

# Joint HOG 特徴を用いた 2 段階 AdaBoost による人検出

## Human Detection by Two Stages AdaBoost with Joint HOG

三井 相和† 山内 悠嗣† 藤吉 弘亘†

Tomokazu MITSUI†, Yuji YAMAUCHI†, Hironobu FUJIYOSHI†

† 中部大学 工学部情報工学科

†Department of Computer Science, Chubu University

E-mail: mitsui@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

### Abstract

近年、統計的学習方法と局所領域の low-level な特徴量を用いた人検出等の物体検出法に関する研究が多く取り組まれている。本稿では、複数の Histograms of Oriented Gradients(HOG) 特徴量を組み合わせた Joint HOG 特徴と、2 段階に構築した AdaBoost による学習法を提案する。本研究では、複数の low-level 特徴である HOG 特徴を AdaBoost により組み合わせることで mid-level な特徴である Joint HOG 特徴を自動生成する。Joint HOG 特徴は、2 つのセル間の HOG 特徴量の共起を表現し、1 段階目の AdaBoost によって組み合わせる。このため、HOG 特徴のみでは捉えることができない、人の対称的な形状や連続的なエッジを捉えることが可能となる。次に、生成された Joint HOG 特徴を入力とした 2 段階目の AdaBoost によって最終識別器を構築する。これにより識別に有効な Joint HOG 特徴のみを選択するため、高精度な検出が可能となる。提案手法の有効性を確認するため、屋外で撮影した映像から作成したデータベースを用いて従来法との比較実験を行った結果、誤検出率 5.0% において、提案手法は従来法より約 24.6% 検出率を向上させることができた。また、本手法で用いる low-level な特徴は、HOG 特徴量以外にも利用することが可能である。HOG 特徴量と時空間特徴に基づく Pixel State Analysis(PSA) 特徴を併用した結果、誤検出率 5.0% において、約 99% の検出率を達成することができた。

## 1 はじめに

近年、セキュリティや ITS(Intelligent Transport System)、マーケティングなどの分野で人検出の実現が期待され、多くの手法が提案されている [1]-[7]。近年提案された人検出法は、Viola と Jones の顔検出法 [8] と同様に、low-level 特徴と Boosting による統計学習からなる手法が多い。人検出に用いられる low-level 特徴とし

ては、エッジベースの特徴量 [1]-[3] が多く提案されている。これらは高い検出性能を示すことから、人の形状を捉えることができるエッジベースの特徴量は、人を検出する際に有効な特徴であるといえる。

本研究の検出対象である人は、大きく分けて 2 つの形状の特徴がある。

1. 頭から肩にかけての  $\Omega$  に似た形状や上半身から下半身にかけての連続的な形状
2. 頭部や肩、胴、足などの左右対称的な形状

1. に対しては、局所領域内の 4 方向のエッジ特徴を AdaBoost により組み合わせることにより局所領域の面の情報を表現した Shapelet 特徴 [7] がある。2. に対しては、AdaBoost の弱識別器が複数の特徴量を同時に観測し、共起を表現する Joint Haar-like 特徴 [9] が提案されている。両手法は、非常に高い検出性能を示すことから、Boosting により low-level な特徴量を組み合わせることで mid-level な特徴量を得ることは、物体の検出に適した特徴量の生成が可能だと考えられる。そのため、特徴をどのように捉えるかが重要である。しかし、1. と 2. の両方の形状特徴を捉えることができる手法は提案されていない。

そこで、本研究では、物体形状の対称性や連続性を自動的に捉えることが可能である mid-level な特徴量として Joint HOG 特徴と 2 段階 AdaBoost による人検出法を提案する。Joint HOG 特徴は、異なる 2 つの領域の HOG 特徴量を AdaBoost により組み合わせることで、単一の特徴量では表現が困難な対称性や連続的なエッジを 1 段階目の AdaBoost により生成する。生成された Joint HOG 特徴量を入力とした 2 段階目の AdaBoost によって最終識別器を構築し、人の検出を行う。

## 2 Joint HOG 特徴

提案手法の概要を図 1 に示す。本手法は、2 段階に AdaBoost を構築する。まず、1 段階目の AdaBoost に

より位置の異なる2つの low-level な特徴である HOG 特徴を組み合わせた Joint HOG 特徴 POOL を作成する。これにより、複数の HOG 特徴を同時に観測するため、従来の単一の HOG 特徴のみでは捉えることができない、対称性や連続的なエッジを表現する Joint HOG 特徴を自動生成することが可能となる。次に、2段階目の AdaBoost により Joint HOG 特徴 POOL から自動的に人検出に最適な Joint HOG 特徴を選択し、最終識別器により人の検出を行う。

