

# 多視点下における 3次元関節座標の歩容特徴による個人識別の精度向上

河野 瞭人\*      小林 太 (神戸大学)

Improving Person Identification Accuracy Using 3D Joint Coordinates  
from Gait Features Under Viewpoint Variations  
Akito Kono\*, Futoshi Kobayashi, (Kobe University)

This research proposes a model-based method for individual identification using time-series 3D skeleton coordinates acquired from gait via an RGB camera. We transform these coordinates to a canonical viewpoint before inputting them into a CNN for training. Our experiments demonstrate that this approach achieves a certain improvement in identification accuracy, even for gait observed under multi-viewpoint conditions.

**キーワード**：歩容認識, 個人識別, モデルベース, RGB カメラ, CNN,  
(Gait recognition, Person identification, Model-based, RGB camera, CNN)

## 1. はじめに

近年, 人の歩き方にそれぞれ独自の個性が存在することを利用し, 歩容から個人を識別する技術が注目されている<sup>(1)</sup>. 歩容による個人識別とは, 異なる場所や時間帯でカメラやセンサーに認識された対象者を, 歩容の特徴から同一人物と識別する技術である<sup>(2)</sup>. この技術は, 不審者を追跡する監視システムや, 商業施設における人の移動経路の統計的分析への活用が期待されている. また, 歩容の特徴を活用した関連技術として, 生体認証のうちの一つである歩容認証が存在する. 歩容認証とは, 歩容の特徴から個人を特定するもので, セキュリティシステムや本人確認方法として活用され始めている. 歩容認証は, その他の生体認証の顔認証, 虹彩認証, 指紋認証, 声紋認証などと比べて, 偽造, 隠蔽が困難であり, 遠距離からの認証が可能という特長がある. 対象人物が遠方にいても, ヘルメットやマスクを着用していても, 歩容をカメラで捉えることで認証が可能である.

個人識別, 歩容認証には, 主にシルエットベースとモデルベースの 2 つの手法がある. シルエットベースの代表的な手法として, Gait Energy Image(GEI)<sup>(3)</sup>と呼ばれる画像を歩容特徴として用いるものがある. GEI は, 歩行中の人物のシルエットを一定時間にわたり平均化した画像である. この手法は, 計算量が小さく, 高い識別精度が出されている. しかし, 歩容のシルエットの情報に依存するため, 撮影角度, 服装, 持ち物により歩容のシルエットに変化がある場合に識別精度が低下してしまう課題がある. 例えば, カ

バンを持った場合やコートを着た場合など, シルエットが大きく変化するシナリオでは, 精度が低下しやすい. 一方で, モデルベースでは歩行中の人物の関節の位置や角度などの人体モデルの動きの情報をを用いる<sup>(4)</sup>. そのため, 歩容の撮影角度が変化する際の影響を受けにくく, 異なる撮影角度でも安定した識別が可能と考えられる. モデルベースの手法では, 3次元骨格座標の時系列データから人体モデルの特徴を抽出するものがあり, 従来は 3次元骨格座標データの取得には, RGB カメラに加えて深度センサーを搭載した Kinect を使用する方法<sup>(5)-(7)</sup>や, 複数のカメラで異なる角度から同時に対象者を撮影する方法<sup>(8)</sup>が存在する. しかし, 近年の骨格推定技術の進歩により, MotionBERT<sup>(9)</sup>と AlphaPose<sup>(10)</sup>を活用することで, 単一の RGB カメラのみで高精度な 3次元骨格座標の時系列データを取得できるようになった. そのため, Kinect のような特殊な機材を必要とせず, より低コストかつ広範囲な実用が可能である.

本研究では, RGB カメラのみを用いて歩容から 3次元骨格座標の時系列データを取得し, モデルベースで個人識別する手法を提案する. 3次元骨格座標の時系列データは, 歩容の撮影角度が変化する際の影響を受けにくい歩容特徴である. この特性を活かして, 撮影角度の変化に対してロバストな個人識別システムの開発を目的とする. また, 従来の Kinect を利用する方法<sup>(5)-(7)</sup>や複数台のカメラを同時に用いる方法<sup>(8)</sup>ではなく, 単一の RGB カメラを用いて, MotionBERT<sup>(9)</sup>と AlphaPose<sup>(10)</sup>から 3次元骨格座標を推定する. これにより, より低コストかつ広範囲での実用が可能である. 実験では, 10名の被験者を対象に, 4つの異なる

