヒトは複雑な環境の中で物体とのインタラクションを通じて様々なタスクを行うことができる．このとき周囲の空間情報を知覚し環境と身体の相互作用を把握する能力など，脳部位でいえば大脳皮質連合野で処理されるようなヒト特有の高次の機能が重要になる．これらの機能を実現させるとき，我々は自己の身体表現を基盤にしていると考えられる．自己の身体表現とは，知覚可能な種々の感覚 (modality) を空間的・時間的に統合することで得られるものである．

身体表現は，身体像 (自分自身の体について意識的に持つ表象) や身体図式 (身体の姿勢や動きを制御する際に働く無意識のプロセス) と呼ばれ，脳神経科学等の分野で多くの研究がなされてきた．中でもサルの道具使用実験において，サルが道具使用を学習する前後でのサルの頭頂葉のニューロン活動の変化から，サルが道具使用によってその身体図式を変化させている可能性が示唆されている [1)2)3)]．この知見は，生物の身体表現は柔軟で，また上述したように異なる感覚を空間的・時間的に統合したものであることを示す．しかしこのような表現がいつどのように獲得されるのか，更に「道具」の概念がより一般的な「オブジェクト」の概念からどのようにして創発されるのかは大きなミステリーである．

生物の身体表現は迅速にかつ柔軟に獲得されるものであるのに対し，ロボティクス分野の多くの研究では，ロボットの身体表現は初めに設計者によって与えられ変更されることがない．認知発達ロボティクス [4)] の分野では，ヒトの身体表現獲得の過程の理解だけでなくそのモデルをロボットに実装することを目指し，適応的な身体表現のモデルが提案されてきた [5)6) 7)]．それらの研究では異なるセンサモダリティの同期性に基づき身体表現を確立している．この表現は多くのタスクの実行に必要不可欠だが，道具使用等の行動計画においては更に最終効果器の位置や動きが重要な要素である．既存研究では最終効果器の位置とその動きを環境中から発見できることは既知としている．また身体発見において重要であると考えられる注意について考慮していない．最終効果器に注目するには，顕著度等ボトムアップな要素と自身の経験から得るトップダウンな要素の両方から成る，生物の視覚的注意のメカニズムが有効だと考えられる．本研究では適応的な身体表現獲得への第一段階としてボトムアップな要素を考慮した注意メカニズムを用いて，視覚，触覚，体性感覚から cross-modal な身体表現を獲得する手法を提案する．Saliency Map に基づいた視覚的注意により視覚受容野を発見し，ロボットが触覚情報を検知したときを契機に視覚受容野と体性感覚情報を統合する．

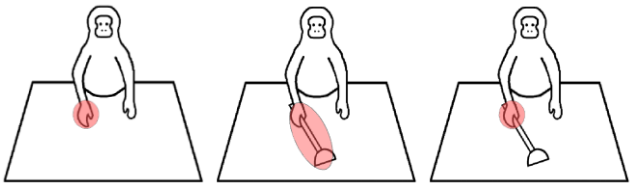
2 神経生理学における知見

脳神経科学等の分野でのこれまでの研究によって，感覚の統合は頭頂連合野において行われるという知見が得られている．中でもサルの頭頂間溝前壁においては，体性感覚刺激と視覚刺激の両方に対して反応する bimodal neuron が発見されている．入來 [1)] はサルの手の平に受容野を持つ bimodal neuron の活動を調べた．このときニューロンの受容野の定義の仕方としては，まずサルの手を動かすなどして体性感覚受容野を計測した．そして体性感覚受容野 (手) の回りの空間を視覚刺激用プローブ (レーザーポインタ) でくまなく走査し，ニューロンが活動するときのプローブの位置を求めた．その位置は手の周囲に密集しており，この空間をニューロンの視覚受容野とした．以上のようにして受容野を計測したニューロンの活動を，道具使用の学習の前後で計測した．その結果，サルが道具使用を学習した後は，手の平だけでなく道具への刺激にもニューロンが反応するようになることが観察された．そのイメージを図 1 に示す．

図 1(a) は道具使用を学習する前のニューロンの視覚受容野のイメージで，手の周囲に広がっている．サルが道具使用可能になった後は，視覚受容野は図 1(b) のように道具全体を含むように広がる．しかし道具使用可能となっても，単に道具を握らせただけでサルに餌をとる意図がない場合，図 1(c) のように視覚受容野は道具に沿って広がらず，手の周囲のままである．

