

歩行中の仮想運動中心の生成

Generating Virtual Motion Center during Walking

佐々木智也^{1,2)}, 鮎澤光³⁾, 吉田英一^{1,4)}

Tomoya SASAKI, Ko AYUSAWA, and Eiichi YOSHIDA

1) 東京理科大学 先進工学部 (〒125-8585 東京都葛飾区新宿 6-3-1)

2) 東京大学 先端科学技術研究センター

3) 産業技術総合研究所 人間拡張研究センター

4) 産業技術総合研究所 AL ラボ

概要: 人間とロボットの協調作業において、インタラクションのための運動中心の設定は重要である。単純に骨格情報や重心を中心として使用すると、ロボット制御時の振動や人間の身体との干渉が生じる。そこで本研究では、人間の身体的特性とロボットの動作制約を考慮した座標として、仮想運動中心を提案する。最適化による軌道計算と深層学習を用いて、歩行中の全身動作から将来の仮想運動中心を予測する手法を開発する。

キーワード: 動作予測、ヒューマンロボットインタラクション、深層学習

1. 緒言

人間とロボットの協調作業において運動の中心をどのように設定するのかは、円滑なインタラクションを実現する上で重要である。適切な座標や特徴点を設定できなければ、ロボットが期待通りに動かないだけでなく、人間とロボットの干渉や衝突の原因になる。しかしながら、協調作業のために人間の身体的位置や姿勢を直接参照するだけでは、ロボットの制御に不具合が生じることがある。

センサで取得した人間の骨格情報や重心をそのままロボットの制御に用いると、ロボットの動作制約を無視した人間の身体動作に影響を受けて制御の精度が低下するリスクがある。例えば、人間の位置に追従して移動する人追従ロボットの研究に、Human Coincident Robot (HCR) [1] がある。この人間と座標を共有して協調する移動ロボットの研究では、腰の位置に取り付けたマーカの座標に追従して移動制御を行うが、歩行中の身体の周期的な揺れがロボットの動作軌道に直接反映されてしまう。身体動作が由来の振動は、ロボットの制御性を低下させるだけでなく、人間とロボットが協調する際にロボットの操作性を低下させ、作業時の負荷を高める原因にもなる。

本研究では、人間の身体動作の特徴を表現し、かつロボットの制御において扱いやすい座標となるような、仮想運動中心 (Virtual Motion Center: VMC) を提案する (図 1)。仮想運動中心の未来の軌道を予測して、それに基づきロボットの移動制御を行うことで振動が抑えられ、人間にとっても扱いやすいロボットの動作が期待できる。

本稿では、歩行中の動作を対象に仮想運動中心の時間軌道を生成することを目的とし、歩行モーションデータから

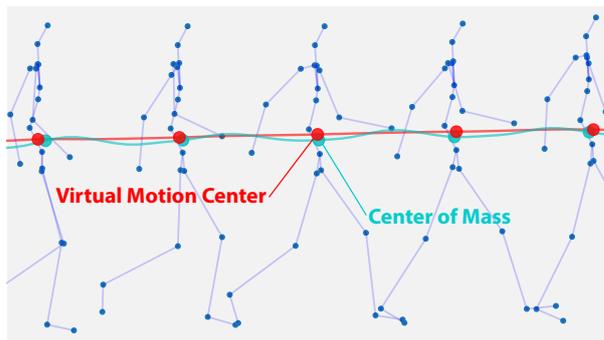


図 1: 歩行中の仮想運動中心と重心

最適化した仮想運動中心の軌道を事前に設計する。モーションデータと仮想運動中心の軌道データを用いた深層学習モデルを構築し、骨格情報から仮想運動中心の未来の軌道を予測する。構築したモデルを用いた未知の歩行モーションデータに対するシミュレーションにおいて、0.5s 先の仮想運動中心が最適化された軌道に対して平均二乗誤差の平方根 (RMSE) 0.051 m で生成できることを確認した。

2. 関連研究

2.1 歩行動作と重心

歩行動作は人間の基本的な全身動作のひとつであり、医療・臨床における歩行分析やヒューマノイドロボットの動作計画など、様々な分野で研究されている。歩行分析において力学的に根拠のある特徴量として用いられるものに重心 (Center of Mass: COM) がある。Iida と Yamamuro による運動学的解析 [2] によれば、歩行時の重心は上下方向に平均 3.2 ± 0.8 cm、左右方向に平均 3.5 ± 0.9 cm 移動する。こ

