

局所特徴量の関連性に着目した Joint 特徴による物体検出

藤吉 弘亘

† 中部大学 工学部 情報工学科

hf@cs.chubu.ac.jp

あらまし 本稿では、複数の HOG 特徴量間の共起を表現する Joint 特徴を用いた動画からの物体検出法について述べる。Joint 特徴は、組み合わせられた 2 つのセル間の 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。次に、生成された Joint 特徴候補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、単一の HOG 特徴量のみでは捉えることができない物体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。Joint 特徴の有効性を示すために、人と車両を検出対象として評価実験を行い、Joint 特徴の有効性を述べる。さらに、異なる解像度の HOG 特徴量間の共起や時空間特徴量の共起、TOF カメラから得られるデプス情報との共起の効果についても述べる。

Object Detection by Joint Feature Based on Relations of Local Features

Hironobu Fujiyoshi

†Dept. of Computer Science, Chubu Univ.

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

Abstract This paper presents a method for detecting objects in a video using a Joint feature, which represents co-occurrence between multiple HOG features. Joint features represent the co-occurrence of the HOG features of two cells combined by the first-stage Real AdaBoost. Next, the generated Joint features are input to the second-stage Real AdaBoost, which constructs the final classifier. In this way, it is possible to capture shape symmetry and edge continuity, which single HOG features cannot do, so highly accurate detection is possible. We report experiments involving the detection of humans and vehicles performed to test the effectiveness of the proposed method. Furthermore, we also describe effectiveness of multi-resolution HOG representation, co-occurrence of spatio-temporal features, and co-occurrence of depth information obtained by a TOF(Time of Flight) camera.

1 はじめに

画像中から対象物体を自動的に検出する技術は、監視システムや ITS など多岐に渡る分野での実現が期待されている。また、物体追跡や姿勢認識の前処理としても必要であるため、近年盛んに多くの研究が取り組まれている。近年の物体検出法は、Viola と Jones が提案した顔検出 [1] に代表されるように low-level な特徴と統計的学習法の組み合わせによる手法

が多く提案されている。物体検出に有効な low-level な局所特徴量には、領域の累積エッジ強度の比を特徴量とする Edge Orientation Histograms (EOH) 特徴量 [2] や、局所領域における勾配方向をヒストグラム化した特徴量である Histograms of Oriented Gradients (HOG) 特徴量 [3]、エッジ同士をつなげた短い線、カーブを特徴量として表現する Edgelet 特徴量 [4] などが提案されている。これらのエッジに

着目した特徴量は，局所領域内の物体形状を表すことができるため，顔だけでなく人や車両の検出に用いられる．

人の形状には，大きく分けて下記に示す2つの特徴がある．

(1) 頭から肩にかけての Ω 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により局所領域の面の情報を表現した Shapelet 特徴量 [5] 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することにより，共起性を表現する Joint Haar-like 特徴量 [6] が提案されている．両手法は，複数の low-level な特徴量をブースティングにより組み合わせることで特徴量間の関連性を捉えることができ，高精度な検出を実現している．このように近年は，ブースティングを用いて，複数の特徴量の関連性をどのように捉えるかが重要となっている．

我々は，これまでに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 Joint 特徴と2段階 Real AdaBoost による物体検出法を提案した．Joint 特徴は，異なる2つの領域の HOG 特徴量を Real 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 Real AdaBoost により生成することが可能となる．生成された Joint 特徴を入力とした2段階目の Real AdaBoost によって最終識別器を構築し，検出ウィンドウ内の画像が検出対象/非検出対象であるか判別する．本稿では，特徴量間の共起表現を用いた Joint 特徴による物体検出法とその効果について述べる．

2 局所特徴量とブースティング

従来の物体検出法は，入力画像と背景画像の差分を計算する背景差分ベースの手法が多く用いられてきた．このような動体検出をベースとした動画像理解のアプローチは，移動体同士が画像上で重なった場合にセグメンテーションに失敗するため，その後の処理である物体識別が不可能となる問題がある．

これに対して，Viola と Jones は Haar-like 特徴量と呼ばれる局所特徴量と統計的学習を組み合わせた高速かつ高精度な顔検出法 [1] を提案した．この

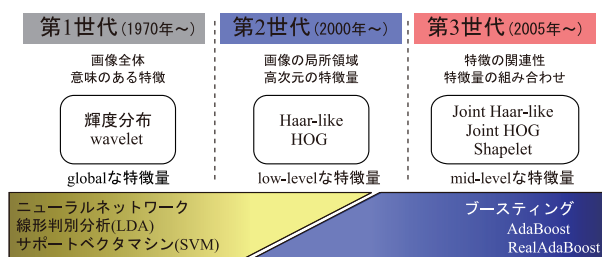


図 1: 学習法と特徴量の変遷

手法は，入力画像に対して検出ウィンドウをラスタスキャンし，AdaBoost の学習アルゴリズムにより構築された識別器を用いて，検出ウィンドウ内の画像を顔と顔以外に判別するため，前処理として動体検出を必要としない．顔検出の分野では，ラスタスキャン方式の手法が多く提案され実用化が進んでいる．近年では，検出対象が顔から形状変化が大きいため検出がより難しいとされている人などへと変わりつつある．本章では，局所特徴量と統計的学習の変遷と，HOG 特徴量を用いた人検出法について述べる．

2.1 識別器と特徴量の変遷

物体検出に用いられる特徴量は，学習手法の進化の側面から特徴量を3つの世代に分類することができる [7]．特徴量は，図1に示すように第1世代の global な特徴量，第2世代の low-level な特徴量，第3世代の mid-level な特徴量に分けることができる．以下に各世代毎の特徴量について述べる．

第1世代の特徴量 物体検出に用いられる第1世代の特徴量は，輝度分布や wavelet [8] など画像全体から得られるものや，形状の複雑度のような意味のある特徴量を研究者の経験に基づき決定していた．第1世代の学習手法としては，ニューラルネットワークが用いられており，次元数が大きな特徴量ではニューラルネットワークのサイズが大きくなり，識別器の構築が困難となる．そのため，入力ユニットの数を小さくする必要があり，低次元数の特徴量が用いられていた．

