

# 歩行者の姿勢に基づいた移動経路の予測手法の構築

## Trajectory Prediction Based on Pedestrian Posture

○学 阿部 翼 (工学院大)

正 禹 ハンウル (工学院大)

Tsubasa ABE, Kogakuin University  
Hanwool WOO, Kogakuin University

This study proposes a system to predict a trajectory through detecting pedestrian postures for a mobile robot. The posture information, which is the position coordinates and quaternions of 32 joints, is obtained as time series data using a near-infrared camera when pedestrians are walking straight or turning left and right. They are used as inputs for machine learning. Since the length of the input data greatly affects the prediction speed and accuracy, this paper discusses the effect of the length of input data on the performance of trajectory prediction. The objective is then to construct a system for detecting instantaneous changes in the posture of pedestrians.

**Key Words:** Pedestrian Trajectory Prediction, Pedestrian Posture Recognition, Neural Network

### 1. 精言

近年、コロナ感染防止から宅配需要が増え、配達におけるサービスを目的とした自律移動ロボットの需要が高まってきている<sup>[1]</sup>。自律移動ロボットを実用化するためには、人や動的障害物などの時々刻々と変化する環境下においても、ぶつかることなく効率的な移動を可能にする制御システムが求められる。

一般的にロボットの安全な自律移動を実現するためには、動作環境の地図情報とロボットの位置・姿勢情報が必要である。また、歩道や屋内環境でロボットの安全な走行を行うために歩行者の移動経路や進行方向を認識することも重要な要素の1つである。歩行者の移動経路を推定する手法として、画像に映る物体を認識する技術だけでなく、リアルタイムで人体の骨格を検出することで、歩行者の姿勢情報を推定することができる<sup>[2][3]</sup>。倉元らは、車載カメラにより撮影された画像から歩行者を認識し、2次元の姿勢データを取得するシステムを構築した<sup>[4]</sup>。この2次元姿勢データを多層パーセプトロンで学習を行うことで歩行者の進行方向を8種類に分類した。しかし、この手法は画像の大きさによって精度が変化している。また、特徴点位置が15点以上検出されたものを使用しているが、適切な特徴点を抽出し、人が突如方向転換した時に対応できるシステムが必要である。

そこで本研究では、歩行者の直進、左、右方向の3パターンの関節位置32点分の姿勢情報を時系列データとして取得する。その入力データの長さは予測時間と精度に大きく影響するため、最適な値に設定する必要がある。なので、この研究では入力データ数が精度に及ぼす影響を調べ、最適な値を見つけるとともに歩行者の瞬間的な姿勢の変化を検知するシステムの構築を目的とする。

### 2. 機械学習を用いた軌道予測手法の構築

複数枚の画像を入力する方法は、一般的な時系列データの機械学習である。この複数枚の画像を使って学習する手順を図1に示す。まず、複数枚の画像から姿勢データである3次元座標と4次元座標、信頼度、関節番号の時系列データを取得し、そのデータを予測器に書けることで人の移動経路を予測するシステムの構築を行う。

この方法は、データ数が多く、移動経路の予測方法は正確に行えるが、人の方向転換した直後に認識ができず、時間遅れが発生する。そのため、できるだけ少ない入力データ数で姿勢の

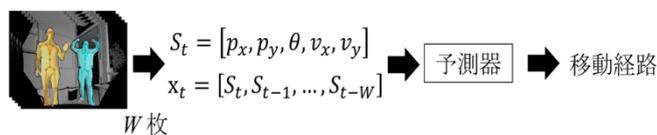


Fig. 1 General learning method for time-series data



Fig. 2 Single image of learning method

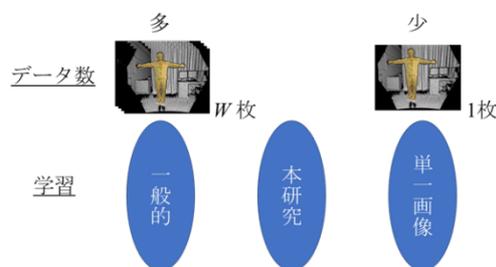


Fig. 3 Conceptual image of proposed method

変化を検知する必要がある。

次に、単一画像を使う手法の手順を図2に示す。車載カメラにより撮影された画像から歩行者を抽出し、姿勢推定をすることで2次元姿勢データを取得し、そのデータを予測器に直接入力して受け取り、人の移動経路の予測するシステムの構築を行った。この方法は、入力データ数が少なく、移動経路の判断は速く行えるが、突如変化した時の方向予測には精度が低くなってしまいう問題が考えられる。

