

時系列形状の主成分分析に基づく 人体・着衣の形状解析

浮田 宗伯^{†1} 辻 良介^{†1} 木戸出 正継^{†1}

人体の3次元復元形状に対して、復元誤りの修正および各体節領域の特定を同時にオンライン実行できる手法を提案する。提案手法は、任意の人体や形状変化の大きな非剛体着衣にも適用可能であるという特徴を持つ。本手法は、詳細解析により得られる対象の高精度形状の時系列変化を体節ラベル付きで事前に学習し、この学習データと入力データとの比較による形状解析を行う。また、学習データ探索の高速化のため、時系列形状は主成分分析して固有空間上の多様体として記録する。その固有空間上にオンラインで計算される入力形状を投影し、学習データ中から類似データを探索する処理を階層的に行うことにより、各体節情報を持った高精度形状を安定に獲得する。この獲得形状とオンライン復元形状の比較により、目的とする各体節領域の特定および復元誤りの修正が可能となる。提案手法の有効性を示すため、形状変化の大きい着物を着衣とした実験を行った。

Shape Analysis of a Human Body Wearing Clothing Based on PCA of Time-series Volume Data

NORIMICHI UKITA,^{†1} RYOSUKE TSUJI^{†1}
and MASATSUGU KIDODE^{†1}

We propose a method for simultaneously refining the reconstructed volume of a human body with loose-fitting clothing and identifying body-parts in it online. A set of temporal target volumes, acquired by a sophisticated 3D reconstruction algorithm, with body-part labels is learned in advance. The temporal volume data is learned using PCA and stored as a manifold in the eigenspace. Each volume reconstructed online is projected into the eigenspace and compared with the manifold in order to find similar high-precision data with body-part labels. Experimental results demonstrate that our method can refine the volume including loose-fitting clothes and identify its body-part labels online.

1. はじめに

人体の姿勢・運動情報の獲得により、ジェスチャなどの動作を入力としたインタフェース、ヒューマン・ロボットインタラクションや、観測人物の位置・姿勢に応じてカメラを制御する能動観測、運動情報に基づく映像のコンテンツ化、スポーツ・伝統技能などの自動撮影・学習支援システムなど様々な応用が期待される。任意のタスク・環境における人体の運動情報獲得には、着衣を含めた身体中の体節領域の特定が要素技術として必要となる。こうした技術としては、用途に合わせた様々な手法^{1),2)}が提案されている。

単眼カメラによる2次元画像情報に基づいた手法では、対象領域に対して人体の近似モデルをあてはめて姿勢推定を行う。しかし、1方向からの観測情報のみを参照しているため、遮蔽に対し不安定である。

この問題に対して、複数カメラにより対象を多方向から撮影して得られる3次元形状（ボリュームデータ）に基づく手法では、遮蔽に対する頑健性が向上している。さらに、モデルあてはめによる姿勢推定にとどまらず、その後の形状解析への展開も可能となる。計算コストを要する3次元復元も、高速かつ安定な視体積交差法^{3),4)}の利用により実時間実行が可能であり、それに続く姿勢推定の高速化によりオンラインシステムの実現も十分に可能となってきている。

3次元形状に基づく手法では、文献5)などで行われているように、各体節を円柱などのシンプルな剛体で表現し、その集合により近似された人体モデルと復元ボリュームの重なりが最大になるモデルパラメータを求めることにより姿勢推定が行われる。こうした手法を洗練することにより、実時間処理を可能にした研究も存在する^{6),*1}。また、文献7),8)のように、人体の各体節（頭、胴、前腕など）がそれぞれ異なる剛体運動をするという仮定に基づいて、時系列ボリュームから各体節領域を抽出する手法も存在する。しかし、これらすべての手法は「各体節は剛体として近似可能」という大きな仮定に基づいている。より詳細な形状・姿勢推定を目的として、(i) レーザレンジファインダで得られる詳細ボリュームから人体モデルを得ることにより円柱などの近似モデルによる誤差を軽減させる、(ii) 姿勢変化に応じた関節部分の変化もモデル化する、といった手法も存在するが（文献9)など）、着物のように変化の大きな着衣の形状まで表現することはできていない。そのため、人体の

^{†1} 奈良先端科学技術大学院大学

Nara Institute of Science and Technology

*1 文献6)では、人体ボリュームを128³個のボクセルに分割して60fpsを実現している。

運動による着衣の変化が大きい場合、各体節を剛体近似できず上記手法を適用することは困難である。

まったく剛体近似を必要としない、または変化の大きな着衣を含んだ姿勢推定やその形状解析を行う従来手法はきわめて少ない。手法 10) では、各体節の形状を剛体ではなくパラメトリックな空間分布として表現し、各体節の体積保存や体節ラベルの空間的連続性を制約条件として各体節形状の最適パラメータを推定している。しかし、(1) 最適パラメータ推定における反復計算による速度低下、(2) パラメトリックモデルによるボリューム表現の限界、(3) 3次元復元誤差による性能低下、などの問題が残る。

本手法では、上記問題 1, 2 についてはそれぞれ、反復を要さない解探索手法、3次元空間の離散表現であるボクセルの集合そのものによって複雑な体節形状も表現できるモデル、を実現することによって解決する。問題 3 に関して、3次元復元結果から誤差を完全に除去することは困難である。特に、視体積交差法による3次元復元を行う限り、復元結果は真の形状を含んだ凸包であり凹形状を表現できない、すなわち、凹部に偽のボリュームデータが含まれてしまうという問題は不可避である。この結果、カメラ配置と対象の位置・姿勢によっては大きな復元誤りが生じてしまう。この偽のボリューム領域を除去する手法として、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 Carving¹¹⁾ や、時系列ボリュームデータ表面の動的変形を行う弾性メッシュモデル¹²⁾ などが提案されている。しかし、それぞれ観測対象中の一様なテクスチャ部分に対応困難、オンライン処理が不可能、という問題がある。文献 13) のように、凹凸が激しく変化する非剛体着衣の正しい形状推定を行うためにコードパターンを写した特殊な着衣を観測する手法も存在する。しかし、この手法も高精度復元を行うためには処理速度(実験¹³⁾では Pentium4 3.2 GHz でメッシュ数 3,000 に対して 30 sec/frame) が実時間にはほど遠いものになってしまう。また、布に力をかけた際の動きを実際に観測し、その動きモデルを推定して人体にあわせた着衣の形状推定を可能にする取り組みも存在するが(文献 14), 15) など)、人体の動きが既知であることを仮定するため、人体の姿勢・形状も同時推定する本稿の目的に対して利用することは不可能である。このように、従来手法により複雑な着衣の形状をオンラインで正しく復元することは難しい。

以上の議論から、本稿では以下の特徴を備えた「視体積交差法により生成される時系列ボリュームデータからの人体・着衣の形状解析」を目的とする。

- (1) オンライン実行可能
- (2) 着衣の大きな変形を許容するボリュームデータ中の体節領域の特定

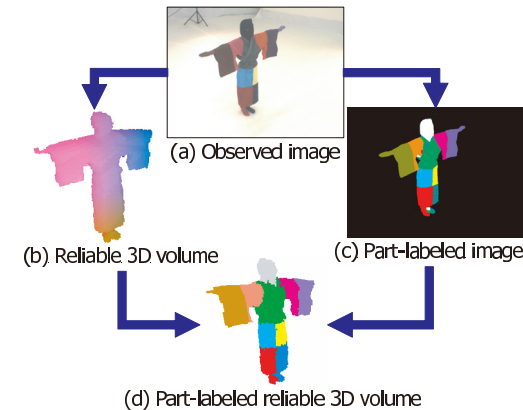


図 1 体節ラベル付き高精度ボリュームデータの生成処理
Fig. 1 Process flow for generating reliable volume data with part-labels.

