ガウス過程回帰を用いた階層的学習手法の提案と 人物3次元動作推定への応用

松本 鮎美 ウ 小軍 松浦 宣彦

†日本電信電株式会社 NTTサイバースペース研究所

〒 239-0847 神奈川県横須賀市光の丘 1-1

E-mail: †{matsumoto.ayumi,wu.xiaojun,matsuura.norihiko}@lab.ntt.co.jp

あらまし 動的システムをモデル化する際には,一つのシステムに対して一つのモデルを用いるのが一般的である.しかし,このような事前モデルを用いて観測データからシステムの状態推定を行う場合,観測条件の変動などにより推定誤差が大きくなる傾向にある.これは,観測データの欠落や変動のためである.本研究では,観測データの変動によって起こる推定誤差の低減を目指し,事前モデルを構築する際に,学習データから部分的なデータを取り出し,全体データと部分データのそれぞれに対してモデル化を行う.さらにそれらの複数のモデルに対して関連付けを行うことで,一つの動的システムに対して階層的なモデル構築を行う方法を提案する.本論文では,この階層的モデルを人物の 3 次元動作推定に適用し,観測データの欠落や観測条件の変動にも安定して 3 次元動作推定が可能であることを確認した.

キーワード 動的システムの学習,階層的学習,モデルベース人体姿勢推定,ガウス過程,人物トラッキング

1. はじめに

株価の変動やある地域の天気の変化,音声認識に用いられる時系列の音声特徴量のように時間的に変化するシステムを動的システムという.動的システムをモデル化して解析したり,過去の観測値をもとに次の時刻の値を予測することは多くの分野に共通な興味対象である[1].ビジョンの分野では,例えば映像から人物の動作を解析・推定することで,映像監視やヒューマンインタラクション,ロボットの運動制御,CGアニメーションへの動作あてはめ,医療分野等,幅広い応用が期待される[2]~[4].

近年では、少数カメラ映像から如何に3次元的な動作情報の推定を実現するかに注目が集まっている。画像情報、特に、少数カメラ映像から、3次元動作を推定する場合、3次元情報の欠如による姿勢の多義性や手足や障害物による遮蔽などが問題となる。これらの問題に対処するため、推定の事前情報として人体の3次元的な動きや姿勢に関する動作データから汎用的な動作モデルを構築し、それを事前知識として用いる方法が有効となる。

動作モデルを事前知識として,2次元映像から人体の3次元動作を推定する従来の方法では,学習時に全身の3次元動作データから一つの動作モデルを構築する.このため,推定時に全身の情報が観測されることが理想的である.一般に,実環境下の2次元映像から,画像特徴量等を用いて人物の全身情報を安定して検出することは本質的に困難な問題である.

そこで,本研究では,人物動作をモデル化する際に, 人体構造の全身情報から構成される全身動作モデルと部 分的情報から構成される部分動作モデルという複数の動的システムの結合として,階層的動作モデルを考える.これにより,動作の部分的な情報のみが観測される状況下では,部分動作モデルを事前モデルとすることで,安定して全身動作が推定可能となる.ここで,複数のモデルを階層的に結合する際に,モデル間の関連を如何に表現するかが重要な設計項目となる.本論文では,二つの動作モデル間の関連をガウス過程回帰により定式化し,部分動作モデルの情報を既知として与えたときに,適切な回帰関数を選択することで全身の動作情報を再構築出来ることを確認した.さらに,提案する階層的動作モデルが実際の映像から人物の3次元動作推定を行う際の事前モデルとして有効であることを確認した.

2. 研究背景

動的システムをモデル化する際,一般的に観測データはシステムの潜在的な状態に依存するとして次式の状態空間モデルを想定する.

$$\mathbf{x}(n) = f(\mathbf{x}(n-1)) + \eta_x(n) \tag{1}$$

$$\mathbf{y}(n) = g(\mathbf{x}(n)) + \eta_u(n) \tag{2}$$

この時,式 (1) を状態方程式と呼び $\mathbf{x}(n)$ は時刻 n のシステムの状態(潜在変数)を, $\eta_x(n)$ はノイズを表す.また,式 (2) を観測方程式と呼び $\mathbf{y}(n)$ は時刻 n のシステムからの出力データを, $\eta_y(n)$ はノイズを表す.動的システムをモデル化するためには,式 (1) の状態方程式と式 (2) の観測方程式のパラメータを学習データから推定することが必要である.

近年では,高次元の出力データからシステムの本質的な状態(低次元潜在変数)を見つけ出すことの重要性が増している[1].高次元データを次元圧縮する手法としては,主成分分析 (PCA) 等の線形な手法や ISOMAP等の非線形な手法が一般的に広く用いられているが,より汎用的な次元圧縮の枠組として,非線形かつ確率分布に基づく GPLVM(Gaussian Process Latent Variable Model) [5], [6] が提案されている.さらに,この手法を潜在変数のダイナミクスのモデリングへと拡張したGPDM(Gaussian Process Dynamical Models) [7] が注目を浴びている.

動作解析の分野では,モーションキャプチャ等のシステムにより獲得される高次元な 3D 情報 (各関節の 3D 回転角や位置,骨格データ等) から如何に本質的かつ汎用的な動作モデルを構築するかが重要な検討項目である.この目的で GPDM を用いる先行研究も多数あり,歩行やジョギング,ゴルフのスイング等の周期的動作に関して人物動作のモデリング [8] や単眼カメラ映像からの人物 3 次元動作推定 [9] にも有効であることが示されている.GPDM は高次元な 3 次元動作データ系列を潜在空間上でのなめらかな状態遷移をモデル化することが可能であり,時系列データである人物 3 次元動作推定の事前モデル構築に適している.また,同一動作に対する多様なスタイルや複数人物の 3 次元動作推定に適用可能である [10],[11] との報告もあり,GPDM は周期的な人物の 3 次元動作をモデル化する汎用的な手法であると言える.

