

動作特徴量を用いたモーションキャプチャデータの検索 Retrieval of Motion Capture Data Using Motion-characteristic Quantities

○大和田慧*, 三浦武*, 松尾健史*, 田島克文*

○Satoru Owada*, Takeshi Miura*, Kenshi Matsuo*, Katsubumi Tajima*

*秋田大学

*Akita University

キーワード： モーションキャプチャ (motion capture), データベース (data base),
検索 (retrieval), 動作特徴量 (Motion-characteristic quantity)

連絡先： 〒010-8502 秋田県秋田市手形学園町 1-1 秋田大学大学院 理工学研究科
三浦 武, Tel. : (018)889-2329, Fax. : (018)837-0406, Email : miura@ipc.akita-u.ac.jp

1. はじめに

現在, アニメーション制作の分野においては, モーションキャプチャ技術によって人体の複雑な動きを精度よく作成できるようになった¹⁾. 一方で, モーションキャプチャデータの測定には高価な専用機器が必要であり, さらに, 人体にセンサあるいはマーカを取り付けるという手間がある. また, 測定するためにある程度広さのある場所を要するため, 測定環境を整えるのが困難かつ費用面での負担も小さくないという問題点を持つ¹⁾.

これらの問題点を解決するため, 一度取得したデータの再利用を可能としたモーションキャプチャデータベースの利用が提案されており¹⁾, 実際にいくつかの大規模なモーションキャプチャデータベースが存在している^{2), 3)}. これらのデータベースには膨大な量のデータ数が保存されているため, その管理手法が重要と

なる. データベースの管理手法の一つとして, 膨大なデータの中から必要なデータを見つけ出すための情報検索⁴⁾がある. 情報検索には種々のアプローチがあるが, その中で索引付けを行う方法は高速な検索が可能な手法として知られている.

モーションキャプチャデータの検索に関しては, 既にいくつかのアプローチが提案されている. それらは大きく **content-base** な方法と **style-base** な方法に分類できる⁵⁾. 後者については, 各々のモーションキャプチャデータに索引付けを行った方法がいくつか提案されている^{6), 7)}. その中で, 文献⁷⁾においては, 人体の姿勢変化に対応した複数の特徴量により構成される特徴ベクトルを索引として用いており, 少ない次元数の特徴ベクトルによって, 比較的精度の良い検索が可能となることが示されている. しかし, 姿勢変化の空間的な変化は考慮さ

れているものの、その時間的な変化である速度の情報は取り入れられておらず、よって、動作速度のみが異なる類似姿勢の動作の判別が難しいことも指摘されている。

上記の欠点を改善するためには、動作速度を取り入れた特徴量の導入が望ましい。幸いにも、文献8)において、動作速度に対応した(空間的特徴ではなく)時間領域における特徴量の導出が提案されており、これを取り入れることによって、上記の改善が可能となることが期待できる。よって、本研究においては、空間的な姿勢変化の特徴量に加えて、時間領域における動作速度の特徴量を導入し、その効果と問題点を検討する。

2. 動作特徴量の抽出法

2.1 動作特徴量の概要

前章で述べたように、本研究では文献7)の空間的特徴量から成る特徴ベクトルを用いる。ベクトルの各成分は、3つの運動軸(前額軸、垂直軸および矢状軸)⁹⁾および3つの運動面(前額面、矢状面、横断面)⁹⁾のそれぞれに対応した人体の広がり具合を定量的に示す。一方で、文献8)で提案されている時間的特徴量は動作時のリズムの特性を動作速度に基づいて抽出したものであり、「拍の強さ」と「リズムの複雑さ」の2つより成る。その導出過程を以下に示す。

