

人を観る技術：People Image Analysis -インビジブルロボットの視覚機能の実現-

藤吉 弘亘

中部大学工学部情報工学科

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

あらまし インビジブルロボットは、環境に配置されたカメラ等のセンサ群から刻々と変化する人の状態を認識し、ユーザである人に対して快適な空間をアシストすることが重要な課題である。このようなインビジブルロボットの実現には、特に、人画像解析 (People Image Analysis) として、動画像からの人の検出、追跡、顔の検出、顔の部位の追跡、モーション理解が不可欠な技術要素となる。現在までに、固定カメラによる背景差分に基づく人等を対象とした動体検出法は数多く提案されているが、このような動体検出をベースとしたアプローチでは、対象とする動体の検出に失敗すると次段の処理である物体識別が不可能となる問題がある。この問題を解決するアプローチとして、近年のコンピュータの高速化に伴い、画像全体を検出ウィンドウによってラスタスキャンし、low-level な特徴量と統計的学習手法の組み合わせによる物体検出法が提案されている。本稿では、インビジブルロボットのためのビデオ解析として、VSAM プロジェクトで開発された動画像理解アルゴリズムと、その実用化例について紹介する。また、人画像解析のための新しいアプローチとして、検出ウィンドウのラスタスキャンベースによる人検出法とモーション解析についても紹介する。

キーワード Video Surveillance and Monitoring(VSAM), 人画像解析, 人検出, モーション解析

1 はじめに

ロボットは、外界のデータを取り込み (感覚)、その意味を理解し (認識)、何をすべきかを判断し (決定)、結果として人に役立つように外界に働きかける (行動) システムとして定義される。環境の中に多数のロボットと人間がいて情報と空間を共有している状態を考えると、環境自体が記憶、通信、センサ、効果器としての機能を果たしている。このような環境システムもロボットであり、環境ロボット、または、姿形の必ずしも見えない、あるいは姿形にとらわれるべきでないという意味で、インビジブルロボットと呼ばれている [1]。

例えば、屋外に設置した多数のカメラを用いた自動監視システムもロボットの一つであり、インビジブルロボットの屋外への応用例として挙げられる。既にイギリスでは 100 万台以上の監視カメラが設置、国内においても新宿歌舞伎町において 50 台の監視カメラが導入され、犯罪防止や事件の検挙に活用されており、これらをインビジブルロボットへのセンサ入力として用いることで、安全安心な空間を提供することが期待できる。このような屋外における物体検出や追跡等の知的な映像監視に関する研究は 1997 年の DARPA(Defence Advanced Research Projects Agency) の自動ビデオ監視システムの研究プロジェクトであった VSAM(Video Surveillance and Monitoring) プロジェクト [2] を一つのきっかけに、以来いっそう進展し、現在はこれらの技術を基にした製品の実用化が盛んに取り組

まれている。

さらに、オフィスや家、公共施設などの空間において、その空間内の人の意図を理解し行動を支援する技術への展開が期待されている。ジョージア工科大学の Aware Home[3] プロジェクトでは、生活空間である家にカメラをはじめとする数多くのセンサ群を埋め込み、24 時間を通して生活空間における人の動きをセンシングする研究が取り組まれている。また、リビングルームを対象とした Microsoft Research の Easy Living[4] のように、センシングにより得られた情報を基に、ユーザである人に対して快適な空間をアシストする研究が盛んである。このような活力生活技術は QoLT(Quality of Life Technology) と呼ばれる技術の一環で、QoLTのためのセンシング技術は、刻々と変化する人の状態を実時間で認識する必要がある。特に、人画像解析 (People Image Analysis) として、動画像からの人の検出、追跡、顔の検出、顔の部位の追跡、モーション理解が不可欠な技術要素となる。中でも人検出は、人の形状変化が大きいため難しい問題とされてきたが、近年のコンピュータの高速化に伴い、画像全体を検出ウィンドウによってラスタスキャンし、low-level な特徴量と統計的学習手法の組み合わせによる検出法が提案されている。

本稿では、インビジブルロボットにおけるビデオ解析として、VSAM プロジェクトで開発された動画像理解アルゴリズムとその実用化について紹介する。また、近年の人画像解析技術のアプローチとして、検出ウィンドウのラスタスキャンベースによる人検出法とモーション解析について

も述べる．

2 動画像理解技術を用いたビデオ監視システム: VSAM

重要施設の入退出管理を目的とした従来のビデオ監視システムは，監視カメラ映像を記録するものや，監視員が複数のカメラ映像を同時にモニタリングするものが多い．監視する範囲が広く 24 時間監視となると監視員への負担が大きくなり，問題とされている．これに対し，米国では 1997 年より 2000 年の 3 年間，DARPA の下，画像理解技術を用いたビデオ監視システムの研究プロジェクト VSAM が行われた．CMU(Carnegie Mellon University) では，キャンパスに 12 台のカメラを配置しテストシステムを構築した [2]．本システムは，動画像理解技術により検出した侵入物体を複数のカメラが協調してトラッキングし，その状況をリアルタイムで監視員に提示する．これにより，監視員の負担軽減と作業効率化に大きく貢献できる新しいビデオ監視システムとして注目された．

2.1 VSAM における動画像理解技術

図 1 に VSAM における動画像理解の流れを示す．まず，入力画像と背景画像との差分を計算し，動体検出を行う．次にセグメンテーションした結果から特徴量を抽出し，この特徴量を基に識別を行う．ここでは，VSAM システムに用いられている動画像理解技術として，動体検出と物体クラス識別について述べる．