第2世代の特徴量 2000年以降に提案された第2世代の特徴量は，Haar-like 特徴量や HOG 特徴量などの局所領域から得られる局所特徴量である．これらの low-level な特徴量は，ブースティングを用いて特徴選択を行うことで高次元(数千～数十万バ

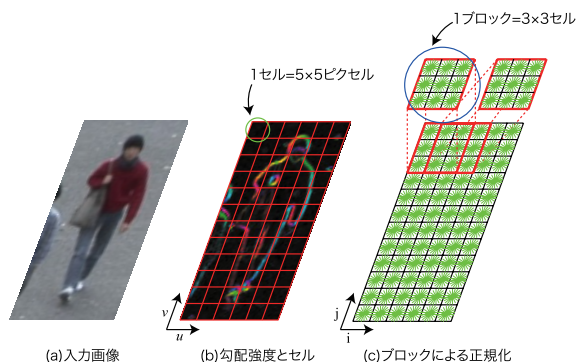


図 2: HOG で用いる領域の構造

ターン) な特徴量を扱うことが可能となった．これにより，第 1 世代では研究者が特徴量を決定していたが，第 2 世代では研究者の事前知識なく，多数の特徴量候補の中から有効な特徴量を自動的に選択することができる．

第 3 世代の特徴量 第 3 世代 (2005 年以降) の特徴量では，第 2 世代の low-level 特徴量をブースティングの特徴選択を利用して組み合わせ，識別に有効な関連性や共起性を表現した Shapelet 特徴量 [5] や Joint Haar-like 特徴量 [6]，Joint 特徴 [9]，共起確率特徴量 [13] などが登場した．これらの特徴量は，第 2 世代の特徴量をベースとしおり，1 つ 1 つでは勾配情報のみを表す 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 mid-level な特徴量に拡張している．

2.2 HOG 特徴と人検出

第 2 世代の局所特徴量である Histograms of Oriented Gradients(HOG)[3] は，Dalal 等により提案された．HOG 特徴量は，局所領域における輝度の勾配方向をヒストグラム化した特徴量であり，ある一定領域に対する特徴量の記述を行う．そのため，大まかな物体形状を表現することが可能であり，人検出 [3][10][11] や車両検出 [12] 等の物体検出に有効であることが報告されている．以下に，HOG 特徴量の算出法と HOG 特徴量を用いたブースティングによる人検出法について述べる．

2.2.1 HOG 特徴量の算出アルゴリズム

HOG 特徴量は，局所領域における輝度の勾配方向をヒストグラム化した特徴量であり，物体の形状

を表現することができる．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長がある．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 から勾配強度 m と勾配方向 θ 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 θ よりセル毎 (図 2(b)) に輝度の勾配方向ヒストグラムを作成する．得られた勾配方向を 20° ずつに分割することにより，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ごと (図 2(c)) に特徴量を正規化する．

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

ここで， v は HOG 特徴量， k はブロック内の HOG 特徴量の数， ϵ 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．

2.2.2 ブースティングによる人検出

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 (5) に示すように， T 個の弱識別器 $h_t(x)$ を線形結合した強識別器 $H(x)$ として学習する．

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

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[14] は 0 または 1 の 2 値，Real AdaBoost[15] は実数で出力する．

図 3 に，HOG 特徴量を用いた際のブースティングによる学習の流れを示す．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

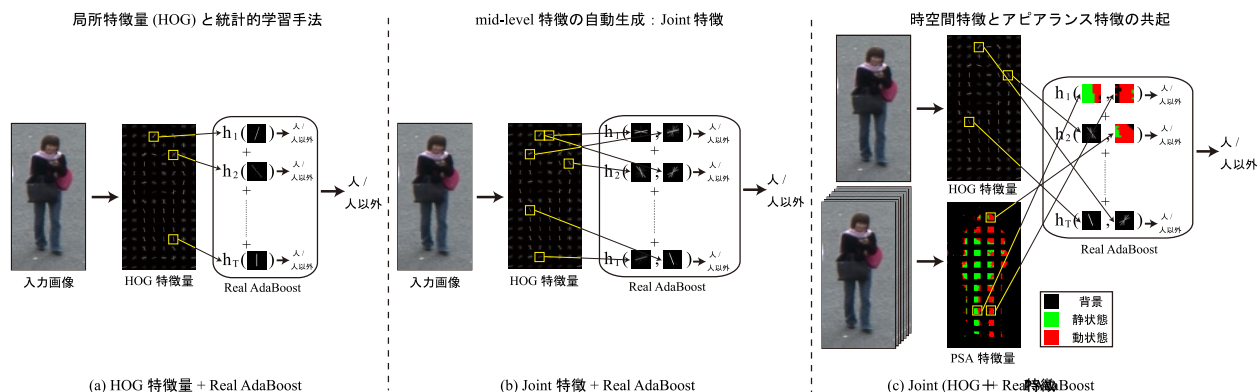


図 4: 人検出における特徴量の捉え方

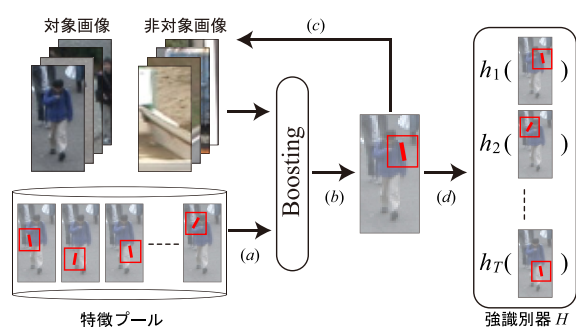


図 3: ブースティングによる学習の流れ

特徴プールには位置や勾配方向が異なる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、HOG 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 HOG 特徴量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) . 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 . 詳細な学習アルゴリズムの処理については、文献 [1][14] を参考にいただきたい .

図 4(a) に、学習した識別器を用いた識別の流れを示す . はじめに、入力画像から HOG 特徴量を算出する . 次に、学習により弱識別器として選択された HOG 特徴量を用いて人と人以外に判別する . こ

の際、AdaBoost では 2 値の出力に弱識別器の信頼により重みつけた線形結合、Real AdaBoost では実数の出力を線形結合する . 最後に、線形結合された数値をしきい値処理することにより人と人以外に判別する .

