情報処理学会論文誌　Vol. 51　No. 4　1238–1247 (Apr. 2010)

確率的な動的モデルと視体積制約に基づく複雑形状からの姿勢推定

平井迪郎[†1]　浮田宗伯[†1]　木戸出正継[†1]

本稿では同期ビデオ群より獲得した人体の 3 次元形状から姿勢推定を行う手法を提案する．本手法は，1）着衣の非剛体・剛体などを選ばず複雑な形状変化に対応可能，2）複雑形状の復元結果に含まれる推定誤差を修正した頑健な姿勢推定が可能，3）従来法と比較して高速処理可能，という特長を持つ．提案法では，オンライン復元形状を誤差修正済みの形状変化の動的モデルと比較することで誤差修正を行う．この際，視体積領域の持つ幾何学的な制約を導入することで復元誤差に頑健な形状推定を可能とする．最後に，各時刻の形状と姿勢（関節角度）を対応付けて学習し，復元形状から姿勢推定を行う．形状変化が複雑で復元誤差の多い非剛体着衣を対象として実験を行い，提案手法が従来法と比べて高精度な姿勢推定ができることを示す．

Complex Pose Tracking with Probabilistic Dynamical Volume Models and Visual Hull Constraints

Michiro Hirai,[†1] Norimichi Ukita[†1]
and Masatsugu Kidode[†1]

We propose a method for estimating the pose of a human body using its 3D volume obtained from synchronized videos. Our method can cope with complex shape variations of loose-fitting clothing, which produce non-rigid motions and critical reconstruction errors including phantom volumes in a visual hull. To this end, the probabilistic dynamical model of human volumes is learned from training temporal volumes refined by error correction. The dynamical model of a body pose (joint angles) is also learned with its corresponding volume. By comparing the volume model with an input visual hull and regressing its pose from the pose model, the input volume is refined online and then pose estimation can be realized. Comparative experiments demonstrated that our method is superior to similar methods in pose estimation of a human body with loose-fitting clothing.

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．これらの研究により，ロボットの運動制御[1)]，CG アニメーション[2)] など，様々な応用が実現可能となる．一般に用いられることの多いモーションキャプチャによって取得される姿勢情報は高精度であるが，装置を身に付けていなければならず，利用できる場面は限られる．これに対して，画像情報に基づくモーションキャプチャシステムはマーカを人体に取り付ける必要がなく，ヒューマンコンピュータインタラクション[3)] やサーベイランスシステム[4)] などの様々な応用の拡大が見込めるため，種々の研究が行われている[5)]．

特に，近年では Shape-From-Silhouette（SFS）により，形状復元が実時間で可能となり[6),7)]，3 次元形状に基づく姿勢・運動解析が望まれている．3 次元形状の利用により，従来の 2 次元画像からの姿勢推定では達成が困難な複雑な姿勢・運動解析が可能となる．

一般的な 3 次元形状からの姿勢推定（文献 8)，9) など）では，剛体近似された 3 次元人体モデルと復元形状を比較し，各剛体パーツの重なりが最大になるパラメータを求める．また，対象が非剛体着衣（着物，スカートなど）の場合には対象の剛体近似が難しくなる．このため，こういった観測対象では剛体近似によって姿勢推定を行うことは容易ではない．

これに対して，文献 10) では，形状特徴量から姿勢への写像を回帰により学習し，復元形状から姿勢推定を実現している．このような回帰学習ベースの姿勢推定は対象を選ばず，緩い着衣にも適用することが可能である．

文献 10) を含めほとんどの類似手法では人体形状を SFS で復元している．SFS は高速で安定に形状復元が可能であるが，その原理上，凹領域に大きな復元誤差（**図 1** の Reconstruciton errors）を含むことがある．この誤差は，四肢が重なり合う状態などの複雑な姿勢や着物などの体を隠す着衣において大きくなる．誤差が大きくなると，その時々で同じ姿勢に対応する形状が変化してしまったり，同じ姿勢であっても形状が異なったりするという曖昧さが生じるため，回帰による姿勢推定が難しくなる．このように大きな誤差を含んだ**視体積**（**visual hull**）と呼ばれる復元形状から姿勢推定を行うことは簡単ではない．本稿では，この復元誤差をファントムボリュームと呼ぶ．ファントムボリュームの発生箇所は対象形状だけでなく，撮影対象とカメラ間の位置関係，前処理（カメラキャリブレーション，シ

†1 奈良先端科学技術大学院大学
Nara Institute of Science and Technology

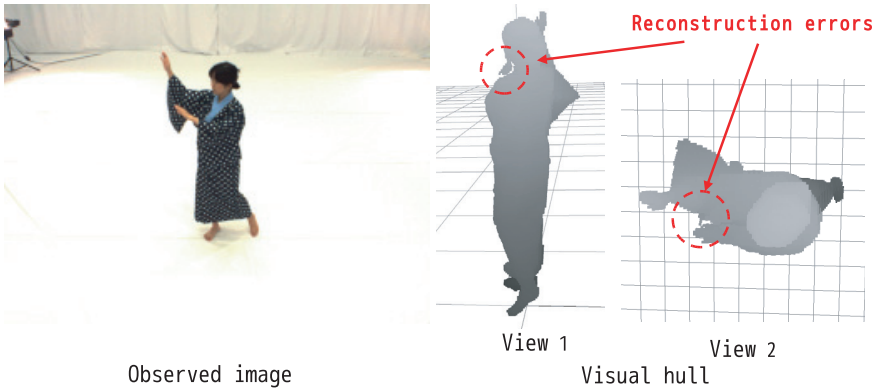

図 1　Shape-From-Silhouette による復元誤差の例
Fig. 1　Reconstruction errors by Shape-From-Silhouette.

ルエット抽出）によっても変化するため，推定することが困難である．つまり，学習ベースの手法であっても，時々刻々と変化し，様々な位置にファントムボリュームを含む視体積から着衣に隠れた姿勢を推定することは難しい問題である．

ファントムボリュームなどのエラーの少ない形状復元法として，異なる視点間での対応画素の色の同一性を用いる space carving [11),12)]，色の同一性に加えて，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，表面形状の動的変形を行う弾性メッシュモデル[13)] などが提案されている．しかしながら，いずれの手法においても計算コストの問題から視体積にオンラインで適用することができない（1 min/frame 以上）．

回帰ベース以外の手法でも，緩い着衣に対応可能な手法はいくつか提案されている．文献 14)，15) では，変化着衣形状と人体姿勢を同時に復元することに成功している．しかし，詳細な形状まで得られる一方で長い計算時間（5 min/frame [14)]，16 sec/frame [15)]）を要する．また，事前に詳細な着衣の物理パラメータや観測対象人物そのものの詳細形状を獲得しなければならないという点で汎用性にやや欠けるという問題もある．文献 16) では，非剛体着衣の 3 次元形状から体節の特定を行う手法を提案している．着衣の物理パラメータなどの取得困難なデータを必要としないため上記手法よりも汎用性は高い一方，姿勢そのものを推定することはできない．