図3にニューラルネットワークの一般的な学習方法と単一画像を使う手法のデータ数の位置関係について示す。一般的なやり方はデータ数としては多く、先行研究は少ないデータ数で移動経路の予測を可能にした。しかし、単一画像を使う手法では人が方向転換した直後の認識が遅れている。また、データ数が多すぎても処理が遅くなり、回避までの判断が遅れてしまうため、少ないデータ数で処理を早くさせ、回避までの判断を速くさせるがある。そこで本研究では、できるだけ少ない

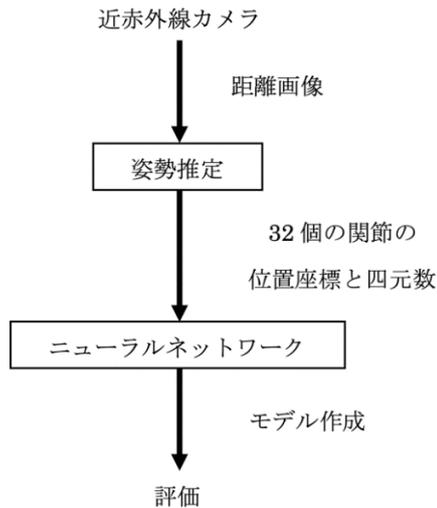


Fig. 4 Schematic diagram of proposed method

データ数、時間で移動経路を予測することが可能なシステムの構築を目指す。また、暗闇における環境下では情報が取れないカメラの弱点を補うために近赤外線が内蔵されているカメラを使用する。

本研究の流れを図4に示す。近赤外線カメラを起動し、歩行者の歩行を開始し、直進方向、左方向、右方向の3パターンから関節位置の32点分の3次元位置座標、4元数、信頼度である姿勢情報を時系列データとしてcsvファイルとして取得する。その後、csvファイルから時系列の関節情報を抽出し、入力を姿勢情報、出力を直進、左、右方向としたニューラルネットワークに抽出したデータを入力する。そのデータを用いてモデル作成を行う。次に、モデル用とは別の評価用に取得したcsvファイルを入れ、評価を行う。

### 3. 実験

#### 3.1 学習時と同様の動きに対する予測モデルの評価

Azure Kinect を用い、取得したデータを元にニューラルネットワークを用いた機械学習を行い、モデルの構築を行った。本実験では、そのモデルを元に新しく取得した姿勢データを使用し、モデルの評価を行った。カメラと人との距離が2mの位置関係でプログラムが動き始めた瞬間から歩行を開始し、データを取得した。図5に示すように左、右、直進方向の3パターン、各100個データを取得した。データを取得した。全データを1つにまとめ、深層学習でモデルの構築を行った。同様に、評価用のデータを図5に示すように3パターンの各10個取得し、評価を行った。本実験で使用する近赤外線カメラの計測周期は15Hzであり、1秒に取得する画像数は15個となる。そこで、1度の入力に使用する画像数を1, 2, 3, 5, 10, 15個に変化させ、入力画像数による移動経路の予測精度を検討した。その結果から高精度かつ瞬時に移動経路を予測可能なモデルの構築を行った。

図6から入力データの計測時間が0.200s(入力画像3枚)の時に精度が高くなったことが分かる。情報量が多くなるたびに精度が上がらなかったのはデータ数を多くしたことにより、みる値が多くなり、逆に精度が下がってしまったことが考えられる。それは、骨格の取得方法として、映ってない関節の位置座標も予測し、そのデータがファイルに出力されているからである。また、0.200sから秒数が短くなるたびに精度が下がったのはデータ数が少なすぎることが考えられる。先ほどの映ってない部分でも関節位置座標を出力できる点から外

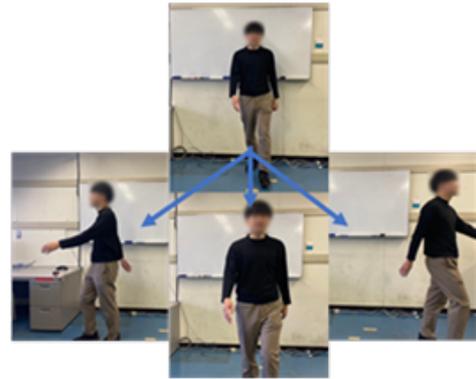


Fig. 5 Three walking directions with same movement as model

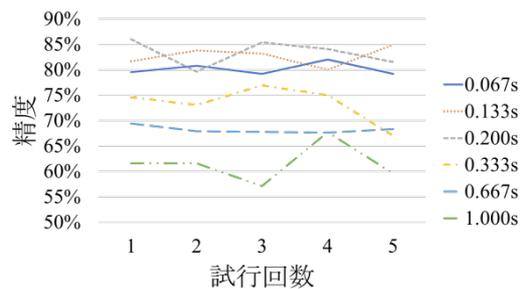


Fig. 6 Result of doing same movement as model

れ値のみで学習してしまうことがあるので精度が下がってしまったと考えられる。

#### 3.2 学習時と異なる動きに対する予測モデルの評価

前章では、モデルの学習時と同じような動きを取得し、そのデータを評価した。しかし、人の移動経路がモデルの学習時と全く同じとは限らないため、歩行者が異なる動きをした際の予測精度を検証する必要がある。本実験では、直進から左方向、右方向、直進方向の3パターンの方向転換を行い、データを取得し、直進についてはモデルと同じ方法でデータを取得した姿勢データを使用し、モデルの評価を行った。