3 特徴量の共起表現

高精度な物体検出を実現するために、局所特徴量や統計的学習法の改良などの手法が提案されている . しかし、局所特徴量や統計的学習法の改良だけでは検出性能の向上に限界があり、このような問題を解決するために、特徴量間の共起性を表現する特徴量 [5][6][9][13] が提案されている .

3.1 従来の共起表現法

特徴量間の共起性を表現する手法として、複数の Haar-like 特徴の共起性を表現する Joint Haar-like 特徴量 [6] がある . Joint Haar-like 特徴量は、図 5(a) に示すように複数の Haar-like 特徴量を 2 値に識別し、2 値の符号を AdaBoost により最適に組み合わせることで表現される . 識別時には、複数の Haar-like 特徴量を検出対象と非検出対象に判定し、その同時確率によって識別を行う . これにより、Joint Haar-like 特徴では、物体の構造に基づいた特徴量の関係性を捉えることができる . Joint Haar-like 特徴は、Viola と Jones の顔検出法 [1] と比較して、高精度かつ高速に処理することが可能である . しかし、組み合わせる特徴量の中に誤った 2 値符号が含まれていた場合、Joint Haar-like 特徴量に大きな影響を

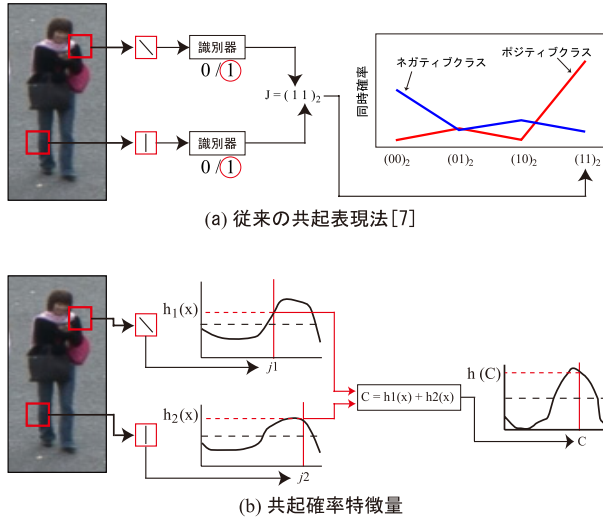


図 5: 特徴量の共起表現

与える．特に検出対象が人など形状変化の激しいものや画像中の重なりによるオクルージョンが発生した場合にこれらの問題が発生すると考えられる．この問題を解決するために、我々は図 5(b) に示すように Real AdaBoost の弱識別器から得られる実数の確率を演算子によって結合する共起表現に展開した．

3.2 共起確率特徴量

共起確率特徴量は、弱識別器の出力が連続値である Real AdaBoost を利用し、弱識別器の出力を複数の演算子によって結合することで得られる特徴量である．そのため、検出対象の形状変化が激しいものやオクルージョンに対応することができる．

3.2.1 共起確率特徴量の算出アルゴリズム

共起確率特徴量は、Real AdaBoost の弱識別器の出力を演算子によって結合した特徴量である．まず、学習用の検出対象画像（ポジティブクラス）と非検出対象画像（ネガティブクラス）から low-level な特徴量を算出する．ここでは、物体検出に用いられることが多い HOG 特徴量を使用する．次に、得られた HOG 特徴量から確率密度分布 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

て次式より作成する．

$$W_+^j = \sum_{i: j \in J \wedge y_i = +1} D_t(i) \quad (6)$$

$$W_-^j = \sum_{i: j \in J \wedge y_i = -1} D_t(i) \quad (7)$$

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．作成した確率密度分布 W_{\pm} は、クラス毎の確率密度分布の総和が 1 となるように正規化する．

作成した確率密度分布 W_+ , W_- に対応した弱識別器の出力 $h(x)$ は次式より算出される．

$$h(x) = \frac{1}{2} \ln \frac{W_+ + \epsilon}{W_- + \epsilon} \quad (8)$$

ここで、 ϵ 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．

算出した Real AdaBoost の弱識別器の出力 $h(x)$ を用いて共起を表現する．検出ウィンドウから得られた HOG 特徴量を x 、弱識別器の出力を $h1(x)$, $h2(x)$ 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量 $C1$, $C2$ を得る．

$$C1 = h1(x) + h2(x) \quad (9)$$

$$C2 = h1(x) \times h2(x) \quad (10)$$

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、片方の特徴量が隠れ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、共起確率特徴量 $C2$ は、弱識別器の出力の積をとるため、両方の出力が非常に高い場合のみ高い値の出力が得られる．

共起確率特徴量 $C1$, $C2$ は、Real AdaBoost の学習アルゴリズムにより、最適な low-level 特徴量の組み合わせが自動的に選択される．学習アルゴリズムの詳細は、文献 [13] を参照して頂きたい．

3.2.2 共起確率特徴量の効果

共起確率特徴量の有効性を示すために、人画像データベースを用いて比較実験を行った．図 6 に、従来の共起表現法 [6] と共起確率特徴量の比較結果を示す．共起確率特徴量は、従来の共起表現法と比

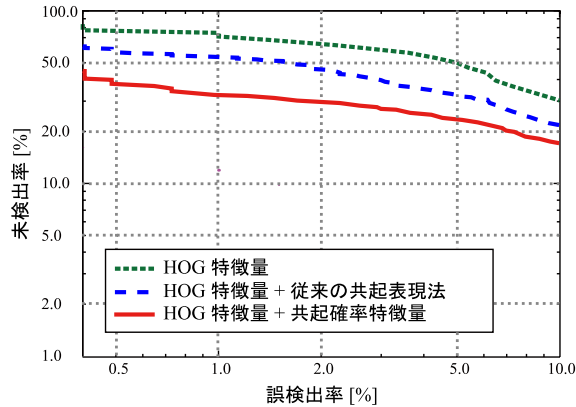


図 6: DET カーブによる検出性能の比較

べて高い検出性能であることがわかる．この理由として，従来の共起表現法は特徴量を 2 値に識別していたが，共起確率特徴量は連続値で表現しているため，誤識別した特徴量を組み合わせた場合の影響が少ないと考えられる．また，共起確率特徴量は連続値で表現される検出対象クラスの確率を複数の演算子により共起を表現するため，従来の共起表現法よりも多様な共起表現が可能となったためだと考えられる．

