解説論文

画像からの統計的学習手法に基づく人検出

山内　悠嗣[†a)]　　山下　隆義[††b)]　　藤吉　弘亘[†c)]

Human Detection Based on Statistical Learning from Image

Yuji YAMAUCHI[†a)], Takayoshi YAMASHITA[††b)], and Hironobu FUJIYOSHI[†c)]

あらまし　人検出とは，画像中から人が存在する位置と大きさを自動的に求める技術である．人検出は古くから取り組まれてきた顔検出の研究がベースとなっている．近年では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，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，本論文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人画像データベースと統一的な評価指標についても紹介する．

キーワード　人検出，特徴量，統計的学習手法，サーベイ

1. まえがき

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，人の位置を特定する人検出技術は，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．本論文における人検出は，画像から人が存在する位置座標，大きさを出力する技術を指す．類似技術として特定の人物を認識する人物照合があるが本論文では触れないため，人物照合については [1] を参照されたい．

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 [2] 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，目や口の濃淡や位置の関係を利用した研究者があらかじ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 [3]～[5] が主流であった．これらの方法は，顔画像から低次元な特徴ベクトルを抽出して比較的簡単な処理で顔と背景を判別する．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するための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 [6]～[9] 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 [7] や SVM を用いた手法 [10]，Naive Bayes に基づく手法 [8]，AdaBoost による手法 [9], [11] 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 [9], [11] 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，検出がより困難とされる人に検出対象が遷移した．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，一般的に画像局所特徴量と統計的学習手法に基づくアプ

† 中部大学，春日井市
Chubu University, 1200 Matsumoto, Kasugai-shi, 487–8501 Japan
†† オムロン株式会社，草津市
OMRON Corporation, 2–2–1 Nishikusatsu, Kusatsu-shi, 525–0035 Japan
a) E-mail: yuu@vision.cs.chubu.ac.jp
b) E-mail: takayosi@omm.ncl.omron.co.jp
c) E-mail: hf@cs.chubu.ac.jp

ローチがとられる [12]〜[14]．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，人領域の隠れの影響により，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困難になる．この問題を解決するために，現在までに画像局所特徴量の抽出と統計的学習手法による識別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，本論文では人検出を難しくする要因について体系的に整理した上で，その解決方法に焦点を当てる．これまでに人検出のサーベイは幾つか行われている [15]〜[17]．[15] や [16] は，それぞれが画像データベースを提案し，そのデータベースを用いて人検出の代表的な手法の優越関係を実験的に示した．[17] では運転支援システムとしての歩行者検出に重点を置き，車載カメラ映像の前処理，画像分割，識別，照合，追跡の各処理についての手法を述べている．上記のサーベイは，本論文とは異なる視点から人検出をサーベイしており，本論文では扱わない内容も含むため参考にして補って頂きたい．

本論文の構成を下記に示す．**2.** では人検出の概要と人検出の基本フレームワークについて述べた後，人検出を難しくする要因について整理する．各要因について解決するアプローチとして，**3.** では画像局所特徴量，**4.** では統計的学習手法による識別について述べる．**5.** 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.** 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.** 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，多くの手法では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とられている．本章では，本論文が対象とする人検出の範囲を明確にするために，問題設定や目標，システム構成について述べる．そして，近年では一般的に採用されている画像局所特徴量と統計的学習手法に基づく人検出の基本フレームワークについて述べる．最後に人検出の課題として人検出を難しくする要因について整理する．

2.1 問題設定と目標

画像から対象の物体を検出する技術を物体検出と呼び，現在までに多くの研究者によって精力的に取り組まれてきた．物体検出という技術が指す研究は，その問題設定とアプローチから下記の三つの問題に大別することができる．一つ目は，前景（対象物体）と背景を分離する問題であり，背景以外の領域を検出する問題として取り組まれている．動的に変化する背景に対して不変となるようにモデル化し，対象とする前景領域のみを抽出する技術である [18], [19]．二つ目は，三次元形状等の情報が既知である特定物体を検出する問題である．対象となる物体の三次元的な形状を保持しておき，そのモデルと入力画像を照合する [20]．三つ目は，検出対象のクラスに属する物体を検出する問題である [11], [14], [21]．検出対象のクラスに属する物体の特徴に着目し，統計的学習手法により識別する．本論文が対象とする物体検出は，三つ目の問題設定の範囲に属する．

物体検出の初期の研究では，人の顔を検出対象とした研究が盛んに取り組まれた．そして，高速かつ高精度な顔検出法が確立された後，顔よりも検出が難しい人に検出対象が遷移した．近年では，顔や人以外を検出対象とした研究も進んでおり，2010 年の画像認識に関するベンチマークワークショップ Visual Object Classes Challenge（VOC2010）では，20 クラスを対象とした物体検出に関するコンテストが行われた．

本論文が扱う人検出の目標は，「人の見た目の個体差」や「背景のテクスチャ」，「人の向き」，「姿勢」，「人領域の隠れ」，「視点」にかかわらず人の存在する位置と大きさを求めることである．しかしながら，上記に挙げた要因により人の見えが大きく変化するため，あらゆる環境に対応して人を検出することは現状では難しい．このような状況ではあるが，人検出を行う環境を固定するなどの制約を課すことにより人検出の問題を簡単化することで実用化が進んでいる．

2.2 システム構成

低コストかつ高い汎用性をもつ人検出システムを実現するためには，可能な限り安価なシステムが望まれる．このような理由もあり，多くの研究では可視光カメラから得られる画像を用いて人検出を行うことを想定している．しかしながら，特殊な環境下では可視光では十分な性能を満たせないため，可視光カメラ以外のカメラを想定した人検出法も提案されている．

夜間における人検出では，可視光カメラでは明瞭な人の輪郭を捉えることができないため，赤外カメラを用いた人検出法 [22], [23] が提案されている．赤外線カ

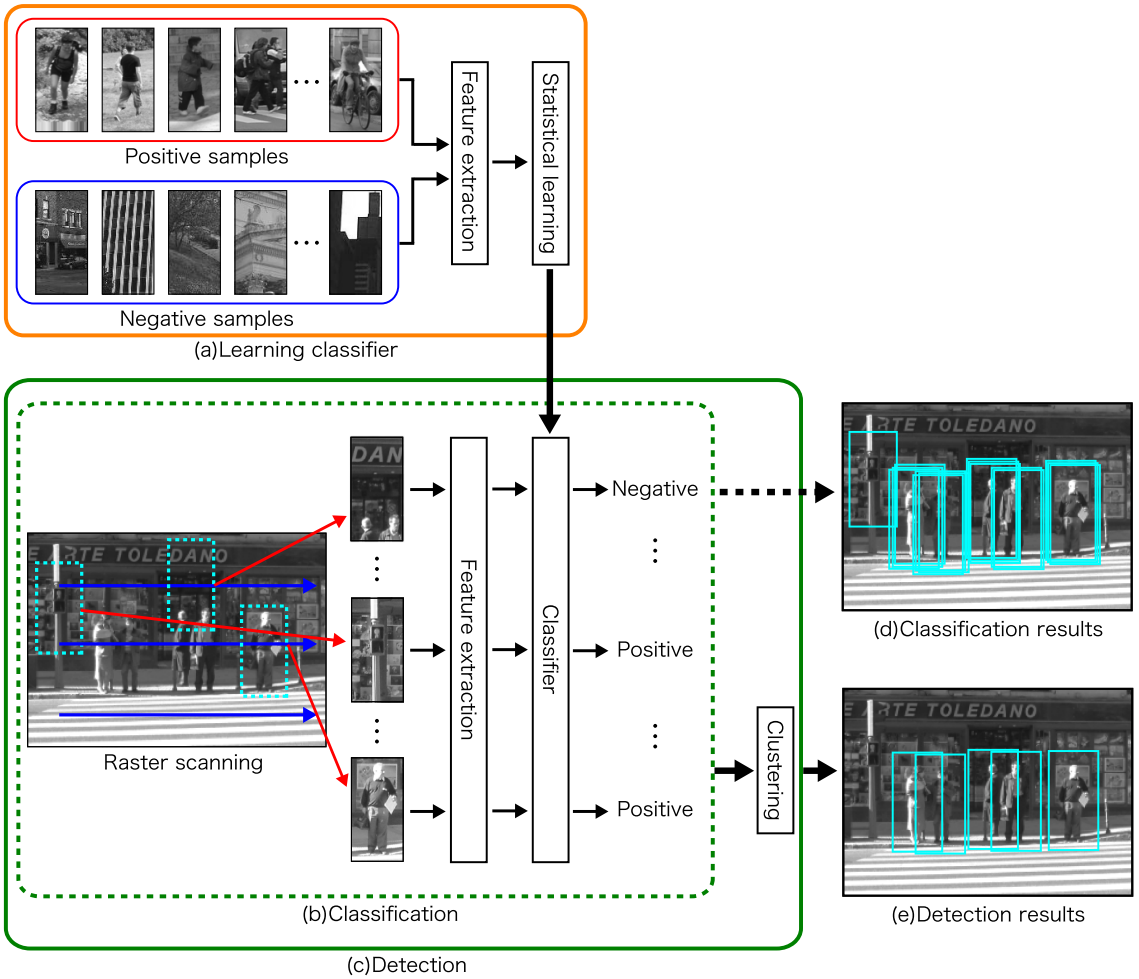

図 1　学習と検出の流れ
Fig. 1　Process of training and detection.

