

部位スコアマップの多段統合による隠れにロバストな人体検出

木下 航^{†,††a)} 村瀬 洋^{††}

Layered Parts Score Map Integration for Partially Occluded Humans Detection

Koichi KINOSHITA^{†,††a)} and Hiroshi MURASE^{††}

あらまし 部位スコアマップの多段統合に基づく、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もたないため、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、部位スコアマップの情報を多段的に統合する手法を提案する。スコアマップを用いることで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。更に一部の部位が隠れた状況に対応するための多段統合を導入することにより、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

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二つに大別できる。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もった 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こう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っていた。

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合せと捉え、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。

初期の研究では色及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。例えば [9] では部位間の関連性が 'body plans' と呼ばれる枠組

[†] オムロン株式会社技術・知財本部, 木津川市
Technology & Intellectual Property H.Q., 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,
Furo-cho, Chikusa-ku, Nagoya-shi, 460-3807 Japan

a) E-mail: kino@ari.ncl.omron.co.jp

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二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、例えば [22], [23] では色及び輝度こう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 及び 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 MCMBoost 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。

いずれの手法も部位検出に基づく手法ではあるが、例えば [3], [10], [11] では顔若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。これに対して我々は、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スコアマップの組合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更に統合を行う。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. 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. で提案するスコアマップによる多段部位統合手法について説明を行う。4. では実験結果を示し、5. 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捉え、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る。この手法をもとにして、特徴量の表現力や処理速度を更に向上させるための様々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合せを表現可能な Joint

Haar-like 特徴量 [17]、離れた部分同士の組合せに着目した 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挙げられる。

(1) 衣服と背景との組合せによって、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、方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もつ。

(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もつ。

(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、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こう配強度 m とその方向 θ 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、縦方向の輝度こう配であり、 $f_u = I(u+1, v) - I(u-1, v)$, $f_v = I(u, v+1) - I(u, v-1)$ である (I は輝度値)。次に $N_c \times N_c$ ピクセルをセルとし、それぞれのセル領域において輝度こう配方向の重み付投票を行うことにより輝度のこう配方向ヒストグラムを作成する。更に各セルで作成した輝度のこう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。正規

化前の特徴量を f とし、あるブロックの正規化前の特徴量ベクトルを \mathbf{V} するとき、正規化後の特徴量は

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

によって得られる。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 $N_c = 3$, $N_b = 2$, こう配方向ヒストグラムの bin 幅は 20° とした。 $0^\circ - 180^\circ$ を 20° ずつに分割するため、9 方向のこう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もつが、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、複数の部位のスコアマップを段階的に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. 部位スコアマップ統合による人体検出

本章では、部位スコアマップ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とおりである。

(1) 部位スコアマップ生成器により人体各部位のスコアマップ作成。

(2) スコアマップ統合。複数の異なる部位の組合せにより部分的なスコアマップ統合を行った上で、それらの情報を更に統合し、人体を検出。

本研究では (2) において、Chen らの手法 [7] のように、一度に全ての情報を統合せずに、異なった複数部位情報の組合せに対する、段階的な統合を行う手法を提案する。これにより、一部の部位が隠れた状態に対して、よりロバストな人体検出を実現する。

以下では各処理について、具体的な手順を示す。

3.1 部位スコアマップ生成

本研究では [7] に従い頭部、胴体、下半身を人体部位として使用するほか、左肩、右肩を使用した。肩を使用した理由としては頭部から肩、腕に至るラインが特徴的であり、更に形状が大きく変化せず比較的安定した検出が可能と考えられることによる。図 2 に例を示す。

これらの部位に対して、部位スコアマップ生成器 \mathcal{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 \mathbf{M}_p を作成する。

なお本研究の部位定義では肩と胴体等、領域に重畳

している部分がある。そのため、例えば左肩が隠れた場合、同時に胴体領域の一部にも隠れが生じることとなり、胴体のスコアマップにも影響が発生する。部位の隠れに対するロバスト性を向上させる意味からは、各部位領域は重畳なく定義されることが好ましいと考えられるが、その一方、重畳を避けるために各部位のサイズを小さくした場合、多くの部位に対してスコアマップ生成器の性能は低くなり、結果として有効な人体検出が構成できなくなる。