4 Joint 特徴による物体検出

3 章では，特徴量間の共起を表現することにより，高精度な検出が可能であることを述べた．本章では，共起表現法を用いて局所領域間の関連性を捉える Joint 特徴 [9] について述べる．

4.1 Joint 特徴と 2 段階ブースティング

Joint 特徴の生成と最終識別器の構築を図 7 に示す．Joint 特徴の生成と学習は，2 段階の Real AdaBoost により行われる．ここでは，Joint 特徴を生成する 1 段階目の Real AdaBoost と，最終識別器を学習する 2 段階目の Real AdaBoost にわけて述べる．

4.1.1 1 段階目の Real AdaBoost による Joint 特徴の生成

Joint 特徴を生成するために，2 つの異なる局所領域から 1 個ずつの特徴量を抽出し，3.2 節で述べた

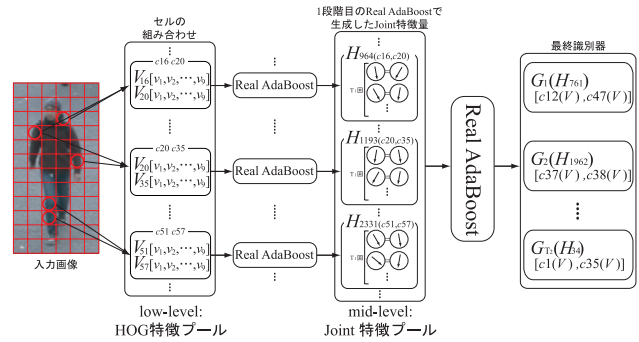


図 7: Joint 特徴による 2 段階 Real AdaBoost

共起表現法により HOG 特徴量の共起を表現する．

まず，位置の異なる 2 つのセル c_m, c_n において HOG 特徴量を算出する．次に，2 つのセルから 1 個ずつの HOG 特徴量を用いて共起確率特徴量を生成する．その際に，共起確率特徴量は HOG 特徴量の全ての組み合わせに対して求め，Real AdaBoost により最も良い組み合わせを弱識別器として自動的に選択する．この処理を T 回繰り返す，1 段階目の Real AdaBoost により次式で表される Joint 特徴である強識別器 $H(x)$ を学習する．

$$H^{c_m, c_n}(x) = \sum_{t=1}^T h_t^{c_m, c_n}(x) \quad (11)$$

上記の処理を全てのセルの組み合わせに対して行ない，組み合わせ数と同数の Joint 特徴候補を生成する．例えば，入力画像が 30×60 ピクセル，セルサイズを 5×5 ピクセルとした場合，72 個のセルに分割され，組み合わせ数は ${}_{72}C_2 = 2,556$ となるため，2,556 個の Joint 特徴候補 $H^{c_m, c_n}(x)$ を生成する．生成したすべての Joint 特徴候補を 1 つのプールとし，後述する 2 段階目の Real AdaBoost の入力とする．

4.1.2 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 特徴候補のプールを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 Joint 特徴を自動的に選択することが可能となる．

