

# 深度センサによる歩容特徴量の DTW 距離に基づく 個人識別手法の提案

森 駿文 (菊池研究室)

概要: 近年, カメラやセンサを用いた個人識別・追跡の試みが始まってきているが, 顔画像は個人識別符号に分類されているためプライバシーに関する大きな問題が存在する. 本研究は, 顔認証に代わる手法として歩容に注目し, モーションキャプチャを用いて複数の関節の動きを観測し, その時系列データに DTW を適用することでより高精度な識別手法を提案する. そして, 歩容に固有の問題である荷物や履き物によって生じる外乱に対する提案手法の精度低下を調査する.

## 1. はじめに

近年, カメラやセンサなどを用いて個人を識別・追跡し, マーケティングや防犯などに活用する試みが始まってきている.

歩行を用いた人物識別の手法は, 2017 年に Muaaz ら [1] がスマートフォンから得られる加速度情報に対して Dynamic Time Warping(DTW)[2] を適用する手法があるが, 彼らの手法はスマートフォンを用いるため, 対象者の協力が必要, センサの数が限られるためカバンを持つなどの外乱に弱いなどといった問題点がある.

本研究では, 深度センサから得られる人体の関節の 3 次元座標に DTW を適用することで, Muaaz らとは異なり外側のセンサから識別可能な手法を提案する. 提案手法を実装したプロトタイプを用いて実験を行い, 通常の歩行に加えて大股出歩く, 箱を持ちながら歩く, スマートフォンを操作しながら歩くという, 3 つの外乱を含む歩行に対する精度を評価した. 145 人の被験者を用いた実験では, 通常の歩行では各関節の標準偏差 (SD) で重みを付けたユークリッド距離を用いた場合では, 5 つの特徴量を組み合わせることで False Rejection Rate(FRR) が 0.07 まで向上するが, 外乱を含む歩行ではそれぞれの特徴量を単体で用いた場合よりも悪化してしまうという結果になった. 対して  $k$ -NN を用いた場合は, 通常歩行は重み付きユークリッド距離よりも精度は低下するが, 外乱を含む歩行に対しては頑強となることを示した.

## 2. 関連研究

### 2.1 加速度, 角度センサーを使った歩容識別手法

Muaaz らは, スマートフォンの加速度センサを用いた個

人認証手法 [1] を提案している. Muaaz らの手法では, ポケットに入れた Android 端末の加速度情報を特徴量としており, 歩行 1 サイクル分を 1 つのテンプレートし, テンプレートを複数登録する. 認証時はそれぞれのテンプレートとの DTW 距離を計算し, 50%以上のテンプレートとの DTW 距離が閾値以下であれば本人と判定している.

### 2.2 DTW

DTW(Dynamic Time Warping)[2] は, 音声識別などに使用されるパターンマッチングの手法で, 長さの異なる 2 つの時系列データの距離を与える. 一部のデータが損失していても, 時間軸を収縮させてデータの類似距離を算出する. 長さ  $n_p \neq n_q$  の 2 つの時系列データ  $P = (p_1, p_2, \dots, p_{n_p})$ ,  $Q = (q_1, q_2, \dots, q_{n_q})$  の DTW 距離  $d(P, Q)$  は  $d(P, Q) = f(n_p, n_q)$  と定義する. ここで,  $f(i, j)$  は次の様に再帰的に定める.

$$f(i, j) = \|p_i - q_j\| + \min \begin{cases} f(i, j-1), \\ f(i-1, j), \\ f(i-1, j-1), \end{cases}$$
$$f(0, 0) = 0, f(i, 0) = f(0, j) = \infty.$$

## 3. 提案手法

本研究では, モーションキャプチャセンサなどから得られた関節の 3 次元座標 [3] を測定し, 一歩分の時系列データの DTW 距離を算出することによって個人識別を行う. 提案手法は次の方法で行う.