さらに近年では,少数カメラ映像からの人物 3 次元動作推定への適用を目的として,より汎用的な動作モデルを構築するために GPDM の拡張手法が提案されている.特に,複数のダイナミクスをもつ対象をどのようにモデル化するかに関心が高まっており,複数の動作を一つのモデルに統合する手法 [12] や異なる動作毎の複数モデル間の切り替えを行う手法 [13], [14] 等がある.また,階層的潜在変数モデルである H-GPLVM [15] を人体構造の階層性に適用し,上半身と下半身で異なる動作の結合を可能としている例等がある [16], [17].

このように,少数カメラからの人物3次元動作推定のための事前動作モデル構築手法の研究は盛んに行われている.しかし,いずれの手法においても,全身の3次元動作データから動作モデルを構築するため,全身の観測が不安定となる場合には,3次元動作と画像からの観測データとの比較の際に曖昧性が増し,推定精度が低下する

そこで本研究では、観測データとして、2次元映像から人体の一部の情報のみが与えられた場合にも安定して全身の3次元動作推定を実現することを目指す.

3. 階層的学習

動的システムの基本となる状態空間モデル (式 (1),式 (2)) の模式図を図 (2) に示す.このモデルを推定問題

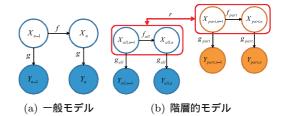


図 1 動的システムのグラフィカルモデル . X は状態 (潜在) 変数 . Y は出力データ

の事前モデルとする場合,観測データをIとして,Iと システムの候補出力データ Y を比較することで状態推 定を行う.このため,観測データIがノイズや観測条件 の変動等でシステムの出力 Y に比べてデータ次元が欠 落する場合,比較の曖昧さが増し,状態の推定誤差が大 きくなる. そこで, 事前のモデル化の際に, 推定時の観 測データの欠落を見越し,部分的なシステム出力に対応 する部分データから部分モデルを構築することを考える (式(3),(4)). これにより,部分出力データ Y_{part} と部分 観測データ I_{part} を比較することで,システムの部分状 態 X_{part} を推定できる.システム全体の状態を推定する ためには,式(5)のように全体の潜在変数Xと部分潜在 変数 X_{part} を関連付ける.これにより,精度よく部分的 観測データからシステム全体の状態を推定することが可 能となる.式(3),(4),(5)の模式図を図1(b)に示す.部分 データの分類や選択という課題は残るが,本論文ではま ずこの枠組の有効性を検証する.

$$\mathbf{x}_{part}(n) = f_{part}(\mathbf{x}_{part}(n-1)) + \eta_{x,part}(n)$$
 (3)

$$\mathbf{y}_{part}(n) = g_{part}(\mathbf{x}_{part}(n)) + \eta_{y,part}(n) \tag{4}$$

$$\mathbf{x}(n) = r(\mathbf{x}_{part}(n)) \tag{5}$$

以下,人物動作のモデル化に関して具体的な学習手法を述べる.3.1 節で従来手法である GPDM を用いた動作モデルの学習方法を説明し,3.2 節で動作モデル間の関連付けの方法を新たに提案する.そして,提案した階層的動作モデルを推定に適用する方法を3.3 節で示す.

3.1 Gaussian Process Dynamical Model に よる動作モデルの学習

GPDM は,D 次元の観測データ $Y=[\mathbf{y}(1),\mathbf{y}(2)$,…, $\mathbf{y}(N)]$ を,q(< D) 次元の潜在空間での位置 $X=[\mathbf{x}(1),\mathbf{x}(2)$,…, $\mathbf{x}(N)]$ からのガウス過程による回帰問題 $\mathbf{y}(n)=g(\mathbf{x}(n);B)+\eta_y(n)$ としてモデル化することに加え,潜在空間上のモデルのダイナミクスを $\mathbf{x}(n)=f(\mathbf{x}(n-1);A)+\eta_x(n)$ として考慮したモデルである.ここで,A,B はモデルパラメータであり, $\eta_x(n),\eta_y(n)$ はサンプルn に対するノイズである.一般に,ノイズは平均 0,共分散 $\alpha^{-1}I,\beta^{-1}I$ のガウス分布を事前分布として与える.学習のため,式 (6) で示す結合

確率の対数尤度(式 (7))を目的関数として潜在変数 X とハイパーパラメータを最適化することで,モデルを学習する.

$$p(\mathbf{Y}, \mathbf{X} \ \alpha \ \beta) = p(\mathbf{Y}|\mathbf{X} \ \beta)p(\mathbf{X}|\alpha)p(\alpha)p(\beta)$$
 (6)

$$L = \frac{q}{2} \log |\mathbf{K}_X| + \frac{1}{2} tr(\mathbf{K}_X^{-1} \mathbf{X}_{2:N} \mathbf{X}_{2:N}^T)$$
$$-N \log |\mathbf{W}| + \frac{D}{2} \log |\mathbf{K}_Y| + \frac{1}{2} tr(\mathbf{K}_Y^{-1} \mathbf{Y} \mathbf{W}^2 \mathbf{Y}^T)$$
$$+ \sum_{i} \log \alpha_i + \sum_{i} \log \beta_i (7)$$

ここで,N は学習データのサンプル数であり, \mathbf{W} は学習データの次元の違いをスケーリングする行列である.