(3) 復元 3次元ボリュームの修正

ボリュームデータ中の各体節領域の特定は、着衣も含んだ復元ボリューム中の各ボクセルに対して各体節に相当するラベルを割り当てる処理である。以降本稿では、一般的に体節と呼ばれる関節で区切られた人体領域に「その人体領域の動きにともない形状を変化させる周辺着衣部分(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、本研究における特定対象と見なす。各体節の色分け例を図 1 に示す。本研究で目的とするボリューム修正とは、シルエット抽出などが原因の大きな復元誤りや視体積交差法の原理上発生する大きな偽のボリュームの除去を指す。

上記した体節領域の特定処理は、従来手法^{1),2)}が目的とする人体そのものの位置や関節角を示した姿勢推定結果ではない。しかし、これら従来手法の多くでは、各関節・肢の位置・角度を推定するために大まかな初期値を求めた後に最適解を探索している^{5),8),16)}。しかし、着物のように人体が着衣によって大きく隠されているとその初期値の推定が困難である。そこで、各関節・肢の存在領域を我々の提案手法により推定される体節内に限定できれば、従来の姿勢推定法^{5),8),16)}により着物を身につけた状態での動作も推定対象とできる可能性が広がる。また、我々の手法の結果から得られる各体節の形状パラメータ(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¹⁷⁾などへの応用も可能である。特に我々は、着物のような緩い着衣を着て行われることの多い伝統舞踊・

技能の3次元デジタルアーカイブ¹⁸⁾・コンテンツ化¹⁹⁾に興味を持っており、我々の提案手法はこうしたデータにおける身体動作の解析にも有用であると考えられる。

2. 処理概要

人体の体節位置・姿勢の推定手法では、人体に関する様々な知識を利用することが精度や頑健性を向上させるための一般的なアプローチである。特に近年では、実際に観測した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多く提案されている(関節可動角²⁰⁾や姿勢変化の確率的表現^{21),22)}など)。こうした事例ベースの手法では、パラメトリック表現では困難な人体の複雑かつ微妙な変化を正確に表現できる。

これらの従来法はすべて、学習データをモーションキャプチャシステムにより得ている。モーションキャプチャでは、マーカを各関節に相当する箇所に装着し、これを観測した結果からその3次元座標を計算することにより、実際の関節の動き、すなわち姿勢データを正確かつ容易に獲得できる。しかし、本稿では人物が着物のように緩い着衣を着ていることも想定している。この場合、観測できるようにマーカを関節に相当する着衣上に装着すると、人の動きにあわせて着衣ずれが生じてしまい、関節の位置を正しく計測することができない。磁気式モーションキャプチャなどでは、外部からマーカを観測する必要がなく人体上に直接マーカを貼り付けることができるが、人の自然な動きを阻害してしまうといった欠点はまぬがれない。さらに本質的な問題として、マーカの3次元座標だけを計測しているため、着衣を含む観測対象のボリューム変化は表現できない。3次元CGモデルを利用して学習データを生成する手法²³⁾も存在するが、実際の観測に基づいて生成された学習データと比較して実際の形状・動きの再現性に欠ける。特に、人体が速く・大きく動いた際の着物の袖・裾のような生地のダイナミックな動きを再現することは難しい。

そこで、本稿では事例ベースの長所を備えたまま本稿の目的を実現するため、以下のデータ学習を行う。

- 人体・着衣の高精細3Dボリュームの時系列変化を学習
- 各時刻の3Dボリュームと体節ラベルを対応づけて記録

学習しておくボリュームデータは、その生成にオフラインで時間をかけても問題ないので、手法11)、12)のような従来法によって正しい対象ボリュームにより近いデータを獲得しておく。体節ラベルデータは、特定したい体節ごとに着色された着衣の3次元復元結果から直接獲得する。この学習データと入力の時系列ボリュームデータとの比較により形状解析を行う。また、学習データ探索の高速化のため、時系列ボリュームデータは主成分分析

され、形状変化は固有空間上の多様体として記録される。この固有空間上に入力ボリュームデータを投影し、多様体中から類似データを探索することにより、体節情報を持った高精度形状を獲得する。この多様体は、時系列的に連結されたボリュームデータにより構成されている。すなわち、パラメトリック固有空間法²⁴⁾が姿勢・照明パラメータに応じた物体の見えの変化を多様体により表現しているように、我々の手法では対象のモーションダイナミクスを多様体表現している。よって、(1)動作予測や、(2)時系列的に連続な入力データの投影軌跡と学習多様体との比較による安定な推定、が可能になる。

2次元画像を対象にしている手法²⁴⁾などと比較した際の本稿の目的の難しさには、多彩な姿勢変化に対応する学習データ収集の困難さなどもあるが、大きな問題の違いは以下の2点である。

巨大な次元数のデータ 次元数が(画像の横方向×縦方向のピクセル解像度)のベクトルで表される画像データと比較して、次元数が(ボリュームを表現する空間の縦×横×高さのボクセル解像度)のベクトルで表現されるボリュームデータの次元数はきわめて大きい。この次元数はボリュームの解像度を高くするほど大きくなってしまふ。そこで、低解像度の全身ボリューム解析を行った後、その結果に基づいて特定された各体節領域の高解像度解析を行うことにより探索の安定性と高速性を両立させる(詳細は3.2、4.4、4.5節)。

同じ姿勢の学習・入力データの間の変遷 「正しく形状修正済みの信頼できる学習データ」と「視体積交差法で復元された大きな誤りを含むデータ」の比較を行うため、これらの差を考慮した学習データ獲得法とマッチング手法が要求される(詳細は4.2節)。

本研究では、入力のボリュームデータの各ボクセルに対して10種の体節ラベル(頭部、胴体部、右上腕部、右下腕部、左上腕部、左下腕部、右上脚部、右下脚部、左上脚部、左下脚部)のいずれかを割り当てる。さらに、特殊なラベルとして非対象を用意する。非対象ラベルは、対象ボリューム以外のボクセルに割り当てられるラベルである。視体積交差法による復元ボリューム以外のボクセルに割り当てられた後、提案手法により入力ボリューム中の偽のボリュームに対してもこの非対象ラベルを割り当てる。また逆に、復元ボリュームにおいて誤って非対象と判定されているボクセルには、人体に対応する10種類の体節ラベルのいずれかを割り当てる。この非対象ラベルと10種類の体節ラベルの上書きが、本稿の目的とするボリューム修正に相当する。すなわち、この特殊ラベルを含めた全11種類のラベル割当てによって、全身の復元ボリュームに含まれる全ボクセルに対する体節ラベル割当てとボリューム修正を同時に実現できる。

3. 時系列ボリュームの学習

3.1 体節ラベル付き高精度ボリュームデータ生成法

図 1 中の (a) のように特定したい体節ごとに着色された着衣の対象人物の形状を視体積交差法³⁾により復元する．視体積交差法の原理上、このボリュームデータには凹部に偽のボリュームが含まれてしまう．そこで、弾性メッシュ変形¹²⁾を利用して復元形状を真の形状(図 1 中の (b))に近づける．この手法では、ボクセルデータを 3 角形メッシュデータに変換し、複数カメラの時系列撮影画像と時系列復元ボリューム間でシルエット・テクスチャ・形状のスムーズさ・時間方向のスムーズさなどの整合性をとるように作用する力をメッシュ頂点にかけることで、より正しく高精度な形状復元結果を得る．この手法の出力が真の形状に近づくことは文献 18)でも定量的に評価されており、その出力は本稿の目的に合致している．

この高精度ボリュームデータの各ボクセルに対して、体節ラベルを割り当てる．そのために、まず色検出によって着色された体節領域ごとにラベル化された画像(図 1 中の (c))を生成する．この色検出には手法 27)を利用した．次に、多視点ラベル画像から復元ボリュームに対してラベルの逆投影を行うことにより、表面ボクセルへの体節ラベルの割当てを行う．最後に、各内部ボクセルにおいて最近傍の体節ラベル化された表面ボクセルを探索して、内部ボクセルの体節ラベルの割当てを行う．以上の処理により、図 1 中の (d) のような体節ラベルを持った高精度形状を生成できる．この結果を学習データとして利用する．

