

## ランダムフォレスト法を用いた人体動作検索

浜田 祐一<sup>†1</sup> 栗山 繁<sup>†1</sup> 向井 智彦<sup>†1,\*1</sup>

本稿では人体動作データをクエリとして用い、類似した動作データを効率的に検索する手法を提案する。本手法では同じ意味を含むデータを検出するために、演技者の体形差や体力差等の影響を受けにくい相対的な多数個の動作特徴を用いる。そして、個人的な揺らぎや測定誤差を含むデータを高精度に識別できる機構をランダムフォレスト法を用いて構築する。大量のデータを用いた実験により、時間長 100 分程度の様々な動作に対して高い識別精度を確認した。

### Human motion retrieval using random forest

YUICHI HAMADA,<sup>†1</sup> SHIGERU KURIYAMA<sup>†1</sup>  
and TOMOHIKO MUKAI<sup>†1,\*1</sup>

This article proposes a method of efficiently retrieving similar motions from a motion sample (or query). For detecting the data of the same contents, our method introduces many relative features that are independent of the difference in performer's body such as proportion and strength. We then compose the mechanism with the random forest method for accurately classifying the data that have personal fluctuation and measurement error. We confirmed the high classification accuracy through the experiment with a variety of motions.

#### 1. はじめに

近年、ゲームや映画等の CG アニメーション制作では、モーションキャプチャシステムを用いて測定された人体動作データが広く利用されている。既に測定されたデータの中から類

似する動作のデータを検索して収集する事により、動作データの補間の素材として用いたり動きの様々なバリエーションの編集に役立つ。ゆえに、本研究では互いに類似する動作データ群を効率的かつ漏れなく検出できる動作データの検索システムを構築する。

単純な検索方法としては、各データの内容や意味を表すテキストデータを付与して用いる事が考えられるが、複雑な動作データに対して意味内容を自動的に付与する方法は未だに確立されていない。また、特殊な動作の解釈は人間の主観にも依存するので、他者に対しては意図しない検索結果となる可能性がある。ゆえに、意味内容を表すテキストを用いずに、複数の教師データから算出した動作固有の特徴を用いる様々な検索手法が提案されている。Müller らは姿勢情報に含まれる関節間の幾何学的関係を姿勢特徴として 2 値ベクトルで表現し、この関係性を用いて人体動作を検索する手法<sup>1)</sup>を提案している。また、その改良手法として動作カテゴリ毎のインデックスの生起確率を学習して作成するモーションテンプレートをを用いた検索手法<sup>2),3)</sup>、さらには帰納論理を用いることにより動作の意味的な構造に着目した検索手法<sup>4)</sup>等が考案されている。これらの研究を踏まえ、本報告では独自の動作特徴を用いて所望する人体動作を効率的に検索する手法を提案する。

人体動作データは計測時の雑音による誤差を含み、さらには演技者の年齢、性別、体形、体力等の差異によって、同じ意味内容のデータ間でも動きに微妙な揺らぎを生じる。ゆえに本手法では、この揺らぎに対して頑健な動作特徴量と識別器を導入する。識別器の構成に導入するランダムフォレスト法は、集団学習法と呼ばれる多数決による判定アルゴリズムであり、多数定義した動作特徴の中から最適な特徴のみを選出するので、不要な判定を除去した高速な検索が期待できる。

以下、2 章でシステムの検索対象となる人体動作データと検索処理に適した動作表現方法について述べた後、3 章で動作検索システムについて説明する。次に、4 章で提案手法を用いた実験結果と考察を述べ、最後に 5 章でまとめと今後の課題について述べる。

#### 2. 人体動作の特徴量

##### 2.1 人体動作のデータ表現形式

人体動作データは、微小時間間隔での 3 次元空間上の人体関節の位置や角度を測定して得られ、一般的には時系列の姿勢情報として表される。人体動作は、データが表現されている関節数や時間解像度を増加させることで精度を向上できるが、検索処理に要する計算コストはデータの情報量に伴い増大するので、空間的および時間的な解像度を低く表現するのが望ましい。また、同じような意味を持つ動作であっても演技者の個性や試行毎の揺らぎ(例

<sup>†1</sup> 豊橋技術科学大学

Toyohashi University of Technology

\*1 現在、株式会社スクウェア・エニックス

Presently with SQUARE ENIX CO., LTD.

