学位論文 博士（工学）

画像局所特徴量の性能改善と次元圧縮による人物検出とその応用

2013 年度

慶應義塾大学大学院 理工学研究科

片岡 裕雄

目次

1 序論

本章の概要

本章ではまず，コンピュータビジョン分野における人物検出の現状や問題点を挙げる．次に人物検出のための局所特徴量に関するサーベイを記載し，最新の局所特徴量についても言及することで本論文にて提案する技術の位置付けを明確にする．さらには研究目的や提案手法のアプローチについても述べる．

1.1 研究背景

近年では，スマートフォン，ウェブカメラや監視カメラなどの導入で簡易的に映像を取得できる装置が増え，我々の生活空間内においても映像を使用する機会が増加してきた．映像が増加するにつれてエンターテインメントや記録媒体としてだけでなく，映像を解析して人物の見守りや監視等を実行する技術は我々の生活空間において安心や安全を提供するための主要技術と変わりつつある．その要素技術として注目を集めるのがコンピュータビジョン技術である．コンピュータビジョンでは「人間の目の代替をコンピュータ上で実現する」という目的の基で創出された学問分野であり，その起源は 1960 年代にまで遡る．それ以来，画像センサ，3 次元画像処理，認識・識別技術や情報の提示技術など数多くの技術が生み出されてきた．そんな中，現在注目を集めているのが人物行動解析技術である．人物行動解析では，映像中に映り込んだ人物を対象として画像中での位置を特定する検出 (Detection)，見つけた人物位置を時系列で対応付ける追跡 (Tracking) など様々な要素を取得するに至っている．ここで，人物行動解析の実例を図 1.1 に示す．検出・追跡の他にも人物の顔検出，年齢や笑顔度など属性を取得する顔認識 (Face Recognition)，顔認識後，目領域の状態から人物の視線方向を抽出する視線推定 (Gaze Estimation)，人体のパーツ毎に検出・追跡処理をして姿勢を把握する姿勢推定 (Posture Estimation) や画像中の人物が「何をしているか」に着目してタグを出力する行動理解 (Activity Recognition) について示している．

コンピュータビジョン分野における人物行動解析研究においては，数多くの技術が提案されてきた [1]．サーベイ論文も数多く発表されており，人物の追跡 [2]，姿勢推定 [3] や顔認識 [4] が

その代表的な例である．その中でも顔認識技術は 2001 年に Viola&Jones が Haar-like 特徴量とカスケード型識別器を適用し高速かつ高精度に顔を認識する技術を提案してから数年でデジタルカメラに顔検出技術が搭載されるなど一挙に盛んになった [5]．Haar-like 特徴量とカスケード型識別器による顔検出技術は顔の年齢 [6] や性別の推定 [7] や笑顔度推定 [8] など，顔認識の研究が急激に進むためのきっかけとなった．人物行動解析技術の適用範囲は非常に広く，顔認識技術以外でも生活空間内の見守り・映像監視・マーケティング分析・スポーツ映像解析・交通分野における歩行者予防安全等，人が存在するほとんどの空間において応用できる可能性を持っている．以下に，実環境で利用されている研究の一例を示す．

見守り/映像監視 [9]：見守りや映像監視では画像中で人物の位置を特定する「検出」やその時系列対応である「追跡」処理にフォーカスしている．現在では多数のカメラが街中や店舗に備えられているため，人手による見守り・監視が困難になっているため，自動で人物位置を特定する意義は大きいと言える．常に人物位置をモニタリングすることにより，安心や安全を提供することにつながる．現在までの成果として，人物画像を大量に準備した統計的学習による人物検出や統計モデルを適用した人物追跡の発展により，人物の位置を高精度に捉えることに成功している．位置の特定だけでなく，身体の状態を把握する姿勢推定や詳細な行動を “standing”や “walking”などタグとして出力する行動理解の研究も取り入れられており，今後も大きな成果が出てくる分野であると期待できる．

スポーツ映像解析 [10]：戦術分析や任意の方向からプレーを可視化する技術である 3 次元の自由視点映像のための技術が提案されてきた．スポーツ映像解析において一番重要なのが選手やボールの検出や追跡である．位置情報は選手の移動距離や選手間の相対関係，ボールにより戦況がいかに変化するかといった情報を抽出するために使用される．選手同士の重なり発生時に追跡が困難となる点が課題に挙げられる．

歩行者予防安全 [11]：歩行者予防安全技術とは，自動車のセーフティシステムの一種であり，自動車に取り付けたカメラから前方を撮影し，歩行者を検出して必要に応じて警報やブレーキを制御する技術である．予防安全の主な機能としては (i) 歩行者検出 (ii) 衝突判定 (iii)(衝突しそうな場合) ブレーキ制御という流れでドライバーへのアシストを行う．そのため，車載カメラから車外の歩行者を検出する精度が非常に重要である．現状ではステレオカメラを用いて距離画像を得られれば領域制限や検出モデルがテクスチャに影響しないなど利点が多数あり実利用化に

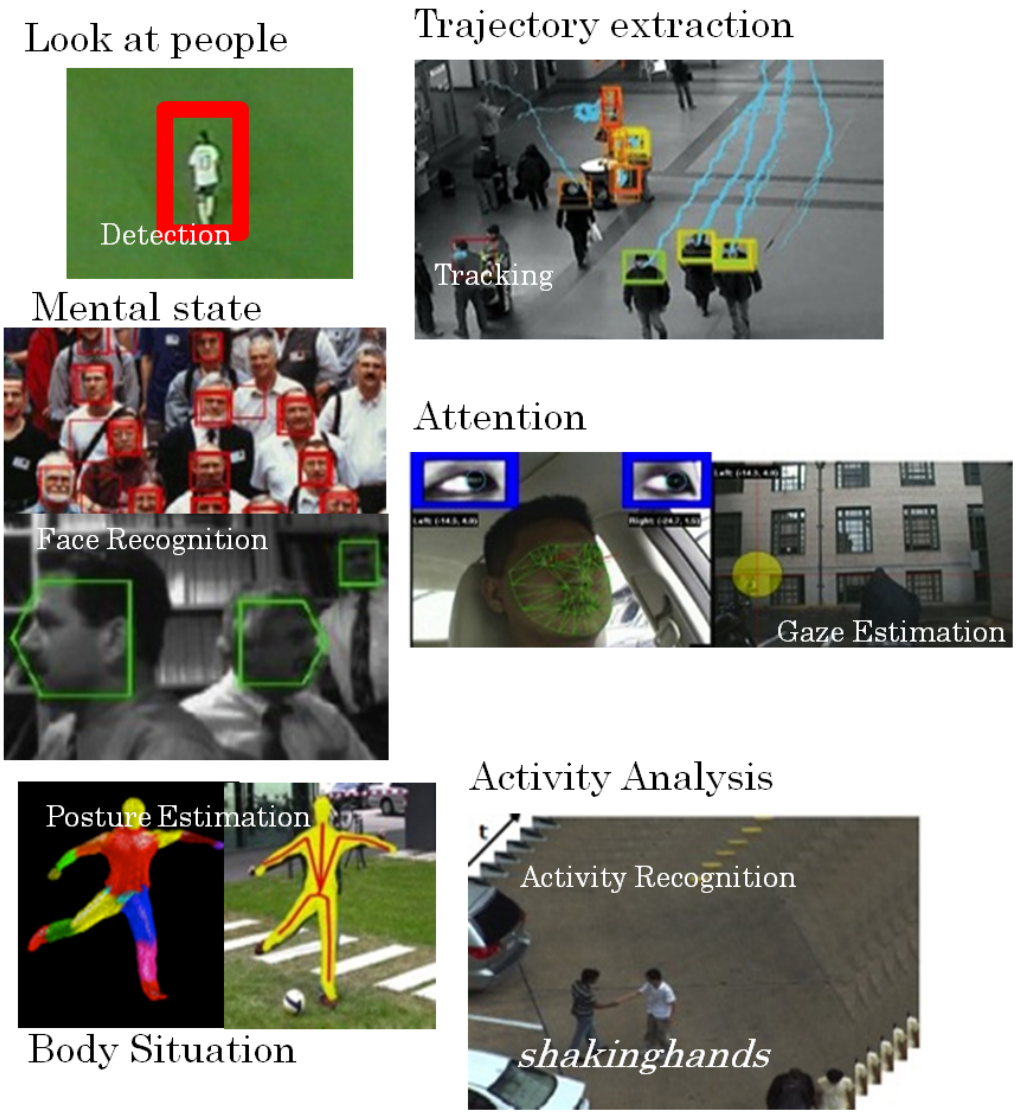

図 1.1　人物行動解析の例 [1]：検出，追跡，顔認識，視線推定，姿勢推定，行動理解

至っている．しかし，今後もさらなる安全のために検出技術の向上が望まれるだけでなく，より早く歩行者を検出する技術が必要である．

人物の検出を始めとした人物行動解析技術では，対象となる物体を認識するためにコンピュータ上にモデルを生成する必要がある．コンピュータビジョン分野では以前からコンピュータにモデルを覚えさせる研究が行われており，形状の組み合わせやテンプレート画像を準備する研究がその一例として提案された．形状の組み合わせでは円形や矩形など簡易的な形状の大小や傾きを変えた組み合わせにより物体の形状を表現するモデルである．テンプレート画像は実際の画像か

ら物体を切り抜いて時系列で変化する物体の追跡や違う場面に存在する物体を検出する手法である．しかし，何れの手法においても物体の表現能力に乏しく，モデルのバリエーションを持たせられないため「照明変動」「スケール変化」「背景や物体による遮蔽」等困難な場面においては著しく精度が低くなってしまう．この問題を解決するために採用されたのが多数のサンプルから分散を考慮してモデルを学習する「機械学習」である．機械学習では学習画像に照明や姿勢変化，複雑背景等，問題に合わせてモデリングの手法を変えることで精度を高めることができる．現在，コンピュータビジョンにおいては機械学習を適用し，統計的なモデルを保有した識別器を構成する方法が一般的である．

機械学習とは人間の学習機能をコンピュータ上で実現する仕組みのことであり機械にモデルを覚えさせ識別器を構成する「学習フェーズ」と，学習フェーズで生成した識別器を用いて対象物体を認識する「検出フェーズ」に分けられる．これは事前に与えられたデータを基にモデルを作るため教師あり学習とも言われている．

図 1.2 に，学習フェーズと検出フェーズに分けた，人物検出の流れを示す．学習フェーズでは，用意した学習画像から局所特徴量を取得して機械学習を行う．局所特徴量は画像の局所的な矩形領域から取り出すベクトルのことであり，動作ベースの特徴量やエッジベースの特徴量など対象物体から特徴を取り出す．ここで，学習画像は対象となる物体が映っている正解画像と対象物以外の背景が映っている非正解画像を用意する．機械学習においては一般的に数千オーダの学習画像枚数が必要である．また，正解画像よりも非正解画像の特徴空間の方が圧倒的に大きいため，通常は正解画像よりも非正解画像の枚数を多く入力する．学習には後述の Real AdaBoost [12] を適用しており，正解画像と非正解画像のラベルが付いている特徴量から識別器を構成する．検出フェーズでは特徴取得ウインドウを画像の左上から右下に走査するラスタスキャンにより画像内を探索する．

ここで，人物検出など人物行動解析のためには対象となる物体を詳細に捉える局所特徴量の適用が非常に重要である．物体の特徴を効果的に捉える局所特徴量の考案は，人物検出だけでなく追跡中の尤度評価や行動理解の精度向上のために適用可能である．学習やモデリングの方法を変更させれば良いので，特徴を捉えることは基礎研究となり得る．人物検出で問題とされるのは “Positive(正解)” と “Negative(非正解)” を分類することであり，2 値分類問題として知られる．正解画像と非正解画像から識別器を生成し，画像から取得した特徴量から識別器により出力値を

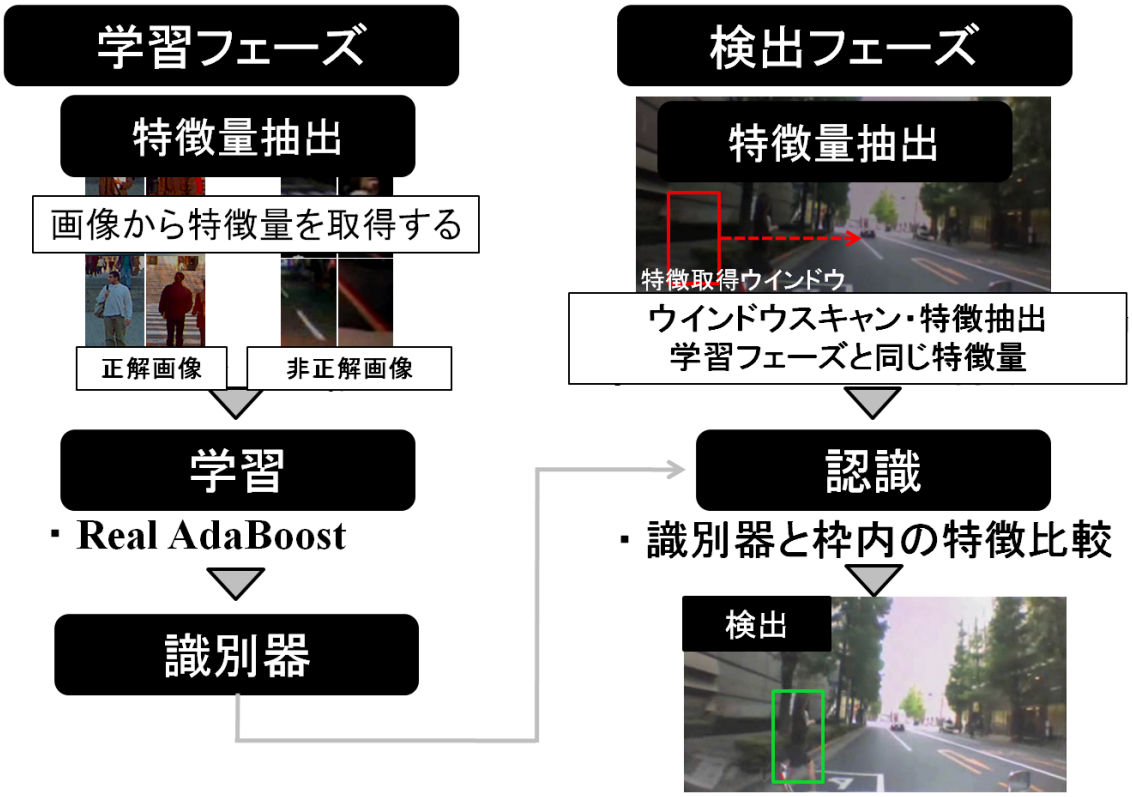

図 1.2　機械学習による人物検出の流れ：学習フェーズでは対象物体を含む学習画像から識別器を生成し，検出フェーズでは識別器により対象となる物体を検出する．機械学習にとって非常に重要な課題として，物体の特徴を詳細に記述する局所特徴量の考案が挙げられる．

得る．出力値の評価により正解か非正解かを分離する．追跡中の尤度評価では，あらかじめ対象となる物体をモデリングし，尤度の観測位置上においてモデルとの比較を行い尤度を得る．尤度マップを作成し，重み付き平均や最大尤度位置に重心を設定して毎フレームでの物体位置を特定することで追跡を実現している．行動理解では複数クラス分類問題として，抽出した矩形領域から特徴量を抽出し，人物の行動タグを出力する．

局所特徴量を用いた人物行動解析においては，複数の課題が存在する．主な課題を以下に示す．

- 環境の変化：複雑背景，背景の変化，照明変動，天候
- 人物のバリエーション：服装，スケール変化，姿勢の変化
- カメラとの関係：撮影方向による見えの角度変化，カメラパラメータと撮像系の違い，カメラとの相対的な位置

- 遮蔽：人物-人物の遮蔽，人物-物体の遮蔽，人物-背景の遮蔽
- 動作：複雑な動きへの対応，素早い動作により生起する画像のぼけ

全ての課題を同時に解決することは非常に困難であるため，問題に合わせて変更することが一般的である．いずれにしても人物の形状をより詳細に記述する特徴量が必要とされている．これらの問題を解決するために，特徴点の周囲で特徴量を記述するキーポイントベースの特徴量や時系列動作の差分から特徴を記述する動作ベースの特徴量が提案されてきた．しかし，人物全体の特徴を詳細に記述できる特徴量としては，形状ベースの特徴量が挙げられる．キーポイント，動作，形状の主に 3 種類の局所特徴量についての関連研究を以下に記載する．

1.2 局所特徴量の関連研究

コンピュータビジョン分野における局所特徴量において，いくつかのサーベイ論文が存在する．Gandhi らは 2006 年 [13]，Geronimo らは 2010 年 [14] と 2014 年 [11]，Dollar らは 2012 年 [15] にそれぞれ人物検出を対象としたサーベイ論文を発表している．何れのサーベイにおいても局所特徴量に関する調査が中心となっており，コンピュータビジョン技術で取り扱う重要技術の一つとして数えられている．ここでは関連する局所特徴量の研究を，「キーポイント特徴量」「動作特徴量」「形状特徴量」に分けて以下に挙げるが，最重要とされる形状特徴量に関しては特に密に述べることとする．

1.2.1 人物検出のための局所特徴量

キーポイント特徴量：画像の回転や拡大・縮小にロバストな特徴量としては SIFT (Scale-Invariant Feature Transform) が挙げられる [16]．SIFT 特徴量は特徴点 (キーポイント) の検出，特徴量の記述の 2 段階からなる． DoG(Difference of Gaussian) 画像を用いてスケールの変化に対応，さらに画像中の特徴点周りのパターンの勾配であるオリエンテーション計算，さらには特徴量取得を行う．DoG の応答値判断によるスケールの変化やオリエンテーション向きからの回転を考慮した特徴取得方法として広く使われている．SIFT では 4×4 にブロック分割した領域から 8 方向のエッジ方向を捉えるので，128 次元の特徴ベクトルを抽出可能である．現在までにも SIFT は様々な形で改良されており，SIFT の特徴次元を PCA(主成分分析：Principal

Component Analysis) により圧縮して低次元での特徴表現や頑健性を向上させた PCA-SIFT [17] や Random Forests [18] を用いてキーポイント検出を高速かつ高精度化した手法 [19]，キーポイントの周辺領域を対数極座標に変換し半径方向と角度方向に分割したグリッド領域内の記述により頑健なマッチングを実現した GLOH (Gradient Location and Orientation Histogram) [20] も提案されている．その中でも，Speeded-Up Robust Features (SURF) は高速化の手法であるインテグラルイメージの利用により SIFT よりも高速に精度も落とすことなくマッチングを可能にし，広く使われるに至っている [21]．インテグラルイメージは原点から座標位置 (x, y) までの値の総和を記録した画像であり，矩形領域の角に位置する画素の計算だけで画素値の総和が計算できるという意味で計算コストを削減可能としている．Bay らの実験によると SURF は SIFT の 10 倍近くの高速化を実現している．

キーポイントベースの手法をベクトル化して特徴量として扱うために，Bag-of-words を用いた手法が用いられる [22]．Bag-of-words はキーポイント周辺の領域から取得した特徴量を量子化してひとつのワード (visual word) として扱う技術である．画像特徴量から visual word にクラスタリングする際には k-means クラスタリングが適用されている [23]．k-means クラスタリングは始めにランダムで割り振られたクラスから中心移動とクラスの評価によりクラスタを分類する手法である．画像から取得したキーポイント全てに対して周辺から特徴量を抽出し visual word の集合をベクトル化することでクラス分類する．特徴ベクトルそのものを識別に用いるのではなく，クラスタリングして得られる visual word の分散により識別に用いることで高精度な識別を実現している．また，Bag-of-words はキーポイントベースの特徴量をベクトル化するために必要な技術である．また， k-means のクラスタ数がそのまま次元数になるが，クラスタ数が多ければ多いほど精度が上がると言われている．これは，その分類する物体のクラスに特有の visual word に分割することができ，他の物体のクラスと混同することが少なくなるからである．一般的にはクラスタ数は数千から数万オーダに設定する．

動作特徴量：画像中の動作を取得する手法としては，Lucas-Kanade 法 (LK 法) が広く知られている [24]．LK 法はオプティカルフローの一種であり，物体のコーナーを前後フレーム間で追跡する特徴点追跡手法である．従来，オプティカルフローが特徴点周辺のブロックをテンプレートとしてマッチングを繰り返すブロックマッチングであるのに対し，LK 法ではフレームでのコーナー点の位置に対して類似性を評価し，その微分変化量を考慮しながら特徴点の動きを計

Author(s)	Type	Approach
Lowe [16]	キ	回転・スケール変化に頑健
Dalal *et al.* [25]	動	前後画像の動作から境界抽出
Dalal *et al.* [30]	形	方向量子化ヒストグラムにエッジ強度累積
Viola *et al.* [5]	形	矩形領域の明暗差組み合わせ
Ojala *et al.* [31]	形	輝度差比較によるバイナリコード化
Levi *et al.* [34]	形	エッジ方向・強度成分をヒストグラム化
Wu *et al.* [35]	形	登録のエッジ繋がりパターンのカウント
Mita *et al.* [37]	形	Haar-like 特徴の共起関係を表現
Mitsui *et al.* [38]	形	識別器の出力結果から HOG の共起表現
Watanabe *et al.* [43]	形	2 画素間の共起ペアのカウント

表 1.1　コンピュータビジョン分野で適用される主な局所特徴量：著者，局所特徴量のタイプ分け (キ：キーポイント特徴量，動：動作特徴量，形：形状特徴量)，アプローチ

