

## 特定物体認識に有効な特徴量

山下 隆義†, ††      藤吉 弘亘†

† 中部大学, †† オムロン株式会社

E-mail: takayosi@ari.ncl.omron.co.jp, hf@cs.chubu.ac.jp

**あらまし** 物体認識において、特徴量は画像の中から顔や人などの特定の物体を認識するために重要な要素の1つである。特徴量は学習手法と密接な関係があり、新たな学習手法の提案に伴い、学習手法を有効に活用する新たな特徴量が提案されている。このような学習手法の進化に合わせた特徴量の進化により、物体認識の性能が向上しており、顔検出などの特定物体認識が実用化されている。本稿では、学習手法の進化の側面から特徴量を3つの世代に分類し、各世代の特徴量について、捉え方や算出手法などの特長を述べる。また、複数の検出対象を例に、特徴量による性能を比較し、どのような特徴量が適しているかを比較実験した結果を紹介する。

## Efficient Features for Object Recognition

Takayoshi Yamashita†, ††      Hironobu Fujiyoshi†

† Chubu University, †† OMRON Corporation

E-mail: takayosi@ari.ncl.omron.co.jp, hf@cs.chubu.ac.jp

**Abstract** Feature extraction is one of the most important components for detecting objects such as face and human. It has a strong relationship with learning algorithms such as Boosting, NN and SVM. State-of-art learning methods have been proposed, and new feature extraction methods have also been proposed to improve the performance of object detection. In this paper, we describe the characteristics of these feature extraction methods, which are classified into three generations by the aspect of learning method's evolution. We present the results of a performance comparison using the feature extraction methods.

### 1 はじめに

物体認識はコンピュータビジョンの分野で最も重要なトピックスの1つであり、古くから研究が行われている。物体認識は検出する対象をあらかじめ定義している特定物体認識と、検出対象を定義せず不特定多数の物体を対象とする一般物体認識がある。特定物体認識の代表的な対象は顔であり、近年では、顔検出がデジタルカメラや携帯電話などへ搭載され、物体認識の実用化が進んでいる。物体認識を実現するための重要な要素は2つあり、物体をモデリングするための特徴量と、その特徴量を選択するための

学習手法である。これらには密接な関係があり、古くから研究されている中で、新たな学習手法が提案されると、それに伴って性能を向上させる特徴量が提案されている。このように、学習手法の進化に伴い、特徴量も進化している。筆者等は統計的学習手法の進化に合わせて、図1のように特徴量を3つの世代に分類している。第一世代の特徴量は輝度分布やwaveletのように、対象物体全体の情報を特徴量として用いる。第二世代は、検出対象の局所的な領域の情報に着目する特徴量である。第二世代の特徴量は、ブースティングと密接な関係があり、ブースティングを組み合わせることで、性能を大幅に向上

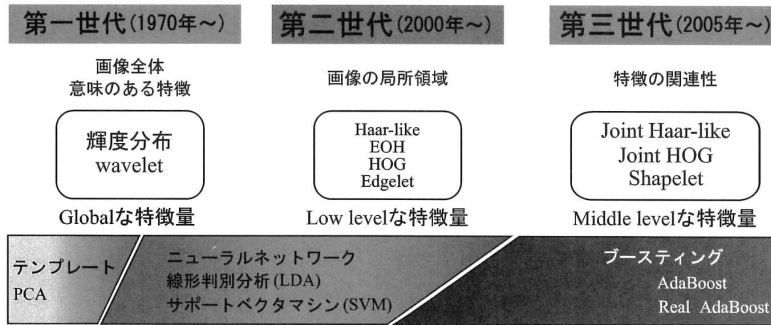


図 1: 特徴量の進化

させる手法が多数提案されている [1][2]. 第三世代は、第二世代のように局所領域の情報を用いるが、複数の局所領域の関連性に着目した特徴量である. 次章以降で、各世代の特徴量の詳細と各特徴量を用いた評価実験について述べる.

## 2 第一世代の特徴量 (画像全体に着目した特徴量)

第一世代の特徴量は、対象物全体を特徴量として捉える. 例えば、対象物全体の輝度や色の分布をヒストグラムとし、多次元ベクトルで表現する. また、Wavelet 特徴量も対象物全体の空間周波数に着目しており、第一世代の特徴量であるといえる.

第一世代の特徴量を用いた学習手法は、二つに分けることができる. 1 つ目は、研究者の経験に基づいた知識ベースの学習手法であり、初期の物体検出の手法に多く用いられている. 研究者の知識ベースの手法として、顔のモデルを作成するために、顔のエッジのある領域とない領域にわけ、このモデルを用いて顔検出を行う手法 [3] や複数の顔テンプレートを用いた顔検出法 [4]、顔の器官の関係に基づいた顔検出法 [5] などが挙げられる. これら知識ベースの手法は、複雑なモデルを作ることが難しいため、比較的少ない特徴量を用いた手法が多い.

2 つ目は、ニューラルネットワークや線形判別分析法 (LDA)、サポートベクターマシン (SVM) など統計的学習手法を用いた学習手法である. これらは、知識ベースの手法よりも大量の学習サンプルを用いることで高精度な検出を行うことができる. また、知識ベースの学習手法では、研究者の知識に基づいてどこに着目するかを定義していたが、統計的学習

手法を用いた手法では、検出対象に共通する情報に着目した特徴量が自動的に定義される. これらの統計的学習を用いた代表的な手法として、Rowley らのニューラルネットワークを用いた顔検出法 [6] や Papageorgiou らの人体検出法 [7] がある.

## 3 第二世代の特徴量 (局所領域に着目した特徴量)

第二世代の特徴量は対象物体の局所領域を捉える. 特徴量は輝度情報をベースにしたものと、エッジ情報をベースにしたものに分類することができる. 輝度に着目する特徴量としては、Haar-like[1] 特徴量やピクセル差分特徴量 [8]、エッジに着目する特徴量としては HOG(Histograms of Oriented Gradients) 特徴量 [2] や EOH(Edge of Orientation Histograms) 特徴量 [9]、Edgelet 特徴量 [10] がある. これらの特徴量は多数の局所領域の特徴量を Low level な特徴量として組み合わせて対象物体を形成する.

第二世代の特徴量はブースティング [11] と密接な関係がある. ブースティングは与えられた学習サンプルを検出するのに適した特徴量を選択することができる. 例えば、顔をサンプルとして与えた場合、位置や大きさを変えた特徴量候補の中から顔に共通する特徴量、目の付近に着目したもの、口の周りに着目したものなどを自動的に選択することができる. すなわち、研究者の事前知識なく、自動的に最適な特徴量を選択することができる. これらの特徴量について、詳しく説明する.

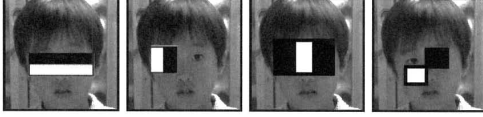


図 2: Haar-like 特徴量

### 3.1 輝度に着目した特徴量

#### 3.1.1 Haar-like 特徴量

Haar-like 特徴量は、図 2 に示すように、白の領域  $r_1$  と黒の領域  $r_2$  の 2 つの領域の輝度差を式 (1) より算出し Haar-like 特徴量としている。

$$H(r_1, r_2) = S(r_1) - S(r_2) \quad (1)$$

ここで、 $S(r)$  は、領域  $r$  の輝度和を算出する関数である。2 つの領域のパターンは多数あり、図 2 に示すようなパターンが一般的に用いられる。これらのパターンは、縦方向の明暗差、横方向の明暗差、斜め方向の明暗差を捉えることが可能である。他にも、3 つの領域に着目することで線分を捉えるパターンもある。これらのパターンをベースとして、様々な形状が提案されている [12]。Haar-like 特徴量は Viola らが顔検出 [1] に用いて以降、非常に多くの物体認識手法に用いられている [13][14][15][17]。また、Haar-like 特徴量は、積分画像を用いることで高速に特徴量を算出することが可能である。高速化の詳細については 5 章で述べる。