る撮影角度(45°, 90°, 225°, 270°)から歩行データを撮影し、撮影角度の異なるデータ間で個人識別を行うことで精度を検証する。

## 2. 歩容による個人識別

〈2・1〉 概略 個人識別システムでは、RGB カメラで撮影した歩行動画から 3 次元骨格推定技術(MotionBERT<sup>(9)</sup>と AlphaPose<sup>(10)</sup>)を用いて 3 次元骨格座標の時系列データを取得し、これを正規化した後に骨格座標画像に変換する。その後、得られた画像を CNN で学習し、個人識別を行う。この手順の全体図の構成を Fig. 1 に示す。

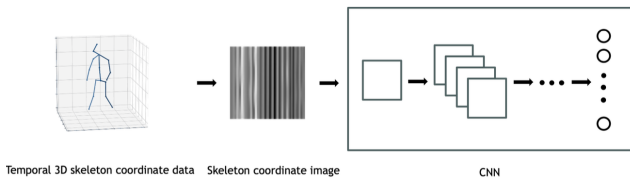


図 1 システム概要図

Fig. 1. System Overview Diagram.

〈2・2〉 3 次元骨格座標の取得 RGB カメラで撮影した歩行動画から 3 次元骨格推定システムである MotionBERT<sup>(9)</sup>を用いて 3 次元骨格座標を取得する。MotionBERT は、2 次元骨格座標データから元の 3 次元骨格座標データを推定できる。そこで、入力に必要な 2 次元骨格座標データは AlphaPose<sup>(10)</sup>を用いて歩行動画から推定する。MotionBERT から出力される 3 次元骨格座標データは、17 個の関節の 3 次元座標で構成され、keypoint の Mapping を Fig. 2, 対応する keypoint 名を Table. 1 に示す。

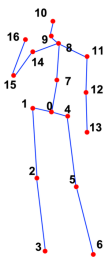


図 2 MotionBERT 出力形式  
Fig. 2. MotionBERT  
Output Format.

表 1 MotionBERT Keypoint 名

Table 1. MotionBERT  
Keypoint Name.

Index	Keypoint	Index	Keypoint
0	MidHip	9	Nose
1	RHip	10	Head
2	RKnee	11	LShoulder
3	RAnkle	12	LElbow
4	LHip	13	LWrist
5	LKnee	14	RShoulder
6	LAnkle	15	RElbow
7	Torso	16	RWrist
8	Neck		

〈2・3〉 正規化 3 次元骨格座標の時系列データを 2 通りの手法で正規化します。

・スケール調整のみ行う方法

keypoint の MidHip を基準点として、MidHip と Head との距離を用いたスケール調整を各フレームごとに行う。具体的には、推定された骨格座標を  $(x, y, z)$ 、正規化後の座標を  $(x', y', z')$  とし、 $i$  はフレームインデックスであり  $i = 1, \dots, N$  で変動し、 $j$  は keypoint のインデックスであり  $j = 0, \dots, 16$  で変動するとして、正規化後の座標は以下の式で表す。

$$\begin{aligned} x'_{i,j} &= (x_{i,j} - x_{i,0}) \cdot \frac{0.95}{y_{i,10} - y_{i,0}}, \\ y'_{i,j} &= (y_{i,j} - y_{i,0}) \cdot \frac{0.95}{y_{i,10} - y_{i,0}}, \\ z'_{i,j} &= (z_{i,j} - z_{i,0}) \cdot \frac{0.95}{y_{i,10} - y_{i,0}}. \end{aligned} \quad (1)$$

・同一視点からの座標に変換する方法

スケール調整および移動方向の統一を行うことで、同一視点から見た座標に変換する。これにより、撮影角度が異なるデータ間においても一貫した特徴量を抽出しやすくなる。本手法では、keypoint の MidHip を基準点とし、他の関節座標を回転および正規化することで、同一視点からの座標に変換する。具体的には、推定された骨格座標を  $(x, y, z)$ 、正規化後の座標を  $(x'', y'', z'')$  とし、 $i$  はフレームインデックスであり  $i = 1, \dots, N$  で変動し、 $j$  は keypoint のインデックスであり  $j = 0, \dots, 16$  で変動するとして、以下の通りに計算する。

$$p_i = (x_{i,0}, z_{i,0}) \quad (2)$$

$$v_i = \begin{cases} (-1, 0) & (i < 3) \\ p_i - p_{i-2} & (i \geq 3) \end{cases} \quad (3)$$

$$d = (-1, 0) \quad (4)$$

$$\theta_i = \arccos\left(\frac{v_i \cdot d}{\|v_i\| \|d\|}\right) \quad (5)$$

$$x'_{i,j} = (x_{i,j} - x_{i,0}) \cdot \frac{0.95}{y_{i,10} - y_{i,0}}, \quad (6)$$