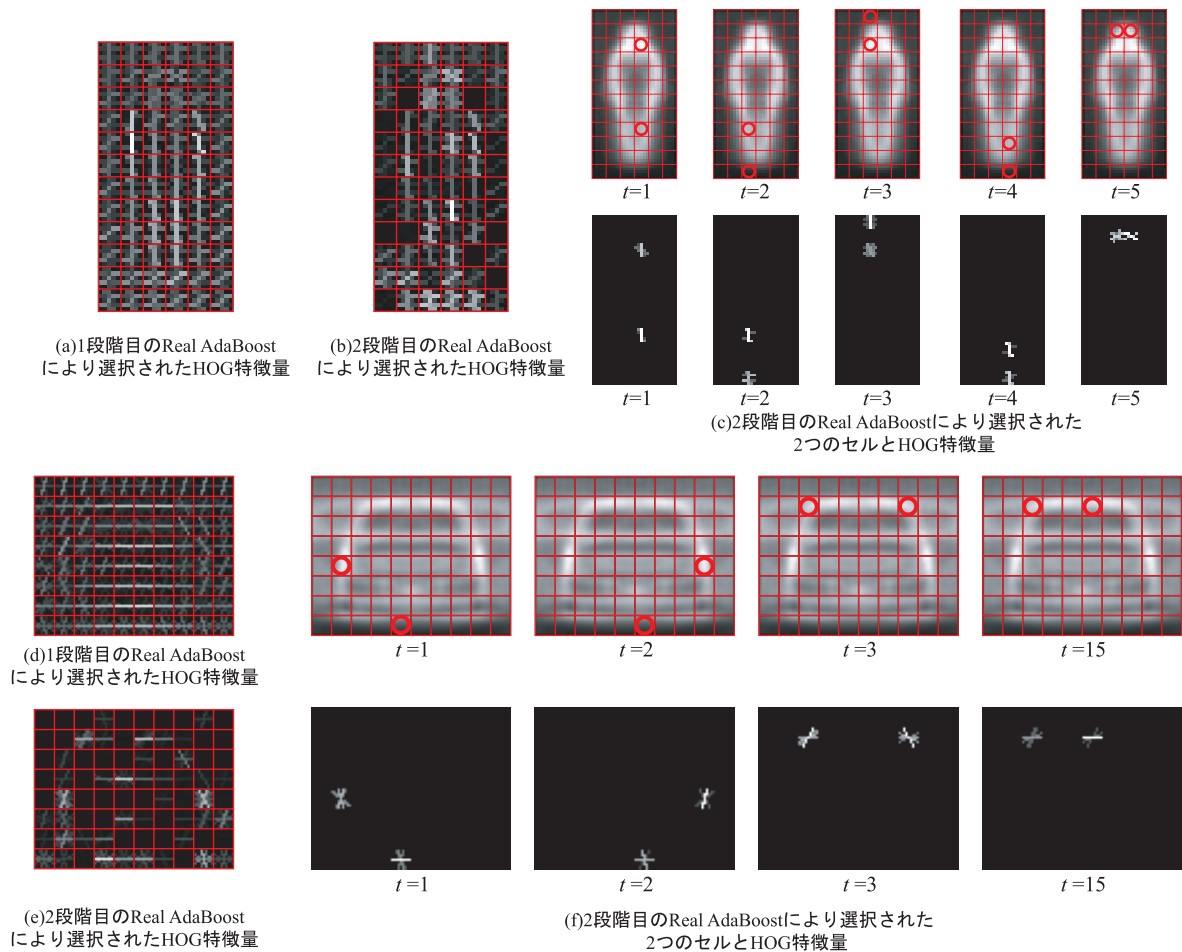


図 10: 選択された Joint 特徴の可視化

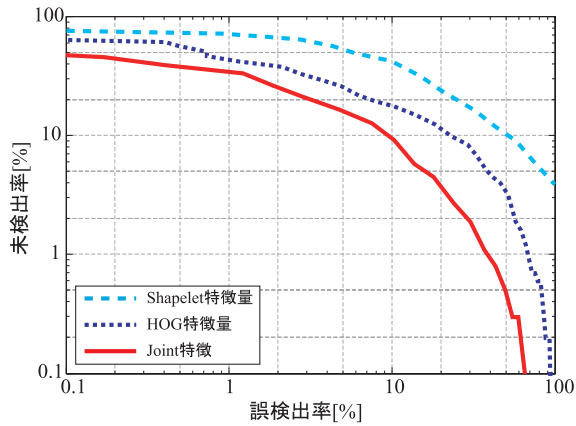
4.1.3 検出性能の評価

Joint 特徴の有効性を確認するために、人と車両を検出対象として評価実験を行った。実験結果を図 8 に示す。提案する Joint 特徴は、従来法である HOG 特徴量 [3] や Shapelet 特徴量 [5] と比較して、人と車両の両方のデータセットに対して高い検出性能であることがわかる。これは、HOG 特徴量のようなエッジ特徴のみでは識別困難なパターンにおいても、位置の異なる 2 つのセル内の HOG 特徴量を組み合わせることにより、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 9 に Joint 特徴を用いて人と車両の検出を行った例を示す。検出対象のスケールが異なる場合や、オクルージョンに対して頑健な検出ができていていることがわかる。

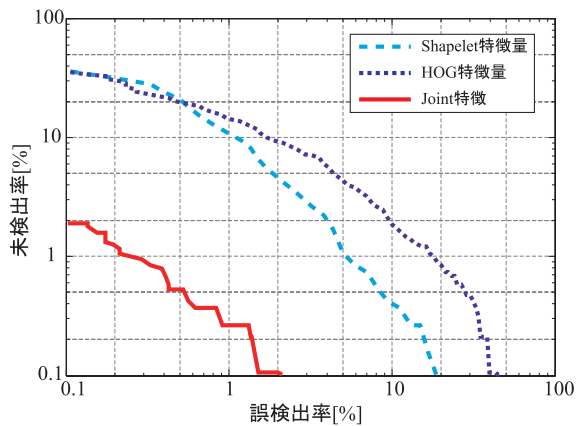
4.1.4 Joint 特徴の効果

Real AdaBoost による Joint 特徴の捉え方を図 4(b) に示す。HOG 特徴量と Real AdaBoost では、1 個の弱識別器が 1 個の HOG 特徴量を用いて識別したのに対し、Joint 特徴では、1 個の弱識別器が位置の異なる 2 つの領域内に含まれる複数の HOG 特徴量を用いて識別を行う。これにより、従来の単一の HOG 特徴量のみでは捉えることができない物体形状の対称性や連続的なエッジを自動的に捉えることができるため、高精度な人検出が可能となる。

図 10(a)、図 10(d) に 1 段階目の Real AdaBoost、図 10(b)、図 10(e) 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 10(c)、図 10(f) に、2 段階目の Real AdaBoost により選択された 2 つのセルと Joint 特徴をラウンドごとに示す。また、HOG 特徴量の勾配方向を人は 9 方向、車両は 18 方向で表現してお



(a)人



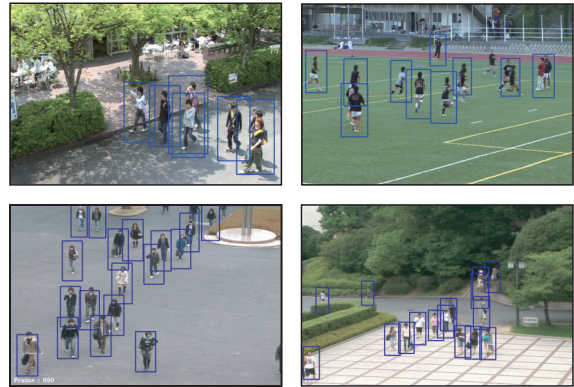
(b)車両

図 8: DET カーブ

り、輝度が高いほど Real AdaBoost における弱識別器の評価値が高く、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 10(b) では、図 10(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 10(c) に注目する。2 段階目の Real AdaBoost により選択された Joint 特徴は、人の輪郭に沿ったセルが選択されていることがわかる。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(d) では、1 段階目の Real AdaBoost の Joint 特徴候補により、車両内部の水平方向の勾配、車両の輪郭部分に沿った勾配が多く選択されている。次に図 10(e) では、図 10(d) で選択された HOG 特徴量の中から、2 段階目の Real AdaBoost により得られる最終識別器により、車両



(a)人



(b)車両

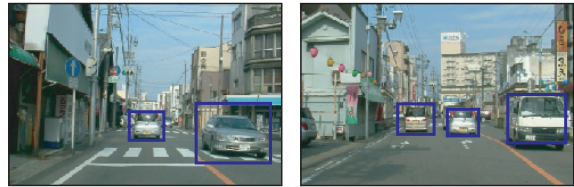


図 9: Joint 特徴による物体検出例

の輪郭部分に沿った Joint 特徴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わかる。最後に図 10(f) に注目する。学習の 1, 2 ラウンド目では垂直勾配と水平勾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捉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わかる。提案した Joint 特徴は、車両形状の対称性や連続性を捉える特徴量を予め用意しなくても、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。

