

力学系の切替モデルの尤度時系列情報を用いた 日常動作のオンライン分節化法

○ 大島 佳菜 (東京大学)
森 武俊 (東京大学)

下坂 正倫 (東京大学)
佐藤 知正 (東京大学)

歩く、座るなどの人の日常動作に対し、教師データを用いてオンラインで分節化する手法を提案する。提案手法では、動作の分節を区切れ目となる分節点で2つの力学系が切り替わるモデルとして表現する。教師データから予め学習した力学系の切替モデルに対する入力運動の尤度の時系列パターンに着目することで、漏れなく大量に分節化候補点を発見する。そして、時系列尤度間のマッチングを用いて候補点を絞り込んだ後、オンライン処理における遅延性を考慮した分節点の確定を行う。モーションキャプチャデータに対する適用実験の結果、本手法により日常動作に対する人の判断に近い分節化が実現できたことを確認した。本手法は画像データからの動作分節化の基盤を担うものとして期待される。

Online Segmentation of Daily Actions Using Time Series Likelihood of Switching Dynamical Systems

*Kana Oshima, Masamichi Shimosaka, Taketoshi Mori and Tomomasa Sato
The University of Tokyo

This paper presents an online supervised segmentation method of daily actions like walking and sitting. The proposed method utilizes switching models of two dynamical systems to represent segmental boundary of human motion. It consists of three phases based on the pattern of time-series likelihood of the switching models. 1) Candidates of the segmental points are detected without omission. 2) The candidates are refined by matching algorithm using time-series likelihood of the correct segmental boundary. 3) The segmental points are output from the candidates by considering latency due to the decision. The experimental results using real motion capture data show that the proposed method are effective for online segmentation of daily actions.

1 緒論

システムにより人を円滑に支援するには、オンラインによる人の判断に近い日常動作の認識 [1–3] が重要である。動作認識の際、一連の動作をまとまりごとに区切る分節化が必要となる。本研究ではオンラインで日常動作に対する人の判断に近い分節化の構築を目指す。

舞踊動作を対象とした Shiratori ら [2] の手法は、舞踊に特徴的に見られる止め動作というものを有効的に利用することで教師なしでも人の判断に近い分節化を実現している。しかし、本研究が対象とする日常動作は動作の区切れ目に止め動作のような明確な指標がないため、人の判断に近い分節化を実現するためには人の教示に基づく教師データの利用が重要となる。また日常動作に対して人が教示する区切れ目には時間方向の誤差があるため時間幅を持った区間を扱い分節化する必要がある。そこで「区切れ目を境に力学系が切り替わる」力学系の切替モデルを用いてモデル化を行う [4–6]。力学系の切替を仮定したモデルを利用する手法の中でも、教師なしの手法 [5, 6] とは異なり、Segawa ら [4] は教師データを効率的に用いることで人の判断に近い分節化を実現している。具体的には予め計測した運動データに対して力学系の切替モデルの一連の時系列尤度全体から閾値処理を用いてピークを算出することで人の判断に近い分節化を実現している。しかし、この手法はオフラインでのみ利用可能であるため、手法の適用範囲が限られる。

本手法では教師データを有効に利用してオンラインで日常動作を分節化する手法を構築することを目指す。オンラインで分節化処理を行うためには、基本的に現在時刻までに得ら

れた運動データのみから分節点の判定を行わなければならないため、単一手掛かりのみから分節点を確認する方法では不安定である。よって、本手法では大量に分節点候補を発見してから絞り込んで分節点を決定する方法を提案する。Fig. 1 に提案手法の流れを示す。提案手法は、力学系の切替モデルを予め獲得しておく段階と、入力運動データをオンラインで分節化する段階に分けられる。

