

## 部位スコアマップ統合による隠れにロバストな人体検出

木下 航一<sup>†,††</sup> 村瀬 洋<sup>††</sup>

† オムロン株式会社技術本部, 京都府  
†† 名古屋大学大学院情報科学研究科, 愛知県  
E-mail: †kino@ari.ncl.omron.co.jp

あらまし 部位検出結果のスコアマップ統合に基づく, 隠れにロバストな人体検出手法を提案する. 一般的に部位検出に基づく人体検出では, ある領域内での部位の存在の有無や位置の情報を確率分布モデル等によって統合することによって判定処理が行われる. しかし個別の部位は, 人体全体ほど明瞭な特徴を持たないため, 複雑な背景では, 人体以外の領域でも部位検出器が誤検出を発生する可能性が比較的高く, 従来手法ではこのような場合に十分対応できなかった. また, 主要な部位に隠れが生じた場合性能が大きく劣化するという課題があった. 本論文ではこれらの課題に対応するため (1) スコアマップ統合 (2) 多段統合による隠れ対応, の2つの手法を提案する. (1) によって人体画像から生じる検出パターンとそれ以外の画像から生じる検出パターンをより明確に区別することができる. また (2) によって, 隠れのある状況に対するロバスト性が向上する. 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, 特に隠れがある状況でも高い性能を維持することが示された.

キーワード 人体検出, 部位検出, 人体モデル

## Parts Score Map Integration for Partially Occluded Humans Detection

Koichi KINOSHITA<sup>†,††</sup> and Hiroshi MURASE<sup>††</sup>

† Research & Development HQ, Omron Corporation, 9-1 Kizugawadai, Kizugawa-city, Kyoto, 619-0283 Japan

†† Graduate School of Information Science, Nagoya University, Furou-cho, Chikusa-ku, Nagoya, 460-3807  
Japan

E-mail: †kino@ari.ncl.omron.co.jp

**Abstract** We propose a human body detection framework based on parts score map integration, which is robust to partially occluded human image. Generally part based body detection methods utilize part detectors information such as presence or absence within certain region or their positions, by integrating them using e.g. probability distribution model. However partial body regions itself tend to be not having distinctive feature like whole body and it likely to occur false positives in case of e.g. complex background region. Another problem is deterioration of performance in case of major parts occlusion. In this paper, we propose (1) score map integration method, (2) multilayered integration method to address partial occlusion. (1) enables to distinguish part detection result of human body from them of non human body more clearly. (2) increases detection performance for partial occluded images. The evaluation result shows better performance especially in case of partial occlusion images in comparison to the traditional method.

**Key words** body detection, part based body detection, part detection

### 1. まえがき

静止画像から人体を検出する技術は, 映像監視による不審者検知や危険行動検知, 車載機器での歩行者検知など, 幅広いアプリケーションでの活用が期待され, 盛んに研究が行われている. 人体は姿勢, 衣服や角度によって見え方が大きく変化する

のみならず, 照明変化, 隠れ, 背景等の影響によっても大きく特徴が変化し, それゆえこれまで行われてきた顔検出や自動車検出等の課題と比較しても, 高精度な検出の実現が困難な分野とされている [1], [2].

静止画からの人体検出技術は, 全身検出に基づく手法と, 部位検出に基づく手法の2つに大別できる. 前者は全身を一度に

検出可能な識別器を使用する。Viola ら [12] が顔検出問題に対して Haar-like 特徴量とカスケード構造を持った boosting 識別器を導入し大きな成功を収めて以来、同様の枠組みが人体検出に対しても適用され [13]、活発に研究されてきたが、これらの多くは全身検出に基づく手法である。Dalal ら [16] は特徴量として輝度勾配の局所的なヒストグラムを活用することを提案し、人体検出に対して優れた効果を示した。Zhu ら [21] は HOG 特徴量をカスケード構造の AdaBoost によって処理し、処理の高速化が図られた。[8] では、多様な特徴量を AdaBoost の各ステージでの検出処理に組み込むことにより、検出能力の一層の向上が図られている。全身検出に基づく手法は比較的単純で高速化しやすい反面、隠れが生じた際に検出能力が低下する問題点が指摘されている。しかしながら人体検出が必要となる多くのアプリケーションにおいて、全身が常に見えていることが保証されるようなケースは少ないと考えられ、実用上大きな課題となる。

