

## 論文

## 動作プリミティブ抽出と舞踊符割当ての自動化

湯川 崇\*\*・小原直子\*\*\*・玉本英夫\*\*\*

## Extraction of Motion Primitives and Automated Buyo-fu Assignment

Takashi Yukawa\*\*, Naoko Obara\*\*\* and Hideo Tamamoto\*\*\*

## Abstract

We have proposed a human motion description method using Buyo-fu, which aims at establishing a new recording method of human motions and re-using a time series of three-dimensional human motion data(mocap data) obtained by using a motion capturing system. The mocap data can be partitioned into some basic motions. We assign a code to each basic motion, and call it Buyo-fu. We can efficiently describe the human motions by using Buyo-fu. At present, we have a problem that it takes much time and effort to make Buyo-fu, because only a person who has professional knowledge about the concerned motions can extract Buyo-fu.

In this paper, we propose a method which can automatically extract Buyo-fu from the mocap data. In our proposed method, 1) a series of the human motions is partitioned into the basic motions based on the speed of the motions, 2) since we consider that similar basic motions appear several times in a series of motions, we do clustering the obtained basic motions, 3) we assign a label to each clustered basic motion, and call it a primitive motion, 4) we can express a series of motions by means of the concatenation of the primitive motions, 5) when we can find out the same pattern of the primitive motions in the human motions, we define this pattern to be a new primitive. Each primitive motion which can be obtained using our proposed method can become Buyo-fu.

Performing the experiment with the typical form of "tensho" and "sanchin" of Karate-do, we show the proposed method is a good candidate for automatically extracting Buyo-fu from the mocap data.

## 1. はじめに

情報通信技術の急速な進歩により、映画、テレビ、ビデオゲーム等をはじめとするさまざまな場面で3次元コンピュータグラフィックスの映像が用いられている。その中でも特に人間の動きを表現するライブアニメーションは上記の分野以外にも、人間工学、福祉工学、安全工学などの広い分野にわたって利用されている<sup>(1)</sup>。現実感のある人間の動きを生成するために広く用いられている方法として、モーションキャプチャリングシステムがある<sup>(1),(2)</sup>。これは磁気や光等を利用して、身体各部に取り付けたセンサの位置や角度の3次元時系列情報として人間の動きを記録するシステムである。

筆者らは、1)新しい身体動作の記録手法の確立、2)モーションキャプチャリングシステムにより記録した身体動作の3次元時系列データの再利用を目指し、舞踊符による動作記述法を提案している<sup>(3)</sup>。民俗芸能の舞踊等の動作は個々の基本動作に分割することができる。舞踊符とは分割した基本動作に割り当てた符号である。舞踊符の作成は、現在はビューアで動作データを再生しながら基本動作の境界を目視により決定して行うため、時間と労力を要する作業である。基本動作の自動的な切り出しができれば、記録した動作データを効率的にデータベース化し再利用することができる。

人間の動作データを自動的に基本動作に分割する試みには、センサの加速度の変化をもとにするもの<sup>(4)</sup>、速度の変化をもとに分割点を求めるもの<sup>(5)</sup>等がある。これらの研究では、抽出した動作データの境界が人間の考える動作の境界と一致するかどうかについては考慮していない。分割した基本動作が人間の考える基本動作と異なるため、舞踊符として再

2002年7月25日受理

\*\*秋田経済法科大学短期大学部商経情報学科, Department of Information for Commerce and Economics, Akita Keizaihoka University Junior College. (秋田大学大学院鉾山学研究所修士課程電子工学専攻, 平成6年度修了)

\*\*\*秋田大学工学資源学部情報工学科, Department of Computer Science and Engineering, Faculty of Engineering and Resource Science, Akita University.