メラは，近赤外線カメラと遠赤外線カメラの 2 種類がある．近赤外線カメラは，環境光や近赤外線投光器の近赤外線光が物体に反射した光を観測する．一方，遠赤外線カメラは物体が発する熱（遠赤外線）を観測する．両者共に，暗闇でも人の輪郭を明瞭に捉えることができる利点がある．

他にも三次元距離画像センサを用いた人検出法 [24], [25] が提案されている．三次元距離画像センサとしては，Time of Flight（TOF）カメラや Microsoft が開発した Kinect が利用され，これらは物体からカメラまでの距離を正確に取得することができる．三次元距離画像センサから得られる距離画像にはテクスチャ情報が含まれないため，複雑な背景テクスチャの悪影響を受けない．そのため，このような距離画像からは非常にはっきりとした人のシルエットを観測できるため，高精度な人検出を実現できる．

また，画像からの人検出とは若干逸脱するが，レーザレーダを用いた人検出法 [26], [27] が提案されている．

2.3　人検出の基本フレームワーク

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判定する処理，人と判定されたウィンドウを統合する処理の三つに分けられる．以下に各処理について簡単に述べる．

2.3.1　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) に識別器の学習の流れを示す．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

表 1　人検出を困難とする要因と対応策
Table 1　Factors to complicate the human detection, and its countermeasures.

要因	対応策		代表的な手法
	特徴量	識別器	
見えの個体差	こう配，色，動き，距離	-	HOG [14]，CSS [28]，HOF [29]
複雑な背景	特徴量の共起	-	Joint Haar-like [30]，CoHOG [31]，Joint HOG [32]
向きの変化	-	複数の識別器	Cluster Boosted Tree [33]
姿勢の変化	-	パーツベース	Deformable Parts Model [34]，Hough Forest [35]
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合せ [36]
視点の変化	-	幾何学情報の利用	シーンのモデリング [37]，転移学習 [38]

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．

2.3.2　識別器による判定

未知入力画像から人を検出するには，図 1 (b) に示すように未知入力画像中に検出ウィンドウを設定し，画像中をラスタスキャンしながら人若しくは背景と判定する．このとき，人とカメラの距離に応じて画像中の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，画像ピラミッドからラスタスキャンする．若しくは幾つかの大きさの違う検出ウィンドウを設定し，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．この場合，識別器側で画像の大きさの違いを対応する必要がある．画像中をラスタスキャンしながら識別器により判定することで図 1 (d) に示すような結果を得ることができる．

2.3.3　ウィンドウの統合

人が存在する領域周辺には，多数の検出ウィンドウが人として判定される．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．そこで，多くの場合には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．更に，クラスタリングする際に統合された検出ウィンドウの個数をしきい値処理等をすることで，図 1 (d) の左上の孤立した検出ウィンドウを除去することができる．検出ウィンドウの統合には，計算コストの低さから Mean Shift [39] が用いられることが多い．

2.4　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．その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 に示すように六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．更に，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二つに分類し，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二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も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いて着目する情報の違いから下記の四つに分けることができる．

- こう配情報に基づく特徴量
- 色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 こう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 [9], [40] 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こう配に着目した特徴量が提案されている [14], [41], [42]．Chen らは，こう配に着目した特徴量として Edge of Orientation Histograms（EOH）特徴量 [41] を提案している [43]．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 [42], [44] を提案している．ほかにも，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 [45] を利用した人検出法も提

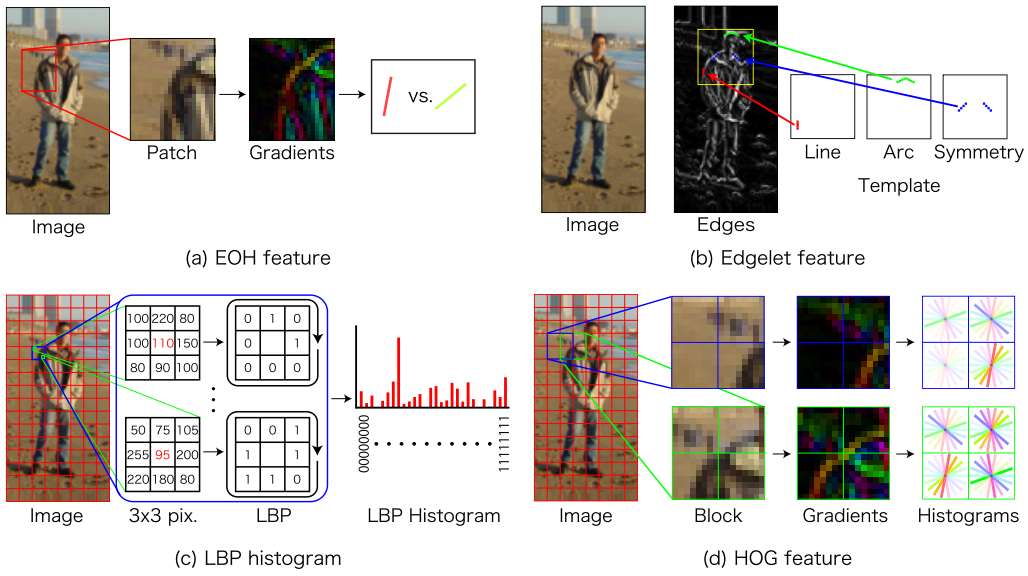

図 2 こう配情報に基づく特徴量
Fig. 2 Gradient-based features.

案されている [36], [46]～[48].

こう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 [14] が最も利用されている．HOG 特徴量は，局所領域（セル）ごとに輝度のこう配方向ヒストグラムを作成し，複数のセルから構成される領域（ブロック）ごとに正規化される．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，局所領域においてこう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 特徴量のようにこう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 [49] 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 [28], [29], [34], [36], [50] 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 [51] や異なる大きさの局所領域から HOG 特徴量とする Pyramid HOG（P-HOG）特徴量 [52]，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 [53]，エッジ強度の相関を考慮した Edge Similarity-based-HOG（ES-HOG）特徴量 [54] 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こう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，領域の輝度和やこう配和を高速に求められる積分画像 [9] や積分ヒストグラム [55] のメリットを生かし，こう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 [56] 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 [28] が提案されている．Walk らは入力画像における二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，二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

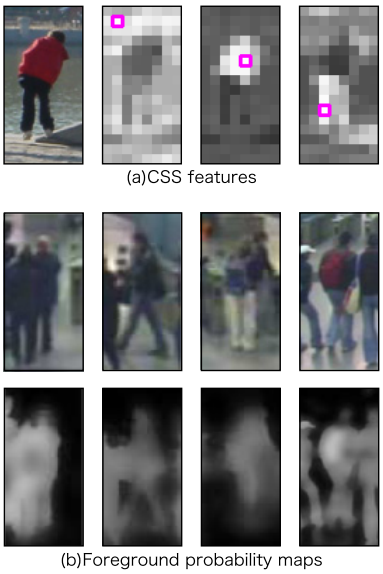

図 3 CSS 特徴量 [28] と前景ゆう度マップ [58] の可視化画像の例

Fig. 3 Visualized images of CSS features [28] and foreground probability maps [58].

