社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

[サーベイ論文] 統計的学習手法による人検出

山内　悠嗣† 　山下　隆義†† 　藤吉　弘亘 †††

†, ††† 中部大学　〒 487–8501 愛知県春日井市松本町 1200
†† オムロン株式会社　〒 525–0025 滋賀県草津市西草津 2 丁目 2-1
E-mail: †yuu@vision.cs.chubu.ac.jp, ††takayosi@omm.ncl.omron.co.jp, †††hf@cs.chubu.ac.jp

あらまし　物体検出とは，画像中から物体が存在する位置と大きさを自動的に求める技術である．物体検出は古くから取り組まれてきた顔検出がベースとなっており，近年では顔画像から多様な見えの変化が生じることから検出が困難とされている人画像に研究対象が遷移している．こうした中で，近年では人検出を難しくする要因を克服するような手法が数多く提案されている．そこで，本稿では人検出を難しくする要因を整理し，この要因を克服するための特徴抽出と統計的学習手法による識別器の二つの観点から手法をサーベイする．また，人検出法を定量的に評価するために利用されている統一的な評価指標と人画像データベースについても紹介する．

キーワード　サーベイ，人検出，特徴量，統計的学習手法

[Survey paper] Human Detection Based on Statistical Learning

Yuji YAMAUCHI†, Takayoshi YAMASHITA††, and Hironobu FUJIYOSHI†††

†, †††Chubu University　1200 Matsumoto-cho, Kasugai, Aichi, 487–8501 Japan
†† Omron Corporation　2-2-1 Nishikusatsu, Kusatsu, Shiga, 525–0025 Japan
E-mail: †yuu@vision.cs.chubu.ac.jp, ††takayosi@omm.ncl.omron.co.jp, †††hf@cs.chubu.ac.jp

Abstract　Object detection is detecting and localizing generic in an image. In object detection, the basis is face detection, which has been researched since early times. In recent years, the detection target has changed to the human image in various different appearances. Under these circumstances, a lot of methods have been proposed for resolving the factors that complicate detecting humans. In this paper, we discuss the factors that complicate human detection and survey human detection methods from the viewpoint of two approaches, feature extraction and classification by statistical learning, to overcome these factors. In addition, we summarize the evaluation methodologies and image databases that spurred development of human detection.

Key words　Survey, Human detection, Feature, Statistical learning

1. はじめに

人々の生活の利便性向上や安心・安全な社会の実現に向けて，人を観る技術が必要とされている．人を観る技術とは，画像から人の位置の特定や人の追跡，人の動作を認識する技術である．特に，人検出は映像中から人の位置を特定するために必要な技術であり，人物追跡や動作認識を実現するためには前処理として必要不可欠である．

人検出技術は，古くから研究されてきた顔検出技術をベースとしている．顔検出の研究は，1969 年の Sakai らの研究 [1] が始まりとされ，以来多くの研究者によって精力的に取り組まれてきた．顔検出の初期の研究では，研究者が目や口の濃淡関係や位置関係等の予め決めたルールに従って顔と顔以外の画像を判定する手法 [2]〜[4] が主流であった．これらの方法は，顔を低次元な特徴量ベクトルで表現し，比較的簡単な処理で顔と背景を判別できる．しかしながら，ルールを決めるには研究者の熟練した知識が必要となる．また，あらゆる環境に対応できる複雑なルールを作成することが難しいため，汎用性の高い検出器を作成することが困難であった．

1990 年代後半に入ると汎用コンピュータの進化に伴い，大量のデータを高速に処理できるようになったことから，画像から高次元の特徴量ベクトルを抽出し，統計的学習手法による識別する手法 [5]〜[8] が主流となった．統計的学習手法は，クラスラベル付きの大量の学習サンプルを必要とするが，ルールベースの手法のように研究者がルールを作成する必要がないため，汎用性の高い識別器を学習できる．統計的学習手法を利用した代表的な顔検出法には，Neural Network を用いた手法 [6] や SVM を用いた手法 [9]，Naive Bayes に基づく手法 [7],

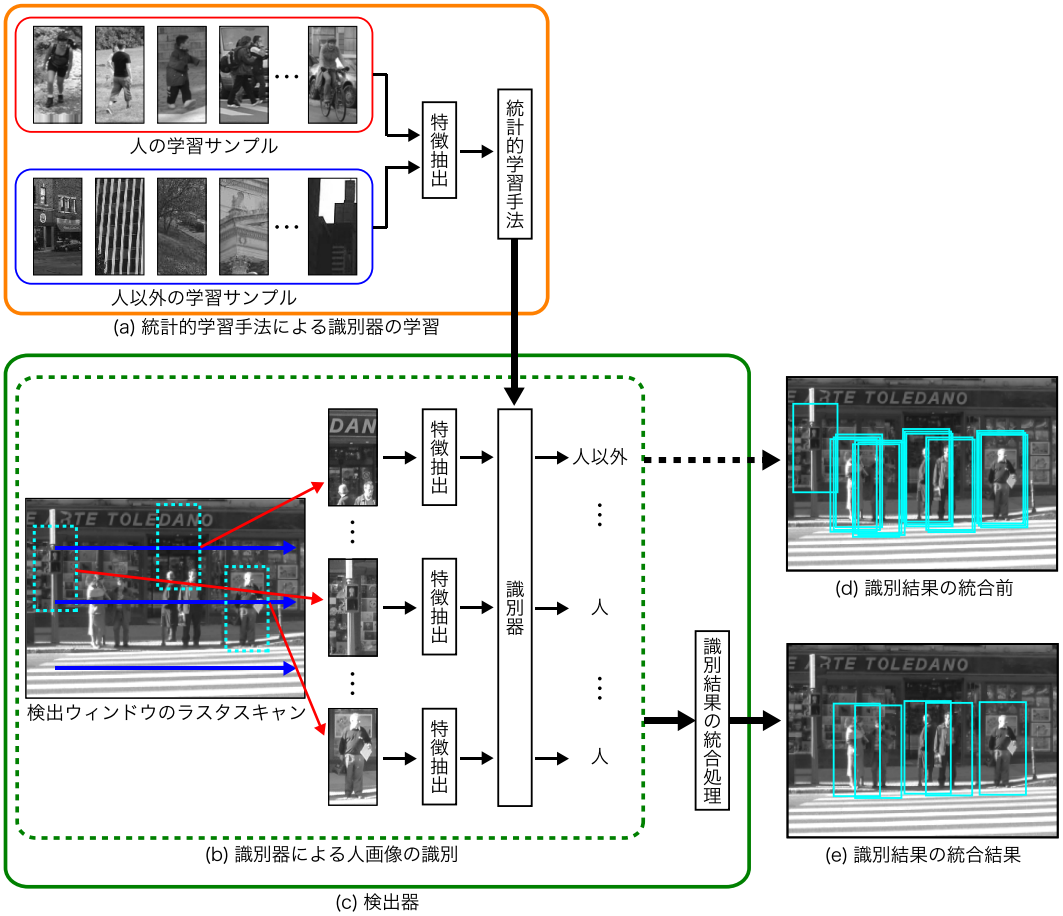


図 1　学習と検出の流れ.

AdaBoost による手法 [8], [10] があり，これらの顔検出法は高い検出性能を実現した．中でも，2001 年に Viola と Jones によって提案された顔検出法 [8], [10] は，高精度かつ高速な顔検出を実現したことから，以降の物体検出の研究に大きな影響を与えた．

高速かつ高精度な顔検出法が確立されると，検出対象がより困難とされる人に遷移した．人検出は，顔検出の研究から得られた知見や技術を引き継ぎ，画像局所特徴量と統計的学習手法に基づくアプローチが一般的である [11], [12]．しかしながら，人の画像は顔の画像と比べると，衣服や体格，向き，姿勢などの個人差，視点の変化，照明の影響により，検出対象である人の見えの変化が大きくなる．そのため，多数の人画像に共通する特徴を抽出することが難しくなり，その結果，正しく人を検出することが難しくなる．この問題を解決するために，現在までに画像局所特徴量と識別器の処理において人の見えの変動を吸収するような様々な工夫が試みられている．

上記を踏まえ，本稿では人検出を難しくする要因について整理した上で，要因を解消するためのアプローチ毎に手法を分類し，解説する．2 章では，人検出の基本的な流れについて述べ，人検出を難しくする要因について整理する．各要因について，3 章では画像局所特徴量により吸収するアプローチ，4 章では識別器を工夫して悪影響を低減するアプローチについて述べる．5 章では，人検出を実用化する際に発生する問題とその解決するための手法について述べる．6 章では，人検出器の性能を測るためのデータベースと評価指標について述べる．最後に 7 章にてまとめる．

2. 統計的学習手法による人検出と課題

人検出は顔検出の研究から得られた知見や技術を引き継ぎ，画像局所特徴量と統計的学習手法に基づく手法が一般的である．本章では，画像局所特徴量と統計的学習手法に基づく人検出のフレームワークについて述べる．そして，人検出を難しくする要因について整理する．

2.1 人検出の学習と識別

人検出は，識別器を学習する処理と学習した識別器により人を検出する処理の 2 つの処理に分けられる．本節では，学習と検出の各処理について簡単に述べる．

2.1.1 識別器の学習

一般的な人検出法は，統計的学習手法に基づき学習した識別器により画像を人と人以外に判定する．図 1(a) に識別器を学習する流れを示す．統計的学習手法により識別器を学習するために，事前に人と人以外の大量の学習サンプルを用意する．そして，これらの学習サンプルから特徴量を抽出し，統計的学習手法により識別器を学習する．

2.1.2 識別器による人検出

未知入力画像から人を検出するには，図 1(b) に示すように

表 1　人検出を困難とする要因と対応策.