2.2 空間的特徴量の算出法

まず、空間的特徴量の導出法を示す。ここでは、人体の関節位置の分布を、Fig. 1に示すように人体(ここでは骨盤)に固定した座標系によって記述する。ただし、体格の違いの影響を減らすために関節位置の値は身長で規格化される。次に、関節座標の広がり具合を分散および共分散を用いて(1)式のように示す。

$$\Sigma(n) = \begin{bmatrix} \sigma_{xx}(n) & \sigma_{xy}(n) & \sigma_{xz}(n) \\ \sigma_{yx}(n) & \sigma_{yy}(n) & \sigma_{yz}(n) \\ \sigma_{zx}(n) & \sigma_{zy}(n) & \sigma_{zz}(n) \end{bmatrix} \quad (1)$$

$$\sigma_{\gamma\eta}(n) = \frac{1}{J} \sum_{j=1}^J \{p_{j,\gamma}(n) - \bar{p}_{\gamma}(n)\} \{p_{j,\eta}(n) - \bar{p}_{\eta}(n)\}$$

$$\bar{p}_{\gamma}(n) = \frac{1}{J} \sum p_{j,\gamma}(n) \quad (\gamma, \eta : x, y \text{ or } z)$$

ここで $p_{j,\gamma}(n)$ は第 n フレームにおける第 j 関節の γ 座標、 J は選んだ関節の総数であり、ここでは Fig.1 において大きな黒点で示した 19 関節を用いた。これらを用いて各フレームにおける人体姿勢を特徴づける特徴ベクトルを、Fig.1 に示すように 3 軸 3 平面に対応した特徴量を用いて(2)式によって表す。

$$\begin{aligned} \mathbf{f}(n) &= [f_1 \quad f_2 \quad f_3 \\ &\quad f_4 \quad f_5 \quad f_6]^T \\ &= [\sigma_{xx}(n) \quad \sigma_{yy}(n) \quad \sigma_{zz}(n) \\ &\quad \sigma_{xy}(n) \quad \sigma_{yz}(n) \quad \sigma_{zx}(n)]^T \end{aligned} \quad (2)$$

さらに、動作の時系列データ全体に対して、以下の 12 次元ベクトル \mathbf{F} を求め、これを 1 つ

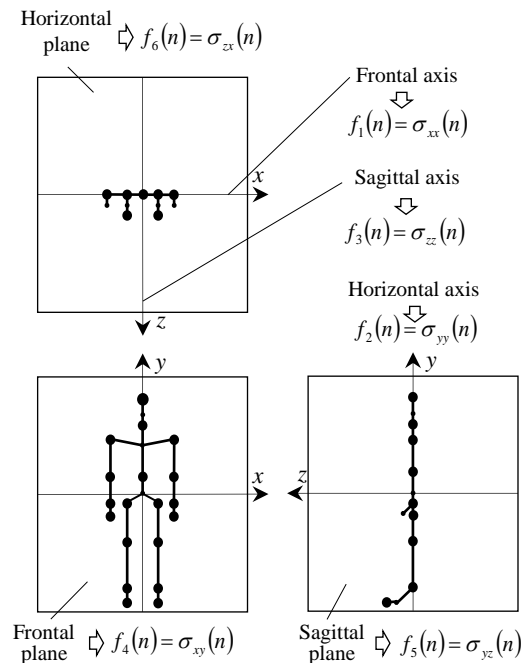


Fig.1 Quantification of the spatial distribution of body segments.

のモーションキャプチャデータ全体を特徴づける特徴ベクトルとする.

$$\begin{aligned}
\mathbf{F} &= [F_1 \ F_2 \ \dots \ F_{12}]^T \quad (3) \\
&= [\bar{f}_1 \ \bar{f}_2 \ \dots \ \bar{f}_6 \ \bar{s}_1 \ \bar{s}_2 \ \dots \ \bar{s}_6]^T \\
&= [\sigma_{xx \text{ mean}} \ \dots \ \sigma_{zx \text{ mean}} \ \sigma_{xx \text{ SD}} \ \dots \ \sigma_{zx \text{ SD}}]^T \\
\bar{f}_i &= \frac{1}{N} \sum_{n=1}^N f_i(n) \quad (= \sigma_{\gamma \text{ mean}}) \\
\bar{s}_i &= \sqrt{\frac{1}{N} \sum_{n=1}^N \{f_i(n) - \bar{f}_i\}^2} \quad (= \sigma_{\gamma \text{ SD}})
\end{aligned}$$

\mathbf{F} の各成分は、3つの運動軸および3つの運動面に対応する.

