回帰分析とパーティクルフィルタを用いた 単眼画像からの人体3次元姿勢推定

大西　克則†　　滝口　哲也††　　有木　康雄††

† 神戸大学大学院工学研究科 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@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

あらまし　本論文では，単眼画像からマーカー等を用いずに人体の 3 次元姿勢を推定する手法を提案する．このような条件で推定を行う場合に画像特徴は非常に重要であり，本論文では物体の大まかな形状を表すことが可能な HOG 特徴を用いる．3 次元姿勢は，人体を多関節モデルで表現しその各関節角で記述する．従来，姿勢推定方法は大きく分けて二つのアプローチがある．一つは画像を観測し，3D 姿勢を推定する Bottom-Up アプローチであり，もう一つは逆向きの推定を行う Top-Down アプローチである．本論文では，Bottom-Up で大まかな姿勢を推定し，Top-Down でより精度を向上させる，双方向から推定する手法を提案する．Bottom-Up アプローチでは，重回帰分析を用い，Top-Down アプローチでは，パーティクルフィルタを用いることで，より高精度な推定を行う．実験には CMU Graphics Lab Motion Capture Database のデータを用いた．
キーワード　姿勢推定，HOG，重回帰分析，パーティクルフィルター

3D Human Pose Estimation from Monocular Image Using Regression Analysis and Particle Filter

Katsunori ONISHI†, 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@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

Abstract　In this paper, we propose a method to estimate 3D human pose from a monocular image without using markers. The framework requires that the visual features should be powerful enough to discriminate the subtle differences between similar human poses. We use HOG feature to describe the visual features. A 3D human body is expressed by a multi-joint model, and a set of the joint angles describes a pose. Generally, the human pose estimation method has two approaches. One is bottom-up approach that observes the image, and estimates the 3D pose. The other is top-down approach that estimates it in the opposite direction. In this paper, we propose a method to estimate the human pose by integrating the bottom-up and top-down approach to improve the pose precision. In the bottom-up approach, we use linear regression analysis to model the mapping from visual observations to human poses. On the other hand, in the top-down approach, the particle filter is employed. We use the CMU Graphics Lab Motion Capture Database for training.
Key words　Pose estimation, HOG, linear regression analysis, particle filter

1. はじめに

近年，人体の 3 次元姿勢推定に関する研究が盛んにおこなわれている [2]．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，ゲームや映画等のエンターテイメント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．こういった仮想世界の構築を行うためには人体全身に対する姿勢推定が必要である．

姿勢推定問題において，人体の姿勢を正確に表す特徴量を抽出することが非常に重要となる．従来手法では，人体の特徴量に，シルエットを用いる手法 [6] [7] [8] [9] [12] [13]

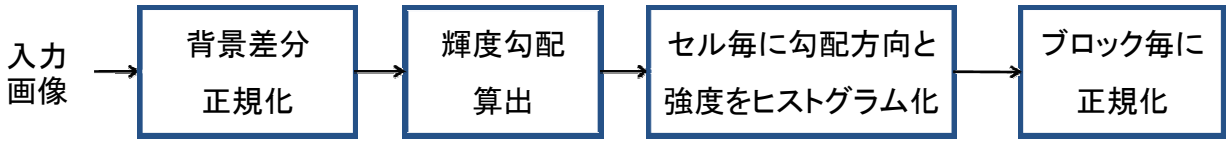


図 1　HOG 特徴抽出の流れ

が最も多く用いられている．この手法の利点として，特徴抽出の際にアピアランスや照明変化に頑健であるという点があげられる．また，シルエット画像の抽出には背景差分法を用いることが一般的である．しかし，シルエット画像は輪郭形状のみの情報になるため，内部のエッジ情報を失うことにより，推定の際にあいまいな部分が生じてしまう．このシルエットの欠点を解決するためには，内部のエッジ情報も考慮した特徴量が必要となる．そこで本論文では，アピアランスに基づく特徴として，物体の形状を表すことが可能な Histograms of Oriented Gradients (HOG) [3] を用いる．HOG 特徴は内部のエッジ情報も考慮しているため，3D 姿勢を安定して推定可能である．また，HOG 特徴は照明変化にも頑健な特徴となっている．