例えば歩行動作の場合、歩幅や進行方向、歩行周期、動作タイミング等の差異)によりデータの値は多様な意味に解釈されうる。ゆえに高い識別精度を得るには、動作データを揺らぎの影響を受け難い形式に変換しておく必要がある。

## 2.2 相対情報による姿勢表現

人体動作データの骨格情報は演技者の基準姿勢を基に決定されるので、身長等の異なる体形の演技者間では同一の動きでも人体関節位置に差異が生じる。そこで、各人体動作データを身長 $h$ の値に基づいて正規化することにより、この個人差を取り除く。まず、直立姿勢から全関節の位置情報  $J = (j_{head}, j_{neck}, \dots)$  を算出し、次に抽出した関節位置情報  $J$  から Y 座標が最小の関節  $\min(J_y)$  と最大の関節  $\max(J_y)$  の差分を基に身長  $h$  を算出する(式(1))。そして、算出した身長を基に骨格情報の全関節の位置情報を以下の式(2)により正規化する。

$$h = \max(J_y) - \min(J_y) \quad (1)$$

$$\tilde{j}_i = \frac{1}{h} \times j_i \quad (2)$$

骨格情報の正規化後、各フレーム時刻の姿勢情報を相対的な値に変換する。動作データが有する姿勢情報は、同じ意味を持つ動作であれば局所的な関節間の関係性は類似する可能性が高いが、姿勢全体が類似しているとは限らない。ゆえに、関節間の関係性を人体を基準として構築される局所座標空間上での相対的な関節位置と2関節間のユークリッド距離、および体の傾き(人体基準座標系での特定の座標軸と体軸との角度)で構成し、その時系列情報として動作データを表現する。

## 2.3 動作特徴の算出

相対的な姿勢情報で表現された動作データから検索処理に用いる動作特徴を算出する。各々の人体動作は継続時間やタイミングが異なるために直接的な比較は困難であるので、指定したフレーム数になるようにデータを等間隔に間引く。次に、同一の意味を持つ動作でも動作中の姿勢が同じタイミングで変化するとは限らないので、本手法では瞬時の姿勢間を比較するのではなく、ある程度幅を持たせた局所区間内での平均姿勢を比較する<sup>5)</sup>。ただし、平均の姿勢情報  $\tilde{p}_i$  は、算出した相対的な姿勢情報を局所区間  $[t_0, t_1]$  で積分することにより以下の式(3)のように得られる。

$$\tilde{p}_i = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} p_{i,t} dt \quad (3)$$

しかし、上式の姿勢情報の積分は多くの計算量を要するので、動作の最初からある時刻  $t$  までの積分姿勢情報  $g_{i,t}$  を事前に算出しておき計算時間を削減する。式(4)、(5)のように、