選択した方形領域との色の類似度が高いことを表している．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分かる．ほかにも，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 特徴量を抽出する CS-HOG 特徴量 [57] も提案されている．CSS 特徴量は色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情報を効果的に利用した特徴量といえる．

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三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．三つ目は，空間的変化と時間的変化の両方を捉える時空間特徴量である．

空間的変化を捉える特徴量 [58] は固定カメラを想定している．Yao らは，背景差分をベースとした手法 [59] から得られる前景のゆう度を特徴量として利用している [58]．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．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をベースとした手法が提案されている [13], [29]．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 [13]．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 [29]．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．

空間的な変化と時間的な変化の両方を捉える時空間特徴量を用いた人検出法 [60]～[62] が提案されている．村井らは STpatch 特徴量 [63] を用いた人検出法を提案している [61]．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．

3.1.4 距離情報に基づく特徴量

近年，簡単に距離情報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（Time of Flight）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 [24] を提案している．図 4 に示すように二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 [64]．Xia らは，Kinect から得られる距離画像を用いて 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 [25]．

TOF カメラや Kinect から得られる距離情報は，カメラから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を観測しない．そのため，人の形状のみ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できる．

3.1.5 人らしさを捉える局所特徴量のまとめ

本項では，人らしさを捉える局所特徴量の性質から四つのアプローチに分類し，それぞれのアプローチの代表的な手法について述べた．最後に，それぞれのアプローチ間の得失について述べる．表 2 に分類した四つの特徴量の性質を整理した．こう配に基づく特徴量は，抽出する際の条件がなく，高い検出性能が得やすいために汎用性が高い．一方，色や動きに基づく特徴量は，幾つかの条件があるため利便性は高くない．ま

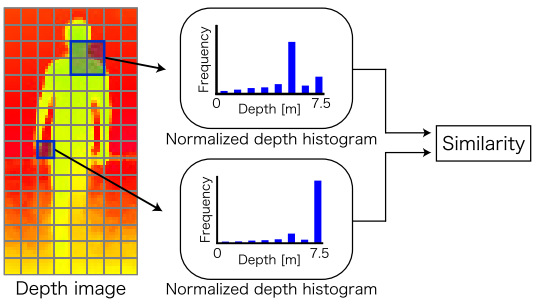

図 4 距離ヒストグラムの類似度 [24]
Fig. 4 Relational Depth Similarity Feature [24].

表 2 人らしさを捉える局所特徴量の性質の比較
Table 2 Comparison of property of local features.

局所特徴量	検出性能	利便性	環境変動に対する頑健性	条件
こう配	中	高	高	
色	低	中	中	カラー画像
動き	中	低	中	人若しくはカメラの動きに制約あり 時間的に連続した複数枚の画像
距離	高	低	中	距離計測システム

た，その特徴量単体で使用すると十分な検出性能が発揮できない場合がある．そのため，こう配に基づく特徴量との併用して用いられることが多い．距離に基づく特徴量については，現在までに提案されている手法では，単体の特徴量でも十分な検出性能が得られることが報告されている．しかしながら，距離を得るためにはステレオカメラや距離計測デバイスを必要とする．

これらの特徴量は，併用することで高い検出性能が得られる一方，計算コストやカメラのためのコストが高くなる．そのため，要求される人検出性能との兼ね合いで用いる特徴量を選択する必要がある．

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，人に似た形状をもつ複雑なテクスチャをも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，あらかじ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 [31], [65] 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 に示すようなある局所領域における 2 点の画素から得られるこう配方向の組合せを累積した同時生起行列である．同時生起行列で表される CoHOG 特徴量は，局所領域においてこう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 [66]．また，同様のアプローチで Local Binary Pattern（LBP）特徴量 [45] の共起を表現した手法も

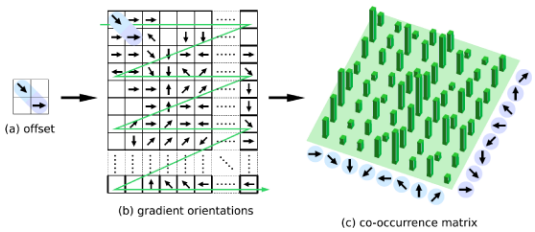

図 5　CoHOG 特徴量 [65]
Fig. 5　CoHOG features [65].

提案されている [67]．ほかにも，Tuzel らは画素の位置や輝度こう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 [68]．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，あらかじめ決められた特徴量の組合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せる手法 [30], [32], [69]～[71] 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 [30] を提案している．この方法は，まず複数の Haar-like 特徴量をしきい値処理することで 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．これにより，検出に有効な Haar-like 特徴量を共起させた Joint Haar-like 特徴量を生成できる．また，特徴量を 2 値化する以外にも，特徴量からクラスゆう度を計算し，得られた複数のゆう度を演算子により共起する手法 [72] 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 [69] 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

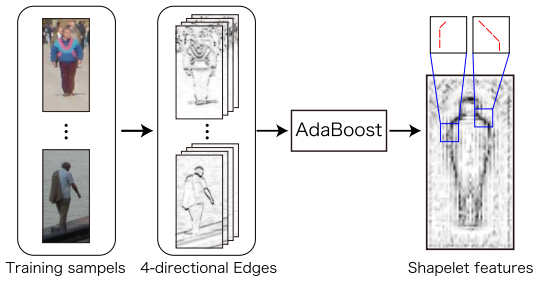

図 6　Shapelet 特徴量 [69]
Fig. 6　Shapelet features [69].

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こう配の共起性の両方を捉える Joint HOG 特徴量 [32] 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合せを統計的学習手法により評価することで，人検出に有効な特徴量の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．

3. 2. 3　人の構造を捉える特徴量のまとめ

本項では，人の構造を捉える特徴量として複数の局所特徴量の共起性を表現した特徴量について述べた．共起を表現する特徴量は，統計的学習手法を利用する方法と利用しない方法の二つに分けられる．

統計的学習手法を利用しない方法は，共起を表現した局所特徴量を統計的学習手法により識別器を学習する．そのため，生成された特徴量が非常に高次元な特徴量となるため，識別に多くの計算時間を要する．一方，統計的学習手法を利用する方法は，共起を表現する際に組み合わせる局所特徴量を評価する．これにより，全ての局所特徴量を組み合わせる必要がなくなるため高速な識別が可能であるが，学習に多くの時間を要する．学習及び識別の時間に大きな影響を与える特徴量の組合せ数は局所特徴量の次元数に依存しているため，どちらのアプローチを採用するか検討する際には局所特徴量の特徴次元数が参考になる．

4.　識 別 手 法

3. 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とられる．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

三次元物体である人は，正面から撮影した場合と横から撮影した場合で見えが異なる．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よい．このように各向きに対応した複数の識別器を用いる考え方は，顔の回転や向きに対応した顔検出の研究において既に提案されており，人検出においても同様のアプローチが適用されている．

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 [73] 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．そして，正面顔に変換した画像を識別する．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 [74], [75]．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 [51], [76], [77] 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）[33] と呼ぶ図 7 に示すような階層的な識別器を提案している．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，人の向きに合わせて学習サンプルを分割し，分割された学習サンプルを用いて識別器を構築するため，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 [78] を提

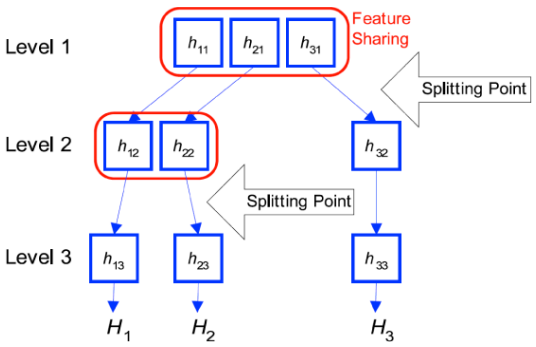

図 7　Cluster Boosted Tree [33]
Fig. 7　Cluster Boosted Tree [33].

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，あらかじ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．

向きに対応した手法についてまとめる．人の向きの変化に対応した手法は，それぞれの向きに対応した複数の識別器を用いることで対応することができる．初期の研究では，人手で学習サンプルの人画像に対して付与した向きのラベルに基づき，向きごとの識別器を学習した．しかしながら，向きのラベルを人手で付与するのは手間がかかるため，向きのラベルを付与する必要がない手法が提案された．この手法は，識別結果に応じて学習サンプルをクラスタリングするため，必ずしも人の向きに対応した識別器が学習されるとは限らないが，識別に適した形で学習サンプルが分割されるため，人手でクラスタリングした場合よりも高い検出性能が得られることが報告されている [33], [78].

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．

本節では，まずパーツベースの手法で採用される四つのパーツの定義について述べる．そして，三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
  - 統計的アプローチ
- 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，足などの一つひと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，いか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

人画像ごとにあらかじめ研究者が決めたルールに従いパーツを分割する [79]．例えば，人画像を頭部と右胴体，左胴体，下半身の四つに分割する．人画像ごと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 [44], [80]．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，人画像ごとに頭部や足，胴体の位置や形状は異なるため，あらかじ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 [35], [81]．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

図 8　Poselet の例 [82]．左図：Poselet の平均画像．右図：人のポーズの一部分を表す Poselet
Fig. 8　Examples of Poselet [82]. left: Mean poselet image, right: Poselet represents a part of the human pose.

