

固有空間でのモデルフィッティングによる 単眼画像からの人体 3 次元姿勢推定

大西 克則[†] 博 格利[†] 滝口 哲也^{††} 有木 康雄^{††}

[†] 神戸大学大学院システム情報学研究科 兵庫県神戸市灘区六甲台町 1-1

^{††} 神戸大学自然科学系先端融合研究環 兵庫県神戸市灘区六甲台町 1-1

E-mail: [†]{katsu0920,bogeli}@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki}@kobe-u.ac.jp

あらまし 従来の 3D 人体姿勢推定の研究におけるアプローチは、学習ベースの手法と、モデルフィッティングによる手法の 2 つに分類することができる。前者の手法は高速な推定が可能であるが、精度が低いという欠点がある。一方後者の手法では、高精度に推定できるが、計算コストがかかってしまう。本論文では、学習ベースの手法で高速に推定を行った後、その推定結果を用いることでモデルフィッティングの計算コストを抑え、精度を向上させる手法を提案する。学習ベースの手法では、回帰分析による推定を行い、その結果を用いてパーティクルフィルタを適用し、モデルフィッティングを行う。その際、入力画像の人体動作の種類に応じて固有な空間を構築し、各空間で探索を行うことで効率よくフィッティングを行う。評価実験では、公開データベースの CMU Graphics Lab Motion Capture Database を用いて行った。実験の結果、提案手法では関節角度の推定誤差を 6.2 度に抑えることができ、固有空間を構築せず、回帰分析から得られる結果よりも約 0.9 度改善された。

キーワード 姿勢推定, 回帰分析, パーティクルフィルタ, 主成分分析

3D Human Pose Estimation from Monocular Image by Model Fitting in Eigen Space

Katsunori ONISHI[†], Geli BO[†], Tetsuya TAKIGUCHI^{††}, and Yasuo ARIKI^{††}

[†] Graduate School of Science and Technology, Kobe University, 1-1, Rokkodai, Nada, Kobe, Hyogo

^{††} Organization of Advanced Science and Technology, Kobe University, 1-1, Rokkodai, Nada, Kobe, Hyogo

E-mail: [†]{katsu0920,bogeli}@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki}@kobe-u.ac.jp

Abstract Generally, the human pose estimation method has two approaches. One is training-based approach, and the other is model-fitting approach. Though the former method can be quickly estimated, that has disadvantages with low accuracy. While the latter method can be estimated in high accuracy. However, its computational cost is expensive. In this paper, we propose a method to integrating both of them to improve the estimation precision. In the training-based approach, we use regression analysis to model the mapping from visual observations to human poses. In the model-fitting approach, the particle filter is employed by using the result of the regression analysis. In that case, the fitting is efficiently searched in respective eigen space corresponding to the different human motions. Finally, the proposed method was evaluated under CMU Graphics Lab Motion Capture Database. The RMS error of human joint angles was 6.2 degree by the proposed method, and it has been improved up to about 0.9 degree by compared to the method without eigen space.

Key words pose estimation, regression analysis, particle filter

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 [2] が行われている。人体の姿勢や運動情報の獲得により、ジェスチャ等の動作を入力とした計算機のインターフェースの構築、ロボットとのインタラクションや画像監視など多くの分野で応用が期待されている。また、遠隔講義

や遠隔会議等のシーンにおいて、現実世界に存在する人や物体の形状を計算機に取り込み、その形状や姿勢を認識することによって仮想世界を構築する取り組みがなされている。

一般的な姿勢推定の手法は、2 種類のアプローチが提案されている。1 つは、画像から得られる形状特徴量から姿勢への写像を回帰により学習し、その復元された形

状から姿勢推定を実現している [3] [4] [5] [6]. このような学習ベースの推定方法は、一度学習を行えば非常に高速な姿勢推定が可能となる. しかしながら、学習していない姿勢が入力されると、推定精度が低下してしまう場合がある. もう 1 つには、あらかじめ用意しておいた 3 次元人体モデルと復元形状を比較し、重なりが最大になるパラメータを、ベイズ則に基づく方法や、観測される尤度を用いて事後確率をモデル化することにより求める方法がある [7] [8] [9] [10]. しかしながらこれらの手法は、高い精度が得られる利点がある一方、一般的に計算コストがかかるという欠点がある.

以上の議論から、本研究では姿勢推定の手法に関して、学習ベースに基づく手法とモデルフィッティングによる推定方法の両方を用いることで、推定精度を向上させる手法を提案する. 一般的なモデルフィッティングに基づく推定方法では、モデルの初期値を決定する方法が問題となる. そこで我々は、学習ベースの手法の 1 つである回帰分析を用いてモデルの初期値を高速に推定し、その後パーティクルフィルタを用いてモデルフィッティングを行うことで、精度を向上させる. その際、通常の人体 3D モデルは高次元な特徴となっており、そのまま探索を行うとうまく推定できない場合がある. 我々は、あらかじめ動作ごとに固有空間を主成分分析を用いて構成しておき、その固有空間で効率よく探索することで、精度を向上させる.

