

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

森 駿文¹ 菊池 浩明²

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.

Person Identification method based on DTW Distance of Gait sequence and its Evaluation of Robustness against Obstacles

TAKAFUMI MORI¹ HIROAKI KIKUCHI²

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. しかし, 2017年から施行された改正個人情報保護法では, カメラで取得した顔画像は個人識別符号に分類されており, 種々のデータ取得の取得の同意やオプトアウトなどのプライバシーに関する配慮を行う必要がある. これらの配慮を欠いて大規模な実験を計画したが大きな批判を受けた例もあった [1].

そこで, 本研究では, 顔認証に代わる生体認証手法として歩容に注目する. なぜならば歩容には, 精度がさほど高くなく, 対象者の心理的抵抗が少ないなどといった特徴があるためである.

歩容を用いた生体認証には, 表1に示されるいくつかの手法が提案されている. 我々が2017年に提案した手法 [2] では, 身体のいくつかの2箇所の関節間の距離を特徴量として個人を識別した. 例えば, 左右の手の間の距離はその

人の手の振り方の特徴を表しており, 個体差が生じることを応用した方法である. 10人の被験者を用いた実験により, 両足の脛長と腿長の統計量からなる4次元のベクトルを特徴量とすることがEER(等価エラー率)の観点で最適であることを示した. しかし, 最適の時でさえ0.25の誤差率であった. 誤差の主な原因としては,

- 左右の手間の距離は変動が大きく, 同一人物でも歩く毎に変化する.
- 本質的に静的な特徴量であり, 身体や手足長が類似の被験者を識別不能である.

ことが考えられた.

そこで, 我々は, これらの問題に対して, 時系列データのあいまいなパターン認識を行うDynamic Time Warping(DTW)を導入する. DTWは, 時間軸上での関節の変動パターンを効率よく比較することを可能とするので, [2]の統計量の様な静的な特徴量からは分からない手足の動的な振り方を識別し, 精度の向上が期待できるためである. DTWを用いた最新の研究には, Muaazらの[3]が挙げられる. 彼らは, スマートフォンの加速度センサのデータに対しDTWを用いることで認証を行っている. そのため, 対象者がセンサを所持する必要があるためセンサ数が限られ, また, 対象者の協力が不可欠である. スマートフォンなどのセンサはポケットなどに装着することを仮定するた

¹ 明治大学大学院先端数理科学研究科
Graduate School of Advanced Mathematical Sciences, Meiji University

² 明治大学総合数理学部
School of Interdisciplinary Mathematical Science, Meiji University

表 1 先行研究との違い

| | Muaaz[3] | 森 [2] | 本研究 |
|---------|----------|-------|------|
| 特徴数 | 1 | 1-36 | 1-24 |
| センサ | 内側 | 外側 | 外側 |
| 観測期間 | ○ | △ | △ |
| テンプレート数 | 複数 | 単数 | 単数 |
| 対象者の協力 | 要 | 不要 | 不要 |
| 手法 | DTW | 統計量 | DTW |
| 被験者数 | 35 人 | 10 人 | 31 人 |

め、手の振り方や足の運び方などの特徴を見ることができない。

本研究は、Kinect などのモーションキャプチャセンサを用いて身体の複数の関節の動きを外側から測定し、その時系列データに DTW を適用する方式を提案する。Muaaz らの方式は単一のスマートフォンから EER=0.13 を達していたが、我々の提案方式は任意個の身体の部位の変化を統合して個人を識別できるので精度の向上が期待できる。被験者にセンサを装着してもらう必要もなく、負担が小さい。以上の先行研究との比較を表 1 に整理する。

本稿では、3 章において、DTW 距離を定義し、個人識別方式を提案する。提案方式は複数の関節の部位の動きを統合するものだが、どの関節をいくつまで組み合わせればよいのかは自明ではない。そこで、提案方式を Kinect を用いて実装し、31 人の被験者を用いた実験結果に基づいて最適なパラメータを同定し、4 章にて報告する。更に、5 章においては、歩容に固有の問題である荷物や更に履き物によって生じる外乱に対する提案方式の精度低下を調査する。最後に、実験の誤差の原因や普遍的な個人識別とするために必要な条件などを 6 章で考察する。

2. 関連研究

2.1 外乱を考慮した顔認証システム

我々は、マスク、帽子、サングラスなどの外乱を含んだ場合の顔認証の精度を検証した [4]。単体の外乱では帽子が最も識別率を下げる原因であるという結論になった。

2.2 DTW[5]

DTW(Dynamic Time Warping) 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、時間軸を取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 $d(P, Q)$ は以下の様に定義する。

$$d(P, Q) = f(n_p, n_q)$$

ただし、 $f(i, j)$ は次の様に再帰的に定める。

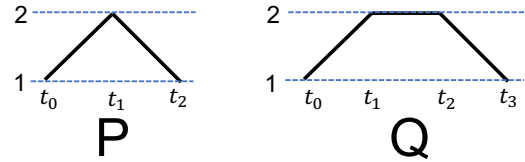


図 1 例に用いる DTW の波形

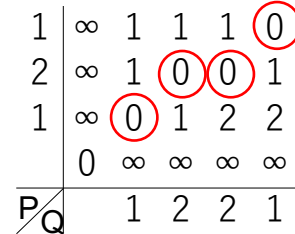


図 2 DTW の計算例

$$f(i, j) = \|p_i - q_j\| + \min(f(i, j - 1), f(i - 1, j), f(i - 1, j - 1)) \quad (1)$$

$$f(0, 0) = 0, f(i, 0) = f(0, j) = \infty$$

DTW の例として、図 1 の 2 つの時系列データ $P = (p_1, p_2, p_3) = (1, 2, 1)$, $Q = (q_1, q_2, q_3, q_4) = (1, 2, 2, 1)$ の類似度を考える。Q の時刻 t_2 における値を P では損失したと考えれば、両者はほぼ等しいデータである。