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．

- 統計的アプローチ

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 [34], [82]．Bourdev らは，あらかじ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 [82] 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，関節位置の情報付加コストがかかる．ほかにも，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 [34]．

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，下記の三つに分けられる．

- 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) [79] 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，下半身の四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を吸収できる．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あらかじ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）[83] を導入した手法 [84]～[86] 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二つに分けることができる．

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，グラフを最ゆう法により学習している [87]．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ときのゆう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ゆう法により求める．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，

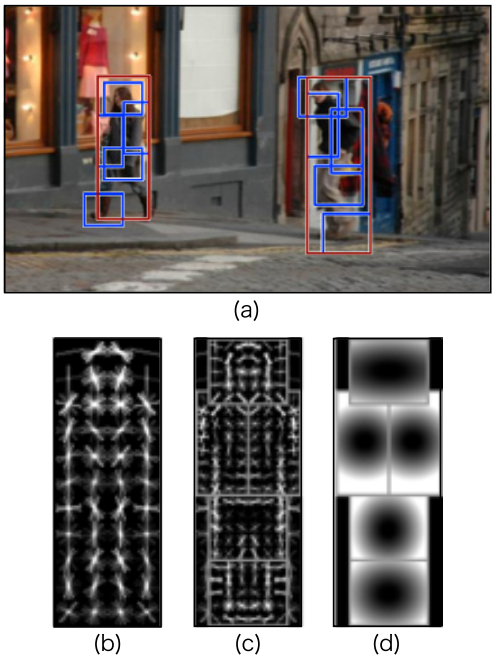

図 9　Deformable Parts Model を用いた人検出の例と各フィルタ [34]．(a) 人検出の例，(b) ルートフィルタ，(c) パーツフィルタ，(d) 各パーツからルートまでの位置関係

Fig. 9　Detection results using Deformable Parts Model and human model [34]. (a) Detection results, (b) Root filter. (c) Parts filter. (d) A spatial model for the location of each part relative to the root.

Constellation Model [88] 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 [89]．

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 [34], [90] 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め，Deformable Parts Model をベースとした人検出法 [91]～[94] や物体認識法 [95], [96] 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について述べる．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．

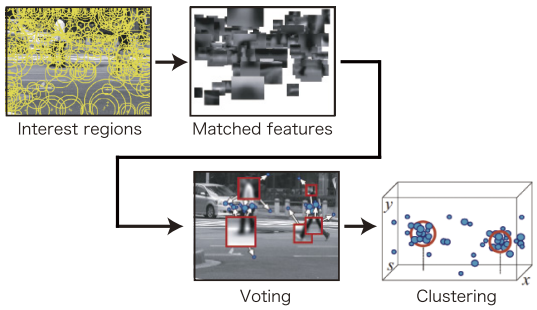

図 10　ISM による物体検出の流れ [97]

Fig. 10　Flow of human detection by ISM [97].

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）[81], [97], [98] 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あらかじ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 [63] 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 [60]．

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いか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 [35] を提案している．Hough Forests では，多クラスの識別が可能な Randam Forest [99] 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 [100]～[102] も提案されている．

4.2.3 姿勢の変化に対応した手法のまとめ

本項では，姿勢の変化に対応した手法としてパーツベースの人検出法について述べた．パーツベースの手法は，人画像を見えの変動が少なくなるように小さなパーツに分割し，人の形状構造を考慮してパーツを統合する．パーツベースの人検出法は，パーツの統合方法の観点から三つに分類することができる．現在までに多数のパーツベースの人検出法が提案されているが，その中でも，Deformable Parts Model はソースコードを Web 上で公開していることから誰もがテストすることができ，検出性能も非常に高いため，人検出の分野におけるデファクトスタンダードな手法として確立されつつある．

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つながるため，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 [36]．Wang らの手法は，識別時に局所領域ごと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 [39] 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分かる．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 [24]．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 [103]．

人領域の隠れに対応した手法についてまとめる．隠れに対応するために人領域の隠れに対するためには隠れ領域を推定した後，その隠れ領域を識別器で考慮するアプローチがとられる．このような 2 段階のアプローチをとる場合，いかに隠れ領域を正確に検知できるかが重要である．そのため，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．

4.4 視点の変化に対応した手法

カメラの高さや俯角が変わると，その大きさに応じて人の見えが変化する．このような見えの変化に対応する手法として，幾何学情報を利用する手法と転移学習を利用した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 [37]．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

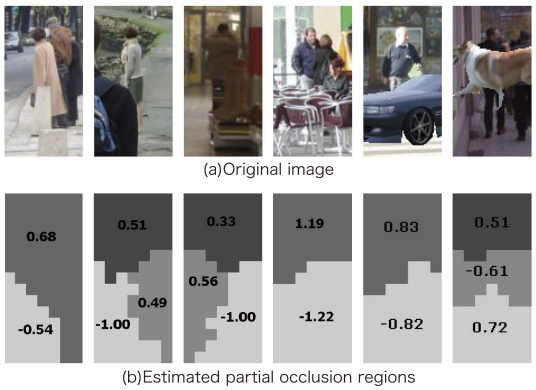

図 11 隠れ推定の結果 [36]．(a) 入力画像．(b) 隠れの推定結果．各領域は背景らしさを表すスコア

Fig. 11 Estimated partial occlusion regions [36]. (a) Original images. (b) Corresponding segmented occlusion likelihood images. For each segmented region, the negative sore.

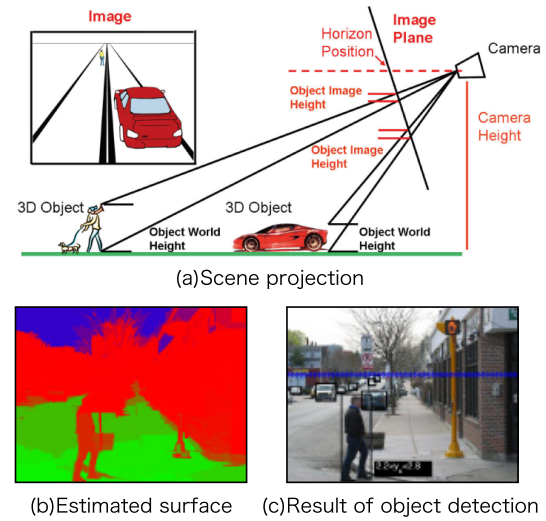

図 12 幾何学情報を利用した人検出 [37]

Fig. 12 Human detection using geometry information [37].

とで視点の変化に対応する．図 12 (c) からも分かるように，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三つの要素を考慮する．そのために，三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 [38] 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．最適化は二つの処理に分けられ，一つ目の処理では 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．二つ目の処理では，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．Pang らの転移学習は，高速に識別器を最適化するために転移学習を行う際の特徴プールの大きさを限定していたが，より視点の変化に対応するために，視点変化による見えの違いに応じて特徴プールのサイズを自動的に決定するハイブリッド型転移学習 [104] も提案されている．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処理が必要ではあるが，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．

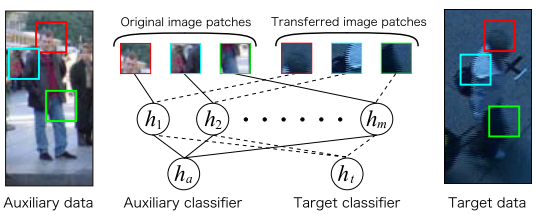

図 13　転移学習による識別器の最適化 [38]
Fig. 13　Optimizing classifier by Transfer learning [38].

