

身体動作データのマルチストリーム性 を考慮した類似・非類似検索

矢部 武志[†] 田中 克己^{††}

近年、センシング技術の発達により、身体動作を3次元時系列データとして取得することが可能となっている。一方、音声や数値列のような時系列データの認識に、DPマッチング法が良く知られている。DPマッチングにより時系列データ間の類似性を測ることができる。しかし、身体動作データは時間同期された多数の時系列データ、即ち、マルチストリーム型データであるので時系列データ間の類似性だけでなく、動きの同期についても考慮する必要がある。本研究では、ユーザの動作と模範となる動作の類似性を測ることでユーザの動作判定を行うことを目的としている。そのために、時系列データ間の類似度と動きの同期度を考慮して、動作の類似性を測る手法について提案する。

Similar and Dissimilar Retrievals of Human Motion Data by its Multistream Nature

TAKESHI YABE[†] and KATSUMI TANAKA^{††}

It is possible to acquire human motion as 3-dimensional time series data by development of sensing technology in recent years. Moreover, the DP matching method is well known by recognition with the time series data, such as voice and a numerical value sequence. The similarity between time series data can be measured by DP matching. However, since human motion data consists of many time series data with the time synchronization, the similarity between human motions need to be taken into consideration not only about the similarity between time series data but about a synchronization of a motion. It aims at performing a judgment of a user motion by measuring the similarity of a user motion and a model motion in this research. Therefore, in consideration of the degree of similar between time series data, and the degree of synchronization of a motion, it proposes about the technique of measuring similarity of a motion.

1. はじめに

近年、モーションキャプチャシステムに代表されるセンシング技術が発達し、身体運動のデータ（以下、モーションデータ）を3次元時系列データとして取り扱うことが可能となっている。一方、時系列データ間の類似性を測ることで音声や数値列の認識や類似検索を行う研究が多数なされている。そこで、これらの時系列データの認識・検索技術にみられる手法を応用して身体動作の認識を行うことを考える。

音声や数値列などのシングルストリーム型データに

おける認識には、DPマッチング法が良く知られている。DPマッチングにより時系列データ間の類似性を測ることができる。しかし、モーションデータはマルチストリーム型データであるので、時系列データ間の類似度だけでは、全体の動作の類似度を表わすことができない場合がある。例えば、時系列データ間の類似度だけで類似動作検索を行うと体の部位の動きは類似しているが、部位間での動きのタイミングが異なる動作が検索される可能性がある。

そこで、本研究では部位の動きの類似度と部位間の動きの同期度を考慮して、身体動作の類似性を測ることで、ユーザの動作判定を行う手法を提案する。

2. 動作判定システムの概要

本研究では、モーションデータを用いてユーザの動作と模範動作とを比較し、その差異をユーザに認識させる動作判定システムを提案する。本システムは、

[†] 神戸大学大学院自然科学研究科情報知能工学専攻
Division of Computer and Systems Engineering, Graduate
School of Science and Technology, Kobe University
^{††} 神戸大学大学院自然科学研究科情報メディア科学専攻
Division of Information and Media Sciences, Graduate
School of Science and Technology, Kobe University

ユーザの動作技術の向上を支援することを目的としている。主に動きが重要な意味をもつスポーツの分野で有用であると考え。ここでは、スポーツの動きの中でも野球、テニス、ゴルフなどのスイングの動作を対象とする。

模範動作の動作内容は基本動作である。ここでの基本動作とは、人間がみて基本的な一連の動きと判断する動きを指す。例えば、ゴルフでは構えた状態からテークバック、スイングまでの動作を指す。ユーザの動作としては、同じ動きを連続的に行った連続動作を入力とする。

2.1 処理の流れ

図1にシステムの処理の流れを示す。

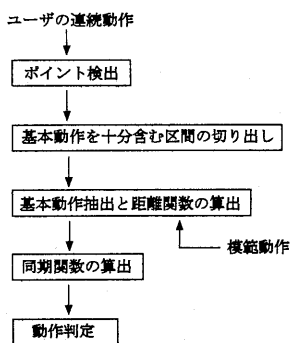


図1 システムの処理フロー

まず、ユーザの連続動作から基本動作に共通するポイントを検出する。ポイントというのは、例えばスイングの開始点、終了点などの時間を指す。次にそのポイントの情報をもとに基本動作を十分含む区間を切り出す。切り出した区間から模範動作とマッチングさせることで、基本動作抽出と時系列データ間の類似度を表わす距離関数について算出する。そして、動きの同期を表わす同期関数を算出してから距離関数と同期関数をもとに動作判定を行う。