また, $\mathbf{K}_X,\mathbf{K}_Y$ は,各要素が $(\mathbf{K}_X)_{i,j}=k_X(\mathbf{x}(i)$, $(\mathbf{x}(j)),(\mathbf{K}_Y)_{i,j}=k_Y(\mathbf{x}(i)$, $(\mathbf{x}(j))$)となる $\mathbf{N}\times\mathbf{N}$ のカーネル行列であり,本研究では,先行研究と同様,ダイナミクスに対して式 (8) の \mathbf{RBF} +線形カーネルを,潜在空間 \mathbf{X} から観測データ空間 \mathbf{Y} のマッピングに対して式 (9) の \mathbf{RBF} カーネルを用いる.

$$k_X(\mathbf{x}, \mathbf{x}')$$

$$= \alpha_1 \exp\{-\frac{\alpha_2}{2} \|\mathbf{x} - \mathbf{x}'\|^2\} + \alpha_3 \mathbf{x}^T \mathbf{x}' + \alpha_4^{-1} \delta_{\mathbf{x}, \mathbf{x}'} (8)$$

$$k_Y(\mathbf{x}, \mathbf{x}') = \beta_1 \exp\{-\frac{\beta_2}{2} ||\mathbf{x} - \mathbf{x}'||^2\} + \beta_3^{-1} \delta_{\mathbf{x}, \mathbf{x}'}$$
 (9)

本手法では,D 次元の全身学習モーションデータ $\mathbf{Y}_{all}=[\mathbf{y}_{all}(1)$, $\mathbf{y}_{all}(2)$,…, $\mathbf{y}_{all}(N)]$ を,GPDM により 3 次元の低次元潜在空間表現を全身動作潜在変数 $\mathbf{X}_{all}=[\mathbf{x}_{all}(1)$, $\mathbf{x}_{all}(2)$,…, $\mathbf{x}_{all}(N)]$ とし,K(<D) 次元の部分学習モーションデータ $\mathbf{Y}_{part}=[\mathbf{y}_{part}(1)$, $\mathbf{y}_{part}(2)$,…, $\mathbf{y}_{part}(N)]$ を,GPDM により 3 次元の低次元潜在空間表現を部分動作潜在変数 $\mathbf{X}_{part}=[\mathbf{x}_{part}(1)$, $\mathbf{x}_{part}(2)$, …, $\mathbf{x}_{part}(N)$] とした二つの低次元潜在空間表現を学習する.ここで,全身動作データと部分動作データは,モーションキャプチャデータより,任意の関節点を選択することで構成する.これらの動作モデル学習により,全身動作モデルと部分動作モデルとして,

- ullet 全身 $3\mathrm{D}$ 動作データ \mathbf{Y}_{all} の低次元潜在空間表現 \mathbf{X}_{all}
- ◆ 全身動作知識モデルから全身 3D データへの写像
- 全身動作知識モデル内でのダイナミクスを得る写像 f_{all}
 - ullet 部分動作データ \mathbf{Y}_{part} の低次元潜在空間表現 \mathbf{X}_{part}
 - 部分動作知識モデルから部分動作データへの写像

 g_{all}

ullet 部分動作知識モデル内でのダイナミクスを得る写像 f_{part}

を得る.

3.2 ガウス過程による複数の動作モデル間の関連 付け

次に,全身動作モデルと部分動作動作モデルを関連付ける.関連付けの目的は,部分潜在空間 \mathbf{X}_{part} 上での位置 \mathbf{x}_{part} が入力として与えられた場合に,全身潜在空間 \mathbf{X}_{all} 上での位置 \mathbf{x}_{all} を推定することである.この問題を,部分動作潜在空間 \mathbf{X}_{part} から全身動作潜在空間 \mathbf{X}_{all} へ回帰問題と考え定式化を行う.

人物動作が非線形性を持つことと,部分情報から全身情報への写像は多義性を持つことを考慮し,この問題を非線形・確率的な回帰を表現できるガウス過程により定式化する.

学習データを $\{\mathbf{x}_{part}(i)$, $\mathbf{x}_{all}(i)\}_{i=1}^N$ で与える.このとき, \mathbf{x}_{part} , \mathbf{x}_{all} は q 次元ベクトルである.ここで, $\mathbf{X}_{part}^{N\times q}=[\mathbf{x}_{part}(1)$,..., $\mathbf{x}_{part}(N)]$, $\mathbf{X}_{all}^{N\times q}=[\mathbf{x}_{all}(1)$,..., $\mathbf{x}_{all}(N)$] とすれば,ガウス過程の定義により,以下の式でモデル化できる [18].

$$p(\mathbf{X}_{all}|\mathbf{X}_{part}) = \prod_{k=1}^{q} N(\mathbf{X}_{all}(:,k)|0,\mathbf{C}_{N})$$
 (10)

$$(\mathbf{C}_{N})_{i,j} = C(\mathbf{x}_{part}(i), \mathbf{x}_{part}(j))$$

$$= k(\mathbf{x}_{part}(i), \mathbf{x}_{part}(j)) + \gamma^{-1}\delta(i,j) \quad (11)$$

ガウス過程は,カーネル関数 $k(\mathbf{x}, \mathbf{x}')$ とハイパーパラメータ γ だけで表すことができる. γ はノイズの精度パラメータである.カーネル関数は,扱うデータ特性を考慮して選択する必要がある.

