



# 対向歩行者の動作予測結果の 3 次元提示による歩行動作への影響

Influence on Walking Motion by 3D Presentation of  
Motion Prediction Results of Oncoming Pedestrians

佐々木大祐<sup>1)</sup>, 平井龍之介<sup>1)</sup>, 小山功太郎<sup>2)</sup>, 上島直登<sup>2)</sup>, 許超舜<sup>1)</sup>, 牧野泰才<sup>1)</sup>, 篠田裕之<sup>1)</sup>  
Daisuke SASAKI, Ryonosuke HIRAI, Koutarou KOYAMA, Naoto UEJIMA, Chaoshun XU,  
Yasutoshi MAKINO, and Hiroyuki SHINODA

1) 東京大学 情報理工学系研究科 (〒 113-0033 東京都文京区本郷 7-3-1)

2) 東京大学 工学部 (〒 113-0033 東京都文京区本郷 7-3-1)

**概要:** 本研究では、深層学習を用いて人間の身体動作を予測し、その結果を HMD を用いてリアルタイム 3 次元提示し活用するシステムを提案する。このシステムにおける動作予測結果提示の効果を検証するため、事前収録した歩行データを用いてリアルタイムでない実験用のシステムを構成し、未来の動作を現実の景色に重畳させた際に回避行動に変化が見られるかを確認した。結果として被験者の左右回避が早くなる傾向が見られ、この提示形式の有用性が示唆された。また、リアルタイムシステムの実装について最後に述べる。

**キーワード:** 動作予測, AR, HCI, 深層学習

## 1. 序論

近年、深層学習を利用し、数百 ms 先の未来の人間の姿勢状態を推定する試みが行われている [1][2][3][4]。本研究ではこれらの技術を活用し、動作予測結果を HMD を用いてリアルタイム 3 次元提示 (AR 提示) するシステムを提案する。これが達成されれば、向かい合っすれ違う歩行者同士が 2 人とも同じ方向に避けて衝突してしまうような状況に応用できるほか、発展的には自動運転技術における歩行者の動作の予測など、社会的需要の大きい分野に応用できる可能性がある。

本研究では、このような予測映像と現実映像の重畳表示システムの有用性を検証するために、歩行時の対向者とのすれ違い体験を実現した。対向歩行者の動作を予測しそれを表示することで、衝突を回避するための補助となることを期待し、その効果を実験用デモを用いて検証する。深層学習による動作予測技術の応用における 1 つの可能性を示す。

## 2. 関連研究

人間の動作予測に関する研究は多数行われている。Horiuchi らは、深層学習により 0.5 秒後の人間の姿勢を推定する手法を提案した [1]。本研究の 4. で動作予測と呼んでいる技術は、この論文で提案された手法を踏襲している。板井らは、被験者の身体動作を向かい合うスクリーン上の棒人間状のモデルに提示し、その提示タイミングを-0.4 秒から+0.3 秒の範囲で遅らせるまたは早めることで被験者にもたらされる感覚について分析した [2]。ここで身体動作の提示タイミングを早めるために動作予測技術が用いられてい

る。渡部らは、人間が左右どちらかに曲がる歩行データを用意し、どちらに曲がるかを動作予測により予測している [3]。このとき、歩行データに含まれる関節数を減らして予測を試みることにより、胸と両足の 3 点でも十分な正確性で予測できることを確認し、この部分が動作予測へ特に本質的な影響を及ぼしている可能性を示唆した。本研究の実験設定にある 4 歩の歩行データの作成は、この研究を参考にしている。Wu らは、武道において相手の攻撃動作を予測し、その結果を HMD を通して MR 提示することで武道のトレーニングに役立てることを提案している [4]。動作予測の結果を HMD を用いて現実の景色に重ねて表示しているという点が、本研究で制作したシステムと共通している。

## 3. 深層学習による動作予測

### 3.1 動作予測技術

本章では深層学習を用いた動作予測技術について説明する。人体の動作には慣性や重心、関節の可動域などの影響によりある程度の予備動作や重心移動を必要とするものが存在する。例えば、ジャンプの前にはしゃがみ込むような動作が必要となる、歩行時には軸足に重心を移動させる必要があるため、8 の字に重心が移動する性質があるなどが挙げられる。これらの動作に伴う予備動作から、数百 ms 後の人体の姿勢を推論する事ができる。今日では深層学習を用いることによって数 cm 単位の高精度な動作予測を行うことが可能であることが知られている。Horiuchi らは直前 15 フレーム (0.5 秒) の人体関節 32 個の位置情報を入力とした 5 層の全結合層からなる深層学習ネットワークによって 0.5

秒後の動作予測が可能であることを示した [1].