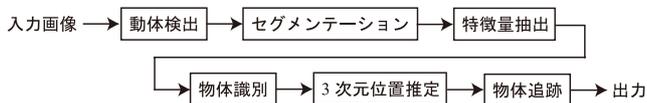


図 1 VSAM システムの処理の流れ

2.1.1 レイヤー型動体検出

侵入物体の検出には，検出すべき物体が存在しない背景画像を予め用意しておき，入力画像と背景画像の差分を計算する背景差分処理が多く用いられている．人と自動車のアクティビティ認識するには，画像上の複数物体の重なりを検出する必要があるが，背景差分処理と領域クラスタリングを組み合わせた手法では隣接する複数物体を 1 つの領域として検出してしまうという問題がある．これに対して，我々はピクセル分析とリージョン分析の 2 つの処理からなるレイヤー型検出法を提案した [5]．ピクセル分析では，各画素の輝度値の時間変化を観測し，その変化軌跡により画素の状態を静もしくは動と判定する．リージョン分析では，動とラベル化された画素領域を移動体と判定する．静とラベル化された領域は静止物体と判定し，背景上のレイヤー

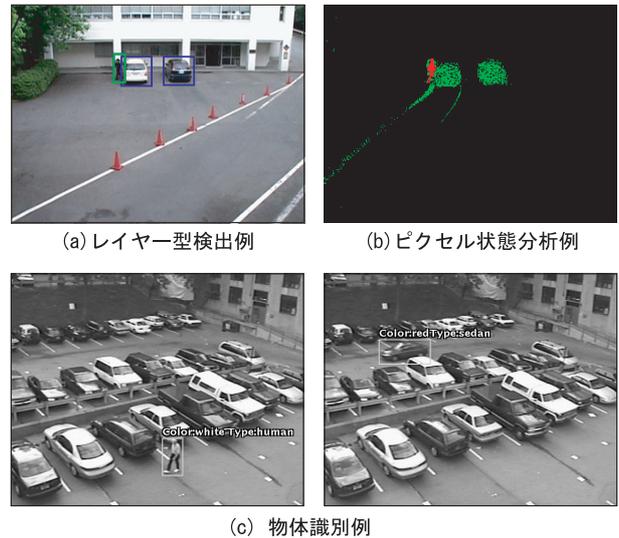


図 2 レイヤー型動体検出と物体識別例

として記憶する．一度停止した物体は，再び動き出すまでレイヤーとして登録されているため，レイヤー上を通過する移動体を区別して検出することが可能となる．図 2 は，停止した車から人が降りる動画に対してのレイヤー型検出例 (図 2(a))，ピクセル状態分析例 (図 2(b)) である．停止した車は，動き出すまでレイヤーとして登録されているため，車と人が画像上で重なった場合にも区別して検出することが可能である．

2.1.2 物体クラスの識別

レイヤー型検出により検出した物体は，対象物体の姿勢変化，天候による照明変化，カメラの位置等の要因により，その「見え」は逐次変化するため，人と自動車の識別は容易でない．VSAM システムでは，12 台のカメラを見えの大きく異なるカメラ毎にグループ化し，人工ニューラルネットワーク (ANN) を用いて，それぞれの識別器を作成することで対処している．検出した画像から計算した形状の複雑度 (周囲長²/面積)，面積，縦横比，カメラのズーム倍率を求め，ANN への入力パラメータとした．出力は，人 (一人)，人のグループ (二人以上)，自動車，その他の 4 クラスとした．小規模な ANN (入力層 4，中間層 16，出力層 4) で構成しているため，リアルタイムでの識別処理を実現することができた (図 2(c))．

2.2 複数カメラによる協調動作

広域の屋外監視を対象とした場合，カメラの設置場所により木や建物等によるオクルージョン領域が発生し，1 台のカメラセンサのみで長時間の物体追跡は困難である．この問題を解決するには，ネットワーク化された複数カメラセンサの協調動作が考えられる．VSAM システムでは，以下に述べる二つの手法による追跡を実現している．

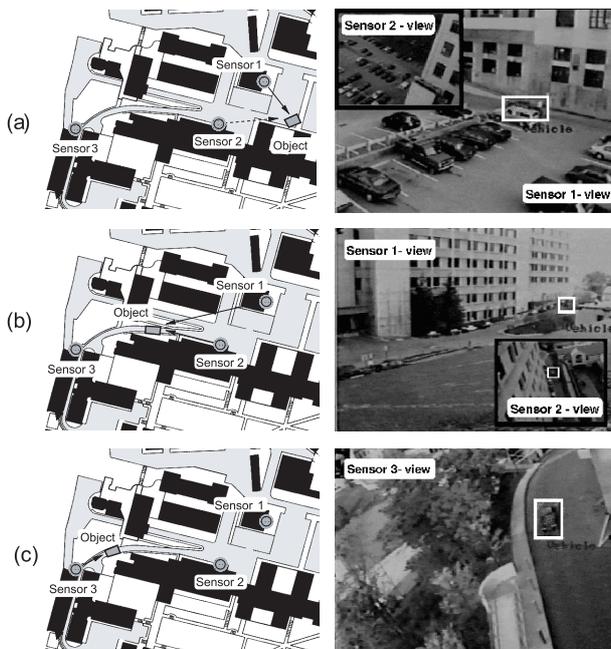


図3 ハンドオフ動作例

2.2.1 複数カメラセンサによるハンドオフ

侵入物体を広域で追跡するには、一度、対象物体を捉えたカメラセンサが pan/tilt により追跡を行い、他のカメラ視野に移動した際にそのカメラセンサに追跡タスクの受渡しを行う。その際、異なるカメラにおいて物体の見えは異なるため、同一物体であるかの判定が重要となる。各カメラセンサは動画像理解技術により物体の識別クラス、色、3次元位置情報を得ているので、それぞれの整合度を計算し同一であるかを判定する。同一物体と判定されたなら、受け渡されたカメラセンサは物体の移動速度、大きさを考慮して pan/tilt/zoom を制御し追跡する。図3は、1台の自動車が約400[m]の距離を3分程度で移動したときの3台のカメラによるハンドオフ動作例である。