このような課題に対応するため、近年いくつかの 'part-based' な人体検出手法が提案されている [3], [4], [10]。これらの手法は人体を複数部位からなる組み合わせととらえ、各部位に対する検出を独立に実行し、その情報を総合することにより人体であるかどうかの判断を行う。

初期の研究では色およびテクスチャ情報を直接使用することで人体を表現する試みがなされた。たとえば [9] では部位間の関連性が 'body plans' と呼ばれる枠組みによって表現されるが、ここでは部位検出は平行に並んだ 2 つのエッジを探すことによるのみ行われたため、限定された条件以外で高い性能を発揮することは難しかった。近年では部位の表現に識別器を活用することが主流となっており、たとえば [22], [23] では色および輝度勾配情報を活用した検出器により部位を検出し、検出結果は動的計画法により統合され人体かどうかの判断が行われる。Mohan ら [4] は [5] で提案された検出器を部位検出器として用い、線形 SVM によって検出結果の統合を行った。

[3], [10], [11] ではそれぞれ position-orientation histograms, edgelet および rectangle features が特徴量として利用され、boosting 手法により部位検出器が学習される。これらの手法はどれも部位検出結果の統合に同時確率分布モデルを利用している。Chen ら [7] は部位検出器として MCMIBoost を使用し、ある領域内での検出結果から各部位の存在確率を求め、最終的に SVM により統合する手法を提案した。近年では Bourdev ら [24] によって、各部位のさまざまな姿勢に対する検出器 (poselets) を学習し、この結果を重みつき投票することにより人体検出する枠組みが提案され、大きな姿勢変動に対して高い検出能力を持つことが示されている。

いずれの手法も部位検出に基づく手法ではあるが、たとえば [3], [10], [11] では顔もしくは頭部が検出できることを前提としており、他の技術でも主要な部位の隠れに対しては依然として大きな性能劣化が見られる。これに対してわれわれは、隠れに対してよりロバストな人体検出のフレームワークを提案する。提案手法は異なった部位検出器の組み合わせに対する部分統合検出器を導入し、それらの結果に対してさらに統合を行う。こ

れにより一部の部位が隠された場合と人体以外の画像との区別をより明確に区別することが可能である。本論文では以下、2 章において現在主流となっている HOG 特徴量による全身検出手法について概説し、3 章で提案する部位統合手法について説明を行う。4 章では実験結果を示し、5 章において考察を行う。

## 2. HOG 特徴量による人体検出

顔検出に対する Haar-like 特徴の有効性が Viola ら [12] によって示されて以来、多くの検出アプリケーションで Haar-like 特徴が使用されてきた。Haar-like 特徴量によって、顔の局所的な領域の明暗差をとらえ、それらを多数組み合わせることにより照明や表情の変化にロバストな特徴量集合を得ることができ。この手法を元にして、特徴量の表現力や処理速度をさらに向上させるためのさまざまな応用手法が開発されている。複数局所領域の明暗差の組み合わせを表現可能な Joint Haar-like 特徴量 [17]、離れた部分同士のを組み合わせに着目した Sparse 特徴量 [18] などはその例である。

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いる。その理由としては主に以下のような点があげられる。

- 1) 衣服と背景との組み合わせによって、濃淡パターンの方向は一定でなく、逆転するケースも生じる
- 2) 人体は姿勢、向きにより形状が大きく変形するため、矩形内の特定の場所の輝度差は一定ではなく、特徴量としては不適切。

これらの課題に対応するため、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の必要性が早くから指摘されてきた。具体的には Edgelet [19] や Shapelet [20] などの特徴量が提案され、それぞれ人体検出における有効性が確認されてきた。このような中で、Dalal ら [16] は、各局所領域でのエッジ方向分布をヒストグラムとして表現し、これらの集合を特徴量として活用することを提案した。この手法は以下に示すような利点を持つ。