図5に示すように左、右、直進方向の3パターン、各100個データを取得した。全データを1つにまとめ、深層学習でモデルの構築を行った。同様に、モデルの評価用のデータを図7に示すように左方向、右方向は直進から左方向、右方向、直進方向の3パターンに方向転換したデータを取得し、モデルと同様に直進方向の3パターンの各10個取得した。スタート位置はカメラから4m、方向転換をする位置はカメラから2mの位置とした。今回は4パターンの評価を行った。

- 入力パターン1  
直進から左、右、直進方向の3パターンに方向転換したデータをそのまま評価した。
- 入力パターン2  
入力パターン1の左方向と右方向のデータの中から直進と思われるデータを抜き取り、評価を行った。
- 入力パターン3  
入力パターン2からデータごとの取得した秒数を抜き取った後もそのままにしていたので0からに直し、評価を行った。
- 入力パターン4  
入力パターン3は左方向と右方向よりも直進方向の1個ごとに取得したデータ数が多いので、直進方向のデータ数を半分にし、評価を行った。

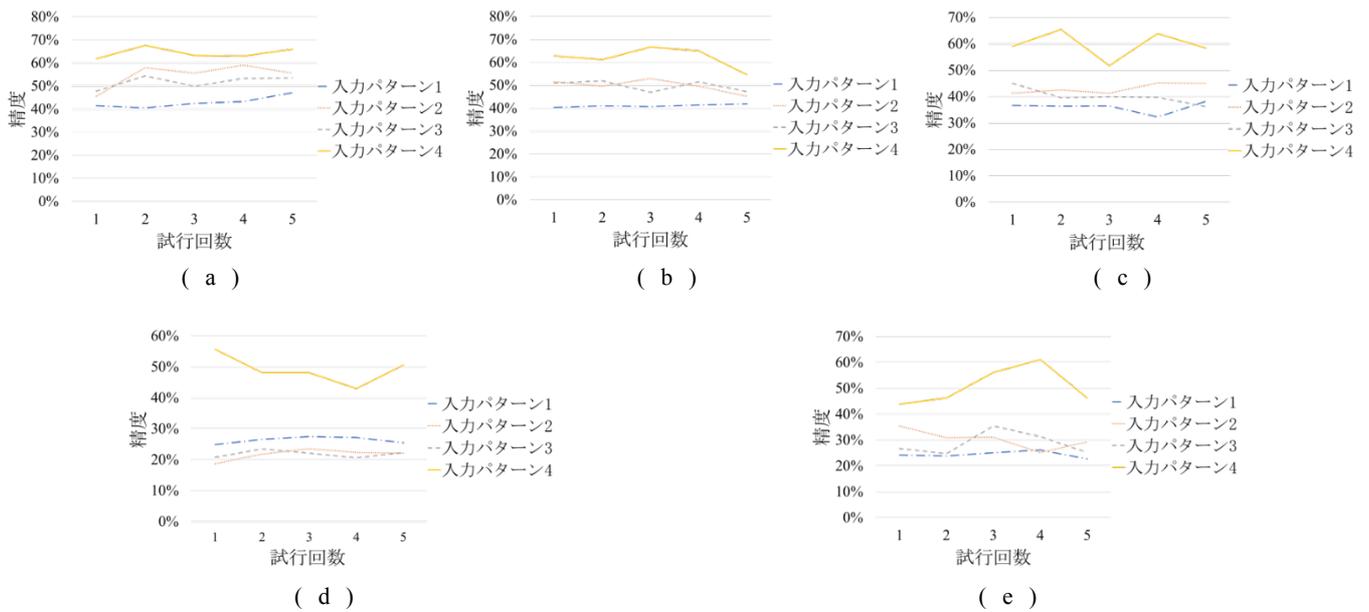


Fig.8 Evaluation results

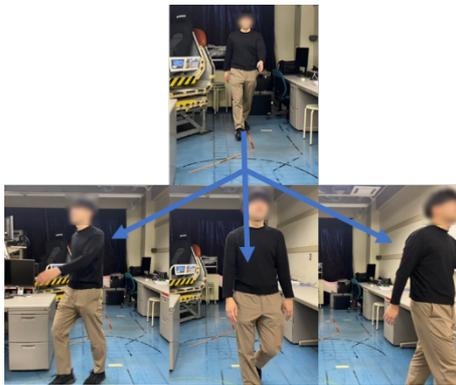


Fig. 7 Three walking directions with change direction from straight ahead

本実験で使用する近赤外線カメラの計測周期は 15 Hz であり、1 秒に取得する画像数は 15 個となる。そこで、1 度の入力に使用する画像数を 2, 3, 5, 10, 15 個に変化させ、入力画像数による移動経路の予測精度を検討した。その結果から高精度かつ瞬時に移動経路を予測可能なモデルの構築を行った。