2.2 動作の判定事項

動作判定の内容を以下に示す。

- 類似動作・非類似動作の判定
 - 全体傾向による類似・非類似判定
- 以上の項目を判定結果として出力することを考える。

3. マルチストリーム型データによる類似度

ここでは、時系列データ間の類似度を表わす距離関数と各部位の動きの同期度を表わす同期関数の算出について述べる。

3.1 DP マッチングによる基本動作抽出と距離関数

3.1.1 人間の動作特性によるポイント検出

人間の自然な動きとして、ある一つの動きは動き始めは遅く、徐々に速度が上がり最大速度を迎えると、動きが終わるまで速度が落ちていく。これはスポーツの動きについても同様である。そこで、動きの速度変化によってスイングの最大速度のポイント検出について述べる。

モーションデータは、一定時間間隔でサンプリングされた離散的なデータであるので、点間の距離を速度を表わす指標として考える。時系列のデータ数を n 個とする1次元データ S が、以下のように与えられたとする。

$$S = s_1, s_2, \dots, s_i, \dots, s_n$$

しかし、モーションデータは、3次元時系列データであるので、これを空間座標 $s_i = (x_i, y_i, z_i)$ とする。ある時間 i から $i+1$ までの距離を $v(i)$ とすると、以下のように与えられる。

$$v(i) = |s_i - s_{i+1}|$$

$$= \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2 + (z_i - z_{i+1})^2}$$

図2に2回スイングしたテニスの手先についての速度関数 $v(i)$ のグラフを示す。

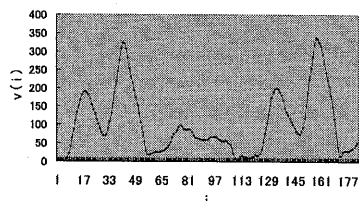


図2 右手先の速度関数

1回目の基本動作は、 $i=1$ から $i=55$ までであり、 $i=1$ から $i=30$ までがテークバックの動作で、 $i=30$ から $i=55$ までがスイングの動作である。図2より、1つの動作はグラフ的に1つの山として表れる。2回目の基本動作についても同じである。また、スイング動作の最大速度の点は他の動作に比べて速いことを利用して、スイング動作の最大速度の時間という基本動作に共通したポイントを検出できる。具体的には、速度関数の最大値を閾値に設定し、そこからスイング回数分の山が検出されるまで、閾値を下げる。そして、すべてのスイングの最大速度の時間の情報をもとに、基本動作を十分含む区間を切り出す。

3.1.2 DP マッチングとその問題点

DP マッチングは、音声認識の分野等で広く用いら

れているパターンマッチング法で、2つの時系列データの長さが異なる場合にも、その差の影響を吸収する方式である。

時系列データ n 個のパターン S と m 個のパターン T が以下のように与えられる。

$$S = s_1, s_2, \dots, s_i, \dots, s_n$$

$$T = t_1, t_2, \dots, t_j, \dots, t_m$$

以下、 s_i, t_j をデータの要素、 i, j を要素番号とする。

また、 S と T のパターン間の距離 $d(i, j)$ を次式で定義する。

$$d(i, j) = |s_i - t_j|$$

DP マッチングの計算は2次元のマッチング行列 $g(i, j)$ を用いて行う。計算手順としては、A, B, C, D の順で、 $g(i, j)$ を計算し、マッチング行列を求める。以下に計算内容を示す。

- A の計算

$$g(1, j) = d(1, j) \quad (1 \leq j \leq m)$$

- B の計算

$$g(i, 1) = d(i, 1) + g(i-1, 1) \quad (2 \leq i \leq n)$$

- C の計算

$$g(i, j) = d(i, j) + \min \begin{pmatrix} g(i-1, j) \\ g(i-1, j-1) \\ g(i, j-1) \end{pmatrix} \quad (2 \leq i \leq n) \quad (2 \leq j \leq m)$$