要因	対応策		代表的な手法
	特徴量	識別器	
見えの個体差	勾配，色，動き，距離	-	HOG [13]，CSS [14]，HOF [15]
複雑な背景	特徴量の共起	-	Joint Haar-like [16], CoHOG [17]，Joint HOG [18]
向きの変化	-	複数の識別器	Cluster Boosted Tree [19]
姿勢の変化	-	パーツベース	Deformable Parts Model [20], Hough Forest [21]
人領域の隠れ	-	隠れの推定	グローバル及びパーツベース識別器の組み合わせ [22]
視点の変化	-	幾何学情報の利用	シーンのモデリング [23], 転移学習 [24]

未知入力画像中に検出ウィンドウを設定し，画像中を走査しながら人か背景かを区別する．このとき，人とカメラの距離に応じて人の大きさが変化するため，このような人の大きさの違いに対応するためには，1 枚の画像を多重解像度で表現したピラミッド画像からラスタスキャンする，もしくは幾つかの検出ウィンドウの大きさを設定し，それぞれの大きさの検出ウィンドウにより画像中を走査する．これにより，図 1(d) に示すような結果を得ることができる．

人領域周辺には，人として識別された多数の検出ウィンドウが存在する．また，人が存在しない領域にも人と誤識別した検出ウィンドウが存在する場合がある．人検出の結果を用いて，例えば人を追跡する場合や人の入退出数をカウントする際には，1 人に対して一つの検出ウィンドウを出力する必要がある．また，人が存在しない領域を人と誤って識別した検出ウィンドウは，最終的な人検出結果に反映させないことが望ましい．そこで，検出ウィンドウの統合処理を行い，図 1(e) に示すような人検出結果を得る．検出ウィンドウの統合には，計算コストの低さから Mean Shift [25] が用いられることが多い．

2.2　人検出を難しくする要因

顔検出の研究から得られた知見や技術を引き継ぎ，人検出においても画像局所特徴量と統計的学習手法に基づく手法が一般的である．しかしながら，顔画像と比べて人画像は多様な見えの変化が生じることから，検出がより困難となる．人の検出を難しくする原因は，人画像の見えの変動が大きく，人と背景を区別するための情報が不足していることである．そのため，見えの変動が大きい人画像より抽出された特徴量から，多くの人画像に共通する要素を統計的学習手法で見つけることが難しくなり，その結果，人と背景を正確に区別する識別器を学習することが困難となる．

人画像の見えの多様性を生じさせる要因は，表 1 に示すように 6 つに分けることができる．以下に人検出を難しくする各要因について述べる．

- 見えの個体差

人の衣服や体格などの個人差により見えが異なる．特に，衣服は様々な色や模様であるため，人検出の特徴量として利用することが難しい．さらに，体格は大人と子供，性別により異なるため，頭部や胴体，足等の見えや位置が異なる．

- 複雑な背景

背景画像が複雑なテクスチャを含むことがある．特に，人画像のように連続した縦エッジを含む背景画像は，部分的な領域のみを見る場合は人画像に似る．

- 向きの変化

人を正面から撮影する場合と横から撮影する場合では人の見えが異なる．

- 姿勢の変化

人は非剛体な物体であり，自由な姿勢をとることができる．そのため，姿勢により人の形状が大きく変化する．

- 人領域の隠れ

人とカメラの間に物体が存在する場合，画像上の人領域が隠れるため部分的に観測できなくなる．そのため，部分的に欠損した人画像から人を検出することになる．

- 視点の変化

人の向きの違いと似ているが，ここではカメラの俯角の違いによる人の見えの違いを表わす．人を正面から撮影した人画像と斜め上から撮影した人画像では，人の見えは大きく異なる．

このような人検出を難しくする各要因に対して，表 1 に示すように特徴量や識別器を工夫することで高精度な人検出を実現している．以降では，人検出法を特徴抽出法に焦点を当てた手法と識別手法に焦点を当てた手法の 2 つに分類し，それぞれの手法について概説する．

3.　特徴抽出法

特徴抽出法に焦点を当てた手法においては，人の見えの個体差を吸収しつつも，多くの人に共通する人らしい特徴を捉えるための特徴量が数多く提案されている．これらの特徴量は，人の局所領域を捉える特徴量と局所領域間の共起性を捉える特徴量の 2 つに分けることができる．前者は人の見えの個体差を吸収しつつも万人に共通する特徴を捉え，後者は人と人に似た見えを持つ背景との違いを捉えるために人の構造に着目する．以下に，それぞれの特徴量について述べる．

3.1　人らしさを捉える局所特徴量 (見えの個体差に対応した手法)

人画像は，衣服や体格などの個体差により見えが大きく異なる．高精度な人検出を実現するには，人画像の見えの変動を小さくする必要があるため，これらの個体差に対してロバストな特徴量を設計する必要がある．そのために多くの特徴量が提案されており，これらの特徴量は特徴抽出における着目する情報の違いから下記の 4 つに分けることができる．

- 勾配情報に基づく特徴量
- 色情報に基づく特徴量

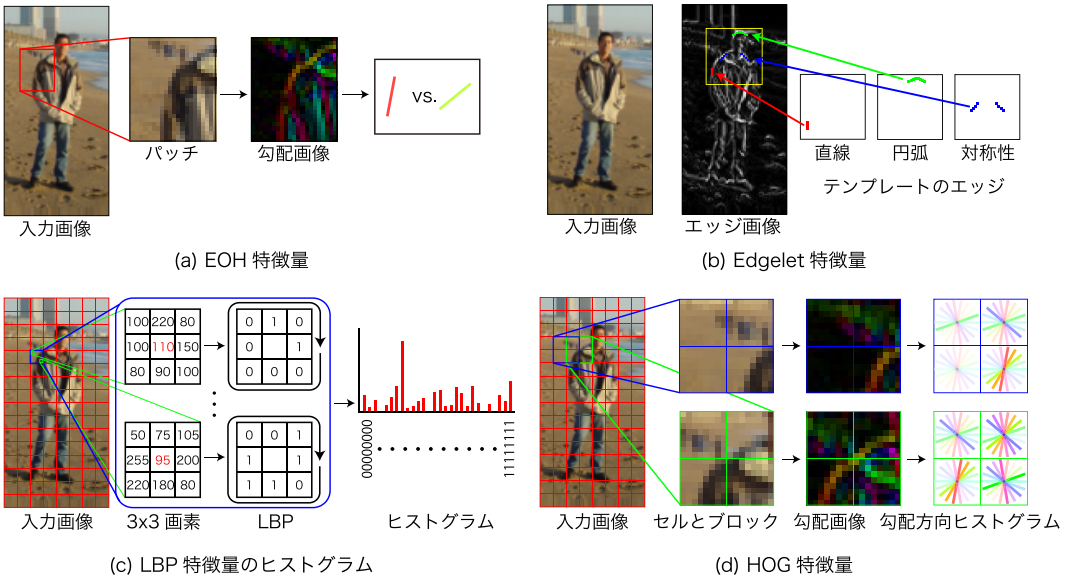


図 2　勾配情報に基づく特徴量.

- 動きに基づく特徴量
- 距離に基づく特徴量

以下に各特徴量について述べる.

3.1.1　勾配情報に基づく特徴量

顔検出では，目や鼻，口などの明暗差を捉える特徴量 [8], [26] が利用されている．人検出の場合では，人の衣服や背景の色が変化するため，輝度に着目した特徴量では人に共通する要素を捉えることが難しい．そこで，人検出に有効な特徴量として，人の形状を捉えるために輝度の勾配に着目した特徴量が提案されている [13], [27], [28]．Chen らは，勾配に着目した特徴量として Edge of Orientation Histograms(EOH) 特徴量 [27] を提案している [29]．EOH 特徴量は，図 2(a) に示すような局所領域の累積したエッジ強度の頻度比で表わされる．Wu らは，図 2(b) に示すようなエッジの直線的な繋がりや円弧及びそれらを対称的に捉える Edgelet 特徴量 [28], [30] を提案している．他にも，図 2(c) のように注目画素と近傍画素の輝度の大小関係により 2 値符号列化する Local Binary Pattern(LBP) 特徴量 [31] を利用した人検出法も提案されている [22], [32]〜[34].

勾配に着目する特徴量が多数提案されているが，その中でも Dalal らが提案した Histograms of Oriented Gradients(HOG) 特徴量 [13] が最も利用されている．HOG 特徴量は，局所領域 (ブロック) 毎に作成される輝度の勾配方向ヒストグラムを線形結合した多次元の特徴量として表される．ブロックは，さらに小さな領域 (セル) から構成され，複数のセルにより正規化されている．これにより，照明の変動に対して頑健な特徴量となる．また，局所領域において勾配情報をヒストグラム化することから，多少の位置ずれや回転に対しても頑健となるため，人の体格の個体差も吸収できる．なお，HOG 特徴量のように勾配をヒストグラム化する過程は，1987 年に文字認識に用いられた加重方向指数ヒストグラム [35] と非常に似ている.

HOG 特徴量は，簡単な処理ながらも優れた人検出性能を達成したため，人検出法のデファクトスタンダードな特徴量 [14], [15], [20], [22], [36] として利用されている．また，HOG 特徴量の正規化処理を簡単化することで高速化した Extended HOG(EHOG) 特徴量 [37] や異なる大きさの局所領域から HOG 特徴量する Pyramid HOG(P-HOG) 特徴量 [38]，色情報から前景と背景のソフトセグメンテーションの考えを導入した Color-HOG(C-HOG) 特徴量 [39]，エッジ強度の相関を考慮した Edge Similarity-based-HOG(ES-HOG) 特徴量 [40] も提案されている.

3.1.2　色情報に基づく特徴量

人検出には，形状を捉えることができる勾配ベースの特徴量が利用されることが多いが，色情報に基づく特徴量と併用することにより人検出性能の高精度化を図る手法も提案されている.

Dollar らは，領域の輝度和や勾配和を高速に求められる積分画像 [8] や積分ヒストグラム [41] のメリットを生かし，勾配方向ヒストグラムや輝度，LUV 表色系により表現される色などを用いた高速な人検出法 [42] を提案している．しかしながら，色は衣服に左右されるため，このように陽に利用することは少ない.