ガウス過程回帰の学習では,対数尤度を表す以下の式 (12) を最大化するカーネル関数のパラメータとハイパーパラメータを求める.

$$L = \log p(\mathbf{X}_2|\theta)$$

= $-\frac{1}{2}\log |\mathbf{C}_N| - \frac{1}{2}\mathbf{X}_2^T\mathbf{C}_N^{-1}\mathbf{X}_2 - \frac{N}{2}\log (2\pi)(12)$

ここで, θ はガウス過程のモデルのハイパーパラメータである.これは非凸最適化問題であり,式 (13) で示されるハイパーパラメータについての勾配を用いる共役勾配法や準ニュートン法等の一般的な勾配法によって解くことができる.

$$\frac{\partial L}{\partial \theta_i} = -\frac{1}{2} Tr(\mathbf{C}_N^{-1} \frac{\partial \mathbf{C}_N}{\partial \theta_i}) + \frac{1}{2} \mathbf{X^T}_2 \mathbf{C}_N^{-1} \frac{\partial \mathbf{C}_N}{\partial \theta_i} \mathbf{C}_N^{-1} \mathbf{X}_2(13)$$

提案手法では,部分動作潜在変数 \mathbf{X}_{part} から全身動作潜在変数 \mathbf{X}_{all} へのガウス過程回帰を学習することで,

ullet 部分動作潜在変数から全身動作潜在変数への写像 $r: (\mathbf{X}_{part}
ightarrow \mathbf{X}_{all})$ を得る .

3.3 階層的動作モデルを事前知識とした 3 次元人 物姿勢推定

候補となる部分潜在変数 \mathbf{X}_{part} から写像 g_{part} により候補部分動作データ \mathbf{Y}_{part} を再構築する.この候補部分動作データ \mathbf{Y}_{part} を観測データ \mathbf{I} と比較し,最も尤度の高くなる候補を観測データに対する推定値 \mathbf{X}_{part}^* とする.さらに学習した \mathbf{GP} 回帰を用いることで全身潜在変数 \mathbf{X}_{all}^* が予測でき,写像 g_{all} により,3 次元全身動作データ \mathbf{Y}_{all}^* が予測できる

3.3.1 部分潜在変数の推定

2 次元映像から観測される人物動作の部分的な画像特徴量,たとえば関節位置を観測系列 $\mathbf{I}(1:t)\equiv(\mathbf{I}(1),\dots,\mathbf{I}(n))$ とする.学習した部分動作モデル $M\equiv(\mathbf{X}_{part},\mathbf{Y}_{part}\,\bar{\alpha}$, $\bar{\beta}$)を事前知識として,観測系列から状態 $\Phi(n)=[\mathbf{y}_{part}(n),\mathbf{x}_{part}(n),S(n)]$ を推定する.ここで,S(n) は $2\mathbf{D}$ 写像に影響するスケールパラメータである.状態推定は,式 (14) の事後確率を最大化する状態を求めることと等しく,これは $\mathrm{MAP}(\mathrm{Maximum}\ a\ posteriori)$ 推定やパーティクルフィルタ等のサンプリング手法によって実現される.

$$p(\Phi(n)|\mathbf{I}(1:n), M)$$

$$\propto p(\mathbf{I}(n)|\Phi(n))p(\Phi(n)|\mathbf{I}(1:n-1), M)$$
(14)

ここで,右辺の第1項は尤度(観測モデル)を表し,第 2項は予測分布(状態遷移モデル)を表している.

パーティクルフィルタを用いて式 (14) を解く. 状態 $\Phi(n)$ を推定する問題を , 事後確率の期待値を求める問題とする .

$$E[p(\Phi(n)|I(1:n),M)]$$

$$= \int \Phi(n)p(\Phi(n)|I(1:n))d\Phi(n)$$
(15)

観測モデルと状態遷移モデルをランダムサンプリング によりモンテカルロ近似し,期待値を式 (16)のサンプルの重み付き平均としてとして表す.

$$E[p(\Phi(n)|I(1:n),M)] \propto \sum_{i=1}^{N} \omega^{(i)}(n) \Phi^{(i)}(n)$$
 (16)

$$\Phi^{(i)}(n) \propto p(\Phi(n)|\Phi^{(i)}(n-1))$$
 (17)

$$\omega^{(i)}(n) \propto p(I(n)|\Phi^{(i)}(n)) \quad (18)$$

このとき,式 (17) が状態遷移モデル,式 (18) は観測モデルとなる.

状態遷移モデル

状態サンプル $\Phi^{(i)}(n)=[\mathbf{x}_{part}^{(i)}(n),\mathbf{y}_{part}^{(i)}(n),S^{(i)}(n)]$ を学習した部分動作知識モデルに従い生成する.

 GPDM の学習より,未知の潜在変数 $\mathbf{x}_{part}(n)$ と部分動作データ $\mathbf{y}_{part}(n)$ はそれぞれ下記の式で与えられる.

$$\mathbf{x}_{part}(n) = \mu_X(\mathbf{x}_{part}(n-1)) \tag{19}$$

$$\mathbf{y}_{part}(n) = \mu_Y(\mathbf{x}_{part}(n)) \tag{20}$$

このとき,

$$\mu_X(\mathbf{x}^*) = X_{2:N}^T K_X^{-1} \mathbf{k_X}(\mathbf{x}^*) \tag{21}$$

$$\mu_Y(\mathbf{x}^*) = Y^T K_V^{-1} \mathbf{k}_Y(\mathbf{x}^*) \tag{22}$$

ここで, $\mathbf{k_X}(\mathbf{x}^*)$, $\mathbf{k_Y}(\mathbf{x}^*)$ は i 番目の要素に $k_X(x^*,x_i)$, $k_Y(x^*,x_i)$ をもつベクトルである.