3.2 階層的なボリュームの学習

ボリュームデータの解像度が高いほど、最終的な形状解析の精度が向上する．しかし、前述したとおりボリュームデータは解像度に応じて次元数が急増するため、高解像度データをそのまま全身比較すると処理速度が低下してしまう．そこで、データ学習を低解像度な全身ボリュームデータと高解像度な各体節領域のボリュームデータとに分けて行う．この結果、高解像度解析の出力は各体節独立の形状解析結果となってしまうが、同一座標系におけるそれぞれの位置・姿勢は低解像度解析の結果から既知のため、すべてを組み合わせると全身形状を復元することができる．

3.3 主成分分析による形状変化の記録

高次元ベクトルであるボリュームの時系列データを、主成分分析により低次元空間の多様体として表現する．主成分分析を行うためには、各フレームにおけるボリュームデータの次元数(ボクセル数)を一致させる必要がある．そこで、各フレームにおけるボリュームデー

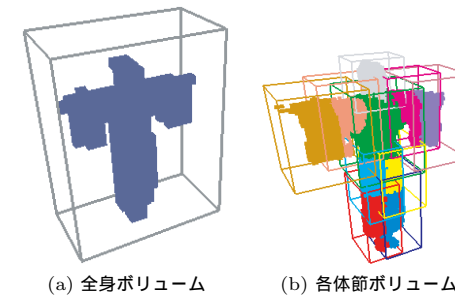


図 2 バウンディングボックスの例

Fig. 2 Bounding boxes.

タの 3 次元重心を求め、それを基準に 3 次元の対象物体を囲む一定サイズの直方体(以下、バウンディングボックスと呼ぶ)を定めた．このバウンディングボックス内のボクセルの集合を主成分分析にかける．バウンディングボックスのサイズは、全フレームにおいて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ボリュームを対象としたバウンディングボックスの例を図 2 に示す．ある体節のバウンディングボックス内には、他の体節のボリュームデータも含まれることに注意されたい．

ある時刻 t における d 次元のボリュームデータのベクトル表現を $v_t = (v_{t,1}, v_{t,2}, \dots, v_{t,d})^T$ ($v_{t,i} \in \{0, 1\}$) とする．ただし、1 は視体積交差法によって復元された対象ボリュームに含まれるボクセル、0 はそれ以外のボクセルとする．このベクトルは、世界座標系の下でバウンディングボックス内のボクセルをある順序で走査 (xyz 昇順など) することにより得る．全フレームのボリュームデータの集合 V は、次のように表現される．

$$V = [v_1 - m, v_2 - m, \dots, v_T - m] \quad (1)$$

ただし、 T は学習データとして利用するボリュームデータ数(すなわち、撮影フレーム数)、 m は T 個のボリュームデータの平均ボリュームデータである．次に、 V からボリュームデータの共分散行列 $S = VV^T$ を算出し、この固有ベクトルの集合 $\{e_i | i = 1, \dots, d\}$ を得る．ただし、 e_i は対応する固有値 λ_i の大きい順に並んでいるものとする．この固有ベクトルを基底とした空間が固有空間となる．

d 次元のボリュームデータ v_t は、 d より十分に小さい k 個の固有ベクトルによって近似表現できる．すなわち、 k 個の固有ベクトルからなる行列 $E = [e_1, e_2, \dots, e_k]$ を用いて、下式の線形射影により、 k 次元の固有空間上の点 y_t に変換できる．

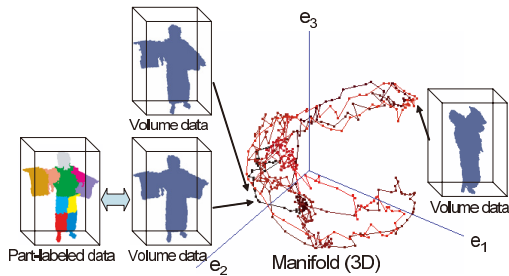


図 3 学習空間内の体節ラベル付き多様体：固有空間の次元数 3 の例
Fig. 3 Manifold with part-labels in a (3D) eigenspace.

$$y_t = E^T (v_t - m) \quad (2)$$

また、時系列に連続なボリュームデータの投影点を結ぶことにより、時系列的な形状変化を軌跡として表現することができる。

上述の処理により、学習データを固有空間上の多様体として表現できる。この多様体は、学習時系列ボリュームの投影点の集合 $\{y_1^L, \dots, y_T^L\}$ で構成される。ただし、 y_t^L は時刻 t における学習データの投影点を表すとする。さらに、多様体中の各点（あるフレームのボリュームデータ）に対して、対応する体節ラベルデータを記録しておく。図 3 にその様子を示す。

本節の処理は、図 3 に示した全身ボリュームデータだけでなく、各高解像度体節ボリュームデータに対してもそれぞれ適用される。したがって、最終的には 11 個（全身 + 10 個の体節）の固有空間中に、体節ラベル付きボリュームデータの時系列変化が多様体として記録される。

4. 形状解析：体節ラベル割当てとボリューム修正

オンライン形状解析では、視体積交差法により生成される時系列ボリュームデータを入力とし、事前に作成済みの学習データを参照した形状解析を行う。図 4 に処理の流れを示す。図 4 中の数字は、各処理の詳細を示す節番号である。

4.1 時系列ボリューム学習データからの解探索

本節では、全身・体節データの解析に共通した類似解探索による形状解析の概要を述べる。時刻 t において現フレームと過去 n フレームからなるオンライン時系列ボリュームデータ $\{v_t, v_{t-1}, \dots, v_{t-n}\}$ を、それぞれ固有ベクトルによって張られる空間である学習空間に

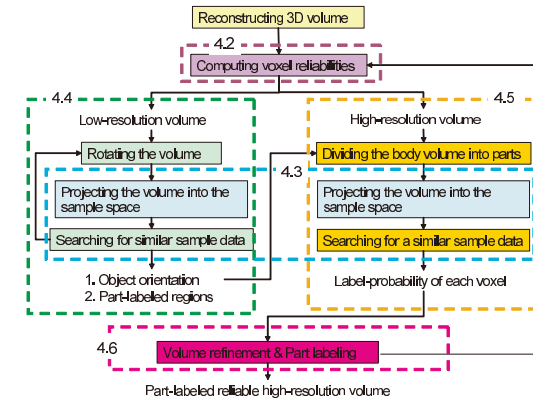


図 4 形状解析の処理の流れ
Fig. 4 Process flow of shape analysis.

投影する。投影計算は式 (2) で示したとおりであり、時系列ボリュームデータの投影点群 $Y_t^L = \{y_t^L, y_{t-1}^L, \dots, y_{t-n}^L\}$ からなる軌跡パターンが得られる。ただし、 y_t^L, Y_t^L は、それぞれ時刻 t における入力データの投影点、時刻 t から始まる $n+1$ 個の時系列的に連続な投影点の集合を表すものとする。この軌跡パターン $Y_t^L = \{y_t^L, y_{t-1}^L, \dots, y_{t-n}^L\}$ と学習データの軌跡パターン $Y_s^L = \{y_s^L, y_{s-1}^L, \dots, y_{s-n}^L\}$ ($s \in \{n+1, \dots, T\}$) との比較により類似データを探索する。よって、参照する過去の履歴数 n が大きいほど、同一動作の探索成功率は向上する。一方、探索コストは大きくなってしまふ。また、 n が大きすぎると、長時間にわたって学習データとまったく同じ動きをしている入力動作の解析しかできなくなってしまう。そこで、タスクにあわせて処理時間、および短い類似動作の組合せからなる動きへの適用可能性を考慮にいれて、 n を決定するのが望ましい。本稿で示す実験では $n=5$ とした。この処理は、各観測フレーム独立の処理である。すなわち、直近数フレームの軌跡探索により、類似姿勢の多数学習データ中から各入力フレームのボリュームに対応する学習データを探索する処理の頑健化をはかっているが、前フレームの推定結果をそのフレーム以降における推定に利用するような追跡処理ではない。