また直接手元を見るのではなく，モニタを通して手元を見て道具使用する実験も行われた．この実験では，サルが手元やテーブルを直接見ることができないようサルの鼻の高さに不透明な板を固定し，代わりにサルの眼前に設置したモニタにサルの鼻先につけたカメラからの映像を映し道具使用させた．この場合，ニューロンの視覚受容野の変化は直接見る場合と同様の結果が得られた．更に，サルの姿勢が一定であっても，モニタ上の手の視覚像に対し拡大，縮小，移動等の加工が施されると，その変化に合わせてニューロンが反応した．

以上の結果は，道具使用によってサルの身体図式が変化している可能性を示すものである．



(a) Before tool-use (b) After tool-use (c) Passive holding

Fig.1 Changes in bimodal visual receptive field[1)]

3 提案システム

前述の知見を踏まえ，以下の考えに基づいた身体表現獲得のシステムを考える．

1. 視覚受容野を獲得する上で重要なのは視覚的注意である．視覚的注意によって学習時には視覚的に顕著な特徴を抽出でき，また検証時には，視覚受容野は視覚的注意が導かれた時活動するからである．
2. 将来道具のカテゴリ化を考えると，道具の最終効果器は学習過程で自身が発見することが重要である．
3. 視覚と体性感覚の統合の契機として触覚を利用する．

提案するシステムの概要を図 2 に示す．

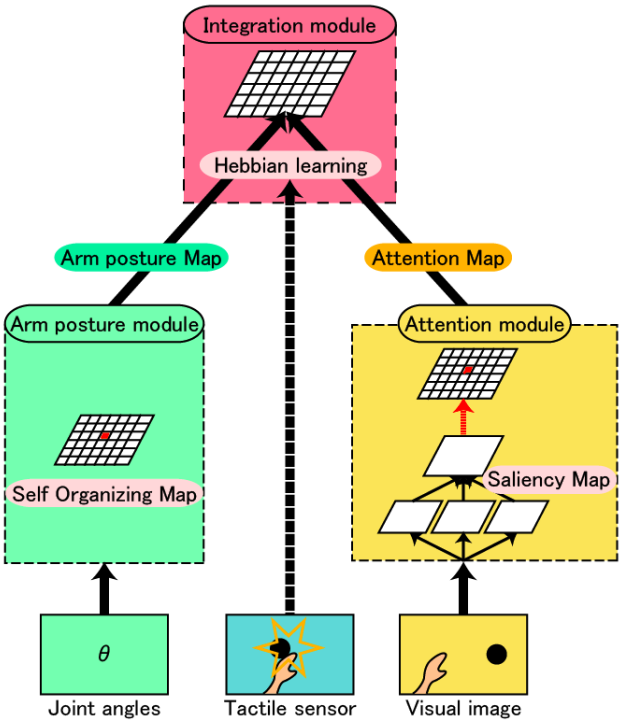
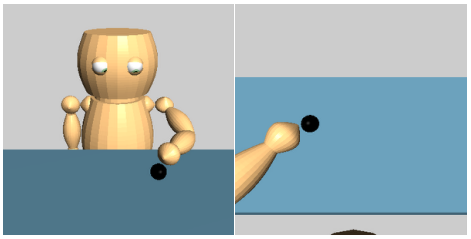


Fig.2 Overview of system

システムは 3 つのモジュールから成る．関節角度モジュールは体性感覚に相当しており，様々な姿勢における腕の関節角度を表す．関節角度は SOM として構成されている．注意モジュールでは，Saliency Map のアルゴリズムに基づき視覚画像の顕著な対象を注意点に決定する．決定された関節角度と注意点は，ロボットの手に触覚刺激が与えられたとき感覚統合モジュールによって統合される．統合には Hebb 学習を用いる．

3.1 ロボットモデル概要

システムの評価はシミュレーション実験によって行う．実験環境の様子を図 3 に示す．ロボットは 5 自由度の両腕を持つが，今回は簡略化のために左腕のみを使用する．また両眼ではなく単眼とし，その位置は両眼の中心とする．ロボットの前方にはテーブルが配置されている．テーブルの高さはロボットの胴体の重心と同じ高さとなっている．ロボットの視線先はこのテーブルの中心に固定している．実験中，ロボットはテーブル上で左手 (左手で把持された道具) をランダムに動かす．またテーブル上には物体をランダムな位置に配置し，ロボットの左手 (左手で把持された道具) が物体に接触すると，物体は新たな位置に配置される．



