

## 時系列ボリュームの非線形確率モデルを用いた 人体・着衣の部位形状追跡

平井 迪郎, 浮田 宗伯, 木戸出 正継<sup>†1</sup>

同期ビデオ群から着物などの緩い着衣を含めた人体の3次元部位形状を追跡する手法を提案する。本稿における部位形状追跡とは、各フレームにおける各体節部位の形状復元と3次元復元誤差の修正を表す。本手法では、対象となる人体・着衣の時系列形状の高精度サンプルを用意し、これを事前知識とすることにより複雑な形状変化の復元と修正を頑健に実現する。サンプルの学習には非線形・確率的な次元削減手法を適用することで、学習サンプルのノイズや入力と学習サンプルの細かな変化を許容し、複雑・急激な形状変化の追跡が可能となる。

形状変化の大きい着物を着衣とした実験により、提案手法の有効性を確認した。

### 3D Body-part Tracking of a Human in Clothing using Probabilistic Non-linear Time-series Volume Learning

MICHIRO HIRAI, NORIMICHI UKITA, AND MASATSUGU KIDODE<sup>†1</sup>

We propose a method for tracking 3D human body parts and clothing from synchronized video sequences. Our objective is (1) to identify each body region with its corresponding clothing in the entire body volume and (2) to refine the volume. Time-series sample volumes are acquired by a sophisticated 3D reconstruction algorithm off-line. The history of these samples are stored as low dimensional trajectories with probabilistic nonlinear representation. With this representation, body-part tracking becomes robust to various noises and delicate, complicated, and quick shape variations. Experimental results using loose-fitting clothing demonstrated the effectiveness of our method.

#### 1. はじめに

近年、人体の姿勢・運動解析に関する数多くの研究が行われている。これらの研究は、ヒューマンコンピュータインタラクション(HCI)やジェスチャ認識、ロボットの運動制御、CGアニメーション、能動カメラ観測システムなど、さまざまな応用が期待される。特に、計算機の進歩によって3次元形状の復元が実時間で可能となり<sup>3),4)</sup>、3次元形状に基づく姿勢・運動解析が注目されている。3次元形状解析には、計算量の増加というデメリット以上に、従来の2次元画像ベースの手法における自己遮蔽や観測結果と解が一体一に対応しないことによる曖昧性の解消が期待される。また、対象の詳細な3次元形状も得られるため、3次元ビデオやCGアニメーションの生成・編集などの応用範囲の拡大も期待される。

しかし、関節数の多さからも分かるように、人体の構造は複雑であるために、3次元物体の姿勢推定や運

動解析は困難な問題といえる。その上、緩やかな着衣を含んだ形状の推定となると、対象の形状変化はさらに複雑なものとなる。これまで数多くの研究が行なわれている剛体運動の解析と比較して、こうした非剛体形状の変化を追跡している文献はまだ少ない。一般的な人体の形状・姿勢推定では、人体の各体節を剛体近似することによって姿勢推定を行なった後に表面形状の詳細な解析<sup>5),6)</sup>を行なっている。しかし、着物のように完全な非剛体着衣に対して剛体モデルのフィッティングを行なうことは不可能である。

そこで我々は、不規則かつ複雑に変化する非剛体着衣の着用者に対しても適用可能な体節領域の追跡および形状復元を行なう手法を提案する。本研究において目的とする体節領域追跡と形状復元を以下に定義する：**体節領域追跡** 関節で区切られて剛体近似可能な人体領域に「その人体領域の動きに伴い形状変化する周辺着衣部分(例:腕における袖や脚における裾)」を加えた領域を体節と呼び、本研究における追跡対象とみなす。

**形状復元** Shape-From-Silhouetteにより高速に獲得可能な対象の近似形状(visual hull)に含まれる大きな復元誤りを修正し、より正確な3次元形状

<sup>†1</sup> 奈良先端科学技術大学院大学情報科学研究科  
Graduate School of Information Science, Nara Institute  
of Science and Technology

を復元する。

上記の体節領域は、従来手法<sup>1),2)</sup>が目的とする人体の位置や関節角そのものを示した姿勢推定結果ではない。しかし、この結果を制約とすることにより（各関節は我々の手法で得られた体節領域間の境界に存在する、など）、従来法の頑健性向上や着物を身につけた対象の姿勢推定への拡張も可能となる。さらに、各体節の形状復元結果から、そのパラメータ（重心座標・体積・面積・シルエットなど）を学習・入力データとして実行可能なジェスチャ認識<sup>12)</sup>やCG編集<sup>13)</sup>への応用も考えられる。

## 2. 関連研究

### ● 体節領域追跡

3次元姿勢・運動解析において、実際の観測対象の動きをそのまま学習しておく事例ベースの手法が数多く提案されている（関節可動角<sup>15)</sup>や姿勢変化の確率的表現<sup>16),17)</sup>など）。これらの研究ではモーションキャプチャを用いて対象を観測しているためパラメトリックな表現では困難な人体の複雑かつ微妙な変化を表現することが出来る。さらに、3次元CGモデルを用いる手法<sup>18)</sup>と比較して、リアルな表現が可能である。