具体的には、入力・学習両軌跡パターンの各点間の距離総和を探索の評価値とし、下式で表される最小距離総和 D_t に対応する Y_s^L 中の学習データ y_s^L を入力データ y_t^L と最も類似した学習データと見なす。

$$D_t = \min_{s=n+1}^T \sum_{i=-n}^0 \| \mathbf{y}_{s+i}^L - \mathbf{y}_{t+i}^I \|, \quad (3)$$

ただし、 $\| \cdot \|$ はベクトルのノルムを表す。こうして決定された学習データ \mathbf{y}_s^L は、凹部分のボリューム除去などの処理が施された高精度形状に相当し、その形状と対応する体節ラベルデータも参照できる。

しかし、入力ボリュームデータには偽のボリュームデータが含まれており、この偽のボリュームデータの影響により正しい学習データ、すなわち入力ボリューム観測時の実際の人体・着衣の姿勢・形状と一致する学習データを探索できない恐れがある。また、学習空間内の全データと入力データとの比較を行うことは、学習したボリュームデータの数に比例して処理時間が増大してしまうという問題をかかえている。これら偽のボリュームへの対処法、および探索の効率化について、それぞれ 4.2, 4.3 節で述べ、4.4 節以降でこれらの処理を利用したオンライン形状解析を実際の手順に沿って説明していく。

4.2 偽のボリュームデータへの対応

形状復元誤差である偽のボリュームデータの影響に対して頑健な処理を行うため、各ボクセルにおける偽のボリュームの発生を確率的に表現・獲得しておく。この値をボクセル信頼度と呼ぶ。ボクセル信頼度は、学習データ中に含まれる各形状の各ボクセルに対して与えられる。

4.2.1 ボクセル信頼度の算出法

偽のボリューム領域の発生箇所は、対象形状だけでなく撮影カメラ群との位置関係によって大きく変化する。この発生箇所を学習するために、以下の処理を行う。

- step.1 仮想的な 3 次元空間中において実システムと同様の位置・姿勢でカメラを設置し、学習ボリュームデータの位置・方位を変化させながら画像平面に投影することより、様々な人体位置・方位における各カメラでのシルエット画像を生成する。
- step.2 この仮想的なシルエット画像から視体積交差法による形状復元を行うことにより、様々な人体位置・方位において獲得されるはずの「視体積交差法による偽のボリュームを含んだ仮想ボリューム」を生成する。
- step.3 step.2 で生成された仮想ボリューム集合と元の学習ボリュームを比較することにより、学習ボリュームの人体・着衣形状における偽のボリューム領域の発生箇所を特定できる。
- step.4 step.1 ~ step.3 を学習データ中のすべてのボリュームデータに対して行うことによ

り、様々な人体・着衣の姿勢・形状における偽のボリューム領域の発生箇所を特定する。 T フレームからなる学習データ中、 t フレームの学習ボリュームの i 番目のボクセルの信頼度 $c_{t,i}$ を下式のように定義する。

$$c_{t,v} = \frac{1}{N_{var}} \sum_i^{N_{var}} \delta_{t,v}^i \quad (4)$$

ただし、 N_{var} は t フレームの学習ボリュームを仮想空間中の様々な位置・方位において観測して視体積交差法を行う際の観測回数、 $\delta_{t,v}^i$ は時刻 t の学習ボリュームとその i 番目の仮想ボリュームの間でボクセル v の復元結果が異なる場合に 0、それ以外の場合 1 をとる変数である。時刻 t の学習データに対するボクセル信頼度 C_t は下式で表現する。

$$C_t = \begin{bmatrix} c_{t,1} & & 0 \\ & \ddots & \\ 0 & & c_{t,d} \end{bmatrix} \quad (5)$$

4.2.2 ボクセル信頼度を利用した学習空間への投影

式 (2) にボクセル信頼度を適用した式、すなわち入力データを学習空間へ投影する式

$$\hat{\mathbf{Y}}_t^I = E^T C_t (\mathbf{v}_t^I - \mathbf{m}) \quad (6)$$

により得られる点 $\hat{\mathbf{Y}}_t^I$ を探索に用いる。これにより、偽のボリュームの影響を軽減させた探索を実現できる。しかし、ボクセル信頼度 C_t は、入力データ \mathbf{v}_t が対応する学習データが既知でないと求まらず、鶏と卵の関係にある。そこで、本手法では時系列的に連続な入力データ間の差は微小であることを仮定し、解析処理時の時刻 t に利用するボクセル信頼度は時刻 $t-1$ で探索された学習データから決定する。

4.3 学習データ探索の効率化

探索の高速化のためには、学習空間中の全データに対して類似度の評価を行うのではなく、入力データの近傍の学習データに対してのみ類似度の評価を行えばよい。そこで本手法では、事前に学習空間全体を一定間隔の部分領域に区切り、各部分領域中に存在する学習データを調べておく。推定時には、まず入力データの投影点が含まれる部分領域を選択し、推定された部分領域内でのみ探索を行う。例を図 5 に示す。図 5(a) のように部分領域のサイズを小さくするほど、探索を行うデータ数は減少するが、投影点 P のように入力データが投影された部分領域 (図 5, 図 6 に示すように、投影点 P を含む部分領域を SR_P と表

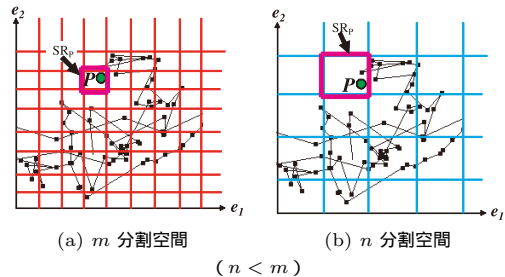


図 5 探索空間の階層化による効率化
Fig. 5 Efficient hierarchical search.

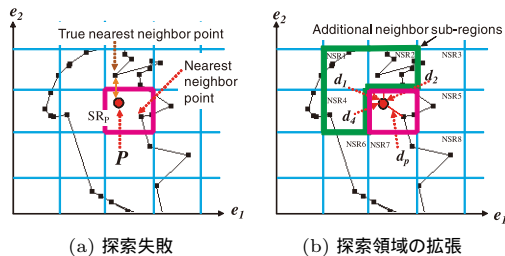


図 6 探索範囲の拡張による最近傍探索
Fig. 6 Additional search regions.

記する)に学習データが存在しない可能性が上がる．そこで，図 5 (b) のようにさらに大きなサイズに分割された部分領域を用意しておき，小さいサイズから順に入力データの投影・探索を行う．

しかし，部分領域 SR_P のみを探索すると，部分領域 SR_P 外に真の最近傍学習データ (図 6 (a) の “True nearest neighbor point”) が存在する可能性がある．そこで，投影点 P と部分領域 SR_P における最近傍学習データ (図 6 (a) の “Nearest neighbor point”) との距離 d_P を閾値にして，投影点 P からの距離 d_i が $d_i < d_P$ を満たす隣接部分領域 NSR_i (図 6 (b) の太線で囲まれた “Additional neighbor sub-regions”) を探索範囲に加える．

このように，探索範囲を階層的に設定した効率的な探索により，高速性と確実に類似解を探索できる安定性を両立させた．

4.4 低解像度全身ボリュームデータの形状解析

本節から，ここまで述べた処理を利用したオンライン形状解析を実際の手順に沿って説明していく．

オンラインで計算される低解像度時系列ボリュームデータを全身ボリュームの学習空間に投影し，類似データの探索により以下の 2 点を推定する．

対象の向き ボリュームデータのベクトル表現は，人体の姿勢とは無関係に世界座標系の下でボクセル空間をつねに同じ順で走査 (xyz 昇順など) することにより得られている．よって，対象の姿勢・形状が同じであっても，対象の向きが異なれば，まったく異なるデータと見なされてしまう．そこで，同じ形状・姿勢の学習データを基準とした入力データの回転角を推定する．

