

人の骨格点と物体の空間的関係の時系列データ解析による人物行動予測の基礎検討

○横田 雅恵（中央大） パトハック サーサク（中央大）
新妻 実保子（中央大） 梅田 和昇（中央大）

We aim to estimate the intention behind the actions of humans in order to enable flexible HRI. By estimating the intention behind a humans actions, the robot can estimate the context and make necessary adjustments. As the first step towards his goal, in this paper, we analyze time series data of 3D human skeletal points with respect to interactable objects for intention estimation and action prediction. We conducted an experiment to observe human movement and determined the necessary parameters for estimation. As a result, we were able to predict action with 91% accuracy using a Temporal Convolutional Network.

1. 緒言

近年、自動化においてロボットのタスクは複雑化し高機能化している。そのため、不確実性の高い実環境ではロボットのみで作業を完結させるのではなく、人とロボットが補完的に協力し合い作業することで適応している。そこで、ロボットと人との効率的な共生を実現する必要がある、柔軟かつ人に負荷の少ないヒューマンロボットコミュニケーション手法が研究されている [1]-[3]。この共生において、人がロボットに対して作業ごとに指示や操作を行うことで人の作業負担を新たに増やしてしまうことを避けるため、ロボットが能動的に動くことで人間同士でのような自然な連携による効率化を達成することが望ましい。そのため、ロボットが人や周囲環境の情報を取得し解釈する課題に対して、様々なアプローチで研究が行われている [3]-[6]。人からの明示的な指示を待つのではなく、人の動きからロボットが動作の意図や文脈を理解し、自身の役割を理解し先回りすることで人の作業負担を真に軽減することが出来る。例えば、人がドアに向かって歩いている時、これまでの人や周囲の状態や今の状態からロボットは人の外出の意図を推定し、先んじてドアを開けに向かうなどの積極的な行動が望ましい。ここで、動作の意図とは人の持つ任意の欲求を解決するために人がこれから取る行為系列を指す。この行為系列を推測し、より柔軟なヒューマンロボットコミュニケーションを実現することが本研究の目指すところである。

意図の推定には、その環境下で実現可能な意図と、達成のための複数の行動を考慮する必要がある。そのため、ある環境下で実現可能な行動を確実に予測できなければならない。そこで、人と物体との間の空間的関係の時系列データから人が物体に対して起こす行動を予測できるかを確認したい。

本論文では、行為系列の推測の基礎検討として、時系列を考慮した骨格点情報と物体情報を用いて、人が椅子に座る行動と座らずに通過する行動の予測に必要な情報を明らかにし、動作意図の推定への貢献を確認することを目的とする。また、ロボットと人の共生の場においては様々なシーンを想定することが出来るが、本研究では日常生活を想定した。

2. 関連研究

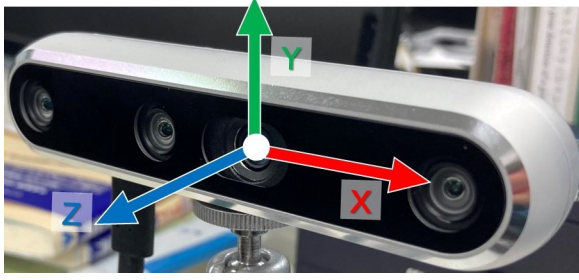
人とロボットが協働する際に、ロボットが人の行動やその目的や傾向などを理解するために様々なセンシング手法が研究されている。その中でも、コンピュータビジョン分野においては人とロボットの間の情報交換を促進し、自然で直感的なコミュニケーション方法を構築することで協働の有効性を向上させている [7]。コンピュータビジョンでのインタラクション手法の中では、人の行動を観測することで人の動きの意図を理解し、ロボットが積極的に行動して人の作業を助ける研究がされている [8]。画像情報を用いることで空間的な解釈を踏まえた人の意図を解析することが出来るが、人の曖昧な意図を画像情報のみで汲み取ることは課題が多い。そこで、本論文では人の確定的な意図に基づいた行動を観測し、意図によって異なる行動の特徴を明らかにすることで、動作意図の推定手法へ発展させることを目的とした行動予測を行う。

