論　文

摂動特徴量による人体形状モデル高速フィッティング

木下　航一[†,††a)]　村瀬　洋[††]

A Fast and Robust Human Body Shape Model Fitting Based on the Perturbation Feature

Koichi KINOSHITA[†,††a)] and Hiroshi MUSASE[††]

あらまし　人体画像に対する，摂動特徴量に基づく高速・高精度なモデルフィッティングを実現した．摂動特徴量に基づくフィッティング手法は，特徴量とモデルパラメータ誤差の関連を学習することにより，モデルを画像上の物体にフィッティングする手法である．この手法を顔画像に対して適用することにより，高速・高精度な顔モデルフィッティングが実現できることが示されている[1]．人体に対してこの手法を適用することができれば，顔の場合と同様，高速・高精度な部位検出や姿勢推定が実現できることが期待される．しかしながら検討の結果，人体のもつ形状自由度や衣服等による特徴量変化がフィッティング性能に悪影響を与え，この手法をそのまま人体に適用することは困難であることが明らかになった．我々はこの問題に対して，特徴量サンプリング手法を改良し，HOG 特徴量をモデルの各ノード上でサンプリングすることにより，高精度なフィッティングを実現した．更に位置ばらつきや，大きな姿勢変動に対応するため，Coarse to Fine の考えに基づく段階的フィッティング手法を導入した．実験の結果，提案手法は従来技術に比べて高い検出精度を示し，フィッティング成功率が 50.0%から 74.0%に向上した．また処理時間は従来手法が 21.5 s であるのに対して，提案手法は 0.28 s であり，大幅な高速化を実現した．

キーワード　人体姿勢推定，人体モデル，モデルフィッティング

1. まえがき

カメラ画像を対象とした人体姿勢推定技術は，幅広い分野での活用が期待され，活発な研究が行われている．しかし人体のもつ形状自由度の多さ，服装や持ち物，隠れ等による特徴変動の多様さから，高速，高精度にその姿勢を推定することは，非常に困難な課題である．画像ベースで人体の姿勢を推定する手法としては体の関節にマーカーを装着し，この位置を検出するものや，複数カメラ間の情報を用いて 3D 形状を推定するものなどが存在する．しかし，これらの手法は画像撮影に特殊な装置や環境を必要とし，通常の環境での一般的なカメラによるアプリケーションには適していない．一般的な単一カメラ画像を対象とした人体姿勢推定技術としては特徴量として色を使うもの，輪郭を使うもの，あるいはこれらの複合的なもの，また部位位置の推定手法としてもシルエット情報を用いるもの，確率モデルによるもの，木構造モデルを用いるもの等，様々な手法が提案されている．

Chen [9] や Thuring [10] らは，背景除去により得られた人物画像のシルエットに基づき，シルエットにマッチするような姿勢を計算する手法を提案している．しかし，シルエットのみしか使用していないため，部位同士が接触したり，ある部位が別の部位で隠されているような状況に対応できない．

Mori ら [11] は入力画像に領域分割を適用し，あらかじめ用意した各部位のテクスチャ画像とのマッチングを行うことで，各部位位置の推定を行った．

このような背景除去や領域分割に基づく手法に対して，人体形状を木構造等でモデル化し，これを画像にフィッティングさせる手法の研究も広く行われてきた．Felzenszwalb ら [5] は部位の関連を，シンプルなスプリングモデルによる木構造で表現し，その形状変化に対するコストと各部位のアピアランスに対するコスト

† オムロン株式会社技術・知財本部，木津川市
Technology & Intellectual Property HQ, Omron Corporation, 9–1 Kizugawadai, Kizugawa-shi, 619–0283 Japan
†† 名古屋大学大学院情報科学研究科，名古屋市
Graduate School of Information Science, Nagoya University, Furou-cho, Chikusa-ku, Nagoya-shi, 460–3807 Japan
a) E-mail: kino@ari.ncl.omron.co.jp

の和を最小化することでモデルのフィッティングを行う枠組みを提案した．部位間の関係を木構造でモデル化し，同時に各部位に対してアピアランスモデルを適用する手法はPictorial Structureと呼ばれ，同様のアプローチはその後多くの研究で使用された．例えば[6]は部位関連性及びアピアランスの記述に確率表現を導入し，事後確率最大化によってフィッティングを行う手法を提案した．また[7]は部位特徴を表す特徴量としてカラーヒストグラムを使用するとともに，探索段階に応じて使用する特徴量を変化させ，探索を高速化させる手法を提案した．Ferrariら[13]は，Grabcut Segmentation [8]によりPictorial Structureの探索領域を限定することにより効率的にフィッティングする手法を提案した．

このように人体モデルのフィッティングに関して数多くの手法が提案されているものの，これらは全て探索空間内でコスト関数を徐々に最適化していく探索手法に基づいており，フィッティングにはなお多くの繰返し計算を必要とするため，リアルタイム処理は難しい．

