

7

人を見る画像認識技術

鷲見 和彦

三菱電機（株）

人を見る画像認識技術について

画像認識技術は、1950年代にリモートセンシングやロボット視覚の研究からスタートし、初期は原理的な方式の研究や蓄積された画像の分析が主体であったが、カメラの半導体化とマイコンやASIC技術の進歩に伴い、1980年代後半から1990年代にかけて半導体製造や電子機器組立工場で使われる微細な位置決め作業や外観検査の自動化などの分野で実用化が進んだ。1990年代に入ると、ヒューマンマシンインタフェースや映像監視の分野での技術開発が急激に進んだ。特に、1996年から2000年にかけて米国で実施されたVSAM（Video Surveillance and Monitoring）プロジェクトによりコンピュータビジョンの個別の研究が「監視」という目的・機能を核として統合されたことで、実用化・商用化の新たな局面を迎えた。さらに、2000年代にはテロ対策を目的とした公共映像監視の導入、新会社法の施行による企業内セキュリティの強化、環境保護・省エネルギーの重視などのニーズが牽引力となって、さまざまな応用製品が普及するに至っている。本稿では、人を見る画像認識技術を、要素技術と応用技術の両面から整理・解説する。要素技術としては、背景・人物モデル（背景差分）、人物同定、移動物体追跡を取り上げ、応用技術としては、警備や公共分野および家電製品での実用化を例に、画像認識技術の役割を説明する。

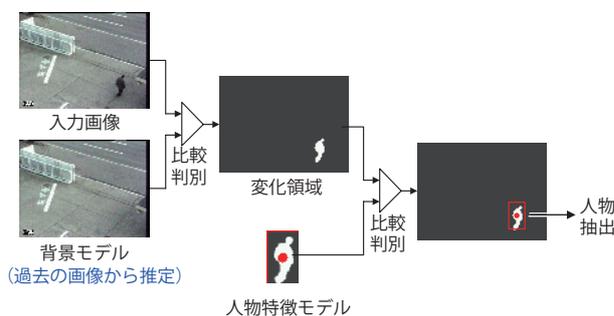


図-1 背景差分による人物抽出の基本的考え方

人物の検出と追跡

人を見る画像認識においては、まず、人物の検出と追跡が必要となる。対象を人に限定しない場合には、物体検出・物体追跡とも呼ばれる基礎技術であり、画像認識の基礎技術の1つである。

◆ 背景差分と物体検出の基本的な考え方

画像から物体を抽出するには、画像中のすべての画素について、その画素が物体に属するのか、それとも背景に属するのかを判別するという問題に帰着する。

たとえば、図-1に示すように、あらかじめ背景となる入力画像をある程度の期間観察して背景モデルを学習しておき、現在の入力画像を背景モデルと比較し、背景モデルに合致しない画素または画素のグループを変化領域として抽出することが基本になる。

しかし、現実には背景にもさまざまな変化が生じており、単純に背景画像と差を取るだけではさまざま

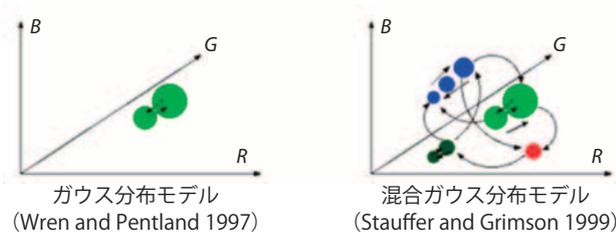


図-2 ガウス分布を仮定した背景画素のモデル



観測画像の例 木の揺れ (2倍強調)

図-3 揺れる背景と不感帯 (右側画像の白い部分)

まな誤検出を発生させてしまう。たとえば日照変動、カメラケースやレンズの水滴付着、監視領域近傍を走行する車などから照射され動く照明、犬・猫・虫・鳥などの小動物、樹木やその影の揺れ、水面に発生する揺らぐ反射、成長する草木、朝日・西陽の光の侵入や光輪などが背景を複雑に変化させ、物体抽出を混乱させることが知られている。

