

## 論 文

## Joint 特徴量を用いた 2 段階 Boosting による物体検出

三井 相和<sup>†a)</sup> 山内 悠嗣<sup>†b)</sup> 藤吉 弘亘<sup>†c)</sup>

## Object Detection by Two-Stage Boosting with Joint Features

Tomokazu MITSUI<sup>†a)</sup>, Yuji YAMAUCHI<sup>†b)</sup>, and Hironobu FUJIYOSHI<sup>†c)</sup>

あらまし 本論文では、複数の Histograms of Oriented Gradients (HOG) 特徴量を 2 段階に構築した Boosting により組み合わせた Joint 特徴量による特定対象 (例えば人や車両) の物体検出法を提案する。近年、統計的学習手法と局所領域より得られる low-level 特徴量を組み合わせた物体検出法に関する研究が多く取り組まれている。本手法では、複数の low-level 特徴量である HOG 特徴量を Real AdaBoost により組み合わせることで Joint 特徴量を自動生成する。Joint 特徴量は、複数のセル間の HOG 特徴量の共起を表現し、1 段階目の Real AdaBoost により組み合わせる。このため、単一の HOG 特徴量のみではとらえることのできない物体の対称的な形状や連続的なエッジをとらえることが可能となる。次に、生成された Joint 特徴量のプールを入力とした 2 段階目の Real 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 特徴量のみを選択するため、高精度な検出が可能となる。本論文では、提案手法の有効性を確認するために、検出対象を人と車両として、評価実験により提案手法の有効性を示す。また、提案手法では HOG 特徴量同士のみでなく異なる low-level 特徴量間での組合せも可能である。本論文では、歩行者のアピランスを表す HOG 特徴量と、動きを表す時空間特徴である PSA 特徴量との共起を表現することで、効果的な識別器を構築することが可能となることを示す。

キーワード 物体検出, Histograms of Oriented Gradients, Real AdaBoost, Joint 特徴量

## 1. ま え が き

物体検出とは、画像中より特定の対象物体を探索する技術である。この物体検出は、コンピュータビジョンの分野において重要な課題の一つであり、既に多くの手法が提案されている。近年の物体検出法の多くは、多くの学習サンプルから抽出した局所領域における特徴量の統計的学習手法により識別を行う手法である。このような物体検出法に用いられる low-level な特徴量としては、Haar-like 特徴 [1] や Edge Orientation Histograms (EOH) [2], Edgelet [3] などの特徴量が提案されている。中でも Histograms of Oriented Gradients (HOG) [4] は照明変化や幾何学的変化に頑健な特徴量であり、その有効性が示されている。しかし、これらの low-level 特徴量のみでは識別性能に限

界がある。そこで、近年では高精度な物体検出を行うために、low-level 特徴量を AdaBoost [5] などの統計的学習手法により組み合わせることで特徴量間の関連性をとらえる特徴量を生成する手法が提案されている。

本論文の検出対象の一つである人の形状には、下記に示すように大きく分けて二つの特徴がある。

(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で局所領域の面の情報を表現した Shapelet 特徴量 [6] がある。しかし、low-level 特徴量が非常に多いため、局所領域内に含まれる特徴量のみでの組合せに限定している。そのため、局所領域外の特徴量を同時に観測することができず、検出対象の物体の特徴をとらえる上で重要な形状の対称的な形状や連続的なエッジをとらえることができない。(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 特徴量 [7] が提案されている。この手法は、

<sup>†</sup> 中部大学工学部情報工学科, 春日井市  
Dept. of Computer Science, Chubu University, 1200  
Matsumoto, Kasugai-shi, 487-8501 Japan

a) E-mail: mitsui@vision.cs.chubu.ac.jp

b) E-mail: yuu@vision.cs.chubu.ac.jp

c) E-mail: hf@cs.chubu.ac.jp

従来の顔検出と比べ高速、高精度に識別することが可能であるが、特徴量を輝度値でとらえるため、テクスチャが多様な物体には不向きである。

両手法は、複数の low-level 特徴量を Boosting により組み合わせることで特徴量間の関連性をとらえることができ、高精度な検出が可能となる。そのため、特徴量をどのようにとらえるかが重要であるが、(1)と(2)の両方の形状特徴をとらえることができる手法は提案されていない。

そこで、本論文では、物体形状の対称性や連続性を自動的にとらえることが可能である Joint 特徴量による特定対象の物体検出法を提案する。Joint 特徴量は、まず、異なる複数の領域の HOG 特徴量を 1 段階目の Boosting により組み合わせることで、Joint 特徴量のプールを生成する。次に、生成された Joint 特徴量のプールを入力とした 2 段階目の Boosting によって最終識別器を構築する。これにより、単一の特徴量では表現が困難な対称性や連続性を表現する Joint 特徴量を自動生成し、対象物体の検出を行う。本論文では、検出対象を非剛体である人と視点の違いによる見え方の違いが大きい車両とし、評価実験により提案手法の有効性を示す。また、更に高精度な人検出を行うために、形状特徴以外の情報が必要である。本論文では、Joint 特徴量のフレームワークを利用し、歩行者のアップランスを表す HOG 特徴量と、動きを表す時空間特徴である PSA 特徴量との共起を表現する。これにより動きと見えの情報を同時に観測するため、より高精度な検出が期待できる。HOG 特徴量を組み合わせ

る Joint 特徴量と同様に、評価実験により提案手法の有効性を示す。

## 2. Joint 特徴量

提案手法の学習の流れを図 1 に示す。本手法は、2 段階の Boosting に分けて最終識別器を構築する。Boosting には弱識別器の出力値が連続値で得られる Real AdaBoost [8] を用いる。Real AdaBoost は、AdaBoost と比べ少ない弱識別器で高精度な検出が行える。まず、1 段階目の Real AdaBoost により位置の異なる複数の low-level 特徴量である HOG 特徴量を組み合わせた Joint 特徴量のプールを生成する。これにより、複数の HOG 特徴量を同時に観測するため、従来の単一の HOG 特徴量のみではとらえることのできない、対称性や連続性が表現可能となる。次に、2 段階目の Real AdaBoost により Joint 特徴量のプールから自動的に人検出に最適な Joint 特徴量を選択し、最終識別器により物体の検出を行う。

### 2.1 low-level 特徴量：HOG 特徴量

本論文では、low-level 特徴量として Dalal らが提案した Histograms of Oriented Gradients (HOG) 特徴量 [4] を用いる。HOG 特徴量はセル (図 1 (b)) と呼ばれる局所領域において輝度のこう配方向をヒストグラム化した特徴量であり、物体の形状をとらえることが可能である。HOG 特徴量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量となる。下記に HOG 特徴量の算出の手順を示す。