以上の議論から，非剛体着衣の複雑動作から高速な姿勢推定を行っている関連研究は存在しないといえる．整理すると，このような姿勢推定実現のため，1）3 次元形状からの姿勢回帰により人体形状の剛体近似仮定にとらわれず任意の着衣に対して適用可能なオンライン手法を実現できる可能性があるが，2）実際には姿勢と形状の間には曖昧性が存在し，高速処理による復元形状には誤差が含まれているため（特に SFS を形状復元に使用すると凹領域の誤差は理論上不可避），オンライン処理可能な速度で正しい形状推定を実現することは難しい．言い換えると，回帰ベースの手法では，入力（3 次元形状）と出力（姿勢）が 1 対 1 で対応し，入力がノイズなしで得られていれば，つねに正しい姿勢を得ることができるため，「誤差が小さく，人体姿勢を曖昧性なく表現できる精度（解像度）で人体形状を高速復元する」ことが技術的な課題となる．

そこで本稿では，1）複雑に変化する対象形状から回帰により精度良く姿勢推定を行うこと，2）オンラインで復元された形状に含まれる誤差を逐次的に推定し，姿勢推定の精度を向上させること，を目的とする．この目的のため，復元形状を回帰に利用しやすい表現に変換し（4 章），学習形状モデルとの比較により復元形状中に含まれる誤差を修正したのち（5，6 章），修正形状から姿勢回帰を行う（7 章）．

2. 関 連 研 究

撮影画像から人体の姿勢を推定する手法において，形状（シルエットやボリューム）を表現する適切な特徴量，すなわち観測画像からの抽出誤差が小さく，高い姿勢表現精度を備え，かつ高速処理に必要な効率性の高い特徴量を画像から抽出することが重要な課題となっている．たとえば，文献 17) ではシルエット特徴量として shape context [18)] を抽出し，Relevance Vector Machine（RVM）による回帰でシルエットから姿勢の写像をモデル化している．また，これを 3 次元に拡張した 3D shape context があり，文献 10) では 3D shape context をボリューム表現に適した形に改良し，Multivariate RVM [19)] によってボリュームから姿勢への写像をモデル化している．このような 3 次元形状表現は文献 20) などの 3 次元モデル検索の分野でも多数提案されている．これらの特徴量は効率的に形状を表現することができ，ノイズにも頑健であるといった特長がある．我々の手法においてもこのような 3 次元形状モデルを利用し，**volume descriptor** と呼ぶ．

近年の研究においては，姿勢推定の高精度化・頑健化のために観測対象の運動情報が事前知識（motion prior）としてよく利用される．時系列情報の利用により一時的な遮蔽などの曖昧性の解消が可能となる．しかし，motion prior を利用する場合においても，復元誤差を含み，姿勢内分散を持つ対象形状の変化を効果的にモデル化することは容易ではない．

そこで，motion prior に内在する低次元な特徴を獲得してノイズに対して頑健かつ効率

的な姿勢推定を実現するため，主成分分析（PCA）のような次元削減が用いられる．さらに，Gaussian Process Latent Variable Models（GPLVM）[21] などの非線形で確率的な次元削減法が登場し，より効果的な運動学習に用いられている[9),22)]．

また，安定な追跡により形状・姿勢間の曖昧性低減を実現するためにパーティクルフィルタが広く利用されており，関節（30〜60 次元程度）の動きをそのまま追跡する手法も提案されている[24)]．しかし，高次元な（100 次元以上）ボリューム特徴量からの姿勢推定ではボリュームに対応する高次元パーティクルをうまく遷移させることが難しく，追跡に失敗しやすい．このため，前述の低次元化との組合せが有効な手段となる．特に，GPLVM の時間拡張である Gaussian Process Dynamical Models（GPDM）[23] は低次元空間上での状態遷移を得ることができるため，この状態遷移情報を追跡処理に組み込むことで効率的な追跡が可能である[25)]．

3. 処 理 概 要

我々の提案手法の概要をまとめる．

オフライン学習　まず，同期された対象のボリュームと姿勢の関係，さらに motion prior を学習する．文献 16) と同様に，視体積ではなく文献 12) の手法で得られる誤差の少ない**修正形状**（**refined volume**）を生成し，学習に用いる．修正形状から姿勢への写像を学習することにより，回帰の結果が復元誤差の影響を受けなくなるといった利点がある．ボリュームからの特徴抽出には文献 10)，20)，26) で提案されている volume descriptor を利用する（4 章）．また，GPDM により修正形状の確率的な動的モデル（**図 2** 左下）を得る（5 章）．このモデルより，ボリュームの低次元特徴である潜在空間（X）とボリューム空間（V）の双方向の写像（$f(x) : X \to V$，$f^{-1}(v) : V \to X$）が得られる．さらに，姿勢も PCA によって低次元化を行い，ガウス過程（Gaussian Process，GP）[27] を用いた回帰によってボリュームの潜在空間から姿勢の潜在空間への写像をモデル化する．

オンライン姿勢推定　オンラインにおける処理は，形状追跡（6 章）とボリュームから姿勢への回帰（7 章）で構成される．形状追跡は潜在空間中の形状潜在変数を対象にしたパーティクルフィルタリングによって行われる．このパーティクルは写像 $f(x)$ によって修正形状の volume descriptor（図 2 の Refined volume descriptor）に変換できる．パーティクルは X 上のあらゆる場所に遷移する可能性があり，モデル生成に利用した学習サンプルと類似したあらゆる姿勢の推定を可能にする．また，各時刻において SFS によって得られた視体積（図 2 右上の Visual hull）をその volume descriptor へ変換する．単純なパーティクルフィルタでは，この volume descriptor を $f^{-1}(v)$ によって X に写像し，各パーティクルとの尤度を計算する．この際，誤差の含まれた視体積と修正形状を比較して尤度を算出するのだが，3 次元空間中での視体積と修正形状の関係が分からないため，X 上で正しい尤度を得ることは不可能である．ここで，修正形状は視体積に内包されるという幾何学的な制約を持つ．我々はこの制約を**視体積制約**と呼ぶ．この視体積制約をパーティクルフィルタに適用することによってファントムボリュームの影響を軽減することができる．つまり，volume descriptor 空間（V）中で視体積とモデルから推定される修正形状（$f(x)$）を比較し，視体積の修正を行う．さらに，すべてのパーティクルによって修正された視体積（図 2 の Estimated volume descriptor）を X へ投影し，尤度を計算する（図 2 の Likelihood evaluation）．最終的に尤度の重み付き平均により推定された $\bar{\boldsymbol{x}}$ から，GP により学習した写像を用いて姿勢を推定する．

