

人体姿勢追跡のための複数動作モデル間の遷移経路生成

浮田 宗伯^{†a)} 市田 憲明[†] 木戸出正継[†]Synthesizing Transition Paths among Multiple Motion Models
for Human Pose TrackingNorimichi UKITA^{†a)}, Noriaki ICHIDA[†], and Masatsugu KIDODE[†]

あらまし 人体の動きを動作ごと（歩行，ジョギングなど）にモデル化した学習データを参照し，映像からその姿勢を追跡する手法を提案する．学習は基本的な動作ごとに行うため，(1) 各動作モデルを適宜追加可能，(2) 各モデルはその動作に特化した最適化が可能となる．追跡時は，観測動作に合わせて動作モデルを自動的に切り換えながら姿勢を追跡する．この切換が遅れると，動作間をまたぐ人の動きに対する追従性が損なわれてしまう．提案手法では，モデル間で動作の遷移が発生する可能性のある箇所をデータの類似性に基づいて特定し，その遷移候補点間を滑らかにつなぐ姿勢変化を仮想的に生成することにより，モデル間の即応的な遷移及び遷移時の姿勢追跡精度の向上を実現する．実際の計測データとこれらの仮想的に追加したデータを学習データとして用いることにより，姿勢追跡の精度向上が可能であることを示す．

キーワード 人体姿勢追跡，姿勢変化の動的モデル，複数動作モデル，動作モデル切換，モデル間遷移経路

1. ま え が き

コンピュータビジョンにおける人体の姿勢推定法は，モデルフィッティングをベースとする手法 [1], [2] と回帰モデルをベースとする手法 [3], [4] に大別できる．シミュレーション可能に簡易化された人体モデル（円筒モデルなど）と入力画像から得られる人体の特徴量（シルエットなど）との相関を調べるモデルフィッティングと比較して，回帰ベースの手法では形状の近似を介さず直接特徴量から姿勢を得ることができる．

しかし，回帰モデルを得るためには事前に同期した特徴量と姿勢データを獲得しておく必要があり，観測される人の動きが獲得済みのモデルから大きく外れると姿勢推定できないという短所がある．回帰ベースの手法においてこの短所を抑えるためには，多くの従来法のように単一動作のみをモデル化するのではなく，複数の動作を統合的にモデル化しておくことが有効である．こうした複数動作の姿勢データは，動きのモデリングや推定の精度評価などを目的としたデータベ

スが数多く公開されている．

これらのデータベースでは，複数動作がそれぞれ独立に計測・記録されている．各データはそれぞれの動作の推定に適用できるが，異なる動作をまたぐ際にそのまま当てはめると推定精度が落ちてしまう．しかし，個人内・個人間の動きのばらつきも含めて全ての動作間で遷移動作を計測して学習データに含めることは現実的ではない．本論文では，こうした動作間の遷移に対応するデータを計測動作集合から仮想的に生成してモデル化しておくことにより，動作をまたいで変化する人の姿勢を継続的に追跡する手法を提案する．

2. 関連研究

各時刻における姿勢推定ではなく，過去の姿勢を参照した追跡を行うことでより安定に姿勢を得る研究が数多く存在する [5]．追跡のために近年広く用いられているパーティクルトラッキングは，ある程度の次元のデータには有効であるが [6], [7]，人体姿勢のような高次元データ（全身の関節角度は 50 次元以上）に対して適用すると精度・安定性が低下してしまう．

高次元性に対応するためには，データの低次元化が有効である．低次元化には，線形な主成分分析や非線形写像によって効率の高い低次元化を実現する GPLVM [8]

[†] 奈良先端科学技術大学院大学，生駒市

Nara Institute of Science and Technology, Ikoma-shi, 630-0192 Japan

a) E-mail: ukita@is.naist.jp

などがある．GPLVM の拡張には，GPDM（時系列拡張）[9] や SGPLVM（複数データの共通低次元空間推定）[10] などがある．GPDM は，低次元空間におけるダイナミクス（時刻 t から $t + 1$ におけるデータの予測）を得るため，姿勢追跡時の動き予測に活用できる．SGPLVM を利用すると低次元空間を介して画像特徴量と人体姿勢データを同期対応付けできるため，回帰モデリングとの親和性が高い．

多様な姿勢に対応した回帰モデルを得るためには，複数の動作を学習しておく必要がある．文献 [11] では複数動作から一つの低次元空間を生成しているが，多様な姿勢を単一モデリングすることは難しく回帰精度が低下してしまう．文献 [12] では，異なる動作ごとの低次元空間化が提案されている．各動作に合わせて独立に複数のモデルを学習することにより，観測動作に応じたモデル切替が必要になる一方，各動作に最適化されたモデルが得られるというメリットがある．

しかし，従来モデル [11], [12] は動作間の遷移状態を含まないので，(1) その間の姿勢推定精度が落ち，(2) その間のダイナミクスが未知である（動き予測ができない）ため動きへの追従性が落ちてしまう．特に動作ごとの複数モデルを利用する場合，最悪の場合モデルの切替が行われず観測動作とは異なる動作モデルでの追跡が続く可能性もある．この問題は，動作ごとの計測データから遷移動作を生成して学習モデルに組み込むことにより回避できる．

遷移動作を自動的に生成する手法はキャラクターアニメーションの分野で数多く研究されている [13]．この研究では，様々な動作の中で類似している動き同士をつないで補間することにより遷移経路を生成している．