式 1 に代入すると以下の様になる。

$$\begin{aligned} d(P, Q) &= f(3, 4) \\ &= \|p_3 - q_4\| + \min(f(i, j - 1), f(i - 1, j), f(i - 1, j - 1)) \\ &= |1 - 1| + \min(f(3, 3), f(2, 4), f(2, 3)) \\ &= \dots = 0 \end{aligned}$$

この式に従って計算すると、 $f(i, j)$ の算出に 3 通りの式の算出があり、 i か j が 0 になるまで繰り返すので、最大で $3^{\max(n_p, n_q)}$ 回かかる。しかし、同じ式を何度も重複して算出していて効率が悪いので、ボトムアップに $i = 1, j = 1$ から計算すれば、 $3 \cdot n_p \cdot n_q$ 回で済む。図 2 に、この計算過程を図示する。赤丸の要素が 1 式の min で選択されている f の要素である。結局のところ、この例は、 $f(1, 1), f(2, 2), f(2, 3), f(3, 4)$ の経路が最小であり、 $D(P, Q) = 0$ である。

2.3 加速度センサを用いた歩容認証なりすまし実験

Muaaz ら [3] 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 端末の加速度情報の特徴量としており、歩行 1 サイクル分を 1 つのテンプレートとし、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。

本研究の特徴的な点は、通常の他人受入率に相当する誤りを、意図せず他人と識別されてしまう zero-effort attack と位置づけ、更に、歩行者の動きを見て真似る minimal-effort attack を導入したところにある。zero-effort attack では 35 人の被験者を用いて EER=0.13, minimal-effort attack では 5 人の被験者を用いて FAR=0 と評価している。

本実験に基づき、指紋や虹彩など他の生体認証と比べて、時系列データを用いる歩容による認証は、意図的になりすましに対して頑強であることが示された。

2.4 Kinect を用いた個人識別手法

我々は、[2] にて、Kinect を用いて身体の部位の 3 次元座標におけるスケルトンデータを測定し、歩行 1 サイクル分の最大値、平均値、中央値などの統計値の差分によって定めている 25 個の関節間の距離と角度を 36 個の特徴量とする個人識別方式を提案した。この方法では、特徴量を組み合わせたベクトルのユークリッド距離を用いて本人を認識する。10 人の被験者を用いた実験により、EER=0.25 であることを報告した。組み合わせる特徴量増やしても、EER を 0.25 より小さくすることはできなかった。部位間の距離を特徴量とすると、両手の間の様に変動が大きく、意図的に量を操作できる特徴量が含まれてしまい、十分な精度が得られないことを明らかにした。

3. 提案手法

本研究では、モーションキャプチャーセンサなどから得られた関節の 3 次元座標を測定し、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 4 つのステップから成る。

- (1) サイクル切り出し
- (2) 関節座標の相対座標化
- (3) DTW 距離の計算
- (4) 本人識別

3.1 1 サイクル切り出し

身体の部位 (図 4 参照) ℓ の時刻 t における 3 次元空間の絶対座標を $a_\ell(t) = (x, y, z)$ とする。ただし、原点はモーションキャプチャーセンサの位置とする。測定時間の絶対座標の時系列データ $\{a_\ell(t_1), a_\ell(t_2), \dots\}$ をスケルトンデータと呼ぶ。

スケルトンデータから歩行の 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

$$\Delta(t) = \text{sign} \|a_{RF}(t) - a_{LF}(t)\|$$

により計算する。ここで、 sign は $\{+1, -1\}$ の値を取る符

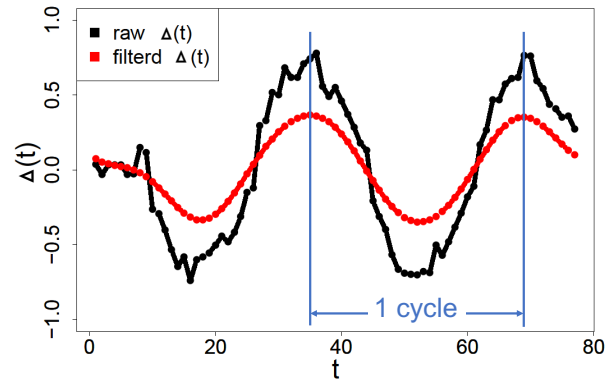


図 3 両足間の距離 (黒) とそのローパスフィルタ後の距離 (赤)

号であり、右足が前の状態を正とする。

次に、両足間の距離 $(\Delta(1), \dots, \Delta(n))$ の時系列データにフーリエ変換を適用し、全周波数成分の $1/30$ の低周波数成分のみを残して、残りを 0 とする。すなわち、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。以上の処理を行って求めたフィルタ前後の両足間距離の変化を図 3 に示す。

ここで時刻 t は、モーションキャプチャーのフレームレートに対応した単位である。例えば、フレームレート 30fps のセンサでは、 $1/30[s]$ の間隔を示しており、図 3 は約 2.6 秒の変化である。小さく変動し、高周波成分を含んでいる $\Delta(t)$ が、ローパスフィルタにより、ゆるやかな変化の時系列データ (赤色) に変換されていることが分かる。このピーク (時刻 37 と 70) を抽出し、それを (t_1, \dots, t_{35}) の 1 サイクルの時系列データと正規化する。