算していくことで追跡対象の位置を把握していく．マッチングを繰り返して誤差が最小になるようにフローの移動を行うことにより，高速な特徴点の探索が可能となる．Dalal らは Motion Boundary Histograms を用いて，密なオプティカルフロー推定画像から対象物体と背景の境界を求めて識別に有効な特徴量を抽出した [25]．映像中の前後フレーム $(framet, andframet+1)$ のオプティカルフローを求め，x, y 方向成分に独立な画像特徴として取得する．オプティカルフローの位置を画像中の位置，オプティカルフローの強度を輝度として x, y 方向成分の画像に記述する．フローを 2 つの特徴成分 x, y 方向に分割することで表現能力を高めており，それぞれの画像においてエッジの方向と強度を特徴ベクトル化することで密なフローから，形状特徴も同時に取得した効果が得られる．

形状特徴量：局所特徴量としては，形状特徴量が非常に多く提案されている．ここでは，形状特徴量をさらに 4 種類に分類して関連研究を示す．ここで，4 種類とは (i) テンプレート特徴 (ii)Holistic 特徴 (iii) 共起特徴 (iv) パーツ特徴である．

(i) テンプレート特徴：テンプレート特徴は形状を取得する特徴のうち，統計的学習を行っていない手法を示す．Gavrila は人物のシルエットを方向や姿勢などバリエーションを階層的に表現してマッチングする手法を提案した [26]．予めシルエット画像からエッジ検出，二値化画像としておき，人物の見えのバリエーションにフィッティングするために階層構造にしている．シル

図 1.3　HOG 特徴量の記述 (Dalal *et al.*[30] より)：(左) 原画像と各セル 8 方向に量子化したエッジ方向と各成分におけるエッジ強度 (中) 正解画像から取得した統計的形状特徴 (右) 非正解画像から取得した統計的形状特徴

エットが非常に類似している場合の比較が非常に困難な課題となっていたが，エッジの距離値を記録した画像でのマッチング手法 Chamfer distance transform [27] の適用により精度が向上した．Broggi らもテンプレートを用いて人物をマッチングしているが，歩行により見えの変化が激しい下半身よりも姿勢の変動が小さい上半身に着目して識別している [28]．Bertozzi らは赤外線カメラを用いており，二値画像の人物マスクを作成してテンプレートマッチングを行っている [29]．人物の二値画像マスクはカメラの角度や姿勢などによる幾何学的な変化も考慮することで単純なマスクからは精度を高めている．

(ii)Holistic 特徴：Holistic 特徴は「局所領域全体から取得する」特徴量のうち，共起性を考慮していない特徴と定義される．Viola らはシンプルな矩形領域を複数組み合わせて輝度差を算出し，矩形の明暗差の組合せを特徴とする Haar-like 特徴量を提案した [5]．識別に有効な特徴の組合せやスケールは AdaBoost により選択し，Cascade 型の識別器を用いて物体検出に応用した．しかし，明暗のパターンの組み合わせによる特徴表現である Haar-like は個人差が少なく，形状の変動がほとんど無い顔検出において用いられているが，人物検出には向かないとされている．そこで，代表的な人物検出特徴量として，Histograms of Oriented Gradients (HOG) が挙げられる [30]． HOG 特徴量はエッジの方向を量子化したヒストグラムに，エッジ強度を累積して物体の大まかな形状を表現する特徴量である (図 1.3)．画像の矩形領域内をピクセルの集合による矩形領域であるセルと，さらに複数のセルから構成される矩形のブロックに分割し，オーバーラップさせながら特徴を抽出する．矩形領域内ではブロック内の方向ヒストグラムに

エッジ強度を累積する．オリジナルの HOG において，ブロックは 3 × 3 のセルで構成される．エッジ方向ヒストグラムは 9 方向に量子化されるため，1 セルから取得される特徴次元数は 9 次元となる．1 ブロックには 9 セルあるため，ブロックあたりの特徴次元数は 81 次元となる．特徴量の取得には，このブロックを移動させ，セルをオーバーラップさせながら特徴を取得する．画像内の明るさの変動に対応するため，特徴取得後にヒストグラムを正規化する．正規化は矩形領域内のブロックごとに行う．Ojala らは局所的な輝度の比較によりバイナリパターンを算出する特徴である Local Binary Patterns (LBP) を提案した [31]．LBP 特徴量は 3 × 3 のウインドウ中において，中心画素とその周囲 8 画素の輝度の大小 (0 か 1) を比較することにより 8 つの 2 進数を得る．取得した 8 つの 2 進数を並べ，10 進数化した数値をそのウインドウの数値として特徴量を構成する．LBP 特徴量では画像全体が明るいか暗いかに関係なく，物体においてはバイナリのパターンの変動が少ないことに着目して特徴を構成するので，画像の照明変動の影響を受けにくい．数値の大小比較と 2 進数から 10 進数への変換によって特徴量を計算するため，高速な特徴取得が可能であり，リアルタイムでの計算も実現しやすい．実装の簡便性や高速な処理が可能なことから，Multi-scale LBP [32] や Extended LBP [33] など応用例も幅広く提案されている．LBP においても Haar-like と同じように，画素の明暗差を用いた符号化を適用して居るので歩行やその他の姿勢変動による特徴パターンの変動が大きく，精度の低下が見られてしまう．画素のエッジのつながりを見る特徴量としては Edge Orientation Histograms (EOH) [34] や Edgelet [35] が挙げられる．EOH 特徴量は，1 つの局所領域内におけるエッジ勾配の関係に着目した特徴量である．ソーベルフィルタによりエッジ画像を生成し，エッジ強度とエッジ勾配を算出する．EOH 特徴量では，異なるエッジ方向の強度の比により特徴量を記述する．Edgelet 特徴量はエッジの分布ではなく，エッジの部分的な繋がりによる特徴記述である．あらかじめ定義された形状パターンと入力画像の局所領域内でのエッジ方向の差異をベースに特徴量を算出している．定義された特徴としては，直線，円弧，またそれらの対称となるパターン組み合わせである．

(iii) 共起特徴：共起特徴は「異なる複数画素のエッジ共起」を表現する特徴量である．識別器の出力値を用いた共起表現や特徴取得時に同時に生起する特徴量等が存在する．画素またはエッジの共起性により物体を検出する手法としては，Shapelet 特徴量 [36]，Joint Haar-like [37]，Joint HOG [38] が挙げられる．Shapelet 特徴量は複数画素のエッジ情報の組合せにより正解か

どうかを判断する．局所領域内で対称物体に共起するエッジを記述可能である．Joint Haar-like は複数の Haar-like 特徴量の共起関係を表現する．取得した Haar-like 特徴が検出対象か否かで符号化の組合せを表現する．同時にエッジが存在していなければ検出することがないため，過検出を減らすことに成功している．Joint HOG は局所領域間や局所領域内の共起性を考慮して物体検出する特徴量である．これは，複数の局所領域内の HOG 特徴量が物体検出に有効な特徴かどうかを符号化により表現しているためである．Joint HOG は HOG より高い検出性能を示すことが報告されている．さらに，山内らは共起確率特徴量 (Cooccurrence Probability Feature: CPF) を提案して，形状と動作など，種類の異なる特徴量の組合せに成功した [39][40]．CPF では Real AdaBoost [12] を用いて，識別器から得られる出力値を基にして統合するアプローチである．さらに後藤らは CPF を適用して HOG 特徴量と物体の色の類似性を捉える特徴量である Color-Similarity [41] を組み合わせてさらに高精度な特徴量を提案している [42]．最近では Co-occurrence Histograms of Oriented Gradients (CoHOG) が考案され，監視カメラや車載ステレオ映像に適用されている [43]．CoHOG は 2 つの異なる位置でのエッジ勾配のペアの出現頻度を表現する．CoHOG はエッジ勾配の共起性により特徴を記述するため，HOG で問題視された過検出を低減することができ，高い性能を持つことが報告されている．実際の車載ステレオ映像から処理領域を絞り込み，CoHOG を適用して人物を検出する研究 [44] や，CoHOG をカスケード型識別器により対象物体を探索して [45] 高精度な人物検出を実現している．

(iv) パーツ特徴：パーツ特徴は「人体の部位領域間のつながりを表現するモデル」により識別することと定義する．手法としてはそれほど多くないが，人体の情報を取り込んでいるので非常に高い性能を誇ることで知られる．たとえば Implicit Shape Model(ISM) は特徴点を中心にして，その周囲の領域 (各部位) を統合して物体を検出する手法である [46]．ISM では特徴点の周囲の小領域をベクトル量子化し，あらかじめ登録したコードブックを基にして物体の重心位置へ投票する．投票点をクラスタリングすることで最終的な検出位置と設定する．Andriluka らの Pictorial Structure も強力なパーツモデルである [47]．Pictorial Structure ではパーツ毎に識別器を用意，部位毎に尤度を計算して姿勢推定ベースの人物検出を実現した．その中でも現在最も精度が高いと言われているのが Felzenszwalb らの Deformable Part Model (DPM) である [48]．DPM は物体の全体と部位毎の組み合わせをモデリングした手法であり，全体とパーツの形状およびパーツの位置ずれや見えの変化による変形を評価可能である．それぞれの形状を

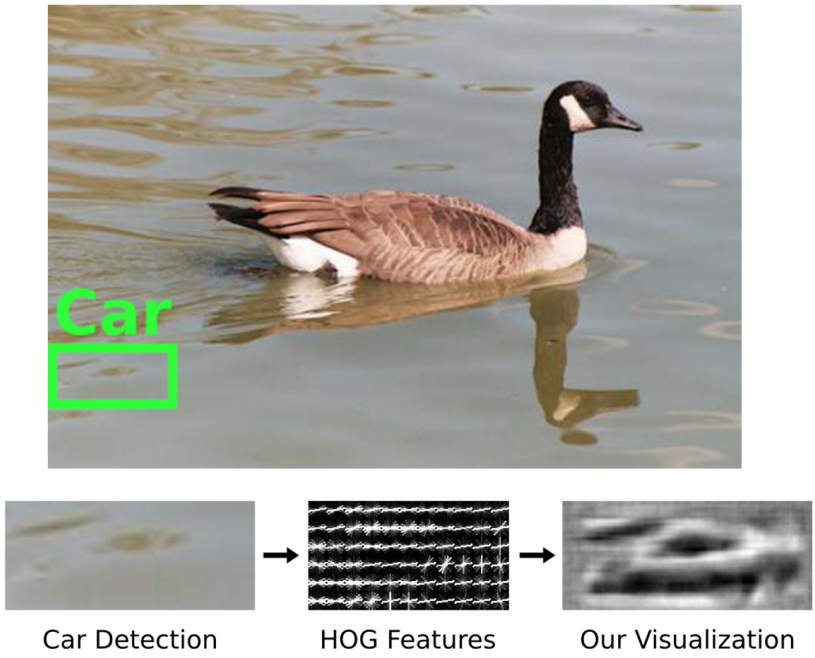

図 1.4　HOG 特徴量の可視化と誤りの原因 (Vondrick *et al.* [49])：(上) 物体認識において，車と誤認識されてしまったウインドウ (緑枠) (下) 車と判断された領域→ HOG のエッジ空間 → Vondrick らの可視化：全く違う物体であるにも関わらず，形状の組み合わせにより車に見える．HOG の記述ではまだ問題を含んでいると考える．

判断する特徴量が HOG であり，識別器には Support Vector Machine(SVM) を適用している．Vondrick らは HOG を可視化した HOGgles(HOG goggles) を提案しており，HOG がなぜ誤りを含むのかを追及している [49]．図 1.4 に HOG の可視化と誤認識を含んだ画像の例を示す．上図では水面の領域を車と誤認識してしまった．可視化手法を施してみると水面のエッジの組み合わせが車の形状に見える．全く形状に関係がないと思われていた背景にも誤認識する原因が含まれていることが実証された．このようなことからも HOG では形状記述能力が十分とは言えず，二つ以上のエッジ特徴の共起性を考慮した共起特徴量では過検出を減らせるということで期待される．

部位領域間のつながりを評価可能であるパーツ特徴は非常に頑健な検出結果を示すが，Pictorial Structure や DPM など用いている特徴量が HOG であり，局所特徴量自体の精度ではなく，モデリングの部分において精度を高めていると言える．ここで，Holistic 特徴の一種である HOG よりも高い精度を示す共起特徴に着目して改善する必要がある．共起特徴の中でも CoHOG は特徴の表現方法や高次元であるという点で改善する余地があると考え，本論文では

CoHOG にフォーカスして改善を実行する．

1.2.2 局所特徴量の関連手法

- **キーポイントベース**：画像上で取得した特徴点上で特徴量を取得し，Bag-of-features により特徴ベクトル化する手法である．SIFT や SURF などがその代表であり，人物中の特定の部位の記述とマッチングを可能にするが，人物は必ずしも一定のサイズが保証されているわけではなく，十分な特徴点が取得できないという問題点があるため人物検出には向かない．
- **動作ベース**：Optical Flow を画像化して方向ヒストグラムを記述する MBH が主な手法として知られている．MBH は取得したフローにより物体間のエッジを取得するため，移動する人物と背景を分離可能である．人物中や背景に存在するテクスチャに依存しないため，移動が見られる場合には距離画像のように特徴が取得できるが，人物が静止している場合には特徴が取得困難である．
- **形状ベース**：(i) テンプレート特徴 (ii)Holistic 特徴 (iii) 共起特徴 (iv) パーツ特徴に分類してそれぞれ説明した．テンプレート特徴では予めシルエットやエッジなどの特徴でテンプレートを作成して画像中の人物とマッチングする手法である．階層構造やピラミッド構造にするなどの対応策は存在するが，スケールや位置ずれには頑健ではないという問題がある．Holistic 特徴では統計的学習の概念を取り込み，識別器の重み付けにより位置ずれ問題を解決している．スケール対応においてはピラミッド構造にして逐次マッチングが必要であるが，テンプレート特徴からは精度を向上させている．さらに頑健にするために共起特徴が提案された．共起特徴では異なる位置に存在するエッジを組み合わせる手法であり，組み合わせの方法としては識別器の応答値から良好な組み合わせを選択する方法や特徴選択ウインドウにより 2 画素の組み合わせを取得する手法が見られる．単一の画素ではなく，複数の組み合わせが共起する場合にのみ反応する特徴量であるため，Holistic 特徴からは特に過検出を減らすことができるため，精度が向上している．パーツ特徴においては，画素単位ではなく領域毎のつながりに関して評価を行っている．人物に関して言えば姿勢変動に対して頑健であり，高い精度を誇ることでも知られる．部位間の繋がりのモデリングに関して精度を向上させているが，抽出している特徴量が HOG であるため，特徴

量に関する改良が望める．

キーポイントベースの手法はあらかじめモデルとなるテンプレートを準備する必要があり，マッチング対象のバリエーションも広くないという問題があるため人物の検出用には向いていない．また，動作ベースの特徴に関しても，動きがある場合は効果的であるが，人物が静止している際には背景との差分が出せないため適用場面に依存する．人物の形状には姿勢変動やスケールなど見え方にバリエーションが存在するが，形状ベースが人物検出向きである．その中でも共起特徴 CoHOG は記述方式や特徴ベクトルの変換などさらに改善の余地がある．CoHOG はエッジ方向ペアの数がどの程度分布するかをヒストグラムに記述する手法である．エッジの方向を参照してはいるがその他の成分を考慮していないため，エッジが存在するかどうかによりヒストグラムへ挿入してしまう．エッジ強度の強弱に依らずにヒストグラムへ累積するため，例えば人物と椅子等，方向成分が対象物体と少しでも似ている物体が画像中に存在する場合に過検出が発生してしまう．その他，強度が弱いながらも複雑なテクスチャを含む物体が存在する際には誤検出する可能性がある．

1.3 研究目的

本論文では共起特徴量 CoHOG に着目した改善手法を提案し，人物検出を始めとした人物行動解析の様々な場面に適用してその有効性を示すことを目的とする．局所特徴量の改善はコンピュータビジョンの課題とされていた諸問題を解くことと考え，人物検出だけでなく追跡や行動理解といった応用に広く適用できる．

CoHOG はエッジ方向のペアを記述する特徴量であるが，エッジ方向ペアのカウントにより特徴を記述していることや，オリジナルの CoHOG では約 35000 次元と高次元の特徴量であった．提案手法ではエッジ強度ペアの累積により，人物特有の曲率度や強度成分を表現する．CoHOG ではエッジ方向ペアの数のカウントで特徴を記述しているが，ECoHOG では共起特徴量としてエッジ強度の分布を記述する．エッジ強度の累積ではエッジが存在するということだけでなく，強弱まで含めて記述できる．弱いエッジ成分は重要度が低く，強いエッジ成分を識別に重要視できると考える．また，主成分分析の適用により，効果的に次元を圧縮し識別に有効な特徴ベクトルへと変換する．実験では特徴量の設定について記述するだけでなく，関連手法との比較を行

い，考察する．

局所特徴量の応用として，「歩行者予防安全」「サッカー映像解析」「行動理解」の場面において適用する．「歩行者予防安全」では複雑背景や動作の中での歩行者検出・追跡を，「サッカー映像解析」では選手の密集による多数の重なりがある場面での複数選手追跡，「行動理解」では人体の姿勢変動を含む，時系列に生起する動作に対して人物の行動識別を解決する．

1.4 本論文の構成

本論文の構成を以下に示す．

2 章では人物検出のための局所特徴量に関する改良方法を提案する．共起特徴である CoHOG に着目して，その特徴と課題を述べた上で，特徴記述，ヒストグラム正規化といった特徴量強化の手法について述べる．また，効率的な特徴表現のため，特徴次元圧縮方法について述べる．

3 章では，人物検出についてデータセットや実環境下で撮影した映像データを用いて実験し，結果を示すと同時に考察を加える．局所特徴量を用いた人物検出について多角的に実験をしている．

4 章では局所特徴量を適用した応用例として「歩行者予防安全」「サッカー映像解析のための複数選手追跡」「行動理解」の場面において適用する．それぞれの問題点や局所特徴量を用いた解決策を提案する．

5 章にて本論文の結びとする．まとめと現在人物検出が抱えている課題について説明する．さらには局所特徴量の改善のための展望についても述べる．

1.5 本章のまとめ

本章ではコンピュータビジョンの生起とその技術を適用した人物行動解析，局所特徴量の関連研究に関して記述した．その上で本論文における研究目的を説明した．

コンピュータビジョンにおける人物行動解析においては画像中の人物位置を特定する検出，見つけた人物位置を時系列で対応付ける追跡，画像中の人物が「何をしているか」に着目して行動タグを付加する行動理解など，様々な要素を抽出する方式が存在する．そのためのキーとなるのが統計的に特徴を分析し，人物をモデリングする機械学習である．機械学習では学習画像から統計的に識別器を生成する「学習フェーズ」と生成した識別器により認識する「検出フェーズ」に

分けられる．いずれにも共通して，対象物体を効果的に表現する局所特徴量の設定が非常に重要な問題となるため，局所特徴量の研究は非常に重要なテーマとなる．形状特徴量が人物の特徴を表現するために非常に効果的な手法であることが分かっており，関連研究を (i) テンプレート特徴 (ii)Holistic 特徴 (iii) 共起特徴 (iv) パーツ特徴の 4 種に分けて紹介している．部位領域間のつながりを評価可能であるパーツ特徴は非常に頑健な検出結果を示すが，Pictorial Structure や DPM などにおいて用いている特徴量が HOG であり，局所特徴量の改善により更なる改善が期待される．ここで，Holistic 特徴の一種である HOG よりも高い精度を示す共起特徴に着目して改善する必要がある．共起特徴の中でも CoHOG は特徴の表現方法や高次元であるという点で改善する余地があると考え，本論文では CoHOG にフォーカスして改善を実行する．CoHOG はエッジの共起性を捉える手法であり，過検出を低減しているという点において人物検出のスタンダードな手法である HOG よりも高い性能を示す．

本論文では共起特徴量 CoHOG に着目し改善策を提案する．CoHOG はエッジ方向のペアを記述する特徴量であるが，エッジ方向ペアのカウントにより特徴を記述していることや，オリジナルの CoHOG では約 35000 次元と高次元の特徴量であった．提案手法ではエッジ強度ペアの累積により，人物特有の曲率度や強度成分を表現する．また，主成分分析の適用により，効果的に次元を圧縮し識別に有効な特徴ベクトルへと変換する．実験では特徴量の設定について記述するだけでなく，関連手法と比較する．さらには人物検出を始めとした人物行動解析の様々な場面に適用してその有効性を示す．局所特徴量の改善はコンピュータビジョンの課題とされていた諸問題を解くことと考え，人物検出だけでなく追跡や行動理解といった応用に広く適用する．

2 提案手法

本章の概要

本章では人物検出のための局所特徴量の改善について記述する．人物を検出するためには局所特徴量を設定する必要があるが，ここでは共起性を考慮した特徴量 Co-occurrence Histograms of Oriented Gradients (CoHOG) を改善した Extended CoHOG (ECoHOG) を提案する．CoHOG では特徴取得ウインドウを用意し，中心となる画素とペアを作るための対象画素から取得するエッジ方向ペアのカウントにより特徴を表現していた．一方，ECoHOG では 2 つの画素から取得するエッジ強度をペアとして累積し，特徴を記述する．エッジ強度は明るさによって変動するためヒストグラムを正規化する．さらには高次元特徴量の空間サイズを効率的に削減するために，主成分分析を適用した次元圧縮の処理を加えて改良する．まずは代表的な人物検出手法である HOG について説明した後，CoHOG，提案手法の ECoHOG について説明する．