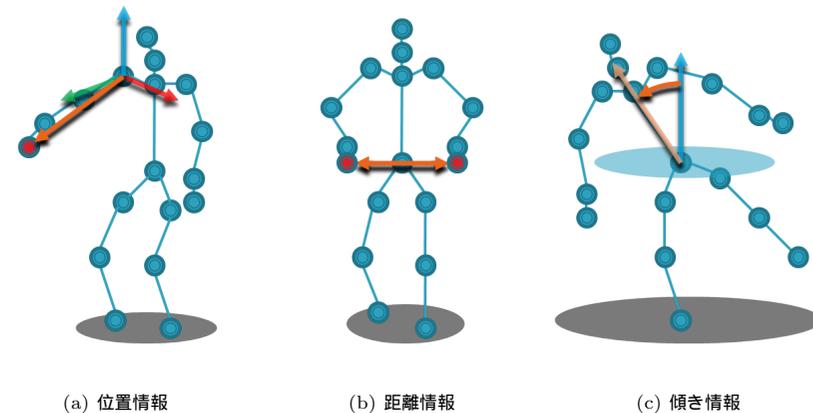


図 1 姿勢表現に用いる相対情報

Fig. 1 Relative variables for representing poses

積分情報で表現された動作データは、 $t_1$  の積分値情報から  $t_0$  の積分値情報を 1 回差分するのみで平均の姿勢情報  $\tilde{p}_i$  を算出することができるので、動作特徴を高速に算出できる。

$$g_{i,t} = g_{i,t-1} + p_i \quad (4)$$

$$\tilde{p}_i = \frac{1}{t_1 - t_0} (g_{i,t_1} - g_{i,t_0}) \quad (5)$$

この平均姿勢情報と閾値との大小関係によって 2 値の動作特徴量を得る。ただし、閾値の値の決定方法は次章での動作検索器の構築過程で説明する。

## 3. 動作検索システム

本章では、人体動作データを効率的に検索する機構について述べる。ここでいう検索とは、意味的に同等と見なせる動作データを検出する機構に相当する。本手法では、高い識別性能が期待できるランダムフォレスト法を用いることにより、高精度かつ高速な検索システムを構築する。

### 3.1 動作検索システムの概要

提案する検索システムは、人体動作データそのものをクエリとして使用し、動作特徴量が

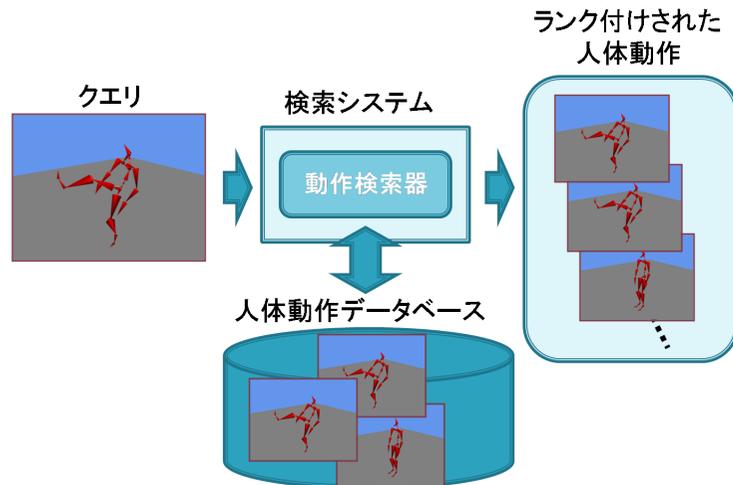


図2 動作検索システム  
Fig.2 Motion retrieval system

ら判定されるカテゴリと一致するデータの一覧を出力する。検索器にクエリとして入力される動作データは、多数の識別器を用いて計算される動作カテゴリの生起確率ヒストグラムにより表現される。一方、検索対象の動作データ群に対しても、同様の変換処理を施して各データの生起確率ヒストグラムを作成しておき、その値から判定されるカテゴリのIDを各データに登録しておく。そして、このカテゴリIDを参照することにより人体動作データの一覧を出力する。

### 3.2 ランダムフォレスト法

ランダムフォレスト法とは、多数の決定木を組み合わせることで精度の高い識別器を得る集団学習法<sup>6)</sup>の一種である。集団学習法は精度が低い複数の識別器を組み合わせることで精度向上させる手法であり、他にもバギングやブースティング等の手法が考案されている。本手法で用いるランダムフォレスト法は、このような集団学習法の中でも特に計算速度が速く、識別の精度が高い手法として知られている。