本論文では，複数動作をそれぞれモデル化する際，それらの間の遷移経路を学習データに加えることで動作をまたいだ姿勢推定の高精度化を実現する．

3. 複数動作モデルの独立学習とそれらの切替による姿勢追跡

複数動作モデルを適宜選択的に参照する提案手法の概要を，学習と追跡の 2 段階に分けて述べる．

3.1 動作モデルの学習

学習の流れを図 1 に示す．まず，複数台のカメラ及びモーションキャプチャシステムを用い，各動作の多視点時系列画像 (Multi view image sequence) と時系列姿勢データ (Pose sequence) $Y = [y_1 \dots y_N]^T$ (N

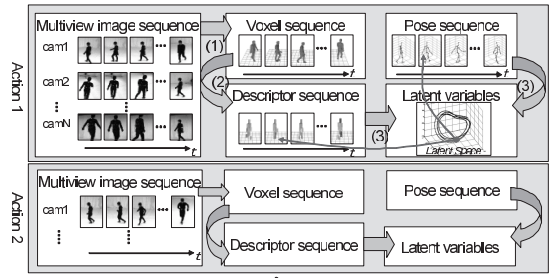


図 1 学習処理 (1) 形状復元，(2) 形状記述子計算，(3) GPDM による姿勢データの潜在変数獲得，(4) 形状記述子から潜在変数への写像関数の獲得．細い曲線矢印は対応する同期データを表す．

Fig. 1 Learning scheme. (1) Volume reconstruction, (2) Shape descriptor computation, (3) Acquiring pose latent variables by GPDM, (4) Acquiring the mapping function from the shape descriptors to the latent variables. Two thin arrows indicate corresponding synchronized data.

は同期データのフレーム数を表す) を同期獲得し，各動作データについて独立に以下の処理を施す．

得られた時系列画像には視点依存性がある．そこで，視体積交差法 (SFS: Shape From Silhouettes) によりボクセル形式の時系列 3 次元人体形状を復元する (図 1 の (1))．得られたボクセルデータ (Voxel sequence) は高次元であるため，姿勢追跡や運動解析においては人体形状を効率的に表現できる特徴量を抽出することが多い (3D Shape Context [14] など)．我々は Volume Descriptor [15] を用いてこのボクセルデータから形状記述子 $V = [v_1 \dots v_N]^T$ (Descriptor sequence) を計算する (図 1 の (2))．

時系列姿勢データに対しては，Gaussian Process Dynamics Models (GPDM) [9] を適用し，時系列姿勢を低次元に表現する潜在変数 $X = [x_1 \dots x_N]^T$ を得る (図 1 の (3))．このとき，同時に潜在空間から姿勢空間への写像及び潜在空間におけるダイナミクスを計算するための写像がガウス過程回帰として得られる．最後に潜在空間から形状記述子への写像を別途ガウス過程回帰により求める (図 1 の (4)) ことにより，文献 [10], [16] で提案されているような「形状空間と姿勢空間により共有される潜在空間」が得られる．

以上の処理により，以下のデータが得られる．

- 姿勢データの潜在空間 X
- X におけるダイナミクス (時刻 $t - 1$ から t への遷移) を得る写像 $m : (X \rightarrow X)$

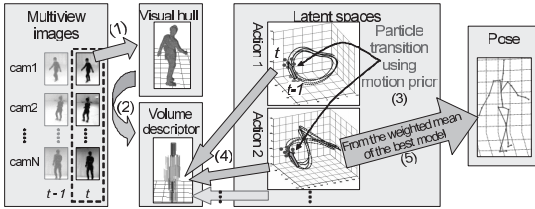


図2 追跡処理 (1) 形状復元, (2) 形状記述子計算, (3) パーティクルの時間遷移, (4) パーティクルから形状記述子への写像, (5) 推定された現在状態を表す潜在変数から姿勢への写像

Fig. 2 Pose tracking scheme. (1) Volume reconstruction, (2) Shape descriptor computation, (3) Temporal drift of particles, (4) Mapping from the particles to the shape descriptors, (5) Mapping from the latent variable of the current state to the pose.

- X から姿勢空間 Y への写像 $f: (X \rightarrow Y)$
- X から形状記述子 V への写像 $g: (X \rightarrow V)$

これらをまとめて本論文では動作モデルと呼ぶ。動作モデルは動作ごとにそれぞれ生成する。動作モデルの分け方は GPDM の特性から同じ動作を繰り返せる程度の長さの動作とした。同じ動作を繰り返し行うことでガウス過程回帰の汎化性能の向上が期待でき、最終的な姿勢追跡精度の向上につながると考えられる。

3.2 動作モデルの切替を伴う姿勢追跡

追跡時の処理の流れを図2に示す。

まず、同期カメラ群により人体の動きを撮影する。学習時と同様に、各時刻 t において画像群から視体積交差法により人体3次元形状を復元し(図2の(1))、この形状データから形状記述子 v_t を計算する(図2の(2))。各潜在空間中で前時刻 $t-1$ におけるパーティクルは写像 m により遷移され、時刻 t におけるパーティクル $x_{t,1}^p \dots x_{t,Nparticle}^p$ が得られる(図2の(3))。各パーティクルは写像 g により形状記述子に変換され(図2の(4))、入力形状との類似度を参照した以下の式によりパーティクルのゆう度が得られる：

$$c_{t,i} = \exp(-w_{var}\sigma_{t,i}^2 - w_{dis}\|v_t - g(x_{t,i}^p)\|) \quad (1)$$

w_{var} と w_{dis} は重み、 $\sigma_{t,i}^2$ は潜在空間におけるパーティクルを姿勢空間に写像する際の分散(GPDMにより獲得済み)である。このゆう度は $\sigma_{t,i}$ が小さいほど、また各パーティクルから計算した形状記述子と観測された形状記述子が近いほど大きくなる。