2.1 人物検出のための局所特徴量

HOG 特徴量が提案されてから人物検出の研究が急激に進んでいる．HOG 特徴量は現在でも最も一般的に用いられる手法であるが単一の画素からの特徴記述なために人物の未検出や背景部分に過検出が発生してしまう．HOG 特徴量の改良研究は多数提案されてきたが，最も高精度に歩行者を検出する特徴量として CoHOG が知られている．しかし，著者は CoHOG にも未だ改良の余地は残されていると判断した．本論文では形状ベースの局所特徴量について代表的な手法である HOG について述べた後，局所特徴量である CoHOG を，そして提案手法である ECoHOG について順を追って紹介する．

2.1.1 Histograms of Oriented Gradients (HOG) [30]

Dalal らは 2005 年の CVPR において HOG 特徴量を発表した [30]．人物検出や物体検出等，コンピュータビジョンの認識においてはすでにスタンダードな手法の一つになっている．HOG 特徴量は，エッジ方向を対象として量子化したヒストグラムにエッジ強度を累積する特徴量であ

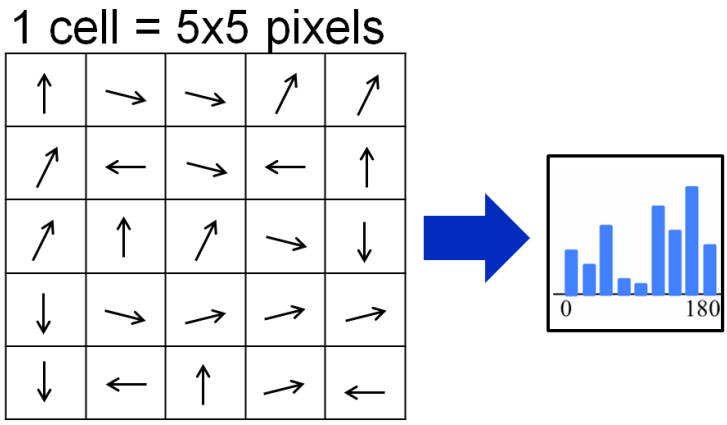

図 2.1　セルの構成：5×5 ピクセルで構成されている．各画素から抽出した方向に対応するヒストグラム位置にエッジ強度を累積する．180 度のエッジ方向を 9 つに量子化してヒストグラムを構成する．

り，物体の大まかな形状を表現可能である．

HOG 特徴量のための前処理としてガウシアンフィルタを施しており，画像をぼかすことによりエッジ強度を分散させ多少の位置ずれに対応する．HOG ではエッジの強度情報による特徴表現のためノイズに敏感であり，少しでもノイズが発生すると強度情報として蓄積されてしまう．位置ずれへの対応やノイズ除去の意味で画像を平滑化してからの方が精度が高く，前処理として採用されている．

まずはセルとブロックに分割してエッジ情報を取得する．セルは図 2.1 に示すように 5×5 ピクセル，ブロックは 3×3 セルと設定し，分割した様子を図 2.2 に示す．セルからは 180 度のエッジ方向を 9 つに量子化したヒストグラムを抽出する．5×5 ピクセルの各画素から抽出した方向に対応するヒストグラム位置にエッジ強度を累積する．さらに，ブロックでは 3×3 のセルを連結して 9(セル)×9(方向) で 81 次元の特徴次元数を得る．

以下に特徴取得で使用するエッジ方向と強度の取得方法を示す．

$$m(x,y) = \sqrt{f_x(x,y)^2 + f_y(x,y)^2} \tag{1}$$

$$g(x,y) = tan^{-1}\frac{f_y(x,y)}{f_x(x,y)} \tag{2}$$

$$f_x(x,y) = I(x+1,y) - I(x-1,y) \tag{3}$$

$$f_y(x,y) = I(x,y+1) - I(x,y-1) \tag{4}$$

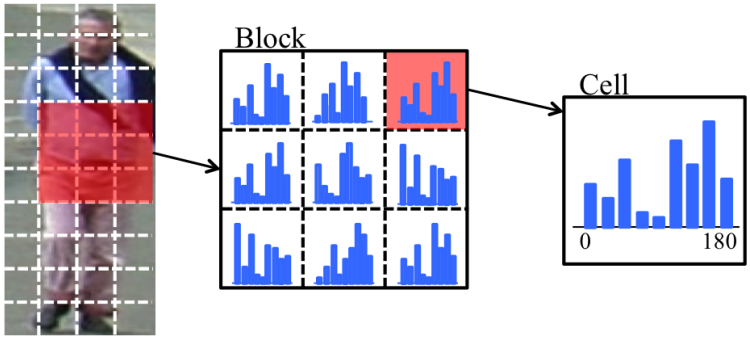

図 2.2　人物画像からのブロックとセルへの分割：人物画像は 5×5 のセルに分割される．セルは 3×3 セルで構成され，特徴量はセルをオーバーラップさせることにより取得する．

ここで，x, y は画像中における座標位置，$m(x, y)$ は位置 (x, y) におけるエッジ強度，$g(x, y)$ は (x, y) におけるエッジ方向である．$f_x(x, y), f_y(x, y)$ はそれぞれ x, y 方向の微分を示す．$I(x, y)$ は画素値である．

一般的に，特徴量の取得には，セル領域をオーバーラップしながら特徴を抽出する．特徴取得後，ブロック毎の正規化処理によりヒストグラムの形状を整え最終的な特徴量とする．正規化処理により，明るさの変動に対応可能であると考えられる．正規化方法を以下に示す．

$$h' = \frac{h}{\sqrt{\sum_{i=0}^{k} h_i^2 + \epsilon}} \tag{5}$$

h, h' はそれぞれ正規化前と後の特徴ベクトル，k はブロック内の次元数，ϵ は除算により分母が 0.0 にならないようにする係数であり，ここでは $\epsilon = 1.0$ に設定する．

HOG の次元数は画像サイズに依存して変化する．例えば 30×60 ピクセルの画像から HOG 特徴量を取得する場合，1 セルは 5×5 ピクセル，1 ブロックが 3×3 セルと仮定して，x, y 方向それぞれオーバーラップさせて特徴取得可能なブロック数が 4(x 方向) と 10(y 方向) である．1 ブロックからは 81 次元の特徴が取得できるので (4×10)×81=3240(次元) の特徴ベクトルが取得可能である．

図 1.3 には Dalal らの HOG 可視化を示しているが，図中輝度の高いほどエッジ強度が高い位置である．歩行中は脚部の運動が大きいので，上半身のエッジの累積が安定しているため，頭部や胴体位置の方が強度が高い傾向にあると言える．背景画像から取得した強度の分布に関しては学習画像に依存する．

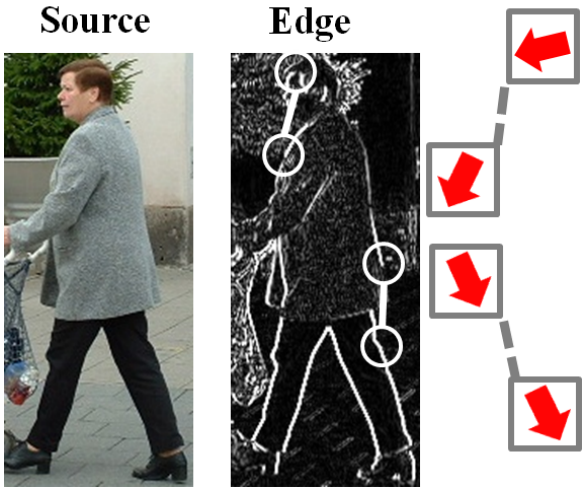

図 2.3　勾配方向のペア：(左) 人物の原画像 (中) エッジの強度や方向を示した画像 (右) 頭部と肩部，腰部と脚部から取得する共起性の例

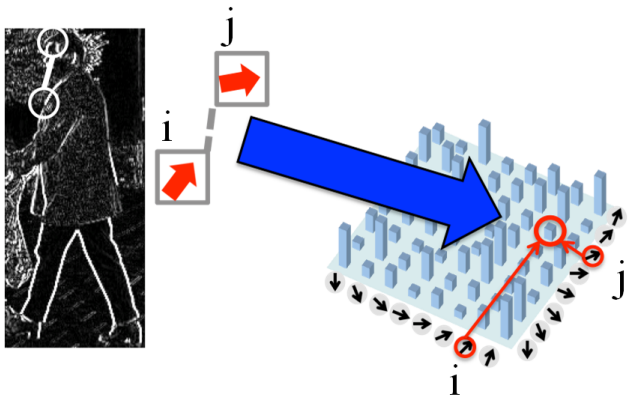

図 2.4　共起表現とヒストグラムへの累積：共起性を取得する 2 つの対象画素からそれぞれエッジ方向を抽出．8 つに量子化した量子化番号 i, j それぞれに対応するヒストグラム位置に特徴を累積する．CoHOG の場合にはペア数のカウントにより特徴を表現し，ECoHOG の場合にはエッジの強度累積により特徴を記述．

2.1.2　Co-occurrence Histograms of Oriented Gradients (CoHOG) [43]

CoHOG は，Watanabe らが 2009 年に発表した手法であり，基本的なアイディアとして離れた位置にある 2 つの画素のエッジ方向のペアをカウントしたヒストグラムにより特徴を記述する．HOG では，局所的に表れるエッジの方向と強度をヒストグラムに累積し特徴量としており単一画素の特徴記述であった．CoHOG では単一の画素ではなく，ペアとして 2 つの画素の特徴

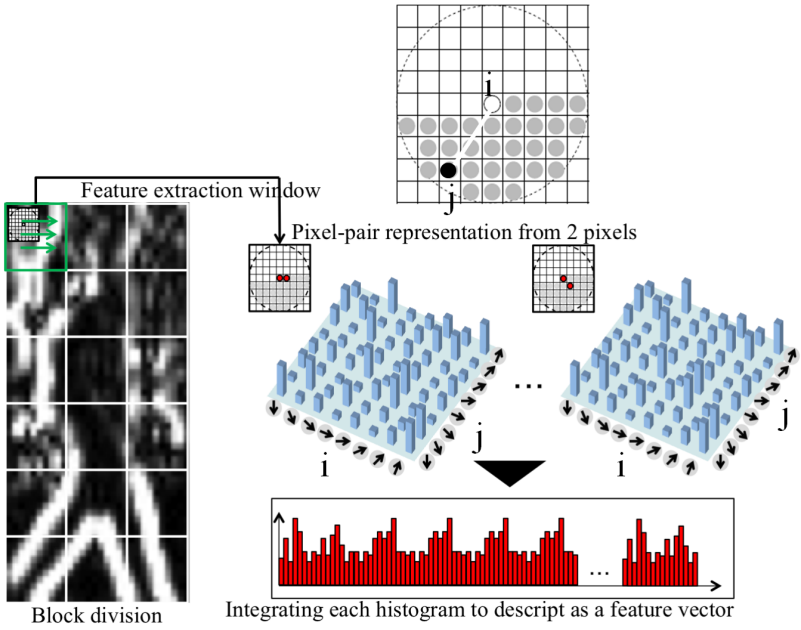

図 2.5　CoHOG，ECoHOG のヒストグラム取得方法：初期設定ではブロック分割数，オフセットでは特徴取得画素数やオフセットサイズを与える．その後，画像をブロック分割してそれぞれのブロック内でオフセットをスキャンさせて共起特徴を取得する．オフセット内の中心画素 (i) と共起の対象となる画素 (j) のエッジ方向を取得して共起ヒストグラムの対応する位置にカウントして累積する．オフセット内全ての対象となる画素 (j) が取得できた際にブロック内でオフセットをずらして再び共起特徴を取得する．ブロック内を全探索するまで繰り返し，全てのブロックに対して同じように共起ヒストグラムを取得する．特徴が取得できた際，全ブロック内，オフセット内全ての特徴取得画素の共起ヒストグラムを連結させて特徴ベクトルを構成する．

を同時に記述することにより，過検出を大幅に削減している．これは，頭部と肩，腰と脚部などの輝度勾配方向が同時に存在する際のみ検出するためである (図 2.3)．CoHOG のヒストグラム表現も図 2.4 に示す．CoHOG では共起性を取得するため，対象となる 2 つの画素からエッジ方向を抽出する．共起ヒストグラムでは 8 方向に量子化した二つのエッジ方向 (i, j) のそれぞれにヒストグラムのビンを対応させ 64 次元のヒストグラムを得る．2 つの異なる画素から取得した量子化番号を基にして，共起ヒストグラムの格納位置を求める．

ここで，CoHOG の特徴量取得の流れを図 2.5 に示す．CoHOG の特徴取得ではオフセットと呼ばれる特徴取得ウインドウを用意し，画像内をスキャンして特徴量を取得する．オフセットは矩形形状をした特徴取得ウインドウであり共起特徴を取得するため，オフセット (図 2.5 上) 中心の白円と黒円 (灰色円) のペアの勾配方向を求め，ヒストグラムにサンプリングする．オフセッ

トの矩形領域において中心から右下側しかペアを取得しないのは，スキャンした際に左上側とペアを重複して取得しないように設定しているからである．エッジ方向は 8 等分に区切り，注目画素とペアとなる画素の 2 つを組み合わせるので，一つの共起ヒストグラムで 64 次元の特徴量が得られる．64 次元の共起ヒストグラムはオフセット内の灰色の画素毎に用意されるため，オフセットに依存して次元数が変化する．また，次元数を左右する要素としては特徴を取得する画像のブロック分割数が挙げられる．CoHOG では位置に関係なく共起特徴を取得してしまうため，分割しない画像においては全ての特徴の位置関係を無視して特徴を累積してしまう．これを，ブロック分割することで大まかに上半身や下半身，右半身と左半身など位置毎に特徴累積を可能にできる．オフセット内の特徴取得画素数が 18，ブロック分割数が x, y 方向にそれぞれ 2 分割だとすると，64×18×2×2 で 4608 次元の共起特徴量となる．画像が大きくてもスキャンして累積，ブロック内に含まれるエッジの共起成分を累積するので次元数には変化が無い．

CoHOG の特徴取得の流れを以下に示す．

- 初期設定：ブロック分割数，オフセット (特徴取得画素数，オフセットサイズ)

1. ブロック内のスキャン
2. オフセット中心画素 (i) のエッジ方向取得
3. オフセット内対象となる画素 (j) のエッジ方向取得
4. 2 つの画素から取得したエッジ方向を共起ヒストグラムに累積
5. 1 に戻り，ブロック内を全探索するまで繰り返し

エッジ方向は以下に示す式により画像から計算している．ここで，$I(x, y)$ は画像の輝度値，$g(x, y)$ はエッジ方向をそれぞれ示す．

$$g(x, y) = tan^{-1} \frac{f_y(x, y)}{f_x(x, y)} \tag{6}$$

$$f_x(x, y) = I(x + 1, y) - I(x - 1, y) \tag{7}$$

$$f_y(x, y) = I(x, y + 1) - I(x, y - 1) \tag{8}$$

ここで，x, y は画像中の画像位置，$g(x, y)$ は位置 (x, y) におけるエッジの方向を示す．特徴記述の式についても以下に示す．特徴は，注目画素とペアとなる対象画素に勾配方向を量子化した値

を割り当てる．2 つの画素にそれぞれ 8 通りの方向をもつので，64 次元となる．

$$C_{x,y}(i,j) = \sum_{p=1}^{n}\sum_{q=1}^{m}\begin{cases} 1 & \text{if I(p, q) = i and I(p + x, q + y) = j} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

x, y は対象画素の位置，p, q は特徴取得ウインドウ中における注目画素の位置を示す．i, j は画素のヒストグラム値のことであり，n, m は画素のヒストグラム値を表す．さらに $C_{x,y}(i, j)$ は共起ヒストグラムであり，(x, y) 位置におけるヒストグラム (i, j) を示す．

CoHOG では勾配方向の組み合わせ数をカウントすることで特徴量を構成する．注目画素と対象画素の勾配方向を量子化した値の対応するヒストグラム位置に 1 をプラスしている．

CoHOG では 2 つの異なる画素での勾配方向ペアの出現頻度によりヒストグラムを構成しているので，共起性が見られない対象物体は検出されない．よって，HOG で問題視された過検出を低減することができ，非常に高い精度を誇ることが報告されている．しかし，CoHOG では共起する方向ペアの数を成分としてどの程度含んでいるかを記述方式としているため，その強弱に関係なくヒストグラムに累積されてしまうという問題点が挙げられる．エッジ方向ペアの「数」では単純に方向ペアが存在する場合に累積してしまうので形状が同じ場合にはほぼ間違いなく過検出が発生する．例えば屋内環境では人物と椅子，車載映像でも歩行者とスケールの近い木や電柱など，同じようなエッジ方向ペアの成分を含む物体に関しては過検出が発生してしまう．さらにはテクスチャが複雑な際には多数の方向成分が存在するため，人物の服装や保持する物体等見え方の変化に依存してヒストグラムの形状が変化するという問題点を含んでいる．加えて，CoHOG では高次元な特徴量であるため特徴空間内での分離や，学習した特徴ベクトルと識別の際の特徴ベクトルに乖離があり識別に非常に不利である．学習画像を膨大に増やさないと同じような特徴ベクトルが出現しないため，学習には向かない．著者は以上のような点から改善の余地が残されていると判断した．

2.2 共起特徴量の改善手法

本論文では人物検出に用いられる手法として，CoHOG を改良した ECoHOG を提案する．エッジの共起性を捉える特徴量であることは共通するが，エッジの強度累積，ヒストグラムの正規化，次元圧縮を施す点で CoHOG と異なる．エッジの強度累積に関しては 2 つの画素から取

得したエッジ強度を和算するか積算するかで異なる特徴累積手法となる．エッジ強度を累積する際には画像の明度状況により変化してしまうため，ヒストグラムをノルムで除算することにより正規化する．次元の圧縮では主成分分析を適用して効果的に次元を圧縮する．

2.2.1 エッジ強度累積による特徴記述方法の強化

エッジの強度累積．特徴量の累積方法はエッジ方向をペアとしてカウントするのではなく，エッジ強度の累積により特徴を記述している．累積方法も和算，積算を試みる．和算による累積ではエッジ強度が弱かったとしても，特徴を全て残すことができるため，背景との位置関係も含めた学習ができると考える．一方，積算による特徴記述では片方のエッジでも弱かった場合には強度情報としてほとんど残らないため，人物の外輪郭に存在する比較的強い強度のエッジのみを残す方法であると言える．

まずは積算による累積方法を示す．

$$m_1(x_1, y_1) = \sqrt{f_{x1}(x_1, y_1)^2 + f_{y1}(x_1, y_1)^2} \tag{10}$$

$$m_2(x_2, y_2) = \sqrt{f_{x2}(x_2, y_2)^2 + f_{y2}(x_2, y_2)^2} \tag{11}$$

$$C_{x,y}(i,j) = \begin{cases} \sum_{p=1}^{n} \sum_{q=1}^{m} \sqrt{m_1(x_1, y_1) m_2(x_2, y_2)} \\ (\text{if I(p, q)} = \text{i and I(p} + \text{x, q} + \text{y)} = \text{j}) \\ 0 \qquad\qquad (\text{otherwise}) \end{cases} \tag{12}$$

m はエッジ強度，(x_1, y_1) はオフセットの中心画素の座標，(x_2, y_2) はオフセット内，中心画素とペアとなる対象の画素から取得した座標である．i, j はエッジ方向の量子化番号 $(0 \leq i, j \leq 7)$ であり，n, m はブロック画像のサイズ，x, y はオフセット内の対象となる画素の位置，$C_{x,y}(i,j)$ は 64 次元の共起ヒストグラムである．

次に和算による累積方法を示す．ここで，エッジ強度の計算方法は上記 m と同様であり，他の引数についても同様である．

$$C_{x,y}(i,j) = \begin{cases} \sum_{p=1}^{n} \sum_{q=1}^{m} m_1(x_1, y_1) + m_2(x_2, y_2) \\ (\text{if I(p, q)} = \text{i and I(p} + \text{x, q} + \text{y)} = \text{j}) \\ 0 \qquad\qquad (\text{otherwise}) \end{cases} \tag{13}$$

特徴記述には注目画素と対象画素のエッジ強度ペアの積または和を累積している．積算の方式では 2 つの画素におけるエッジ強度の積を取るので，どちらも強度値が高い場合でないと弱い特徴

