

## DTW 法を用いた動作間の類似性評価手法の検討 A Similarity Evaluation Between Two Motions Using Dynamic Time Warping

堀川 康生<sup>†</sup> 市川 嘉裕<sup>†</sup> 山口 智浩<sup>†</sup>  
Kosei Horikawa Yoshihiro Ichikawa Tomohiro Yamaguchi

### 1. はじめに

近年、スポーツやダンスなどの動画がインターネットを通じて共有されるようになり、動画内の動作を目標として真似ることで、自身の動作を改善しようとする需要が高まっている。しかし、カメラの位置や動作時間の違いが存在し、単純には比較できない問題点や、目視では定量的な動作の違いが把握できないという問題点がある。

この問題に対して、動作のタイミングの違いやあるフレームでの動作の違いを評価・フィードバックして支援を行う研究[1]が存在するが、上記の問題を全て解決することは困難である。そこで本研究では、撮影時のカメラの位置や時間が異なる場合の類似性を適切に評価することを目的とする。

そのため、音声認識の分野で幅広く使用されており、時間軸のずれに頑健な DTW 法を用いることで解決を図る。本稿では、DTW 法を用いた類似度評価手法を提案し、ビデオカメラで撮影された映像から得られる動作データについて時間的・空間的に加工を施し、それらに対して手法を適用することで有効性を検証する。

### 2. 関連手法

#### 2.1 MediaPipe Pose

MediaPipe Pose[2]とは、Google が提供しているソリューションの一つで、動画や画像中の人物の姿勢推定が可能である。具体的には、33 箇所の部位（関節や目など）の横方向  $x$ 、縦方向  $y$ 、奥行方向  $z$  の値を取得できる。 $x, y$  は  $[0.0, 1.0]$  で正規化され、 $z$  は左右の腰関節の midpoint を原点とし、カメラに近いほど値が小さく、遠いほど大きくなる。

#### 2.2 ガウス過程回帰

ガウス過程回帰とは、入力変数  $x$  から出力変数である実数値  $y$  への関数  $y = f(x)$  を推定するモデルの一つである。その特徴の一つは非線形性であり、線形回帰ではうまくフィッティングできない場合にも有効である。もう一つの重要な特徴はベイズ推定を用いる点である。推定される関数は一つの関数ではなく、関数の分布として得られるため、推定の不確実性を表現することができる[3]。

#### 2.3 Dynamic Time Warping

Dynamic Time Warping (DTW)とは、速度が異なる二つの時系列データの類似性を評価するアルゴリズムの一つであり、音声認識の分野で開発された[4]。二つの時系列データ  $A = a_1, a_2, \dots, a_i, a_{i+1}, \dots, a_n$ ,  $B = b_1, b_2, \dots, b_j, b_{j+1}, \dots, b_m$  があるとき、二つの要素間の距離を  $d(i, j) = \|a_i - b_j\|$  と定義すると、DTW 距離は次式で定義される。

$$DTW(i, j) = d(i, j) + \min(DTW(i, j-1), DTW(i-1, j), DTW(i-1, j-1))$$

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

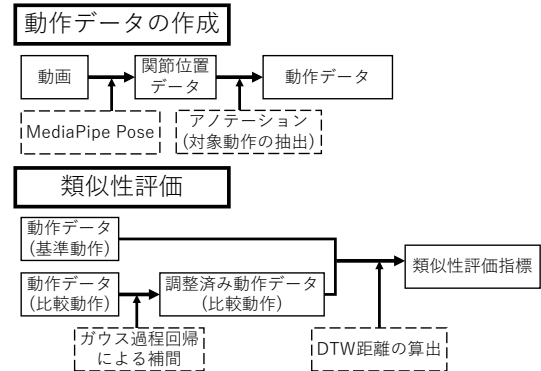


図1 提案手法の概略図

### 3. 提案手法

動作間の類似性を評価するための提案手法の概略を図1に示す。まず、動作データ生成の手順として、(1)動画からMediaPipe Poseを用いて関節位置データを取得し、(2)アノテーションを行い、対象動作中の関節位置データのみ抽出する。類似性評価の手順として、(3)対象となる二つの動画（基準動作と比較動作）からそれぞれ動作データを生成し、(4)データの長さを基準動作に揃えるため、比較動作にガウス過程回帰による補間を適用した後、(5)DTW法を用いて以下の二つの指標を取得し、類似性を評価する。