2 力学系の切替モデルの獲得

オンライン分節化処理を行う前に予め、人の教示に基づく区切れ目前後をモデリングした力学系の切替モデルを獲得しておく。以下にその詳細を説明する。

2.1 日常動作の運動データ

本研究では入力として人の骨格の構造と各関節の姿勢情報が記録されたモーションキャプチャデータを使用する。人体の骨格モデルとして Fig. 2 に示すような全部で 11 の身体部位からなる、計 36 自由度のモデルを用いた。30Hz のサンプリングレートでモーションキャプチャにより取得された身体部位ごとの運動データを、身体構造の中心となる Hips の位置 3 次元と、11 身体部位の姿勢 4 次元（クォータニオン表現による）の合計 47 次元の特徴量で表現する。この 47 次元の特徴量を入力データとして分節化処理を行う。時刻 t での特徴量を o_t とし、 o_t において身体部位 b に該当する次元を抜き出して $o_t^{(b)}$ とする。

2.2 力学系の切替モデル

人による区切れ目区間の抽出と力学系の切替モデルの概要

入力時系列運動情報に対して、人が区切れ目であると判断した分節点の前後一定期間 ($2T + 1$ サンプリングフレーム)

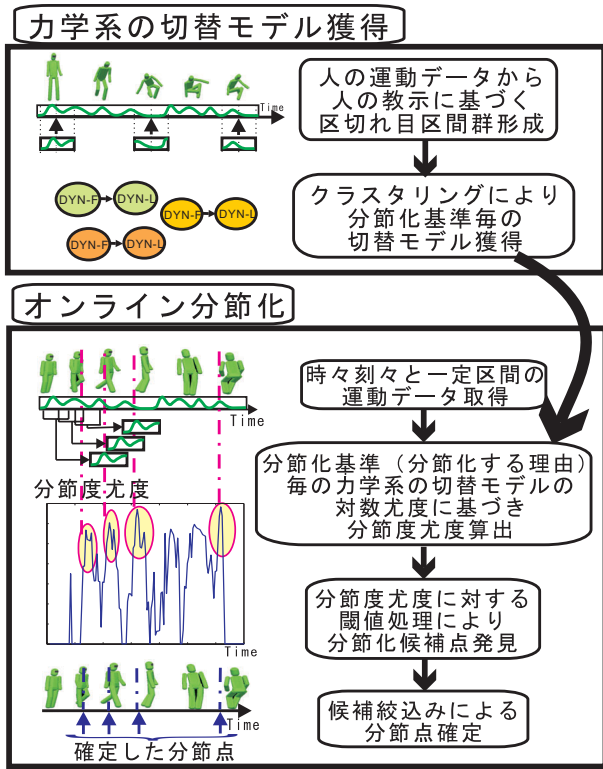


Fig. 1: モデル獲得とモデルによるオンライン分節化の流れ

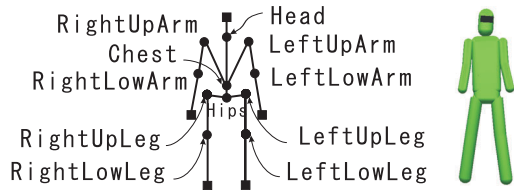


Fig. 2: 身体構造モデル

を区切れ目区間として抽出し、その区間中に力学系が切り替わるものとして表現するのが力学系の切替モデルである．具体的には、人の教示に基づいて抽出した運動データの途中の時刻（時刻 τ とする）を境に、前区間と後区間に分け、前半の $\tau - T$ から τ フレームまでを一つ目の力学系（以下 DYN-F）で、後半の $\tau + 1$ から $\tau + T$ フレームまでを二つ目の力学系（以下 DYN-L）でモデル化する．それぞれの力学系は線形力学系（LDS）の集合体として表現する．線形力学系の定式化は以下で述べる．

線形力学系（LDS）

DYN-L, DYN-F を構成する LDS は、以下のように定式化されるものである．

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{v}_t \quad (1)$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{w}_t \quad (2)$$

