

モーションキャプチャデータからの特徴抽出による舞踊動作のセグメンテーション

園田 真史[†] 吉村 ミツ[‡] 八村 広三郎[†]

[†]立命館大学 理工学研究科 [‡]立命館大学 COE 推進機構

概要 モーションキャプチャシステムによって計測された身体動作のデータは、無形文化財としての舞踊などのデジタルアーカイブ化や動作解析の研究に利用されるようになってきている。本報告では、モーションキャプチャによる舞踊の身体運動データに対して、全体の身体動作をセグメンテーションするための手法を提案し、これを用いたクラシックバレエとストリートダンスの動作を対象とした実験と評価について述べる。身体の速さや移動方向の変化、身体の作る空間の大きさの変化など複数の特徴から抽出する手法について検討し、良好な結果が得られることが分かった。

Segmentation of Dancing Movement by Extracting Features from Motion Capture Data

Masafumi Sonoda[†], Mitsu Yoshimura[‡], and Kozaburo Hachimura[†]

[†]Graduate School of Science and Engineering, Ritsumeikan University

[‡]Centre for Promotion of the COE, Ritsumeikan University

Abstract Motion capture data come to be used for digital archive of traditional dance as an intangible cultural asset and a study of movement analysis by a motion capture system. This paper describes a method of segmentation of dancing movement from motion capture data and some of the experimental results for dancing movements. A method of segmentation to extract from some characteristics have been proposed by using information about velocity of body parts, a change of a movement direction, and a change of size of space made the body. Experiments showed the promising results.

1 はじめに

モーションキャプチャシステムによって計測された身体動作のデータは、伝統芸能や舞踊などの無形文化財のデジタルアーカイブ化や動作解析の研究に利用されるようになってきている。このようなアーカイブの中から、必要なデータを効率よく見つけだすためのインデックスや要約の作成、また、舞踊動作データの中から見せ場となる特徴のある一連の動作を見つけて出すことなどが求められる。

本研究は、モーションキャプチャによる舞踊の身体運動データに対して、以上のようなさまざまな処理を行うための基礎的処理として、全体の身体動作をいくつかの単位となる動作に区分（セグメンテーション）することを目標とし、そのための手法について考察する。また、その手法を実験により評価する。

動作の区切り方としては、純粋に身体の動きの物理情報から動きの切り替わった点を単位動作として区分

する物理レベルのセグメンテーションとその物理レベルで得られた結果からプリミティブな動作に区分する意味レベルのセグメンテーションがあるが、本論文ではセグメンテーションの第一段階として物理レベルでのセグメンテーションを対象とする。すなわち、必ずしも舞踊の表現内容や意味、対象舞踊に特有のスタイル等に基づいたセグメンテーションとはならない。意味レベルのセグメンテーションは物理レベルの結果を利用してその次のステップとして行うものとする。

2 セグメンテーションの手法

一般的には、身体全体の動作が停止した時点など、身体運動の速さに着目してセグメンテーションの候補を求めることが行われる [1, 2]。また [3] では、角速度が0となる点を見つけて出すことにより、動作をセグメンテーションしている。一方 [4] では、動作の最初から注目点までの身体各部の位置座標の系列に対

して主成分分析を行い、結果の固有値の変動から運動の性質が大きく変わるところを検出している。しかしながら、この手法では、各時点で主成分分析を行う必要があり計算コストがかかる。[5]では、身体は階層的に分割できるとして、各々の部分における運動エネルギーなど3つの運動パラメータを用いて検出している。

本研究では、身体の停止した時点、さらに、身体の動きの方向の変化する時点、および身体の形作る空間の大きさの変化する時点も検出し、最終的にこれらの結果を統合してセグメンテーションを行う。以下で、このそれぞれについて述べる。

なお、ここで取り扱う身体動作データは、図1に示すようなマーカーのXYZ座標値の時系列で表現されている。本研究で実験に用いるデータでは、クラシックバレエについては図1(a)のモデル、ストリートダンスについては図1(b)のモデルを使用している。