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 ℓ 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。身体の中心 c は 4 章の実験では図 4 の Spine-Mid (脊椎の中心) を用いる。

3.3 DTW 距離の計算

類似度は DTW を用いて定める。本手法では 1 つの関節につき xyz 座標の 3 軸が存在するため、MD-DTW (Multi-Dimensional Dynamic Time Warping)[8] を用いる。MD-DTW は 2.2 節のノルムとして、3 次元ベクトルのユークリッド距離

$$\|p_i - q_j\| = \sqrt{(p_{i,x} - q_{j,x})^2 + (p_{i,y} - q_{j,y})^2 + (p_{i,z} - q_{j,z})^2}$$

を用いる。

1 サイクルの関節 ℓ 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. DTW なので $n = n'$ である必然性はないが, 3.1 のサイクル切り出しで正規化はされているため, ほぼ同じ長さに揃っている.

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 ℓ と m についての 2 つの時系列データ (R_ℓ, R_m) と (R'_ℓ, R'_m) があるとき, 統合 DTW 距離 $D((R_\ell, R_m), (R'_\ell, R'_m))$ は, ℓ と m についての DTW 距離の L2 ノルム (ユークリッド距離), すなわち, $\sqrt{d(R_\ell, R'_\ell)^2 + d(R_m, R'_m)^2}$ とする. 同様に, k 種の関節を統合した場合も, k 次元のユークリッド距離で類似度を定める. (どの k 個の特徴的な関節を定めるかは 4 章の実験で述べる.)

3.4 本人識別

U を全ユーザの集合とする. あるユーザ u の k 個の特徴的な関節から成る, 正規化された相対座標の時系列データ (R_1, \dots, R_k) を $\mathbf{R}^{(u)}$ とする. 1 サイクル分の正規化時系列データが s 個分 $(\mathbf{R}_1^u, \dots, \mathbf{R}_s^u)$ 与えられたとき, その 1 つをテンプレート $\mathbf{R}_*^{(u)}$ とする.

時系列データ $\mathbf{R}^{(u)}$ と $\mathbf{R}^{(v)}$ の統合 DTW 距離 $D(\mathbf{R}^{(u)}, \mathbf{R}^{(v)}) \leq \theta$ であるとき, $u = v$ と判断する.

閾値 θ_ℓ^* は, 次の様にして定まる EER によって定める. u の時系列データの集合を $W^{(u)} = \{\mathbf{R}_1^{(u)}, \dots, \mathbf{R}_s^{(u)}\}$ とする. このとき, 本人拒否率 FRR は,

$$FRR(\theta, u) = \frac{|\{\mathbf{R}^{(u)} \in W^{(u)} | D(\mathbf{R}^{(u)}, \mathbf{R}_*^{(u)}) > \theta\}|}{|W^{(u)}|}$$

$$FRR(\theta) = \frac{1}{|U|} \sum_{u \in U} FRR(\theta, u)$$

$$FAR(\theta, u) = \frac{|\{R \in W/W^{(u)} | D(R, \mathbf{R}_*^{(u)}) \leq \theta\}|}{|W^{(u)}|}$$

ここで, W は全ユーザの全時系列データの集合である. $FAR(\theta_\ell^*) = FRR(\theta_\ell^*)$ となる閾値 θ_ℓ^* についての平均誤差率 EER とする.

4. 実験 1(提案手法の精度評価)

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の最適なパラメータ (関節数 k , 上位 k 個の関節, 閾値 θ^*) を同定する.
- 提案手法の基本的な識別精度を明らかにする.

4.2 モーションキャプチャデバイス

データ取得には Kinect v2 を用いる. Kinect v2 は, Microsoft 社から発売されたモーションキャプチャデバイス

表 2 Kinect v2 の仕様

| 属性 | 値 |
|-----------|-----------------|
| RGB 解像度 | 1920×1080 pixel |
| 深度 解像度 | 512×424 pixel |
| フレームレート | 30 [fps] |
| 同時測定可能人数 | 6 人 |
| 同時測定可能骨格数 | 6 人 |
| 推定可能な関節数 | 25 個 |
| 推定可能距離 | 0.5~4.5m |

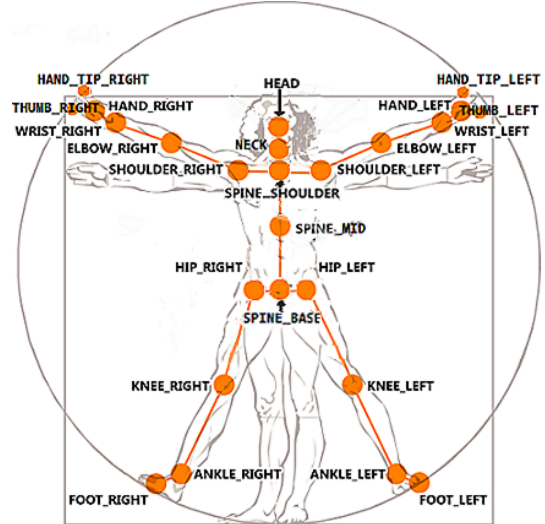


図 4 Kinect v2 により認識可能な 25 個の関節 [7]

である. ユーザの身体の動きによりゲーム中におけるプレイヤーの操作を行うことを主目的として設計されている. NUI(Natural User Interface) と呼ばれるインタフェースの一つである.