式中 $\mathbf{x}_t$ は時刻 t における状態変数、 $\mathbf{A}$ は状態空間における遷移行列、 $\mathbf{C}$ は状態空間から観測空間へのマッピングのための行列、 $\mathbf{y}_t$ は観測変数、 $\mathbf{v}_t$ はシステムノイズ、 $\mathbf{w}_t$ は観測ノイズを意味する．また、 $\mathbf{v}_t$ と $\mathbf{w}_t$ はそれぞれ平均 0、共分散行列 $\mathbf{Q}$, $\mathbf{R}$ のガウス分布から生成されたものとして仮定する．LDS のパラメータとは、 $\mathbf{A}, \mathbf{C}, \mathbf{Q}, \mathbf{R}$ で示されるものであり、これらパラメータをまとめて θ とする．本モデルに対する時

刻 t から時刻 $t + \tau$ の観測データの時系列情報 $\mathbf{y}_{t:t+\tau}$ の対数尤度は Kalman フィルタを援用し効率的に算出される [7]．本稿では関数 f を用いて

$$f(\mathbf{y}_{t:t+\tau}, \theta) = \ln p(\mathbf{y}_{t:t+\tau} | \mathbf{A}, \mathbf{C}, \mathbf{Q}, \mathbf{R}) \quad (3)$$

として表現する．

身体構造の考慮

本研究では身体構造を考慮し、11 の身体部位について、それぞれに対応する力学系の切替モデルを作成する．具体的には、11 身体部位それぞれについて、各部位に相当する観測変数 $\mathbf{o}_t^{(b)}$ を用いて切替点前後の DYN-F, DYN-L を構築する．このように、力学系の切替モデルを構成する切替点前後の DYN-F, DYN-L は、各身体部位に対応した 11 個の LDS の集合として定義される．以下、力学系の切替モデルに基づく尤度の計算方法を説明する．

力学系の切替モデルの尤度

力学系の切替モデルを構成する切替点前後の DYN-F, DYN-L による対数尤度はそれぞれ、その要素である 11 個の LDS の対数尤度情報をもとに計算される．身体部位による観測変数の次元の大きさの違いを考慮して、11 個の LDS それぞれから計算される対数尤度を、それぞれの LDS の観測変数の次元で正規化した後、後述の重み付けベクトル $\mathbf{w} \in \mathbb{R}^{11} (|\mathbf{w}|_1 = 1, \{\mathbf{w}\}_b \geq 0, b = 1, \dots, 11)$ を利用し算出する．例えば時刻 $\tau - T$ から τ までの運動情報 $\mathbf{o}_{\tau-T:\tau}$ の DYN-F に対する対数尤度 $l(\mathbf{o}_{\tau-T:\tau})$ は

$$l(\mathbf{o}_{\tau-T:\tau}) = \mathbf{w}^T \mathbf{q}(\mathbf{o}_{\tau-T:\tau}) \quad (4)$$

$$\{\mathbf{q}(\mathbf{o}_{\tau-T:\tau})\}_b = \frac{1}{\dim(\mathbf{o}_\tau^{(b)})} f(\mathbf{o}_{\tau-T:\tau}^{(b)}, \theta^{(b)}) \quad (5)$$

として計算される．ただし $\theta^{(b)}$ は、身体部位 b に相当する LDS のパラメータであり、 $\dim(\mathbf{o}_t^{(b)})$ は $\mathbf{o}_t^{(b)}$ の次元数を指す．

2.3 クラスタリングによる分節化基準毎の力学系の切替モデル獲得

人が運動を分節化する際に区切れ目と感ずる理由、すなわち運動の分節化基準は、動作の文脈により異なる．例えば、椅子に座った状態で上体を反らしはじめる部分は、異なる動作が始まるという意味では区切れ目と捉えられるが、座るという一連の動作で考えれば区切れ目とは捉えられない．よって、人が教示した区切れ目区間群を分節化基準ごとにグルーピングして、それぞれの分節化基準ごとに力学系の切替モデルを獲得することで、人に近い日常動作の分節化の実現を考える．以下、分節化基準ごとに区切れ目区間群をグルーピングする手法をクラスタリングと呼ぶ．