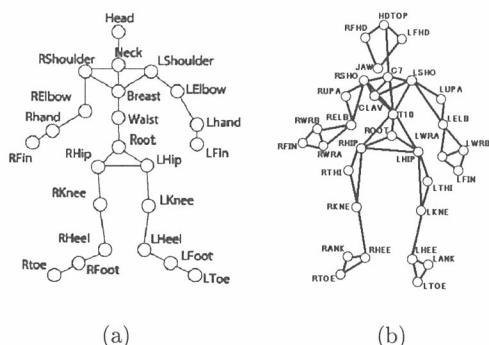


図 1: 用いた人体モデルとそのマーカー位置

2.1 速さを利用したセグメンテーション

本研究では身体の各部位で最も速い部位の速さが閾値以下の場合に動作が停止しているとみなす。このために、まず、フレーム i におけるマーカー k の速さ v_i^k について、このフレームにおける最速のマーカーの速さ V_i を求める。そして、 V_i が極大値を与える付近で最も速く動いているマーカーの速さの変化から動作の状態の変化を判断する。すなわち、 V_i を用いて以下の手順でセグメンテーション点の候補を求める。

1. V_i が十分に大きく、 $(V_{i-1} < V_i) \wedge (V_i > V_{i+1})$ を満たすフレーム i とそのとき v_i を与えるマーカー k に着目する。
2. i に最も近い、 $(v_{j-1}^k > v_j^k) \wedge (v_j^k < v_{j+1}^k)$ を満たすフレーム j を求める。ただし、 $i < j$ とする。
3. V_j が十分に小さい場合、フレーム j では動作は止まっているとみなして、このフレームをセグメンテーションの候補とする。

そうでない場合はまだ一連の動作が続いていると判断して現在の j を i にして 1. に戻る。

上述の手法は、ある時点における身体の各部位で最も速い部位に着目し、その部位の動作が停止し、かつその時点で全ての部位が停止するところまでを一連の動作とみなし、セグメンテーションの候補を求める手法である。

他の手法として、例えば速さ V_i が閾値以下で $(V_{i-1} > V_i) \wedge (V_i < V_{i+1})$ を満たすフレーム i を候補とすることが考えられるが、身体で最も速い部位が頻繁に変わった場合やほとんど動作がみられない時点がある場合に、上の条件を満たすフレームが数多く現れてしまい、実際には必要ではない候補も検出されてしまう。しかし、本手法では最も速い部位の速さが閾値以下になるセグメンテーションの候補を、常に上述 1. の部位に着目して求めることで、検出されないようにすることができる。

身体の各部位で最も速い部位の速さ V_i のグラフと、上記手法により得られるセグメンテーションの候補の例を図2に示す。図2中のグラフは横軸が時間を表し、縦軸が速さである。グラフの下にある垂直のバーがセグメンテーションの候補である。

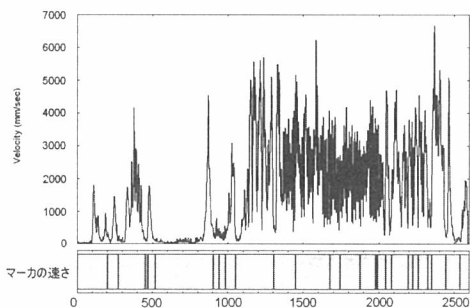


図 2: 速さを利用したセグメンテーション

そして、図2の一部(1200~1350フレーム)を拡大したものを図3に示す。図3中のグラフは、図2に示した速さ V_i のグラフと、セグメンテーションの候補となった垂直のバーの時点で着目していた部位、ここでは LToe マーカーの速さ v_i^{LToe} のグラフである。グラフの下にある垂直のバーがセグメンテーションの候補である。