(a) Front view (b) From robot's eye

Fig.3 Experimental environment

3.2 SOM を用いた関節角度モジュール

学習前，ロボットに左手をテーブルの上の様々な位置へランダムに伸ばす運動を行わせ，その間の左腕の関節角度を記録しておく．そして記録した関節角度データを参照ベクトルの値として SOM(自己組織化マップ) を作成する．今回作成した SOM は 8×8 のユニットをもつ 2 次元 SOM である．作成した SOM を図 4 に示す．

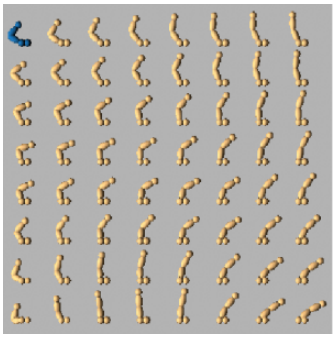


Fig.4 Arm posture map

学習中には，SOM の各ユニットの参照ベクトル $\mathbf{\Theta_i} = (\theta_1^i, ..., \theta_n^i)$ と現在の関節角度 $\mathbf{\Theta} = (\theta_1, ..., \theta_n)$ とを比較して最も類似度の高いユニットを勝者ユニットとし，勝者ユニットの参照ベクトルを現在の関節角度として決定する．n は関節角度の自由度で，今回は $n = 5$ である．通常 SOM では類似度の高さとは距離の近さ，すなわち入力関節角度とユニットの参照ベクトルとの距離である．ここではベクトル間のユークリッド距離とし，その距離が最も小さいユニットが勝者として発火する．そして各ユニットは勝者ユ

ニットからの距離 d_i に応じて活性する．勝者ユニットの番号を c とすると，各ユニットの活性度 a_i^{arm} は

$$a_i^{arm} = e^{-\beta d_i^2}, \tag{1}$$

$$d_i = \|\mathbf{\Theta_i} - \mathbf{\Theta_c}\|,\ c = \arg\min_i \|\mathbf{\Theta} - \mathbf{\Theta_i}\|, \tag{2}$$

となる．今回は $\beta = 100$ としている．

3.3 Saliency Map を用いた注意モジュール

注意モジュールは Saliency Map のアルゴリズムを利用している．Saliency Map は，Itti et al.[8] が提案したヒトのボトムアップな注意機構のモデルである．

そのアルゴリズムの流れは，まず入力視覚画像から特徴量それぞれのマップを作り，それらのマップをガウスピラミッドを基に分割する．今回の実験では入力画像のサイズは 512 × 512 であり，その分割段階は 9 段階である．分割段階を $\sigma = [0...8]$ とする．特徴量は次の 5 つである．

1. 明度：RGB 成分の合計
2. 色：RGB 成分それぞれと，RGB 成分から計算されるその他の色
3. フリッカー：連続する 2 つのフレーム間の差分
4. 傾き：ガボールフィルタによって計算した $\psi = [0°, 45°, 90°, 135°]$ の 4 種類
5. エッジの法線フロー：ガボールピラミッドの空間的移動の差分

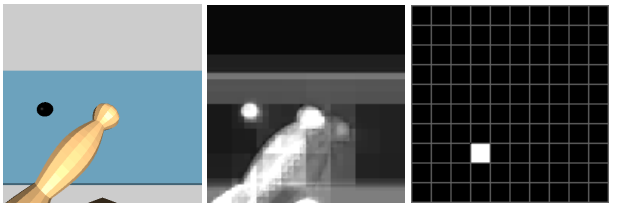
次に計算した Saliency 画像を 10 × 10 のユニットに区切る．ユニットの座標 $\mathbf{x_j}$ は，領域の中心とする．そして各ユニットについて，領域内の全ての画素値を足し合わせたものをそのユニットの SaliencyS_j とする．この S_j の度合いによって確率的に注意点座標 $\mathbf{x_k}$ を選択する．そして各ユニットは注意点座標からの距離 h_j に応じて活性する．その活性度 $a_j^{attention}$ は，

$$a_j^{attention} = e^{-\gamma h_j}, \tag{3}$$

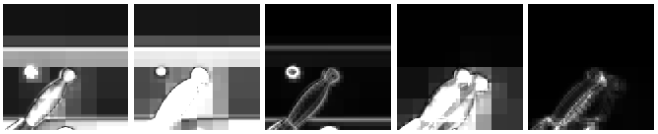
$$h_j = \|\mathbf{x_j} - \mathbf{x_k}\|, \tag{4}$$