- 1) 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持つ。
- 2) 局所領域内での方向分布を利用するため、姿勢、向きによる位置の変化をある程度吸収することができる。

この手法は HOG 特徴量と呼ばれ、幅広い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。

HOG 特徴量の算出は以下の手順で行う。まず各ピクセルにおいて輝度の勾配強度  $m$  とその方向  $\theta$  を以下により求める。

$$m(u, v) = \sqrt{f_u(u, v)^2 + f_v(u, v)^2} \quad (1)$$

$$\theta(u, v) = \tan^{-1} \frac{f_v(u, v)}{f_u(u, v)} \quad (2)$$

ここで  $f_u, f_v$  はそれぞれ画像横方向、縦方向の輝度勾配であり、 $f_u = I(u+1, v) - I(u-1, v)$ ,  $f_v = I(u, v+1) - I(u, v-1)$  である ( $I$  は輝度値)。次に  $N_c \times N_c$  ピクセルをセルとし、そ

それぞれのセル領域において輝度勾配方向の重み付投票を行うことにより輝度の勾配方向ヒストグラムを作成する。さらに各セルで作成した輝度の勾配方向ヒストグラムを  $N_b \times N_b$  セルを1ブロックとして正規化を行う。あるブロックでの正規化前の特徴量ベクトルを  $f$  とするとき、正規化は

$$v = \frac{f}{\sqrt{\|f\|_2^2 + \epsilon^2}} \quad (3)$$

によって行う。最後に各ブロックごとに正規化された特徴量を1列のベクトルにまとめることでHOG特徴量が得られる。本研究では  $N_c = 3$ ,  $N_b = 2$ , 勾配方向ヒストグラムのbin幅は  $20^\circ$  とした。  $0^\circ - 180^\circ$  を  $20^\circ$  ずつに分割するため、9方向の勾配方向ヒストグラムが得られる。また  $\epsilon = 1$  とした。HOG特徴量による人体検出手法は従来の手法に比べ高い性能を持つが、一部が隠れた際に大きく性能が低下する。本論文ではこの課題に対応するため、全身を一度に検出するのではなく、複数の部位の検出結果を統合することにより隠れに対する検出性能の低下を抑える手法を提案する。

### 3. 部位検出統合による人体検出

本章では、部位検出統合による人体検出のフレームワークについて説明する。全体のフローは以下の通りである。

- (1) 部位検出器により人体各部位のスコアマップ生成。
- (2) 統合検出器により各部位のスコアマップ情報を統合し、人体を検出。

一度に全身の検出を行うのではなく、各部位の検出結果を統合する枠組みにより、一部が隠れたような状況に対しても高い検知性能を維持することができる。部位検出の結果に基づく人体検出(いわゆるボトムアップベースの手法)はすでにいくつか提案されているが、本手法は以下の点で従来手法と異なる。

(1) 部位検出の結果として、ある領域内での2値の識別結果ではなく、スコアマップを使用する。これにより、複雑な人体の動きや、隠れが生じた状況などに対しても、それが人体の一部であるかについて、より適切な統合判断を行うことが可能となる。

(2) 一部の部位が隠れた状態に対してロバストな多段階統合手法。具体的には、各部位  $p$  ( $p = 1 \dots N$ ,  $N$  は検出部位の数) に対して、この部位を除く  $N - 1$  部位による部分統合識別器を学習する。これら部分統合識別器の情報をすべて統合することで最終結果を得る。

以下では全体フロー中の各項目について、具体的な手順を示す。

#### 3.1 部位検出器

本研究では頭部、左肩、右肩、胴体、下半身の5部位を人体部位として使用する。図2に例を示す。部位検出器  $B_p$  によって、人体各部位の存在可能性を示すスコアマップ  $M_p$  を作成する。検出器については、本研究では統合手法フレームワークによる性能向上確認を目的とするため、人体検出の分野で広く使われている Real AdaBoost [14] を用いた。

