

# 物体検出

## — 背景と検出対象のモデリング —

鷲見 和彦<sup>†</sup> 関 真規人<sup>‡</sup> 波部 斉<sup>††</sup>

<sup>†</sup>{sumi,habe}@vision.kuee.kyoto-u.ac.jp <sup>‡</sup> Seki.Makito@wrc.melco.co.jp

<sup>†</sup> 京都大学大学院 情報学研究科 知能情報学専攻

<sup>‡</sup> 三菱電機株式会社 先端技術総合研究所

<sup>††</sup> 京都大学大学院 工学研究科 電子工学専攻

物体検出は実世界の画像を入力とする画像認識において基本的な問題である。物体検出の問題は、画像を対象物体とそれ以外(背景)との二つの領域に分離する識別問題あるいは領域分割の問題とも定義できる。対象物体の見え方を基準に対象物体を探す場合と、背景の見え方を基準に背景ではないものを対象物体とする場合があり、それらの基準はそれぞれ検出対象モデル・背景モデルと呼ばれる。また、このモデルは見え方に基づいていることが多く、先見的知識だけでなく実際の設置環境における背景と物体の見え方を学習することにより構築されることが多い。このサーベイでは、最近の様々な画像認識研究事例から、背景モデルと検出対象モデルに焦点をあてて、物体検出技術とは何かを論じ、加えて、モデルの学習方法や識別器の利用などについても分析する。

# Object Detection

## — Model of Foreground and Background —

K. Sumi<sup>†</sup> M. Seki<sup>‡</sup> H. Habe<sup>††</sup>

<sup>†</sup> Department of Intelligence Science and Technology  
Graduate School of Informatics, Kyoto University

<sup>‡</sup> Advanced Technology Research and Development Center,  
Mitsubishi Electric Corporation

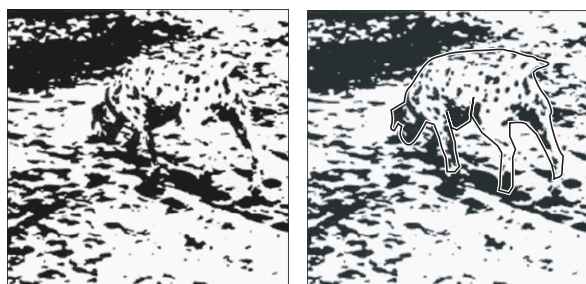
<sup>††</sup> Department of Electronic Science and Engineering  
Graduate School of Engineering, Kyoto University

**Abstract.** Object detection is a basic problem of image understanding in a real environment. It can be defined as means to segment an image into foreground regions and background regions. There can be two approaches in object detection. One is to use characteristics of appearance of target objects. The other is to user characteristics of background. Those characteristics are referred as to foreground model and background model respectively. Those models are obtained not only by prior knowledge but also by examples gathered on line under operation. In this survey, we focus on background model and foreground model proposed in recent object detection researches. We also discuss model acquisition, learning, and classifier employed in them.

# 1 まえがき

物体の検出は、実世界の画像を入力とするほとんどのすべての画像認識において必要な処理である。実世界の画像認識では、認識対象とそれ以外の物体が混在しているのが通常である。ある認識対象について、その位置や姿勢を求め、あるいは対象から詳細な情報を得ようとした場合、対象とそれ以外が混在していると正しく情報を抽出できない。そのため、認識の初期の段階で認識対象が画像上で占める領域を検出し、非対象領域を除外した上で、対象領域ごとに必要な認識を行うことが望ましいからである。

しかし、物体の正確な位置・姿勢・見え方がわからない状態で、物体を背景から正確に抽出することは一般的に困難な問題である。たとえば図 1(a) において、そこに何が写っているか予備知識のない状態で犬とその境界を認識することは人にとってもたいへん難しい問題である。「草地に白黒のまだら模様の犬がいる」と聞けば、多くの人は数秒でこの絵を解釈することができる。それは、人が犬の形・模様・取り得る姿勢などと、地面や草木の模様とを予備知識にして、図 1(b) のように仮に輪郭を想定し、その解釈の尤もらしさを判断できるからに他ならない。それでもなお、その解釈が正解かどうかは簡単には判断がつかない。



(a) Dalmatian Dog  
(Richard Gregory)

(b) Estimated boundaries of the dog

図 1: 隠し絵 ダルマチア犬 (Richard Gregory)(a) は予備知識なしには何が描かれているのか分かり難いが、その解釈が与えられると犬の輪郭 (b) が感じられる。

このように、物体検出とは「何 (形状や内部の模様)」があるかがわかっているときに、それが「あるか」あるとすれば「どこ」に「どんな姿勢」であると考えれば良いかを求める問題であると言える。あるいは、画素単位で考えたときには、

ある画像が与えられたときに、それに含まれる個々の画素をそれぞれ前景と背景とのいずれかに識別する問題であるとも考えることができる。

物体検出を画素をそれぞれ前景と背景とのいずれかに識別する問題であると考え、画像処理の知識のある人は画像の二値化を真っ先に思い浮かべることであろう。画素を 1 または 0 で表現する二値画像は、一方が前景で他方が背景として扱えば物体検出の一種である。そのため、輝度や色が一律な背景を想定することができる文書画像や初期の工業用ビジョンでは二値画像処理が好んで用いられた。一般的には、背景の輝度や色が一律でなかったり、模様やパターンを持っていたり、またそれらが時間的に変化したりするなど現象はもっと複雑であり、それに対処するために、背景差分や物体検出などの技術が発展した。

このサーベイでは、まず、関連する応用事例の歴史をたどって物体検出とは何かを考え、ついで、物体検出のフレームワークを改めて定義した。その中で、筆者らが物体検出の本質であると考えた背景モデルと検出対象モデルとを中心にこれまでの技術を整理した。さらに、関連するセンシング手法・学習・識別器の利用についても物体検出の視点から論評を加えた。なお、物体検出の対象として人物検出は大きなテーマであるが、人物全体の抽出や人物のパーツの抽出については本稿にて取り扱うこととするが、本サーベイシリーズにおいて「顔画像認識」が単独で調査・整理されている [1] ため、顔検出については本稿では触れない。また、学習・識別技術そのものについても、本サーベイの範囲外と考えている。学習識別については、本シリーズ [2] および関連文献 [3, 4] を参照されたい。

## 1.1 物体検出技術の発展と応用

物体検出はコンピュータによる画像理解とともに進歩してきた。この中でも物体検出技術に大きく関わるのはロボットビジョン、医用画像処理、文書画像処理、および最近では画像圧縮とデータベースである。

まず、初期の画像認識技術においては、そのアプリケーションとして工場における組み立て作業の自動化などに用いられる産業用ロボットビジョンが注目された。これら産業用ロボットビジョン [5] [6] [7] では、既知の二次元・三次元形状と外観とを持つ工業用部品を、位置・姿勢が自由な条件で検出し、その位置と姿勢とを求めるものが多い。対象は工業部品のようにほぼ同じ形状と模様を持つので、認識対象の見え方をモデルとして持つことが一般的であった。見え方のモデルとしては、人工的な照明を当ててその反射パターンを二値画像として粒子解析するも

のが多かったが、見え方を二次元のパターンとしてパターン間の距離計算を行うテンプレートマッチング [8] など現在のアピアランスによる対象モデルの原型がすでに 1970 年代に国内で実用化されている。

一方、工業用ビジョンのもう一つの応用は外観検査である。外観検査においては欠陥や異常が検出対象となるが、欠陥や異常は発生頻度が少なく実例を用いてその外観をモデル化することができないため、主に、正常な外観を持つサンプルがモデルとして用いられ、モデルから離れた外観を持つ領域を欠陥として抽出するという枠組みが考案された。この考え方は現在の背景差分に極めて類似しており、[9][10] に初期の技術を見ることができる。

しかしながら、工業製品の世界では、認識対象の回りをとり囲む安定した環境が作りやすく、対象の形や模様が均質であることが多い上に、高いスループットとコストパフォーマンスが求められたため、1980 年代には、対象モデル・背景モデルに対する一般的な発展よりも、可視化技術や照明技術など撮像と利用ノウハウの蓄積に主眼がおかれ物体検出アルゴリズムに関する進展は少ない。1980 年代後半には画像処理専用のハードウェアの実用化が進み、それに伴って認識対象のアピアランスをモデルに持つ二次元の画像テンプレートマッチングが広く普及した。

1980 年代以降、研究としての画像理解の対象は形状や条件の安定したオートメーションの世界から、移動ロボットナビゲーション・自動走行・ヒューマンインタフェイス・セキュリティなど一般室内や屋外の非整備環境での認識技術へと展開した。特に、米国の DARPA によって主導された、ALV (Autonomous Land Vehicle)[11]、VSAM (Video Surveillance and Monitoring)[12] では、屋外環境における道路や地形の認識、人や車両の認識などが研究され、検出対象や背景のモデルが進化するきっかけとなった。日本においても、ITS への期待や、セキュリティに対する関心の高まりを受け、非整備環境における物体検出・背景差分の性能向上は、大学など研究機関でも企業でも盛んになっている。

医用画像処理においては、まず、レントゲン画像、CT/MRI などの再構成画像、超音波エコー反射像などの画像を対象に臓器や器官の抽出が行われた。ただし、医用画像の場合には診断対象の臓器が存在する画像上の位置が撮影時の設定によってほぼ決まっていることが多く、臓器の存在と概略の位置を前提として、その輪郭や境界を正確に求めるという事例 [13] が多い。X 線画像における骨組織のように高コントラストに撮影される場合は少なく、多くの認識対象の境界は他の器官と重なって映っていたり、ノイズと見

分けるのが困難だったりしている。そのため、多くの研究では低コントラストの画像への既知の形状の輪郭の当てはめ問題として取り組まれた。もう一つの物体検出技術は、腫瘍などの病変部位の抽出 [14] である。この場合には、大きさや形状がまちまちであることが多く、テクスチャなど内部のパターンのもつ特徴や輪郭の特徴などを用いて正常組織との識別問題として取り組まれることが多い。

文書画像 [15] においては、まず無地の背景の上に記された文字や記号を正確に抽出する技術、特に、照明の輝度の不均一を補いつつ、文字の掠れを許容するための二値化アルゴリズムとして技術が発展した。この場合、文字・図・記号が物体であり、それ以外の部分が背景であると解釈できる。1980 年代になって、従来の無地背景ではなく写真やイラストの上に書かれた複雑な背景の文書画像からの文字抽出が行われた。また、画像と文字の混在する複合文書が扱われ、文書画像を文字領域とその他の領域とに分割する手法が研究された。この場合には、個別の文字ではなく、その並びや領域がそれぞれ物体として扱われていることになる。

最近 10 年間の傾向として、物体検出は画像圧縮や画像データベース検索においても重要な技術となってきている。たとえば、動画像圧縮 [16] においては、カメラから見て前後関係が存在するシーンを移動するカメラから撮影した画像シーケンスや移動する物体を撮影したシーケンスを複数のレイヤーに分離してそれぞれを圧縮し、再生時にレイヤーの隠蔽関係を考慮して合成することにより、圧縮率を高めたり精度を最適に制御することができる。また、画像検索においてはあらかじめ映っている物体や背景に分解した上でそれぞれの特徴を記述することで、検索精度を高めることができる。

## 2 物体検出概論

本章では、これまでに説明した物体認識に関する技術の展開を見直し、物体検出に関する共通概念を整理する。まず、物体検出とは何かという問題定義を行い、次に、物体検出における技術課題について整理する。最後に、物体検出の性能評価法についての指針を示す。

### 2.1 フレームワーク

第 1.1 節で物体認識技術の発展と適用事例を概観したが、それらに共通する考えとして、物体検出とは入力画像を物体の映っている領域 (前景) とそれ以外 (背景) とに分類する問題である — と考えられる。この場合、問題は前景と背景との 2 クラス識別問題となり、その枠組みは図 2 で表される。このような枠組みで表現できる物体検

出とは、物体と背景とが単一の特徴空間において分離可能な分布を持つ場合であり、たとえば、明るさ・カラー・時間変化など画素ごとの特徴量や、フィルタバンクや周波数解析によって得られる空間的パターン特徴などが評価基準(メトリック)として用いられる。