各ピクセルの輝度  $L$  からこう配強度  $m$  とこう配方

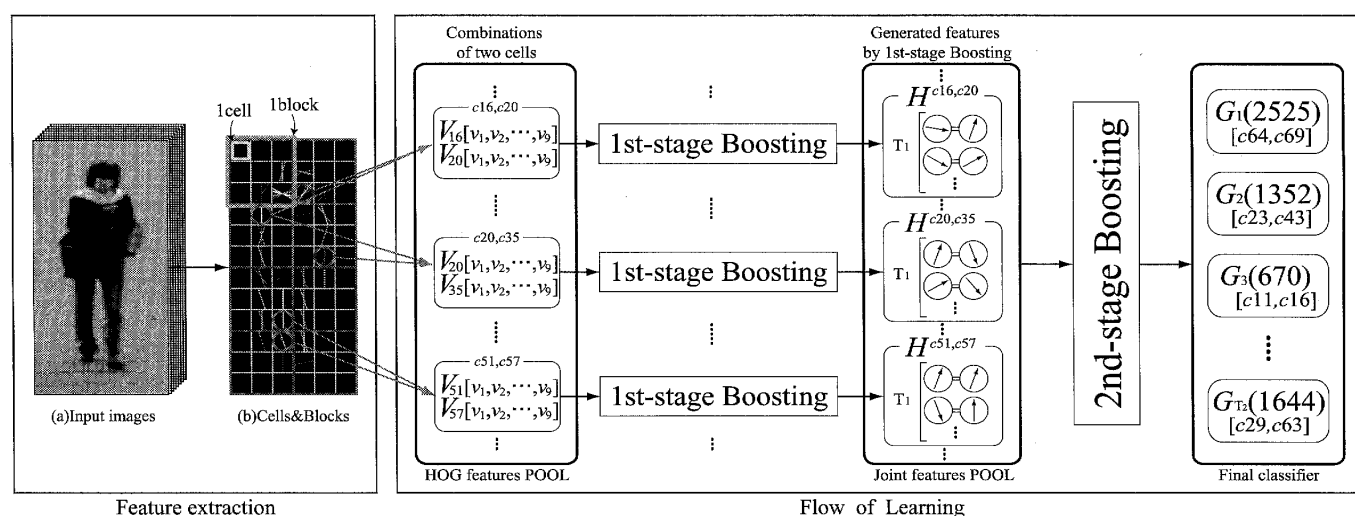


図 1 提案手法の学習の流れ  
Fig. 1 The flow of learning by the proposed method.

向  $\theta$  を次式より算出する.

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

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

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

算出したこう配強度  $m$  と, こう配方向  $\theta$  よりセルごとに輝度のこう配方向ヒストグラムを作成する. 得られたこう配方向を  $20^\circ$  ずつに分割することにより, こう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ごと (図 1(b)) に特徴量を正規化する.

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

ここで,  $v$  は正規化後の HOG 特徴量,  $j$  は注目している HOG 特徴量のブロック内での番号,  $k$  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. 本論文で用いたブロック領域のサイズは, 文献 [4] により有効性が報告されている  $2 \times 2 \sim 3 \times 3$  セルをそれぞれ車両と人の検出に用いた.

## 2.2 HOG 特徴量の共起

Joint 特徴量を生成するために, 複数の HOG 特徴量の共起 [7] を表現する. この共起表現法を用いることにより, 複数の特徴量を同時に観測することが可能であるが, 本節以降は二つの特徴量を組み合わせる場合について述べる.

まず, 次式から検出対象, 若しくは非検出対象を表す 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,  $o$  は注目している HOG 特徴量のこう配の方向,  $p$  は不等号の向きを決定する符号であり  $p \in \{+1, -1\}$  とする.  $\mathbf{V} = [v_1, v_2, \dots, v_q]$  は一つのセルから算出される HOG 特徴量,  $q$  は一つのセルに含まれるこう配の全方向数を表す. これにより, 得られた 2 値化符号を二つ組み合わせることで共起を表現した特徴量  $j$  を得る. 例えば, 図 2 のよう

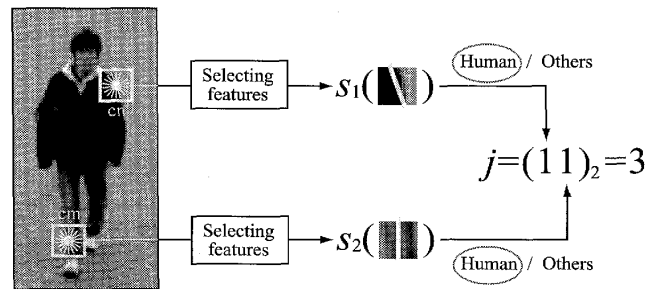


図 2 特徴量の共起表現

Fig. 2 Co-occurrence of features.

な入力画像において, HOG 特徴量を 2 値化した符号  $s(v_a) = 1, s(v_b) = 1$  を観測したとき, 共起を表現した特徴量  $j$  は

$$j = (s(v_a) s(v_b))_2 = (11)_2 = 3 \quad (6)$$

となる.  $j$  は 2 進表現された特徴量の組合せのインデックス番号であり, 二つの特徴量の組合せとなるため, 4 通りの値をとる.

## 2.3 Joint 特徴量

2.2 にて算出した HOG 特徴量の共起を用いて, 1 段階目の Real AdaBoost により Joint 特徴量を生成する. 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とらえることが可能となる.

まず, 位置の異なる二つのセル  $c_m, c_n$  において, 共起を表現した特徴量の中から, 識別に有効な特徴量を Real AdaBoost により選択する. ここで,  $N$  個の学習サンプルは  $(x_1, y_1), \dots, (x_N, y_N)$  とし,  $i$  は学習サンプルの番号,  $y_i$  はクラスのラベル  $y_i \in \{+1, -1\}$  を表す. 学習サンプル  $x_i$  から得られる HOG 特徴量の共起を観測する関数を  $J_t(x_i)$  と表し, 特徴量  $J_t(x_i) = j$  を観測したとき, 1 段階目の Real AdaBoost の弱識別器  $h_t(x)$  を次式のように表す.

$$h_t(x) = \frac{1}{2} \ln \frac{P_t(y = +1 | j) + \epsilon}{P_t(y = -1 | j) + \epsilon} \quad (7)$$

ここで,  $t$  は学習回数,  $\epsilon$  は分母が 0 の場合に計算不能になることを防ぐための係数である. 本論文では, 予備実験 1 より  $\epsilon = 1/N$  とした. 予備実験 1 の詳細については, 2.4 にて述べる.  $P_t(y = +1 | j)$  及び  $P_t(y = -1 | j)$  は, それぞれ HOG 特徴量の共起を表現した特徴量  $j$  を観測したときの条件付き確率分布であり, 次式より算出する.