そこで，色そのものを特徴量として使用せず，色の類似度を特徴量とする手法 [14] が提案されている．Walk らは入力画像における 2 つの局所領域の色ヒストグラムの類似度を Color Self-Similarity(CSS) 特徴量として人検出に利用している．色の類似度を利用することで，2 つの局所領域が同一の部位であるかを陰に表現することが可能である．例えば，図 3(a) は CSS 特徴量を可視化した画像であり，輝度が高いほど選択した矩形領域との色の類似度が高いことを表わしている．このように，人の胴体や足などの同一パーツの領域内においては高い類似度が得られていることがわかる．他にも，CSS 特徴量と同様の方法で色の類似度を計算した後に，各画素の類似度から求めた HOG 特徴量を利用する CS-HOG 特徴量 [43] も提案されている．CSS 特徴量は色情報そのものに対する依存性はなく，人検出において扱いにくい色を有効に利用した特徴量といえる.

図 3　CSS 特徴量 [14] と前景尤度マップ [44] の可視化画像の例.

3.1.3　動き情報に基づく特徴量

画像の見えに基づく特徴量のみでは検出性能に限界があるため，動き情報に基づいた特徴量を加えた人検出法が提案されている．動き情報に基づいた特徴量は，動きの性質から特徴量を下記の 2 つに分けることができる．一つ目は，背景差分により物体領域を抽出することで画像内の空間的変化を捉える特徴量である．二つ目は，フレーム間差分により画像内の時間的変化を捉える特徴量である．

空間的変化を捉える特徴量 [44] は固定カメラを想定している．背景差分を行うことで前景領域のみを抽出できるため，前景領域を人と人以外に判別する簡単化した問題となる．Yao らは，背景差分をベースとした手法 [45] から得られる前景の尤度を特徴量として利用している [44]．図 3(b) からも分かるように，背景差分を行った後の人画像は人の輪郭を表現すると同時に，各画素が人体に含まれているかを表わすことができる．人体にはドーナツのような穴は存在しないため，人の輪郭内に存在する画素を捉えることができれば，人検出に有効な特徴量として利用できる．

時間的変化を捉える特徴量はフレーム間差分やオプティカルフロー，STpatch をベースとした手法が提案されている [12], [15], [46]．Viola らは，2 フレーム間の差分画像と Haar-like をベースとしたモーションフィルタにより，人の輪郭や移動方向，動きに関する特徴量を抽出し，見えに基づく特徴量と組み合わせることで高精度な人検出を実現している [12]．Dalal らは 2 フレーム間の画像から得られるオプティカルフローをベースとした動きの特徴量を提案している [15]．カメラと物体の距離に応じて観測される画像上での動きベクトルを局所領域においてヒストグラム化し，近隣の領域で作成したヒストグラムとの差分を抽出する HOF(Histogram of Flow) 特徴量を提案している．Dalal の手法は，相対的な動きを観測するため，移動カメラでも人検出に有効な動きの特徴量を得ることができる

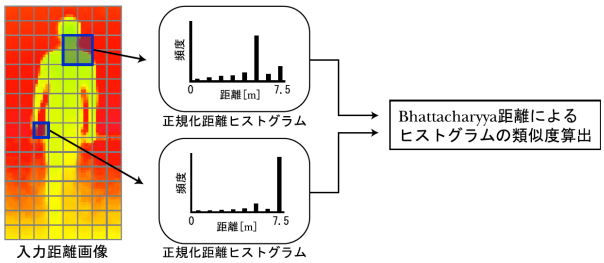


図 4　距離ヒストグラムの類似度を計算する流れ [50].

が，カメラ及び人が動いていない状態では識別に有効な動きの特徴量を得られない問題もある．また，村井らは STpatch 特徴量 [47] を用いた人検出法を提案している [48]．STpatch 特徴量は，局所領域における見えと動きの時間的変化を表現するため，人の見えに加えて人の移動方向と移動量を捉えることができる．他にも，空間的変化と時間的変化を同時に捉える特徴量を用いた人検出法 [49] も提案されている．

3.1.4　距離情報に基づく特徴量

近年，簡単に距離画像を取得できるデバイスが開発されたことを受け，距離情報を用いた高精度な人検出を目的とした手法が提案されている．池村らは，TOF カメラより得られる距離画像から，図 4 に示すような背景と人の距離差を捉える Relational Depth Similarity Feature(RDSF) 特徴量による人検出法 [50] を提案している．図 4 に示すように 2 つの局所領域から得られる距離ヒストグラムの類似度を RDSF 特徴量として抽出している．また，Shotton らは距離画像から 2 点間の距離差分を特徴量としている [51]．Xia らは，距離画像を用いて Chamfer Matching による人の頭部領域の推定と，頭部の 3D モデルとのフィッティングにより人を検出している [52]．TOF カメラや Kinect から得られる距離情報は，各部位までの距離を利用できる他，可視光カメラとは異なり物体や背景のテクスチャの影響を受けない．そのため，人の形状を明確に捉えることが可能であり，可視光カメラを使用する人検出法よりも高精度に人を検出することができる．

3.2　人の構造を捉える特徴量 (複雑な背景に対応した手法)

人の形状を捉えるために適した特徴量を用いたとしても，人に似た形状を持つ複雑なテクスチャを持つ背景画像を正しく識別することは容易ではない．より高精度な検出を実現するためには，例えば，頭から肩の Ω のような人の構造を捉える必要がある．しかしながら，予め構造的な形状を捉えるような特徴量を設計するにしても，顔検出の初期の研究のように研究者の経験に基づくルールベースの手法では，汎用性の低い手法となる問題がある．

そこで，人の構造を捉える手法として，複数の特徴量の共起性を表現する手法が提案されている．この手法は，複数の局所領域を同時に観測することで，局所特徴量の関係性を捉える．例えば，肩の斜めエッジと胴体の横エッジを同時に観測し，各領域から観測されるエッジの共起性を考慮して識別する．このような特徴量の共起に着目した手法は，統計学習を利用する手法としない手法の二つに分けることができる．以下に各手法に

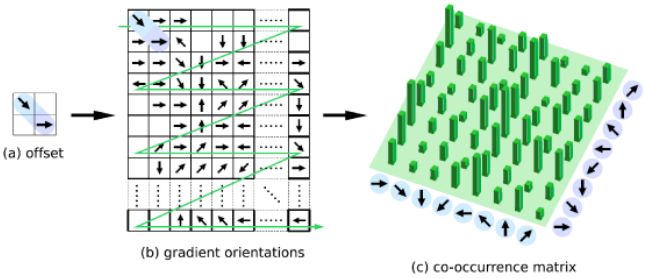


図 5　CoHOG 特徴量の算出の流れ [53].

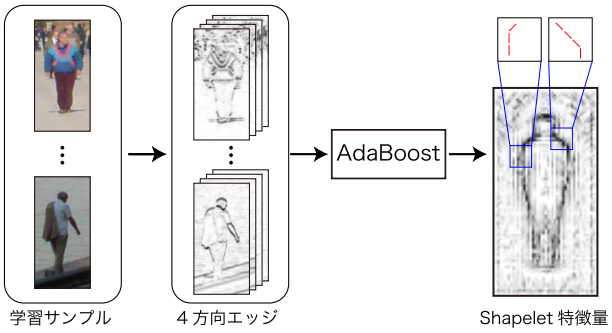


図 6　Shapelet 特徴量の生成の流れ [57].

ついて述べる．

3.2.1　統計的学習手法を用いない特徴量の共起

統計学習を用いない共起表現として，Watanabe らが提案した Co-occurrence Histograms of Oriented Gradients (CoHOG) 特徴量 [17], [53] のような同時生起行列を用いたアプローチがある．CoHOG 特徴量は，図 5 に示すように，ある局所領域における 2 点の画素から得られる勾配方向の組み合わせを累積した同時生起行列である．局所領域において勾配のペアの出現頻度を表現している．このアプローチを改良し，ステレオ画像から得られる距離情報を利用した手法も提案されている [54]．また，同様のアプローチで Local Binary Pattern(LBP) 特徴量 [31] の共起を表現した手法も提案されている [55]．Tuzel らは，画素の位置や輝度勾配等を特徴ベクトルとし，局所領域の各画素から計算されたベクトルを分散共分散行列で表現し，これを特徴量として利用している [56]．分散共分散行列は対角成分が分散，それ以外は共分散を表わすため，特徴ベクトルを構成する各要素との相関関係を表わす．

3.2.2　統計的学習手法を用いた特徴量間の共起

3.2.1 で述べた手法は，予め決められた特徴量の組み合わせにより特徴量の共起を表現し，それら共起を表現した全ての特徴量により識別器を学習する．そのため，共起を表現したい局所特徴量の次元数が高い場合には，組み合わされる特徴次元数が膨大となるため，学習することが困難となる．そこで，二つ目の手法として，共起を表現した特徴量を統計的学習手法により評価することで，人の識別に有効な特徴量を組み合わる手法 [16], [18], [57]〜[59] が提案されている．

三田らは，検出対象の構造に基づいた共起関係を表現する Joint Haar-like 特徴量 [16] を提案している．この方法は，複数の Haar-like 特徴量を閾値処理することで 2 値化する．そして，それら 2 値符号を組み合わせることで生成した Joint Haar-like 特徴量を AdaBoost により評価する．また，特徴量を 2 値化する以外にも，特徴量からクラス尤度を計算し，得られた複数の尤度を演算子により共起する手法 [60] も提案されている．