#### 3.1.2 ピクセル差分特徴量

ピクセル差分特徴量は任意の画素間の輝度差に着目した特徴量である [8]。前述の Haar-like 特徴量は隣接する局所領域間での輝度差に着目している。一方、ピクセル差分特徴量は、図 3 に示すように、離れた任意の画素間での輝度差を捉えることができる。ピクセル差分特徴量  $P(I)$  は、式 (2) のように、2 つの画素間  $I_1, I_2$  での輝度差を特徴量として捉えている。

$$P(I) = I_1 - I_2 \quad (2)$$

ここで、 $I$  は入力画像の輝度値を表す。

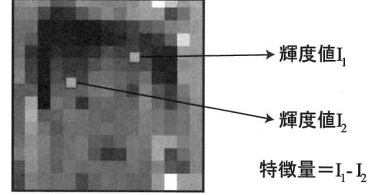


図 3: ピクセル差分特徴量

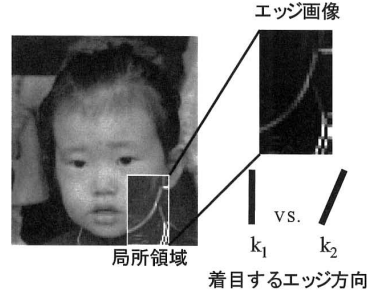


図 4: EOH 特徴量

### 3.2 エッジに着目した特徴量

#### 3.2.1 EOH 特徴量

EOH 特徴量は、1 つの局所領域内におけるエッジ勾配の関係に着目した特徴量である [9]。EOH 特徴量の算出には、図 4 の頬の領域のようにソーベルフィルタによりエッジ画像を作成し、式 (3)(4) のように各画素のエッジ強度  $G(x, y)$  とエッジ方向  $\theta(x, y)$  を算出する。

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (3)$$

$G_x(x, y)$  は、 $x$  方向のエッジ画像、 $G_y(x, y)$  は  $y$  方向のエッジ画像である。

$$\theta(x, y) = \arctan\left(\frac{G_x(x, y)}{G_y(x, y)}\right) \quad (4)$$

式 (5) より、エッジ方向  $\theta$  を  $k$  段階に量子化し、各方向のエッジ強度画像を作成する。

$$\psi_k(x, y) = \begin{cases} G(x, y) & \text{if } \theta(x, y) \in \text{bin}_k \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

領域  $R$  における方向  $k$  の累積エッジ強度は式 (6) のように算出する。

$$E_k(R) = \sum_{(x, y) \in R} \psi_k(x, y) \quad (6)$$

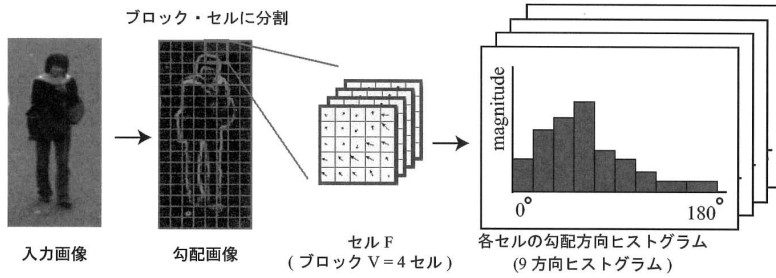


図 5: HOG 特徴量

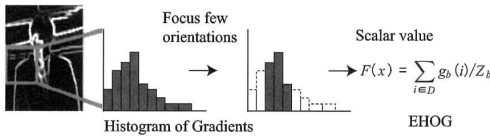


図 6: EHOH 特徴量

EOH 特徴量は、式 (7) のように方向  $k_1$  に対する累積エッジ強度と方向  $k_2$  に対する累積エッジ強度の累計の比を算出する。

$$A_{k_1, k_2}(R) = \frac{E_{k_1}(R) + \epsilon}{E_{k_2}(R) + \epsilon} \quad (7)$$

ここで、 $E_k(R)$  は、 $R$  領域における  $k$  方向の勾配を求める関数、 $\epsilon$  は 0 による除算を防ぐための係数である。エッジ方向は 4 から 8 方向に量子化している。

### 3.2.2 HOG 特徴量

HOG 特徴量は 1 つの局所領域内におけるエッジ方向ごとのエッジ強度に着目した特徴量である。局所領域は図 5 に示すように、ブロックとして表現され、ブロックには複数のセルと呼ばれるさらに小さな局所領域から構成される。特徴量は、セルのエッジ方向毎のエッジ強度ヒストグラムを結合した多次元ベクトルで表現される。Dalal らはセルサイズを  $8 \times 8$  画素、エッジ方向を 9 方向とし、ブロックは  $2 \times 2$  セルで構成される 36 次元ベクトルとしている。HOG 特徴量はブロックに対して、L1 ノルムまたは L2 ノルムにより、ブロックを正規化している。これにより、全体的な明暗条件によるエッジ強度がばらつく場合でも特徴量が大きく変化しないようにしている。

HOG 特徴量は、複数のセル内に含まれる勾配方向ごとの情報を特徴量とする高次元特徴量である。

Zhu ら [18] は SVM, LDA を用いて次元圧縮する方法を提案しているが、計算コストが大きいなどの問題がある。計算コストを削減する方法として、HOG 特徴量を改良した EHOH (Extend Histograms of Oriented Gradients) 特徴量が提案されている [19]。EHOH は図 6 のように、1 つのセル内における 1 つのエッジ方向のみに着目する特徴量である。そのため、セルから得られる特徴量が 1 次元となり、HOG 特徴量と比べ高速に高精度な検出を行うことが可能となる。特徴量  $F_b(D)$  は、式 (9) のように算出される。

$$F_b(D) = \sum_{i \in D} g_b(i) / Z_b \quad (8)$$

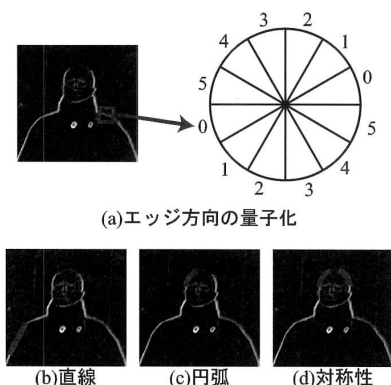
$$Z_b = \sum_{i=0}^n g_b(i) \quad (9)$$

このように、エッジ方向  $D$  の内、着目する方向  $b$  のエッジ強度  $g_b(i)$  を累積し、全方向のエッジ強度  $Z_b$  で正規化する。

### 3.2.3 Edgelet 特徴量

EOH 特徴量や HOG 特徴量は局所領域内におけるエッジの分布に着目しているが、Edgelet 特徴量は図 7 のように、エッジの部分的な繋がり捉える特徴量である。特徴量の形状は、直線的な繋がりや、円弧およびそれらの対称性を捉えるものである。

Edgelet 特徴量は、 $K$  個の画素から構成されており、各画素は位置  $u_i$  とエッジ方向  $V_i^E$  からなる。エッジ方向  $V_i^E$  は図 7 (a) のように 6 方向に量子化された値である。入力画像  $I$  の注目領域  $p$  における Edgelet 特徴量  $S(p)$  は、式 (10) に示すように、 $i$  番目の要素のエッジの強度  $E^I(u_i + p)$  と近似類似度関数  $I[\cdot]$  の内積を  $K$  個の要素位置について、累積



(a)エッジ方向の量子化



(b)直線 (c)円弧 (d)対称性

図 7: Edgelet 特徴量

し正規化した値である。

$$S(p) = \left(\frac{1}{K}\right) \sum_{i=1}^K E^I(u_i + p) l[V^I(u_i + p) - V_i^E] \quad (10)$$

ここで、 $u_i + p$  は注目領域  $p$  における Edgelet 特徴量の  $i$  番目の要素位置である。エッジ強度  $E^I$  はソーベルフィルタを用いて作成したエッジ画像  $E$  の画素値から得ることができる。また、近似類似度関数  $l[\cdot]$  は式 (11) に示すように、入力画像  $I$  の注目領域  $p$  における  $i$  番目の要素のエッジ方向  $V^I(u_i + p)$  と Edgelet 特徴量の  $i$  番目の要素のエッジ方向  $V_i^E$  との差をもとに算出される。

$$l[\cdot] = \begin{cases} 1 & x = 0 \\ 0.8 & x = \pm 1, \pm 5 \\ 0.5 & x = \pm 2, \pm 4 \\ 0 & x = \pm 3 \end{cases} \quad (11)$$