今回は,スケールパラメータ S(t) に関しては,観測データと人体構造を元にしたサンプル生成を行った. 観測モデル

尤度は下記で規定する.

$$p(I(n)|\Phi(n)) = exp(-\frac{1}{2\sigma^2} \sum_{j=1}^{J} \|\hat{I^j}(n) - y^j(n)\|^2)(23)$$

ここで, $\hat{I^j}(n)$ は,観測映像から $2{\rm D}$ ビデオトラッキング によって検出された j 番目の関節位置であり, $y^j(n)$ は 状態 $\Phi(n)$ の時の $2{\rm D}$ 観測のうち j 番目の関節を示す.

3.3.2 部分潜在空間から全身動作データの推定

ガウス過程により関連付けられている入出力空間において,入力データ $\mathbf{x}^*_{part}(n)$ が与えられたときに出力 $\mathbf{x}^*_{all}(n)$ の推定値は予測分布 $p(\mathbf{x}^*_{all}|\mathbf{X}_{all})$ の期待値として求める.

$$\mathbf{x}_{all}^* = \mathbf{C}^* \mathbf{C_N}^{-1} \mathbf{X}_{all} \tag{24}$$

ここで, \mathbf{C}^* は入力データ \mathbf{x}_{part}^* に対する学習データ \mathbf{x}_{part} についてのカーネル行列である.

$$\mathbf{C}^* = (C(\mathbf{x}_{part}^*, \mathbf{x}_{part}(1))...C(\mathbf{x}_{part}^*, \mathbf{x}_{part}(n))) (25)$$

これらの定式化により,部分潜在空間での位置 \mathbf{X}^*_{part} から全身潜在空間での位置 \mathbf{X}^*_{all} を予測することが可能

全身潜在空間での位置 \mathbf{X}^*_{all} が予測できれば, GPDM の学習により求めた写像 g_{all} を用いることで,全身動作知識モデルから全身の $\operatorname{3D}$ 動作データ \mathbf{Y}^*_{all} を再構築できる.

4. シミュレーションによるアルゴリズムの妥 当性の検証

3.2節で示した部分潜在空間から全身潜在空間へのガウス過程回帰の妥当性を検討する.部分潜在空間での変数 \mathbf{X}_{part} を既知として与えたときに,学習したガウス過程回帰によって全身潜在変数 \mathbf{X}_{all} を予測する.適切なカーネル関数を選択するために,3 種類のカーネル関数を用いて比較を行った.

4.1 ガウス過程回帰精度の検証

4.1.1 学習データ

学習に用いる動作データとして、CMU(Carnegie Mellon University) の公開モーションキャプチャライブラリから'walk' と'jog' を用いた.部分潜在空間の学習には人体構造から、下半身、上半身、右半身、左半身、末端部分(手足先と頭)等の複数種類の部分を選択して用いた.

4.1.2 カーネル関数

二つの潜在空間間のガウス過程を規定するカーネル関数は,学習の精度に大きく影響を与えるため,データ特性を考慮して適切に選択することが必要となる.本検討では,以下の3つを用いてカーネル関数の選択による推定結果の比較を行った.

- (1) 線形カーネル $k(\mathbf{x}, \mathbf{x}') = \theta_0 \mathbf{x}, \mathbf{x}'$
- (2) RBF カーネル $k(\mathbf{x}, \mathbf{x}') = \theta_0 exp(-\theta_1 || \mathbf{x} \mathbf{x}' ||^2)$
- (3) 線形 + RBF カーネル $k(\mathbf{x}, \mathbf{x}') = \theta_0 \mathbf{x}, \mathbf{x}' + \theta_1 exp(-\theta_2 ||\mathbf{x} \mathbf{x}'||^2)$

4.1.3 検討手順と結果

検討は,下記の手順に基づいて行った.ここで,部分は4.1.1に記載した5種類それぞれで検討を行った.どの部分に対しても同様の傾向の結果を得たため,本論文では,末端部分(手足先と頭)の結果を記載する.

動作モデル学習 一つのモーションキャプチャデータから全身動作モデルと部分動作モデルの二つをそれぞれ GPDM により学習する.'walk 'のデータを GPDM により学習した潜在空間を図2に示す.

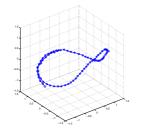
動作モデルの関連付け 部分潜在空間 X_{part} から全身潜在空間 X_{all} への回帰を 3 種類のカーネルによって規定したガウス過程によって学習する. 具体的には,式 (12) を共役勾配法を用いて最適化し,カーネル関数のハイパーパラメータを求める.

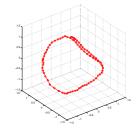
部分潜在変数から全身動作データの推定 シミュレーションデータとして式 (19) を用いて,学習データより部分潜在変数 \mathbf{X}^*_{part} を 30 フレーム分生成する.学習したガウス過程回帰 r により全身潜在変数 \mathbf{X}^*_{all} を予測する.カーネルの違いによる予測全身潜在変数の軌道を図 3 の $(\mathbf{b}) \sim (\mathbf{d})$ に示す.式 (20) により推定した全身潜在変数 \mathbf{X}^*_{all} から全身に関する 3 次元動作 \mathbf{Y}^*_{all} を求める.

推定誤差の算出 Y_{part}^* と推定した Y_{all}^* うち, Y_1^* と重複する関節の値が近い値をとっているかを RMS(Root Mean Square) 誤差により評価する.この評価尺度は,関節角度の絶対誤差を測定するものであり,今回の検討では評価値として関節各の次元での平均値を時間平均したものを用いた.各カーネル関数による RMS 誤差を図 3 に記載する.