しかしながら、着物のような緩い着衣を対象とした場合にはマーカの位置ずれによって体節の境界部分の再現が困難なこと。さらに、着物全体の表面形状変化を表現できないことが問題となる。加えて、着衣の形状や動きをシミュレートしている文献（例えば、CG<sup>10)</sup>によるモデルパラメータの推定や、非剛体物体の追跡<sup>11)</sup>）も存在する。これらの手法では人体の動きに関する事前知識を用いて推定を行っている。従来法<sup>29)</sup>は、これらのことを踏まえ人体やそれに伴う着衣の変化をそのまま学習サンプルとして蓄え、事例ベースの事前知識として利用することでこれらの問題を解決している。また、効率的に学習したサンプルから類似形状を見つけるために、主成分分析を用いて低次元な固有空間上で類似サンプルの探索を行っている。

同様のアプローチとして文献、19) 24)ではシルエットから低次元の特徴を抽出し、時系列的な人間の動きが低次元の多様体として獲得できることを示している。これは3次元形状であってもその特徴的な動きや形状の変化が低次元の特徴として獲得できることを示唆している。実際に従来法<sup>29)</sup>では、ボクセルデータを主成分分析することによって、低次元な特徴として固有空間での軌跡を獲得できている。

### ● 形状復元

一方で、Shape-From-Silhouetteによって得られたvisual hullは真の形状を内包する凸形状であり、原理的に凹形状を含むような形状復元が難しい。加えて、撮影対象とカメラ間の位置関係によっても形状復元誤差が発生する。これらの形状復元誤差は一般に偽ボ

リュームと呼ばれる。また、カメラキャリブレーションやシルエット抽出など、体積復元以前の前処理においても誤差の影響は避けられない。これらの形状復元誤差を修正する方法としては異なる視点間で対応画素の色の同一性をチェックするspace carving<sup>7)</sup>や、それに加えてシルエット制約や時間的な形状変化の滑らかさを利用して表面形状の動的変形を行う弾性メッシュモデル<sup>8)</sup>などが提案されている。

また緩い着衣の復元に関しては表面にカラーコードパターンと呼ばれるマーカをプリントした布を用いて凹形状の復元を行う研究<sup>9)</sup>も存在する。さらにコンピュータビジョン以外の分野でも3次元形状の高精細化に関して数多くの研究<sup>26)</sup>が存在する。しかしながら、いずれの手法においても計算機コストの問題からvisual hullにオンラインで適用することが困難である。そこで、本稿では学習サンプルの形状情報を利用して復元された形状から誤差の修正、さらに凹形状の復元を目的とする。

## 3. 処理概要

人体の複雑な動きやそれに伴う着物の動きは非常に複雑で非線形であることが予測される。さらに、同様の動きであっても手先の位置や袖の変化などは確率的な分散を持っていると考えられる。つまり、線形な主成分分析の結果に基づく従来手法<sup>29)</sup>では、形状変化が複雑である場合には予測誤りが生じることがあった。これらの問題に対して、時系列ボリュームの学習に非線形・確率的な次元削減手法であるGaussian Process Latent Variable Model<sup>20)</sup>(GPLVM)を用いることを提案する。GPLVMを用いることにより、非線形な写像が可能となり、複雑な形状変化にも対応可能な形状解析ができる。

さらに従来手法<sup>29)</sup>においては、各時刻において復元された形状に対して独立に形状解析を行っていた。しかしながら、ある時刻における復元結果に大きな復元誤りが発生した場合にはその影響を受けて形状解析が失敗することがある。そこで本稿ではこれまでの手法が**形状検出**であったのに対して、時系列情報を用いた**形状追跡**へと拡張する。つまり、過去の形状復元結果を考慮した形状修正が可能となり、前述のような一時的な大きな復元誤りに対しても頑健な形状修正が可能となる。また、GPLVMを用いることでボリュームの変化に対して確率的なモデリングが可能となり、細やかな形状の変化に対しても頑健な追跡が可能となる。

ここから、本稿における体節領域追跡と形状復元の全体の流れを述べていく。

本手法では入力形状の各ボクセルに対して10種類の体節（頭、胴体、右上腕、右前腕、左上腕、左前腕、右もも、右下肢、左もも、左下肢）を設定する。加えて、特殊なクラスとして**非対象**を用意する。これは入

力形状における復元誤りを表す。入力形状に対して非対象を含めた11種類の体節を割り当てることにより、体節領域追跡と復元誤差修正を同時に実現できる。

本システムは時系列ボリュームの学習と部位形状追跡の2つに分けることができる。

#### 時系列ボリュームの学習

まず、信頼できる学習サンプルの生成のために、SFSによって得られた visual hull に対して、弾性メッシュモデル<sup>8)</sup>を用いて高精細なボリュームデータを生成する。得られた高精細ボリュームに GPLVM を適用して低次元の特徴を獲得する。

#### 部位形状追跡

オンラインで撮影された同期画像から SFS を用いて形状復元を行い、復元された大まかな低解像度全身形状を特徴空間に投影する。特徴空間上で近傍に存在する学習サンプルを用いてボリュームデータをバウンディングボックスに分割する。各バウンディングボックスに含まれる詳細な体節部位の復元形状を体節ごとの特徴空間へと写像し、パーティクルフィルタを用いて形状追跡を行う。最後に近傍の学習サンプルを用いてラベル割当てとボリューム修正を行う。