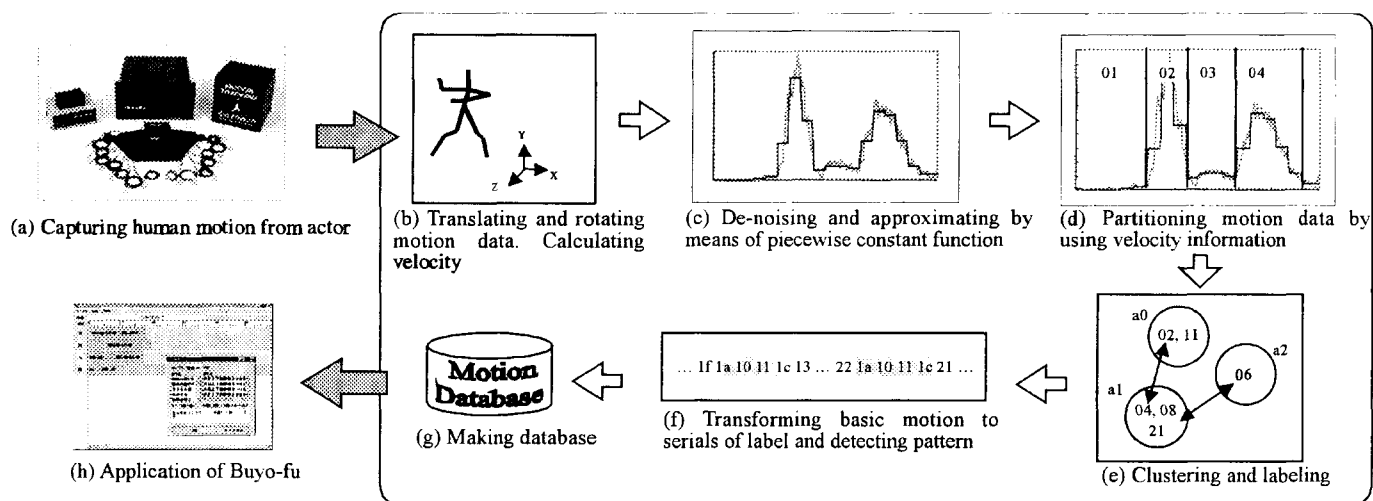


Fig.1 Overview of our project.

利用するには困難が伴う。

本研究では現在手作業で行っている舞踊符作成の自動化を目指し、検出する動作の境界が目視による境界と一致することを目標とする。

動作データを調査することにより、目視による境界が動作データの中でどのような特徴をあらわしているのかを明らかにする。その特徴を強調するために、動作の速さの時系列データを区分的に定数である階段関数で近似し、各区分の差をもとに基本動作の境界を決定する。つぎに、基本動作をクラスタリングすることにより、類似した基本動作を同じクラスタにまとめラベルを割り当てる。この基本動作を動作プリミティブと呼ぶ。以上の手続きによって一連の動作から動作プリミティブを抽出し舞踊符を作成することができる。

以下2.では本研究の概要について述べる。3.では、動作データの取得と基本動作の抽出について述べる。4.では基本動作をクラスタリングすることにより動作プリミティブを求める手法について述べる。5.では動作データを用いて実験により動作プリミティブを求め、提案する手法の有効性を確認する。

## 2. 提案する手法の概要

提案する手法の概要を Fig.1 に示す。モーションキャプチャリングシステム(Fig.1(a))により収集した3次元時系列データを体の向きや高さに影響されないようにするために座標変換を行い、速さの時系列データを計算する(Fig.1(b)).

動作データを調査したところ、基本動作の境界は(1)速さが減少しその後一定になる部分、(2)速さが一

定でその後増加する部分、(3)速さが減少しその後増加する部分、の3つの部分に分類されることが分かった。これらを検出するためには区分的に定数である階段関数で近似することが有効であることが考えられる。そこで、速さの時系列データを区分的に定数である階段関数で近似し(Fig.1(c)), 各区分の差をもとに基本動作の境界を決定する(Fig.1(d)).

一連の動作(舞踊等が開始してから終了するまでの動き)には同一の基本動作が複数回現れる場合があると考えられるので、基本動作同士を比較して類似する基本動作を同一のクラスタに分類する。基本動作間の類似性は、音声認識の分野で用いられている Dynamic Time Warping<sup>(5)</sup>の手法を用いることにより判定できる。同じクラスタに分類された基本動作には同一のラベルを割り当てる(Fig.1(e)). ラベルにより区別される基本動作を動作プリミティブと呼ぶ。

以上の手続きにより生成した動作プリミティブを連結して、一連の動作を表すことができる。一連の動作の中に動作プリミティブの並びが同一のパターンが現れることがある。その場合は、そのパターンが動作のまとまりを表していると考え新しいラベルを割り当てる(Fig.1(f)).

抽出した動作プリミティブはデータベース化することにより(Fig.1(g))舞踊符として再利用可能になる(Fig.1(h)).