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

$$P_t(y = -1 | j) = \sum_{i: J_t(x_i) = j \wedge y_i = -1} D_t(i) \quad (9)$$

$$D_{t+1}(i) = D_t(i) \exp[-y_i h_t(x_i)] \quad (10)$$

$D_t(i)$  は学習サンプル  $x_i$  の重みであり, 式 (10) のように表す. 条件付き確率分布  $P_t(y = +1 | j)$  及び  $P_t(y = -1 | j)$  は, 一次元のヒストグラムにより表現され, 共起を表現した特徴量の値から, 対応する一次元ヒストグラムの要素 (BIN) の番号  $j$  へ学習サンプルの重み  $D_t(i)$  を加算することで作成する. 1 段階目の Real AdaBoost では, 要素の番号  $j$  は共起を表現した特徴量のインデックス番号に対応しているため, 要素数は四つとなる.

次に条件付き確率分布を用いて, 1 段階目の Real AdaBoost における分布の分離度を表す評価値  $z_1$  を次式より求める.

$$z_1 = 1 - 2 \sum_j \sqrt{P_t(y = +1 | j) P_t(y = -1 | j)} \quad (11)$$

この評価値  $z_1$  は, 大きいほどポジティブクラスとネガティブクラスの分布が分離していることを示す. 多くの弱識別器候補から, ラウンドごとに一つの弱識別器を選択する際にこの評価値  $z_1$  が最大となるものを採用する.

最後に, Joint 特徴量である 1 段階目の Real AdaBoost の強識別器  $H^{c_m, c_n}(x)$  を次式より構築する.

$$H^{c_m, c_n}(x) = \sum_{t=1}^T h_t^{c_m, c_n}(x) \quad (12)$$

上記の処理をすべてのセルの組合せにおいて行い, 組合せ数と同じ数の Joint 特徴量を生成する. 例えば, 入力画像が  $30 \times 60$  ピクセル, セルサイズを  $5 \times 5$  ピクセルとした場合, 72 個のセルに分割され, 組合せ数は  ${}_{72}C_2 = 2,556$  となり, 2,556 個の Joint 特徴量  $H^{c_m, c_n}(x)$  を生成する. 生成したすべての Joint 特徴量を一つのプールとし, 後述する 2 段階目の Real AdaBoost の入力として最終識別器を構築する.

#### 2.4 2 段階目の Real AdaBoost による最終識別器の構築

2 段階目の Real AdaBoost では, 1 段階目の Real AdaBoost により生成した Joint 特徴量のプールを入力として最終識別器を構築する (図 3). これにより, 識別に有効な Joint 特徴量を自動的に選択することが可能となる.

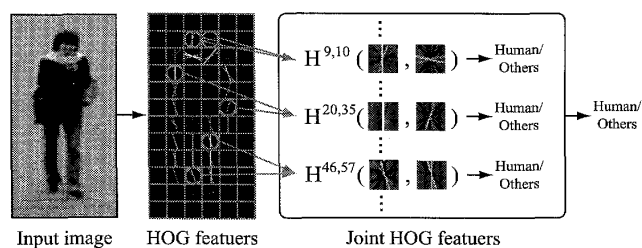


図 3 Joint 特徴量  
Fig. 3 Joint features.

まず, ポジティブクラスとネガティブクラスの確率密度関数  $W_+$ ,  $W_-$  を作成する. 確率密度関数  $W_+$ ,  $W_-$  は, 一次元のヒストグラムにより表現され, 学習サンプルの重み  $D_t(i)$  に基づいて次式より作成する.

$$W_+^k = \sum_{i: k \in K \wedge y_i = +1} D_t(i) \quad (13)$$

$$W_-^k = \sum_{i: k \in K \wedge y_i = -1} D_t(i) \quad (14)$$

ここで,  $t$  は学習回数,  $i$  は学習サンプルの番号,  $k$  は一次元ヒストグラムの要素の番号,  $y_i$  はクラスのラベル  $y \in \{+1, -1\}$  を表す. 学習サンプルの重み  $D_t(i)$  の更新については 1 段階目と同様に式 (10) を用いる. 学習サンプル  $x$  から特徴量を算出し, 特徴量の値に対応する一次元ヒストグラムの要素の番号  $k$  へ学習サンプルの重み  $D_t(i)$  を加算することで, 確率密度関数  $W_+$ ,  $W_-$  を作成することができる. 一次元ヒストグラムの要素数は, 扱う特徴量によって適切な値にする必要がある. そこで, 予備実験 1 として, 提案手法 (Joint(HOG(2))) において, 要素数を変更した場合の人の検出精度の比較を行った結果を図 4 に示す. 実験方法などについては, 3.2 に従って行う. また, この場合の  $\epsilon$  の値は文献 [8] より,  $1/N$  とする. 人検出については, 誤検出率 1.0% のときに対する未検出率で評価を行っている. 誤検出率 1.0% の場合, 要素数が 64 が最も高精度であるため, 一次元ヒストグラムの要素数を 64 とする.

作成した確率密度関数  $W_+$ ,  $W_-$  は, クラスごとの確率密度関数の総和が 1 となるように正規化する. この確率密度関数を用いて, 2 段階目の Real AdaBoost の分布の分離度を表す評価値  $z_2$  を次式より求める.

$$z_2 = 1 - 2 \sum_j \sqrt{W_+^j W_-^j} \quad (15)$$

この評価値  $z_2$  は, 大きいほどポジティブクラスと

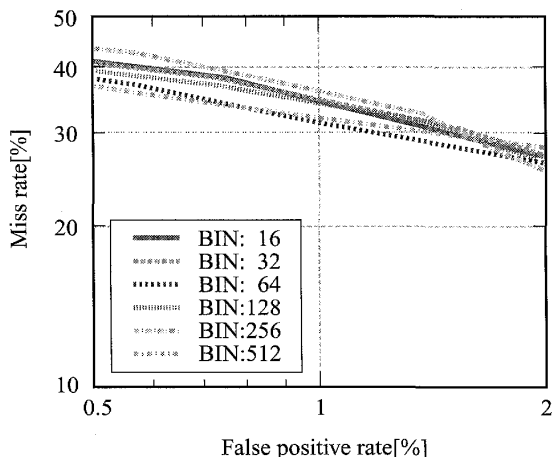


図 4 予備実験 1: 要素数を変更した提案手法による人検出結果

Fig. 4 Exploratory experiment 1: Human detection by the proposed method changed BIN.