2.2.2 2台のカメラセンサによるスレーブ動作

より詳細の侵入物体の画像（ズーム画像）を得るために、2台のカメラセンサによるスレーブ動作を実現した。屋上に設置してある広角の視野を持つ固定カメラをマスタとして物体を検出し、3次元位置を推定する。このとき、物体に最も近いカメラをスレーブとして、マスタカメラ上で検出した物体の3次元位置を画像中心となるように逐次スレーブカメラの pan/tilt/zoom を制御する。図4に2台のカメラセンサによるスレーブ動作例を示す。

2.3 VSAM 動画像理解技術の応用

1997年に始まったVSAMプロジェクトでは、当時Pentium II 300Mhz相当のPCを用いて、約5~10fpsのリアルタイム動画像処理を実現していた。10年後の現在では、

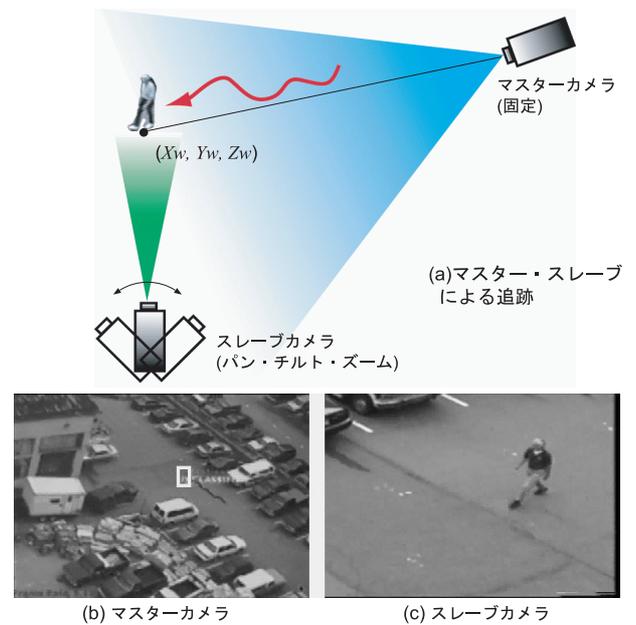


図4 2台のカメラセンサによるスレーブ動作

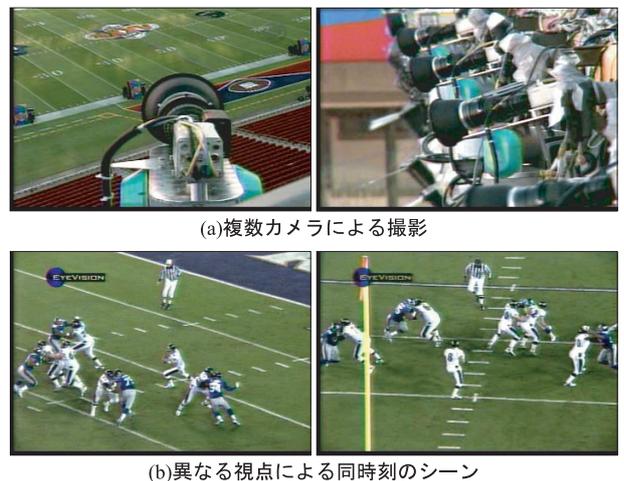


図5 EyeVision システム

計算処理能力が大幅に向上していることからFPGA(Field Programmable Gate Array)によるハードウェア化やDSPを用いた高速処理の実現が可能である。また、VSAMプロジェクトは自動ビデオ監視技術の実用化だけでなく、Eye-Visionシステム等のエンターテインメントの分野に応用されている。

2.3.1 VSAM 技術の実用化

VSAM プロジェクトマネージャが起業した米 ObjectVideo 社 [6] では、国境、空港、石油精製施設、化学工場、核燃料施設など、重要な施設や高リスクの環境下における警備、防犯用途の製品を提供している。VSAM 技術を基にした物体の検出、識別、追跡等を実用化し、さら

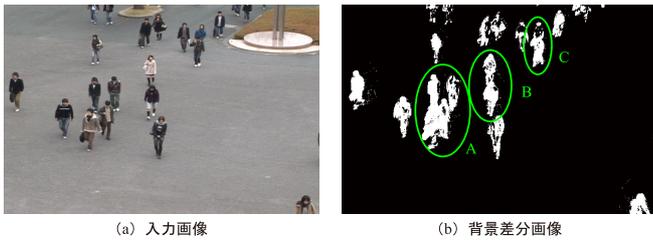


図 6 背景差分による動体検出の失敗例

に VSAM プロジェクトから発展した技術として、ユーザがセキュリティ規則(アクティビティ)を作成し、監視の自由度を高めている。ユーザは進入する物体に対して、進入領域や物体の種類などをセキュリティ規則として作成することにより、その事前に定義された規則に反する状態のみをユーザにリアルタイムに通知することが可能である。近年では、Texas Instruments 社によって、ObjectVideo 社の技術を組み合わせたインテリジェンス機能付きビデオ監視カメラ等も開発されている。

2.3.2 VSAM 技術の他分野への応用

2.2.2 で述べたカメラ間のマスタースレーブ技術は、エンターテインメント分野へ応用されている。アメリカのテレビ局 CBS が、映画 Matrix で使用された映像効果をリアルタイムで提供する EyeVision システム(図 5 参照)を CMU と開発した。EyeVision システムは 32 台のロボットカメラ [8] を活用し、複数の角度から撮影された映像を連続的に合成した 3 次元的なテレビゲームのような感覚でユーザーに映像が提供できる。カメラマンが撮影した映像中心の対象物の 3 次元位置を求め、32 台のスレーブカメラを自動制御している。2001 年の第 35 回スーパーボール(Super Bowl XXXV)にて実用化され、テレビ放送で用いられた。