◆ 人物検出に有効な背景差分法

そこで、一般によく使われる手法は、背景の統計的なモデルを学習することである。たとえば、1997年に Wren らによって発明された P-finder という人物検出を説明する。図-2左は、画像中のある画素を長時間観測した値の分布を RGB 色空間に累積したものである。ガウス分布モデルは、画素の値は、図中の円を中心に円の半径を標準偏差とする正規分布で変動し、中心(平均)と分散(標準偏差の二乗)が時間とともにゆっくり動いてゆくと仮定した。1999年に Stauffer らによって発明された混合ガウスモデルでは、図-2右のように、ある画素は繰り返し現れる複数の物体(図では色分けされた4種)のいずれかが観察され、各物体がガウス分布で表現できると拡張され、繰り返し現れる背景にも対応できるようになった。

しかし、背景の画素モデルを複雑にすればするほど、背景だと判別される画素の明るさ範囲は広がることになり、人物の抽出感度が低下してしまう。検出感度と誤検出抑制の両立は簡単ではない。たとえば図-3左側の画像のように、風で背景が揺れている場合、右側の画像の白い部分においては画像の変動量が大きく、この部分では前景に人が現れても検

出できないか、人のシルエットが抜けてしまう。

現在、最も信頼して使える背景差分アルゴリズムは、これらの画素単位の背景モデルを、空間的關係および時間的關係も記述できるように拡張したものである。

まず、空間的な拡張を行った正規化距離に基づくテクスチャ背景差分法¹⁾を説明する。周囲の明るさなど環境が変化すると背景の見え方も変化するが、画像中のある画素と周りの画素との比率(テクスチャ)は変化しないという性質がある。これを利用し、図-4左に示すように、画像をある一定の大きさのブロックに分割し、ブロック単位で入力画像を背景と比較する。比較の際には明るさ(画素の値)の差ではなく、ブロック内の画素の集まりをベクトルとして解釈し、入力画像ベクトル Z_{Ax} と背景画像ベクトル Z_{Bx} とをそれぞれ長さ1に正規化した際の距離(正規化距離)をとる。図から分かるように一定の比率の明るさ変化があっても正規化距離は変化しない。ブロック内の一部の画素だけが周りとは違う変化をした場合に感応する背景差分を実現することができる。

次にテクスチャ背景差分をさらに時間的に拡張する。あるブロックの画素ベクトルと近傍の画素ブロックとでは、変化がペアになって生じるという性質を利用する。たとえば揺れる背景があり、ある背景物体がブロック B_x と隣接ブロック B_{x-1} を行き来する様を考えると分かりやすい。そこで図-4右のように、ブロックごとの変化を Z_{Bx} と Z_{Bx-1} をペアにして記憶する。入力画像のブロック Z_{Ax} が前景か背景かを判定するには、まず、 Z_{Ax-1} の観測値から、過去の学習結果をたどり現在値を予測し、