Sabzmeydani らは，局所領域内の 4 方向エッジを AdaBoost により組み合わせることにより特徴量の共起を表現する Shapelet 特徴量 [57] を提案している．Sabzmeydani らの手法は，2 段階の AdaBoost により識別器を構成している．1 段階目の AdaBoost では，図 6 に示すように局所領域における各画素の 4 方向のエッジから識別に有効なエッジを選択することで Shapelet 特徴量を生成する．そして，2 段階目の AdaBoost では，生成した Shapelet 特徴量から特徴量を選択することで識別器を学習する．AdaBoost により識別に有効な Shapelet 特徴量を選択することで，より人の形状を捉える識別器が学習できる．また，Joint Haar-like 特徴量の特徴量間の共起性と Shapelet 特徴量の局所領域内の勾配の共起性の両方を捉える Joint HOG 特徴量 [18] も提案されている．これらの手法は，識別に有効な特徴量の組み合わせを統計的学習手法により評価することで，人検出に有効な共起を自動的に表現できることから，高精度な人検出を実現している．

4.　識 別 手 法

3. 章にて述べた特徴抽出法では人の個体差を吸収できる一方，人の向きや姿勢，人領域の隠れ，視点の違いによる人の見えの変動を特徴量のみで吸収するには限界がある．そこで，これらに対して頑健な検出を実現するために，識別器を工夫する手法が採られる．以下に，それぞれの要因を解決するための手法について述べる．

4.1　向きの変化に対応した手法

人は 3 次元物体であるため，人を正面から撮影した場合と横から撮影した場合では見えが異なる．そのため，人という同一クラスでありながら，クラス内での見えの変動が大きく異なる．この変動を小さくするためには，例えば，正面や横向きに対応した各々の識別器を学習し，各々の識別器により人を検出すれば良い．このような考え方は，顔の回転や向きに対応した顔検出の研究において採用されており，同様のアプローチを人検出でも適用している．

正面顔の回転に対応する代表的な手法として Rowley らの顔検出法 [61] がある．この手法は，まず顔がどれほど回転しているかを推定してから，その角度に合わせて顔画像を回転させることで正面顔に変換する．多くの手法が顔の回転や向きに対応するためにこのアプローチを採用している [62], [63]．しかしながら，Rowley らの手法に代表されるような顔の回転や向きを推定した結果に応じて識別する手法は，回転と向きの推定に失敗した場合は次段の識別処理も失敗する問題がある．この問題を解決するために，回転や向きに対応した複数の識別器を用意し，これら識別器を階層的に構築する手法 [37], [64], [65] が提案されている．

ここまでに取り上げた手法は，検出対象の回転や向きに対応

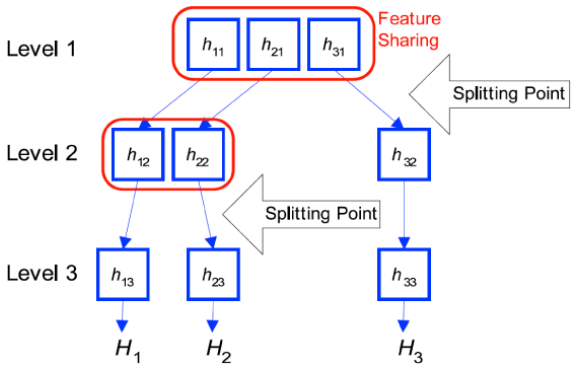


図 7　Cluster Boosted Tree の構造の例 [19].

することができる一方，回転や向きの情報を事前に研究者が付与する必要がある．そのため，学習サンプルに対して人と背景のラベル以外に，回転角と向きの情報を与えなければならない．この問題を解決するために，Wu らは自動的に人の向きに対応した手法として Cluster Boosted Tree(CVT) [19] と呼ぶ階層的な識別器を提案している．CVT は，図 7 のように識別器 h を木構造で表現することで人の向きに対応した識別器を構築する．この手法は，識別器を学習する過程において学習誤差が収束状態に陥ると k-means 法により学習サンプルを分割する．これにより，学習サンプルが人の向きに合わせて分割されて識別器が構築されるため，人の向きに対応した識別器を自動的に学習することができる．同様に，土屋らも自動的に人の向きに対応した識別器を学習する方法 [66] を提案している．土屋らの手法は，ソート問題で利用される分割統治法のアプローチに基づき，予め統計的学習手法により識別した結果から学習サンプルを分割し，多クラス分類問題を扱う Joint Boosting をベースとした改良型 Joint Boosting により識別器を学習する．両手法共に，研究者の知識によって向きをクラスタリングするわけではなく，学習結果により学習サンプルをクラスタリングする．そのため，人手で向きに対するラベルを付与する必要がない利点がある．また，識別結果に応じてクラスタリングするため，必ずしも検出対象の向きに対応した識別器が学習されるとは限らないが，識別に適した学習サンプルの分割が可能であることから，人手でクラスタリングした場合よりも高い検出性能を得ることができる．

4.2　姿勢の変化に対応した手法

人間は複雑な姿勢をとることができるため，同一クラスの複数の人画像であっても見えの変動が大きくなる．そのため，特徴量から人に共通した要素を見つけることが難しくなり，正しく検出することが困難な場合がある．そこで，人体のパーツ (頭部や胴体，手，足など) を検出し，各パーツに対応した識別器の結果を統合することで人を検出するパーツベースの手法が提案されている．

パーツベースの人検出法は，パーツを検出する処理とパーツの識別結果を統合する処理で構成おり，姿勢の変動に頑健な人検出を実現するために重要な点は，パーツの定義とパーツの統合方法である．パーツベースの人検出法は，何をパーツと定義するかにより人検出性能と計算コスト，そしてパーツの統合方法に大きな影響を与える．また，パーツの統合方法も識別結果に大きな影響を与えるため，様々な統合方法が提案されている．

本節では，まずパーツベースの手法で採用される 4 つのパーツの定義について述べる．そして，3 つに分けられる識別器によるパーツの統合方法について述べる．

- パーツの定義
  - ルールベース
  - 画像の位置
  - 特徴点ベース
  - 見えの情報に基づくクラスタリング
- パーツの統合方法
  - 識別器によるパーツの統合
  - グラフモデルによるパーツの統合
  - 投票処理によるパーツの統合

4.2.1　パーツの定義

人は複雑な姿勢をとるため見えの変化が生じるが，頭部や胴体，手，足などの一つ一つのパーツに着目すると，パーツ自体の見えの変化は少ない．パーツベースの人検出法では，如何に人の見えの変動が少ないパーツに分割するかが重要となる．パーツの定義方法は，以下の四つに分けられる．

- ルールベース

人画像毎に予め研究者が決めたルールに従いパーツを分割する [67]．例えば，人画像を頭部と右胴体，左胴体，下半身の 4 つに分割する．人画像毎に異なる位置に存在するパーツを適応的に分割できるが，人手でパーツのラベルを付与する必要がある．

- 画像の位置

人画像の位置によってパーツを分割する [30], [68]．例えば，人画像の上部 3 割を頭部，下部 5 割を足，それ以外を胴体とすることで，パーツの分割を簡単化することができる．しかしながら，人画像毎に頭部や足，胴体の位置や形状は異なるため，予めパーツの位置を決定する方法ではパーツを適応的に捉えることができない．

- 特徴点ベース

パーツよりも小さな小領域に区切る [21], [69]．ハリスのコーナー検出等により検出した特徴点を中心とした小さな領域をパーツとする．必ずしも小領域が人のパーツを捉えるとは限らない．

- 見えの情報に基づくクラスタリング

学習サンプルから人画像に共通する領域を自動的に定義してパーツとする [20], [70]．Bourdev らは，予め人の関節位置を付与しておき，関節位置のずれが学習サンプルで最小となる領域を Poselet [70] として定義している．図 8 に自動的に定義された Poselet を示す．Poselet は，検出に適したパーツを自動的に定義できる一方，関節位置の情報不可コストがかかる．他にも，Latent SVM により学習サンプルに共通する見えの情報と位置の情報から自動的にパーツを分割する手法も提案されている [20].

図 8　自動的に定義された Poselet の例 [67]．左図は Poselet(平均画像)，右図は学習サンプルにおける間接位置のずれが最小となる領域を表わす．

4.2.2　パーツの統合方法

姿勢の変化に対応するためには，検出した幾つかのパーツから，それぞれパーツの見えや位置の関係性を考慮して人を検出する必要がある．そのために，パーツベースの人検出法では検出したパーツを統合する．パーツの統合方法は，下記の 3 つに分けられる．

- 識別器によるパーツの統合

識別器によるパーツの統合手法は，各パーツの検出結果を新たに特徴量と見立てて識別器に入力する方法である．各パーツを識別器により統合する手法として，Mohan らは人間の各パーツを識別する識別器と，それらの識別器の結果を統合する識別器の 2 段階で構成される Adaptive Combination of Classifiers(ACC) [67] を提案している．Mohan らの手法は，1 段階目の識別器において，人の頭部と右胴体，左胴体，下半身の 4 つのパーツを識別し，この識別結果を 2 段階目の識別器に入力することで各パーツの識別結果を統合する．各パーツの位置や大きさは，検出ウィンドウを基準として，ある程度の変動を許容することができるため姿勢の違い吸収できる．しかしながら，この手法は各パーツの位置と大きさを人間が予め定義しているため，全ての学習サンプルに対して人手でラベルを付与する必要がある．

この問題を解決する一手法として，半教師付き学習の一種である Multi-Instance Learning(MIL) [71] を導入した手法 [72]〜[74] が提案されている．MIL を導入した検出法では，まず人画像の位置によりパーツを区分する．そして，各パーツを検出する際には，定義したパーツの位置周辺を適当にサンプリングし，この中に定義したパーツが含まれていればパーツを検出できるアプローチとなっている．これにより，学習サンプルの人画像に対してパーツのラベルを付与する必要がなくなる．しかしながら，これらの方法は統合する識別器に入力する情報は各パーツの識別結果のみであるため，各パーツ間の位置の関係性を捉えるまでには至っていない．

- グラフモデルによるパーツの統合

識別器によりパーツを統合する手法は，パーツを識別する際にパーツの位置や大きさの変動を許容することができるが，パーツ間の関係性を捉える枠組みがないため，人検出に有効なパーツの位置関係を捉えることができない．そこで，パーツの位置関係を利用するアプローチとして，パーツ間の関係性をグラフモデルにより表現する手法が提案されている．グラフモデルによりパーツを統合する手法は，各パーツをノードとしてグラフで表現し，人検出に有効なグラフを学習する．グラフモデルによるパーツを統合する手法は，生成モデルを用いた手法と判別モデルを用いた手法の 2 つに分けることができる．