3. 人と物体の時系列空間的関係を用いた行動予測の検討

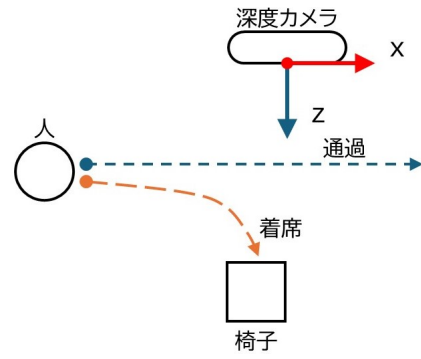
3.1 手法概要

本論文では、行為系列推測を目指した行動予測のための予備実験を行った。着席と通過の2種類の確定的な意図のものと行動を観測し、人と物体との空間的関係の時系列データを解析し、行動予測を行った。

深度カメラである Intel 社の Intel RealSense Depth Camera D455 と骨格点検出に LIPS 社の LIPSense 3D Body Pose SDK を用いて、人の骨格点の3次元座標を取得し観察した。また、時系列情報から予測を行う手法として Temporal Convolutional Network (TCN) [9] を用いる。行動予測には様々なネットワークが用いられるが、自然言語処理や時系列情報の処理には TCN が注目されている [10][11]。また、[12] では動きが激しく、柔軟性の高い点運動の予測にも TCN が用いられており、実環境下での人の行動にも有用であると考えた。そこで、本論文では人が物体に対して何らかの行動を起こすことを「働きかけ」と定義し、TCN を用いて人の骨格点と椅子との空間的関係から人から椅子への働きかけを予測する。

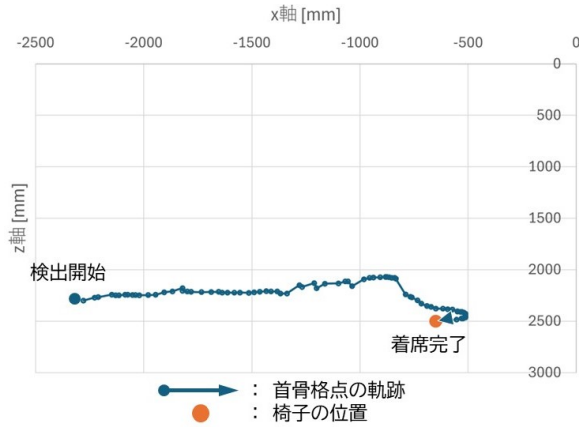


(a) RealSense 座標系

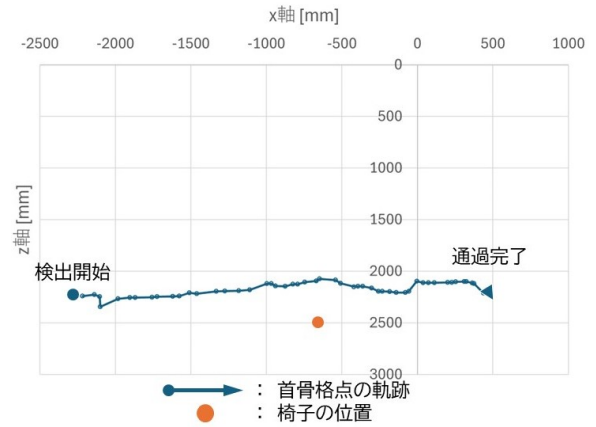


(b) 実験環境

図 1: データ取得時の環境



(a) 着席時



(b) 通過時

図 2: 首の骨格点の軌跡俯瞰図

ここで、TCN での処理に実際の座標値をそのまま使ってしまうと計測誤差による影響や、人と椅子それぞれの初期位置に依存してしまうと考えた。そのため、人と椅子の位置関係を利用したパラメータを生成する必要がある。そこで、実験 1 では様々な被験者の骨格点情報を取得し、椅子との位置関係の時系列変化を観察し、パラメータを検討した。続いて、実験 2 では実験 1 を通して選定したパラメータによる TCN での行動予測を行い、結果を観察した。

3.2 実験 1: パラメータの検討

3.2.1 実験概要

本項では骨格点情報を取得し椅子との位置関係の時系列変化を観察することで、TCN での行動予測に用いるための適切なパラメータを検討することを目的とする。深度カメラの座標系は図 1a に示すとおりである。