$$y'_{i,j} = (y_{i,j} - y_{i,0}) \cdot \frac{0.95}{y_{i,10} - y_{i,0}},$$

$$z'_{i,j} = (z_{i,j} - z_{i,0}) \cdot \frac{0.95}{y_{i,10} - y_{i,0}}.$$

$$\begin{bmatrix} x''_{i,j} \\ z''_{i,j} \end{bmatrix} = \begin{bmatrix} \cos \theta_i & -\sin \theta_i \\ \sin \theta_i & \cos \theta_i \end{bmatrix} \cdot \begin{bmatrix} x'_{i,j} \\ z'_{i,j} \end{bmatrix}, \quad (7)$$

$$y''_{i,j} = y'_{i,j}.$$

### 〈2・4〉 3次元骨格座標の画像変換

前項で正規化した3次元骨格座標の時系列データを画像変換する方法を述べる。まず、動画ごとに縦軸をフレーム数、横軸をkeypointとして座標データを格納したCSVファイルをpng形式の256階調のグレースケール画像に変換する。横軸はkeypointごとにX,Y,Z座標を並べる。このとき、keypointのMidHipの座標は正規化によりすべてのフレームにおいて0となっているので、画像変換の際には除外している。結果、1本の動画につき1枚の画像が生成される。次に、1枚の画像をある程度の重なりを持たせて正方形の画像に切り取る。具体的には、横軸のkeypointの座標数が48個であることから、縦方向に1pxずつずらしながら47pxの重なりを持たせて(48,48)のサイズに切り取る。この切り取った画像を分類モデルの入力として使用する。例えば、1本の動画から大きさ(48,140)の画像を生成した場合、この画像から(48,48)のサイズの画像を93枚切り取ることになる。

〈2・5〉 骨格座標画像の分類モデル 個人識別における分類にCNNを用いる。CNNは畳み込み層、活性化関数、プーリング層、バッチ正規化、ドロップアウト層、全結合層から構成する。

## 3. 実験

〈3・1〉 実験概要 本学習と推定に用いる歩行の動画は、10人の被実験者に4mの距離を10往復してもらい、2台のRGBカメラをFig.3のように、人が歩行する方向に対して45°、90°の角度から撮影できるように1.5mの高さに配置した。ここで、復路を45°、90°から撮影した歩行の動画は本研究では225°、270°から撮影したものとして扱う。それぞれのカメラで5往復を撮影したため、撮影した歩行の動画は1人あたり20本の計200本となった。そこで、未学習の撮影角度におけるデータをテストデータとして用い、個人識別システムの精度を検証することで、歩容の撮影角度変化に対する性能を評価する。また、2パターンの正規化手法を用いて検証する。学習データには、一人あたり4つの撮影角度のうち3つの撮影角度の動画からそれぞれ4本使用する。したがって、10人×3角度×4本で、合計120本の動画を学習データとして用いる。各動画は1枚の骨格座標画像に変換され、さらに複数枚の正方形画像に切り取り、CNNの入力データとする。テストデータには、学習データとして使用しなかった残りの1つの撮影角度の5本の動画を使用する。つまり、10人×1角度×5本で、合計50本の動画を学習データとして用いる。このようにして、4つの撮影角度からのデータをそれぞれテストデータとする場合を検証する。

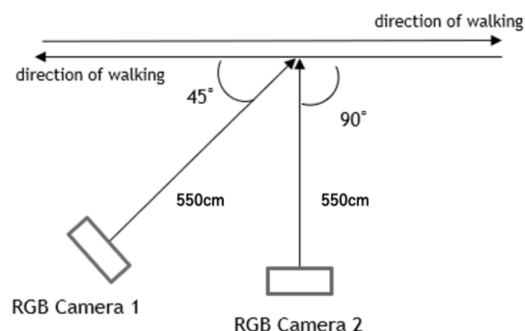


図3 カメラ配置と歩行方向

Fig. 3. Camera Placement And Walking Direction.