そこで我々は，人体を特徴点分布の部分空間に基づく形状モデル[3]で表現し，更に摂動特徴量に基づく手法により高速にフィッティングを行う手法を提案する．形状モデルと摂動特徴量によるモデルフィッティングは，[1]によって提案され，比較的形状自由度や特徴量変動の少ない顔画像処理に対して有効性が示されているが，これまで人体を対象とした検討はなされてこなかった．

本論文ではこの手法をベースとした高速・高精度な人体モデルフィッティングの実現のための，具体的手法を述べる．我々は本手法を人体モデルに適用するにあたって，以下の項目について改良を行った．

特徴量サンプリング：顔と人体では有効な特徴量，及びそのサンプリング手法が異なると考えられ，人体に適した手法を適用する必要がある．検討の結果，[1]で使用されたHaar-like特徴及びStructured Retinotopic Samplingの手法は，衣服や背景の影響を受けやすく，人体に対して不適当であることが明らかになった．人体モデルに対してはHOG特徴量をモデルの各ノード上で取得する手法が有効であることが示された．

フィッティング手法：人体は顔と比較して，部位間の動き，姿勢の変動が非常に大きい．我々は大きな姿勢変動に対しても高精度なフィッティングを実現するため，Coarse to Fineの考えに基づいた段階的フィッティング手法を開発した．

これらの改良の結果，我々は摂動特徴量の手法に基づく，高速，高精度な人体モデルフィッティングの実現に成功した．性能検証の結果，提案手法は従来技術と比較して精度，速度共に優れた性能を示した．

なお本研究においては，フィッティングの対象を頭部及び胴体としたが，その理由は，人体の大まかな姿勢が必要となる多くの用途では，頭部と胴体のモデルフィッティングだけでも十分に役立つためである．また，動きの自由度の非常に大きい腕及び脚については，別の手法が必要になると考えられるために今回は対象外とする．

本論文では以下，**2.** で形状モデルについて概説し，**3.** で提案手法について全体の流れを示す．**4.** では人体モデルに適した特徴量サンプリング手法について検討を行い，**5.** において段階的フィッティング手法の提案を行う．最後に **6.** にて考察を行う．

2. 形状モデル

本研究では頭部，肩，腰等のパーツを人体の‘特徴点’と捉え，その位置を特定することを目的とする．本研究では人体パーツ間の関連を，一般に形状モデルと呼ばれる手法によって表現する．形状モデルはCootesらが[3]において顔画像に対するモデルフィッティングを目的として導入した．形状モデルを用いることにより，複数の特徴点配置を少数のパラメータによって近似的に表現することが可能となる．本研究においては頭部，肩等の人体各部位の位置をモデルのノードとすると，このノード配置は，適切な方法で移動，回転，縮小の操作を行って正規化した後，主成分分析（Principal Component Analysis：PCA）を用いることにより低次元で表現可能となる．今，ある点の座標を$[x_m, y_m]$とし，全点の座標をまとめて$\hat{\mathbf{x}} = [x_1, y_1, x_2, y_2, \ldots, x_M, y_M]^T (\in \mathcal{R}^{2M})$と書く．ここで$M$はノードの数である．また，ある画像$n$についてのノード座標を$\hat{\mathbf{x}}_n$とし，正規化を行った後のノード座標を$\mathbf{x}_n$と表す．分散共分散行列$\mathbf{\Sigma}$は，

$$\mathbf{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T \tag{1}$$

となる．Nは人体画像の数である．任意の正規化座標ベクトルは標準固有値問題，

$$\mathbf{\Sigma\Phi} = \mathbf{\Phi\Lambda} \quad (\mathbf{\Phi^T\Phi} = \mathbf{I}) \tag{2}$$

の解として得られる正規直交基底$\mathbf{\Phi}$を用いて，以下

のように表せる．

$$\mathbf{x} \approx \bar{\mathbf{x}} + \tilde{\mathbf{\Phi}}\tilde{\mathbf{b}} \tag{3}$$

ここで，$\tilde{\mathbf{\Phi}}$ は上位 k 個の固有値に対応する固有ベクトルから構成した基底ベクトル，$\tilde{\mathbf{b}}$ はそれに対応するパラメータベクトルである．したがって，元画像上のノード座標は，

$$\begin{aligned}\hat{\mathbf{x}} &\approx F(t_x, t_y, t_\theta, t_s, \tilde{\mathbf{b}}) \\ &= F(\mathbf{p})\end{aligned} \tag{4}$$

として相似変換パラメータと基底ベクトルに関するパラメータの関数として近似的に表現することが可能である．ただし，$\mathbf{p}^T = [t_x, t_y, t_\theta, t_s \mid \tilde{\mathbf{b}}^T]$ である．ここで，相似変換に関するパラメータを “Pose パラメータ”，基底ベクトルにかかるパラメータを “Shape パラメータ” と呼ぶ．