2.3 時間的特徴量の算出法

次に、時間的特徴量の導出を示す. この場合も Fig. 1 に示す人体モデルを用いる. ただし、人体に固定された座標系において実際に変位を与える関節である両肩、両肘、両手首、両手の指先、両膝、両足首、両つま先、首および頭の16関節のみを使用する⁸⁾. モーションキャプチャデータから得られた関節位置データにガウシアンフィルタ(遮断周波数: 9.0Hz)を適用し、適用後に改めて得られた関節位置データから(4)式を用いて動作速度時系列データの導出を行う⁸⁾.

$$v(n) = \frac{\sqrt{\sum_{j=1}^{J'} \sum_{\gamma=x,y,z} \{p_{j,\gamma}(n+1) - p_{j,\gamma}(n)\}^2}}{\Delta t} \quad (4)$$

ここで、 $p_{j,\gamma}(n)$ は第 n フレームにおける第 j 関節の位置座標、 Δt はサンプリング時間、 J' は関節の総数($J'=16$)を表している. 得られた時系列データから(5)式を用いて拍の強さ q_{BI} を算出する. ここでは、拍の強さが速度変動の要素および拍の間隔の2つの要素から構成されるとみなし⁸⁾、それぞれの重みを変更できることを考慮し、文献8)のEq.(2)を修正した文献10)のEq.(2)を用いる.

$$q_{\text{BI}} = \frac{1}{2} \log \frac{\sum_{n=1}^N \{v(n) - v_0(n)\}^2}{N} - A \log(\tau \Delta t) \quad (5)$$

ここで、 τ は拍子の間隔($v(n)$ の自己相関関数の最初の正のピークを示すフレーム番号を用いている)、 $v_0(n)$ は $v(n)$ の移動平均(周期 τ)、 A は定数($A=0.25$)である. q_{BI} の値が大きいほど拍が強いことを示す.

リズムの複雑さ q_{ApEn} は $v(n)$ から得られる Approximate Entropy¹¹⁾より(6)式を用いて算出される.

$$\begin{aligned}
\boldsymbol{\mu}(n) &= [v(n) \ v(n+\tau') \ \dots \ v(n+(m-1)\tau')]^T \\
d(\boldsymbol{\mu}(n), \boldsymbol{\mu}(j)) &= \max_{k=1,2,\dots,m} (|v(n+(k-1)\tau') - v(j+(k-1)\tau')|) \\
C_n^m &= \frac{\sum_{j=1}^{N-(m-1)\tau'} \theta(r - d(\boldsymbol{\mu}(n), \boldsymbol{\mu}(j)))}{N - (m-1)\tau'} \\
\Phi^m &= \frac{\sum_{n=1}^{N-(m-1)\tau'} \log C_n^m}{N - (m-1)\tau'} \\
q_{\text{ApEn}} &= \Phi^m - \Phi^{m+1} \quad (6)
\end{aligned}$$

ここで、 $\tau' = \text{round}(0.2\tau)$ 、 $m=3$ 、 $r=0.25\sigma$ (σ は標準偏差)であり、 $\theta(x)$ はHeaviside関数である. q_{ApEn} が大きいほどより複雑なリズム特性を示す.

実際のモーションキャプチャデータについては、対象となる動作領域を抽出するため、開始および終了フレーム付近の動作の少ない領域($v(n)$ がその全体の平均値から標準偏差の1/4だけ引いた値より小さい領域)をカットし、取り出したデータの開始および終了フレームをそれぞれ第1フレームおよび最終フレーム(第 N フレーム)と改めて定める.