4.2 多重解像度間の共起

人検出に有効な形状を Joint 特徴により捉えることで検出精度を向上させることができた。しかし、人の頭や肩などの形状は、特徴を最も最適に表現す

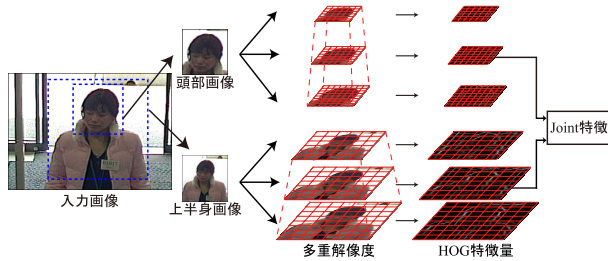


図 11: 多重解像度 HOG 特徴量の共起

ることができる解像度が同一であるとは限らない。そこで、入力画像の大きさを変えながら HOG 特徴量を抽出することで、異なる解像度間の共起を表現する。解像度を変化させることで、人の頭や肩などの部位毎で最適な解像度の HOG 特徴量を選択することができる。図 11 に、多重解像度間の共起表現の流れを示す。まず、入力画像から頭部と上半身にわけ、それぞれダウンサンプリングすることで多重解像度画像を作成し、HOG 特徴量を抽出する。そして、各セルより得られた HOG 特徴量を用いて Joint 特徴を計算する。

図 12(a), (b), (c) は Real AdaBoost の各ラウンドにおいて選択された HOG 特徴量の可視化画像、図 12(d) は各学習ラウンドにおいて選択された HOG 特徴量の累積を可視化した画像である。特徴選択の傾向が出やすい学習の初めでは、頭部は高解像度の HOG 特徴量が選択され、上半身は低解像度の HOG 特徴量が選択されていることがわかる。これは、学習サンプルの平均勾配画像からもわかるように、頭部の勾配はばらつきが少ないため高解像度の HOG 特徴量が選択され、上半身のように勾配が分散していると低解像度の HOG 特徴量が選択されるためである。低解像度画像からの HOG 特徴量は、広い範囲でヒストグラムをまとめることになるため、このばらつきを吸収することができたと考えられる。また、2 回目の弱識別器では左頭部と右の肩のライン、3 回目の弱識別器では右頭部と左の肩のラインを捉えている。これらのことから、多重解像度において対称性を捉えることにより、部分的なオクルージョンに影響を受けない特徴量が選択されている。

4.3 時空間特徴との共起

Joint 特徴のフレームワークでは、人のアピランスを表す HOG 特徴量 (図 13(b)) に、他の特徴を

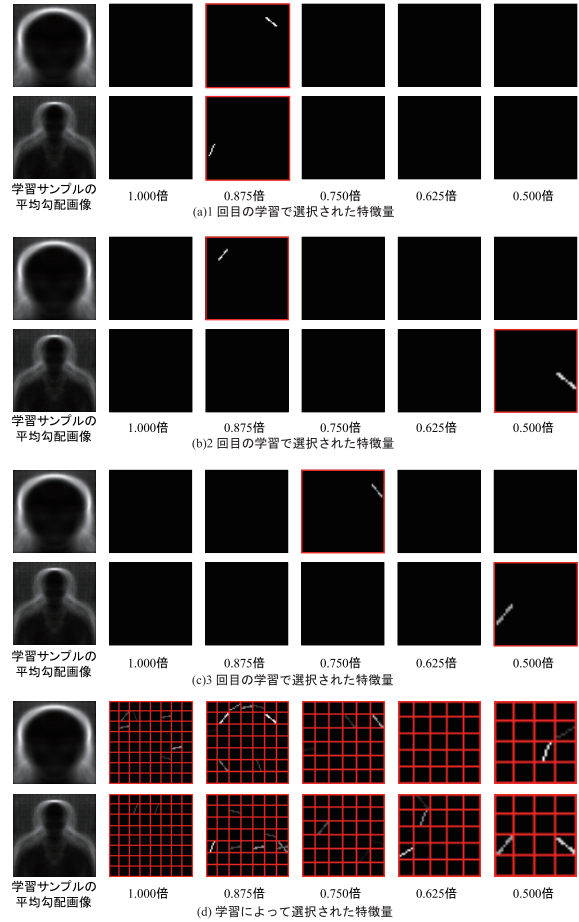


図 12: 選択された HOG 特徴量の可視化

追加することが可能である。我々は、動体検出に用いられてきた時空間特徴に基づく特徴量として図 13(c) に示すピクセル状態分析 (PSA) の結果を加えることにより、さらに高精度な人検出を実現した [10]。

4.3.1 ピクセル状態分析による時空間特徴量

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。前後 5 フレームの画像を用いて急激な輝度の変化 (Motion Trigger) と安定度 (Stability Measure) を算出し、各ピクセルを 3 状態に判別する。これにより、物体の動きを表現することができる。図 14 の T フレーム目の画像に注目した場合、1 枚の画像のみでは人が静止しているか歩行しているか判断できないが、前後数フレームの画像に着目することで、右足を軸に