図3でセグメンテーションの候補となった時点では V_i の値も小さく、またその直前で、 V_i のグラフと v_i^{LToe} のグラフが重なっており、LToe マーカーに着目して検出している。また、図3中にはその前後にも V_i のグラフと v_i^{LToe} のグラフが重なっている時点があるが、その直後において V_i の値が大きくなっているので、検出されていない。

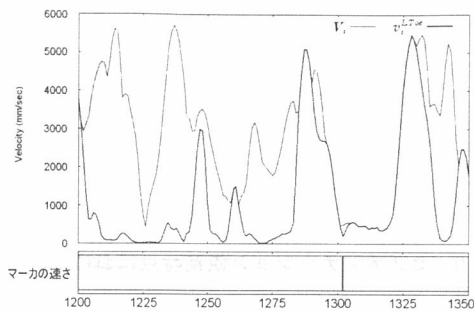


図 3: 速さを利用したセグメンテーション

2.2 移動方向の変化を利用したセグメンテーション

体の移動方向の変化は、腰付近に付けた Root マーカの水平面内の移動の軌跡から求められる離散曲率に基づいて検出する。フレーム i での Root マーカの水平面内の位置と $i-L$ フレームでの位置を結んで得られるベクトルを $\mathbf{a}_i$ 、フレーム i での位置と $i+L$ フレームでの位置を結んで得られるベクトルを $\mathbf{b}_i$ とする。離散曲率 C_i は、ベクトル $\mathbf{a}_i$ と $\mathbf{b}_i$ のなす角 θ_i の余弦

$$C_i = \cos \theta_i = \frac{\mathbf{a}_i \cdot \mathbf{b}_i}{\|\mathbf{a}_i\| \|\mathbf{b}_i\|}$$

で与えられる。ただし、 $\|\mathbf{a}_i\| < T$ または $\|\mathbf{b}_i\| < T$ の場合は身体 (Root マーカ) がほとんど移動していないとして離散曲率の計算を行わない。ここでは、 $L = 15$ 、 $T = 100\text{mm}$ としている。そして、離散曲率 C_i が閾値以上の大きさの極大値をもつ点をセグメンテーション点の候補とする。

離散曲率 C_i のグラフと、上記手法により得られるセグメンテーションの候補の例を図 4 に示す。図 4 中のグラフは横軸が時間を表し、縦軸が離散曲率 C_i である。図 2 同様、グラフの下にある垂直のバーがセグメンテーションの候補である。

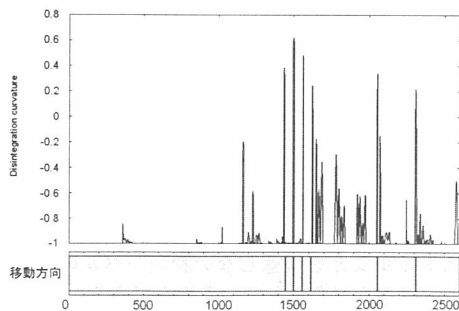


図 4: 移動方向の変化を利用したセグメンテーション

この方法によりセグメンテーションの候補として抽出された結果の例を、図 5 に示す。平面上の点が、Root マーカの水平面内の軌跡であり、黒丸印がセグメンテーションの候補として検出された時点を示している。

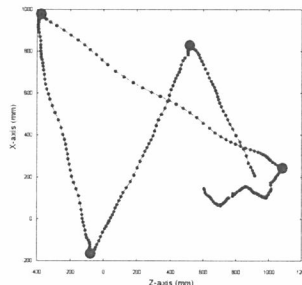


図 5: Root マーカの軌跡とセグメンテーション点 (1400~1700 フレーム)

2.3 身体を作る空間の変化を利用したセグメンテーション

身体を作る空間の大きさを求める方法として、文献 [6] では身体全体につけたマーカを包み込む 3 次元凸包を用いている。本研究においても、この手法を利用して、3 次元凸包で表した身体を作る空間 (図 6) の体積のフレームごとの変化に基づいてセグメンテーション点の候補を求める。