### 3.1 1 サイクル切り出し

身体関節  $l$  の時刻  $t$  における, モーションキャプ

チャセンサの位置を原点とする 3 次元空間の絶対座標を  $a_\ell(t) = (x, y, z)$  とする。測定時間の絶対座標の時系列データ  $\langle a_\ell(t_1), a_\ell(t_2), \dots \rangle$  をスケルトンデータと呼び、その歩行の 1 サイクル分を抽出する。観測した 2 歩程度の歩行から 1 サイクルを抽出する。

まず、時刻  $t$  の左右の足の絶対座標  $a_{LF}(t), a_{RF}(t)$  から、両足の間の距離を  $\Delta(t) = \pm \|a_{RF}(t) - a_{LF}(t)\|$  により計算する。右足が前の状態は正、左足が前のときは負とする。

次に、両足間の距離  $(\Delta(1), \dots, \Delta(n))$  の時系列データにローパスフィルタをかけることでノイズを除去し、そのピーク間を 1 サイクルとする。

### 3.2 関節座標の相対座標化

歩行中の各関節の座標について、身体の中心付近の安定している関節を原点とした相対座標を計算する。

関節  $\ell$  の時刻  $t$  の絶対座標を  $a_\ell(t)$ 、中心の関節の絶対座標を  $a_c(t)$  とすると、相対座標  $r$  は  $r_\ell(t) = a_\ell(t) - a_c(t)$  と定める。 $c$  は 4 章の実験では SpineMid を用いる。

### 3.3 DTW 距離の計算

1 サイクルの関節  $\ell$  の 2 つの時系列データ  $R_\ell = \langle r_\ell(t_1), \dots, r_\ell(t_n) \rangle$  と  $R'_\ell = \langle r'_\ell(t_1), \dots, r'_\ell(t_{n'}) \rangle$  の DTW 距離  $d(R, R')$  を  $R$  と  $R'$  の類似度とする。 $R = R'$  のとき、 $d(R, R') = 0$  である。

### 3.4 複数特徴量の統合と本人識別

#### 3.4.1 重み付きユークリッド距離

複数の部位を用いたときの類似度は、次の様に定める。関節  $\ell$  と  $m$  についての 2 つの時系列データ  $(R_\ell, R_m)$  と  $(R'_\ell, R'_m)$  があるとき、統合 DTW 距離は

$$D((R_\ell, R_m), (R'_\ell, R'_m)) = \sqrt{\left(\frac{d(R_\ell, R'_\ell)}{\sigma_\ell}\right)^2 + \left(\frac{d(R_m, R'_m)}{\sigma_m}\right)^2}$$

と定める。 $\sigma_\ell$  は、関節  $\ell$  の DTW 距離の SD である。同様にして、 $c$  種の関節を統合した場合も、 $c$  次元のユークリッド距離で類似度を定める。各ユーザのテンプレートと比較をし、最も近いユーザがテストデータの持ち主であると判定する。

#### 3.4.2 $k$ 近傍法 ( $k$ -NN)

$U$  を全ユーザの集合、未知のユーザ  $u$  の関節  $\ell$  の歩行 1 周期を  $R_\ell^{(u)}$  とする。登録されている  $n$  人の  $U$  の全テンプレートの全関節と DTW 距離を計算する。得られた DTW 距離を関節  $\ell$  の  $SD\sigma_\ell$  で正規化する。 $u$  の全関節データを  $R^{(u)} = \{R_1^{(u)}, \dots, R_{24}^{(u)}\}$ 、テンプレート集合を  $T = \{R_1^{(u_1)}, R_2^{(u_2)}, \dots, R_{24}^{(u_n)}\}$  とすると、上位  $k$  個の特徴量は  $N(R^{(u)}) =$

$$\left\{ v \in U \mid \frac{d(R_\ell^{(u)}, R_\ell^{(v)})}{\sigma_\ell} \leq \frac{d_\ell^k}{\sigma_\ell}, R_\ell^{(u)} \in R^{(u)}, R_\ell^{(v)} \in T \right\}$$

表 1 被験者の情報

Table 1 Subjects of the experiments

項目	期間 1	期間 2
実験日時	2018 年 4 月 19 日 12 時 40 分から 2 時間	2018 年 7 月 26,27,30 日 11 時 00 分から 7 時間
人数	31 名	114 名
男女比	男性 26 名 女性 5 名	男性 77 名 女性 44 名
年齢	18~51 歳	18~70 歳
場所	本学教室	本学体育館

と定める。 $\frac{d_\ell^k}{\sigma_\ell}$  はテンプレートの中で小さい順から  $k$  位の特徴量とする。 $N(R^{(u)})$  の中で最も多く含まれているユーザがテストデータの持ち主であるとする。