以降では、4章で GPLVM を用いて時系列ボリュームを低次元な事前知識として獲得する方法について、さらに5章ではオンラインで部位形状を追跡する方法について詳しく述べる。

### 4. 時系列形状の非線形確率モデル学習

一般に人体の動きや、それに伴う非剛体着衣の形状変化は非線形である。つまり、従来法<sup>29)</sup>の様な線形モデルでは入力形状のノイズに影響されてしまうこと、着物の複雑な動きに対応できないことが問題となる。このような非剛体着衣に対しても頑健なモデルを獲得する必要がある。そこで、非線形・確率的な次元削減手法として知られる Gaussian Process Latent Variable Models<sup>20)</sup> (GPLVM) を適用する。

本章では、信頼できる学習サンプルの生成法と GPLVM を用いて時系列的な形状の変化を事前知識として表現する方法について述べる。

#### 4.1 体節ラベル付き高精度形状の生成

信頼できる学習サンプルの生成と、それに付随した体節の特定を行うため、以下の処理を行う。

- step.1 Shape-From-Shilhouette により (図 1-(a)) の様な体節ごとに着色された着物を着た対象の visual hull を得る。
- step.2 SFS による復元誤差補正のため、弾性メッシュモデルを visual hull に適用し、高精細形状 (図 1-(b)) を得る。
- step.3 色検出<sup>27)</sup>を用いて多視点体節ラベル化画像 (図 1-(c)) を生成する。
- step.4 多視点ラベル画像から step.2 で得られた形

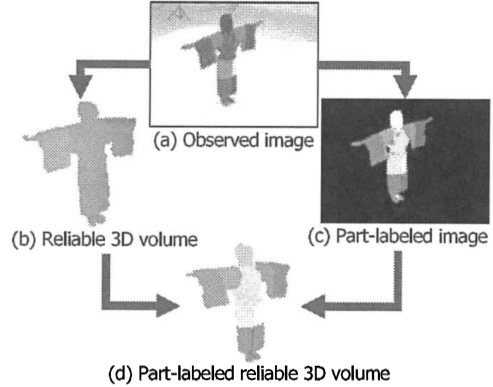


図 1 体節ラベル付き高精度ボリュームデータの生成処理

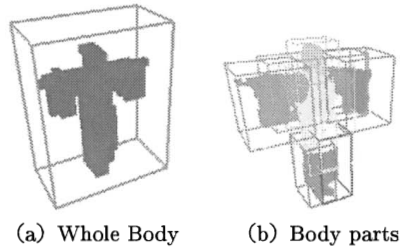


図 2 バウンディングボックス

状に対して体節ラベルの逆投影を行い、表面ボクセルの体節割り当てを行う。

- step.5 表面ボクセルからの最近傍探索によって内部ボクセルの体節割り当てを行う。

#### 4.2 学習の階層化

対象の体節領域追跡のために、階層的な学習を採用する。つまり、1) 低解像度の全身形状、2) 高解像度の体節部位形状、の2段階処理によって学習を行う。

また、形状をボリュームデータに変換して GPLVM を適用するための前処理として、バウンディングボックスと呼ばれる物体を囲む一定サイズの直方体を定める。バウンディングボックス内のボクセルデータを列ベクトルに変換し、データの次元数を一致させることで GPLVM を適用することが出来る。実際の処理では、まず各フレームにおけるボリュームデータの3次元重心を求め、それを基準に対象を囲むバウンディングボックスを設定する。本稿においてはバウンディングボックスのサイズは、学習サンプルにおける全フレームで対象のボリュームを囲うことのできるサイズとする。全身ボリュームおよび各体節のバウンディングボックスの例を図 2 に示す。しかしながら、ある体節のバウンディングボックス内には他の体節のボクセルデータも含まれてしまうため、人体の事前知識を使用して領域内に含まれる不必要な体節情報を除去する。ここで不必要な体節とは、例えば、頭領域に含まれる

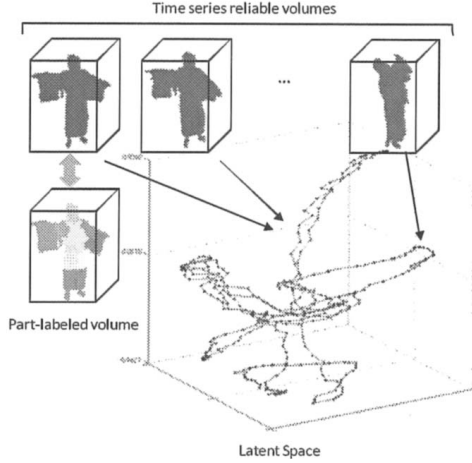


図3 時系列ボリュウムデータによる潜在空間

左上腕や、右前腕領域に含まれる左前腕などの構造上繋がっていない体節のことを指す。また、逆に胴領域に含まれる左前腕などの繋がっている体節領域については必要な情報として残す。この処理によって、後述する体節ラベル割り当ての際に悪い影響を及ぼす体節情報を除去することが出来ると考えられる。