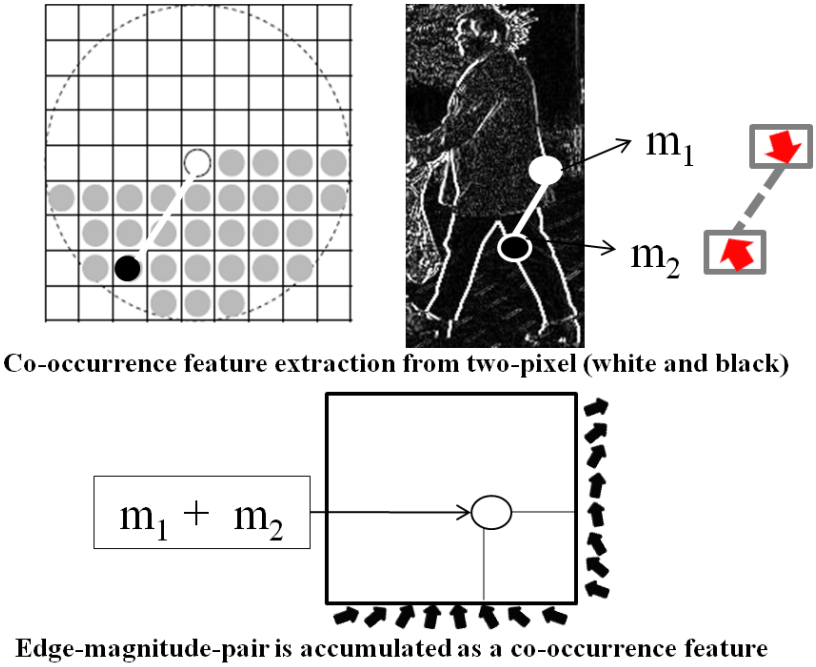

図 2.6　ECoHOG のエッジ強度累積：特徴抽出ウインドウであるオフセットの中心 (矩形中心の白画素) とエッジ共起を取得する対象画素 (矩形下半分の灰色もしくは黒画素) の 2 画素から強度を取得する．ヒストグラムへの累積については CoHOG と同様，エッジ方向を量子化した番号から対象のビン位置に累積する．図中の累積手法は $m_1 + m_2$ と和算であるが，積算の場合には $\sqrt{m_1 m_2}$ を累積していく．

になることから，強い特徴のみを取り出せると考えた (図 2.6)．また，和算の方式では 2 つの画素におけるエッジ強度の和を取るので，前景・背景問わず総合的に特徴量を残すことができる．積算の方式では $\sqrt{m_1 m_2}$ を，和算の方式では $m_1 + m_2$ をヒストグラムに蓄積していく．

ECoHOG では共起するエッジ強度をペアとしてヒストグラムに累積しているため，方向ペア毎にどの程度の強度成分を含んでいるか記述する．強度は方向ペアが存在するだけで一定の値を累積するのではなく，強度の成分を累積する．弱い成分が多数内在する場合 (人物の着用する服や車両など人工物内部のテクスチャ等) には弱い特徴として蓄積されるのに対して，外輪郭や上半身・下半身の境目など，検出の際には主に強いエッジ特徴に着目して識別するので精度が向上すると考える．曲率だけでなく，直線においてペアの方向成分が存在するというだけでなく，その度合も表現できるため数で表現する CoHOG よりも強度で特徴記述する ECoHOG の方が有利である．

ヒストグラム正規化．画像の明るさの変動によるエッジ強度の変化に対してロバストに検出で

きるよう，ヒストグラムを正規化する．正規化する範囲は，共起ヒストグラム単位の 64 次元とした．正規化は以下に示す式により計算する．

$$C'_{x,y}(i,j) = \frac{C_{x,y}(i,j)}{\sum_{p=1}^{8}\sum_{q=1}^{8} C_{x,y}(p,q)} \tag{14}$$

C, C' はそれぞれ正規化前と後の共起ヒストグラムを示す．

2.2.2 次元圧縮

主成分分析 (Principal Component Analysis : PCA) により ECoHOG の次元数を圧縮し，特徴空間内で正解と非正解の空間を分離しやすくする．従来の CoHOG[43] では約 35,000 次元と膨大な次元数を要したが，本論文では数十から数百次元に圧縮し特徴空間内で正解と非正解を分離しやすくする．識別する際の特徴空間が膨大になると準備する学習画像数が膨大になることや正解/非正解の分離が困難であると考える．実験では膨大になった特徴空間を，情報を保存しながら特徴空間が最適な特徴次元数を模索する．これは，パターン認識では一般的に言われている“次元の呪い”と呼ばれる問題を解決できると考える．次元の呪いでは，特徴次元数が増えるほどに学習サンプルと識別対象の物体の乖離が激しく，(学習画像と識別画像で同じような特徴をもつ物体が極めて少なくなり) 識別率が下がるとされている．情報を保存しながら次元削減を実現する主成分分析は有効な対策手段であると言える．

主成分分析は，次元削減に使われる手法の一種である．主成分分析は 2 つの異なる使用法があり，ひとつは部分空間と呼ばれる低次元の線形空間上への，直交射影として定義できる．また，もとのデータと射影した点の間の 2 乗距離の平均値で定義される射影のコスト関数の期待値を最小化するような線形射影としても定義できる．ここでは前者の低次元空間への射影を対象にして解説する．主成分分析では分散最大化による定式化を与える．

サンプル集合を $x_n(n = 1, ..., N)$，x_n は D 次元のユークリッド空間の変数である．これを，圧縮して $M(M < D)$ 次元の空間に射影することが主成分分析を用いた次元圧縮の目的である．まずは 1 次元空間 $(M = 1)$ 上への射影を考える．この空間の方向を D 次元ベクトル u_1 として表す．このベクトルは単位ベクトルと仮定 $(u_1^T u_1 = 1)$ する．各データ点 x_n はスカラー値 $u_1^T x_n$ の上に射影される．射影されたデータの平均値は $u_1^T \overline{x}$ である．$\overline{x}$ はサンプル集合の平均で

下式により示される．

$$\overline{x} = \frac{1}{N}\Sigma_{n=1}^{N} x_n \tag{15}$$

で与えられる．また，射影されたデータの分散は下式である．

$$\frac{1}{N}\Sigma_{n=1}^{N}\left(u_1^T x_n - u_1^T \overline{x}\right)^2 = u_1^T S u_1 \tag{16}$$

S は共分散行列であり，次のように与えられる．

$$S = \frac{1}{N}\Sigma_{n=1}^{N}\left(x_n - \overline{x}\right)\left(x_n - \overline{x}\right)^T \tag{17}$$

射影された分散 $u_1^T S u_1$ を u_1 に対して最大化する．最大化は $||u_1|| - > \infty$ を防ぐような制約付き最大化にならねばならない．適切な制約は正規化条件 $u_1^T u_1 = 1$ から来る．この制約を課すためラグランジュ乗数を導入し，制約なしに最大化する．

$$u_1^T S u_1 + \lambda_1(1 - u_1^T u_1) \tag{18}$$

u_1 に関する微分を 0 とおくことにより，

$$S u_1 = \lambda_1 u_1 \tag{19}$$

において停留点を持つ．これは u_1 が S の固有ベクトルでなければならないということである．上式を変形すると

$$u_1^T S u_1 = \lambda_1 \tag{20}$$

となる．したがって分散は u_1 を最大固有値 λ_1 に属する固有ベクトルに選んだときに最大となる．この固有ベクトルは第 1 主成分と呼ばれる．第 2 主成分以下もすでに得られている主成分ベクトルに直交するという条件のもとで射影分散を最大にするような方向を選ぶことで逐次的に得られる．主成分分析ではデータ集合の平均 $\overline{x}$ と共分散行列 S が必要であり，さらに S の上位 M 個の固有値に対応する M 個の固有ベクトルを求める必要がある．

図 2.7　階層的クラスタリングによる検出枠の統合：(左) ラスタスキャンによる人物検出では人物領域周辺に複数の検出枠が配置される (右) 階層的クラスタリングによる検出枠の統合，ユークリッド距離を指標として近傍の枠を統合する

2.2.3　クラスタリングによる検出位置特定

識別器が人物と判断した場合，検出枠が表示されるが，人物の周辺には複数のウインドウが表示されてしまう．そこで，検出枠が集中する箇所を統合することで最終的な人物検出とする．ここでは，近傍の検出枠をまとめクラス数を決定する階層的クラスタリングを適用する [54]．階層的クラスタリングは，事前情報を設定しなくても自動的にクラス数や統合するデータを決定するクラスタリング手法である．以下に，階層的クラスタリングの手順を示す．

1. 個々のデータをひとつのクラスタとして設定
2. クラスタ間の類似度を計算，最も類似しているクラスタを併合
3. クラスタ併合を終了条件になるまで処理を繰り返す

ここでは検出の中央位置の座標 $X = (x_i, y_i)(i = 1, 2, ..., N)$ を入力とする．座標間はユークリッド距離により計算してクラスタリングする．ユークリッド距離には閾値を設け，閾値距離以内の領域においてクラスタリングを実行する．階層的クラスタリングによる検出枠の統合を，図 2.7 に示す．図 2.7 左では検出枠が複数出力されているが，階層的クラスタリングの実行により，図 2.7 右のように検出枠が統合される．階層的クラスタリングはあらかじめクラス数を設定する必要がなく，画像中に複数の人物がいても処理が可能である．また，階層的クラスタリングの適

用により，複数の検出枠が集中する位置を最終的な人物検出位置にできるので，過検出をさらに低減可能である．

2.3 本章のまとめ

本章では人物検出のための局所特徴量について，HOG，CoHOG，提案手法である ECoHOG を記載した．HOG は人物検出の分野ではスタンダードな手法の一つになっている．HOG はエッジ方向を対象として量子化したヒストグラムにエッジ強度を累積する特徴量であり，物体の大まかな形状を表現している．HOG では画像をブロックとセルに分割してエッジ強度成分を累積する特徴である．CoHOG では離れた位置にある 2 つの画素のエッジ方向のペアをカウントしたヒストグラムにより特徴を記述する．特徴取得ウインドウであるオフセットを準備して共起ペアを取得し，それぞれの画素に対応する量子化番号のヒストグラム位置にカウントする．CoHOG を改良した ECoHOG では，エッジ強度の累積・ヒストグラムの正規化・主成分分析を適用した次元圧縮を加えた．エッジ強度の累積では和算による累積，積算による累積を提案している．和算では特徴量のペアを総合的に判断できるように，積算では強い特徴量が残るような戦略で特徴量を構成する．エッジ強度は画像の明るさにより変動してしまうので，明るさに依存しないようヒストグラムの正規化を加えており，共起ヒストグラム 64 次元単位で正規化した．また，学習を有利にするため，そして特徴空間のサイズを縮小してクラスの分類を容易にするよう主成分分析により次元を圧縮している．

3 実験・評価及び考察

本章の概要

本章では，提案した局所特徴量を適用した人物検出の精度を検証する．まずは使用するデータセット INRIA person dataset や Daimler pedestrian benchmark dataset について概説する．局所特徴量の実験では従来法との比較だけでなく主成分分析や和算・積算によるヒストグラム累積について提案した特徴量の設定を検証する．

3.1 実験概要

本論文で使用した INRIA(Institut National de Recherche en Informatique et en Automatique：フランス国立情報学自動制御研究所) person dataset，Daimler pedestrian benchmark dataset ついて解説する．

3.1.1 データセット

INRIA Person Dataset． INRIA person dataset は人物検出用のデータセットであり，正解画像：人物画像と非正解画像：それ以外の背景画像を学習させて正解か非正解かを判定する (図 3.1，図 3.2) [55]．ここでは局所特徴量の表現能力を比較するために INRIA person dataset を使用する．学習用に配布されている画像はそれぞれ正解が 2415 枚，非正解が 12180 枚，テスト用の画像は正解 1132 枚，非正解 453 枚である．画像サイズは正解が 64×128 ピクセル，非正解画像は $214 \times 320 - 648 \times 486$ ピクセルの画像からランダムで画像を切り抜いて使用する．

Daimler Pedestrian Benchmark Dataset． Daimler pedestrian benchmark dataset は人物検出用に人物の全身を切り出しているデータセットである [56]．自動車前方の人物を検出する目的で作成されたデータセットであるため，人物のサイズが 18×36 ピクセルに設定されている (図 3.3)．正解画像 4800 枚，非正解画像 5000 枚が用意されている．

3.1.2 実装

ECoHOG の設定について，特徴取得ウインドウ中の特徴取得画素数は 18 ピクセル，量子化数は 8 次元，共起ヒストグラムの次元数は $8 \times 8 = 64$ 次元，ブロック分割数は x, y 方向それぞ

図 3.1　INRIA person dataset の正解画像例

図 3.2　INRIA person dataset の非正解画像例

れ 2×2，総次元数は 4608 次元である．また，主成分分析を用いた次元圧縮については実験により $5 - 200$ 次元のうち最適な次元数を決定する．和算/積算による特徴抽出も同様である．また，CoHOG の設定も ECoHOG と同様であり，総次元数は 4608 次元に設定した．

実装した環境はラップトップ PC であり, Intel Corei7-3520M, CPU 2.90GHz, RAM 8.00GB である．また，システムは C++ と OpenCV2.4 を用いた環境下で実装されている．

解像度は 640×480 ピクセル，フレームレートは $30fps$ に設定した．

図 3.3　Daimler pedestrian benchmark dataset の正解画像 (上) と非正解画像 (下)

3.2　人物検出実験と考察

3.2.1　特徴量累積手法の選択

まず，ECoHOG の設定について述べる．ECoHOG のエッジ強度ペアの累積では和算による手法と積算による手法が存在する．まずはどちらの手法が有効であるかについて検証した．図 3.4 に 2 つの手法から取得した Detection Error Tradeoff (DET) カーブを示す．ここで，和算による累積を ECoHOG plus，積算による累積を ECoHOG multiple と表記している．DET カーブでは縦軸に未検出率 (Miss rate) を，横軸は過検出率 (False positive rate) を示す図である．よって，左下の原点に近ければ近いほど高い性能を示す．図 3.4 からは左下の原点に近い手法は ECoHOG plus であることが確認できる．詳細に見ると，ECoHOG plus が過検出を抑える手法であり，ECoHOG multiple が未検出を抑える手法であることが分かる．和算による手法では特徴を総合的に見る特性があるため，見落とすエッジ特徴が少なく過検出が抑えられた．一方で積算による累積では強い特徴量のみを残すため，過検出は見受けられたが，未検出が抑えられたと言える．人物検出では過検出をしてしまうとシステムとして成り立たないため，過検出を抑えられる ECoHOG plus を適用し，性能を出来る限り向上させるように改良を重ねた．以降，ECoHOG plus を ECoHOG と省略して呼称することとする．

3.2.2　圧縮次元数と検出性能

次に，PCA-ECoHOG の次元数と検出率との関係を示した実験 (図 3.5) においては，5〜200 次元に圧縮した特徴ベクトルとその識別性能を示している．図中では 5 次元が最低，100 次元で

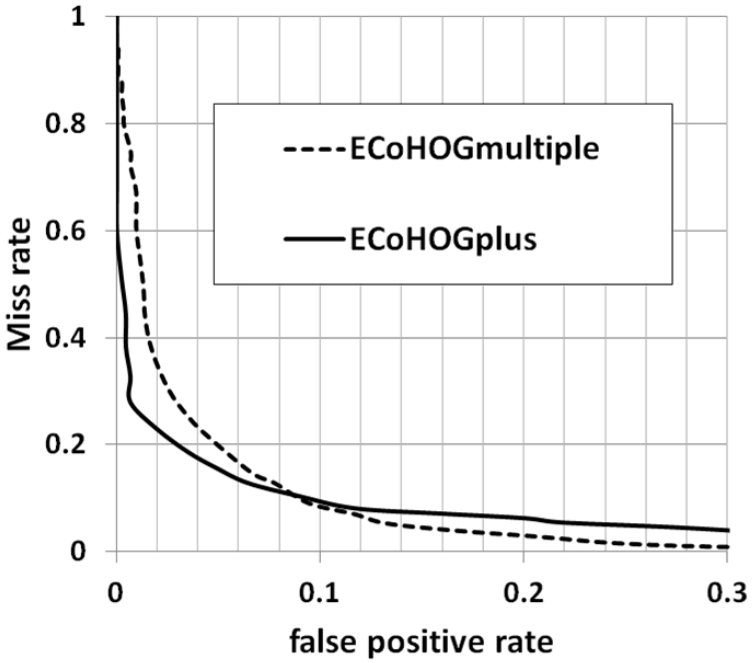

図 3.4　ECoHOG plus と ECoHOG multiple の DET カーブ：和算による累積を ECoHOG plus(図中の実線)，積算による累積を ECoHOG multiple(図中の点線) と呼称する．DET カーブは左下の原点に近いほど高い精度を示すため，ECoHOG plus が高い精度を誇る．途中で精度が入れ替わっているが，ECoHOG plus は過検出を抑える手法，ECoHOG multiple は未検出を抑える手法であることが読み取れる．総合的な制度や多くのシステムでは誤報を抑えるシステムが理想的であり，過検出を減らす必要があることからも，ECoHOG plus を ECoHOG のエッジ強度ペア累積手法として採用する．

最高性能を示している．5 次元では圧縮したベクトルに情報量が小さいために識別性能が下がり，200 次元では特徴空間が広くなり識別性能が悪くなったと考えられる．ここで，情報の保持率も重要であるがその他の要素として，特徴空間サイズも重要な要素として挙げられる．情報の保持率は高いほど良く，特徴空間サイズは小さい方が正解/非正解を分離しやすいとされるが，両者のバランスも非常に重要であると考える．4608 次元の ECoHOG においては 100 次元が特徴空間や特徴ベクトルに含む情報量が最適であると言える．

提案した特徴量 ECoHOG，PCA-ECoHOG の比較 (図 3.6) において，ECoHOG で得られた特徴ベクトルを圧縮した PCA-ECoHOG が高い性能を示した．図 3.6 の ECoHOG，PCA-ECoHOG の比較は見やすさのため，対数表示にしている．PCA-ECoHOG では 4608 次元から 100 次元に圧縮しているため，特徴空間内で正解と非正解の分離が容易になったためと考えられ

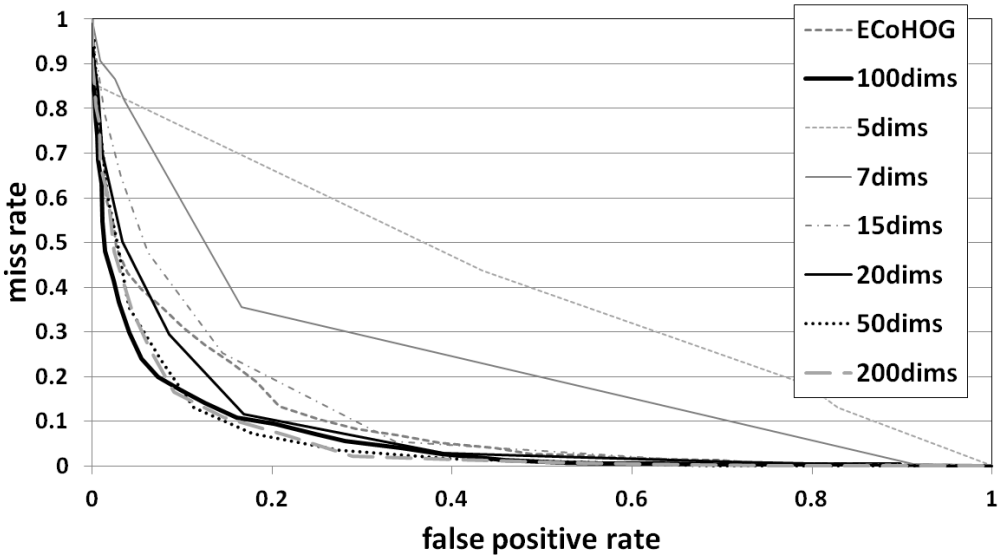

図 3.5　PCA による圧縮次元数と DET カーブ：5-200 次元 (5,7,15,20,50,100,200) と次元数を変更して精度を比較している．5 から 100 次元までは増加の傾向にあるが 200 次元になると精度が下がっている．これは，情報量 (累積寄与率) の指標が大事であるが一方で特徴空間サイズの大きさが正解・非正解の分離に重要な要素であることを示す．

る．PCA は多数の特徴次元間の共分散を少数の合成関数で説明する方法であり，効果的に特徴次元を圧縮できている．一般的に識別問題においては少数の次元であるほど分離が容易であるため，人物を検出する際の多数次元から少数次元に圧縮する識別方式は効果的であると言える．主成分分析に関して，今回は人物を対象としてモデリングしている．人物は正面や側面など，向いている方向に依らず頭部から肩部にかけて見られるΩ型の形状を捉える事ができる．人物の識別においても，このΩ形状を基にして検出している．この特徴に関しては服装や体型などに依存せず，スケールを固定すればデータセットによる差異は非常に少ないので，背景が学習できていれば主成分分析のモデリングは汎用的であると考える．依って，交差検定法や背景画像さえ変えれば学習の汎用的な問題はクリアできていると考える．

3.2.3　従来手法との特徴量比較

従来手法である HOG [30]，CPF [40]，CoHOG [43] と提案手法の ECoHOG を比較 (図 3.7) した．従来手法との比較において対数表示にしており，ECoHOG が最も高い精度を示している．これは，共起性を考慮した特徴取得だけでなく，エッジの強度を捉えていることや，明るさ

