

霊長類における身体動作時系列の分節推移構造推定

Body Motion Segmentation in Non-Human Primate Based on Gaussian Process Hidden Semi-Markov Model

三村 喬生 *1

Koki Mimura

中村 友昭 *2

Tomoaki Nakamura

松本 惇平 *3

Junpei Matsumoto

西条 寿夫 *3

Hisao Nishijo

須原 哲也 *1

Tetsuya Suhara

持橋 大地 *4

Daichi Mochihashi

南本 敬史 *1

Takafumi Minamimoto

*1量子科学技術研究開発機構

National Institutes for Quantum and Radiological Science and Technology

*2電気通信大学

The University of Electro-Communications

*3富山大学

University of Toyama

*4統計数理研究所

The Institute of Statistical Mathematics

Understanding the nature of nonverbal communication (eye contacts, face expressions, body postures, hand gestures, body motions, etc...) is one of the core issue in behavioral neuroscience. In this study, we demonstrated the data-driven dynamical segmentation of the body expressions in free moving small non-human primate, common marmoset. We developed a new marker-less 3D motion tracking system optimized to marmoset. Then, we proposed unsupervised segmentation using a Gaussian process-hidden semi-Markov model (GP-HSMM). As a result, we succeeded to classify three types of marmoset feeding behavior (high position feeding, low position feeding, and low position feeding with hands) only based on body parts positions, face direction, and body angle information. This result suggested that proposing system could represent high versatility to quantify the animal nonverbal body expressions without qualitative teacher labels.

1. はじめに

「目は口ほどに物を言う」という慣用句があるが、視線・表情・姿勢・動作などの非言語表現は、意図や感情などヒトの内部状態の伝達手段として機能している [Coulson 2004]。ヒトにおいて特異的に複雑化した言語表現の進化的起源は、非ヒト霊長類など社会集団を構成する動物種においても広く観察される身体表現を用いた非言語表現に求められ [Burrows 2008, Kemp 2013]、社会的コミュニケーションの本質的要素を成していると考えられる。しかし、言語表現に紐づけて解釈が可能な例外 (手話やジェスチャーなど) を除き、非言語表現を定量的に解析する有効な手段は知られていない [Knapp 2014]、コミュニケーションダイナミクスの包括的な記述には至っていない。

小型霊長類コモン・マーモセット *Callithrix jacchus* は、他個体に対する協力性や公平性を認識するなど豊かな社会性を示す [Yasue 2018]。また、マーモセットは餌が設置される状況に応じて手や歯を使った多様な摂餌行動パターンを使い分ける事が知られ、他個体の摂餌行動を観察して模倣するなど社会的機能の基盤的動作として報告されている [Voelkl 2000]。しかし、自由な運動を妨げずに視線や身体運動情報を定量的に測定することはこれまで困難であった。

ヒトの言語表現に関する研究 (自然言語処理) においては、連続する文字列から形態素や統語構造を推定する構成的アプローチが試みられている [Mochihashi 2009]。内部状態の伝達という機能的相同性に着目すると、非言語表現にもこれと同様の最小単位や統語的な時間構造が仮定できる。既に、自然言語処理を応用し、淀みなく連続するヒトの動作の観測情報から時系列的遷移構造を推定する試みも報告されているが [Taniguchi 2011, Nakamura 2017]、非ヒト動物の動作解析では自己回帰

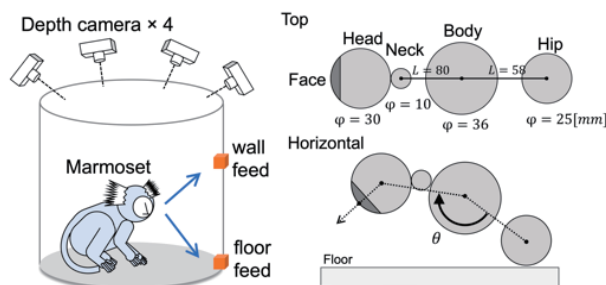


図 1: 摂餌行動の測定および骨格モデルのデザイン。

隠れマルコフモデル (AR-HMM) など単純な時系列構造推定に留まり、モデルの表現力は限定的である [Wiltchko 2015]。

一方、動物行動研究分野からは多くの動物種に対応したモーショントラック技術が報告されており [Mathis 2018]、ヒトのモーションを 2D 画像からトラッキングする技術も高精度化している [Tome 2017 など]。これらの技術によりトラッキングログ取得のハードルが低下しつつあり、集積されたデータから行動エソグラム (動物固有の行動パターン) やその文脈的な意味構造に対する解析手法が求められている。

