

時系列動作情報からの頻出動作パターンの抽出について

荒木, 裕
九州大学システム情報科学研究院知能システム学部門

有田, 大作
九州大学システム情報科学研究院知能システム学部門

谷口, 倫一郎
九州大学システム情報科学研究院知能システム学部門

<http://hdl.handle.net/2324/5865>

出版情報 : 電気学会研究会資料. IIS, 産業システム情報化研究会. 6, pp. 39-44, 2006-03. 電気学会
バージョン :
権利関係 :

時系列動作情報からの頻出動作パターンの抽出について

荒木 裕*, 有田 大作, 谷口 倫一郎 (九州大学 大学院システム情報科学府・研究院)

Motif Extraction from Multi-dimensional Motion Information

Yutaka Araki*, Daisaku Arita, Rin-ichiro Taniguchi (Department of Intelligent Systems, Kyushu University)

Abstract

Recently, there are a lot of researches on virtual environments for distant human communication. Real-time Human Proxy (RHP), which is a concept for communication system based on a virtual environment, has been proposed. For realizing natural communication by RHP, it is necessary to recognize human actions essential for human communication. However, it is difficult for system developers to decide which human actions should be recognized. For supporting the decision, we propose a human motion analysis method which automatically extracts frequent motion patterns as human action candidates. In this paper, we outline our idea and show some experimental results for extracting human action.

キーワード：モチーフ, 時系列データ
(motif, time-series data)

1. はじめに

近年、バーチャルリアリティ技術の発展により、リアリティの高い仮想空間を用いた遠隔コミュニケーションの研究が注目されている。それは主に、3次元の仮想環境内において、アバタと呼ばれるCGキャラクター同士のインタラクションにより、遠隔地にいる人同士のコミュニケーションの実現を目指すものである。仮想環境内では、各参加者はアバタとして視覚化され、誰がどこにいてどういう行動をとっているかなど、実世界における空間的情報を擬似的に再現できるため、誰が誰に向けて話しているなどのコミュニケーションの発信とその対象者を直感的に理解でき、実世界における自然なコミュニケーションに近い臨場感を創り出すことができると考えられる。

仮想環境を用いたコミュニケーションシステムでは、NICE(Narrative Immersive Constructionist / Collaborative Environments)⁽¹⁾ や Nessie World⁽²⁾ がある。しかしながら、これらは二つの問題を抱えている。一つはシステムへの入力に特別な装置が必要であるということ、もう一つは用いているアバタが単純すぎて自然なコミュニケーションの妨げになっているということである。これらの問題により、仮想環境でのコミュニケーションに実世界との違和感を感じてしまう。

これらの問題を解決するために、仮想環境による遠隔コミュニケーションのフレームワークとして、著者らは Real-time Human Proxy (RHP)⁽³⁾ という概念を提唱している。それは、実世界にいる人間が発信する情報を実時間で仮想環境内のアバタへ投影するための概念である。このRHPによる遠隔コミュニケーションを達成するためには、参加者のさまざまな情報を獲得、認識し、仮想環境内のアバタをその参加者として振舞わせるという処理を実時間で行う必要がある。このための第一歩として、著者らは参加者の情

報の中から動作情報に着目し、その獲得、認識のための研究を行っている。人物の動作情報をアバタに反映させる上で、動作の認識は、欠かすことの出来ない処理である。ここで、動作認識のためには、遠隔コミュニケーションで利用される動作を認識すべき動作として全て洗い出しておく必要がある。

本稿では、認識用動作の獲得のためのアプローチとして、あらかじめ記録された人物の動作情報から認識のために必要になる動作パターンを自動的に抽出するための手法を提案する。