このような理由により、部位領域の重畳によって有効なスコアマップ生成器の数を増やすことは、人体検出の信頼性を向上させる上で有効と考えられるが、それが隠れに対するロバスト性に悪影響を与えないことを確認する必要がある。

本研究においては予備実験により、頭部、胴体、下半身の重畳のない 3 部位による人体検出と比較して、ここで定義した 5 部位による人体検出が高い性能を示すことを確認した。

スコアマップ生成器については、本研究では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とおりである。

学習データセット: $(x_1, y_1), \dots, (x_m, y_m); x_i \in \mathcal{X}, y_i \in \{-1, +1\}$ 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、識別を行う際は $H(x) = \text{sign}(f(x))$ によって 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。

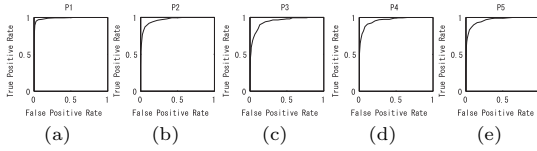


図1 部位検出器性能評価結果. (a) 頭部, (b) 下半身, (c) 胴体, (d) 左肩, (e) 右肩

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

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_t に基づいて次式より作成する.

$$W_b^j = \sum_{i: x_i \in X_j \wedge y_i = b} D_t(i) \quad (6)$$

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最も信頼度の高い弱識別器を選択する. なお ϵ の値は [1] に従い 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, 独立した検出器によっ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スコアマップ生成器を検出器として独立に使用した際の性能評価結果を示す. 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. 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分かる.

3.2 統合検出器

次に各部位の情報を統合することにより, ある方形内に人体が存在するかどうかの最終判断を行う. 3.2.1 において部位スコアマップを統合することにより人体検出を行う方法について示し, その上で 3.2.2 にて提案手法であるスコアマップ多段統合手法について説明を行う.

3.2.1 スコアマップ単純統合

各部位のスコアマップ \mathbf{M}_p を 1 列のベクトル \mathbf{m}_p に変換し, 全ての部位のスコアマップをまとめた統合ベクトル \mathbf{v} を以下によって得る.

$$\mathbf{v} = [\mathbf{m}_1^T, \mathbf{m}_2^T, \dots, \mathbf{m}_N^T]^T \quad (8)$$

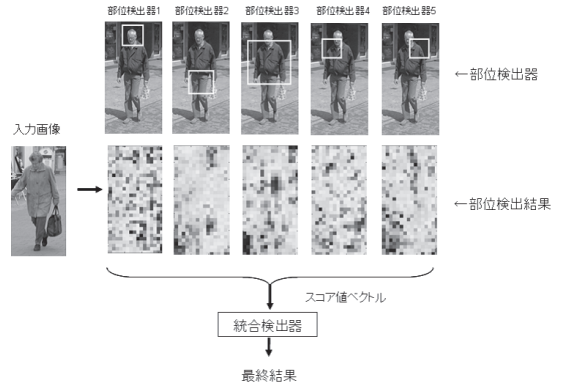


図2 スコアマップ単純統合手法フレームワーク

Fig. 2 Framework of parts score integration detection method.

得られた統合ベクトル \mathbf{v} を統合検出器 \mathcal{C} への入力とし, 対象方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 \mathcal{C} としては Real AdaBoost を用いた.

この手法は対象方形の中での各部位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」と呼ぶ.