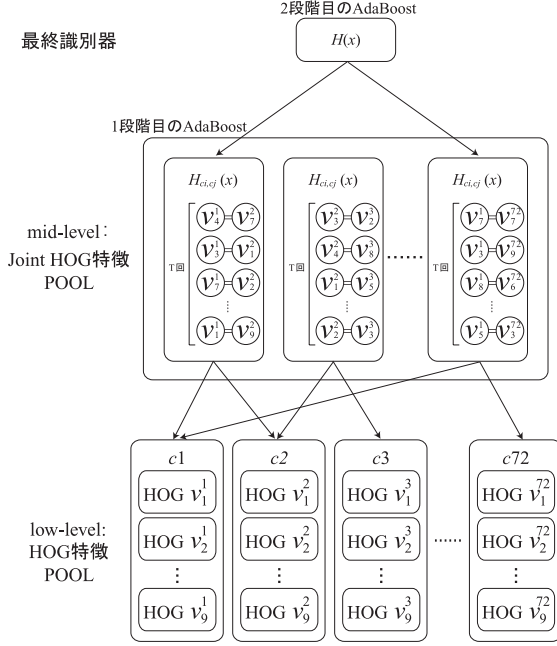


図 1 提案手法の流れ

## 2.1 low-level 特徴：HOG

本研究では、low-level な特徴として Dalal らが提案した Histograms of Oriented Gradients(HOG)[2]を用いる。HOG 特徴はセル (図 2(b)) における勾配方向をヒストグラム化した特徴であり、物体の形状を表すことが可能である。HOG 特徴は照明の変化を受けにくく、局所的な幾何学変化に頑健な特徴となる。下記に HOG 特徴算出の手順を示す。

各ピクセルの輝度  $L$  から勾配強度  $m$  と勾配方向  $\theta$  を次式より算出する。

$$m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \quad (1)$$

$$\theta(x, y) = \tan^{-1} \frac{f_y(x, y)}{f_x(x, y)} \quad (2)$$

$$\begin{cases} f_x(x, y) = L(x+1, y) - L(x-1, y) \\ f_y(x, y) = L(x, y+1) - L(x, y-1) \end{cases} \quad (3)$$

算出した勾配強度  $m$  と、勾配方向  $\theta$  よりセル (5×5 ピクセル) ごとに輝度の勾配方向ヒストグラムを作成する。ただし、勾配方向  $\theta$  は  $0^\circ \sim 360^\circ$  となるが、本研

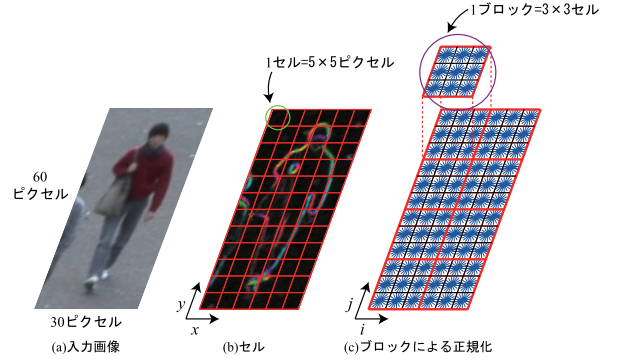


図 2 提案手法で用いる領域の構造

究の検出対象である人の場合には、勾配の向きを考慮する必要がないため  $0^\circ \sim 180^\circ$  とする。勾配方向を  $20^\circ$  ずつに分割することにより、9 方向の勾配方向ヒストグラムを得る。

最後に、次式によりブロック領域毎 (図 2(c)) に特徴量を正規化する。

$$v = \frac{v}{\sqrt{\left(\sum_{i=0}^k v(i)\right)^2 + \epsilon}} \quad (\epsilon = 1) \quad (4)$$

ここで、 $v$  は HOG 特徴量、 $k$  はブロック内の HOG 特徴量の数、 $\epsilon$  は分母が 0 の場合に計算不能になることを防ぐ係数である。

## 2.2 HOG 特徴量の共起

Joint HOG 特徴を生成するために、複数の HOG 特徴量により共起 [9] を表現する。まず、次式から人もしくは人以外を表す 2 値化符号  $s$  を算出する。

$$s(\mathbf{V}) = \begin{cases} 1 & p \cdot v_o > p \cdot \theta \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

ここで、 $\theta$  はしきい値、 $p$  は不等号の向きを決定する符号であり、1 もしくは 0 をとる。 $\mathbf{V} = [v_1, v_2, \dots, v_9]$  は 1 つのセルから算出される特徴量、 $o$  は勾配の方向を表す。これにより、得られた 2 値化符号を 2 つ組み合わせることで共起 [9] を表現した特徴  $j$  を得る。例えば、図 3 のような入力画像において、HOG 特徴を 2 値化した符号  $s_1 = 1$ 、 $s_2 = 1$  を観測したとき、共起を表現した特徴  $j$  は  $j = (11)_2 = 3$  となる。共起を表現した特徴  $j$  は 2 進表現された特徴の組み合わせのインデックス番号であり、今回は 2 つの特徴の組み合わせとなるため、4 通りの値をとる。

## 2.3 mid-level 特徴：Joint HOG 特徴

ここまで算出した HOG 特徴量の共起を表現した特徴量を組合わせて mid-level な特徴量として表現する。2.2 にて算出した HOG 特徴の共起を用いて、2 つのセルの low-level 特徴から算出した共起を表現した特徴量