従来用いられている人体の姿勢推定方法は，*bottom-up* と *top-down* のふたつのアプローチに分けることができる．bottom-up アプローチ [6] [8] [10] では，学習データを用いて事後確率密度関数をモデル化し，画像から 3D 姿勢に直接写像するという手法がある．一度学習をしてしまえば，非常に高速な姿勢推定が可能となる．しかし，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．一方，top-down アプローチ [11] [12] [13] では，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方法などがある．これらの手法は，一般的に計算コストがかかるという欠点がある．そこで，これらの問題に対応するために，我々は bottom-up と top-down を組み合わせて推定する手法を提案する．Top-down において問題点となる初期値の決定を bottom-up で高速に推定し，その後推定精度を向上させるために，top-down に切り替えて推定を行う．Bottom-up では重回帰分析を，top-down ではパーティクルフィルタを用いて推定を行う．本手法では，bottom-up である程度の推定を行っているため，top-down でも計算コストを抑えることが可能である．

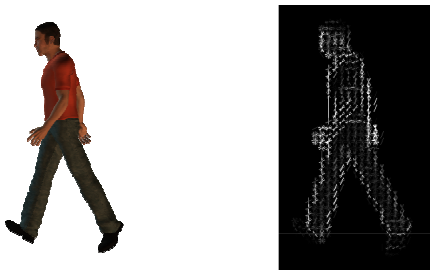
学習データを作成する際に CMU Graphics Lab Motion Capture Database [1] で公開されているモーションキャプチャデータを用いた．このデータからコンピュータグラフィックス (CG) を用いて実験を行った．

2. 特　徴　量

本章では，1 枚の画像から抽出する特徴量 HOG と，3 次元の人体を表現するための 3D 人体ボーンモデルについて述べる．

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) [3] や Scale-Invariant Feature Transform (SIFT) [5] が提案されている．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 2 に表すように，HOG は物体の大まかな形状を表すことが可能となる．図 1 に特徴抽出の簡単な流れを示す．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．



(a) 入力画像　　(b)HOG 特徴

図 2　人体形状特徴

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．

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

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．

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

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

このようにして得られた勾配方向 $\theta(x,y)$ は，$[-180°, 180°]$ 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°,180°] の範囲にする．

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

2.1.2 セルによるヒストグラム化

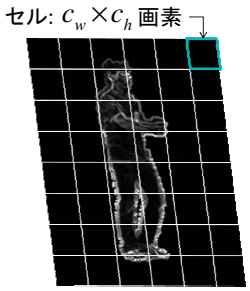


図 3　セル分割画像

算出された輝度勾配画像を，図 3 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

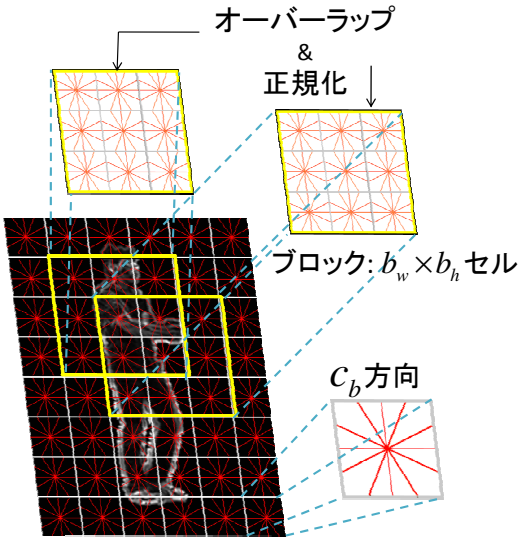


図 4　ブロックによる正規化

セルごとに作成された輝度勾配ヒストグラムを，図 4 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数は $d_b = b_w \times b_h \times c_b$ 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 $(i,j), \{1 \leqq i \leqq b_w, 1 \leqq j \leqq b_h\}$ 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．

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

正規化の際，ブロックは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 人体 3 次元モデル

図 5　3 次元人体モデル

人体は，複数の関節をも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 次元モデル特徴では，人体の構造に基づいて各関節 (肘，腰，膝など) の角度を特徴量 $\mathbf{x}$ とする．つまり，この関節角を変更することにより，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．このデータは人体を 56 次元で表現しているが，指，手首，つま先など，姿勢推定に不必要な関節角は用いず，38 次元とする．図 5 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

本章では，3D 姿勢の推定方法について述べる．一般的に，人体の姿勢推定方法は大きく分けて二つのアプローチがある．一つは画像を観測し，3D 姿勢を推定する Bottom-Up アプローチであり，もう一つは逆向きの