Kinect には RGB カメラ, 深度センサ, マイクなどが搭載されており, 体の関節の 3 次元座標を推定し, その変化に基づいて人の動きを認識する. 関節の座標データの時系列データであるスケルトンデータは Kinect SDK を用いて参照することができる. Kinect v2 の仕様を表 2[6] に, Kinect v2 により推定できる関節の情報を図 4 に示す.

4.3 実験方法

4.3.1 概要

Kinect v2 を用いて通常の歩行データを収集し, 提案手法の識別率を求める. 被験者は本研究室の学生, 教員である. 実験に関する情報を表 3 に示す. 実験 1 の 31 人の被験者の被験者 ID を $U1 \sim U31$ とする

4.3.2 データ収集

図 5 に示す実験環境において, 直進歩行中のスケルトンデータ $(a_1(t), \dots, a_{25}(t))$ を正面から測定する. Kinect は床から 0.9m の位置に水平を向いて固定する. Kinect より 5.5m の離れた位置から歩行を開始し, 1m の位置で終了する. 歩容の測定は 4.5m 地点から 2m 地点までの区間で行う. 一人あたり 5 回測定する.

表 3 被験者の情報

| 項目 | 実験 1 | 実験 2 |
|--------|-----------------|-----------------|
| 実験日 | 2018 年 4 月 19 日 | 2018 年 3 月 26 日 |
| 実験開始時刻 | 12 時 40 分 | 19 時 00 分 |
| 実験終了時刻 | 14 時 50 分 | 21 時 15 分 |
| 人数 | 31 人 | 5 人 |
| 男女比 | 男性 26 人:女性 5 人 | 男性 5 人 |
| 測定回数 | 5 回 / 人 | 2 回 / 人 × 外乱 |
| 年齢 | 18~51 歳 | 21~24 歳 |
| 場所 | 本学教室 | 本研究室 |

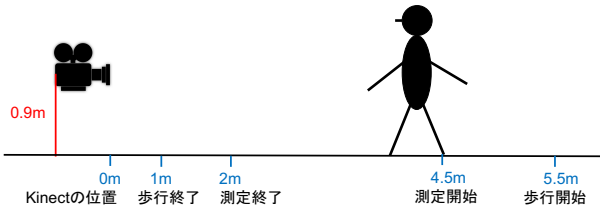


図 5 実験 1 の実験環境

4.3.3 各関節の DTW 距離の算出

本実験では Kinect から取得した関節座標のうち Spine-Mid を中心 c とする。 c 以外の 24 個の関節それぞれについて c を原点とした相対座標を計算する。それぞれの関節 ℓ について DTW 距離 $d(R_\ell^{(u)}, R_\ell^{(v)})$ を計算し、EER となる閾値 θ_ℓ^* を求める。

4.3.4 統合 DTW 距離の算出

24 個の関節を、4.3.3 節で求めた EER の低い順に組み合わせる。組み合わせる数 k と EER の関係を求める。

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 6 に示す。24 個の関節のうち主要な 11 個 (Head, SpineShoulder, ShoulderRight, ShoulderLeft, HandTipRight, HandTipLeft, SpineBase, HipRight, HipLeft, FootRight, FootLeft) 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見てわかる。時間は 1 サイクルに正規化している。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 7, 図 8 に示す。赤線が被験者 1, 青線が被験者 2 の HandTipLeft の動きの 3 次元座標の 2 次元のみの軌跡であり、1 式の 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 7 は被験者 1 と 2 が同一人物同士で比較した場合。図 8 は異なる人物間で比較した場合の図である。

図 7 の本人間の関節の変化は似通っており、図の灰色線分の総和で与えられる DTW 距離は $d(R_{HTL}^{(u)}, R_{HTL}^{(u)}) = 0.45$ である。従って、1 周期が 30 フレームであり距離の単位は

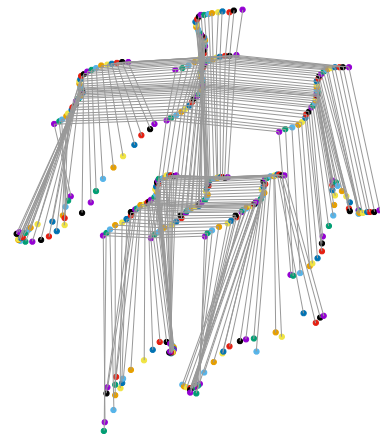


図 6 1 サイクルのスケルトンデータ $a(t)$ の変化

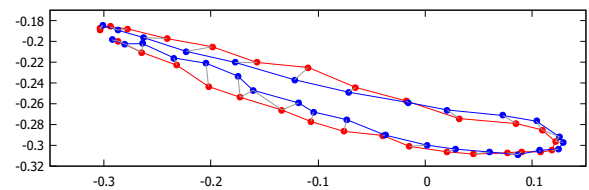


図 7 関節 HandTipLeft の DTW 距離 (本人間) ($d(R_{HTL}^{(u)}, R_{HTL}^{(u)})$)

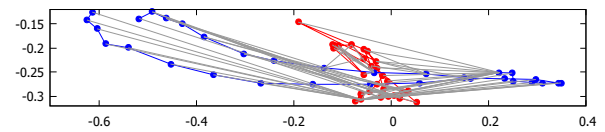


図 8 関節 HandTipLeft の DTW 距離 (他人間) ($d(R_{HTL}^{(u)}, R_{HTL}^{(v)})$)

m であるため、1/30 秒の間に左手の変動は平均 1.5cm である。

それに対して、ユーザ u (青) と異なる v (赤) との間では、どの様に時間軸を変化させても手の振りの小さい v と大きい u との間には大きな差が生じる。 $d(R_{HTL}^{(u)}, R_{HTL}^{(u)}) = 12.0$ であった。