また、階層処理による形状追跡については5.3で詳しく述べる。

### 4.3 GPLVM による時系列ボリュウムの事前知識表現

#### 4.3.1 Gaussian Process Latent Variable Model (GPLVM)

Gaussian Process Latent Variable Model (GPLVM) は入力データから低次元な潜在空間への滑らかな確率的写像を学習する効果的な次元削減手法として知られており、姿勢・運動解析の分野においても注目を浴びている。

本手法においては学習サンプルとしてボクセルデータ  $V = [v_1, \dots, v_N]^T$  が与えられた時、低次元の特徴量である  $X = [x_1, \dots, x_N]^T$  への写像、

$$x_i = g(v_i), \quad i \in (1, \dots, N) \quad (1)$$

を学習することで3次元形状の時系列的な変化をモデル化する。ここで、 $v_i = (v_{i,1}, \dots, v_{i,D})^T$  は各時刻の3次元形状をボクセルとしてを表したものであり、時刻  $i$  における  $n$  番目のボクセルデータ  $v_{i,n}$  ( $v_{i,n} \in 0, 1$ ) は非対象なら0、対象なら1である。また、 $x_i = (x_{i,1}, \dots, x_{i,d})^T$  は潜在変数と呼ばれる。

#### 学習

GPLVM の学習はモデルパラメータに関して、以下の  $\mathcal{L}$  を最小化することによって行う。

$$\mathcal{L} = \frac{d}{2} \ln |K| + \frac{1}{2} \text{tr}(K^{-1} V V^N) + \frac{1}{2} \sum_m^N \|x_m\|^2 + \text{const.} \quad (2)$$

ここで、 $K$  は学習サンプルによって計算される  $N \times N$  のカーネル行列を表し、 $K_{i,j} = k(x_i, x_j)$  である。負の対数尤度関数  $\mathcal{L}$  の最小化によって得られた、ボクセルデータ ( $D$  次元) から潜在空間 ( $d$  次元:  $d \ll D$ ) への写像は適切なカーネルを用いることで、非線形化することができる。動径基底関数 (RBF) を用いればカーネル関数は、

$$k(v_i, v_j) = \alpha_{rbf} \exp\left(-\frac{\gamma}{2} \|v_i - v_j\|^2\right) + \alpha_{bias} + \beta_{-1} \delta_{i,j} \quad (3)$$

と表せる。

$\alpha_{rbf}, \gamma, \alpha_{bias}, \beta$  はカーネルパラメータであり、 $\delta_{i,j}$  はクロネッカーのデルタ関数である。

#### テストデータの写像

今、新しいボクセルデータ  $v^*$  が与えられた時、

$$\mathcal{L} = \frac{\|v^* - \mu(x^*)\|^2}{2\sigma^2(x^*)} + \frac{M}{2} \ln \sigma^2(x^*) + \frac{1}{2} \|x^*\|^2 \quad (4)$$

であり、平均と分散は

$$\mu(x^*) = \mu + V_T K^{-1} k(x^*) \quad (5)$$

$$\sigma(x^*)^2 = k(x^*, x^*) - k(x^*)^T K^{-1} k(x^*) \quad (6)$$

で与えられる。

$x^*$  を求めるためには、 $\arg \min_{x^*} \mathcal{L}$  を満たす  $x^*$  を求めればよい。

図3に GPLVM によって得られた潜在空間を示す。時系列に従って滑らかな軌跡が得られていることが分かる。

#### 4.3.2 Back-constrained GPLVM

Back-constrained GPLVM<sup>21)</sup> (bc-GPLVM) はデータ空間において近いデータ同士が潜在空間においても近い距離に投影される様に GPLVM に対して逆制約を加えたもので、潜在空間において局所的な距離が保存される。GPLVM における式 (2) の  $X$  の代わりに以下の  $x_j = g_j(v_n)$  について最適化を行う。

また bc-GPLVM においては  $\arg \min_{x^*} \mathcal{L}$  を求める代わりに、以下の  $g(v^*)$  を求めることで、データ空間から潜在空間への明確な写像が定義できるといった特長も持っている。

$$x_j^* = g_j(v^*, a) = \sum_{n=1}^N a_{jn} k_{bc}(v^* - v_n) \quad (7)$$

ここで、 $A = \{\{\alpha\}_{n=1}^N\}_{j=1}^d$  は制約を与えるパラメータであり、 $x_j^*$  は  $x^*$  の  $j$  番目の値である。

RBF カーネルを用いる場合、

$$k_{bc}(v^* - v_n) = \exp\left(-\frac{\gamma_{bc}}{2} \|v^* - v_n\|^2\right) \quad (8)$$

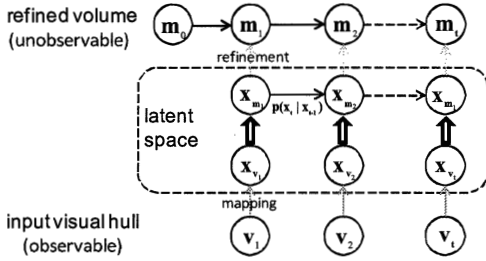


図 4

となる。

本手法では、bc-GPDM を用いることでボクセルデータの類似性を局所的に保存したまま、潜在空間へ写像することができる。つまり、潜在空間で局所的に近いデータがボクセル空間でも近いデータとなるので、以降の形状追跡に有効である。