### 3.2 対向者の動作予測結果のリアルタイム 3次元提示

我々はこの動作予測技術にHMDを用いたAR技術を組み合わせることで、次のようなシステムを考案した。このシステムでは、対向歩行者の動作予測をリアルタイムで行い、その結果を3次元提示する。これによりHMDの装着者は、あたかも数百ms後の未来の対向歩行者の動作を現実の景色に重畳したような景色を見ることができる。

以降、本研究ではリアルタイムでの表示処理は行わず、記録されたデータに対してオフライン処理で得られた結果を提示したものを利用する。リアルタイムシステムの実装については、後の6.章で述べる。

## 4. 歩行動作予測結果を用いた対向者回避実験

3.2節で示したシステムが、有用性のある行動予測の応用例の1つであることを示すため、本章で解説するような実験を行った。この実験では3.2節で提案したシステムをベースにした検証用デモを使用している。この検証用デモはリアルタイムに動作予測をするものではなく、事前収録した歩行データと、事前に推論した予測データを用いている点で3.2節で示したシステムとは異なる。この実験は3.2節で示したシステムにおける、現実に予測結果を重ねて表示する機能の意義を示すことを目的としたものであるため、対向者の動作の計測・推論がリアルタイムでなくとも問題ない。また、計測・推論を事前に行うことで、同じ対向者の動作を複数の被験者に提示して実験することができる。

本実験は3人の被験者を対象として行う。被験者は図1のように、HMDを装着して足踏みをしなが、正面に現れる3Dモデルの動作を注視する。この3Dモデルには事前収録した歩行データが適用されており、被験者に向かい合って近づく向きに歩行する。この歩行データは図2のように、被験者との距離が3.5mの地点から1.0mの地点まで4歩で移動する歩行を記録したものである。この歩行データでは、その歩行の中で左右どちらかに体を避ける動作を行っている。実験において被験者には、3Dモデルが左右どちらに避けるかを出来るだけ早く見極め、3Dモデルと逆の方向に避けてもらう。この一連の動作は、狭い歩道において、ある歩行者とその対向者が互いに避けるような状況を想定している。

3Dモデルに適用する歩行データを収録する際、以下の3つの要素を変えて12種類の歩行動作を行った。

- 体を避ける方向 (左/右)
- 1歩目に踏み出す足 (左/右)
- 何歩目で体を避けるか (2/3/4)

全ての種類の歩行動作を2人で2回ずつ収録し、48個の歩行データを作成した。

歩行動作の収録はAzure Kinectによって行った。歩行動作のデータはAzure Kinectによって取得可能な32点の身体の各部位から、鼻・耳などの微細なもの (NOSEの1点



図1: 対向者回避実験の様子



図2: Azure Kinectを用いた歩行データ計測の様子

と、EYE,EAR,CLAVICLEの左右2点ずつ)を除いた25点の位置情報をフレームレート30fpsで記録した。位置情報は、25点のうち腰骨に相当する位置に相当する点のみを絶対座標で記録し、残りの位置情報は腰骨からの相対座標で記録した。

その後、これらのデータをリアルタイムに動作予測モデルへ入力したと仮定したときの、各瞬間における9フレーム後の予測結果を記録したデータを用意した。実験で用いた行動予測モデルはHoriuchiらの研究と同様に5層の全結合層を用いた深層学習モデルである [1]。先述の手法で収録したデータから抽出した15フレームの時間的に連続する人体関節位置データとその9フレーム後の人体関節位置データのペアによって構成されるデータセットを用いて1000Epochの学習をさせたものを用いた。更に、身体動作の持つ平行移動不変性をネットワークに学習させるため、データ拡張としてデータセットをネットワークに入力する際にランダムな量のオフセットを行った。

また、今回は東京大学VR教育研究センターで撮影したデータを基に作成した図3のようなアバターを使用した。このようなアバターの使用は3.2節で提案したシステムにおいて、予測結果を重ね表示する際に使用できる1つの手段