2. Real-time Human Proxy (RHP)

2-1 RHP RHPとは、実世界に存在する人間を実時間で仮想化し、その仮想化した情報をネットワークを介して通信することにより、遠隔地であたかも実の人間がその場に存在するかのように振舞わせるという概念である。人間の仮想化とは、その人間が他の人へ発信する情報(手や足の動き、体格、人間の表情、発せられる音声、着用している衣服など)をシンボルと呼ばれる表現形式へ変換することをいい、仮想化された情報を再現するとは、変換されたシンボルから、CGアバタという媒介を介して人間の表現をつくり出すことをいう。

ここで、仮想化すべき人間の情報には多くのものが考えられるが、特に体の動きに関する情報はコミュニケーションにおいて重要な役割を担っていると考えられる。これは、人と人のインタラクションでは、身振り手振りのようなボディランゲージが、言葉と同様に、時には言葉以上に様々なメッセージを伝達しているからである。そこで、本研究では、人物の動作情報からシンボルを生成し、それをもとにアバタの振る舞いとして参加者の情報を表現することを目指す。

2-2 動作情報獲得システム 人間の動作の獲得に

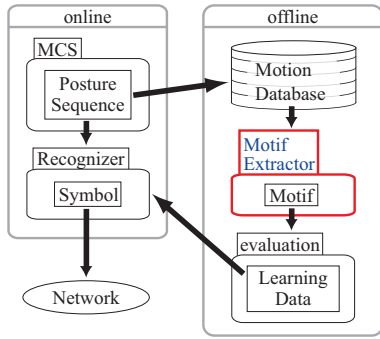


図1 動作情報獲得システムの処理の流れ

あたっては、モーションキャプチャシステム (MCS)⁽⁴⁾ を利用する。MCS を利用することにより、人間の体の動きそのものを入力として用いることが可能となり、普通のしぐさやその人の癖、無意識のうちに行う動作なども入力として利用することが可能となる。そしてモーションキャプチャシステムにより獲得された動作から、アバタを操るための動作シンボルを生成する。このとき、動作シンボルとして認識したい動作を決定しておく必要があるが、たとえコミュニケーションの状況が特定されていたとしても、必要な全ての動作を決定することは容易なことではない。そのため、コミュニケーションに必要な動作を効率的に決定し、その動作から動作シンボルを生成する枠組みが必要となる。これまで著者らは、コミュニケーションに必要な動作をあらかじめ列挙しておき、それらの動作を認識していくようなトップダウン的なアプローチをとってきた。このトップダウン的な方法では、コミュニケーションに必要な動作を列挙しておく必要があるが、これを人間が手作業で行うことを考慮すると、過不足なく全て列挙するのは困難である。そこで本稿では、人間が頻繁に行う動作には意味があるという考えのもと、実世界で行われるコミュニケーションで用いられた動作情報をあらかじめ取得しておき、得られた動きの時系列データから、頻出する動作（モチーフと呼ばれる⁽⁵⁾）を自動的に発見し、その後それらがコミュニケーションに必要などうかを判定するボトムアップ的なアプローチを試みる。

このための動作情報獲得システムの概要を図1に示す。オンラインの処理として、モーションキャプチャシステム (MCS) により実時間で動作情報を獲得し、認識器で動作シンボルを生成する。その際、認識すべき動作内容は、オフラインで得られた学習データを用いる。学習データの取得はMCSにより得られた動作情報をデータベースに記録しておき、それらをモチーフ抽出器で解析することで頻出動作をモチーフとして抽出する。そして得られたモチーフからさらに重要なものを学習データとして用いる。