本研究では、分節化基準の多様性の主たる要因として「どの部位に着目したか」という着目部位の違いを、各部位の尤度の重要度として表現する．前節において、身体全体の対数尤度の算出に $\mathbf{w}$ を用いたのはこのためである．このベクトルはクラスタリングの際に区切った理由（分節化基準）ごとのクラスタを得るため、クラスタ $C_k (k = 1, \dots, K)$ ごとに各部位に対応した重みベクトル $\mathbf{w}_k$, ($k = 1, \dots, K$), 対応する LDS 群のパラメータ集合 Θ_k を定義する．各クラスタに対応する LDS 群のパラメータ集合 Θ_k は、クラスタリングにより最適なクラスタ結果が求まった後に、各クラスタに属する区切れ目区間を用いて推定される．

人が教示したそれぞれの区切れ目が属する分節化基準を固定した下での重みベクトル推定と、重みベクトルを固定した下でそれぞれの区切れ目がどの分節化基準に属するかという推定を、交互に繰り返しながら、最適なクラスタリング結果を求める．クラスタリングの流れを Fig. 3 に示す．

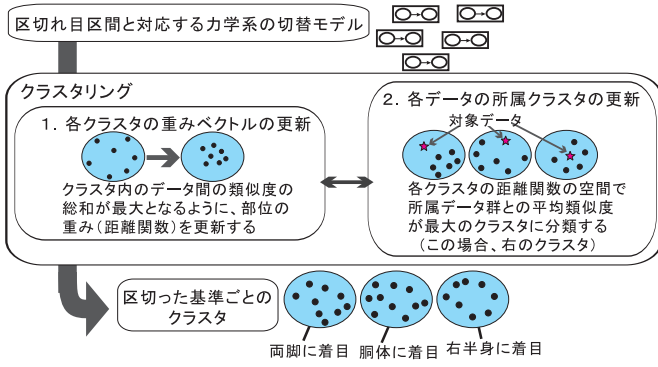


Fig. 3: クラスタリングの流れ

3 切替モデルの時系列尤度によるオンライン分節化

分節化手法ではまず、時々刻々と入力される運動データに対し、分節化基準ごとに得られた力学系の切替モデルの尤度を計算する。最も重要となる分節化基準に着目して各時刻ごとに分節化基準ごとの時系列尤度の最大値を分節化の手掛かりとする。この最大尤度を分節度尤度と呼ぶ。分節度尤度は分節点付近で力学系の切替モデルにフィットして急激に上昇すると共に、似た形状をして高い値を保つという特徴がある。この特徴を利用し以下の手順からなる分節化処理を行う。

分節度尤度から漏れなく分節点を発見することを狙い、分節度尤度の差分値を用いて正解となる分節点を含むそれ以外の点も大量に分節点候補として発見する。次に精度を上げるために、時系列尤度のマッチングを用いて候補を絞り込み、分節点を確定する。

3.1 力学系の切替モデルに基づく分節度尤度の計算

まず、予め獲得した人の分節化の教示による力学系の切替モデル DYN-F, DYN-L のうち、分節点前の DYN-F のみを用いて、オンラインで入力される運動データに対する尤度を計算する。この際、分節点より過去側の DYN-F のみを用いるのは、オンライン処理における遅延を考慮したためである。