動作検索器は図3に示す様に多数の決定木により構成される。各決定木は判定する動作特徴をランダムに選出するため単体での識別精度はあまり高くないが、多数の決定木の出力

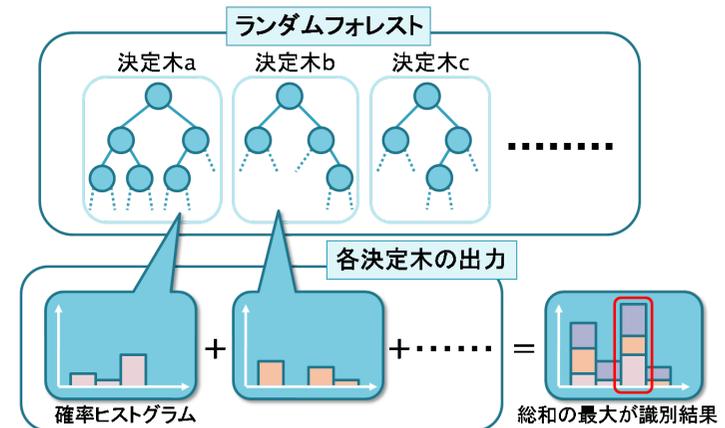


図3 ランダムフォレストの模式図  
Fig.3 Schematic representation of random forest

結果を組み合わせる判定により、揺らぎを多く含むデータに対しても高精度な識別が期待できる。また、構築される決定木は必要最小限の動作特徴判定を行うため、計算回数が少ない高速な判定が期待できる。

### 3.3 動作検索器の構築

動作検索器には、教師データが有する動作カテゴリを精度良く識別できる状態を学習させる。この学習のために、人体動作を動作カテゴリ毎に分割してラベルを付加した教師データ集合  $D$  を使用する。動作検索器を構成する各決定木は、ノードが2.3で定義した動作特徴を判定する局所動作識別器により構成されている。局所動作識別器は平均姿勢情報を出すためのパラメータ  $r$  と動作特徴の判定に使用する閾値  $s$  を有しており、算出した平均姿勢情報  $\bar{p}_i$  と閾値  $s$  の比較により動作特徴を判定する単純な構造である。このような構造を有する決定木は、局所動作識別器のパラメータ  $r$  をパラメータ集合  $R = (r_1, r_2, \dots, r_N)$  の中からランダムに選出しながら構築するため、動作検索器を構成する全ての決定木は異なる構造を持つ。以下に決定木の具体的な構築手順を示す。

ステップ1: 特徴量算出パラメータ集合  $R$  からランダムにパラメータ  $r_i$  を選出する。

ステップ2: 選出したパラメータ  $r_i$  を用いて局所動作識別器を構築する。

ステップ3: 教師データ集合  $D$  を用いて情報エントロピー  $E(D)$  を最大にする(すなわち、付与した動作カテゴリを精度よく分割する)局所動作識別器の閾値  $s$  を決定する。

ステップ4: ステップ1~3を指定回数繰り返し、情報利得  $\Delta I$  を最大にする局所動作識別器を選出する。

ステップ5: 選出した局所動作識別器を用いて、教師データ集合  $D$  を動作特徴が存在する集合  $D_1$  と存在しない集合  $D_2$  に分割する。

ステップ6: 集合  $D_j$  の情報エントロピーが指定値以下の場合、 $D_j$  を構成する動作カテゴリの確率ヒストグラムを決定木のリーフノードが持つ出力情報として保存する。

ステップ7:  $D_1, D_2$  それぞれに対してステップ1以降の処理を行い、構築された局所動作識別器をステップ4で選出した局所動作識別器の子ノードとして登録する。