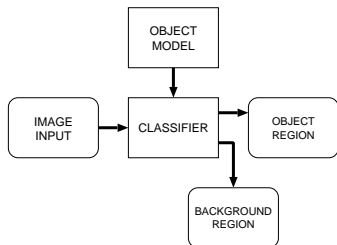


図 2: 物体検出の基本的アーキテクチャ。

ところが、現実の応用事例において物体検出を単純なクラス識別問題として定義することは難しい。その理由は、基準データまたは学習サンプルを集める難しさに原因がある。たとえば、図 3 において、物体と背景とを識別する識別器に与えられるモデルは、入力画像から得られる特徴ベクトルとそれに対する正解との組み合わせとして学習アルゴリズムに与えられ、識別結果の誤りを最小にするようなモデルが生成される。識別の精度を良くするためには、適切なメトリック(すなわち特徴ベクトルと距離の計算式)と、最適な識別境界を決めるための基準データまたは統計的に十分有意な多くの事例を必要とする。

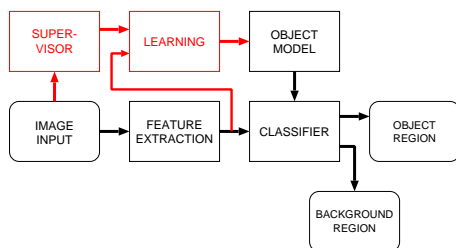


図 3: モデル学習を考慮したアーキテクチャ。

ところが、現実には以下に述べるような理由で、これらを獲得することが困難なことが多い。まず、第一に、モデル学習時にすべてのバリエーションを持つサンプルが集められない。たとえば、人物を検出する場合に、人物には体格・姿勢・着衣など多くのバリエーションがあり、ある視点から見た全変化範囲を網羅する画像例を集

めることは困難である。人のように姿勢や形状の変化の少ない人工物である自動車であっても、車種・姿勢・彩色・積載物などで予想できないアピランスの変化がある。たとえば学習後に発売された新製品は実例としてはモデリングできない。次に、モデルの精度を高めようとして情景に関する制約条件をうまく使おうとしても、物体と背景とでは利用できる制約条件が異なっている。たとえば、固定視点から見た静止背景は画素毎の重なりが大変良い。したがって同じ画素は背景上の同一の点を表していると考えれば都合が良いが、前景物体は移動するのでこの制約条件は当てはまらない。

このように考えると、一つの識別器だけで物体と背景とを誤りなく識別するよりも、背景と前景に関する制約条件をうまくモデルに取り入れて、背景と物体とを別々のメトリックでモデル化して個別の識別器を構成し、順次候補を絞り込んで行くカスケード検出のフレームワーク(図 4)や、並列検出器の統合(図 5)が解決策として考えられる。カスケードかパラレルかのアーキテクチャ選択は、個別の検出器の性能と処理コストのトレードオフによって決められる。カスケード構成では個別の識別器で誤って検出対象物体がふり落とされると誤りを回復不可能なのに対して、並列アーキテクチャではそれぞれの識別器の出力信頼度の強さによって回復可能であるというメリットがあるが、共通する学習サンプルを集めなければならない上に、最適化すべきパラメータが多いので学習のコストも高いことが欠点である。そのため多くの実用化事例ではカスケード構成を取ることが多い。なお、顔検出をはじめとして最近多用される多数の弱識別器にブースティング学習を組み合わせたアーキテクチャ[17]の改良では、カスケードと並列との併用[18]もみられる。

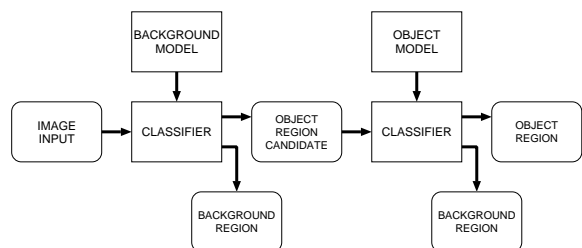


図 4: 実用的な物体検出のためのカスケード構成。

## 2.2 物体検出における課題

環境が制御しにくい実世界での物体検出では、机上や実験室環境で想像している以上に、物体

表 1: 物体検出に影響する要因の整理

項目	現象の説明	識別への影響	代表的対処法
緩やかな照明変化	時間とともに変化する日照など。	単純に記憶した背景との差分をとると大きな残差が残り誤検出する。	背景をゆっくりした変化に追従させて更新する。誤検出率を最低にするために更新速度の調節が難しい。
急激な照明変化	雲や遮蔽物の移動・照明の点灯/消灯による瞬時的照明変化。	画面全体に物体が現れたかのような過検出を生じる。	輝度の正規化による線形変化の補償。照明基底の導入による任意の照明条件の予測。
背景の構造変化	移動体が通り過ぎたあとに痕跡が残り背景が変化する。背景物体が急に移動する。	新たな物体が出現した様に誤検出される。	背景更新の改良。変化した領域と検出対象の識別を追加する。
揺れる・点滅する背景	木や旗がはためくことによる画素の繰り返し変動、信号や広告照明など短周期で点滅する照明とその反射、水面のさざ波や反射など。	周期的な現象である場合が多いが変化が画像全体で一様ではないことが多く、過検出を発生しやすい。	周期性のある変動を履歴で記憶したり、基底展開などでパラメータ化する。
検出対象ではない前景	カメラの前に現れた虫や鳥。背景に投射されたスポットライト(視野外の車のヘッドライトなど)。	過検出	検出対象と過検出対象の識別をカスケードに行う。
検出対象のアピランス変化	検出対象の形状・色・模様時間がともに変化する。立体的な対象の視点移動や対象の個体差など。	未検出	形状モデルの導入
対象物体や他の物体による影	検出対象や背景上に他の物体や自分自身の影が投射される。	影の部分を誤検出する。	影を補償する。不変特徴や距離の利用。
検出対象の長時間静止と再移動	検出対象が静止して画像上の変化が止まる。	静止後に見失う。移動後に背景を過検出。	背景の構造化(レイヤーを持つ背景モデル)
検出対象の隠蔽	検出対象の手前にそれを隠蔽する物体が存在。	部分的にしか見えない。検出不能・誤検出	部分-全体モデルの導入
背景と類似した対象	輝度・色・テクスチャなどの特徴をとっても背景と前景が良く似ている。	検出不能。	距離画像の差分や動きの違いで検出できる場合がある。
実例のない前景や背景	めったに発生しない背景内の現象。システム稼働後に新たに生じる検出対象。	検出不能・誤検出	実サンプルに基づくモデルではなく先見的・知識的モデルを用いる。
背景の部分的隠蔽	歩行者や交通などが途切れることがなく、完全な背景が観測できない。	単純に背景を求めると平均化された検出対象が背景に混入し、対象の検出感度が低下する。	背景学習における移動体検出やアウト라이어除去による前景物体のマスキング。
移動するプラットフォームからの撮影	自律移動ロボットや自動車などカメラを搭載したプラットフォームが進むため、背景も検出対象も大きく動く。	通常背景差分では対処不能。	背景の画像上での移動に関する拘束条件を求め、自己運動によらない成分を持つ対象を検出する。
カメラの非線形性	黒つぶれ・白飛びなどカメラの非線形現象	照明が変化しただけなのに新たな物体が出現したような誤認識を生じる。	過去の輝度変化履歴の記憶。黒つぶれ・白飛びの例外処理。
カメラの揺れ	風や振動によってカメラが揺れることにより、背景を含めすべての情景が動く。	誤検出または検出感度の低下。	背景の移動を補償する(背景画像の安定化)
視界不良(ノイズ・悪天候)	極度のノイズ(豪雨・濃霧・降雪)や照明の直射による光斑。	検出不能。	低コントラストでも物体と背景とを識別できるモデルに切り替える。



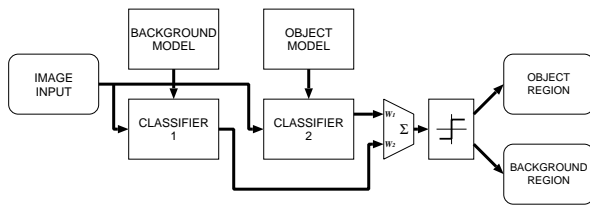


図 5: 個別の識別器の出力を並列統合する枠組み.

検出の識別器を惑わせる様々な現象が発生する。本節では、物体検出を困難にする種々の現象や制約条件について整理する。表 1 に、良く知られている課題と影響および、その対策をまとめた。なお、対策の項は一般的な方針であり、個別の技術に関しては次章以降の技術分析を参照されたい。

### 2.3 性能評価

第 2.2 節で述べた様に、物体検出には多くの課題があり、利用する側にとってはある応用においてどのアルゴリズムを選択すべきか決定することが難しい。また、新しいアルゴリズムを研究しそれを公表する場合にも、どのようにその性能を表現すべきかをあらかじめ理解していなければならない。本節では物体検出の性能評価について、その指針を明らかにする。

性能評価に関しては、まず誤認識の定義を行わなければならない。物体検出においては誤認識には二つのタイプがある。

**タイプ 1** 統計的検定における第 1 種の過誤で、物体検出においては物体が存在するのに検出できない未検出あるいは検出漏れを意味する。fail to detect, または false negative とも表現される。

**タイプ 2** 統計的検定における第 2 種の過誤で、物体検出においては物体が存在しないのに検出した過検出を意味する。false detection, false alarm, または false positive とも表現される。

仮に、物体検出が一次元の類似度  $Q$  に基づいているとすれば、検出対象物体とそれ以外の物体に関する  $Q$  の出現頻度は図 6 の様に分布することが一般的である。ある物体検出のしきい値  $T_h$  を変数として、 $T_h$  以上を検出対象としそれ未満を非対称と判断すれば、図上にハッチングで示した領域がそれぞれ未検出 (False Negative) と過検出 (False Positive) の発生度数を表している。このように、未検出誤りを少なくしようと  $T_h$  を下げると過検出誤りが増加するというトレードオフがある。物体検出が類似度ではなく実例の

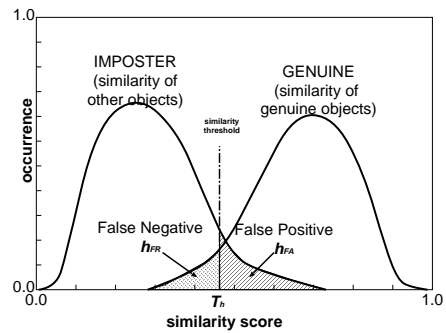


図 6: 物体検出における類似度分布の例

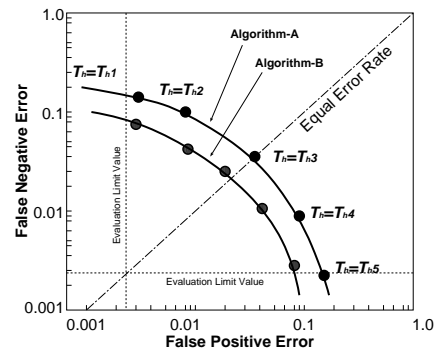


図 7: 物体検出における照合精度曲線 (ROC) の例

サンプルに基づいて異種サンプルへの距離が等距離になるように識別境界を決定する境界学習型のアルゴリズム (SVM など) の場合には、しきい値  $T_h$  という概念が存在しないが、識別マージンの取り方がしきい値の調整に相当する。二つのアルゴリズムを比較するには、どのしきい値  $T_h$  における未検出率と検出率を比べて良いのか分かりにくい。そこで、しきい値  $T_h$  を媒介変数として過検出率に対する未検出率をプロットした照合精度曲線 ROC (Receiver Operator Characteristics) カーブ (図 7) を用いる。ROC は DET (Detection Error Tradeoff) とも呼ばれる。ROC カーブにおいて、二つのアルゴリズムの振る舞いをプロットした場合、エラー率の低くなる (すなわち左下側) にプロットされるアルゴリズムが優れていると言える。

未検出率 (false negative) ではなく検出率 (detection rate) で表現したものも見受けられるが、アルゴリズムの性能が高まってくるとその値は 1.00 に近くなるので図 7 に示すように、検出率ではなく未検出率で表現し、さらに両軸とも対数で表現することが望ましい。