単一動作モデルであれば、上記処理でそのモデル内の全パーティクルを処理し、それらから時刻 t における状態(形状記述子の推定値)を得る。しかし、動作

モデルが複数ある場合には、観測動作に応じた適切なパーティクルの処理と動作モデル選択が必要になる。

パーティクルのリサンプリング：モデルごとではなく、全モデル中のパーティクル全体で行う。ゆう度の低いパーティクルを消してゆう度の高いパーティクルをもつモデル内に移動させることにより、観測動作に対応するモデルに多くのパーティクルを配置する。

モデルの選択：各モデルにおいて式(1)で得られるパーティクルゆう度の総和を「このモデルが観測動作に対応している評価値」とする。ただし、一時的なノイズの影響などによる不要なモデル切替を抑制するため、 C_{change} 回連続で評価値が最大となったモデルが現れたときにのみモデルを切り換える。

最後に、選択した動作モデルにおいて状態ベクトル \hat{x}_t (パーティクルのゆう度重み平均)を求め、 $y_t = f(\hat{x}_t)$ を計算して時刻 t の姿勢を得る(図2の(5))。

4. モデル間遷移経路の生成による複数動作の姿勢追跡の追従性・精度の向上

3.の処理により、複数モデルを用いた姿勢追跡が可能になった。更に提案手法では、学習データに動作モデル間の遷移形状・姿勢を加えることにより、動作遷移時の姿勢推定精度の向上を実現する。本章では姿勢空間において動作が切り換わるフレームを特定し、そのフレームから動作が切り換わる途中の姿勢を仮想的に生成して学習データに加える手法を提案する。

4.1 姿勢空間における遷移経路の生成

4.1.1 遷移開始・終了する姿勢の抽出

文献[13]のように、動作の遷移は遷移元動作と遷移先動作の類似した姿勢付近で起きると仮定する。“姿勢が類似”とは各関節の3次元位置が近いこととみなし、二つの姿勢 y_i, y_j の類似度を以下のように定義した。

$$s_{i,j} = - \sum_k^{D/3} \|J(y_i - y_j)^T\| \quad (2)$$

J_k は全身姿勢から k 番目の関節の3次元座標を得る行列で、 $J_{3k-2,3k-2} = J_{3k-1,3k-1} = J_{3k,3k} = 1$ 、その他の要素は0である。 $s_{i,j}$ は、姿勢 i, j 間における各関節位置の距離の和に -1 を掛けた値である。

学習データ中の動作1(N^1 フレーム)と動作2(N^2 フレーム)の間の遷移箇所候補を選ぶ際、式(2)を $1 \leq i \leq N^1, 1 \leq j \leq N^2$ を満たす全ての i, j の組合せについて計算する。この $\{s_{i,j}\}$ を2種類の舞踊動

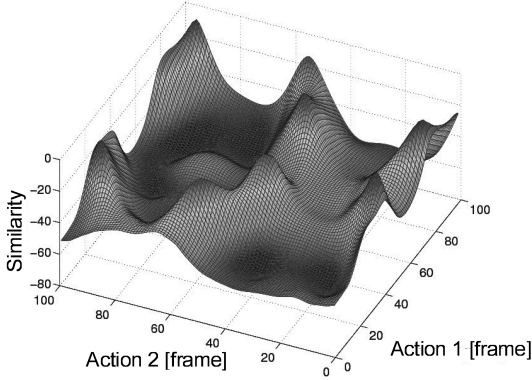


図 3 二つの姿勢シーケンスの類似度を示したグラフ。底面の 2 軸が動作 1 と動作 2 のフレーム、縦軸が各フレームの組合せにおける類似度を表す。

Fig. 3 The graph shows similarity between two pose sequences. Two axes of the base indicate the frame IDs of the sequences of actions 1 and 2. The vertical axis indicates the similarity.

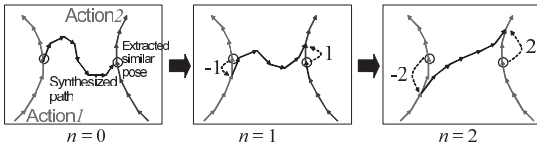


図 4 姿勢空間における補間の様子。両端の矢印は動作 1, 動作 2 の計測姿勢の軌跡を表す。その上の点が抽出された遷移開始・終了の姿勢, それらを結ぶ中央の実線矢印が補間姿勢の軌跡である。

Fig. 4 Interpolating pose data from potential transition points (two circles) from actions 1 to 2.

作^(注1)について計算した結果を図 3 に示す。

$s_{i,j}$ の最大, 最小値を s_{max}, s_{min} とする。 $s_{i,j}$ が

$$s_{i,j} > s_{max} - \alpha(s_{max} - s_{min})$$

満たし, かつ極大値 (図 3 のピーク) となる i と j の組合せを, 遷移が開始・終了する姿勢の候補とする。定数 α を大きくすると候補姿勢が増加する。

4.1.2 遷移経路の生成

次に, 遷移候補点間をつなぐ時系列姿勢データの生成について, 図 4 の動作 1 と動作 2 を例にして述べる。

図 4 の黒色の点 (動作 1 の i フレーム目と動作 2 の j フレーム目とする) を遷移箇所候補とし, 動作 1 から動作 2 へ移り変わる場合を考える。姿勢は動作 1 の i フレーム目から動作 2 の j フレーム目に瞬間的に切り換わるのではなく, 図 4 の右図のように動作 1 と動作 2 の“中間的な動き”を経て滑らかに切り換わる。このような“中間的な動き”を遷移経路とみな