そこで本研究では、まず、マーモセットに最適化された独自の高精度度モーショントラック技術を開発・実装した。深度カメラデータ解析と物体検出アルゴリズムを組み合わせることで、ボディマーカを用いずに摂餌行動時の一連の動作について身体パーツの空間配置と顔の方向ベクトルを推定することに成功した。続いて、既にヒトの身体運動パタンの解析において優れた成績を示しているガウス過程隠れセミマルコフモデル (Gaussian Process Hidden Semi-Markov Model, GP-HSMM, [Nakamura 2017]) による時系列分節構造の推定を応用し、モーショントラック情報からデータ駆動に、動作パターンに含まれる単位構造とその時系列遷移を解析した。

連絡先: 三村 喬生, 量子科学技術研究開発機構 放射線医学総合研究所, 千葉県千葉市稲毛区穴川 4-9-1, 電話番号 043-206-3251, mimura.koki@qst.go.jp

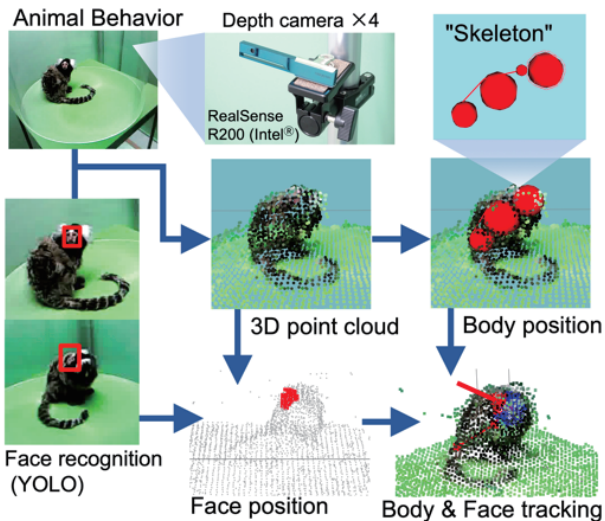


図 2: マーカーレス 3 次元モーショントラッキング技術の概要。

2. 方法

2.1 行動記録

本研究では、実験環境に馴染みのマーマセット (成体オス 1 頭, メス 3 頭) を用い、計 25 回の摂餌行動を計測した。摂餌行動は設置される餌の位置によって 2 種類の分類が含まれる (床 floor, 17 回; 壁 wall, 8 回)。実験には死角のない透明なアクリル円柱型行動試験ケージ (径 40cm; 高 50cm) を用い、動物はケージ中で自由に行動しながら壁面もしくは床に設置された餌 (約 5mm 角に成形されたタマゴ蒸しパン片) を任意のタイミングで摂餌した [図 1 左]。測定には、深度カメラ RealSense R200(Intel 社) を用いた。この R200 はパターン投射用の近赤外プロジェクタと 2 台の近赤外カメラを搭載し、視差を用いて深度距離を 1cm スケールで算出することができる (時間解像度: 30 frame/sec)。行動試験ケージを 4 方向から撮影し、ソフトウェア上で空間キャリブレーションを行い、統合することにより 3 次元点集合 (point cloud) を得た [図 2]。