2. 画像特徴と 3D 人体モデル

2.1 Histograms of Oriented Gradients

一般物体認識のための gradient ベースの特徴量として、Histograms of Oriented Gradients (HOG) [11] や Scale-Invariant Feature Transform (SIFT) [12] など、局所特徴を用いて物体を認識する手法がある. HOG は SIFT と同様に局所領域における輝度の勾配方向をヒストグラム化した特徴量である. SIFT と類似した特徴量の記述を行うが、SIFT は特徴点 (keypoint) に対して特徴量を記述するのに対し、HOG ではある一定領域に対する特徴量の記述を行う. そのため、図 1 に表すように、HOG は大まかな物体の形状を表すことが可能となる. 以下に HOG 特徴の具体的な算出アルゴリズムについて述べる.

2.1.1 輝度勾配算出

HOG 特徴を抽出する前に、入力画像に対してあらかじめ背景差分法を用いて、人物領域のみの画像を取得する. このとき背景を削除すると同時に画像サイズの正規化も行い、人物が画像の中央部に位置するようにする. 画像の位置 (x, y) における輝度値を $I(x, y)$ として輝度勾配を次式より算出する.

$$\begin{cases} f_x(x, y) = I(x+1, y) - I(x-1, y) & \forall x, y \\ f_y(x, y) = I(x, y+1) - I(x, y-1) & \forall x, y \end{cases} \quad (1)$$

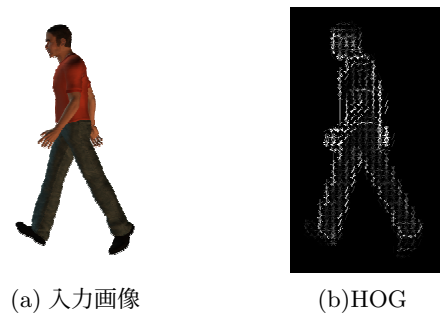


図 1 人体形状特徴

ここで f_x と f_y は、それぞれ画像の x と y 方向の輝度勾配をあらわす. 得られた輝度勾配画像から、勾配強度 $m(x, y)$ と勾配方向 $\theta(x, y)$ を次式より求める.

$$m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \quad (2)$$

$$\theta(x, y) = \tan^{-1}(f_y(x, y)/f_x(x, y)) \quad (3)$$

このようにして得られた勾配方向 $\theta(x, y)$ は、 $[-180^\circ, 180^\circ]$ となる. しかし姿勢推定を行う場面においては、表情や服装といった細かい変化に対しては影響されにくいことが望ましい. そのため、勾配方向の向きに関しては考慮しないよう、次式より $[0^\circ, 180^\circ]$ の範囲にする.

$$\tilde{\theta}(x, y) = \begin{cases} \theta(x, y) + \pi, & \text{if } \theta(x, y) < 0 \\ \theta(x, y), & \text{otherwise} \end{cases} \quad (4)$$

2.1.2 セルによるヒストグラム化

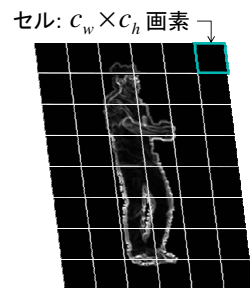


図 2 セル分割画像

算出された輝度勾配画像を、図 2 のようにセルと呼ばれる $c_w \times c_h$ 画素からなる小領域に分割する. それぞれの領域において、算出された勾配強度 $m(x, y)$ と勾配方向 $\tilde{\theta}(x, y)$ からヒストグラムを作成する. $\tilde{\theta}(x, y)$ の角を c_b 方向になるよう量子化し、各方向に $m(x, y)$ を重みとして与える. すなわち、1 セルあたり c_b 方向の勾配方向ヒストグラムができる.

2.1.3 ブロックによる正規化

セルごとに作成された輝度勾配ヒストグラムを、図 3 に示すような、セルよりも大きな領域 $b_w \times b_h$ セルを 1 ブロックとして正規化を行う. 1 セルあたり c_b 方向の特徴を持っているため、1 ブロックあたりの特徴次元は

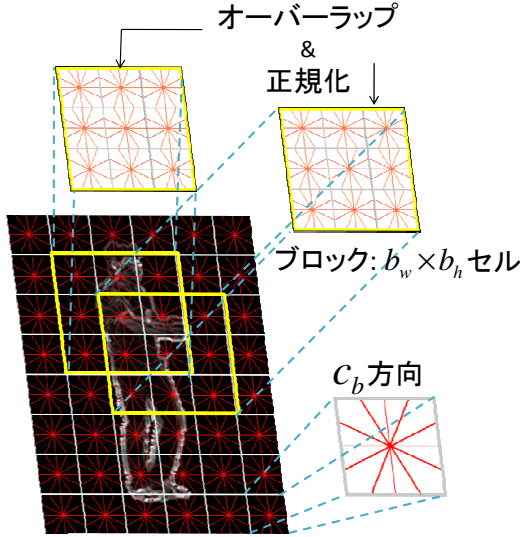


図3 ブロックによる正規化

$d_b = b_w \times b_h \times c_b$ となる．あるブロックの特徴ベクトルを $\mathbf{v}$ ，ブロック内で位置 (i, j) ， $\{1 \leq i \leq b_w, 1 \leq j \leq b_h\}$ にあるセルのヒストグラムを h_{ij} としたとき，次式により正規化を行う．

$$h'_{ij} = \frac{h_{ij}}{\sqrt{\|\mathbf{v}\|_2^2 + \epsilon}} \quad (\epsilon = 1) \quad (5)$$