の移動は周期的に起こり、正面から見ると 8 の字を横にしたような軌道になる。一般に、人間の重心を直接計測することは難しい。そこで、予め身体部位ごとに質量を計測し、数学モデルを用いて推定する方法が提案されている [3]。稲垣ら [4] は数学モデルによる関節推定法を用いて、Kinect センサで取得した骨格情報から重心を算出している。

本研究では、重心を参照点としつつも周期的な振動を含まない座標を考えることで、身体動作の特徴の表現とロボット制御時の扱いやすさの両立を目指す。

2.2 全身運動の動作予測

歩行中の関節の位置や速度から将来の動作を予測する試みが行われている。古川 [5] は、3 層全結合型ニューラルネットワークを用いて、1 フレーム分の関節速度の入力に対して、1 歩先に相当する 0.5 s 後の頭部速度を予測できることを示した。Kurai ら [6] は 4 層の中間層をもつ 6 層全結合型ニューラルネットワークを用いて、3 フレーム分の関節位置を入力に 0.5 s 後の重心と各関節位置を予測した。また、この予測器を追従ロボットに用いたもの [7] や進行方向の提示に応用した研究 [8] が行われている。

コンピュータグラフィクス分野では、全身運動の動作予測のために大規模なモーションデータセットを学習に用いた CNN [9] や RNN [10] の研究がある。これらは歩行動作だけでなく、食事や喫煙動作といった複数の全身動作を予測できる。ただし、コンピュータグラフィクス用途では実時間での動作を要求していないため、処理時間の検討は十分に行われていない。

本研究では、仮想運動中心の座標点を考え、現在のフレームから将来の連続的な軌道を予測するモデルを検討する。時間軌道の予測を導入することでロボットの制御が行いやすくなる利点が考えられる。

3. 深層学習を用いた仮想運動中心の生成

3.1 仮想運動中心の軌道最適化

本研究における仮想運動中心は、歩行中の各身体部位や重心の動きと比較して振動が小さく、移動ロボットの制御に好ましい軌道を通る座標とする。そこで、仮想運動中心の軌道を B スプライン曲線として以下のように近似する。

$$\mathbf{x}_{\text{vmc},t} = \mathbf{B}(t)\boldsymbol{\alpha} \quad (1)$$

ただし、 $\mathbf{x}_{\text{vmc},t}$ は時刻 t における仮想運動中心、 $\boldsymbol{\alpha}$ は B スプライン曲線に対する係数である。 $\boldsymbol{\alpha}$ を以下の目的関数の最適化問題として計算する。

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \sum_{t=1}^T \{ & w_1 \| \mathbf{e}_t \|^2 + w_2 \| \dot{\mathbf{x}}_{\text{vmc},t} \|^2 + w_3 \| \ddot{\mathbf{x}}_{\text{vmc},t} \|^2 \} \\ \text{subject to } & -d_{\text{margin}} \leq \mathbf{e}_t \leq d_{\text{margin}} \end{aligned} \quad (2)$$

ここで、 $\mathbf{e}_t = \mathbf{x}_{\text{com},t} - \mathbf{x}_{\text{vmc},t}$ であり、 $\mathbf{x}_{\text{com},t}$ は時刻 t における重心である。第一項は重心と仮想運動中心の位置誤差、第二項は仮想運動中心の速度、第三項は仮想運動中心の加速度をそれぞれ制約条件下で最小化することを意味する。

w_1, w_2, w_3 は重み係数であり、今回はそれぞれ 0.005, 0.5, 0.5 とした。位置誤差が d_{margin} の範囲に収まることを制約条件とし、今回は $d_{\text{margin}} = 0.01$ とした。最適化の計算には MATLAB を用いた。

図 2 に x 軸に平行に直進する歩行モーションデータにおける重心 (COM) と仮想運動中心 (VMC) のプロットを示す。上段は x-y 平面および x-z 平面における位置の変位、下段は x 軸および z 軸の速度の時間変化を表す。なお、使用した歩行モーションデータおよび重心の導出に用いたパラメータは 3.2 に後述する。