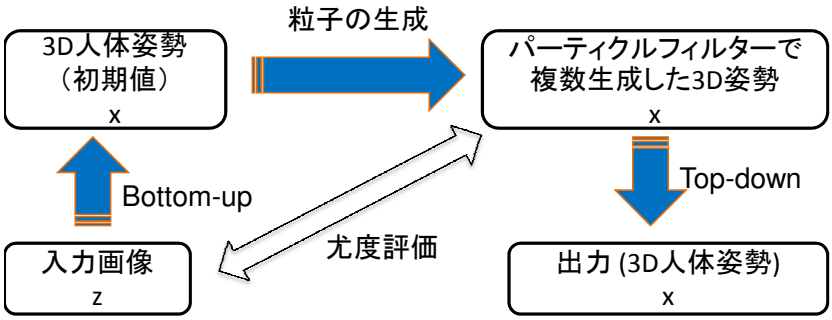


図 6　姿勢推定システム

推定を行う Top-Down アプローチである．本論文では，Bottom-Up で大まかな姿勢を推定し，Top-Down でより精度を向上させる，双方向から推定する手法を提案する．図 6 に推定システムの概要を示す．

3.1　Bottom-up アプローチ

Bottom-up アプローチでは，[8] の手法と同様に，回帰分析を用いて姿勢推定を行う (図 7)．この手法は，画像から得られた特徴量から，直接 3 次元姿勢を復元できるという利点があり，非常に高速に推定が可能である．

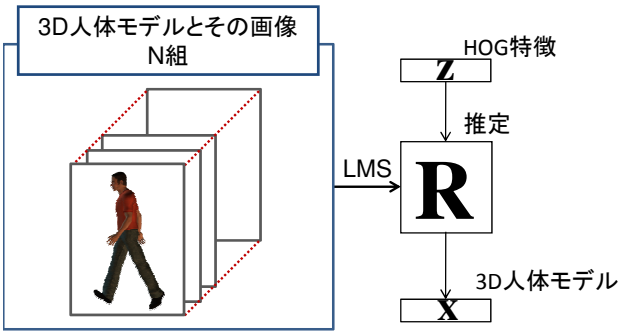


図 7　姿勢推定方法

画像から得られた HOG 特徴ベクトル $\mathbf{z} \in \mathbb{R}^d$ と人体 3 次元モデルの特徴ベクトル $\mathbf{x} \in \mathbb{R}^m$ の関係を，次式で近似する．

$$\mathbf{x} = \mathbf{R}\mathbf{z} + \varepsilon \tag{6}$$

このとき，$\mathbf{R}$ は $m \times d$ 行列，ε は残差ベクトルである．画像特徴 $\mathbf{z}$ が入力されれば人体 3 次元モデル特徴 $\mathbf{x}$ 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 $\{(\mathbf{x}_i, \mathbf{z}_i) \mid i = 1 \cdots n\}$(この場合，3D 人体関節角と画像の HOG 特徴を指す) を用いて，次式の最小自乗推定法により，$\mathbf{R}$ の値を決定する．

$$\mathbf{R} := \arg\min_{\mathbf{R}} \sum_{i=1}^{n} \|\mathbf{R}\mathbf{z}_i - \mathbf{x}_i\|^2 \tag{7}$$

学習データの特徴ベクトルを $m \times n$ 行列の $\mathbf{X} \equiv (\mathbf{x_1}\ \mathbf{x_2}\ \cdots\ \mathbf{x_n})$ と $d \times n$ 行列の $\mathbf{Z} \equiv (\mathbf{z_1}\ \mathbf{z_2}\ \cdots\ \mathbf{z_n})$ と表すことにより，推定問題は以下のように表せる．

$$\mathbf{R} := \arg\min_{\mathbf{R}} \|\mathbf{R}\mathbf{Z} - \mathbf{X}\|^2 \tag{8}$$

画像から入力された HOG 特徴ベクトル $\mathbf{z}$ に対して，求められた変換行列 $\mathbf{R}$ を用いて変換することにより $\mathbf{x}$ を推定する．これにより人体の 3 次元姿勢推定を行う．この推定結果を，top-down で推定する際の初期値として用いる．

3.2　Top-down アプローチ

3.2.1　パーティクルフィルタ

Top-down アプローチでは，パーティクルフィルタ (CONDENSATION 法) [14] [15] を用いて推定を行う．[14] の手法と同様に，時間 t における状態ベクトルを x_t，同時刻に観測される画像特徴を z_t とする．また，時刻 t までに得られる画像特徴の系列を $Z_t = (z_1, \cdots, z_t)$ とする．このとき，x_t の事後確率分布 $p(x_t|Z_t)$ は，ベイズの定理により式 (9) のようになる．

$$p(x_t|Z_t) \propto p(z_t|x_t)p(x_t|Z_{t-1}) \tag{9}$$

ここで，$p(z_t|x_t)$ は z_t の尤度，$p(x_t|Z_{t-1})$ は事前分布（予測確率）である．このとき，$p(x_t|Z_{t-1})$ は前の時刻 $t-1$ の事後分布 $P(x_{t-1}|Z_{t-1})$ と，時間が経過するときの分布の推移確率 $p(x_t|x_{t-1})$ から，次式のように求めることができる．

$$p(x_t|Z_{t-1}) = \int_{x_{t-1}} p(x_t|x_{t-1})p(x_{t-1}|Z_{t-1})dx_{t-1} \tag{10}$$