Real AdaBoost の学習手順は以下の通りである。

学習データセット:  $(x_1, y_1), \dots, (x_m, y_m); x_i \in \mathcal{X}, y_i \in$

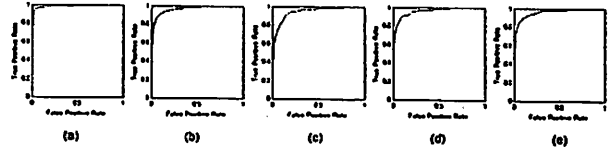


図1 部位検出器性能評価結果。(a) 頭部, (b) 下半身, (c) 胴体, (d) 左肩, (e) 右肩。

Fig.1 Evaluation result for individual part detectors. (a) head, (b) lower body, (c) torso, (d) left shoulder, (e) right shoulder.

$\{-1, +1\}$  とする。  $\mathcal{X}$  はデータ空間を示し、  $y$  はそのデータに対するラベル値である。

学習データに対する重みの初期化。  $D_1(i) = 1/m$ 。

For  $t = 1, \dots, T$ :

- step 1 重み  $D_t$  を用いて弱識別器の学習。
- step 2 弱仮説  $h_t: \mathcal{X} \rightarrow R$  を得る。
- step 3 重みの更新。

$$D_{t+1}(i) = \frac{D_t(i) \exp(-y_i h_t(x_i))}{Z_t} \quad (4)$$

ここで  $Z_t$  は  $D_{t+1}$  の総和を1にするための正規化項である。

入力  $x$  に対する強識別器の出力は

$$f(x) = \sum_{t=1}^T h_t(x) \quad (5)$$

によって得られ、  $H(x) = \text{sign}(f(x))$  によって最終的に  $y$  が  $-1$  か  $+1$  かの判断がなされる。本論文では式5によって得られる値をスコアと呼ぶ。

上記 step 1 の弱仮説  $h_t$  の選択は、各特徴量次元の確率密度分布を考慮し、以下の手続きによって行う。

各特徴量次元において確率密度分布  $W$  を学習サンプルの重み  $D_i$  に基づいて次式より作成する。

$$W_b^j = \sum_{i: x_i \in \mathcal{X}_j \wedge y_i = b} D_i(i) \quad (6)$$

ここで  $b \in \{-1, +1\}$  である。弱識別器の出力  $h_t(x)$  は次式により算出する。

$$h_t(x) = \frac{1}{2} \ln \frac{W_+^j + \epsilon}{W_-^j + \epsilon} \quad (7)$$

これを各特徴量次元について求め、step 2においてこれらの中でもっとも信頼度の高い弱識別器を選択する。なお  $\epsilon$  の値は [1] にしたがって  $10^{-7}$  とした。

一般的に人体の部位は、全身を対象としたときと比較して固有の特徴が乏しい。そのため複雑な背景等で誤検出を起こしやすく、独立した検出器として十分な識別性能を得ることは難しい。図1に本研究で学習した部位検出器の単体での性能評価結果を示す。頭部付近は形状に明瞭な特徴があるためよい性能を示しているが、それ以外の部位では相対的に低い性能を示していることがわかる。

#### 3.2 統合検出器

次に各部位の検出結果を統合することにより、ある矩形内に

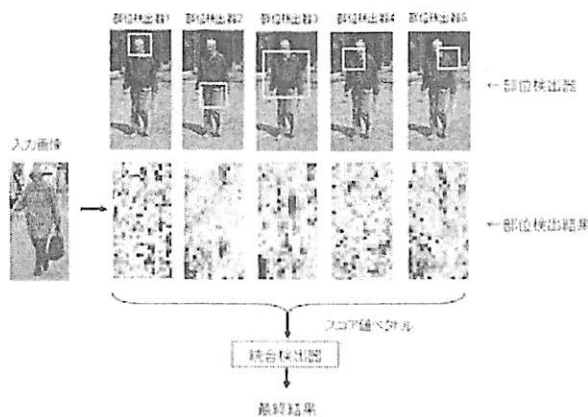


図2 スコアマップ単純統合手法フレームワーク。

Fig. 2 Framework of parts score integration detection method.