体節領域 入力ボリュームデータの中から各体節の存在領域を特定できれば，その領域のボリュームと学習済の高解像度体節ボリュームとを比較できる．そこで，入力全身ボリュームデータ中の各体節ボクセルが占める領域を含んだバウンディングボックスの特定を行う．

上記 2 点を推定するための低解像度全身ボリュームデータの解析処理を以下に示す．

- step.1 時刻 t において入力ボリュームデータの 3 次元重心を計算．
- step.2 重心を原点とし，鉛直上向きの軸回りにボリュームデータを θ 回転．
- step.3 学習データ作成時と同サイズのバウンディングボックスを設定し，その内部のボリュームデータを学習空間に投影．
- step.4 step.1 ~ step.3 を時刻 $t-1$ から $t-n$ のボリュームに対しても実行して，学習空間における時系列ボリュームの軌跡 $Y_t^i(\theta)$ を獲得．ただし， $Y_t^i(\theta)$ は Y_t^i の各要素を θ 回転させた $\{y_t^i(\theta), y_{t-1}^i(\theta), \dots, y_{t-n}^i(\theta)\}$ である．
- step.5 式 (3) を拡張して，学習データ y_s^L と入力データ $y_t^i(\theta)$ との類似度評価値 $D_t(\theta) = \min_{s=n+1}^T \sum_{i=-n}^0 \|y_{s+i}^L - y_{t+i}^i(\theta)\|$ を得る．この評価値 $D_t(\theta)$ により，学習データ中から投影点の最類似データを探索．
- step.6 回転角度 θ を変化させ step.2 ~ 5 を反復．
- step.7 $D_t(\theta)$ の θ を対象の向き $\hat{\theta}$ と見なす．
- step.8 $D_t(\hat{\theta})$ に対応する全身学習データ中の体節ラベル付き高精度ボリュームデータを参照し，入力データ中の各体節領域を含むバウンディングボックスを決定．

4.5 高解像度体節ボリュームデータの詳細形状解析

低解像度な全身形状解析により，入力全身ボリューム中の各体節領域に対応するバウン

ディングボックスを特定できた．次に，高解像度復元された全身形状から各体節領域のボリュームデータを抽出する．これらをそれぞれの体節ボリューム学習空間に投影し，各ボクセルのラベル確率を決定する．ラベル確率とは，各ボクセルにおけるラベル候補を確率的に表現したものである．以下に，処理の流れを示す．

- step.1 時刻 t において入力の高解像度時系列ボリュームデータを対象の向き θ に回転．
 step.2 各体節に相当するバウンディングボックス内のボリュームデータをそれぞれの体節ボリューム学習空間に投影．
 step.3 step.1 と step.2 を時刻 $t-1$ から $t-n$ のボリュームに対しても実行して，学習空間における時系列ボリュームの軌跡を獲得．
 step.4 各体節において $D_t(\theta) < (\text{閾値})$ を満たす類似データを選択．
 step.5 各体節のデータに対してそれぞれ以下の処理を行う．ある体節において選択されたデータの個数を N_{data} とする．各データ d が step.3 で選択された際の距離の逆数をそのデータ d の尤度と見なし， L_t^d と表記する（ただし $d \in \{1, \dots, N_{data}\}$ ）．この尤度を用い，各ボクセル v におけるラベル l ($l \in \{1, \dots, N_{label}\}$ ； N_{label} は体節ラベルの種類数であり，本研究では $N_{label} = 11$) のラベル確率 $P^{v,l}$ を下式により決定．

$$P^{v,l} = \sum_{d=1}^{N_{data}} \frac{L_t^d}{S_t} c_{v,l}^d \quad (7)$$

ただし， S_t は尤度の総和 $\sum_{d=1}^{N_{data}} L_t^d$ であり， $\epsilon_{v,l}^d$ はデータ d のボクセル v のラベルが l の場合 1，それ以外の場合 0 を返す変数．

4.6 体節ラベル割当ておよびボリューム修正

体節ごとの詳細な形状解析結果を組み合わせ，各ボクセルがラベル確率を持った全身の高解像度ボリュームデータを生成する．ただし，体節のバウンディングボックスの間には重なりがあり，異なるバウンディングボックス中の同一ボクセルのラベル確率は異なる値を持つ．そこで，以下の処理によって最終的に各ボクセルに 1 つのラベルを割り当てる．

- step.1 異なる体節のバウンディングボックスが重複する領域中で，体節 p の解析結果におけるボクセル v の体節ラベル l のラベル確率を $P^{p,v,l}$ とする．まず，各ラベル確率の平均値 $\bar{P}^{v,l} = \frac{1}{N_{part}} \sum_{p=1}^{N_{part}} P^{p,v,l}$ を計算する．ただし， N_{part} はバウンディングボックスを持つ体節数，すなわち本研究では $N_{part} = 10^{*1}$ ．

各ボクセルで，得られた平均値 $\{\bar{P}^{v,1}, \dots, \bar{P}^{v,N_{label}}\}$ 中の最高値に対応するラベルを単純に選んでしまうと，隣接ボクセル間で異なるラベルが振り分けられてしまう可能性がある．人体構造上，孤立して異なるラベルが振り分けられることはほぼありえないので，次の step で全体的な整合化をはかる．

- step.2 そのボクセルを中心にした $N_{nv} \times N_{nv} \times N_{nv}$ の近傍ボクセルのラベル確率の距離重み付き平均 $\hat{P}^{v,l}$ を計算する．

$$\hat{P}^{v,l} = \frac{1}{\sum_{i=1}^{N_{nv}^3} w_i} \sum_{i=1}^{N_{nv}^3} w_i \bar{P}^{i,l} \quad (8)$$

ただし， w_i は中心ボクセル v と近傍ボクセル i とのユークリッド距離．本稿の実験では $N_{nv} = 3$ とした．

- step.3 $\hat{P}^{v,l}$ が最大値をとるラベル $l = L$ を，このボクセル v の体節ラベルとして決定する．

5. 実験と考察

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の舞踊動作（会津磐梯山踊り）を観測対象として実験を行った．こうした舞踊では，裾・袖のある着物を着用した状態で腕を上下左右に大きく動かし，また，ときには素早い動きも入る．よって，緩い着衣の速く大きな動きを観測できるため，提案手法の有効性の確認に適している．

観測対象を囲むように天井に設置された 7 台の同期撮影カメラ（Pointgrey 社 Flea：1,024 × 768 pixel，8 bit bayer）により 30 fps で撮影される時系列画像集合を利用した実験を行った．すべての処理は Opteron 1.8 GHz の PC で行った．

5.1 実験結果

学習データは，ある 1 人の観測データとした．被験者はビデオを見て舞踊動作を習得した．学習データは 1,000 フレームからなる時系列ボリュームにより構成され，低解像度な全身ボリュームおよび高解像度な各体節ボリュームのボクセルサイズはそれぞれ 60 mm^3 ， 20 mm^3 として， $138 \text{ cm} \times 66 \text{ cm} \times 186 \text{ cm}$ の範囲の形状復元を行った．すなわち，ボリュームデータの次元数は低解像度の全身解析において $23 \times 11 \times 31 = 8,743$ である．すべてのフレームにおいて確実に各体節をバウンディングボックスに収まるように，余裕を持ってサイズ設定した．高解像度の体節解析は，各体節の体積と動きの大きさによってバウンディングボックスの大きさが変わる．それぞれの総ボクセル数は表 1 のように設定した．手や足

*1 非対象ラベルはバウンディングボックスを持たないので，この体節数には含まれない．

表 1 高解像度体節データの総ボクセル数 (次元数)

Table 1 The number of voxels in high-resolution body-part volumes.