視点の変化に対応した手法についてまとめる．幾何学情報を利用する手法では，人と周辺の背景の幾何学的な関係を利用することで多少の視点の変化に対応することができる．大きな視点の変化に対応するいちアプローチとしては転移学習を利用した手法があり，対象とするシーンの学習サンプルと再度識別器を学習する必要があるが大きな視点の変化に対応することができる．このように視点の変化に対応する手法は幾つか提案されているが，現在までに一つの識別器であらゆる視点の変化に対応する手法は知られていない．設置するカメラの高さや俯角によって容易に視点の変化が生じて人の見えが変化するため，複数の場所に人検出システムを設置することを考えた場合，それぞれの環境で視点の変化に対応した手法を適用するしかない．そのため，今後はあらゆる視点の変化に対応することができるような利便性の高い手法が望まれる．

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ができる．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三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
- ひずみへの対応

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，ラスタスキャンの三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 [9] が利用されている．こう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 [55] を利用することで，高速な人検出を実現できる [105]．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 [105]．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こう配情報を計算するだけで済むため，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 [56] 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 [9] が採用されている．Zhu らは，局所領域ごとに 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 [105]．ほかにも，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 [43], [87], [91] が提案されている．これらの手法は，同じ情報量を観測する識別器をカスケード状に並べるが，異なる情報量をもつ識別器を並べることで高速な識別を実現した手法が提案されている．Zhang らは解像度の異なる複数の識別器を用いた効率の良い識別手法を提案した [106]．Zhang らは，低解像度画像から計算する特徴量は低計算量で得られることに着目し，まず低解像度画像から得られる特徴量により識別し，人と判定された場合のみ高解像度画像の特徴量を用いて識別する．これにより約 8 倍の高速化を実現した．

人を検出するために画像中を網羅的に検出ウィンドウをラスタスキャンさせるため，膨大な数のウィンドウを処理することになる．そのため，効率的にラスタスキャンすることで高速化を実現する方法が提案されている．Lampert らは，分枝限定法による効率的なラスタスキャン方法を提案した [107]．Lampert らの手法は，まず画像を探索する最小範囲と最大範囲を定義し，その範囲における最大スコアを求める．そして，最大範囲の中から逐次的に探索する際，最大スコアの値を利用して枝狩りしながら効率的にラスタスキャンすることができる．ほかにも，幾何的な情報を用いて検出ウィンドウのラスタスキャンする範囲を限定する方法も提案されている．池村らは，TOF カメラから得られる距離情報を用いて検出ウィンドウのラスタスキャンする範囲を限定した [24]．実空間における人の身長を仮定し，検出ウィンドウの大きさを固定しながら三次元空間をラスタスキャンした．これにより，複数の解像度に対してラスタスキャンする必要がなくなるため，高速な検出を実現している．Benenson らは，人は地面に立っているという仮定に基づき，ステレオ視に基づき推定した地面とカメラから人までの距離から検出ウィンドウのラスタスキャン範囲を限定した [108]．

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 [109]～[111] 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．

最後に，これまでに提案されている人検出法の計算速度について紹介する．Dalal らの HOG 特徴量と SVM に基づく人検出法 [14] では，640 × 480 画素の画像に対して人検出した場合，0.24 FPS で処理可能である．Dollar らの Integral Channel Features 特徴量と Soft cascade Boosting に基づく人検出法 [56] では 1.18 FPS，効率的にマルチスケール画像をラスタスキャンする方法 [112] では 6.4 FPS で処理できる．更に，ステレオ視に基づくラスタスキャンエリアの制限，GPU の使用を用いた手法 [111] では 135 FPS で処理することができる．これらのフレームレートは，画像サイズや検出ウィンドウのラスタスキャン幅，検出ウィンドウのスケール幅にかなり影響を受けることに注意されたい．

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

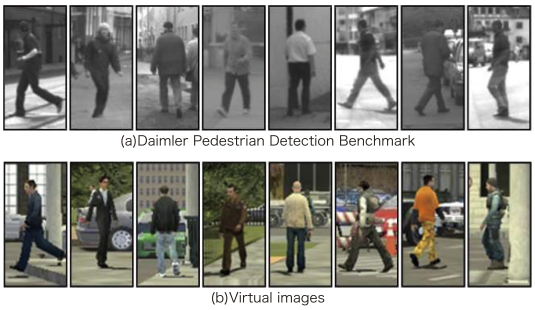

図 14　(a) 実画像．(b) 仮想画像 [114]
Fig. 14　(a) Real images. (b) Virtual images [114].

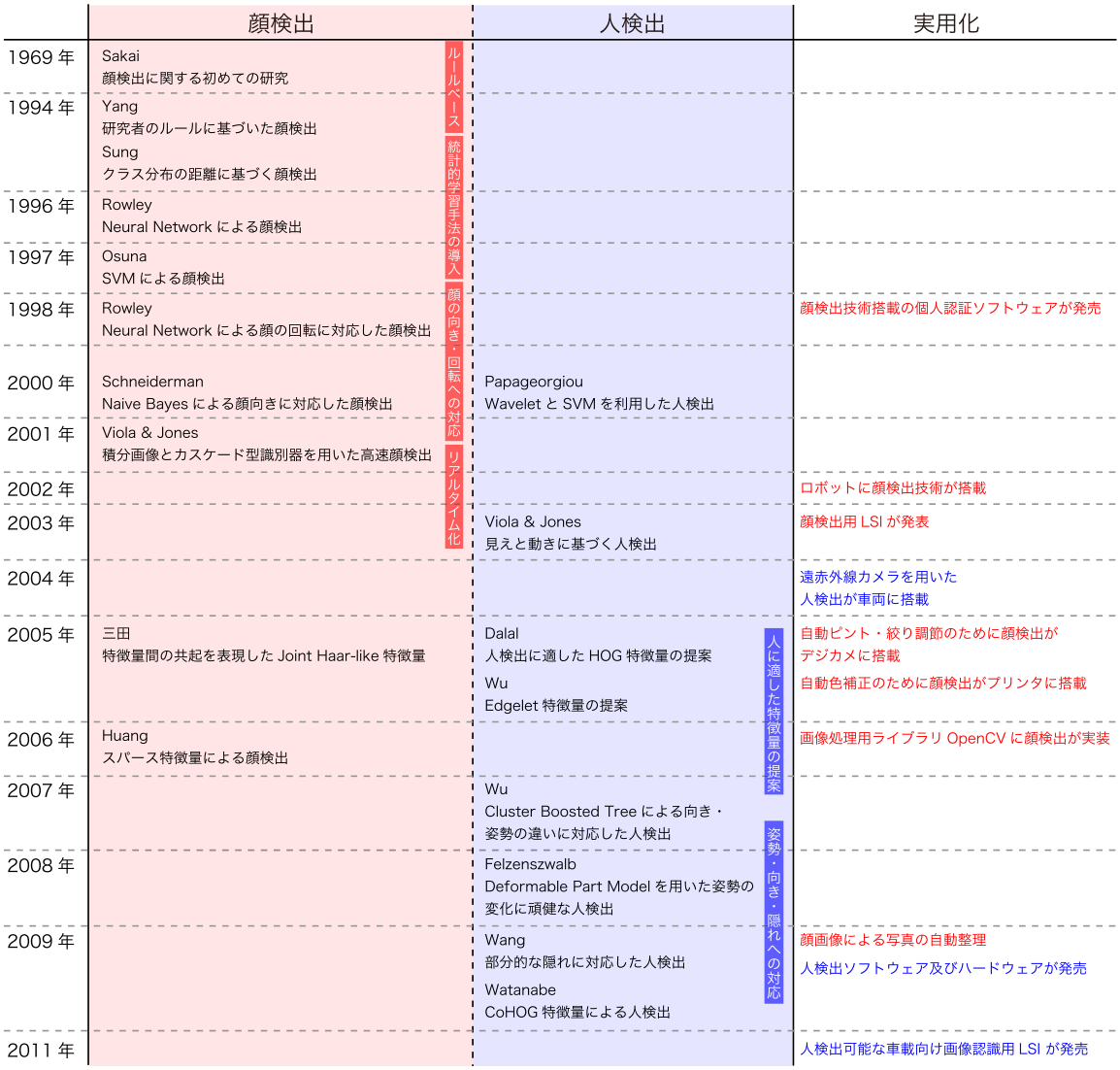

図 15　物体検出技術の遷移
Fig. 15　History of object detection.

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別器を学習する生成型学習 [113] を利用した人検出法 [114]～[116] 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 [114]．このとき，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 [115]．

5.3　ひずみへの対応

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ひずみが発生する場合がある．ひずみ

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，ひずみの影響を取り除く必要がある．

Li らは，遠近法によるひずみの影響で画像の 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 [117]．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，レンズによるひず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．遠近法によるひず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ひずみのない画像を用いて識別する．これにより，遠近法によるひず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ひずみの影響を抑制した手法として，Smart Window Transform [118] 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多くない．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．図 15 に物体検出技術の変遷と実用化例を示す．

2004 年には，自動車に設置された遠赤外線カメラを用いて歩行者を検出し，歩行者の存在を運転者に伝える機能が世界で初めて自動車に搭載された [119]．これにより，人間が視認することが難しい夜間の運転の安全リスクを減らすことができる．また，2008 年には近赤外線カメラを用いて人を検出する機能が搭載された自動車が発売されている [120]．2010 年には可視光カメラのみを用いた人検出技術が搭載された歩行者衝突警報が発表された [121]．

また．近年では人検出技術をハードウェア化する取組みも盛んである．2010 年には LSI や FPGA に実装可能な人検出機能ハードウェア論理回路が発売された [122]．2011 年には人検出可能な車載向け画像認識用 LSI を発売されている [123]．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．本章では，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 3 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，若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 [124] がある．MIT CBCL Pedestrian Data は，人を正面若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ともに INRIA Person Dataset [14] 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．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，人がディジタルカメラ

表 3　人画像データベースの比較
Table 3　Comparing human image databases.

