第 26 回日本バーチャルリアリティ学会大会論文集（2021 年 9 月）

急な動作変更に頑健な予測器作成に向けた予測挙動評価と動作予測アルゴリズムの基礎的検討

板井俊樹 [1)]，砥出悠太郎 [1)]，牧野泰才 [1)2)]，篠田裕之 [1)]
Toshiki ITAI, Yutaro TOIDE, Yasutoshi MAKINO and Hiroyuki SHINODA

1) 東京大学 新領域創成科学研究科 (〒 277-8561 千葉県 柏の葉 5-1-5, t.itai@hapis.k.u-tokyo.ac.jp)
2) JST さきがけ

概要: VR 空間内でアバタを表示する際に生じる描画遅れを解消するために, 動作を予測し表示する手法が提案されている. 既存の予測手法では，フェイントや急な切り返し行動等を行った場合，予測軌道が不連続に変化し, 違和感が生じる．この問題解消のため，１）予測誤差の極力小さい予測アルゴリズムの確立，２）多様なフェイント動作を含む学習データの計測，の 2 点が必要と考えた．本稿では１）に着目し，現在の予測方法における急な動作変更時の挙動を詳細に確認した．また既存の動作予測モデルを改良し，予測精度を向上できるかを検証した.

キーワード： 動作予測，ニューラルネットワーク

1. 序論

VR 空間上でのリアルタイムな身体動作コミュニケーションを行うためには，レンダリングに伴う計算負荷や通信による遅延に影響されずにリアルタイムでアバタが表示されるのが望ましい．これら遅延を解消するための一つの手段として，身体動作を事前に予測し，レンダリング計算や通信による遅延を解消する方法に着目する.

の２つのポイントに注目した．これまでに骨格情報を元に，過去 10 フレームの情報から，0.5 秒程度先の未来の位置と姿勢を予測するという方法が提案されている [1][2]．これら先行研究では，歩行やジャンプのような全身運動において，重心位置の平均二乗誤差が数 cm 程度で予測できるということが示されている.

一方で，平均的には数 cm 程度の誤差であっても，瞬間的には予測誤差が 30cm 程度と大きくなることが確認されていた．これは特に，急な動き出しや動きの変化を伴う動作時に生じる．既存のアルゴリズムでは，このような急な動きの変化の予測においてに，静止状態から特定の姿勢へと遷移する瞬間に，予測される位置が大きく変動していた．このため，予測された身体動作を連続的に動画として見ると，動き出しの瞬間に予測値が不連続に遷移するような映像となる．このような大きな遷移を伴う予測を，アバタのレンダリングのための情報として利用すると，アバタが不自然に瞬間移動することになる.

本研究では，予測情報を利用したより自然なアバタ動作生成のために，フェイントや切り返しのような急な動作変更に強い予測器の作成を目的とする．これを達成するために，１）予測誤差の極力小さい予測アルゴリズムの確立，２）多様なフェイント動作を含む学習データの計測，の 2 点が必要であると考えた．本稿ではまず最初のステップとして，１）に着目し，急な動作変更時に現在の予測方法ではどのような挙動を示すかを詳細に確認した．また既存の動作予測モデルを改良し，予測精度を向上できるかを検証した.

実験の結果，次の動作に対する予備動作が数フレーム入力されるまでは，予測が次の動作に対応できないことがわかった．また，既存の動作予測モデルの正解データとの比較フレーム数を変更しても，予測精度向上には寄与しないことが確認できた.

2. 急な動作変更における予測

ある一連の動作をしている状態から，別の動作に瞬時に切り替える動作を，本研究で「急な動作変更」と呼ぶ．解析のしやすさから，本稿では特に，最初静止している状態からの急な動作変更に着目し，先行研究でも用いられていた大きく跳躍するジャンプ動作を対象にした．急な動作変更のタイミングに対して予測結果がどのような挙動を示すかを調査した.

2.1 入力に用いるデータの詳細

ジャンプ動作の骨格データは，Kurai ら [2] が Kinect V2 を用いて計測したものを使用した．計測方法は以下のように行われている.