3 人を観る技術:PIA

VSAM における動体検出法は、入力画像と背景画像の差分を計算する背景差分ベースの手法であった。このような動体検出をベースとした動画理解のアプローチは、移動体同士が画像上で重なった場合にセグメンテーションに失敗するため、その後の処理である物体識別が不可能となる問題があった。図 6 に背景差分による動体検出の失敗例を示す。図 6 の A, B, C のように人同士が画像上で重なった場合、1 つの動体として検出される。

これに対して、Viola と Jones は Haar-like(図 7(a)) と呼ばれる局所特徴と統計的学習の組み合わせによる高速かつ高精度な顔検出法 [9] を提案した。この手法は、入力画像に対して検出ウィンドウを図 7(b) のようにラスタスキャンし、図 7(c) に示すように、AdaBoost を用いて検出ウ

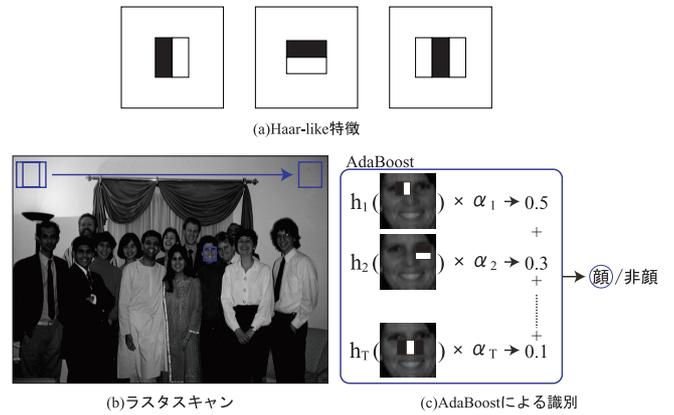


図 7 ラスタスキャンと AdaBoost による顔検出

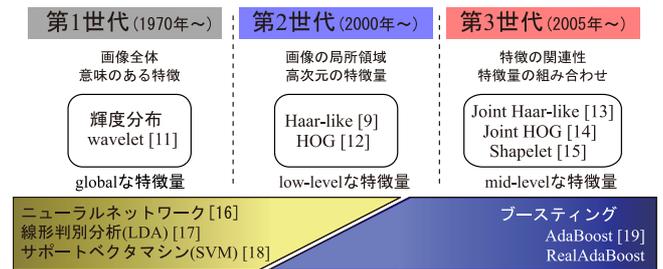


図 8 識別器と特徴量の変遷

ンドウ内の画像が顔/非顔として判別するため、前処理として動体検出を必要としない。顔検出の分野では、ラスタスキャン方式による手法が多く提案され実用化された。近年では検出対象が顔から人へと変わり、形状変化が大きいことにより検出がより難しいとされている人検出の研究が盛んになっている。本章では、近年のアプローチである局所特徴量と統計的学習を組み合わせた人検出法と、高精度化のための有効な特徴量の捉え方について述べる。

3.1 識別器と特徴量の変遷

図 8 に物体認識に用いられる識別器と特徴量の変遷を示す [10]。物体認識に用いられる第 1 世代の特徴量は、輝度分布や wavelet など画像全体から得られるものや、研究者・開発者が特徴量の種類を決定し、2.1.2 の物体認識で述べたような形状の複雑度のような意味のある特徴量を抽出していた。これは、リアルタイム性を求めるために、ニューラルネットワーク全体のサイズ(結合数重み)をコンパクトにするために、入力ユニットの数を小さくする必要があり、そのため、低次元数の特徴量が用いられていた。それに対し、2000 年以降に提案された第 2 世代では、Haar-like 特徴や HOG 特徴などの low-level な局所特徴量から、統計的学習法であるブースティングを用いて特徴選択を行うため、高次元(数千~数十万パターン)の特徴量を扱うことが可能である。さらに、第 3 世代(2005 年以降)では、第 2 世代の low-level 特徴量を AdaBoost の特徴選択を利用して組み合わせ、識別に有効な mid-level 特徴量を自動生成