ここで、 $V^I(u_i + p)$  は  $V_i^E$  と同様に、6 段階に量子化しており、 $x$  はエッジ方向  $V^I(u_i + p)$  と  $V_i^E$  の差である。 $x$  は  $-5$  から  $5$  までの範囲の値であり、それぞれの値に応じて、式 (11) のような値をとる。

### 3.3 第二世代の特徴量とブースティング

第二世代の特徴量は、ブースティングがコンピュータビジョンで用いられるようになったことで提案された特徴量であり、両者は密接な関係がある。ブースティングは式 (12) に示すように、 $T$  個の弱識別器  $h_t(x)$  を重みつき線形結合した強識別器  $H(x)$  と

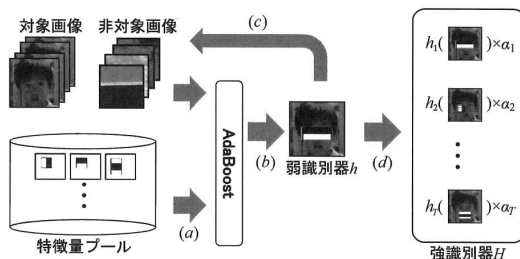


図 8: ブースティングによる学習の流れ

して学習する。

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

弱識別器  $h_t(x)$  は対象物体であるかどうかを出力する。出力値はブースティングの手法により異なるが、AdaBoost[11] は 0 または 1 の 2 値、Real AdaBoost[16] は実数で出力する。第二世代の特徴量を用いて、ブースティングにより学習する流れを図 8 に示す。

検出対象と非検出対象の学習サンプルを多数用意し、各サンプルに重みを設定する。初期時はすべて同じ重みとする。また、特徴量の位置や形状を変えた特徴量候補を作成し、特徴量プールとして用意する (a)。次に、特徴量プールから学習サンプルを最も正しく識別できる特徴量候補を弱識別器として選択する (b)。選択された弱識別器により、サンプル画像の重みを更新する (c)。正しく識別できたサンプルの重みは小さくし、誤って識別したサンプルには大きな重みとする。これにより、次の選択時に正しく識別できなかったサンプルについて、正しく識別することが可能な弱識別器が選択されるようになる。これらの処理を繰り返し行い、強識別器を学習する (d)。詳細な処理については、[11] や [1] を参考にしたい。

このように、ブースティングを用いることで研究者の知識を必要とせず、多数の特徴量候補の中から対象物体の検出に有効な特徴量を学習することができる。図 9 に顔検出において、ブースティングによって選択された Haar-like 特徴量の例を示す。学習回数 1 ラウンド目では額の明暗差を捉え、2 ラウンド目では目の明暗差、3 ラウンド目では鼻柱の明暗差、4 ラウンド目では頬の明暗差を捉えていることがわかる。このようにブースティングを用いることで、顔と非顔を判別するために有効な Haar-like



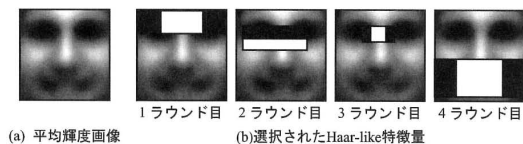


図 9: 顔画像における平均輝度画像と選択された Haar-like 特徴量

特徴量を自動的に選択している。

#### 4 第三世代の特徴量 (関連性に着目した特徴量)

前章で述べた第二世代は局所領域を捉える特徴量であり、各局所領域を独立として考えている。第三世代の特徴量は局所領域間の関連性や共起性を捉える特徴量である。代表的なものとして、第二世代の特徴量で輝度に着目した Haar-like 特徴量に共起性を考慮した Joint Haar-like 特徴量 [15] や Haar-like 特徴量とピクセル差分特徴量の特徴を持ち合わせたスパース特徴量 [17]、複数の局所領域におけるエッジの関連性に着目した Shapelet 特徴量 [20] がある。また、エッジに着目した HOG 特徴量の関連性に考慮した Joint HOG 特徴量 [21]、共起する確率を演算子により結合する共起確率特徴 [22] なども提案されている。

これらは Low level な特徴量である第二世代の特徴量をベースとし、それらの関連性に着目することにより、1 つ 1 つでは意味のない Low level な特徴量に対して、関連性を考慮することで何らかの意味をもつ Middle level な特徴量に拡張している。

第三世代の特徴量も第二世代と同様にブースティングと密接な関係がある。ブースティングにより、複数の特徴量を選択することで検出器を学習するが、関連性を考慮することで、第二世代の特徴量よりも少ない特徴量数で、検出器を学習することが可能である。

次に、代表的な第三世代の特徴量について説明する。

##### 4.1 Joint Haar-like 特徴量

Joint Haar-like 特徴量は、複数の Haar-like 特徴量の関係性に着目した特徴量である。例えば、図 10

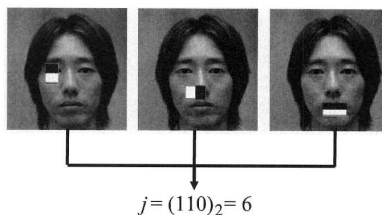


図 10: Joint Haar-like 特徴量

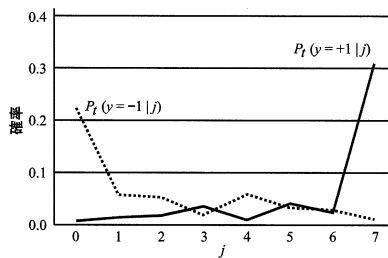


図 11: 同時確率

に示すように、複数の Haar-like 特徴を組み合わせ、検出対象物体の構造に基づいた特徴の共起関係表現することが可能である。そのため、単一の特徴だけでは捉えることができない関係性を捉えることができるため、検出性能を大きく向上させることができる。

特徴の共起を表現するためには、同時確率を用いる。同時確率を求めるために、1 つ 1 つの Haar-like 特徴量が検出対象物か否かの 2 値化を行う。検出対象クラスと非検出対象クラスに対応する 1 と 0 の符号  $s$  により特徴量を表現する。符号  $s$  は、サンプル  $x$  から、式 (13) により算出する。

$$s(x) = \begin{cases} 1 & p \cdot z(x) > p \cdot \theta \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

ここで、 $z$  は、1 つの Haar-like 特徴から算出された特徴量であり、 $\theta$  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。 $\theta$  及び  $p$  は、Boosting により、学習サンプルに対する識別エラーが最小となるように自動的に決定される。

次に、この 2 値化符号を多数組み合わせ、Joint Haar-like 特徴を生成する。図 10 は、3 つの Haar-like 特徴を組み合わせ、Joint Haar-like 特徴  $j$  を算出した例である。

最後に、AdaBoost により、共起を表現した Joint Haar-like 特徴を用いて学習を行う。AdaBoost の弱

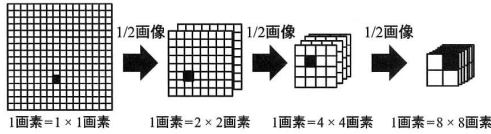


図 12: スパース特徴量

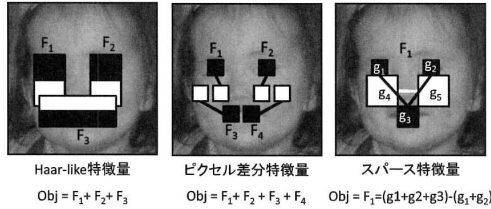


図 13: 特徴量の表現力の比較

識別器には、複数の Haar-like 特徴の同時確率を表現することが可能な Joint Haar-like 特徴の確率を用いる。弱識別器  $h_t(x)$  において、入力画像  $x$  から Joint Haar-like 特徴を観測する処理を関数  $J_t(x)$  と表わす。  $x$  から特徴量  $J_t(x) = j$  を観測したとき、  $h_t(x)$  は式 (14) のようになる。

$$h_t(x) = \begin{cases} +1 & P_t(y = +1 | j) > P_t(y = -1 | j) \\ -1 & \text{otherwise} \end{cases} \quad (14)$$