## 3. 動作プリミティブの抽出

### 3.1 動作データの取得と速さの計算

本研究では、磁気式のモーションキャプチャリングシステムである Ascension Technology 社の

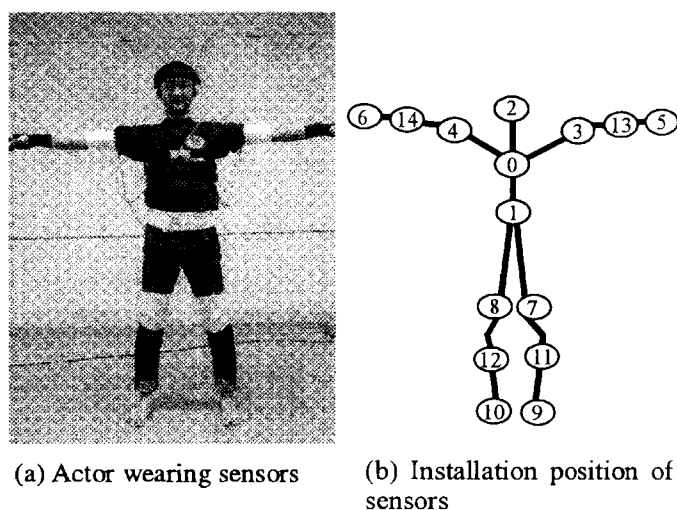


Fig.2 Position of sensors.

MotionStar Wireless を用いて身体動作の記録を行った。磁気式のモーションキャプチャリングシステムでは、身体各部にコイルを内蔵したセンサを取り付ける。トランスミッタから発生される磁界がセンサ内のコイルに誘起する電流の大きさをもとに、センサの位置と角度の 3 次元情報を計測する。Fig.2(a)は演技者にセンサを取り付けた様子である。Fig.2(b)に示す 15 箇所の位置にセンサを取り付け、1 秒間に 30 フレームのサンプリングレートで動作データの記録を行った。腕部と頭部 (Fig.2(b):0, 2, 3, 4, 5, 6, 13, 14) のデータは体が向いている方向や地面からの距離に影響されないようにするため、常に正面を向いた状態での腰 (Fig.2(b):1) のセンサからの相対的な距離を表すように座標変換を行う。脚部 (Fig.2(b):7, 8, 9, 10, 11, 12) のデータは体が向いている方向に影響されないようにするため、常に正面を向いた状態になるように座標変換を行う。

腰のセンサの 3 次元の位置と角度をそれぞれ  $P_1$ ,  $R_1$ , 腕部と頭部のセンサの 3 次元の位置を  $P_n(n=0,2\sim6,13,14)$  とする。  $M$  を  $R_1^{-1}$  の回転行列とする。腰が正面を向いたとき ( $R_1=0$ ) ,  $P_n$  の腰からの相対的な位置  $P_n'$  は、式(1)により求めることができる。

$$P_n' = M^T * (P_n - P_1)^T \quad (1)$$

脚部のセンサの位置を  $Q_n$  とする。腰が正面を向いたとき、  $Q_n$  の腰からの相対的な位置  $Q_n'$  は、式(2)により求めることができる。

$$Q_n' = M^T * Q_n^T \quad (2)$$

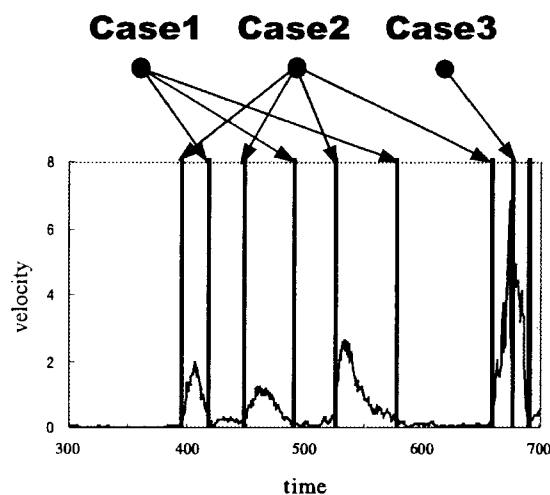


Fig.3 Velocity of right hand and boundary of basic motion.

これらの式を記録したすべてのデータに対して適用することにより、変換後の座標を得ることができる。変換した位置の時系列データから 5 点近似法を用いて各部位の速さを求める。

### 3.2 動作データの目視による解析

Fig.3 は、右手 (Fig.2(b):6) の動作データの速さの時系列データと目視により分割した基本動作の境界 (縦線) を示した例である。これより人間の基本動作の境界には、着目する部位の速さが

- (ケース 1) 減少から一定に変化する時点
- (ケース 2) 一定から増加に変化する時点
- (ケース 3) 減少から増加に変化する時点

の 3 つの場合があることが分かる。Fig.3 中の矢印はそれぞれの境界がどのケースに該当するかを示している。一連の動作中の基本動作の境界を決定するためにはケース 1~3 の 3 種類の点を検出する必要がある。