ステップ8: 分割した集合の情報エントロピーが全て指定値以下になるまで繰り返す。

上述した処理で用いている情報エントロピー  $E(D)$  は、事象がどれほど起こりにくいかという曖昧性を表す尺度であり、式(6)のように人体動作データ集合  $D$  に対して割り当てられている動作カテゴリ集合  $C(D) = (c_1, c_2, \dots)$  上の確率分布  $P$  から得られる。また、情報利得はデータ集合  $D$  を  $m$  個のデータ集合  $(D_1, D_2, \dots, D_m)$  に分割した際の情報エントロピーの減少量であり、データ集合を分割することにより曖昧性をどの程度削減できたかを表す指標である。ゆえに、式(7)に示す様に分割前のデータ集合  $D$  の情報エントロピー  $E(D)$  と分割後のデータ集合  $D_j$  の情報エントロピー  $E(D_j)$  の総和の差分を算出する事により得られる。ただし、本手法では決定木の構造に2分木を用いるため、分割数は  $m = 2$  となる。

$$E(D) = - \sum_{c \in C(D)} P(c) \log_2 P(c) \quad (6)$$

$$\Delta I = E(D) - \sum_{j=1}^m \frac{|D_j|}{|D|} \times E(D_j) \quad (7)$$

### 3.4 検索方法

構築した動作検索器に対して人体動作データがクエリとして入力されると、入力データは各決定木のルートノードからリーフノードに向かって中間ノードである局所動作識別器を用いて判定計算が実行され、リーフノードで得られる動作カテゴリ名が各決定木の出力となる。図3に示す様に、クエリとして与えた動作データはこの動作カテゴリ名を基に構成した確率ヒストグラムで最大の出現確率となるカテゴリに属すると判定される。この動作検索器の判定回数は、各決定木でリーフノードに到達するまでの識別器の数と等しいので、高速な識別が期待できる。

### 3.5 検索結果に対するランク付け

前節で述べた識別方法では、入力されたクエリが学習したどの動作カテゴリに属するかという判定はできるが、同じカテゴリ内の動作データ間での順位付けはできない。そこで、得られた確率ヒストグラムから類似度を算出してランク付けを行う方法を提案する。

クエリ  $q$  を与えて出力された確率ヒストグラム  $H(q)$  と検索対象の動作データ群  $D$  の各データ  $d_i$  の確率ヒストグラム  $H(d_i)$  間の類似度  $s(q, d_i)$  は、各要素(動作カテゴリ名)間のユークリッド距離を求めることにより算出する(式8)。この類似度を降順に並び替えることによりランク付けされた検索結果を得る。

$$s(q, d_i) = \sqrt{\sum_{c \in C(D)} (H(q)_c - H(d_i)_c)^2} \quad (8)$$

## 4. 検索性能評価実験

本章では、3章で提案した検索システムの検索性能に関する評価実験の結果を示す。

### 4.1 実験データ集合

実験で使用する人体動作データ集合には、Müllerらが測定して構築した人体動作データ集合[HDM05]<sup>7)</sup>を用いた。このデータ集合は、高解像度な時空間情報を有する人体動作データにより構成されているが、複数の意味を持つ動作情報が連続した状態でファイルに保存されているので、手作業で1ファイルに唯一の意味の動作を含むように分割し、各動作ファイルを78種の動作カテゴリに分類した。このカテゴリ分類には、動作データ集合の仕様書<sup>8)</sup>に記載されている測定動作情報を基準とし、どの動作カテゴリにも属さない動作データは今回の実験では破棄した。また、動作情報の分割点は著者らが主観に基づいて決定した。この分割作業により、表1に示す実験用のデータ集合を構築した。

動作検索器の構築には学習用のデータ集合が必要なため、このデータ集合を表2の様に学習用と評価実験用に分類した。ただし、各動作カテゴリ毎に学習用と評価実験用のデータ数の比率が同等になるようにランダムに振り分けた。

### 4.2 動作検索器に対する識別性能の検証

まず、提案した動作検索器の識別性能について検証する。ランダムフォレスト法を用いて学習した識別器は多数の決定木により構成されるので、決定木単体とランダムフォレスト法を用いた識別器の識別性能に対して比較を行い、識別精度が改善されているかを確認する。ただし、比較用の決定木にはランダム要素を含むものと含まないものの2種類を用い、3.3