前節の方法で定義される、分節化基準ごとの時系列尤度は、それぞれの分節化基準に着目した場合、各時刻が分節点らしいかどうかを表す指標となっている。よって、人はある時刻のもとではある単一の分節化基準に基づいて分節化を行うという仮定の下、各時刻ごとに最も重要な分節化基準にのみ着目した分節度尤度を分節化の手掛かりとする。時刻 t における分節度尤度 $\mathcal{L}_t$ は分節化基準ごとの時系列尤度の最大値として、以下のように表される。

$$\mathcal{L}_t = \max_k l_k(o_{t-T:t}) \quad (6)$$

ただし $l_k(\cdot)$ は k 番目のクラスターにおける DYN-F に対する運動情報の対数尤度であり、(4) から算出される。

3.2 分節度尤度の変化値に基づく分節化候補点の発見

人の教示に基づく区切れ目付近をモデル化した力学系の切替モデルに基づき計算された分節度尤度 $\mathcal{L}_t$ は Fig. 4 に示すように、分節点近傍では値が急激に上昇する特徴を持つ。本研究ではこの特徴を利用し分節度尤度の2次差分量に対して閾値処理を行い、分節化の候補点を大量に発見する。具体的には、分節度尤度の2次差分量

$$d_t = \mathcal{L}_t - 2\mathcal{L}_{t-1} + \mathcal{L}_{t-2} \quad (7)$$

が時刻 $t-1$ では閾値 α 未満かつ、時刻 t では閾値 α 以上であれば、時刻 t を分節化候補点 $u(=t)$ とする。閾値 α は予め検証データを用いることで、最適な値を取得しておく。発見

された分節化候補点はすぐには分節点とは確定せずに一定時間保留し、保留中に新しい候補点が発見された場合には次に示す候補点の絞込みにより候補点の取捨選択を行う。

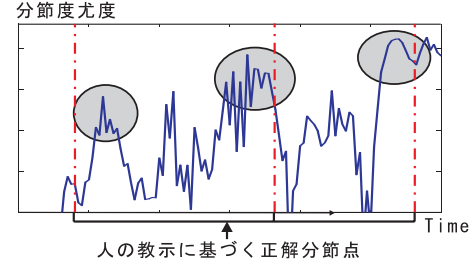


Fig. 4: 床から起き上がる動作に対する分節度尤度の一例

3.3 候補点の絞込みと分節点の確定

大量に発見された分節化候補点に対して時系列分節度尤度のマッチングを用いて絞込みを行い、一定時間以上候補点として残ったものを分節点と確定して出力する。正解分節点付近の時系列分節度尤度のマッチングを行う際には、分節度尤度が分節点付近で似た形状をして高い値を保つという特徴に基づき、正解分節点付近の分節度尤度の特徴を最もよく表している代表パターン $h_{1:W_f+W_l+1}$ を比較基準とする。

具体的には、オンラインで逐次分節化候補点を発見し、候補点が複数存在する場合には、候補点 u の前後 W_f, W_l フレームの時系列分節度尤度 $\mathcal{L}_{u-W_f:u+W_l}$ を抽出し、区切れ目を表す代表的な、時系列分節度尤度のパターン $h_{1:W_f+W_l+1}$ との距離が小さいほど候補点 u が分節点らしいとして、複数の候補点の中で絞込みを行う。時系列尤度のマッチングに用いる距離は、正規化ユークリッド距離を用いる。正規化ユークリッド距離とは、比較する同一幅の部分時系列同士を共に正規化した後、ユークリッド距離をとったものである。また、代表パターンとの比較する時系列幅 W_f, W_l の値については予め最適な値を検証しておく。

そして、分節化候補点のうち、発見されてから一定時間 W_z 以上保留されて絞込みを経てもなお候補点として残っているものを分節点として確定し、出力する。候補点が発見されてから分節点として確定するまでの保留時間 W_z は、予め計測した様々な日常動作を含む運動データを用いて 0 から 30 フレームまで値を変化させて分節化処理を行い、精度が最も高くなる最適値を検証しておく。