31 人の 5 回の歩行について、それぞれの関節毎に総当たりで DTW を計算した。例として同人物の歩行と他人の歩行との HandTipLeft (HTL) についての DTW 距離 $d(R_{HTL}^{(u)}, R_{HTL}^{(v)})$ の分布を図 9 に、ShoulderLeft (SL) の分布 $d(R_{SL}^{(u)}, R_{SL}^{(v)})$ を図 10 に示す。どちらも本人の変動 (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 FAR と FRR の和を表す。両者が同面積になる DTW 距離が θ_ℓ^* であり、図 9 の HTL の関節については $\theta_{HTL}^* = 2.19$ である。HTL, SL 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 をソートして表 4 に示す。

表 4 から次の観察ができる。

- (1) 中央値は SpineBase (SB) である。身体の内中心とも一致

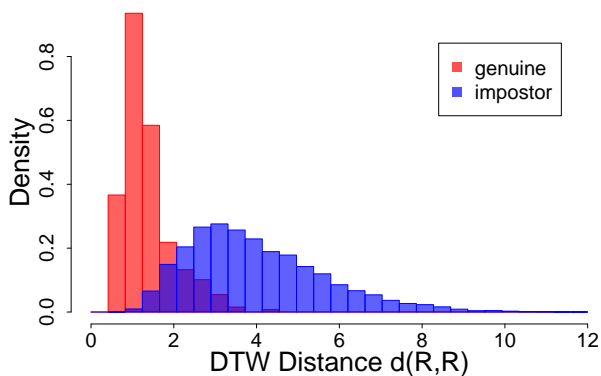


図 9 HTL の DTW 距離の分布

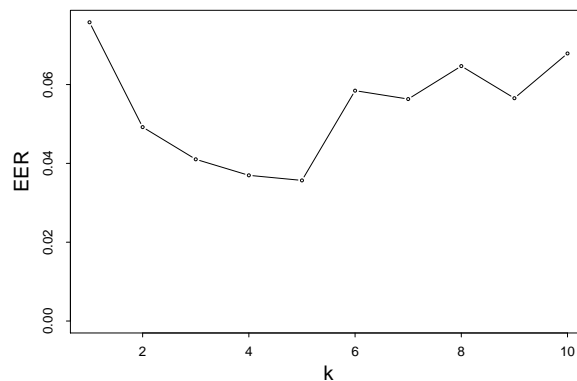


図 11 特徴量を組み合わせた際の EER

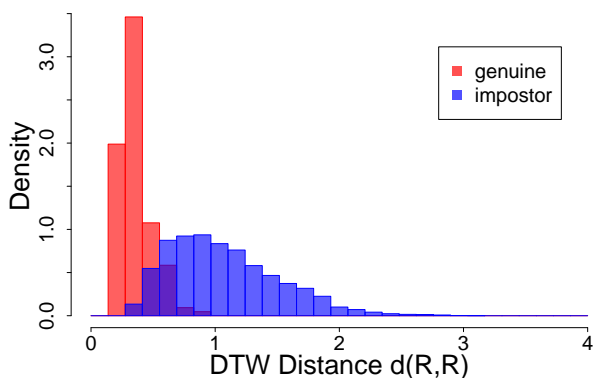


図 10 SL の DTW 距離の分布

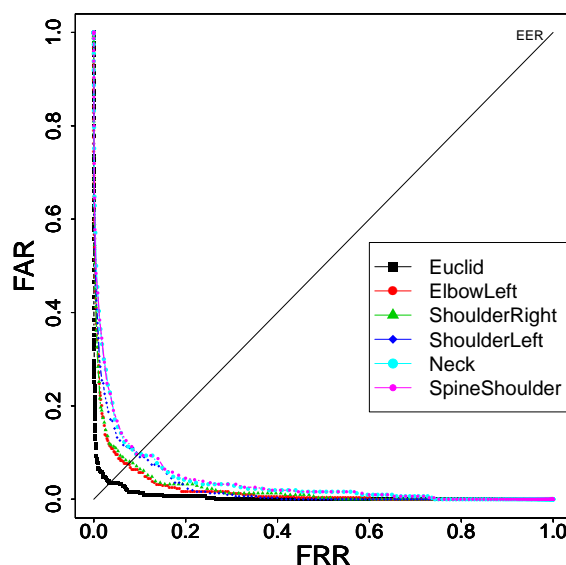


図 12 ROC カーブ

表 4 24 関節の EER

| 関節 | EER | 関節 | EER |
|---------------|-------|--------------|-------|
| ElbowLeft | 0.076 | HandRight | 0.124 |
| ShoulderRight | 0.081 | HipLeft | 0.127 |
| ShoulderLeft | 0.095 | WristRight | 0.133 |
| Neck | 0.100 | HandTipRight | 0.133 |
| SpineShoulder | 0.100 | FootRight | 0.144 |
| WristLeft | 0.107 | KneeRight | 0.145 |
| HipRight | 0.107 | AnkleRight | 0.148 |
| HandLeft | 0.108 | KneeLeft | 0.155 |
| Head | 0.110 | ThumbRight | 0.177 |
| HandTipLeft | 0.112 | ThumLeft | 0.187 |
| ElbowRight | 0.113 | AnkleLeft | 0.187 |
| SpineBase | 0.123 | FootLeft | 0.192 |