もう一点注意すべきは、評価限界値 (Evalu-

tion Limit Value) の存在である．評価試験に用いられた試行の数  $N$  における統計的評価の限界を意味する．試行が互いに独立であると仮定すると，未検出率  $p_N$  あるいは，過検出率  $p_P$  の信頼度  $\beta$  での信頼区間の上限  $p_H$  と下限  $p_L$  は，あるしきい値における誤りの個数  $M$  を用いて，それぞれ次式で与えられる

$$\beta = \sum_{X=0}^M \frac{N!}{(N-X)!X!} p_H^X (1-p_H)^{N-X} \quad (1)$$

$$\beta = 1 - \sum_{X=0}^M \frac{N!}{(N-X)!X!} p_L^X (1-p_L)^{N-X} \quad (2)$$

統計的評価の限界である  $p_{min}$  は式 2 において誤りなし ( $M=0$ ) の場合の  $p_H$  となるので

$$\beta = 1 - (1-p_{min})^N \quad (3)$$

一般的に統計的信頼区間として 95% 信頼区間が広く用いられるので  $\beta = 0.95$  において  $p_{min}$  が十分小さいと仮定して次の近似を用いる．

$$N \approx \frac{3}{p_{min}} \quad (4)$$

これは，たとえば未検出率 1% を統計的に証明するには，最低 300 回の試行を行わなければならないことを意味している．

### 3 背景のモデリング

本章では，背景のモデリングやそれを利用した背景と前景の識別技術（広義には背景差分）について，研究事例を分析し，整理する．これまでにいろいろなモデルや識別尺度が提案されているが，それらの比較については文献 [19] や [20] にも一部記載があるため，参照頂きたい．

表 1 から分かるように，画像内で観測される背景には，見掛け上の変化も含めてさまざまな背景変化がある．このような動的背景に対するモデリングには主に 4 つのアプローチが採られている．1 つは背景変化に対して不変な特徴を利用したアプローチである．この方法では，背景変化の種類を陽に（または暗に）限定するため適用先が限られるものの，モデルを時間に沿って更新する必要がなく，背景差分法の実現は容易である．2 つ目はサンプル画像をもとに背景変化を統計量で記述するモデリングである．背景変化の種類を問わず適用できる可能性を秘めていると考えられるが，検出感度やモデル更新に対する柔軟性，実装面での課題を残すものも多い．3 つ目は背景変化に時間的，空間的な相関関係があると仮定し，その特徴を記述するものである．仮定が強いいため，第 1 のものと同様に適用先が限られるが，仮定が成り立つ場合には検出感度が高いという特長を有する．4 つ目は直近または基底の画像系列から現在の背景を推定するもので，リアルタイム的な要素を持っている．

### 3.1 不変特徴モデル

#### 3.1.1 テクスチャ

緩急の照明変化に不変な（あるいは頑健な）画像特徴にテクスチャパターンがある．最も単純には，画像をベクトル表記し，ベクトルの大きさを正規化することでテクスチャパターンが記述できる．長屋らは部分画像ごとの正規化ベクトルで背景をモデル化し，入力画像と背景画像における正規化ベクトル間の距離を測ることで背景差分を実現している [21]．この距離は正規化距離 (Normalized Vector Distance) と呼ばれ，後に，松山らもこの正規化距離による背景差分を改良し，後述する照明条件の推定に基づく背景差分と組み合わせている [22, 23]．

厳密に言えば，部分画像内が一様な倍率で変化している場合にのみ，正規化ベクトルは不変な特徴となる．これに対し，佐藤らが提案した周辺増分符号 (Peripheral Increment Sign) [24, 25] による記述は，着目画素 16 近傍の明度増減を 0,1 に符号化するため，良くも悪くも非一様な変化に対して若干の許容範囲を持ち合わせていると考えられる．また，入力画像と背景画像における周辺増分符号の相関を求めることで実現した背景差分法は，計算コストが低く実用性が高い．なお，着目画素と放射線上の適切な位置の画素との間の明度増減を符号化する派生モデルも提案されている [26, 27, 28]．

画像から得られる空間周波数もテクスチャを記述できる特徴量である．伊藤らは FFT により背景画像系列から部分画像ごとにパワースペクトルが極大となる 2 つの空間周波数成分を求め，それら周波数成分とそのパワースペクトルにより背景をモデル化した [29]．照明変化や背景物体の微小な動きに対して頑健であるのはもとより，2 つの空間周波数成分に着目することで 2 つの背景状態を記述できるようになり，例えば点滅する背景や，木の揺れで枝と葉が交互に出現する場合にも対応できるようになっている．

また，空間周波数を具体的に求めることなく，その特性だけを簡便に利用した方法もある．阿部らは，対象の出現による変化に比べ日照変化やノイズによる画像変化が特定帯域の空間周波数成分にのみにしか影響を与えないことに着目し，Sobel フィルタによる高周波強調画像の背景差分結果と平均化フィルタによる低周波強調画像の背景差分結果から両方に変化が現れている領域を抽出している [30]．

そのほか，小領域における照明強度の分布が変化しないという仮定のもとで，画像間の照明強度比を推定し，このときに生じた残差を用いた  $\chi^2$  検定で画像間差分を行う方法も提案されている [31, 32]．

### 3.1.2 視差，距離

ステレオ画像などから得られる視差や距離情報も照明変化に頑健な特徴である。また、足跡や轍などの背景構造変化のほか、カメラ前に現れた虫や鳥など検出対象ではない前景物体を排除するのも役に立つ。そのため、背景を視差でモデル化し、背景差分を行うものも提案されている。単純には、複数視点の入力画像から毎回視差画像を算出し、背景の視差画像と比較すればよいと考えられるが、これには以下の問題がある。

- 視差得るための対応点探索に時間がかかる
- 均一または低コントラストな領域では視差が得られない
- ノイズによる影響を受ける

視差の計算時間については、特別な H/W を用意することで解決することもできるが [33]，対応点探索を省略する方法も提案されている [34],[35]。例えば、Ivanov らはステレオカメラから得られる背景画像間の対応点マップ (warp map) を事前に生成しておき、マップに基づいて入力画像間の差分を行うことで視差に変化が現れた領域を抽出する背景差分法を提案している。この方法には、前景物体とあわせて occlusion shadow も抽出されるという欠点があるが、3 台以上の多視点カメラがあれば、それが軽減できることも示唆されている。

一方、均一または低コントラストな領域については、周辺から内挿したり、多様な照明条件やカメラゲインで撮影された画像から視差を計算したりすることで解決できる場合がある [35],[36]。また、色情報など別の尺度を用いる背景差分と組み合わせることで解決できる [37]。

ノイズについては、後述する統計モデルの概念を採り入れ、ノイズによる視差の変動幅を平均値と分散値でモデル化したり [38]，視差ヒストグラムの最頻値周辺を背景モデルとしたり [39] する方法が提案されている。

## 3.2 統計モデル

一方、照明変化が均一ではない、背景物体に動きがある、あるいはステレオカメラが使えない場合など、不変な特徴でモデリングできない背景変化に対しては、統計量によるモデリングがよく行われる。また、その統計量はサンプル画像として与えられた背景画像系列より計算されることが多い。

### 3.2.1 平均値，中央値，最頻値

最も簡単な方法は、過去に観測された背景画像系列から画素ごとに平均値，中央値 [40]，あるいは最頻値 [41] を求め、背景画像とすることであ

る。また、その時系列的な重みとして新しい画像に重みを持たせる方法もよく行われる [42],[43]。時間的に緩やかな変化であれば、変化に追従させてモデルを更新することで対応可能であるが、現実にはこのモデル単体で背景差分がうまく動作する環境は少ない。

### 3.2.2 最小・最大値，振幅

画素ごとに背景変化の範囲を記述するものも提案されている。例えば、Haritaoglu らは画素値の最小・最大値およびフレーム間変化の最大値で背景をモデル化している [44, 45, 46]。また、Yamazawa らも画素値の平均値と振幅で背景をモデル化している [47]。背景差分は、これら範囲を超えた画素を検出することで実現される。サンプル画像など事例ベースにモデル化する場合は、混入した前景物体やノイズの影響を強く受けるため、注意が必要である。

### 3.2.3 正規分布

画素値に対する背景としての確率を正規分布パラメータで表現するものも提案されている。最終的に閾値処理にて背景の範囲が決定されることを考えれば、先のものと同種と解釈できるが、アウトライヤには強い。このモデリングには、画素ごとに YUV カラー値の分布をモデル化 [48] したり、部分画像ごとにパターンの分布をモデル化 [49, 50] したりする方法が提案されている。また、画素値だけでなく特徴量を使うものとして、各画素での移動ベクトルと画素値の変化量をモデル化 [51] したり、正規化距離をモデル化 [22, 23] するものもある。

### 3.2.4 混合正規分布

振幅や正規分布でのモデル化は、背景変化の確率分布が単峰である場合に限られ、草木の揺れなど複雑な背景変化を表現できない。これに対し、Grimson らは各画素のカラー値に対し、 $K$  個 (3 ~ 5) の正規分布成分からなる混合正規分布 (Mixture of Gaussians) でモデリングを行っている [52, 53, 54]。唯一の欠点は、背景に関する事前知識に基づいて成分数を予め設定しなければならないことである。しかし、多峰性の分布が記述でき、 $K$ -means 近似等によるオンラインでのモデル更新もできるという利点は大きく、実用性が高いといえる。そのため、現在では広く用いられており、色と距離の空間でのモデル化 [55] やパンチルトカメラ画像への適用 [56],[57] も報告



されている。

### 3.2.5 ヒストグラム

混合正規分布でも近似できない、より複雑な背景変化には、サンプル画像系列から画素ごとに画素値のヒストグラムを求め確率分布とする non-parametric な方法がある [58, 59]。ただし、統計的に意味のある分布とするためには大量のサンプルが必要となる。

これに対し、Elgammal らは、比較的少数のサンプルで滑らかな確率分布ヒストグラムを構成するために、カーネル密度推定 (kernel density estimation) により画素値の発生確率を推定する方法を提案している [60]。この方法は計算時間がかかるという欠点がありモデルの更新は事実上できないが、後に Han らによって提案された再帰的カーネル密度近似 (Sequential kernel density approximation) による方法 [61] は、度数の最頻値予測を用いて再帰的に確率密度を近似するため、計算時間の遅さが改善されている。なお、最頻値予測には、variable-bandwidth mean shift 法 [62, 63] が用いられている。

### 3.3 相関モデル

本節では、背景変化の時間的・空間的相関関係をモデル化する 3 つ目のタイプについて述べる。

#### 3.3.1 時間的相関

足跡や轍といった背景に構造変化が生じたとしても暫くはそのまま定常状態が続くことなる。長屋らはこれに着目し、部分画像ごとの時間的な相関値の一定性により背景をモデル化している [21]。この方法では、ある特定時刻の部分画像と各時刻の部分画像の相関値が一定値を保っている時間帯を背景時間帯とみなし、その値の大きさにより背景構造変化の有無を判定している。

また、波部らは部分画像ごとに時刻  $t$  と  $t + \Delta t$  における正規化距離に相関があると仮定し、それら共起頻度の分布で背景変化をモデル化している [64]。この方法では、共起頻度の分布形状を解析することにより、背景変化を 5 つのクラスに分類し、クラスごとに異なる背景差分が行われている。

#### 3.3.2 空間的相関

一方、空間的な相関関係に着目したのもある。画像内での明るさの変化率が一定であると仮定し、予め設定された画素や時間差分、背景差分

の結果から背景領域における画素値の変化率を求め、前景物体で隠された背景も同じ変化率で更新する方法が提案されている [65], [66], [67, 68], [69]。同種のものとして、背景差分と時間差分で変化領域を抽出し、各変化領域における変化量の分散が小さい場合には照明変化とみなすものもある [70]。

また、関らは画像全体に渡る仮定ではなく、局所領域間での関係として、背景変化の空間的連続性を仮定し、隣接した 2 つのブロックに現れる部分画像パターンの共起関係により背景をモデル化した [71, 72]。この方法は、木の揺れなどの背景物体の変動だけでなく、日照や天候の変化など、空間的共起性の強い照明条件の変化にも対応することができる。

### 3.4 推定モデル

4 つ目のタイプとして、直近または基底の画像系列からリアルタイムに生成される背景モデルについて述べる。

#### 3.4.1 履歴画像

島井らはロバスト統計手法として知られている M 推定を利用し、直近 N フレームの入力画像から現在の背景を推定している [73]。このとき、時間軸にそって指数的に累積した重みつき推定誤差を最小にするように背景の画素値を推定することで、背景の時間的変化への追従性を高めている。M 推定はアウトライヤに対する頑健性を持っているため、前景物体が混入する入力画像系列からも背景を推定することができるという特長をもっている。