人体が存在するかどうかの最終判断を行う。本研究では以下の2つの統合手法について検討を行った。

### 3.2.1 スコアマップ単純統合

各部位検出器のスコアマップ  $M_p$  を1列のベクトル  $m_p$  に変換し、統合ベクトル  $v$  を以下によって得る。

$$v = [m_1^T, m_2^T, \dots, m_N^T]^T \quad (8)$$

得られた統合スコアベクトル  $v$  を統合検出器  $C$  への入力とし、対象矩形が人体であるかどうかの最終結果を得る。図2にこの手法の概略を示す。統合検出器  $C$  としては RealAdaboost を用いた。

この手法は対象矩形の中での各部位検出器のスコアマップをそのまま統合する。そのためここでは本手法を「スコアマップ単純統合手法」とよぶ。

### 3.2.2 スコアマップ多段階統合

本研究ではさらに一部の部位が隠れた画像に対する検知性能の低下を軽減するため、統合段階を複数階層化する手法を検討した。本手法では、全部位のスコアマップから統合ベクトルを得るのではなく、一つの部位  $k$  を欠いた部分統合ベクトル  $v^{(k)}$  を以下のように作成する。

$$v^{(k)} = [m_1^T, \dots, m_{k-1}^T, m_{k+1}^T, \dots, m_N^T]^T \quad (9)$$

すべての  $k$  に対して上記を求めることにより、 $N$  種類の部分統合ベクトルが得られる。これらの部分統合ベクトルそれぞれに対して、1段階目の統合検出器  $C^{(k)}$  を学習する。いま統合検出器  $C^{(k)}$  の出力を  $r^{(k)}$  で表し、これを一列のベクトルとしたものを

$$r = [r^{(1)}, \dots, r^{(N)}]^T \quad (10)$$

とする。これを入力として2段階目の統合検出器  $D$  を学習する。本研究では2段階目の統合検出器としては SVM を用いた。

この手法では、人体のある部位  $p$  が隠れた場合でも、その影響を大きく受けない統合検出器  $C^{(p)}$  が存在する。人体以外の画像に対してはすべての統合検出器が低いスコアを出力するため、これと比較してスコア分布に明確な差が生じる。そのため状況によっては人体の一部が隠れた場合でも、人体画像であるかど

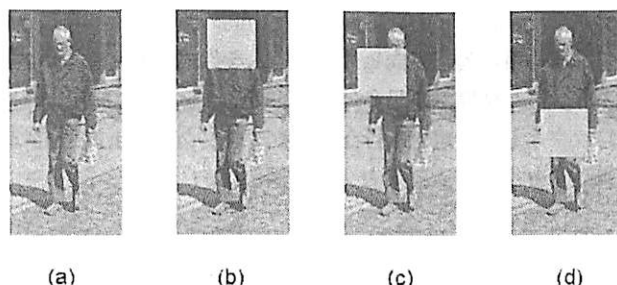


図3 評価画像例。(a) 隠れなし、(b) 頭部隠れ、(c) 肩隠れ、(d) 下半身隠れ。

Fig. 3 Evaluation image. (a) without occlusion, (b) head occlusion, (c) shoulder occlusion, (d) waist occlusion.