ここで、  $y$  は正解ラベル  $y \in \{+1, -1\}$  である。  $P_t(y = +1 | j)$  及び  $P_t(y = -1 | j)$  は、それぞれ Haar-like 特徴量の共起を表現した特徴  $j$  を観測したときの確率であり、学習サンプル  $i$  の重み  $D_t(i)$  に基づき式 (15)(16) より算出される。

$$P_t(y = +1 | j) = \sum_{p: J_t(x_p) = j \wedge y_i = +1} D_t(i) \quad (15)$$

$$P_t(y = -1 | j) = \sum_{p: J_t(x_p) = j \wedge y_i = -1} D_t(i) \quad (16)$$

図 11 に、3 つの Haar-like を組み合わせた際の  $P_t(y = +1 | j)$  と  $P_t(y = -1 | j)$  の分布を示す。

## 4.2 スパース特徴量

第二世代のピクセル差分特徴量は、画素単位で輝度差を算出している。一方、スパース特徴量は Granular Spase という局所領域を 1 つの画素に射影した空間で特徴量を算出することができる Haar-like 特徴量とピクセル差分特徴量を組合わせたような特徴量である。

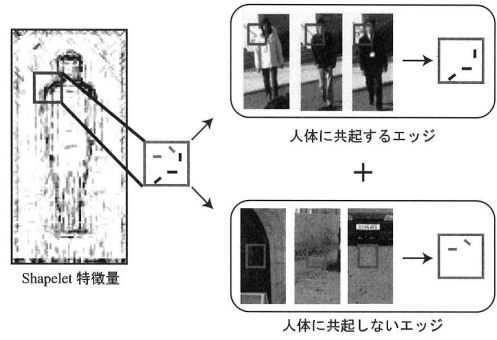


図 14: Shapelet 特徴量

Granular Spase は図 12 に示すように、複数の解像度の縮小画像を作成する。スパース特徴量は複数の局所領域での明暗差を特徴量とし、局所領域はピクセル差分特徴量に相当する  $1 \times 1$  画素と、  $2 \times 2$  画素、  $4 \times 4$  画素、  $8 \times 8$  画素から構成される。これらの局所領域から特徴量は式 (17) のように算出することができる。

$$F(\pi) = \sum_i \alpha_i p_i(\pi; x, y, s), \alpha \in \{+1, -1\} \quad (17)$$

ここで、  $x$  および  $y$  は局所領域の位置、  $s$  は局所領域の大きさであり  $\pi$  は局所領域の輝度である。また、  $\alpha$  は加減算の符号であり、特徴量  $F(\pi)$  は  $i$  個の局所領域における輝度の加減算から算出される。

ピクセル差分特徴量は 2 つの画素の明暗差を特徴量としている。一方、スパース特徴量は複数の Granular の明暗差を特徴量としている。これにより、複数の局所領域の輝度情報から総合的に判別が可能な特徴量である。すなわち図 13 に示すように、目と口は暗く、両ほほは明るいというような特徴を表現する場合、Joint Haar-like 特徴量は 3 つ、ピクセル差分特徴量は 4 つの特徴量が必要であるが、スパース特徴量は 1 つの特徴量で表現することができる。

## 4.3 Shapelet 特徴量

Shapelet 特徴量は、図 14 に示すように、各画素のエッジ情報を Low level な特徴量とし、複数画素のエッジ情報の関連性に着目した Middle level な特徴量である。これにより、局所領域内で人に共起するエッジと共起しないエッジを同時に捉えることができる。

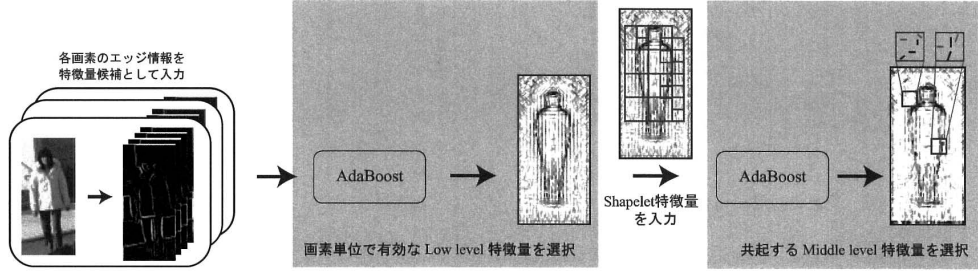


図 15: Shapelet 特徴量の生成と検出器構築の流れ

図 15 に, Shapelet 特徴量をブースティングにより選択する処理について示す. まず, 入力画像から 4 方向のエッジ画像を求める. 各画素の値を Low level な特徴量  $g(\pi; x, y, d)$  とする. ここで,  $x$  および  $y$  は位置,  $d$  はエッジ方向である. Shapelet 特徴量を作成するために, これらの Low level な特徴量を局所領域  $w_i \in W$  に分割する. 局所領域  $w_i$  ごとに, Shapelet 特徴量を学習するために Low level な特徴量を AdaBoost により選択する. Low level な特徴量は弱識別器として, 式 (18) のように算出される.

$$g_t(x) = \begin{cases} 1 & p_t \cdot f_t(x) > p \cdot \theta \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

ここで,  $f_t(x)$  はエッジ強度,  $\theta$  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.  $\theta$  及び  $p$  は, AdaBoost での学習課程において, 学習サンプルに対する識別エラーが最小となるように決定される. Low level 特徴量を AdaBoost により複数選択することで, 局所領域  $w_i$  における Shapelet 特徴量は式 (19) のように算出される.

$$H_i(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t^i g_t^i(x) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

ここで,  $\alpha$  は弱識別器の重み,  $i$  は局所領域の番号を表わす. このように, 局所領域内  $w_i$  で検出対象とそれ以外を判別するのに有効な Low level 特徴量  $g(x)$  を複数含む Shapelet 特徴量を学習する. Sabzmeydani らは Shapelet 特徴量に含まれる Low level 特徴量の個数  $T$  を, 局所領域  $w_i$  のサイズにより 10 から 30 個と定義している.

局所領域ごとに構築された識別器は, 図 15 に表わすように, 検出対象の物体の形状を捉えるような特徴量として学習されている. 局所領域ごとに算出した Shapelet 特徴量を 2 回目の AdaBoost を用い

て最終識別器  $C(s)$  を式 (20) により構築する.

$$C(s) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(s) > \lambda \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

ここで,  $C(s)$  は 2 回目の AdaBoost の弱識別器を表わし,  $T$  個の Shapelet 特徴量  $h_t$  から構成される. 2 回の AdaBoost を通して, 検出対象を判別するのに有効な Low level な特徴量が Middle level な特徴量として選択されている.

#### 4.4 Joint HOG 特徴量

Joint Haar-like 特徴量は局所領域間の共起性, Shapelet 特徴量は局所領域内のエッジ勾配の共起性を捉える特徴量である. Joint HOG 特徴量はこれら 2 つの特徴量の特長を組み合わせた特徴量である.

図 16 に Joint HOG 特徴量を用いた学習の流れを示す. Joint HOG 特徴量は, 2 つの局所領域内の HOG 特徴量を組み合わせることで, 局所領域間の共起性とエッジ勾配の共起性を捉える. Joint HOG 特徴量を生成するために, Joint Haar-like の共起表現法と同様に, 複数の HOG 特徴量の共起を表現する. まず, 式 (21) から検出対象物体と非検出対象物体を表す 2 値化符号  $s$  を算出する.

$$s(\mathbf{V}) = \begin{cases} 1 & p \cdot v_o > p \cdot \theta \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

ここで,  $\mathbf{V}$  は 1 つの局所領域から得られる HOG 特徴量,  $o$  は HOG 特徴量の勾配の方向,  $\theta$  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. 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を表現した特徴量  $j$  を得る.