正規化の際，ブロックは1セルずつオーバーラップさせながら移動させる．つまり，セルのヒストグラム h_{ij} は異なるブロック領域によって繰り返し正規化されることになる．

このようにして得られた特徴ベクトルは，照明や影の影響を受けにくく，局所的な幾何学変化に頑健となる．

2.2 3次元人体モデル



図4 3D 人体モデル

人体は複数の関節を持ち，様々な形に変形する多関節物体であると考えられる．しかし，各関節間を連結している体節部分は剛体であるときみなすことができる．そのため，体節を関節によって互いに接続させておけば，関節角を決定することで人体モデルを表現することが可能となる．つまり，人体の姿勢を表現するためには関節角の値が重要になる．

そこで人体3次元モデルの構造を表す方法として，人

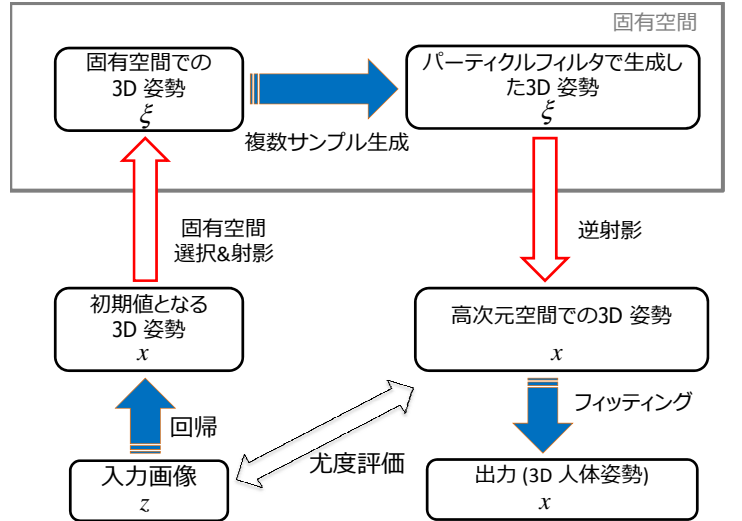


図5 提案システム

体の構造に基づいて各関節(肘，腰，膝など)の角度を用いる．つまり，この関節角を変更することにより，人体のさまざまな姿勢を表現することが可能となる．我々は，CMU Graphics Lab Motion Capture Database [1] で公開されているモーションキャプチャデータを実験に使用する．このデータでは人体を62次元ベクトルで表現しており，人体のワールド座標系でのポジションと方位(6次元)と20箇所の関節(56次元)から構成されている．図4に，関節角による3次元人体モデルの一例を示す．

3. 姿勢推定方法

3.1 概要

本章では，3D姿勢の推定方法について述べる．一般的な人体の姿勢推定方法は，学習ベースの手法とモデルフィッティングによる手法が存在する．学習ベースの推定方法は，一度学習を行えば非常に高速な姿勢推定が可能であるが，テスト画像の種類によっては精度低下が懸念される．一方モデルフィッティングに基づく手法では，より高い精度が期待できるが，高い計算コストや，モデルの初期値の設定方法に問題がある．

本研究では，学習ベースとモデルフィッティングによる推定方法の両方を用いることで，推定精度を向上させる．まず，回帰分析により高速に推定を行い，その推定された姿勢を初期値としてパーティクルフィルタを用いてモデルフィッティングを行うことで，高精度な推定を実現する．モデルフィッティングを行う際，2.2節で述べたように，人体モデルの自由度が高いため，収束に時間がかかりすぎるのが考えられる．この問題に関して，固有空間を構成することにより高速に収束させる方法を提案する．図5に推定システムの概要を示す．

3.2 学習ベースの推定

学習ベースの推定では，回帰分析を用いる．図6に示

すように、2.1 節で抽出した特徴空間から、姿勢空間へ写像することで姿勢を推定する。文献 [4] [6] でも述べられているように、姿勢空間と特徴空間との関係を線形に近似して表すことが可能である。

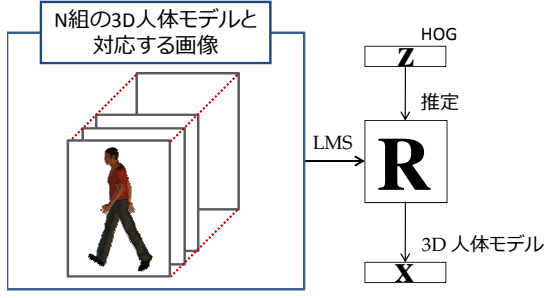


図 6 回帰分析に基づく推定

画像から得られた HOG 特徴ベクトル $z \in \mathbb{R}^{D_z}$ と人体 3 次元モデルの関節角度から構成される特徴ベクトル $x \in \mathbb{R}^D$ の関係を、次式で近似する。

$$x = Rz + \varepsilon \quad (6)$$

このとき、 R は $D \times D_z$ 行列、 ε は残差ベクトルである。画像特徴 z が入力されれば人体 3 次元モデル特徴 x に変換され、姿勢推定が可能となる。