3. 摂動特徴量による人体モデルフィッティング

本章では摂動特徴量による人体モデルフィッティングについてその全体構成を示す．なおモデルの学習方法及び基本的なフィッティング手法は [1] に基づく．

形状モデルのモデルパラメータに対して，正解位置から様々な摂動を加えることによってずれを生じさせることを考える．このとき，モデルパラメータ摂動量と，その状態での特徴量（摂動特徴量）との相関関係を学習することによって，特徴量からモデルパラメータ摂動量を推定することが可能となる．これによって形状モデルがどれだけ正解形状からずれているかの推定ができ，大幅な形状修正による高速なフィッティングが実現できる．特徴量と形状モデルパラメータはどちらも多次元であるため，両者の相関関係の推定には多次元変量の回帰学習手法が必要となる．[1] では回帰学習として正準相関分析（Canonical Correlation Analysis：CCA）が用いられた．以下，CCA について概説した後，学習手順とフィッティング手順について述べる．

3.1 正準相関分析

p 次元の変量ベクトル $\mathbf{x} = [x_1, \cdots, x_p]^T$ と q 次元の変量ベクトル $\mathbf{y} = [y_1, \cdots, y_q]^T$ があるとき，この同時分布を考え，その分散共分散行列を

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{XX} & \mathbf{\Sigma}_{XY} \\ \mathbf{\Sigma}_{YX} & \mathbf{\Sigma}_{YY} \end{bmatrix} \tag{5}$$

とする．これらの変量ベクトルの任意の線形結合によって生成される新変量

$$u = \mathbf{a}^T\mathbf{x}, \quad v = \mathbf{b}^T\mathbf{y} \tag{6}$$

を考えたとき，両者の相関が最大になるような係数ベクトル $\mathbf{a}$，$\mathbf{b}$ を求める．そのためには共分散

$$Cov(u, v) = \mathbf{a}^T\mathbf{\Sigma}_{XY}\mathbf{b} \tag{7}$$

を最大にするような $\mathbf{a}$，$\mathbf{b}$ を求めればよい．この問題は，両者の分散を 1 に標準化し，ラグランジュの未定乗数法を用いることで一般固有値問題を解く問題に帰着される．

今，この固有値問題を解いて得られた第 i 固有値に対応する固有ベクトルを $\mathbf{a}_i, \mathbf{b}_i$ と書く．$p > q$ とし，第 q 正準変量まで求めるとすれば，元の変量空間 $\mathbf{x}, \mathbf{y}$ から新変量 $\mathbf{u}, \mathbf{v}$ への変換はそれぞれ

$$\begin{aligned}\mathbf{u} &= [\mathbf{a}_1, \cdots, \mathbf{a}_q]^T\mathbf{x} \\ &= \mathbf{A}^T\mathbf{x}\end{aligned} \tag{8}$$

$$\begin{aligned}\mathbf{v} &= [\mathbf{b}_1, \cdots, \mathbf{b}_q]^T\mathbf{y} \\ &= \mathbf{B}^T\mathbf{y}\end{aligned} \tag{9}$$

となる．$\mathbf{u}$ から $\mathbf{v}$ への線形回帰式は，

$$\begin{aligned}\mathbf{v} &= diag[\lambda_1, \cdots, \lambda_q]\mathbf{u} \\ &= \mathbf{\Lambda}\mathbf{u}\end{aligned} \tag{10}$$

で与えられる．以上より，$\mathbf{x} \Rightarrow \mathbf{y}$ の写像変換は

$$\mathbf{y} = \mathbf{G}\mathbf{x} \quad (\text{ただし}, \mathbf{G} = (\mathbf{B}^T)^{-1}\mathbf{\Lambda}\mathbf{A}^T) \tag{11}$$

のように表される．

3.2 学習手順

学習は，以下の手順で行う．なお，学習画像にはあらかじめ特徴点の正解座標が入力されているものとする．

1. $i = 1, n = 1$ とする．
2. n 枚目の画像に対して，式 (3) の関係を用い正解座標群をパラメータ空間に射影し，正解位置でのモデルパラメータ $\mathbf{p}_{\text{fit}}$ を求める．
3. 乱数により摂動 $\mathbf{\Delta p}_i$ を加えたパラメータ $\mathbf{p}_{\text{err}} = \mathbf{p}_{\text{fit}} + \mathbf{\Delta p}_i$ をもつ摂動モデルを生成する．
4. $\mathbf{p}_{\text{err}}$ の状態での特徴量 $\mathbf{f}_i$ を取得する．
5. $i \Leftarrow i + 1$
6. 3. 以降を R 回繰り返す．R は乱数の発生回数．

7. 2. 以降を $n = N$ まで繰り返す（N は画像の枚数）．

8. $\mathbf{\Delta p}$ と $\mathbf{f}$ の同時分布を考え，CCA により変換行列 $\mathbf{G}$ を求める．

ここで 4. では，モデルパラメータ $\mathbf{p}_{\mathrm{err}}$ によって生成されるモデル上での特徴量サンプリングを行う．具体的には，生成されたモデルの各ノード点を中心とした方形領域における HOG 特徴量をサンプリングする．これについては **4.** において詳述する．