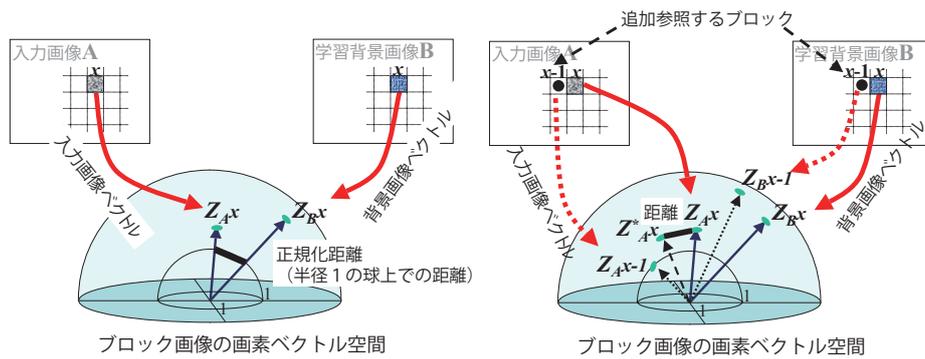


図-4 画像ブロック間の正規化距離を利用した背景モデル（左）と、同時生起性を利用した背景モデル（右）

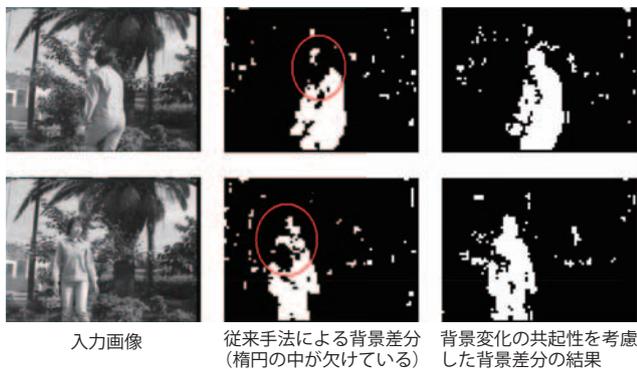


図-5 背景変化の共起性を利用した背景差分の効果

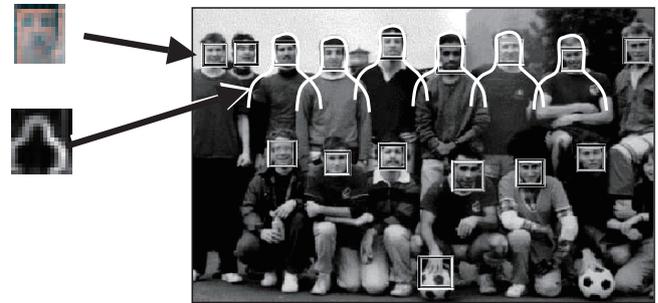


図-6 物体モデル（顔・上半身輪郭）を用いた人物検出

予測値 Z_{Ax}^* と実測値 Z_{Ax} との距離が大きければ前景物体が現れたと考える。

この手法を用いるとたとえば図-5のような揺れる背景の前に現れた人のシルエットを従来よりも実際の形に忠実に抽出することができる²⁾。

一方、背景のモデルではなく、人に特有な形状パターンを用いて人物を抽出することも行われている。たとえば図-6に示すように、人の上半身は頭部・肩・腕と胴など万人に共通の形状を持つので、それを輪郭形状モデルとして用いた人物の検出や、デジカメなどに広く採用されている顔検出器を用いて人物の顔を抽出し、それを基点にして人物全体を抽出するなどの方法が使われる。

◆ 人物の追跡

物体追跡とは、入力画像中のある領域が連続する画像系列の中でどの部分に対応するかという対応関係を解く問題である。画像系列中の物体像の対応関

係が分かることにより、移動物体に対して注目し続けることや、物体の移動に関するパラメータ（速度・加速度）を求めたり、移動に関するパラメータから物体の位置・姿勢・形状などの3次元情報を復元したりすることが可能になる。

発見された人物・移動体などの物体を追跡することで、行動を理解・予測し、複数のフレームにわたって同一の物体を観察していることが確認されれば、追跡中の画像や認識結果の融合によって、追跡対象に対するより詳しい情報を取得することも可能になる。

人物追跡に用いられる手法⁴⁾としては、参照領域と探索領域との相関が最大になる位置を探索する相関法、領域内の特徴量を計算し類似した領域を探索する領域特徴追跡法、参照画像上の特徴点と類似した特徴点を探索する特徴点追跡法などの手法が用いられる。

相関法では、たとえば図-7に示すように着目し

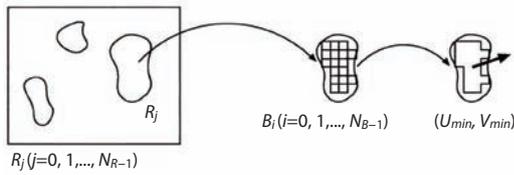


図-7 相関法による追跡の原理

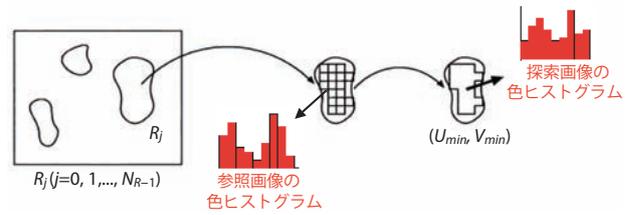


図-8 色ヒストグラムを用いた追跡

ている領域 R_j に対して内部に複数の参照ブロック B_i を設定し、それぞれのブロックが連続する画面でどの位置において最も相関が高くなるかを探索し、それらを統合して移動ベクトル (u, v) を求める。ここでいう相関とは、たとえば、参照ブロック B_i と探索画像中のブロック B'_j との間で画素ごとの差をとった総和 (Sum of Absolute Difference = SAD) が小さいときに相関が高いと決めた指標である。SAD 以外にも、差の二乗和の総和 (Sum of Square Difference) や相互相関係数などが指標として使われる。