#### 4.3.3 Gaussian Process Dynamical Models

Gaussian Process Dynamical Models<sup>23)</sup>(GPDM) は GPLVM に対して、潜在空間で  $x_{t-1}$  から  $x_t$  への遷移が滑らかなる様に制約を加えたもので、

$$v_t = \sum_i b_i \phi_i(x_t) + n_{v,t} \quad (9)$$

$$x_t = \sum_i a_i \psi_i(x_{t-1}) + n_{x,t} \quad (10)$$

である。

制約には RBF カーネルを用いて  $k_{dy}(v_t^* - v_n) = \alpha_{dy} \exp(-\frac{\gamma_{dy}}{2} \|v_t^* - v_n\|^2) + \beta_{dy}^{-1} \delta_{t,t-1}$  (11)

と表せる。GPDM は対象の状態遷移をモデル化できるので時系列的な形状の変化が潜在空間上で滑らかなるといった利点がある。この滑らかな軌跡は、複雑な形状変化のモデリングを効果的に行う事を可能にする。

### 5. 部位形状追跡：体節領域追跡と形状復元

#### 5.1 非線形確率モデルを用いた形状追跡

一般の追跡手法においては高次元データをそのまま扱うことは、次元の呪いによって局所解に陥るといった問題がある。パーティクルフィルタ<sup>14)</sup>を用いることでこれらの問題を改善してきたが、頑健な追跡には多数のパーティクルが必要となり、計算コストが高いことが問題となる。

我々の手法においては GPLVM によって獲得した事前知識を利用すること<sup>22),23),25)</sup>で高次元ボクセルデータを効率的に追跡する。

つまり、復元形状から真の形状を推定する問題を時刻  $t$  において復元された形状  $v_t$  を観測値、弾性メッシュ変形によって得られる高精細形状  $m_t$  をを真の形状とみなし、一般の追跡問題の様に観測値  $v_t$  から状

態  $m_t$  を求める問題に帰着させる。

状態空間モデルは

$$m_t = A m_{t-1} + n_1 \quad (12)$$

$$m_t = v_t + n_2 \quad (13)$$

となる、 $n$  は  $v_t$  の形状復元誤差となる。

対象の状態遷移  $P(m_t|m_{t-1})$  は近似的に潜在空間上での状態遷移  $P(x_{m_t}|x_{m_{t-1}})$  を用いることができるので式 (12) は、 $x_t = g(v_t)$  によって得られる潜在変数  $x_t$  を用いて、

$$g(v_t) = \hat{A}g(v_{t-1}) + \hat{n}_1 \quad (14)$$

として、形状の状態遷移を記述することができる。

本稿では、潜在空間における状態遷移  $P(x_{m_t}|x_{m_{t-1}})$  に対して、線形な自己回帰モデル (AR model) を当てはめ、式 (14) の状態遷移を記述した。

#### 尤度

尤度  $p(m_t|v_t)$  は観測値  $v_t$  が与えられた時に  $m_t$  がどれほど尤もらしいかを表し、単純には  $m_t$  と  $v_t$  の幾何学的な形状の近さを表す。ここで、4.3.2 において、bc-GPLVM ではデータ空間において近いデータ同士が潜在空間でも近い距離に保存されることを述べた。

そこで、潜在空間上でのデータ同士の距離  $d_t = \|g(v_t) - s_n\|^2$  が近ければ尤度が高くなるという仮定を置き、パーティクル  $s_n$  が与えられた時の尤度関数を近似的に以下の様に設定する。

$$p(g(v_t)|x_n = s_n) \propto 1 + \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{d_t^2}{2\sigma^2}) \quad (15)$$

$$d_t = \|g(v_t) - s_n\| \quad (16)$$

以上の処理によって高次元なボクセルデータの形状追跡を効率的に実現することが出来る。

#### 5.2 形状復元誤差への対応

形状復元誤差の影響に対して安定な処理を行なうため、対象の各ボクセルにおける復元誤差の発生し易さを確率的に表現したボクセル信頼度<sup>29)</sup>を設定する。ボクセル信頼度は「弾性メッシュモデルにより得られた高精細形状」と「SFSによって得られる復元形状」の比較によって推定することができる。

**step.1** 外部パラメータが既知である実カメラ群を用いて仮想的な空間を設定し、各カメラごとに仮想空間上で高精細形状から画像平面へシルエット投影を行う。

**step.2** 得られたシルエット画像群から SFS によって復元形状を得る。ここで、高精細形状と復元形状 (visual hull) の差分が SFS によって生じる形状復元誤差である。

**step.3** 高精細形状の位置・姿勢を変化させて step.1 - step.2 を行う。

**step.4** それぞれお復元形状と高精細形状の差の平均からボクセル信頼度を求める

時刻  $t$  の学習サンプルに対するボクセル信頼度は下

式で表現される。

$$R_t = \text{diag}(r_{t,1}, \dots, r_{t,d}) \quad (17)$$

ここで、 $r_{t,d} = \frac{1}{S} \sum_i^S \delta_{i,v}^d$  であり、

$S$  は仮想的に観測されたサンプルデータ数、 $\delta_{i,v}^d$  は  $t$  番目のサンプルにおいて  $d$  個目のボクセルデータが高精度形状と異なる値をとれば、1、そうでなければ、0 となる。