表 1 実験に用いた人体動作データ集合  
Table 1 Experimental motion dataset

データ集合	ファイル数	総動作時間 [sec]	データ容量 [MB]
分割前データ集合	293	11805	1,354
分割後データ集合	4660	6192	532

表 2 実験用データ集合の分割  
Table 2 Division of experimental dataset

データ集合	ファイル数	総動作時間 [sec]	データ容量 [MB]
学習用データ集合	2346	3126	269
実験用データ集合	2314	3066	263

節で説明した手順で決定木を構築した。ランダム要素を含む決定木は、ランダムフォレスト法で構築する決定木を一本のみに限定する。ランダム要素を含まない決定木は、全てのパラメータ値の中から最良の値を逐次選択することにより単体の決定木を構築する。これら 2 種類の決定木とランダムフォレスト法により構築した動作検出器の検出性能を比較する。ただしランダムフォレスト法は、本実験では 100 本の決定木により構成された識別器を用いている。また、動作特徴の算出に用いる相対姿勢情報として 56 種類の情報を定義し、全人体動作データを 10 フレームで正規化した。

評価実験を行った結果を図 4 に示す。この図は、評価実験用データ集合に含まれる全ての人体動作データに対して識別処理を行った結果から平均の識別精度を算出した結果であり、評価指標として再現率 (recall)、適合率 (precision)、F 値 (F-measure) の三種類の検出指標を用いている。この結果から、特にランダムフォレスト法を用いた識別器は、他の 2 手法に比べて高い識別性能を有している事が確認できるので、ランダムフォレスト法は識別性能を向上させる手法として有効であると結論できる。

#### 4.3 既存手法との比較

本節では、4.2 節で検証した動作検出器を動作検出システムに組み込み、システムの検出性能を評価する。ただし、ここでの検出性能とはクエリとして与えた動作データとの間での動作カテゴリの一致割合を表す識別精度を基に検証する。

既存手法と提案手法の識別精度の比較を行った結果を図 5 に示す。ここでは、既存手法として類似した動作特徴を用いて検出を行うモーションテンプレート手法と比較する。この結果から、提案手法はモーションテンプレート手法と比較してより高い識別精度を得ていることが確認できる。しかし、この実験は人体動作データ集合の分割方法、動作カテゴリの種類