LIPSense 3D Body Pose SDK で検出可能な 18 の骨格点の内、本論文では首の骨格点を抽出し観察した。首の骨格点の選定理由は 2 点ある。まず、人と物体の位置関係を明らかにするにあたり体動によるノイズと、安定して検出できることを考慮したためである。身体の末端であるほど体動によるノイズが大きいと考えられるため、身体の中心部や体軸に近い骨格点が望ましい。次に、カメラと被写体の位置関係に依存せず安定

して検出される骨格点は数少ない。これらの理由から、人の移動軌跡の観察のため首の骨格点に注目することとした。

3.2.2 実験結果

続いて、実際に得られた骨格点の時系列情報を整理した。データを取得した実験環境を図 1b に示す。人から椅子への働きかけには着席のみが考えられる。被験者には経路について詳細を指定せず、着席と通過の 2 種類の行動を依頼した。本実験は成人 9 名を対象に行い、全員に計測の許可を取った。

床と並行である xz 平面での首の骨格点の軌跡の一例は図 2 のようになった。この図より、着席に際して、被験者によって進行方向に向かって左右に触れる動きの個人差はあったが、弧を描く軌跡ではなく概ね直線的な経路であった。また、身体を横にスライドさせるように座る場合も、一度椅子の正面に立ち止まってから座る場合も、首の位置は一度椅子の前を少し通り過ぎてから戻る動きを示していることが分かった。

続いて、首の骨格点の物体との距離値の時系列変化は図 3 のようになった。図 3a では椅子の付近で一定になった後急激に下がる。図 3b では一定の傾きで椅子に近づいていくが、どの被験者の場合も約 500mm より近くなることはなかった。