- ① DTW 距離が最小となる基準動作と比較動作のフレーム同士の対応関係
- ② 基準動作と比較動作のフレーム毎の DTW 距離指標①は横軸を基準動作のフレーム、縦軸を比較動作のフレームとする二次元のパスで表現され、始点の(0, 0)から 3 方向（横、縦、斜め）への遷移を繰り返して終点(n, m)まで繋がる形となる。縦・横に延びるパスはあるフレームが一对多の対応関係にあることを示す。これはユークリッド距離では同時刻の値で距離を算出することに対して、時間を超えて最小な距離となる適切な対応関係を表す情報である。指標②は 0 に近いほど類似性が高く、遠いほど類似性が低いことを表す。

以上を踏まえて、指標①によって動作間の時間のずれを明らかにし、指標②によって動作間の動きの大きさの違いを明らかにすることが提案手法である。

### 4. 実験

#### 4.1 内容

実験では、バスケットのフリースローを評価対象の動作として、その動画を取得し、生成した動作データに時間的・空間的な加工を施して類似性を評価する。なお、本稿では変化が最も大きい  $y$  軸方向の値のみを扱うこととした。時間的な加工として、以下の 2 種類の変更を行う。

1. 基準・比較動作のタイミングの変更
2. 比較動作のデータ点数の増減

<sup>†</sup> 奈良工業高等専門学校 National Institute of Technology, Nara College

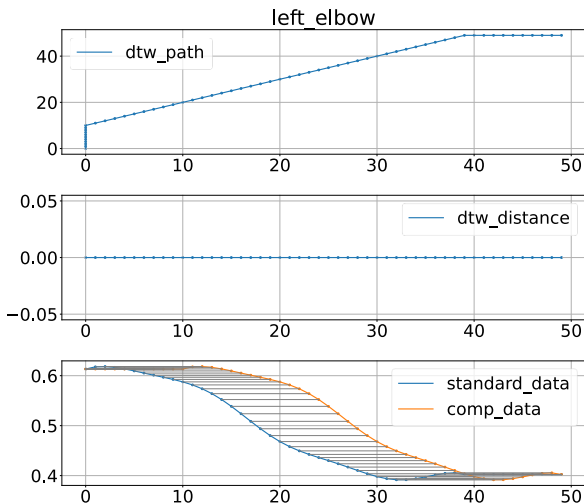


図 2 加工 1 データでの類似性評価

加工 1 では、基準動作の終点から 10 フレームに終点と同じデータを追加し、比較動作の始点より前の 10 フレームに始点と同じデータを追加した。これにより、動作のタイミングのみ異なるデータに対する類似性評価の性能を確認する。加工 2 では、基準動作のデータを基にしてデータ数を  $b$  倍にしたものを比較動作のデータとした。これにより、動画のフレームレートが異なる場合や動作の形は同じであるが速度のみ異なるデータに対する性能を確認する。なおデータを増やす加工にはガウス過程回帰を利用した。

空間的な加工として、以下の 2 種類の変更を行う。

3. 比較動作の振幅（動作の大きさ）の変更
4. 比較動作の動作データ全体の平行移動

加工 3 では、基準動作のデータを基にして、動作データの中央値を軸に、データの値を  $c$  倍したものを比較動作のデータとした。これにより、動作の大きさが異なっても、指標①の対応関係が変化しないことを確認する。加工 4 では、基準動作のデータを基にして、動作データの値を全て  $d$  だけ大きくしたものを比較動作のデータとした。これにより、カメラの撮影位置が異なる場合でも、指標①の対応関係が変化しないことを確認する。