N 個の学習データのセット $\mathcal{D} = \{(z_i, x_i) \mid i = 1 \dots N\}$ (この場合、画像の HOG 特徴と 3D 人体関節角を指す) を用いて、次式の最小自乗推定法により、 R の値を決定する。

$$R := \arg \min_R \sum_{i=1}^N \|Rz_i - x_i\|^2 \quad (7)$$

学習データの特徴ベクトルを $D_z \times N$ 行列の $Z \equiv (z_1 \ z_2 \ \dots \ z_N)$ と $D \times N$ 行列の $X \equiv (x_1 \ x_2 \ \dots \ x_N)$ と表すことにより、推定問題は以下のように表せる。

$$R := \arg \min_R \|RZ - X\|^2 \quad (8)$$

画像から入力された HOG 特徴ベクトル z に対して、求められた変換行列 R を用いて変換することにより x を推定する。これにより人体の 3 次元姿勢推定を行う。

3.3 モデルフィッティングに基づく推定

3.3.1 パーティクルフィルタによるモデルフィッティング

パーティクルフィルタを用いて入力画像と 3D 人体モデルのマッチングにより姿勢推定を行う。パーティクルフィルタを用いる場合、初期状態の決定方法が問題となる。文献 [13] [14] は、パーティクルフィルタを適用した推定手法であるが、初期状態は手動で与えている。

本手法では、回帰分析に基づく推定で得られた結果を初期状態として探索を開始する。これにより、手動で与える作業は一切なくなり、すべて自動で高精度な姿勢推定が可能となる。

回帰分析により得られた初期姿勢の周囲にパーティクルを生成し、それぞれのサンプルの尤度評価を行い、各サンプルの重みを求める。得られた重みに基づいて、各パーティクルのリサンプリング、状態遷移を繰り返し行うことで、人体モデルを入力画像にフィッティングさせる。ある一定回数繰り返した後、最も高い尤度を持つ状態を最終的な 3D 姿勢として出力する。

3.3.2 状態空間

パーティクルの状態量には、2.2 節で述べた 3D 人体モデルを用いる。しかし 2.2 節でも述べたように、人体は多関節物体であり、さまざまな姿勢を取ることができるため、姿勢変形の自由度が非常に高い。そのため、3D 人体モデルをそのまま用いると状態空間が大きくなりすぎてしまい、パーティクルフィルタではうまく推定することができない。

一方、顔認識等で用いられている Active Appearance Model (AAM) [15] という手法では、顔の形状に PCA を適用して次元圧縮を行い、少ないパラメータで顔の形状変化を表現できるようにしたモデルを構築している。AAM の特徴は、顔の特徴点の分布が次元圧縮された顔形状モデルで表現されるため、顔らしい形状を保つための拘束条件として有効に働いている点である。

そこで我々は、入力動作の種類ごとに PCA を行うことで次元削減すると同時に、各動作ごとに固有な空間を生成する手法を提案する (図 7)。各動作ごとに PCA を用いることで、3D 人体モデルがその動作らしい形状を保つ拘束条件として働き、次元圧縮も同時に行うことで探索の効率化を図る。

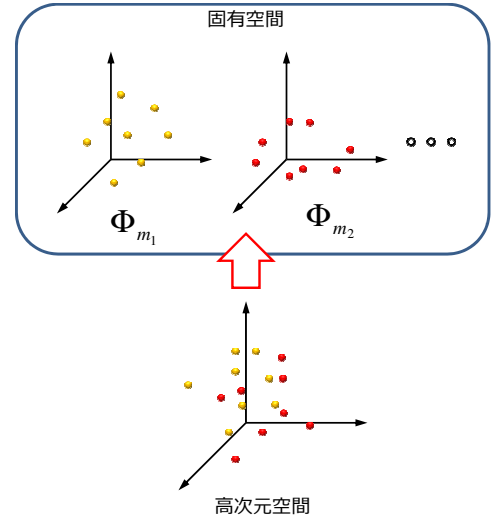


図 7 高次元空間と固有空間

動作の種類数を $\mathcal{M}$ 個とすると、ある動作 $m \in \mathcal{M}$ の全学習データに対して PCA を行った際の 3 次元モデルを表すベクトル x_m は、

$$\xi_m = P_m(x_m - \bar{x}_m) \quad (9)$$

と表現される。ただし、 $\bar{x}_m$ はある動作 m の学習データに対する平均 Pose ベクトルであり、 P_m は基底ベクトル、 ξ_m は固有空間での Pose ベクトルである。学習データにある M 個の全動作に対して個別の空間を作成するため、 M 個の固有空間が形成される。このように PCA を行い、Pose ベクトルをそれぞれの動作の固有空間に写像したものを状態量として用いる。