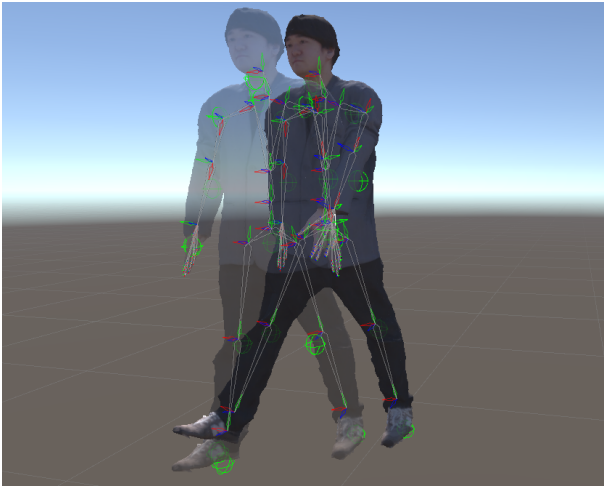


図 3: 東京大学 VR センターで作成したアバターと重畳表示された動作予測結果 実験では実験室の背景映像にこのアバターを重畳表示した映像を提示した。

である。実際に 3.2 節で提案したシステムを実装する際はカメラで取得した映像からリアルタイムに対向者の姿を再構成するのが理想であるが、今回はその代わりにこのようなアバターを使用し、対向者の姿のリアルタイム再構成ができていない状態を想定した。関節位置情報からアバターの姿勢情報への変換には、逆運動学によって関節の角度情報などを補完するソフトウェアである SAFullBodyIK を利用した [5]。

被験者実験では、48 個の歩行データをそのまま提示する 48 パターンと、その上に 9 フレーム後の予測結果を図 3 のように半透明にして重ねて表示する 48 パターンの合わせて 96 パターンを用意し、被験者に対向者を避ける行動を取ってもらった。この動作をする 3D モデルを、Quest Pro のシースルー機能によって読み込まれた現実の背景に重畳し、その映像が HMD に出力される。実験結果として、被験者の装着している Quest Pro の頭部座標をフレームレート 30 で出力した。実験の評価ではこのデータを解析し、3D モデルの動作に対してどれだけの遅延時間で被験者が回避行動を取ったかを集計する。予測結果を重ねて表示したときの方が、より少ない遅延時間で回避行動を取れるという結果を期待して実験を行った。仮にこの仮説が正しければ、動作予測結果を現実を重ねて表示することに意義があるといえる。

## 5. 評価

### 5.1 実験結果

本実験では、実験開始前に毎回 HMD の座標軸のキャリブレーションをしており、HMD 座標のうち X 座標を見ることによって被験者の左右移動を判断できるようになっている。左右移動をした時刻は、実験開始から終了までの左右移動による X 座標の最大振幅の 50% を初めて超えた時刻と定義した。以上を踏まえ、3 人の被験者に対して 96 回ずつ測定した、全 288 個の HMD 座標データをまとめた箱ひ

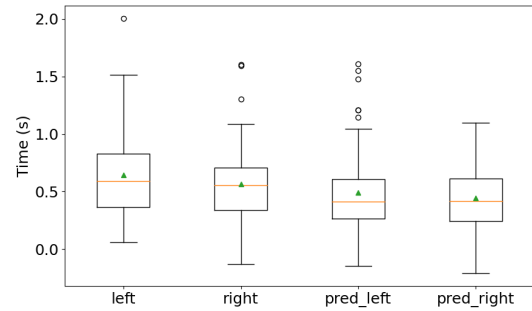


図 4: 被験者と 3D モデルの左右移動時刻の差分

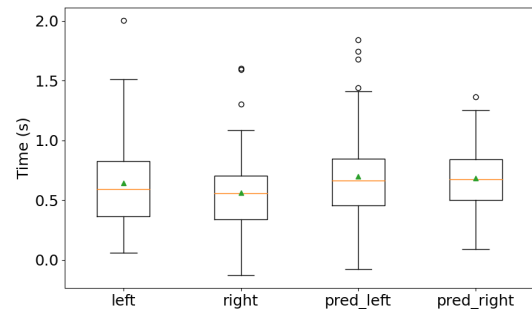


図 5: 被験者と 3D モデルの予測の左右移動時刻の差分

げ図を図 4, 図 5 に示す。緑三角の点は平均値である。また、これらのデータの、平均値と標準偏差を表 1, 表 2 に示す。これらの図における左右は被験者の視点から見た被験者が避ける方向を示している。図 4 における left, right は、予測を表示しなかった場合の (被験者の左右移動時刻)-(3D モデルの左右移動時刻) を示している。pred\_left, pred\_right は、予測を表示した場合の (被験者の左右移動時刻)-(3D モデルの左右移動時刻) を示している。図 5 における left, right は、図 4 と同じである。pred\_left, pred\_right は、予測を表示した場合の (被験者の左右移動時刻)-(3D モデルの予測の左右移動時刻) を示している。

### 5.2 考察

図 4 と表 1 を見ると、動作予測結果を表示することによって被験者の左右移動時刻が平均で、左方向の場合 153ms, 右方向の場合 120ms 早くなった。今回の実験は被験者が少なく、有意差は出なかったものの、動作予測結果を表示することにより被験者の左右移動が早くなる傾向が見られた。図 5