3. モチーフ抽出

ここでは、人物動作情報からモチーフを抽出する手法を提案する。モチーフを抽出するために、以下の3段階の処

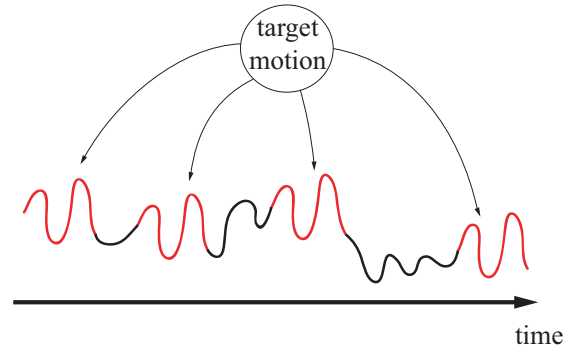


図2 1次元データからのモチーフ抽出例

理を行う。第一に高次元の時系列データである動作データを直接扱うことで生じる問題である計算量と情報の冗長性の低減のため、Principal Feature Analysis(PFA)⁽⁶⁾ によって多次元の動作データを低次元のものに圧縮する。第二に低次元化した各次元の動作情報を時系列データからラベル列に変換する。第三にデータマイニングでよく用いられる最小記述長 (MDL) 基準⁽⁷⁾ を用いてラベル列から検出される頻出パターンからモチーフを選出する。詳しくは次節より述べる。

3.1 PFAによる低次元化 人間の動作データのような高次元データを低次元化することは、パターン認識やデータの分類、圧縮のための前処理としてよく行われる。主成分分析 (PCA) や独立成分分析 (ICA) は、この低次元化のためによく用いられる手法である。これらの手法は、元の特徴空間から低次元化した空間への射影を算出するために用いられる。しかしながら、これらの手法により動作情報を低次元化した場合、低次元化された空間での値の変位がもとの特徴空間のどの次元の変位かわからないため、動作が起こったかどうかはわかるが体のどの部位が動いたのかわからず、モチーフ抽出では有効ではないと考えられる。そこで、元の特徴空間から重要な特徴の部分集合を決定できる Principal Feature Analysis(PFA) を用いることで、この問題に対応する。人物動作情報は (x, y, z) の3次元で表される体の観測部位の集合で表現される。以後、各観測部位を動作ベクトル $f_i = [f_{1i} f_{2i} \dots f_{ni}]^T \in \mathbb{R}^n$ 、動作データ全体を行列 $M = [f_1 f_2 \dots f_s] \in \mathbb{R}^{n \times s}$ で表す。ここで、 s は次元数 (観測部位の数 $\times 3$)、 n は動作データの長さを示す。低次元化は次の4つの処理により選出される。

第一に行列 M の共分散行列を求める。

第二に、 M の主成分分析を次式により算出する。

$$C(M) = \mathbf{V} \Sigma \mathbf{V}^T \dots \dots \dots (1)$$

$$\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_s]$$

$$\Sigma = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_s)$$

ここで、 $C(M)$ は M の共分散行列を、 $\mathbf{v}_i \in \mathbb{R}^s$ は M の主成分を、 Σ は対角成分に $C(M)$ の固有値 $\lambda_1 \geq \lambda_2 \geq \dots \lambda_s$ を持つ対角行列である。

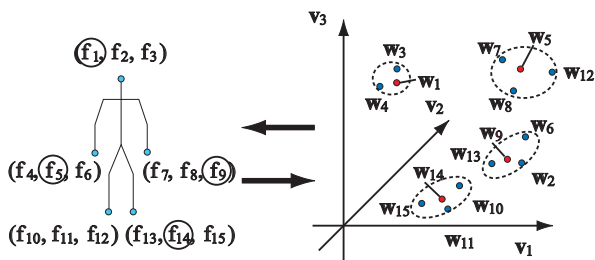


図3 PFAにより重要な軸を選出する例。ここで、軸の数は15, $q = 3, p = 4$, (v_1, v_2, v_3) は主成分空間の座標系, 線を引く w_i が各クラスタの中央値に最も近いことを表し, 丸で囲まれた f_i が対応する軸で, この軸が選出される。