## 4. 評価実験

### 4.1 実験目的

本実験の目的を以下に示す。

- Kinect から得られるスケルトンデータと DTW を用いた提案識別手法において、各特徴量毎の精度を明らかにする。
- 複数特徴量の統合手法の精度を評価し、最も有効な手法を明らかにする。
- 統合する特徴量数  $c$  による精度の変化、最適な値を明らかにする。
- 被験者数  $n$  による精度の変化を明らかにする。
- 深度センサから得られる関節の部分集合で識別に有効な関節の最適な組み合わせを求める。
- 箱を持つなどいくつかの外乱により識別精度がどの程度変化するかを検証する。

### 4.2 実験内容

#### 4.2.1 データ収集

Microsoft Kinect V2 を用いて本学の学生・職員 145 名の歩行時の関節データを正面から撮影する。実験に関する情報を表 1 に示す。期間 2 のうち 111 名からは通常歩行に加え、それぞれ大股歩く、箱を持ちながら歩く、スマートフォンを操作しながら歩くという 3 つの外乱のうちランダムで 1 つを行ってもらった。

### 4.3 提案手法の精度評価

以下の 3 項目について精度評価を行う。評価基準には FRR を用いる。FRR は本人拒否率の意味で、 $FRR = \frac{\text{他人と識別された回数}}{\text{識別試行回数}}$  により求める。

#### 4.3.1 特徴量を単体で用いた際の FRR

Kinect により得られる 25 個の関節のうち、 $a_c$  として用いている SpineMid 以外の 24 個の関節それぞれの軌跡について DTW を用いて距離を計算する。計算した DTW 距離

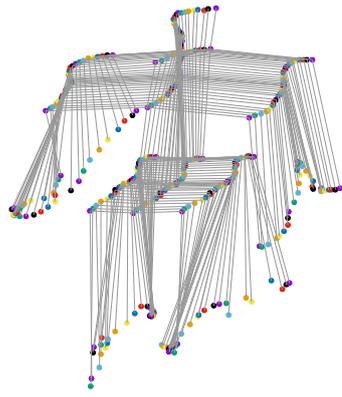


図 1 1 サイクルのスケルトンデータ  $a(t)$  の変化

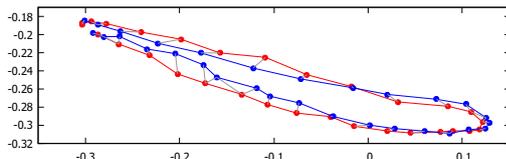


図 2 関節 HandTipLeft の DTW 距離 (本人間) ( $d(R_{HTL}^{(u)}, R_{HTL}^{(u)})$ )

を用いて EER と、他人同士のデータで計算した DTW 距離の SD を求める。EER とは当誤り率の意味で、他人受入率と FRR が一致するように閾値を設定したときのエラー率である。

#### 4.3.2 複数特徴量を組み合わせた際の FRR

複数の特徴量を組み合わせた際の FRR を計算する。特徴量の組み合わせ方は SD による重み付きユークリッド距離と  $k$ -NN の 2 パターンを行う。

### 4.4 実験結果

#### 4.4.1 データ収集

収集したスケルトンデータ  $a(t)$  の一部を図 1 に示す。24 個の関節のうち主要な 11 個 (Head, SpineShoulder, ShoulderRight/Left, HandTipRight/Left, SpineBase, HipRight/Left, FootRight/Left) のみをプロットしている。これは 24 歳男性の歩行であり、頭を左右に振りながら歩いている特徴がわかる。本例は、 $t_1, \dots, t_{33}$  の 1.1 秒が 1 サイクルであった。

#### 4.4.2 各関節の DTW 距離の算出

例として関節 HandTipLeft の DTW 距離の算出過程の結果を図 2, 図 3 に示す。赤線が被験者 1, 青線が被験者 2 の関節の動きの 3 次元座標の 2 次元のみの軌跡であり、DTW のマッチしている時系列データ間を灰色線で結んでいる。図 2 は被験者 1 と 2 が同一人物同士で比較した場合、図 3 は異なる人物間で比較した場合の図である。図 2 の DTW 距離は 0.45, 図 3 の DTW 距離は 12 であり、大きく差があることがわかった。

例として同人物と他人の歩行との HandTipLeft (HTL) についての DTW 距離  $d(R_{HTL}^{(u)}, R_{HTL}^{(v)})$  の 31 人の 5 回の歩

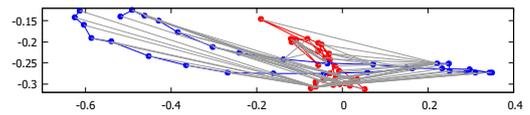


図 3 関節 HandTipLeft の DTW 距離 (他人間) ( $d(R_{HTL}^{(u)}, R_{HTL}^{(v)})$ )