す。“滑らかである”とは姿勢空間における時系列姿勢の軌跡において時刻が $t, t+1, t+2$ と進むとき, $y_{t+1} - y_t$ と $y_{t+2} - y_{t+1}$ のなす角 $\theta_{t,t+1}$ が小さいこととする。そこで, 遷移経路において $\theta_{t,t+1}$ が学習データから以下のように統計的に決定されるしきい値 θ_{max} 以下になるように姿勢データを生成する。まず $1 \leq i \leq N^1 - 1, 1 \leq j \leq N^2 - 1$ の範囲で

$$\theta_{i,i+1} = \arccos \frac{(y_{1,i+2} - y_{1,i+1}) \cdot (y_{1,i+1} - y_{1,i})}{|y_{1,i+2} - y_{1,i+1}| |y_{1,i+1} - y_{1,i}|}$$

$$\theta_{j,j+1} = \arccos \frac{(y_{2,j+2} - y_{2,j+1}) \cdot (y_{2,j+1} - y_{2,j})}{|y_{2,j+2} - y_{2,j+1}| |y_{2,j+1} - y_{2,j}|}$$

を計算する。各動作例におけるこれらの平均 $\bar{\theta}_1, \bar{\theta}_2$, 分散 σ_1^2, σ_2^2 からしきい値を決める。

$$\theta_{max} = \bar{\theta} + \beta \bar{\sigma}^2$$

$\bar{\theta}$ はフレーム数を重みとした $\bar{\theta}_1, \bar{\theta}_2$ の重み付き平均, $\bar{\sigma}^2$ はフレーム数を重みとした σ_1^2, σ_2^2 の重み付き平均, β は姿勢変化の平均値からの逸脱の程度を指定するパラメータであり, β を大きくすれば遷移経路は短くなり, 小さくすれば遷移経路は長くなる。

具体的な遷移経路生成法は以下のとおりである。遷移経路の始点及び終点を $y_{1,i-n}, y_{2,j+n}$ とし, n を 0 から順に 1 ずつ大きくしながら $y_{1,i-n}, y_{2,j+n}$ を補間する。遷移経路における $\theta_{t,t+1}$ は一定になるように計算する。計測姿勢から生成姿勢へのつながりも滑らかになるように, $\theta_{t,t+1}$ は始点と終点から $-1, +1$ ずつずらした $y_{1,i-n-1}$ から $y_{2,j+n+1}$ まで評価することに注意されたい。この $\theta_{t,t+1}$ が θ_{max} 以下になったとき, その遷移経路を「類似姿勢間を滑らかにつなぐ経路」とみなして学習データに追加する。

以降, n を大きくすると滑らかさの制約を満たしたより長い遷移経路を生成できる。遷移経路の本数は, タスク (観測環境, 動作の種類など) に応じて適切に設定すべきである。本論文の実験では各遷移候補姿勢間において 1 本の遷移経路のみを生成した。

補間点数を人手により決める従来法 [13] と異なり, 提案手法ではサンプルデータ同士を滑らかにつなぐように自動的に補間点数が決まるという特徴をもつ。

4.2 形状空間における遷移経路の生成

形状空間から遷移経路を含んだ低次元空間への写像を得るため, 前節で生成した遷移経路に対応する形状

(注1): 5.2 の実験で用いる動作集 1 の動作 1 と動作 2。

データを追加する。

まず、ガウス過程回帰により遷移経路を追加していない姿勢データから形状データへの回帰モデルを得る。この回帰モデルを用いて遷移経路を形状記述子空間に写像する。得られた形状記述子空間中の点を遷移経路に対応する仮想形状データとみなす。この処理により、遷移経路における形状記述子から潜在空間を介して姿勢データを得る回帰モデルを得ることができる。

4.3 拡張性の高い複数動作のモデリング

上述の処理により、以下の手順で N^A 個の動作シーケンス $(1, \dots, N^A)$ から動作モデルの集合を学習する。

L1: 変数 i を 1 に初期化する。

L2: $j < i$ となる全ての動作シーケンス j に対して、動作シーケンス i との間で遷移経路を生成し、その経路上の姿勢データと形状記述子を得る。これらの生成データを全て動作シーケンス i に組み込んで GPDM を適用することにより、動作 i のモデルを得る。

L3: もし $i < N^A$ であれば $i = i + 1$ として L2 に戻る。 $i = N^A$ であれば学習処理を終了する。

新たに動作シーケンス $N^A + 1$ が得られた際には、 $i = N^A + 1$ として L2 を実行するのみである。この際、GPDM による最適化済みの $i = 1, \dots, N^A$ の動作モデルは再最適化不要である。モデル最適化における主要な計算量はサンプルデータ数 N^S の 3 乗のオーダ [17] なので、一つのシーケンスとそれに付随する短い遷移経路のみから各動作モデルを得る提案手法は、全シーケンスを単一モデルで学習する手法 [11], [12] と比較して学習コストを大幅に低減できる。

4.4 複数モデルの切替による連続動作の姿勢追跡

パーティクルトラッキングのイメージを図 5 に示す。この例は、動作が Action1 から Action2 に切り換わるときの潜在空間の様子である。時刻 t が遷移の開始時点であり、この段階では遷移経路の有無にかかわらずパーティクルの配置は同じである。開始から少し時間が経過した時刻 t 以降では、遷移経路の有無によりパーティクルの分布と状態変数の位置が異なってくる。