第三に, 式(2)において M の主成分分析による低次元化された表現を $M_{pc} \in \mathbb{R}^{n \times q}$, 求めたい主成分の数を $q < s$, 行列 V^T の第 q 列までを $W = [w_1 \ w_2 \ \dots \ w_s] \in \mathbb{R}^{q \times s}$ としたとき,

$$M_{pc} = \tilde{M}W^T \dots \dots \dots (2)$$

$$\tilde{M} = [\tilde{f}_1 \ \tilde{f}_2 \ \dots \ \tilde{f}_s]$$

$$\tilde{f}_i = [f_{1i} \ f_{2i} \ \dots \ f_{ni}]^T$$

$$\tilde{f}_{ji} = f_{ji} - \frac{1}{n} \sum_{k=1}^n f_{ki}$$

ユークリッド距離を用いた K-Means 法によってベクトル $|w_i|$ を $p \geq q$ 個のクラスにクラスタリングを行う。ここで, w_i は f_i の主成分分析による低次元化された空間への射影を示している。

最後に, 各クラスタ毎に, その中央値に最も近いベクトル w_i を選出する。そしてそのベクトルに対応する f_i が重要な特徴部分集合のうちの1つの次元として選出される。似たような次元を集めた各クラスタから各代表 f_i を選出するため, 低次元化とともに次元の冗長性を低減させることができる。図3にPFAの例を示す。

3.2 時系列データからラベル列への変換 PFAにより選出された特徴のデータは時系列データである。このデータをそのまま入力データとして用いた場合, モチーフ抽出には多大な計算時間が必要となる。この問題を解決するために, PFAで選出された時系列データ f を Symbolic Aggregate Approximation(SAX)⁽⁸⁾ によってシンボル表現(ラベル列)に変換する。SAXは Piecewise Aggregate Approximation(PAA)を用いた低次元化, 量子化によって時系列データをラベル列に変換するための手法である(図4)。第一にPAAによって長さ n の時系列データ f を w 次元のベクトル $\tilde{f} = [f_1, \dots, f_w]$ に変換する(式(3))。

$$\tilde{f}_i = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} f_j \dots \dots \dots (3)$$

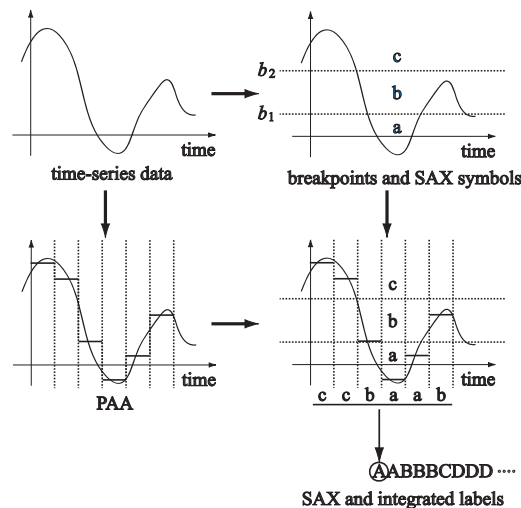


図4 時系列データからSAXを用いてラベル列への変換処理。

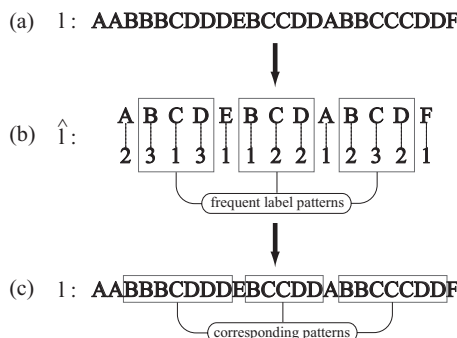


図5 ラベル列の表現変換と対応する頻出パターン:(a)元のラベル列,(b)変換されたラベル列,(c)変換されたラベル列で見つかる頻出パターンに対応するものとのラベル列でのパターン。