この手法では，事前分布に基づいてランダムサンプリングを行い，各サンプルに対して尤度を求めることによって，離散的に事後確率分布を推定する．これは，各時刻 t における，N 個のサンプル群 $s_t^{(n)}(n = 1, ..., N)$ とその尤度評価によって得られる重み $\pi_t^{(n)}$ を用いて，式

11 のように表される．

$$p(x_t|Z_t) \approx \sum_{n=1}^{N} \pi_t^{(n)} \delta(s_t^{(n)}) \qquad (11)$$

ただし，δ はディラックのデルタ関数である．
パーティクルフィルタの概念図を図 8 に示す．

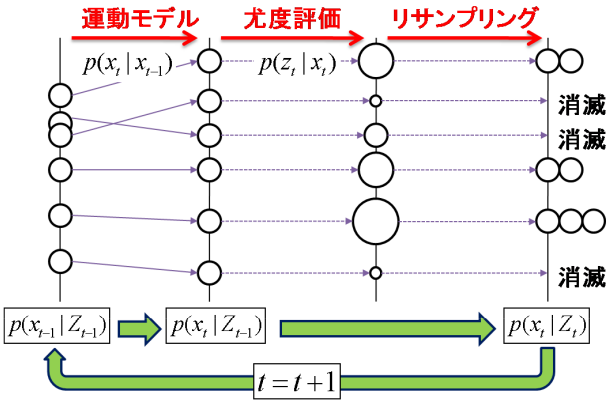


図 8　パーティクルフィルタ

3.2.2　パーティクルフィルタによる姿勢推定

Bottom-up アプローチの重回帰分析を用いた推定により得られた 3D 姿勢を初期値とし，その値の周囲に粒子をサンプリングする．このとき，各粒子の状態量は 3D 人体の間接角度を表すものであるため，人体の関節可動域を超えない範囲でサンプリングを行う．各粒子の尤度を計算し，それぞれの重み $\pi_t^{(n)}$ と重み付き平均を求める．それぞれの重みに基づいて，各粒子のリサンプリング，状態遷移を繰り返し行う．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．

4.　実　　験

4.1　実 験 方 法

本手法を用いた実験方法について述べる．実験は CMU のモーションキャプチャデータベース [1] のデータから，CG を用いて画像を生成して行った．CG 画像では，640 × 480 画素の解像度を持つ映像を作成した．bottom-up アプローチでは，重回帰分析における変換行列 **R** を決定するための学習データが必要となる．そこで，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 方向から見たデータを学習データとして用いる．それぞれの方向に対して，歩く，走る，キックの 3 動作を行い，学習に用いる．使用したデータの詳細を，表 4.1 に示す．

生成した映像に対し，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 に正規化する．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．

表 1　学習データ数

姿勢	フレーム数	
	1 方向	Total (8 方向)
歩く	316	2528
走る	148	1184
キック	801	6408
Total	1265	10120

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する．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，3600 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の数は 800，繰り返し数は 10 回とした．

推定した 3D 人体姿勢から CG を用いて生成した画像と，入力画像の一致度を測ることで，bottom-up と top-down を組み合わせた提案手法の有効性を示す．評価実験に用いるテストデータも，学習データと同様に CMU のデータベースを使用する．テストデータには，歩行画像 80 枚，走る画像 60 枚を用いた．また，市販されている一般的なカメラで撮影された映像を用いての実験も行った．

4.2　実 験 結 果

実験により，本手法が有効であることを確かめる．実験は，歩行，走る画像を合わせて 140 枚用いて行った．最終的に推定された 3D 人体姿勢から CG を用いて画像を作成し，入力画像と画素毎に一致度を測ることで評価を行った．図 9 にその実験結果のグラフを示す．