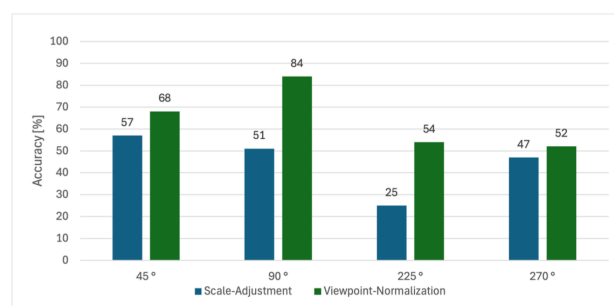


図4 識別結果

Fig. 4. Experimental Results

〈3・1〉 実験結果と考察 2パターンの正規化手法別に分類モデルを作成し、未学習の撮影角度別のテストデータに対する精度をFig.4に示す。

Fig.4より、未学習の撮影角度からのデータをテストデータとした場合は、どの角度においても“Scale-Adjustment”よりも“Viewpoint-Normalization”の手法の精度が高いことから、同一視点から見えるように座標変換することに効果があることがわかる。“Viewpoint-Normalization”の90°からのデータが最も高い精度で84%を達成しており、撮影角度の変化における精度低下の課題に対して、一定の成果がみられる。しかし、その他の撮影角度の精度は低く、撮影角度に対して不変な歩容の特徴量の学習は十分ではない。その原因として、正規化前の座標はカメラ座標系であるため、レンズの歪みの影響を受け、正規化の“Viewpoint-Normalization”による同一視点から見た座標への変換の精度が高くないことが考えられる。また45°、90°のデータの精度が225°、270°より高い。これは45°から撮影したデータを骨格推定の方が体を正面から捉えるため、225°より精度が高い可能性がある。その場合、45°とそれに近い角度である90°における精度が高くなると考えられる。以上のことから、同一視点から見た座標

変換は、撮影角度の変化に対して一定の効果があることを示した。さらにレンズの歪みを考慮した骨格座標データの正規化、骨格推定精度の向上に取り組むことで識別精度をより高くすることができると考えられる。

#### 4. まとめ

3次元骨格座標データを用いることで、撮影角度が変化しない場合の個人識別においては高い精度を達成することができた。しかし、撮影角度が変化した際の識別精度は実用的な段階までは向上していない。この問題を解決するには、レンズの歪みを考慮した骨格座標データの正規化、骨格推定精度の向上に取り組む必要がある。これらの取り組みを通じて、撮影角度が変化した際の個人識別精度の向上が期待される。

#### 文 献

- (1) Md. Khaliluzzaman, Ashraf Uddin, Kaushik Deb, and Md Junayed Hasan : "Person Recognition Based on Deep Gait: A Survey", *Sensors*, Vol. 23, No. 10 pp. 4875 (2023)
- (2) 森 駿文・菊池 浩明 : 「深度センサによる歩容特徴量を用いた個人識別・追跡方式の提案」, *Computer Security Symposium*, pp. 23-25 (2017)
- (3) Han, J. and Bhanu, and Bir : "Individual recognition using gait energy image ", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 2 pp. 316-322 (2006)
- (4) Daisuke IMOTO, Kenji KUROSAWA, Ken'chi TSUTIYA, Kenro KUROKI, Norimitsu AKIBA, and Hidetoshi KAKUDA : "Study on Model-Based Gait Recognition Based on Body Points and Local Shape Information", *JSPE*, Vol. 83, No. 1 pp. 94-100 (2017)  
井元 大輔・黒沢 健至・土屋 兼一・黒木 健郎・秋葉 教充・角田 英俊 : 「身体部位の特徴点と形状情報に基づくモデルベース歩容認証の検討」, *精密工学会誌*, Vol. 83, No. 1 pp. 94-100 (2017)
- (5) 宮島 春菜・山本 正信 : 「Kinect からの方向動作による個人識別」, *一般社団法人映像情報メディア学会誌*, Vol. 67, No. 11 pp. J414-J420 (2013)
- (6) Rashmi M., and Ram Mohana Reddy Guddeti : "Human identification system using 3D skeleton-based gait features and LSTM model", *Journal of Visual Communication and Image Representation*, vol. 82, pp. 103416 (2022)
- (7) 三堀 裕・花泉 弘 : 「Kinect V2 を用いる歩容認識に基づく個人識別手法」, *情報処理学会 第 79 回全国大会講演論文集*, Vol. 2017, No. 1 pp. 439-440 (2017)
- (8) Ryosuke BABA, Yumi IWASHITA, Koichi OGAWARA, and Ryo KURAZUME : " Gait identification robust to changes in walking direction by 4D gait database", *IEICE Technical Report*, (2012)  
馬場 亮輔・岩下 友美・小川原 光一・倉爪 亮 : 「4 次元歩容データによる歩行の向きの変化に頑強な個人識別」, *電子情報通信学会技術研究報告 (信学技報)*, (2012)
- (9) Wentao Zhu, Xiaoxuan Ma, Zhaoyang Liu, Libin Liu, Wayne Wu, and Yizhou Wang "MotionBERT : "A Unified Perspective on Learning Human Motion Representations", *IEEE Conference on Computer Vision*, pp. 15039-15053 (2023)
- (10) Fang HS, Xie S, Tai YW, and et al : "RMPE: Regional multi-person pose estimation[C]", *the IEEE International Conference on Computer Vision*, pp. 2334-2343 (2017)