2.4 データ間距離の算出法

第2.2節で述べたように、空間的特徴量の特徴ベクトルは12次元となり、一方で、時間的な特徴量は第2.3節で述べた拍の強さとリズム

の複雑さで構成される 2 次元特徴ベクトルにより表される. 結果として, 各モーションキャプチャデータは 12+2=14 次元空間にプロットされることになる.

類似検索においては, 検索質問 (クエリ) ⁴⁾ として, ユーザが任意のサンプル動作を与え, それに対応する特徴ベクトルを索引として類似特徴ベクトルを持つデータが抽出される. 各データは前述のように 14 次元空間にプロットされるが, このうち 12 次元は空間的特徴量が占めているため, 2 次元のみである時間的特徴量に比べてその比重が極端に大きくなってしまふ. これを避けるため, 本研究では以下のような処理を行う.

与えられたサンプルの特徴ベクトルとデータベース内の各データの特徴ベクトルの類似度は, ベクトル間の距離によって評価できる. 14 次元空間における通常のユークリッド距離を用いると, 前述のように空間的特徴量が過剰な影響を及ぼす. よって, ここではこれを緩和するため, 次式のような重み付けユークリッド距離を用いる.

$$d_{ij} = \sqrt{d_{Sij}^2 + d_{Tij}^2} \quad (7)$$

$$d_{Sij}^2 = \frac{w_S}{12} \sum_{k=1}^6 \left\{ (\bar{f}'_{ki} - \bar{f}'_{kj})^2 + (\bar{s}'_{ki} - \bar{s}'_{kj})^2 \right\}$$

$$d_{Tij}^2 = \frac{w_T}{2} \left\{ (q'_{BIi} - q'_{BIj})^2 + (q'_{ApEni} - q'_{ApEnj})^2 \right\}$$

ただし, \bar{f}'_{ki} は第 i モーションキャプチャデータの空間的特徴ベクトル \mathbf{F} の第 1~第 6 成分, (は全データに渡って平均 0, 標準偏差 1 に標準化されていることを示す), \bar{s}'_{ki} は同じく第 7~第 12 成分, q'_{BIi} は標準化された第 i データの拍の強さ, q'_{ApEni} は同じくリズムの複雑さ, w_S は空間的特徴量の重み, w_T は時間的特徴量の重みを示す.

上式において, w_S と w_T の比を変化させることによって空間的特徴量と時間的特徴量の重みを変化させることができる. 次章において,

この比をどのような値にすれば良いかを, 実際に多数のカテゴリに分類される大量のモーションキャプチャデータを用いた解析によって検討する. 具体的には, これらの値を

$$w_S + w_T = 1, \quad w_S \geq 0, \quad w_T \geq 0$$

を満足する範囲で変化させ, その時の検索精度の変動によって最適値の検討を行う.

3. 評価

前章で述べたように, ここでは多数のモーションキャプチャデータを用いる. 具体的には, Table 1 に示す 17 カテゴリに分類される 138 個のデータ (Carnegie-Mellon Mocap Database²⁾

Table 1 Mocap data used in the evaluation.

Label	Category	Data	Number of data
A	Walk	07_01-07_03 , 07_06-07_11	9
B	Walk(slow)	07_04, 07_05, 08_04, 37_01	4
C	Walk (on uneven terrain)	36_10-36_20	11
D	Marching	138_01-138_10	10
E	Run	09_01-09_09	9
F	Jump	118_01-118	10
G	Climbradder	13_33, 13_34, 14_33-14_35	5
H	Golf(swing)	64_01-64_10	10
I	Soccer (kick ball)	10_01-10_03 , 10_05 , 10_06, 11_01	6
J	Basketball (forward dribble)	06_02-06_05	4
K	Boxing	14_01-14_03 , 15_13 , 17_10	5
L	Modern dance	05_02-05_14	13
M	Chicken dance	18_15, 19_15, 20_01, 21_01, 143_34	5
N	Salsa dance	61_01-61_10	10
O	Breaking	85_01-85_08, 85_10	9
P	Charleston	93_03-93_06, 93_08	5
Q	Indian dance	94_01-94_13	13

Total : 138

Downloaded from <http://mocap.cs.cmu.edu.2>

Table 2 Evaluation of motion-capture data retrieval.