第二に, \tilde{f}_i を SAX シンボルへ量子化する。量子化には閾値(breakpoint)を用いて次のように \tilde{f}_i を SAX シンボル \hat{f}_i に変換する。

$$\hat{f}_i = \text{sym}_j, \text{ iff } b_{j-1} \leq \tilde{f}_i \leq b_j \dots \dots \dots (4)$$

$$(j = 1, \dots, a)$$

ここで, b_j は閾値, sym_j は SAX シンボルである。閾値は変換前の時系列データ f の分布をガウス分布と仮定したとき, 量子化された各レベルで時系列データが等確率で生成されるように決定される。第三に, 連続した c 個の SAX シンボル " $\hat{f}_i \dots \hat{f}_{i+c-1}$ " をユニークな1つのラベル l_i に変換し, ラベル列 $l = [l_1 \dots l_{w-b+1}]$ を生成する。各ラベル l_i はモチーフの最小構成要素となる。第四に, 動作速度の影響を低減させるため, l をランレングス符号化により \hat{l} に変換する(図5)。 \hat{l} は2種類の情報, ラベル列とそれぞれのラベルの連続繰り返し数で表現される。

3.3 MDL 基準によるモチーフ抽出 図5(b)に示



図6 繰り返し処理：(a)元のラベル列，(b)1番目のモチーフ抽出，(c)2番目のモチーフ抽出，(d)最終的に検出されたセグメント

されるようなラベル列 \hat{l} が与えられたとき，“BCD”はモチーフとして抽出したい頻出パターンである．しかし，人物動作情報からはこのような頻出パターンが数多く検出される．そのため，それら頻出パターンからモチーフを選択するための評価用の指標が必要となる．そこで，この評価指標にMDL基準を用いる．MDL基準の考えを頻出パターンの評価に用いると，頻出パターンが一つのラベルに置き換えられたとき，もとのラベル列を最も圧縮できるようなモデルを探すことになる．このときのモデルがモチーフとして適切なものを表していると考えられる．

MDL基準を用いて頻出パターンを評価する場合，ラベル列 \hat{l} の記述長を評価指標として用いることになる．記述長 DL の式はモデル h とラベル列 \hat{l} それぞれの記述長を表すように以下の式として表現される⁽⁹⁾．

$$DL = DL(\hat{l}|h) + DL(h) \dots\dots\dots (5)$$

$$DL(\hat{l}|h) = \sum_i^m \sum_j -w_{ij} \log_2 \frac{w_{ij}}{t_i} \dots\dots\dots (6)$$

$$DL(h) = \sum_i^m \log_2 t_i + m \log_2 \left(\sum_i^m t_i \right) \dots\dots\dots (7)$$

ここで，モデル h はラベル列 \hat{l} のセグメンテーションを， m はセグメントの数を， w_{ij} は i 番目のセグメントの j 番目のラベルの繰り返し回数を， t_i は i 番目のセグメントの数を示している．図5(b)のラベル列 \hat{l} の場合， $h = \{1, 4, 5, 8, 9, 12\}$ ， $m = 7$ ， $w_{11} = 2$ ， $t_1 = 2$ ， $w_{21} = 3$ ， $w_{22} = 1$ ， $w_{23} = 3$ ， $t_2 = 7$ ，...となる．

ここで，ラベル列 \hat{l} の長さを t_L とすると，全てのセグメンテーションの数は $O(2^{t_L})$ となり，多大な計算量となってしまふ．そこで，次のような繰り返し処理によってこの問題を近似的に解決する．

第一に，全ての頻出パターンをモチーフ候補として探索窓を用いて抽出し，それらの中から式(5)を最小にするようなパターンを1番目のモチーフとして選出する(6(a))．このときの計算量は $O(t_L^3)$ である．

