

# 骨格データを用いた人物の行動認識手法における汎用性に関する検討

雲 河晨<sup>†</sup> 中村 悦郎<sup>†</sup> 景山 陽一<sup>†</sup> 石沢 千佳子<sup>†</sup>  
 加藤 伸彦<sup>‡</sup> 五十嵐 健<sup>‡</sup> 鈴木 守<sup>‡</sup>  
 秋田大学<sup>†</sup> エイデイケイ富士システム<sup>‡</sup>

## 1. 背景・目的

近年，労働現場では労働者の高齢化が進行している．特に，建設業における高齢労働者は，全産業と比較して高い割合を占める<sup>[1]</sup>．高齢労働者は，事故発生率が若年者と比較して高いため，建設現場などにおける従事者の安全管理を行うことが必要不可欠である．したがって，作業管理の可視化や作業従事者の労働環境向上を実現させ，作業の効率化および安全性向上を図る必要があると考える．特に，危険行動および疲労を自動で予測し，作業者に注意喚起を促すシステムは，安全かつ安心な労働環境の構築に寄与すると考える．

本研究グループではこれまでに，各被験者が5種類の動作をしている動画像から骨格特徴点の時系列変化を抽出し，Long short-term memory<sup>[2]</sup>(以下，LSTM と略記する)を用いて，人物の動作を認識する手法に関して検討を行った．このとき，正しく取得できなかった骨格特徴点に対して補間処理を行い，補間済みの骨格特徴点から算出した特徴量を使用して検討を行った．その結果，骨格特徴点の時系列変化を学習させた LSTM を用いることは，動作の判別を可能にすることを明らかにした．しかしながら，提案手法の汎用性に関して十分な検討を行うまでには至っていない．

本稿では，建築現場などにおける複数の作業動作および異常動作を想定した人物の動作を対象とし，提案手法の実用性と汎用性に関して検討を加えた．具体的には，被験者10名が12種類の動作をしている動画データを対象として，動作判別手法における特徴量の汎用性に関して検討を加えた．

## 2. 使用データ

一般的な蛍光灯(照度 500~900lx)の下，3台の単眼カメラ(FDR-AX60：ソニー社製，HC-VX2M：パナソニック社製×2台)を用いて，被験者10名が下記の12種類の動作を行う様子を動画データ(1920×1080pixel, 60fps)として3回ずつ取得した(図1参照)．

- ・歩行：手を振りながら一定の速度で歩行する動作
- ・しゃがみ，立ち上がり：屈伸運動をする動作
- ・上り，下り：脚立を被験者が上り下りする動作
- ・座り，起立：椅子に被験者が着座起立する動作

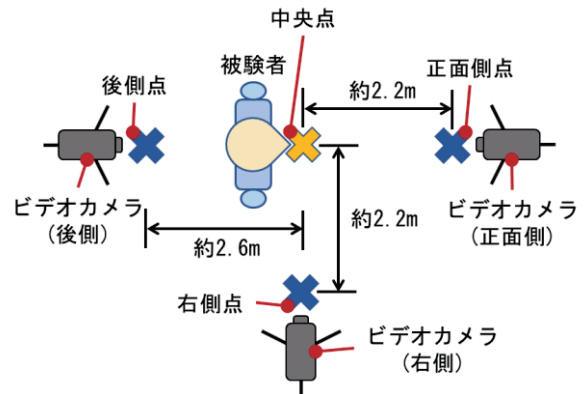


図1 データ取得環境

- ・持ち上げる，落とす(前・後)：物体を頭上に持ち上げ，体の前または後ろに落とす動作(安全面を考慮し，柔らかい物体を使用)
- ・倒れる，起きる：マットレスに倒れ，起きる動作(安全面を考慮し，ゆっくり倒れるように指示)