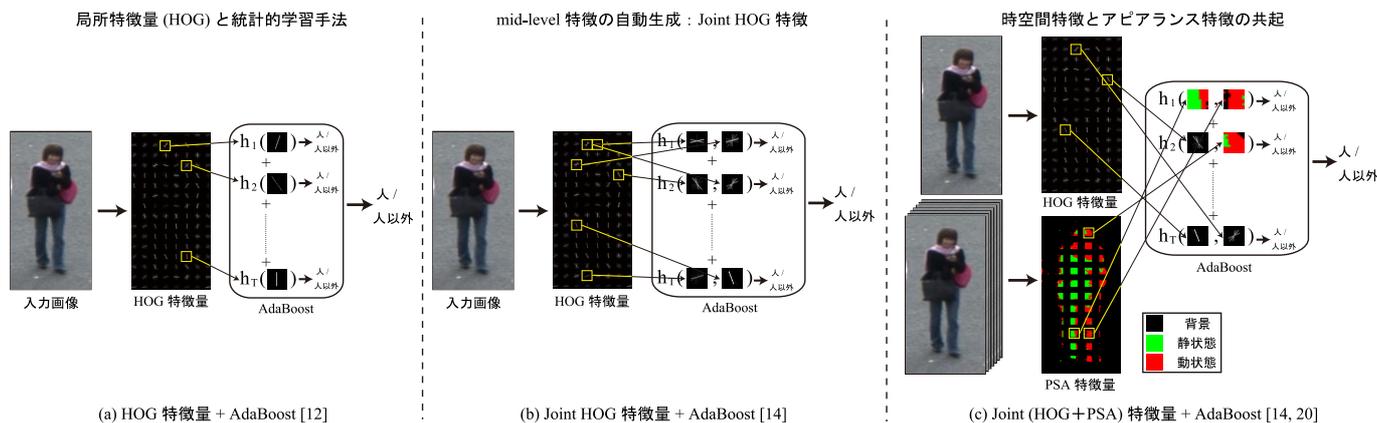


図 9 人検出における特徴量の捉え方

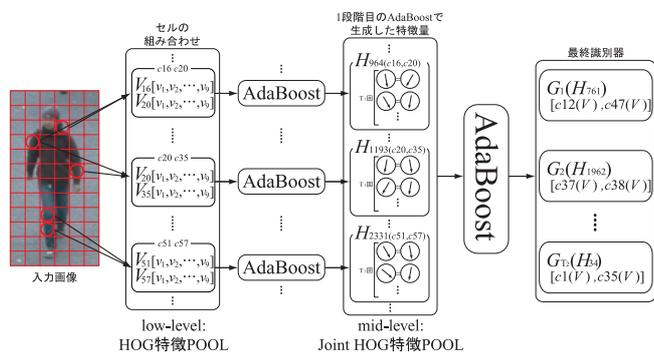


図 10 2段階 AdaBoost による Joint HOG

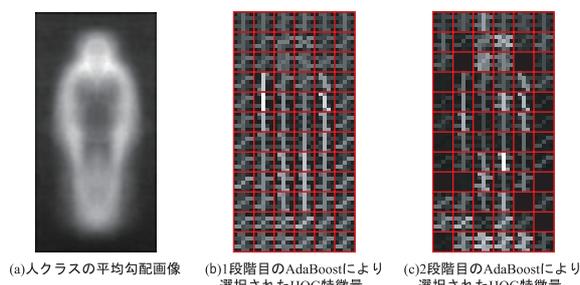


図 11 選択された HOG 特徴量の可視化

する Joint-HOG[14] や Shapelet[15] が登場した。

3.2 局所特徴量 (HOG) と統計的学習手法による人検出

人は動きとともに形状が変化する非剛体な物体であるため、顔検出と比較して難しい問題である。また、画像中での人同士の重なりによるオクルージョンの発生や衣服の違い、照明や影の影響も検出を困難とする要因である。このような問題に対して、Dalal 等により局所領域における勾配方向をヒストグラム化した Histograms of Oriented Gradients (HOG) 特徴量と統計的学習手法を組み合わせた人検出法 [12] が提案された。HOG 特徴量は、照明の変動による影響が少なく、局所的な幾何学的変化に頑健であるため高精度な人検出を可能とした。

図 9(a) に AdaBoost による HOG 特徴量の捉え方を示す。AdaBoost の弱識別器により 1 個の HOG 特徴量が選択され、最終的に多数ある弱識別器の重み付き多数決により人と人以外に判別する。

3.3 mid-level 特徴の自動生成: Joint HOG 特徴

人には、形状の左右対称性や連続したエッジがあり、これらの特徴を捉えることで検出精度を向上させることができると思われる。我々は、人独特の形状を捉えるために、複数の HOG 特徴量を組み合わせた Joint HOG 特徴量と、2 段階に構築した AdaBoost による学習法 [14] を提案した。

複数の low-level な特徴量である HOG 特徴量を AdaBoost により組み合わせることで mid-level な特徴量である Joint HOG 特徴の POOL を自動的に生成し、この Joint HOG 特徴を入力とした 2 段階目の AdaBoost により、識別に有効な Joint HOG 特徴を自動的に選択して最終識別器を構築する (図 10)。

図 9(b) に、AdaBoost による Joint HOG 特徴の捉え方を示す。HOG 特徴量では、1 個の弱識別器が 1 個の HOG 特徴量を用いて識別したのに対し、Joint HOG 特徴では、1 個の弱識別器が位置の異なる 2 つの領域内に含まれる複数の HOG 特徴量を用いて識別を行う。これにより、従来の単一の HOG 特徴量のみでは捉えることができない物体形状の対称性や連続的なエッジを自動的に捉えることが可能となり、高精度な人検出法を実現した。図 11(a) に人の平均勾配画像、図 11(b), (c) に AdaBoost により選択された HOG 特徴量を可視化した結果を示す。HOG 特徴量の勾配方向を 9 方向で表現しており、輝度が高いほど AdaBoost における弱識別器の重みが高いことを表す。1 段階目で選択された HOG 特徴量 (図 11(b)) は全ての局所領域において選択されているが、2 段階目で選択された HOG 特徴量 (図 11(c)) では、人の輪郭に沿った HOG 特徴量が自動的に選択され、高い重みを持つことがわかる。