Weightings ($w_S:w_T$)	Error										
	1:0	0.9:0.1	0.8:0.2	0.7:0.3	0.6:0.4	0.5:0.5	0.4:0.6	0.3:0.7	0.2:0.8	0.1:0.9	0:1
A_e [%]	92.75	94.93	94.20	94.20	94.93	93.48	91.3	89.86	89.86	85.51	57.97
A	1	1	1	1	1	1	1	1	1	2	2
B	3	2	2	2	2	2	2	2	2	2	4
C	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	1	5
F	0	0	0	0	0	0	0	0	1	1	5
G	0	0	0	0	0	1	1	3	1	1	2
H	0	0	0	0	0	0	0	0	0	0	3
I	0	0	0	0	0	0	0	0	0	0	4
J	0	0	0	0	0	0	0	0	0	1	3
K	1	0	0	0	0	0	1	1	1	1	1
L	2	2	2	2	1	1	1	1	2	4	10
M	0	0	0	0	0	0	0	1	0	1	3
N	0	0	0	0	0	0	0	0	0	1	2
O	0	0	1	1	1	1	1	1	1	1	7
P	3	2	0	2	2	3	5	4	5	4	4
Q	0	0	0	0	0	0	0	0	0	0	3

よりダウンロード)を用いている。これらに対して、最近傍法¹²⁾を用いた 1-fold cross validation を実施する¹²⁾。このとき、(w_S, w_T) を (0, 1), (0.1, 0.9), ..., (0.9, 0.1), (1, 0) と 0.1 刻みで変化させ、それに対して上記の処理を実施して Empirical Accuracy¹²⁾ A_e の値を求めた。

得られた結果を Table 2 に示す。 A_e の値を見ると、(w_S, w_T) を (1, 0) とした空間的特徴量のみを用いた場合と比較して、 $w_T > 0$ とした時間的特徴量を導入した場合の方がその値が高くなっている。特に、(w_S, w_T) を (0.6, 0.4) とした場合に最大値 94.93% を示している。 (w_S, w_T) が (0.9, 0.1) の場合にも同様に $A_e=94.93\%$ となっているが、その前後の (1, 0) および (0.8, 0.2) においてはそれぞれ $A_e=92.75, 94.20\%$ となっており、(0.6, 0.4) の前後の (0.7, 0.3) および (0.5, 0.5) においてそれぞれ $A_e=94.20\%, 93.48\%$ となっているのに比べると変化が急峻である。このことを考慮すると、より安定した (0.6, 0.4) 付近を最大値と見なすのが妥当であると考えられる。

また、エラーの数を見ると Walk(slow) (B), Boxing (K), Modern dance (L) および Charleston (P) に関しては、時間的な特徴量を導入することによって減少している。特に Boxing (K) および Charleston (P) においては一部で 0 となっている。Walk (A), Walk(slow) (B) および Modern dance (L) においては、(w_S, w_T) をいかなる値に変更させてもエラーは 0 とはならなかったが、それでも Walk(slow) (B) および Modern dance (L) に関しては、適切な重みを与えて時間的特徴量を導入することにより、エラーを減少させることが可能となった。

また、 A_e の最大値を与えた (w_S, w_T) = (0.6, 0.4) の場合に関しては、大半のカテゴリにおいてエラー数は改善されているが、Breaking (O) のように一部のカテゴリにおいては若干の悪化が見られるものもあった。