次に, 位置の異なる 2 つの局所領域であるセル  $cm, cn$  において, 共起を表現した特徴量の中から,



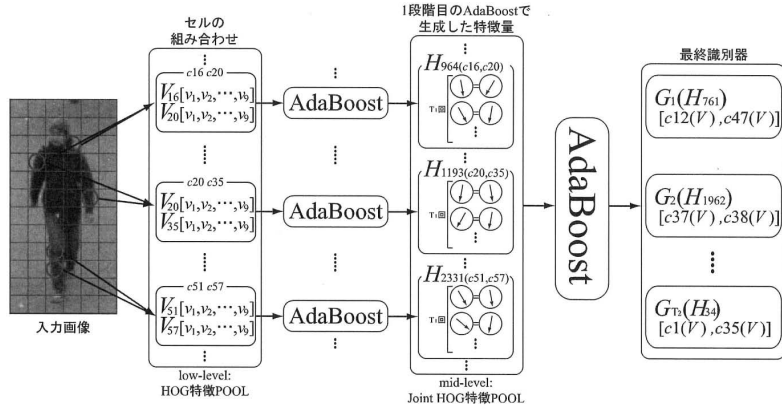


図 16: Joint HOG 特徴量

識別に有効な特徴量を選択する。入力画像  $x$  から HOG 特徴量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  $h_t(x)$  を式 (22) により表す。

$$h_t(x) = \begin{cases} +1 = P_t(y = +1|j) > P_t(y = -1|j) \\ -1 = \text{otherwise} \end{cases} \quad (22)$$

ここで、 $y$  は正解ラベル  $y \in \{+1, -1\}$  を表す。 $P_t(y = +1 | j)$  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  $j$  を観測したときの確率であり、学習サンプル  $i$  の重み  $D_t(i)$  に基づき式 (23)(24) により算出される。

$$P_t(y = +1|j) = \sum_{p: J_t(x_p)=j \wedge y_i=+1} D_t(i) \quad (23)$$

$$P_t(y = -1|j) = \sum_{p: J_t(x_p)=j \wedge y_i=-1} D_t(i) \quad (24)$$

1 回目の強識別器である Joint HOG 特徴量  $H_{cm,cn}(x)$  を式 (25) より構築する。

$$H_{cm,cn}(x) = \begin{cases} 1 \sum_{t=1}^T \alpha_t^{cm,cn} \cdot h_t^{cm,cn}(x) > 0 \\ 0 \text{ otherwise} \end{cases} \quad (25)$$

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

上記の処理を全ての局所領域の組み合わせにおいて行い、Joint HOG 特徴量 POOL を作成する。最後に、2 回目の AdaBoost により、Joint HOG 特徴量 POOL から、Joint HOG 特徴量を 2 段階目の

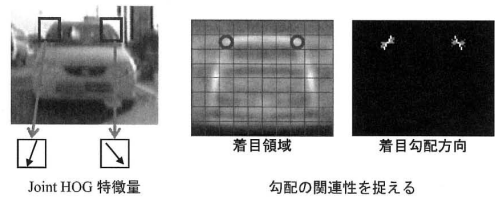


図 17: Joint HOG 特徴量による特徴量の捉え方

AdaBoost の入力とし、式 (26) より最終的な強識別器  $G(c)$  を得る。

$$G(c) = \begin{cases} 1 \sum_{t=1}^T \alpha_t \cdot g_t(c) > \lambda \\ 0 \text{ otherwise} \end{cases} \quad (26)$$

ここで  $\alpha$  は、1 段階目の AdaBoost より得られる強識別器  $g_t(c)$  の重み、 $\lambda$  は検出器の閾値、 $c$  はセルの組み合わせを表す。2 段階目の AdaBoost により、Joint HOG 特徴量 POOL から特徴量を選択して、検出器を学習する。これにより、検出対象を判別するために有効な Joint HOG 特徴量を選択される。

図 17 に、Joint HOG 特徴量による特徴量の捉え方を示す。このように、Joint Haar-like 特徴量では、局所的な領域内の勾配情報を用いることで物体の形状を捉えることができる。

#### 4.5 共起確率特徴量

共起確率特徴量は、複数の第二世代の特徴量が出力する物体らしさをもとに、共起確率を算出することができる特徴量である。すなわち、Joint Haar-like

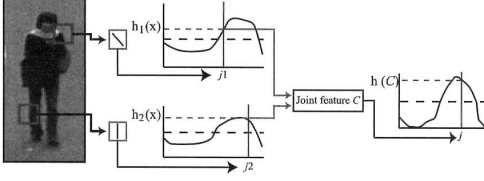


図 18: 共起確率特徴量

特徴量の 0 または 1 の 2 値による値の共起表現を、図 18 に、示すように実数で得られる確率を演算子により結合する共起表現に展開した手法である。このとき、実数の確率を得るために、Real AdaBoost により弱識別器を学習する。

Joint Haar-like 特徴量では、組み合わせる特徴量から検出対象と非検出対象に判別することで 2 値化するが、この際にノイズやオクルージョンの影響により 2 値符号化を誤った場合、間違っただけの符号を組み合わせることになるため、同時確率に基づく識別が間違える可能性がある。一方、共起確率特徴は、連続値を演算子に組み合わせるため、ノイズやオクルージョンの影響により誤った場合においても、それ以外の組み合わせる特徴量が正しい識別を行うことにより、正しい識別を行うことができる。また、共起確率特徴は演算子によって結合するため、4 方向のエッジに基づく識別を線形和によって組み合わせる Shapelet 特徴量と同様の共起表現法である。しかし、Shapelet 特徴量はサンプルの重みが更新を行っても、Shapelet 特徴量の POOL が更新されないため、学習が進む過程で識別に適した Shapelet 特徴量を得ることができないなどの違いがある。

Real AdaBoost の弱識別器の出力を演算子で結合する共起確率特徴を算出するために、確率密度分布と弱識別器の評価値を算出する。まず、Low level な特徴量から検出対象と非検出対象の確率密度分布  $W_+$ ,  $W_-$  を作成する。確率密度分布  $W_{\pm}$  は、1 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  $D_t$  に基づいて式 (27)(28) より作成する。

$$W_+^j = \sum_{i: j \in J \wedge y_i = +1} D_t(i) \quad (27)$$

$$W_-^j = \sum_{i: j \in J \wedge y_i = -1} D_t(i) \quad (28)$$

ここで、 $t$  は学習回数、 $i$  は学習サンプルの番号、 $j$  は 1 次元ヒストグラムの BIN 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  $i$  から

特徴量を算出し、特徴量の値に対応する 1 次元ヒストグラムの 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、確率密度分布  $W_{\pm}$  を作成することができる。

次に確率密度分布  $W_{\pm}$  を用いて、検出対象と非検出対象の確率密度分布の分離度を表す評価値  $\alpha$  を式 (29) より算出する。

$$\alpha = 1 - \sum_j \sqrt{W_+^j W_-^j} \quad (29)$$

この評価値  $\alpha$  が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意味する。

そして、作成したポジティブクラスとネガティブクラスの確率密度分布  $W_+$ ,  $W_-$  を用いて、弱識別器の出力  $h(\mathbf{x})$  を算出する。学習サンプルより得られる特徴量  $\mathbf{x}$  の値から、1 次元ヒストグラムの BIN の番号  $j$  を算出し、それに対応した確率密度分布  $W_{\pm}$  から弱識別器の出力  $h(\mathbf{x})$  を式 (30) より算出する。

$$h(\mathbf{x}) = \frac{1}{2} \ln \frac{W_+ + \epsilon}{W_- + \epsilon} \quad (30)$$

算出した Real AdaBoost の弱識別器の出力  $h(\mathbf{x})$  を用いて共起特徴を算出する。low-level な特徴量を  $\mathbf{x}$ , 弱識別器の出力を  $h1(\mathbf{x}1)$ ,  $h2(\mathbf{x}2)$  とした場合、下記の 2 パターンの結合により共起確率特徴  $C1$ ,  $C2$  を式 (31)(32) より表現する。

$$C1 = h1(\mathbf{x}) + h2(\mathbf{x}) \quad (31)$$

$$C2 = h1(\mathbf{x}) \times h2(\mathbf{x}) \quad (32)$$

まず、共起確率特徴量  $C1$  は、弱識別器の出力の和であるため、2 つの特徴量を総合的に捉える特徴量となる。そのため、片方の特徴量がオクルージョンやノイズ等の外乱の影響を受けた場合でも、もう 1 つの特徴量がクラスをよく表現していればそのクラスを反映した特徴量となる。次に、共起確率特徴  $C2$  は、弱識別器の出力の積をとるため、両方の重みつき出力が高いときは高い共起確率特徴の値、それ以外は小さい共起確率特徴の値となる。共起確率特徴は、この捉え方が異なる 2 つの演算子によって生成する。