#### ボクセル信頼度を利用した潜在空間への投影

得られた復元形状  $v$  に対して  $R$  との積をとることで、復元誤差を低減した形状  $\hat{v} = R_t v_t$  を得る。よって学習空間への射影式 (7) には  $v$  の代わりに  $\hat{v}$  を用いる。

これにより、偽のボリュームデータの影響を低減させた写像を行うことができる。しかし、実際にはボクセル信頼度  $R_t$  は、復元形状  $v_t$  が学習サンプル中のどのデータと対応づいているかがわからないと与えることができない。そこで、時系列的に連続な入力データ間の差は微小であると仮定し、時刻  $t$  におけるボクセル信頼度には  $R_{t-1}$  を用いる。

### 5.3 階層処理による詳細部位形状追跡

本節では、実際の追跡処理概要を順に紹介する。大まかな流れとしては、まず、低解像度全身追跡によって、対象の方位、体節領域の推定を行う。次に、高解像度部位追跡によって、部位形状ごとの体節ラベル割り当て、復元誤差修正を行う。

直感的には、低解像度追跡によって対象の大まかな位置と姿勢を推定し、高解像度追跡によって対象の細やかな形状変化を推定することに相当する。

時刻  $t$  において SFS で計算される復元形状  $v_t^*$  とそれに対応するボクセル信頼度  $R_t$  から、 $\hat{v}_t^* = R v_t^*$  を得る。この  $\hat{v}_t^*$  を観測データと考え、GPLVM によって得られた潜在空間への写像を行う。

$$\hat{x}_t^* = g(\hat{v}_t^*) \quad (18)$$

さらに、時刻  $t-1$  において推定された高精度形状  $m_{t-1}$  から得た  $x_{t-1}$  と潜在空間周辺でのダイナミクスから次状態  $x_t$  を推定する。

最後に推定された次状態  $x_t$  の  $k$  近傍学習サンプル  $v_k^{sam}$  からラベルの推定を行う。

#### 5.3.1 低解像度全身追跡

**step.1** 低解像度復元形状  $v_t^{low}$  の 3 次元重心を原点とし、 $z$  軸周りに  $\theta$  回転。

**step.2** 復元形状  $v_{\theta,t}^{low}$  とボクセル信頼度  $R_{\theta,t}^{low}$  から  $\hat{v}_{\theta,t}^{low}$  を計算。

**step.3** 各体節形状  $\hat{v}_{\theta,t}^{low}$  を潜在空間に投影し、 $x_{v_t}^{low}$  を得る。

**step.4** パーティクルフィルタによって  $x_{v_t}^{low}$  から  $x_{m_t}^{low}$  を推定する。

**step.5**  $\theta$  を変化させて  $\min \|x_{sam} - x_{m_t}^{low}\|$  となる学習サンプル  $v_n^{sam}$  から、対象方位  $\hat{\theta}$  とバウンディングボックスを決定。

#### 5.3.2 高解像度部位追跡

**step.1** 高解像度復元形状を  $\hat{\theta}$  回転させ、バウンディングボックスに分割。

**step.2** それぞれの体節形状  $v_t^{part}$  を潜在空間に投影し、 $x_{v_t}^{part}$  を得る。

**step.4** パーティクルフィルタによって  $x_{v_t}^{part}$  から  $x_{m_t}^{part}$  を推定する。

**step.3**  $x_{v_t}^{part}$  の  $k$  近傍学習サンプルから、体節割り当てと復元誤差修正を行い、時刻  $t$  における体節ラベル付修正形状  $m_{\theta,t}^{labeled}$  を得る。

#### 5.3.3 部位統合による詳細全身形状の獲得

それぞれの体節バウンディングボックスはその他のバウンディングボックスと重複していること、また  $k$  近傍の類似サンプルから体節を決定するために、1つのボクセルに対して複数のラベル候補が存在することになる。さらに、体節領域は塊で存在することから、周辺領域のラベルを用いることも有効である。

よってこれらの重みつき平均によってもっともラベル確率の高いものをそのボクセルの体節として割り当て、形状全体としての整合化をはかり、詳細全身形状を獲得する。

## 6. 実験と考察

### 6.1 実験結果

提案手法の有効性を確認するため、形状変化の大きい着物を身に付けた人物を観測対象として実験を行った。観測対象を囲むように天井に設置された 8 台の同期撮影カメラにより 30fps で撮影される画像群を利用した実験について以下に示す。

400 フレームからなる時系列ボリュームサンプルを利用し、低解像度な全身形状および高解像度な部位形状のボクセルサイズは世界座標系でそれぞれ 80[mm]、40[mm] とした。また、低解像度全身形状は 6240 次元で高解像度部位形状は 1521 - 18879 次元を満たすものだった。