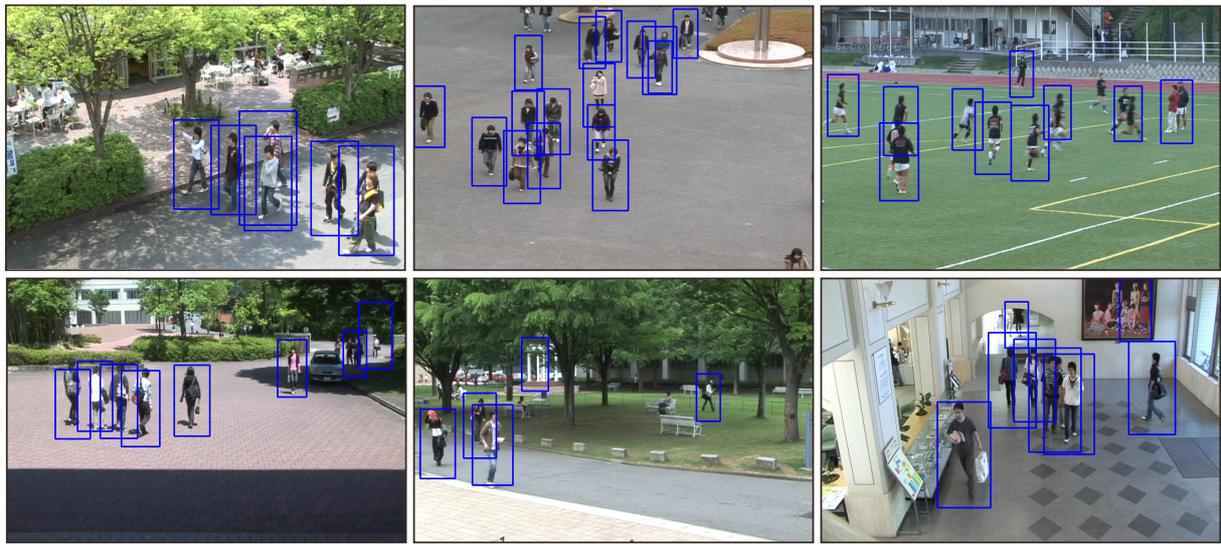


図 14 人検出例

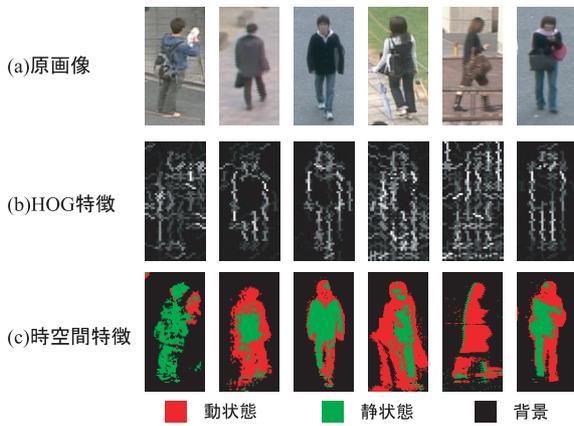


図 12 HOG 特徴とピクセル状態分析の例

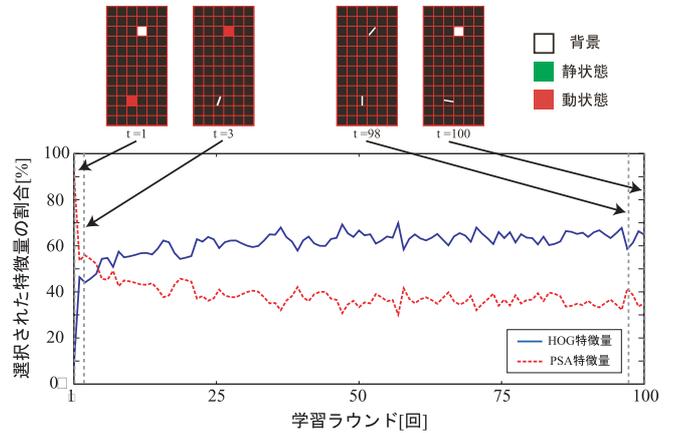


図 13 選択された特徴量の割合

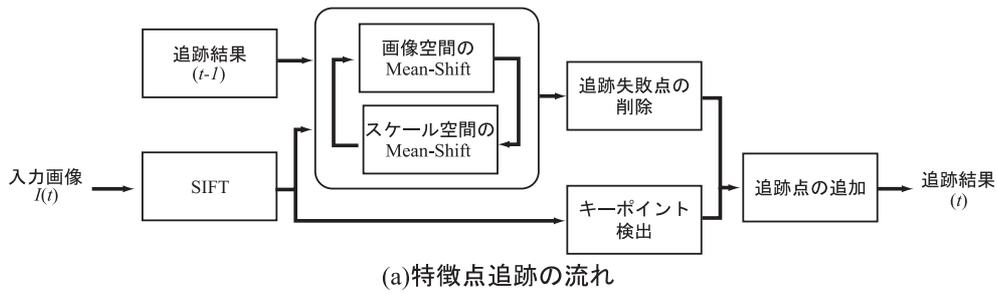
3.4 時空間特徴とアピランス特徴の共起

Joint HOG 特徴のフレームワークでは、人のアピランスを表す HOG 特徴量 (図 12(b)) に、他の特徴を追加することが可能である。我々は、従来動体検出に用いられてきた、時空間特徴に基づく特徴量として図 12(c) に示すピクセル状態分析 (PSA) の結果を加えることにより、より高精度な人検出を達成した [20]。ピクセル状態分析とは、2.1.1 で示したレイヤー型検出に用いられた手法であり、ピクセル状態の時間変化をモデル化し、各ピクセルを背景と動状態、静状態に判別する手法である。この時空間特徴とアピランス特徴を図 9(c) に示すように同時に捉える手法を示す。時空間特徴を加えることにより、AdaBoost の弱識別器は、人のアピランスと動きの情報を捉えることが可能となる。これにより、アピランスの情報のみでは誤検出する人に似た物体に対して誤検出を抑制することができる。

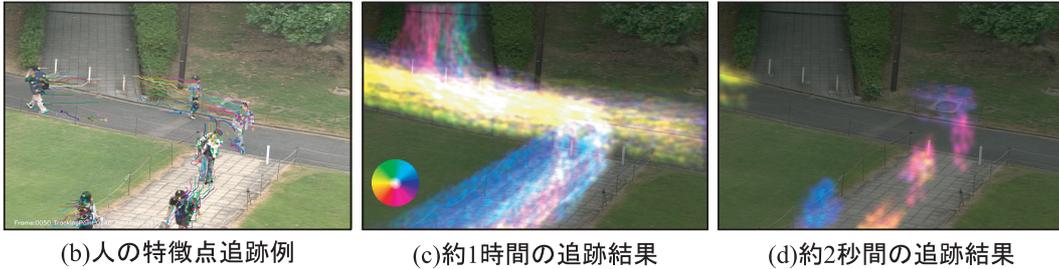
AdaBoost により選択された特徴量に着目するために、図 13 に各学習ラウンドにおける HOG 特徴量と PSA 特徴