	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 [124]	924	-	-	-	-
INRIA [14]	2,416	1,218	288	1,132	453
USC-A [44]	-	-	205	303	-
USC-B [44]	-	-	54	271	-
USC-C [33]	-	-	100	232	-
ETH [125]	1,578	-	1,803	9,380	-
Daimler2006 [126]	14,400	150,000*	-	1,600	100,000
Daimler2009 [15]	15,660	6,744	21,800	56,492	-
NICTA [127]	18,700	5,200*	-	6,900	50,000
TUD [128]	400	-	250	311	-
Caltech [16]	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す．

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 [15], [16], [126] もある．その中でも，Caltech Pedestrian Detection Benchmark [16] 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，下記の二つの指標がデータベースとともに提案されている．

（1） Miss rate VS. False Positive Per Window (FPPW) [14]

（2） Miss rate VS. False Positive Per Image (FPPI) [16]

（1）の FPPW は，一つの検出ウィンドウ当りの誤検出した検出ウィンドウ数を表す．例えば，検出ウィンドウをラスタスキャンした際に 10,000 ウィンドウについて識別し，10 ウィンドウの背景画像を人と誤識別した場合には 0.001 FPPW となる．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 枚の画像当りの誤検出した数を表す．例えば，100 枚の画像に対して人検出を行い，10 ウィンドウの切り出された背景画像を人と誤識別した場合には 0.1 FPPI となる．FPPI

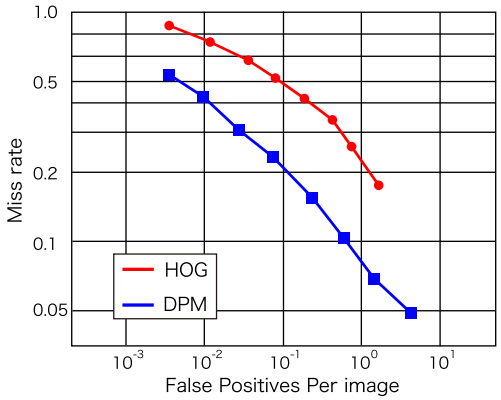

図 16　DET カーブの比較の例
Fig. 16　Example of comparison of DET curves.

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．二つの評価指標は一見同じようにもとれるが，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（Miss rate）と誤検出率（FPPW 若しくは FPPI）を比較する．一般的に人検出の性能の比較においては，識別器のしきい値を変化させたとき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す．図 16 に DET カーブの例を示す．図 16 の DET カーブは，HOG+SVM [14] と Deformable Parts Model（DPM）+ HOG + Latent SVM [34] の結果である．結果から，DPM のカーブの方がグラフの左下に位置しているため，DPM の性能が HOG 単体よりも性能が高いことが分かる．

7. むすび

本論文では，人検出法のサーベイと人検出用データベース並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六つに分け，各要因を解決するための特徴抽出法と識別手法の二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．

最後に，最新の研究動向と社会情勢から人検出の今後について考える．2012 年の CVPR にて幾つかの人検出に関する研究が発表された．中でも，検出するシーンに合わせて識別器を最適化する等の新しいテーマを扱った研究も発表されている [111], [129]～[132]．これは人検出技術が向上したため実用化し得るという見通しがたち，実用化する際に発生し得る問題を解決するような研究に遷移していると考えられる．

また，これは社会情勢からみても同様のことがいえる．例えば，ヨーロッパで実施されている自動車衝突安全テスト EURO NCAP では，自動車の安全性を測るためのテストの一つとして，パッシブセーフティ機能による歩行者保護性能カテゴリーがある．消費者団体によって実施されるテストの結果を消費者が見て，自動車選択の指針とすることができるため，非常に重要視されているテストである．そのため，より一層と人検出技術が必要とされていることが読み取れる．この時流は自動車業界のみならず，他の業界においてもこの流れを追従すると思われる．

このように，人検出の研究は著しい進化を遂げ，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は実用化されている．今後は，実用化する際に発生するような問題を解決する取組みも増え，多様な分野で人検出技術の実用化が進むと考えられる．

文　献

[1] 井尻善久，川西康友，美濃導彦，村瀬　洋，“サーベイ論文：視野を共有しない複数カメラ間での人物照合，” 信学技報，PRMU2011-21 (PRMU), pp.117–124, 2011.

[2] T. Sakai, M. Nagao, and S. Fujibayashi, “Line extraction and pattern detection in a photograph,” J. Pattern Recognition, vol.1, pp.233–248, 1969.

[3] V. Govindaraju, S.N. Srihari, and D.B. Sher, “A computational model for face location,” IEEE International Conference on Computer Vision, pp.718–721, 1990.

[4] G. Yang and T.S. Huang, “Human face detection in a complex background,” J. Pattern Recognition, vol.27, no.1, pp.53–63, 1994.

[5] C. Kotropoulos and I. Pitas, “Rule-based face detection in frontal views,” International Conference on Acoustics, Speech, and Signal Processing, vol.4, pp.2537–2540, 1997.

[6] K.K. Sung and T. Poggio, “Example-based learning for view-based human face detection,” Technical Report MIT AI Lab, 1994.

[7] H.A. Rowley, S. Baluja, and T. Kanade, “Neural network-based face detection,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.203–208, 1996.

[8] H. Schneiderman and T. Kanade, “A statistical method for 3D object detection applied to faces and cars,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2000.

[9] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.511–518, 2001.

[10] E. Osuna, R. Freund, and F. Girosi, “Training support vector machines: An application to face detection,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.130–136, 1997.

[11] P. Viola and M. Jones, “Robust real-time object detection,” Int. J. Comput. Vis., vol.57, no.2, pp.137–154, 2004.

[12] C. Papageorgiou and T. Poggio, “A trainable system for object detection,” Int. J. Comput. Vis., vol.38, no.1, pp.15–33, 2000.

[13] P. Viola, M. Jones, and D. Snow, “Detecting pedestrians using patterns of motion and appearance,” IEEE International Conference on Computer Vision, pp.734–741, 2003.

[14] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp.886–893, 2005.

[15] M. Enzweiler and D.M. Gavrila, “Monocular pedestrian detection: Survey and experiments,” IEEE Trans. Pattern Anal. Mach. Intell., vol.31, no.12, pp.2179–2195, 2009.

[16] P. Dollár, C. Wojek, B. Schiele, and P. Perona, “Pedestrian detection: An evaluation of the state of

the art," IEEE Trans. Pattern Anal. Mach. Intell., vol.34, no.4, pp.743–761, 2012.

[17] D. Gerónimo, A.M. López, A.D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," IEEE Trans. Pattern Anal. Mach. Intell., vol.32, no.7, pp.1239–1258, July 2010.

[18] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.246–252, 1998.

[19] T. Bouwmans, "Recent advanced statistical background modeling for foreground detection — A systematic survey," Recent Patents on Computer Science, pp.147–176, 2011.

[20] P.J. Besl and N.D. McKay, "A method for registration of 3-D shapes," IEEE Trans. Pattern Anal. Mach. Intell., vol.14, no.2, pp.239–256, 1992.

[21] O. Chum and A. Zisserman, "An exemplar model for learning object classes," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 2007.

[22] F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi, "Pedestrian detection using infrared images and histograms of oriented gradients," IEEE Intelligent Vehicles Symposium, 2006.

[23] L. Zhang, B. Wu, and R. Nevatia, "Pedestrian detection in infrared images based on local shape features," IEEE International Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum (in conjunction with CVPR2007), pp.1–8, 2007.

[24] 池村　翔，藤吉弘亘，"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（D），vol.93-D, no.3, pp.355–364, March 2010.

[25] L. Xia, C.C. Chen, and J. Aggarwal, "Human detection using depth information by kinect," International Workshop on Human Activity Understanding from 3D Data (in conjunction with CVPR), pp.15–22, 2011.

[26] L. Navarro-Serment, C. Mertz, and M. Hebert, "Pedestrian detection and tracking using three-dimensional ladar data," Int. J. Robotics Research, vol.29, no.12, pp.1516–1528, 2010.

[27] 城殿清澄，内藤貴志，三浦　純，"高解像度レーザレーダと画像の統合による歩行者認識，" 画像センシングシンポジウム，2012.

[28] S. Walk, N. Majer, K. Schindler, and B. Schiele, "New features and insights for pedestrian detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.1030–1037, 2010.