1. 被験者は Kinect に正対し，正面にあるモニタを見る
2. モニタ上にジャンプの挙動についての指示を出す．具体的には,「方向 n に距離 S」の形式で，$n =$ 左，前，右の 3 方向，$S = 50, 100, 150$ cm の 3 つのジャンプ距離.
3. 被験者は指示が出た後できる限り早く指定された方向にジャンプする．実験開始位置およびジャンプの目

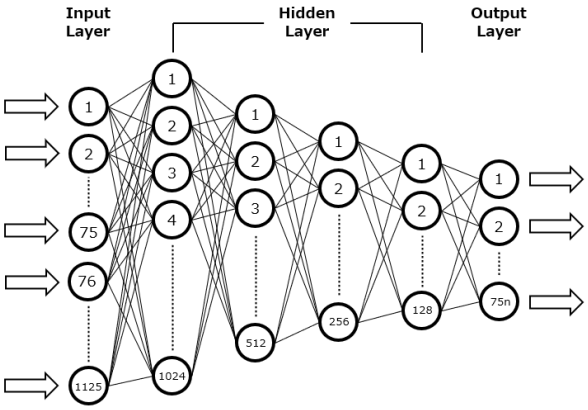

図 1: 学習に用いたニューラルネットワーク

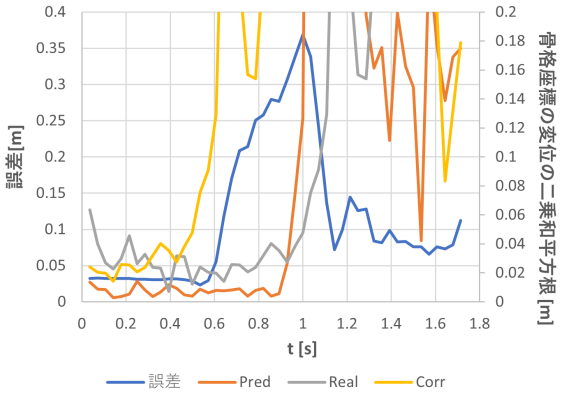

図 2: 0.5 秒後の予測と実測との誤差を各フレームごとに比較したグラフ

標位置は足元に目印がある．

4. 1〜3 を，各指示がそれぞれ同数になるように，ランダムな順序で提示し繰り返す．

20 代男性 5 人の被験者数について，一人の被験者ごとにジャンプ方向 (3 パターン) × ジャンプ距離 (3 パターン) × 繰り返し (10 セット) の計 90 回分，28 fps のジャンプデータが収集されている．

2.2 学習方法とネットワーク

既存の研究 [1][2] では, 順伝搬型ニューラルネットワークに過去 10 フレーム程度の骨格座標データを入力し，0.5 秒先の骨格データを出力として，身体動作予測を行っていた．本研究では，入力フレーム数を 15 フレーム（0.5 秒）とし，入力に用いた最後のフレームから 0.5 秒後の 25 点関節情報を出力とした. ネットワークは，図 1 の構造を利用した．モデルは全てのジャンプ距離/方向のデータを学習させた．出力された予測値と実際の観測値を比較し，誤差評価には，胸の座標の予測値と実測値の距離を用いた．

学習は 10 回行い，予測結果の平均で誤差を評価した．ニューラルネットワーク (NN) の誤差関数に平均二乗誤差 (MSE)，活性化関数に ReLU，オプティマイザーに Adam を用いた．バッチサイズは 20，エポック数は 100 とした．計測した跳躍データのうち 80 %を学習に用い，残りの 20 %をテストに用いた．

2.3 結果

図 2 は一連のジャンプ動作から得られた骨格データにおいて，0.5 秒後の予測と実測との誤差を各フレームごとに比較したグラフである．予測は人の予備動作を観測してそれに基づいて行われるため，人が全く動いていない状態では予測は不可能である．