している。

- (2) 首 (Neck), 頭 (Head), 肩 (ShoulderRight/Left) は, SB より上位に位置する安定した相対位置を取る。
- (3) 腕に関する関節 (Elbow, Wrist, Hand) は, 左側が SB より上位に, 右側が下位にある。
- (4) 足に関する関節 (Foot, Knee, Ankle) は SB より下位に分類される。

4.4.3 統合 DTW 距離の算出

表 4 の上位 k 個 ($1 \leq k \leq 10$) を組み合わせた統合 DTW 距離 $D(W^{(u)}, W^{(v)})$ を用いて識別したときの EER を求

めた結果を図 11 に示す。 $k = 5$ までは EER が単調減少し, 最小で 0.036 となった。 $k = 6$ 以降はそれを下回ることにはなかった。 よって, $k = 5$ を最適値とみなし, 表 4 の ElbowLeft (EL), ShoulderRight (SR), SL, Neck (NK), SpineShoulder (SS) の関節を以降の実験で用いる。

表 4 の上位 5 つの関節を組み合わせた際の ROC カーブを図 12 に示す。 ここで, $k = 5$ の統合したときの値を Euclid (黒) で, 各部位の識別率も合わせて示している。 図の対角線が EER を表しており, 各部位だけで識別したときよりも, 統合したときが誤差を削減していることが観測できる。

平均的な被験者として U31 を選択し, U31 以外の全被験者について U31 のテンプレートとの DTW 距離をの箱ひげ図を図 13 に示す。

30 人の DTW 距離は四分位数の区間と共に表し, 平均値でソートしている。 数名似通ったユーザもいるが, 概ね識別可能な差が生じていることが分かる。

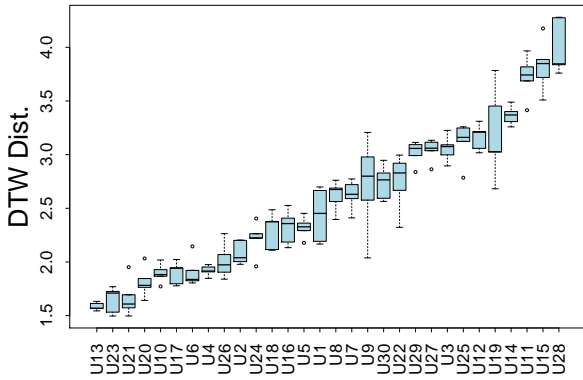


図 13 全被験者の統合 DTW 距離の分布

5. 実験 2(外乱実験)

5.1 実験目的

Muaaz[3]らは、意図的に特定の人物になりすます攻撃に対する識別精度を明らかにしたが、歩容においては、意図せずカバンを持つなどの、通常と異なる歩き方をする外乱の存在を考慮すべきである。そこで、本実験は代表的ないくつかの外乱を想定し、それに対して次を目的とする。

- カバンなどのいくつかの外乱を加えると識別精度がどの程度変化するかを検証する。
- 外乱に対して頑強な関節を明らかにする。

5.2 実験方法

カバンを持つ、サンダルを履くなど図 14 に示す例を含む 11 種類の外乱を用意し、その外乱下で歩容のスケルトンデータを採取する。外乱の種類を以下に挙げる。

- (1) 通常(外乱なし)
- (2) 大股&手大振り (b-swing)
- (3) 小股&手小振り (s-swing)
- (4) ポケットに手を入れる (pocket)
- (5) 歩きスマホ (phone)
- (6) カバンを手を持つ (hand bag)
- (7) カバンを肩にかける (shoulder bag)
- (8) リュックを背負う (zack)
- (9) 傘を持つ (umbrella)
- (10) 大きな箱を持つ (box)
- (11) サンダルを履く (sandals)
- (12) スーツケースを引っ張る (suitcase)

被験者の情報を表 3 に、撮影環境を図 15 に示す*1 通常歩行をテンプレートとし、11 種類の外乱を含む場合のデータの識別をする。 k 個の関節は実験 1 で求めたものを使用する。

*1 図 5 と測定の角度などが異なるのは実験環境の都合の為であり、他意はない。Kinect の座標を変換することで両者の差はないように正規化が可能である。



図 14 外乱 (2(大股),4(ポケット),5(スマホ),8(肩掛けカバン),9(傘),10(箱),11(サンダル),12(スーツケース)) の例

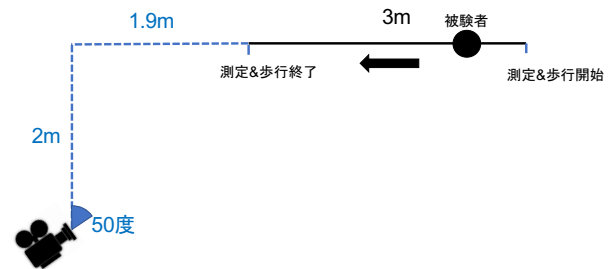


図 15 実験 2 の撮影環境 (平面図)