### 3.3 ウェーブレット変換を用いた雑音の除去と階段関数による近似

モーションキャプチャリングシステムによる動作記録時には、雑音や動作の揺れがデータに混入する。速さの計算時には動作データが含んでいる雑音の影響が強調される。従って、雑音と基本動作の境界を区別するために時系列データから前もって雑音を除去する必要がある。

基本動作の境界を決定するためには、動作中の動きのある部分と一定の姿勢を保ったまま静止している部分とを区別する必要がある。そのためには速さの時系列データを区分的に定数である階段関数で近似することが有効であると考えられる。

スケーリング関数に Haar 関数を用いたウェーブレット変換を行うことにより，もとの時系列データから雑音を取り除き，区分的に定数である階段関数で近似することができる<sup>(6)</sup>。

Fig.4 に示す手続きを用いて時系列データ  $C_n$  に対し Haar ウェーブレット変換を行う<sup>(6)</sup>。しきい値の決定には Minimax 法<sup>(7),(8)</sup>を用いる。Fig.5 に示す手続きを用いてウェーブレット係数を再構成することにより， $C_n$  から雑音を取り除かれ，区分的に定数で近似した時系列データを得る<sup>(9)</sup>。

Fig.6 に速さの時系列データとウェーブレットを用いて雑音を除去し区分的に定数である階段関数で近似した速さの時系列データの例を示す。

### 3.4 動作データの分割

以上の考察に基づいて，次の手続きで基本動作の境界を選択する。

[手続き 1] (基本動作の境界決定)

ステップ I: 3.3 で求めた速さの時系列データから各区間の差を計算する。差がある値以上の箇所を境界の候補とする。

ステップ II: 立下りと立上がりが連続している箇所以外を候補から削除する。

ステップ III: 隣り合う候補の間隔が狭い箇所はその中間の時点境界とする。

ステップ IV: それ以外の候補を境界とする。

ステップ I では速さに変化が生じた時点候補とする。ステップ II ではある動作が終了する時点(ケース 1)，およびその後ある時間静止して次の動作を開始する時点(ケース 2)を候補から選出する。ステップ III では 2 種類の動作が続いて起こる境界(ケース 3)を選出する。

## 4. 舞踊符の割り当て

3.4 で述べた手続き 1 を実行することによって基本動作の境界を求めることができる。一連の動作の中には類似した基本動作が複数回現れる場合があるため，クラスタリングを行い類似した基本動作をクラスタにまとめる。

### 4.1 Dynamic Time Warping<sup>(5)</sup>

クラスタリングを行うためには基本動作間の類似度を距離として求める必要がある。基本動作のフレーム数はそれぞれ異なるために音声認識の分野で使われる手法である Dynamic Time Warping (DTW) を用いて 2 つの基本動作間の距離を求める。

ここで，2 つの基本動作  $A, B$  を

```

procedure Decomposition(c: array[1..2j] of reals)
  c ← c / (2j)1/2
  g ← 2j
  while g ≥ 2 do
    DecompositionStep(c[1 .. g])
    g ← g / 2
  end while
end procedure

procedure DecompositionStep(c: array [1..2j] of reals)
  for i ← 1 to 2j/2 do
    c'[i] ← (c[2i - 1] + c[2i]) / 21/2
    c'[2j/2 + i] ← (c[2i - 1] - c[2i]) / 21/2
  end for
  c ← c'
end procedure

```

Fig.4 Haar Wavelet decomposition procedure<sup>(6)</sup>.

```

procedure Reconstruction(c: array[1..2j] of reals)
  g ← 2
  while g ≤ 2j do
    ReconstructionStep(c[1..g])
    g ← 2g
  end while
  c ← c (2j)1/2
end procedure

procedure ReconstructionStep(c: array[1..2j] of reals)
  for i ← 1 to 2j/2 do
    c'[2i - 1] ← (c[i] + c[2j/2 + i]) / 21/2
    c'[2i] ← (c[i] - c[2j/2 + i]) / 21/2
  end for
  c ← c'
end procedure

```

Fig.5 Haar Wavelet reconstruction procedure<sup>(6)</sup>.