図 5 上部に示す遷移経路がない状態では、学習データ（図中の実線）の方向にしかダイナミクスが学習されていないため、パーティクルは学習データに沿って移動していく（時刻 $t + 1$ ）。また、その時刻における観測形状の特徴量 v_t に近いパーティクル $x_{t,i}^p$ が存在しても、GPDM の潜在空間において学習データから離れた位置は分散が大きく、そのパーティクルのゆわ度が下がってしまう。その結果、ゆわ度重み付きの

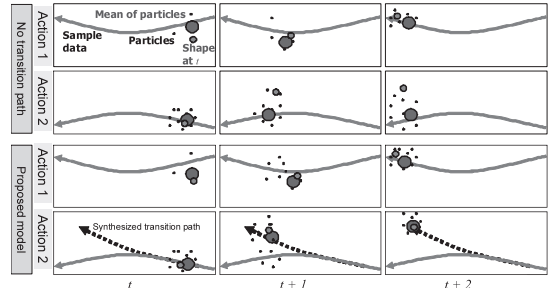


図 5 生成された遷移経路の有無によるパーティクルトラッキングの変化。実線は学習データの潜在変数、点線は生成された遷移経路、小点はパーティクル、中点と大点はその時刻の観測形状と状態変数を表す。

Fig. 5 Transition of particles between different actions. Solid and dotted lines show latent variables of real sample data and synthesized data, respectively. Small, middle, and large points depict particles, mapping points from the volume descriptors, and current states, respectively.

パーティクル平均である状態変数は、 $t + 2$ のように時刻が進んでも学習データに近接した位置になる。

これに対して遷移経路を追加した動作モデルでは、パーティクルが学習データと遷移経路上にそれぞれ進む（時刻 $t + 1$ ）。その結果、形状特徴量より写像された点が遷移経路上にあればゆわ度の計算によって最終的に状態変数の位置は遷移経路付近になる。そして Action1 と Action2 の中間の姿勢を経て Action2 へ切り換わる。このように中間的な姿勢を経由することにより「遷移中の姿勢・形状を補間学習することによる姿勢回帰精度の向上」と「観測される動きに即応した滑らかな追跡」が行われる。

5. 実験と評価

複数の基本的な動作を計測・学習した後、これらの動作を組み合わせた動作の姿勢追跡を行った。

5.1 実験環境

- 同期多視点時系列画像

1024 × 768 pixel, 30 fps の同期カメラ 12 台を撮影対象の人間を取り囲むように配置して撮影した。各カメラの内部及び外部パラメータは、事前にキャリブレーション [18] しておいた。

- 時系列姿勢データ

ジャイロ式モーションキャプチャシステム IGS-190 を用いて計測した。計測データから全身 25 箇所の部位の腰からの相対座標を計算して姿勢データとした。

全姿勢データは画像データと同期されている。

5.2 実験データ

[学習データ] 2種の動作集を用意した。

動作集1(舞踊動作): 基本的な動作が組み合わされている動作の例として舞踊動作を選んだ。学習した動作は以下の4種類である。

- 動作1 両手を右上左上に動かす動作
- 動作2 両手を右上左下に動かす動作
- 動作3 両手を右下左上に動かす動作
- 動作4 両手を右下左下に動かす動作

動作集2(歩行等): 日常的な動作である歩行群の学習データを用意した。具体的には次の6種類である。

- 動作1 歩き出してすぐに立ち止まる
- 動作2 ゆっくり歩く
- 動作3 普通で速度で歩く
- 動作4 早く歩く
- 動作5 大股で歩く
- 動作6 ジョギング

各動作の学習データは300フレームからなり、1名(被験者1)の動きデータから構成されている。

[テストデータ] 動作集1と2の各々について、学習データとして計測した動作を様々な順で組み合わせた動作を観測した。テストデータは、被験者1に2名(被験者2と3)を加えた動きデータからなる。

[実験で用いたパラメータ] 実験で用いるパラメータは経験的に以下のように設定し、全実験において同じ値を用いた。 $N^{particle} = 500$, $w_{var} = 0.5$, $w_{dis} = 10$, $C_{change} = 3$, $\alpha = 0.1$, $\beta = 2.0$ 。

[誤差の計算方法] 推定誤差は、全身計25箇所関節と端点についてモーションキャプチャにより得た真値との差分(全箇所における誤差のユークリッド距離の平均)により評価した。追跡結果を $\mathbf{y}_1 \cdots \mathbf{y}_{N^{frame}}$, 対応する真の姿勢を $\mathbf{y}_1^* \cdots \mathbf{y}_{N^{frame}}^*$ としたときの誤差計算式は以下のとおりである。

$$\frac{1}{N^{frame} \cdot D/3} \sum_i \sum_k \|\mathbf{J}_k(\mathbf{y}_{i,k} - \mathbf{y}_{i,k}^*)^T\| \quad (3)$$

N^{frame} はテストシーケンスのフレーム数、 \mathbf{J}_k は全身の姿勢を表すベクトル \mathbf{y}_i から k 番目の関節の3次元座標を抜き出す行列である(式(2)参照)。

提案法で学習・推定される姿勢データは関節の3次元座標であるが、ジャイロ式モーションキャプチャ

からは関節角が得られる。この角度と各被験者に対する事前のキャリブレーションで得られる「関節間の長さ」から関節の3次元座標を得ている。被験者2と3の観測時には、モーションキャプチャで得られた関節角に対して被験者1の関節間長を適用して真値の3次元座標を得て、この値を式(3)により推定3次元座標と比較した。また、形状データは身長で正規化した。