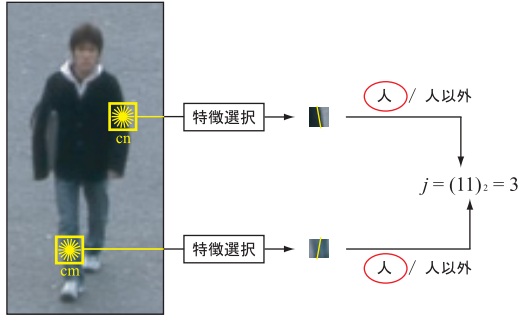


図3 HOG 特徴量の共起

と1段階目 AdaBoost により Joint HOG 特徴を生成する。これにより、物体形状の対称性やエッジの連続性だけでなく、セルとセルの関係を捉えることが可能となる。

まず、位置の異なる2つのセル  $cm$ ,  $cn$  において、共起を表現した特徴の中から、識別に有効な特徴を選択する。入力画像  $x$  から HOG 特徴の共起を観測する関数を  $J_t(x)$  で表す。入力画像  $x$  からの特徴量  $J_t(x) = j$  を観測したとき、1段階目の AdaBoost の弱識別器  $h_t(x)$  を次式により表す。

$$h_t(x) = \begin{cases} +1 & P_t(y = +1|j) > P_t(y = -1|j) \\ -1 & \text{otherwise} \end{cases} \quad (6)$$

ここで、 $y$  は正解ラベル  $y \in \{+1, -1\}$  を表し、 $P_t(y = +1|j)$  及び  $P_t(y = -1|j)$  は、それぞれ HOG 特徴の共起を表現した特徴  $j$  を観測したときの条件付き確率である。条件付き確率は、学習サンプル  $i$  の重み  $D_t(i)$  に基づき次式により算出される。

$$P_t(y = +1|j) = \sum_{p: J_t(x_p) = j \wedge y_i = +1} D_t(i) \quad (7)$$

$$P_t(y = -1|j) = \sum_{p: J_t(x_p) = j \wedge y_i = -1} D_t(i) \quad (8)$$

1 段階目の強識別器である Joint HOG 特徴  $H_{cm, cn}(x)$  を次式より構築する。

$$H_{cm, cn}(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t^{cm, cn} \cdot h_t^{cm, cn}(x) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

ここで、 $T$  は学習回数、 $\alpha_t^{cm}$ ,  $\alpha_t^{cn}$  は1段階目の弱識別器  $h_t^{cm, cn}$  の重みを表す。

上記の処理を全てのセルの組み合わせにおいて行う。本手法では、検出ウィンドウ ( $30 \times 60$  ピクセル) に対して 72 個のセルがあるため、2 つのセルの組み合わせは  ${}_{72}C_2 = 2,556$  となり、2,556 個の Joint HOG 特徴 POOL が作成される。この Joint HOG 特徴 POOL から、後述する2段階目の AdaBoost により最終識別器を構築する (図4)。

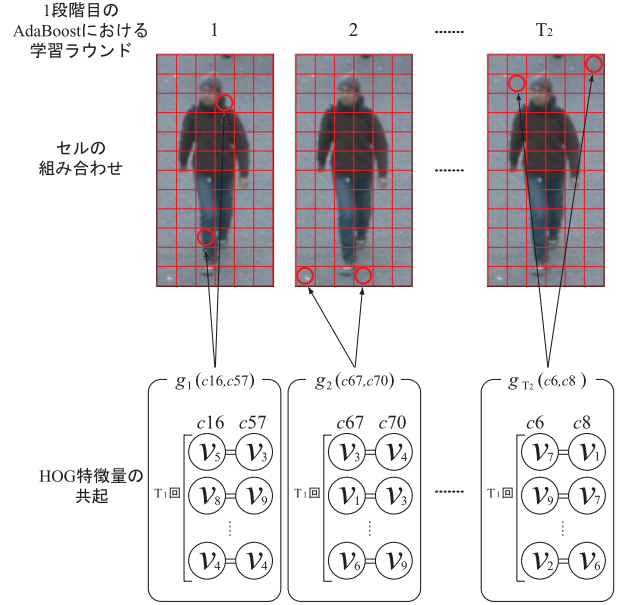


図4 Joint HOG 特徴

## 2.4 2段階目の AdaBoost による最終識別器の構築

2段階目の AdaBoost では、1段階目の AdaBoost により生成した Joint HOG 特徴 POOL を AdaBoost の入力として最終識別器を構築する。これにより、識別に有効なセルとセルの関係を表現した Joint HOG 特徴を自動的に選択することになる。2段階目の AdaBoost により得られる最終的な強識別器  $G(c)$  を得る。

$$G(c) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t \cdot g_t(c) > \lambda \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

ここで  $g_t(c)$  は1段階目の AdaBoost より得られる強識別器であり、 $\lambda$  は検出器の閾値である。 $cm$  と  $cn$  はセルの組み合わせを表し、 $c = \{cm, cn : 1 \leq m \leq 72, 1 \leq n \leq 72\}$  である。2段階目の AdaBoost により、Joint HOG 特徴 POOL から識別に有効な特徴量のみを選択して、検出器を構築する。