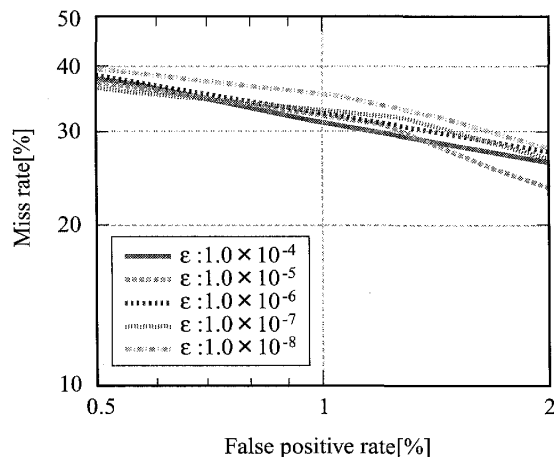


図 5 予備実験 2:  $\epsilon$  を変更した提案手法による人検出結果

Fig. 5 Exploratory experiment 2: Human detection by the proposed method changed  $\epsilon$ .

ネガティブクラスの分布が分離していることを示す。多くの弱識別器候補から、ラウンドごとに一つの弱識別器を選択する際にこの評価値  $z_2$  が最大となるものを採用する。

次に、作成した確率密度関数  $W_+$ ,  $W_-$  を用いて、Real AdaBoost の弱識別器の出力  $g_t(c)$  を算出する。入力された Joint 特徴量の値から、一次元ヒストグラムの要素の番号  $k$  を算出し、それに対応した確率密度関数  $W_+$ ,  $W_-$  から弱識別器の出力  $g_t(c)$  を次式より算出する。

$$g_t(c) = \frac{1}{2} \ln \frac{W_+^k + \epsilon}{W_-^k + \epsilon} \quad (16)$$

$c$  はセルの組合せを表す通し番号、 $\epsilon$  は分母が 0 の

場合に計算不能になることを防ぐための係数である。係数  $\epsilon$  は、学習サンプルの数に応じて適切な値にする必要がある。そこで、予備実験 2 として、提案手法 (Joint(HOG(2))) において、1 段階目と 2 段階目の Real AdaBoost の  $\epsilon$  を変更したときの人の検出精度の比較を行った結果を図 5 に示す。実験方法などについては、3.2 に従って行う。また、このときの要素数は 64 とする。人検出については、誤検出率 1.0% のときに対する未検出率で評価を行っている。図 5 より、誤検出率 1.0% の場合、 $\epsilon$  が  $1.0 \times 10^{-4}$  が最も高精度であることが確認できる。ここで、 $1.0 \times 10^{-4} \doteq 1/N$  であり、また、文献 [8] より、 $\epsilon = 1/N$  とされているため、1 段階目と 2 段階目の Real AdaBoost における  $\epsilon$  を  $1/N$  とする。

最後に、次式より 2 段階目の Real AdaBoost により得られる Joint 特徴量である最終的な強識別器  $G(c)$  を得る。

$$G(c) = \begin{cases} 1 & \sum_{t=1}^T g_t(c) > \lambda \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

$\lambda$  は識別器のしきい値を表す。2 段階目の Real AdaBoost により、Joint 特徴量のプールから識別に有効な特徴量のみを選択することにより、識別器を構築する。

### 3. Joint 特徴量を用いた識別実験

本手法の有効性を示すため、人と車両を検出対象とした評価実験を行う。

#### 3.1 データベース

図 6 に使用したデータベースの一部を示す。

[人用データベース] ポジティブサンプルは、異なる複数の場所で撮影された映像から歩行者を切り出した画像である。ネガティブサンプルは、背景領域をランダムに切り出したものである。学習用データベースとして、ポジティブサンプル 2,054 枚、ネガティブサンプル 6,258 枚、評価用データベースとして、ポジティブサンプル 1,024 枚、ネガティブサンプル 1,234 枚を用いる。

[車両用データベース] ポジティブサンプルは、車載カメラの映像から後方車両領域を切り出した画像である。ネガティブサンプルは、背景領域をランダムに切り出したものである。学習用データベースとして、ポジティブサンプル 2,464 枚、ネガティブサン

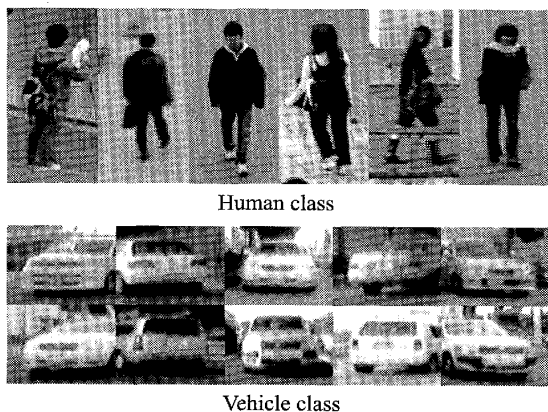


図 6 学習に用いたデータベース  
Fig. 6 Training databaces.

表 1 実験パラメータ  
Table 1 Experiment parameter.

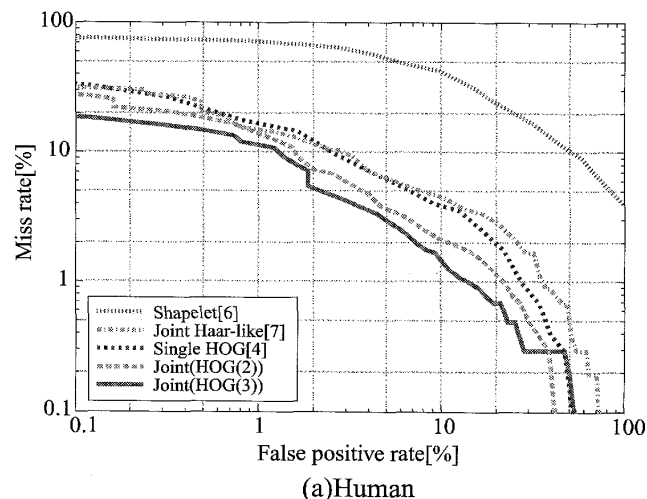
	Human	Vehicle
# of orientations	9(0° ~ 180°)	18(0° ~ 360°)
Patch size (pixels)	30×60	90×72
Cell size (pixels)	5×5	9×9
Block size (cells)	3×3	2×2
# of cells	72	80
# of 2 combinations	2,556	3,160
# of 3 combinations	59,640	82,160

ル 16,158 枚, 評価用データベースとして, ポジティブサンプル 1,900 枚, ネガティブサンプル 5,153 枚を用いる.

