

HOG 特徴に基づく単眼画像からの人体 3 次元姿勢推定

3D Human Posture Estimation Using the HOG Features from Monocular Image

大西 克則† 滝口 哲也‡ 有木 康雄‡
Katsunori Onishi Tetsuya Takiguchi Yasuo Ariki

1. はじめに

本研究では、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 HOG 特徴[1]を用いて姿勢推定を行う方法を提案する。しかし通常の HOG 特徴は画像全体の特徴量であるため、人物以外の背景領域にも多くの特徴次元をもつことになる。そこで、HOG のブロック領域ごとに PCA を行うことにより、背景領域の特徴次元のみを大きく削減する方法も提案する。このようにして得られた特徴を、3 次元人体モデルの特徴ベクトルに射影することで姿勢推定を行う。

2. 特徴量

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 [1]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、[1]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 $81 = 3 \times 3 \times 9$ 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。

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

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。

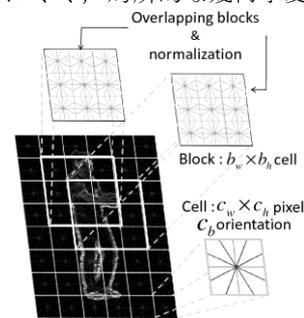


図 1 ブロックによる正規化

2.2 PCA による背景領域の次元削減

HOG 特徴はブロック領域ごとに特徴量が得られている。背景差分を行っているので、背景領域の特徴量のほとんどは一定の値をもっている。つまり PCA をすることにより多くの特徴次元を削減できる。逆に、人体領域の特徴量はさまざまに変化するため、特徴次元はそれほど削減されない。よって PCA をすると、図 2 に示すように、背景領域の特徴量は削減され、人体領域の特徴量を多く用いることができるようになる。この PCA による背景領域の特徴量削減では、ブロック毎に独立して PCA を行い、設定した累積寄与率を超えた次元を削減することで HOG 特徴ベクトルの次元削減を実現している。

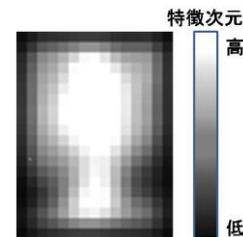


図 2 ブロック毎の PCA

2.3 人体 3 次元特徴

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)の角度を要素とする特徴ベクトル y で特徴量を表現する。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。 y は 24 次元の特徴ベクトルとして得られる。

†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

3. 回帰分析に基づく推定

本章では、画像特徴から人体姿勢を3次元として復元する方法について述べる。[2]と同様に、回帰分析により姿勢推定を行う。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る。

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

$$y = Ax + \varepsilon \quad (2)$$

n 個の学習データのセット $\{(y_i, x_i) \mid i=1, \dots, n\}$ 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 A を決定する。

$$A = \arg \min_A \sum_{i=1}^n \|Ax_i - y_i\|^2 \quad (3)$$

4. 評価実験

本章では、従来研究であるシルエット画像から Shape Contexts 特徴[3]との比較実験の結果を示し、考察を加える。

4.1 実験方法

実験には 640×480 画素の解像度を持つ単眼のカメラで撮影した映像を用いた。固定したカメラに対して、水平方向に被験者1名の人体を回転させ、8方向から撮影した。それぞれの方向に対して、直立、手をあげる、手を広げる、歩く、走るの5動作を行い、姿勢推定を行う。あらかじめ各姿勢に対して手動で教師データを与えておき、推定結果との平均二乗誤差を求めることで評価する。

学習データには、8方向のうち各方向について5つの動作全体で画像30枚を、またテストデータとしては、学習データ撮影時と同条件の画像(a)、撮影条件を変えノイズが多く乗った画像(b)、<http://www.nada.kth.se/~hedvig/data.html> の画像(c)の全123枚の画像を使用した(表1)。

表1 使用したデータ数

姿勢	画像枚数			
	学習データ	テストデータ		
		(a)	(b)	(c)
直立	16	8	8	0
手を上げる	40	8	8	0
手を広げる	24	8	8	0
歩く	80	16	16	11
走る	80	16	16	0

ここで、歩く、走るといった姿勢は、他の姿勢と比較すると変形がはげしく、ひとつの動作において様々な変化をするのでデータ数を多くしている。また今回の実験では、姿勢の推定のみを行っており、認識は行っていない。

画像から得られる HOG 特徴は、18954次元となる。得られた HOG 特徴を、累積寄与率 90%で 81次元のブロックごとに PCA を行う。PCA を行った結果、4867次元に削減された。

4.2 実験結果

実画像を用いて実験することにより、本手法が有効に働くことを確かめる。図3に従来手法との比較実験の結果を示す。評価は平均二乗誤差で行った。

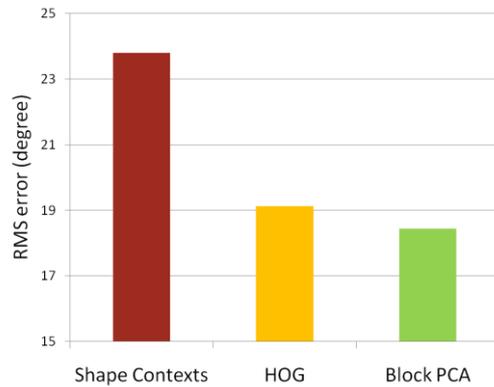


図3 比較実験結果

従来手法よりも、本手法のほうが誤差が約 5.35 度軽減された。シルエット画像とは異なり、HOG 特徴では内部のエッジまで考慮していることから、誤差が減ったものと考えられる。また通常の HOG よりもわずかではあるが、ブロック毎に PCA を行った方が誤差を小さく抑えることができています。

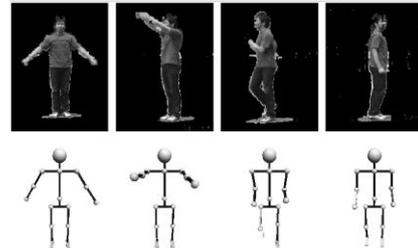


図3 姿勢推定結果

図3に3次元姿勢推定の結果の一例を示す。図3の上段は入力画像を、下段は3次元姿勢を表している。推定結果に多少のずれはあるものの、概ね推定できている。

5. 結論

単眼画像から人体3次元姿勢を推定する方法について提案した。本論文では従来、人検出に用いられてきた HOG 特徴を用いること、ブロック毎に PCA を行うことで、背景領域の特徴次元を削減して姿勢推定する方法を提案した。シルエット画像と比較すると、内部のエッジ情報も考慮することから、推定精度を上げることができた。今後は、姿勢推定の精度を向上させるために、推定方法を改良していく予定である。

文献

- [1] N.Dalal and B.Triggs, "Histograms of Oriented Gradients for Human Detection" IEEE Computer Vision and Pattern Recognition, 886-893, 2005.
- [2] A.Agarwal and B.Triggs, "3D Human Pose from Silhouettes by Relevance Vector Regression" IEEE Conference on Computer Vision and Pattern Recognition, VOL.2, 882-888, 2004.
- [3] G.Mori and J.Malik, "Estimating Human Body Configurations using Shape Context Matching" European Conference on Computer Vision, 150-180, 2002.