5.3 実験結果

5.4 外乱を含む場合の識別精度の評価

外乱を含む歩容データの DTW 距離 ($d(R_{normal}^{(u)}, R^{(u)})$) を計算した。外乱ごとの DTW 距離の平均値の一部を表 5 に示す。各関節の中で最も DTW 距離を大きくする外乱に下線を引いた。11 種類の全ての外乱について、normal よりも DTW 距離を大きくしていることが示されている。下線が轆かれている箇所が最も多い外乱は box(箱)であるが、足先 (FootRight/Left) に対しては b-swing(大股)が、肩 (SR/SL) には suitcase が大きな影響を与えている。変動の大きさも、例えば、左手 (HTL) の場合で box は 3.46 から 14.278 に拡大しており、約 4.1 倍になっている。平均すると、normal と比較した box の DTW 距離の差は 1.13 であり、2.95 倍増加している。

例として、外乱ごとの HTL の DTW 距離 $d(R_{HTL}^{(u)}, R'_{HTL}^{(u)})$ の分布を図 16 に、SR の分布を図 17 に示す。

また、外乱毎の DTW 距離の平均値 $\overline{d(R, R')}$ を図 18、関節毎の平均値を図 19 に示す。

表 5 各関節の外乱毎の DTW 距離 $d(R, R')$ の平均値

| | normal | b-swing | s-swing | pocket | phone | hand bag | shoulder bag | zack | umbrella | box | sandals | suitcase |
|---------------|--------|-------------|---------|--------|-------|----------|--------------|------|----------|--------------|---------|-------------|
| ElbowLeft | 1.83 | 5.41 | 1.83 | 3.06 | 2.36 | 2.18 | 1.86 | 1.88 | 2.48 | <u>5.61</u> | 1.68 | 2.28 |
| ShoulderRight | 0.97 | 1.76 | 0.99 | 0.96 | 1.06 | 1.31 | 1.33 | 1.14 | 1.12 | 1.60 | 0.99 | <u>1.63</u> |
| ShoulderLeft | 0.98 | 1.38 | 1.00 | 1.23 | 1.43 | 1.44 | 1.31 | 1.24 | 1.03 | 1.59 | 1.01 | <u>1.71</u> |
| HipRight | 0.95 | 1.44 | 1.13 | 1.11 | 1.13 | 1.56 | 1.62 | 1.04 | 1.18 | <u>1.89</u> | 1.13 | 1.67 |
| Head | 0.92 | 1.60 | 1.18 | 1.20 | 1.82 | 1.56 | 1.74 | 1.09 | 1.37 | <u>2.57</u> | 1.09 | 1.43 |
| HandTipLeft | 3.46 | 12.85 | 3.60 | 8.59 | 3.90 | 3.79 | 3.40 | 3.29 | 5.98 | <u>14.27</u> | 3.44 | 3.47 |
| ElbowRight | 1.84 | 4.55 | 2.28 | 2.92 | 4.20 | 3.17 | 2.76 | 2.47 | 3.52 | <u>6.90</u> | 2.15 | 5.83 |
| HipLeft | 1.09 | 1.39 | 1.16 | 1.14 | 1.14 | 1.23 | 1.50 | 1.16 | 1.27 | <u>2.21</u> | 1.10 | 1.39 |
| HandTipRight | 3.18 | 9.28 | 3.84 | 6.48 | 20.85 | 7.76 | 5.61 | 4.02 | 8.19 | <u>18.06</u> | 3.94 | 12.50 |
| FootRight | 3.31 | <u>6.91</u> | 4.39 | 3.48 | 3.68 | 3.95 | 4.12 | 4.41 | 4.03 | 4.46 | 4.13 | 4.55 |
| FootLeft | 2.96 | <u>6.32</u> | 3.89 | 3.15 | 2.88 | 3.27 | 3.21 | 3.70 | 3.69 | 4.14 | 4.06 | 3.80 |