ボリュームの学習には Back-constrained Gaussian Process Dynamical Models(bc-GPDM) を用いた。潜在空間の次元数  $d$  は全身形状では 50、部位形状では 10 を用いた。カーネル関数には共に RBF を使い、 $\gamma_{bc} = 1 \times 10^{-3}$ 、 $\gamma_{dy} = 1 \times 10^{-3}$  を用いた。

パーティクルフィルタに関しては粒子数 1000、 $\sigma$  については部位ごとに経験的に設定した。

形状追跡結果を図 5 に示す。着物の複雑な変化に対しても体節割り当てができていことが伺える。実行速度は 4sec/frame 程度であった。

### 6.2 考察

最終的な追跡結果はボクセルの空間解像度の設定によって大きく変化するが、本手法においてはボクセルデータに対して直接次元削減を適用していることに起因して、膨大な計算コストが発生する。特に GPLVM

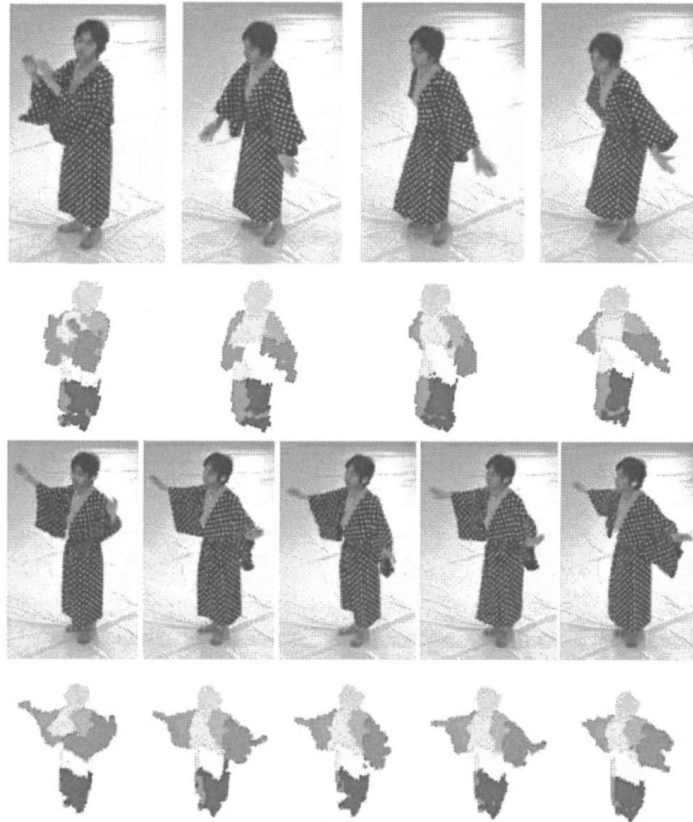


図 5 実験結果：上段は観測画像，下段は形状解析結果

の学習には入力次元に関して  $O(n^3)$  の計算量がかかってしまう。今回の実験では計算量の理由から、従来法の実験と比較して全身形状および部位形状のボクセルサイズをそれぞれ 60[mm] から 80[mm] へ、20[mm] から 40[mm] へ変更する必要がある。それゆえ、形状追跡結果は従来法に比べて解像度が荒くなってしまった。また、本来ならば従来法と比較をすべきであるが今回はできなかった。よって従来法との比較を含めた性能評価は今後の課題とさせていただく。

本稿では AR model によって潜在空間内での状態遷移を記述したが、式 (10) によって得られた  $a_i$  を用いて形状追跡を行うことも考えられる。

また、今後大量データの学習を行い、さまざまな動きの学習を行うことを考えた場合、ボクセルデータを直接用いることは好ましくない。そこで、シルエットから Shape context<sup>28)</sup> によって特徴抽出を行っているようにボクセルデータから 3 次元の幾何学情報を抽出し、その特徴量に関して GPLVM を行うことが考えられる。さらに、6) の様に類似形状ごとに分割を行い、学習させることで、オフラインでの学習における計算量爆発の問題を解決できると考える。合わせ

て、これらは今後の課題とする。

## 7. おわりに

本研究の一部は、文科省プロジェクト「知的資産の電子的な保存・活用を支援する支援するソフトウェア基盤技術の構築」における研究開発課題「大型有形・無形文化財の高精度デジタル化ソフトウェアの開発」の補助を受けた。また、色検出<sup>27)</sup>、形状修正<sup>8)</sup>、については、それぞれ和歌山大学和田俊和教授、京都大学延原章平助教にソフトウェアを提供して頂いた。さらに、GPLVM<sup>20)</sup> については Dr. Neil D. Lawrence の公開コードを用いて行った。深謝致します。

## 参考文献

- 1) R. Poppe, "Vision-based human motion analysis: An overview," *CVIU*, Vol.108, No.2, pp.4-18, 2007.
- 2) T. B. Moeslund and E. Granum, "A Survey of Computer Vision-Based Human Motion Capture," *CVIU*, Vol.81, No.3, pp.231-268, 2001.
- 3) G. K. M. Cheung, T. Kanade, J.-Y. Bouguet,