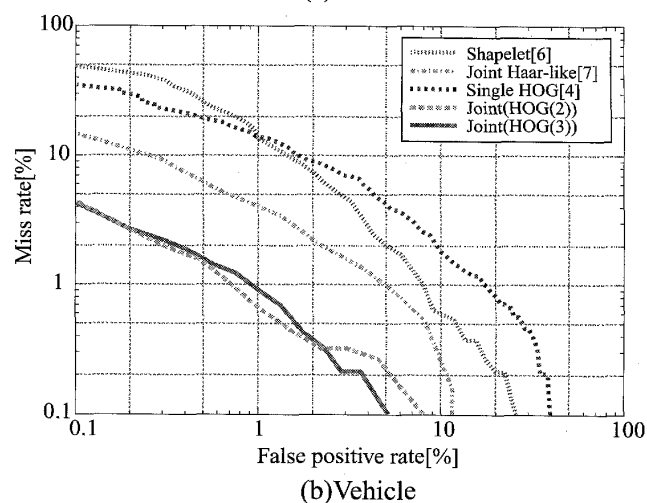
### 3.2 実験概要

評価用データベースを用いて, 人と車両の識別実験を行う. 表 1 に, 人と車両の実験パラメータを示す.

表 1 のこの配方向は, 車両では 0°~360° の 18 方向を用いるが, 人の場合には 0°~180° の 9 方向とする. この変換を行う理由として, 人の場合には衣服と背景の色が逆の関係になる場合があるため, 人の服装に影響しない方向を得るためである. 本手法では組み合わせる特徴量は複数個可能だが, 処理コストが爆発的に増加するため, 今回は 2 個 (Joint(HOG(2))) と 3 個 (Joint(HOG(3))) で実験を行った. 比較対象は, 従来法である HOG 特徴量 (Single HOG) [4], Shapelet 特徴量 (Shapelet) [6], Joint Haar-like 特徴量 (Joint Haar-like) [7] とする. また, 識別精度は, 用いる low-level 特徴量数に依存するため, すべての手法において low-level 特徴量数を 1000 個に統一した. HOG 特徴量は単一の特徴量のため, 弱識別器数が 1000 となる. 提案手法では, 組み合わせが二つのときは, 1 段階目の弱識別器数を 10 (HOG 特徴量 20 個), 2 段階目の弱識



(a) Human



(b) Vehicle

図 7 識別実験の DET カーブ

Fig. 7 DET curves of classification result.

別器数を 50 の合計 1000 個とし, 組み合わせが三つのときは, 1 段階目の弱識別器数を 10 (HOG 特徴量 30 個), 2 段階目の弱識別器数を 34 の合計 1020 個とした. 評価には, Detection Error Tradeoff (DET) カーブを用いる. DET カーブとは, 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る. これにより, 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 カーブは原点に近いほど性能が良いことを示す.

### 3.3 実験結果

図 7(a) に人の識別結果, 図 7(b) に車両の識別結果を示す. 従来法である HOG 特徴量, 比較手法である Joint Haar-like 特徴量や Shapelet 特徴量と比較し, 提案手法は DET カーブが原点により近くなっているため識別精度が向上していることが分かる. 人については, 誤検出率 1.0% の場合, 従来法である HOG 特徴量は未検出率 14.8% であるのに対し, 提案手法で

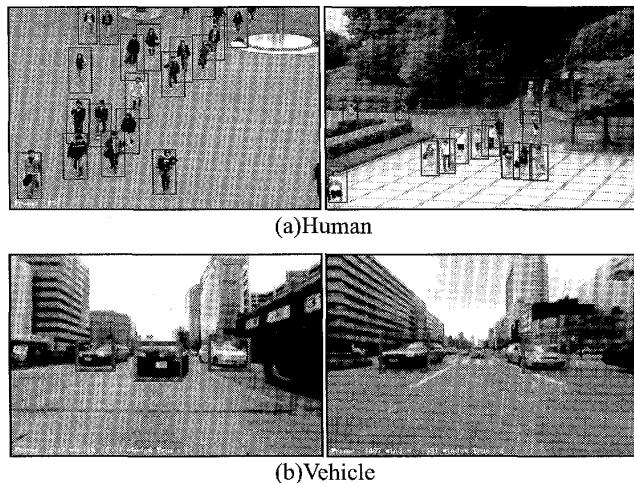


図 8 物体検出の例

Fig. 8 Some examples of object detection.

ある Joint(HOG(3)) 特徴量は未検出率 10.7%となり、4.1%検出精度を向上させることができた。車両については、誤検出率 1.0%の場合、従来法である HOG 特徴量は未検出率 14.2%であるのに対し、提案手法である Joint(HOG(2)) 特徴量は未検出率 0.7%となり、7.2%検出精度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おいても、位置の異なる複数のセル内の HOG 特徴量を組み合わせることにより、新たに識別に有効な特徴量を自動的にとらえることができたためである。

次に、特徴量の数と検出精度について述べる。従来法である HOG 特徴量と HOG 特徴量のみを用いた Joint 特徴量では、同一の low-level 特徴量を 1000 個を用いているが、提案手法がより原点に近く、高精度であることが確認できる。これは、単一の HOG 特徴量を用いるよりも、共起を表現した特徴量が一つの識別器の表現能力が高向上していることが確認できる。

最後に、構築した識別器を用いた物体検出の結果を図 8 に示す。対象の検出を行うために、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。検出対象として検出されたウィンドウは、最後に Mean Shift [9] クラスタリングによるウィンドウの統合処理を行う。人検出、車両検出どちらにおいても、未検出や誤検出がなく、高精度な検出ができていたことが分かる。

### 3.4 考察

人と車両の識別実験において、Real AdaBoost により選択された Joint 特徴量について考察を行う。図 9(a)、図 10(a) に 1 段階目の Real AdaBoost、

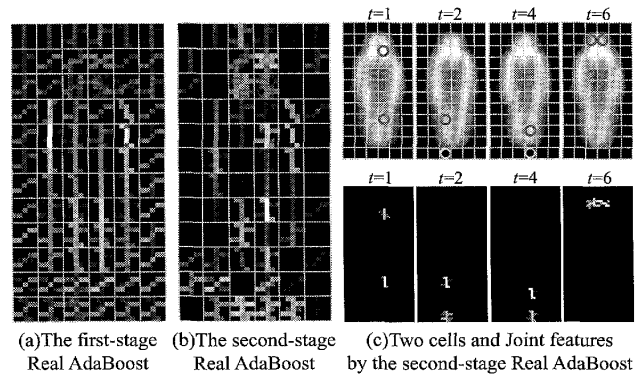


図 9 人検出における選択された Joint 特徴量の可視化  
Fig. 9 Visualization of selected Joint features for human detection.