## 3 Joint HOG 特徴を用いた人検出実験

本手法の有効性を示すため評価実験を行う。

### 3.1 データベース

データベースには [6] で用いたデータベースを利用する。データベースに含まれる画像は、固定カメラにより撮影された多様な照明、背景、人の歩行方向など、異なる複数の場所で撮影された画像である。[6] と同様に、学習用ポジティブサンプルは2,054枚、学習用ネガティブサンプルは6,258枚、評価用ポジティブサンプルは1,000枚、評価用ネガティブサンプルは1,234枚用いる。学習に用いたデータベースの一部を図5に示す。

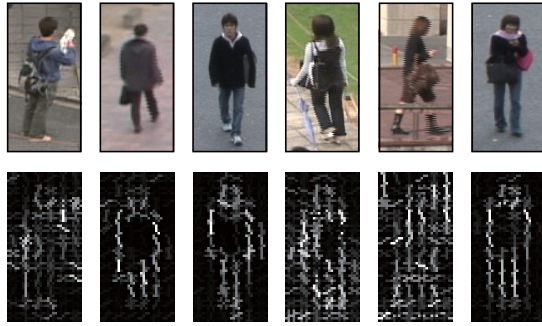


図5 学習に用いたデータベース

### 3.2 実験概要

評価用データベースを用いて、人の検出実験を行う。提案手法 (Joint HOG) と、HOG 特徴 + AdaBoost (HOG) [2] との比較を行う。評価には Detection Error Tradeoff (DET) を用いる。DET とは、横軸に誤検出率、縦軸に未検出率をとり、両対数グラフにより表したものであり誤検出率に対する未検出率の比較を行うことができる。また、DET は原点に近いほど性能が良い。

### 3.3 実験結果

図6にDETによる検出結果を示す。提案手法は従来法と比較し、検出精度が向上しているのがわかる。誤検出率5.0%の場合、約24.6% 検出率を向上させることができた。これはHOG特徴量のみでは識別困難なパターンに対しても、Joint HOG特徴を用いることにより識別困難なパターンを捉えることが可能となったためである。

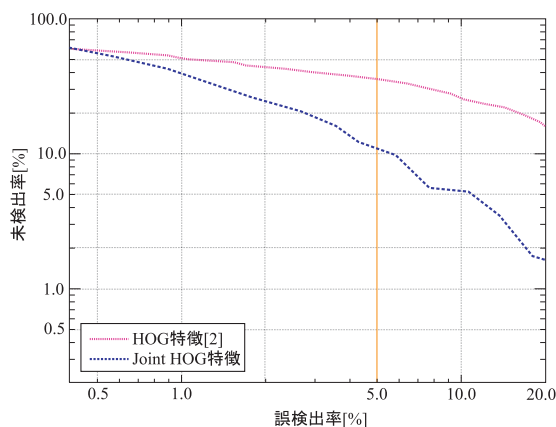


図6 DET カーブ

### 3.4 考察

2段階のAdaBoostにより選択されたJoint HOG特徴量について考察を行う。図7(a)に1段階目のAdaBoostにより、選択されたHOG特徴を可視化した結果、図7(b)に2段階目のAdaBoostにより、選択されたHOG特徴を可視化した結果を示す。また、図7(c)において、2段階目のAdaBoostにより選択された2つ

のセルとJoint HOG特徴をラウンド毎に示す。HOG特徴の勾配方向を9方向で表現しており、輝度が高いほどAdaBoostにおける弱識別器の重み $\alpha$ が高いことを表す。図7(a)では、全てのセル領域においてHOG特徴が選択されるが、特に人の形状に沿った勾配が高い弱識別器の重みであることがわかる。次に図7(b)に注目する。図7(a)で選択されたHOG特徴量であっても人の輪郭以外は選択されにくい傾向がある。これは、2段階目のAdaBoostの特徴選択において、識別に有効ではないと判断されたためである。そのため、2段階のAdaBoostは有効だと考える。最後に図7(c)に注目する。2段階目のAdaBoostにより選択されたJoint HOG特徴は、人の輪郭に沿ったセルが選択される傾向があることがわかる。

## 4 Joint HOGと時空間特徴を用いた人検出実験

本手法に用いるlow-level特徴は、HOG特徴量以外にも利用可能である。本章では、HOG特徴と[6]により有効性が確認された時空間特徴に基づくピクセル状態分析 (Pixel State Analysis : PSA) を併用した人検出法について述べる。

### 4.1 時空間特徴の併用

文献[6]により有効性が確認された時空間特徴に基づく特徴としてピクセル状態分析 (PSA) [10] を用いる。ピクセル状態分析とは、ピクセル状態の時間変化をモデル化することにより、各ピクセルを背景 (Background)、静状態 (Stationary)、動状態 (Transient) の三状態に判別する手法である (図8)。

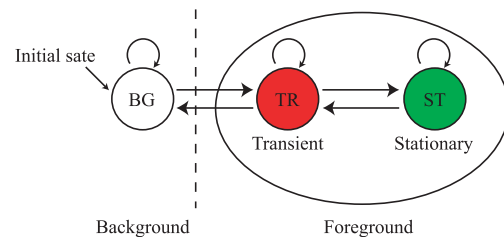


図8 ピクセルの状態遷移図

三状態に判別するためには、前後数フレームを用いて急激な輝度の変化 (Motion Trigger) と安定度 (Stability Measure) を用いる (図9)。