うかの判断が、よりの確に行えるようになることが期待される。

本論分ではこの手法を「スコアマップ多段階統合手法」とよぶ。

## 4. 性能評価実験

提案手法について人体検出性能評価を行った。実験画像としては "TUD Multiview Pedestrians" [15] を用いた。このデータセットは 4732 枚の学習用画像と 248 枚のテスト用画像を含んでおり、さまざまな場所で撮影された歩行者画像からなる。それぞれの画像に対しては 8 種類の方向別のタグ付けがされている。本研究ではこのうち 'front', 'left-front', 'right-front' とタグ付けされた画像を使用した。画像枚数は学習画像が 1790 枚、テスト画像が 101 枚である。学習用画像に対しては目視で頭部、肩、腰および足先の位置を入力し、これらの点を基準に矩形を切り出し、各部位の学習画像を作成した。またこのデータセットは人体の写っていない画像を含まないことから、ネガティブ画像としては "INRIA Person Dataset" [16] の画像を使用した。

また、人体の一部が隠れた状態に対する検出性能を評価するため、上記ポジティブテスト画像に対して、頭部、肩、下半身のそれぞれの部位が隠れた画像を人工的に生成し、これらに対する検出性能の評価を行った。図3に、評価に用いた画像の例を示す。なおこれらの人工画像は性能評価のみに用い、学習には用いていない。

性能評価結果を図4に示す。各図は縦軸に検出率、横軸に誤検出率をとった ROC 曲線である。各手法とも隠れなしの画像に対してはほぼ同等の検出性能を示しているが、通常の全身検知手法では隠れが生じた際に大きな性能劣化が見られる。これに対して、本論文で提案する2つの手法はいずれも隠れに対して全身検知よりも高い性能を示している。またスコアマップ単純統合手法と多段階統合手法を比較した場合、「shoulder occlusion」および「waist occlusion」で、多段階統合手法がより高い性能を示していることが分かる。

次に、既存のボトムアップベースの人体検出手法との性能比較を行った。ここでは Chen ら [7] の手法を取り上げる。彼らは頭部、胴体、下半身の3部位の検出器を用いている。まず第1段階としてこれらの部位の検出を行う。次に第2段階として

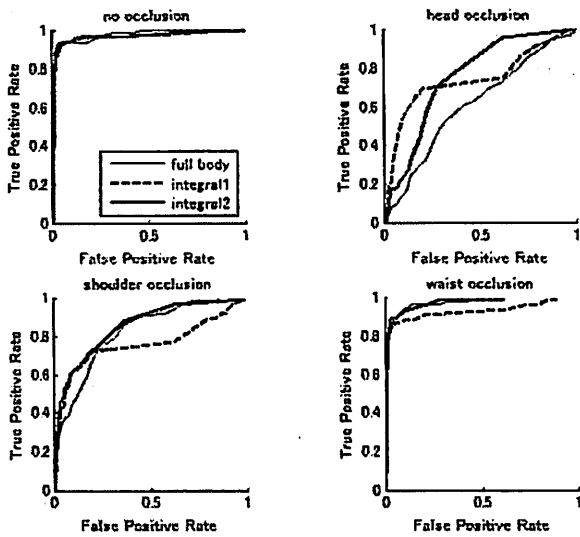


図4 ROC 曲線による各手法の検出性能評価. full body は通常の全身検出手法, integral1 はスコアマップ単純統合, integral2 はスコアマップ多段階統合手法を示す.

Fig.4 ROC curves for each integration method.

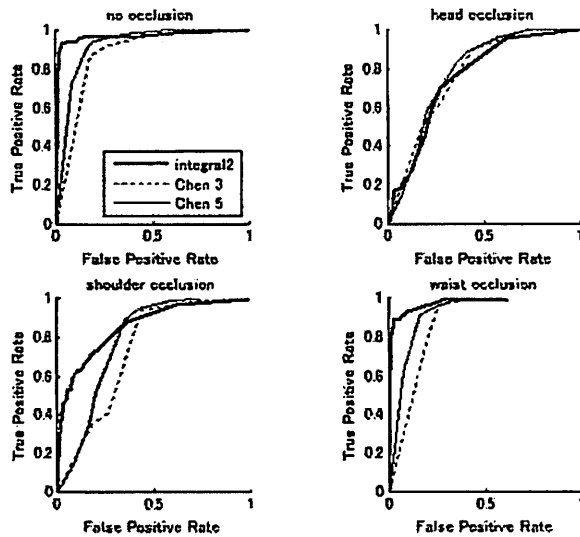


図5 ROC 曲線による各手法の検出性能評価. chen 3 は3部位を使用したChenらの手法, chen 5 は5部位に拡張したChenらの手法, integral2 はスコアマップ多段階統合手法を示す.