## 5 特徴量算出の高速化

物体検出は、入力画像中から検出ウィンドウをラスタスキャンし、検出ウィンドウより得られる特徴

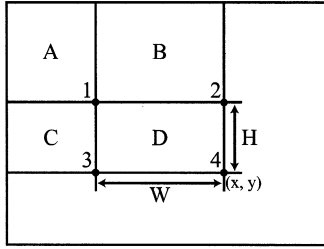


図 19: 積分画像

量を用いて物体検出を行う。そのため、非常に多くの検出ウィンドウから特徴量を算出することになるため、高速に特徴量を算出する方法が求められている。高速に特徴量を算出するために、積分画像または積分ヒストグラム画像を用いることが一般的である。積分画像は画素の値が画像の左上からの輝度値を累積した値であり、入力画像  $I(x, y)$  から求められる積分画像  $ii(x, y)$  は式 (33)(34) から算出できる。

$$s(x, y) = s(x, y - 1) + i(x, y) \quad (33)$$

$$ii(x, y) = ii(x - 1, y) + s(x, y) \quad (34)$$

$s(x, y)$  は行の画素の総和、 $ii(x, y)$  は列の  $s$  の総和を表す。ただし、 $s(x, -1) = 0$ 、 $ii(-1, y) = 0$  とする。図 19 の D の領域から輝度値の和を求める場合、式 (35) より算出することができる。

$$D = (ii(x, y) + ii(x - W, y - L)) - (ii(x - W, y) + ii(x, y - L)) \quad (35)$$

エッジに着目した特徴量の場合、積分画像を勾配方向ごとに作成することで同様に特徴量を高速に算出することができる。

## 6 画像データベース

特定物体認識の研究が盛んになると共に、共通データベースを用いて統一の評価実験を行うため、多くのデータベースが公開されている。データベースの多くは、画像中から検出対象の領域が示されていることが多く、同一のデータベースを用いることにより、容易に他の手法との比較を行うことができる。図 20 に Web 上で公開されているデータベースの一部を示す。

### 6.1 顔画像データベース

顔のデータベースの場合、正面から撮影した正面顔や横から撮影した横顔、大きく照明条件を変えて撮影した顔など多くのデータベースが公開されている。正面から撮影した顔では、Rowley らのニューラルネットワークによる顔検出 [6] に用いられた CMU テストセット [23] が多く用いられる。このデータセットは、人の正面顔 130 枚と回転を含んだ顔 50 枚のデータセットになっており、顔領域を示す座標データも公開されている。そのため、多くの顔検出法の有効性を示すために用いられてきた。MIT CBCL では、既に顔領域を切り出した顔と非顔のデータセット [24] を公開している。また、他にも顔検出以外の用途にも利用できるデータベースとして、Yale Database[25] が公開されており、このデータベースは、顔を正面より照明条件を大きく変えながら撮影した高解像度の顔データベースとなっている。他にも、顔の向きや照明条件が異なる顔画像を収集した PIE データベース [27] も公開されている。

### 6.2 人画像データベース

人のデータベースの場合、可視光カメラにより撮影したデータベースや、赤外線カメラを用いたデータベースが公開されている。MIT 歩行者データベース [28] は可視光で撮影しており、正面または後方から撮影した画像から構成されている。この MIT 歩行者データベースは、車からの歩行者検出をターゲットとして考えているため、画角や人の向きにバリエーションが少ない。そのため、検出が比較的容易であり、Dalal らによって INRIA データベース [29] が公開された。INRIA データベースは、正面や後方から撮影した画像だけでなく、横から撮影したサンプルも含まれており、照明条件、人の姿勢などが多様であり、非常に難しいデータベースとなっている。他にも、Wu らによる人検出 [30] に用いられた USC Pedestrian Set[31] や Munder らによる歩行者の識別 [32] に用いられた Gavrila's Pedestrian Set[33] などが公開されている。

これらのデータベースは、可視光カメラにより撮影したデータベースとなっているが、赤外線カメラを用いて撮影したデータベース [36] も公開されており、このデータベースを用いた歩行者検出法 [34]

データベース名	検出対象	向き	タイプ※	対象クラスの枚数	非対象クラスの枚数	備考
MIT+CMU Frontal Face Image	顔	正面	1	正面顔 130枚 回転顔 50枚		回転した顔画像を含む
CMU Profile Face Image	顔	横	1	208枚		
MIT CBCL Face Data	顔	正面	2	学習用 2,429枚 評価用 472枚	学習用 4,548枚 評価用 23,573枚	
Yale Face Database B	顔	正面	3	576枚	0枚	照明条件が大きく異なる
PIE	顔	正面, 横	3	41,368枚	0枚	視点, 照明, 表情が異なる
INRIA Person Dataset	人	正面, 後方, 横	1	学習用 614枚 評価用 288枚	学習用 1,218枚 評価用 453枚	カラー画像
USC Pedestrian Set	人	正面, 後方, 横	1	355枚		
MIT CBCL Pedestrian Data	人	正面, 後方	2	924枚		
Gavrila's Benchmark Dataset	人	正面, 後方, 横	2	4,000枚	5,000枚	
NICTA	人	正面, 後方, 横	2	25,551枚	5,207枚	カラー画像
OTCBVS	人	正面, 後方, 横	3	284枚		赤外線カメラ
UIUC	車	正面, 後方	学習 2 評価 1	学習用 550枚 固定スケール 170枚, マルチスケール 108枚	学習用 500枚	
MIT CBCL Car Data	車	正面, 後方	2	516枚	0枚	
CMU	車	複数視点	1	104枚		

※ 1 Ground Truthを含むデータベース  
 2 切り出し済みデータベース  
 3 Ground Truth及び切り出した画像がないデータベース

図 20: データベース

も提案されている。また、近年では、人体検出を行うためにどのようなデータセットを用意すれば良いかを検証した研究 [35] も行われており、これに基づいた NICTA 歩行者データベース [37] も公開されている。

### 6.3 車画像データベース

車のデータベースでは、視点の異なるデータベースが公開されている。車の正面もしくは後方から撮影したデータベースとして、MIT 車データベース [38]、車の横から撮影したデータベースとして、UIUC 車データベース [39] が公開されている。また、複数の視点から撮影したデータベースとして、CMU 車データベース [40] が公開されている。

## 7 特徴量の比較実験

本章では、複数の検出対象を例に、特徴量の性能を評価するために比較実験を行う。特徴量は検出対象により、向き不向きがあるため、今回は顔と車、人の3つを検出対象とし、検出対象ごとにどのような特徴量が有効であるか検証する。比較を行う特徴量は、Haar-like 特徴量と EOH 特徴量、Edgelet 特徴量、HOG 特徴量の4つとする。

### 7.1 実験概要

学習と評価には、公開されているデータベースを利用する。利用するデータベースを下記に示す。

- 顔データセット
  - 学習: Web から収集した約 5000 枚

- 評価 : MIT+CMU Frontal Face Image[23]

● 車データセット

- 学習 : UIUC Cars [39]

- 評価 : UIUC Cars [39]

● 人データセット

- 学習 : INRIA Person Dataset[29]

- 評価 : USC Pedestrian Detection Test Set[31]

今回の比較実験では, Viola 等が提案したカスケード型 AdaBoost により学習を行う. カスケード型 AdaBoost は, 段階的に計算量の異なる強識別器を直列に繋げた識別器である. カスケード構造により, 非検出対象のサンプルは, 比較的早い段階で判別することができるため, 計算効率が非常に良い.

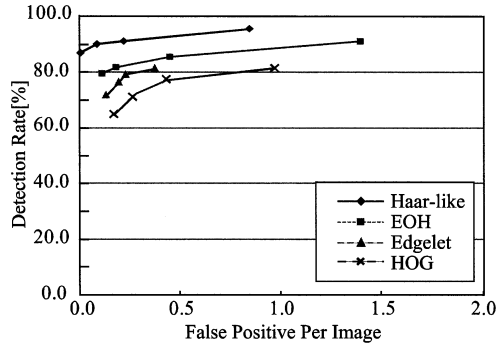
カスケードの各段が, 学習用検出対象サンプルを通過させる割合を  $D_r (0 < D_r < 1)$ , 学習用非検出対象サンプルを通過させる割合を  $F_p (0 < F_p < 1)$  としたとき, カスケード  $n$  段通過後は,  $(F_p)^n$  だけ通過することになる. 今回は, 特徴量を統一した評価を行うために,  $D_r = 0.995$ ,  $F_p = 0.300$  に固定し, カスケードの段数の最大  $n_{max} = 16$ , 1つの強識別器に用いる弱識別器の最大数  $h_{max} = 100$  に統一した.