5.3 実験結果と考察

最初に、全動作の単一モデル化と各動作に分けたモデル化(提案法)におけるモデル学習に要する時間を計測した。動作集1を対象にすると、単一モデル化(300フレーム×4種=1200フレームのモデル化)には88分、分割モデル化(300フレームのモデル化×4回)には8分を要した。このことから、学習時間における分割モデル化の優位は明らかである。

5.3.1 同一人物の学習データと推定データ

動作モデルを複数にすることが単一動作シーケンス(動作遷移のないシーケンス)における追跡精度に与える影響を確かめるため、1種類の動作のみを観測したテストシーケンスに対して下記3種のモデルを利用した姿勢追跡の誤差を評価した。

- その単一動作のみを学習(1 model)
- 各動作集中の全動作を遷移経路なしで学習(no path)
- 各動作集中の全動作に遷移経路を付加して学習(proposed method)

動作集1と2の動作1をそれぞれ観測したテストシーケンスにおける誤差を表1に示す。

このテストデータは動作遷移を含まないので、学習データが複数種類の動作を含んでいると無駄な動作モデル切替により誤差が大きくなると予想した。動作集2では予想どおりの結果が得られたが、動作集1では提案手法を含めた複数動作モデルの誤差の方が小さかった。これは、4種の動作が類似姿勢を重複して含むため、モデル全体としてその姿勢の学習データが増えて回帰写像が汎化されたためと考えられる。

次に、遷移経路の追加が動作遷移の発生する際の姿勢追跡精度に与える影響を検証する。動作集1と2の

表1 単一動作における姿勢追跡の平均誤差 [cm]/分散
Table 1 RMS error [cm]/variance with single-action sequences.

	1 model	no path	proposed method
action set 1	4.3/1.0	4.0/1.2	4.0/0.9
action set 2	3.1/0.9	3.6/1.1	3.3/0.7

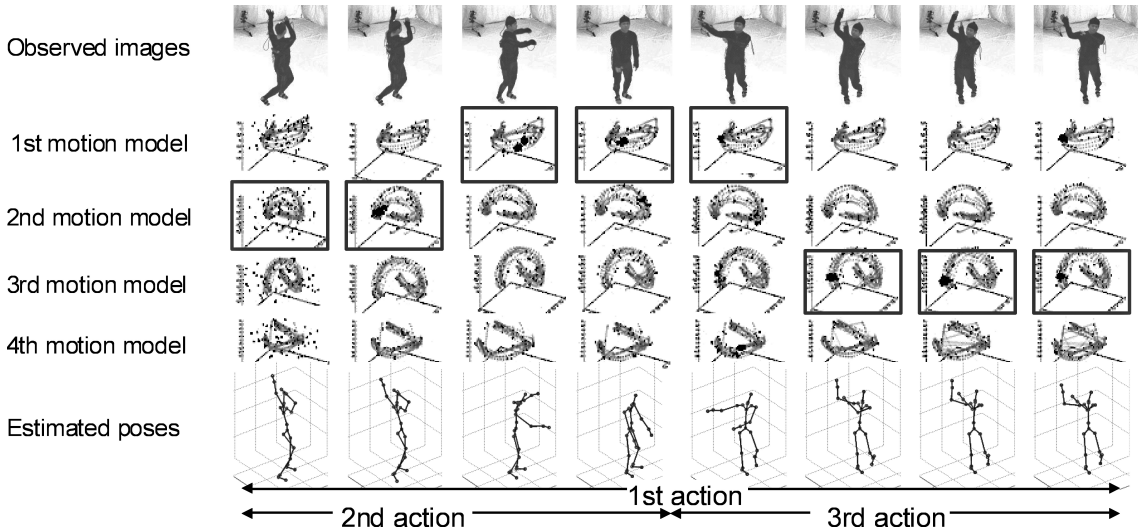


図 6 動作集 1 の混合シーケンスにおける姿勢追跡結果．方形で囲ったモデルが各時刻の姿勢回帰で選ばれた動作モデルを表す．モデル内の学習データとパーティクルをそれぞれ灰と黒の点で示す．最下段は人手により得た対応する動作の種類である．

Fig. 6 Pose tracking results in action set 1. Gray and black points in models depict sample data and particles, respectively. The selected model at each moment is enclosed by a rectangle.

表 2 混合動作における姿勢追跡の平均誤差 [cm]/分散
Table 2 RMS error [cm]/variance with multi-action sequences.

	no path	proposed method
action set 1	5.4/1.9	5.0/1.4
action set 2	4.6/1.4	3.9/1.2

表 3 モデル切替発生時のフレームの前後 5 フレームの平均誤差 [cm]/分散

Table 3 RMS error/variance with multi-action sequences: the frames in which the model selected was changed \pm five frames.

	no path	proposed method
action set 1	9.6/2.7	9.0/2.1
action set 2	10.8/2.6	8.5/1.5

それぞれにおいて、遷移経路のないモデル (no path) とあるモデル (proposed model) で複数動作を組み合わせたテストシーケンスの姿勢追跡を行った。結果を表 2 に示す。いずれのモデルでも単一動作のみを観測した場合 (表 1) よりも誤差が大きいが、遷移情報を付加した提案モデルの方が誤差が小さい。このときの入力画像、モデル (潜在空間) 内におけるパーティクルの様子、姿勢推定結果を図 6 と図 7 に示す。モデル内の黒、灰の点がそれぞれパーティクル、学習データの潜在変数を示す。方形で囲ったモデルがその時刻に選択された動作モデルを表す。

動作集 1 の結果 (図 6) を例に追跡処理を確認する。追跡開始直後である最左のフレームではパーティクルは広く分布しているが、その後は観測動作の周辺に集まるように分布を変えている。パーティクルが集まっている位置と姿勢の変化に注目すると、姿勢変化が大きいとき (図 6 の左から 3, 4, 5 番目) はパーティク