そのほか、ウィナーフィルタにより背景を予測する方法 [19] やカルマンフィルタにより背景画像を推定する方法 [74], [75], [76], [77]、ロバストカルマンフィルタにより背景画像を推定するものも提案されている [78]。

#### 3.4.2 基底画像

一般的に、照明変化などをともなった画像系列は、画像空間において、より低次元の部分空間中に分布する [79], [80]。例えば、

- 物体表面は完全拡散面である
- すべての光源は無限遠点にある
- 影が生じない

が成り立つとき、任意照明下でとりうる画像はたかだか 3 次元の部分空間に分布することが知られている [81, 82]。また、

- 物体が凸形状で、attached shadow も生じる

場合は、画像空間の原点を頂点とする凸錐を形成するも知られている。この凸錐は“*Illumination Cone*”と呼ばれる [83]。なお、実シーンではこれら仮定を完全に満たすことは少なく、画像を拡散反射成分のみに線形化することも行われている [84]。

いずれにしても、こういった部分空間は少数の基底ベクトルで構成される固有空間として近似(定義)することができる。Oliverらはこれを利用し、背景画像系列から構築された固有空間に入力画像を射影することで、その基底ベクトル(通称“*Eigenbackgrounds*”と呼ばれる)の線形結合により入力画像に対する背景画像を推定している [85]。また、Incremental PCA (IPCA) [86] 等を使うことで、この固有空間をオンライン生成、更新する方法も報告されている [87],[88],[89]。

同じく、土田らも予め学習された複数の背景画像を基底画像とし、それらの線形和で入力画像に対する背景画像を推定している [90]。この方法では、逐次モンテカルロ法を利用することで、緩やかな照明変化に起因する線形係数の変化を予測している。

しかし、これらの方式には、入力画像における前景領域が小さくなければならないという制約があり、前景領域が大きい場合に推定誤差が大きくなるという問題がある。これに対しては、

**Step1:** 推定背景画像を用いて背景差分により前景領域を抽出

**Step2:** 入力画像中の前景領域を除去、または(前)背景画像で置換したものに対し背景画像を推定

の2ステップを繰り返すことにより、領域抽出結果/背景推定結果を最適解へ収束させることができるという報告がある [22, 23],[91]。

なお、ブロック単位に固有空間を構成し、固有空間への射影ベクトルと入力ベクトルとのなす角を評価したり [88]、固有空間への射影ベクトルと固有空間原点とのマハラノビス距離(固有ベクトルの寄与率を考慮した距離)を評価したりして [49, 50]、各ブロックが前景領域に含まれるか否かを判定する背景差分も実現できる。

### 3.4.3 イントリンシック画像

イントリンシック画像 (*intrinsic images*) である照明画像  $L$  (*illumination images*) と反射画像  $R$  (*reflectance images*) を用いると、すべての視覚的な画像  $I$  は  $I = L \cdot R$  により表現できる [92]。このとき照明画像  $L$  は入射光分布を表し、反射画像  $R$  は物体の反射成分を表している。

短時間の入力画像系列からこれらイントリンシック画像を推定し、照明画像  $L$  を利用して入力画像の照度を正規化するものがある [93, 94]。照明画像にはシーンの照度成分がすべて含まれるため、入力画像を照明画像で割ることにより照度の正規化が実現できる。この手法では、照明変化に伴う反射成分の変化を考慮することにより、より正確に照明画像を推定する。また同時に、照明画像系列を固有空間で学習し、入力画像に対する照明画像を直接推定する方法も示されている。

### 3.4.4 その他

そのほかには、各光源の光量により背景を推定するものがある。新宮らは、光量を示すパラメータを各照明装置から取得し、予め測定しておいた各光源の光量と画像輝度との関係をもとに背景画像を推定している [95]。また、森田らは、入力画像中からランダムに選ばれた画素の組み合わせで、各光源からの光の到達度を推定し、光量と画像輝度との関係をもとに背景画像を推定する [96]。このとき、到達度は画像全体で一様であると仮定している。

また、高橋らは、水面など空間的な特性がある背景に対し、周波数スペクトルに関するパラメータ推定を行い、その特性を打ち消す白色化フィルタを入力画像に適用することで、特性から外れる前景物体領域を抽出する手法を提案している [97]。

## 4 検出対象のモデリング

検出対象となる物体が既知であれば、その知識を用いて検出処理を行うことができる。検出対象に関する知識は検出処理にとって非常に大きな助けになりうるが、一般環境では、検出対象の見え方のバリエーションを完全に網羅することは容易でない。一般的なパターン認識と共通するが、与えられた教師パターン(検出対象物体の実例)に含まれないものでも正しく検出する、汎化能力の高い検出アルゴリズムが望まれている。

表1に示す課題の中から、検出対象のモデリングに関するものを整理すると、

- シーン中を移動する物体は、任意の位置・姿勢・スケールをとる。
- 照明条件の変化や個体差などで検出対象の色・模様が変化する。
- 歩行する人間など、検出対象自身の形状が変化する場合がある。
- 検出対象が複数存在する場合は、それら相互の干渉(オクルージョンや影)が発生する。

のようになる。その一方で、前章に述べた背景のモデリングに基づく手法の多くでは、カメラの位置が変化しないことを前提としていたが、検出対象のモデリングを行う場合はその制約がないため、カメラが移動するような場合に適した手法であるといえる。

先述の課題を解決するために採られているアプローチは、以下の4つに大別できる。

- 対象の3次元形状モデルを保持しておき、入力画像と照合することで対象を検出するもの。
- 対象を2次元画像で観測したときに現れる特徴量に着目し、画像からの特徴抽出処理によって対象を検出するもの。この手法はさらに、空間的な特徴量に着目するものと時間的な特徴量に着目するものに分けることができる。
- 2次元画像上の見え方をそのまま対象モデルとして保持し、入力画像とのマッチングを行うもの。
- 先に述べた課題に適した識別器を採用して性能を向上させたもの。

本章では、以上の分類にそって研究事例を紹介していく。

## 4.1 形状モデル

検出対象の幾何形状があらかじめ分かっており、とりうる姿勢も限定できる場合には、幾何形状と姿勢から検出対象物体の見え方を推定し、観測画像と照合して対象物体を検出できる。1.1節で述べたように、このようなアプローチは古くから工業用ビジョンで広く用いられている。

道路上の移動車両は大きさや走行方向がある程度限定されているために、形状モデルを利用することができる。車両の場合は輪郭部のエッジをはっきりと観測できるので、ワイヤフレームモデルで形状を記述してマッチングをとれば車両を検出できる [98, 99]。移動車両に搭載されたカメラの場合の場合は、まず、路面上の特徴点からカメラ自身の位置を求め、検出対象車両の取りうる姿勢を推定する手法が提案されている [100]。

## 4.2 特徴抽出

形状モデルは事例の収集にコストがかかり、汎化の面からみても不利である。これに対して、検出対象がもつ本質的な(変動に対して不変な)特徴に着目し、観測画像でその特徴を観測できるか否かで対象の有無を判定するアプローチがとられている。ここでは本章冒頭で述べたような課題に対して頑健な特徴量を選択することが重要になる。

### 4.2.1 空間的特徴

画像から抽出できる空間的特徴として最も単純なもの、空間微分演算によって得られるエッジ特徴である。エッジは影や天候変化の影響を受けにくいので、車両の検出 [101, 102, 103, 104] に多く利用されている。人物の検出では、肌や衣服の色も有用な特徴となる。領域 (blob) 中の色を YUV 空間での正規分布で表現したもの [48] や、Kernel Density Estimation によって確率分布をノンパラメトリックに表現したものが提案されている [105]。エッジと色ヒストグラムと併用して競技場での選手の検出を行う手法も提案されている [106]。また、Schneiderman らは、線形位相フィルタを用いた Wavelet 変換の係数からヒストグラムを求める手法 [107] を提案している。

また、ガボール変換を利用して、検出対象の画像パターンの方向性を反映したモデルを構築する手法も提案されている。具体的な応用例として、車両検出 [108] や人物検出 [109] が報告されている。Rajagopalan らは高次統計量を利用して検出対象モデルの表現能力を向上させる手法を提案している [110]。

背景差分などによって得られたシルエットの形状を解析して対象検出を行う手法も提案されている。人物のシルエットを垂直軸に投影して脚の動きをキャンセルし、歩行人物を検出する手法が提案されている [111]。また、Hasegawa らはシルエットのモーメント、面積などを特徴量とすることで、車両検出だけでなく車種の判別を可能とする手法を提案している [112, 113]。

人物や車両以外を検出する例としては、画像中の煙領域の検出が試みられており、煙領域の持つフラクタル性を利用し、観測画像に対するフラクタル符号化の結果から煙領域を検出する手法 [114] が提案されている。

### 4.2.2 時間的特徴

2フレーム以上の画像があれば、その間の動き情報を抽出することができる。動き情報を得る手段としては、フレーム間差分とオプティカルフローが挙げられる。フレーム間差分からは各画素における動きの有無が分かり、オプティカルフローからは画像上のある点や領域がどこに移動したかを求めることができる。オプティカルフローの方が情報量は大きい、対応付けの曖昧さが残る場合があり信頼性は必ずしも高くない。

フレーム間差分は連続するフレームを単純に引き算するだけであるので、光環境の変化が激しい場合や、コントラストが低い場合でも比較的安定に動き情報を検出できる。この長所を生かして、トンネル入り口などでの車両検出に応用

したもの [115] がある。また、フレーム間差分の性能を大きく左右する検出閾値を過去の履歴に応じて変化させる手法 [116] が提案されている。

オプティカルフローを用いた例としては、白井らが、移動対象は等速直線運動をすると仮定して、複雑に運動する背景の中から移動対象を検出する手法を提案している [117, 118]。計算されるオプティカルフローの精度は必ずしも高くないため、位置  $x, y$ ・時間・移動方向の 4 次元空間への投票によって移動対象を検出している。また、映像中の火災や煙を検出するために、オプティカルフローの大きさ・方向・大きさのばらつき具合を利用する手法 [119, 120] も提案されている。

MPEG 符号化データのモーションベクトルも、オプティカルフローにほぼ類するものと考えることができる。Babu らは数フレームのモーションベクトルを蓄積し、EM アルゴリズムを利用してクラスタリングを行う手法を提案している [121]。

対象検出・追跡の結果得られた軌跡の中で、ノイズなどによる偽の軌跡を排除するアプローチも行われている。伊藤らは得られた軌跡を階層的に解析し、軌跡の分離・交差などの補正や、木々のざわめきや波による軌跡と侵入者による軌跡の識別を行った [122]。羽下らは、動きの強さの空間平均・時間平均と動きの一様性を評価する指標を定義し、対象による軌跡を検出する手法を提案している [123, 124]。

#### 4.2.3 時空間特徴

ここまで述べてきた空間的特徴と時間的特徴のどちらかのみを利用するのではなく、の双方を利用して精度を向上させようとするのは自然な発想である。

藤吉らが提案した手法 [125, 126] では、検出対象物体に含まれる各画素の時間変化をモデル化し、さらにその空間的隣接関係も用いて移動対象を検出している。前項で述べた MPEG のモーションベクトルを用いた検出でも、モーションの空間的連続性を用いるものが提案されている [127]。また、 $x, y, t$  の 3 次元ボリューム中のエッジのふるまいを解析して対象検出を行う手法が提案されている [128]。他には、動き情報を用いて検出した対象領域の中で、対象の色情報と合致するものを最終的な検出領域とする手法 [129] も提案されている。

また、Velocity Hough Transformation も空間・時間の双方の特徴を利用したものと位置づけることができる。Nash らによって当初提案された手法 [130] では、円形の輪郭を持った対象に限定し、形状と速度ベクトルを示すパラメータを投

票して対象を検出している。この手法を任意形状の物体に拡張したものも提案されている [131]。

#### 4.3 アピアランスモデル

幾何モデルや特徴抽出を用いず、観測画像の見え（アピアランス）をそのまま用いて対象モデルを構築するアプローチも考えられる。アピアランススペースの手法は顔や物体の認識に広く用いられているが、対象検出への応用を考えると一部分が他の物体に隠蔽されてしまった場合に非常に弱いという欠点をもっている。