2.2 モーショントラッキング

測定された試験フィールドの point cloud から色相フィルタを用いてマーマセットの身体概形を抽出し、ボールとパスからなるマーマセット骨格モデルを物理シミュレーションによりフィッティングする事で、身体パーツ (頭・首・胴・腰) の空間位置を推定した。骨格モデルは球体で代表される身体パーツとそれを結ぶ骨格パスからなる。骨格パスには長さや角度の律則があり、身体パーツ間の距離と相互配置を緩く制限している [図 1 右]。また、身体パーツには、それぞれ point cloud に対して引力範囲と斥力範囲が中心からの距離 (半径) として設定されており、この釣り合いが取れる安定な空間配置に身体パーツがアトラクトされる事で推定が完了する。2D 画像における物体認識を用いた frame-by-frame のモーショントラッキング (例えば openpose や deeplabcut などの手法) に対し、本手法は frame を独立とせず、ある frame における推定結果が次の frame における推定の初期配置として用いられるため、時系列に対して自然で連続した身体運動を表現する事ができる。一連の計算アルゴリズムは公開されている齧歯類のトラッキング解析プログラム^{*1} [Matsumoto 2013] をマーマセット用に最適化して実装された。更に、顔の向きを検出するため frame-by-frame に 2D

*1 <http://3dtracker.org/>

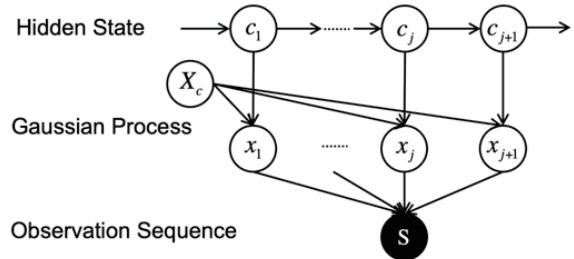


図 3: GP-HSMM のグラフィカルモデル。

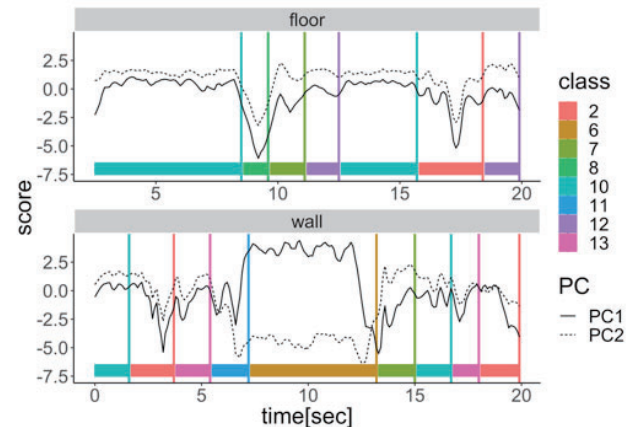


図 4: GP-HSMM を用いた分節構造推定結果の代表例。

RGB 画像上で顔の座標を推定し、point cloud に投影することで、空間座標を推定した。2D 画像における顔検出には深層学習を用いたオブジェクト検出アルゴリズム YOLO [Redmon, 2016] を用い、教師データ 2500 セットに対し学習を行った (学習率 0.1%, 積算回数 5 万回)。YOLO により検出された顔範囲に含まれる point cloud を 4 方向から抽出し、頭中心座標からの距離で抽出 ($d \leq 5 \text{ cm}$) した後、重心を算出することで空間中の顔の座標を身体マーカを用いずに推定した [図 2]。

2.3 身体動作の分節構造推定

身体動作の分節化のために、部位 (顔・頭・胴・腰) ごとの床からの高さおよび運動速度と、骨格モデルの屈折角 [図 1 左, θ] を含めた計 9 種類の特徴量を抽出し、Loess 回帰により時間軸に対して 10 Hz に平滑化した後、主成分分析により 2 次元に圧縮したデータを用いた [図 4, 6]。分節構造推定の実行は、ガウス過程セミマルコフモデル (GP-HSMM) を用いた [図 3]。当該手法は、連続量のパラメータによる時系列多次元軌道に対し、ガウス過程を用いたクラスタリングによるセグメントの分類と同時に、セミマルコフモデルによりセグメント長を推定し、さらにセグメント間の時間的推移構造をマルコフモデルにより考慮するものである。解析プログラムは公開されているもの^{*2} をカスタマイズして用いた。

学習のプロセスでは、まず全ての摂餌行動時系列をランダムに分節・分類し初期化する。今回のデータでは平均長 20 frame に設定した。次に、ある摂餌行動時系列 s_n を分節化して得られた単位系列 $x_{nj} (j = 1, 2, \dots, J_n)$ を、それらが分類されているクラス $c_{nj} (j = 1, 2, \dots, J_n)$ から全て削除し、各クラスタのガウス過程パラメータ X_c と遷移確率 $P(c|c')$ を更新する。更新されたパラメータを用い、新たな単位時系列とそれらのクラスを推定する事で、パラメータの更新を繰り返す (Forward