3.2.2 スコアマップ多段階統合

本研究では更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 $\mathbf{v}^{(k)}$ を以下のように作成する.

$$\mathbf{v}^{(k)} = [\mathbf{m}_1^T, \dots, \mathbf{m}_{k-1}^T, \mathbf{m}_{k+1}^T, \dots, \mathbf{m}_N^T]^T \quad (9)$$

全ての 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, 一段階目の統合検出器 $\mathcal{C}^{(k)}$ を学習する. 今, 統合検出器 $\mathcal{C}^{(k)}$ 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

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

とする. これを入力として二段階目の統合検出器 \mathcal{D} を学習する. 本研究では二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 $\mathcal{C}^{(p)}$ が存在する. 人体以外の画像に対しては全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．

本論文ではこの手法を「スコアマップ多段階統合手法」と呼ぶ．

4. 性能評価実験

提案手法の効果を検証するために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，様々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，腰及び足先の位置を入力し，これらの点を基準に方形を切り出し，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．ネガティブデータセットから方形をランダムに切り出すことにより，ネガティブ画像 2500 枚を生成した．

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，下半身の一部及び左半身付近が隠れた画像を人工的に生成し，これ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．

評価を行った手法は以下の 4 種類である．

(1) 全身一括検知手法



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，(e) 左半身隠れ

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

(2) 単純統合手法 (スコアマップ使用せず)

(3) 単純統合手法 (スコアマップ使用)

(4) 多段階統合手法 (スコアマップ使用)

(1) 全身一括検知手法は，RealAdaboost によって全身画像をポジティブ，それ以外の画像をネガティブとして学習したものである．特徴量としては HOG 特徴量を利用した．(2) のスコアマップを使用しない単純部位統合手法は，部位情報を統合する人体検出手法として一般的なものであり様々な手法が存在するが，ここでは Chen ら [7] の手法を取り上げた．(3) は 3.2.1 で示した手法であり，(4) が提案手法である．

Chen らの手法の概要は次のとおりである．まず第一段階として部位検出器により部位の検出を行う．次に第二段階として人体検出に相当する方形をスライディングしながら，各部位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 $p_b^Q = \frac{1}{1 + \exp(-\mathbf{H}_S^Q(b^Q))}$ によって計算される．ここで $\mathbf{H}_S^Q(b^Q)$ 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 N 次元 (N は検出部位数) 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 らは，部位検出器として MCMIBoost と呼ばれる学習・識別手法を利用している．MCMIBoost はポジティブ学習インスタンスの自動選択が可能な点，及びマルチクラス対応可能な点が Real AdaBoost と異なる．本研究の主旨は，新たな統合フレームワークの提案にあるため，個々の部位検出器としては，幅広く活用されている Real AdaBoost を利用した実装・評価を行った．また利用部位についても [7] では頭部，胴体，下半身の 3 部位を利用しているが，本研究では比較のために，我々と同じ 5 部位に拡張した実装を行い，評価を行った．

性能評価結果を表 1 に示す．表の数値は誤検出率 10% 時の検出率である．通常の全身一括検知 (1) では隠れのない画像に対する検出率が 88.1% であるのに対して隠れのある画像ではいずれも性能が低下しており，特に頭部，肩，及び左半身の隠れ画像においてそれぞれ 39.6%，50.5%，35.6% と大幅に性能が低下している．またスコアマップを使用しない単純統合手法である (2) は，上述の隠れ画像に対してそれぞれ 55.5%，55.5%，42.6% と，(1) と比較して高い性能を示した