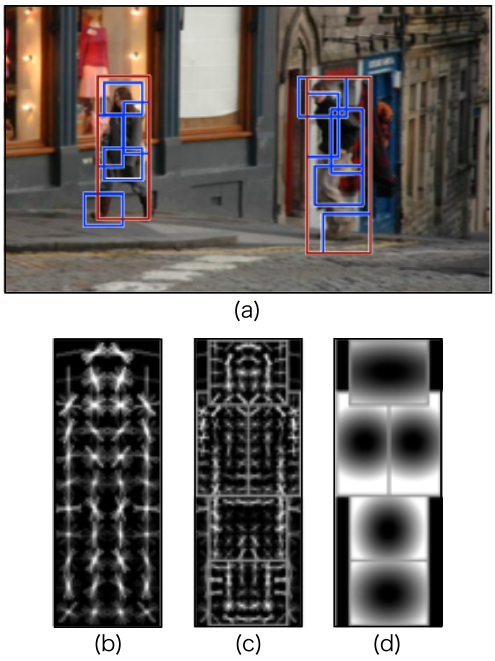


図 9　Deformable Parts Model を用いた人検出の例と各フィルタ [20]．(a) は人検出の例，(b) はルートフィルタ，(c) はパーツフィルタ，(d) は各パーツの移動範囲と移動コストを表わしている．

まず，グラフの学習に生成モデルを用いた手法について述べる．Xia らは，各パーツの位置関係を Star Model により表現し，グラフを最尤法により学習している [75]．Xia らの手法は，検出ウィンドウと各パーツの位置関係を Star Model に表現し，パーツの識別結果を入力した時の尤度が学習サンプルにおいて最大となるよう，各パーツの位置のパラメータを最尤法により求める．Star Model は，基準となる検出ウィンドウと各パーツの関係性を表現するが，Constellation Model [76] のように各パーツ間の関係性を表現するモデルも利用されている [77]．

次に，判別モデルを用いた手法について述べる．判別モデルを利用してグラフモデルを学習する手法として，Felzenszwalb らが提案した Deformable Parts Model [20], [78] がある．Deformable Parts Model は，図 9 に示すように人全身を捉える検出ウィンドウをルートフィルタとパーツを捉えるパーツフィルタの位置関係を Star Model により表現する．そして，ルートフィルタとパーツフィルタの見えに対するスコアと，パーツフィルタの位置に対するスコアの和が学習サンプルにおいて最大となるように Latent SVM により学習する．Deformable Parts Model は姿勢の大きな変化に対応することができるた

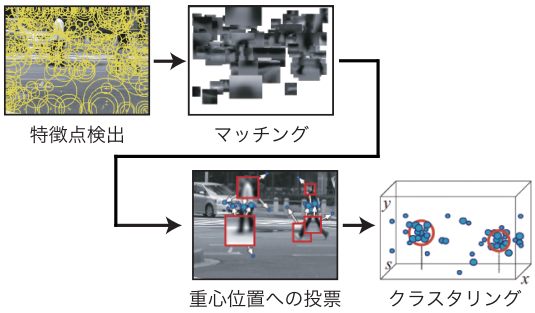


図 10　Leibe らの物体検出の流れ [84].

め，Deformable Parts Model をベースとした人検出法 [79]〜[81] や物体認識法 [82], [83] が多数提案されている．

- 投票処理によるパーツの統合

最後に，投票処理を利用したパーツベースの人検出法ついて述べる．グラフモデルによりパーツを結合する手法は，頭や肩，胴体，足など人体を構成する各パーツの見えと位置の関係性を考慮して人を検出する．これに対して，パーツよりも小さなパッチと呼ばれる領域を検出し，一般化ハフ変換を利用した投票処理に基づきパーツを統合する手法が提案されている．

Leibe らは，投票処理によるパーツを統合する物体検出法として，検出とセグメンテーションを同時に行う Implicit Shape Model(ISM) [69], [84], [85] を提案している．Leibe らの手法による人検出の流れを図 10 に示す．この手法では，まずハリスのコーナー検出により特徴点を検出し，特徴点を中心とした小領域をパッチとして抽出する．そして，抽出したパッチから予め作成したコードブックを参照してベクトル量子化し，同一のコードブックとなる学習サンプルのパッチと重心位置の関係より，物体の重心位置に投票する．最後に，投票結果をクラスタリングすることで物体を検出する．Leibe らの手法では，物体の見えのみに基づいてベクトル量子化していたが，Space-Time patch [47] と呼ばれる時空間情報を利用した投票処理による手法も提案されている [46].

これらの手法では，抽出したパッチがどのコードブックに属するかを如何に精度良く求められるかによって，人検出器の性能が変化する．より高精度にパッチを識別するために，Gall らは人画像と背景画像を用いて判別モデルにより識別器を学習する Hough Forests [21] を提案している．Hough Forests では，多クラスの識別が可能な Randam Forest [86] を利用し，パッチを識別する決定木を学習する．そして，学習された決定木によりパッチを識別し，人と識別されたパッチを重心位置へ投票することで人を検出する．近年では Hough Forests を発展させた手法 [87]〜[89] も提案されている．

4.3　人領域の隠れに対応した手法

物体の密度が高い環境や多くの人が存在する混雑下では，画像上で人は他の物体と重なり，部分的に人画像が隠れる．このような人領域の隠れは，人の特徴を欠損させることに繋がるため，正確に検出することが困難となる．多くの手法は，人領域の隠れに対応するために隠れを予測するモデルを導入している．

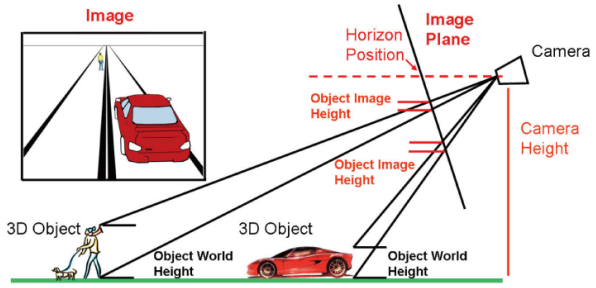
Wang らは人画像を識別する過程において隠れ領域の判定処理を導入した検出法を提案している [22]．Wang らの手法は，識別時に局所領域毎で人と背景を判別し，隠れ領域が存在すると判定した場合には，まず Mean Shift クラスタリング [25] により隠れ領域をセグメンテーションし，隠れた領域を考慮したパーツ検出器を適用する．図 11 に隠れ領域の推定結果の例を示す．Wang らの手法は，HOG 特徴量と LBP 特徴量の画像の見えに関する特徴量のみを用いているが，高精度に隠れ領域を推定できていることがわかる．池村らも，TOF カメラから得られる距離情報を利用して隠れ領域を推定し，隠れが発生している場合にはその隠れ領域を考慮した識別をしている [50]. Enzweiler も同様に，ステレオから得られる距離情報と動き情報から隠れ領域を推定し，隠れ領域に応じて各パーツ領域に対応する識別器に重み付けするアプローチを採っている [90]．どの手法にも共通な点は，まず隠れ領域を推定し，隠れ領域がある場合には隠れを考慮して識別する点である．如何に隠れ領域を正確に検知できるかが重要であり，画像の見えだけではなく距離，動き等の豊富な情報を利用するアプローチに発展している．

図 11　Wang らの手法による隠れ推定の結果 [22]．上段は入力画像，下段は隠れの推定結果を表わしている．数字は人クラスの尤度を表わし，数字が小さいほど隠れ領域の可能性が高い．

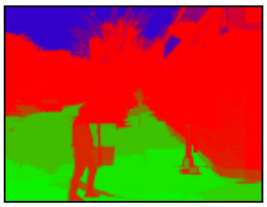
4.4　視点の変化に対応した手法

カメラの俯角が変わると，俯角の大きさに応じて人の見えが変化する．この見えの変動を吸収するために，幾何学的な情報を利用する手法が提案されている．

幾何学情報を利用した代表的な人検出法として，Hoiem らの手法がある [23]．この手法は，図 12(a) に示すような人とカメラの幾何学的な関係を利用することで視点の変化に対応する．図 12(c) からもわかるように，地面の上に立っている人は水平線よりも下に検出ウィンドウの底が位置する．このような幾何学的な関係を利用するために，Hoiem らの手法はカメラ視点 (高さや俯角) の推定結果，地面や空，垂直物のジオメトリの推定結果 (図 12(b))，人識別器の検出結果の 3 つの要素を考慮する．そのために，3 つの要素をグラフィカルモデルにより表現し，ベイジアンネットワークを用いることで各々の関係性を考慮して人を検出する．人検出器の結果のみでなく，カメラと人の幾何関係や周囲のジオメトリ情報までを含めた情報から識別するため，単純な識別器のみと比べて高精度な検出が可能で



(a) 物体とカメラの幾何学的関係



(b) ジオメトリの推定

(c) 幾何情報を利用した検出結果

図 12　幾何学情報を利用した人検出 [23].

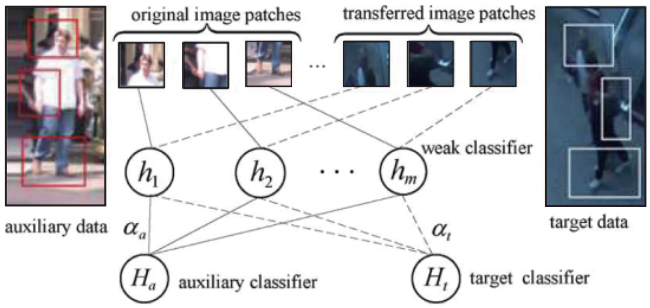


図 13　特徴の転移と識別器の最適化の概要 [24].

