



静止状態からの動作における予測可能性の検討及び 予備動作の推定

Investigating the Predictability and Estimating the Preliminary Movement of
Movements from a Quiescent State

許 超舜¹⁾, 藤原 正浩²⁾, 牧野 泰才^{2) 3)}, 篠田 裕之²⁾

Chaoshun XU, Masahiro FUJIWARA, Yasutoshi MAKINO, and Hiroyuki SHINODA

1) 東京大学 情報理工学系研究科 (〒 113-8656 東京都文京区本郷 7-3-1, xu@hapis.k.u-tokyo.ac.jp)
2) 東京大学大学院 新領域創成科学研究科 (〒 277-8561 千葉県柏市柏の葉 5-1 - 5) 3) JST さきがけ

概要: 近年, 機械学習や NN を用いて人間の動作を分類・予測する研究が注目を集め, 特に骨格座標から未来の動作を予測する研究が発展段階にある. 人間の動作の研究において, 静止状態からの動作に際して予備動作がどの程度前から観測されるのかということは定量的に調査されていなかった. 本稿では 100fps で 3 次元計測した骨格座標を用いて, 各動作開始前の予備動作の有無を検証し, 左右分類に必要な予備動作の量を確認した. 結果として, 腕を振る動作では予備動作は見られなかったが, 並進ジャンプ動作では動作開始 0.4 秒前からジャンプの方向に関係する身体動作が見られた.

キーワード: 動作予測, ニューラルネットワーク, 計測・認識

1. 背景

身体動作を計測できる VR 機器の普及と共に, 身体動作によって VR 空間内や遠隔地のアバターを操作する機会は今後増えていくと思われる. その際の課題の一つとして, レンダリングや通信による遅延の存在がある. そこで事前に人間の動作を推定し, レンダリングを予め行っておくことで, 遅延を無くすことができると考えられる.

人間が何かの動作を行う際, 動作の種類に応じて予備動作が発生する. 予備動作は各部位の質量や筋肉量, 動作の向きや速度など様々な条件によって左右されるが, 人間はこれらの予備動作を元に無意識の内に次の動作を予測し, 適切な行動をとることができる. これまでこの予測能力をコンピューターで再現しようとする研究が行われてきた [1]-[4]. Martinez ら [1] や Chiu ら [2] の研究では, Human3.6M [5] のデータに対し, 回帰型ニューラルネットワーク (RNN) を用いて未来の動作を予測した. また, Horiuchi ら [3] や Kurai ら [4] は Kinect で測定したジャンプのデータを対象に, 順伝搬型ニューラルネットワークを用いて動作を予測した. このように人間の動作を予測する研究は行われてきたが, ある動作を予測する際に予備動作がどの程度前から観測されるのか, ということに関して定量的に調査した研究は少ない.

人間の動作における予測限界を議論するためには, 完全に静止した状態からの動作推定をターゲットとするのが望ましい. というのも, 人間の動作には直前までの動作によるシーケンスとしてのダイナミクスが含まれている. これ

までに取り扱ってきた動作においては, 単一の動作を行おうとした際の最小の予測限界を, 動作シーケンスのダイナミクスから切り離して予測できるかを検証するのが本研究の目指すところである.

静止状態から単一の動作を行う際に予備動作として必要なものは, 動き始めてからの動作情報であり, それ以前の情報は必要ない. そのため, 本研究では動作の直前までのすべての動作の情報が含まれる回帰型ニューラルネットワークではなく, 入力する時系列情報を陽に指定できる Horiuchi ら [3] や Kurai ら [4] の順伝搬型ニューラルネットワークを利用した. また, 時間軸においてより精度の高い議論を行うため, Human3.6M や Kinect よりもフレームレートの高い OptiTrack を使用し, 100fps で人間の静止状態からの動作を測定した.