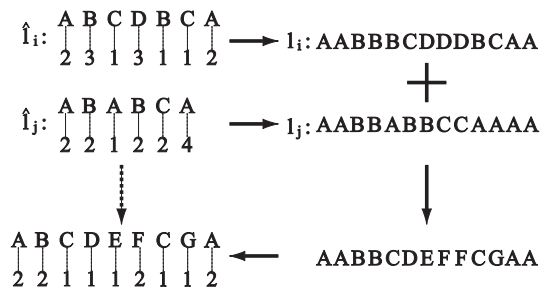


図7 \hat{l}_i と \hat{l}_j の統合例．

第二に，選出された頻出パターンをユニークなラベルに置き換える．ここで長さが2以上の頻出パターンが選出されなかった場合，繰り返し処理を終了する．

第三に，第一，第二の処理を繰り返し， $i(\geq 2)$ 番目のモチーフを選出していく．このとき，各頻出パターンにモチーフとして置き換えられたラベルを含まないようにする(図6(b))．処理が全て終われば，図6(d)のようにセグメンテーションも終了している． t_M を繰り返し処理により抽出されたモチーフの数とすると，繰り返し処理全体の計算量は $O(t_M t_L^3)$ となる．

3.4 軸の統合

これまででは，動作データ内の1次元のデータからモチーフ抽出するための手法を述べてきた．全身動作を解析するためには，PFAによって選出された各次元を統合した処理が必要となる．本論文では，単純に，PFAにより選出された動作ベクトル f_i に対応する \hat{l}_i の全ての組み合わせを用いて統合したラベル列を生成し，それらからMDL基準によりモチーフ抽出を行う．ラベル列 \hat{l}_i と \hat{l}_j を統合する場合を例として説明すると，組み合わせ (l_{it}, l_{jt}) に対し，ユニークなラベルを割り当てることで統合されたラベル列を生成する(図7)．ここで， l_{it} は \hat{l}_i に変換する前の l_i の t 番目の要素に対応する．

4. 実験と考察

提案手法を用いてモチーフ抽出実験を行った．動作データはカーネギーメロン大学の Graphics Lab motion-capture database[†] (<http://mocap.cs.cmu.edu/>) のデータを利用した．ここに記して，感謝する．実験に用いた動作データは25箇所の観測部位情報を持つ，2486フレーム分のデータである．それぞれの観測部位は (x, y, z) で表現されるため，75次元データとなる．動作内容は，3種類のバスケットボールで用いられる動作(A)，(B)，(C)が含まれている(図8(a)，(b) and (c))．それぞれの動作の合い間には直立姿勢(図8(d))を経由する．それぞれ(A)は手を上下に振る，(B)は肩に手を置く，(C)は手を前に押し出す動作を表現している．この実験では，座標系の原点位置と回転を図9にある“root”と呼ばれる観測部位に固定している．