現実には人が動くとき人の形が変化したり大きさが変化したりする、また、人が向きを変えることによって見えている部分が変わってしまう。そのため、相関は必ずしも高いとは限らず、また、画面が進むにつれて相関値はどんどん低下していきってしまう。そこで、探索中の画面ごとに参照ブロックを更新してゆかねばならない。一方、参照ブロックの更新は背景の参照ブロックへの混入という問題を生じることがあり、最悪の場合には人物が背景に置き換わって追跡に失敗するという不具合を生じてしまう。人物追跡では、このような追跡失敗をいかに防ぐかが性能の決め手である。

このような見かけの変化に対しても変化しにくい特徴量があれば、それを用いて追跡を行うと有利である。そのような特徴量として色ヒストグラムがしばしば用いられる。色ヒストグラムとは、あらかじめ色空間 (r, g, b) を (m, m, m) 個の空間 (ビン) に区切っておき、ある注目領域の全画素について、その画素がどのビンに属するかを頻度表として表現したものである。2つのヒストグラムの論理積をとることでヒストグラムの類似度を算出することができる。

この色ヒストグラムの類似度を用いた探索手法に図-8に示す Mean-shift 法がある。Mean-shift 法は、ヒストグラムを特徴量とし、特徴量の類似度が高くなる方向に探索を行う。

Mean shift の探索は、類似度が高くなる方向への一方向であるので、たとえば急激な見え方の変化や隠蔽などにより探索方向に対象が存在しない場合、しばしば対象物が背景とすり代わってしまうという失敗も生じる。そこで、探索対象が存在する可能性のある空間を、存在する可能性が高いところを丁寧に、そうでないところもそれなりに粗く探索する方法が考案され、広く使われている。この手法は探索対象の存在確率分布を伝播させてゆくという意味で Condensation あるいは、候補点 (パーティクル = 粒子) を確率に応じてふるい分けてゆくという意味でパーティクルフィルタ (Particle filter) と呼ばれている。パーティクルフィルタでは、図-9のように探索する場所を、見つかる確率に応じて密度を変えて探索するので、たといったん見つけそこなっただとしても回復することができるという特徴が好まれている。

◆ 追跡を利用した人物抽出の精度向上

背景差分の性能向上、検出対象のモデルによる対象検出性能の向上、および、検出した対象の追跡について解説したが、検出と追跡を組み合わせると、さらに精度を向上させることができる。すなわち、追跡した対象の動きを求め、人の動きモデルに適合しない検出を抑制することで、誤検出を抑制することができる。

たとえば、表-1のように検出された人物候補領域を複数フレームにわたって追跡し、その動きベク

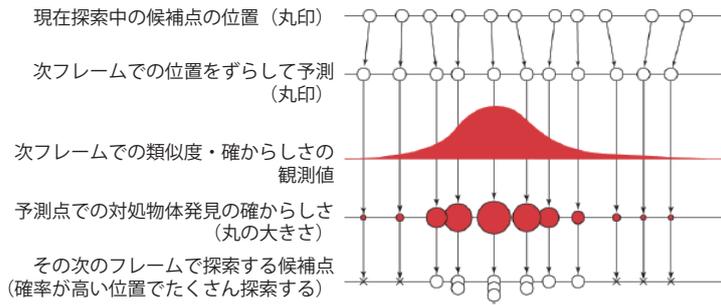


図-9 パーティクルフィルタの探索点更新手法
簡単のため、2次元画像の探索を1次元で表現してある。

トルの分布をもとに識別を行えば、実際に動いている物体 (グループ A) と、草木や影の揺れ・光の反射 (グループ B)、光る水面・西陽や水滴による光の模様 (グループ C)、日照変動と照明点滅による模様の出現や消失 (グループ D) とに識別することができ、結果として図-10のように誤検出や見逃しのない高精度人物抽出ができるようになる。