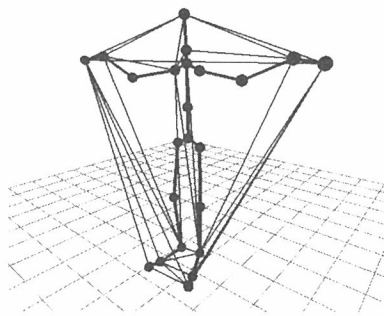


図 6: 3 次元凸包を用いた身体を作る空間

あるフレーム i における 3 次元凸包の体積を V_i とする。 $D_i = |V_i - V_{i-1}|$ とするとき、以下の 2 つの条件を満たすフレーム i を、セグメンテーション点の候補とする。

1. $(D_{i-1} > D_i) \wedge (D_i < D_{i+1})$ となること
2. t を閾値として $(D_i < t) \wedge (D_{i+1} > t)$ であること。

3次元凸包の体積の変化 D_i を用いてセグメンテーションすることは、体積そのもので判断するのではなく、体積の変化に着目することで、体積の変化の小さい時点から急激に大きく変化していく時点と判断することが容易であると考え。

3次元凸包の体積の変化 D_i のグラフと、上記手法により得られるセグメンテーションの候補の例を図7に示す。図7中のグラフは横軸が時間を表し、縦軸が3次元凸包の体積の変化 D_i である。図2や図4同様、グラフの下にある垂直のバーがセグメンテーションの候補である。

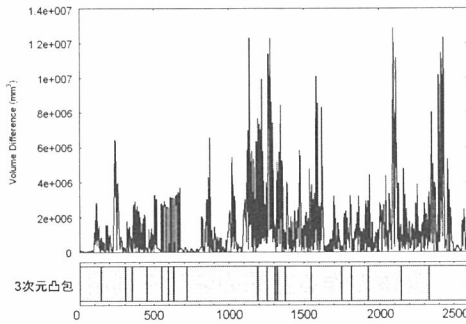


図7: 身体を作る空間の変化を利用したセグメンテーション

また、図7においてセグメンテーションの候補となった時点を拡大したものを図8に示す。グラフの下にある垂直のバーがセグメンテーションの候補である。

図8中には、上述の条件1.である $(D_{i-1} > D_i) \wedge (D_i < D_{i+1})$ を満たす箇所は数箇所あるが、その中で条件2.を満たすのが垂直のバーの2箇所のみである。

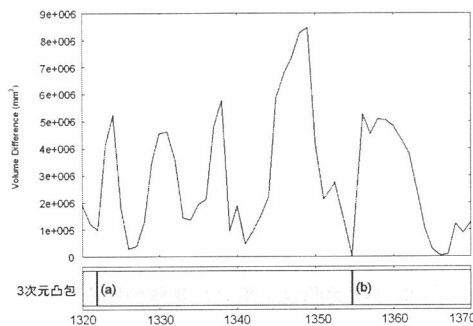


図8: 身体を作る空間の変化を利用したセグメンテーション (部分を拡大したもの)

そして、図8においてセグメンテーションの候補となった時点の3次元凸包を図9に示す。左側の図(a)が図8の左側の垂直のバー(a)の時点、右側の図(b)が図8の右側の垂直のバー(b)の時点である。

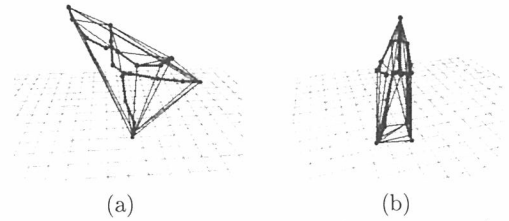


図9: セグメンテーション候補時点における3次元凸包

2.4 各結果の統合

上述した各手法により得られたセグメンテーション候補を統合する。ここでは、それぞれの手法によって求まる候補点の集合の和集合を求める。ただし、得られた候補点間のRootマーカの移動距離が閾値より小さいものはこれらを1つの候補点とみなすものとする。