表 1: 3D モデルの左右移動時刻の差分 (図 4) の平均値・標準偏差

[sec]	left	right	pred_left	pred_right
Mean	0.644	0.563	0.491	0.443
Stdev	0.363	0.324	0.386	0.284

表 2: 3D モデルの左右移動時刻の差分 (図 5) の平均値・標準偏差

[sec]	left	right	pred_left	pred_right
Mean	0.644	0.563	0.701	0.686
Stdev	0.363	0.324	0.386	0.284

と表 2 から、重ねて表示された予測結果を見てから動き出しまでの時間は、予測がない時の 3D モデルの動作を見てから動き出しまでの時間よりも少し長いことが分かる。これは、予測は薄く半透明にした状態で重ねて表示しているため、3D モデル本体の動きよりは少し見づらいことが原因として考えられる。また、動作予測は最終的に左右どちらに曲がるかは間違えないものの、左右へ避ける動きを開始する前の直進している間の時間に少し左右に揺れるような挙動をすることがあった。そのため被験者としては予測が動き始めても、その時点では直進時にも発生するブレとの区別がつかず、予測が左右どちらかに大きく移動するのを待ってから避ける動作をしていたとも考えられ、それが図 5 と表 2 の結果に出ている可能性がある。

以上の考察を踏まえると、今回の実験結果から、動作予測結果の表示によって被験者の左右への回避動作が早くなる傾向を読み取ることができた。被験者を増やした詳細な実験の必要はあるが、動作予測の重畳表示は効果的であるという可能性が示された。

## 6. 結言

### 6.1 本研究の結論

本研究では、動作予測技術の応用例の 1 つとして、対向歩行者の動作予測結果を HMD を用いて観測者へリアルタイム 3 次元提示し、観測者の動作の補助とするシステムを提案した。その中で重要な点である、動作予測結果の提示方法として、現実の景色に重畳表示する方法を提案し、その有用性を確かめるために実験を行った。この実験では対向歩行者の動作をリアルタイム計測・推論・提示するのではなく、事前に計測・推論を済ませておき、その歩行データを複数の被験者に提示して実験を行った。その結果、動作予測結果を現実の景色に重畳表示することで被験者の左右への回避動作は平均で、左方向の場合 153ms、右方向の場合 120ms 早くなり、手法は有効であるという可能性が示された。

### 6.2 今後の展望

本研究ではオフライン処理した結果を提示する形で実験を行った。一方、3.2 節で提案したリアルタイムシステムを実装するための技術要素は揃っており、今後以下の手順でシステムを実現する予定である。具体的なシステム構成要素について、図 6 を参照して解説する。このシステムは、動作を計測される対向者と、HMD を装着した観測者の 2 名の人物を対象とする。使用する機器は PC, Azure Kinect,

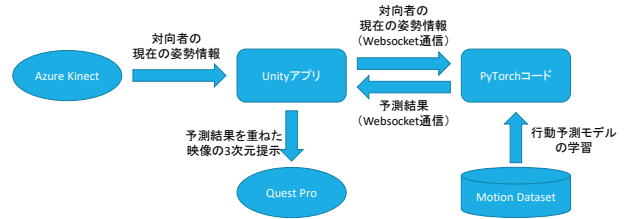


図 6: リアルタイム動作予測システム システム図

Quest Pro である。対向者は、Azure Kinect の計測範囲内で運動することとする。これにより、対向者の現在時点での姿勢状態を計測する。この姿勢情報を蓄積し、現在から 15 フレーム前までの姿勢情報を学習済みの動作予測モデルに入力することで、9 フレーム後の姿勢状態の予測結果を得る。この予測結果は Unity アプリケーション上で Humanoid 形式の 3D モデルに適用できるように変換され、3D モデルの動作として予測結果を現実の景色に重ねて Quest Pro 上に表示する。以上、実験の拡充とリアルタイム動作予測提示システムの構成が今後の展望となる。

謝辞 本研究は科研費 21H03479 の支援を受けて行われた。

## 参考文献

- [1] Yuuki Horiuchi, Yasutoshi Makino, and Hiroyuki Shinoda. 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] 板井俊樹, 牧野泰才, 篠田裕之. 動作予測情報を利用したアバター動作変調による身体感覚操作. ロボティクス・メカトロニクス 講演会 2023, pp. 2A2–G17, 2023.
- [3] 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. ロボティクス・メカトロニクス 講演会 2023, pp. 2A1–H28, 2023.
- [4] Erwin Wu and Hideki Koike. Futurepose-mixed reality martial arts training using real-time 3d human pose forecasting with a rgb camera. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1384–1392. IEEE, 2019.
- [5] Stereoarts. Safullbodyik. <https://github.com/Stereoarts/SAFullBodyIK>.