C-02

Bottom-up と Top-down アプローチの組み合わせによる 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‡
Katsunori Onishi Tetsuya Takiguchi Yasuo Ariki

１．はじめに

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 で公開されているデータを用いた．

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，HOG [2]が提案されている． 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，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自動的に取得する．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．輝度の勾配強度と勾配方向を画像中のすべての画素で求める．

2.1.2 セルによるヒストグラム化

算出された輝度勾配画像を，セルと呼ばれる 10×10 画素からなる小領域に分割し，各領域においてヒストグラムを作成する．勾配方向の角を 9 方向になるよう量子化し，各方向に強度を重みとして与える．すなわち，1 セルあたり 9 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

輝度勾配ヒストグラムを，[2]と同様に正規化する．図 1 に示すような，セルよりも大きな領域 3×3 セルを 1 ブロックとして正規化を行う．1 セル当たり 9 方向の特徴を持っているため，1 ブロック当たりの特徴次元数は 81 = 3×3×9 となる．あるブロックの特徴ベクトルを v，ブロック内で位置 (i, j) 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．

$$h_{ij}' = \frac{h_{ij}}{\sqrt{\|v\|_2^2 + \varepsilon}} \qquad (\varepsilon = 1) \qquad (1)$$

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．
得られた特徴ベクトルは，照明や影，服装の変化などの影響を受けにくく，局所的な幾何学変化に頑健となる．

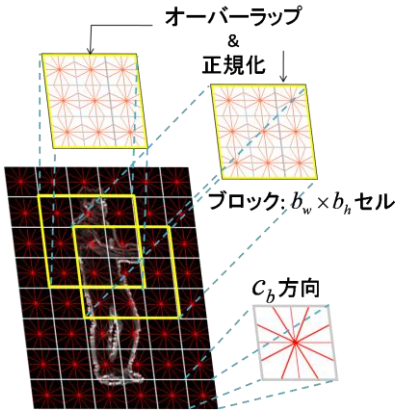

図１　ブロックによる正規化

2.2 人体３次元モデル

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 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

†神戸大学大学院工学研究科
‡神戸大学自然科学系先端融合研究環

56 次元の関節角度で表現されている．図 2 に，関節角による 3 次元人体モデルの一例を示す．

図 2　3 次元人体モデル

3. 姿勢推定方法

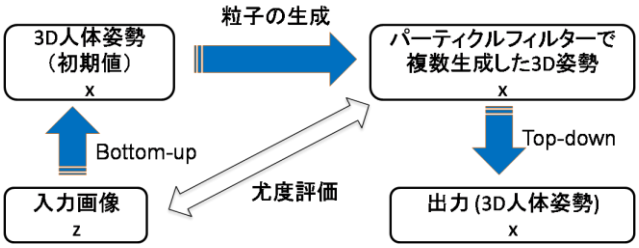

図 3　姿勢推定システム

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 3 に推定システムの概要を示す．

3.1　Bottom-up アプローチ

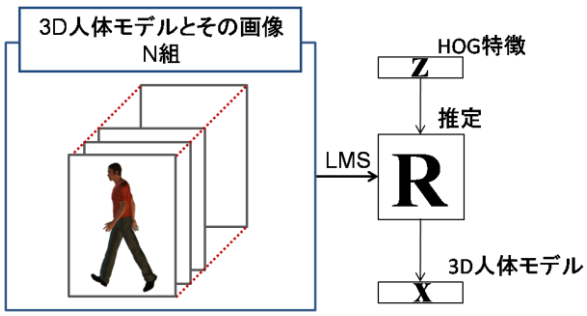

図 4　回帰分析による推定

Bottom-up アプローチでは，[3]の手法と同様に，回帰分析により姿勢推定を行う(図 4)．この手法は，画像から得られた特徴量から，直接 3 次元姿勢を高速に復元できるという利点がある．

画像から得られた HOG 特徴ベクトル z と人体 3 次元モデルの特徴ベクトル x の関係を，次式で近似する．

$$x = Rx + \varepsilon \qquad (2)$$

n 個の学習データのセット$\{(x_i, z_i) \mid i= 1,\cdots n \}$を用いて，学習モデルを作成する．次式より，最小自乗推定法を用いてモデルパラメータである行列 R を決定する．

$$R = \arg\min_R \sum_{i=1}^{n} \left\| Rz_i - x_i \right\|^2 \qquad (3)$$

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)[5]を用いて推定を行う．[5]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 $p(x_t \mid Z_t)$は，ベイズの定理により式(4)のようになる．

$$p(x_t \mid Z_t) \propto p(z_t \mid x_t) p(x_t \mid Z_{t-1}) \qquad (4)$$

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

$$p(x_t \mid Z_{t-1}) = \int_{x_{t-1}} p(x_t \mid x_{t-1}) p(x_{t-1} \mid Z_{t-1}) dx_{t-1} \qquad (5)$$

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．

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しており，非常に高次元なデータとなっているため，そのまま計算を行うと探索範囲が広くなってしまい，推定精度が低下してしまう．そこで，本手法では PCA を行い，部分空間で探索を行う．これにより，人らしい姿勢をとるような範囲内のみを探索することが可能となる．

各粒子の尤度を計算し，それぞれの重みを求める．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　実験条件

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 1 に示す．

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

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．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，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 5 にその実験結果のグラフを示す．

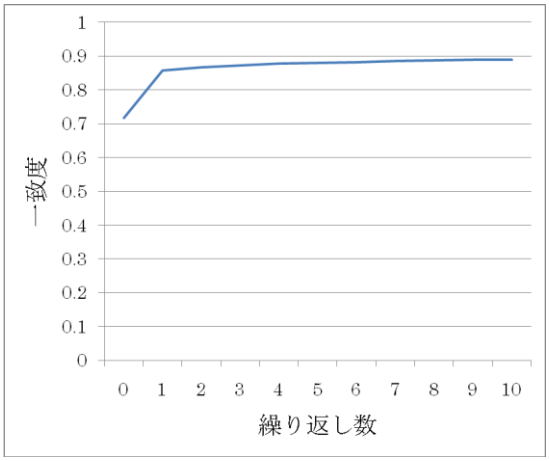

図 5　評価実験

図 5 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 6 に示す．

図 6　実験結果画像

4. 結論

本論文では，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．

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する場合や，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] N.Dalal and B.Triggs, ” Histograms of Oriented Gradients for Human Detection” IEEE Conputer Vision and Pattern Recognition, 886-893, 2005.

[3] 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.

[4] 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.

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