2. 計測する動作の種類と詳細

本システムでは, OptiTrack によって計測された 21 箇所の骨格情報を計測し利用した (図 1). 人間の動作において比較的大きく動くかつインタラクションに重要な動作として, 腕の動作, 脚の動作, 全身の動作を計測した.

具体的には,

- 利き腕を掲げ, 左右斜め 45 度の空中位置にある仮想のボタンをタッチする動作
- 脚を伸ばしたまま, 右足あるいは左足を約 60 度蹴り上げる動作

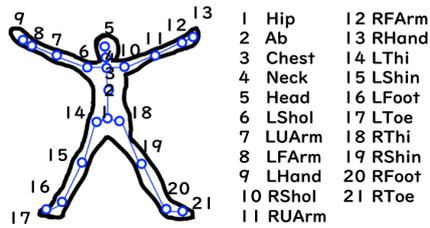


図 1: OptiTrack で測定する身体 21 点の図

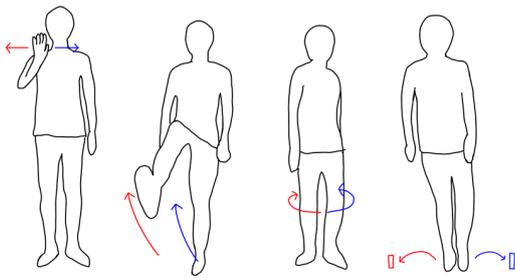


図 2: 各種動作の図

- その場で右あるいは左方向に 180 度回転するジャンプ
- 左右いずれかの方向に 75cm ジャンプ

の 4 種類である (図 2)。それぞれの動作はなるべく自然な動きで、かつなるべく早く動くように指示を出した。

被験者にはモニターを通じて左右方向の指示を出し、対応した方向に動作を行ってもらった。左右方向の指示はランダムに提示され、一人の被験者ごとに各動作 120 回ずつを 100fps で収集した。被験者数は 20 代の右利き男性 7 名である。実験開始時の立ち位置及びジャンプの目標位置は足元に目印を示した。

3. 実験

3.1 動作の前の予備動作の発生瞬間

ある意図した動作 (これを本動作とする) を行うにあたって、バランスを取るためなどの理由で無意識の内に行われる動作がある。これらの動作の内、本動作よりも前に発生するものを予備動作とする。予備動作を観測することで本動作が行われるよりも早く本動作の動きを予測することができる。本研究ではこの予備動作の出現箇所を特定することを目標としている。そのため、本動作を開始する瞬間を定義する必要がある。また、静止している状態は予備動作ではないので、各身体部位が動き出す瞬間を計算することで、予備動作の出現領域を絞り込むことができる。

本動作開始点は、対象の部位を見ることで左右どちらに動こうとしているのか分かる点が適切であると考えた。そこで、ボタンタッチ動作とキック動作では手や足が 3 フレーム連続で指示方向に動いておりかつ累計で 3mm 以上動いていた点、ジャンプ動作 2 種では重心が最も低くなった点を本動作開始点とした。一方で、身体の各部位が動き出しているかどうかは、それぞれの部位において静止時の速度の平均と標準偏差を求め、平均の 3 倍と標準偏差の 3 倍を

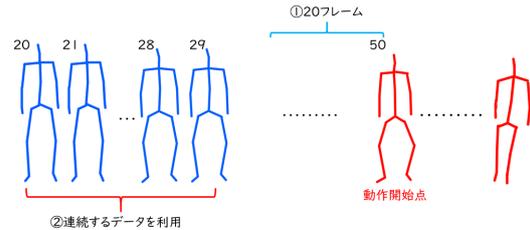


図 3: 骨格情報の入力

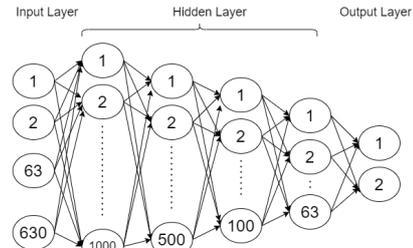


図 4: Horiuchi ら [3]Fig.4 が用いたニューラルネットワーク

足したものを速度の閾値とし、3 フレーム連続で速度がこの閾値を超えた点を動作開始点とした。

3.2 入力フレーム数と何フレーム前のデータを入れているかで左右分類

動作開始点を基準として、

1. 何フレーム前までの情報か (frames before movement, FBM とする)
2. 使用フレーム数 (used frames, UF とする)