実験の結果，PFAにより5次元の軸が選出された(図9)．PFAでのクラスタリングの結果を表1に示す．この実験で

[†] NSF EIA-0196217 の支援で開発された．

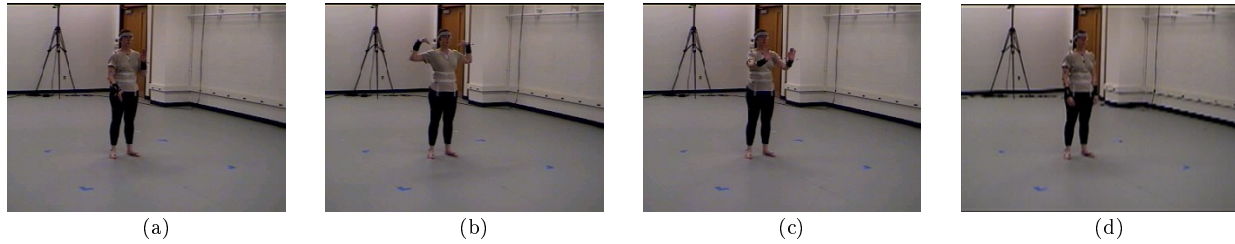


図 8 動作の情景：(a) 手を上下に振る，(b) 手を肩に置く，(c) 手を前に押し出す，(d) 直立姿勢．

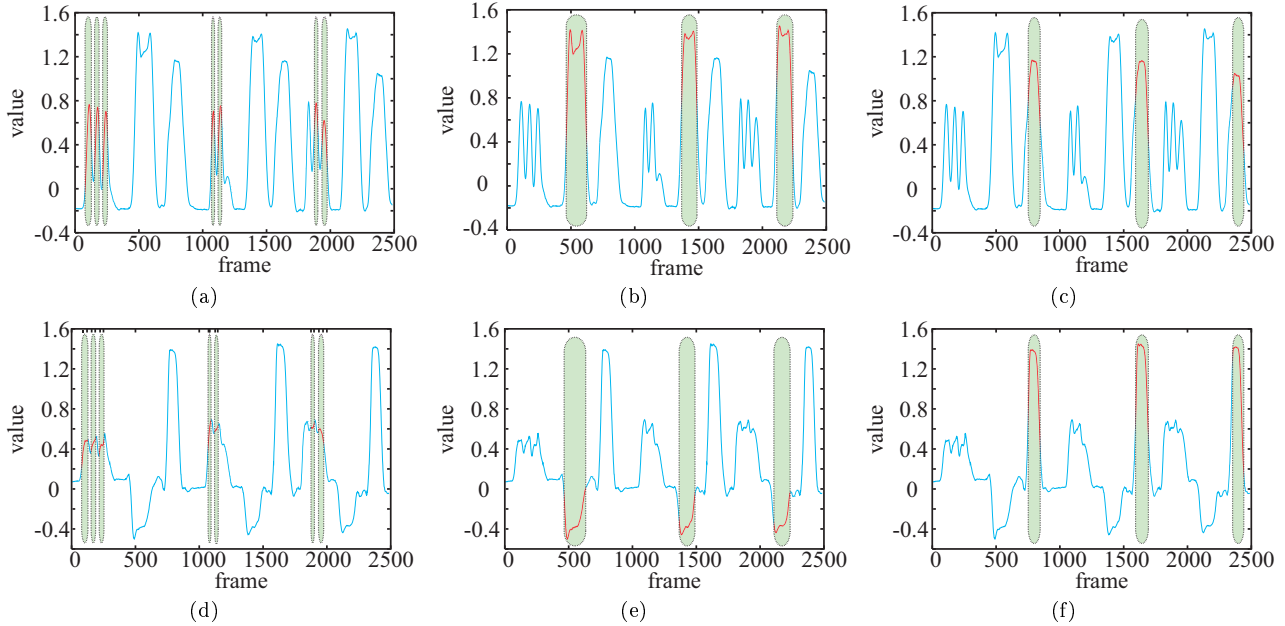


図 10 right elbow's y-axis (上段) と z-axis (下段) を統合したラベル列から抽出された 3 番目 (中央)，5 番目 (左)，7 番目 (右) のモチーフ．

表 1 PFA により選出された軸と各クラスターの要素

Cluster	Selected axis	Others
(1)	left hand's x	left elbow's x
(2)	left clavicle's y	head's x
		head's y
(3)	right humerus' z	∴
		left humerus' z
(4)	right elbow's y	right elbow's x
		left hand's y
(5)	right elbow's z	left elbow's y
		right hand's y
		left elbow's z
		left hand's z
		right hand's x
		right hand's z

は、モチーフ抽出の繰り返し処理回数として入力動作データから全ての頻出動作を抽出するのに十分な回数として経験的に 8 回と固定している。その結果、全ての軸の組み合わせを用いて統合されたラベル列から合計 $248 = (32 - 1) \times 8$