[29] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," European Conference on Computer Vision, vol.2, pp.428–441, 2006.

[30] T. Mita, T. Kaneko, B. Stenger, and O. Hori, "Discriminative feature co-occurrence selection for object detection," IEEE Trans. Pattern Anal. Mach. Intell., vol.30, no.7, pp.1257–1269, 2008.

[31] T. Watanabe, S. Ito, and K. Yokoi, "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.

[32] 三井相和，山内悠嗣，藤吉弘亘，"Joint 特徴量を用いた 2 段階 Boosting による物体検出，" 信学論（D），vol.J92-D, no.9, pp.1591–1601, Sept. 2009.

[33] B. Wu and R. Nevatia, "Cluster boosted tree classifier for multi-view, multi-pose object detection," IEEE International Conference on Computer Vision, pp.1–8, 2007.

[34] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part based models," IEEE Trans. Pattern Anal. Mach. Intell., vol.32, no.9, pp.1627–1645, 2009.

[35] J. Gall and V. Lempitsky, "Class-specific Hough forests for object detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.1022–1029, 2009.

[36] X. Wang, T.X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," IEEE International Conference on Computer Vision, 2009.

[37] D. Hoiem, A.A. Efros, and M. Hebert, "Putting objects in perspective," Int. J. Comput. Vis., vol.80, no.1, pp.3–15, 2008.

[38] J. Pang, Q. Huang, S. Yan, S. Jiang, and L. Qin, "Transferring boosted detectors towards viewpoint and scene adaptiveness," IEEE Trans. Image Process., vol.20, no.5, pp.1388–400, 2011.

[39] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol.24, no.5, pp.603–619, 2002.

[40] 佐部浩太郎，日台健一，"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，" 画像センシングシンポジウム（SSII），2004.

[41] K. Levi and Y. Weiss, "Learning object detection from a small number of examples: The importance of good features," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.2, pp.53–60, 2004.

[42] B. Wu and R. Nevatia, "Detection and segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses," Int. J. Comput. Vis., vol.82, no.2, pp.185–204, 2009.

[43] Y.T. Chen and C.S. Chen, "A cascade of feed-forward classifiers for fast pedestrian detection," Asian Conference on Computer Vision, pp.905–914, 2007.

[44] B. Wu and R. Nevatia, "Detection of multiple,

partially occluded humans in a single image by Bayesian combination of edgelet part detectors," IEEE International Conference on Computer Vision, pp.90–97, 2005.

[45] W. Li and H. Dong-Chen, "Texture classification using texture spectrum," J. Pattern Recognition, vol.23, no.8, pp.905–910, 1990.

[46] Y.D. Mu, S.C. Yan, Y. Liu, T. Huang, and B.F. Zhou, "Discriminative local binary patterns for human detection in personal album," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.1–8, 2008.

[47] 阿部　厳，岡谷貴之，出口光一郎，"カーネル部分最小二乗法分析を用いた歩行者認識," 情処学研報（コンピュータビジョンとイメージメディア），CVIM-21, 2010.

[48] 西村洋文，曹　芸芸，南里卓也，黒河　久，司岡　兼，"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.

[49] 鶴岡信治，栗田昌徳，原田智夫，木村文隆，三宅康二，"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（D），vol.J70-D, no.7, pp.1390–1397, July 1987.

[50] Z. Lin and L.S. Davis, "A pose-invariant descriptor for human detection and segmentation," European Conference on Computer Vision, 2008.

[51] C. Hou, H. Ai, and S. Lao, "Multiview pedestrian detection based on vector boosting," Asian Conference on Computer Vision, pp.210–219, 2007.

[52] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," International Conference on Image and Video Retrieval, 2007.

[53] P. Ott and M. Everingham, "Implicit color segmentation features for pedestrian and object detection," IEEE International Conference on Computer Vision, 2009.

[54] 河合富美，H.C. Keng，"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," 画像の認識・理解シンポジウム（MIRU），pp.2084–2091, 2010.

[55] F. Porikli, "Integral histogram: A fast way to extract histograms in cartesian spaces," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp.829–836, 2005.

[56] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," British Machine Vision Conference, 2009.

[57] 後藤雄飛，山内悠嗣，藤吉弘亘，"色の類似性に基づいた形状特徴量 CS-HOG の提案," 画像センシングシンポジウム，2012.

[58] J. Yao and J.M. Odobez, "Fast human detection from videos using covariance features," Visual Surveillance Workshop (in conjunction with ECCV2008), 2008.

[59] J. Yao and J.M. Odobez, "Multi-layer background subtraction based on color and texture," Computer Vision and Pattern, Recognitionisual Surveillance Workshop, 2007.

[60] 村井泰裕，藤吉弘亘，金出武雄，"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（コンピュータビジョンとイメージメディア），vol.1, no.2, pp.21–31, 2008.

[61] 村井泰裕，藤吉弘亘，数井誠人，"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," 信学技報，PRMU2008-42, 2008.

[62] Y. Yamauchi, H. Fujiyoshi, Y. Iwahori, and T. Kanade, "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.

[63] E. Shechtman and M. Irani, "Space-time behavior-based correlation-or-how to tell if two underlying motion fields are similar without computing them?," IEEE Trans. Pattern Anal. Mach. Intell., vol.29, no.11, pp.2045–2056, 2007.

[64] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 2011.

[65] T. Watanabe, S. Ito, and K. Yokoi, "Co-occurrence histograms of oriented gradients for pedestrian detection," Pacific-Rim Symposium on Image and Video Technology, pp.37–47, 2009.

[66] H. Hattori, A. Seki, M. Nishiyama, and T. Watanabe, "Stereo-based pedestrian detection using multiple patterns," British Machine Vision Conference, vol.243, 2009.

[67] R. Nosaka, Y. Ohkawa, and K. Fukui, "Feature extraction based on co-occurrence of adjacent local binary patterns," Pacific-Rim Symposium on Image and Video Technology, 2011.

[68] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on riemannian manifolds," IEEE Trans. Pattern Anal. Mach. Intell., vol.30, no.10, pp.1713–1727, 2008.

[69] P. Sabzmeydani and G. Mori, "Detecting pedestrians by learning shapelet features," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.1–8, 2007.

[70] C. Huang, H. Ai, Y. Li, and S. Lao, "Learning sparse features in granular space for multi-view face detection," International Conference on Automatic Face and Gesture Recognition, pp.401–406, 2006.

[71] G. Duan, C. Huang, H. Ai, and S. Lao, "Boosting associated pairing comparison features for pedestrian detection," International Workshop on Visual Surveillance (in Conjunction with International Conference on Computer Vision), 2009.

[72] 山内悠嗣，山下隆義，藤吉弘亘，"Boosting に基づく特徴量の共起表現による人検出," 信学論（D），vol.J92-D,

no.8, pp.1125–1134, Aug. 2009.

[73] H.A. Rowley, S. Baluja, and T. Kanade, "Rotation invariant neural network-based face detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.38–44, 1998.

[74] M. Jones and P. Viola, "Fast multi-view face detection," Mitsubishi Electric Research Lab Technical Report, 2003.

[75] S. Li, X.H. Peng, X.W. Hou, H. Zhang, and Q. Cheng, "Multi-view face pose estimation based on supervised ISA learning," International Conference on Automatic Face and Gesture Recognition, pp.100–105, 2002.

[76] S.Z. Li, L. Zhu, Z. Zhang, A. Blake, H. Zhang, and H. Shum, "Statistical learning of multi-view face detection," European Conference on Computer Vision, 2002.

[77] C. Huang, H. Ai, Y. Li, and S. Lao, "Vector boosting for rotation invariant multi-view face detection," IEEE International Conference on Computer Vision, vol.1, pp.446–453, 2005.

[78] 土屋成光，藤吉弘亘，"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，" 信学技報，PRMU2009-13, 2009.

[79] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-based object detection in images by components," IEEE Trans. Pattern Anal. Mach. Intell., vol.23, no.4, pp.349–361, 2001.

[80] Z. Lin, L. Davis, and D. Doermann, "Hierarchical part-template matching for human detection and segmentation," IEEE International Conference on Computer Vision, 2007.

[81] B. Leibe and B. Schiele, "Interleaved object categorization and segmentation," British Machine Vision Conference, pp.759–768, 2003.

[82] L. Bourdev and J. Malik, "Poselets: Body part detectors trained using 3D human pose annotations," IEEE International Conference on Computer Vision, 2009.

[83] T. Dietterich, R. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," Artif. Intell., vol.89, pp.31–71, 1997.

[84] Z. Lin, G. Hua, and L. Davis, "Multiple instance feature for robust part-based object detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.1–8, 2009.