の二つを変化させ、ニューラルネットワークへの入力とする。例えば、本動作開始点が 50 フレーム目、FBM が 20、UF が 10 の場合、20 から 29 フレーム目のデータが入力となる (図 3 参照)。

ニューラルネットワークについては、先行研究 [3] と同様に、全結合層のみで構成される 5 層のネットワークを用いる。

入力層のユニット数は $63 \times (\text{使用フレーム数})$ 、中間層は順に 1000, 350, 100, 63, 出力層は 2 とする。中間層の各ユニットが持つ活性化関数は正規化線形関数とし、出力層には恒等写像を用いる。また、誤差関数には Softmax Cross Entropy, オプティマイザには Adam を用いる。入力データには OptiTrack で計測した 21 点の 3 次元座標を、複数フレームまとめて入力とする。

7 人のデータの内、4 人を訓練データ、2 人をバリデーション、1 人をテストデータとして使用した。学習はミニバッチ学習で行い、ミニバッチ数は 100、エポック数は 1000 とした。

4. 結果

4.1 動作の前の予備動作の発生瞬間

以降、図の中ではボタンタッチ動作を BT_Rhand, キック動作を Kick.f, 回転ジャンプ動作を Jump_Rot.180, 並進ジャンプ動作を Jump_LR と表す。

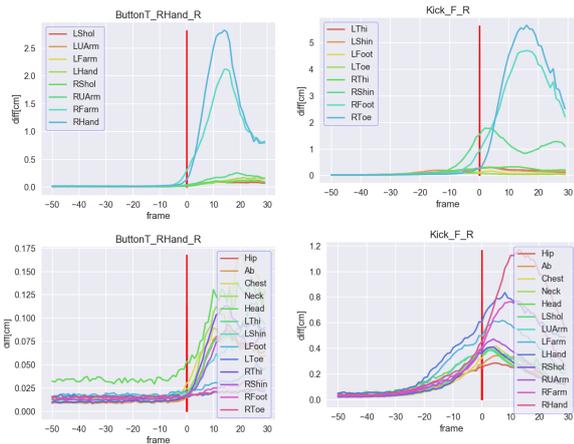


図 5: ボタンタッチ動作

図 6: キック動作

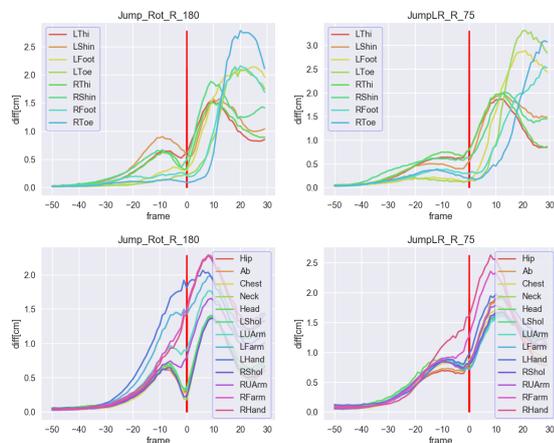


図 7: 回転ジャンプ動作

図 8: 並進ジャンプ動作

本動作開始点と各身体部位の動作開始点をまとめたのが表 1 である。

また、本動作開始点を基準に、各フレームにおける身体部位の動作速度の平均を取ったのが図 5～図 8 である。動作の大きい身体部位を上側に、そうでないものは下側に置いている。これらの図は右側へ動かす場合のものであるが、左側に動かす動作もほぼ同様であった。