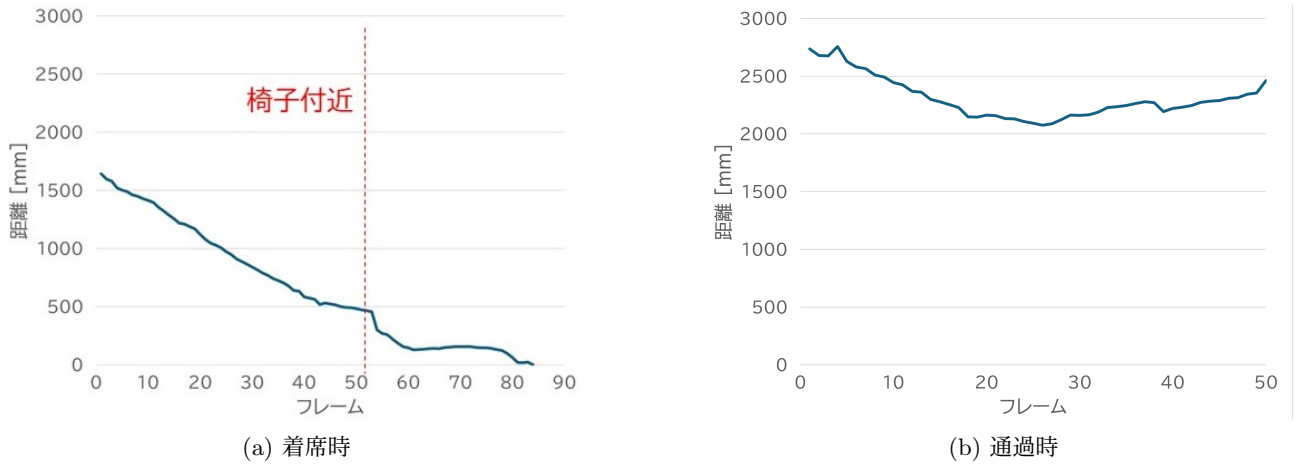


図 3: 首の骨格点と椅子との距離の時系列変化

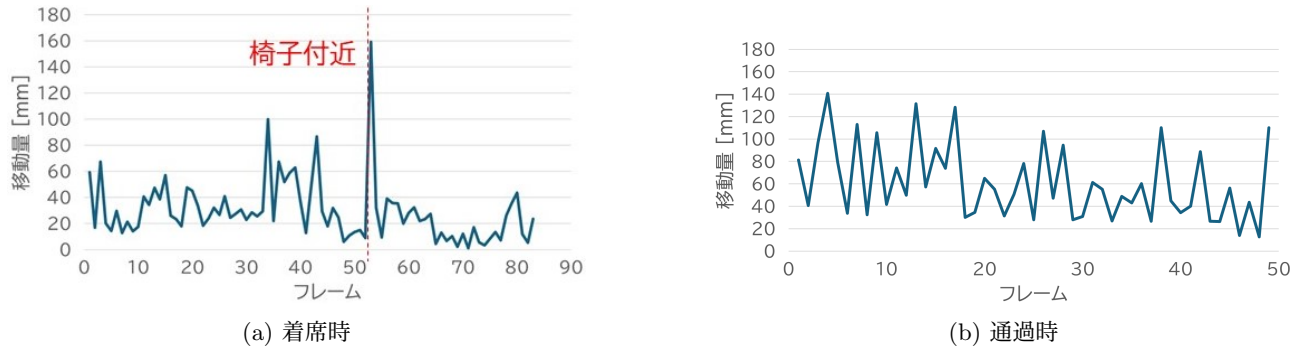


図 4: xz 平面上の首の骨格点の移動量の時系列変化

次に、xz 平面上でのフレーム間移動量の時系列変化は図 4 のようになった。図 4a では椅子に近づくにつれ移動量が小さくなり、図 3a で椅子との距離が一定になった直後に移動量は急増し、その後急激に下がることが分かる。一方、図 4b では椅子との距離が近づいていっても移動量はあまり大きくは変動しないことが分かる。

以上の結果より、首の骨格点に着目することで、椅子に働きかける意図を持つ時とそうでない時とを見分けることができると考えられる。

3.3 実験 2：TCN での行動予測

3.2 より、次の二つの特徴量を用いて着席と通過を判断することとする。 f は現在のフレーム時点であることを示す。また、特徴量を表 1 のように解釈した。

- xz 平面上の首骨格点と椅子との距離： d_f mm
- xz 平面上の首骨格点のフレーム間移動量： m_f mm

これらの解釈をもとに、表 2 のような TCN モデルを用いて複数人のデータから着席か通過かの予測を行ったところ、91%の精度のモデルを得られた。本実験では簡単な TCN モデルを用いて行動予測を行った。畳み込み層を増やすと約 100 エポック以降から損失関数が増大してしまったため、最も増大しにくかった 2 層を採用した。また、ウィンドウサイズを変化させて観察したところ、10 より小さくすると損失関数が十分に下がりきらず、10 より大きくすると約 120 エポック以

表 1: 特徴量の値の解釈

各特徴量の値 [mm]	解釈
$d_f > d_{f-1}$	椅子に近づく
$d_f \leq d_{f-1}$	椅子から遠ざかる
$d_f \leq 50$	椅子に到達
$m_f > 30$	移動速度が大きい
$10 < m_f \leq 30$	移動速度が小さい
$m_f \leq 10$	ほぼ静止している

表 2: TCN モデルの構造

畳み込み層	2
バッチサイズ	64
ウィンドウサイズ	10
カーネルサイズ	2

降から損失関数が増大したため、安定して損失関数が下がっていく値を採用した。

4. 考察

実験 1 では首の骨格点のみに着目し、体動に影響されにくい骨格点であったため、人が物体に働きかける

明確な意図を持っている際の物体への近づき方について観察することができた。まず、図 2a では椅子の付近でさらに椅子に近づき、回り込むような軌跡を描いていることが分かった。また、図 3a では、椅子の付近で動きが停滞し、その後急激に近づくという変化の様子を見られた。そして、図 4a では、椅子に座るという明確な意図を持っているほうがフレーム間移動量に大きな差が出やすいことが分かった。体動に影響されにくい骨格点である分、行動予測に必要な情報量を得られにくいと考えていたが、着席と通過の 2 種の行動については、首の骨格点から十分な差異を観察することができた。一方で、首の骨格点情報のみでは顔や体の向きを読み取ることができないため、より多様な物体への働きかけの意図を推定するためには他の骨格点の採用も検討する必要がある。

実験 2 では、91%の精度の着席と通過を予測するモデルが得られた。この精度の値から、着席と通過の行動予測に適した特徴量を採用できたと考えられる。しかし、本実験で扱った行動は着席と通過の二種のみであり、着目した物体も座る用途のみで扱われる椅子であった。そのため、単純かつ体の動きが必然的に大きくなる行動の予測には首の骨格点のように体動の影響が小さい特徴量で充分であったと考えられる。人の様々な動作の意図を予測するためには、人から物体への可能な働きかけの多様性を加味し、その様子を観察した上で適切な特徴量を選定する必要がある。また、本実験では椅子のある方向へ歩いてきてそのまま座るというシナリオであったが、通常、人は様々な物体に働きかけて生活しているため、一つの物体と人との関係を見るだけでは動作意図を正確に捉えることはできない。そこで、ある物体に働きかける前にその人がどの物体にどのように働きかけていたかという情報も必要である。人から多様な物体への働きかけを物体ごとに整理するだけでなく、それらの働きかけ同士の関係性を明らかにすることで、人の動作意図を理解し、予測することができると考える。