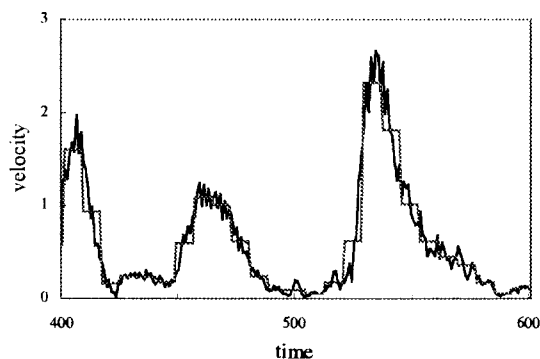


Fig.6 Example of velocity waveform of right hand and piecewise-constant approximations.

$$\begin{cases} A = a_1, \dots, a_i, \dots, a_M, a_i = (x_{ai}, y_{ai}, z_{ai}) \\ B = b_1, \dots, b_i, \dots, b_N, b_j = (x_{bj}, y_{bj}, z_{bj}) \end{cases} \quad (3)$$

とする。ただし、 $a_i, b_i$  は各基本動作の  $i$  番目のフレームの座標である。このとき  $A, B$  の距離  $D(A, B)$  は DTW を用いて式(4)で与えられる<sup>(5)</sup>。

$$\begin{cases} d_{i,j} = ((x_{ai} - x_{bj})^2 + (y_{ai} - y_{bj})^2 + (z_{ai} - z_{bj})^2)^{1/2} \\ S_{i,j} = d_{i,j} + \min(S_{i,j-1}, S_{i-1,j-1}, S_{i-1,j}) \\ S_{1,1} = d_{1,1} \\ D(A, B) = S_{M,N} \end{cases} \quad (4)$$

## 4.2 基本動作のクラスタリング

DTW を用いて計算した各基本動作間の距離をもとに次の手続きで階層型クラスタリングを実行する。

[手続き 2] (基本動作クラスタの作成)

ステップ I :  $n$  個の基本動作から  $n$  個のクラスタを作成する。

ステップ II : クラスタ間の距離を求め、最も近くにあるクラスタを結合して新しいクラスタを作成する。どのクラスタが結合したかを記録する。

ステップ III : クラスタ数が 1 ならステップ IV へ。2 以上ならステップ II へ戻る。

ステップ IV : ステップ II で記録したクラスタ間の結合関係を、Dendrogram と呼ばれる木構造で表す。この Dendrogram を切断する木の高さによりクラスタの数を選べる。

Fig.7 に 6 個の個体 (基本動作) をクラスタに分類する例を示す。ステップ IV で作成した Dendrogram を図中の点線の高さで切断することにより 6 個の個体を 4 つのクラスタに分類している。

クラスタ間の距離の判定には最近隣法を用いる。最近隣法は 2 つのクラスタ間の最も近い個体の距離をクラスタ間の距離とする手法である。クラスタ  $r$  中の個体数を  $n_r$ 、クラスタ  $s$  中の個体数を  $n_s$ 、クラスタ  $r$  中の  $i$  番目の個体を  $x_{ri}$ 、クラスタ  $s$  中の  $j$  番目の個体を  $x_{sj}$  とすると、クラスタ間の距離  $d(r, s)$  は式 (5) で求められる。

$$d(r, s) = \min(D(x_{ri}, x_{sj}), i \in (1, \dots, n_r), j \in (1, \dots, n_s)) \quad (5)$$

クラスタリングの結果、分類された各クラスタにラベルを割り当てる。ラベルにより区別される基本動作を動作プリミティブと呼ぶ。

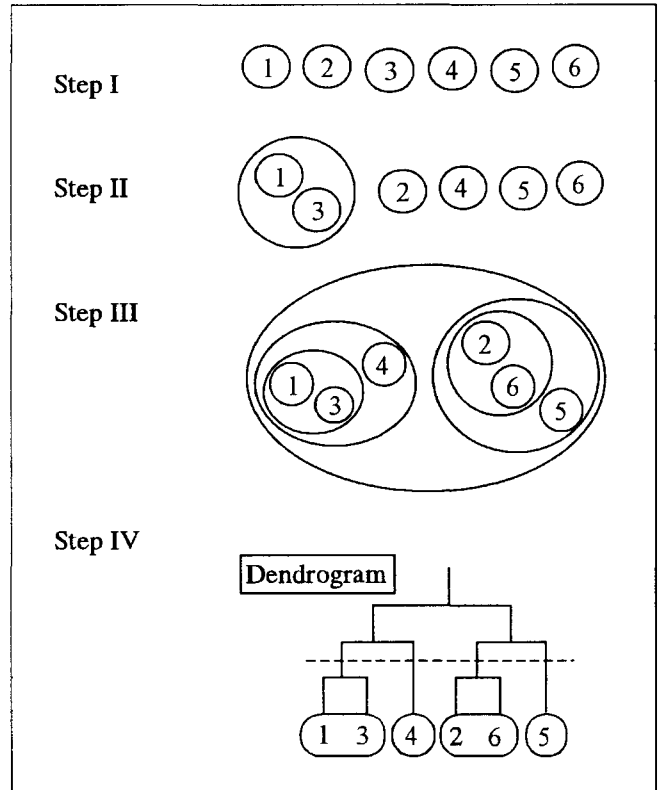


Fig.7 Clustering procedure.

```
procedure findpat(pri:array 1..plen) of character,
  minpatlen: integer)
```

```
  for n = 1 to prilen-1
    A ← pri[n+1..plen]
    B ← pri[1..plen-n]
    c ← getseq(A==B);
    p ← find(c>minpatlen-1);
    for m ← 1 to length(p)
      pos ← p[m]
```