3 実験と評価

以上で述べたセグメンテーション手法の妥当性を確認するため、ここではクラシックバレエの動作データ、およびストリートダンスの動作データを用いて実験を行った。

3.1 クラシックバレエ

まず、クラシックバレエについて知識を有していない8人の被験者にクラシックバレエの動作データをスティックフィギュアによるアニメーションとして提示し、それぞれの目視によって各人の直感的な判断に基づいてセグメンテーションを行ってもらう。この8人の被験者が定めた区切り位置に基づき、さらに以下を満たすものを、人間が判断した「真の」セグメンテーション点の集合(以下、リモートセンシングや文字認識等の用語に習い“ground truth”と呼ぶ)とする。

- 4人以上が挙げた箇所
- 4人未満の人が挙げた箇所、例えば動作が完全に止まっているなど特徴的な箇所

なお、バレエの知識を有しない者による結果を用いるのは、1章で述べたように、本研究では舞踊の意味やスタイルなどを考慮しない物理レベルのセグメンテーションを対象としているからである。

このようにして得られたデータセグメンテーション点の“ground truth”と、提案手法により得られるセ

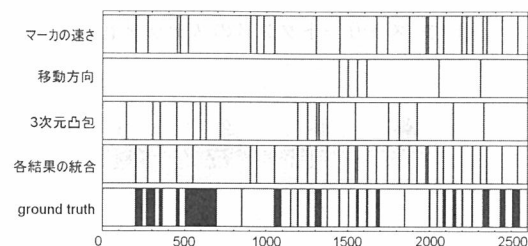


図 10: バレエ動作のセグメンテーション

図 10 の各結果の統合として得られた結果のセグメンテーション点での姿勢と、その中の一連の動作のスナップショットを図 11 に示す。図 11 の一段目左端、三段目右端がセグメンテーション点での姿勢、その他がその一連の動作である。なお、図 11 のスナップショットの図は一連の動作の概略を把握するために示しており、各スナップショットは一定の時間間隔ではない。以下のスナップショットの図についても同様である。

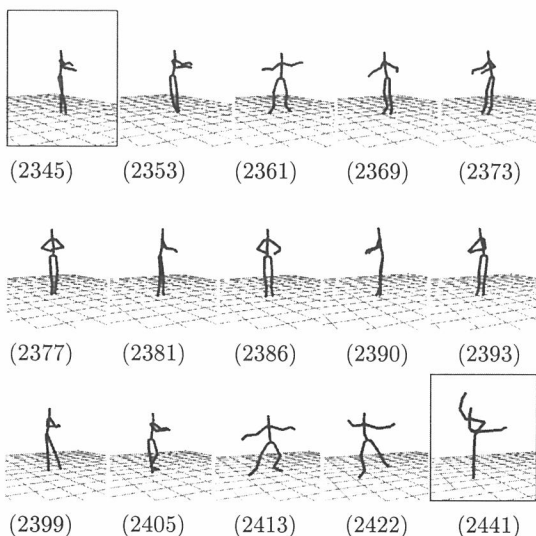


図 11: バレエ動作から抽出された 1 つのセグメントの例

() 内の数字はフレーム番号

“ground truth”を用いて，セグメンテーションの処理手法の評価を行った．評価方法には，情報検索の分野で使われる，recall(再現率)と precision(適合率)

の尺度を用いた [4]. “ground truth” とした位置の集合を R , 処理手法で得られた結果の集合を A とすると, recall と precision は,

$$precision = \frac{|A \wedge R|}{|A|}, recall = \frac{|A \wedge R|}{|R|}$$

で表される。ただし、 $|\cdot|$ は集合の要素の個数を表す。この時、処理手法で得られた結果と“ground truth”との相違が5フレーム以下の場合、および、2つのフレームの間の姿勢がほぼ同じである場合は、同一のフレームであるものとする。評価の結果を表1に示す。

表 1: セグメンテーションの処理手法の評価

処理手法	precision	recall
マーカの速さ	0.64	0.59
移動方向	0.83	0.19
3次元凸包	0.74	0.52
各結果の統合	0.72	0.85