5. 結言

本論文では、首の骨格点と椅子の位置関係を考慮した時系列情報と、首の骨格点のフレーム間移動量から TCN を用いた着席と通過の行動予測を行い、人が物体に働きかける意図の予測のための時系列情報の基礎検討を行った。しかし、予備実験の環境や着目した行動は極めて単純であったため、物体の多様化や、人の行動や意図の複雑化にも対応できることが望ましい。そこで、以降は解析・手法の検討とともに以下の二点に留意する必要がある。

- 人からの物体ごとへの多様な働きかけと、物体間での働きかけの関係性を考慮する。
- 物体の用途ごとに選定した他の骨格点の時系列情報も特徴量として扱うことで、より汎用性の高いモデルを得られると考えられる。

将来的には、一つの部屋の中で多様な物体への人の働きかけを時系列的に整理・解釈し、行為系列を予測することで人に対し機器やロボットが先回りをして能動的に動作することを目指す。

参 考 文 献

- [1] D. Mukherjee, K. Gupta, L. H. Chang and H. Najjaran: "A Survey of Robot Learning Strategies for Human-Robot Collaboration in Industrial Settings," *Robotics and Computer-Integrated Manufacturing*, Volume 73, 102231, 2022.
- [2] D. Wei, L. Chen, L. Zhao, H. Zhou and B. Huang: "A Vision-Based Measure of Environmental Effects on Inferring Human Intention During Human Robot Interaction," *IEEE Sensors Journal*, vol. 22, no. 5, pp. 4246-4256, 1 March1, 2022
- [3] Y. Zhang and T. Doyle: "Integrating intention-based systems in human-robot interaction: a scoping review of sensors, algorithms, and trust," *Frontiers in Robotics and AI*, Volume 10, 2023, DOI: 10.3389/frobt.2023.1233328.
- [4] W. Wang, R. Li, Y. Chen, Y. Sun and Y. Jia: "Predicting Human Intentions in Human-Robot Hand-Over Tasks Through Multimodal Learning," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 3, pp. 2339-2353, July 2022, DOI: 10.1109/TASE.2021.3074873.
- [5] I. Jacoby, J. Parron and W. Wang, "Understanding Dynamic Human Intentions to Enhance Collaboration Performance for Human-Robot Partnerships," 2023 IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, MA, USA, 2023, pp. 1-6, DOI: 10.1109/URTC60662.2023.10535035.
- [6] N. K. Gajjar, K. Rekik, A. Kanso, R. Müller: "Human intention and workspace recognition for collaborative assembly," *IFAC-PapersOnLine*, Volume 55, Issue 10, 2022.
- [7] N. Robinson, B. Tidd, D. Campbell, D. Kulić, and P. Corke: "Robotic vision for human-robot interaction and collaboration: A survey and systematic review," *J. Hum.-Robot Interact.*, 12(1), feb 2023.
- [8] E. V. Mascaro, D. Sliwowski, D. Lee: "HOI4ABOT: Human-Object Interaction Anticipation for Human Intention Reading Collaborative roBOTs," *Proceedings in Conference on Robot Learning*, 2023.
- [9] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, H. Austin and G. D. Hager: "Temporal convolutional networks for action segmentation and detection," *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [10] S. Bai, J. Z. Kolter, V. Koltun: "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," 2018, <https://arxiv.org/abs/1803.01271>.
- [11] K. Lyu, H. Chen, Z. Liu, B. Zhang and R. Wang: "3D Human Motion Prediction: A Survey," *Neurocomputing*, 489, pp.345-365, 2022.
- [12] T. Sawahata, A. Moro, S. Pathak and K. Umeda, "Instance Segmentation-Based Markerless Tracking of Fencing Sword Tips," 2024 IEEE/SICE International Symposium on System Integration (SII), pp. 472-477, 2024.