Here, we can notice that sequence  $pri[pos..pos+c[pos]-1]$  appears at position  $pos$  and  $pos+n$ .

```
  endfor
endfor
end procedure
```

```
procedure a = getseq(a:array[1..alen] of character)
```

```
  for n ← alen to 2 step -1
    if a[n-1] > 0
      a[n-1] ← a[n-1] + a[n]
      a[n] ← 0
    end if
  end for
end procedure
```

Fig.8 Pattern detection procedure.

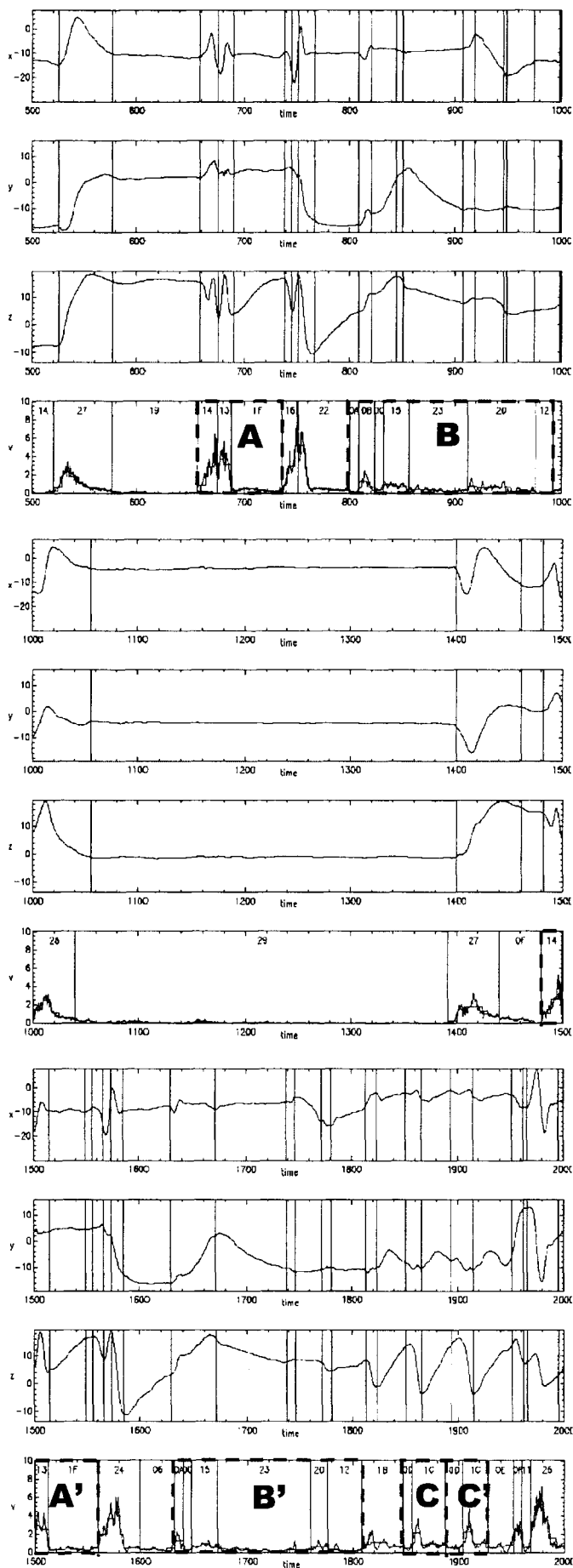


Fig.9 Result of motion analysis 1 (tensho).