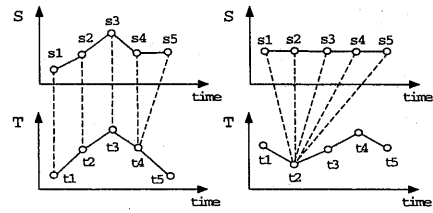
- D の計算

$$D(S, T) = \min\{g(n, j) | 1 \leq j \leq m\}$$

$g(n, j)$ は、 S の要素 s_n が T の要素 t_j と対応づけたときの最適なマッチングによる距離を表わす。また、 $D(S, T)$ の値は S と最適に対応する T の部分的なパターンとの距離であり、パターン S と T の部分的なパターンとの類似度、または距離関数を表わす。

しかし、音声の場合は、1つの時系列データであるが、モーションデータは、体の各関節につけられるマーカー^{*}が24個あるので、24本の3次元時系列データである。しかも、すべてが運動しているので、独立に考えることができない。独立に考えて、マッチングさせると、2つのパターン間の対応づけが、それぞれの体の部位で異なる。当然、始終端の対応づけも異なる。図3に2つのパターン S と T における体の部位 G, H の2つの対応づけの例を示す。

図3の(a)では、 S 全体は、 t_1 から t_4 とマッチングし、(b)では、 t_2 の一点でマッチングしている。 S に対応する T の始終端の要素も異なるし、対応する幅も



(a) Gでのマッチング (b) Hでのマッチング

図3 G,Hでの対応づけ

異なる。DP マッチングでは、2つのパターン間の距離が最小となるように対応づけるので、(b)のように一点に対応することもある。

このように、体の部位ごとに対応する始終端、または、対応幅が異なると、動きの類似度として求めることができない。よって、体の各部位ごとに対応する始終端、対応する幅をそろえる必要がある。

3.1.3 2段階DP マッチング

3.1.1節で求めた連続動作のポイントの情報に基づき、基本動作を十分含む区間で時系列データを連続動作から切り出す。

模範動作の時系列データは3.1.2節におけるパターン S 、連続動作から切り出したデータはパターン T に相当する。切り出したデータから基本動作を抽出するためにDP マッチングを用いる。また、モーションデータは、3次元時系列データであるため、2つの要素の距離 $d(i, j)$ は、空間座標 $s_i = (s_{x_i}, s_{y_i}, s_{z_i}), t_j = (t_{x_j}, t_{y_j}, t_{z_j})$ のユークリッド距離で、次式で表わされる。

$$d(i, j) = \sqrt{(s_{x_i} - t_{x_j})^2 + (s_{y_i} - t_{y_j})^2 + (s_{z_i} - t_{z_j})^2}$$

このように、3次元空間的に距離の近い要素とマッチングをするように行われる。

まず、DP マッチングにより各部位ごとに、 $g(n, j)$ が求まる。そこで、各部位の番号を $l = \{1, 2, \dots, 24\}$ とし、各部位の $g(n, j)$ を $h_l(j)$ とする。そして、要素番号 j に対して、すべての部位の $h_l(j)$ を加算する。

$$G(j) = \sum_{l=1}^{24} h_l(j) \quad (1 \leq j \leq m)$$

こうして、求められた $G(j)$ は、体全体の要素 s_n が t_j に対応づけたときの最適なマッチングによる距離を表わす。 $G(j)$ が最小である時の j を基本動作の終端とする。

次に、その終端を固定した状態で、逆方向にDP マッチングを行う。具体的には、終端の T の要素を t_b とすると、以下のように変更する。

- T の最大要素を、 t_b とする。

^{*} マーカーとは、体の各関節部分に取りつけて、モーションキャプチャーシステムが各関節の3次元の座標データとして採取するためのものである。

- A の計算の変更

$$g(1, j) = d(1, j) + g(1, j - 1) \quad (1 \leq j \leq b)$$

- 距離関数の変更

$$d(i, j) = \sqrt{(s_{x_{i'}} - t_{x_{j'}})^2 + (s_{y_{i'}} - t_{y_{j'}})^2 + (s_{z_{i'}} - t_{z_{j'}})^2}$$