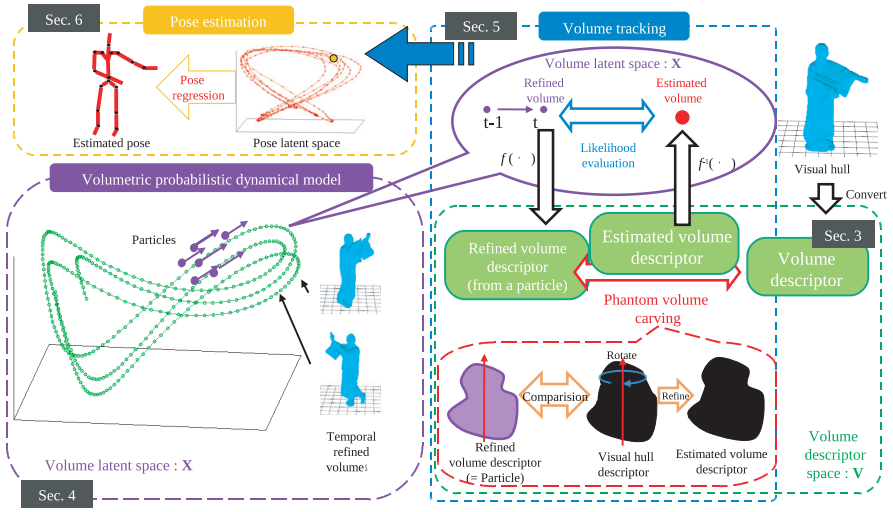

図 2　処理全体の流れ．Sec. は対応する章を指す
Fig. 2　Process flows. The process described in each section is indicated.

4. ボリュームデータの表現

ボリュームデータからの特徴抽出では，視体積制約を利用したファントムボリュームの推

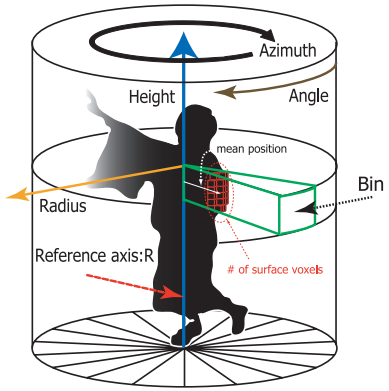

図 3　Bin モデルを用いたボリュームからの特徴抽出
Fig. 3　Volume descriptor with a Bin model.

定を実現するために，volume descriptor を利用する．

Bin の決定　**図 3** のように対象の基準軸（$\boldsymbol{R}$）中心の円柱領域を設定する．高さ方向を n_z 個，回転方向を n_θ 個に分割する．各分割領域を Bin と呼び，この集合を $\boldsymbol{B} = [b_1, \cdots, b_{N_{bin}}]$ とする（$N_{bin} = n_z n_\theta$）．

Volume descriptor への変換　各 Bin に存在する表面ボクセルの基準軸からの “位置” の平均値（$r_{i \in B}$）と各 Bin に存在する表面ボクセルの “個数”（$n_{i \in B}$）を記録する．各 Bin に記録する値は 2 つとなり，volume descriptor は $D = 2\,N_{bin}$ 次元の特徴量となる．

円柱領域は対象の身長 H に応じて決定し，高さを $1.23H$，半径を $0.6H$ とした．このため，スケール不変特徴量となる．Bin 内での平均を用いるために細かいノイズに頑健である．また，対象形状表面の位置を記録することで，対象表面に発生するファントムボリュームの有無・大小を評価可能なままデータ次元数を削減できている．ファントムボリュームの評価は，視体積制約を導入した提案手法を実現するうえで重要な特徴である．

5. 形状変化の確率的な動的モデル

GPDM を用いて形状変化のモデル化を行う．GPDM は 1）潜在空間における時刻 $t-1$ から t への滑らかな写像，2）潜在空間から形状空間への確率的な写像，をもたらす．特に，1）の写像は形状追跡に非常に有用である．この 2 つの写像は以下の式でモデル化される．

$$\boldsymbol{x}_t = \sum_i \boldsymbol{a}_i \phi_i(\boldsymbol{x}_{t-1}) + n_{x,t} \tag{1}$$

$$\boldsymbol{v}_t = \sum_j \boldsymbol{b}_j \psi_j(\boldsymbol{x}_t) + n_{v,t} \tag{2}$$

だだし，時刻 t において，平均を 0 とした volume descriptor を $\boldsymbol{v}_t \in \Re^D$，潜在変数を $\boldsymbol{x}_t \in \Re^d$（$d \ll D$）とし，$\phi_i$，$\psi_k$ は基底関数で $\boldsymbol{\mathcal{A}} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots]$，$\boldsymbol{\mathcal{B}} = [\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots]$ で表せる重みを持つ，また，$n_{x,t}$，$n_{v,t}$ で表されるノイズが平均 0 の正規分布に従うと仮定すると式 (2) の基底関数は周辺化することができ，以下の確率分布を得る．

$$p(\boldsymbol{V}|\boldsymbol{X}, \alpha) = \frac{1}{\sqrt{(2\pi)^{ND}|\boldsymbol{K}_V|^D}} \exp\left(-\frac{1}{2} tr(\boldsymbol{K}_V^{-1}\boldsymbol{V}\boldsymbol{V}^T)\right) \tag{3}$$

ただし，$\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N]^T$，$\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]^T$ である．$\boldsymbol{K}_V$ は各要素が $(\boldsymbol{K}_V)_{i,j} = k_V(\boldsymbol{x}_i, \boldsymbol{x}_j)$ となるカーネル行列でありハイパーパラメータ α を持つ．本稿では，非線形な動径基底関数を用いた．同様に，式 (1) も，

$$p(\boldsymbol{X}|\beta) = \frac{p(\boldsymbol{x}_1)}{\sqrt{(2\pi)^{(N-1)d}|\boldsymbol{K}_X|^d}} \exp\left(-\frac{1}{2} tr(\boldsymbol{K}_X^{-1}\boldsymbol{X}_{out}\boldsymbol{X}_{out}^T)\right) \tag{4}$$

となる．また，$\boldsymbol{X}_{out} = [\boldsymbol{x}_2, \ldots, \boldsymbol{x}_N]^T$ であり，$\boldsymbol{K}_X$ は $(N-1) \times (N-1)$ のカーネル行列でハイパーパラメータ β を持ち，$\boldsymbol{X}_{in} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{(N-1)}]$ から構成される．

学習　文献 23) に従いハイパーパラメータの事前分布を導入すると，GPDM の事後確率分布は以下のようになる．

$$p(\boldsymbol{X}, \alpha, \beta|\boldsymbol{V}) \propto p(\boldsymbol{V}|\boldsymbol{X}, \alpha) p(\boldsymbol{X}|\beta) p(\alpha) p(\beta) \tag{5}$$