固有空間で表現されたデータは、固有値が大きいものに対応する固有ベクトルを基底とする成分に主要な情報が含まれる性質がある。人体の姿勢・運動では各関節の動きに強い相関があるため、この性質がよく表れていると考えられる。そのため、固有空間で表された D 次元 Pose ベクトルは、累積寄与率

$$Q(D_m) = \frac{\sum_{i=1}^{D_m} \lambda_i}{\sum_{i=1}^D \lambda_i} \quad (10)$$

が 95% 以上になるように D_m 次元に削減する。これにより、それぞれの動作に固有な低次元空間を構築することが可能となる。

3.3.3 尤度の評価方法

尤度の評価方法は、パーティクルフィルタの性能を左右するものであり、どのような画像特徴を用いるかが重要な問題となる。評価に用いる画像特徴は、様々な条件下でも安定して、かつ容易に抽出できるものが望ましい。本研究では、形状を表す HOG と、人体領域を評価するためのシルエットの 2 種類の画像特徴を用いる。

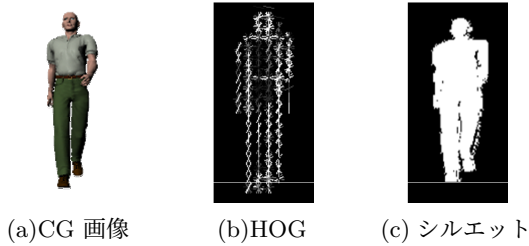


図 8 特徴抽出

Configuration:

HOG は人体の形状を表すことができ、色や服装、照明変化に頑健であるという点からも、尤度評価に適した特徴であるといえる。パーティクルの状態量から CG を用いて画像を生成し、HOG を抽出する (図 8 (a), (b))。こうして得られた HOG と入力画像から得られた HOG の距離を、以下の式を用いて計算する。

$$E_{HOG}(z, \xi) = \frac{1}{D_z} \chi^2(z, I(\xi)) \quad (11)$$

$$\chi^2(z, z') = \sum_{c=1}^{D_z} \frac{(z^c - z'^c)^2}{z^c + z'^c} \quad (12)$$

ここで、 D_z は HOG の次元数、 $I(\xi)$ はパーティクル

の状態量 ξ から CG を用いて画像特徴を抽出する関数、 $\chi^2(z, z')$ は HOG 特徴である z と z' のカイ二乗距離であり、 z^c は特徴ベクトル z の c 番目の要素を表す。

Region:

シルエット画像は、背景差分法により前景部分を取り出し、画像サイズを正規化して抽出する (図 8 (c))。シルエット特徴も HOG と同様に、服装や照明変化に頑健な特徴であり、安定して評価に用いることができる。シルエット画像を抽出した後、前景領域の画素値を 1、背景領域の各画素の値を 0 として、以下の式により入力画像との距離を計算する。

$$E_{Region}(z, \xi) = \frac{1}{M} \sum_{i=1}^M (1 - p_i(z, I(\xi))) \quad (13)$$

M は画素数を表しており、 $p_i(z, I(\xi))$ は入力画像と状態量の AND を出力する関数となっている。

入力画像と状態量の画像上での適合度 $C(\xi)$ を、式 (11)、式 (13) より、以下の式で求める。

$$C(\xi) = \exp \{-(E_{HOG}(z, \xi) + E_{Region}(z, \xi))\} \quad (14)$$

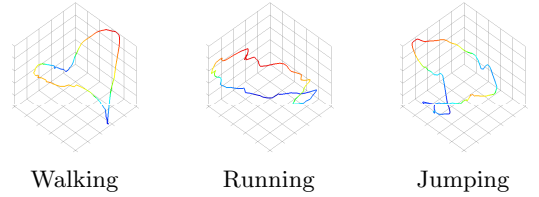


図 9 学習データから作成したトラジェクトリ

本手法では、各動作ごとに固有な空間を構築して探索を行っているため、その動作に特化した空間的制約を加えていることになる。しかし、その空間内でもその動作を表す部分はさらに制約を加えることが可能であると考えられる。そこで、各動作の学習データからトラジェクトリ特徴を抽出し、さらに制約を加えることとする。動作のトラジェクトリの一例を図 9 に示す。図 9 は、Pose ベクトルの上位 3 次元を表したものである。状態量 ξ とトラジェクトリの距離を算出し、その値をペナルティ $P(\xi)$ として尤度評価に用いる。最終的に出力する姿勢は、式 (15) が最大となる値を用いる。

$$\mathcal{L} = \lambda C(\xi) P(\xi)^{-1} \quad (15)$$

3.3.4 固有空間の選択方法

モデルフィッティングでは、各動作ごとに固有空間の上で行うため、事前に各動作の固有空間を構築する、入力画像の動作の種類によって最適な空間を選択する必

要がある．まずそれぞれの動作空間に多数のサンプル点 $\{\xi_m^i | m \in \mathcal{M}, i = 1, \dots, S_m\}$ をランダムに生成する．そのサンプル点と，回帰分析で得られた初期姿勢 $\mathbf{x}$ の平均距離を求め，最も近い空間を入力された動作として推定する．