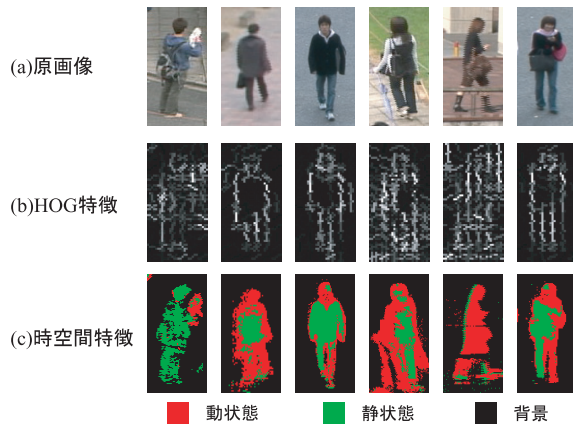


図 13: HOG 特徴とピクセル状態分析の例

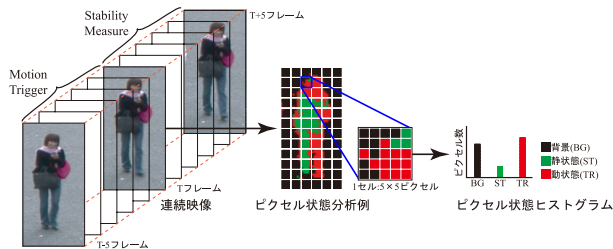


図 14: ピクセル状態分析と状態ヒストグラム

して左足を前に出していることがわかる。この連続映像を用いてピクセル状態分析を行うことにより、歩行中の動きを動状態と静状態により表現することが可能となる。

図 14 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。

4.3.2 時空間特徴との共起による効果

図 15 に、HOG と PSA の Joint 特徴による人検出の実験結果の DET カーブを示す。従来の HOG 特徴量に比べて Joint+PSA 特徴量は、誤検出率 5.0%において検出率を 99%以上にすることができた。図 16 に本手法による人検出の例を示す。アピランス特徴のみでは人に似た物体を誤検出しているが、時空間特徴量を加えることにより、誤検出を抑制することができる。

Real AdaBoost により選択された特徴量に着目するために、図 17 に各学習ラウンドにおける HOG

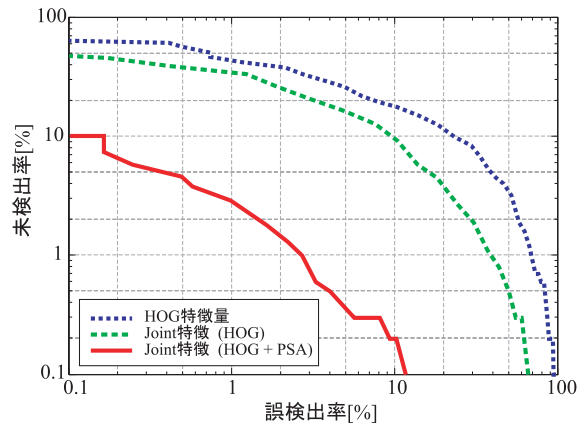


図 15: DET カーブ

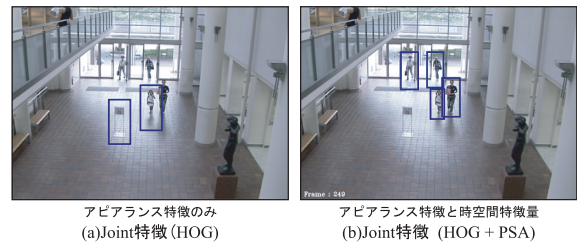


図 16: 時空間特徴量を用いた人検出による効果

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示す。初期ラウンドでは PSA 特徴量が多く選択され、学習ラウンド数が進むにつれて HOG 特徴量が選択される割合が多い。これは、まず物体の動きを表すことが可能な PSA 特徴量により、大まかに人と人以外を判別し、その後アピランスの情報をもつ HOG 特徴量を用いて、より細かな識別境界を形成していると考えられる。

4.4 デプス情報との共起

可視光カメラにより取得した画像から人の検出を行う場合、背景のテクスチャの複雑さによって、人検出に有効なアピランス情報を取得することが困難となる場合がある。そこで我々は、カメラから物体までのデプス情報を取得できる Time of Flight カメラ (TOF カメラ) を用いた人検出を提案している。TOF カメラとは、カメラの前面に付いた複数の LED より照射される赤外光が物体に反射し、カメラで観測されるまでの時間を計測することによ

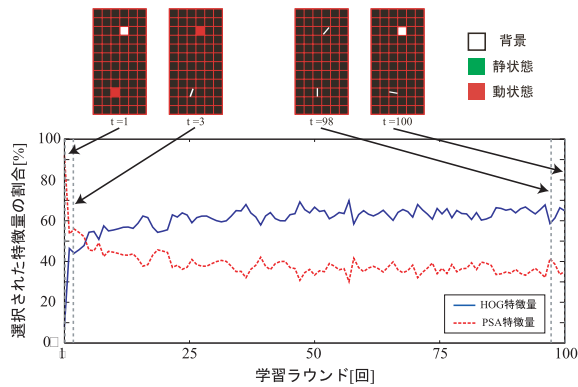


図 17: 選択された特徴量の割合

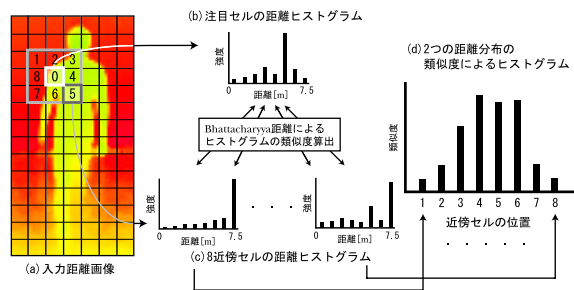


図 18: 距離ヒストグラムから得られる特徴量

り、物体までのデプスを取得するカメラである。この TOF カメラから得られるデプス情報と、アピランス特徴を同時に捉えることにより高精度な人の検出を行う。デプス情報を加えることにより、Real AdaBoost の弱識別器は物体と背景の距離関係と物体のアピランスを捉えることが可能となり、オクルージョンや背景の複雑さの影響を抑制することができる。

デプス情報から得られる特徴量は、図 18 に示すように、距離画像をセル分割し、注目セルの距離ヒストグラムと 8 近傍のセルの距離ヒストグラムとの Bhattacharyya 距離による類似度を算出し、各セルとの距離を特徴量とする。この特徴量は注目セルに対する近傍セルの相対的な距離関係を表現した特徴量となっている。図 19 にアピランス情報である HOG 特徴量と距離ヒストグラムから得られる特徴量の共起による人検出の流れを示す。これにより図 20 に示すように、人の重なりがある場合でも高精度な検出が可能となる。