これらの事後確率を最大化する潜在変数 $\boldsymbol{X}$ とハイパーパラメータ α，β は，以下の負の対数事後確率 $\mathcal{L}$ を最小化することで求められる．

$$\begin{aligned}\mathcal{L} = &\frac{d}{2}\ln|\boldsymbol{K}_X| + \frac{1}{2} tr(\boldsymbol{K}_X^{-1}\boldsymbol{X}_{out}\boldsymbol{X}_{out}^T) \\ &+ \frac{D}{2}\ln|\boldsymbol{K}_V| + \frac{1}{2} tr(\boldsymbol{K}_V^{-1}\boldsymbol{V}\boldsymbol{V}^T) + \sum_i \ln\alpha_i + \sum_i \ln\beta_i + \mathcal{C}^l\end{aligned} \tag{6}$$

ここで，$\mathcal{C}^l$ は定数を表す．以上の処理により，ボリュームデータの系列から**図 4** に示すような潜在空間を得る．ここでは可視化のため $d = 3$ とした．

予測　学習後，$\boldsymbol{v}_t^{(*)}$ の平均 $\mu_V(\boldsymbol{x}_t^{(*)})$ と分散 $\sigma_V^2(\boldsymbol{x}_t^{(*)})$，$\boldsymbol{x}_t^{(*)}$ の平均 $\mu_X(\boldsymbol{x}_{t-1}^{(*)})$ は以下で与えられる．

$$\mu_V(\boldsymbol{x}_t) = \bar{\mu_v} + \boldsymbol{V}^T\boldsymbol{K}_V^{-1}k(\boldsymbol{x}_t) \tag{7}$$

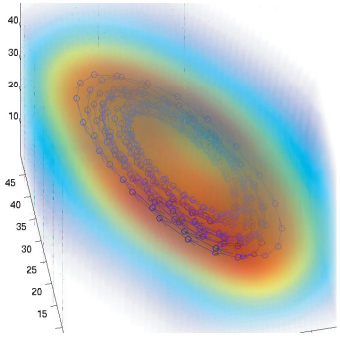

図 4　形状変化の動的モデル．色は各座標における分散の大小を表し（色：暖色ほど分散小），矢印は状態遷移を表す
Fig. 4　Dynamical model of a shape variation. Circles and arrows depict latent variables and temporal mapping, respectively. Background colors denote the variance at each point; lower (red) to higher (blue).

$$\sigma_V^2(\boldsymbol{x}_t) = k(\boldsymbol{x}_t, \boldsymbol{x}_t) - k(\boldsymbol{x}_t)^T \boldsymbol{K}_V^{-1} k(\boldsymbol{x}_t) \tag{8}$$

$$\mu_X(\boldsymbol{x}_{t-1}) = \bar{\mu_x} + \boldsymbol{X}_{out}^T \boldsymbol{K}_X^{-1} k(\boldsymbol{x}_{t-1}) \tag{9}$$

また，$\bar{\mu_v}$，$\bar{\mu_x}$ は $\boldsymbol{V}$，$\boldsymbol{X}$ の平均値，$\boldsymbol{I}$ は単位行列を表す．つまり，$\mu_V(\boldsymbol{x})$ によって $f(\boldsymbol{x}) : X \to V$，$\mu_X(\boldsymbol{x})$ によって x_{t-1} から x_t への状態遷移を定義する．

入力された視体積を潜在空間へ投影する必要がある．しかしながら，GPDM では入力空間から潜在空間への逆写像（$V \to X$）が明確に定義できない．一般には $\arg\min_{\boldsymbol{x}} \mathcal{L}$ を満たす $\boldsymbol{x}$ を求めればよいのだが，これをオンラインで行うと非常に低速である．また，逆写像を同時に学習することのできる Back-constrained GPLVM [28] も存在するが，扱う変数が増えるため最適化が難しくなり，滑らかな潜在空間を得ることができない．そこで，サンプルボリューム $\boldsymbol{V}$ と最適化後の潜在変数 $\boldsymbol{X}$ に関して GP を適用し，回帰によって $f^{-1}(\boldsymbol{v}) : V \to X$ の写像をモデル化する．

6. 視体積制約と動的形状モデルを利用した形状の高速追跡

回帰による姿勢推定の高精度化のためには，視体積に含まれる誤差の修正が必要である．本手法では，各時刻の入力視体積の誤差修正を行いながら，GPDM で得られたボリュームの潜在空間上で形状潜在変数を対象にしたパーティクルフィルタを行う．パーティクルフィルタによる形状追跡内の各処理を以下に述べる．

対象方位の推定　学習時には実空間上のある方向を基準として対象の方位方向 ϕ が定まっていた．このため，学習データと同じ方位の各パーティクル（$\boldsymbol{x}_i^p | i \in 1, \cdots, N^p$，$N^p$ はパーティクル数）と入力視体積の方位を一致させる必要がある．視体積を ϕ 回転させた volume descriptor $\boldsymbol{v}^\phi$ から以下の重み付き正規化相互相関が最大となる $\hat{\phi}$ を決定し，対象の方位方向を推定する．

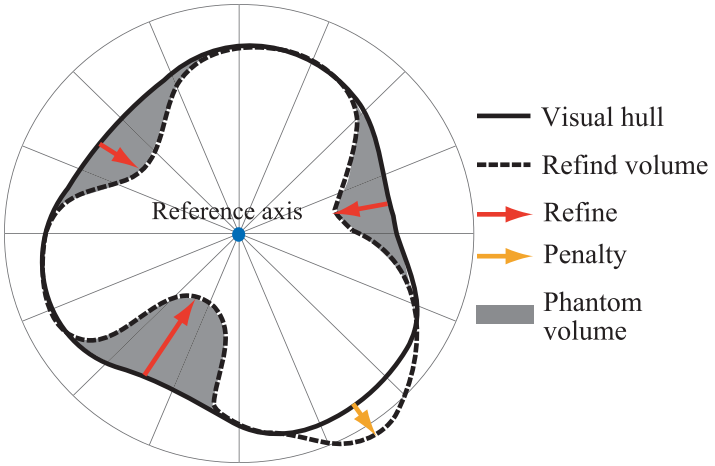

図 5　視体積制約を用いた形状の修正とペナルティ
Fig. 5　Volume refinement and penalty in visual hull constraints.

$$\hat{\phi} = \max_{\phi \in [\phi_{t-1}-m, \phi_{t-1}+m]} \sum_{i=1}^{N_p} c(\boldsymbol{x}_i^p) \frac{\boldsymbol{v}^\phi \cdot f(\boldsymbol{x}_i^p)}{\sqrt{(\boldsymbol{v}^\phi)^2 (f(\boldsymbol{x}_i^p))^2}} \tag{10}$$

$$c(\boldsymbol{x}) = \exp(-\sigma_V^2(\boldsymbol{x})/w) \tag{11}$$

$c(\boldsymbol{x}_i^p)$，ϕ_{t-1} はそれぞれ各パーティクルの信頼度，1 時刻前の対象方位を表し，対象方位の探索幅（Bin の方位分割方向の探索数）は $m = 3$ とした．w は重みであり，潜在変数 $\boldsymbol{x}$ の分散に応じて決定する．