3.3 フィッティング手法

以上の学習によって，特徴量と摂動によるパラメータ誤差を関連づける変換行列が得られた．本節では，これを用いてフィッティングを行う手法について述べる．なおここでは何らかの検出手段により既に画像上での人体位置は検出できているものとする．

1. $i=1$ とし，$\mathbf{p}_i = \mathbf{p}_{\mathrm{init}}$ とする（例えば $\mathbf{p}_{\mathrm{init}} = [\hat{t}_x, \hat{t}_y, \hat{t}_\theta, \hat{t}_s \mid \mathbf{0}^T]^T$，$\hat{t}_x, \hat{t}_y, \hat{t}_\theta$ は人体検出器によって求まった方形中心座標と方形の傾き，$\hat{t}_s$ は方形サイズに応じた平均的な値）．

2. $\mathbf{p}_i$ の状態での特徴量 $\mathbf{f}_i$ を取得する．

3. 変換行列 $\mathbf{G}$ と式 (11) より誤差の推定値 $\mathbf{\Delta p}_i$ を得る．

4. $\mathbf{p}_{i+1} = \mathbf{p}_i - \delta\mathbf{\Delta p}_i$ によりパラメータの修正を行う．

5. $i \Leftarrow i+1$

6. 2. 以降を終了条件を満足するまで繰り返す．

なお，終了条件としては $i > R$（R は繰返し上限回数）とする方法，$\|\mathbf{\Delta p}_i\| < \varepsilon$ とする方法，またあらかじめ正解パラメータでの特徴量識別器を学習しておき，この識別器の出力により終了判定する手法等が考えられる．

以上が基本的なフィッティングの流れであるが，人体は顔と比較して部位間の動き，姿勢の変動が大きく，また特徴量サンプリングの際に背景の影響を受けやすい．そのため単一のモデルでは十分な精度を得ることが難しい．本研究ではこの課題に対応するため，複数のモデルを切り換えて段階的にフィッティングする手法を提案する．これについては **5.** で詳述する．

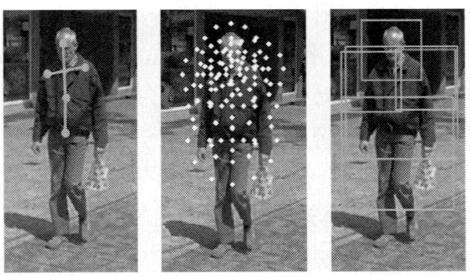
4. 人体モデルフィッティングのための特徴量

4.1 特徴量サンプリング方式

特徴点の周りに放射状にサンプリング点を配置する Retinotopic Sampling 手法は，Smeraldi ら [4] において顔特徴点検出を目的として提案され，特徴量としては Gabor 特徴量が使用された．[1] では顔モデルフィッティングの問題に対して，Retinotopic Sampling と Haar-like 特徴量が組み合わせて適用され，精度，処理速度ともに高い性能を示すことが示された．しかしながら人体を検出対象として考えた場合，Haar-like 特徴のように局所的な輝度差に着目した特徴量では十分な性能を発揮できないことが多くの研究により指摘されており，人体検出においては輝度差よりも，エッジ情報を使用すること必要性が指摘され，様々な特徴量が提案されている [15], [16]．中でも Dalal ら [2] の提案した HOG 特徴量は，人体検出に対して高い性能を示すことが示され，現在多くの活用研究が進められている．

4.2 特徴量比較実験

そこで本研究ではまず，摂動量推定に基づく人体モデルフィッティングに適した特徴量を検討するため，Gabor，Haar-like，及び HOG の各特徴量について，検出性能の比較を行った．比較に用いるための形状モデルは頭部，左肩，右肩，胴体中央及び腰の計 5 点から生成した．図 1 (a) に例を示す．TUD multiview データベース [19] の学習用画像のうち ‘front’，‘left-front’，‘right-front’ の 3 方向，計 1790 枚の画像を用いて学習を行った．画像一枚当りの学習回数は 10 回である．学習画像は頭部–腰間の距離が 24 pix となるようにサイズを正規化したのち，**3.2** に記述した手順によって摂動特徴量サンプリング，CCA による学習を行った．比較にあたっては，各手法間で条件に差が生じないように，モデル全体での特徴量次元数がほぼ同一になるように Retinotopic Sampling の点数及び特徴量パラメータを設定した．なお HOG 特徴量は，局所領域



(a) (b) (c)

図 1 (a) 人体形状モデル，(b) サンプリング点配置（Gabor 及び Haar-like），(c) サンプリング領域（HOG）

Fig. 1 (a) Body shape model, (b) Sampling point layout (for Gabor and Haar-like), (c) Sampling area (for HOG).