表 1 の recall のデータによれば、それぞれの処理手法では、人間の求めたものの一部しか検出できていないが、各手法の結果を統合することにより、良好な結果が得られるようになったことが分かる。一方、precision に着目すると、マーカーの速さによる手法では人間の求めたものの以上に多くの点を検出していることが分かる。

3.2 ストリートダンス

次にストリートダンスの動作データについてのセグメンテーションを行った。今回は、ストリートダンスの中のブレイクダンスについて、基本ステップを数種類組み合わせで踊ったデータを使用する。

今回使用したストリートダンスのデータは、いくつかのステップをそれぞれ異なった組み合わせで踊っている A, B, C の 3 種類の動作データである。

動作データ A は、両手を広げる動作（トップロック）、スキップをするような動作（ブロンクスステップ）、腰をひねるような動作（エントリー）を順に踊ったデータである。

動作データ B は、立ったりしゃがんだりする動作（アップロック）、ブロンクスステップ、手を床につける動作（2 ステップ）、しゃがんで両足を交互に動かす動作（6 歩）、後ろに手をついて片足を上げる動作（シフト）を順に踊ったデータである。

動作データ C は、2 ステップ、エントリー、6 歩、2 ステップ、エントリー、トップロックの順番で踊ったデータである。

これらの7種類のステップのスナップショットを、
図12～図18に示す。

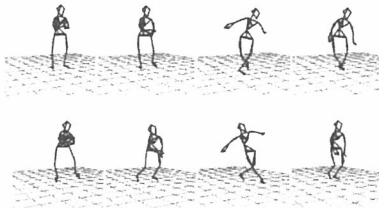


図 12: ストリートダンスのステップ (トップブロック)

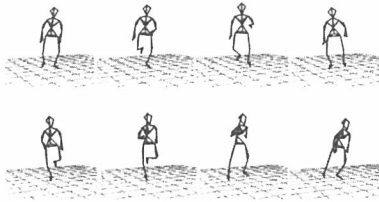


図 13: ストリートダンスのステップ (ブロンクスステップ)

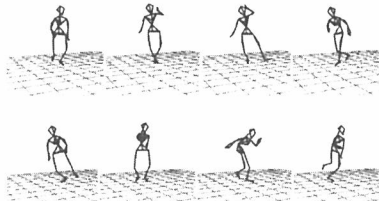


図 14: ストリートダンスのステップ (エントリー)

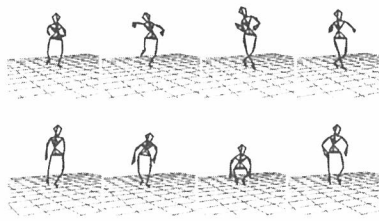


図 15: ストリートダンスのステップ (アプロック)

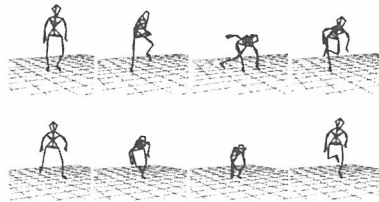


図 16: ストリートダンスのステップ (2 ステップ)

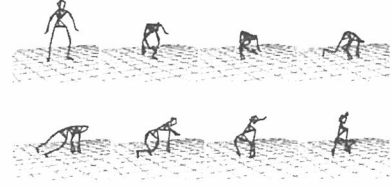


図 17: ストリートダンスのステップ (6 歩)

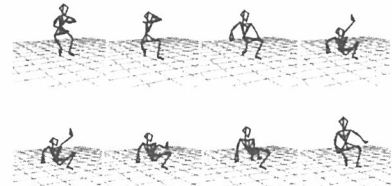


図 18: ストリートダンスのステップ (シフト)

クラシックバレエ同様に得た、データセグメンテーション点の“ground truth”と、提案手法により得られるセグメンテーション結果を比較したものを図 19 に示す。