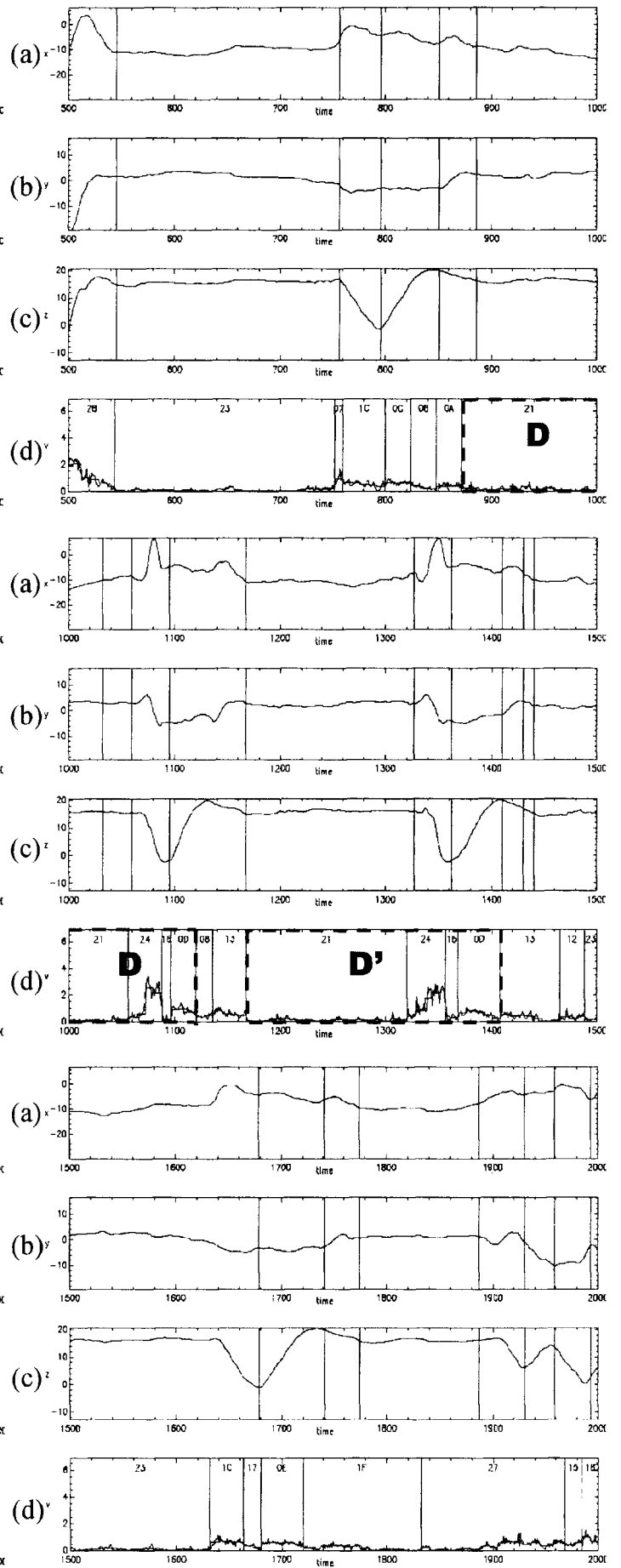


Fig.10 Result of motion analysis 2 (sanchin).

Table 1 Experimental results

motion name	tensho	sanchin
# motion boundaries (by hand)	63	46
# motion boundaries (detect)	61	56
# motion primitives	41	41
# ≤5frames(0.17sec)	36 (55.7%)	23 (47.7%)
# ≤10frames(0.33sec)	48 (75.4%)	29 (61.4%)
# ≤20frames(0.67sec)	59 (93.4%)	36 (77.3%)
# ≤30frames(1.00sec)	63 (100.0%)	43 (93.2%)

#### 4.3 動作プリミティブ列からのパターン検出

前節の手続き2を実行して得られる動作プリミティブを接続することにより、一連の動作を表現することができる。この表現の中に動作プリミティブの並びが同一のパターンが現れているかどうかを調べ、見つかった場合にはそのパターンを新しい動作プリミティブとする。これを実現する手続きを Fig.8 に示す。

### 5. 実験

提案する手法の有効性を確認するために、動作データを用いて実験を行った。使用したデータは、剛柔流空手道の代表的な形である「転掌」と「三戦」<sup>(10)</sup>の右手の動きである。

Fig.9(a)~(c)と Fig.10(a)~(c)に「転掌」と「三戦」の右手の位置の3次元時系列データをそれぞれ示している。Fig.9(d)と Fig.10(d)にそれぞれの速さを示している。Fig.9,10(a)~(c)中の垂線は目視により決定した基本動作の境界を表している。Fig.9,10(d)中の垂線は提案する手法で検出した基本動作の境界を表している。Fig.9,10(d)の各区間には、それぞれの区間の動作プリミティブに割り当てたラベルを記している。図中の太線で囲んである区間は複数回現れるラベルの並びを表しており、同じアルファベット(AとA'等)で示している。x,y,zの位置の変化の波形を比較すると、同じ基本動作に同じ動作プリミティブが割り当てられていることが分かる。