領域内の動き分布	一様な分布	一様な分布	不均一な分布	すべてゼロ
領域内の平均的動き	動きあり	動きあり	動きゼロ	動きゼロ
平均的な動きの時間的分布	一方向に連続運動	ランダムに運動	動きゼロ	動きゼロ
事象の分類	A 侵入者および車両、小動物、鳥等の一部	B 草木、影の揺れ、光の反射、ヘッドライトの一部	C 水面の揺らぎ、西陽、光輪、ヘッドライト、レンズ面の水滴	D 日照変動および影の出現と消失

表-1 抽出された対象の動きによる対象物の識別方法

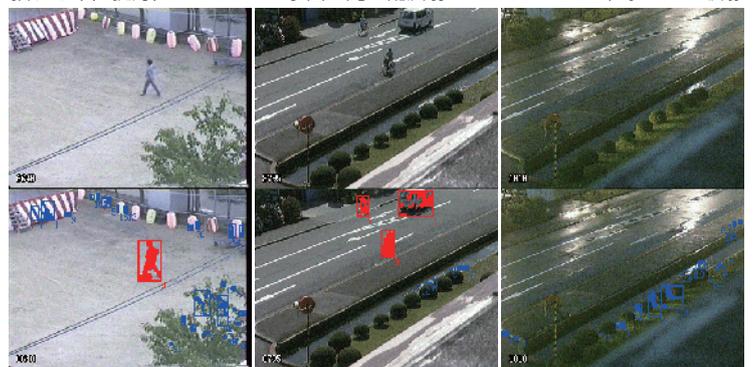
応用事例紹介

前章では人を見るための画像処理の基礎技術を解説したが、本章ではその応用事例を紹介する。

◆ 屋外警備への応用 (人物発見, 追尾, 行動パターンの解析, 特定行動検索システム)

立ち入りが禁止されている区域への立ち入りや、認証されていない人物の立ち入り (侵入) を発見・通報・排除することは、警備システムの重要な機能の1つである。常時監視員をおいてこれを行うことは監視個所の数だけの人員を必要とし、警備対象の価値が高くない限り経済的に実現できない。そこで、画像処理を使って機械的に警備を行うのはきわめて自然である。侵入を検知する方法は、赤外線ビームセンサ、赤外線熱源探知センサ、振動センサ、重量センサ、静電容量センサ、

揺れる木、提灯、カーテン 水面の光の乱反射 ヘッドライト、水たまりの反射



上段：原画像，下段：処理結果 (赤：人物，青：疑似物体)

雨によるレンズ面の汚れ 木の枝葉の揺れ 鳥



上段：原画像，下段：処理結果 (赤：人物，青：疑似物体)

図-10 対象の動きモデルを用いた高精度人物抽出

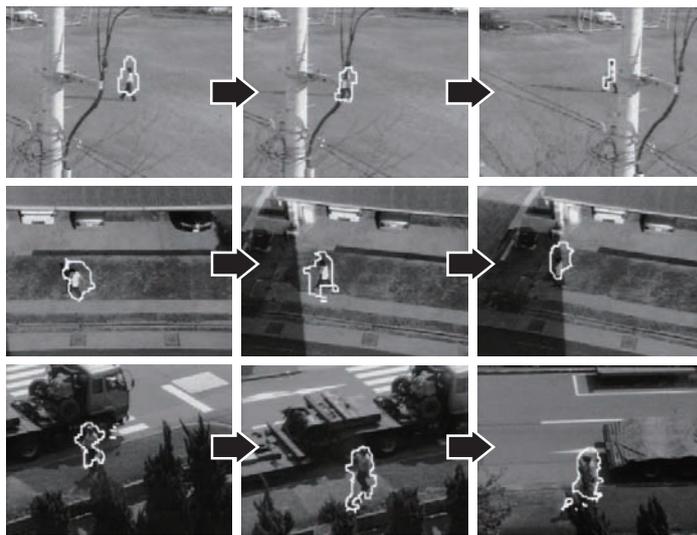


図-11 検出した人物をパン・チルト・ズームして追尾する監視カメラ。追尾困難な状況でも成功している事例
 上段：電柱による全隠蔽の発生、
 中段：影境界の通過によるコントラスト急変、
 下段：揺れる樹木による隠蔽。