歩行動作は，被験者正面に配置した1台のカメラに向かって歩く様子を，左・右方向，接近する方向，ならびに遠ざかる方向を撮影した．他の2台のカメラに対しても同様の撮影を行った．また，歩行以外の動作は，3台のカメラで正面，側面，ならびに背面からそれぞれ撮影した．なお，本研究は「秋田大学手形地区におけるヒトを対象とした研究に関する倫理規程第6条第2項」に基づき，被験者の同意を得た上でデータを取得している．

## 3. 提案手法

### 3.1. 骨格特徴点の取得処理

画像から人間の骨格データを検出可能なオープンソースライブラリのOpenPose<sup>[3]</sup>とBODY\_25<sup>[3]</sup>モデルを用いて，動画像データの各フレームから25点の骨格特徴点を取得した．これらの骨格特徴点の時系列変化を以降の検討に使用した．

### 3.2. 特徴点の補間処理

動作姿勢によっては側面や背面から撮影した場合，体や脚立で検出対象が隠れることに起因し，特徴点の値に欠損値が存在することがある．特徴点の補間処理手順を下記に示す．

- (1) 対称性がある特徴点(手や足など)については，欠損した点の座標に，もう片方の検出可能な点の座標を設定して補間した．
- (2) 対称性の無い特徴点に対しては，着目フレームから，次に特徴点検出されるまでのフレームにおける特徴点座標の時系列変化に基づ

A Study of Versatility in Action Recognition Method Based on Human-Skeleton Data

Hechen Yun<sup>†</sup>, Etsuro Nakamura<sup>†</sup>, Yoichi Kageyama<sup>†</sup>, Chikako Ishizawa<sup>†</sup>, Nobuhiko Kato<sup>‡</sup>, Ken Igarashi<sup>‡</sup>, Mamoru Suzuki<sup>‡</sup>

<sup>†</sup>Akita University, <sup>‡</sup>ADK Fuji System

いて線形補間を行った。また、前、あるいは後の5フレームにおける特徴点座標の時系列変化に基づいて、線形補間を行なった。

- (3) 両目および鼻の欠損値については、耳の特徴点の座標の中間に鼻の特徴点を設定した。さらに、三角関数を使用して、両目の特徴点座標を設定した。

### 3.3. 特徴量の算出

補間した特徴点を利用し、①～④に示す数値の時系列変化を特徴量としてそれぞれ算出した。

- ① 座標：骨格特徴点の座標
- ② 角度：2つの骨格特徴点を結ぶ直線の角度
- ③ 距離：2つの骨格特徴点を結ぶ直線の長さ
- ④ 移動軌跡：前後フレームにおける骨格特徴点の移動ベクトルデータ

### 3.4. LSTMの学習処理

特徴量を用いて、被験者が行なっている動作を認識するために、LSTMの学習処理を行った。具体的には、入力層306次元、中間層4層(各層128, 64, 64, 32次元)、出力層12次元の6層構造のLSTMを用い、任意の連続した15フレームにおける特徴量の時系列変化から動作の種類を分類するように、LSTMの学習を行った。このとき、過学習を防止するために、中間層第2層～第3層間および第3層～第4層間にDropout層(ドロップ率: 0.2)を設け、中間層第2層および第3層において、L1(レート: 0.02)およびL2(レート: 0.02)の正則化を行った。なお、勾配法にはAdamを用いた。

### 3.5. 動作の判別処理

15フレームのフィルタを時系列順に1フレームずつシフトし、動画データにおける各区間の特徴量の時系列変化を取得した。さらに、各特徴量の時系列変化を学習済みのLSTMに入力し、各フレーム間における12種類の動作を判別した。

## 4. 動作判別実験

### 4.1. 実験内容

提案手法の汎用性を評価するために、被験者10名を対象として、動作の判別実験を実施した。具体的には、被験者9名の全ての動画データと教師データ、残り1名の全ての動画データをテストデータに設定し、教師データとテストデータの被験者の組み合わせを変更して、10データセットの組み合わせ(No.1～No.10)に対して交差検証を実施した。LSTMに入力した時系列特徴量が手動で分類した動画データのラベルと同じ動作であると判別された場合に、判別成功とした。なお、LSTMの学習回数を10～100回まで1刻みで変更し、各動作の判別成功率を算出して比較した。