ただし、 $i' = n - i$ $j' = b - j$

そして、同様の操作で基本動作の始端を見つける。このようにして、体全体に対応する始末端をそろえることで距離関数を算出する。体の各部位についての距離関数は、その始端の要素を t_a とすると、 $h_l(b - a)$ の値となる。また、基本動作の距離関数 H は以下のように与えられる。

$$H = \sum_{l=1}^{24} h_l$$

3.2 同期関数

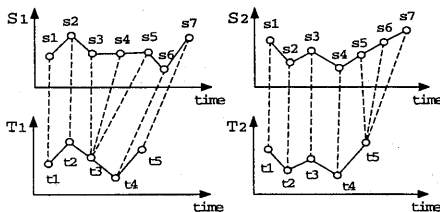
2段階のDPマッチングにより、連続動作からすべての基本動作の区間が決定される。同時に2つの動作における時系列のすべての要素は対応づけられている。この対応づけを利用して同期関数を求める。

模範動作の各部位 l が n 個の時系列データ S_l をもち、抽出されたユーザの動作の各部位 l が m 個の時系列データ T_l もつとすると以下のように表わされる。

$$S_l = s_1, s_2, \dots, s_i, \dots, s_n$$

$$T_l = t_1, t_2, \dots, t_j, \dots, t_m$$

$n = 7, m = 5$ 、のときの S_1, S_2 と T_1, T_2 の対応づけの例を図4に示す。



(a) S_1 と T_1 の対応関係 (b) S_2 と T_2 の対応関係

S_1, S_2	s_1	s_2	s_3	s_4	s_5	s_6	s_7
	↑	↑	↑	↑	↑	↑	↑
T_1	t_1	t_2	t_3	t_3	t_3	t_4	t_5
T_2	t_1	t_2	t_3	t_4	t_5	t_5	t_5

(c) S_1, S_2 と T_1, T_2 の対応関係

図4 S と T の対応関係

S_1, S_2 の要素 s_i に対応する T_1, T_2 の要素番号が同じであれば同期、異なれば非同期とみなす。図4では

s_1, s_2, s_3, s_5 に対応する要素の要素番号が T_1, T_2 において同じであるので、 s_1, s_2, s_3, s_5 の地点では T_1, T_2 は同期しているといえる。しかし、 s_4, s_6, s_7 の地点では、 T_1, T_2 において要素番号が異なるので同期していないとみなす。特に、 s_5 の地点が最も T_1, T_2 の対応する要素番号の差が大きいため、 s_5 に対応するユーザの動作が非同期であると考えられる。

そこで、模範動作の要素 s_i に対応する要素を T_1, T_2, \dots, T_{24} について求め、それぞれの要素番号のばらつきにより、同期具合を測る。ばらつきが小さいほど同期しており、ばらつきが大きいほど非同期であるとする。

模範動作の要素 s_i に対応する T_1, T_2, \dots, T_{24} の要素番号をそれぞれ u_1, u_2, \dots, u_{24} とし、その分散を求める。 u_1, u_2, \dots, u_{24} の平均値を U 、分散を r とすると以下のように与えられる。

$$U = \frac{u_1 + u_2 + \dots + u_{24}}{24}$$

$$r = \frac{(u_1 - U)^2 + (u_2 - U)^2 + \dots + (u_{24} - U)^2}{24}$$

r の値が小さければ、要素番号のばらつきが少なく体全体の同期がとれていることを表わす。

また、模範動作の複数の要素に対して、それぞれ分散を計算する。模範動作の要素を k 個指定し、それに対する分散の値が $r_1, r_2, \dots, r_q, \dots, r_k$ とし、その合計を R とすると、以下ようになる。

$$R = \sum_{q=1}^k r_q$$

この R の値を基本動作の同期を表わす同期関数の値とする。また、 r_q の値を区分 q の同期関数の値とする。

4. 類似・非類似判定

3章で述べた距離関数と同期関数を用いてユーザ動作の類似・非類似判定を行う。

4.1 距離と同期の総合判定

基本動作の距離関数と同期関数を総合した判定を行うことで、類似動作・非類似動作の判定を行う。

図5に距離関数 H と同期関数 R を軸として4回の基本動作の距離関数 H_1, H_2, H_3, H_4 と同期関数 R_1, R_2, R_3, R_4 を座標値として表わした例を示す。