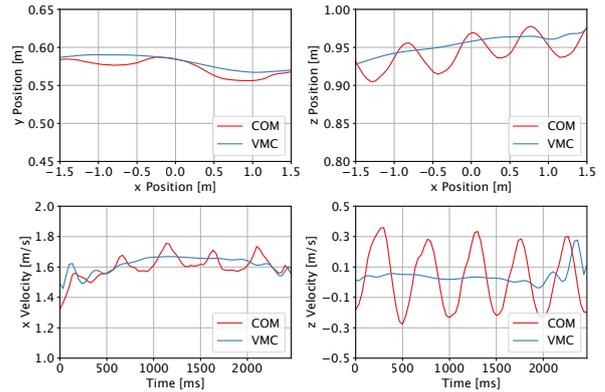


図 2: 重心から最適化した仮想運動中心の位置と速度

図 2 上段に示すとおり、歩行中の重心に特徴的な水平および高さ方向の振動が抑制されている。また、図 2 下段では、移動開始と停止を除く歩行動作中の x 軸および z 軸の速度が等速になるような運動であることが確認できる。これは、例えば x-y 平面上を曲がるような軌道を描く場合、x-y 平面上での速度が一定になる。

3.2 データセット

学習用の歩行モーションデータに 3 次元モーションデータベース AMASS [11] を使用する。AMASS は SMPL [12] 形式で標準化されたモーションデータベースであり、歩行動作以外にも座位や立位での作業、走る動作やダンスなどのモーションを含む。今回は、その中から典型的な歩行動作として直進、右折、左折、転回、後退を含む 31 モーションを学習用に選定した。このうち直進のみの歩行は 12 モーションである。

モーションデータの骨格情報には、各フレームで 22 点の関節座標がある。本研究では、阿江ら [3] の身体部分係数を用いて質量を仮定し、骨格情報から重心を計算した。重心の時系列データに対して、前述した仮想運動中心の軌道を計算し、これをモデル学習時の教示データとする。

モーションデータの骨格情報は、絶対座標で記録されているため、各関節座標を骨盤の座標との相対位置に変換した。また、骨盤の座標を床面に投影した点を root 座標とした。root 座標の x, y 要素はフレーム間の変位、z 要素は床面から骨盤までの高さとする。また、教示データの仮想運動中心座標も root 座標に対する相対位置に変換した。これにより、任意のフレームの root 座標から、現在の骨格情報、

過去フレームの骨格情報、および仮想運動中心座標がそれぞれ復元できる。

選定した元のモーションデータは 120 Hz で計測されている。本研究では、最終的に Kinect センサなどの RGB-D カメラを用いた実装を想定している。Kinect センサの更新周期は約 30 Hz である。そこで、元データを 1/4 にダウンサンプリングし、学習用にデータ数を 4 倍に拡張した。

3.3 モデルの学習

深層学習モデルは、Kurui ら [6] の提案手法を参考にした 5 層の全結合ニューラルネットワークを用いる。今回使用したモデルの概要を図 3 に示す。実装には Python 3.10, PyTorch 2.0.1 を用いた。

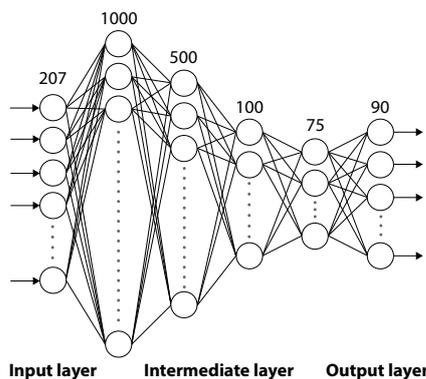


図 3: 使用したニューラルネットワークモデル

入力、現時点から過去 2 フレームを含めた合計 3 フレームの 22 点の 3 次元座標と 1 点の root 座標とし、出力は、現時点から将来 29 フレームを含めた合計 30 フレームの仮想運動中心 1 点の 3 次元座標とした。フレームの更新周期は 30 Hz として、過去 0.1 s の入力に対して将来 0.97 s 先までを予測するモデルである。入力層は 207 次元 (23 点 × 3 次元 × 3 フレーム)、出力層は 90 次元 (1 点 × 3 次元 × 30 フレーム) となる。中間層は順に 1000, 500, 100, 75 とした。活性化関数は、ReLU を採用し、各全結合層ではバッチ正規化を行なっている。

学習プロセスでは、最適化アルゴリズムに Adam、損失関数に平均二乗誤差 (MSE) を採用した。バッチサイズを 128、エポック数を 300、初期の学習率を 0.001 とした。使用したデータセットは合計 31456 フレームで、8 割を学習用、2 割を検証用とした。