ある．

Hoiem らの手法は，識別する際に推定した視点情報を利用しているため，多少の視点の違いに対応できるが，人画像の見えが大きく変動するような場合には対応することが難しい．そこで，見えの変化が大きく変わるようなシーンに対応するための手法として，Pang らの転移学習を用いた人検出法 [24] がある．この方法は，事前に学習した識別器と人検出を動作させる環境から収集した少数のサンプルを用いて転移学習により識別器を最適化する．最適化は 2 つの処理に分けられ，1 つ目の処理では Boosting により学習した弱識別器 h_m の捉えている局所領域が，人検出器を動作させる環境ではどこに位置するかを推定する．これを特徴の転移と呼び，図 13 に示すように視点の異なる人画像においても，弱識別器 h_m が同一のパーツを捉えることが可能となる．2 つ目の処理では，特徴の転移を行った弱識別器 h_m の選択と弱識別器の信頼度 α_m を Covariate Boost により最適化する．人検出を行う環境から収集した人画像と識別器を最適化する計算が必要ではあるが，カメラの視点の変化の影響で人画像の見えが大きく変動する問題に対応することができる．

5.　人検出の実用化に向けて

人検出を困難とする要因に対して適切なアプローチを採ることにより，高精度な人検出を実現することできる．その一方，人検出を実用化するためには下記に示すような 3 つの項目に対して考慮する必要がある．

- 人検出の高速化
- 学習サンプルの収集
- 検出ウィンドウのラスタスキャン

ここでは，各項目に対して取り組まれている研究について述べる．

5.1　人検出の高速化

人を検出するためには，膨大な検出ウィンドウを人と人以外に判定する必要があるため計算コストが非常に高い．そのため，人検出技術を実用化するには，汎用的なパーソナルコンピュータよりも性能が低いハードウェアでも動作する低計算量かつ演算効率の良い人検出アルゴリズムが必要不可欠である．人検出アルゴリズムは，特徴抽出と識別の 2 つの処理に分けることができ，それぞれの処理において高速化が取り組まれている．

特徴抽出の高速化手法は，局所領域内の輝度和を高速に求められる積分画像 [8] が利用されている．勾配ベースの特徴量を計算する際には，積分画像を応用した積分ヒストグラム [41] を利用することで，高速な人検出を実現できる [91]．Zhu らは，HOG 特徴量の計算に積分ヒストグラムを利用することにより高速に HOG 特徴量を計算した [91]．積分ヒストグラムを利用することで，各画素において一度の勾配情報を計算だけで済むため，計算コストを大幅に削減できる．また，積分画像や積分ヒストグラムは，輝度和やヒストグラムを高速に求められることから，このメリットを生かした Integral Channel Features [42] も提案されている．

識別の高速化手法には，効率の良い識別が可能なカスケード型識別器 [8] が採用されている．Zhu らは，局所領域毎に HOG 特徴量と SVM により多数の識別器を学習し，それらの識別器をカスケード型に構成することで高速な識別器を構築している [91]．他にも，カスケード型識別器を採用することで識別に必要な計算コストを削減した手法 [29], [75], [79] が提案されている．

また，近年では Graphics Processing Unit(GPU) の性能が劇的に進化したことを受け，GPU を用いた高速な人検出法 [92]〜[94] が提案されている．GPU を用いた人検出法は，GPU の特性に合わせてアルゴリズムを最適化することにより，HOG 特徴量と統計的学習手法による人検出法を数十から数百倍の高速化を実現している．

5.2　学習サンプルの収集

検出性能の高い識別器を学習するには，人検出器を稼働させる環境の人画像を大量に収集する必要がある．しかしながら，実際にはプライバシーや収集コストの面から人画像を収集することが困難な場合が多い．この問題を解決する手法として，少数の学習サンプルから実環境で測定されうる変動を含むように変形させた学習サンプルを生成し，生成したサンプルにより識

図 14　上段は実画像，下段は CG から生成された画像の例 [96].

別器を学習する生成型学習 [95] を利用した人検出法 [96]〜[98] が提案されている．

Mar らは仮想空間中にゲームで使用されている人体モデルを配置し，仮想カメラから撮影される映像から人画像を生成した [96]．この時，人や仮想カメラの三次元位置は既知であるため，映像から自動的に大量の人画像を収集することができる．図 14 に CG から生成した人画像の例を示す．CG から生成した人画像は，リアリティに欠けて見えるが，生成した人画像を用いて学習した識別器は，人の実画像を用いて学習した識別器と同等の検出性能が得られたと報告されている．同様に，Yamauchi らも三次元人体モデルを用いて人画像を生成し，背景画像も実環境の画像からランダムで切り出すことで，全ての学習サンプルを自動的に生成している [97].

5.3　検出ウィンドウのラスタスキャン

人検出を実用化する場合，設置するカメラのレンズの影響で画像に歪みが発生する場合がある．歪みが大きい場合には人の見えの変化を大きくし，人検出に悪影響を及ぼすため，歪みの影響を取り除く必要がある．

Li らは，遠近法による歪みの影響で画像の y 軸に平行して直立していないように見える人物を検出する手法を提案している [99]．Li らの手法は，まずカメラキャリブレーションによりカメラパラメータを取得し，レンズによる歪みを補正した画像の地面と人物は地面に直立していることを仮定する．そして，地面を基準として三次元空間をラスタスキャンする．遠近法による歪みを除去するために，人物に円柱を当てはめ，円柱に内接しながらもカメラから正面を向く平面を求め，擬似的に人を正面から観測するように平行投影した画像を取得する．識別する際には，擬似的に平行投影した歪みのない画像を用いて識別する．これにより，遠近法による歪みの影響を受けにくい検出が可能となる．李らも同様の考え方により画像の回転や歪みの影響を抑制した手法として，Smart Window Transform [100] を提案している．

5.4　人検出の実用化例

人検出技術は，一般環境下で十分な性能を満たすことが難しいことから実用化例が少ない．しかしながら，人検出を行うシーンの限定や特殊なカメラを使用することにより，人検出技術の応用先で要求される検出性能を満たすことで実用化されている．

2004 年に発表された本田技研工業株式会社のインテリジェント・ナイトビジョンシステムは，自動車に搭載された遠赤外線カメラを用いて歩行者を検出する．そして，歩行者の存在をドライバーに伝えることで夜間運転のリスクを減らすことが可能となり，運転支援システムとして実用化されている．また，2008 年にはトヨタ自動車株式会社からナイトビューシステムが搭載された自動車が発売されている．ナイトビューシステムでは，自動車に搭載された近赤外線カメラから人を検出し，ドライバーに伝えることで運転を支援する．

近年では人検出技術をハードウェア化する取り組みも盛んである．2010 年には三栄ハイテックス株式会社から FPGA による人検出システム ODEN(Object Detect ENgine) が発売されている．この人検出システムは，ハードウェアとソフトウェアの両方で人検出が可能な製品である．2011 年には株式会社東芝から人検出可能な車載向け画像認識用 LSI を発売されている．このように人検出技術をハードウェア化することで，人検出技術が様々な用途で利用され始めている．

6.　評価データと評価指標

多数の人検出法が提案されるにつれ，これらを平等に評価するために画像データベースと統一的な評価指標が整備されている．本節では，人画像データベースと評価指標について述べる．

6.1　人画像データベースと作成法

多数の人検出法を比較するためには，同一の人画像データベースで評価する必要がある．そのため，人検出法を評価するための多くのデータベースが作成され，Web 上で公開されている．

6.1.1　人画像データベース

表 2 に公開されている人画像データベースを示す．表中の正規化とは，ある一定の大きさに正規化された画像を表わす．正規化人画像の場合には，画像から人領域を切り出し，一定の大きさに正規化された画像である．正規化画像以外は，切り出し等の処理が行われておらず，画像中に人が存在する画像を表わす．

これらのデータベースの多くは，撮影された人を一定サイズの矩形で切り取られている，もしくは画像中で人が存在する位置や大きさなどが記述されているアノテーションデータが付随されている．これらのデータベースを用いることで，他の手法と容易に比較することができるため，研究の信頼性を高めるためにも，一般公開されているデータベースを使用することが望ましい．

人検出法を評価するために，人検出の分野で初めて公開された画像データベースとして，MIT CBCL Pedestrian Data [101] がある．MIT CBCL Pedestrian Data は，人を正面もしくは後ろから撮影した画像で構成されている．また，撮影した場所が限定されている他，多数の画像において人の見えが類似しており，比較的簡単な人画像データベースといえる．その後，Dalal らは HOG 特徴量と SVM による人検出法の提案と共に INRIA Person Dataset [13] を公開している．HOG 特徴量と線形 SVM による人検出法は，MIT CBCL Pedestrian Data では人と背景を完全に区別できたとして，より人の向きや背景に制限を設けずに作成した INRIA Person Dataset を公開した．

表 2　人画像データベースの比較.

	学習用		評価用		
	正規化人画像	背景画像	人画像	正規化人画像	背景画像
MIT [101]	924	-	-	-	-
INRIA [13]	2,416	1,218	288	1,132	453
USC-A [30]	-	-	205	303	-
USC-B [30]	-	-	54	271	-
USC-C [19]	-	-	100	232	-
ETH [102]	1,578	-	1,803	9,380	-
Daimler2006 [103]	14,400	150,000*	-	1,600	100,000
Daimler2009 [104]	15,660	6,744	21,800	56,492	-
NICTA [105]	18,700	5,200*	-	6,900	50,000
TUD [106]	400	-	250	311	
Caltech [107]	192,000	61,000	56,000	155,000	5,600

∗ は一定の大きさで切り出された画像を表わす.