[85] P. Dollár, B. Babenko, S. Belongie, P. Perona, and Z. Tu, "Multiple component learning for object detection," European Conference on Computer Vision, pp.211–224, 2008.

[86] Y.T. Chen and C.S. Chen, "Multi-class multi-instance boosting for part-based human detection," International Workshop on Visual Surveillance (in conjunction with ICCV2009), pp.1177–1184, Sept. 2009.

[87] X. Xia, W. Yang, H. Li, and S. Zhang, "Part-based object detection using cascades of boosted classifiers," Asian Conference on Computer Vision, 2009.

[88] M. Burl and P. Perona, "Recognition of planar object classes," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.223–230, 1996.

[89] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.2, pp.264–271, 2003.

[90] P.F. Felzenszwalb, D. Mcallester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2008.

[91] P.F. Felzenszwalb, R. Girshick, and D. McAllester, "Cascade object detection with deformable part models," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.2241–2248, 2010.

[92] P. Ott and M. Everingham, "Shared parts for deformable part-based models," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010.

[93] M. Pedersoli, A. Vedaldi, and J. Gonzalez, "A coarse-to-fine approach for fast deformable object detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2011.

[94] H. Cho, P.E. Rybski, A. Bar-Hillel, and W. Zhang, "Real-time pedestrian detection with deformable part models," IEEE Intelligent Vehicles Symposium, pp.1035–1042, 2012.

[95] L. Zhu, Y. Chen, A. Yuille, and W. Freeman, "Latent hierarchical structural learning for object detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010.

[96] M. Sadeghi and A. Farhadi, "Recognition using visual phrases," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.1745–1752, 2011.

[97] B. Leibe, A. Leonardis, and S. Bernt, "Robust object detection with interleaved categorization and segmentation," Int. J. Comput. Vis., vol.77, no.1-3, pp.259–289, 2008.

[98] B. Leibe, A. Leonardis, and B. Schiele, "Combined object categorization and segmentation with an implicit shape model," Statistical Learning in Computer Vision, (in conjunction with ECCV), 2004.

[99] L. Breiman, "Random forests," Mach. Learn., vol.45, no.1, pp.5–32, 2001.

[100] K. Vijay and I. Patras, "A discriminative voting scheme for object detection using Hough forests," British Machine Vision Conference Postgraduate

Workshop, 2010.
[101] 波部 斉，ロベルト チポラ，“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,” 画像の認識・理解シンポジウム MIRU2011, 2011.
[102] 波部 斉，“スケール不変 Hough Forests による物体検出,” 画像センシングシンポジウム，2011.
[103] M. Enzweiler, A. Eigenstetter, B. Schiele, and D. Gavrila, “Multi-cue pedestrian classification with partial occlusion handling,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.990–997, 2010.
[104] 土屋成光，山内悠嗣，山下隆義，藤吉弘亘，“ハイブリッド型転移学習による物体検出における学習の効率化,” 信学技報，PRMU2013-52, 2013.
[105] Q. Zhu, S. Avidan, M.C. Yeh, and K.T. Cheng, “Fast human detection using a cascade of histograms of oriented gradients,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.1491–1498, 2006.
[106] W. Zhang, G. Zelinsky, and D. Samaras, “Real-time accurate object detection using multiple resolutions,” IEEE International Conference on Computer Vision, 2007.
[107] C.H. Lampert, M.B. Blaschko, and T. Hofmann, “Beyond sliding windows: Object localization by efficient subwindow search,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 2008.
[108] R. Benenson, M. Mathias, R. Timofte, and L.V. Gool, “Fast stixel computation for fast pedestrian detection,” IEEE Workshop on Computer Vision in Vehicle Technology: From Earth to Mars (in Conjunction with ECCV2012), 2012.
[109] B. Bilgic, B.K.P. Horn, and I. Masaki, “Fast human detection with cascaded ensembles on the GPU,” IEEE Intelligent Vehicles Symposium, pp.325–332, 2010.
[110] V.A. Prisacariu and I. Reid, “fastHOG — A real-time GPU implementation of HOG,” Technical Report, Oxford University, 2009.
[111] R. Benenson, M. Mathias, R. Timofte, and L.V. Gool, “Pedestrian detection at 100 frames per second,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.2903–2910, 2012.
[112] P. Dollár, S. Belongie, and P. Perona, “The fastest pedestrian detector in the west,” British Machine Vision Conference, 2010.
[113] 村瀬 洋，“画像認識のための生成型学習,” 情処学論（コンピュータビジョンとイメージメディア），vol.46, no.15, pp.35–42, 2005.
[114] J. Marín, V. David, D. Gerónimo, and Antonio M-López, “Learning appearance in virtual scenarios for pedestrian detection,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.137–144, 2010.
[115] Y. Yamauchi and H. Fujiyoshi, “Automatic generation of training samples and a learning method based on advanced MILboost for human detection,” Asian Conference on Pattern Recognition, pp.603–607, 2011.
[116] 吉田英史，出口大輔，井手一郎，村瀬 洋，後藤邦博，木村好克，内藤貴志，“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,” 信学技報，PRMU2011-37, 2011.
[117] Y. Li, B. Wu, and R. Nevatia, “Human detection by searching in 3D space using camera and scene knowledge,” International Conference on Pattern Recognition, 2008.
[118] 李 媛，伊藤誠也，三好雅則，藤吉弘亘，金子俊一，“Smart Window Transform とエッジベ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.
[119] HONDA インテリジェント・ナイトビジョンシステム，http://www.honda.co.jp/tech/auto/night-vision/.
[120] TOYOTA ナイトビュー，http://www.toyota.co.jp/jpn/tech/safety/technology/technology_file/active/night_view.html.
[121] Mobileye C2-270, http://www.mobileye.com/products/mobileye-c2-series/mobileye-c2-270/.
[122] 三栄ハイテックス株式会社 ODEN, http://www.sanei-hy.co.jp/product/index.html.
[123] 株式会社東芝画像認識用 LSI, http://www.toshiba.co.jp/rdc/rd/fields/11_t06.htm
[124] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio, “Pedestrian detection using wavelet templates,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.193–199, 1997.
[125] A. Ess, B. Leibe, and L. Van Gool, “Depth and appearance for mobile scene analysis,” IEEE International Conference on Computer Vision, 2007.
[126] S. Munder and D.M. Gavrila, “An experimental study on pedestrian classification,” IEEE Trans. Pattern Anal. Mach. Intell., vol.28, no.11, pp.1863–1868, 2006.
[127] G. Overett, L. Petersson, N. Brewer, L. Andersson, and N. Pettersso, “A new pedestrian dataset for supervised learning,” The Intelligent Vehicles Symposium, 2008.
[128] M. Andriluka, S. Roth, and B. Schiele, “People-tracking-by-detection and people-detection-by-tracking,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2008.
[129] M. Wang and X. Wang, “Automatic adaptation of a generic pedestrian detector to a specific traffic scene,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.3401–3408, 2011.
[130] P. Sharma and C. Huang, “Unsupervised incremental learning for improved object detection in a video,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.3298–3305, 2012.

[131] M. Wang and X. Wang, “Transferring a generic pedestrian detector towards specific scenes,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.3274–3281, 2012.
[132] X. Wang and T.X. Han, “Detection by detections: Non-parametric detector adaptation for a video,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.350–357, 2012.

（平成 24 年 11 月 15 日受付，25 年 5 月 6 日再受付）

山内　悠嗣　（正員）

2012 中部大学大学院博士後期課程了．博士（工学）．2012 より同大学博士研究員．2010 独立行政法人日本学術振興会特別研究員 DC．コンピュータビジョン，パターン認識の研究に従事．

山下　隆義　（正員）

2002 奈良先端科学技術大学院大学情報科学研究科修士課程了．同年オムロン（株）入社．2011 中部大学大学院博士後期課程了．顔及び人画像センシング技術の研究に従事．IEEE-CS，情報処理学会各会員．

藤吉　弘亘　（正員）

1997 中部大学大学院博士後期課程了．1997～2000 米カーネギーメロン大学ロボット工学研究所 Postdoctoral Fellow．2000 中部大学講師，2004 同大准教授を経て 2010 より同大教授．2005～2006 米カーネギーメロン大学ロボット工学研究所客員研究員．計算機視覚，動画像処理，パターン認識・理解の研究に従事．2005 ロボカップ研究賞．2009 情報処理学会論文誌コンピュータビジョンとイメージメディア優秀論文賞，2009 山下記念研究賞．2010，2013 画像センシングシンポジウム優秀学術賞．情報処理学会，電気学会，IEEE 各会員．