	総ボクセル数 (縦 × 横 × 高さ)
頭	7,695 (19 × 15 × 27)
胴体	15,939 (21 × 23 × 33)
右上腕	8,925 (25 × 17 × 21)
右下腕	26,825 (29 × 25 × 37)
左上腕	10,143 (23 × 21 × 21)
左下腕	37,625 (35 × 25 × 43)
右上脚	8,211 (17 × 23 × 21)
右下脚	11,781 (17 × 21 × 33)
左上脚	7,497 (17 × 21 × 21)
左下脚	11,781 (17 × 21 × 33)

の総ボクセル数は左右でも大きく違うが、これは舞踊による手や足およびそれらの動きにもなった着衣の動きの大きさの違いに依存するものであり、この総ボクセル数は 3.3 節で述べた基準に基づいて学習データから自動的に求められた。

学習空間の次元数 k は主成分分析で得られる固有値 λ の累積寄与率 a を基に決定した。

$$a = \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^d \lambda_j} \quad (9)$$

本実験では、十分な形状解析率を保てる最小の次元数として、累積寄与率が 75% を満たす次元数を各学習空間の次元数 ($k = 35$) とした。

上記の学習データを利用した形状解析を行った。被験者は 2 人で、うち 1 人は学習データと同一人物であった。2 人とも、学習データ中の着物と同じ形状・素材の着物を着用し、学習データと同じビデオを見て覚えた舞踊を踊った。ただし、学習データに含まれる被験者 (155 cm) ともう 1 人の被験者 (175 cm) は大きく体格が異なるため、後者の視体積交差法の結果は身長比にあわせて全体形状をサイズ変更した。

形状解析結果の例を図 7 に示す。全フレームにわたり、低解像度解析では手先などの細かな形状を無視した探索が行われ、高解像度解析では低解像度解析で無視された細かな形状まで正しく推定できている。また、人体の運動による着衣の大きな変形に対しても、安定に各体節領域の特定ができていることも確認できる。また、図 8 に示すように、入力ボリュームデータ中に含まれる偽のボリューム領域が、除去されていることも確認できた。

上記結果の精度を定量的に評価するための実験を行った。この実験では、学習時と同じ着

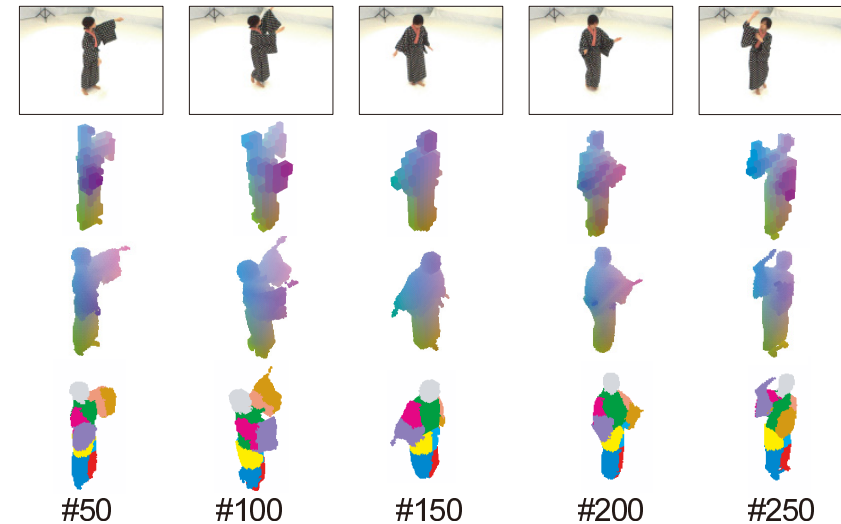


図 7 実験結果：上段から、観測画像、入力の高解像度ボリューム、形状解析結果
Fig. 7 Experimental results (1st-row: Observed images, 2nd-row: Input low-resolution volumes, 3rd-row: Input high-resolution volumes, 4th-row: Shape-analysis results).

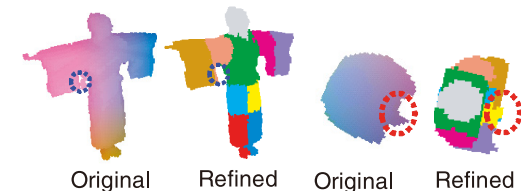


図 8 ボリューム修正結果

Fig. 8 Volume refinement results shown from two viewpoints.

色された着衣での舞踊動作を観測し、この観測画像系列から得られる以下 2 種類のデータを比較した。

正解データ 着色情報を利用して学習データ取得と同様の手法で得られる各体節の高解像度ボリューム

解析結果 着色情報を利用することなく提案手法によって得られる各体節の高解像度ボリューム

72 時系列形状の主成分分析に基づく人体・着衣の形状解析

表 2 各体節の復元結果の定量的評価：正解データの総ボクセル数を 100 に正規化した場合の誤検出・未検出ボクセルの割合および体積 cm^3 (値はすべて全フレームの平均値)

Table 2 Error analysis: The percentages of false-positive (FP) and false-negative (FN) voxels in proportion to the size of each body-part are shown.

体節 (正解体積)	誤 (割合/体積)	未 (割合/体積)
頭 (10,040)	7.2/202	5.2/47
胴体 (45,016)	3.9/1,765	3.1/1,376
右上腕 (19,344)	9.9/1,099	6.4/1,231
右下腕 (27,840)	11.4/3,176	10.8/3,010
左上腕 (21,304)	6.5/1,386	11.1/2,356
左下腕 (30,784)	9.5/2,917	10.0/3,068
右上脚 (26,584)	5.1/1,360	7.1/1,883
右下脚 (37,664)	11.8/4,474	8.7/3,259
左上脚 (28,928)	3.3/964	6.3/1,816
左下脚 (39,216)	7.5/2,946	6.4/2,518

比較内容は、正解データ以外の領域を検出してしまった誤検出と正解データの領域を検出できなかった未検出である。比較結果を表 2 に示す。表中の体積は、ボクセル 1 つの体積を 8cm^3 として計算した。この結果から、直観的にも類推できるとおり、動きの少ない頭・胴体と比較して着衣の揺れが激しい下腕・下脚のほうが誤差が大きいが分かる。この問題については、5.2 節において考察を行う。

次に、学習データと違う被験者における実験結果を図 9 に示す。この例では、背景差分の失敗により視体積交差法の復元結果に大きな誤りが含まれる場合においても、正確な学習データとの比較により良好な復元結果が得られていることも確認できた (図 9 の赤円で囲まれた領域)。先の実験と同様、正解データと解析結果との比較結果を表 3 に示す。図 9 を見ただけでは確認できないが、表 2 と表 3 の結果と比較すると、学習データと異なる被験者の解析結果の方が誤差が全体的に大きい。これも容易に想像できることではあるが、原因としては (1) 身長比に応じたサイズ変更だけでは体格の違いへの対応は不十分、(2) 個人によって動作が微妙に異なる、などが考えられる。この問題についても 5.2 節において考察を行う。

通常の姿勢推定手法で対象とされるタイトな着衣でも実験を行った。この実験では、被験者はラジオ体操第 2 を踊った。この体操も全身、特に腕の動きが激しく、提案手法の有効性の確認に適している。学習データのフレーム数が 4,000 である以外は、先に示した着衣における実験と同様の条件で実験を行った。形状解析結果の例を図 10 に、正解データと解析結果との比較結果を表 4 に示す。これらの結果から、手足の動きは激しくてもそれらの動き

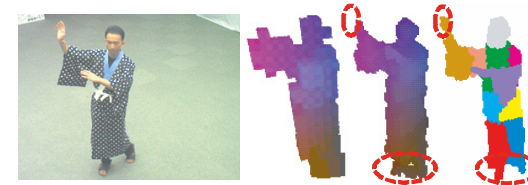


図 9 異なる被験者における形状解析結果：左から観測画像，低解像度復元結果，高解像度復元結果，形状解析結果
Fig. 9 Results in the other subject.

表 3 異なる被験者における各体節の復元結果の定量的評価
Table 3 Error analysis in the other subject.