これに対し、Kagesawa ら [132, 133] は、固有窓法を採用し、微小なウィンドウの見え情報とその相互の位置関係をモデル化している。局所パターンの見えと大局的な幾何形状を分けて考えることで、オクルージョンへの対処を可能としている。この手法は先述の幾何モデルとアピアランススペースの手法を組み合わせたものともでき、学習時に様々な車両の形状に対応したモデルを用意する必要があるという、形状モデルと同様の問題をばらんでる。この問題を解決するために、CG モデルを利用する試みも行われている [134]。

#### 4.4 識別器の選択・改良

ここまでは、検出に利用する特徴によって手法を分類してきたが、ここでは識別器を取り上げる。当然、ここまで述べてきた全ての手法は識別器を用いているが、その中でも識別器の選択・改良によって本章冒頭に述べた課題を解決しようとしている手法に焦点をあてる。

識別器の汎化能力を高めるため、SVM と Harr Wavelet の組み合わせを用いる手法 [135] が提案されている。この手法を部分的なオクルージョンに対応するために拡張し、体の各部位（頭、脚、右腕、左腕）を検出する SVM による識別器を用意し、その結果から最終的に人物か否かを判断する手法も提案されている [136]。オクルージョンへの対処としては、他に、リカレントニューラルネットワークを用いたものも提案されている [137]。

動き特徴を捉えた処理を行うために、移動する人物の足のパターンを TDNN (Time Delay Neural Network) を使って識別する手法が提案されている [138]。また、Viola らは、自らが提案した AdaBoost による顔検出手法 [17] を拡張して、見え方だけでなく動きにも基づいて歩行者を検出する手法を提案している [139]。

また最近傍識別器によって特定の色領域を検出する手法も提案されており [140]、検出対象が単一の色領域である場合には非常に有効であることが示されている。

## 5 複数のモデルを用いた検出

単一のモデルだけを用いた検出手法では性能に限界があるため、実際の応用では複数のモデルを組み合わせて検出を行うことが多い。複数のモデルの組み合わせ方は、

1. 背景と検出対象を示すモデルを単一の特徴空間内に構築し、観測データの識別を行うもの、
2. 様々な特徴に基づく背景・検出対象のモデリングを行い、それらを用いた検出の結果を統合するもの、

に大別することができる。1. は2.1節で論じた図2に相当し、2. は図4および図5に相当する。

1. を用いる場合は、単一の特徴空間内で背景と検出対象が識別可能な分布をする必要があり、そのようなケースはあまり多くないが、特徴抽出にかかるコストが少ないというメリットがある。
2. の場合は、様々な観点で背景・検出対象をとらえた特徴量を用いるので、処理コストは増大するが、より高精度な検出の実現が期待できる。以下、それぞれの手法の具体例を紹介する。

### 5.1 共通の特徴を用いた検出

背景と検出対象に共通した特徴としては、オプティカルフローや色などが考えられるが、物体検出の観点からは、(1) 実画像から安定して検出でき、(2) 特徴空間中での分布が背景と検出対象に分離可能、であることが必要になる。

上記の要件を満たすものとして、画像上のオプティカルフローが挙げられる。静止背景を移動するカメラで撮影したとき、画像上のオプティカルフローはエピソード拘束条件を満たす。これに対して、背景中に移動物体が存在する場合には、拘束条件を満たさないフローが得られる。一般にオプティカルフローの計算は不安定であるため、ある一定以上の大きさをもつ領域である、などの検出対象に対する制約条件を加えて検出処理を行うことが多い。

このようなアプローチは古くから行われており、Thompson はエピソード拘束に加えて、奥行きによってフローの見かけの大きさが変化することを利用して検出を行う手法を提案している [141]。同じような考え方による手法は多く提案されており、ロボットなどの移動観測系に用いたもの [142, 143] が提案されている。このような手法では、閾値決定の方法が実用上大きな問題となるが、太田らはモデルのよさを評価する基準を導入し、理論的に閾値を決定する手法を提案している [144, 145]。

フローから、エピソード拘束を明示的に用いずに対象を検出する手法として、運動視差の違いから検出対象を検出するものがある [146]。この手法は、走行車両のカメラで前方の障害物を

検出するもので、路面に平行・垂直な2つの仮想平面を考え、平面の3次元中での向きを固定したまま画像内で追跡し、路面に平行な仮想平面での輝度の変化しなければ路面、垂直な仮想平面での輝度の変化しなければ障害物とみなすものである。また、画面の輪郭部のフローからカメラワークを求めてそのフローを補償し、検出対象が画面内で相対的に静止している状況に対処したのものもある [147]。

色に着目した手法としては、背景画像と検出対象の色を与えておき、最近傍識別器で識別する方法 [148] が提案されている。さらにその拡張として、画素間の共起性を考慮するために  $xy-YUV$  の5次元空間を用いる手法 [149] も提案されている。

また、加藤らは画像上の輝度値とその微分値を特徴量とし、背景・移動物体・影の各状態における特徴量の確率分布を正規分布と一様分布で表現し、さらに各状態間の遷移をHMMで記述した [150, 151]。HMMで時系列的な変化を表現し、輝度の微分値を考慮することで画像の空間的な特徴も利用していることになる。

同じように、画像の時空間での特徴に着目したものとしては、何らかの方法でセグメンテーションした画像を初期入力として与え、時間方向、あるいは時空間双方でセグメント同士の類似度を評価する尺度を計算し、類似したセグメントの統合を行っていく手法が提案されている [152, 153]。これらの手法では、セグメンテーション結果のうちどこが検出対象に相当するかを知ることができないが、検出対象を特徴づける情報を別途付加することで、物体検出にも応用可能であると考えられる。

### 5.2 異なる特徴を用いた検出結果の統合

異なる特徴を用いた検出結果を統合する場合には、図4のような、前段の識別器の結果を後段の識別器に入力して検出を行うカスケード統合か、図5のような2つの識別器で独立に検出した結果を統合する並列統合の組み合わせで全体が構成される。

図4のようなカスケード接続の単純なものとしては、背景差分で検出された前景を処理対象領域として検出対象モデルによる検出を行うものがあり、非常に多くの例がある [48, 112, 113, 123, 124]。より複雑なものとしては、マルチレベルでの解析を行うことで性能向上を図っている研究事例がある。背景差分を例にとると、例えば、Toyamaらは時間方向の拘束条件に基づき画素単位で背景差分の結果を用いて空間方向の拘束条件による領域単位での解析で前景領域を抽出している [19]。さらに、フレーム間での解析により、突然の大域的な変化にも対応できるよう



になっている．同様の解析方法は Javed らの研究にも見られる [154]．この方法では，色ベースの背景差分で検出された前景画素候補をグルーピングし，勾配ベースの背景差分で領域としての検証を行っている．またフレームレベルで大域的な照明変化を検出している．藤吉らも画素単位での解析と領域単位での解析を併用している [125, 126]．一方向のデータの流れだけではなく，Harville らの研究では，領域レベルでの解析結果を画素レベルの背景差分にフィードバックするアプローチがとられている [155]．

また，異種の手法を組み合わせることで性能向上を図っているものもある．例えば，境田らは背景差分による変化領域抽出と watershed 法による領域分割を組み合わせ，前景領域を検出している [156]．Wang らは HMM による時間的な処理と MRF による空間的な処理を組み合わせている [157]．

一方，図 5 に示す並列統合の場合には，統合の際の判断基準の設定が問題となり，各識別器の性能や想定しているタスクに応じて判断基準が設定されている．

例えば，Darrel らは，距離画像，肌の色検出，顔検出を用いて人物の検出・追跡を行う手法を提案している [158] が，そこでは，もっとも精度のよい顔検出器の結果を優先させ，補助的な役割として距離画像や色検出を用いている．

Maki らの人物検出システム [159] では，奥行き，フローの向き，対象の動きをそれぞれ用いた 3 つの識別器を並列に接続しているが，対象の追跡を続けるモードと，新たな対象を検出するモードを切り替え，それぞれのモードで識別器の組み合わせを変化させている．

Chen らはイメージモザイクに基づく時間的なセグメンテーションと色情報に基づく空間的なセグメンテーションを組み合わせた物体検出手法を提案している [160]．この手法では，時間的なセグメンテーションで頻繁に発生する欠落を空間的なセグメンテーションで補償している．

## 6 システム構成技術

最後に，関連技術として，性能向上のための体系的なアプローチについて簡単に述べる．

### 6.1 全方位視覚への対応

広域監視，遠隔監視，自律移動ロボット，テレプレゼンス，バーチャルリアリティ等のシステムでは，カメラ周辺の広範囲にわたって物体を検出しなければならない．しかし単体の固定カメラでは視野に限界があるため，これには，専用の光学系（双曲面ミラー）を備えた全方位カメラ（Hyper Omni Vision） [161] や多方向に向けた複数台のカメラからなる入力系 [162, 163]，周囲

を巡回撮影するパン・チルトカメラが用いられている．それぞれに得失があり，例えば，全方位カメラは，一度の撮影でシームレスな全方位パノラマ画像が得られるという特長を持っており，その点では物体検出に有利であるが，物体に対する解像度が十分に確保できないという欠点がある．これに対し，マルチカメラは解像度を十分に確保できるが，装置構成が大掛かりになるうえ，カメラ間で入力情報をやりとりするためにカメラ間の性能誤差や視点の違いに対する処理を付加する必要がある．一方，パン・チルトカメラは，全方位を同時には撮影できないものの，巡回中心と視点を一致させておけば，比較的容易にパノラマ画像を生成できるという特長をもっており，それを利用した背景差分も提案されている [164, 165], [166]．また，巡回中に照明条件が変化した場合でもパノラマ画像が生成できる方法も報告されている [167]．ついでに言えば，移動するカメラから撮影された画像系列からパノラマ背景画像を生成する方法も示されている [168, 169]．

### 6.2 高速移動物体への対応

走行車両や飛行物体など高速移動物体の検出性能を向上させるために高速度カメラが用いられている例もある．ただし，高速度カメラにより撮影された画像は，露光時間が短いため，通常の光量では暗い画像になってしまうという問題がある．そのため，周辺画素の値を足すことによって，仮想的に露光面積を広げてセンサの感度を向上させるといった工夫が行われている [170]．また，高レートに伴う伝送や処理の負荷を軽減するため，独自のセンサも開発されている．例えば，浜本らは，現在の画素値とメモリ内に蓄えられた画素値とを比較し，その差が閾値以上の画素のみを有意画素としてアドレス情報とともに出力するセンサを開発している [171]．また，石川らは，光検出器（PD）と処理回路（PE）を画素ごとに直結したものを 1 チップに集積化することにより，高速リアルタイムビジョンシステムを実現している [172]．

### 6.3 照明不良環境への対応

夜間などの視界不良環境下での物体検出性能を向上させるために赤外カメラが利用されることも多い [173]．事実，車載への応用としては，運転者の視覚支援として赤外画像を表示する装置 [174] や，歩行者を検出してドライバーに知らせ注意を喚起する装置が実用化されている [175]．物体検出にとっては，照明条件を問わず同じアルゴリズムを適用することができるため有用である．

そのほか，トンネルの出入り口など，照明が急激に変化するような環境下では，広ダイナミッ

クレンジカメラが利用されている。広ダイナミックレンジカメラを使えば、通常のカメラで発生する黒つぶれや白飛びが軽減される。照明が制御できる場合には、検出された前景領域をもとに、前景の画像輝度がダイナミックレンジ内に入るように照明光量を動的に制御するといったフィードバック制御が行われることもある [95]。

## 7 まとめ

本稿では、物体検出手法の応用事例の歴史をたどってそのフレームワークを改めて定義したあと、モデル化の対象に着目し、背景のモデリング、検出対象のモデリングそれぞれについて最新の研究動向を紹介し、それらの組み合わせによって高精度な対象検出を実現した事例を紹介した。また、システム構成技術の観点から、様々な課題を解決するための入力系の設計についても触れた。

実環境に適用する対象検出手法では、背景や検出対象に起こりうる変化をすべて知ることができない中で、どのように高い性能を確保するかが大きな問題となる。これは、一般のパターン認識と共通する問題であり、物体検出の事例でも、パターン認識の他分野で得られた知見を応用する試みが行われている。