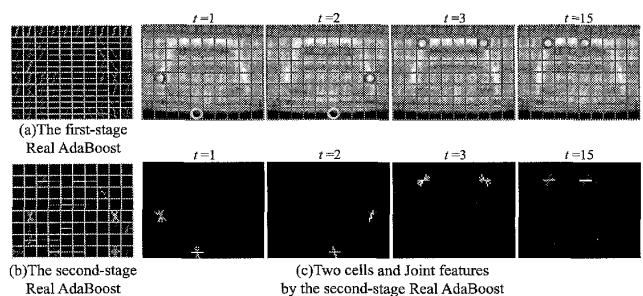


図 10 車両検出における選択された Joint 特徴量の可視化  
Fig. 10 Visualization of selected Joint features for vehicle detection.

図 9(b)、図 10(b) に 2 段階目の Real AdaBoost により、選択された HOG 特徴量を可視化した結果を示す。また、図 9(c)、図 10(c) に、2 段階目の Real AdaBoost により選択された二つのセルと Joint 特徴量をラウンドごとに示す。また、HOG 特徴量のこの配方向を人は 9 方向、車両は 18 方向で表現しており、輝度が高いほど Real AdaBoost における弱識別器の評価値  $z$  が高く、識別に有効な特徴量であることを表す。

検出対象が人の場合、図 9(b) では、図 9(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 Real 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。次に図 9(c) に注目する。2 段階目の Real AdaBoost により選択された Joint 特徴量は、人の輪郭に沿ったセルが選択される傾向があることが分かる。これにより、非剛体である人に対して本手法が有効だと考えられる。

検出対象が車両の場合、図 10(a) では、1 段階目の Real AdaBoost の Joint 特徴量により、車両内部の水平方向のこう配、車両の輪郭部分に沿ったこう配が

多く選択されている。次に図 10 (b) では、図 10 (a) で選択された HOG 特徴量の中から、2 段目の Real AdaBoost により得られる最終識別器により、車両の輪郭部分に沿った Joint 特徴量が選択されていることが確認できる。これにより、車両の輪郭部分に沿った HOG 特徴量が車両と背景を識別するのに有効な特徴量であることが分かる。最後に図 10 (c) に注目する。学習の 1, 2 ラウンド目では垂直こう配と水平こう配の位置関係が選択されている。また 3 ラウンド目では左右対称となる位置関係のセルが選択されている。15 ラウンド目では連続性をとらえた位置関係のセルが選択され、水平方向の特徴量が選択されていることが分かる。提案した Joint 特徴量は、車両形状の対称性や連続性をとらえる特徴量をあらかじめ用意しなくても、学習により自動的に対称性や連続性を表現したセルの位置関係を選択し、物体の識別に有効な特徴セットを得ることができる。

#### 4. HOG 特徴量と PSA 特徴量を用いた Joint 特徴量による歩行者検出実験

2. で述べた HOG 特徴量を組み合わせた Joint 特徴量による物体検出法は、従来法と比較し、高精度な検出が可能であることを示した。更に高精度な検出を行うためには、形状特徴以外の情報を用いることが考えられる。そこで、提案手法と同様のフレームワークを利用し、HOG 特徴量にピクセル状態分析 (Pixel State Analysis : PSA) [10] に基づいた時空間特徴である PSA 特徴量を組み合わせることで、人の見えと動きを同時にとらえることが可能となる。これにより、更なる検出精度の向上が期待できる。本章では、HOG 特徴量と PSA 特徴量を用いた Joint 特徴量による歩行者検出について述べる。

##### 4.1 ピクセル状態分析

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、静状態、動状態の 3 状態に判別する手法である (図 11)。3 状態に判別するためには、以下の二つに着目しピクセルの輝度の自然な変化をとらえる必要がある。

- 輝度値の急激な変化 (Motion Trigger)
- 不安定な状態から安定した状態への推移 (Stability Measure)

まず最初に、急激な輝度変化をとらえるモーショントリガーを考える。ここで、 $I_t$  を現在のフレームから  $k$  フレーム前の輝度値とする。輝度値の変化量  $T$  を求

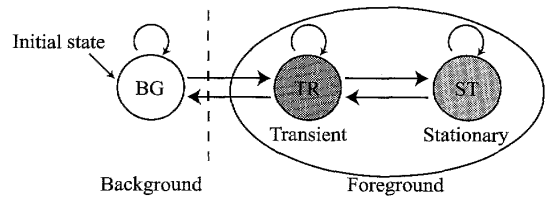


図 11 ピクセルの状態遷移図

Fig. 11 Diagram of state transition for a pixel.

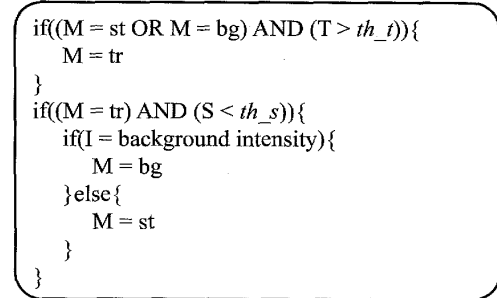


図 12 ピクセル状態分析のアルゴリズム

Fig. 12 Algorithm for pixel state analysis.

めるには、 $t$  よりも前 (過去) のフレームに着目する。変化量  $T$  は次式により求められる

$$T = \max\{|I_t - I_{(t-j)}|, \forall j \in [1, 5]\}. \quad (18)$$

急激な輝度値の変化がピクセル上に生じたとき、 $T$  の値は大きくなる。

次に、ピクセルの安定度を表す  $S$  について考える。安定状態の検出には  $t$  より後 (未来) のフレームに着目する。安定度  $S$  は次式により求められる。

$$S = \frac{k \sum_{j=0}^k I_{(t+j)}^2 \left( \sum_{j=0}^k I_{(t+j)} \right)^2}{k(k-1)}. \quad (19)$$

これはフレーム  $t$  から  $t+k$  までの分散であり、安定した状態のとき  $S$  は小さくなる。

ここで、各ピクセルの状態を表す  $M$  は、背景のとき  $bg$ 、動状態のとき  $tr$ 、静状態のとき  $st$  をとるように、図 12 に示すアルゴリズムにより決定される。 $I$  は、時刻  $t$  における輝度値とする。 $Th_t$  は急激な変化を判定するしきい値、 $Th_s$  は安定性を判定するしきい値である。background intensity は背景画像として、あらかじめ用意しておく。