(i)入力データの長さ 0.133 s (入力画像数 2 枚)

学習回数は 100 回行った。入力パターン 1~4 の 0.133 s の結果を図 8(a)に示した。

(ii)入力データの長さ 0.200 s (入力画像数 3 枚)

学習回数は 60 回行った。入力パターン 1~4 の 0.200 s の結果を図 8(b)に示した。

(iii)入力データの長さ 0.333 s (入力画像数 5 枚)

学習回数は 50 回行った。入力パターン 1~4 の 0.333 s の結果を図 8(c)に示した。

(iv)入力データの長さ 0.667 s (入力画像数 10 枚)

学習回数は 30 回行った。入力パターン 1~4 の 0.667 s の結果を図 8(d)に示した。

(v)入力データの長さ 1.00 s (入力画像数 15 枚)

学習回数は 25 回行った。入力パターン 1~4 の 1.000 s の結果を図 8(e)に示した。

入力パターン 4 の左、右方向の直進のデータ数、直進方向のデータ数を半分にし、秒数を修正したデータに関しては 0.133 s, 0.200 s, 0.333 s, 0.667 s, 1.000 s のどのグラフについても入力パターン 1~3 の評価データよりも精度が高くなった。また、入力パターン 2 と 3 の直進方向のデータを減らさずに行った評価は精度が低くなった。その結果から直進方向のデータ数が左方向と右方向のデータ数よりも多すぎると精度が下がることがわかった。

また、入力パターン 2 と 3 については左方向と右方向のデータ数を減らし、入力パターン 2 は抜き取った場所の秒数を修正していないデータと入力パターン 3 は抜き取った場所の秒数を 0 からに修正したデータなのだが、0.133 s, 0.200 s, 0.333 s, 0.667 s, 1.000 s のどのグラフについても精度の差はなく、近い値であったので秒数は精度において影響を及ぼさないことがわかった。

直進から左、右、直進方向に方向転換した入力パターン 1 に関しては、0.133 s, 0.200 s, 0.333 s, 1.000 s においては、入力パターン 2 と 3 よりも精度の値が小さくなったが、0.667s については入力パターン 2 と 3 よりも精度の値が大きくなった。入力パターン 2 と 3 の精度が小さくなってしまった原因は、1 度に入力するデータ数を多くしたときに情報量が多くなり、1 度に見る値が多くなったことだと考えられる。1.000s のときのグラフも見てみると、入力パターン 2 と 3 については精度が安定しおらず、値が上下している。1.000 s の 15 個のデータ数のときでみると安定していたときと安定していないときがあったが、0.667 s の 10 個のデータ数でみると不都合なデータが多く存在していたのではないかと考えられる。

#### 4. 結言

近赤外線カメラで取得した人の姿勢データをニューラルネットワークに入れ、移動経路予測モデルの構築を行った。また、突如出現する場合と直進から突如方向転換する場合の歩行者に対する移動経路予測モデルの評価を行った。予測モデルの入力データ数としては、0.200 s が最も精度が高くなることがわかった。直進から方向転換をした歩行情報を入れた評価実験については、左、右方向の直進のデータ数、直進方向

のデータ数を半分にし、秒数を修正したデータである入力パターン4が精度の値が高くなることがわかった。

今後の課題として、精度によってモデルの精度と評価をすることはできたが、ロボットに実装したときに左、右、直進方向に適切に回避を行えるかは未知数である。今後の予定として、ロボットに今回作成したモデルを実装し、適切に回避行動を行えるかを確認していく。

### 参考文献

- [1] 国土交通省, “自動走行ロボットとの連携について”, バリアフリーナビプロジェクト説明資料, 2022, <https://www.mlit.go.jp/sogoseisaku/soukou/content/001479667.pdf>, (閲覧日: 2022年3月25日)
- [2] Redmon, J. and Farhadi, A., “YOLOv3: An Incremental Improvement,” arXiv, 2018, <https://pjreddie.com/darknet/yolo/>
- [3] Cao, Z., Simon, T., Wei, S. and Sheikh, Y “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” IEEE Computer Vision and Pattern Recognition Conference, pp.7291-7299, 2017
- [4] 倉元 昭季, “単眼カメラ画像上の2次元姿勢を用いた歩行者の進行方向推定”, 自動車技術会 2022年秋季大会学術講演会予稿集, 文献番号 20226122, 2022.