*2 <https://github.com/naka-lab/GP-HSMM>

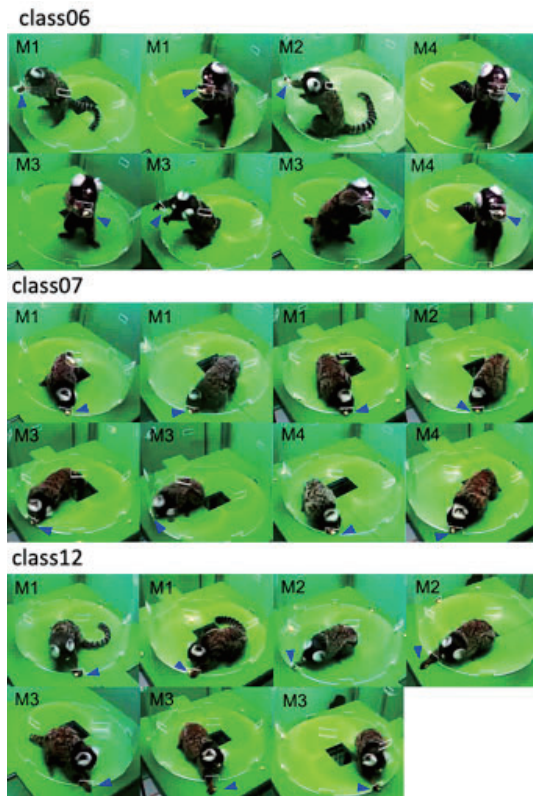


図 5: 異なるクラスに分類された摂餌行動の代表例。

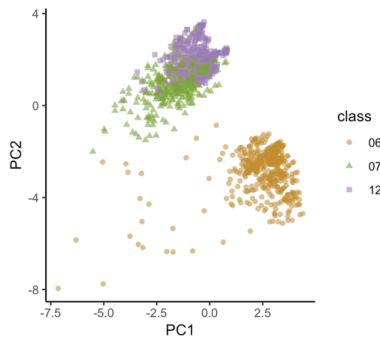


図 6: 特徴量主成分平面におけるクラスの分離。

filtering-Backward sampling)。これにより、観測系列の分節点とガウス過程のパラメータを交互に最適化する事ができる [図 3, 4, Nakamura 2017 に詳述]。

得られた餌設置条件ごとのクラス出現率を集計し、その差異パターンを階層的クラスタ解析 (完全連結法) により解析した [図 7B]。このクラスタ構造に従ってクラスの遷移確率行列を Hinton diagram により可視化した [図 7C]。

3. 結果

3.1 摂餌行動セグメント

25 セットの摂餌行動 (計 500 秒, 10fps) から抽出された 9 次元の行動特徴量の第 1, 2 主成分スコア時系列に対し、クラス数を 15 種類に設定した GP-HSMM による分節構造推定 (分節平均長 20 frame) を行った。結果 [図 4]、11 種類が採用され、4 種類のクラスは頻度 0 となった。これは最適クラス数の近似的な推定結果とみなす事ができる。