INRIA Person Dataset は，人を横から撮影した画像だけでなく，直立した人の様々な姿勢の人画像を含む．これ以降，人検出に関する多く研究が評価実験に INRIA Person Dataset を利用している．

INRIA Person Dataset は，人がデジタルカメラで撮影した画像がほとんどであったが，自動車やロボットに取り付けたカメラで撮影した映像より作成したデータベース [103], [104], [107] もある．その中でも，Caltech Pedestrian Detection Benchmark [107] は最も汎用性が高く，非常に多くの画像から構成されるデータベースである．このデータベースは，車両に取り付けたカメラより得られる映像から作成されている．大量の画像が用意されているだけでなく，人領域の隠れに対するラベルの付与，時系列の映像データが使用できるなど，非常に汎用性の高いデータベースとなっている．

6.2 評価指標

複数の人検出法を平等に評価するためには，同一の画像データベースを使用する他に，評価指標も統一する必要がある．評価指標としては，下記の 2 つの指標がデータベースと共に提案されている．

（1） Miss rate VS. False Positive Per Window(FPPW) [13]

（2） Miss rate VS. False Positive Per Image(FPPI) [107]

(1) の FPPW は，1 つの検出ウィンドウあたりの誤検出した検出ウィンドウ数を表わす．FPPW による評価には，画像から人の領域に合わせて切り出された人画像を用いる．(2) の FPPI は，1 枚の画像あたりの誤検出した数を表わす．FPPI の評価には，人画像として切り出し等の処理が行われていない画像中に人が存在する画像を用いる．2 つの評価指標は一見同じようにも取れるが，検出ウィンドウの統合処理の有無が異なる．(2) では検出ウィンドウの統合処理を行うことで，誤検出となった明らかに孤立した検出ウィンドウを除去する．人検出結果を用いるアプリケーションでは，検出ウィンドウの統合処理を行うことから，人検出システムを稼働させる場合と同じ条件により評価できる．そのため，より直感的に性能を把握できるため，近年提案される人検出法では FPPI により評価する傾向がある．

(1) と (2) では，トレードオフの関係にある未検出率と誤検出率を比較する．一般的に人検出の性能の比較においては，識別器の閾値を変化させた時の未検出率と誤検出率をプロットした Detection Error Tradeoff(DET) カーブを用いて比較する．未検出率と誤検出率をプロットした際，カーブの面積が小さい方 (グラフの左下にあるカーブ) がより高い検出性能をもつ手法であることを表わす.

7. おわりに

本稿では，人検出法のサーベイと人検出用データベースならびに検出性能の評価法について述べた．まず，人検出が困難となる要因を 6 つに分け，各要因を解決するための特徴抽出法と識別手法の 2 つに分けてサーベイを行った．そして，人検出法を評価するためのデータベースと評価指標について紹介した．

2005 年に Dalal らが HOG 特徴量と SVM による人検出法を提案した以降，非常に多数の人検出法が提案され，人検出の研究は著しい進化を遂げた．近年では条件や用途を限定できるような環境に限れば，人検出技術の実用化が進みつつある．

一方で，実用化は人検出に関する研究が新たな局面を迎えたと言える．例えば，パーツベースの人検出法 [20] は，人の姿勢変動を吸収できることから，視点の変化や姿勢の変化が大きい動物や乗り物等の一般的な物体カテゴリにも適用されている．さらに，近年では一般的な人検出法とは異なり，より実用化する際に想定される問題に合わせた研究も発表されている．例えば，人検出器の性能を最大限に高めるためには，実際に人を検出する環境から学習サンプルを大量に収集することが好ましいが，現実的には難しい．そこで，少数のサンプルもしくは自動的にサンプルを収集し，その環境下に特化した人検出器を学習する方法 [108]〜[111] が提案されている．より実用化する際に想定される問題に合わせた研究が多く発表されていることからも，最近の研究動向がうかがえる．

文　献

[1] T. Sakai, et al., “Line Extraction and Pattern Detection in a Photograph,” Journal of the Pattern Recognition, vol.1, pp.233–248, 1969.

[2] V.Govindaraju, et al., “A Computational Model for Face

Location," ICCV, pp.718–721, 1990.
[3] G. Yang, et al., "Human Face Detection in a Complex Background," Journal of the Pattern Recognition, vol.27, no.1, pp.53–63, 1994.
[4] C. Kotropoulos, et al., "Rule-Based Face Detection in Frontal Views," International Conference on Acoustics, Speech, and Signal Processing, vol.4, pp.2537–2540, 1997.
[5] K.-K. Sung, et al., "Example-Based Learning for View-Based Human Face Detection," Technical Report MIT AI Lab, 1994.
[6] H.A. Rowley, et al., "Neural Network-Based Face Detection," CVPR, pp.203–208, 1996.
[7] H. Schneiderman, et al., "A Statistical Method for 3D Object Detection Applied to Faces and Cars," CVPR, 2000.
[8] P. Viola, et al., "Rapid Object Detection Using a Boosted Cascade of Simple Features," CVPR, pp.511–518, 2001.
[9] E. Osuna, et al., "Training Support Vector Machines: an Application to Face Detection," CVPR, pp.130–136, 1997.
[10] P. Viola, et al., "Robust Real-Time Object Detection," IJCV, vol.57, no.2, pp.137–154, 2004.
[11] C. Papageorgiou, et al., "A Trainable System for Object Detection," IJCV, vol.38, no.1, pp.15–33, 2000.
[12] P. Viola, et al., "Detecting Pedestrians Using Patterns of Motion and Appearance," ICCV, pp.734–741, 2003.
[13] N. Dalal, et al., "Histograms of Oriented Gradients for Human Detection," CVPR, vol.1, pp.886–893, 2005.
[14] S. Walk, et al., "New Features and Insights for Pedestrian Detection," CVPR, pp.1030–1037, 2010.
[15] N. Dalal, et al., "Human Detection Using Oriented Histograms of Flow and Appearance," ECCV, vol.2, pp.428–441, 2006.
[16] T. Mita, et al., "Discriminative Feature Co-Occurrence Selection for Object Detection," PAMI, vol.30, no.7, pp.1257–1269, 2008.
[17] T. Watanabe, et al., "Co-occurrence Histograms of Oriented Gradients for Human Detection," Information Processing Society of Japan Transactions on Computer Vision and Applications, vol.2, pp.39–47, 2010.
[18] 三井相和 他, "Joint 特徴量を用いた 2 段階 Boosting による物体検出," 信学論, vol.J92-D, no.9, pp.1591–1601, 2009.
[19] B. Wu, et al., "Cluster Boosted Tree Classifier for Multi-View, Multi-Pose Object Detection," ICCV, pp.1–8, 2007.
[20] P.F. Felzenszwalb, et al., "Object Detection with Discriminatively Trained Part Based Models," PAMI, vol.32, no.9, pp.1627–1645, 2009.
[21] J. Gall, et al., "Class-Specific Hough Forests for Object Detection," CVPR, 2009.
[22] X. Wang, et al., "An HOG-LBP Human Detector with Partial Occlusion Handling," ICCV, 2009.
[23] D. Hoiem, et al., "Putting Objects in Perspective," IJCV, vol.80, no.1, pp.3–15, 2008.
[24] J. Pang, et al., "Transferring boosted detectors towards viewpoint and scene adaptiveness," IEEE Transactions on Image Processing, vol.20, no.5, pp.1388–400, 2011.
[25] D. Comaniciu, et al., "Mean Shift : A Robust Approach Toward Feature Space Analysis," PAMI, vol.24, no.5, pp.603–619, 2002.
[26] 佐部浩太郎 他, "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," SSII, 2004.
[27] K. Levi, et al., "Learning Object Detection from a Small Number of Examples: the Importance of Good Features," CVPR, vol.2, pp.53–60, 2004.
[28] B. Wu, et al., "Detection and Segmentation of Multiple, Partially Occluded Objects by Grouping, Merging, Assigning Part Detection Responses," IJCV, vol.82, no.2, pp.185–204, 2009.
[29] Y.T. Chen, et al., "A Cascade of Feed-Forward Classifiers for Fast Pedestrian Detection," ACCV, pp.905–914, 2007.
[30] B. Wu, et al., "Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors," ICCV, pp.90–97, 2005.
[31] W. Li, et al., "Texture Classification Using Texture Spectrum," Journal of the Pattern Recognition, vol.23, no.8, pp.905–910, 1990.
[32] Y.D. Mu, et al., "Discriminative local binary patterns for human detection in personal album," CVPR, pp.1–8, 2008.
[33] 阿部厳 他, "カーネル部分最小二乗法分析を用いた歩行者認識," CVIM, 2010.
[34] 西村洋文 他, "オートモーティブ分野向け画像センシング技術の開発," パナソニック技報, vol.57, no.3, pp.62–67, 2011.
[35] 鶴岡信治 他, "加重方向指数ヒストグラム法による手書き漢字・ひらがな認識," 信学論, vol.70-D, no.7, pp.1390–1397, 1987.
[36] Z. Lin, et al., "A Pose-Invariant Descriptor for Human Detection and Segmentation," ECCV, 2008.
[37] C. Hou, et al., "Multiview Pedestrian Detection Based on Vector Boosting," ACCV, pp.210–219, 2007.
[38] A. Bosch, et al., "Representing Shape with a Spatial Pyramid Kernel," International Conference on Image and Video Retrieval, 2007.
[39] P. Ott, et al., "Implicit Color Segmentation Features for Pedestrian and Object Detection," ICCV, 2009.
[40] 河合富美 他, "オブジェクトの同一性を考慮したエッジ特徴による人物検出," MIRU, pp.2084–2091, 2010.
[41] F. Porikli, "Integral Histogram: a Fast Way to Extract Histograms in Cartesian Spaces," CVPR, vol.1, pp.829–836, 2005.
[42] P. Dollár, et al., "Integral Channel Features," British Machine Vision Conference, 2009.
[43] 後藤雄飛 他, "色の類似性に基づいた形状特徴量 CS-HOG の提案," SSII, 2012.
[44] J. Yao, et al., "Fast Human Detection from Videos Using Covariance Features," Visual Surveillance Workshop(in conjunction with ECCV2008), 2008.
[45] J. Yao, et al., "Multi-Layer Background Subtraction Based on Color and Texture," Computer Vision and Pattern, Recognitionisual Surveillance Workshop, 2007.
[46] 村井泰裕 他, "Space-Time Patch を用いた物体の移動方向識別とセグメンテーション," 情処学論 CVIM, vol.1, no.2, pp.21–31, 2008.
[47] E. Shechtman, et al., "Space-Time Behavior-Based Correlation-OR-How to Tell if Two Underlying Motion Fields are Similar without Computing Them?," PAMI, vol.29, no.11, pp.2045–56, 2007.
[48] 村井泰裕 他, "時空間特徴に基づくエスカレータシーンにおける人の異常行動検知," PRMU, pp.247–254, 2008.
[49] Y. Yamauchi, et al., "People Detection Based on Co-occurrence of Appearance and Spatio-temporal Features," National Institute of Informatics Transactions on Progress in Informatics, vol.1, no.7, pp.33–42, 2010.
[50] 池村翔 他, "距離情報に基づく局所特徴量によるリアルタイム人検出," 信学論, vol.93-D, no.3, pp.355–364, 2010.
[51] J. Shotton, et al., "Real-time human pose recognition in parts from single depth images," CVPR, June 2011.
[52] L. Xia, et al., "Human Detection Using Depth Information by Kinect," International Workshop on Human Activity Understanding from 3D Data(in conjunction with CVPR), pp.15–22, 2011.
[53] T. Watanabe, et al., "Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection," Pacific-Rim Symposium on Image and Video Technology, pp.37–47, 2009.
[54] H. Hattori, et al., "Stereo-Based Pedestrian Detection Using Multiple Patterns," British Machine Vision Conference, vol.243, 2009.
[55] R. Nosaka, et al., "Feature Extraction Based on Co-occurrence of Adjacent Local Binary Patterns," Pacific-Rim Symposium on Image and Video Technology, 2011.