4.1.4 考 察

図 2 と図 3 を比較すると , 図 2 の (a) と図 3 の (c) の 軌道が似ていることが見て取れる . これは , 部分潜在空間から全身潜在空間へのガウス過程回帰が , RBF カーネルによって適切に学習されたことを示している . 線形のカーネルでは , 部分から全身への回帰がうまくモデル化できていない . 'jog 'のデータを用いた時にも同様の傾向が見られた . このことを定量的に示すため , 誤差を見ると , RBF カーネルを用いた場合には , 時間平均で 5 。以下の誤差であるのに対し , 線形カーネルを用いた場合には , 20 。以上の誤差となっている . この結果によると , 今回検討した walk や jog のような単純かつ周期的な動

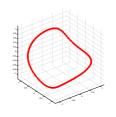




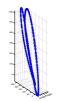
(a) 学習した全身動作の低次元 表現

(b) 学習した部分動作の低次元 表現

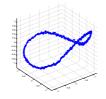
図 2 CMU モーションデータ「歩く」を GPDM によって学 習した低次元潜在空間



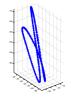
(a) シミュレーションデー タ



(b) 線形カーネル GP 回帰で予測した全身動作の低次元表現.3次元動作データの推定誤差:38.0°(フレーム・関節自由度平均)



(c) RBF カーネル GP 回 帰で予測した全身動作の 低次元表現.3次元動作 データの推定誤差:1.0° (フレーム・関節自由度 平均)



(d) 線形+ RBF カーネル GP 回帰で予測した全身動作の低次元表現 .3次元動作データの推定誤差:39.2°(フレーム・関節自由度平均)

図 3 シミュレーションデータ (部分潜在空間)から GP 回帰 により予測した全身潜在空間

作に関しては、部分潜在空間と全身潜在空間の間の回帰を規定するには、RBF カーネルが適当であることがわかる.ガウス過程を規定するカーネル関数は、一般的に、入力空間 x での相関関係が出力空間 y での相関関係に対応するように扱うデータ特性を考慮して選択する.今回用いた二つの潜在空間は同一のダイナミクス(歩くなど)により規定されているため、入力である部分潜在空間での距離が近いほど出力である全身潜在空間でも類似度が高くなる.よって、入出力間で局所距離を保存する特性をもつ RBF 関数が適当であったと考えられる.

4.2 3次元人物姿勢推定の事前動作モデルとして の有効性の検証

前節で部分動作モデルから全身動作モデルへの回帰が、 RBF カーネルで規定した GP 回帰で精度よく行えるこ とを確認した.次に,この階層的動作モデルを 2D 映像 から人物の一部分の動作情報しか得られない場合の、全 身 3D 動作推定に適用した際の効果を検証する.

4.2.1 検討と結果

動作モデル学習 学習データは、CMU モーションライ ブラリの'walk' を用いた.従来手法(GPDM[10])では, 全身の 3D 動作データを用いて全身動作モデルを構築す る.これを従来動作モデルとする.提案手法では,従来 と同様に全身動作モデルを構築することに加え,部分と して手足の先と頭の5点を選択し,それを仮想視点に 2D 投影して生成した腰の位置を原点とする部分 2D 位 置データから部分動作モデルを構築する.

動作モデルの関連付け 部分動作モデルから全身動作モ デルへの回帰を学習する.これを単独の動作モデルと合 わせて階層的動作モデルとする.

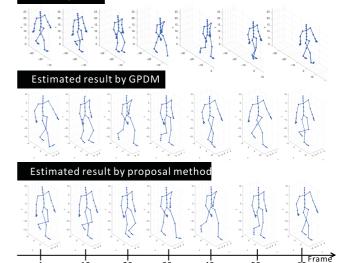
シミュレーションデータ生成 2D ビデオトラッキング データを模擬したシミュレーションデータを用いて検証 を行った. 学習時の 3D 動作データから区間切り出しを 行い,部分として,手足の先と頭の5点を選択した.そ れを 2D 投影して生成した 2D 位置データ (腰の位置を原 点とする相対位置)をシミュレーションデータに用いた. 3次元人物姿勢推定 従来動作モデルと階層的動作モデ ルをそれぞれ事前知識とし、3.3節のパーティクルフィル タを用いて推定を行った.推定結果を図4に示す.また, テストデータ生成に使用した正解の全身 3D 動作データ との RMS 誤差 (時間・関節平均)を図5に示す. 状態 サンプル数 N=100 として 5 試行行った際の平均であ る.図4をみると従来手法であるGPDMと比較して, 提案手法では,おおよそ正解の3Dモーションに近い姿 勢が推定できていることが確認できる.また,図5を見 ると,提案手法では従来手法と比較してRMS 誤差が小 さくなっており, 定量的にも提案手法の有効性が確認で きた.

5. 実映像への適用

ここまで,シミュレーションデータで階層的動作モデ ルの有効性を確認した.次に,階層的動作モデルを実映 像からの 3D 動作推定へ適用することを考える.

人物の 3D 動作を記述する際,一般的にモデルの汎用 性を高めるために,空間中に基準点を置くいわゆる世界 座標系ではなく,人体の一部分,たとえば腰の位置(ルー ト)を基準とした相対的な変化量を動作データとして用 いる.しかし実際の映像では,カメラは実空間に置かれ, 対象人物の動作により,このカメラと人体にある基準点 との相対位置は常に変動する.このため,実映像を観測





上段: 学習データ(正解データ). 中段: GPDM によ リ部分的観測データから推定した全身 3D 動作,下段: 提案手法により部分的観測データから推定した全身 3D 動作.

RMS error [degree]

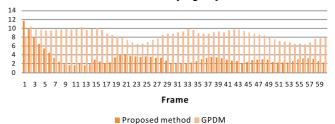


図 5 部分的観測データから推定した全身 3D 動作の RMS 誤

差(5回の平均). GPDM と提案手法による結果の比 較.提案手法では誤差が小さくなっている.