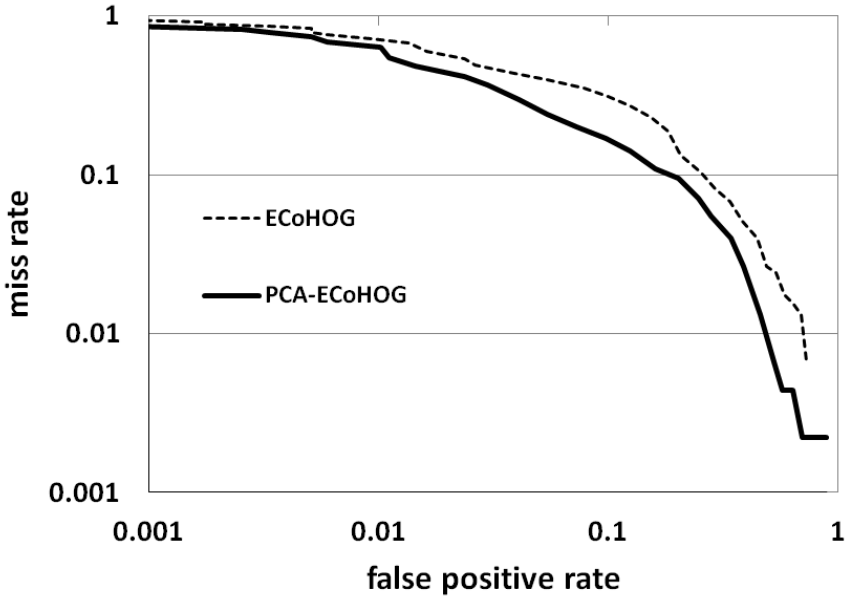

図 3.6　2 つの提案手法 (ECoHOG, PCA-ECoHOG) の DET カーブ：INRIA person dataset を対象として ECoHOG と主成分分析 (PCA) により圧縮した PCA-ECoHOG の比較を行った．主成分分析では射影されたサンプルの分散を最大化しながら，圧縮された空間へ射影することにより，効率的に次元数を減らすことが可能である．グラフを見ても，結果は明らかであり，精度が向上している．

の変動に頑健な正規化処理を加えていることから得られた結果である．CoHOG と比べても，格段に精度が上がっていることは自明である．CoHOG は方向ペアの数を成分として記述しているのに対して ECoHOG では共起特徴としてどの程度のエッジ強度成分が分布するかを特徴量として記述している．CoHOG から ECoHOG に精度が向上したということは数だけでは捉えられない特徴を，強度成分の累積により捉えられたということである．強度は存在するだけでなく，その成分も特徴として捉えるので全ての特徴を一律に扱う CoHOG とは本質的に違う．数により一律にヒストグラムに累積する手法 (CoHOG) では，背景や人物の服装・保持する物体のテクスチャに依存して特徴ベクトルの形状が大きく変わってしまう．一方で強度により累積する手法 (ECoHOG) では，弱い特徴は重要度の低い成分として累積されるので，人物検出の場合は服装や物体に内在する弱い特徴成分は重要度が低く，外輪郭や上半身・下半身の境目など重要な特徴を重要視して識別ができる．また，曲率や直線などにおいても度合を特徴として表現可能である．ECoHOG の強度累積では，同じようなエッジ方向成分を持つ物体でも，人物特有の強度成分や曲率・直線の度合を評価し検出性能を高められる．

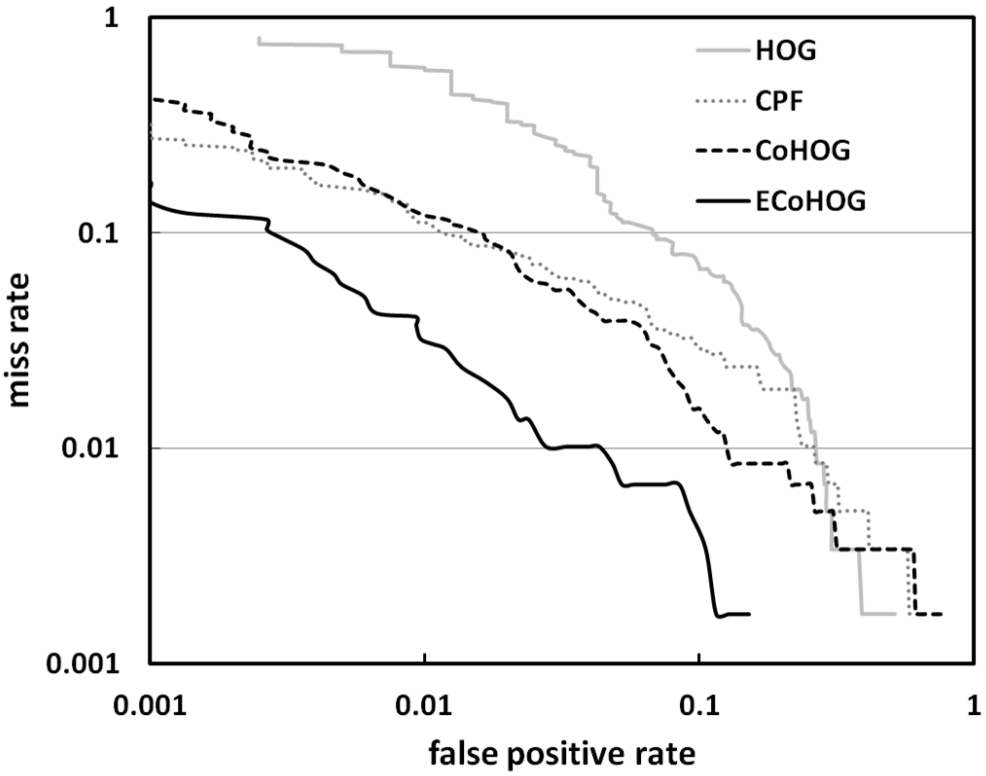

図 3.7　提案手法 (ECoHOG) と従来手法 (HOG, CPF, CoHOG) の DET カーブ：HOG から CPF/CoHOG には共起性の考慮により精度が上がっている．特に，図中の左側に顕著に現れているように過検出の低減が著しい．また，CoHOG からさらに ECoHOG の方が精度向上が見られた．エッジ強度の累積と正規化により，人物に特有のエッジ強度や曲率度合を表現可能であり姿勢変動や明度変化など見えの変化に一部対応したためである．

ここで，エッジ強度の累積による有効性，つまり CoHOG と ECoHOG の違いについて検証する．戦略として，(i) 見えの変動による 2 人の人物のヒストグラム解析，(ii) 人物と人物に似たエッジ成分を持つ背景との比較とする．(i) ではヒストグラム成分が類似する方が，(ii) ではヒストグラム成分が出来る限り離れている方が望ましい．図 3.8 と図 3.9 にそれぞれ ECoHOG と CoHOG の (i) における比較を，図 3.10 と図 3.11 にそれぞれ ECoHOG と CoHOG の (ii) における比較を示す．

図 3.8 は人物の服装テクスチャの違いによる ECoHOG の共起ヒストグラムの比較である．ECoHOG では強度により累積するため，重要度が高い (エッジ強度が強い) エッジの重要度が高くなる．人物検出では人物のシルエットに見られるエッジの重要度が高くなる．図 3.9 に示す二つの CoHOG ヒストグラムの距離よりも，ECoHOG のヒストグラムの方が近いことが分かる．

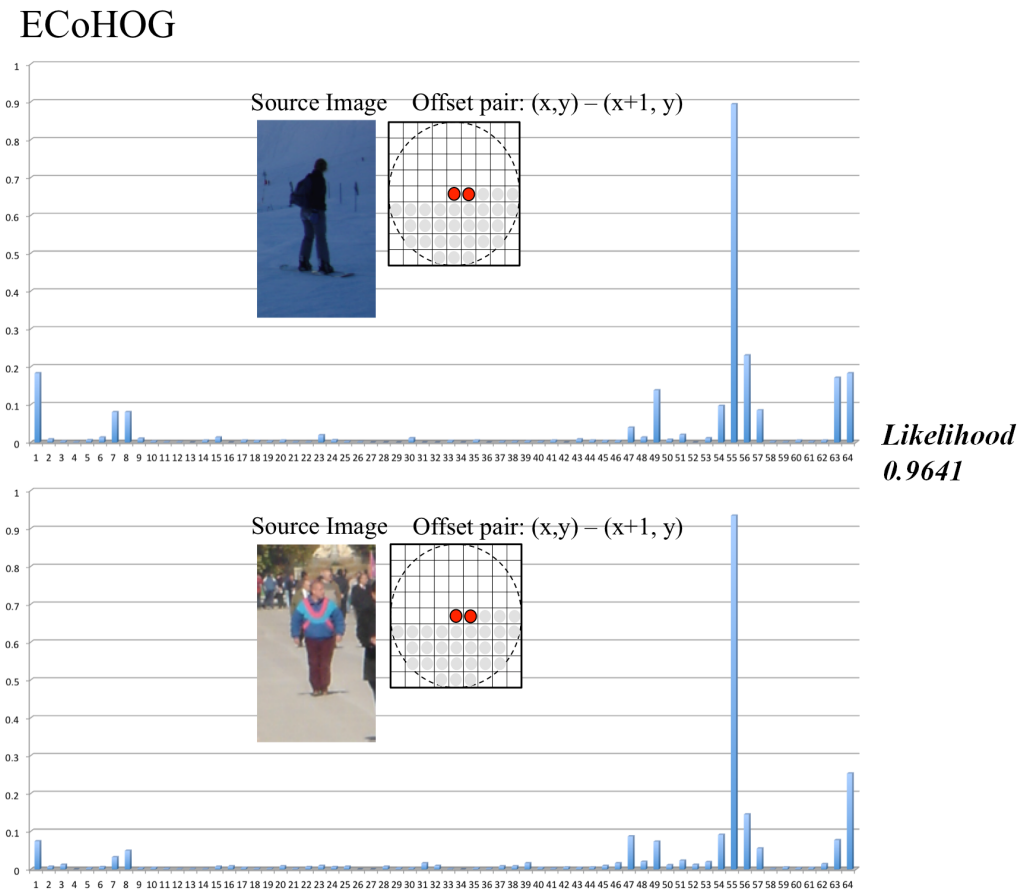

図 3.8　人物の服装テクスチャの違いによる ECoHOG のヒストグラム解析．横軸に 64 次元の共起ヒストグラムの要素，縦軸は 0.0〜1.0 で正規化されたエッジペアの出現確率．(上) 服装のテクスチャが比較的単純な人物 (下) 服装のテクスチャが比較的複雑な人物：ECoHOG では横軸 55 次元の位置にピークがあるが，CoHOG に比較してピーク値の差に変動が少ない．これは，形状ベースの人物検出では強い特徴とされる外輪郭形状を評価できたことによる．その他の成分についてもエッジ強度累積により重要度を評価しているため，変動が少ない．Battacharyya 係数によるヒストグラム類似度は 0.9641 である．

具体的には，ピーク位置 (共に 55 次元) での差分が小さいことやその他重要度の高い成分がほぼ均等にそろっている．一方で CoHOG では服装のテクスチャや背景に含まれる弱いエッジも一様に特徴として累積しているため，ノイズを含んだヒストグラムになってしまう．CoHOG と ECoHOG の人物と背景の成分についてもそれぞれ図 3.10 と図 3.11 に示す．背景については一概に ECoHOG の分離能力が優れているとは言い難いが，CoHOG に関して言えば，エッジが多ければ多いほど背景エッジを一様に捉えてしまうため ECoHOG に比べると安定しないヒストグラムになった．

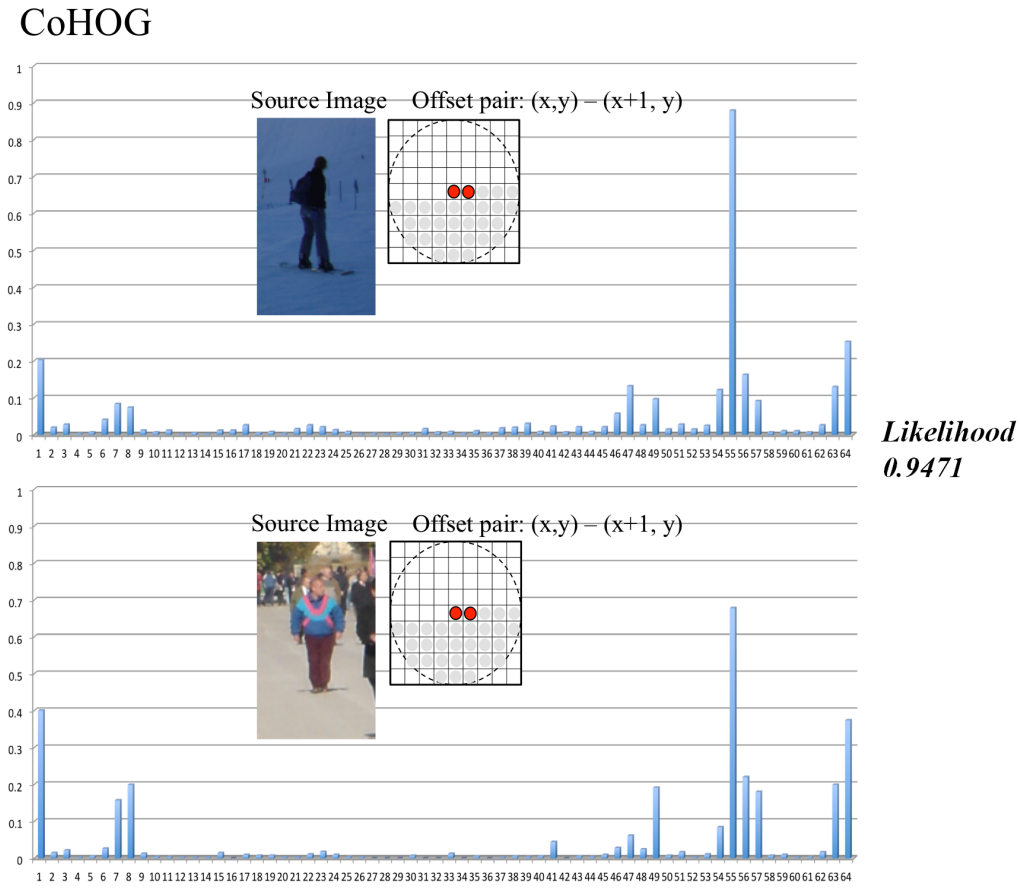

図 3.9　人物の服装テクスチャの違いによる CoHOG のヒストグラム解析．横軸に 64 次元の共起ヒストグラムの要素，縦軸は 0.0〜1.0 で正規化されたエッジペアの出現確率．(上) 服装のテクスチャが比較的単純な人物 (下) 服装のテクスチャが比較的複雑な人物：CoHOG では横軸 55 次元の位置にピークがあるが，ECoHOG に比較してピーク値の差が離れている．その他の成分についても，CoHOG はエッジ成分が存在するだけで同様に累積してしまうためテクスチャに依存して変化している．Battacharyya 係数によるヒストグラム類似度は 0.9471 である．

2 つのヒストグラムを判断する際には，ヒストグラムの類似度計算の指標として Bhattacharyya 係数を適用した [53]．Bhattachayya 係数はヒストグラムを正規化して類似度を計算するため，このヒストグラム計算では人物形状の類似度をエッジまでの距離によらずに計算することが可能である．類似度は 0.0〜1.0 で示される．ヒストグラムが一致すると 1.0 になる．Bhattacharyya 係数を以下に示す．

$$S = \sum_{u=1}^{m} \sqrt{h_u^1 h_u^2} \tag{21}$$

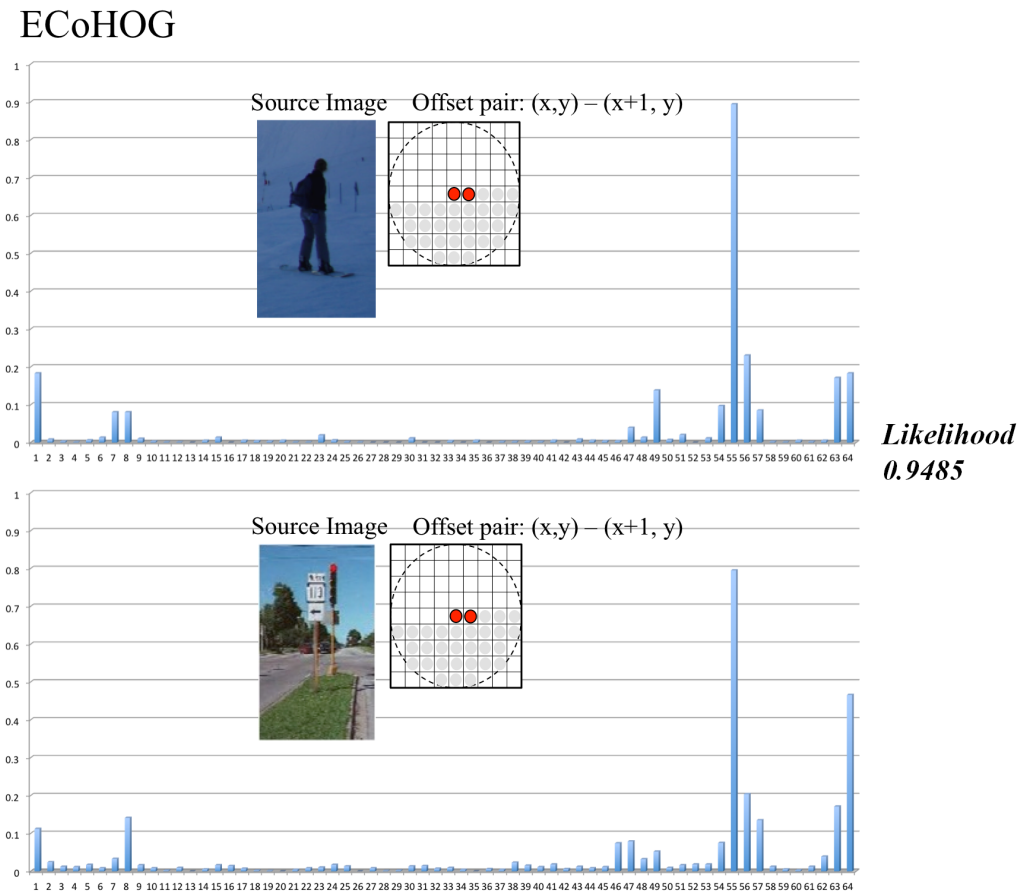

図 3.10　人物と類似する成分を保持する背景画像による ECoHOG のヒストグラム解析．横軸に 64 次元の共起ヒストグラムの要素，縦軸は 0.0〜1.0 で正規化されたエッジペアの出現確率．(上) 人物画像の成分 (下) 人物と類似する成分を保持する背景の成分：ECoHOG は似たような成分を保持する背景が入力された際，CoHOG に比較すると全体の値が抑えられている．強度の累積により重要度が低い (強度が弱い) 遠方に存在する背景エッジについて CoHOG に比較して排除することができる．Battacharyya 係数によるヒストグラム類似度は 0.9485 である．

$$\sum_{u=1}^{m} h_u^1 = 1 \tag{22}$$

$$\sum_{u=1}^{m} h_u^2 = 1 \tag{23}$$

S は Bhattacharyya 係数により得られる類似度，h^1, h^2 は縦軸に距離，横軸にウインドウの y 座標を示したヒストグラム，m はヒストグラムのビン番号を示す．Bhattacharyya 係数によるヒストグラムの比較には，正規化したヒストグラムを用いる．モデルとなる画像からモデルとな

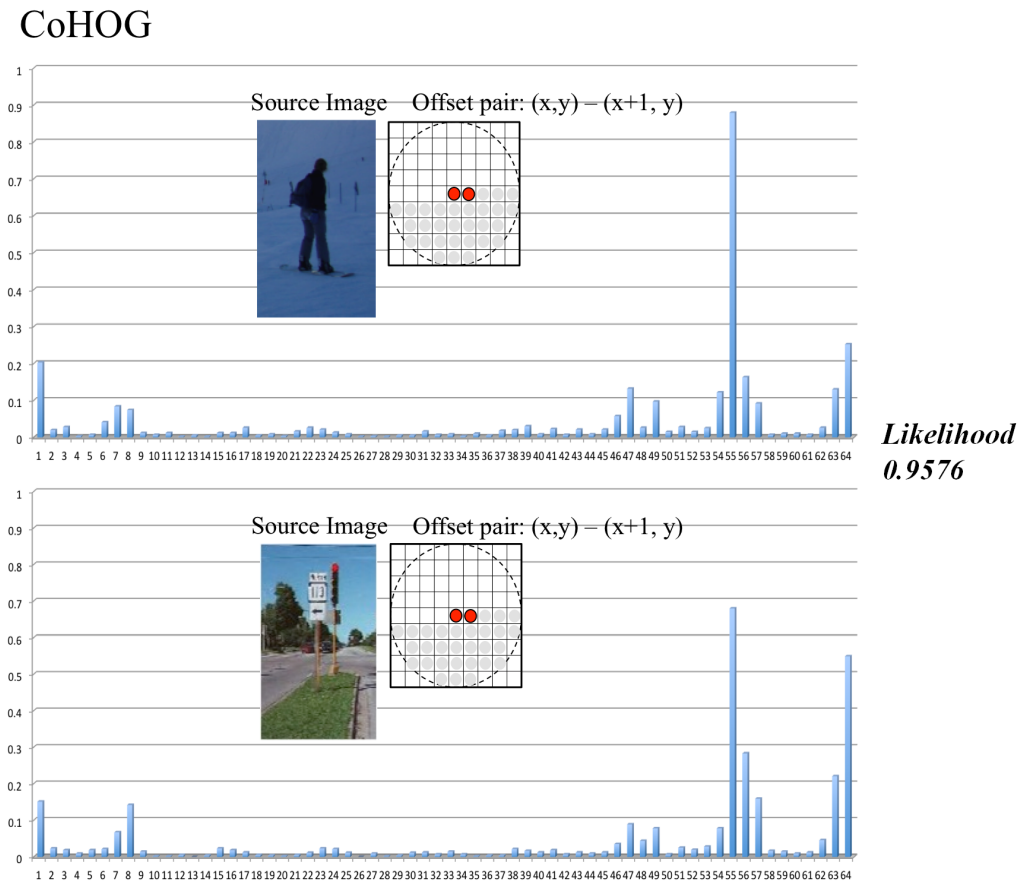

図 3.11　人物と類似する成分を保持する背景画像による CoHOG のヒストグラム解析．横軸に 64 次元の共起ヒストグラムの要素，縦軸は 0.0〜1.0 で正規化されたエッジペアの出現確率．(上) 人物画像の成分 (下) 人物と類似する成分を保持する背景の成分：CoHOG は似たような成分を保持する背景が入力された際，エッジの存在に起因してヒストグラム累積が大きくなってしまう．エッジが多ければ多いほど背景のエッジを一様に捉えてしまうため ECoHOG に比べると安定しないヒストグラムとなった．Battacharyya 係数によるヒストグラム類似度は 0.9576 である．