距離を計算する際には，それぞれの固有空間と，各サンプル点を高次元空間へ写像した空間の 2 つの空間で以下の式を用いて計算を行う．

$$\omega_m = \frac{1}{S_m} \sum_{i=1}^{S_m} \frac{1}{D_m} \|\xi - \xi_m^i\|_2 \quad (16)$$

$$\Omega_m = \frac{1}{S_m} \sum_{i=1}^{S_m} \frac{1}{D} \|\mathbf{x} - \mathbf{x}_m^i\|_2 \quad (17)$$

ξ は $\mathbf{x}$ をある動作の固有空間へ写像した低次元 Pose ベクトルである．式 (16)，式 (17) を用いて，各空間で求めた距離の和が最小となる空間が入力画像の動作であると推定する． $f(m)$ をある動作 m の固有空間 Φ_m への写像関数とすると，以下の式を用いて固有空間 Φ_m を決定する．

$$\Phi_m = f(\arg \min_{m \in [1, \dots, \mathcal{M}]} \{\omega_m + \Omega_m\}) \quad (18)$$

3.3.5 対象方位の推定

回帰分析で推定をする際には，人体の方位 ϕ が未知の状態でも，画像の特徴量に変化が表れるため姿勢推定が可能である．しかしモデルフィッティングを行う場合には，方位 ϕ を推定しなければ画像上でのマッチングがうまくとれない．

そこで，回帰分析で得られた姿勢を 360 度のあらゆる方向から観測した画像を CG で生成し，式 (14) を用いて以下の式で方位を推定する．

$$\hat{\phi} = \arg \max_{0 \leq \phi < 360^\circ} C(\mathbf{x}^\phi) \quad (19)$$

$C(\mathbf{x}^\phi)$ は，ある方位 ϕ の姿勢 $\mathbf{x}$ から生成された CG 画像と入力画像の適合度を表す．式 (19) で得られた $\hat{\phi}$ を用いてモデルフィッティングを行う．

4. 評価実験

4.1 データセット

評価実験では CMU Graphics Lab Motion Capture Database [1] のデータを用いる．データベースには様々な動作のモーションキャプチャデータが収められているが，人体の関節角の変動を一通り表現できる，“walking”，“running”，“jumping”の 3 動作のデータを使用する．データベースにはモーションキャプチャデータのみで，それに対応する画像データは公開されていない．そのため，本研究では画像データを CG で生成し，実験を行った．それぞれの動作の様子の一例を図 10 に示す．

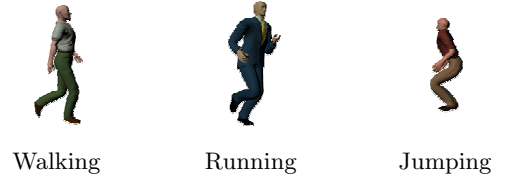


図 10 用いた姿勢の一例

4.2 実験条件

データセットの各動作は，それぞれ複数回の試行が収められている．実験では，1 回目の試行データを学習データに用い，2 回目の試行データをテストデータとして使用する．

実験に用いる CG 画像では， 640×480 画素の解像度を持つ映像を生成した．映像を生成する際，固定したカメラに対して水平方向に CG で作成した人体を回転させ，8 方向から観測したデータを用いる．それぞれの方向に対して， $\mathcal{M} = \{\text{“walking”}, \text{“running”}, \text{“jumping”}\}$ の 3 動作を行い，学習に用いる．テストデータは 8 枚の画像を各動作，方向ごとに用いた．なお，学習データとテストデータに用いる人物は，テクスチャを変えることで様々な服装をとっているものを使用する．使用したデータの詳細を，表 1 に示す．

表 1 実験に用いたデータ

Pose	画像枚数	
	学習データ	テストデータ
Walking	720	64
Running	1512	64
Jumping	1536	64
Total	3768	192

HOG の抽出の手順は，背景差分法を用いることで人体領域を抽出する．切り出した画像の大きさは 64×128 正規化する．HOG 特徴の各パラメータの値は $c_w = 10, c_h = 10, c_b = 9, b_w = 3, b_h = 3$ とした．すなわち，大きさ 10×10 画素のセルに 9 方向のヒストグラムを作成し， 3×3 セルの大きさのブロックで正規化する．その際，ブロックは 1 セルずつオーバーラップさせながら正規化した．各ブロックの特徴次元は $d_b = b_w \times b_h \times c_b = 81$ となり，画像全体で 40 個のブロックができるため，3240 次元の HOG 特徴画像ができる．

パーティクルフィルタの粒子数，および固有空間の選択の際に用いるサンプル数は，ともに実験的に 800 とした．

4.3 評価尺度

姿勢推定の精度評価には，文献 [4] と同様に root-mean-square error (RMS error) を用いる．この評価尺度は，真値 x_i と推定値 $\hat{x}_i$ の関節角度の絶対誤差を測定するのである．関節角は D 次元 (= 56) あるため，その平均値を以下の式より算出する．

$$D(x, \hat{x}_i) = \frac{1}{D} \sum_{i=1}^D |(x_i - \hat{x}_i) \bmod \pm 180^\circ| \quad (20)$$