量の選択された割合と、その際に選択された特徴量の可視化の例を示す。初期ラウンドにおける弱識別器では PSA 特徴が多く選択され、学習ラウンド数が進むにつれて HOG 特徴量が選択される割合が多い。これは、まず物体の動きを表すことが可能な PSA 特徴により、大まかに人と人以外を判別し、この後アピランスの情報をもつ HOG 特徴量を用いて、より細かな識別境界を形成していると考えられる。

図 14 に本手法による人検出の例を示す。人の大きさの変化や人同士の画像上での重なりによる部分的な隠れに対しても高精度な人検出が可能である。図 15 は、評価実験の結果を示す DET (Detection Error Tradeoff) カーブであり、原点に近いほど識別器の性能が高いことを示す。従来人検出に用いられている HOG 特徴量のみよりも、Joint HOG 特徴の方が識別性能が高いことがわかる。また、人のアピランス特徴と時空間特徴量を同時に捉えることで、さらに高精度な人検出ができている。従来の HOG 特徴量に比



(a)特徴点追跡の流れ



(b)人の特徴点追跡例

(c)約1時間の追跡結果

(d)約2秒間の追跡結果

図 16 特徴点追跡と動きの可視化

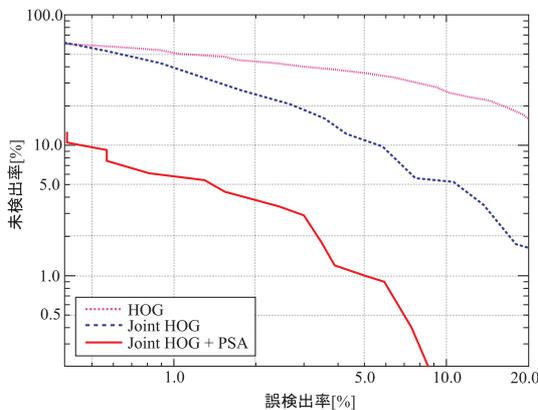


図 15 DET カーブ

べて Joint HOG+PSA 特徴は、誤検出率 5.0%において検出率を 99%まで向上させることができた。

3.5 SIFT 特徴量に基づく Mean-Shift 探索による特徴点追跡

刻々と変化する人の状態を認識するには、人がどのような動きをしているかを知ることが重要である。既に、モーション情報を獲得する手法として、オプティカルフローや KLT[23] 等の特徴点追跡法が提案されている。しかし、長時間の映像に急激な照明の変化や対象物体の運動に回転やスケール変化を含む場合には、特徴点の追跡は困難となる。そこで、我々は回転・スケール変化・照明変化による画像の変化に不変な SIFT(Scale Invariant Feature Transform) 特徴量 [21] を用いることで、この問題を解決する手法を提案した [22]。図 16(a) に特徴点追跡の流れを示す。まず画像より初期追跡点を検出し、各点の SIFT 特徴量を抽出する。次フレームより特徴点の追跡を行うが、画像空間の移動量と SIFT 特徴量の特徴を抽出する範囲を表すスケ

ルの各空間で Mean-Shift 探索を行う。Mean-Shift 探索によって求められた画像空間とスケール空間の移動量を、現フレームでの最も最適な位置とスケールに移動するよう計算を繰り返す。移動前の SIFT 特徴量と Mean-Shift 探索によって求められた移動後の SIFT 特徴量から、しきい値により追跡失敗点を削除する。また、新たに画像内に進入した移動体に対応するために、特徴点の追加を毎フレーム行う。人に対しての特徴点追跡を行った例を図 16(b) に示す。図 16(b) より、人などの非剛体に対しても頑健に特徴点を追跡できていることがわかる。

この特徴点追跡を応用して、特徴点追跡結果のビジュアライゼーションが可能である。点の密度とスケールにより色の強度を変えることで、セグメンテーションを必要とせず、動いている物体の流れが表現可能となる。図 16(c) は長時間の追跡結果から動線をビジュアライゼーションした結果、図 16(d) は短時間の追跡結果から動線をビジュアライゼーションした結果を示す。この手法により、物体のセグメンテーションを行うことなく、人の流れを表すことができる。また、このような信頼性の高いモーション情報は、定常状態の動きのモデル化による異常行動検出 [24] の精度向上に貢献することができる。

4 おわりに

本稿では、インビジブルロボットの視覚機能の実現として、VSAM プロジェクト以外で開発した動画理解技術と、人を観る技術である人検出法とモーション解析について紹介した。環境が刻々と変化する人の状態を認識するには、図 17 に示すように、人がどこにいるのか(人検出)とどのような動きをしているのか(モーション解析)を知る必要がある。このような認識技術は、ユーザである人の行動

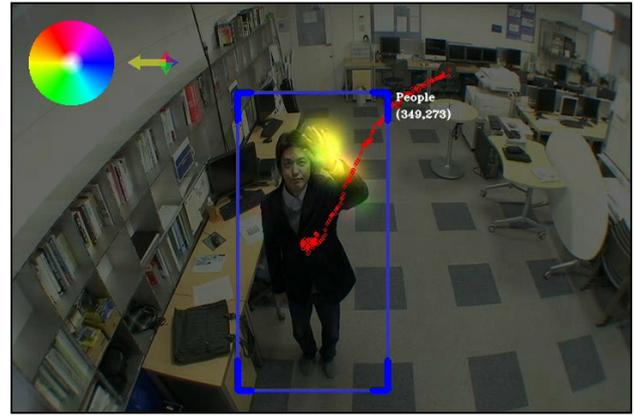
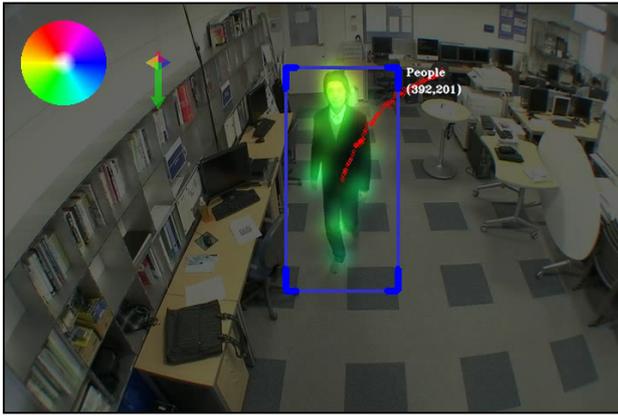