3.4 区切れ目区間における代表的な分節度尤度パターンの獲得

比較基準となる代表パターン $h_{1:W_f+W_l+1}$ を選出する際には、人の教示による正解分節点群を教師データとして用いる。各正解分節点の前後 W_f, W_l の時系列分節度尤度を抽出して、正解分節点付近の部分時系列尤度群を形成する。この中から自他との距離の総和が最小となるものを、代表パターン $h_{1:W_f+W_l+1}$ として予め選出しておく。これは、似た形状をした部分時系列群の中で自他との距離の総和が最小である部分時系列は、その部分時系列群を最もよく近似するものであると考えられるからである。この際にも、部分時系列同士の距離として正規化ユークリッド距離を用いる。このように時系列尤度から特徴的なパターンを抽出する手法として、Linらの EMMA アルゴリズム [8] を援用した。ただし、Linらは部分時系列を記号化し DP 法により効率的にマッチングを行い時系列データから頻出パターンを抽出しているが、本手法においては比較する部分時系列群が同一幅であることや、尤度計算時にある種の時間伸縮性を考慮しているため、計算がさらに簡便である正規化ユークリッド距離を用いた。

4 提案手法の評価実験

提案手法を用いて運動データに対してオンラインで分節化処理を行い、手法の精度と遅延時間を評価する。

4.1 実験の条件

実験には11身体部位、総計36自由度からなるモーションキャプチャデータを用いた。対象動作は起き上がる、寝転ぶ、座るなどを含む動作であり、サンプリングレートは30Hzとした。そして、200個程度の正解分節点を持つ約300秒の長さの運動データセットを学習用、検証用、実験用の合計3セット用意した。予め、学習用データセットを用いて力学系の切替モデルと代表パターン $h_1: W_f + W_l + 1$ を学習し、検証用データセットを用いて分節化に必要なパラメータ（閾値 α 、時系列尤度マッチングに用いる前後幅 W_f , W_l 、保留時間 W_z ）の最適値も予め検証しておく。最適値の検証については、パラメータ値を少しずつ変化させて最も分節化精度が高くなるパラメータ値を採用する。これらを用いて実験用データセットに対してオンラインによる分節化を行い、手法の精度を評価する。遅延時間についてはミリ秒単位で評価を行う。

精度の評価指標にはF値を用いた。F値とは、Recall（再現率）とPrecision（適合率）の調和平均である。F値が大きいほど分節化精度が高いといえ、最大値は1である。人が教示した分節点の総数を M_a 、提案手法により発見された分節点の総数を M_m 、提案手法により発見された分節点のうち正解である分節点の数を M_c とし、Recall 値を R 、Precision 値を P とすると、 R , P , F はそれぞれ以下のように表される。

$$R = \frac{M_c}{M_m}, P = \frac{M_c}{M_a}, F = \frac{2RP}{R+P} \quad (8)$$

本実験では手法による分節点と人の教示した分節点を対応付け、対応する点同士が一定のズレの範囲内に含まれるときを正解とみなして、精度を算出する。

4.2 実験結果

正解とみなすズレの許容範囲を0から30フレーム（0から1秒）まで変えて評価を行った。区切れ目区間については、 $T = 15$ フレームとして分節点前後1秒間を抽出した。クラスタ数は事前に分節化基準数を数え $K = 27$ とした。また、予め分節化に必要なパラメータの最適値を検証した結果、閾値は $\alpha = 0.12$ 、時系列尤度マッチングに用いる前後幅は $W_f = 7$ フレーム、 $W_l = 1$ フレーム、分節点の確定を保留する時間長は $W_z = 28$ フレームとなった。

この最適パラメータ値を用いて実験データセットに対してオンラインで分節化を行った結果、ズレの許容範囲を0から30フレームまで変えた場合の精度はFig. 5のようになった。正解とみなすズレの許容範囲を15フレーム（0.5秒）とした場合の再現率、適合率、F値は、0.80、0.83、0.81となった。