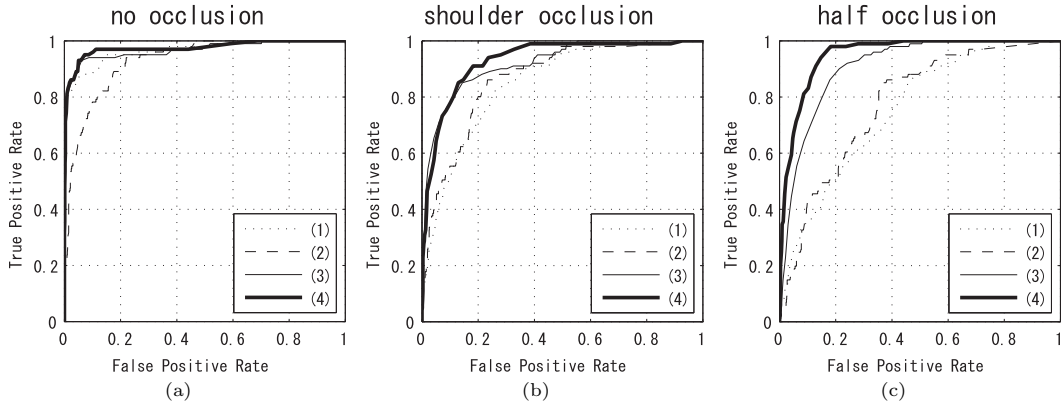


図4 ROC 曲線による各手法の検出性能評価. (a) 隠れなし, (b) 肩隠れ, (c) 左半身隠れ. 各図で (1) は通常の全身検出手法, (2) はスコアマップを使用しない統合手法 (Chen ら), (3) はスコアマップ単純統合, (4) はスコアマップ多段統合手法を示す.

Fig. 4 ROC curves for each method. (a) without occlusion, (b) shoulder occlusion, (c) half occlusion. (1) Full body detection, (2) Body parts integration method without using scoremaps, (3) Simple scoremap integration method, (4) Layered scoremap integration method.

表1 単純統合手法と多段統合手法検出率, FPR = 10%時. (1) は通常の全身検出手法, (2) はスコアマップを使用しない統合手法 (Chen ら), (3) はスコアマップ単純統合, (4) はスコアマップ多段統合手法を示す.

Table 1 Detection rate of simple integration and layered integration method, FPR = 10%. (1) Full body detection, (2) Body parts integration method without using scoremaps, (3) Parts scoremap integration method (simple), (4) Parts scoremap integration method (layered).

	(1)	(2)	(3)	(4)
no occlusion	88.1%	78.2%	94.1%	96.0%
head occlusion	39.6%	55.5%	59.5%	53.5%
shoulder occlusion	50.5%	55.5%	76.2%	78.2%
waist occlusion	84.2%	76.2%	92.1%	94.1%
half occlusion	35.6%	42.6%	64.4%	83.2%

が、隠れのない画像では検出性能が低下している。

これに対して、スコアマップを使用する統合手法である (3), (4) は、いずれの画像に対しても (1) と比較して高い性能を示しているが、中でも提案手法である (4) では隠れのない画像に対する検出率が 96.0% に向上し、更に肩、下半身、左半身隠れの画像についてもそれぞれ 78.2%, 94.1%, 83.2% と高い性能を保っており、4 手法中最も高い検出率を示している。なお頭部隠れの画像に関しては、いずれの手法も低い検出率を示した。提案手法は全身検知と比較して高い性能を示すものの、その検出率は従来の統合手法 (2) と同程度にとどまった。また本実験では、下半身隠れについ

ては、いずれの手法においても大幅な性能低下は見られなかった。

以上より、提案手法は、頭部付近の完全な隠れに対して従来の統合手法と同程度の性能向上を示し、それ以外の頭部隠れを除く様々な状況においては、大幅に高い性能を示すことが確認された。これらのことから、部位統合にスコアマップを用い、更にそれを多段統合する提案手法の有効性が確認できる。

図 4 に、それぞれの手法の隠れなし、肩隠れ、左半身隠れに対する ROC 曲線を示す。提案手法である (4) が最もよい性能を示していることが確認できる。

なお実験は、MATLAB によって実装したコードを、PC (Intel corei5, 2.67 GHzCPU) 上で実行することにより行った。処理時間は画像 1 枚当たり平均 26.2s であり、そのうち 99% がスコアマップの作成時間であった。現在のところコードは最適化されておらず、高速化の工夫もなされていないため比較的長い処理時間がかかっている。部位検出器にカスケード構造を取り入れることにより、スコアマップ作成に要する処理時間は大幅な削減が可能であると考えられる。その他、色情報を用いた探索領域の削減などの対策を行うことで、よりいっそうの高速化が可能であると考えられる。