電波センサなど多くの方法があるが、それらのセンサが侵入を検知しても、現場の状況を得るために結局TVカメラを必要とすることが多く、画像処理による侵入検知はセンサの数を減らし、検出結果を人が再確認できるという点できわめて合理的である。

画像処理によって人の侵入を発見するシステムの応用として侵入者の行動を詳しく観察・記録するためにカメラのパン・チルト・ズームを自動制御して侵入者を追尾するシステム³⁾が使われている。このようなシステムによって撮影される画像シーケンスの例を図-11に示す。図-8に示した相関追尾のテンプレート更新および図-10に示した背景差分と動きベクトル分布による識別とによって、前景物体が人物を完全に隠すシーン(上段)、影の境界を通過しコントラストが急変するシーン(中段)、揺れる樹木の後ろを通過するシーン(下段)などでも追尾を失敗せずパン・チルト・ズーム制御に成功している。

さらに追尾した人物の行動履歴を軌跡として抽出することにより、多くの監視対象の人物から不審な行動を行った人物だけを抽出することができる。人の歩いた軌跡や滞留時間にはばらつきがあるので、多少の時間的な伸び縮みや位置・方向の変動があることを許容しながら、標準パターンと比較すると、たとえば、攻撃対象を物色している犯罪者の発見や、店舗における動線や品定め傾向の可視化など、防犯やマーケティングに役立つ情報が得られる。

たとえば図-12に示すシステムでは、検出・追跡した人物に対して、検索時のキーとなる識別番号や特徴記述(メタデータ)を求めて、映像とは別に映像監視サーバに記録する。そうすることで、行動パターンを検索キーとして、検索キーにマッチする画像シーケンスを(画像を参照することなく高速に)検索することができる。まず、検出された人物領域をフレーム間で追跡し、時空間的に連続した人物オブジェクトとして記述する。オブジェクトには、オブジェクトの識別子(ID)、出現時刻・消滅時刻(映像フレーム番号)、追跡中のオブジェクト中心座標や外接四角形、どこからどこに移動したかという移動特徴などからなるメタデータが与えられ、さらにインデックス表示に使う代表的なフレームが付加される。このメタデータは、映像シーケンスとは別のデータとして映像監視サーバに記録される。検索用のユーザインタフェースでは、非直線的移動、特定方向への移動、立ち止まり、のろのろなどの行動パターンと、画像中のある一定のラインを超えたかどうかなどの位置を指定することができる。これらの条件にマッチするオブジェクトをメタデータから検索し、長時間の記録映像から特定の行動をした人の画像だけを効率よく表示することができる。