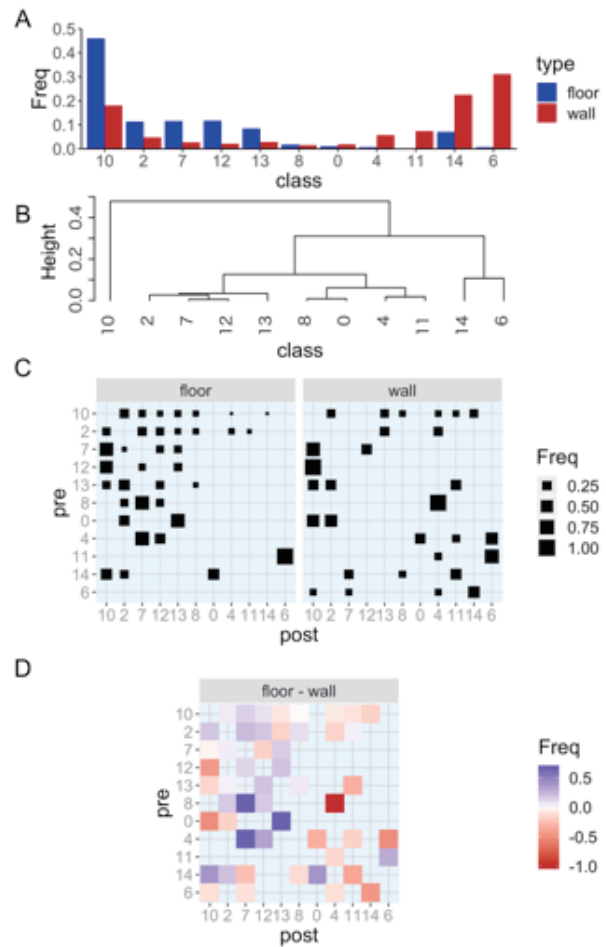


図 7: 餌の設置条件に応じた時系列遷移構造の変化。

最も多く検出されたクラス 10 は、あまり身体を動かさず顔の向きの変化のある行動が含まれていた。分節化後にビデオログから摂餌タイミングが含まれるクラスを検討したところ、大きく 3 つのクラスが得られた。それぞれ、床に置かれた餌を口を使って直接摂餌する場合の行動 [図 5 中, class07]、壁に置かれた餌を手を使って摂餌する行動 [図 5 上, class06]、および床に置かれた餌を手を使って摂餌する行動 [図 5 下, class12] を含んでいた。行動特徴量の主成分平面において、これら 3 種類の摂餌行動クラスについて分布を確認したところ、壁・床の条件差異は当該主成分得点においても認められたが、床条件における 2 種類の摂餌行動 (class07, 12) の分布は重複しており、動作クラスの分離にはパラメータの時系列構造を考慮したモデル解析が有効である事が示唆された [図 6]。

3.2 摂餌条件による遷移構造変化

検出された 11 種類の行動クラスの出現率は、餌の設置場所の条件 (壁・床) に応じて異なっていた [図 7A]。条件ごとのクラス出現率に対する完全連結法 (complete linkage method) による階層的クラスタ解析 [図 7B] に基づいた順序で遷移確率および条件間差異を可視化した [図 7CD]。結果、共通して高頻度で観測されたクラス 10 の他、床条件 (2, 7, 12, 13) と、壁条件 (0, 4, 6, 11, 14) のそれぞれにおいて相互に強く関連づけられたクラス群を検出した。分節化後にビデオログを参照したところ、それぞれの摂餌行動における特徴的な行動要素が含まれていた [表 1 に詳述]。

表 1 餌設置条件に応じて特徴的に検出された行動分節

餌条件	クラス	特徴
床	2	前傾になり頭を床に近づける
床	7	摂餌行動 (床, 手を使わない)
床	12	摂餌行動 (床, 手を使う)
床	13	視線を周囲に大きく動かす
壁	0	視線を上から下に移動する
壁	4	大きく動き状態を起こす
壁	6	摂餌行動 (壁, 手を使う)
壁	11	壁にある餌にむけて立ち上がる
共通	10	大きく動かさず周囲を見渡す

4. まとめ

本研究では、非言語表現を高解像度かつ汎用的に解析する手法の開発および実装として、小型霊長類コモン・マーマセットの典型的な摂餌行動のログデータを取得し、身体動作時系列の分節推移構造推定を行った。データ取得には深度カメラとオブジェクト検出器を組み合わせた新規のマーカーレス・3次元モーショントラッキング技術を開発・実装し身体部位のトラッキング情報を抽出した。分節推移構造の推定には、ガウス過程の導入により多次元連続量を取り扱える拡張を施した隠れマルコフモデル GP-HSMM を用いた。

当該手法を用い、多くの動物種において自然に観察可能な代表的な目的指向型行動 (goal-directed behavior) である摂餌行動を解析した。マーマセットの摂餌行動は、手や指を駆使した複雑な行動パターンを示す。提案手法により、手の運動情報・餌の空間配置情報・あるいはヒトの観察による教師ラベルを用いずに、身体部位 (顔・頭・胴・腰) の時系列情報のみから摂餌行動パターンの分離が得られたことは、本解析手法が高解像度な行動エソグラム分離器として機能したことを示唆している。