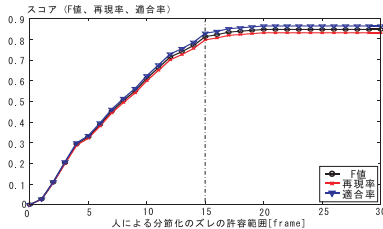


Fig. 5: ズレの許容範囲による精度

本手法の適用例として、寝転ぶ運動に対する分節化結果をFig. 6に示す。提案手法によって人の判断した3箇所の正解分節点を漏れなくほぼ同じタイミングで検出できることを確認した。

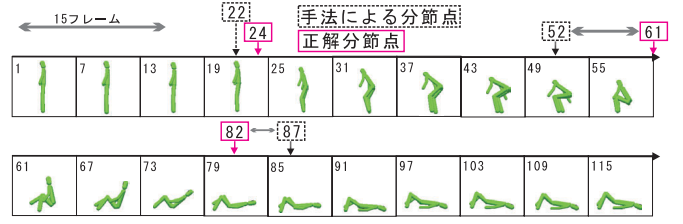


Fig. 6: 寝転ぶ動作に対する分節化結果

また、遅延時間については、オンラインで運動データが入力されてから分節点確定されるまでに平均996msの遅延が生じた。本手法においては候補点を分節点として確定するまでに一定時間 W_z だけ保留していると共に、時系列尤度のマッチングのために候補点より未来側 W_l 分の情報も利用するため、処理時間に依存せず必ず $W_z + W_l = 29$ フレーム（約967ms）だけの遅延が生じることを考えると、ほぼ必ず生じてしまう遅延のみでオンライン分節化処理を行うことができていることを確認した。動作認識に伴う分節点の自動抽出への手法の実応用を考えた場合、約1秒の遅れは、オンラインで入力されるデータに対する出力の遅延としては、許容できるものと考えられる。

5 結論

本手法では、人による分節化の教示を利用し、力学系の切替モデルにより区切れ目区間をモデル化した後、力学系の切替モデルの時系列尤度を利用してオンラインによる分節化手法を構築した。モーションキャプチャデータを用いた分節化適用実験により、提案手法の有効性を明らかにした。このように、力学系の切替モデルによって日常動作の区切れ目区間を適切にモデル化できたと共に、分節点付近の時系列尤度パターンの特徴に着目することで、高速性と遅延性を考慮した分節化を実現することができた。将来課題としては、着目身体部位の違いに基づくクラスタリング結果を有効に利用して精度を向上させることが挙げられる。

文 献

- [1] A. Bobick and J. Davis. The Recognition of Human Movement Using Temporal Templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 3, pp. 257–267, 2001.
- [2] T. Shiratori, A. Nakazawa, and K. Ikeuchi. Detecting Dance Motion Structure through Music Analysis. In *Proc. of the 6th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 857–862, May 2004.
- [3] 川嶋宏彰, 堤公孝, 松山隆司. 動的イベントの分節化・学習・認識のための Hybrid Dynamical System. 第3回情報科学技術フォーラム (FIT2004), pp. 175–178, September 2004.
- [4] Y. Segawa, T. Mori, M. Shimosaka, and T. Sato. Human Like Segmentation of Daily Actions based on Switching Model of Linear Dynamical Systems and Human Body Hierarchy. In *Proc. of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5859–5865, October 2006.
- [5] Y. Li, T. Wang, and H. Shum. Motion Texture: A Two-Level Statistical Model for Character Motion Synthesis. In *Proc. of ACM SIGGRAPH 2002*, pp. 465–472, July 2002.
- [6] V. Pavlovic, J. M. Rehg, and J. MacCormick. Learning Switching Linear Models of Human Motion. In *Proc. of Neural Information Processing Systems*, pp. 626–632, December 2000.
- [7] R. Kalman. A New Approach to Linear Filtering and Prediction Problems. In *Trans. of the ASME-Journal of Basic Engineering*, pp. 35–45, March 1960.
- [8] J. Lin, E. Keogh, P. Patel, and S. Lonardi. Finding Motifs in Time Series. In *Proc. of the 2nd Workshop on Temporal Data Mining, at the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 301–302, 2002.