◆ 公共警備への応用(混雑度計測・警報システム)
 駅やイベント会場など人が常に流動している場所

ビル入口での行動検索画面例

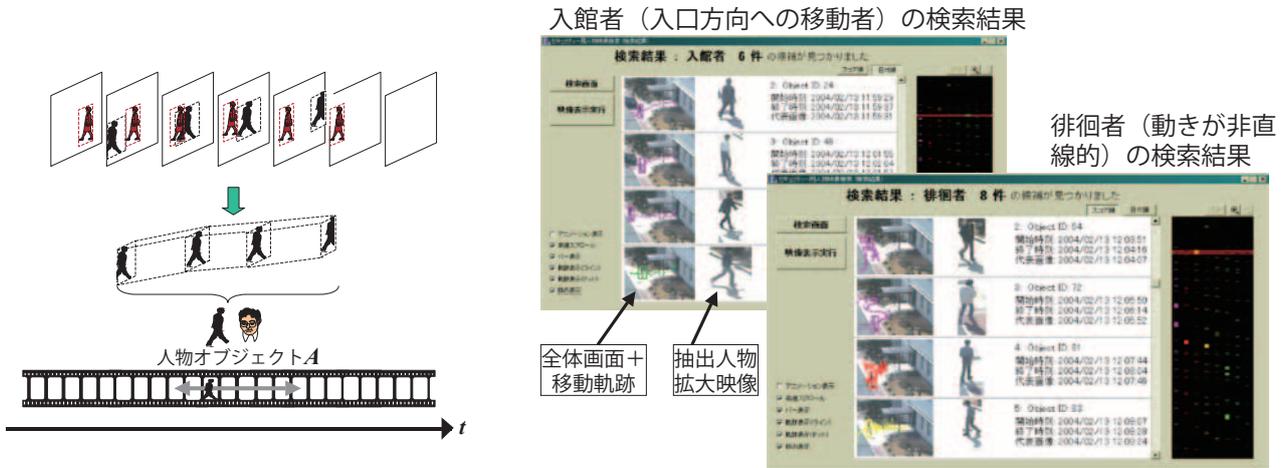


図-12 人物追跡を応用した監視画像サーバの特定行動検索機能の実装例

で、万一人が密集してしまうと事故に至る可能性がある。たとえば、エスカレータの降り口に人が密集してしまうと、後から来る人は行き場がなくなって将棋倒し事故になりかねない。このような状況を自動的に検知して通報することができれば事故を防止することができる。

たとえば、図-13に示すシステムでは、画像から人らしい領域を抽出し、その領域上の人特有の特徴点を抽出、さらにカメラの透視変換パラメータと画面上の位置から計算される人物像の大きさ予測を考慮して人数をカウントしている。

◆ 屋内警備への応用 (エレベータ内の見守り・人物サムネイルシステム)

屋内においても屋外と同じような警備システムが用いられるが、人の存在がより大きく撮影される代

わりに、人物行動についてのより詳しい解釈が求められることが多い。

たとえば、エレベータかご内犯罪事例を分析すると過半数の犯罪では加害者の暴力や被害者の抵抗といった動作が伴い、犯罪に至る前には待ち伏せといった不審状態がある。これらの行動が検知できれば、犯罪抑止の強化や犯罪に遭遇した人の被害軽減に貢献できると考えられる。さらには、昨今の高齢者や子供による単独利用機会の増加を考慮すると、急病で動けなくなったり閉じ込められたりした場合にも対応することで安心確保につながるものと期待される。これに対応するシステムとして図-14のようなエレベータ監視システムが用いられている⁵⁾。このシステムでは、人物の有無判定と、人物の動きパターンの大きさ・分布・不自然さからエレベータ内で動けなくなった要介助者や、エレベータ内での犯罪

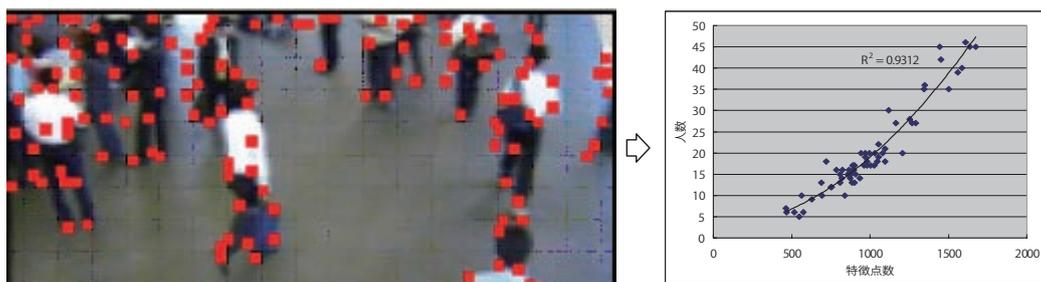
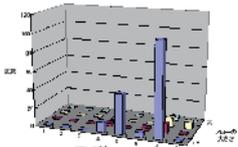


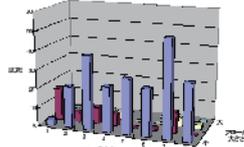
図-13 人物抽出と人物上の特徴点 (図左中の赤点) の分布から求めたホール的人数カウント結果 (右図の縦軸) 真値 (右図の横軸) とよく符合していることが分かる。



正常時の映像シーケンス 異常時の映像シーケンス



フローの分布(正常時)



フローの分布(異常時)