表 1　比較実験における特徴量次元数
Table 1　Feature dimensions for the experiment.

	Gabor	Haar-like	HOG
形状モデル点数	5	5	5
Retina 点数	25	25	1
1 点当り特徴量次元数	12	15	324
モデル特徴量次元数	1500	1875	1620

内のエッジ方向ヒストグラムを表すことから，1 サンプリングがカバーする領域が比較的広く，また特徴量次元も数百～数千と大きくなることが一般的である．そのため Retinotopic Sampling の目的の一つである，ある着目点周りで局所情報を抽出することを特徴抽出の時点で行っていると考えることができる．本実験では HOG 特徴量に対しては Retinotopic Sampling を用いず（Retinotopic Sampling の点数をノードと重なった 1 点とする），特徴量算出領域を他の手法の Retinotopic Sampling と同等の広がりをもつ領域とすることで比較を行った．表 1 に使用したサンプリング次元及び特徴量次元の一覧を示す．

Gabor は 3 周波×4 方向の 12 次元，Haar-lile は 5 種類×3 サイズの 15 次元とし，HOG に関してはサンプリング次元数を 324 とすることで，それぞれのモデル全体の特徴量次元がほぼ同一となるようにした．

性能比較に際しては，まずテスト画像に対して人体検出器を適用し，人体方形の抽出を行った．本実験で用いた人体検出器は，識別器としては Real AdaBoost [17]，特徴量には HOG を用い，上記と同様の TUD multiview データベース学習用画像を用いて学習を行ったものである．

抽出された人体方形のうち，検出成功と判定されたものに対して **3.3** の手順に従い人体モデルのフィッティングを行った．人体検出の正解判定は，PASCAL VOC challenge [17] に従い，正解方形とのオーバラップが 50%以上の物を正解とした．なお，正解方形は目視入力した頭頂点 (x_1, y_1) 及び左右足先中間点 (x_2, y_2) を基準に，$(x_1, \frac{y_1+y_2}{2})$ を方形中心として縦横比が 2:1 となるように作成した．

評価指標としては，頭部，肩，腰の全ての部位を正しく検出した場合にフィッティング成功とした．各部位ごとの正解判定は，検出位置が正解位置を中心とした半径 r の円内に入っている場合に正解とした．なお r は本実験では頭部–腰間の距離の 20%とした．これは頭部と同程度の大きさの円に相当する．

評価には，TUD multiview データベースの評価用

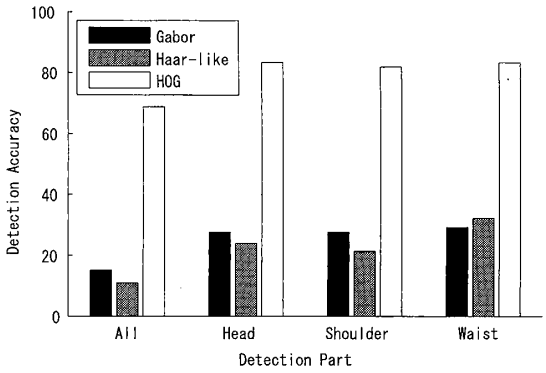


図 2　特徴量サンプリング手法によるフィッティング性能比較結果
Fig. 2　Model fitting accuracy comparison for feature extraction method.

画像のうち 'front', 'left-front', 'right-front' の 3 方向，計 101 枚の画像を用いた．このうち人体検出に成功したのは 96 枚であり，これらに対する検出精度の評価を行った．したがって，検出率算出の際の母数は 96 である．

4.3　特徴量比較実験結果

図 2 に性能評価結果を示す．図で，All はフィッティングの成功率，Head，Shoulder，Waist はそれぞれ頭部，肩，腰の検出成功率である．評価の結果，Gabor 及び Haar-like 特徴による Retina Sampling では，フィッティング成功率が 15%程度にとどまり，モデルフィッティングが有効に動作しないことが明らかとなった．その一方 HOG 特徴を利用した場合では成功率は 70%程度となり，高い性能を示した．以上より，人体モデルフィッティングに関しては，各ノードにおいて HOG 特徴をサンプリングする手法が有効であるという結果が得られた．

5.　段階的フィッティングによる精度向上

5.1　段階的フィッティング

人体は顔と比較して，部位間の動き，姿勢の変動が非常に大きく，また特徴量サンプリングの際に背景の影響を受けやすい．初期配置が正解位置から遠く，更に大きな姿勢変動がある場合でもフィッティング可能なロバストなモデルを学習しようとすると，学習に用いる特徴量の大部分は背景からくるノイズが占めることとなり，結果として学習されたモデルは，大まかな位置・姿勢推定は可能であるが，詳細な位置推定には不向きなものとなる．これに対して詳細な位置推定が可能なモデルを学習するために，正解位置近傍だけの

表 2　モデル学習時の変動量
Table 2　Parameter perturbations for the model learning.