個のモチーフが抽出された。ここで 32 は選出された軸の組み合わせの数、-1 は軸の組み合わせから何も軸を選ばない場合を除くため、8 は繰り返し処理の回数である。抽出結果の例として、図 10 に選出された軸 (4) と (5) を統合したラベル列から抽出されたモチーフを示す。これらのモチーフは、それぞれ 3 番目、5 番目、7 番目の繰り返し処理によって抽出されたもので、動作 (B)、(A)、(C) に対応している。このほかに抽出されたモチーフのほとんどは、基本姿勢である直立姿勢に対応したものであった。図 8(b),(c),(e),(f) に示される 3 番目と 7 番目に抽出されたモチーフは、それぞれ一つの動作全体に対応しているが、図 8(a),(d) に示される 5 番目に抽出されたモチーフは一つの動作の中で繰り返し行われている動きの一つに対応している。これは、動作 (A) が手を上下に振るという繰り返し動作となっており、その繰り返し動作を構成する基本動作が頻出動作として抽出されたためである。本手法では、MDL 基準により、頻出動作のうち、繰り返し回数の多い動作がモチーフとして抽出される傾向にあるため、動作 (A) のような繰り返し動作全体をモチーフとして抽出することは難しい。

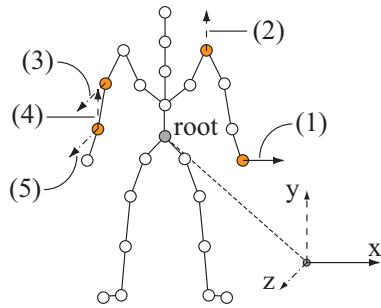


図 9 選出された軸と座標系 : (1)left hand's x-axis, (2)left clavicle's y-axis, (3)right humerus z-axis, (4)right elbow's y-axis, (5)right elbow's z-axis .

5. おわりに

本論文では、人物動作情報からモチーフ抽出のための手法を提案した。この手法は、PFA によって低次元化した動作情報から自動的に頻出動作をモチーフとして抽出する。実験の結果、本手法により全てのモチーフを効果的に抽出することができた。しかし、本手法ではまだ 2 つの問題がある。1 つは繰り返し動作全体を抽出するのが難しいということ。もう一つは、用いる動作情報によって PFA で選出される軸の組み合わせの数が大きくなりすぎ場合があるということである。

参考文献

- (1) M. Roussos, A.E. Johnson, J. Leigh, C.A. Vasilakis, C.R. Barnes, T.G. Moher, "NICE: combining constructionism, narrative and collaboration in a virtual learning environment," ACM SIGGRAPH Computer Graphics, Volum 31, Issue 3, pp.62-63, 1997.
- (2) P. Jeffrey, A, McGrath, "Sharing serendipity in the workplace," Proc. of the third international conference on Collaborative virtual environments, pp.173-179, 2000.
- (3) D. Arita, H. Yoshimatsu, D. Hayama, M. Kunita, R. Taniguchi, "Real-time Human Proxy: An Avatar-based Interaction System," CD-ROM Proc. of International Conference on Multimedia and Expo, 2004.
- (4) N. Date, H. Yoshimoto, D. Arita, R. Taniguchi, "Real-time Human Motion Sensing based on Vision-based Inverse Kinematics for Interactive Applications," ICPR'04, vol.3, pp. 318-321, 2004.
- (5) J. Lin, E. Keogh, S. Lonardi, P. Patel, "Finding Motifs in Time Series," Proc. of the 2nd Workshop on Temporal Data Mining, pp.53-68, 2002.
- (6) Chohen I., Tian Q., Zhou X. S., and Huang T. S. "Feature selection using principal feature analysis," ICIP'02, 2002.
- (7) P.D. Grunwald, "A tutorial introduction to the minimum description length principle," MIT Press, April 2005.
- (8) J. Lin, E. Keogh, S. Lonardi, B. Chiu, "A Symbolic Representation of Time Series, with Implications for Streaming Algorithms," DMKD03, pp.2-11, 2003
- (9) TANAKA.Y., IWAMOTO.K., & UEHARA.K., "Discovery of Time-Series Motif from Multi-Dimensional Data Based on MDL Principle," Machine Learning Vol.58 pp.269-300, 2005.