しかし、基本動作の距離関数は部位の距離の総和であり、同期関数は区分による分散の総和であるので、直接原点からの距離で動作の類似度を求めることができない。値の内容が異なる距離関数と同期関数を正規化する必要がある。

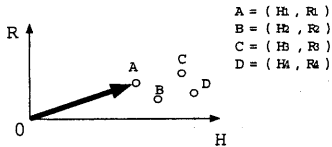


図5 距離関数と同期関数の座標表示

そこで、すべての基本動作の距離関数と同期関数の平均値をそれぞれ、1とした値にすべての距離関数と同期関数の値を変換する。具体的には、距離関数と同期関数の値をそれぞれの平均値で割る。そして、原点から座標点の距離を基本動作の類似度とする。これを、正規化された距離関数と同期関数の座標軸をそれぞれ、 H' 、 R' とし図6に示す。

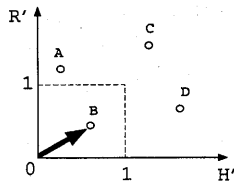


図6 正規化された座標表示

原点と正規化された座標点との距離により類似動作・非類似動作の判定を行う。原点からの距離が最小であれば類似動作と判定され、最大であれば非類似動作と判定される。

4.2 全体傾向による類似・非類似判定

同じ人が連続的に行った動作には、その人の動きの類似・非類似の傾向がでる。例えば、常に左手の動きが間違っている人や部位間の同期がよくとれている人などの類似・非類似の部分が表れる。ここでは、部位または同期についてユーザの傾向的な類似・非類似の判定を行う。

判定としては、以下の項目があげられる。

- 部位の類似・非類似
- 区分の同期・非同期

まず、すべての基本動作における部位ごとに距離関数の平均値を求める。また、すべての基本動作における区分ごとに同期関数の平均値を求める。基本動作を $n = \{1, 2, \dots, N\}$ としたときの、各部位 l の距離関数の h_{nl} とし、その平均値を h'_l とすると以下のように与えられる。

$$h'_l = \frac{1}{N} \sum_{n=1}^N h_{nl} \quad (1 \leq l \leq 24)$$

この h'_l により部位の類似・非類似判定を行う。 h'_l が

最小である部位が類似であり、最大である部位が非類似となる。同様に、各区分 q における同期関数 r_{nq} の平均値 r'_q とすると以下のように与えられる。

$$r'_q = \frac{1}{N} \sum_{n=1}^N r_{nq} \quad (1 \leq q \leq k)$$

この r'_q により区分の同期・非同期判定を行う。

5. 試作システムと実験

4章までに述べたことをもとに動作判定の試作システムを実装し、実験を行った。

5.1 試作システム

図7にシステムの画面を示す。システムは、データ表示部と操作部で構成されている。

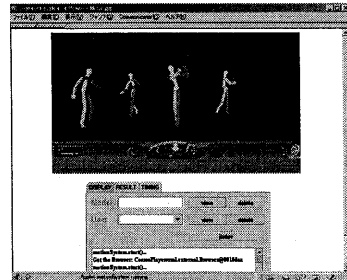


図7 システムの画面

データ表示部では、モーションデータをVRML形式に変換することで動作のブラウジングを行う。また、動きをクリックすることで模範動作の選択を行う。

操作部では、EAI(External Authoring Interface)を利用し、Javaアプレットからデータ表示部の制御を行う。また、選択された模範動作との動作判定を行い、結果を出力する。動作判定のプログラムはJava言語を使用している。

5.2 実験

5.2.1 モーションデータ

本研究では、赤外線を利用したモーションキャプチャーシステムを使い、MoitonAnalysisのEVAというソフトウェアと6台のカメラを制御する計算機を使用している。

モーションデータには、モーションキャプチャーシステムによって決定されたある原点からの絶対座標であるので、その人の体格、背丈の情報も反映されてしまう。よって、同じ動作をしても体格、背丈の差が類似度に影響を及ぼすので、その差を減らすために、モーションデータの体の各部位における最初の座標を初期