##### 4.2 時空間特徴の併用

図 13 に、ピクセル状態分析例と状態ヒストグラムの算出法を示す。

ピクセル状態分析の結果から 1 セルに含まれる各状



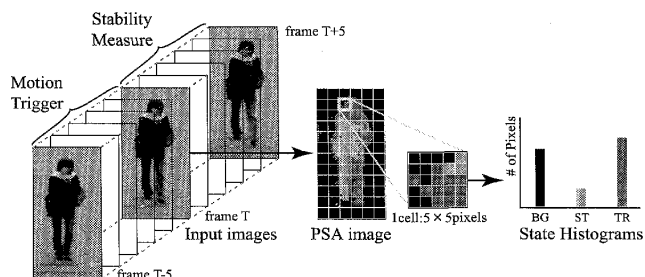


図 13 ピクセル状態分析例と状態ヒストグラム

Fig. 13 An example of pixel state analysis and state histograms.

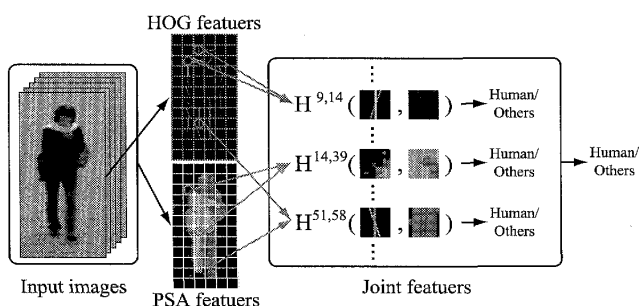


図 14 HOG 特徴と PSA 特徴量を併用した Joint 特徴量

Fig. 14 Human detection by Joint features using HOG features and PSA features.

態のピクセル数を加算することで状態ヒストグラムを作成する。各ピクセルは 3 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。この PSA 特徴量を用いることで、HOG 特徴量とは性質の異なる動きをとらえる特徴量を扱うことが可能となる。例として、図 13 の人は前後 5 フレーム間において右足を軸足とし、左足を前にして歩行している。この一連の動作は HOG 特徴量のようなエッジベースの特徴量からはとらえることはできない。しかし、ピクセル状態分析の結果を見ると右足は静状態、左足が動状態を示しており、1 枚の画像から検出対象がどのような動きを示しているのかをとらえられる。このような人らしい動きを特徴量としてとらえることで、更なる検出精度の向上が期待できる。

HOG 特徴量と PSA 特徴量を併用した Joint 特徴量による最終識別器を図 14 に示す。

#### 4.3 実験概要

3.2 で述べた人のデータベースを用いて評価実験を行う。実験パラメータについては、表 1 と同じものを使用する。実験は、提案手法 (Joint(HOG+PSA)) と、従来法として HOG 特徴量 (Single HOG)、比較手法として Joint 特徴量 (Joint(HOG(2))) との比較を行う。

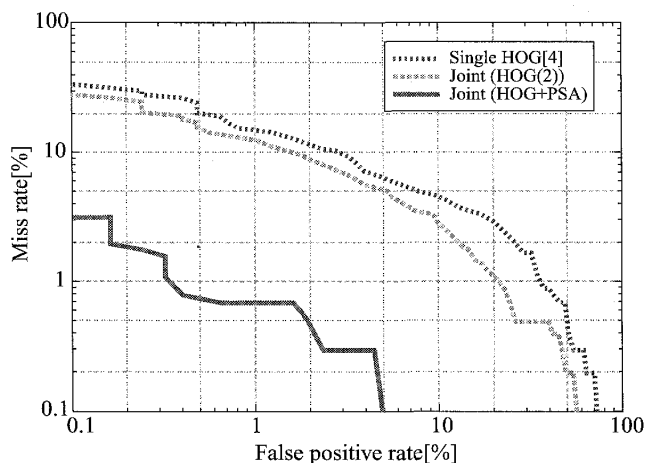


図 15 HOG 特徴量と PSA 特徴量を用いた Joint 特徴量による人識別実験の DET カーブ

Fig. 15 DET curve of classification result for Joint features with HOG features and PSA features.

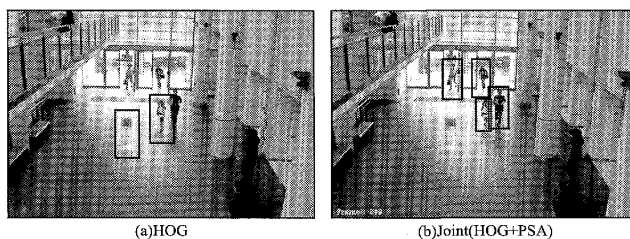


図 16 HOG 特徴量と PSA 特徴量を用いた Joint 特徴量による物体検出の例

Fig. 16 Examples of object detection for Joint features with HOG features and spatiotemporal features.

#### 4.4 実験結果

図 15 に DET による検出結果を示す。誤検出率 1.0% の場合、従来法である HOG 特徴量は未検出率 14.8% であるのに対し、提案手法である Joint(HOG+PSA) は未検出率 0.7% となり、14.1% 検出精度を向上させることができた。

次に、構築した識別器を用いて人検出を行う。検出方法は 3.3 と同様の手法を用いる。比較対照とし、HOG 特徴量による検出結果を図 16 (a)、HOG 特徴量と PSA 特徴量を用いた Joint 特徴量による検出結果を図 16 (b) に示す。

図 16 (a) を見ると画像中央に位置している看板に誤検出がある。しかし、図 16 (b) の結果ではこの誤検出を抑制できている。これは、動きの特徴である PSA 特徴量をとらえることにより、人に似た形状をもつ物体や複雑な背景に対して高精度な検出が可能であることが分かる。

表 2 平均処理時間  
Table 2 Average processing time.

	# of low-level features	time[ms]
single HOG	1000	0.0191
Joint(HOG+HOG)	20(1st stage features)× 50(2nd stage) = 1000	0.0378
Joint(HOG+PSA)	20(1st stage features)× 50(2nd stage) = 1000	0.0607

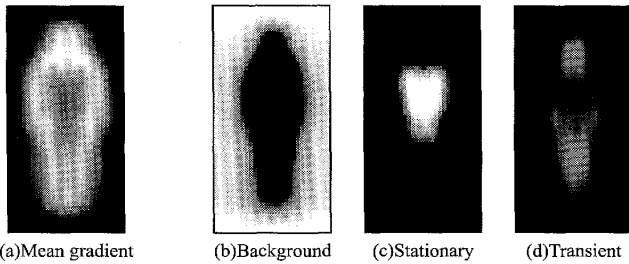


図 17 各状態の出現頻度の可視化画像

Fig. 17 Visualization of occurrence ratio of each states.