3.4 結果と考察

モデルの学習に使用していない未知の歩行モーションデータを用いて、フレームごとに仮想運動中心の生成を行い、モデルの予測精度を検証する。使用したモーションデータは、直進 (walk straight)、直進後に 90 度左折 (turn left)、直進後に 45 度右折 (turn right)、その場旋回を含むランダムな歩行 (wander around) の 4 つである。モーションデータのフレーム数はそれぞれ、58, 182, 143, 1350 であった。

図 4 は、各フレームの入力に対する予測精度を示す。縦軸は生成された仮想運動中心と予め最適化された仮想運動

中心の座標の平均二乗誤差の平方根 (RMSE)、横軸は入力フレームから何フレーム先を予測した点かを示している。グラフは前述した 4 つのモーションと各モーションをフレーム数で重み付き平均した結果 (average) をプロットした。

まず、全体として、予測するフレーム先が未来になるほど、誤差が大きくなる傾向がみられる。例えば、平均した結果では、予測フレーム 15 (0.5 s 先) と予測フレーム 29 (0.97 s 先) の RMSE はそれぞれ 0.051, 0.160 m であった。モーションデータごとに精度をみると、直進と左折の誤差は、右折とランダムな歩行の誤差と比較して小さい傾向がみられる。直進と左折、右折では、予測フレーム 15 と 29 では約 2 倍、ランダムな歩行と平均では約 3 倍となっている。予測フレーム 15 においては右折モーションの誤差が最も大きくなっているが、これは学習データのバイアスが影響している可能性が考えられる。

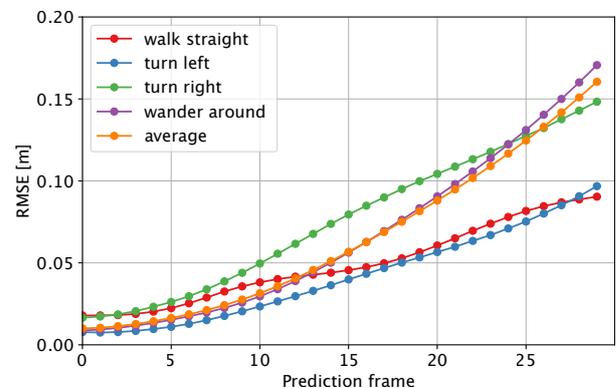


図 4: 予測フレームと推定精度

図 5 は、各モーションに対して 30 フレーム間隔のスナップショットである。橙色の点で示された軌道が予め最適化された仮想運動中心であり、緑色の点で示された軌道が各フレームの入力で生成された仮想運動中心である。赤色の点は 15 フレーム先の予測として出力された点である。青色の点はスナップショット時の骨格情報であり、可視化のために直線で繋いである。それぞれ、矢印の方向に進行している。

図 5(a) の直進では、生成された仮想運動中心が最適化された軌道に沿っている様子が確認できる。これは、図 5(b) や図 5(c) の後半の直進部分でも再現されている。一方で、図 5(b) の左折部分や図 5(d) の旋回部分では、最適化された軌道には沿っていない部分が見られる。また、図 5 にプロットされていないが、曲がる動作を含むフレームでは予測された軌道が不安定に振動し、特に 16 フレーム以降の誤差が大きい傾向がみられた。ただし、その場での旋回や曲率の違う動作においても、15 フレーム程度先までの予測であれば軌道に沿っていることが観察された。

今回の結果は、先行研究 [5, 6] の 0.5 s 先の動作予測の結果を支持するものになっており、0.5 s よりも将来の動作を予測することが難しいことを示唆している。また、モーションデータの種類によって推定精度にばらつきがある点や、曲がり動作において予測が不安定になる点は、学習データを