一方、物体検出に特有な観点としては、

- 映像を撮像するプロセスに介在する、光源・シーン中の物体・カメラの3者はそれぞれ実世界に存在しているものである。そこには物理法則に基づく制約が存在しており、それらの利用によって検出性能の向上が見込まれる。
- 処理対象となる映像は、時間方向・空間方向に大きな相関を持つ。単一フレーム、単一画素に限定した処理ではなく、相関関係を利用した処理による性能向上が望まれる。

などがあげられる。一般のパターン認識のアプローチに加え、このように物体検出に独自の制約・条件を利用した手法も数多く提案されている。強力なパターン認識アルゴリズムと、これら物体検出に特有な観点をうまく組み合わせることが、今後の研究においても重要になると考えられる。

なお、本稿の範囲では網羅し切れなかった研究もあると思われる。コメントなどがあれば筆者までご指摘いただければ幸いです。

## 謝辞

本稿に関する調査活動は21世紀COEプログラム「知識社会基盤構築のための情報学拠点形成」の支援を受けている。

## 参考文献

- [1] 岩井儀雄, 勞世竝, 山口 修, 平山高嗣: 画像処理による顔検出と顔認識, 情処研報 CVIM, Vol. 2005, No. 38, pp. 343-368 (2005).
- [2] 西田健次, 栗田多喜夫: 画像認識におけるカーネル学習法, 情処研報 CVIM, Vol. 2005, No. 38, pp. 333-342 (2005).
- [3] 石井健一郎, 上田修功, 前田英作, 村瀬 洋: わかりやすいパターン認識, オーム社 (1998).
- [4] Duda, R. O., Hart, P. E. and Stork, D. G.: *Pattern Classification*, John Wiley & Sons (2000).
- [5] 白井良明, 井上博允: ビジュアル・フィードバックを利用したロボットの組合せ作業, 電総研彙報, Vol. 35, No. 3, pp. 327-333 (1971).
- [6] 壺井芳昭, 井上驍: 視覚付ロボットによる物体の3次元位置方向測定の一方法, 通信学会論文誌, Vol. 60-D, No. 8, pp. 586-593 (1977).
- [7] 深田陽司, 土井 博, 永峰啓二, 稲荷隆彦: 山積みされたクラックシャフトの認識, 三菱電機技報, Vol. 59, No. 10, pp. 30-33 (1985).
- [8] 依田晴夫, 池田貞弘, 江尻正員: 回転形パターンマッチングによる形状選別の試み, 計測自動制御学会論文誌, Vol. 10, No. 3, p. 284 (1974).
- [9] 河村貞和, 中谷敏雄, 目瀬道弘, 烏野 武: 傷検出のパターン認識とそのディスプレイについて, 通信学会研究会資料, Vol. IE73, No. 3 (1973).
- [10] 原 靖彦, 中川 清: 局所的特徴の抽出と比較による回路パターン欠陥検出法, 電子通信学会論文誌, Vol. 66-D, No. 8, pp. 970-976 (1983).
- [11] Thorpe, C., Hebert, M., Kanade, T. and Shafer, S.: Vision and navigation for the Carnegie-Mellon Navlab, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 10, No. 3, pp. 362-373 (1988).
- [12] Kanade, T., Collins, R., Lipton, A., Burt, P. and Wixson, L.: Advances in cooperative multi-sensor video surveillance, *DARPA Image Understanding Workshop*, Vol. 1, pp. 3-24 (1998).
- [13] Chow, C. K. and Kaneko, T.: Boundary Detection of Radiographic Images by Threshold Method, *Frontiers of Pattern Recognition* (S., W., ed.), Academic Press, pp. 61-82 (1972).
- [14] Giger, M., Karssemeijer, N. and Armato, S.: Guest editorial computer-aided diagnosis in medical imaging, *IEEE Trans. Medical Imaging*, Vol. 20, No. 12, pp. 1205-1208 (2001).
- [15] 美濃導彦, 加藤俊一: 文書画像処理, 別冊 O plus E 「画像処理技術の最新動向」(高木幹夫, 鳥脇純一郎, 田村秀行, 編), 新技術コミュニケーションズ, pp. 158-171 (1988).
- [16] 栄藤 稔: コンピュータビジョンと動画像符号化, コンピュータビジョン技術評論と将来展望(松山隆司, 久野義徳, 井宮 淳, 編), 新技術コミュニケーションズ, pp. 230-243 (1998).
- [17] Viola, P. and Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features, *CVPR01*, pp. I:511-518 (2001).

- [18] Huang, C., Al, H., Wu, B. and Lao, S.: Boosting nested cascade detector for multi-view face detection, *Int. Conf. on Pattern Recognition ICPR2004*, Vol. 2, pp. 415–418 (2004).
- [19] Toyama, K., Krumm, J., Brumitt, B. and Meyers, B.: Wallflower: Principles and Practice of Background Maintenance, *ICCV99*, pp. 255–261 (1999).
- [20] Pless, R., Larson, J., Siebers, S. and Westover, B.: Evaluation of local models of dynamic backgrounds, *CVPR03*, Vol. II, pp. 73–78 (2003).
- [21] 長屋茂喜, 宮武孝文, 藤田武洋, 伊藤 渡, 上田博唯: 時間相関型背景判定法による移動物体検出, 信学論 D-II, Vol. J79-D-II, No. 4, pp. 568–576 (1996).
- [22] Matsuyama, T., Ohya, T. and Habe, H.: Background Subtraction for Non-Stationary Scenes, *ACCV00*, pp. 662–667 (2000).
- [23] 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, 信学論 D-II, Vol. J84-D-II, No. 10, pp. 2201–2211 (2001).
- [24] 佐藤雄隆, 金子俊一, 五十嵐悟: 周辺増分符号相関画像に基づくロバスト物体検出及び分離, 信学論 D-II, Vol. J84-D-II, No. 12, pp. 2585–2594 (2001).
- [25] Kaneko, S., Murase, I. and Igarashi, S.: Robust Image Registration by Increment Sign Correlation, *Pattern Recognition*, Vol. 35, No. 10, pp. 2223–2234 (2002).
- [26] 佐藤雄隆, 丹羽義典, 金子俊一, 五十嵐悟, 山本和彦: Radial Reach Filter によるロバスト物体検出, 技術報告, 信学技報 PRMU2001-41 (2001).
- [27] 佐藤雄隆, 坂上勝彦: Bi-polar Radial Reach Correlation によるロバスト背景差分, 技術報告, 信学技報 PRMU2004-224 (2004).
- [28] 伊藤誠也, 清原数井 誠人, 金子俊一: R R C を用いた組込み向け動体検知に関する検討, 画像センシングシンポジウム SSII05, pp. 161–165 (2005).
- [29] 伊藤竜之介, 岩井儀雄, 谷内田正彦: 空間周波数にもとづく移動物体と背景の分離, 技術報告, 信学技報 PRMU2002-112 (2002).
- [30] 阿部大輔, 中山塩原 守人, 佐々木繁, 菅野展行, 菅野 肇: 屋外監視における広域映像内の小領域抽出技術, 画像センシングシンポジウム SSII05, pp. 165–168 (2005).
- [31] 太田直哉: 不変量の利用に関する考察とそれに基づく照明不変な画像差分処理, 信学論 D-II, Vol. J82-D-II, No. 6, pp. 1009–1017 (1999).
- [32] Ohta, N.: A Statistical Approach to Background Subtraction for Surveillance Systems, *ICCV01*, Vol. II, pp. 481–486 (2001).
- [33] Kanade, T., Yoshida, A., Oda, K., Kano, H. and Tanaka, M.: A Stereo Machine for Video-rate Dense Depth Mapping and Its New Applications, *CVPR96*, pp. 196–202 (1996).
- [34] Ivanov, Y., Bobick, A. and Liu, J.: Fast Lighting Independent Background Subtraction, *IJCV*, Vol. 37, No. 2, pp. 199–207 (2000).
- [35] Darrell, T. J., Demirdjian, D., Checka, N. and Felzenszwalb, P. F.: Plan-View Trajectory Estimation with Dense Stereo Background Models, *ICCV01*, Vol. II, pp. 628–635 (2001).
- [36] 柴山憲文, 埴 圭二, 太田 勝: 踏切障害物検知用のステレオ画像装置の開発, 第 9 回画像センシングシンポジウム (SSII03) (2003).
- [37] Gordon, G., Darrell, T., Harville, M. and Woodfill, J.: Background estimation and removal based on range and color, *CVPR*, pp. 459–464 (1999).
- [38] Eveland, C., Konolige, K. and Bolles, R. C.: Background Modeling for Segmentation of Video-rate Stereo Sequences, *CVPR98*, pp. 266–271 (1998).
- [39] 飯尾 淳, 比屋根一雄: 環境の変化に対して安定した人物検出手法, 画像センシングシンポジウム SSII04, pp. 129–134 (2004).
- [40] Lo, B. and Velastin, S.: Automatic congestion detection system for underground platforms, *International Symposium on Intelligent Multimedia, Video and Speech Processing*, pp. 158–161 (2001).
- [41] 佐藤 敦, 土川 仁, 伴野 明, 石井健一郎: 歩行者計数のための照明変化にロバストな背景画像更新法, 技術報告 D-408, 信学春期全大 (1994).
- [42] 川端 敦, 谷藤真也, 諸岡泰男: 移動物体像の抽出技術, 情処論文誌, Vol. 28, No. 4, pp. 395–402 (1987).
- [43] 谷寄徹也, 上田浩次, 池谷和夫, 堀場勇夫: 背景画像更新処理を用いた路面湿潤状況の検出, 信学論 D-II, Vol. J80-D-II, No. 9, pp. 2270–2277 (1997).
- [44] Haritaoglu, I., Harwood, D. and Davis, L.: W4: Real-Time Surveillance of People and Their Activities, *PAMI*, Vol. 22, No. 8, pp. 809–830 (2000).
- [45] Haritaoglu, I., Harwood, D. and Davis, L.: W4s: A Real-Time System for Detecting and Tracking People in 2 1/2 D, *ECCV98*, pp. 877–892 (1998).
- [46] Haritaoglu, I., Harwood, D. and Davis, L.: A Fast Background Scene Modeling and Maintenance for Outdoor Surveillance, *ICPR00*, Vol. 4, pp. 179–183 (2000).
- [47] Yamazawa, K. and Yokoya, N.: Detecting moving objects from omnidirectional dynamic images based on adaptive background subtraction, *ICIP03*, Vol. III, pp. 953–956 (2003).
- [48] Wren, C., Azarbayejani, A., Darrell, T. and Pentland, A.: Pfister: Real-Time Tracking of the Human Body, *IEEE Trans. PAMI*, Vol. 19, No. 7, pp. 780–785 (1997).
- [49] Seki, M., Fujiwara, H. and Sumi, K.: A Robust Background Subtraction Method for Changing Background, *Workshop on Applications of Computer Vision 2000*, pp. 207–213 (2000).
- [50] 関真規人, 藤原秀人, 鷲見和彦: 背景変動に頑健な背景差分法, 画像の認識・理解シンポジウム MIRU2000, Vol. II, pp. 403–408 (2000).
- [51] 和田俊和, 松山隆司: 動的背景モデルを用いた移動領域の抽出, 情処全国大会, Vol. 2, pp. 141–142 (1994).