ピクセル状態分析結果をセル領域によってヒストグラム化することにより、時空間特徴となる特徴量を算出する。HOGで用いた領域の構造を用いて、HOGと同様にピクセル状態ヒストグラムを作成する。ピクセルは三状態に判別されるので、一つのヒストグラムから三つの特徴量が算出される。HOG特徴量にPSA特徴を加えることにより、1セルあたり12個の特徴量が



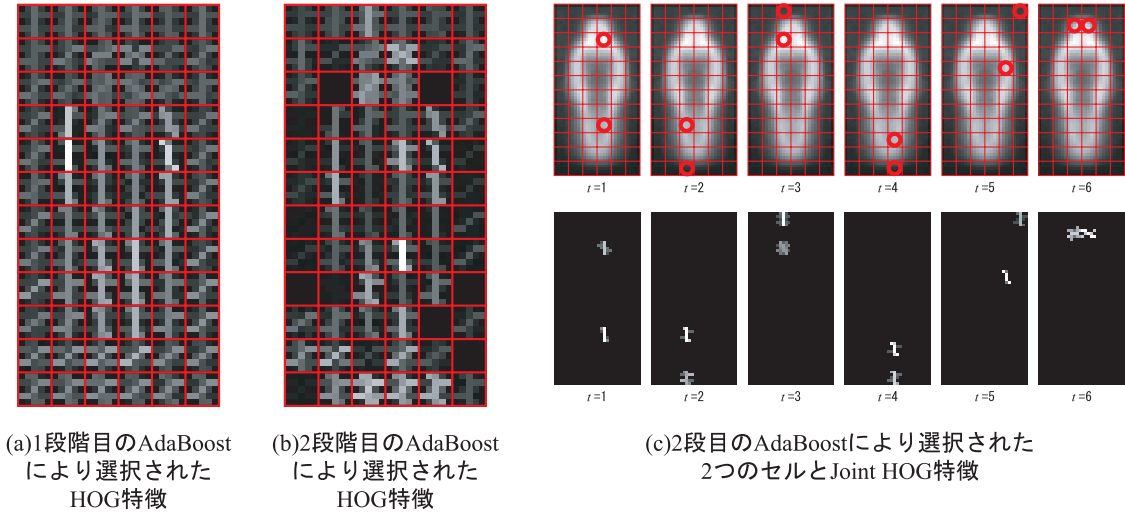


図 7 Joint HOG 特徴の可視化

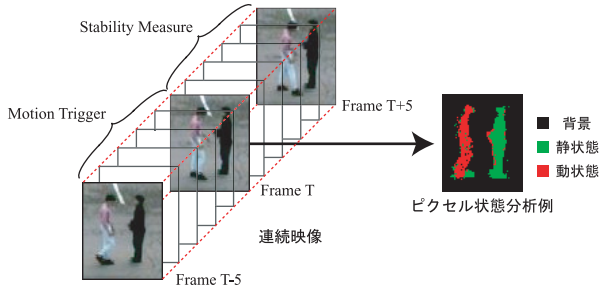


図 9 ピクセル状態分析例

得られる．HOG 特徴と PSA 特徴を併用した人検出法を図 10 に示す．

#### 4.2 実験概要

3.1 で述べたデータベースを用いて評価実験を行う．実験はピクセル状態分析の結果を特徴量として抽出する．学習で使用するデータベースの一部を図 11 に示す．実験は，提案手法（Joint HOG-PSA + 2 段階 AdaBoost）と，従来法として，HOG 特徴と PSA 特徴の共起 + AdaBoost(HOG+PSA(共起))[6] との比較を行う．

#### 4.3 実験結果

図 12 に DET による検出結果を示す．提案手法は，従来法である HOG 特徴と PSA 特徴の共起と比較して，検出性能が向上している．誤検出率 5.0% の場合，約 2.2% 検出率が向上し，検出率約 99% を得ることができた．

#### 4.4 考察

学習時に選択された特徴量から提案手法の考察を行う．図 13(a) に，学習用ポジティブサンプルを用いたときの平均勾配画像，図 13(b)，(c)，(d) に学習用ポジティブサンプルのピクセル状態分析の結果の画像を用いて，各状態毎の出現する頻度を画像として可視化し