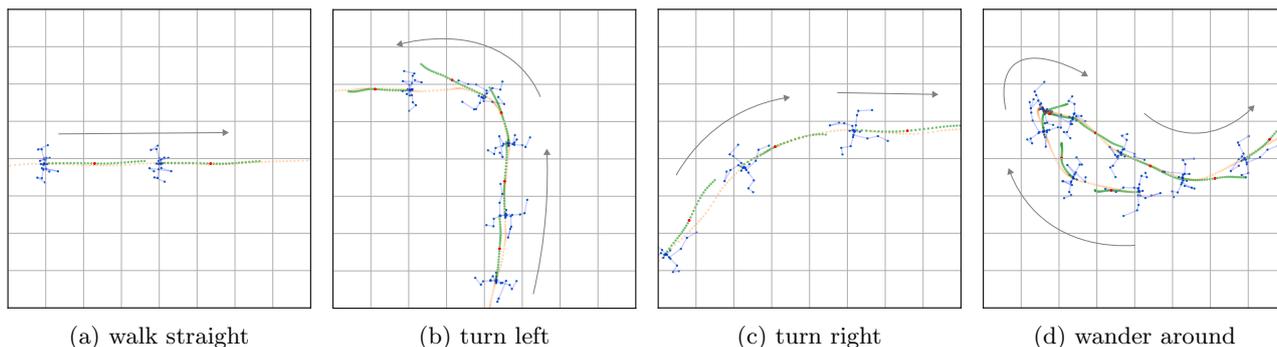


図 5: 入力された骨格情報と生成された仮想運動中心

増やすことで解決できる可能性がある。例えば、右左折や旋回のデータ数を増やすことで、これらの動作時の推定精度を高めることができると考えられる。

4. 結言

本研究では、人間の歩行動作における特徴とロボット制御性を考慮した滑らかな軌道となる仮想運動中心を提案した。歩行中の重心に対して位置誤差、速度、加速度を制約条件に基づいて仮想運動中心の軌道を最適化し、骨格情報から将来の軌道を推定するモデルを構築した。

本稿では、典型的な歩行動作を対象にしており、横歩きなどを含む全身動作には別のモデルや学習データの追加が必要になる。本研究では、人追従ロボットの制御を想定した仮想運動中心の軌道最適化を行い、 x - y 平面での推定について主に考察した。ドローンやヒューマノイドロボットの制御では、異なるモデルや最適化の制約条件が考えられる。

構築したモデルの評価はシミュレーションで行ったが、本モデルの処理速度は実時間で実行可能である。今後は、Kinect センサなどの RGB-D カメラを用いてシステムを実装し、実機のロボットで有効性を検証する。また、モデルの予測や生成を活用した人間拡張の応用も検討する。

謝辞 本研究は、科研費 基盤研究 (S) 22H05002 の助成を受けたものである。また、データセットの選定を補助してくれた星崇宏氏に感謝の意を示す。

参考文献

- [1] Takafumi Watanabe, Tomoya Sasaki, Zendai Kashino, and Masahiko Inami. Human coincident robot: A non-contact surrounding robot sharing the coordinate with a human inside. In *Proceedings of the Augmented Humans International Conference 2023*, March 2023.
- [2] Hirokazu Iida and Takao Yamamuro. Kinetic analysis of the center of gravity of the human body in normal and pathological gaits. *J. Biomech.*, Vol. 20, No. 10, pp. 987–995, January 1987.
- [3] 阿江通良, 湯海鵬, 横井孝志. 日本人アスリートの身体部分慣性特性の推定. *バイオメカニズム*, Vol. 11, pp. 23–33, 1992.
- [4] 稲垣潤, 春名弘一, 昆恵介, 本郷節之, 岡崎哲夫. Kinect センサーを用いた身体合成重心の検出と精度. *臨床歩行分析研究会誌 / 臨床歩行分析研究会 編*, Vol. 4, No. 1, pp. 21–27, 2017.
- [5] 古川正紘. 実時間歩行誘導のための多層パーセプトロンを用いた 1 歩先の歩行運動予測. 第 22 回日本バーチャルリアリティ学会大会予稿集, 2017.
- [6] Takafumi Kurai, Yutaro Shioi, Yasutoshi Makino, and Hiroyuki Shinoda. Temporal conditions suitable for predicting human motion in walking. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, October 2019.
- [7] Ansheng Wang, Yasutoshi Makino, and Hiroyuki Shinoda. Machine learning-based Human-Following system: Following the predicted position of a walking human. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, May 2021.
- [8] 渡部慎太郎, 牧野泰才, 篠田裕之. 人間の歩行動作予測に貢献する重要な身体部位の解明. *ロボティクス・メカトロニクス講演会講演概要集*, 2023.
- [9] Chen Li, Zhen Zhang, Wee Sun Lee, and Gim Hee Lee. Convolutional sequence to sequence model for human dynamics. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, June 2018.
- [10] Julieta Martinez, Michael J Black, and Javier Romero. On human motion prediction using recurrent neural networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, July 2017.
- [11] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. AMASS: Archive of motion capture as surface shapes. In *International Conference on Computer Vision*, October 2019.
- [12] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. SMPL: a skinned multi-person linear model. *ACM Trans. Graph.*, Vol. 34, No. 6, November 2015.