## 4.2 結果と考察

### 4.2.1 加工 1：動作のタイミングの変更

加工 1 を施したデータでの比較結果の一例（左肘の動きについて）を図 2 に示す。なお、図 2 は上から順に指標①、指標②、指標③を実データ上に表示したものを示している。

指標①より、加工前は対応関係にある基準動作の 0~39 フレームと比較動作の 10~49 フレームがそれぞれ結びついており、指標②が常に 0 であることから、動作の大きさには違いがないことが表されていることが確認できた。

### 4.2.2 加工 2：データ点数の増減

加工 2 を施した場合の、フレーム毎の DTW 距離の平均（33 箇所部位の動作データ）を表 1 に示す。なお、指標①はフレームごとに一対一対応を示していた。結果より、DTW 距離が微小であることから、動作は同一であることが表されていることが確認できた。実際には、ガウス過程回帰によるデータ補間が時間軸を歪めていることを考慮する必要がある。

表 1 加工 2 データでの DTW 距離の平均

データ点数の増減割合 $b$	DTW
0.5	0.001816
0.75	0.001973
2.0	0.000071
3.0	0.000059

表 2 加工 3 データでのユークリッド・DTW 距離の平均

振幅倍率 $c$	ユークリッド	DTW
0.5	0.027353	0.020449
0.7	0.038294	0.029807
2.0	0.105789	0.093636
4.0	0.165693	0.153821

表 3 加工 4 データでのユークリッド・DTW 距離の平均

増加値 $d$	ユークリッド	DTW
0.1	0.1	0.084648
0.3	0.3	0.291056
0.5	0.5	0.5
0.7	0.7	0.7

### 4.2.3 加工 3：振幅（動作の大きさ）の変更

加工 3 を施した場合の、フレーム毎のユークリッド距離および DTW 距離の平均を表 2 に示す。この結果から、ユークリッド距離と比べて DTW 距離の値が小さくなっており、同一にならないことがわかった。これは、DTW 法が時間を超えて最小な距離を求めているためである。指標①の対応関係にも変化が生じることが確認された。

### 4.2.4 加工 4：動作データ全体の平行移動

加工 4 を施した場合の結果を表 3 に示す。この結果から、増加値が 0.5, 0.7 の場合にユークリッド距離と DTW 距離が同じになり、対応関係は変化していないことが確認できた。一方で、増加値が 0.1, 0.3 の場合はそのようなならず、指標①の対応関係にも変化が生じることがわかった。これは、動きの激しい関節の場合、0.1, 0.3 より大きい動作を行うため、一対一対応ではなく、時間を超えて最小となる点と結びついたからであると考えられる。

## 5. まとめと今後の課題

動作間のタイミングやデータ点数が異なるという時間的な問題に対しては提案手法が有効であるが、位置や大きさが異なる場合は想定通りにならないことが明らかになった。

後者の問題解決に加え、ガウス過程回帰を用いることで時間軸が歪んでいることを考慮して、動作の違いを示す方法を検討することが今後の課題である。

### 参考文献

- [1] 榎井 昇一, 矢吹 彰彦, 佐々木 和雄, “3d センシングと技認識を用いた体操競技の採点支援技術”, 電気学会誌, Vol.140, No.8, pp.538-541, (2020)
- [2] Google, “Mediapipe pose”, <https://google.github.io/mediapipe/solutions/pose.html>, 2022/06/20 閲覧
- [3] 赤穂 昭太郎, “ガウス過程回帰の基礎”, システム/制御/情報, Vol.62, No.10, pp.390-395, (2018)
- [4] H. Sakoe, S. Chiba, “Dynamic programming algorithm optimization for spoken word recognition”, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol.26, No.1, pp.43-49, (1978)