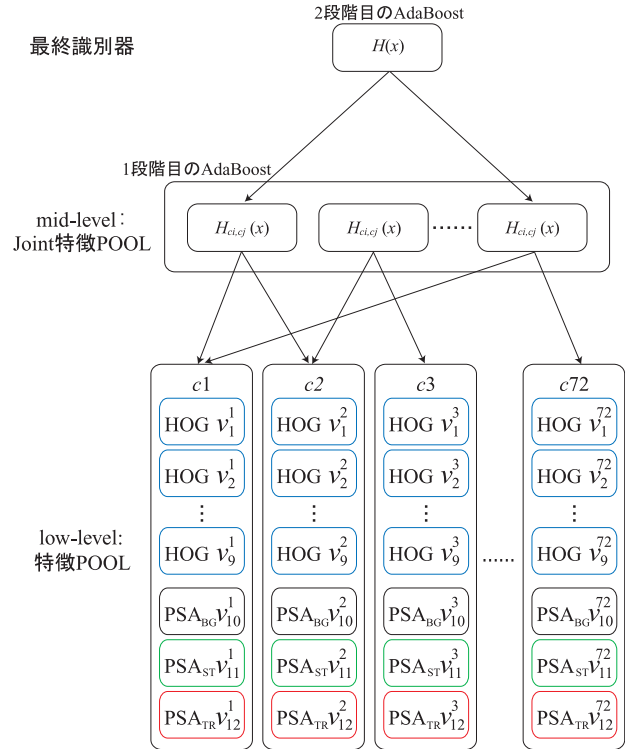


図 10 HOG 特徴と PSA 特徴を併用した人検出法

たものを示す．図 13 は，輝度が高いほど勾配が強い，もしくは頻度が高いことを表す．図 13(b) の背景の状態の平均画像に注目したとき，人のシルエットを表現していることがわかる．これは，ピクセル状態分析が背景と前景を明示的に区別していないが，背景とそれ以外の 2 クラス問題として扱うことにより，背景と前景を区別することが可能となっていることを表す．また，図 13(c)(d) から，静状態のピクセルは人の上半身部分の頻度が高く，動状態のピクセルは人の足の頻度が高

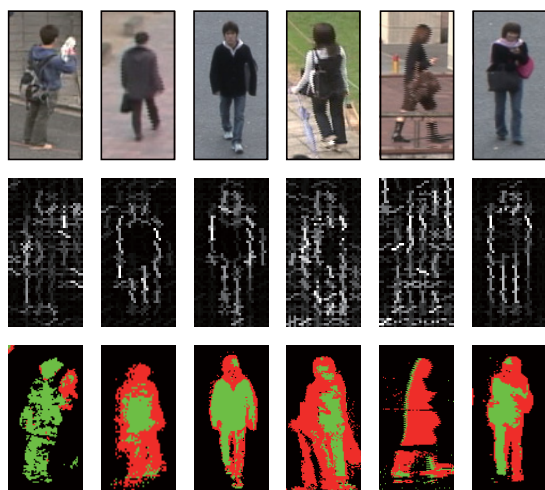


図 11 学習に用いたデータベース

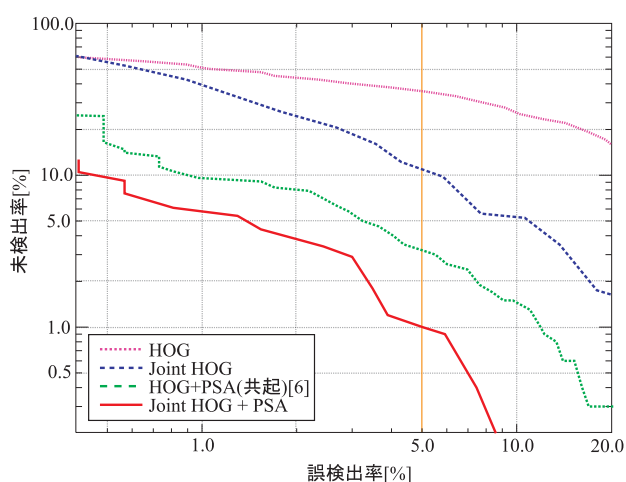


図 12 DET カーブ

いことがわかる。

実際に AdaBoost により選択された特徴量の可視化を図 14 に示す。学習の初期に多く選択されている特徴量として、ピクセル状態分析の背景から得られる特徴が用いられている。これは、背景の状態を占めるピクセルが多い/少ないによって、人を判別していることを表す。

また、可視化結果から、学習のはじめでは、PSA 特徴が選択され、学習の後半では HOG 特徴が選択される傾向があることがわかる。この詳細を調べるために、1 段階目の AdaBoost において学習ラウンド毎に選択された HOG 特徴と PSA 特徴の割合を図 15 に示す。学習の始めでは、非常に高い割合で PSA 特徴が選択され、5 ラウンドを越えた辺りから HOG 特徴量が多く選択されることがわかる。これは、識別時において、まず物体の動きを表すことが可能な PSA 特徴により、おおまかに人と人以外を判別し、この後より詳細に識別境界を形成するためにアピアランス情報である HOG 特徴