るヒストグラム h^1 を取得し，パーティクルの座標周辺から取得したヒストグラム h^2 と比べる．ヒストグラムのビンごとの乗算の平方を類似度としている．この係数は 2 つのヒストグラム h^1 と h^2 の分布が似ているほど大きな値となる．

それぞれの図を参照してみると，Positive 同士の ECoHOG ヒストグラム類似度 (図 3.8) は 0.9641，CoHOG ヒストグラム類似度 (図 3.9) は 0.9471 であった．また，Positive と Negative のヒストグラム類似度は ECoHOG が 0.9485(図 3.10)，CoHOG が 0.9576(図 3.11) であった．これらの結果から ECoHOG は人物画像同士のヒストグラムがより近く，人物と背景から取得す

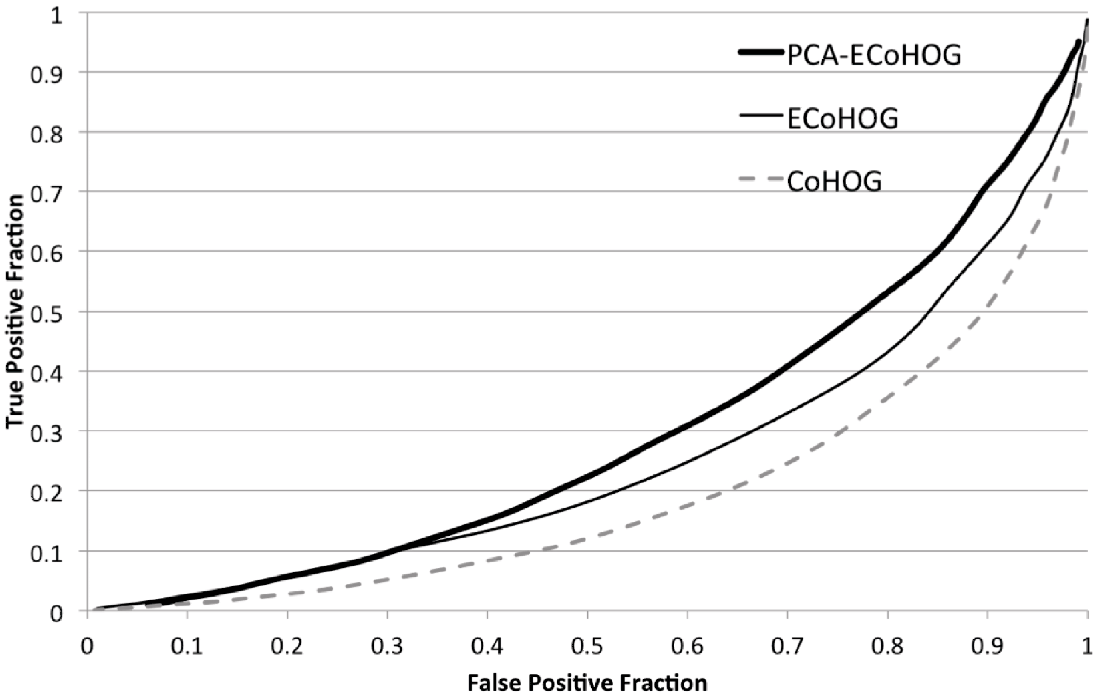

図 3.12　Daimler pedestrian benchmark dataset を用いた提案手法 PCA-ECoHOG, ECoHOG と従来手法 CoHOG の ROC カーブによる比較：ROC カーブはグラフの左上に近いほど高い精度を示す．提案手法の PCA-ECoHOG，ECoHOG，そして CoHOG の順番に精度が高いことを示した．Daimler のデータセットは車載映像中の人物検出であり非常に小さい人物画像からの検出であるが，提案手法は微小な強度を取得できていることや，特徴空間サイズの効果的な削減方法の設定が精度向上へ直結した．

るヒストグラム距離を遠ざけることから識別に有効な特徴量であると言える．

3.2.4　提案手法の汎用性検証

INRIA person dataset だけでなく，Daimler pedestrian benchmark dataset でも人物検出精度を比較した．Daimler pedestrian benchmark dataset を用いて PCA-ECoHOG，ECoHOG と CoHOG を比較している結果を図 3.12 に示す．ここで，評価には ROC カーブを用いている．ROC カーブは縦軸に検出率を，横軸は過検出率を示すためグラフの左上に近いほど高い性能を表している．よって，PCA-ECoHOG，ECoHOG，CoHOG の順に性能が良いということになる．ここで，PCA-ECoHOG や ECoHOG が検出に成功して，CoHOG のみが未検出であった画像の例を図 3.13 に示す．図 3.13 から，CoHOG では姿勢変動した場合や人によってオクルー

図 3.13　ECoHOG や PCA-ECoHOG への改善で検出できた Daimler pedestrian benchmark dataset の画像例：CoHOG では姿勢変動した場合や人によってオクルージョンが発生している場合，複雑な背景状況，傘など物体によって人物形状の一部が変更されている場合などに未検出であった．

ジョンが発生している場面，複雑な背景状況，傘など物体によって人物形状の一部が変更されている場合などに未検出であった．Daimler pedestrian benchmark dataset は遠くの人物でも検出できるよう，画像サイズが 18×36 ピクセルと非常に小さい．そのため，特徴量取得が非常に困難な状況ではあるが，そんな中でも PCA-ECoHOG や ECoHOG は人物を正しく検出できている場面が多い．これは，画像がぼけている場面においても，画像空間内に広がった人物特有のエッジ強度情報を捉えているからである．CoHOG ではどの程度のエッジ方向ペアが存在するのか成分として累積しているため，車載カメラより撮影しブラーを含んでいる不鮮明な画像に対しては正確にエッジ方向成分が取得できていなかった．人物と背景の境目が曖昧になりエッジ方向成分にずれを含んでいるため識別性能が低下した．一方で強度による累積ではブラーにより拡散しているものの強度情報を捉えられているため精度が向上している．また，PCA-ECoHOG の方が精度の面で上回っているのは，次元圧縮による効果であると考える．一般的に次元を効果的に削減した場合には少ないサンプルでの学習ができる上に，特徴空間内での正解/非正解の分離が容易である．同じ学習枚数でも次元が少ない方が有利である．

3.2.5　処理時間

CoHOG，ECoHOG，PCA-ECoHOG の処理時間の面について評価した結果も表 3.1 に示す．処理時間の評価は 1 フレームあたりの平均時間で示している．CoHOG が最も早い処理時間で特徴ベクトルを抽出できるが，ECoHOG も同程度の処理時間である．CoHOG ではエッジ方向

ペアのカウントであったが，ECoHOG ではエッジ強度ペアの累積であるだけでなく，正規化の処理を加えているためわずかながら処理時間が遅くなっている．また，PCA-ECoHOG では特徴ベクトルを抽出した後に固有ベクトルにより低次元空間へ射影している．行列計算の分だけ処理時間が遅くなっているが同程度での処理が可能であるといえる．

表 3.1　局所特徴量 CoHOG，ECoHOG，PCA-ECoHOG の処理時間比較：1 フレームあたりの平均処理時間

局所特徴量	処理時間 (msec)
CoHOG	49.59
ECoHOG	51.67
PCA-ECoHOG	59.51

3.3　本章のまとめ

本章では人物検出の有効性を示すため，提案した局所特徴量の設定を決定する実験や関連研究との比較を行った．実験のデータセットには INRIA person dataset や Daimler pedestrian benchmark dataset を適用した．INRIA person dataset を適用した提案手法の設定の決定においては, 過検出を減らす特徴として和算によるヒストグラム累積が有効であることや, 圧縮次元数においては 100 次元が情報量と特徴空間サイズを保つ上でバランスの良い次元数であることを示した．INRIA person dataset や Daimler pedestrian benchmark dataset を用いた関連手法との比較では，ECoHOG が最も高い性能を示した．共起性を考慮することで HOG から CoHOG が性能向上し，エッジ強度の累積やヒストグラムの正規化により CoHOG から ECoHOG に精度が向上している．ECoHOG から主成分分析により次元圧縮した PCA-ECoHOG にしたことによりさらに識別性能が向上していることも実験により分かった．CoHOG から ECoHOG, PCA-ECoHOG にしたことにより人物の強度情報を含めた表現ができるようになり姿勢変動, 部分的なオクルージョン，複雑背景や物体を保持している場合などの場面において検出できるようになり，精度が向上した．

4 提案技術を適用した人物行動解析への応用例

本章の概要

本章では，提案した手法を適用して歩行者予防安全のための検出，サッカー映像解析のための複数選手追跡や行動理解のための応用についての検討を行う．歩行者予防安全では車載カメラで撮影された前景映像から歩行者を検出する．移動する自車両から撮影する上に複雑背景であることなどが課題として挙げられる．複雑な背景や特徴記述能力の高い ECoHOG を適用することで困難な場面においても検出精度が向上することが見込まれる．サッカー映像解析では複数選手の追跡が非常に重要なテーマとなっているが，オクルージョンへの対応策が課題となっていた．単独のサッカー選手は色ヒストグラムを特徴とした Particle Filter を適用して追跡するが，選手の重なり発生時にはより接近した状態でも検出できるように ECoHOG 特徴量を用いて検出，選手の速度を考慮して重心の再配置を行う．行動理解では身体から取得する形状の精密な記述が重要となっているが，共起性を考慮した特徴記述により特徴記述能力を高められると考える．

4.1 予防安全のための歩行者検出と追跡

日本の交通事故による年間死亡者数は 1970 年の交通事故死者数 16,765 人を境に減少し，2013 年では 4,373 人と最近は 10 年以上連続で減少傾向にある (図 4.1)．しかし，死亡者数に占める歩行者の割合は 3 割強と近年増加の傾向にある [60] (図 4.2)．日本政府は 2018 年までに交通事故死亡者数を 2,500 人以下とする目標を掲げており [61]，今後政府目標を達成するためにも高度道路交通システム (Intelligent Transport Systems: ITS) にフォーカスする必要がある．ITS は主に情報通信技術を用いて自動車をインテリジェント化する技術であり，今後は歩行者への更なる安全対策が必要不可欠である．歩行者事故への対策としては，大きく分けて「衝突安全」「予防安全」と 2 種類の安全技術が挙げられる．以下にそれぞれの特性を記載する．

- **衝突安全 (Crash Safety)**：衝突時の歩行者や自動車乗員の安全確保を主な対象とする．衝突時の衝撃吸収機能の搭載や歩行者との激突衝撃低減，乗員の生存空間の確保などを考慮して自動車の車体を設計した．車内のエアバッグや衝撃を吸収する車体がその一例で

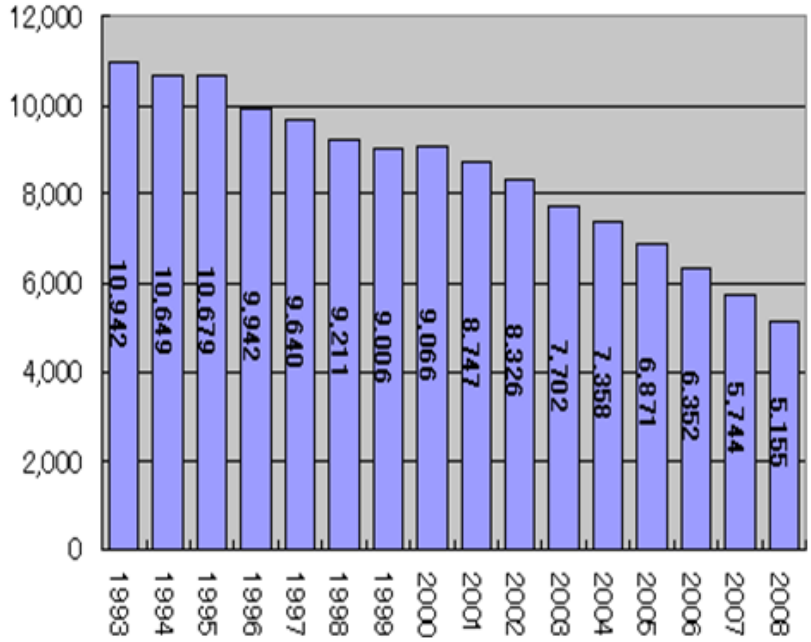

図 4.1　交通事故死亡者数の推移：1993 年から 2008 年の推移を示しており，全体としては減少傾向にある [60]．2013 年の統計でも 4,373 人と 13 年連続で減少となった．

ある．

- **予防安全 (Active Safety)**：自動車と歩行者または周囲環境に対して事前に危険を検出し，衝突を回避する概念である．大きな事故要因のひとつである人間の認知のミスを補うために研究開発されてきた．センサとして主にレーザやステレオカメラを用いたシステムが使用される．

1993 年の「道路運送車両の保安基準」 [62] が改訂され，前面衝突試験が義務付けられたことがきっかけとなり，衝突安全への意識は高まり，その後開発が進められた．しかし，ITS の研究は衝突安全の限界や情報通信技術の発展と共に衝突を回避する可能性を探り，衝突安全から予防安全へと主な研究対象が移り変わることになる．事故が起きてからの衝撃を低減するよりも「ぶつからない」車を開発する方が交通事故による死者数を圧倒的に減少させることが可能であるため，今後は予防安全の研究開発が必要不可欠である．歩行者に対する予防安全では，歩行者を自動車のセンサで検知し，衝突前に警報および制動制御をかける「歩行者事故を未然に予防するための安全装置」の普及が有望視されている．前方障害物を検知する装置には 2 種類，レーザ方式とステレオカメラ方式が存在する．歩行者予防安全システムで頻繁に用いられている，レーザ方

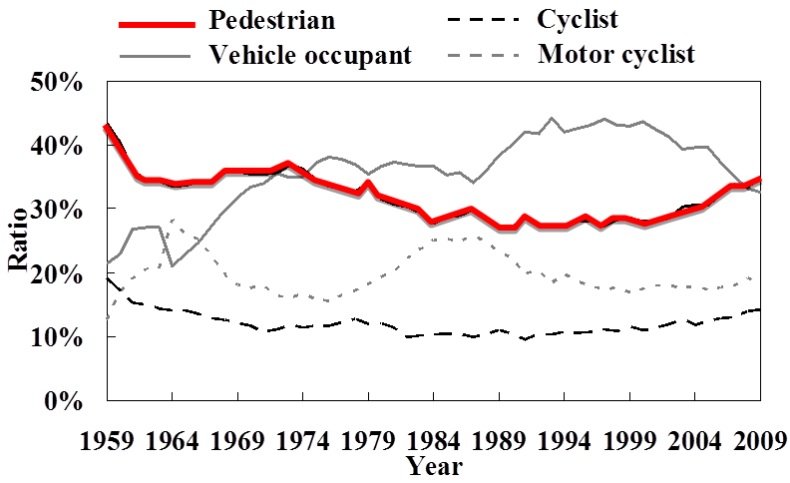

図 4.2　交通事故死亡者数に占める各割合 (歩行者，自転車，車内，オートバイ) [60]：一時は減少傾向にあった歩行者も，近年では全体の約 35% と高い傾向にあるため，歩行者への益々の対策が課題となる．

式とステレオカメラ方式について以下に記載する．

- **レーザ方式**：測定の原理はレーザを照射し，物体に反射して戻ってくるまでの時間 (Time of flight) から物体までの距離を測定する．距離画像が取得できる．車両の周囲環境の認識等にも用いられている．
- **ステレオカメラ方式**：ステレオカメラ方式では基本的にカメラの高さが同じであるという前提があるので，画像を平行にスキャンするだけでステレオマッチングができ，距離画像を得ることができるので非常に効果的である．その上，可視画像を認識に用いることも可能であるため，現在最も有力な手法として期待されている．

富士重工業 (SUBARU) は，ステレオカメラを用いた運転支援システム Eye Sight(アイサイト)® を提案している (図 4.3) [63]．Eye Sight® は，ステレオカメラにより車両前方を監視するシステムであり，障害物までの距離を計測し，危険と判断した際に自動ブレーキによって車両を停止させる．車両と対象物との速度差が約 30km/h 以下の状況では，自動ブレーキによって衝突の回避・衝突被害の軽減を図る．速度差が約 30km/h を超える状況では，自動ブレーキによって減速し，衝突被害の軽減を実現している．その他，予防安全機能として，前の車両に近づきすぎると警告する車間距離警報や，車両が車線から外れそうになると注意を促す車線逸脱警報

図 4.3　SUBARU Eye Sight®：屋内のリアカメラの上部に取り付けたカメラにより車両の前方を撮影して歩行者や障害物を捉える．Eye Sight では歩行者の衝突回避機能だけでなく，車線逸脱の防止や前方車両の自動追尾機能も搭載する．

やふらつきを検知し警告するふらつき警告などが搭載されている．

近年では Eye Sight のようにステレオカメラ方式で予防安全システムを構成できることも分かっており，コンピュータビジョン技術による歩行者検出が必要である．他方，自動車業界では歩行者事故予防安全の普及につながる手段として，自動車にカメラを搭載する傾向にある．タクシーやバスなど商用車にドライビングレコーダの搭載が増加の一途にあることや，米国運輸省道路交通安全局 (National Highway Traffic Safety Administration: NHTSA) が 2014 年までにリアカメラの搭載を義務付けるなどカメラを車両に設置する傾向は今後さらに増加すると考えられる．このような背景からも，コンピュータビジョン技術による歩行者検出技術の向上は必須であると言える．

ここで，コンピュータビジョン技術を適用したセーフティシステムについて説明する．図 4.4 に想定するシステムの概要を示す．まず，車内に設置した単眼カメラにより車載映像を取得する．映像中では歩行者の位置を確認するために検出処理を実行する．本論文で取り扱う技術としては歩行者の検出であるが，その後ドライバーへの注意喚起や自動でブレーキを制御することが考えられる．

歩行者予防安全の目的は単眼カメラより撮影された車両前方の映像から歩行者を検出することである．提案手法の流れを図 4.5 に示す．まず，前処理として歩行者の左右画像から対称性判断を行い，歩行者対象領域を大まかに絞り計算コストを削減する．歩行者の検出方法としては，機械学習による識別器により画像中を探索する．識別器を生成するための特徴量としては，ECoHOG を適用する．歩行者を検出するだけでなく，時系列で検出結果をつなぎ合わせ

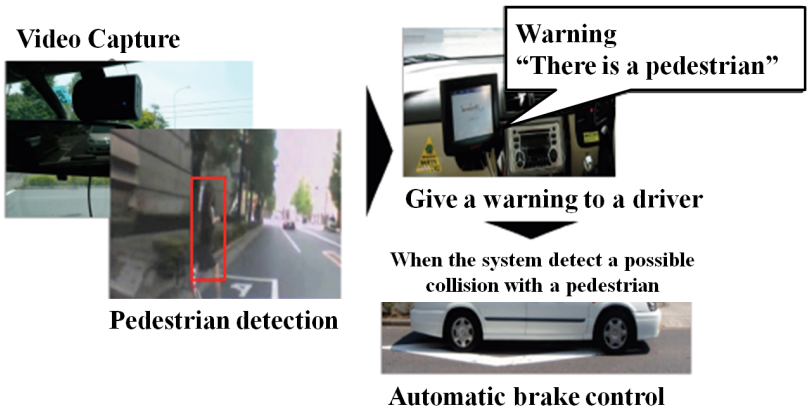

図 4.4　カメラを用いたセーフティシステム：車内に取り付けたカメラにより歩行者を検出する．自車両の速度情報や歩行者の位置から衝突判定をし，ドライバへの注意喚起を行う．衝突する可能性が高いと判断された場合にはオートブレーキシステムを制御する．

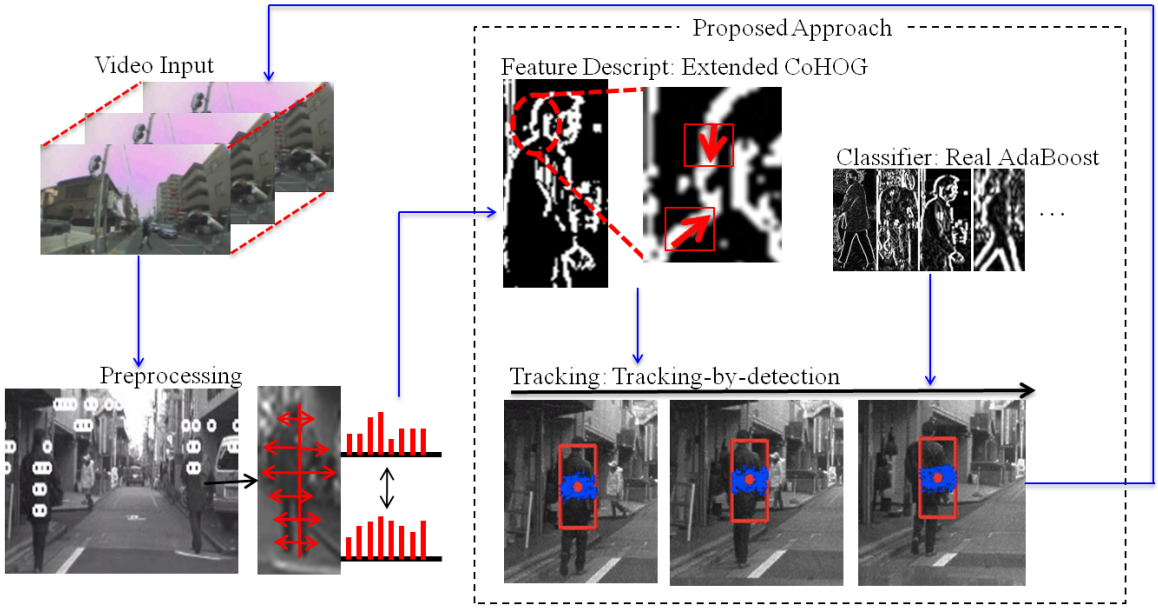

図 4.5　提案手法：歩行者の左右対称性に着目した前処理により，歩行者の候補領域に処理を制限する．歩行者候補領域上では局所特徴量である ECoHOG を適用して歩行者を適用する．その後，Tracking-by-detection の概念により歩行者を追跡する．

る Tracking-by-detection を用いて歩行者追跡にも取り組む．Tracking-by-detection の枠組みには Particle Filter[50] を適用する．Tracking-by-detection にも ECoHOG を適用するが，さらに車両の移動を考慮した状態推定モデルも提案する．

4.1.1 歩行者検出のための前処理

主に，歩行者検出のための前処理について述べる．処理領域を制限するための前処理については歩行者の形状の左右対称性に着目してエッジを評価している．制限された処理領域上においては前述の ECoHOG と Real AdaBoost[12] を適用して最終的な位置を確定する．

まずは画像から輪郭情報を取得するためにエッジ検出を行う．画像の濃淡値が急激に変化する部分を検出するため，たとえば物体と物体の境目などがエッジとして検出される．ここで，エッジ検出には Sobel フィルタを適用した [51]．x, y 方向のエッジ検出が可能で，直立する人物の輪郭をはっきりと抽出するために x 方向のエッジ抽出を試みた．また，Sobel フィルタはノイズを抑える効果があることでも知られ，より滑らかなエッジが抽出できる．エッジ抽出後はエッジ強度が画素値として取得できる．ここではさらに，エッジ強度が高い部分を取り出すために二値化を試みた．二値化には自動閾値決定法として判別分析法 (大津の二値化) を用いている [52]．判別分析法はクラス内とクラス間の分散から閾値を決定する方法であり，クラス内を最小に，クラス間の分散を最大にするような閾値を決定することで，最適な閾値を決定する．画像の入力からエッジ検出，二値化までを図 4.6 に示す．

判別分析法により得られた二値画像から，エッジ形状を検査するために画像をラスタスキャンする．ラスタスキャンする際にはウインドウを用意し，縦から 1 画素ずつ左右にスキャンする (図 4.7)．このとき，頻度情報ではなくエッジまでの距離をヒストグラムとして記録する．左右 2 つのヒストグラムを判断する際には，ヒストグラムの類似度計算の指標として前述の Bhattacharyya 係数を適用した [53].