データとして3次元動作推定を行う際には,この相対位 置関係(視点)の変動を考慮する必要がある.

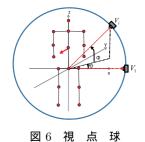
本研究では,前述の階層的学習手法を視点変動の推定 にも適用した. 具体的には, 学習時に視点毎に階層的動 作モデルを構築し,推定時に視点を状態として扱うこと で,この位置関係の変動を考慮する.

5.1 視点球を導入した視点毎階層的動作モデルの 学習

学習時にルート (人体の腰の位置)を基準とした視点 球(図6)を考え,球の表面を一定の間隔でサンプリン グする.サンプリングした各点に仮想的なカメラを置き, 各時刻の人体の関節位置を 2D 投影する. i 番目のサンプ リング点を視点 V_i とし, 各視点に関してそれぞれ階層 的学習を行った.つまり,学習によって視点 V_i に関する 階層的動作モデルを得る(図7).

5.2 視点変化を考慮した状態推定

全ての視点に関する階層的動作モデルを全探索するこ



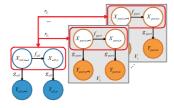


図 7 視点毎の階層的学習 .V は視点

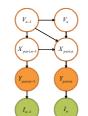


図 8 視点変化を考慮した 3D 姿勢推定のグラフィカルモデル

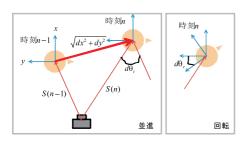


図 9 視点変化の定義 . ルート座標からの並進 t と回転 r から それぞれ視点の変化量 $d\theta_t, d\theta_r$ を求める . 全体としての 視点変化は $d\theta=d\theta_t-d\theta_r$ で定義する . ϕ に関しても 同様 .

とでも推定は可能だが,学習時の視点の粒度を細かくするほど探索範囲が膨大となってしまう.そこで,3.3.1 において,図 8 のように状態に視点 V を加えて $\Phi(n)=[V(n),\mathbf{y}_{part}(n),\mathbf{x}_{part}(n),S(n)]$ として状態推定を行う.このとき,視点 $V(n)=[\theta(n),\phi(n)]$ とする. ここで, $\theta(n),\phi(n)$ は図 6 の視点球における定義と等しい.時刻 n の視点候補を $\theta(n)=\theta(n-1)+d\theta,\phi(n)=\phi(n-1)+d\phi$ を平均とした正規分布に基づきサンプリングする.

本手法では,視点変化 $d\theta(n)$, $d\phi(n)$ は,図 9 のように学習データである $3\mathrm{D}$ 動作データの動きベクトル(モーションキャプチャデータのルートの並進と回転の変化量)を用いて定義した.これは,前の時刻のルートに対する並進と回転は動作の特徴を保持するとの考えに基づく.前の時刻の推定全身 $3\mathrm{D}$ 姿勢 $Y^*_{all}(n-1)$ と最も近い学習データ Y_{all} の並進 t と回転 t からそれぞれ視点の変化量 $d\theta_t$, $d\theta_t$ を求める.全体としての視点変化は $d\theta = d\theta_t - d\theta_t$ で定義する. ϕ に関しても同様に求められる.

5.3 実 験

視点変化を考慮した階層的動作モデルを事前知識として,視点変動が起こる映像に対して,部分的観測データ

から全身 3D 動作を推定可能かの検証を行った.

5.3.1 学 習

学習動作データは,マルチカメラのマーカレスモーションキャプチャ [19] によって獲得した 3D 歩行動作データ $(30\mathrm{fps})$ を用いた.部分として,手足の先と頭を選択し,視点球のサンプル点に投影し,視点 V_i の部分動作モデルとして学習する.さらに,全身動作モデルと視点 V_i の部分動作モデルを RBF カーネルの GP 回帰で関連付けることで,視点 V_i の階層的動作モデルを構築する.サンプリングの間隔は 5 °とし,方位角 θ は 0 ° ~ 355 °,仰角 ϕ は 90 ° ~ 90 ° とした.

5.3.2 推 定

検討1:視点変化のない場合

まず,事前検討として,今回の実験に用いた階層的動作モデルを事前知識として 4.2 節と同様に視点変化のない場合の推定精度をシミュレーションで確認した.シミュレーションデータは、学習データ (5.3.1 節)を $\theta=\pi,\phi=0$ に 2D 投影して生成した.推定方法は 4.2 と同様である.この時の RMS 誤差は,10 試行の平均で 9.9 (関節自由度,時間の平均)であった.この誤差は,4.2 節の図 5 と比較して大きくなっているが,これは,用いた学習データのフレームレートが 30fps と少なかったことと,マーカレスモーションキャプチャ [19] による学習データの 3D 姿勢推定自体の結果精度が低かったことが原因であると考えられる.

検討 2: 視点変化のある場合

学習データの生成時に使用したマルチカメラから一台を選ぶ、その映像に対して、2Dトラッキングで得た部分的な観測データを想定し、手足の先と頭、腰の位置にマーキングをする(図 10 の赤丸)、腰の位置に対する手足の先と頭の位置を観測データ I とする、初期視点 V(0) を大まかに与え、5.2 節に示した方法で全身の 3D 動作推定を行う、あるカメラについての推定結果を図 10 に示す、これによれば、部分的な観測データから全身の 3D 動作が推定できていることが視覚的に確認できる、また、この時の RMS 誤差は 5 試行の平均で 8.9 °であり、検討 1 の視点変化のない場合の結果と同等である、このことから、視点変化のアルゴリズムが妥当である見込みを得た、

次に,様々な視点変化のパターンを想定し,複数位置のカメラ映像を用いて同様に推定を行った.結果を図 11に示す.これにより,ある程度カメラ位置や視点変化のパターンに依存せず映像から 3D 動作推定が可能であるとの見込みを得た.