視体積制約　入力された視体積と修正形状のマッチングの頑健性向上のため，以下に述べる視体積制約を用いる．ファントムボリュームが基準軸から見て表面形状の外側に生じると仮定し，ある Bin において，1）修正形状が視体積の内側に存在する場合にはその差分領域がファントムボリュームとなり，視体積の修正を行う（**図 5** の赤矢印）．逆に，2）視体積が修正形状の内側に存在する場合にはペナルティを与える（図 5 の緑矢印），

という処理を行う．

しかし，パーティクルフィルタは潜在空間 X 上で行われるため，そのままでは視体積制約は評価できない．そこで，すべてのパーティクルを $f(\boldsymbol{x}_i^p) : X \to V$ によって volume descriptor へ変換する．そして，b 番目の Bin における表面ボクセルの平均値（視体積では r_b^{vh}，修正形状では r_b^{rv}）を以下の手順で評価する．

1）視体積の誤差修正　$r_b^{rv} \leq r_b^{vh}$ の場合には，視体積から計算される volume descriptor（$\boldsymbol{v}$）のファントムボリュームを削る．これは r_b^{vh} を r_b^{rv} に近づけることで達成できるが，それぞれの形状の信頼度 $c(\boldsymbol{x})$ を考慮に入れずに r_b^{vh} をそのまま r_b^{rv} へ移動させることはボリュームを削りすぎてしまう危険がある．そこで，修正後の r_b^{vh}（$\hat{r}_b^{rv}$ で表す）は X における $\boldsymbol{v}$ と $f(\boldsymbol{x}_i^p)$ の信頼度 $c(\boldsymbol{x})$ を考慮に入れて以下のように計算する．

$$\hat{r}_b^{rv} = r_b^{vh} - (r_b^{vh} - r_b^{rv})(1 - c(f^{-1}(\boldsymbol{v})))c(\boldsymbol{x}_i^p) \tag{12}$$

$f^{-1}(\boldsymbol{v})$ は 5 章でモデル済みの写像 $V \to X$ である．形状がもっともらしいほど分散が小さいという GPDM によって得られたモデルの分散 $\sigma_V^2(\boldsymbol{x})$（式 (11) 中）の定義から，$c(\boldsymbol{x})$ は X から V への写像が確からしいほど 1 に近づく．この式 (12) により，修正形状値 r_b^{rv} が内側にあるほど，その修正形状の信頼度 $c(\boldsymbol{x}_i^p)$ が大きいほど，逆に入力形状の信頼度 $c(f^{-1}(\boldsymbol{v}))$ が小さいほど，入力形状値が小さくなる（誤差修正により形状が内側に削られる）．

2）パーティクルへのペナルティ　$f(\boldsymbol{x}_i^p)$ が視体積制約を満たさない場合（$r_b^{rv} > r_b^{vh}$）には，そのパーティクル $\boldsymbol{x}_i^p$ にペナルティを与える．ペナルティ $\mathcal{P}$ は r_b^{rv} と r_b^{vh} の距離に応じて決めることができて，

$$\mathcal{P} = \sum_{b=1}^{N_z N_\theta} e(b) + \mathcal{C}^p \tag{13}$$

と定義する．$e(b) = \max(r_b^{rv} - r_b^{vh}, 0)$ で，$\mathcal{C}^p$ は定数である．

視体積制約を用いた尤度計算　入力視体積に $\boldsymbol{x}_i^p$ を参照した修正とペナルティを適用し，$\hat{\boldsymbol{v}}_i$ を得る．これら $\hat{\boldsymbol{v}}_i$ を潜在空間に再射影した $\hat{\boldsymbol{x}}_i$ の尤度重み付き平均 $\bar{\boldsymbol{x}}$ を「視体積からファントムボリュームを削って得られる推定形状」に相当する潜在変数と見なす．

尤度の指標として，推定形状とモデルから推定される修正形状が近ければ尤度は高くなるので，1）潜在空間 X での $\hat{\boldsymbol{x}}_i$ と $\boldsymbol{x}_i^p$ の距離，2）パーティクルの信頼度 $c(\boldsymbol{x}_i^p)$，3）$\mathcal{P}_i$ によるペナルティ，を用いる．2）は GPDM で学習したモデルとどれほど近いかが評価できるため重要である．以上より，尤度関数を以下のように設定する．

$$p(\hat{\boldsymbol{x}}_i|\boldsymbol{x}_i^p) \propto \exp\left(-\frac{|\hat{\boldsymbol{x}}_i - \boldsymbol{x}_i^p|^2}{\nu}\right) c(\boldsymbol{x}_i^p)\mathcal{P}_i^{-1} \tag{14}$$

ν は重みパラメータである．

形状追跡の流れ　以上から形状追跡の流れをまとめる．

（1）　式 (9) に基づいてパーティクルの遷移を行う．

（2）　SFS によってシルエット群から形状復元を行い，視体積を生成する．

（3）　すべてのパーティクルを式 (7) によって X から V へ投影し，$f(\boldsymbol{x}_{i\in N^p}^p)$ を得る．

（4）　視体積の回転方位 $\hat{\phi}$ を推定し，基準軸を中心に $\hat{\phi}$ 回転させ，volume descriptor（$\boldsymbol{v}^{\hat{\phi}}$）へ変換する．

（5）　$f(x_{i\in N^p}^p)$ と $v^{\hat{\phi}}$ の比較により式 (12) の修正，式 (13) のペナルティを評価し．推定形状 $\hat{\boldsymbol{v}}_{i\in N^p}$ を得る．さらに，$f^{-1}(\hat{\boldsymbol{v}}_{i\in N^p})$ により X に投影し，$\hat{\boldsymbol{x}}_{i\in N^p}$ を得る．

（6）　式 (14) より各パーティクルの尤度を計算し，$\hat{\boldsymbol{x}}_{i\in N^p}$ の重み付き平均が $\bar{\boldsymbol{x}}$ となる．

以上の処理により提案する形状追跡が実現できる．

7. 推定形状からの姿勢推定

オフライン処理では，ボリュームデータと同期された姿勢を PCA によって低次元化し，**図 6** に示すように，ボリュームの潜在空間から姿勢の潜在空間への写像を GP で学習する[⋆1]．

オンライン処理では 6 章の追跡結果で得られる推定形状の潜在変数 $\bar{\boldsymbol{x}}$ を姿勢空間の潜在空間へ投影（図 6 の右向き赤矢印）して，姿勢の低次元特徴量 $\boldsymbol{x}^{\bar{J}}$ を得る．最後に，$\boldsymbol{x}^{\bar{J}}$ を姿勢空間へ逆投影（図 6 の上向き赤矢印）して，姿勢（関節角集合）$\bar{\boldsymbol{J}} = [\hat{\theta}_1, \cdots, \hat{\theta}_n]$ を得る．