計算した類似度に関して，閾値処理により歩行者候補位置を決定する．ここでは，正解画像 (人物画像) と非正解画像 (歩行者以外の背景領域) から類似度を計算して，統計的に閾値を決定する．対称性判断の閾値は，歩行者画像と背景画像から決定する．図 4.8 は歩行者画像 3,000 枚，図 4.9 には背景画像 20,000 枚の類似度 (横軸) と頻度 (縦軸) を示すヒストグラムである．ここで，歩行者画像全ての類似度 (対称度) の平均は 0.88，分散は 0.006 であった．また，背景画像の最大頻度位置は類似度が 1.0 のビン位置であった．これは，人工物に見られる，対称性のほぼ一致する画像であると思われる．歩行者画像の分布より，全体の 95% (2 σ:95.45%) を許容する閾値として対称度 0.77〜0.98 を採用した．歩行者の平均値 0.88 から分布を考慮して歩行者の見落としを少なく，さらには背景画像の最大頻度位置である 1.0 を除外し，出来る限り背景を処

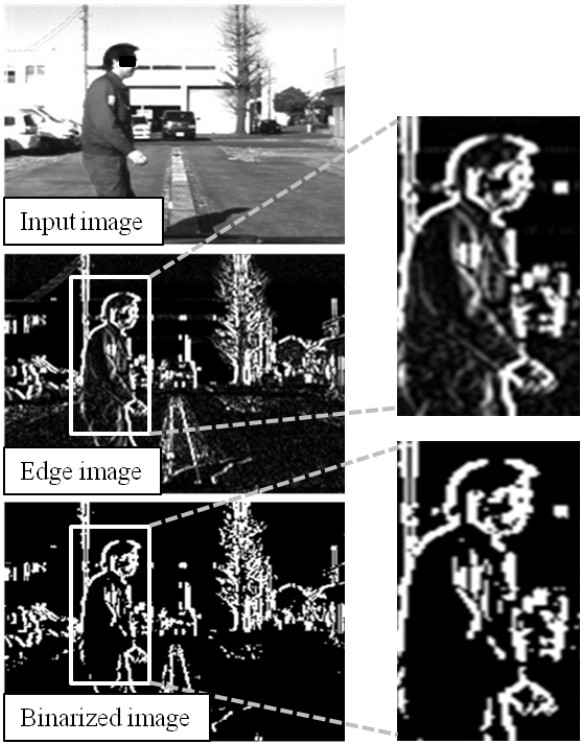

図 4.6　原画像とエッジ検出，二値化：エッジ検出には Sobel フィルタ，二値化には判別分析法を適用.

理領域として含まないよう閾値設定にしている.

歩行者候補位置の絞り込みの結果を図 4.10 に示す．図中の小円で示される位置が歩行者候補位置であり，候補位置において局所特徴量を適用した検出処理を実行する．正解・非正解画像から閾値を取得しているため，歩行者を捉えられる，かつ背景を出来る限り除去できるような閾値に設定した.

4.1.2　歩行者追跡

映像から歩行者を追跡する技術について記述する．Tracking-by-detection はフレーム間の追跡対象位置を対応付ける処理であり，状態推定モデルや尤度計算を設定する必要がある．状態推定モデルには前方映像からフローを取得し，自車両の運動を近似する．尤度計算には前章で提案した ECoHOG を適用して，重み付き平均により歩行者の位置を正確に捉える.

Tracking-by-detection．Tracking-by-detection は推定ベースの追跡器において局所特徴量を用いることで強力な追跡手法に強化している．Tracking-by-detection の追跡手法には，

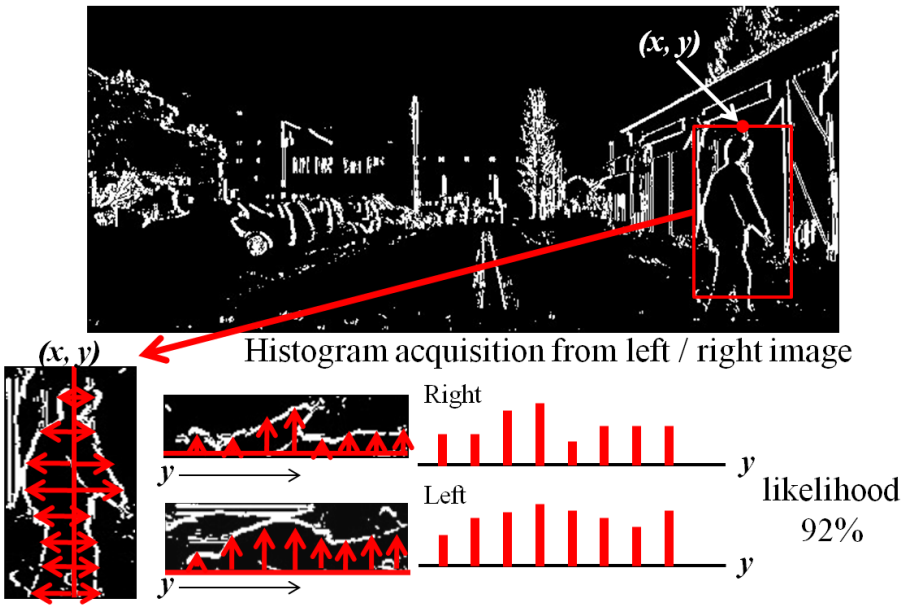

図 4.7　二値画像の走査と対称性判断：二値画像上はウインドウのラスタスキャンにより走査する．ウインドウ内は 1 画素毎に画像上からスキャンし，ウインドウ中心から左右のエッジまでの距離をヒストグラムの縦軸として記録する．左右から取得したヒストグラムの類似度を比較することにより左右の対称度を得る．

Particle Filter [50] を用いる．Particle Filter は，コンピュータビジョンの分野において物体追跡手法として広く用いられている．Particle Filter は推定に基づく時系列追跡手法であり，尤度と状態推定モデルを持つパーティクルを多数用いて追跡を行う．パーティクルがそれぞれ尤度を観測する位置であり，尤度が高い方向にパーティクル群を移動させて物体を追跡している．尤度観測方法や状態推定モデルを自由に設定できることや，多数のパーティクルを配置するので，複雑な動作や姿勢の変動がある場合にもロバストな追跡が可能であると言われている．Particle Filter の概念図を図 4.11 に示す．Particle Filter は初期位置を設定後，状態推定モデルに従ってパーティクルを遷移させる．パーティクルがひとつの観測位置となっており，対象物体周辺のパーティクルほど尤度が大きくなるため，対象物体の位置を確定できる．Particle Filter ではランダムサンプリングにより確率密度を求める際にも全探索する必要がなく，高速な処理を実現できる．

Particle Filter の流れを以下に概説する．

Step1 初期設定：対象物体の位置を指定し，複数個のパーティクルを発生させる

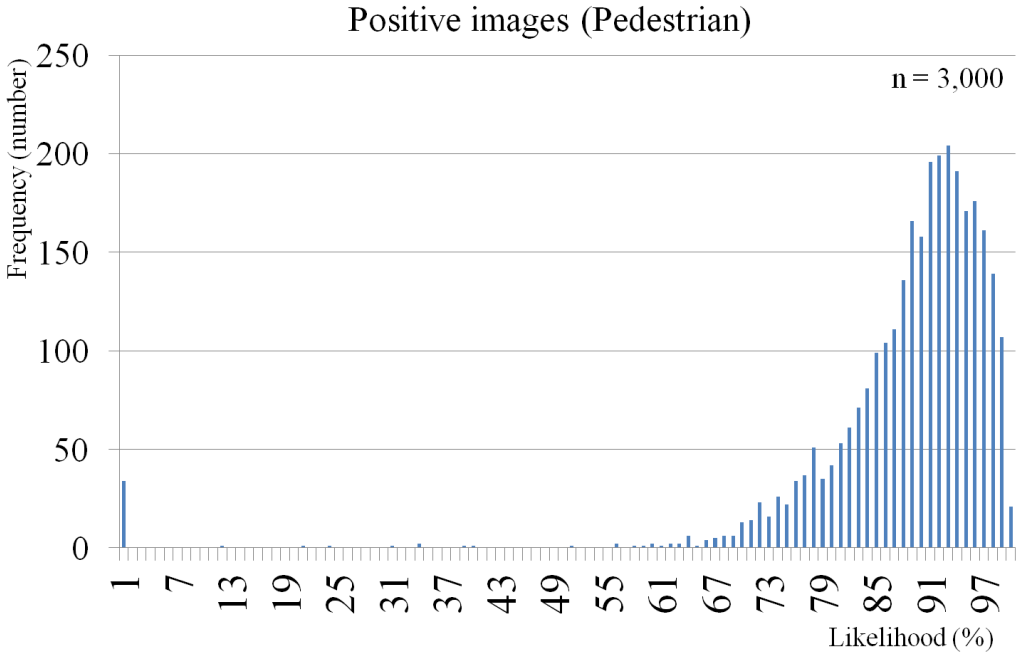

図 4.8　正解画像 3000 枚から取得した対称性：歩行者から取得した対称度の平均値は 0.88, 分散は 0.006 であった.

Step2 状態推定：設定した運動モデルに従いパーティクルを移動させる

Step3 尤度観測：パーティクル位置において特徴量を取得し，あらかじめ用意した対象物体のモデルとの比較により尤度を計算する

Step4 尤度評価：パーティクルの尤度を正規化し，対象物体の重心を求める

(Step1 - Step3 の繰り返しにより物体を追跡する)

以下に，本論文で取り扱う歩行者追跡の詳細を記述する.

Step1:初期化. 検出した歩行者位置を対象にして，Particle Filter のトラッカーを配置する. 検出した歩行者の周辺にパーティクルを，ガウス分布を考慮して散布する.

Step2:状態推定. 次フレームの歩行者位置を推定してパーティクルを動かす. 状態推定モデルには画像中から得たオプティカルフローを用いて，移動方向や速度を推定する. 車載カメラから得た画像の前後移動量 (図 4.12) により，車両の移動量や方向がわかることから，毎フレームの状態推定モデルとする. ここでは画像中から取得したすべてのオプティカルフローの平均を状態推定モデルとする. さらに，速度にランダムノイズを付加することにより，歩行者や車両の不

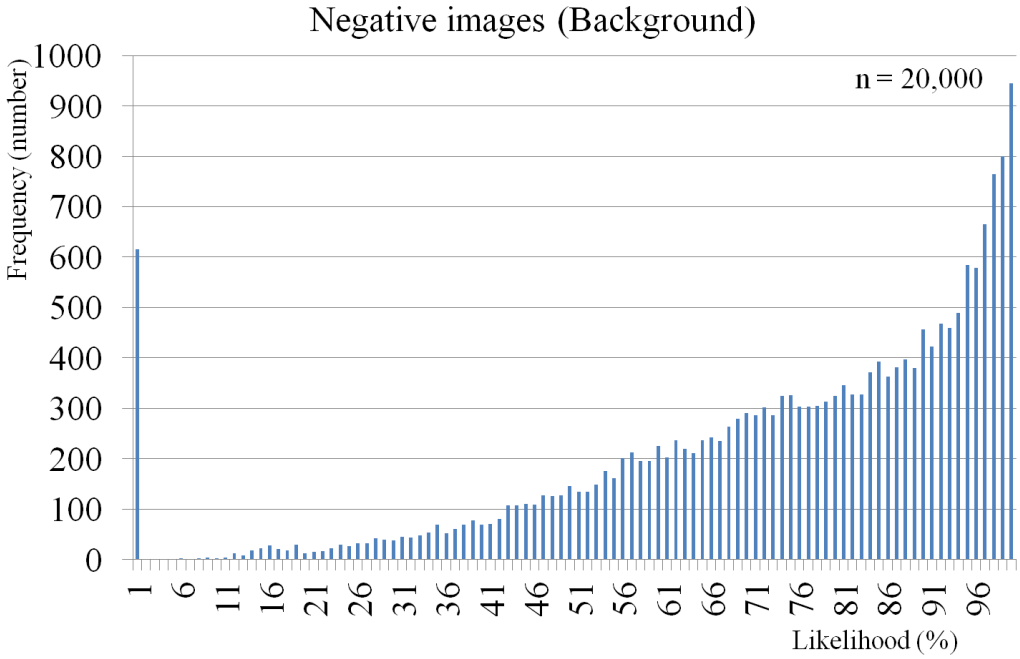

図 4.9　非正解画像 20000 枚から取得した対称性：背景から取得した最大頻度位置は類似度が 1.0 のビン位置であった．人工物に見られる，左右対称性がほぼ一致する画像が多かったためである．

規則な移動にも対応できると考える．また，画像の左右でフローは逆方向になることから，画像中心から左右別々の状態推定モデルを施している．フローが少ないとき，つまり車両の移動量が小さい時や停車している際には等速直線運動に切り替える．運動モデルは以下に示される式により計算されている．

$$X_{t+1} = FX_t + w_t \tag{24}$$

$$X_t = (x, y, v_x, v_y)^T \tag{25}$$

$$F = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{26}$$

x と y はパーティクルの位置，v_x，v_y はフレーム間の速度である．v_x，v_y についてはオプティカルフローにより取得した速度を適用する．w_t はパーティクルに付加するノイズであり，x，y 方向の位置や速度にガウシアンノイズを与えている．

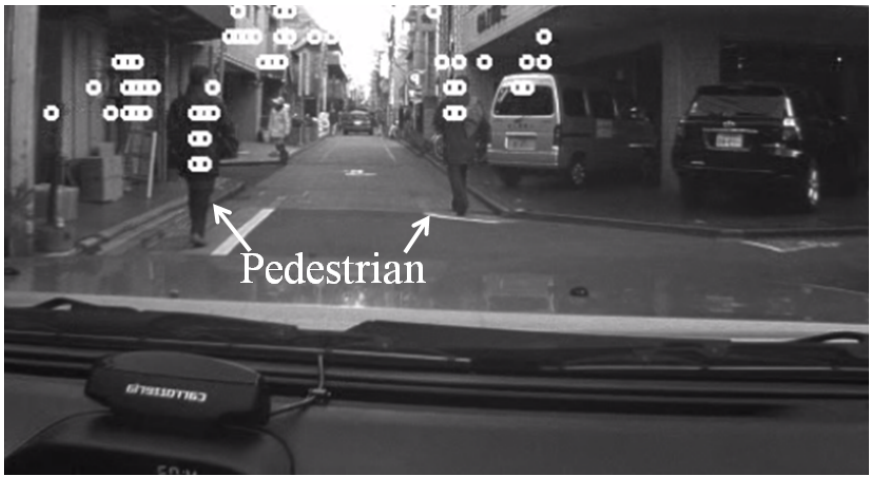

図 4.10　実環境で対称性判断を実行した例：学習画像から閾値を決定し，歩行者の約 95% を受け入れ，背景を出来る限り排除できる閾値に設定しているため左右の対称性から歩行者を候補領域として抽出できている．

Step3:尤度計算．歩行者追跡において，歩行者モデルの作成は重要な課題であった．ここで，我々は歩行者検出にて用いた ECoHOG を用いた尤度計算により追跡精度向上を達成できると考える．機械学習手法 Real AdaBoost との組み合わせにより強力な識別器を作成する．Real AdaBoost [12] は AdaBoost [102] を発展させた学習方法であり，多数の画像を学習することにより複数の識別器を生成する．この学習により得られた識別器を組み合わせて学習するため，対象物体を高精度に検出する手法として知られている．一般に，Boosting では正解クラスと非正解クラスの境目，つまり誤検出の可能性がある部分に対して重みを高くし，検出の誤りを最小にしている．さらに，AdaBoost と Real AdaBoost との違いは，識別器の評価値の出力である．AdaBoost では，閾値処理により 0 か 1 であったが，Real AdaBoost では確率密度により正解/非正解の確からしさを計算する．効果的に評価値を算出することで，さらに高精度な識別器が生成可能である．ここで，著者は Real AdaBoost の出力値を尤度として利用している．以下に Real AdaBoost の出力値を適用した尤度計算を示す．

$$h(x) = \frac{1}{2} ln \frac{W_{positive} + \epsilon}{W_{negative} + \epsilon} \tag{27}$$

$$H(x) = \Sigma_{t=1}^{T} h_t(x) \tag{28}$$

ここで $h(x)$ と $H(x)$ はそれぞれパーティクル上の観測と，Particle Filter 全体としての尤度を

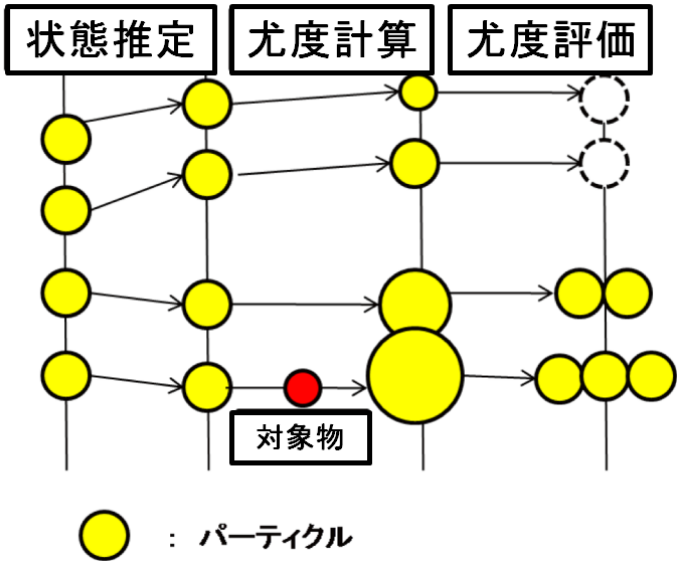

図 4.11　パーティクルフィルタの概念図：パーティクルとは観測点のことであり，適用した状態推定モデルを基にフレーム間でパーティクルを遷移させる．パーティクル上では尤度を計算し，尤度評価では計算した尤度とパーティクルの座標から重み付き平均により対象物の重心を求める．

図 4.12　車載カメラからのフローベクトル取得

示す．$W_{positive}$ と $W_{negative}$ は Real AdaBoost の弱識別器の出力値である．ϵ は学習サンプル数 N の逆数 $\frac{1}{N}$ に設定する．識別器を用いた尤度計算を図 4.13 に示す．パーティクル位置周辺の画像を切り出し，特徴量を取得する．あらかじめ用意した ECoHOG + Real AdaBoost の識別器を用いて評価を行う．識別器の出力を 0.0 - 1.0 に正規化して歩行者の尤度とする．