結果の図を見ると，誤差（青線）が大きくなるのは，予測すべき 0.5 秒後の正解値（Corr：黄線）が変化していくフレームで，かつ現在の身体座標（Real：灰線）が変化していない 0.53 秒間である．一旦身体動作が開始されると，2 3 フレーム後には予測が始まり（Pred:橙線），誤差が 10cm 以下に低下し，正しい予測が行われることが観察された．ここでは代表的な一例を示しているが，ジャンプの方向等に依らず，同様の傾向が得られることを確認している．

2.4 考察

静止状態からスタートした場合，Real が動かない限り予測はできない．0.5 秒先を予測する場合，予測すべき黄線（Corr）が動き出すのは動作開始の 0.5 秒前であるから，実際の身体座標が動くまでの 0.5 秒間予測ができないのは原理上避けられない．つまり，Corr が動き出してから Real が動き出すまでの 0.5 秒は誤差が上がり続ける．これはこの 0.5 秒程度の間に静止状態からジャンプ動作に移行した変位に対応する．0.5 秒経過し，Real が動き出すと，橙線（Pred）が真値に近づき，誤差が急激に減る様子が確認できる．

アバタをこの予測動作に基づいて描画する場合，見た目の違和感を減らすには，瞬間的な動作の飛びを無くすことが重要となる．したがって，誤差の瞬間最大値を減らすことが望ましい．許容される最大誤差を人の知覚特性から決定し，それに応じて予測時間を決定するのが望ましい．

例えば，予測時間を 0.3 秒とした場合の結果を図 3 に示す．0.5 秒のときに比べ，予測開始のタイミングが早く訪れることで誤差が 10cm 程度以下に抑えられることが確認できる．

以上より，予測モデルによる誤差が予測時間に依存した最大誤差よりも常に小さい範囲で予測できるような場合には，

1) 予測時間を短くすると最大誤差は抑えられる．
2) どのような速度の動作を予測するかに応じて最大誤差が変化する．例えばゆっくりした動作の場合，0.5 秒間で移動する変位が小さいので，最大誤差は小さくなる．

ということが確かめられた．一方で，現在の予測器に比べ，より精度の高い予測器が実現された場合，誤差のグラフは，最大誤差に至るまでの推移は同じであるが，実測が少し動き出してからの誤差が急速に小さくなることが期待される．次節より，学習時の正解データの次元を増やすことで，精度が向上するかについて検討した結果を示す．