- M. Holler, "A real time system for robust 3D voxel reconstruction of human motions," In *CVPR*, Vol.2, pp.714–720, 2000.
- 4) X. Wu, O. Takizawa, and T. Matsuyama, "Parallel Pipeline Volume Intersection for Real-Time 3D Shape Reconstruction on a PC Cluster," In *Proc. of The 4th IEEE International Conference on Computer Vision Systems (ICVS)*, 2006.
  - 5) I. Mikic, M. Trivedi, E. Hunter and P. Cosman, "Human Body Model Acquisition and Tracking using Voxel Data," *IJCV*, Vol.53, No.3, pp.199–223, 2003.
  - 6) F. Caillette and T. Howard, "Real-Time Markerless Human Body Tracking with Multi-View 3-D Voxel Reconstruction," In *BMVC*, Vol.2, pp.597–606, 2004.
  - 7) K. N. Kutulakos, S. M. Seitz, "A Theory of Shape by Space Carving," *IJCV*, Vol.38, No.3, pp.199–218, 2000.
  - 8) S. Nobuhara and T. Matsuyama, "Deformable Mesh Model for Complex Multi-Object 3D Motion Estimation from Multi-Viewpoint Video," In *3DPVT*, 2006.
  - 9) V. Scholz, T. Stich, M. Keckeisen, M. Wacker, and M. Magnor, "Garment Motion Capture Using Color-Coded Patterns," *Computer Graphics Forum*, Vol.24, No.3, pp.439–448, 2005.
  - 10) K. Bhat, C. Twigg, J. Hodgins, P. Khosla, Z. Popovic, and S. Seitz, "Estimating Cloth Simulation Parameters from Video," In *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp.37–51, 2003.
  - 11) R. Plankers and P. Fua, "Tracking and Modeling People In Video Sequences," *CVIU*, Vol.81, No.3, 2001.
  - 12) G. Ye, J. J. Corso, and G. D. Hager, "Gesture Recognition Using 3D Appearance and Motion Features," *Workshop on Real-time Vision for Human-Computer Interaction*, 2004.
  - 13) D. Vlastic, I. Baran, W. Matusik, and J. Popovic, "Articulated Mesh Animation from Multi-view Silhouettes," In *ACM SIGGRAPH*, 2008.
  - 14) Isard, M. and Blake, A, "CONDENSATION-Conditional Density Propagation for Visual Tracking," *IJCV*, Vol.29, No.1, pp.5–28, 1998.
  - 15) L. Herda, R. Urtasun, and P. Fua, "Hierarchical implicit surface joint limits for human body tracking," *CVIU*, Vol.99, No.2, pp.189–209, 2005.
  - 16) H. Sidenbladh, M. J. Black, and L. Sigal, "Implicit Probabilistic Models of Human Motion for Synthesis and Tracking," In *ECCV*, Vol.1, pp.784–800, 2002.
  - 17) A. Agarwal and B. Triggs, "Tracking Articulated Motion using a Mixture of Autoregressive Models," In *ECCV*, Vol.3, pp.54–65, 2004.
  - 18) K. Grauman, G. Shakhnarovich, T. Darrell, "Inferring 3D Structure with a Statistical Image-Based Shape Model," In *ICCV*, pp.641–648, 2003.
  - 19) A. Elgammal and C. S. Lee, "Inferring 3D body pose from silhouettes using activity manifold learning," In *CVPR*, Vol.2, pp.681–688, 2004.
  - 20) N. D. Lawrence, "Probabilistic non-linear principal component analysis with Gaussian process latent variable models," *Journal of Machine Learning Research*, Vol.6, pp.1783–1816, 2005.
  - 21) N. D. Lawrence and J. Quiñero-Candela, "Local distance preservation in the GP-LVM through back constraints," In *ICML*, Vol.148, pp.513–529, 2006.
  - 22) R. Ursasun, D.J Fleet and P. Fua, "3D People Tracking with Gaussian Process Dynamical Models," In *CVPR* Vol.1, pp.238–245, 2006.
  - 23) J.M. Wang, D.J. Fleet, A. Hertzmann, "Gaussian Process Dynamical Models for Human Motion" *PAMI*, Vol.30 pp.283–298, 2008.
  - 24) K. Grochow, S. L. Martin, A. Hertzmann, and Z. Popovic, "Style-based Inverse Kinematics," *ACM Transaction on Graphics*, Vol.23, No.3, pp.522–531, 2004.
  - 25) S. Hou, A. Glata and F. Caillette, "Real-time Body Tracking Using a Gaussian Process Latent Variable Model" *ICCV*, Vol.14p.1–8, 2007.
  - 26) J.W.H. Tangelde and R.C. Veltkamp, "A survey of content based 3D shape retrieval methods" In *Proc. of International Conference on Shape Modeling Applications (SMI)*, pp.145–156, 2004.
  - 27) 和田俊和 : "最近傍識別器を用いた色ターゲット検出", 情報処理学会 CVIM 論文誌, Vol.44, No.SIG17, pp.126–135, 2003.
  - 28) A. Agarwal and B. Triggs, "3D Human Pose from Silhouettes by Relevance Vector Regression" In *CVPR*, Vol.2, pp.882–888, 2004.
  - 29) N. Ukita, R. Tsuji, and M. Kidode, "Real-time Shape Analysis of a Human Body in Clothing with Time-series Part-labeled Volumes," In *ECCV*, 2008.