8. 実　　験

提案手法の有効性を確認するため，形状変化の大きな着物（図 1）を身に付けた人物を観測対象として実験を行い，姿勢推定を行った．

実験では，観測対象を囲むように天井に設置された 8 台のカメラ（Pointgrey 社 Flea：1,024 × 768 pixel，8 bit bayer pattern）により同期撮影を行い，動作は舞踊動作 1，舞

⋆1 関節角度 θ での回帰を行うと 360 度付近の連続性が満たされないため，$(x, y) = (\cos\theta, \sin\theta)$ を回帰に用い，$\tan^{-1}(y/x)$ によって θ を復元した．

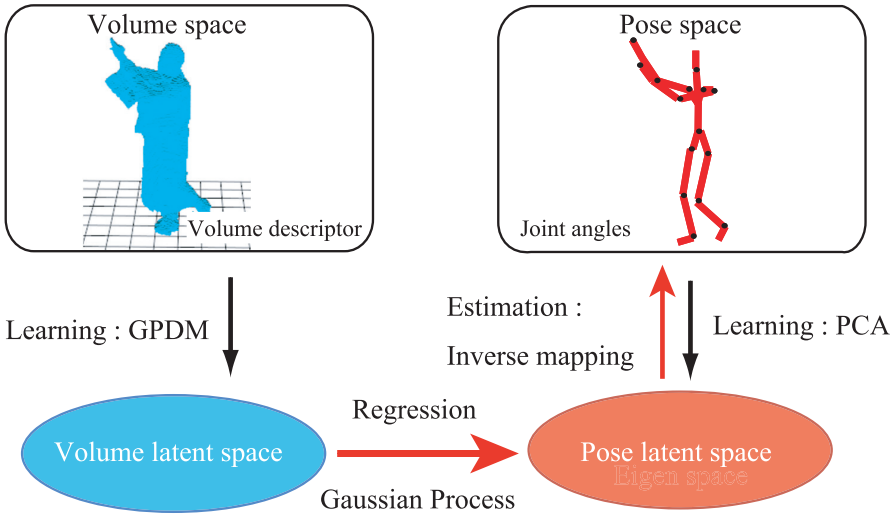

図 6　姿勢推定の流れ
Fig. 6　Process flows in pose estimation.

踊動作 2 の 2 つを選定した．カメラの内部・外部キャリブレーションは事前に行われている．各ボリュームのボクセルサイズは 10^3 [mm] を用いた．また，着物内部の姿勢情報を同時に得るために磁気式のモーションキャプチャシステム（Spice 社 http://www.mocap.jp, IGS-190）を用いて姿勢情報を取得した．画像と姿勢は同期されており，それぞれ 30 [fps] で取得している．人体モデルは 18 関節，5 端点の 54 自由度（関節角）を用いた（図 6 右上）．

オフライン学習では，グラフカットによる形状復元[12] でボリュームデータを生成し，volume descriptor は分割数 $[n_\theta, n_z] = [16, 5]$ で $2 \times 16 \times 5 = 160$ 次元を用いた．学習にはそれぞれ 350 フレームを用い，GPDM における潜在空間 X の次元数 d は経験的に 6 とした．姿勢の低次元化，比較実験におけるボリュームの潜在空間生成については PCA で行い，累積寄与率 $c = (\sum_{j=1}^{p} \lambda_j)/(\sum_{j=1}^{s} \lambda_j) \geq 0.95$ を満たす p 次元を用いた．姿勢の潜在空間は舞踊 1，舞踊 2，ともに 4 次元であった．

オンライン姿勢推定では学習データとは別のシーケンスの視体積を入力とし，以下 5 種の手法の間で比較実験を行った．

1. Direct detection　低次元化を行わず，視体積から直接回帰を行う．

2./3. PCA/GPDM detection　視体積を PCA/GPDM によって生成された潜在空間に写像し，得られた潜在変数から回帰を行う．

4. GPDM tracking　提案手法と同様だが，幾何学制約を用いない．追跡結果から得られる潜在変数から回帰を行う．

5. Proposed method　256 個のパーティクルを用いて追跡を行う．式 (11) において視体積では $w = 1$，パーティクルでは $w = 10$ を用いた．

表 1　姿勢推定の平均誤差 [deg]
Table 1　RMS errors of estimated joint angles (degrees).

	8 camera		4 camera	
	舞踊 1	舞踊 2	舞踊 1	舞踊 2
Direct regression	6.13	10.21	13.77	13.81
PCA detection	6.34	12.77	14.71	16.70
GPDM detection	5.81	7.69	6.72	8.03
GPDM tracking	5.39	6.46	6.37	6.46
Proposed method	**5.04**	**5.03**	**5.59**	**5.54**

推定された姿勢 $\bar{\boldsymbol{J}}$ と実際にモーションキャプチャで計測した真の姿勢（$\boldsymbol{J} = [\theta_1, \cdots, \theta_n]$）を比較し誤差を評価する．各フレームにおける誤差 e を以下で定義する．

$$e = \sqrt{\sum_{n=1}^{54} ((\hat{\theta}_n - \theta_n) \bmod \pm 180^\circ)^2 / 54} \tag{15}$$

表 1 に 300 フレームについて姿勢推定の平均誤差を示す．ファントムボリュームによる変化を示すために，カメラ数を 4，8 と変化させた．カメラ数が少ない場合には，SFS で生成される視体積に含まれるファントムボリュームが増えるため，推定がより難しくなる．

また，**図 7** にフレームごとの誤差を他手法と比較した結果（縦軸は誤差，横軸はフレーム）を示す．

以上の結果より以下のことがいえる．

- 非線形写像を用いた GPDM が線形写像（PCA）よりも優れている．
- Motion prior を用いた追跡が時系列情報を利用しないその他の姿勢検出手法よりも優れている．
- 視体積制約によりすべての結果に対して，推定精度の向上が見られる．特に，カメラ数 4 においてはより精度の向上が大きく，視体積制約がうまく機能している．

テストデータの姿勢推定結果を 10 フレームごとに表示した結果を**図 8** に示す．この推定は学習モデルの生成時とは別の被験者で行った．体格の異なる人物の非剛体着衣から姿勢推