- [52] Grimson, W., Stauffer, C., Romano, R., and Lee, L.: Using Adaptive Tracking to Classify and Monitor Activities in a Site, *CVPR98*, pp. 22–31 (1998).
- [53] Stauffer, C. and Grimson, W.: Adaptive Background Mixture Models for Real-time Tracking, *CVPR99*, Vol. II, pp. 246–252 (1999).
- [54] Stauffer, C. and Grimson, W.: Learning patterns of activity using real-time tracking, *Trans. on PAMI*, Vol. 22, No. 8, pp. 747–757 (2000).
- [55] Harville, M., Gordon, G. and Woodfill, J.: Adaptive Video Background Modeling Using Color and Depth, *ICIP01*, Vol. III, pp. 90–93 (2001).
- [56] Ren, Y., Chua, C.-S. and Ho, Y.-K.: Statistical background modeling for non-stationary camera, *Pattern Recognition Letters*, Vol. 24, No. 1-3, pp. 183–196 (2003).
- [57] Hayman, E. and Eklundh, J.: Statistical background subtraction for a mobile observer, *ICCV03*, pp. 67–74 (2003).
- [58] 中井宏章：事後確率を用いた移動物体検出手法，技術報告，情処研報 CV90-1, (1994).
- [59] Nakai, H.: Non-Parameterized Bayes Decision Method for Moving Object Detection, *ACCV95*, pp. 447–451 (1995).
- [60] Elgammal, A., Harwood, D. and Davis, L. S.: Non-Parametric Model for Background Subtraction, *ECCV00*, Vol. II, pp. 751–767 (2000).
- [61] Han, B., Comaniciu, D. and Davis, L.: Sequential kernel density approximation through mode propagation: applications to background modeling, *ACCV04*, Vol. 2, pp. 818–823 (2004).
- [62] Comaniciu, D., Ramesh, V. and Meer, P.: The variable bandwidth mean shift and data-driven scale selection, *ICCV01*, Vol. 1, pp. 438–445 (2001).
- [63] Comaniciu, D.: An algorithm for data-driven band-width selection, *PAMI03*, Vol. 25, No. 2, pp. 281–288 (2003).
- [64] 波部 齊, 大矢 崇, 松山隆司: 動的環境における頑健な背景差分法, 画像の認識・理解シンポジウム MIRU98, Vol. I, pp. 467–472 (1998).
- [65] 高藤政雄, 北村忠明, 小林芳樹: 空間微分および差分処理を用いた車両抽出法, 信学論 D-II, Vol. J80-D-II, No. 11, pp. 2976–2985 (1997).
- [66] 高橋祐介, 亀井俊男, 溝口正典: 空間相関法による照明変動に頑強な物体検出法, 画像センシングシンポジウム SSII99, pp. 1–6 (1999).
- [67] 奥村晃弘, 岩井儀雄, 谷内田正彦: 屋外における移動物体の検出 - 照明変動と影への対策 -, 画像の認識・理解シンポジウム MIRU2000, Vol. II, pp. 307–312 (2000).
- [68] 伊藤竜之介, 岩井儀雄, 谷内田正彦: 照明変動と影の影響を抑えた移動物体の検出, 技術報告 CVIM-127-17, 情処研報 (2001).
- [69] 福井真二, 石川富康, 岩堀祐之, 伊藤英則: 輝度変化に対する頑健な背景差分法, 画像の認識・理解シンポジウム MIRU2004, Vol. II, pp. 305–310 (2004).
- [70] 天本直弘, 藤井明宏: 画像処理技術による障害物検出と移動物体追跡方法, 信学論 A, Vol. J81-A, No. 4, pp. 527–535 (1998).
- [71] Seki, M., Wada, T., Fujiwara, H. and Sumi, K.: Background subtraction based on cooccurrence of image variations, *CVPR03*, Vol. II, pp. 65–72 (2003).
- [72] 関真規人, 和田俊和, 藤原秀人, 鷺見和彦: 背景変化の共起性に基づく背景差分, 情処論文誌, Vol. 44, No. SIG5(CVIM6), pp. 54–63 (2003).
- [73] 島井博行, 栗田多喜夫, 梅山伸二, 田中 勝, 三島健稔: ロバスト統計に基づいた適応的な背景推定法, 信学論 D-II, Vol. J86-D-II, No. 6, pp. 796–806 (2003).
- [74] Karmann, K. and A., B.: Detection and Tracking of Moving Objects by Adaptive Background Extraction, *Scandinavian Conference on Image Analysis 1989*, pp. 1051–1058 (1989).
- [75] 影広達彦, 大田友一: 動画像からの背景画像の自動生成と適応的更新, 画像の認識・理解シンポジウム MIRU94, Vol. II, pp. 263–270 (1994).
- [76] Ridder, C., Munkelt, O. and Kirchner, H.: Adaptive background estimation and foreground detection using Kalman filtering, *Int. Conf. on Recent Advances in Mechatronics*, pp. 193–199 (1995).
- [77] Gao, D., Zhou, J. and Xin, L.: A Novel Algorithm of Adaptive Background Estimation, *ICIP01*, Vol. II, pp. 395–398 (2001).
- [78] Zhong, J. and Sclaroff, S.: Segmenting foreground objects from a dynamic textured background via a robust Kalman filter, *ICCV03*, pp. 44–50 (2003).
- [79] Hallinan, P.: A low-dimensional representation of human faces for arbitrary lighting conditions, *CVPR94*, pp. 995–999 (1994).
- [80] Basri, R. and Jacobs, D.: Lambertian reflectance and linear subspaces, *ICCV01*, pp. 383–390 (2001).
- [81] Shashua, A.: Geometry and Photometry in 3D Visual Recognition, PhD Thesis, Dept. Brain and Cognitive Science, MIT (1992).
- [82] Shashua, A.: On photometric issues in 3D visual recognition from a single 2D image, *IJCV*, Vol. 21, No. 1/2, pp. 99–122 (1997).
- [83] Belhumeur, P. and Kriegman, D.: What is the set of images of an object under all possible lighting conditions?, *CVPR96*, pp. 270–277 (1996).
- [84] 石井育規, 福井孝太郎, 向川康博, 尺長 健: 光学現象の分類に基づく画像の線形化, 情処論文誌, Vol. 44, No. SIG5(CVIM6), pp. 11–21 (2003).
- [85] Oliver, N., Rosario, B. and Pentland, A.: A Bayesian Computer Vision System for Modeling Human Interactions, *IEEE Trans. on PAMI*, Vol. 22, No. 8, pp. 831–843 (2000).

- [86] Weng, J., Zhang, Y. and Hwang, W.-S.: Candid Covariance-Free Incremental Principal Component Analysis, *Trans. on PAMI*, Vol. 25, No. 8, pp. 1034–1040 (2003).
- [87] Monnet, A., Mittal, A., Paragios, N. and Ramesh, V.: Background modeling and subtraction of dynamic scenes, *ICCV03*, pp. 1305–1312 (2003).
- [88] 杉山賢治, 古賀久志, 渡辺俊典, 横山貴紀: オンライン実時間 PCA を用いた動画からの変動背景の推定, 技術報告, 信学技報 PRMU2004-184 (2005).
- [89] Kahl, F., Hartley, R. and Hilsenstein, V.: Novelty Detection in Image Sequences with Dynamic Background, *Workshop on Statistical Methods in Video Processing*, pp. 117–128 (2004).
- [90] 土田 勝, 川西隆仁, 村瀬 洋, 高木 茂: 背景差分法による物体検出を目的とした逐次モンテカルロ法による背景推定, 信学論 D-II, Vol. J87-D-II, No. 5, pp. 1062–1070 (2004).
- [91] 川端 聡, 日浦慎作, 佐藤宏介: オクルージョン領域を含んだ動画からの繰り返し最適投影法による実時間動的背景推定法, 画像の認識・理解シンポジウム MIRU2004, Vol. II, pp. 30–36 (2004).
- [92] Barrow, H. and Tenenbaum, J.: Recovering intrinsic scene characteristics from images, *Computer Vision Systems* (Hanson, A. and Riseman, E., eds.), Academic Press, pp. 3–26 (1978).
- [93] Matsushita, Y., Nishino, K., Ikeuchi, K. and M., S.: Illumination Normalization with Time-dependent Intrinsic Images for Video Surveillance, *CVPR03*, Vol. 1, pp. 3–10 (2003).
- [94] 松下康之, 西野 恒, 池内克史, 坂内正夫: 時変イントリンシック画像とビデオサーベイランスへのその応用, 信学論 D-II, Vol. J86-D-II, No. 8, pp. 1186–1196 (2003).
- [95] 新宮 淳, 亀田能成, 角所 考, 美濃導彦: 背景画像合成による前景抽出とそれに基づく照明の動的制御, 技術報告 569, 信学技報 PRMU2001-218 (2001).
- [96] 森田順也, 岩井儀雄, 谷内田正彦: 室内における背景画像の推定と影の除去, 情処論文誌, Vol. 44, No. SIG9, pp. 105–114 (2003).
- [97] 高橋行行, 川島康孝, 松下泰雄: 不規則なパターンの背景を持つ実画像からの標的検出, 画像の認識・理解シンポジウム MIRU2004, Vol. II, pp. 13–18 (2004).
- [98] Sullivan, G., Worrall, A. and Ferryman, J.: Visual Object Recognition Using Deformable Models of Vehicles, *Context95*, pp. 75–86 (1995).
- [99] Ferryman, J., Worrall, A. D., Sullivan, G. D. and Baker, K. D.: A Generic Deformable Model for Vehicle Recognition, *Proc. 6th British Machine Vision Conference*, pp. 127–136 (1995).
- [100] Ferryman, J. M., Maybank, S. J. and Worrall, A. D.: Visual Surveillance for Moving Vehicles, *IJCV*, Vol. 37, No. 2, pp. 187–197 (2000).
- [101] 内村圭一, 松島宏典: オクルージョンを考慮した交通流計測, 電学論 C, Vol. 122-C, No. 12, pp. pp.2120–2127 (2002).
- [102] 谷口浩之, 久津間祐二, 浜本隆之: 車載用スマートイメージセンサにおける白線および前方車両 検出, 画像の認識・理解シンポジウム (MIRU2004), Vol. I, pp. 386–391 (2004).
- [103] 許 昭倫, 古賀拓也, 櫻井和之, 岡崎信一郎: 動画画像処理ボード IMAP-VISION による天候に口バストな先行車検出・追跡システム, 信学技報 PRMU, Vol. PRMU98-96, pp. 29–36 (1998).
- [104] Zhao, T. and Nevatia, R.: Car Detection in Low Resolution Aerial Images, *IVC*, Vol. 21, No. 8, pp. 693–703 (2003).
- [105] Elgammal, A., Duraiswami, R., Harwood, D. and Davis, L. S.: Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance, *Proceedings of IEEE*, Vol. 90, pp. 1151–1163 (2002).
- [106] 内海慎久, 浜田玲子, 井手一郎, 坂井修一, 田中英彦: 色頻度・局所エッジ成分を総合的に用いた物体検出手法とサッカー映像における選手検出への応用, 信学技報 PRMU, Vol. PRMU2001-241 (2002).
- [107] Schneiderman, H. and Kanade, T.: A Statistical Method for 3D Object Detection Applied to Faces and Cars, *CVPR00*, pp. I: 746–751 (2000).
- [108] 三谷重知, 呉 海元, 塩山忠義: ガボール変換による自動車検出, 信学論, Vol. J83-D-II, No. 12, pp. 2641–2651 (2000).
- [109] 堀田一弘, 本間康展, 栗田多喜夫: コントラストガボール特徴を用いた人検出法と特徴選択による汎化能力の改善, 画像の認識・理解シンポジウム (MIRU2002), Vol. II, pp. 303–310 (2002).
- [110] Rajagopalan, A., Burlina, P. and Chellappa, R.: Higher Order Statistical Learning for Vehicle Detection in Images, *ICCV99*, pp. 1204–1209 (1999).
- [111] 中上友宏, 松尾啓志: 人物モデルと体軸方向フィルタリングを用いた実時間人物検出・計数システム, 電学論 C, Vol. 122-C, No. 12, pp. 2011–2019 (2002).
- [112] Hasegawa, O. and Kanade, T.: Type classification, color estimation, and specific target detection of moving targets on public streets, *MVA*, Vol. 16, No. 2, pp. 116–121 (2005).
- [113] 長谷川修, 金出武雄: 一般道路映像中の移動物体の識別・色の推定と特定対象の検出, 情報処理学会論文誌, Vol. 44, No. 7, pp. 1795–1807 (2003).
- [114] 藤原伸幸, 寺田賢治: フラクタル符号化を用いた煙領域の検出, 第 10 回画像センシングシンポジウム (SSII04), pp. 151–156 (2004).
- [115] 風間 久, 武田信之, 小野口一則: マルチレート型時間差分法による移動車両の検出, 画像の認識・理解シンポジウム (MIRU2000), Vol. II, pp. 271–276 (2000).
- [116] 藤吉弘亘, 小村剛史, 矢入 (江口) 郁子, 香山健太郎, 吉水 宏: 歩行者 ITS のためのフレーム間差分による移動体検出法とその評価, 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, Vol. 45, No. SIG13(CVIM10), pp. 11–20 (2004).
- [117] 長井 敦, 久野義徳, 白井良明: 複雑変動背景下における移動物体の検出, 信学論, Vol. J80-D-II, No. 5, pp. 1086–1095 (1997).