4.4 実験結果

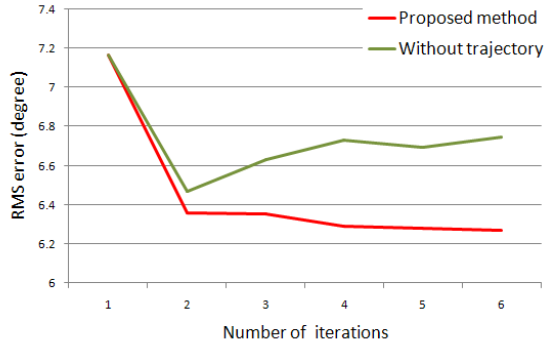


図 11 姿勢推定結果

姿勢推定精度の実験結果を、図 11 に示す。グラフの横軸である iteration で表されている回数は、1 は回帰分析による結果を、2 回目以降はパーティクルフィルタを用いて繰り返し推定を行った値を示している。縦軸は RMS error を表しており、値が小さいほど推定誤差が少なく、精度が良いことを表している。

探索にトラジェクトリの制約を用いない場合、収束できていないことがわかる。提案手法では解が安定して収束しており、トラジェクトリによる制約の有効性が確認できる。画像情報のみの推定では、奥行き情報が欠落してしまうため、どうしても 3 次元での誤差が大きくなってしまふ。トラジェクトリによる制約を用いることで、奥行き情報の制約がうまく働き、誤差軽減につながったと考えられる。

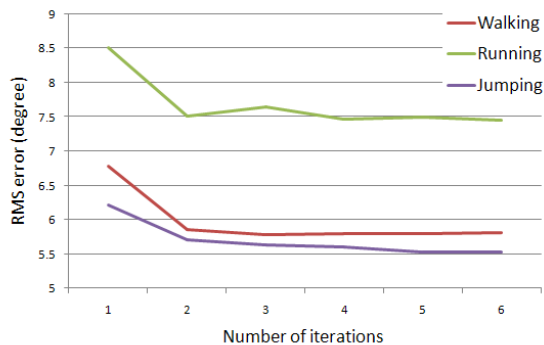


図 12 動作ごとの評価

次に、各動作ごとの結果を図 12 に、固有空間選択の精度を表 2 に confusion matrix で示す。Confusion matrix は、行が Ground truth であり、列が推定された動作空間である。図 12 より、"running" が他の動作と比較すると精度が悪くなってしまった。また、空間選択の精度が

表 2 Confusion matrix [%].

	Walking	Running	Jumping
Walking	84.38	3.12	12.5
Running	9.37	78.13	12.5
Jumping	15.63	4.68	79.69

80% 前後と改善の余地がある結果となった。本研究では、単一フレームのみで空間の選択や姿勢推定を行っている。あるフレームでは図 13 のように、"walking", "running", "jumping" の全ての動作に類似した姿勢が存在しており、どの動作においてもそれぞれの空間に特化した姿勢が出力されないことがある。そのため、そういった画像では空間を正しく選択することは困難であり、精度低下につながってしまう。この問題に関しては、1 フレームではなく動画を扱い、前フレームの状態を知識として利用する方法が考えられる。状態系列を事前知識として導入すれば、類似した姿勢でも知識を用いて区別することが可能となり、空間選択の精度向上につながる。空間選択の精度が向上すれば、図 14 からわかるように、姿勢の推定精度が向上する。



図 13 異なる動作で発生する類似姿勢

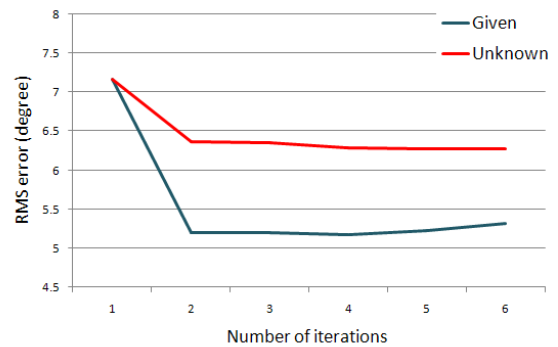


図 14 固有空間が既知の場合

図 14 に、入力画像の動作が既知の場合と、固有空間を推定した場合の結果を示す。表 2 の結果を考慮すると、80% 前後の精度から、残り 20% を改善するだけで推定誤差を大きく減らせることが確認できる。

図 15 に、関節角ごとの推定誤差を示す。グラフより、関節部位によって誤差の大きい部分とそうでない部分があることがわかる。誤差が小さい部位は、胴体や頭などの比較的動きが少ない部分である。一方、腕や脚など激