マーマセットでは既に遺伝子操作/改変・神経薬理操作・神経活動記録・脳イメージングなど神経科学分野の先端的な研究・解析技術の適用が報告されている。提案手法による身体表現の行動エソグラム評価と組み合わせることで、精神・疾患モデル動物の表現系の評価が可能となり、ヒト臨床における言語表現に障害がある精神・神経疾患の病態評価、言語機能が未成熟な発達初期における障害リスクの早期診断など幅広い応用が見込まれる。

謝辞

本研究にあたり島谷健一郎准教授 (統計数理研究所)、谷口忠大教授 (立命館大学)、および鈴木航室長・浅場明莉研究員 (国立精神・神経医療研究センター) に多くの有益なアドバイスをいただいた。また本研究は、科研費 JP17H06040 およびAMED 課題番号 JP18dm0307007 の支援を受けて実施されたものである。ここに感謝の意を表する。

参考文献

- [Burrows 2008] Burrows, M. A.: The facial expression musculature in primates and its evolutionary significance, *BioEssays*, 30 (2008).
- [Coulson 2004] Coulson, M.: Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence, *Journal of Nonverbal Behavior*, 28 (2) (2014).

- [Kemp 2013] Kemp, C., Kaplan, G.: Facial expressions in common marmosets (*Callithrix jacchus*) and their use by conspecifics, *Animal Cognition*, 16 (2013).
- [Knapp 2014] Knapp, L. M., Hall, A. J., and Horgan, G. T.: *Nonverbal Communication in Human Interaction*, Wadsworth, Cengage Learning (2014).
- [Mathis 2018] Mathis, A., Mamidanna, P., Cury, M. K., Abe, T., Murthy, N., V., and Mathis, W., M., DeepLabCut: markerless pose estimation of user-defined body parts with deep learning, *Nature Neuroscience*, 21 (2018).
- [Matsumoto 2013] Matsumoto, J., Urakawa, S., Takamura, Y., Malcher-Lopes, R., Hori, E., Tomaz C., Ono T., and Nishijo, H. A 3D-video-based computerized analysis of social and sexual interactions in rats, *PLoS One*, (2013).
- [Mochihashi 2009] Mochihashi, D., Yamada, T., and Ueda, N.: Bayesian unsupervised word segmentation with nested Pitman-Yor language modeling, *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, (2009).
- [Nakamura 2017] Nakamura, T., Nagai, T., Mochihashi, D., Ichiro, K., Asoh, H., and Kaneko, M.: Segmenting Continuous Motions with Hidden Semi-markov Models and Gaussian Processes, *Frontiers in Neurobotics* (2017).
- [Redmon 2016] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., You Only Look Once: Unified, Real-Time Object Detection, *arXiv* (2016).
- [Taniguchi 2011] Taniguchi, T., Hamahata, K., and Iwahashi, N.: Unsupervised Segmentation of Human Motion Data Using a Sticky Hierarchical Dirichlet Process-Hidden Markov Model and Minimal Description Length-Based Chunking Method for Imitation Learning, *Advanced Robotics*, 25 (2011).
- [Tome 2017] Tome, D., Russell, C., and Agapito, L., Lifting from the Deep: Convolutional 3D Pose Estimation from a Single Image, *CVPR*, (2017).
- [Voelkl 2000] Voelkl, B., and Huber, L., True imitation in marmosets, *Animal Behavior*, 60 (2000).
- [Wiltschko 2015] Wiltschko, A. B., Johnson, M. J., Iurilli, G., Peterson, R. E., Katon, J. M., Pashkovski, S. L., Abaira, V. E., Adams R. P., and Datta, S. R., Mapping Sub-Second Structure in Mouse Behavior, *Neuron*, 88(6) (2015).
- [Yasue 2018] Yasue, M., Nakagami, A., Nakagaki, K., Ichinohe, N., Kawai, N., Inequity aversion is observed in common marmosets but not in marmoset models of autism induced by prenatal exposure to valproic acid, *Behavioral Brain Research*, 343 (2018).