図-14 エレベータ内部のオプティカルフローによる異常行動(暴れ)の検出原理

異常かどうかの判定は、演技を含む多数の事例を用いた学習型識別器を用いている。

行為などを自動検出し、音声案内やエレベータの救助動作などを実現した。

また、監視映像記録システムにおいても、顔検出を用いて検出された人物を分かりやすく表示するような顔サムネイル機能が実用化され映像監視業務の効率向上に役立っている(図-15)。この機能は、映像ストリーム中から人物を抽出・追跡すると同時に各フレームの顔も抽出し、顔の向きと顔の画質の2つを指標にして、その人物が誰であるかを判断するに適した最も顔らしい画像を追跡された1つのオブジェクトに対するサムネイルとして記録するものである。顔をキーとして全記録データからその人物が現れたシーンを検索することもできるし、顔が判別しにくい人物オブジェクトに対して、別画面から抽出されたサムネイルを見ることにより、それが誰であるかを簡単に確認することができる。

まとめ

これまで述べてきたように、人を見る画像処理の基本は、複雑に変化する背景から人物領域を抽出し、前後の画像シーケンスにわたって追跡することで1つの人物オブジェクトとして認識すること、および、人に関連した重要情報として、顔のクローズアップなど個人の同定に役立つ情報や、人物の行動を

- ・顔を検出・追跡、**通行人の顔画像を確実に記録**
- ・顔の向きや画質などの情報を使い、**最も記録に適した画像を選択**

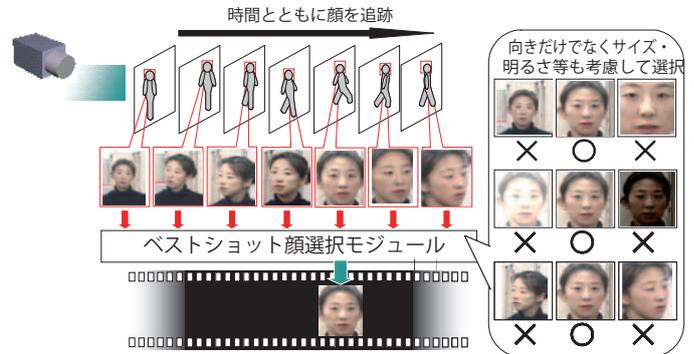


図-15 映像ストリームから人物の顔が最もよく見えているベストショットを検出し、それを検索キーとする顔サムネイル機能の原理

詳しく記述することである。

これらの技術は社会生活における安全安心を高める用途に役立っており、屋外監視、屋内監視を中心に適用が進んでいる。

参考文献

- 1) 松山隆司, 和田俊和, 波部 斉, 棚橋和也: 照明変化に頑健な背景差分, 電子情報通信学会論文誌 D-II, Vol.J84-D-II, No.10, pp.2201-2211 (2001).
- 2) 関真規人, 和田俊和, 藤原秀人, 鷺見和彦: 背景の共起性に基づく背景差分手法, 情報処理学会 CVIM 研究会論文誌 MIRU2002 特集号, Vol.44, No.SIG5 (CVIM.6), pp.54-63 (2003).
- 3) 羽下哲司, 鷺見和彦, 八木康史: 時間平均シルエットを用いた能動カメラによる人の追跡, 電子情報通信学会論文誌, Vol. J88-DII, No.2, pp.291-301 (2005).
- 4) 加藤丈和, 深尾隆則, 羽下哲司: 対象追跡一フレーム間の類似度に着目した手法から動きのモデルに着目した手法まで, 情報処理学会研究報告 CVIM, No.88, pp.185-198 (2005).
- 5) 鷺見和彦, 関真規人, 塩崎秀樹: 安全と安心のための画像処理技術: 3. 画像によるエレベータ内異常検知技術, 情報処理, Vol.48, No.1, pp.17-22 (Jan. 2007).

(平成 22 年 10 月 4 日受付)

鷺見和彦 (正会員) Sumi.Kazuhiko@ds.MitsubishiElectric.co.jp
1982年京大工(電気電子)卒, 1984年同修士(電気電子専攻)修了,
三菱電機(株)入社, 現在先端技術総合研究所所属, その間, 1989
年 Maryland 大客員研究員, 2003~06年京大院情報学研究所客員研究
員, 工博。

† 本稿の執筆に際して三菱電機(株)羽下哲司氏と関真規人氏の研究成果を使用させていただいたことを感謝申し上げます。