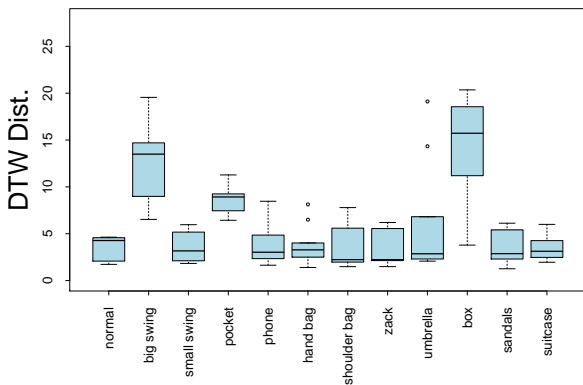


図 16 HTL の外乱ごとの DTW 距離 $d(R_{HTL}^{(u)}, R'_{HTL}^{(u)})$ の分布

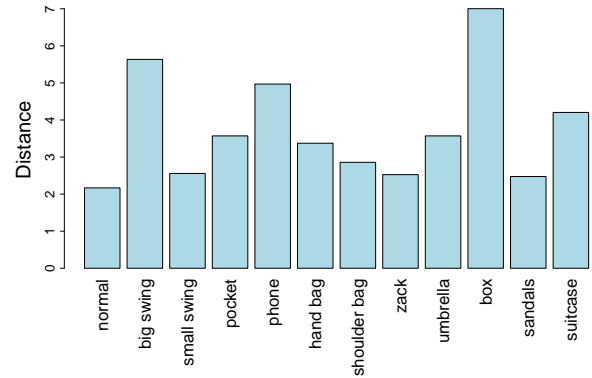


図 18 外乱ごとの DTW 距離 $d(R^{(u)}, R'^{(u)})$ の平均

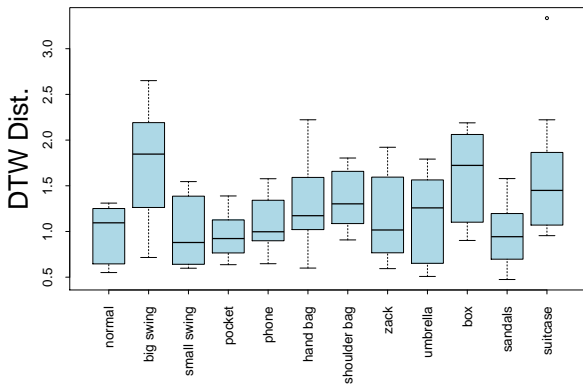


図 17 SR の外乱ごとの DTW 距離 $d(R_{SR}^{(u)}, R'_{SR}^{(u)})$ の分布

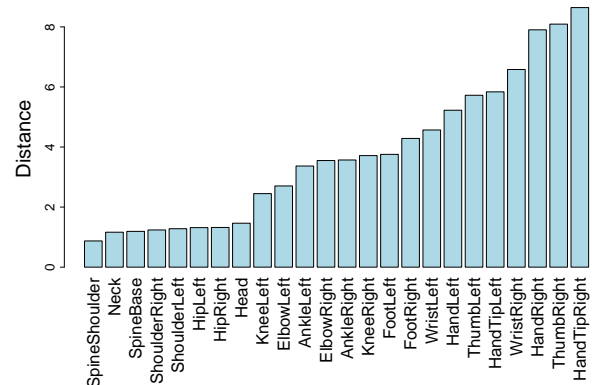


図 19 関節ごとの DTW 距離 $d(R^{(u)}, R'^{(u)})$ の平均

図 18 より、識別精度を最も低下させる外乱は大きな箱 (box) を持つことであった。また図 19 より、外乱に頑強な関節は Shoulder や Head, Hip など歩行中に大きく動かない部位であり、中でも最も頑強な特徴量は SpineShoulder であった。

6. 考察

実験 1 について、足や手などの変動が激しい関節よりも肩や首など、比較的変動の少ない関節の方が EER が低くなった。これは、他人との距離は小さいがそれ以上に本人内の変動が小さく安定していることがあげられる。DTW

により詳細に識別を行ったので、ほぼ動かないと考えられる肩や首などの関節も小さいながら周期的な動きをしていると考えられる。

表 4 より、上半身の関節は右側よりも左側の方が EER が低くなる傾向がみられた。これは、図 8 の赤線の被験者のように、左手をほとんど振らない被験者が数名いたことが原因である。手の振りが少ない場合、肩などと同様に本人の変動が少なくなり特徴量が安定するため EER が低くなったと考えられる。

図 11 について、 $k=5$ までは特徴量の次元が増えることにより他人との差異が大きくなるため FAR が単調減少し

ていたが、 $k = 6$ 以降は特徴量の次元が増えすぎることによって本人内の再現性が低下して FRR が増大するので、結果として EER が増大したと考えられる。

外乱について、大股&手大振りと大きな箱を持ったときが特に DTW 距離が大きくなっていた。また、関節毎に見た場合それにより大きな影響を受ける手や腕の DTW が大きくなる傾向となった。外乱毎の特徴としては、大股で歩く、大きな箱を持つなど日常的にあまり現れない外乱においては識別率が悪化する結果となったが、カバンを持つ、スニーカーの代わりにサンダルを履くなど日常的に頻繁に現れる外乱に対しては提案システムはある程度頑強であるといえる。従って、多様な外乱に影響されずに正確に個人を識別するためには、SpineShoulder などの変動の少ない関節を統合して DTW を適用すればよいことを主張する。

7. 結論

本研究では、深度センサなどにより得られた 3 次元の関節の時系列データに対して DTW を適用することで個人識別を行う手法を提案した。実験の結果、ElbowLeft, ShoulderRight, ShoulderLeft, Neck, SpineShoulder の 5 つの関節を組み合わせることで EER を最低で 0.03 まで下げることができた。先行研究の森ら [2] の 0.25 や Muaaz ら [3] の 0.13 を大きく上回る結果である。

提案システムを用いて、外乱を含む歩容データの DTW 距離の変化を検証した。実験の結果、肩などの変動の少ない部分が外乱の影響を受けにくいことが判明した。

参考文献

- [1] 顔認証の追跡実験延期 J R 大阪駅、市民の不安受け 入手先 (https://www.nikkei.com/article/DGXNASHC1102E_R10C14A3AC8000/) (参照 2018/5/9)
- [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・追跡手法の提案”, コンピュータセキュリティシンポジウム 2017, 2017
- [3] M. Muaaz, R. Mayrhofer, “Smartphone-Based Gait Recognition: From Authentication to Imitation”, IEEE Transactions on Mobile Computing, Vol. 16, pp. 3209-3221, 2017.
- [4] 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, 2018
- [5] 櫻井 保志, 吉川正俊, “ダイナミックタイムワーピングのための類似探索手法”, 情報処理学会論文誌, 2014
- [6] Kinect ハードウェア - Microsoft Developer 入手先 (<https://developer.microsoft.com/ja-jp/windows/kinect/hardware/>) (参照 2017/8/10)
- [7] JointType Enumeration - MSDN - Microsoft 入手先 (<https://msdn.microsoft.com/ja-jp/library/microsoft.kinect.jointtype.aspx>) (参照 2017/8/10)
- [8] G. A. ten Holt, M. J. Reinders, E. A. Hendriks, “Multi-Dimensional Dynamic Time Warping for Gesture Recognition”, Thirteenth annual conference of the Advanced School for Computing and Imaging, June 2007.