	位置 (xy)	回転	スケール	形状
Rough モデル	モデル高さの 15%	±10 deg.	±30%	3σ
Fine モデル	モデル高さの 10%	±5 deg.	±10%	1σ

摂動でモデルを学習すると，大きな位置・姿勢変化に対応できない．

このような課題に対応するため，我々は Coarse to Fine の考えに基づいた段階的フィッティング手法を提案する．手法の概要は次のとおりである．

- 大きな位置/姿勢変動に対応したモデルを学習する．ここでは Rough モデルと呼ぶ．
- 小さな位置/姿勢変動に対応したモデルを学習する．ここでは Fine モデルと呼ぶ．
- 検出の際は，Rough モデルによってフィッティングを始め，途中で Fine モデルに切り換えてフィッティングを行う．

モデルの切換に関しては，繰返し回数を最小化するために，Fine モデルの対応範囲に入ったことを何らかの手段により検知し，適応的に切り換えることが本来望ましいと考えられる．しかしながら本研究では初期検討としてシンプルな手法を採用し，各モデル一定回数の繰返しを行うものとした．具体的には各モデルとも繰返し上限回数を $R = 5$ とし，無条件でこの回数のフィッティングを行うものとした．また更新重みは $\delta = 0.4$ とした．

本研究で Rough，Fine それぞれのモデル学習に用いた変動量を表 2 に示す．

5.2　性能評価実験

提案手法の有効性を確認するために，性能評価を実施した．**4.2** と同様に人体検出器によって人体方形を抽出し，それに対するモデルフィッティングを行い，頭部，肩，及び腰の各点の検出精度及び処理時間を評価した．

性能比較のための従来手法としては，Ferrari らの Progressive Search [13] を取り上げる．この手法は Pictorial Structure をベースとし，色及びエッジ情報をもとにしたモデルの最適配置探索を行う．Grabcut Segmentation による前景抽出を行い，探索領域を削減することによって大幅な高速化が図られていることが特長である．Web 上に作者による MATLAB ソースコードが公開されており [14]，これによって任意の人体画像に対して部位位置推定を実行することが可能である．このプログラムは，上半身画像に対して，頭部，胴体，左右上腕，左右前腕の計 6 部位を推定し，その位置をラインセグメントとして出力する．図 3 に検出結果例を示す．本実験では，このうち頭部セグメントの中心点を頭部として扱い，上腕セグメントの上部の点を肩，胴体セグメントの下部の点を腰とみなして精度比較を行った．評価の際は前提条件を同一とするため，我々と同一の人体検出器を用いた．ただし Ferrari らの手法は入力として上半身方形座標を必要とするため，[14] に記述された上半身方形の定義に基づき，人体方形座標から上半身方形を生成し，これを入力情報として用いた．その他の実験条件は **4.2** に示したものと同一である．

図 3　Ferrari ら [13] による検出結果例
Fig. 3　Detection result by Ferrari [13].

評価を行った手法は以下の 3 種類である．

（1）　摂動量推定によるモデルフィッティング（段階的フィッティングあり）

（2）　摂動量推定によるモデルフィッティング（段階的フィッティングなし）

（3）　Ferrari らの手法

なお（1）及び（2）で使用した特徴量は HOG である．（2）については Rough モデルのみを使用し，（1）と条件を揃えるために繰返し上限回数を $R = 10$ とした．図 4 に評価結果を示す．図で，ASAM multi が（1）の結果を，ASAM single が（2），Ferrari が（3）をそれぞれ表す．フィッティング成功率（図で All と表示）に関して，Ferrari らの手法が 50.0%を示したのに対して，摂動量推定に基づく手法の精度はいずれもこれを上回った．なかでも段階的フィッティングを導入した提案手法は 74.0%と最も良い性能を示し，段階的フィッティングのない場合（64.6%）に比べて 9 ポイント程度の精度向上が見られた．また，各部位に対する検出精度も，段階的フィッティングを導入した提

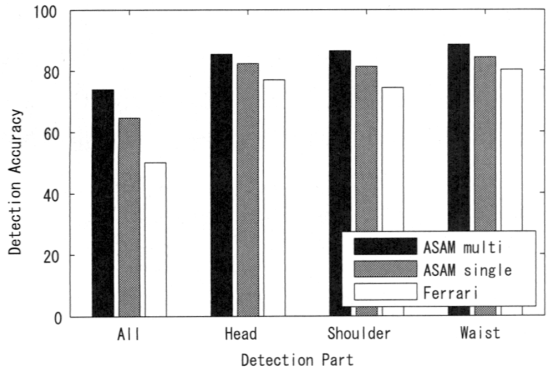


図 4　性能比較結果
Fig. 4　Evaluation results.

図 5　検出結果例．左：ASAM multi，右：Ferrari
Fig. 5　Model fitting result comparison. Left: ASAM multi. Right: Ferrari.