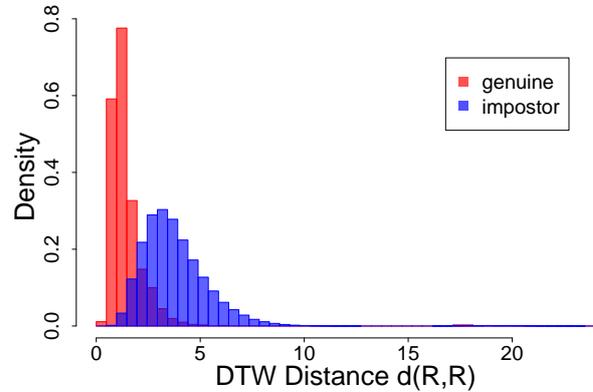


図 4 HTL の DTW 距離の分布

表 2 DTW 距離の統計量

	EER	SD		EER	SD
ElbowL	0.10	0.61	AnkleR	0.13	1.63
ElbowR	0.11	0.64	HandR	0.14	1.58
ShoulderR	0.11	0.38	HipL	0.14	0.33
ShoulderL	0.12	0.40	HandTipR	0.14	1.83
HandL	0.12	1.44	WristR	0.14	1.30
KneeR	0.12	1.07	Head	0.14	0.55
WristL	0.12	1.19	KneeL	0.14	1.14
Neck	0.12	0.31	FootL	0.14	1.79
HandTipL	0.13	1.66	SpineBase	0.14	0.33
SpineShoulder	0.13	0.23	AnkleL	0.14	1.62
HipR	0.13	0.36	ThumbL	0.19	1.48
FootR	0.13	1.78	ThumbR	0.19	1.57

行の分布を図 4 に示す。本人の変動 (赤) は他人との距離 (青) よりも短く、分散も小さい傾向が観察できる。この 2 つの分布が離れているほど、識別に適している関節である。図で濃く表示される面積が False Acceptance Rate (FAR) と FRR の和を表す。HTL 以外の他の部位についてもほぼ同様の分布をしていた。こうして求めた関節毎の EER とその SD を EER でソートして表 2 に示す。ここで SD は他人同士の DTW 距離のみの SD である。

#### 4.4.3 特徴量を単体で用いた際の識別精度

特徴量を単体で用いた際の FRR を表 3 に示す。すべての関節において外乱が加わることで FRR が悪化した。各外乱ごとに最も FRR が小さいところに下線を引いた。

#### 4.4.4 複数特徴量を統合した場合の精度

3.4.1 節の重み付きユークリッド距離 (Euc) を用いたときの FRR と 3.4.2 節の  $k$ -NN を用いたときの FRR を 4 に示す。重み付けユークリッド距離は ElbowLeft, ElbowRight,

表 3 特徴量を単体で用いた際の FRR

	normal	big	box	phone
SpineBase	0.60	0.80	0.94	0.82
Neck	0.57	0.78	0.93	0.82
Head	0.59	0.78	0.97	0.95
ShoulderLeft	0.45	0.85	1.00	0.89
ElbowLeft	0.35	0.91	0.98	0.85
WristLeft	0.45	0.95	1.00	0.90
HandLeft	0.43	0.95	0.98	0.88
ShoulderRight	0.47	0.86	0.97	0.92
ElbowRight	<u>0.33</u>	0.93	1.00	0.99
WristRight	0.45	0.91	0.98	0.96
HandRight	0.46	0.91	1.00	0.94
HipLeft	0.55	0.76	0.96	0.79
KneeLeft	0.44	0.84	0.84	0.62
AnkleLeft	0.52	0.88	0.83	0.64
FootLeft	0.52	0.82	0.84	0.66
HipRight	0.53	0.81	1.00	0.85
KneeRight	0.42	<u>0.75</u>	0.88	<u>0.60</u>
AnkleRight	0.49	0.82	0.85	0.69
FootRight	0.52	0.86	<u>0.82</u>	0.74
SpineShoulder	0.57	0.78	0.94	0.83
HandTipLeft	0.42	0.95	0.99	0.88
ThumbLeft	0.64	0.95	1.00	0.97
HandTipRight	0.47	0.89	1.00	0.93
ThumbRight	0.69	0.94	0.99	0.95
平均	0.50	0.86	0.95	0.84