5. 考 察

本章では複数部位のスコアマップ統合、及びそれを多段に行うことが有効な理由について考察を行う。

前章の実験結果において、スコアマップを使用する統合手法である (3) は、使用しない手法である (2) と比較して高い性能を示した。個別の人体部位は、顔や人体全体ほど複雑な特徴をもたないため、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。(2) 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。スコアマップを使用することにより、各部位について検出の有無だけでなく、その位置情報が統合の際に使用される。そのため (3) においては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している。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、隠された部位により、影響をあまり受けない統合検出器が存在することがある。人体以外の画像に対しては全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と人体以外の画像とを、より高い精度で区別できるものと考えられる。

このことを検証するために、スコアマップ単純統合手法と多段階統合手法について、隠れのある画像に対する、統合検出器の出力比較を行った。図 5 が実験に使用した画像である。これらの画像に対して各手法を適用し、統合検出器の出力を示したものが図 6 であり、 C がスコアマップ単純統合手法の統合検出器、 $C^{(k)}$ は多段階統合手法における部位 k を欠いた第一段統合検出器、 D は多段階統合手法の第二段統合検出器を示している。

隠れのない画像に対しては全ての検出器が高いスコアを出力しているのに対して、隠れのある画像に対してはスコアマップ単純統合手法の統合検出器 C の出力はほぼ 0 となっている。そのためスコアマップ単純統合手法ではこの人体を検出することはできない。しかしながら、多段階統合手法における第一段統合検出器には高いスコアを出力するものがあり、中でも左肩部の



図 5 スコアマップ単純統合手法と多段階統合手法の比較実験に用いた画像。左：隠れなし，右：隠れあり。

Fig. 5 Images for classifier responses comparison test. Left: no occlusion, right: occlusion.

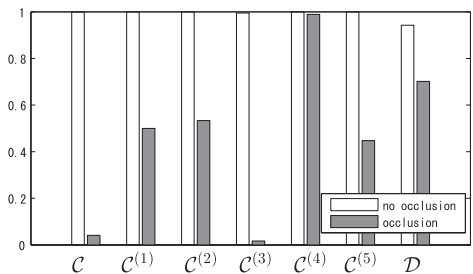


図 6 スコアマップ単純統合手法と多段階統合手法の検出器出力比較。スコアマップ単純統合手法の統合検出器 C 、多段階統合手法における部位 k を欠いた第一段統合検出器 $C^{(k)}$ 、及び多段階統合手法の統合検出器 D 。

Fig. 6 Classifier responses for each integration method. Simple integration method classifier C , 1st layer integration classifiers $C^{(k)}$, which lacking part k information, in the multi integration method framework. Layered integration classifier D .

情報を欠いた $C^{(4)}$ は高い出力を保っている。このため、多段階統合手法の統合検出器 D はこの画像に対しても高い出力値を示し、結果としてこの画像を人体として検出することができる。このような理由により、多段階統合手法は隠れのある画像に対して単純統合手法より高い検出性能を示したものと考えられる。

図 7, 図 8 に多段階統合手法による検出の成功例及び失敗例を示す。誤検出の傾向として、縦に長いエッジがある領域で失敗する例が多く見られた。これは、今回使用した部位のうち多くの部位において、長い縦エッジが領域内にあることが重要な判断要素になっているためと推測される。また、人体の一部のみを検出して、全身を検出できない例も見られた。これは一部の部位が何らかの理由により通常と異なる特徴を示したために、部位検出器が本来の位置で高いスコアを示さなかったこと、及び同時に他の位置で高いスコアを



図 7 検出成功例

Fig. 7 Human detection success examples.



図 8 検出失敗例

Fig. 8 Human detection failure examples.