6. おわりに

本研究では,動的システムの状態推定において,観測条件等により観測データが欠落・変動する際にも安定した状態推定を行うことを目指している.この目的で,部分モデルから全体モデルへの回帰をもつ階層的モデルの

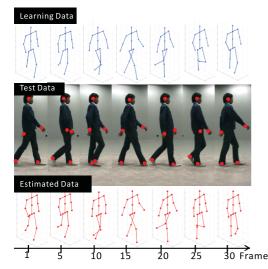


図 10 提案手法を実映像に適用した際の推定結果.上段:学習 3D 動作データ(正解データ),中段:観測映像と観測データとして与えた位置(丸印),下段:推定 3D 動作データ.

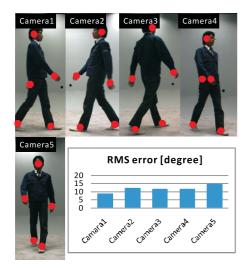


図 11 提案手法を実映像に適用した際の推定結果(複数視点 映像の結果,時間平均).RMS 誤差は5回推定を行っ た際の平均値.

構築手法を提案した.この階層的モデルを人物の3D動作推定問題に適用し,人体の一部分の情報しか観測されない条件下でも,一般的な動作モデルを用いた場合と比較して全身の3D動作を安定して推定可能であることを確認した.さらに,このモデルを実映像からの3D動作推定へ適用するために,視点変動を考慮したモデルへ拡張した.実際に実映像を用いた実験を行い,視点変動が起こるカメラ映像に対してモデルの有効性を確認した.今回の検討では,周期的な動作という単純なシステムを扱い基本枠組みの確認を行った.さらに,部分データの分類や選択方法や,推定時の部分モデル間の切り替え等の課題を検討する必要がある.

文 献

[1] 矢入健久. 次元削減と動的システムの学習. 第1回

- Latent Dynamics Workshop 予稿集, 2010.
- [2] Thomas B. Moeslund, Adrian Hilton, and Volker Krüger. A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.*, Vol. 104, No. 2-3, pp. 90–126, 2006.
- [3] Ronald Poppe. Vision-based human motion analysis: An overview. Comput. Vis. Image Underst., Vol. 108, No. 1-2, pp. 4–18, 2007.
- [4] 伸敬島田, 大作有田, 徹玉木. 関節物体のモデルフィッティング (サーベイ (2)). 情報処理学会研究報告. CVIM, Vol. 2006, No. 51, pp. 375-392, 2006.
- [5] Neil Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. NIPS, 2003.
- [6] Neil Lawrence and Aapo Hyvarinen. Probabilistic non-linear principal component analysis with gaussian process latent variable models. *Journal of Machine Learning Research*, Vol. 6, pp. 1783–1816, 2005.
- [7] Jack M. Wang, David J. Fleet, and Aaron Hertzmann. Gaussian process dynamical models. NIPS, pp. 1441– 1448, 2006
- [8] JM Wang, DJ Fleet, and A. Hertzmann. Gaussian process dynamical models for human motion. *IEEE* transactions on pattern analysis and machine intelligence, Vol. 30, No. 2, pp. 283–298, 2008.
- [9] Feng Guo and Gang Qian. 3d human motion tracking using manifold learning. Proceedings 2007 IEEE International Conference on Image Processing, pp. I– 357–60, 2007.
- [10] D.J. Fleet R. Urtasun and P.Fua. 3d people tracking with gaussian process dynamical model. CVPR, 2006.
- [11] D.J. Fleet R. Urtasun and P.Fua. Temporal motion models for monocular and multiview 3d human body tracking. CVIU, 2006.
- [12] A. Geiger, R. Urtasun, and T. Darrell. Rank priors for continuous non-linear dimensionality reduction. Computer Vision and Pattern Recognition, pp. 880–887, 2009.
- [13] Jixu Chen, Minyoung Kim, Yu Wang, and Qiang Ji. Switching gaussian process dynamic models for simultaneous composite motion tracking and recognition. pp. 2655–2662, Los Alamitos, CA, USA, 2009. IEEE Computer Society.
- [14] Jack M. Wang, David J. Fleet, and Aaron Hertzmann. Multifactor gaussian process models for style-content separation. In *Proceedings of the 24th international* conference on Machine learning, pp. 975–982, New York, NY, USA, 2007. ACM.
- [15] Neil D. Lawrence and Andrew J. Moore. Hierarchical gaussian process latent variable models. NIPS, pp. 481–488, 2007.
- [16] John Darby, Baihua Li, Nicholas Costen, David Fleet, and Neil Lawrence. Backing off: Hierarchical decomposition of activity for 3d novel pose recovery. British Machine Vision Conf., 2009.
- [17] Mykhaylo Andriluka, Stefan Roth, and Bernt Schiele. People-tracking-by-detection and people-detectionby-tracking. In *IEEE Conference on Computer Vision* and Pattern Recognition (CVPR'08), 2008.
- [18] Gaussian Processes for Machine Learning. MIT Press, 2006.
- [19] H.Deng Y.Liu X.Wu K.Wakabayashi S.Shen, M.Tong and H.Koike. Model based human motion tracking using probability evolutionary algorithm. *Pattern Recognition Letters*, Vol. Volume 29, No. Issue 13, pp. Pages 1877–1886, 1 October 2008,.