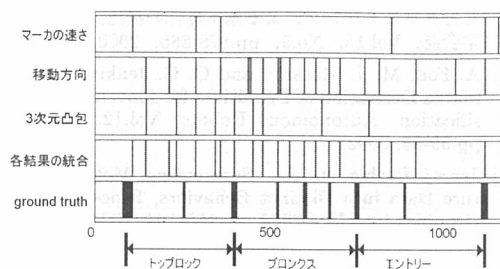
図 19 の各結果の統合として得られた結果のセグメンテーション点での姿勢と、その中の一連の動作のスナップショットを図 20 に示す。図 20 の (a) から (c) それぞれの両端がセグメンテーション点での姿勢である。

クラシックバレエと同様に、セグメンテーションの処理手法の評価を行った。評価について、2つのフレームの間の姿勢がほぼ同じ場合に同一のフレームであるとみなすのはクラシックバレエと同じであるが、クラシックバレエとストリートダンスではサンプリングレートが異なっているので処理手法で得られた結果と“ground truth”との相違は 10 フレーム以下の場合に同一とみなしている。評価の結果を表 2 に示す。

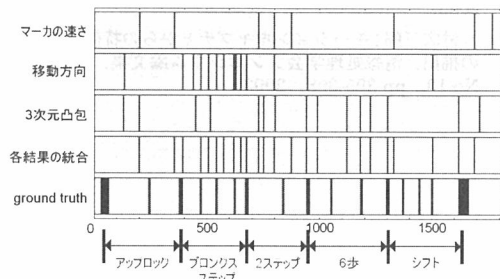
表 2 の recall のデータによると、クラシックバレエの結果と同様にそれぞれの処理手法では、人間の求めたものの一部しか検出できていないが、各手法の結果を統合することにより、良好な結果が得られるようになったことが分かる。一方、precision に関しては、あまり十分な結果を得ることはできなかった。

以上は、対象の舞踊の表現上の意味や特性を考慮しない物理レベルのセグメンテーションについて評価を行ったが、ここでは、念のため、この手法がストリートダンスの各ステップの切り替わる点を抽出しているかどうかを確認してみた。この結果を表 3 に示す。表 3 は、左側にはステップを表しており、2つのステップ間で、動きが切り替わった時点を検出できたかどうかを右側に表している。

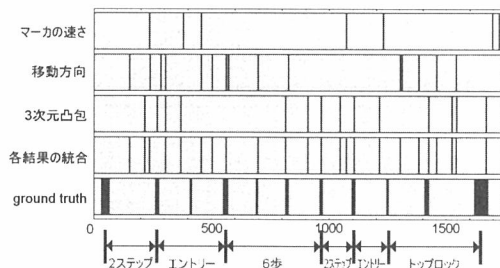
表 3 より、ダンスのステップの切り替わった箇所の



(a) 動作データ A



(b) 動作データ B



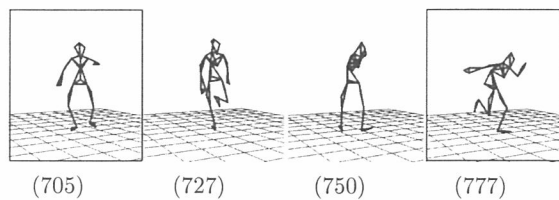
(c) 動作データ C

図 19: ストリートダンスのセグメンテーション

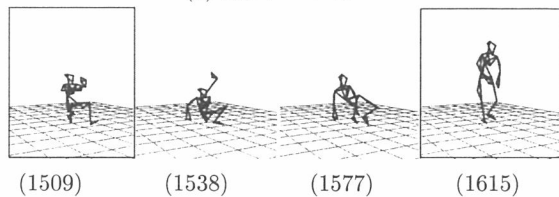
検出についても良好な結果を得ていることが分かる。しかし、ステップがエントリーに切り替わる時点は3箇所のうち2箇所で、検出できていない。また、ステップがエントリーから別のステップに切り替わる時点においても、1箇所検出されていない時点があった。