案手法が最も良い性能を示した．以上より，本論文で提案する摂動特徴量による人体モデルフィッティング手法の優位性，及び段階的モデルフィッティングの効果が確認できる．

なお処理時間は提案手法が 0.28 s，Ferrari らの手法は 21.5 s であった（Core i5，2.67 GHz CPU，4.0 GB RAM，両者とも MATLAB により実装）．Ferrari らの手法が腕位置の推定まで行っていることを考慮しても，提案手法は大幅な高速化を実現できていることが分かる．

図 5 に提案手法及び Ferrari らの手法による検出結果例を示す．Ferrari らの手法がフィッティングに失敗する複雑な背景の画像でも，提案手法は正しくフィッティングする例が複数見られた．

6. 考　　察

本章では HOG 特徴量を用いた摂動特徴量によるモデルフィッティング，及びそれを段階的に行うことが有効な理由について考察を行う．

4.2 の実験結果において，Retinotopic Sampling により Haar-like 特徴量をサンプリングする手法は，人体モデルフィッティングに対して非常に低い性能した．その一方，HOG 特徴量をモデルの各ノード上でサンプリングする手法は，高い性能を示した．前者の手法が顔モデルフィッティングに対しては有効であるにもかかわらず，本研究で低い性能を示した理由としては，特徴量及びサンプリング手法が，それぞれ以下のような理由により人体に対しては適切でなかったためと考えられる．

特徴量：Haar-like 特徴は局所的な濃淡パターンに着目した特徴量であるが，衣服と背景との組合せや衣服の色等によって，濃淡パターンの方向，強度は一定でなく，逆転するケースも生じる．そのため同じような位置，姿勢であっても特徴量は様々な値をとり，有効な情報が得られない．

サンプリング手法：Retinotopic Sampling は着目点を中心とした放射状のサンプリング点配置を行う．そのため多数のサンプリング点は人体領域内部に存在し，それらは衣服の状態に応じてランダムな値をとり，ノイズ要因となる．また残りの大部分の点も背景領域に配置されることでノイズとなり，特徴となり得る人体輪郭に関する情報が占める割合が低い．

これに対して，HOG 特徴量をモデルの各ノード上でサンプリングする手法は，以下のような理由により人体モデルフィッティングに適しているものと考えられる．

特徴量：エッジ情報を利用するため，濃淡パターンの変化に対してロバスト性をもつ．

サンプリング手法：ノードを中心とした局所領域内での方向分布を抽出するため，人体輪郭に関する情報が豊富に得られる．またサンプリング位置，角度，サイズの変化による輪郭の位置，形状変化が特徴量に反映され，この関係が摂動特徴量学習に利用できる．

また **5.2** の評価において，段階的フィッティングによる検出精度は，単一のモデルによってフィッティングする場合と比較して 9 ポイント程度よい性能を示した．これは，提案手法によりフィッティングのロバスト性を保ちながら，位置，形状の微細なずれに対するフィッティング能力が向上した結果であると考えられる．ラスタスキャンによる人体検出は検出結果の位置，大きさにある程度変動をもつが，提案手法によってその変動を吸収しながら，高い検出性能を得ることが可能になっている．

図 6，図 7 に提案手法による検出成功例及び失敗例を示す．失敗の傾向として，人体と重なる背景に強いエッジが存在している場合や，人体の一部でエッジ強度が弱い場合に，フィッティングに失敗する例が多かった．この理由としては，HOG 特徴量は方向分布作成

図 6　検出成功例
Fig. 6　Model fitting success examples.

図 7　検出失敗例
Fig. 7　Model fitting failure examples.

の際，エッジ強度による正規化を行うため，人体輪郭がその局所領域で支配的でない場合，有用な特徴量が得られないためであると考えられる．同様の理由により，衣服の上に強い斜めのライン（鞄の肩紐等）がある場合に，モデルがそれに沿って回転してしまう例が見られた．また人物が重なって映っている場合に，モデルの一部が別な人物上に引きずられる例，肩のラインが全く見えない場合にフィッティングに失敗する例が見られた．このようなケースに対応するためには，特徴量のサンプリングをより人体輪郭上に限定することが有効であると考えられる．例えば各ノードにおけるサンプリングにおいて，方形の周辺領域のみからHOG 特徴量を算出する，などの対策が考えられるが，これについての検証は今後の課題とする．

7. むすび

人体を特徴点分布の部分空間に基づく形状モデルで表現し，更に摂動特徴量に基づく手法により高速にフィッティングを行う手法を提案した．検討の結果，顔モデルフィッティングで有効であった特徴量サンプリング手法は人体に対しては不適切であり，HOG 特徴量をモデルの各ノード上でサンプリングする手法により，人体に対してのモデルフィッティングが実現できることを確認した．更に人体検出による位置ばらつきや，大きな姿勢変動がある状況でも高精度なフィッティングを実現するため，Coarse to Fine の考えに基づく段階的フィッティング手法を提案した．実験の結果，提案手法は従来技術に比べて高い性能を示し，従来技術と比較して，フィッティング成功率が 50.0%から 74.0%に向上し，段階的フィッティングのない場合と比較しておよそ 9 ポイントの性能向上が得られた．処理時間は従来手法が 21.5 s，提案手法は 0.28 s であり，大幅な高速化を実現した．今後は手，足を含むモデルフィッティング，及びより大きな姿勢変動や向きの変動を含む画像への対応が課題である．

文　献

[1] 木下航一，小西嘉典，勞　世竑，川出雅人，村瀬　洋，“摂動特徴量による顔画像に対する形状モデルフィッティング，”信学論（D），vol.J94-D, no.4, pp.721–729, April 2011.