ルの移動速度も大きく、姿勢変化が小さいとき (図 6 の右から 1, 2, 3 番目) はパーティクルの移動速度も小さい。更に、動作の変化に伴ってパーティクルがその動作のモデルへと遷移している様子が分かる。動作集 1 では動作間の重複が大きい、その中で適切な動作モデルのいずれかが選択されている。

最後に、動作遷移時の姿勢追跡性能を検証するため、モデル切替が発生したフレームの前後 5 フレームの誤差を計算した。結果を表 3 に示す。追跡が難しい瞬間であるため表 2 の結果と比較して誤差が大きいが、動作集 1 と 2 のどちらにおいても提案手法の方が特に誤差が小さい。この結果から、動作遷移の瞬間に注目した提案手法の有効性が確認できる。

5.3.2 異なる人物の学習データと推定データ

提案手法では、少数の学習データからでも GPDM による汎化性が期待される。そこで、被験者 1 のみを

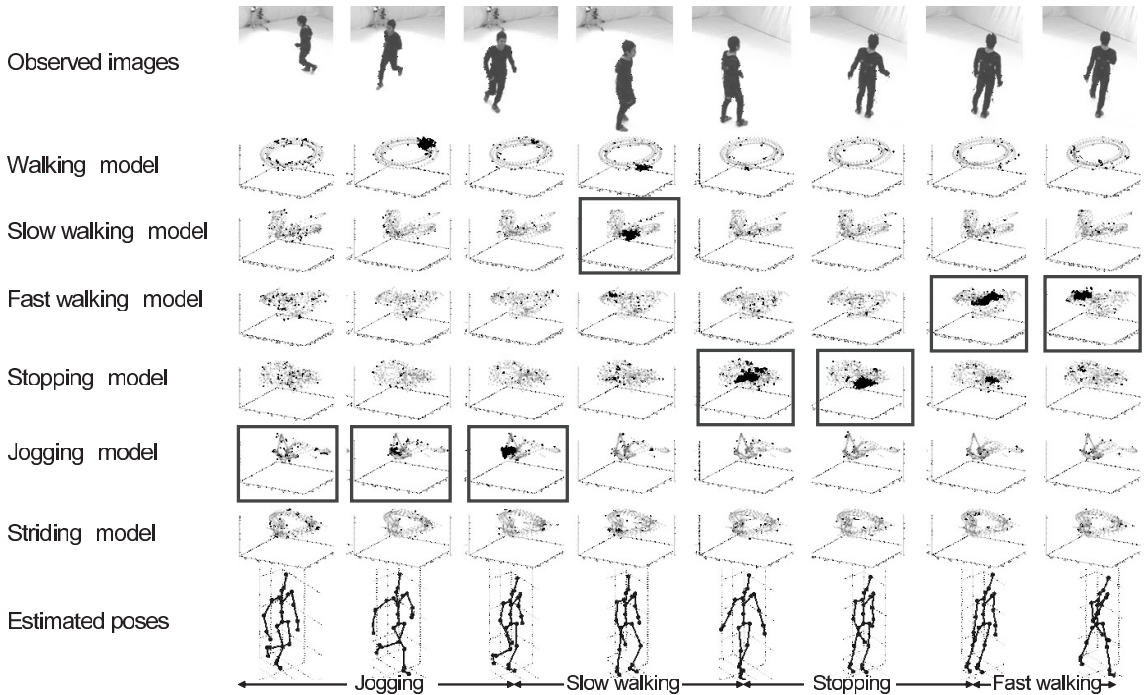


図 7 動作集 2 の混合シーケンスにおける姿勢追跡結果 .
Fig. 7 Pose tracking results in action set2.

表 4 学習データ外の未知の人 (被験者 2 と 3) の動きに対する姿勢追跡の平均誤差 [cm]/分散 . E1 : 単一動作における評価, E2 : 複数動作における評価, E3 : モデル切替前後に限った評価

Table 4 RMS error [cm]/variance with test sequences of other subjects. E1: single-action sequences, E2: multi-action sequences, E3: sub-sequences around model changes.

		1 model	no path	proposed method
E1	action set 1	4.9/1.4	4.8/1.3	4.9/1.3
	action set 2	4.1/1.1	4.1/1.2	3.9/0.8
E2	action set 1	NA	6.1/2.0	5.5/1.4
	action set 2	NA	5.0/1.8	4.5/1.2
E3	action set 1	NA	9.3/3.8	8.3/1.9
	action set 2	NA	11.4/3.1	8.1/1.7

観測した学習データを被験者 2 と 3 のテストデータに適用した姿勢追跡により汎化性能を確認する。復元形状は各被験者の身長に応じて正規化した。

表 4 に、5.3.1 で示した同一人物に対する実験と同様の実験の結果を示す。当然同一人物を対象にした結果よりも全体的に誤差が大きくなってしまっているが、表 4 内の相対的な誤差の大小は同一人物を対象にした結果 (表 1, 表 2, 表 3) と等しい傾向が観測されており、提案手法の有効性が確認できた。

以上、5. の結果から、(1) 各動作に対応する複数モデルを自動的に切り換えて参照することにより、単一動作を単一モデルで追跡する際と同様の推定精度を得ることができ、(2) 学習時に計測していない遷移動作を含むシーケンスを対象にして、補間生成された遷移経路を学習した提案手法の方が (特に動作遷移の前後において) 姿勢推定誤差を抑えることができており、(3) 学習データ外の未知の人間に対しても適用可能な汎化性能を備えている、ことを確認できた。