次に、処理時間の計測を行った。提案手法と従来法において、一つの検出ウィンドウ (30×60 ピクセル) に対する平均処理時間を表 2 に示す。実験環境は CPU: Intel Xeon x5460 3.16 GHz, メモリ: 24 GByte, OS: Microsoft Windows XP にて計測を行った。平均処理時間の算出は 1000 ウィンドウにおける画像の読み込み、特徴量の算出、識別するまでの処理時間ウィンドウ数で割ることにより算出した。また、識別に用いる low-level 特徴量の数は、1000 個で統一した。提案手法である Joint(HOG+PSA) 特徴量では従来法と比較し、約 3 倍の処理時間を必要とする。しかし、QVGA サイズの入力画像を想定し、検出ウィンドウが 3000 個だった場合、動画像に対して約 5.5 fps でのオンライン処理が可能である。

#### 4.5 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 17(a) に、学習用ポジティブサンプルを用いたときの平均こう配画像、図 17(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、各状態ごとの出現する頻度を画像として可視化したものを示す。図 17 は、輝度が高いほどこう配が強い、若しくは頻度が高いことを表す。図 17(b) の背景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分かる。これは、ピクセル状態分析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。

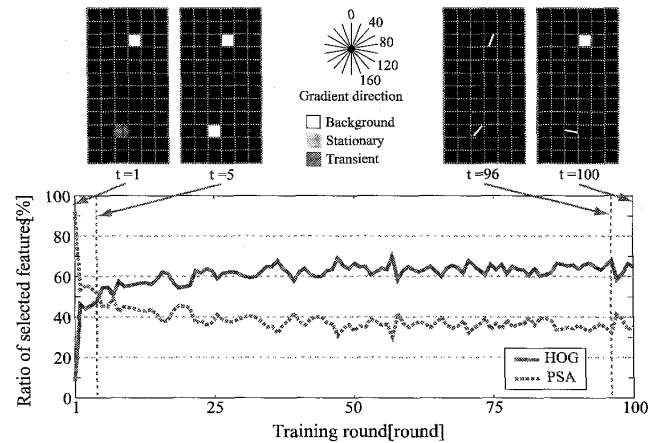


図 18 学習により選択された特徴量と割合  
Fig. 18 Selected Features and the ratio by training.

また、Real AdaBoost では、学習ラウンドの初期段階で選択された特徴量ほど有効な特徴量だということを示す。どのような特徴量が選択されているか傾向を見るために、各学習ラウンドにおける HOG 特徴量と PSA 特徴量の選択された割合とその際に選択された特徴量の可視化の例を図 18 に示す。

学習の初期では、PSA 特徴量が多く選択され、中でもピクセル状態分析の背景から得られる特徴量が用いられている。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、5 ラウンド以降から HOG 特徴量が多く選択されることが分かる。これは、まず物体の動きを表すことが可能な時空間特徴である PSA 特徴量により、大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ランス情報である HOG 特徴量が選択されたと考えられる。

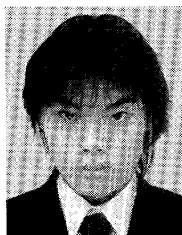
#### 5. む す び

本論文では、物体形状の対称性や連続性を自動的にとらえることが可能である Joint 特徴量による物体検出法を提案した。異なる複数の領域の HOG 特徴量の組み合わせることにより、従来までの単一の HOG 特徴量ではとらえられない領域間の関係性をとらえ、人と車両に対し、有効な特徴量を自動的に得ることが可能となった。Joint(HOG+PSA) 特徴量では、HOG 特徴量から得られる歩行者の形状特徴と、PSA 特徴量から得られる時空間特徴を組み合わせることにより、歩行者特有の特徴をとらえることが可能となり、更なる検出精度の向上を確認した。今後は、シーンコンテキスト情報を用いた車載カメラからの人検出を行う予定

である.

## 文 献

- [1] P. Viola and J. Jones, "Rapid object detection using a boosted cascade of simple features," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.511-518, 2001
- [2] K. Levi and Y. Weiss, "Learning object detection from a small number of examples: The importance of good features," IEEE Comput. Vis. Pattern Recognit., vol.2, pp.53-60, 2004.
- [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] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," IEEE Comput. Vis. Pattern Recognit., pp.886-893, 2005.
- [5] Y. Freund and R.E. Schapier, "Experiments with a new boosting algorithm," Proc. Thirteenth International Conference on Machine Learning, 1996.
- [6] P. Sabzmeydani and G. Mori, "Detecting pedestrians by learning shapelet feature," IEEE Comput. Vis. Pattern Recognit., pp.511-518, 2007.
- [7] T. Mita, T. Kaneko, B. Stenger, and O. Hori, "Discriminative feature co-occurrence selection for object detection," IEEE Trans. Pattern Anal. Mach. Intell., vol.30, no.7, pp.1257-1269, July 2008.
- [8] R.E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," Mach. Learn., no.37, pp.297-336, 1999.
- [9] D. Comaniciu and P. Meer, "Mean shift analysis and applications," IEEE International Conference on Computer Vision, pp.1197-1203, 1999.
- [10] H. Fujiyoshi and T. Kanade, "Layered detection for multiple overlapping objects," IEICE Trans. Inf. & Syst., vol.E87-D, no.12, pp.2821-2827, Dec. 2004.  
(平成 20 年 12 月 10 日受付, 21 年 3 月 31 日再受付)



山内 悠嗣 (学生員)

2007 中部大・工・情報卒. 同年より同修士課程に在籍. 画像処理, パターン認識に関する研究に従事. 情報処理学会会員.



藤吉 弘巨 (正員)

1997 中部大学大学院博士後期退学. 1997~2000 米カーネギーメロン大学ロボット工学研究所 Postdoctoral Fellow. 2000 中部大学講師を経て 2004 より同大准教授. 2006 米カーネギーメロン大学ロボット工学研究所客員研究員. 工博. 計算機視覚, 動画画像処理, パターン認識・理解の研究に従事. 2005 年度ロボカップ研究賞, 2007 年度 SSII ライブデモセッション・オーディエンス賞, 同年 MIRU インタラクティブセッション優秀賞受賞. 情報処理学会, 電気学会, IEEE 各会員.



三井 相和

2007 中部大・工・情報に在籍. 画像処理, パターン認識に関する研究に従事.