これらを見ると、ボタンタッチ動作ではほとんど予備動作が発生しておらず、むしろ腕の動作開始後にバランスを取るために身体部位が動いていることが分かる。また、キック動作では平均 10 フレーム、最大で 20～30 フレーム前からの動作が予備動作であることが分かる。ジャンプ動作 2 種

表 1: 本動作開始点と予備動作開始点

No.	本動作開始 [ms]	各身体部位の動作開始平均 [ms]
ボタンタッチ	379 ± 65	+82
キック	609 ± 115	-109
回転ジャンプ	660 ± 123	-200
並進ジャンプ	733 ± 72	-221

表 2: ボタンタッチ動作

UF \ FBM	-20	-10	0	10
5	0.99	0.86	0.53	0.54
10	0.99	0.89	0.54	0.53
15	0.98	0.80	0.53	0.53
20	0.99	0.67	0.54	0.55
25	0.97	0.79		
30	0.81	0.54		

表 3: キック動作

UF \ FBM	0	10	20	30	40
5	0.74	0.58	0.50	0.49	0.48
10	0.77	0.63	0.49	0.47	0.49
15	0.83	0.68	0.47	0.46	0.46
20	0.86	0.69	0.49	0.45	0.45
25	0.83	0.66			
30	0.85	0.65			
35	0.86	0.61			
40	0.82	0.65			

類については平均 20 フレーム、最大で 30～40 フレーム前からの動作が予備動作であることが分かる。

4.2 入力フレーム数と何フレーム前のデータを入れているかで左右分類

予備動作の可能性のある部分を求めたので、これらのフレームを基準に入力フレームの位置を決めた。キック動作、ジャンプ動作 2 種では本動作開始点の 0～40 フレーム前まで、使用フレーム数を 5～40 (0.05 秒～0.40 秒) のデータを入力とした。ボタンタッチ動作については予備動作が存在する可能性が低いため、本動作開始後 20 フレーム～本動作開始前 10 フレーム前まで、使用フレーム数を 5～30 (0.05 秒～0.30 秒) のデータを入力とした。これらのデータを入力としたとき、テストデータ 120 個 (学習に利用していない 1 名の被験者のデータ) に対しての左右分類の正答率をまとめたのが表 2～表 5 である。

ボタンタッチ動作はやはり手が動く前に分類するのは困難であり、予備動作がほとんど存在しないと考えられる。キック動作は入力に 20 フレーム以前の情報を入れてもほとんど分類できないことから、予備動作は本動作開始点前 20 フ

表 4: 回転ジャンプ動作

UF \ FBM	0	10	20	30	40
5	0.75	0.74	0.56	0.51	0.50
10	0.74	0.75	0.55	0.47	0.50
15	0.75	0.74	0.56	0.49	0.53
20	0.75	0.70	0.55	0.48	0.53
25	0.74	0.72	0.55		
30	0.74	0.68	0.49		
35	0.76	0.74	0.51		
40	0.75	0.70	0.55		

表 5: 並進ジャンプ動作

UF \ FBM	0	10	20	30	40
5	0.99	0.97	0.95	0.73	0.51
10	0.99	0.98	0.86	0.72	0.52
15	0.99	0.99	0.95	0.68	0.52
20	0.99	0.97	0.94	0.68	0.50
25	0.99	0.99	0.89	0.73	
30	0.99	0.97	0.88	0.63	
35	0.99	0.98	0.95	0.70	
40	0.98	0.99	0.93	0.66	

表 6: 入力から右腕を除いた場合のボタンタッチ動作

UF \ FBM	-20	-10	0	10	40
5	0.61	0.55	0.52	0.53	0.51
10	0.59	0.56	0.53	0.53	0.52
15	0.54	0.54	0.53	0.53	0.52
20	0.54	0.54	0.54	0.54	0.50