4 まとめ

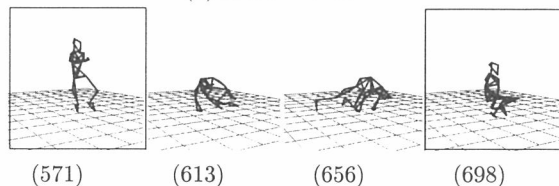
モーションキャプチャによる舞踊の身体運動データから複数の特徴を抽出してセグメンテーションを行う手法を検討し、人間の目視によるセグメンテーションから得られた“ground truth”との比較による評価を行った。これにより、複数の手法により得られる結果を統合することで、人間の評価に近い良好なセグメンテーションの結果が得られることが分かった。



(a) 動作データ A



(b) 動作データ B



(c) 動作データ C

図 20: ストリートダンスの動作データから抽出された1つのセグメントの例

() 内の数字はフレーム番号

表 2: セグメンテーションの処理手法の評価

(a) 動作データ A

処理手法	precision	recall
マーカの速さ	0.50	0.33
移動方向	0.16	0.33
3次元凸包	0.11	0.11
各結果の統合	0.28	0.67

(b) 動作データ B

処理手法	precision	recall
マーカの速さ	0.25	0.13
移動方向	0.50	0.44
3次元凸包	0.33	0.31
各結果の統合	0.46	0.69

(c) 動作データ C

処理手法	precision	recall
マーカの速さ	0.29	0.18
移動方向	0.13	0.18
3次元凸包	0.43	0.55
各結果の統合	0.30	0.64

表 3: ステップが切り替わった箇所の検出結果

(a) 動作データ A

ステップの種類	検出結果
トップロック	○
ブロンクスステップ	
ブロンクスステップ	×
エントリー	

(b) 動作データ B

ステップの種類	検出結果
アップロック	○
ブロンクスステップ	
ブロンクスステップ	○
2 ステップ	
2 ステップ	○
6 歩	
6 歩	○
シフト	

(c) 動作データ C

ステップの種類	検出結果
2 ステップ	×
エントリー	
エントリー	○
6 歩	
6 歩	○
2 ステップ	
2 ステップ	○
エントリー	
エントリー	×
トップロック	

今後の課題としては、更に他の動作データについて実験を重ねること、および、これらの物理レベルの結果をふまえて、それぞれの舞踊の特性や意味等を考慮した意味レベルのセグメンテーション手法の検討を行うことである。

謝辞 本研究は、文部科学省 21 世紀 COE プログラム「京都アート・エンタテインメント創成研究」、文部科学省オープンリサーチセンター補助事業、および、科学研究費補助金基盤研究 (B)16300035 の支援により行われた。

参考文献

- [1] 湯川崇, 小原直子, 玉本英夫: モーションキャプチャデータからの基本動作分割点の自動検出, 情報処理学会論文誌, Vol.45, No.4, pp.1198-1201, 2004.
- [2] 大崎竜太, 嶋田光臣, 上原邦昭: 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, 人工知能学会誌, Vol.15, No.5, pp.878-886, 2000.

- [3] A. Fod, M. J. Mataric, and O. C. Jenkins: Automated Derivation of Primitives for Movement Classification, Autonomous Robots, Vol.12, Issue 1, pp.39-54, 2002.
- [4] Jernej Barbic et al.: Segmenting Motion Capture Data into Distinct Behaviors, Proceedings of Graphics Interface 2004, pp.185-194, 2004.
- [5] Kanav Kahol, Priyamvada Tripathi, Sethuraman Panchanathan: Gesture Segmentation in Complex Motion Sequences, Proceedings IEEE International Conference in Image Processing, Vol.3, pp.105-108, 2003.
- [6] 八村広三郎: モーションキャプチャからの特徴フレームの抽出, 情報処理学会シンポジウム論文集, Vol.2002, No.13, pp.305-308, 2002.