[56] O. Tuzel, et al., “Pedestrian Detection via Classification on Riemannian Manifolds,” PAMI, vol.30, no.10, pp.1713–1727, 2008.

[57] P. Sabzmeydani, et al., “Detecting Pedestrians by Learning Shapelet Features,” CVPR, pp.1–8, 2007.

[58] C. Huang, et al., “Learning Sparse Features in Granular Space for Multi-View Face Detection,” International Conference on Automatic Face and Gesture Recognition, pp.401–406, 2006.

[59] G. Duan, et al., “Boosting Associated Pairing Comparison Features for Pedestrian Detection,” International Workshop on Visual Surveillance(in conjunction with Internationa Conference on Computer Vision), 2009.

[60] 山内悠嗣 他，“Boosting に基づく特徴量の共起表現による人検出,” 信学論，vol.J92-D，no.8，pp.1125–1134，2009.

[61] H.A. Rowley, et al., “Rotation Invariant Neural Network-Based Face Detection,” CVPR, pp.38–44, 1998.

[62] M. Jones, et al., “Fast Multi-View Face Detection,” Mitsubishi Electric Research Lab Technical Report, 2003.

[63] S.Z. Li, et al., “Multi-view face pose estimation based on supervised ISA learning,” International Conference on Automatic Face and Gesture Recognition, pp.100–105, 2002.

[64] S.Z. Li, et al., “Statistical Learning of Multi-View Face Detection,” ECCV, 2002.

[65] C. Huang, et al., “Vector boosting for rotation invariant multi-view face detection,” ICCV, vol.1, pp.446–453, 2005.

[66] 土屋成光 他，“Boosting に基づく分割統治的戦略による高精度な識別器構築手法の提案,” PRMU，pp.81–86，2009.

[67] A. Mohan, et al., “Example-Based Object Detection in Images by Components,” PAMI, vol.23, no.4, pp.349–361, 2001.

[68] Z. Lin, et al., “Hierarchical Part-Template Matching for Human Detection and Segmentation,” ICCV, 2007.

[69] B. Leibe, et al., “Interleaved Object Categorization and Segmentation,” British Machine Vision Conference, pp.759–768, 2003.

[70] L. Bourdev, et al., “Poselets: Body Part Detectors Trained Using 3D Human Pose Annotations,” ICCV, 2009.

[71] T.G. Dietterich, et al., “Solving the Multiple Instance Problem with Axis-Parallel Rectangles,” Artificial Intelligence Journal, vol.89, pp.31–71, 1997.

[72] Z. Lin, et al., “Multiple Instance Feature for Robust Part-based Object Detection,” CVPR, pp.1–8, 2009.

[73] P. Dollár, et al., “Multiple Component Learning for Object Detection,” ECCV, pp.211–224, 2008.

[74] Y.-T. Chen, et al., “Multi-Class Multi-Instance Boosting for Part-Based Human Detection,” International Workshop on Visual Surveillance(in conjunction with ICCV2009), pp.1177–1184, Sept. 2009.

[75] X. Xia, et al., “Part-Based Object Detection using Cascades of Boosted Classifiers,” ACCV, 2009.

[76] M. Burl, et al., “Recognition of Planar Object Classes,” CVPR, pp.223–230, 1996.

[77] R. Fergus, et al., “Object Class Recognition by Unsupervised Scale-Invariant Learning,” CVPR, vol.2, pp.264–271, 2003.

[78] P.F. Felzenszwalb, et al., “A Discriminatively Trained, Multiscale, Deformable Part Model,” CVPR, 2008.

[79] P.F. Felzenszwalb, et al., “Cascade Object Detection with Deformable Part Models,” CVPR, pp.2241–2248, 2010.

[80] P. Ott, et al., “Shared Parts for Deformable Part-Based Models,” CVPR, 2010.

[81] M. Pedersoli, et al., “A Coarse-to-Fine Approach for Fast Deformable Object Detection,” CVPR, 2011.

[82] L.L. Zhu, et al., “Latent Hierarchical Structural Learning for Object Detection,” CVPR, 2010.

[83] M.A. Sadeghi, et al., “Recognition Using Visual Phrases,” CVPR, pp.1745–1752, 2011.

[84] B. Leibe, et al., “Robust Object Detection with Interleaved Categorization and Segmentation,” IJCV, vol.77, no.1-3, pp.259–289, 2008.

[85] B. Leibe, et al., “Combined object categorization and segmentation with an implicit shape model,” Statistical Learning in Computer Vision, (in conjunction with ECCV), 2004.

[86] L. Breiman, “Random Forests,” Machine Learning, vol.45, no.1, pp.5–32, 2001.

[87] K. Vijay, et al., “A Discriminative Voting Scheme for Object Detection using Hough Forests,” British Machine Vision Conference Postgraduate Workshop, 2010.

[88] 波部斉 他，“Joint Hough Forests: 局所パッチ間の共起関係を考慮した投票ベースの物体検出,” MIRU，2011.

[89] 波部斉，“スケール不変 Hough Forests による物体検出,” SSII，2011.

[90] M. Enzweiler, et al., “Multi-Cue Pedestrian Classification with Partial Occlusion Handling,” CVPR, pp.990–997, 2010.

[91] Q. Zhu, et al., “Fast Human Detection Using a Cascade of Histograms of Oriented Gradients,” CVPR, pp.1491–1498, 2006.

[92] B. Bilgic, et al., “Fast Human Detection with Cascaded Ensembles on the GPU,” IEEE Intelligent Vehicles Symposium, pp.325–332, 2010.

[93] V.A. Prisacariu, et al., “fastHOG - a Real-Time GPU Implementation of HOG,” Technical Report Oxford University, 2009.

[94] R. Benenson, et al., “Pedestrian Detection at 100 Frames per Second,” CVPR, pp.2903–2910, 2012.

[95] 村瀬洋，“画像認識のための生成型学習,” 情処学論 CVIM，vol.46，no.15，pp.35–42，2005.

[96] J. Marín, et al., “Learning Appearance in Virtual Scenarios for Pedestrian Detection,” CVPR, pp.137–144, 2010.

[97] Y. Yamauchi, et al., “Automatic Generation of Training Samples and a Learning Method Based on Advanced MIL-Boost for Human Detection,” ACPR, pp.603–607, 2011.

[98] 吉田英史 他，“生成型学習法を用いた姿勢変化に頑健な歩行者検出の検討,” PRMU，pp.127–132，2011.

[99] Y. Li, et al., “Human Detection by Searching in 3D Space Using Camera and Scene Knowledge,” ICPR, 2008.

[100] 李媛 他，“Smart Window Transform とエッジペース識別器に基づく人検出,” 精密工学会秋季学会学術講演会，2011.

[101] M. Oren, et al., “Pedestrian Detection Using Wavelet Templates,” CVPR, pp.193–199, 1997.

[102] A. Ess, et al., “Depth and Appearance for Mobile Scene Analysis,” ICCV, 2007.

[103] S. Munder, et al., “An Experimental Study on Pedestrian Classification,” PAMI, vol.28, pp.1863–1868, 2006.

[104] M. Enzweiler, et al., “Monocular pedestrian detection: survey and experiments,” PAMI, vol.31, no.12, pp.2179–2195, 2009.

[105] G. Overett, et al., “A New Pedestrian Dataset for Supervised Learning,” The Intelligent Vehicles Symposium, 2008.

[106] M. Andriluka, et al., “People-Tracking-by-Detection and People-Detection-by-Tracking,” CVPR, 2008.

[107] P. Dollár, et al., “Pedestrian Detection: An Evaluation of the State of the Art,” PAMI, vol.34, no.4, pp.743–761, 2012.

[108] M. Wang, et al., “Automatic adaptation of a generic pedestrian detector to a specific traffic scene,” CVPR, pp.3401–3408, 2011.

[109] P. Sharma, et al., “Unsupervised Incremental Learning for Improved Object Detection in a Video,” CVPR, pp.3298–3305, 2012.

[110] M. Wang, et al., “Transferring a Generic Pedestrian Detector Towards Specific Scenes,” CVPR, pp.3274–3281, 2012.

[111] X. Wang, et al., “Detection by Detections : Non-parametric Detector Adaptation for a Video,” CVPR, pp.350–357, 2012.