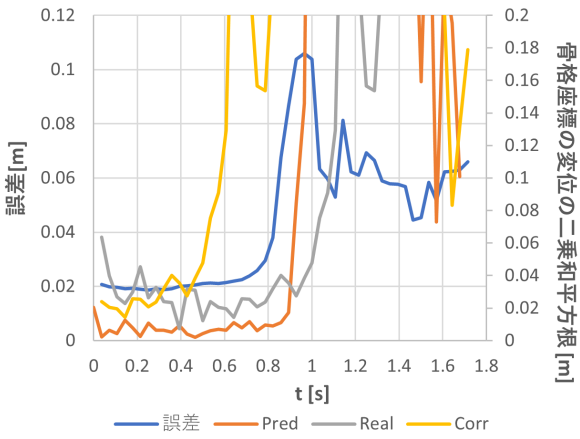

図 3: 0.3 秒後の予測と実測との誤差を各フレームごとに比較したグラフ

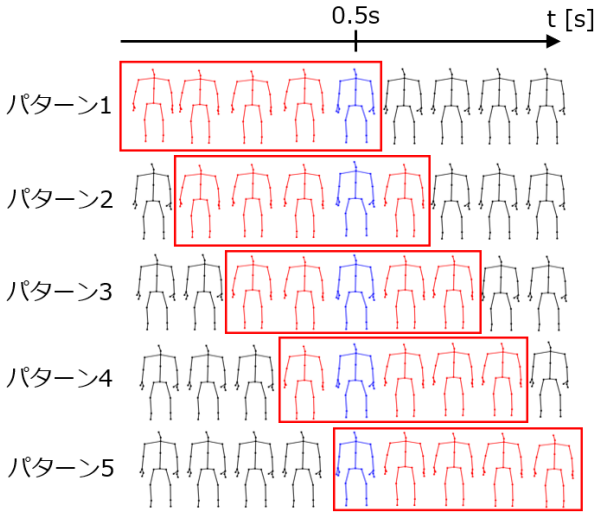

図 4: 正解データを複数選択する場合の選択パターン

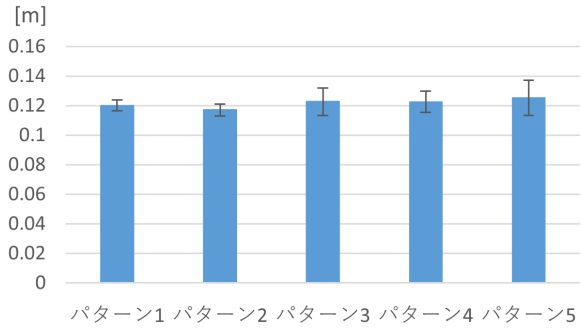

図 5: 各パターンの誤差平均

3. 予測精度向上への試み

これまで，Horiuchi ら [1] や Kurai ら [2] が用いていた予測モデルでは入力に利用するフレーム数を変えたときの出力結果についての議論がなされていた．すなわち，どれくらい過去にデータを遡ったときに予測精度がどう変化するかを議論していた．一方，出力については目的とする単一フレームのみを正解データとして利用していた．本稿では予測精度向上のため，正解データの次元を増やすことを考える．これまで単一フレームのデータのみを正解値として利用していたが，連続する複数フレームのデータを正解値として利用することで，複数フレームに渡る動作の滑らかさが保証され，予測精度が向上すると考えた．

3.1 複数フレームの正解データの利用方法

正解データを複数フレームにする場合，目的とする 0.5 秒先の骨格情報を含む複数フレームのとり方が，図 4 のように複数存在する．例えばパターン 1 は，予測したい 0.5 秒先の骨格が，正解シーケンスの一番最後に入っているのに対して，パターン 5 では正解シーケンスの一番最初のフレームに含まれるというとり方を意味する．この最適なパターンを調べるために正解フレーム数を 5 フレームで固定し，各パターンに対して得られる Pred と Corr との誤差がどのようになるかを検証した．ジャンプ動作の骨格データは前セクションで用いたものを使用した．NN は前セクションで用いた設定に対して，出力層を 5f 分の骨格データに変更したものを使用した．誤差評価は以下 2 つの観点

観点 1: Pred と Corr との誤差の全フレームでの平均値

観点 2: 動作開始 0.5 秒後に誤差が最大値をとってから，予測が開始され誤差が減衰するときの時定数

をもって検証した．学習時のランダムさにより，学習ごとに異なるパラメータが得られるため，評価は初期設定が同じ 10 個の異なる学習モデルに対してそれぞれ行った．

3.1.1 結果

図 5 はそれぞれ，観点 1 において評価した値を パターンごとにまとめたものである．観点 1 において，各パターンで比較した結果，それぞれの平均値の差はおおよそ 1cm 以下であった．人の体格が 1m のオーダーであるのに対して，パターンの違いによる誤差は $\frac{1}{100}$ のオーダーでしか変化しておらず，パターンの差による有意差は見られなかった．

図 6 は，観点 2 で評価するために，それぞれのパターンで Pred と Corr の誤差を各フレームごとに比較したグラフである．各パターンで誤差変化の傾向に大きな違いはなく，最大誤差となる 0.8 秒の瞬間から，予測誤差が下がるのに要する時間は，パターンに依らず 0.2 秒以内程度であり，同じ傾向を示していた．以上より，パターン間に差は見られなかったため，今後のセクションにおいてはパターン 1 のみ評価することとした．

3.2 正解フレーム数と精度の関係

前節において，正解フレーム数が一定（5 フレーム）の場合，予測したいフレームがその連続するフレームのどこにあっても精度が変化しないことを確認した．本節では次に，正解フレーム数を変更した際における，ジャンプの予測誤差について検討する．誤差評価は前セクションと同じ過程で行い，予測したいフレームは，パターン 1 を，正解フレームは 1 から 10 フレームの間で評価した．