座標とし、すべての座標データからその初期座標を引くことにより、体の体格、背丈の差を吸収する。

5.2.2 実験方法

今回、実験の対象とする動作として、テニスのスイングの動作を用いる。実験に用いるユーザの連続動作を表1に示す。

表1 ユーザの連続動作

スイングの種類	1つの連続動作の動作回数	連続動作のパターン数
フォア	4	4
バック	4	2
フォアボレー	4	2
バックボレー	4	2

1つのパターンにおけるユーザの連続動作は、以下の動作が含まれている。

- (a) 模範動作に近い動作
- (b) タイミングのずれた動作
- (c) 間違っている動作

また、スイングの種類ごとに模範動作を1つ用意した。

動作判定の処理の流れは、図1の処理フローに従い、類似動作・非類似動作の判定を行う。全体傾向による類似・非類似判定は同じような動作が連続していることを前提としている。今回用意したモーションデータは(a)、(b)、(c)を混ぜた動作であるので傾向が明確にでないと考ええる。よって、今回の実験では全体傾向による類似・非類似判定は行わない。

また、同期関数における模範動作の指定要素数を $k = 8$ とし、模範動作の時系列データの長さを9等分した区分についてユーザ動作の同期関数を算出する。

5.2.3 実験結果

4種類のスイングごとのすべてのパターンの連続動作において類似動作・非類似動作の判定の実験を行った。いずれのパターンにおいても類似動作として、(a)が検出された。これは、動作の類似度に同期の概念を取り入れることにより、(a)と(b)のタイミングだけずれている動作の区別が明確になったためであると考ええる。非類似動作としては、(b)と(c)が検出された。中でも(b)はタイミングがずれることにより模範動作の軌道から大きくずれる動作が検出された。(c)では、模範動作の軌道から大きくずれることにより、非同期である動作が検出された。いずれの場合も距離関数と同期関数の値が大きくなるので、非類似動作として検出された。

6. おわりに

本研究では、モーションデータのマルチストリーム性を考慮して、動作の類似度を時系列データ間の距離と部位間の同期により決定した。これにより、距離だけでは判別できなかったタイミングのずれた動作についても判別できるようになった。しかし、距離関数と同期関数は全く独立ではなく、相互に作用し合う値である。実験結果からも(b)と(c)のいずれも模範動作と大きく軌道がずれる場合、DPマッチングでは無理にでも距離が近い要素と対応づけを行おうとするので、適切な対応づけが行えない。よって、距離関数と同期関数の値が大きくなる。

本研究では、時系列データ間の距離と部位間の同期を統合するために距離関数と同期関数の正規化によって行った。しかし、この手法はユーザ動作間の相対的な値であるので、決して適切な方法とはいえない。類似動作・非類似動作の判定では、距離関数と同期関数を含めた模範動作との絶対的な類似度により行われるべきであり、今後の課題として検討する必要がある。

謝辞 この研究は、一部、文部省科学研究費重点領域研究「高度データベースNo.275」(課題番号08244103)、および日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア・コンテンツの高次処理の研究」による。ここに記して謝意を表す。また、実験用データの取得に関してモーションキャプチャー装置の利用を許可頂いた通信・放送機構神戸リサーチセンター(平成11年3月終了)に深謝致します。

参考文献

- 1) 大崎竜太, 上原邦昭. Dynamic Time Warping法を用いた身体運動の動作認識. 情報処理学会研究報告, 98-DBS-116(2), pp.116-57, 1998
- 2) 吉野太智, 高木秀幸, 清木康, 北川高嗣. 楽曲データを対象としたメタデータ自動生成方式とその意味的連想検索への適用. 情報処理学会研究報告, 98-DBS-116(2), pp.116-41, 1998
- 3) 西原祐一, 梅田昌義, 紺谷精一. 大規模音楽DBに対する高速ハミング検索方式. 情報処理学会ADBS'98論文集, pp.117-124, 1998
- 4) 柳瀬隆史, 高須淳宏, 安達淳. メロディからの特徴抽出による曲検索システム. 電子情報通信学会データ工学ワークショップ DEWS'99, 2B-5, 1998
- 5) 川島英之, 嶋田総太郎, 安西祐一郎. 効果的な時系列データ検索手法. 電子情報通信学会データ工学ワークショップ DEWS'99, 3A-5, 1998