表 4 重み付きユークリッド距離と  $k$ -NN の FRR

	normal	big	box	phone
Euc( $c = 5$ )	<u>0.07</u>	0.73	0.99	0.88
$k$ -NN( $k = 40$ )	0.12	<u>0.67</u>	<u>0.78</u>	<u>0.52</u>

KneeRight, FootRight, Neck の 5 つを組み合わせただけで、 $k$ -NN は  $k = 40$  の場合である。重み付きユークリッド距離は通常歩行の場合は単体で用いたときよりも FRR が小さくなったが、外乱を含む歩行ではほぼ単体時と同じであった。 $k$ -NN は通常歩行は重み付きユークリッド距離よりも FRR が大きくなったが、外乱を含む歩行では FRR が単体時、通常歩行を大きく下回る結果となった。

#### 4.5 考察

各特徴量の SD を用いて重み付けしたユークリッド距離を用いることで EER が小さくなった。全関節に一律なユークリッド距離では、特徴量ごとに大きさが異なっていたため値の大きな特徴量に引っ張られてしまうが、SD の逆数を重みとすることで、すべての特徴量の値がほぼ公平に扱われるため、誤差率 EER を低下させる効果が生じたと考えられる。

外乱に対する  $k$ NN を用いた手法でも複数の特徴量を用いることで精度が向上した。これは、正規化した DTW 距離でソートし、上位のデータのみを用いることで、手など

外乱に対して大きな影響を受けてしまう関節ではなく、膝など、外乱を含んでいても比較的安定して識別が可能な関節を用いて認証が行えたからである。

## 5. おわりに

本研究では、モーションキャプチャセンサにより人の歩行を観測し、得られた関節の動きに対して DTW を適用することにより人物識別を行う手法を提案した。実験の結果、通常の歩行では各関節の SD で重みを付けたユークリッド距離を用いた場合では、5 つの特徴量を組み合わせることで FRR が 0.07 まで向上するが、外乱を含む歩行ではそれぞれの特徴量を単体で用いた場合よりも悪化してしまうという結果になった。対して  $k$ -NN を用いた場合は、通常歩行は重み付きユークリッド距離よりも精度は低下するが、外乱を含む歩行に対しては頑強となることを示した。

## 参考文献

- [1] M. Muaaz, R. Mayrhofer, "Smartphone-Based Gait Recognition: From Authentication to Imitation", *IEEE Transactions on Mobile Computing*, Vol. 16, pp. 3209-3221, 2017.
- [2] D. Berndt, J. Clifford, "Using Dynamic Time Warping to Find Patterns in Time Series", *The Third International Conference on Knowledge Discovery and Data Mining*, pp.359-370, 1994
- [3] JointType Enumeration - MSDN - Microsoft 入手先 (<https://msdn.microsoft.com/ja-jp/library/microsoft.kinect.jointtype.aspx>) (参照 2017/8/10)

## 業績

- (1) 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", コンピュータセキュリティシンポジウム (CSS 2017), pp.972-979, 2017.
- (2) 脇 一史, 森 駿文, 菊池 浩明, "CNN を用いた顔認証システムの開発と追跡停止に対する評価", 情報処理学会第 80 回全国大会, Vol. 3, pp.543-544, 7W-03, 2018.
- (3) 森 駿文, 菊池 浩明, "歩容データの DTW 距離に基づく個人識別手法の提案と外乱に対する評価", マルチメディア, 分散, 協調とモバイルシンポジウム (DICOMO 2018), pp.672-680, 2018.
- (4) Takafumi Mori, Hiroaki Kikuchi, "Person Tracking Based on Gait Features from Depth Sensors", *The 21st International Conference on Network-Based Information Systems (NBIS-2018)*, pp.743-751, 2018.
- (5) 森 駿文, 菊池 浩明, "歩容データの DTW 距離に基づく個人識別における複数部位のフュージョン手法", コンピュータセキュリティシンポジウム (CSS 2018), pp.1-8, 2018.
- (6) 森 駿文, 菊池 浩明, "複数の歩容特徴量の  $k$  近傍による「歩きスマホ」にロバストな個人識別手法の提案", 暗号と情報セキュリティシンポジウム (SCIS 2019), 2019.
- (7) Takafumi Mori, Hiroaki Kikuchi, "Person Identification based on DTW Distance of Gait sequence", *5th International Conference on Information Systems Security and Privacy*, 2019. (to be presented)