Table 1に実験結果を示す。1~3行目はそれぞれ目視により決定した基本動作の境界(目標境界)数、提案する手法により検出した基本動作の境界(検出境界)数、検出した基本動作をクラスタリングした結果、抽出した動作プリミティブの数を表している。4~7行目はそれぞれ目標境界の5,10,20,30フレーム以内に検出境界が存在する数(検出率)を示してい

る。これを検出境界が目標境界にどれだけ近いかを示す指標とする。検出境界数(2行目)が多くなると検出率の値が増加するため手法の性能が正確に反映されるものではないが、検出境界数と目標境界数が近い値であれば、ある程度の目安になるものと考えられる。

「転掌」に比べ「三戦」の検出率が低い値を示しているのは次の理由によるものと考えられる。「転掌」はゆっくりした動きと速い動きがはっきり分かれているため、基本動作の境界を検出しやすい。一方「三戦」はゆっくりした動きを渾身の力をこめて演じるため、動作の境界が速さの変化に現れにくく、力を入れることにより生じる部位の揺れが雑音と区別しにくい。

提案する手法により検出した境界の約半数が目視により決定した境界の5フレーム以内にあることが分かる。また Fig.9,10 から動作を行っている部分と静止している部分が区別されていることが分かる。以上のことから提案する境界の抽出方法が有効であることを確認できた。

### 6. おわりに

モーションキャプチャリングシステムにより記録した動作データを基本動作に分割するための手法を提案した。本手法の特徴は基本動作に分割する際に目視で決定した境界と同一になるように基本動作の境界を選択することである。目視により決定した境界が動作データの中でどのような特徴を持つのかを調査した結果、人間の動作は(1)減少から一定に変化する時点(2)一定から増加に変化する時点(3)減少から増加に変化する時点、の3通りの時点で分割できることが分かった。動作データをこれらの時点で分割するために、身体の部位の速さの時系列データを区分的に定数である階段関数で近似し、各区間の差の大きさにより境界の判定を行った。

動作データとして空手道の形を用いた実験により、提案する手法が一連の動作からの動作プリミティブ抽出の有効な手法となることを示した。

作成した舞踊符はデータベース化し、モーションコンポーザ<sup>(3)</sup>などのツールを利用することにより再利用可能になる。

今後の課題としては、1)民俗芸能の舞踊等他の動作を用いた解析、2)右手以外の他部位の利用、3)部位の速さ以外の角度などの特徴量の利用などが考えられる。

### 謝辞

本研究を進めるにあたりモーションキャプチャデータの収録に協力していただいた(株)わらび座デジタルアートファクトリのスタッフに感謝いたします。

### 参考文献

ピュータシンポジウム論文集, 25-32 頁.

- (1) 中島正之 (1998) : デジタル映像処理の基礎と応用 [IV] - コンピュータアニメーション入門 -, 信学誌, 81 巻, 3 号, 282-8 頁.
- (2) M. O'Rourke (1997): 3次元コンピュータアニメーションの原理, トップラン
- (3) 湯川崇, 海賀孝明, 長瀬一男, 玉本英夫 (2000) : 舞踊符による身体動作記述システム, 情報学論, 41 巻, 10 号, 2873-80 頁.
- (4) 中澤篤志, 中岡慎一郎, 池内克史 (2001) : 動きのデジタル保存 - 舞踊動作のプリミティブ化とヒューマノイドにより再演 -, 人文科学とコン

- (5) 大崎竜太, 嶋田光臣, 上原邦昭 (2000) : 速度に基づく切り出しとクラスタリングによる基本動作の抽出, 人工知能誌, 15 巻, 5 号, 878-86 頁.
- (6) Stollnitz, E. J., Deroose, T. D. and Salesin, D. H. (1996): Wavelets for Computer Graphics theory and applications, Morgan Kaufmann Publishers, pp.9-20.
- (7) Dohono, D. L. (1995): De-nosing by Soft-Thresholding, *IEEE Trans. on Information Theory*, Vol. 41, No. 3, pp.613-27.
- (8) Donoho, D. L., and Johnstone, I. M. (1994) : Ideal spatial adaptation via wavelet shrinkage, *Biometrika*, 81: pp.425-55.
- (9) Mallat, S. G., Mallat, S. (1999): a Wavelet tour of signal processing, Academic Press, pp.434-74.
- (10) 山口剛玄 (1996) : 空手道教範, 東京書店, 100-23 頁.