## 7.2 評価方法

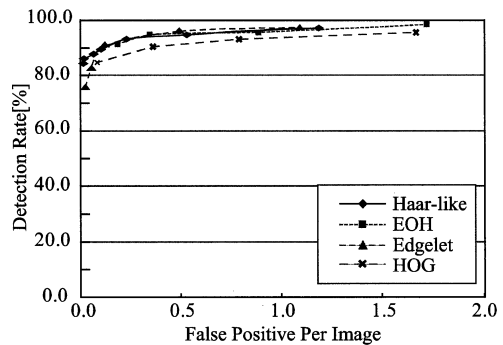
評価実験は, 評価用サンプルを使用し, 対象の物体検出を行う. 検出結果より, 対象物の検出率 (Detection Rate) と1枚の画像中での誤検出率 (False Positive Par Image) を算出し, ROC (Receiver Operator Characteristics) カーブにより比較を行う. ROCカーブの縦軸に検出率, 横軸に誤検出率をプロットすることで, 各特徴量の性能を比較することができる. ROCカーブは, グラフの左上ほど検出器の性能が高いことを表す. 検出率と誤検出率は式 (36)(37) より求めることができる.

$$\text{検出率} = \frac{\text{検出した正解数}}{\text{検出対象数}} \quad (36)$$

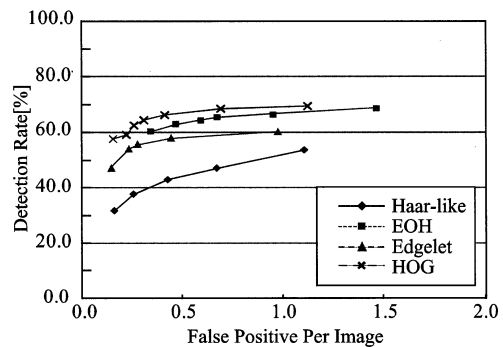
$$\text{誤検出率} = \frac{\text{誤検出した数}}{\text{評価用画像数}} \quad (37)$$



(a) 顔画像の比較



(b) 車画像の比較



(c) 人画像の比較

図 21: 実験結果の ROC

## 7.3 実験結果

図 21 に実験結果<sup>1</sup>を示す. 実験結果より, 検出対象と特徴量によって検出性能が異なることがわかる.

<sup>1</sup>今回の実験は, 各特徴量の特長を比較するための実験である. そのため, 検出精度が提案されている論文よりも若干低い場合がある



顔の場合では、輝度に着目した特徴量である Haar-like 特徴量が最も検出精度が良く、次にエッジに着目した特徴量である EOH 特徴量, Edgelet 特徴量, HOG 特徴量の順となっている。顔の特徴として、目や鼻、口などの器官はどのような人でも同じ位置にあり、比較的剛体に近いので、形状変化が少ない。そのため、顔の器官を捉えることができる Haar-like 特徴量が、最も検出性能が高いと考えられる。

車の場合では、全ての特徴量において同程度の検出精度が得られている。車画像は、顔と同様に形状変化が少なく、テクスチャの多様さも少ない。そのため、エッジに着目する特徴量も輝度に着目する特徴量も高精度な検出を行うことができたと考えられる。

人の場合では、顔や車と比べて検出精度が低いことがわかる。これは、顔や車は比較的剛体に近い物体であるのに対して、人は姿勢の変化による形状変化が大きい。また、画像中での人同士の重なりによるオクルージョンも多く発生するため、顔や車と比べて検出精度が低くなる。このような問題に対しては、エッジに着目した HOG 特徴量が最も検出精度が良く、EOH 特徴量, Edgelet 特徴量, Haar-like 特徴量の順になっている。エッジに着目した特徴量は、輝度に着目した特徴量よりも、衣服の色の違いに対してロバストな特徴量を抽出することができるためである。エッジに着目した特徴量の中でも、HOG 特徴量と EOH 特徴量は検出精度が高い。これは、局所領域内でエッジの方向ヒストグラムを作成することにより、局所領域内での若干の形状変化を吸収できるためである。

今回の実験より、検出対象と特徴量によって、検出性能に違いがあることから、高精度な検出を行うためには、検出対象にあった特徴量を選択することが必要となる。また、本稿で述べた特徴量は、輝度、エッジに着目した特徴量においても、特長に違いがある。そのため、多数の特徴量を用いることや、第三世代の特徴量で取り上げた関連性を表現することにより、さらに高精度な検出を行うことができると考えられる。

最後に、本稿で取り上げた特徴量について、図 22 にまとめる。図 22 の項目にある候補プール数は、 $24 \times 24$  画素の顔画像から得られる特徴量の個数である。他にも、特定物体認識において有効な特徴量の評価は、文献 [41] で比較実験によって確かめられている。文献 [41] では、本稿で紹介した第二世代の

特徴量である Haar-like 特徴量や HOG 特徴量、第三世代の特徴量である Shapelet 特徴量の比較実験を行っている。また、文献 [42] では、第三世代の特徴量のように、Boosting を用いて特徴量を合成させることにより、第二世代の特徴量よりも検出性能が高いことが報告されている。

## 8 おわりに

本稿では、物体認識における学習手法の進化から特徴量の変化を進化として捉え、特徴量を世代として分類した。特徴量は世代を追うごとに、研究者の事前知識無しに大量の特徴量候補の中から、検出対象に最適な特徴量を選択することが可能になっている。これには、ブースティングが大きな役割を果たしている。また、反対にブースティングを生かすように特徴量が進化したともいえる。このような、学習手法と特徴量の両者の進化により、物体認識の高精度化、高速化が実現されたと言える。

## 参考文献

- [1] Viola, P. and Jones, M.: Robust real-time face detection, International Journal of Computer Vision, Vol. 57, no. 2, pp. 137-154, 2004.
- [2] Dalal, N. and Triggs, B.: Histograms of Oriented Gradients for Human Detection, In Proc. IEEE Conf. on Computer Vision and Pattern Recognition 2005, Vol.1, pp.886-893, 2005.
- [3] Sakai, T., Nagao, M. and Fujibayashi, S.: Line extraction and Pattern recognition in a photograph. PR, Vol. 1, pp.233-248, 1969.
- [4] Maio, D. and Maltoni, D.: Real-time face location on gray-scale static images, PR, Vol. 33, No. 9, pp-1525-1539, 2000.
- [5] Leung, T. K., Burl, M. C. and Perona, P.: Finding faces in cluttered scenes using random labeled graph matching, In Proc. IEEE Conf. on International Conference of Computer Vision, pp. 637-644, 1995.
- [6] Rowley, H. A., Baluja, S., and Kanade, T.: Neural Network-Based Face Detection, IEEE Trans on Pattern Analysis and Machine Intelligence, Vol.20, Issue 1, pp.23-38, 1999.
- [7] Papageorgiou, C. and Poggio, T.: A Trainable System for Object Detection, International Journal of Computer Vision, Vol.38, no.1, pp. 15-33, 2000.
- [8] 佐部, 日台: ピクセル差分特徴を用いた実時間任意姿勢顔検出器の学習、第 10 回画像センシングシンポジウム, 2004.