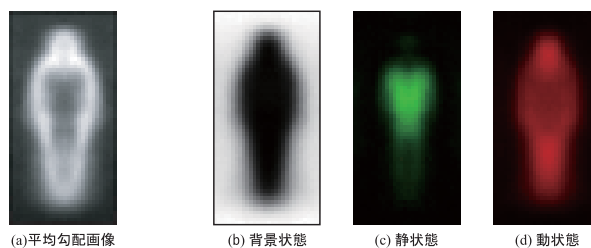


図 13 各状態の出現頻度の可視化画像

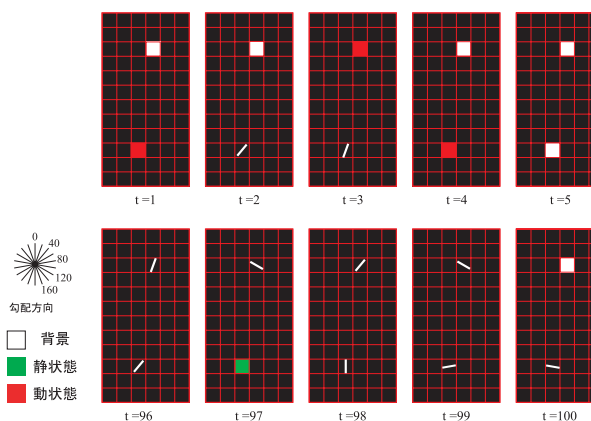


図 14 学習によって選択された特徴量

量が選択されたと考えられる。

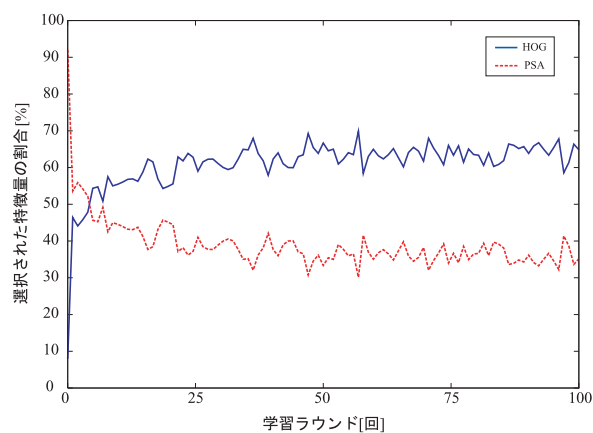


図 15 選択された特徴の割合

#### 4.5 人検出による人数のカウント実験

構築した検出器による人検出の実験を行い、その結果から画像中に存在する人数のカウントを行う。

##### 4.5.1 実験概要

人検出を行うシーケンスは、3 箇所撮影した映像を用いる。人検出を行うために、検出ウィンドウを画像左上からスケールを変化させて複数回ラスタスキャンする。これにより、人のスケールが異なる場合であっても検出が可能となる。人として検出されたウィンドウ