Fig.5 ROC curves for each integration method.

矩形をスライディングしながら、部位検出結果が定められた領域内に存在する確率計算を行う。部位  $Q$  のある位置での検出結果  $b^Q$  に対して存在確率は  $p_b^Q = \frac{1}{1 + \exp(-H_S^Q(b^Q))}$  によって計算される。ここで  $H_S^Q(b^Q)$  は、検出器の出力値である。ある領域内における部位  $Q$  存在確率は Noise-OR モデルに基づき  $P^Q = 1 - \prod_b (1 - p_b^Q)$  によって求められる。最終的に3次元の存在確率ベクトルが得られ、これが SVM に入力されて人体であるかどうかの判断が行われる。

Chen らのオリジナルの手法では、部位検出器として MCMI-Boost と呼ばれる複数クラス間での特徴共用が可能な手法を利用しているが、ここでは統合手法に着目した性能比較を行うため、部位検出器についてはわれわれが用いたものを利用した実

装を行った。また利用部位についてもオリジナルの3部位を利用するものと、われわれと同様の5部位の2種類を実装し、評価を行った。

性能比較を図5に示す。図より、頭部が隠れた状態については従来手法と提案手法はほぼ同等の性能を示しているが、その他の場合ではいずれも、提案手法が従来手法を上回る結果を示している。

## 5. 考 察

今回の実験では、Chen らの手法は頭部が隠れた画像に対しては高い検知性能を示したものの、その他の隠れ画像、および隠れなし画像に対して全体的に低い検出精度を示した。これは彼らの統合手法がある領域内での部位検出の結果を、その位置にかかわらず使用しているためと考えられる。個別の人体部位は、顔や人体全体ほど複雑な特徴を持たないため、複雑な背景画像では、人物以外の部分でも部位検出器が誤検出を発生する可能性は比較的高い。そのため、ネガティブクラスの画像に対しても複数の部位を誤検出する結果が得られることがある。彼らの手法ではそのようなケースと人体画像をうまく区別することができないため、結果として低い検出性能となっているものと考えられる。

それに対してわれわれの提案する手法では、ある領域内における各部位についての有無だけではなく、その位置も含めた統合を行っている。そのため人体画像から生じる検出パターンとそれ以外の画像から生じる検出パターンをより明確に区別することができ、結果としてより高い検出性能を示したものと考えられる。

またスコアマップ単純統合手法と多段階統合手法を比較した場合、多段階統合手法がより高い性能を示すことが多い。一部の部位が隠れたとき、単純統合手法では統合検出器が一つしかないため、検出性能は大きな影響を受ける。それに対して多段階統合手法では、影響をほとんど受けない統合検出器が一つは存在する。人体以外の画像に対してはすべての統合検出器が低いスコアを出力するため、この状況とは明確な差がある。このような理由により、状況によっては多段階統合手法は人体の一部が隠れた画像を人体以外の画像とより高い精度で区別できるものと考えられる。

ただし実験結果より、常に多段階統合手法が優れているわけではなく、この理由の解析に関しては今後の課題である。

なお各部位の位置情報を利用する本手法は、強識別器の出力にポジティブ側に寄与した特徴量位置を求めることで、各部位の位置候補情報が得られる。図6に例を示す。この情報は、人体検出後により詳細な姿勢推定などを行う際、有効であるものと考えられる。

## 6. む す び

部位検出結果に基づく人体検出に関して、(1)スコアマップ統合 (2)多段階統合による隠れ対応、の2つの手法を提案した。実験の結果、提案手法はどちらも従来技術に比べて高い性能を示し、特に隠れがある状況でも高い性能を維持することが示さ



図6 部位位置推定結果. 各部位検出の出力のうち, 統合検出器の判定に寄与した領域を白で表示.

Fig.6 Parts position estimation results. The parts detection results which contributed for integral detectors are indicated white region.