図 17 インビジブルロボットによる人を観る技術 (人検出とモーション解析例)

意図を理解するために、今後一層インビジブルロボットのセンシング技術に欠かせない技術となる。

今後は、インビジブルロボットがよりきめ細かい、かつ安全なサービスを行うために、一般物体認識やシーン理解を通して、人とその他の物体や周辺との関係を利用して人が何をしようとしているかを理解することが重要な課題である。また、従来は計測等の用途に用いられてきた IR カメラ等の特殊なカメラと認識技術を組み合わせることで、可視光の通常のカメラでは捉えられない情報を利用した認識性能の高精度化とその実用化が期待されている。

参考文献

- [1] 井上博允, 金出武雄, 安西祐一郎, 瀬名秀明: “岩波講座 ロボット学 1 ロボット学創成”, p49 - 65, (2004).
- [2] R. Collins, A. Lipton, H. Fujiyoshi and T. Kanade: “Algorithms for cooperative multi-sensor surveillance”, Proceedings of the IEEE, Vol. 89, No. 10, pp. 1456 - 1477, (2001).
- [3] Aware House, <http://www.gatech.edu/innovations/futurehome/>
- [4] Easy Living, <http://research.microsoft.com/easyliving/>
- [5] H. Fujiyoshi and T. Kanade, “Layered Detection for Multiple Overlapping Objects”, IEICE Transactions on Information and systems, vol. E87-D, pp. 2821-2827, (2004).
- [6] ObjectVideo, <http://www.objectvideo.com/>
- [7] Eye Vision, <http://www.eyevision.com/>
- [8] 川内直人, 金澤宏幸, 見持圭一, 宮内礼三, 大西献, 中山淳二, 藤田淳, 大道武生: “新映像システム EYE VISION 向けカメラ制御パンチルトの開発”, ロボティクス・メカトロニクス講演会, p. 59, (2002).
- [9] P. Viola and M. Jones, “Robust Real-Time Face Detection”, Int. Journal of Computer Vision, 57(2), pp. 137-154, (2004).
- [10] 山下隆義, “統計的学習法を用いた物体認識における特徴量の進化”, 第 14 回画像センシングシンポジウム SSII08, OR4-01, (2008).
- [11] H. Schneiderman, T. Kanade, “A statistical method for 3d object detection applied to faces and cars”, IEEE Computer Vision and Pattern Recognition, pp. 746-751, (2000).
- [12] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection”, IEEE Computer Vision and Pattern Recognition, pp. 886-893, (2005).
- [13] T. Mita, T. Kaneko, B. Stenger and O. Hori, “Discriminative Feature Co-occurrence Selection for Object Detection”, IEEE Pattern Analysis and Machine Intelligence, vol. 30, no. 7, pp. 1257-1269, 2008.
- [14] 三井相和, 山内悠嗣, 藤吉弘巨, “Joint HOG 特徴を用いた 2 段階 AdaBoost による人検出”, 第 14 回画像センシングシンポジウム SSII08, IN1-06, (2008).
- [15] P. Szabzmeydani and G. Mori, “Detecting Pedestrians by Learning Shapelet Features”, IEEE Computer Vision and Pattern Recognition, pp1-8, 2007.
- [16] M. Minsky and S. Papert, “Perceptrons”, MIT Press, 1969.
- [17] R. A. Fisher, “The Use of Multiple Measurements in Taxonomic Problems”, Annals of Eugenics, vol.7, Part II, pp.179-188 (1936).
- [18] Bernhard E. Boser, Isabelle M. Guyon and Vladimir N. Vapnik “A Training Algorithm for Optimal Margin Classifiers”, 5th COLT, pp. 144-152 (1992).
- [19] R. E. Schapire and Y. Singer, “Improved Boosting Algorithms Using Confidence-rated Predictions”, Machine Learning, No. 37, pp. 297-336, 1999.
- [20] 山内悠嗣, 藤吉弘巨, Bon-Woo Hwang, 金出武雄, “アビアランスと時空間特徴の共起に基づく人検出”, 画像の認識・理解シンポジウム (MIRU2007), (2007).
- [21] D. G. Lowe, “Distinctive image features from scale-invariant keypoints”, International Journal of Computer Vision, 60(2), pp. 91-110, (2004).
- [22] 都築勇司, 藤吉弘巨, 金出武雄, “SIFT 特徴量に基づく Mean-Shift 探索による特徴点追跡”, 情報処理学会論文誌 Vol. 49, No. SIG 6, pp. 35-45, (2008).
- [23] C. Tomasi and T. Kanade, “Shape and Motion from Image Streams under Orthography: A Factorization Method”, Int. Journal of Computer Vision, Vol. 9, No. 2, pp. 137-154, 1992.
- [24] 都築勇司, 西村孝, 藤吉弘巨, 森次郎, “特徴点追跡結果を用いた動きのモデル化と例外行動検出”, 第 14 回画像センシングシンポジウム SSII08, IN1-07, 2008.