4. 考察

第 3 章で示したように、Walk(slow) (B), Boxing (K) に関しては、時間的特徴量を導入

することによりエラーが減少している。空間的特徴量のみを用いた場合には、これらのカテゴリではそれぞれ Walk (A) および Basket(forward dribble)(J) との混同が見られた。Walk(slow) (B) - Walk (A), Boxing (K) - Basket(forward dribble) (J) のそれぞれの組み合わせに関しては、動作速度は大きく異なるが空間的な姿勢は類似しており、動作速度を考慮して考案された時間的特徴量の導入により、これらの動作の判別精度がある程度改善された。ただし、エラーは完全には解消されておらず、何らかの手段によるさらなる改善が望まれる。

今回使用したデータ全体に関しては(w_s, w_T)を(0.6, 0.4)とした場合に A_e が最大となった。ただし、例えば Breaking (O) や Charleston (P) のように(0.6, 0.4)以外でエラーが最小となるカテゴリも見られた。このことは、検索時のクエリとなるカテゴリに応じて(w_s, w_T)の値を適宜調整しなければならないことを示唆している。

5. おわりに

本研究では、モーションキャプチャデータの検索法に関して、索引として用いられる特徴ベクトルに、空間的特徴量に加えて時間的特徴量も導入した。その際、重み付けユークリッド距離を用いて空間的特徴量と時間的特徴量それぞれの重みを変化させ、検索精度の改善に関する検討を行った。

その結果、時間的特徴量を導入することにより、空間的特徴量のみを用いた場合と比較して検索精度が向上した。このとき、動作カテゴリによってエラーの数が最少となる重みが異なるという現象もみられた。このことは、検索時のクエリとなるカテゴリに応じて重みの値を調整しなければならないことを示唆している。その実装が今後の課題である。

参考文献

- 1) 中澤 篤志：知っておきたいキーワード モーションキャプチャ, 映像情報メディア学会誌 **63-9**, 1224/1227 (2009)
- 2) Carnegie-Mellon Mocap Database, 〈<http://mocap.cs.cmu.edu>〉
- 3) Perfume Global Site, 〈<http://perfume-global.com>〉
- 4) 北 研二, 津田 和彦, 獅々堀 正幹：情報検索アルゴリズム, 共立出版 (2002)
- 5) M. Müller and T. Röder : A Relational Approach to Content-based Analysis of Motion Capture Data. In B. Rosenhahn R. Klette and D. Metaxas (eds.), Human Motion, Springer, 477/505 (2008)
- 6) K. Onuma, C. Faloutsos and J. K. Hodgins : A Fast and Effective Distance Function for Motion Capture Data, EUROGRAPHICS 2008, Short Papers (2008)
- 7) T. Miura, N. Matsumoto, T. Kaiga, H. Katsura, K. Tajima and H. Tamamoto : Indexing of Motion Capture Data Using Feature Vectors Derived from Posture Variation, Journal of Information Processing, **21-2**, 358/361 (2013)
- 8) T. Miura, T. Kaiga, N. Matsumoto, H. Katsura, T. Shibata, K. Tajima and H. Tamamoto : Characterization of Motion Capture Data by Motion Speed Variation, IEEJ Transactions on Electronics, Information and Systems, **133-4**, 906/907 (2012)
- 9) I.P. Herman (著), 齋藤 太郎, 高木 建次 (共訳)：人体物理学 ～動きと循環のメカニズムを探る～, NTS (2009)
- 10) T. Miura, T. Kaiga, H. Katsura, T. Shibata, K. Tajima and H. Tamamoto : Quantitative Motion Analysis of the Japanese Folk Dance “Hitoichi Bon Odori”, IPSJ Symposium, **2013-4**, 167/174 (2013)
- 11) S. M. Pincus : Approximate Entropy as a Measure of System Complexity, Natl. Acad. Sci. USA, **88**, 2297/2301 (1991)

12)M. Mullin and R. Sukthankar : Complete Cross-
Validation for Nearest Neighbor Classifiers,
ICML2000, 639/646 (2000)