第二世代の特徴量

特徴量	候補プール数	長所	短所
Haar-like	18 万個	<ul style="list-style-type: none"> <li>● 顔や車など形状変化の少ない物体に有効</li> <li>● 演算速度が高速 (積分画像を用いた場合)</li> </ul>	<ul style="list-style-type: none"> <li>● 人等の形状やテクスチャが多様な物体に対して不向き</li> </ul>
ピクセル差分	16 万個	<ul style="list-style-type: none"> <li>● 顔など形状変化の少ない物体に有効</li> <li>● 演算速度が高速</li> </ul>	<ul style="list-style-type: none"> <li>● 照明変化などの撮影条件により性能が低下</li> </ul>
EOH	60 万個	<ul style="list-style-type: none"> <li>● 形状変化の少ない顔のみでなく、形状変化を含む人にも有効</li> <li>● 演算速度が高速 (積分画像を用いた場合)</li> </ul>	<ul style="list-style-type: none"> <li>● 低コントラストなどの撮影条件により性能が低下</li> </ul>
HOG	900 個	<ul style="list-style-type: none"> <li>● 人などの形状やテクスチャが多様な物体に有効</li> </ul>	<ul style="list-style-type: none"> <li>● 低解像度な物体や小さい物体の検出には不向き</li> <li>● 演算が遅い</li> </ul>
Edgelet	60万個	<ul style="list-style-type: none"> <li>● 顔や車などの形状変化が少なく、テクスチャが多様な物体に有効</li> </ul>	<ul style="list-style-type: none"> <li>● 輪郭等の形状が多様な物体には不向き</li> <li>● 演算速度が遅い</li> </ul>
Joint Haar-like	18 億個	<ul style="list-style-type: none"> <li>● 少ない特徴量数で顔や車などの形状変化が少ない物体に有効</li> </ul>	<ul style="list-style-type: none"> <li>● 人等の形状やテクスチャが多様な物体には不向き</li> </ul>
スパース	300万個	<ul style="list-style-type: none"> <li>● 演算回数が少ないため、高速</li> <li>● 少ない特徴量数で顔などの形状変化が少ない物体に有効</li> </ul>	<ul style="list-style-type: none"> <li>● 学習の組み合わせが膨大</li> </ul>
Shapelet	116 個	<ul style="list-style-type: none"> <li>● 人などの形状やテクスチャが多様な物体に有効</li> </ul>	<ul style="list-style-type: none"> <li>● 演算速度が遅い</li> </ul>
Joint HOG	300 個	<ul style="list-style-type: none"> <li>● 人などの形状やテクスチャが多様な物体に有効</li> </ul>	<ul style="list-style-type: none"> <li>● 演算速度が遅い</li> </ul>
共起確率	80 万個	<ul style="list-style-type: none"> <li>● 人などの形状やテクスチャが多様な物体に有効</li> </ul>	<ul style="list-style-type: none"> <li>● 演算速度が遅い</li> </ul>

第三世代の特徴量

図 22: 特徴量の比較

- [9] Levi, K. and Weiss, Y.: Learning Object Detection from a Small Number of Examples: the Importance of Good Features, In Proc. IEEE Conf. on Computer Vision and Pattern Recognition 2004, Vol.2, pp.53-60, 2004
- [10] Wu, B. and Nevatia, R.: Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors, In Proc. IEEE Conf. on International Conference of Computer Vision, vol.1 pp. 90-97, 2005.
- [11] Y, Freund and R, E. Schapire : “A decision-theoretic generalization of on-line learning and an application to boosting”, Journal of Computer and System Sciences, No. 1, Vol. 55, pp. 119-139, 1997.
- [12] Lienhart, R. and Maydt, J.: An Extended Set of Haar-like Features for Rapid Object Detection, In Proc. IEEE Conf. on International Conference of Image Processing 2002, Vol.1, pp.900-903, 2002.
- [13] Bartlett, M. S., Littlewort, G., Fasel, I. and Movellan, J. R.: Real time face detection and facial expression recognition: Development and application to human computer interaction, In Proc. IEEE Conf. on Computer Vision and Pattern Recognition WorkShop, pp.139-157, 2003.
- [14] Wu, B., Ai, H., Huang, C. and Lao, S.: Fast Rotation Invariant Multi-view Face Detection Based on Real AdaBoost, In Proc. IEEE Conf. on Automatic Face and Gesture Recognition, 2007.
- [15] Mita, T., Kaneko, T., Stenger, B., T, and Hori, O.: Discriminative Feature Co-occurrence Selection for Object Detection, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.30, no.7, pp.1257-1269, 2008
- [16] Schapire, R. E. and Singer, Y.: Improved Boosting Algorithms Using Confidence-rated Predictions, Machine Learning, No.37, pp.297-336, 1999.
- [17] Huang, C., Ai, H., Yamashita, T., Lao, S. and Kawade, M.: Incremental Learning of Boosted Face Detector, In Proc. IEEE International Conference of Computer Vision, 2007.

- [18] Zhuang, Q., Avidan, S., Yeh, M. and Cheng, K. : Fast Human Detection Using a Cascade of Histograms of Oriented Gradients, In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Vol. 2, pp. 1491-1498 2006.
- [19] Hou, C., Ai, H. and Lao, S.: Multiview Pedestrian Detection Based on Vector Boosting, 8th Asian Conference on Computer Vision, 2007.
- [20] Sabzmejdani, P. and Mori, G.: Detecting Pedestrians by Learning Shapelet Features, In Proc. IEEE Conf. on Computer Vision and Pattern Recognition 2007, pp.1-8, 2007.
- [21] 尾崎, 山内, 藤吉, :Joint HOG 特徴量を用いた 2 段階 AdaBoost による車両検出, 動的画像処理実用化ワークショップ (DIA2008), pp.101-106,2008.
- [22] 山内, 藤吉, 山下, :Boosting に基づく共起表現による人検出, 第 11 回画像の認識・理解シンポジウム (MIRU2008), pp180-187, Jul. 2008.
- [23] CMU Frontal Face Images : [http://vasc.ri.cmu.edu/idb/html/face/frontal\\_images/index.html](http://vasc.ri.cmu.edu/idb/html/face/frontal_images/index.html)
- [24] MIT CBCL Face Data : <http://cbcl.mit.edu/cbcl/software-datasets/FaceData.html>
- [25] Yale Face Database : <http://cvc.yale.edu/>
- [26] CMU Profile Face Images : [http://vasc.ri.cmu.edu/idb/html/face/profile\\_images/index.html](http://vasc.ri.cmu.edu/idb/html/face/profile_images/index.html)
- [27] PIE Database : [http://www.ri.cmu.edu/projects/project\\_418.html](http://www.ri.cmu.edu/projects/project_418.html)
- [28] MIT CBCL Pedestrian Data : <http://cbcl.mit.edu/cbcl/software-datasets/PedestrianData.html>
- [29] INRIA Person Dataset : <http://pascal.inrialpes.fr/data/human/>
- [30] Wu, B. and Nevatia, R.: Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet based Part Detectors, International Journal of Computer Vision, Vol.75, pp. 247-266, 2007.
- [31] USC Pedestrian Detection Test Set : <http://iris.usc.edu/bowu/DatasetWebpage/dataset.html>
- [32] Munder, S. and Gavrilu, D. M.: An Experimental Study on Pedestrian Classification, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 28, no. 11, pp. 1863-1868, 2006.
- [33] DaimlerChrysler Pedestrian Classification Benchmark Dataset : <http://www.gavrila.net/index.html>
- [34] Zhang, L., Wu, B. and Nevatia, R.: Pedestrian Detection in Infrared Images based on Local Shape Features, OTCBVS Workshop, In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 17-22, 2007.
- [35] Gary, O., Lars, P., Nathan, B., Lars, A. and Niklas, P.: A New Pedestrian Dataset for Supervised Learning, IEEE Conf. on Intelligent Vehicles Symposium, pp.373-378, 2008.
- [36] OTCBVS Benchmark Dataset Collection : <http://www.cse.ohio-state.edu/otcbvs-bench/>
- [37] NICTA Pedestrian Dataset : [http://nicta.com.au/research/project\\_list/completed\\_projects/smart\\_cars/computer\\_vision\\_datasets/](http://nicta.com.au/research/project_list/completed_projects/smart_cars/computer_vision_datasets/)
- [38] MIT CBCL Car Data : <http://cbcl.mit.edu/cbcl/software-datasets/CarData.html>
- [39] UIUC Image Database for Car Detection : <http://l2r.cs.uiuc.edu/cogcomp/Data/Car/>
- [40] Testing Images for Car Detection : <http://vasc.ri.cmu.edu/idb/html/car/index.html>
- [41] Wojek, C. and Schiele, B.:A Performance Evaluation of Single and Multi-feature People Detection, 30th DAGM Symposium (DAGM 2008), pp. 82-91, 2008.
- [42] Feng, H., Ying, S., Sawhney, H. S., and Kumar, R.: Discovering class specific composite features through discriminative sampling with Swendsen-Wang Cut, In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1-8 2008.