フレームくらいまでに収まっており、また、入力の長さは 15～20 フレームくらいが適していることが分かる。回転ジャンプ動作も予備動作が本動作開始点前 20 フレームくらいまでに収まっていると考えられる。入力の長さは 5 フレームでも一定の精度で分類できるが、全体的に精度が低く、また本動作直前のフレームを入れても左右分類が難しいことが分かる。並進ジャンプ動作は最も予備動作が大きく、本動作開始点前 40 フレームくらいから予備動作が存在すると考えられる。また入力するフレーム数は 5 フレームでも充分である。

5. 考察

予備動作の内、本動作で重要でない身体部位の影響を調べるため、入力骨格座標から重要と思われる部位を除いた場合についても同様に左右分類を行った。ボタンタッチ動作では右腕の座標を除いた 17×3 点を、他の動作では両足の動作を除いた 13×3 点を入力とした。結果として、キック動作、回転ジャンプ動作、並進ジャンプ動作ではあまり変化は見られなかったが、ボタンタッチ動作の精度は大きく変化した (表 6)。このことから、ボタンタッチ動作は動作開始後の右腕の動きをみて左右分類していると考えられる。

また、ボタンタッチ動作では予備動作が表れず、他の動作では表れた理由について、

- 腕を動かすためのモーメントが比較的小さく、他の部位の力を必要としないから
- キック動作、回転ジャンプ動作、並進ジャンプ動作ではバランスを取る必要があるから

などが考えられる。一つ目の要因については、腕に重りをつける・棒を持つなどした状態で同様の計測を行い、左右分類できるようになったら示すことができると考えられる。二つ目の要因については、バランスが崩れない状態、例え

ば座った状態でのキック動作を行った際に左右分類の精度が下がれば、寄与度が大きいと考えられる。

6. 結論

本研究では静止状態からの単一の動作の予備動作を求めするために、腕の動作、脚の動作、全身を使った動作などを 100fps で計測し、本動作開始点と各身体部位の動作開始点を求めた。また、本動作前の骨格座標について、入力するフレーム数や入力タイミングを変えてニューラルネットワークの入力とし、左右分類を行った。これにより、本動作を行う際に必要な予備動作が現れている箇所を求めた。ボタンタッチ動作は予備動作が見られなかったが、キック動作と 180 度回転ジャンプ動作は本動作開始前 0.2 秒、75cm の並進ジャンプ動作では本動作開始前 0.4 秒までが左右分類に影響を与えており、これらの時間において予備動作が現れている可能性が高い。

今後は身体部位の質量や動力の大きさに注目して、人間の自然な動きを求めるこれは、将来的には遠隔コミュニケーションにおけるジェスチャーの遅延補償やスポーツ応用、全身を利用したエンターテインメントなどに使用できると考えられる。

謝辞 本論文は JST さきがけ 17939983 の支援を受けて行われた。

参考文献

- [1] Martinez, J., Black, M. J., and Romero, J. : “ On human motion prediction using recurrent neural networks. ”, *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2891–2900, 2017.
- [2] Chiu, H. K., Adeli, E., Wang, B., Huang, D. A., and Niebles, J. C. : “ Action-agnostic human pose forecasting. ”, *In 2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1423–1432, 2019.
- [3] Horiuchi, Y., Makino, Y., and Shinoda, H. : “ Computational foresight: Forecasting human body motion in real-time for reducing delays in interactive system ”, *In Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, pp. 312–317, 2017.
- [4] 倉井 敬史, 牧野 泰才, 篠田 裕之, : “ニューラルネットワークを用いた人動作予測モデルにおける最適入力時間長の検討”, 第 20 回計測自動制御学会システムインテグレーション部門講演会論文集 (SI2019), pp. 494–498, サンポート高松, 香川, Dec. 12-14, 2019.
- [5] Ionescu, C., Papava, D., Olaru, V., and Sminchisescu, C. : “ Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. ”, *IEEE transactions on pattern analysis and machine intelligence*, pp. 1325–1339, 2013.