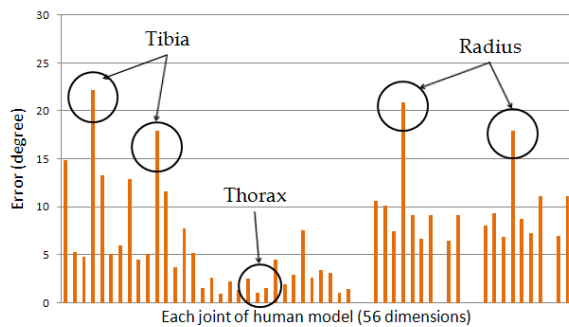


図 15 関節角度ごとの誤差

しく変化する部位の誤差が非常に大きくなっている。そのため、腕や脚で発生する誤差を抑えることができれば、全体的な推定精度はさらに向上できることがわかる。そのため、そういった変化の激しい部位に関して、体の部位検出手法を組み合わせることによって、誤差を軽減できると考えられる。しかしながら、一方で腕や脚は胴体に隠れてしまう状態も頻繁に発生する。そのため、単純な部位検出ではなく、オクルージョンが発生した場合に対応するためには、やはり前フレームの知識を導入する必要があると考えられる。

最後に、実際に我々の手法を用いて推定した結果の一例を図 16 に示す。結果画像は、上段が入力画像、下段が推定結果となっている。

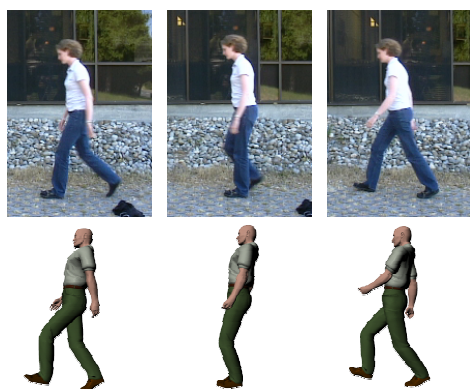


図 16 実験結果の一例

5. 結 論

本論文では、学習ベースとモデルフィッティングによる手法を組み合わせ、さらに動作ごとに固有空間を構築することで、単眼画像から 3 次元人体姿勢の推定手法を提案した。固有な空間を用いることで、単に次元削減するよりも空間制約が働き、効率的な探索が可能となった。実験結果より、初期値を手動で与えなくとも、回帰分析の結果を用いることで、すくないイタレーションの回数で誤差を軽減し、収束できることを示した。

今後は、空間選択の精度を向上させる必要がある。そのため、1 枚の画像ではなく、映像を用いて前フレー

ムの情報を事前知識として用いる方法が考えられる。また、変動の激しい腕や脚の誤差を抑えることで、推定誤差を小さくできることが可能となる。そのため、特徴量に体のパーツディテクションの機能を、現在用いている特徴量とは別に、補助的な情報として加えることも検討中である。

文 献

- [1] CMU Human Motion Capture DataBase. Available online at <http://mocap.cs.cmu.edu/>.
- [2] R. Poppe, "Vision-based Human Motion Analysis: An Overview," *Computer Vision and Image Understanding*, Vol. 108, No. 1-2, pp. 4-18, 2007.
- [3] C. Sminchisescu, A. Kanaujia, D. N. Metaxas, "*BM³E*: Discriminative Density Propagation for Visual Tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2030-2044, 2007.
- [4] A. Agarwal and B. Triggs, "3D Human Pose from Silhouettes by Relevance Vector Regression," *IEEE Conference on Computer Vision and Pattern Recognition*, 882-888, 2004.
- [5] X. Zhao, H. Ning, Y. Liu, T. Huang, "Discriminative Estimation of 3D Human Pose Using Gaussian Processes," *Proc. of 19th Int'l Conf. on Pattern recognition (ICPR08)*, 2008.
- [6] H. Ning, Y. Hu, T. Huang, "Efficient initialization of Mixtures of Experts for human pose estimation," *IEEE ICIP2008*, pp. 2164-2167, 2008.
- [7] T. Jaeggli, E. Koller-Meier, L. V. Gool, "Learning Generative Models for Monocular Body Pose Estimation," *Asian Conference on Computer Vision*, November, 2007.
- [8] M. Lee, I. Cohen, "A Model-Based Approach for Estimating Human 3D Poses in Static Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, VOL.28, No.6, 2006.
- [9] S. Hou, A. Galata, F. Caillette, N. Thacker and P. Bromiley, "Real-time Body Tracking Using a Gaussian Process Latent Variable Model," *In IEEE International Conference on Computer Vision, ICCV2007*, pp. 1-8, 2007.
- [10] Peng Guan, Alexander Weiss, Alexandru O. Blan, Michael J. Black, "Estimating Human Shape and Pose from a Single Image," *In IEEE International Conference on Computer Vision, ICCV2009*, pp. 1381-1388, 2009.
- [11] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *IEEE Computer Society International Conference on Computer Vision and Pattern Recognition*, pp. 886-893, 2005.
- [12] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, Vol.60, No.2, pp. 91-110, 2004.
- [13] J. Deutscher, A. Blake and I. Reid, "Articulated Body Motion Capture by Annealed Particle Filtering," *Computer Vision and Pattern Recognition, CVPR2000*, pp. 126-133, 2000.
- [14] Long Ye, Qin Zhang, Ling Guan, "Use Hierarchical Genetic Particle Filter to Figure Articulated Human Tracking," *International Conference on Multimedia and Expo, ICME2008*, pp. 1561-1564, 2008.
- [15] T. F. Cootes, G. V. Wheeler, K. N. Walker, C. J. Taylor, "View-based active appearance models," *Image and Vision Computing* 20, pp. 227-232, 2002.