示したことが重なったことによって起きる現象であると考えられる。また隠れに対応したことにより、頭部のない人形を誤検出する例が見られた。

なお統合手法に関しては人体全体を一つの部位とみなして使用することも可能であるが、本研究では全身検知手法との差異を明確にするため、統合手法においては人体の部分領域のみを用いている。人体全体を部位として含めた場合の性能評価は、今後の課題とする。

6. む す び

部位検出結果に基づく人体検出に関して、部位検出のスコアマップを多段に統合する手法を提案した。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、通常の全身一括検知と比較して、隠れのない画像に対する検出率（誤検出 10%時）が 88.1%から 96.0%に向上し、更に肩の隠れに対して 50.5%から 78.2%、左半身の隠れに対して 35.6%から 83.2%、それぞれ検出率が向上した。この理由として、スコアマップを用いることで各部位の位置情報が統合に活用され、これ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がより明確に区別可能になったこと、更に多段統合により一部の部位の隠れに対するロバスト性が向上したためと考えられる。なお評価実験の結果、頭部付近が完全に隠された場合の性能には依然として課題があることが明らかとなった。処理速度の向上、及びより大きな姿勢変動を含む画像への対応とともに今後の課題である。

文 献

- [1] 山内悠嗣, 藤吉弘亘, Hwang Bon-Woo, 金出武雄人, “アピラ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 and D.M. Gavrila, “Monocular pedestrian detection: Survey and experiments,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.31, no.12, pp.2179–2195, Dec. 2009.
- [3] K. Mikolajczyk, C. Schmid, and A. Zisserman, “Human detection based on a probabilistic assembly of robust part detectors,” *Proc. ECCV*, vol.1, pp.69–82, 2004.
- [4] A. Mohan, C. Papageorgiou, and T. Poggio, “Example-based object detection in images by components,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.23, no.4, pp.349–361, 2001.
- [5] C. Papageorgiou and T. Poggio, “A trainable system for object detection,” *IJCV*, vol.38, no.1, pp.15–33, June 2000.
- [6] B. Wu and R. Nevatia, “Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors,” *IJCV*, vol.75, no.2, pp.247–266, 2007.
- [7] Y. Chen, C. Chen, Y. Hung, and K. Chang, “Multi-class multi-instance boosting for part-based human detection,” *Proc. International Conference on Computer Vision Workshops*, pp.1177–1184, 2009.
- [8] Y. Chen and C. Chen, “MFast human detection using a novel boosted cascading structure with meta stages,” *IEEE Trans. Image Process.*, vol.17, no.8, pp.1452–1464, 2008.
- [9] D. Forsyth and M. Fleck, “Body plans,” *Proc. 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,” *Proc. BMVC*, 2005.
- [12] P. Viola and M. Jones, “Robust real-time face detection,” *Int. J. Comput. Vis.*, vol.57, no.2, pp.137–154, 2004.
- [13] P. Viola, M. Jones, and D. Snow, “Detecting pedestrians using patterns of motion and appearance,” *Proc. ICCV*, 2003.
- [14] R.E. Schapire and Y. Singer, “Improved boosting algorithms using confidence-rated predictions,” *Mach. Learn.*, no.37, pp.297–336, 1999.
- [15] 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.
- [16] 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.
- [17] T. Mita, T. Kaneko, and O. Hori, “Joint Haar-like features for face detection,” *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,” Proc. IEEE International Conference of Computer Vision, 2007.
- [19] 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.
- [20] P. Sabzmejdani and 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,” Proc. CVPR, 2006.
- [22] P.F. Felzenszwalb and D.P. Huttenlocher, “Efficient matching of pictorial structures,” Proc. CVPR, 2000.
- [23] R. Ronfard, C. Schmid, and B. Triggs, “Learning to parse pictures of people,” Proc. ECCV, 2002.

(平成 24 年 6 月 5 日受付, 8 月 23 日再受付)



木下 航一

平 8 神戸大・工・システム工卒。平 10 同大大学院博士前期課程了。同年オムロン(株)入社。主として画像認識の研究開発に従事, 現在に至る。平 22 より名大・情・メディア博士後期課程所属。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 本会学術奨励賞。

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 13 高柳記念奨励賞, 平 13 本会ソサイエティ論文賞, 平 14 本会業績賞, 平 15 文部科学大臣賞, 平 16 IEEE Trans. MM 論文賞, 平 22 前島賞, 他受賞。IEEE フェロー, 情報処理学会会員。