となる．今回は $\gamma = 1$ としている．このようにして注意マップを作る．

図 5 に入力画像，それぞれの特徴量のマップ，合成された Saliency 画像，Saliency 画像から決定した注意点の例を示す．



(a) Input image (b) Saliency image (c) Attention map



(d) Intensity (e) Colors (f) Orientation (g) Flicker (h) Motion

Fig.5 bottom-up attention

3.4 感覚統合モジュール

感覚統合モジュールでは，ロボットが物体に触れ触覚情報が入力されたとき，この触覚を契機として関節角度マップと注意マップを Hebb 学習によって統合し，統合マップを作成する．ここで，触覚情報とは手への触覚反応のみである．このとき注意マップから統合マップへは 1 対 1 対応に投影させる．すなわち注意マップのユニット j から統合マップ k への結合荷重 w_{jk}^A は，

$$w_{jk}^A = \begin{cases} 1 & (j = k) \\ 0 & (j \neq k) \end{cases}, \tag{5}$$

である．関節角度マップのユニット i から統合マップのユニット k への結合加重 w_{ik}^B は

$$\Delta w_{ik}^B = \epsilon a_i^{arm} a_k^{attention}, \tag{6}$$

$$w_{ik}^B(t+1) = w_{ik}^B(t) + \Delta w_{ik}^B, \tag{7}$$

$$w_{ik}^B(t+1) \leftarrow \frac{w_{ik}^B(t+1)}{\sum_k w_{ik}^B(t+1)}, \tag{8}$$

となる．今回は $\epsilon = 0.05$ とし，また N_1 は注意マップのユニット数で $N_1 = 100$ である．w_{ik}^B は初期値 0.5 で初期化しておく．

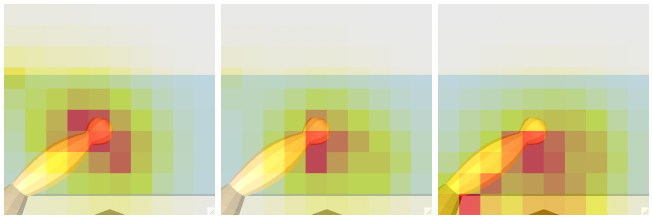
4 実験結果

実験は表 1 に示す条件の下で行った．

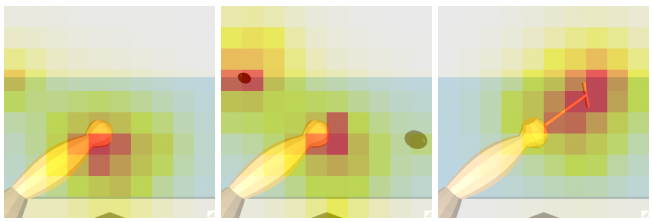
Table 1 Experimental condition

条件	最終効果器	物体の色	物体の数	Saliency
1	手	黒色	1	全て
2	手	机と同色	1	全て
3	手	黒色	1	動きのみ
4	手	黒色	1	動きを除く
5	手	黒色	3	全て
6	道具	黒色	1	全て

関節角度マップのあるユニットに対する統合マップの重みは，ある姿勢において視覚画像の各点を注意している度合いを示す．この統合マップの重みを色のグラデーションによって表したものが図 6 である．



(a) Case1 (b) Case2 (c) Case3



(d) Case4 (e) Case5 (f) Case6

Fig.6 Weights level

基本となる実験条件は条件 1 である．図 6(a) を見ると，手の周囲ほど結合が強くなっており，手先の統合ができて

いると考えられる．この結合を獲得した後は，ロボットはある点を注意した際，そのときの姿勢において結合している座標と注意点を比較することで，注意点が自身の身体にどれだけ近いか判断することができる．すなわち，図 6(a) の表現は図 1 に示すような bimodal neuron の視覚受容野に対応するものであると考えられる．これを検証するため，前述の入來らが bimodal neuron の受容野を計測した方法を参考にした検証実験を，統合マップの結合学習後に行った．すなわち実験環境を暗闇にし，レーザポインタでロボットの手の周囲をくまなく走査したときの統合マップの発火度を計算した．発火度 $a_k^{integrate}$ の計算は以下の通りである．