れた. 一方, スコアマップ単純統合手法と多段階統合手法を比較した場合, 対象とする画像により前者が高い性能を出す場合とそうでない場合とが分かれた. この理由に関しては今後の解析が必要である. またより大きな姿勢変動や向きの変動等を含む画像への対応も今後の課題である.

#### 文 献

- [1] 山内悠嗣, 藤吉弘亘, Hwang Bon-Woo, 金出武雄人, "アピアランスと時空間特徴の共起に基づく人検出", 第10回画像の認識・理解シンポジウム, July, 2007.
- [2] M. Enzweiler, D.M. Gavrila, "Monocular Pedestrian Detection: Survey and Experiments," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 31, No. 12, Dec.2009.
- [3] K. Mikolajczyk, C. Schmid, and A. Zisserman, "Human Detection Based on a Probabilistic Assembly of Robust Part Detectors," in Proc. ECCV (1), 2004, pp.69-82.
- [4] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-based object detection in images by components," IEEE TPAMI, 23(4), pp349-361, 2001.
- [5] C. Papageorgiou, T. Poggio, "A trainable system for object detection," IJCV, 38(1):15-33, June 2000.
- [6] B. Wu, R. Nevatia, "Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors," IJCV, 75(2), 247-266, 2007.
- [7] Y. Chen, C. Chen, Y. Hung, and K. Chang, "Multi-Class Multi- Instance Boosting for Part-Based Human Detection," in Proc. of the International Conference On Computer Vision Workshops, pp. 1177-1184, 2009.
- [8] Y. Chen, C. Chen, "MFAST human detection using a novel boosted cascading structure with meta stages," IEEE TIP, 17(8):1452-1464, 2008.
- [9] D. Forsyth, M. Fleck, "Body plans," in Proc. of CVPR 1997.
- [10] D. Tosato, M. Farenzena, M. Cristani, and V. Murino, "Part-based human detection on Riemannian manifolds," IEEE Int'l Conf. on Image Processing ICIP 2010, Hong Kong, Sept. 2010.
- [11] A. S. Micilotta, E. J. Ong, and R. Bowden, "Detection and tracking of humans by probabilistic body part assembly," in Proc. of BMVC 2005.
- [12] P. Viola, M. Jones, "Robust real-time face detection," International Journal of Computer Vision, Vol.57, no.2, pp. 137-154, 2004.
- [13] P. Viola, M. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in Proc. of ICCV 2003.
- [14] R. E. Schapire and Y. Singer, "Improved Boosting Algorithms Using Confidence-rated Predictions," Machine Learning, No. 37, pp. 297-336, 1999.
- [15] M. Andriluka, S. Roth, 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.
- [16] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection," International Conference on Computer Vision & Pattern Recognition, vol.2, pp. 886-893, June, 2005.
- [17] T. Mita, T. Kaneko, O. Hori, "Joint Haar-like Features for Face Detection," In Proc. IEEE International Conference on Computer Vision, Vol.2, pp.1619-1626, 2005.
- [18] C. Huang, H. Ai, T. Yamashita, S. Lao and M. Kawade, "Incremental Learning of Boosted Face Detector," In Proc. IEEE International Conference of Computer Vision, 2007.
- [19] B. Wu, R. Nevatia, "Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors," in Proc. of IEEE International Conference of Computer Vision, 2005.
- [20] P. Sabzmejdani, G. Mori, "Detecting Pedestrians by Learning Shapelet Features," International Conference on Computer Vision & Pattern Recognition, 2007.
- [21] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in Proc. of CVPR, 2006.
- [22] P. F. Felzenszwalb, D. P. Huttenlocher, "Efficient matching of pictorial structures," in Proc of CVPR 2000.
- [23] R. Ronfard, C. Schmid, and B. Triggs, "Learning to parse pictures of people," in Proc. of ECCV 2002.
- [24] L. Bourdev, J. Malik, "Poselets: Body Part Detectors Trained Using 3D Human Pose Annotations," in Proc. of ICCV 2009.