- [118] Iketani, A., Nagai, A., Kuno, Y. and Shirai, Y.: Real-Time Surveillance System Detecting Persons in Complex Scenes, *RealTimeImg*, Vol. 7, No. 5, pp. 433–446 (2001).
- [119] 寺田賢治, 宮原宏幸, 新居康俊: オプティカルフローを用いた火災煙の検出, *電学論 D*, Vol. 124, No. 4, pp. 413–420 (2004).
- [120] 水上幸子, 寺田賢治, 星川京子: カラー動画をういた火災煙の自動検出, *信学技報 PRMU*, Vol. PRMU2001-111, pp. 15–20 (2001).
- [121] Babu, R., Ramakrishnan, K. and Srinivasan, S.: Video Object Segmentation: A Compressed Domain Approach, *CirSysVideo*, Vol. 14, No. 4, pp. 462–474 (2004).
- [122] 伊藤 渡, 上田博唯: 物体検出状態の階層的解析による複数物体追跡方式を用いた侵入物体検出, 画像の認識・理解シンポジウム (MIRU98), Vol. II, pp. 13–18 (1998).
- [123] 羽下哲司, 鷲見和彦, 八木康史: 変化領域内の動きの時空間特徴に着目した屋外情景における歩行者の検出, *信学論*, Vol. J87-D-II, No. 5, pp. 1104–1111 (2004).
- [124] Haga, T., Sumi, K. and Yagi, Y.: Human detection in outdoor scene using spatio-temporal motion analysis, *ICPR04*, pp. IV: 331–334 (2004).
- [125] 藤吉弘巨, 金出武雄: 複数物体の重なりを理解するレイヤー型検出法, 画像センシングシンポジウム SSI01, pp. 369–374 (2001).
- [126] Fujiyoshi, H. and Kanade, T.: Layered detection for multiple overlapping objects, *ICPR02*, Vol. 4, pp. 156–161 (2002).
- [127] 米山暁夫, 中島康之, 氏原清乃: MPEG 動画データからの移動物体の検出, *信学技報 PRMU*, Vol. PRMU96-101, pp. 63–70 (1996).
- [128] Sarkar, S., Majchrzak, D. and Korimilli, K.: Perceptual Organization Based Computational Model for Robust Segmentation of Moving Objects, *CVIU*, Vol. 86, No. 3, pp. 141–170 (2002).
- [129] Wanderley, J. and Fisher, M.: Spatial-Feature Parametric Clustering Applied to Motion-Based Segmentation in Camouflage, *CVIU*, Vol. 85, No. 2, pp. 144–157 (2002).
- [130] Nash, J., Carter, J. and Nixon, M.: Dynamic Feature-Extraction via the Velocity Hough Transform, *PRL*, Vol. 18, No. 10, pp. 1035–1047 (1997).
- [131] Grant, M., Nixon, M. and Lewis, P.: Extracting moving shapes by evidence gathering, *PR*, Vol. 35, No. 5, pp. 1099–1114 (2002).
- [132] Kagesawa, M., Ueno, S., Ikeuchi, K. and Kashiwagi, H.: Recognizing Vehicles in Infrared Images Using IMAP Parallel Vision Board, *IEEE Trans. ITS*, Vol. 2, No. 1, pp. 10–17 (2001).
- [133] 影沢政隆, 上野信一, 池内克史, 栢木 寛: 並列画像処理ボード IMAP-vision を利用した赤外画像による車両認識システム, *信学論*, Vol. J85-A, No. 5, pp. 597–605 (2002).
- [134] Yoshida, T., Mohottala, S., Kagesawa, M. and Ikeuchi, K.: Vehicle Classification System with Local-Feature Based Algorithm Using CG Model Images, *IEEE Trans. IS*, Vol. E85-D, No. 11, pp. 1745–1752 (2002).
- [135] Papageorgiou, C. P., Oren, M. and Poggio, T.: A General Framework for Object Detection, *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*, Washington, DC, USA, IEEE Computer Society, p. 555 (1998).
- [136] Mohan, A., Papageorgiou, C. and Poggio, T.: Example-Based Object Detection in Images by Components, *PAMI*, Vol. 23, No. 4, pp. 349–361 (2001).
- [137] 稲垣圭一郎, 佐藤省三, 梅崎太造: リカレントニューラルネットワークを用いた車両検出, *信学技報 PRMU*, Vol. PRMU2002-218, pp. 43–48 (2003).
- [138] Wohler, C., Aulanf, J., Portner, T. and Franke, U.: A Time Delay Neural Network Algorithm for Real-time Pedestrian Recognition, *IEEE International Conference on Intelligent Vehicles*, pp. 247–252 (1998).
- [139] Viola, P., Jones, M. and Snow, D.: Detecting Pedestrians Using Patterns of Motion and Appearance, *IJCV*, Vol. 63, No. 2, pp. 153–161 (2005).
- [140] 和田俊和: 最近傍識別器を用いた色ターゲット検出, 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, Vol. 44, No. SIG17(CVIM8), pp. 126–135 (2003).
- [141] Thompson, W. and Pong, T.: Detecting Moving Objects, *IJCV*, Vol. 4, No. 1, January 1990, pp. 39–58 (1990).
- [142] 岩田彩見, 加藤邦人, 山本和彦: Horizon View Camera による物体検出法, 画像の認識・理解シンポジウム (MIRU2002), Vol. II, pp. 385–390 (2002).
- [143] 古賀由紀夫, 片山明伯, 渡邊 睦, 山本美子, 渡部悠紀, 岩田 穆: 移動観測系における移動物体検出・呈示システムの検討, *情処研報*, Vol. 2002-CVIM-136, pp. 91–98 (2003).
- [144] Ohta, N. and Kanatani, K.: Moving object detection from optical flow without empirical thresholds, *IEICE Transactions on Information and Systems*, Vol. E81-D, No. 2, pp. 243–245 (1998).
- [145] 太田直哉, 金谷健一, 木村和広: しきい値を用いないオプティカルフローからの移動物体検出, *情処研報*, Vol. 1997-CVIM-105-8, pp. 55–60 (1997).
- [146] 岡田隆三, 谷口恭弘, 小野口一則: 仮想平面追跡法を用いた単眼車載カメラによる障害物検出, *信学技報 PRMU*, Vol. PRMU2001-89, pp. 29–36 (2001).
- [147] 鈴木賢一郎, 中嶋正臣, 坂野 鋭: 局所動き方向ヒストグラム特徴による映像データからの動物体の検出, *信学技報 PRMU*, Vol. PRMU2002-110, pp. 47–52 (2002).
- [148] 加藤文和, 和田俊和: 最近傍識別による背景差分と色検出の統合-事例に基づく情報統合-, *情処学論: コンピュータビジョンとイメージメディア*, Vol. 45, No. SIG 13(CVIM10), pp. pp.110–117 (2004).

- [149] 浮田宗伯, 加藤丈和: 背景差分と色検出の統合によるターゲット色の自動学習と背景変動に頑健な実時間対象検出, 画像の認識・理解シンポジウム (MIRU2004), Vol. II, pp. pp.24-29 (2004).
- [150] 加藤ジェーン, 渡邊豊英, 米田政明: HMM に基づく交通監視映像の背景・物体・影の分離手法, 情処学論, Vol. 42, No. 1, pp. 1-15 (2001).
- [151] Rittscher, J., Kato, J., Joga, S. and Blake, A.: A Probabilistic Background Model for Tracking, *ECCV00*, pp. II: 336-350 (2000).
- [152] Gelgon, M. and Bouthemy, P.: A region-level motion-based graph representation and labeling for tracking a spatial image partition, *PR*, Vol. 33, No. 4, pp. 725-740 (2000).
- [153] Moscheni, F., Bhattacharjee, S. and Kunt, M.: Spatiotemporal Segmentation Based on Region Merging, *PAMI*, Vol. 20, No. 9, pp. 897-915 (1998).
- [154] Javed, O., Shafique, K. and Shah, M.: A hierarchical approach to robust background subtraction using color and gradient information, *IEEE Workshop on Motion and Video Computing 2002*, pp. 22-27 (2002).
- [155] Harville, M.: A Framework for High-Level Feedback to Adaptive, Per-Pixel, Mixture-of-Gaussian Background Models, *ECCV02*, Vol. II, pp. 543-560 (2002).
- [156] 境田慎一, 苗村昌秀, 金次保明: 背景差分法と時空間 watershed による領域成長法を併用した動画像オブジェクトの抽出, 信学論 D-II, Vol. J84-D-II, No. 12, pp. 2541-2555 (2001).
- [157] Wang, D., Feng, T., Shum, H. and Ma, S.: A Novel Probability Model for Background Maintenance and Subtraction, *International Conference on Vision Interface*, pp. 109-117 (2002).
- [158] Darrell, T., Gordon, G., Harville, M. and Woodfill, J.: Integrated Person Tracking Using Stereo, Color, and Pattern Detection, *IJCV*, Vol. 37, No. 2, pp. 175-185 (2000).
- [159] Maki, A., Nordlund, P. and Eklundh, J.-O.: Attentional Scene Segmentation: Integrating Depth and Motion, *CVIU*, Vol. 78, No. 3, pp. 351-373 (2000).
- [160] Chen, L., Lai, Y., Su, C. and Liao, H.: Extraction of video object with complex motion, *Pattern Recognition Letters*, Vol. 25, No. 11, pp. 1285-1291 (2004).
- [161] 八木康史, 横矢直和: 全方位ビジョン: センサ開発と応用の最新動向, 情処論文誌, Vol. 42, No. SIG13(CVIM3) (2001).
- [162] Shimizu, S., Yamamoto, K., Wang, C., Sato, Y., Tanahashi, H. and Niwa, Y.: Moving object detection with mobile stereo omni-directional system (SOS) based on motion compensatory inter-frame depth subtraction, *ICPR04*, pp. III: 248-251 (2004).
- [163] 清水早苗, 山本和彦, 王 彩華, 佐藤雄隆, 棚橋秀樹, 丹羽義典: 全方向ステレオシステムを搭載した移動体による実環境からの動物体検出, 電学論 C, Vol. 124, No. 6, pp. 1288-1295 (2004).
- [164] Wada, T. and Matsuyama, T.: Appearance sphere: Background model for pan-tilt-zoom camera, *ICPR96*, Vol. A, pp. 718-722 (1996).
- [165] 和田俊和, 浮田宗伯, 松山隆司: 視点固定型パンチルトズームカメラとその応用, 信学論 D-II, Vol. J81-D-II, No. 6, pp. 1182-1193 (1998).
- [166] 波田尚哉, 三宅哲夫: アクティブビジョンシステムによる遮へいを伴う移動物体の追跡, 信学論 D-II, Vol. J84-D-II, No. 1, pp. 93-101 (2001).
- [167] 高田和豊, 佐藤 智, 登 一生: PTZ カメラを用いた物体検出のための画像の線形化による予測背景画像生成, 技術報告 CVIM-148-29, 情処研報 (2005).
- [168] 譲田賢治, 坪内貴之, 菅谷保之, 金谷健一: 移動ビデオカメラ画像からの運動物体の抽出, 技術報告 CVIM-243-6, 情処研報 (2004).
- [169] Sugaya, Y. and Kanatani, K.: Extracting Moving Objects from a Moving Camera Video Sequence, 第 10 回画像センシングシンポジウム (SSII04), pp. 279-284 (2004).
- [170] 岡田隆三: 高速自動追跡カメラシステム, 東芝レビュー, Vol. 59, No. 10, pp. 72-73 (2004).
- [171] 浜本隆之, 長尾祥一: 高機能イメージセンサを用いた高速画像処理, 画像ラボ, Vol. 15, No. 9, pp. 1-5 (2004).
- [172] 石川正俊: 超高速ビジョンの展望, 日本ロボット学会誌, Vol. 23, No. 3, pp. 274-277 (2005).
- [173] 青木正喜: 赤外線画像の人間検出への応用, 技術報告 CVIM-147-14, 情処研報 (2004).
- [174] 田中宏明: ITS 車載システムの現状と課題, 画像センシングシンポジウム SSII05, No. I-2 (2005).
- [175] 辻 孝之: ナイトビジョンによる歩行者検知, 画像センシングシンポジウム SSII05, No. I-4 (2005).