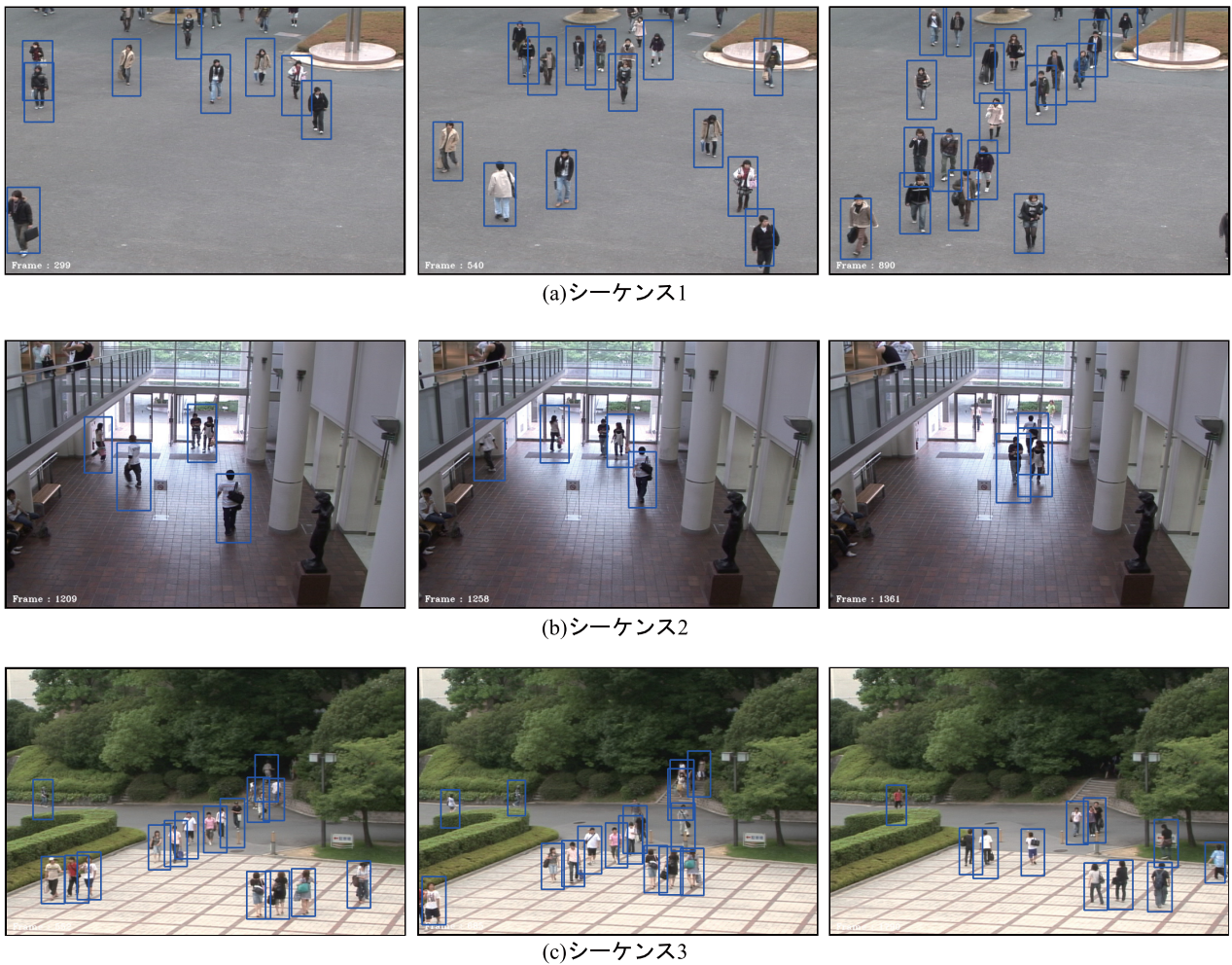


図 16 人検出例

は、最後に Mean Shift クラスタリングによるウィンドウの統合処理を行う [11].

人検出した結果から、人数のカウントを行う。なお、人数のカウントで使用する人検出結果の画像は、3 シーケンスの中から合計 500 フレームを使用する。画像中に人が存在する人数の真値は、目視により計測した値を真値とする。基準としては、画像外に人がはみ出している場合は検出対象としない。また、画像上で人同士の重なりが発生した場合、人の体が半分隠れてしまう場合も検出対象としない。評価は、次式で表される検出性能によって評価する。

$$\text{検出性能} = \frac{\text{正識別数} - \text{誤検出数}}{\text{真値}} \times 100 \quad (11)$$

#### 4.5.2 実験結果

図 16 に Joint HOG と時空間特徴を用いた人検出例を示す。提案手法は、時空間特徴を用いているため、人に似た形状を持つ物体や複雑な背景に対して高精度な検出ができていることがわかる。

次に、人検出の結果から、人数のカウントを行う。人数カウントの結果を表 1 に示す。

表 1 人数カウント結果

| 真値 [人] | 検出数 [人] | 誤検出数 [人] | 検出性能 [%] |
|--------|---------|----------|----------|
| 3,004  | 2,783   | 33       | 93.7     |

提案手法による人数のカウントは、検出性能 93.7% を得た。通常、マーケティングにおいて、人による人数カウントの精度は約 90% と言われている。提案手法では、自動的に人数のカウントを高精度に行うことができるため、マーケティングの分野等での利用が期待できる。

## 5 おわりに

本稿では、複数の HOG 特徴を組み合わせた Joint HOG 特徴と、2 段階に構築した AdaBoost による学習法を提案した。Joint HOG 特徴は、物体形状の対称性やエッジの連続性だけでなく、セルとセルの関係性を特

徴として捉えることが可能となるため、識別に有効な特徴を得ることができた。また、時空間特徴に基づくPSA特徴を併用することにより、さらに高精度な検出を実現することができた。今後は、提案した検出法を利用した、人領域のセグメンテーションを行う予定である。

## 参考文献

- [1] K. Levi and Y. Weiss, “ Learning Object Detection from a Small Number of Examples: the Importance of Good Features. ”, IEEE Computer Vision and Pattern Recognition, vol. 2, pp. 53-60, 2004.
- [2] N. Dalal and B. Triggs, “ Histograms of Oriented Gradients for Human Detection ”, IEEE Computer Vision and Pattern Recognition, vol.1, pp. 886-893, 2005.
- [3] B. Wu and R. Nevatia, “ Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors ”, IEEE International Conference on Computer Vision, vol. 1, pp. 90-97, 2005.
- [4] P. Viola, M. Jones and D. Snow, “ Detecting Pedestrians Using Patterns of Motion and Appearance ”, IEEE International Conference on Computer Vision, pp. 734-741, 2003.
- [5] N. Dalal, B. Triggs and C. Schmid, “ Human Detection Using Oriented Histograms of Flow and Appearance ”, IEEE European Conference on Computer Vision, vol.2, pp. 428-441, 2006.
- [6] 山内悠嗣, 藤吉弘亘, Hwang 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), pp. 1492-1497, 2007.
- [7] P. Sabzmeydani and G. Mori: “ Detecting Pedestrians by Learning Shapelet Feature ”, IEEE Computer Vision and Pattern Recognition, pp. 511-518, 2007.
- [8] P. Viola and M. Jones, “ Rapid Object Detection Using a Boosted Cascade of Simple Features ”, IEEE Computer Vision and Pattern Recognition, vol. 1, pp. 511-518, 2001.
- [9] T. Mita, T. Kaneko and O.Hori: “ Joint Haar-like Features for Face Detection ”, IEEE International Conference on Computer Vision, pp. 1619-1626, 2005.
- [10] H. Fujiyoshi and T. Kanade, “ Layered Detection for Multiple Overlapping Objects ”, IEICE Transactions on Information and systems, vol. E87-D, pp. 2821-2827, 2004.
- [11] D. Comaniciu and P. Meer, “ Mean Shift Analysis and Applications ”, IEEE International Conference on Computer Vision, pp. 1197-1203, 1999.