### 4.2. 実験結果および考察

提案手法における各パターンの平均判別成功率は、学習回数に応じて変動し、学習回数20回以降、最大で79.8%、最小で71.9%となった。特に、全パターンにより、学習回数32回の場合において、判別成功率の平均値が最も高くなり、76.7%以上の平均判別成功率を得た。学習回数を32回に設定した場

表1 動作判別結果(学習回数: 32)

データセット	テストデータ	判別成功率
No.1	被験者 A	79.5%
No.2	被験者 B	73.2%
No.3	被験者 C	78.3%
No.4	被験者 D	75.1%
No.5	被験者 E	78.7%
No.6	被験者 F	77.6%
No.7	被験者 G	78.9%
No.8	被験者 H	74.7%
No.9	被験者 I	74.6%
No.10	被験者 J	76.3%
平均判別成功率		76.7%

表2 真の結果に対する予測結果の割合(データセット No.1) [単位: %]

	予測結果											
	歩行	しゃがみ	立ち上がり	上り	下り	座り	起立	持ち上げる	前に落とす	後に落とす	倒れる	起きる
歩行	98.3	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.7	0.8
しゃがみ	0.2	73.1	26.3	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.2
立ち上がり	1.8	39.9	57.3	0.0	0.0	0.2	0.6	0.0	0.0	0.0	0.0	0.2
上り	1.6	0.0	0.2	70.6	26.1	0.0	1.0	0.6	0.0	0.0	0.0	0.0
下り	9.3	0.0	0.0	16.0	72.4	0.0	1.9	0.1	0.0	0.0	0.0	0.2
座り	0.0	3.5	9.7	0.0	0.0	69.4	17.4	0.0	0.0	0.0	0.0	0.0
起立	0.2	1.0	4.6	0.0	0.0	39.5	54.5	0.0	0.0	0.0	0.0	0.2
持ち上げる	1.0	0.0	0.0	0.0	0.2	0.2	84.7	9.9	3.8	0.0	0.0	0.0
前に落とす	0.2	0.0	0.0	0.0	0.0	0.0	22.6	77.0	0.2	0.0	0.0	0.0
後に落とす	0.0	0.0	0.0	0.2	0.0	0.0	24.7	3.5	71.6	0.0	0.0	0.0
倒れる	0.5	0.3	0.6	0.0	0.0	0.2	0.2	0.0	0.0	59.7	38.5	0.0
起きる	1.1	0.1	0.0	0.0	0.5	0.2	0.3	0.0	0.0	21.3	76.6	0.0

合の動作判別結果を表1に示す。特に、被験者Aをテストデータとして使用した場合、判別成功率が79.5%となり、最大の判別成功率が得られた。

以上の結果は、今回対象とした動作の判別に提案手法が有用であることを示唆している。

また、学習回数32回において、最も高い成功率が得られた事例(被験者A)を対象とし、真の結果に対する予測結果の割合を算出した(表2)。真の結果と予測結果が一致する事例の割合が高いことがわかる。一方、“座り”と“起立”のように、類似した動きが誤判別される事例を認めた。これは、移動方向における特徴量が必ずしも有効でなかったこと、および動作が短時間であることに起因し、学習が困難であったためと考える。したがって、有効な特徴量の選定、およびLSTMに入力する適切なフレーム数に関してさらに検討する必要がある。

### 参考文献

- [1] 国土交通省：“建設産業の現状と課題”，<https://www.mlit.go.jp/common/001187379.pdf> (Accessed 2020/11/16)
- [2] S. Hochreiter and J. Schmidhuber: Long short-term memory, Neural computation, Vol. 9, No. 8, pp. 1735-1780 (Nov. 1997)
- [3] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei and Y. Sheikh: OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 43, No. 1, pp. 172-186 (Jan. 2021)