体節 (正解体積)	誤 (割合/体積)	未 (割合/体積)
頭 (13,472)	7.9/393	5.7/81
胴体 (54,208)	5.2/2,841	5.8/3,121
右上腕 (22,792)	9.1/2,067	8.1/1,838
右下腕 (32,984)	13.2/4,358	12.2/4,019
左上腕 (23,888)	10.1/2,423	9.6/2,292
左下腕 (34,472)	14.8/5,091	11.0/3,783
右上脚 (30,672)	6.4/1,949	4.4/1,364
右下脚 (45,696)	11.6/5,323	9.4/4,287
左上脚 (32,152)	7.7/2,461	5.1/1,649
左下脚 (48,064)	12.9/6,197	8.1/3,914

にともなった複雑な着衣の動きは生じないため、着物着用の場合以上に提案手法が有効であることが確認できた。

以上の実験結果から、(1) 動きの激しい対象を観測する条件下で、緩い着衣の有無にかかわらず各体節について誤検出・未検出ともに最大 15%程度の精度で領域を特定して、(2) シルエット抽出失敗に起因する誤検出・未検出および視体積交差法特有の復元誤差をボリューム修正できた、ことを確認できる。

上記実験において、形状解析の実行速度は約 0.042 秒/フレームであった。この実行速度には三次元復元の時間は含まれていない。しかし、文献 4) において、VGA サイズの画像集合から 5mm^3 解像度での視体積交差法がビデオレート (0.033 秒/フレーム) で実行可能なことが示されているように、形状解析に十分な解像度での復元が実時間実行可能になっている。よって、この形状復元と我々の手法の実行時間をあわせても、ビデオレートに近い実行速度を実現できる。この結果、1 章であげた具体的な応用の中でもオンライン性が必要

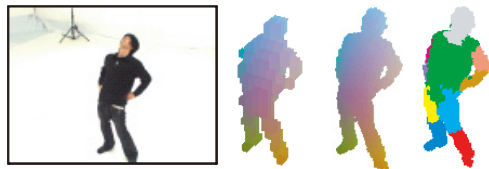


図 10 タイトな着衣における形状解析結果：左から観測画像，低解像度復元結果，高解像度復元結果，形状解析結果
Fig. 10 Results in another dance.

表 4 タイトな着衣における各体節の復元結果の定量的評価
Table 4 Error analysis in another dance.

体節 (正解体積)	誤 (割合/体積)	未 (割合/体積)
頭 (14,008)	4.7/374	5.81/291
胴体 (47,376)	3.2/1,509	3.2/1,478
右上腕 (7,648)	4.9/374	5.6/449
右下腕 (11,216)	7.2/807	6.0/672
左上腕 (7,464)	4.0/301	3.6/266
左下腕 (10,600)	8.3/878	7.1/756
右上脚 (18,392)	4.8/881	5.1/938
右下脚 (11,520)	9.5/1,091	7.6/873
左上脚 (19,648)	6.8/1,327	6.5/1,274
左下脚 (12,560)	11.4/1,429	9.7/1,216

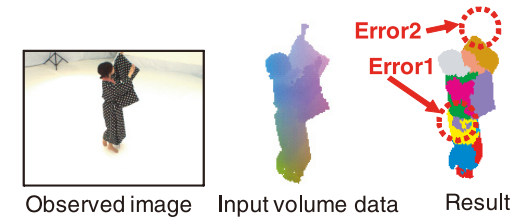


図 11 誤った体節ラベルの割当ての例
Fig. 11 Wrong labeling results.

である HCI や能動カメラ制御なども実現可能であるといえる。よって、先の段落であげた (1)(2) に加えて、(3) オンライン実行という目的も満たせたことを確認できた。以上から、本稿の目的としてあげた 3 条件を満たすことができた。

5.2 考 察

実験結果で示したように、体節境界付近で誤検出・未検出を含むものの多くのフレームでは全体的に正しい体節ラベル割当ておよびボリューム修正が成功している。しかし、図 11 のように体節ラベル割当てを大きく誤っているフレームもいくつかみられた。

図 11 中の Error1 では、右上腕ラベルが右上脚部分に一部割り当てられている。本手法では、各体節の詳細形状解析において、各体節はそれぞれ独立に類似する学習データの探索を行っている。そのため、各体節において類似解として選択された学習データの全身姿勢を比べてみると、大きく異なった全身姿勢から類似解が選択される可能性もある。この場合、選ばれた各体節の形状を組み合わせることにより、学習データ中に含まれない観測対象の全身形状にも対応可能となっている。一方で、各体節がそれぞれ異なる学習データを参照す

ることにより、Error1 のように形状全体として誤った結果が得られてしまう。この誤りは、各体節のバウンディングボックス内に、他の体節を含んでしまっていることにも起因する。今回の実験結果からは確認されなかったが「ある体節 A で選択された学習データ内に他の体節 B の一部分が含まれているが、体節 B ではそれとは大きく異なる姿勢の学習データが選択された」場合、図 11 よりも致命的な問題になる。この問題では、隣接する体節どうし（胴体と上腕など）よりも独立に動く体節どうし（右腕と左腕など）において顕著な誤りを生じてしまう。たとえば、選択された左腕のバウンディングボックス内に右腕が入り込んでいるが、右腕はそれとは大きく異なる姿勢が選ばれた場合、最終的な全身形状中に右腕が 2 つ含まれてしまう。こうした問題に対処するためには、人体に関する制約条件をとり入れ、学習データ生成や体節統合において整合性を確保する必要がある。

また、本稿で示した実験では複雑な舞踊動作などを対象としたが、そのバリエーションはそれほど多くなかった。そのため学習固有空間の次元数を抑えても誤検出・未検出ともに最大 15%以下に収まる解析結果が得られた。しかし、この精度はさらに改善の余地がある。また、図 11 中の Error2 のように体積復元結果では正しく得られている左腕が解析結果では削られてしまうという誤りもみられた。これは、少ない学習データの中から最も類似する解析結果をほぼそのまま参照して形状修正してしまうことが原因であり、学習データと入力データの微小な差に対して敏感すぎるという欠点を生んでいる。実験結果で示したように、より複雑な動きの体節で誤差が大きくなったり、学習データと異なる被験者を観測した際に誤差が大きくなったりしたのも、この問題に起因していると考えられる。これらの問題に対しては、以下の対策が考えられる。

- 提案手法では、ボクセル信頼度によってのみ復元体積の信頼度が評価され、学習・入力データのマッチングに利用されている。これに加えて、多値解析の手法である主成分分

析の特徴を活用するべく、復元体積を 0 と 1 の 2 値で表現するのではなく、復元体積の境界から各ボクセルへの距離や時系列的な形状変化の程度などを考慮し、学習データの記録の時点でボクセルの信頼度を評価する。

- 文献 5) で行われているように、入力ボリュームに合わせてより丁寧に学習ボリュームのサイズ合わせを行う。
- 同じ動きを何度か繰り返し観測し、それらをすべてを多数の学習パターンとして記録しておくことにより、人体・着衣の動きの揺らぎにも対処する。

多数の学習パターンの記録は、提案手法をさらにパリエーション豊かな動作パターンにも対応させ、提案手法の応用範囲を広げることにも可能にする。このように膨大な学習データに対応するためには、学習固有空間の次元数を上げることも有効ではあるが、3 次元形状、特に着物の袖のように複雑かつ急激な変化を表現するためには線形な主線分析の結果では不十分になることが予想される。そこで、陰変数を利用した非線形写像²⁵⁾ などによってボリュームデータをより低次元かつ類似データの識別に優れた表現にできる。手法 25) に対して各パラメータの重みを導入し、2 次元画像上での姿勢推定を行っている手法²⁶⁾ も提案されており、我々の手法にも適用することができれば手法の性能を向上させることができる考えられる。