$$a_k^{integrate} = (\sum_j w_{jk}^A a_j^{attention})(\sum_i w_{ik}^B a_i^{arm}) \quad (9)$$

$$= a_j^{attention}(\sum_i w_{ik}^B a_i^{arm}), \quad (10)$$

これを図にしたものが図 7 である．手先の周囲ほど発火度が高くなっており，確かに bimodal な受容野の表現が獲得できているといえる．

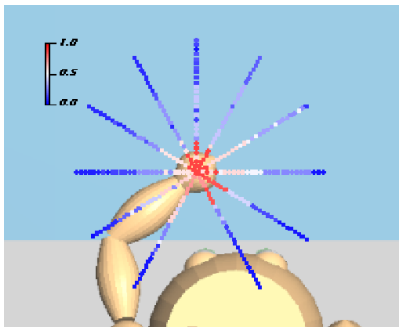


Fig.7 The activation level of bimodal neuron

また統合の際に重要な要素を検討するため，条件 1 と比較する条件として条件 2〜5 の条件下で実験を行った．条件 1 と 2 では物体が視覚的に興味を引く度合いが異なる．物体が興味を引きにくい条件 2 では，手の周囲の物体を注意する頻度が低いため結合範囲が狭いと考えられる．受容野が手の周囲に広がりを持つには，興味を引きやすい物体を用いるべきである．条件 1 と 3,4 では動きの特徴量の扱いが異なる．動きの特徴量のみで注意点を決定する条件 3 では，視覚画像において動きを持つ腕全体に結合が広がっている．よって注意点を動きの情報のみで決定するのでは最終効果器の位置を発見することは難しいと考えられる．条件 1 と 5 では物体の数が違い，条件 5 では図 6(e) に示されているように，2 つの余分な物体が視野内の固定位置に配置されている．正しい統合が行われず，最終効果器を発見できていない．

条件 6 は道具使用の場合で，結合は道具全体を含むように広がっており，これは道具が手先の一部として認識されたということを示していると言える．

5 結言

本研究では，Saliency Map に基づいたボトムアップな注意システムを用いて視覚的注意点を発見し，視覚，体性感覚を触覚を契機に統合することで bimodal な受容野を獲得するシステムを提案した．シミュレーションによってシステムの検証を行い，最終効果器が手先である場合，また道具である場合も同様にシステムが有効であることを示した．

今後は物体と作用する効果器というアフォーダンスや意図など，トップダウンな要素も含めた注意システムを構築すべきである．また今回は触覚反応さえあれば統合を行っているが，条件 5 のような場合正しい統合が行われない．感覚の同期性や報酬系についても考慮し，適応的な身体表現を獲得すると同時に最終効果器の位置をも獲得できるシステムを考える必要がある．

参考文献

[1] 入來篤史. 道具を使うサル. 医学書院, 2004.

[2] Athushi Iriki, Michio Tanaka, Shigeru Obayashi, and Yoshiaki Iwamura. Self-images in the video monitor coded by monkey intraparietal neurons. *Neuroscience Research*, Vol. 40, pp. 163–173, 2001.

[3] Angelo Maravita and Atsushi Iriki. Tools for the body (schema). *Trends in Cognitive Sciences*, Vol. 8, No. 2, pp. 79–86, 2004.

[4] Minoru Asada, Karl F. MacDorman, Hiroshi Ishiguro, and Yasuo Kuniyoshi. Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous System*, Vol. 37, pp. 185–193, 2001.

[5] Yuichiro Yoshikawa. *Subjective Robot Imitation by Finding Invariance*. PhD thesis, Osaka Univ., 2005.

[6] Cota Nabeshima, Yasuo Kuniyoshi, and Max Lungarella. Adaptive body schema for robotic tool-use. *Advanced Robotics*, Vol. 20, No. 10, pp. 1105–1126, 2006.

[7] Alexander Stoychev. Toward video-guided robot behaviors. *Proceedings of the Seventh International Conference on Epigenetic Robotics*, pp. 165–172, 2007.

[8] Laurent Itti, Nitin Dhavale, and Frederic Pighin. Realistic avatar eye end head animation using a neurobiological model of visual attention. *Proceedings of SPIE*, 2003.