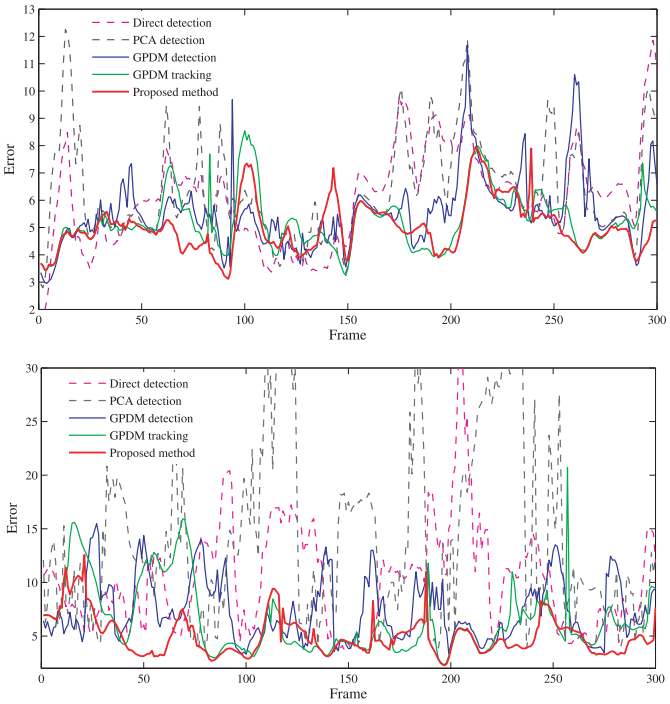

図 7　他手法との推定誤差の比較（上：舞踊 1，下：舞踊 2）
Fig. 7　Comparison of joint angle errors in dance sequences.

定が行えていることが分かる．

潜在空間における追跡結果を**図 9** に示す．緑の点が学習に利用した修正形状の潜在変数（X），青の点が視体積を投影した結果（$\boldsymbol{x}$），赤の点が提案手法を用いて得られた推定形状を投影した結果（$\bar{\boldsymbol{x}}$）を表す．$\boldsymbol{x}$ の一部は学習モデルから離れている．これはファントムボリュームを含んだまま投影したためであり，ファントムボリュームの影響が考慮されない一般的な姿勢推定では一時的な大きな復元誤りによって姿勢推定が失敗してしまうことを示唆している．これに比べて提案手法では，$\bar{\boldsymbol{x}}$ が学習モデルに近づき，投影点も滑らかになっている．

提案手法において，形状追跡に要する実行速度は平均で約 3 fps であった．オフライン学習におけるグラフカットを用いた形状修正の実行速度が約 0.017 fps，緩い着衣中の姿勢推定の従来法が約 0.063 fps であるのに対し，高速な処理が実現できた．しかし，多くのオンラインアプリケーションを実現するうえで十分な速度ではない．たとえば動作認識システムへの応用を目標にした場合，多くの一般的な動作プリミティブ（歩行，手招き，指示動

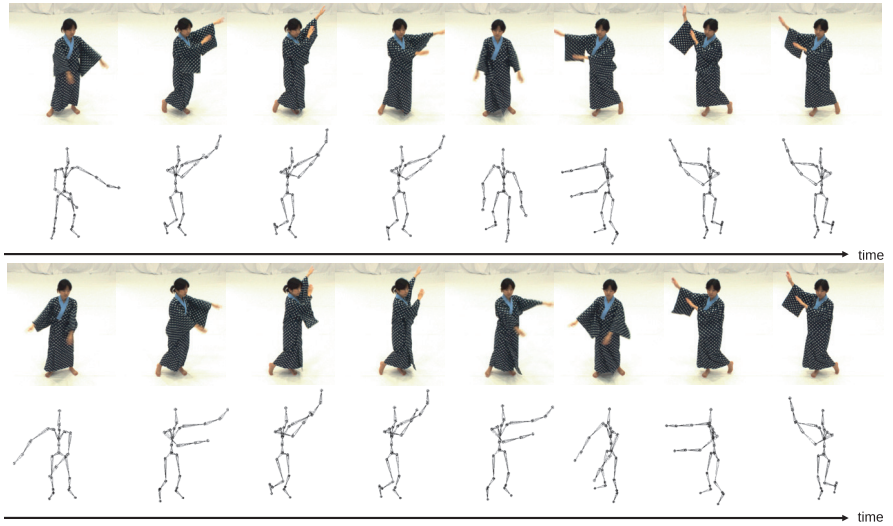

図 8　姿勢推定結果（舞踊 1）：上段　観測画像，下段　姿勢推定結果
Fig. 8　Visualization of the results of the proposed method (dance1). Upper: observed images. Lower: Estimated poses.

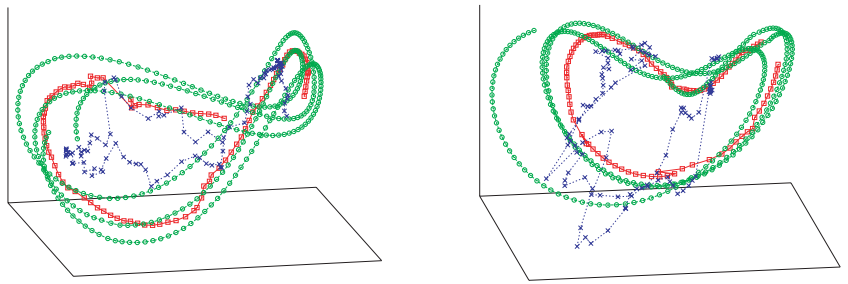

図 9　潜在空間での追跡結果（左：舞踊 1，右：舞踊 2）緑：学習した形状変化のモデル，青：GPDM の写像で得られた結果，赤：提案手法を用いた追跡結果
Fig. 9　Tracking results in the latent space (green: sample refined volumes, blue: input visual hulls, red: our tracking results). Left: dance1, Right: dance2.

作など）の長さは 1 秒程度である：動作認識法の評価に広く使用されている KTH Action Database[29] を参照した．これらの動作認識にビデオレート 30 fps で撮影された動画像が必要と仮定すると，提案手法が 1 秒の姿勢追跡に必要な時間は約 10 秒である．アラームシステムであれば 10 秒の遅れは許されもする可能性があるが，コミュニケーションロボットにおいては 10 秒の遅れは致命的である．システムに必要な撮影レートが 30 fps であれば，その前処理である姿勢推定も同様の速度で実行可能であることが望まれる．

提案手法において最も計算に時間を要しているのは V から X，X から V への写像であった．さらなる高速化のためには，文献 30) のようなさらなる効果的なボリュームのモデリングが必要である．

9. おわりに

本稿では，任意の着衣に適用可能な，3 次元ボリュームからの姿勢推定法を提案し，従来法と比較して高い精度での姿勢推定が達成できた．GPDM により得られた低次元潜在空間でのパーティクルフィルタ，および本稿で新たに提案した視体積制約による対象形状の大きな復元誤差の逐次的修正により，高速で頑健な姿勢推定が可能となる．この手法では，形状・姿勢の確率的な表現を獲得し，また学習データからの単純な探索ではなく回帰を用いることにより，個人差またはその時々の動きのぶれを許容した姿勢推定を実現している．