以上の議論を整理すると、今後の課題として以下のような改良が必要であると考えられる。

- 体節ごとの学習データ探索において、他の体節における探索結果を相互参照する。
- ラベル確率の統合およびラベル決定時に、人体・着衣に関する制約条件（各体節は 1 つの塊である、各体節の体積は一定、時系列的に急激には変化しない、など）を利用する。
- より類似したボリュームデータの識別が可能かつ高速な学習データの学習法を利用する。

6. おわりに

任意の人体・着衣のオンライン形状解析に基づいて、人体・着衣形状中の復元誤りの修正および各体節領域の特定を同時に実行できる手法を提案した。提案手法では、詳細解析により得られる対象の高精度形状の時系列変化を体節ラベル付きで事前に学習し、この正解学習データと入力データとの比較による形状解析を行う。この解析の結果、視体積交差法で得られる対象の 3 次元ボリュームの中から大きな復元誤りを修正し、修正された 3 次元ボリュームの中から定義済の 10 種類の体節領域を得ることができた。

謝辞 本研究の一部は、文部科学省プロジェクト「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」の「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。色検出²⁷⁾、オフライン形状修正¹²⁾には、それぞれ和歌山大学和田俊和教授、京都大学延原章平助手にソフトウェアを提供していただいた。深謝いたします。

参 考 文 献

- 1) Moeslund, T.B. and Granum, E.: A Survey of Computer Vision-Based Human Motion Capture, *CVIU*, Vol.81, No.3, pp.231–268 (2001).
- 2) Poppe, R.: Vision-based human motion analysis: An overview, *CVIU*, Vol.108, No.2, pp.4–18 (2007).
- 3) Cheung, G.K.M., Kanade, T., Bouguet, J.-Y. and Holler, M.: A real time system for robust 3D voxel reconstruction of humanmotions, *CVPR2000*, Vol.2, pp.714–720 (2000).
- 4) Wu, X., Takizawa, O. and Matsuyama, T.: Parallel Pipeline Volume Intersection for Real-Time 3D Shape Reconstruction on a PC Cluster, *Proc. 4th IEEE International Conference on Computer Vision Systems (ICVS)* (2006).
- 5) Mikic, I., Trivedi, M., Hunter, E. and Cosman, P.: Human Body Model Acquisition and Tracking using Voxel Data, *IJCV*, Vol.53, No.3, pp.199–223 (2003).
- 6) Caillette, F. and Howard, T.: Real-Time Markerless Human Body Tracking with Multi-View 3-D Voxel Reconstruction, *BMVC2004*, Vol.2, pp.597–606 (2004).
- 7) de Aguiar, E., et al.: M3: Marker-free Model Reconstruction and Motion Tracking from 3D Voxel Data, *the 12th Pacific Conference on Computer Graphics and Applications*, pp.101–110 (2004).
- 8) Cheung, G., Baker, S. and Kanade, T.: Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture, *CVPR2003*, Vol.1, pp.77–84 (2003).
- 9) Alexandru, O.B., Sigal, L., Black, M.J., Davis, J. and Houssecker, H.W.: Detailed Human Shape and Pose from Images, *CVPR2007*, pp.1–8 (2007).
- 10) 藤田武史, 向川康博, 尺長 健: 多視点カメラシステムによる舞踊動作の獲得と解析, 情報処理学会研究報告, CVIM-2002-132, pp.95–102 (2002).
- 11) Kutulakos, K.N. and Seitz, S.M.: A Theory of Shape by Space Carving, *IJCV*, Vol.38, No.3, pp.199–218 (2000).
- 12) Nobuhara, S. and Matsuyama, T.: Deformable Mesh Model for Complex Multi-Object 3D Motion Estimation from Multi-Viewpoint Video, *3DPVT* (2006).
- 13) Scholz, V., Stich, T., Keckeisen, M., Wacker, M. and Magnor, M.: Garment Motion Capture Using Color-Coded Patterns, *Computer Graphics Forum*, Vol.24, No.3, pp.439–448 (2005).
- 14) Yabuuchi, T., Kakusho, K. and Minoh, M.: Modeling Real Deformable Objects by

- Incremental Observation, *10th International Conference on Virtual Systems and Multimedia (VSMM2004)*, pp.239–248 (2004).
- 15) Bhat, K., Twigg, C., Hodgins, J., Khosla, P., Popovic, Z. and Seitz, S.: Estimating Cloth Simulation Parameters from Video, *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp.37–51 (2003).
- 16) Plankers, R. and Fua, P.: Tracking and Modeling People in Video Sequences, *CVIU*, Vol.81, No.3 (2001).
- 17) Ye, G., Corso, J.J. and Hager, G.D.: Gesture Recognition Using 3D Appearance and Motion Features, *Workshop on Real-time Vision for Human-Computer Interaction* (2004).
- 18) Matsuyama, T., Wu, X., Takai, T. and Nobuhara, S.: Real-Time 3DShape Reconstruction, Dynamic 3D Mesh Deformation, and High Fidelity Visualization for 3D Video, *CVIU*, Vol.96, No.3, pp.393–434 (2004).
- 19) Yamasaki, T. and Aizawa, K.: Temporal 3D Video Segmentation Using Modified Shape Distribution, *IEEE International Conference on Multimedia and Expo*, pp.1909–1912 (2006).
- 20) Herda, L., Urtasun, R. and Fua, P.: Hierarchical implicit surface joint limits for human body tracking, *CVIU*, Vol.99, No.2, pp.189–209 (2005).
- 21) Sidenbladh, H., Black, M.J. and Sigal, L.: Implicit Probabilistic Models of Human Motion for Synthesis and Tracking, *ECCV2002*, Vol.1, pp.784–800 (2002).
- 22) Agarwal, A. and Triggs, B.: Tracking Articulated Motion using a Mixture of Autoregressive Models, *ECCV2004*, Vol.3, pp.54–65 (2004).
- 23) Grauman, K., Shakhnarovich, G. and Darrell, T.: Inferring 3D Structure with a Statistical Image-Based Shape Model, *ICCV2003*, pp.641–648 (2003).
- 24) Murase, H. and Nayar, S.K.: Visual learning and recognition of 3-D objects from appearance, *IJCV*, Vol.14, pp.5–24 (1995).
- 25) Lawrence, N.D.: Probabilistic non-linear principal component analysis with Gaussian process latent variable models, *Journal of Machine Learning Research*, Vol.6, pp.1783–1816 (2005).
- 26) Grochow, K., Martin, S.L., Hertzmann, A. and Popovic, Z.: Style-based Inverse Kinematics, *ACM Trans. Graphics*, Vol.23, No.3, pp.522–531 (2004).
- 27) 和田俊和：最近傍識別器を用いた色ターゲット検出, *情報処理学会論文誌：コンピュータビジョンとイメージメディア*, Vol.44, No.SIG17, pp.126–135 (2003).

(平成 19 年 9 月 22 日受付)

(平成 20 年 3 月 10 日採録)

(担当編集委員 日浦 慎作)



学会論文賞 .

浮田 宗伯 (正会員)

2001 年京都大学大学院博士後期課程修了 . 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 年同准教授 . 2002 ~ 2006 年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 . 現在, 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, 対象追跡に関する研究に従事 . 1999 年電子情報通信



辻 良介

2005 年同志社大学工学部卒業 . 2007 年奈良先端科学技術大学院大学情報科学研究科修士課程修了 . 現在, (株) キヤノン勤務 . 在学中, 人体の姿勢・形状解析の研究に従事 .



木戸出正継 (フェロー)

1970 年京都大学大学院工学研究科修士課程修了 . 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 年奈良先端科学技術大学院大学情報科学研究科教授 . 京都大学工学博士 . パターン認識, 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 . 電子情報通信学会フェロー, IEEE フェロー, IAPR (国際パターン認識協会) フェロー, 電子情報通信学会業績賞, 高柳記念奨励賞等を受賞 . 情報処理学会関西支部長, 電子情報通信学会理事, MVA 国際ワークショップ組織委員長, 電子情報通信学会情報システムソサイエティ会長等を歴任 .