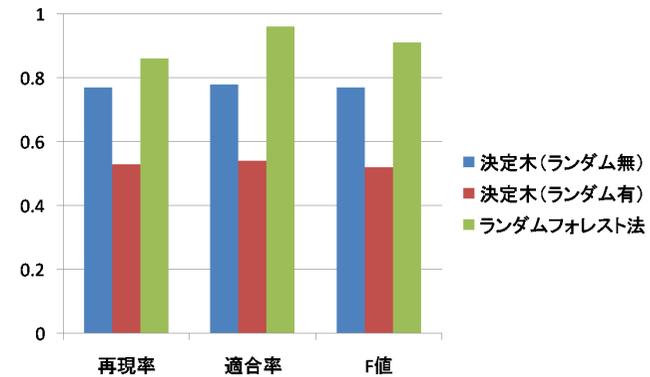


図 4 識別器の精度比較  
Fig. 4 Comparison of classifier's accuracy

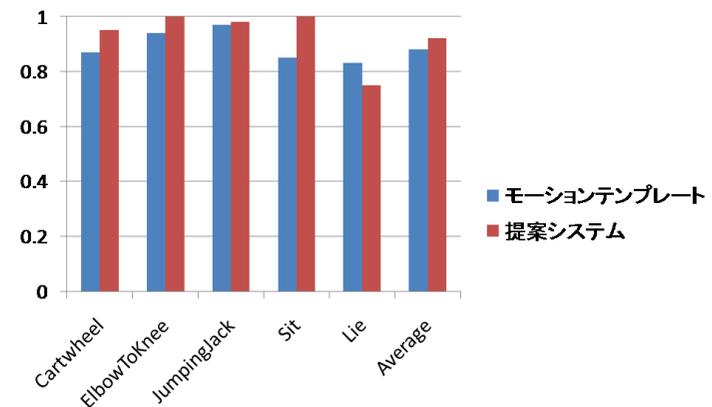


図 5 既存手法との識別精度比較  
Fig. 5 Comparison of classification accuracy with existing methods

が異なる環境で行った実験のため、今後より精密な比較実験を行い提案手法の有効性をさらに検討していくことが必要であると考えます。

## 5. おわりに

本稿では、ランダムフォレスト法を用いて動作検索器を構築することにより高い検索性能を有する動作検索システムを提案した。そして、この動作検索システムが既存手法と遜色ない検索性能を有することを実験により確認した。今後の課題としては、動作検索システムの識別精度をさらに向上させることが挙げられる。今回学習に用いた動作カテゴリは検索システムの識別精度を考慮せずに無作為に選んだため、使用している動作カテゴリが検索システムを構成する上で最適であるとは限らない。さらに、動作特徴の算出に使用する相対姿勢情報は今回定義したもの以外にも、識別精度の向上に貢献するものが存在する可能性がある。これらの点を調査することにより、より有用な動作検索システムを提案することができると考える。また、検索して得られた同一カテゴリに属する動作データ群に対して、類似の度合いに応じてランクを付与する手法を提案したが、このランキング手法の性能の定量的な評価も今後の課題である。

提案したシステムは学習に用いた動作カテゴリに対しては高い識別精度を得ているが、学習していない未知の動作カテゴリに属する動作に対しては再学習が必要となる。しかしながら、データの規模が膨大化するに従い、再学習のコストが増大し、識別精度も低下する可能性がある。本手法のスケーラビリティを実現するためには、新たなカテゴリの動作データに対して再学習が不要となるような識別機能の拡張が必要である。

謝辞 本研究は、日本学術振興会科学研究費補助金・基盤研究(B)18300068「大規模なヒューマンデータの知的探索と再利用」の支援を受けた。

## 参 考 文 献

- 1) Meinard Müller, Tido Röder, Michael Clausen: Efficient Content-Based Retrieval of Motion Capture Data, ACM Transactions on Graphics 24(3), pp. 677-685, (2005).
- 2) Meinard Müller, Tido Röder: Motion Templates for Automatic Classification and Retrieval of Motion Capture Data, Proceedings of SIGGRAPH Symposium on Computer Animation, pp. 137-146, (2006).
- 3) Meinard Müller, Bastian Demuth, and Bodo Rosenhahn: An Evolutionary Approach for Learning Motion Class Patterns, DAGM, pp. 365-374, (2008).
- 4) 脇坂健一, 向井智彦, 栗山繁: 帰納推論を用いた動作のセマンティクス解析, Visual Computing/グラフィクスとCAD 合同シンポジウム 2007, pp. 151-156, (2007).
- 5) 浜田祐一, 栗山繁, 向井智彦: 姿勢変位量の特徴抽出に基づく動作データの高速度検出, 情報科学技術フォーラム講演論文集, I-052 (2008)

- 6) Pang-Ning Tan, Michael Steinbach, Vipin Kumar: Introduction to Data Mining, Pearson Addison Wesley, (2006).
- 7) [www.mpi-inf.mpg.de/resources/HDM05/](http://www.mpi-inf.mpg.de/resources/HDM05/)
- 8) Meinard Müller, Tido Röder, Michael Clausen, Bernhard Eberhardt, Björn Kröger, and Andreas Weber: Documentation Mocap Database HDM05., Computer Graphics Technical Report CG-2007-2, (2007).