6. む す び

複数の動作モデルを用いた人体姿勢追跡を提案した。学習時にモデル間の遷移情報を仮想的に追加することで、動作切替わり時の姿勢追跡精度を向上できた。

本論文では複数視点画像からの復元形状を特徴量として姿勢推定を行ったが、提案モデルはシルエットなどの単眼特徴量からの姿勢推定 [1], [2] に応用できる。

今後の展望として、生成経路のもっともらしさに関する考察が必要である。動作遷移の開始・終了フレームの特定もより実際の動作に近いほど経路生成の精度向上につながる。こうした評価や精度向上のため、人体関節構造の制約や速度・加速度などの力学的な知見

(文献 [19], [20] など) の参照を検討していく。

謝辞 GPDM [9] については Dr. Neil D. Lawrence にソフトウェアを提供して頂いた。深謝致します。

文 献

- [1] R. Urtasun, D.J. Fleet, A. Hertzmann, and P. Fua, "3d people tracking with Gaussian process dynamical models," CVPR, pp.238–245, 2006.
- [2] A. Agarwal and B. Triggs, "Tracking articulated motion using a mixture of autoregressive models," ECCV, pp.54–65, 2004.
- [3] Y. Sun, M. Bray, A. Thayananthan, B. Yuan, and P.H.S. Torr, "Regression-based human motion capture from voxel data," BMVC, 2006.
- [4] A. Fossati, M. Salzmann, and P. Fua, "Observable subspaces for 3d human motion recovery," CVPR, pp.1137–1144, 2009.
- [5] S. Hou, A. Galata, F. Caillette, N.A. Thacker, and P.A. Bromiley, "Real-time body tracking using a Gaussian process latent variable model," ICCV, 2007.
- [6] J. MacCormick and A. Blake, "A probabilistic exclusion principle for tracking multiple objects," IJCV, pp.572–578, 1999.
- [7] J. Sullivan, A. Blake, M. Isard, and J. MacCormick, "Bayesian object localisation in images," IJCV, vol.44, no.2, pp.111–135, 2001.
- [8] N.D. Lawrence, "Gaussian process latent variable models for visualisation of high dimensional data," NIPS, 2003.
- [9] J.M. Wang, D.J. Fleet, and A. Hertzmann, "Gaussian process dynamical models," NIPS, 2005.
- [10] A.P. Shon, K. Grochow, A. Hertzmann, and R.P.N. Rao, "Learning shared latent structure for image synthesis and robotic imitation," NIPS, 2005.
- [11] A. Geiger, R. Urtasun, and T. Darrell, "Rank priors for continuous non-linear dimensionality reduction," CVPR, pp.880–887, 2009.
- [12] R. Urtasun, D.J. Fleet, A. Geiger, J. Popovic, T. Darrell, and N.D. Lawrence, "Topologically-constrained latent variable models," ICML, pp.1080–1087, 2008.
- [13] L. Kovar, M. Gleicher, and F.H. Pighin, "Motion graphs," SIGGRAPH, pp.473–482, 2002.
- [14] M. Kortgen, G.J. Park, M. Novotni, and R. Klein, "3D shape matching with 3D shape contexts," 7th Central European Seminar on Computer Graphics, 2003.
- [15] N. Ukita, M. Hirai, and M. Kidode, "Complex volume and pose tracking with probabilistic dynamical models and visual hull constraints," ICCV, 2009.
- [16] C. Henrik Ek, P.H.S. Torr, and N.D. Lawrence, "Gaussian process latent variable models for human pose estimation," MLMI, pp.132–143, 2007.
- [17] N.D. Lawrence and A. Hyvarinen, "Probabilistic non-

linear principal component analysis with Gaussian process latent variable models," J. Mach. Learn. Res., vol.6, pp.1783–1816, 2005.

- [18] Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. on PAMI, vol.22, no.11, pp.1330–1334, 2000.
- [19] A. Safonova, J.K. Hodgins, and N.S. Pollard, "Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces," ACM Trans. Graph., vol.23, no.3, pp.514–521, 2004.
- [20] M. Vondrak, L. Sigal, and O.C. Jenkins, "Physical simulation for probabilistic motion tracking," CVPR, 2008.

(平成 22 年 9 月 15 日受付, 23 年 1 月 4 日再受付)



浮田 宗伯 (正員)

2001 京都大学大学院博士後期課程了。同年奈良先端科学技術大学院大学情報科学研究科助手。2007 同准教授。2002～2006 まで科学技術振興機構さきがけ(「情報基盤と利用環境」領域) 研究員兼任。2007～2009 までカーネギーメロン大学客員研究員兼任。博士(情報学)。コンピュータビジョン, 分散協調視覚, 対象追跡, 人体運動解析に関する研究に従事。1999 本会論文賞受賞。



市田 憲明

2010 奈良先端科学技術大学院大学修士課程了。現在,(株)東芝勤務。在学中, 人体の動作解析に関する研究に従事。



木戸出正継 (正員:フェロー)

1970 京都大学大学院工学研究科修士課程了。同年東京芝浦電気(現, 東芝)総合研究所入社。同社総合企画部, 関西研究所, 東芝アメリカ社を経て, 2000 奈良先端科学技術大学院大学情報科学研究科教授。京都大学工博。ロボットビジョン, ヒューマンインタフェースに関する研究に従事。情報処理学会, IEEE, IAPR 各フェロー, 本会業績賞, 高柳記念奨励賞, などを受賞。情報処理学会関西支部長, 本会理事, 本会情報システムソサイエティ会長などを歴任。