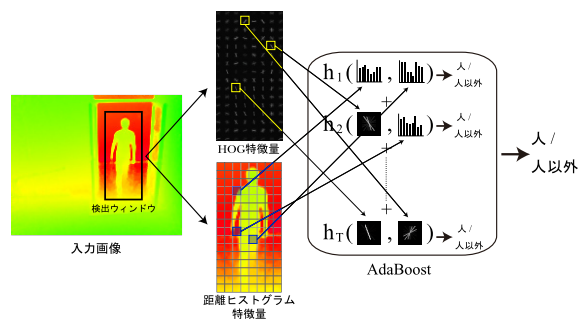


図 19: HOG 特徴量と距離ヒストグラム特徴量の共起

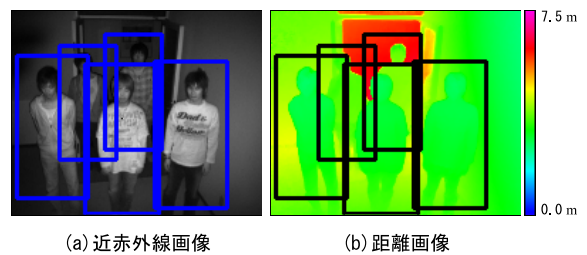


図 20: TOF カメラによる人検出例

5 おわりに

本稿では、図 21 に示すような複数の特徴量間の共起を用いた Joint 特徴による物体検出法について述べた。単一の HOG 特徴量では識別困難なパターンに対して、Joint 特徴は位置の異なる 2 つのセル内の HOG 特徴量を組み合わせることにより、識別困難なパターンを正しく識別することができる。人と車両の画像を用いた評価実験により、Joint 特徴は高精度な人検出が可能であることを示した。また、さらに高精度な検出を実現するために、時空間特徴である PSA 特徴量の共起や多重解像度画像から得られた HOG 特徴量間の共起、距離ヒストグラム特徴量との共起の効果について述べた。今後は、ブースティングにおける学習の効率化について検討する予定である。

参考文献

- [1] P. Viola and M. Jones: "Robust Real-Time Face Detection", IJCV, 57(2), pp. 137-154, (2004).
- [2] K. Levi and Y. Weiss: "Learning Object Detection from a Small Number of Examples: the Importance of Good Features.", CVPR, vol. 2, pp. 53-60, (2004).

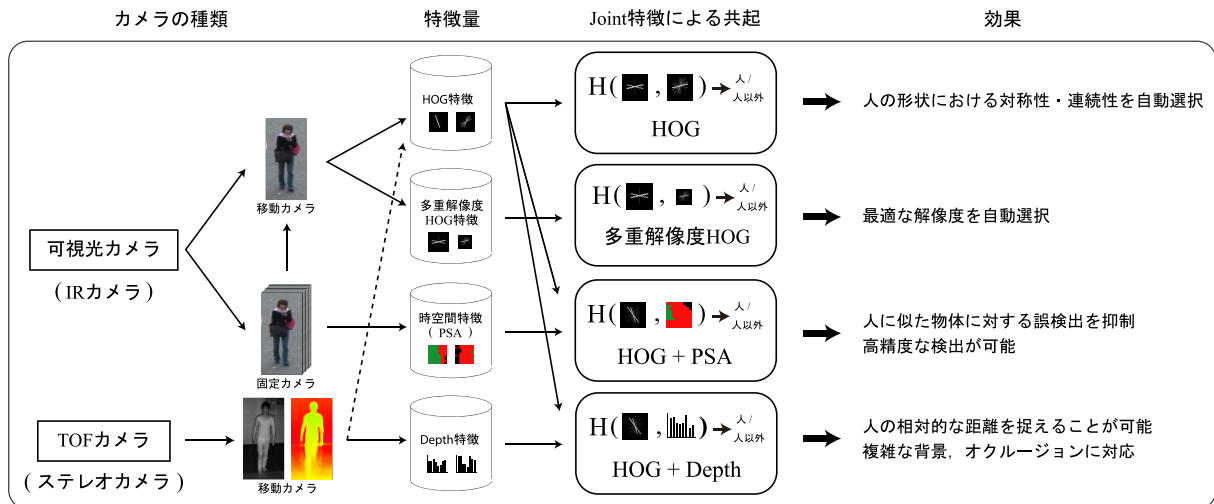


図 21: Joint 特徴による共起特徴表現

- [3] N. Dalal and B. Triggs: "Histograms of Oriented Gradients for Human Detection", CVPR, pp. 886-893, (2005).
- [4] B. Wu and R. Nevatia: "Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors", ICCV, vol. 1, pp. 90-97, (2005).
- [5] P. Sabzmeydani and G. Mori: "Detecting Pedestrians by Learning Shapelet Features", CVPR, pp. 1-8, (2007).
- [6] T. Mita, T. Kaneko, B. Stenger and O. Hori: "Discriminative Feature Co-occurrence Selection for Object Detection", PAMI, vol. 30, no. 7, pp. 1257-1269, (2008).
- [7] 山下隆義, 藤吉弘巨 "特定物体認識に有効な特徴量", 情報処理学会 研究報告 CVIM 165, pp. 221-236, November, 2008.
- [8] H. Schneiderman and T. Kanade: "A statistical method for 3d object detection applied to faces and cars", CVPR, pp. 746-751, (2000).
- [9] 三井相和, 山内悠嗣, 藤吉弘巨, "Joint HOG 特徴を用いた 2 段階 AdaBoost による人検出", 第 14 回画像センシングシンポジウム SSII08, IN1-06, (2008).
- [10] 山内悠嗣, 藤吉弘巨, Bon-Woo Hwang, 金出武雄: "アピ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), (2007).
- [11] C. R. Wang and J. J. Lien: "AdaBoost Learning for Human Detection Based on Histograms of Oriented Gradients", ACCV, p. 485, 2007.
- [12] F. Han, Y. Shan, R. Cekander, H. S. Sawhney and R. Kumar: "A two-stage approach to people and vehicle detection with HOG-based SVM", Workshop on PMIS, pp. 133-140, (2006).
- [13] 山内悠嗣, 藤吉弘巨, 山下隆義 "Boosting に基づく共起表現による人検出", 画像の認識・理解シンポジウム (MIRU2008), pp. 180-187, 2008.
- [14] 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.
- [15] R. E. Schapire and Y. Singer: "Improved Boosting Algorithms Using Confidence-rated Predictions, Machine Learning," No.37, pp.297-336, 1999.
- [16] H. Fujiyoshi and T. Kanade: "Layered detection for multiple overlapping objects", IEICE Transactions on Information and systems, pp. 2821-2827, (2004).