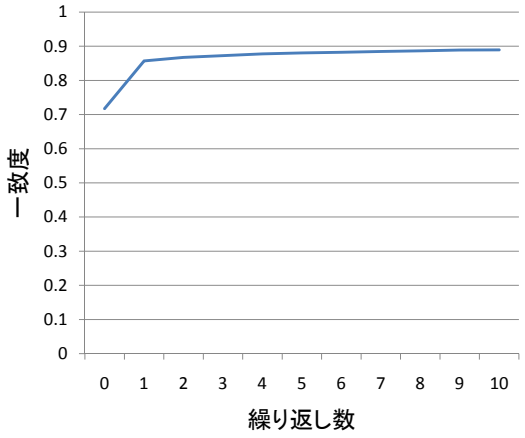


図 9　評 価 実 験

図 9 において，繰り返し数が 0 となっている部分は bottom-up の推定結果となっている．グラフからも読み取れるように，提案手法を用いることで，より入力画像に近い姿勢を得ることが可能となった．重回帰分析を用

いて初期値を決定しているため，top-down の推定に移る前におおまかな推定がなされていることも確認できる．初期値が決定されている分，収束するまでにかかる時間も抑えることができたと考えられる．

また，テスト画像に CMU のモーションキャプチャデータベースから生成した CG 画像だけでなく，一般に市販されている 640 × 480 画素の解像度を持つ単眼のカメラで撮影した映像も用いて実験を行った．この場合も CG 画像での実験と同様に，背景画像を用意しておき，背景差分法により人体領域の抽出を行う．実験結果の画像の一例を，図 10 に示す．

(a) (b) (c)

図 10　実験結果画像

(a) 入力画像，(b) bottom-up アプローチのみの推定結果，(c) 提案手法

CG で生成した画像とは異なり，動きの速い部分で画像にブレなどが生じるが，図 10 に示すように，CG 画像でなくても推定できている．実験結果から，CG 画像，実画像の両方で提案手法が有効であることが確認できた．

5. ま と め

我々は，単眼画像からの人体 3 次元姿勢推定問題において，bottom-up と top-down アプローチを組み合わせる方法を提案した．双方向からアプローチをすることにより，両者の欠点を補うことで推定精度を向上させた．また，従来画像から抽出する特徴量はシルエット画像を用いることが一般的であったが，本論文では HOG 特徴を用いることで輪郭内部のエッジも特徴として抽出している．これにより，変形の激しい人体のオクルージョン問題にも対応できた．

本手法を用いることで bottom-up と top-down は高速に推定可能となったが，CG 画像を用いているため，大部分の計算コスト CG 画像の生成に費やしている．また，本手法では初期値が非常に重要となっており，その値次第では容易に収束したり，またその逆も考えられる．そのため，今後の予定は bottom-up アプローチの推定精度の向上をする必要がある．それにより，パーティクルフィルタにおける繰り返し数を減らすことができ，計算コストを抑えることができ，収束しやすくなると考えられる．それらの問題が解決できれば，動画での姿勢推定に対応していく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] 島田伸敬，有田大作，玉木徹，“ 関節物体のモデルフィッティング，”研究報告-Computer Vision and Image Media, Vol.2006, No.51, pp.375-392, 2006.

[3] N.Dalal and B.Triggs，”Histograms of Oriented Gradients for Human Detection,” IEEE Computer Vision and Pattern Recognition, 886-893, 2005.

[4] 山内悠嗣, 藤吉弘亘, H.Bon-Woo, 金出武雄, “アピアランスと時空間特徴の共起に基づく人検出,” 画像の認識・理解シンポジウム (MIRU2007), 1492-1497, 2007.

[5] D.Lowe, ”Distinctive Image Features from Scale-Invariant Keypoints,” International Journal of Computer Vision, 60(2), 91-110, 2004.

[6] C.Sminchisescu, A.Kanaujia, D.N.Metaxas, ”BM^3E : Discriminative Density Propagation for Visual Tracking,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2030-2044, 2007.

[7] G.Mori and J.Malik, ”Recovering 3D Human Body Configurations using Shape Contexts,” IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL.28, No.7, 1052-1062, 2006,

[8] 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.

[9] R.Poppe, ”Evaluating Example-based Pose Estimation: Experiments on the HumanEva Sets,” CVPR 2nd Workshop on Evaluation of Articulated Human Motion and Pose Estimation, 2007.

[10] 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.

[11] 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.

[12] T.Jaeggli, E.Koller-Meier, L.V.Gool, ”Learning Generative Models for Monocular Body Pose Estimation,” Asian Conference on Computer Vision, November, 2007.

[13] X.Zhao, Y.Liu, ”Tracking 3D Human Motion in Compact Base Space,” Proceedings of the Eighth IEEE Workshop on Applications of Computer Vision, Feb. 2007.

[14] Michael Isard and Andrew Blake, ”CONDENSATION-Conditional Density Propagation for Visual Tracking,” Int. J. Computer Vision, 5-28, 1998.

[15] 樋口知之，”粒子フィルタ，” 電子情報通信学会誌, vol.88, No.12, pp.989-994, 2005.