本手法に限らず学習データから得られる動きモデルを参照する手法では，学習データからかけ離れた動きの姿勢推定は不可能であるが，本手法は対象とする動き・形状に制約がないため，歩行やジョギングなどの一般的な動きに適用できる．しかし，本手法を含めてすべての類似手法では，各動きが独立にモデル化されており，入力画像シーケンスに合わせて手動でモデルが選択されている．そこで，様々な動きを統一的なモデルで表現し，どのような動きが入力されても姿勢推定できるフレームワークが今後の重要な課題となる．

形状修正[12] については京都大学延原章平助教に，GPDM[23] については Dr. Neil D. Lawrence 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) Pollard, N.S., Hodgins, J.K., Riley, M.J. and Atkeson, C.G.: Adapting Human Motion for the Control of a Humanoid Robot, *IEEE International Conference on Robotics and Automation* (2002).
2) Grochow, K., Martin, S., Hertzmann, A. and Popovic, Z.: Style-Based Inverse Kinematics, *SIGGRAPH* (2004).
3) Wu, Y., Huang, T.S. and Mathews, N.: Vision-based gesture recognition: A review, *International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction*, pp.103–115 (1999).
4) Hu, W., Tan, T., Wang, L. and Maybank, S.: A survey on visual surveillance of object motion and behaviors, *IEEE Trans. Systems, Man, and Cybernetics, Part C: Applications and Reviews*, Vol.34, No.3, pp.334–352 (2004).
5) Poppe, R.: Vision-based Human Motion Analysis: An Overview, *Computer Vision and Image Understanding*, Vol.108, No.1–2, pp.4–18 (2007).
6) Cheung, G.K.M., Kanade, T., Bouguet, J.Y. and Holler, M.: A Real Time System for Robust 3D Voxel Reconstruction of Human Motions, *IEEE Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.714–720 (2000).
7) Wu, X., Takizawa, O. and Matsuyama, T.: Parallel Pipeline Volume Intersection for Real-Time 3D Shape Reconstruction on a PC Cluster, *IEEE International Conference on Computer Vision Systems* (2006).
8) Mikić, I., Trivedi, M., Hunter, E. and Cosman, P.: Human Body Model Acquisition and Tracking Using Voxel Data, *International Journal of Computer Vision*, Vol.53, No.3, pp.199–223 (2003).
9) Hou, S., Galata, A., Caillette, F., Thacker, N. and Bromiley, P.: Real-time Body Tracking Using a Gaussian Process Latent Variable Model, *IEEE International Conference on Computer Vision* (2007).
10) Sun, Y., Bray, M., Thayananthan, A., Yuan, B. and Torr, P.H.S.: Regression-based human motion capture from voxel data, *British Machine Vision Conference* (2006).
11) Kutulakos, K.N. and Seitz, S.M.: A Theory of Shape by Space Carving, *International Journal of Computer Vision*, Vol.38, No.3, pp.199–218 (2000).
12) Tung, T., Nobuhara, S. and Matsuyama, T.: Simultaneous Super-resolution and 3D Video Using Graph-cuts, *IEEE Conference on Computer Vision and Pattern Recognition* (2008).
13) Nobuhara, S. and Matsuyama, T.: Deformable Mesh Model for Complex Multi-Object 3D Motion Estimation from Multi-Viewpoint Video, *International Symposium on 3D Data Processing, Visualization, and Transmission*, pp.264–271 (2006).
14) Rosenhahn, B., Kersting, U., Powell, K., Klette, R., Klette, G. and Seidel, H.P.: A System for Articulated Tracking Incorporating a Clothing Model, *Machine Vision and Applications*, Vol.18, No.1, pp.25–40 (2007).
15) Vlasic, D., Baran, I., Matusik, W. and Popovic, J.: Articulated Mesh Animation from Multi-View Silhouettes, *SIGGRAPH* (2008).
16) Ukita, N., Tsuji, R. and Kidode, M.: Real-time shape analysis of a human body in clothing using time-series part-labeled volumes, *European Conference on Computer Vision*, pp.681–695 (2008).

17) Agarwal, A. and Triggs, B.: Tracking articulated motion using a mixture of autoregressive models, *European Conference on Computer Vision*, pp.54–65 (2004).
18) Belongie, S., Malik, J. and Puzicha, J.: Shape matching and object recognition using shape contexts, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.24, pp.509–522 (2002).
19) Thayananthan, A., Navaratnam, R., Stenger, B., Torr, P.H.S. and Cipolla, R.: Multivariate Relevance Vector Machines for Tracking, *European Conference on Computer Vision* (2006).
20) Bustos, B., Keim, D., Saupe, D., Schreck, T. and Vranić, D.: An Experimental Effectiveness Comparison of Methods for 3D Similarity Search, *International Journal on Digital Libraries*, Vol.6, No.1, pp.39–54 (2006).
21) Lawrence, N.D. and Hyvarinen, A.: Probabilistic Non-linear Principal Component Analysis with Gaussian Process Latent Variable Models, *Journal of Machine Learning Research*, Vol.6, pp.1783–1816 (2005).
22) Shon, A.P., Grochow, K., Hertzmann, A. and Rao, R.P.N.: Learning shared latent structure for image synthesis and robotic imitation, *Neural Information Processing Systems*, pp.1233–1240 (2006).
23) Wang, J.M., Fleet, D.J. and Hertzmann, A.: Gaussian process dynamical models for human motion, *IEEE Trans. Pattern Analysis and Machine Intelligence* (2007).
24) Deutscher, J., Blake, A. and Reid, I.: Articulated body motion capture by annealed particle filtering, *IEEE Conference on Computer Vision and Pattern Recognition* (2000).
25) Urtasun, R., Fleet, D.J., Hertzmann, A. and Fua, P.: 3D people tracking with Gaussian process dynamical models, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.238–245 (2006).
26) Sagawa, Y., Shimosaka, M., Mori, T. and Sato, T.: Fast Online Human Pose Estimation via 3D Voxel Data, *International Conference on Intelligent Robots and Systems*, pp.1034–1040 (2007).
27) Mackay, D.J.C.: *Information Theory, Inference and Learning Algorithms* (2003).
28) Lawrence, N.D. and Candela, J.: Local distance preservation in the GP-LVM through back constraints, *International Conference on Machine Learnig*, pp.513–520 (2006).
29) KTH Action Database. http://www.nada.kth.se/cvap/actions/
30) Snelson, E. and Ghahramani, Z.: Sparse Gaussian processes using pseudo-inputs, *Neural Information Processing Systems*, pp.1257–1264 (2006).

(平成 21 年 5 月 27 日受付)
(平成 22 年 1 月 8 日採録)

平井　迪郎

2007 年神戸大学工学部卒業．2009 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）ソニー勤務．在学中，人体の姿勢・形状解析の研究に従事．

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．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 年電子情報通信学会論文賞．

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，そして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．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．情報処理学会フェロー，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．