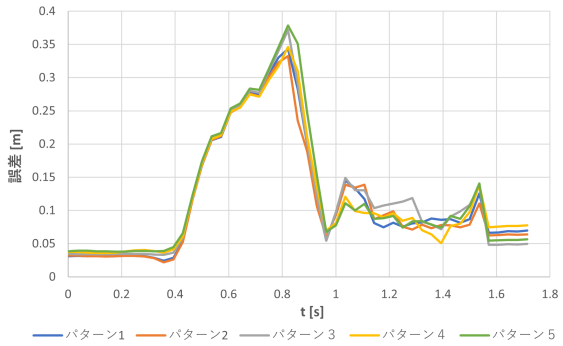

図 6: 各パターンの誤差の時間推移

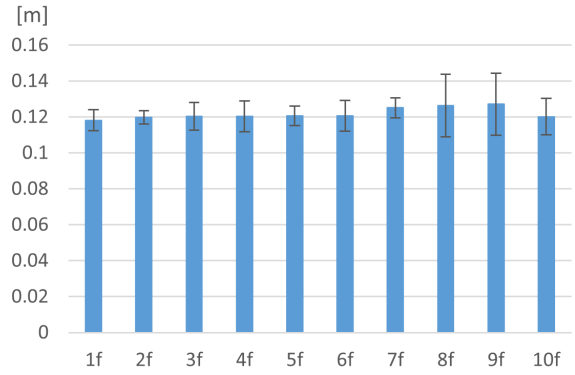

図 7: 正解フレーム条件ごとの, 胸の距離の最大値の平均値

3.2.1 結果

図 7 は，観点 1 に対する全学習モデルの平均値を正解フレーム数ごとにまとめたものである．また，図 8 は，観点 2 で評価するための，それぞれのパターンで Pred と Corr の誤差を各フレームごとに比較したグラフである．観点 1 において，各正解フレーム数で比較した結果，前セクションと同様に，それぞれの平均値の差はおおよそ 1cm 以下であり，最大のオーダーに対して $\frac{1}{100}$ のオーダーでしか変化していなかった．また，観点 2 において，各パターンで誤差が現象する時間は 0.2 秒以内であり，同じ傾向を示していた．

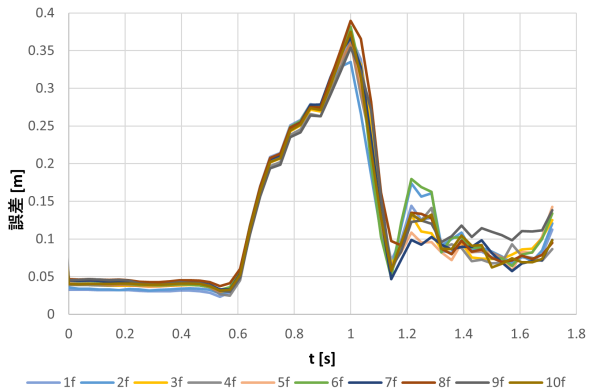

図 8: 各正解フレーム数においての Pred と Corr の誤差を各フレームごとに比較したグラフ

3.3 考察

3.1, 3.2 の結果を確認すると，出力フレームの違いに対する誤差の時間平均間での顕著な差は見られなかった．そのため，正解フレーム数の変更は予測精度向上に寄与しないことがわかった．

4. 結論

本研究では，急な動作変更としてジャンプ動作を対象に現在の予測方法ではどのような挙動を示すかを詳細に確認した．また, 既存の動作予測モデルの正解フレーム数を変更することで, 予測精度を向上できるのかを検証した．その結果，入力に次の動作の予備動作が数フレーム入るまでは，予測が次の動作に対応できないことがわかった．また，正解フレーム数を変更しても，予測の精度は向上しないことが確認できた．そのため, 今後の研究では，予測モデルのアルゴリズムを改変する方針ではなく，多様なフェイント動作を含む学習データを追加することで，急な動作変更に強い予測器の作成を行っていく．

謝辞 本研究は JST さきがけ 17939983 の支援を受けて行われた．

参考文献

[1] 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.

[2] 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.