[2] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” International Conference on Computer Vision & Pattern Recognition, vol.2, pp.886–893, June 2005.

[3] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, “Active shape models - Their training and application,” Computer Vision and Image Understanding, vol.6, no.1, pp.38–59, 1995.

[4] F. Smeraldi and J. Bigun, “Retinal vision applied to facial features detection and face authentication,” Pattern Recognit. Lett., vol.23, pp.463–475, Feb. 2002.

[5] P.F. Felzenszwalb and D.P. Huttenlocher, “Efficient matching of pictorial structures,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2000.

[6] P.F. Felzenszwalb and D.P. Huttenlocher, “Pictorial structures for object recognition,” IJCV, vol.61, no.1, pp.55–79, Jan. 2005.

[7] D. Ramanan, “Learning to parse images of articulated bodies,” NIPS, 2006.

[8] C. Rother, V. Kolmogorov, and A. Blake, “Grabcut: Interactive foreground extraction using iterated graph cuts,” SIGGRAPH, 2004.

[9] Y. Chen, J. Lee, R. Parent, and R. Machiraju, “Markerless monocular motion capture using image features and physical constraints,” Computer Graphics International, pp.36–43, 2005.

[10] S. Thuring, J. Herwig, and A. Schmitt, “Silhouette-based motion capture for interactive VR-Systems including a rear projection screen,” Computer Animation and Virtual Worlds, vol.16, Issue 3-4, pp.245–257, July 2005.

[11] G. Mori, X. Ren, A. Efros, and J. Malik, “Recovering human body configurations: Combining segmentation and recognition,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2004.

[12] D. Ferrari, “Learning to parse images of articulated bodies,” NIPS, 2006.

[13] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, “Progressive search space reduction for human pose estimation,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2008.

[14] http://www.robots.ox.ac.uk/~vgg/research/

pose_estimation/index.html

[15] B. Wu and R. Nevatia, "Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors," Proc. IEEE International Conference of Computer Vision, 2005.

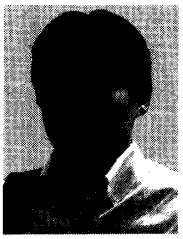
[16] P. Sabzmeydani and G. Mori, "Detecting pedestrians by learning shapelet features," International Conference on Computer Vision & Pattern Recognition, 2007.

[17] M. Everingham, L.V. Gool, C.K.I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," Int. J. Comput. Vis., vol.88, no.2, pp.303–338, 2010.

[18] R.E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," Mach. Learn., no.37, pp.297–336, 1999.

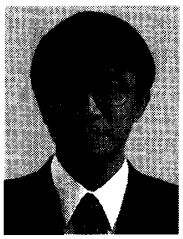
[19] M. Andriluka, S. Roth, and B. Schiele, "Monocular 3D pose estimation and tracking by detection," IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010), San Francisco, USA, June 2010.

（平成 24 年 11 月 30 日受付，25 年 3 月 14 日再受付）



木下　航一

平 8 神戸大・工・システム卒．平 10 同大大学院博士前期課程了．同年オムロン（株）入社，現在に至る．平 22 より名大・情・メディア博士後期課程所属．主として画像認識の研究開発に従事．平 20 MIRU ベストインタラクティブセッション賞受賞．平 20 SSII 優秀論文賞受賞．平 21 SSII 高木賞受賞．



村瀬　洋　（正員：フェロー）

昭 53 名大・工・電気卒．昭 55 同大大学院修士課程了．同年日本電信電話公社（現 NTT）入社．平 4 から 1 年間米国コロンビア大客員研究員．平 15 から名古屋大学大学院情報科学研究科教授．現在に至る．文字・図形認識，コンピュータビジョン，マルチメディア認識の研究に従事．工博．昭 60 電子情報通信学会学術奨励賞，平 6 IEEE-CVPR 最優秀論文賞，平 7 情報処理学会山下記念研究賞，平 8 IEEE-ICRA 最優秀ビデオ賞，平 13 高柳記念奨励賞，平 13 本会ソサイエティ論文賞，平 14 電子情報通信学会業績賞，平 15 文部科学大臣賞，平 16 IEEE Trans. MM 論文賞，平 22 前島賞，平 24 紫綬褒章，他受賞．IEEE フェロー，情報処理学会会員．