Step4:尤度評価．Particle Filter の位置情報と，パーティクルフィルタ上で計算した尤度を

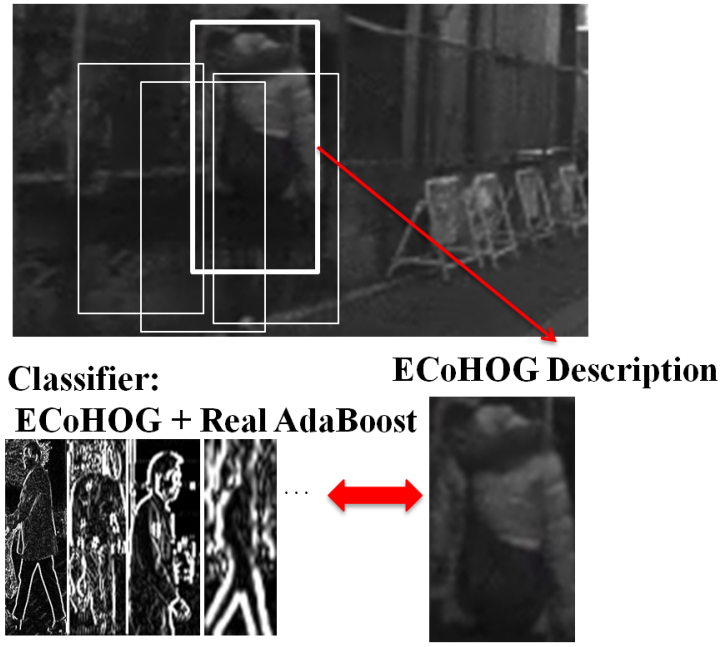

図 4.13　識別器を用いた尤度計算：パーティクル周辺から矩形領域を切り取り ECoHOG を抽出する．抽出した特徴ベクトルは統計的に学習して生成された識別器と比較して尤度を計算する．

用いて歩行者の重心を決定する．重心は以下に示す式で計算する．

$$(g_x, g_y) = (\Sigma_{i=1}^{n} L_i x_i, \Sigma_{i=1}^{n} L_i y_i) \tag{29}$$

g は歩行者の重心，L は歩行者の尤度，x，y はパーティクルの位置を示す．

4.1.3　歩行者検出・追跡実験と考察

まず，歩行者検出と追跡実験で使用する実際の路上で撮影した映像データについて解説する．撮影環境と動画の設定を表 4.1 に示す．東京都内の複数個所で撮影した映像に対して本手法を適用して検出精度や追跡精度を実証する．今回は約 2 時間の動画像から人物が存在するフレームを切り出して処理を施した．画像例を図 4.14 に示す．カメラの設置位置と撮影して得た画像を，図 4.15 に示す．

実際の路上で撮影された車載映像を対象として，人物検出の精度を検証する．検出精度の指標としては，適合率 (Precision)，再現率 (Recall)，F 値 (F mesure) を用いている．ここで適合率はシステムの正解率を，再現率は実際の正解値のうち，システムの検出率を示す．F 値は適合

図 4.14　路上で撮影した映像データの例

率，再現率の調和平均により算出している．適合率，再現率，F 値の計算式を以下に示す．

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{30}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{31}$$

$$Fmeasure = \frac{2 * Recall * Precision}{Recall + Precision} \tag{32}$$

ここで，システムの出力値と実際の正解値の関係においては，人物をシステムが人物と判断することを True Positive，人物をシステムが背景と判断することを False Negative と呼ぶ．また，背景をシステムが人物と判断することを False Positive，背景をシステムが背景とすることを True Negative と呼ぶ (表 4.2).

ここでの特徴量や人物を検出するための設定方法については上記実験と同様であるが，人物を

表 4.1　路上で撮影した映像データの撮影環境と動画の設定

撮影場所	調布，三鷹，小金井
カメラの高さ	地上 124 (cm)
撮影日時	2010 年 2 月 4 日 19:00 - 21:00
天候	晴
フレームレート	30.0fps
画像サイズ	640 × 480 ピクセル

図 4.15　カメラ位置と撮影された画像：車載カメラはルームミラーの上部に取り付け前方を撮影する．

表 4.2　評価する際のラベル名

		Actual class	
		Pedestrian	Background
Output	Pedestrian	True Positive	False Positive
	Background	False Negative	True Negative

検出する際，画像を比較する基になる人物のモデル画像として 5,000 枚の正解画像，20,000 枚の非正解画像，計 25,000 枚の画像を用意した．評価用画像は人物が存在する画像を車載映像から抜き出した．画像は我々が東京都内で撮影した画像である．正解画像の一部を図 4.16 に，非正解画像の一部を図 4.17 に示す．また，評価用画像から人物を検出し，従来手法 (CoHOG) と提案手法 (ECoHOG) を使った人物検出の適合率，再現率，F 値で評価した結果を表 4.3 に示す．

図 4.16　正解画像：路上から撮影した映像データセットから手動により切り取った人物の正解画像．実環境における正解画像は合計 5000 枚入力している．

図 4.17　非正解画像：路上から撮影した映像データセットからランダムで切り取った背景画像．非正解画像は 20000 枚入力している．

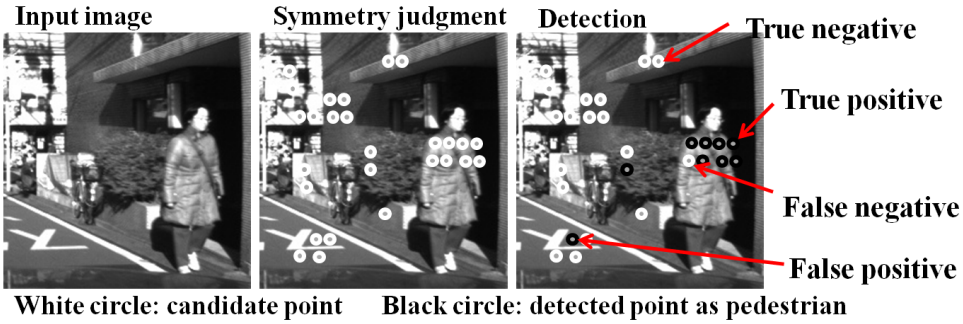

図 4.18　人物検出の評価方法：(左) 原画像 (中) 候補領域 (白円) を表示 (右) 検出位置 (黒円) 表示画像．True Positive は「人物」を「人物」と判断した領域．True Negative は「背景」を「背景」と判断した領域．False Negative は「人物」を「背景」と判断した領域．False Positive は「背景」を「人物」と判断した領域．

表 4.3　路上で撮影した映像データセットに対する検出結果 (適合率，再現率，F 値)

	Precision	Recall	F measure
CoHOG	0.6811	0.5949	0.6321
ECoHOG	0.7922	0.5949	0.6795

実験の結果，本論文の提案手法である ECoHOG が CoHOG よりも F 値が高いことを示し，精度の面で上回っていることがわかった．これは，ECoHOG のエッジ強度累積や正規化の効果が現れたためであると考えられる．エッジ強度累積では画像中の背景に似たような物体が存在した際に，人間とそれ以外の背景を分離した．人間特有の輪郭輝度の強さを統計的にデータ化す

図 4.19　路上データセットでの検出例：(左)CoHOG での検出．過検出が背景に出力されている上に人物を検出できていないため未検出となっている．(右)ECoHOG での検出．過検出がなく，人物領域周辺に複数の検出枠が配置されている．

図 4.20　監視カメラ映像での検出例：(左)CoHOG での検出．未検出はなく，人物領域周辺に複数の検出枠が出力されているが，過検出を含んでいる．(右)ECoHOG での検出．人物領域上に多数の検出枠が出力されている．

ることで，より的確に人物を検出するためのモデルを形成できたと言える．正規化処理では，ヒストグラムの形状を整えるので，明度の強弱によるヒストグラムの大きな変化が無くなる．その結果，人物に影がかかっていたり，時間帯により暗くなってしまっても人物の検出が困難になるケースが少なくなる．図 4.19 は実環境で人物を検出した結果である．画像は比較的暗い環境で撮影しているが，その中でも効率的に強度を捉え，正規化により明るさの変動に対応したものと見て取れる．

しかし，車両が高速に移動している際には，画像中のエッジが不明瞭になり，人物の特徴量が著しく減少してしまう．その結果，人物検出用のモデルとの特徴量のずれが大きくなり，検出が困難になるケースも考えられる．人物モデルに画素値が劣化している状態の人物画像を加えることが対策として考えられる．今回の実験で使用した正解画像枚数は 5,000 枚と比較的少ないため，今後サンプル数をさらに用意することで，検出精度を高めていくことが必要である．

また，屋内環境において検出した例を図 4.20 に示す．ここで，左側が CoHOG による検出，

表 4.4　歩行者追跡精度の比較

Method	Accuracy (%)	Recall
(i)	18.7	62.42
(ii)	72.3	52.00
(iii)	87.2	26.99
(iv)	99.2	5.85

右側が ECoHOG による検出例である．ここでは屋内環境の画像においてぼかしを入れた悪条件下での検出を試した．屋内環境の検出において，CoHOG では背景に過検出が生じている．背景に人物と似たエッジ勾配が存在する場合に過検出してしまうことも多い．また，ぼかしが施されている画像中では，特徴が捉えられていないため検出枠が少なくなっている．ECoHOG では強度の累積や正規化，次元圧縮によりブラーがかかっていても人物特有のエッジ強度を捉えているため，検出枠が人物領域周辺に集中している．

歩行者追跡では，実際の車載映像から撮影された映像に対して提案手法を適用した．比較する手法にはいずれも Tracking-by-detection の枠組みで実装しているが，尤度計算方法と状態推定モデルを変更している．具体的には，(i) ヒストグラムマッチング + 等速直線運動，(ii) テクスチャマッチング + 等速直線運動，(iii) テクスチャマッチング + 車両運動モデルの 3 つを従来手法として比較した．提案手法には (iv)ECoHOG + 車両運動モデルを適用する．ここで，実験の初期位置設定は手動で与えており，ヒストグラムマッチングやテクスチャマッチングは初期位置周辺からヒストグラムやテクスチャをモデルとして取得する．著者のフレームワークでは，あらかじめ歩行者画像や背景画像から生成した歩行者モデル (識別器) を保持しているため，多少の位置ずれやオクルージョン，ノイズには強いと思われる．しかも，初期位置からテクスチャやヒストグラムなどモデルを新しく取得する必要が無い．ここで，歩行者追跡精度の比較を表 4.4 に示す．

比較手法についてまず，ヒストグラムマッチングはほとんど正当な評価ができていなかった．追跡精度を見ても 18.7% であり，歩行者が影領域に入った場合や姿勢変動などあった場合にはほとんど追跡が成功していない．歩行者追跡では時系列的に歩行者の見え方が変化するので，ヒストグラム情報のように位置を無視してしまう方法では尤度の評価が機能していない．手法 (ii) のテクスチャマッチングでは位置情報を含んでテクスチャとして評価しているので歩行者の追

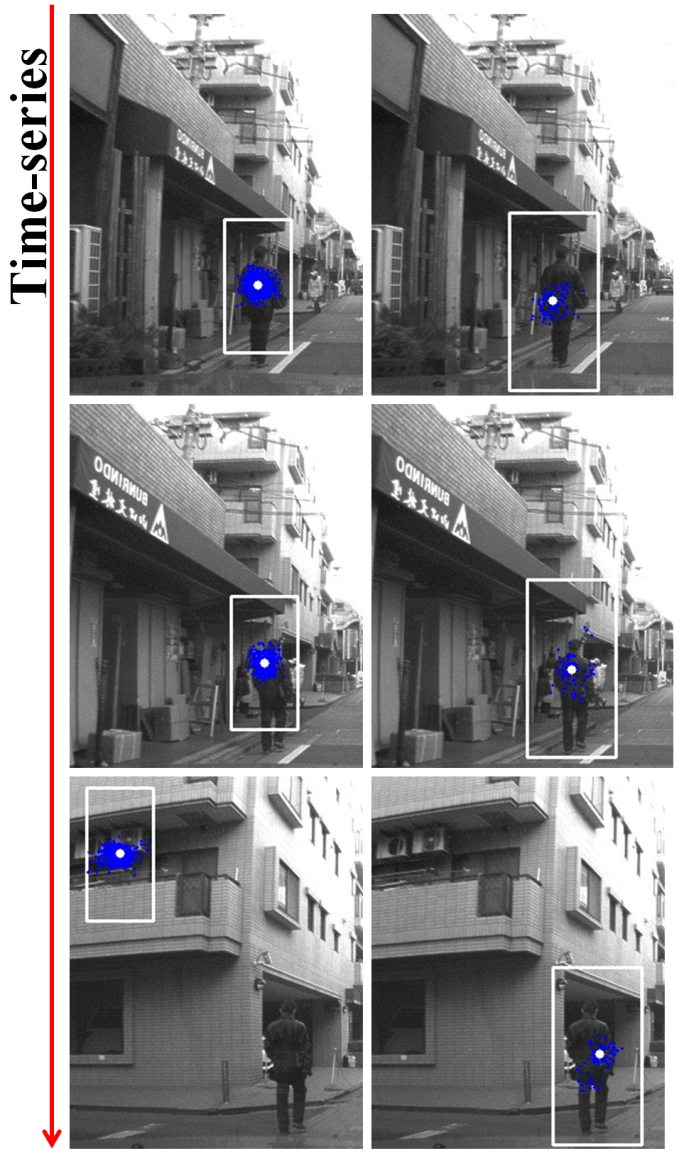

図 4.21　歩行者追跡の例：(左) テンプレートマッチング (右)Tracking-by-detection

跡がヒストグラムマッチングから劇的に向上している．しかし，車両が高速に移動する場合においてはテクスチャがぼけてしまい，歩行者の動作を捉えきれていない場面が発生した．(iii) ではテクスチャマッチングだけでなく，車両の運動モデルも考慮した追跡を実現している．車両の移動方向や移動量をオプティカルフローにより近似しているので等速直線運動よりも自動車の移動に特化した状態推定を可能にした．しかし，テクスチャマッチングでは追跡時間が進むにつれて自動車から歩行者の見え方が変わり，初期に取得したテクスチャとの見えの差異が大きく

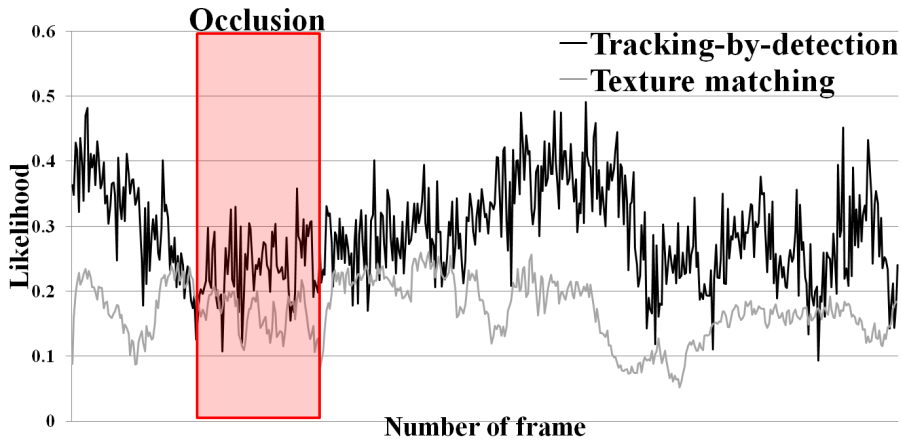

図 4.22　追跡時の尤度の推移：Tracking-by-detection とテクスチャマッチングの比較．テクスチャマッチングは検出時の歩行者から取得したテンプレートなのに対して Tracking-by-detection では統計的学習による歩行者モデルとの比較を行う．見えの変化に頑健であり，尤度もテクスチャマッチングより安定して高い傾向にある．オクルージョン領域内にあっても高い尤度を誇っている．

なってしまう．そのため，テクスチャマッチングを用いた手法では例えば SIFT [16] や SURF [21] などキーポイントマッチング手法を適用して定期的にテクスチャを更新しながら追跡する必要がある．さらに，提案手法 (iv) では ECoHOG による尤度計算と車両の運動モデルを適用し，99.2% の追跡精度を実現した．識別器を適用した手法では歩行者の一般モデルを，ノイズやオクルージョンも考慮した上で含めているため，外乱に強い手法であると言える．学習ベースの歩行者モデルを生成することは，姿勢の変動，カメラの角度変化や照明変動による見えの変化に一部対応しているといえる．これは，正解画像や非正解画像の枚数を増やすことや撮影の場面を増やして学習画像とすることでさらに強力なモデルが構成可能である．

図 4.21 は追跡実験の一例である．図 4.21 左はテクスチャマッチング + 車両運動モデルによる追跡，図 4.21 右は ECoHOG + 車両運動モデルによる追跡を示す．追跡が進んでいく毎にテクスチャマッチングでは徐々にテクスチャと映像中の歩行者の見えの変化が激しくなり尤度が低下し追跡が困難になる．ECoHOG を適用した尤度評価方法では，カメラの角度変化による歩行者の形状モデルも保持しているため，追跡が成功している．また，表 4.4 の通り，5.85fps という処理速度を達成した．通常のラップトップ PC ではリアルタイム処理には不十分であるが，車載ハードを適用した実装に変更するなどの対策で十分リアルタイム処理が可能な範囲である

と考える．図 4.22 は歩行者追跡下における尤度の推移である．ECoHOG による評価とテクスチャマッチングによる評価を並べて表示している．尤度はいずれも 0.0 - 1.0 に正規化して評価している．一連の追跡処理を通して，識別器を用いた尤度評価の方が尤度が高い傾向にある．オクルージョン状況下においても低下が少なくテクスチャマッチングによる評価よりも効果的な尤度関数であると言える．

4.2 サッカー映像解析のための複数選手追跡

近年，サッカーコンテンツを生成するための研究が盛んに行われている [81]-[92]．その例として，戦術解析やハイライトシーンの再現，自由視点映像，シーン検索，自動撮影などが挙げられる．戦術解析は試合中のサッカー選手個人や，チームとしての挙動を記録し戦術を評価し改善するために用いられる．選手やチームとしての傾向，能力や選手一人一人の走行距離などの情報が抽出できるのではないかと期待されている [81]-[92]．ハイライトシーンの再現としては，膨大なサッカー映像を短時間に編集する技術として知られ，たとえばシュートやゴールシーン，試合終了時点などの前後数十秒を抜き出すことが考えられる [87][88]．自由視点映像は大規模な空間で行われるイベントを，多数のビデオカメラで撮影して処理することで視点の選択権がユーザにある 3 次元映像コンテンツである [81]-[85]．サッカースタジアムを模した 3 次元空間中で選手やボールを動かすことで臨場感のある映像を提供する．シーン検索とは，ユーザが見たいシーンを瞬時に検索するコンテンツであり，意味のあるシーン毎にまとめ，自動でタグ付を行う [86]．ユーザ側はテキストや画像情報を基に検索を容易にする．自動撮影では，ボールや選手など試合展開の中心となるオブジェクトを追跡して自動でカメラを制御する [89]-[91]．しかし，これらのコンテンツを実現するためには，サッカー試合映像から選手やボールなどのオブジェクトを抽出する必要がある．当初，選手の移動軌跡は映像から人手により抽出してきた．90 分間の試合の，22 人の選手やボールの軌跡を手動で記録するために，膨大な時間と労力を要する作業である．そのため，サッカー選手やボールの自動追跡が行われ，現在までに様々な追跡手法が提案されている．

サッカー映像解析をしている研究は多数存在するが，ここでは特に，総合的な解析を実現しているサッカー映像解析システム Automated Sports Game Analysis Model (ASPOGAMO)[95]-[99] や，最近のサッカー映像解析手法である Kim らの手法について記述する [93][94]．ASPOG-

AMO では，グラウンド上に存在する全ての選手やボールを撮影し，追跡している．グラウンド上で選手がどこにいるかやボールがどこにあるかの位置を算出しているので，どの選手がボールを保有しているかや，グラウンド上のどこに選手がいるのかを表示することが可能になる．また，プレーがグラウンド上のどこで主に行われているかを記録して，チームのグラウンド上での使い方の解析や，選手の移動速度の推移グラフの生成，試合の状況理解を可能にしている．さらにはボールの移動した履歴や，選手同士の位置関係から戦術を解析することもできる．具体的には，ボールにインタラクションした選手が誰であるのか，またドリブル/パス/シュートいずれかによりボールを運んだか，その時のグラウンド上の位置を求めている．しかし，尤度を求めるときにサッカー映像の全探索による Maximum a Posteriori を適用しているため，選手の大部分に重なりが発生している場合には追跡が困難であるという課題が残されている．Hamid らの手法では，グラウンド全体の選手フローから，ボールの次の位置を推定し，カメラワークを制御する方式について検討している．しかし，選手個人を追跡できるわけではないので戦術解析の用途では使用できないという問題点がある．

ここでは，単眼カメラで撮影したサッカー映像から複数選手を追跡する手法を提案する．一台の単眼カメラではグラウンド全体の撮影は困難であるため，カメラを水平にスイングさせプレーの中心位置を撮影した映像を得る．そのため，カメラの水平スイング動作や選手によるオクルージョンを考慮してサッカー選手とボールを追跡する必要がある．選手の追跡には，色ヒストグラムを特徴とした Particle Filter を適用しているが，特徴量の類似した物体，つまり同じユニフォームの選手が接近した場合の追跡は困難であった．そこで，映像中でオクルージョン領域を判断し，重心を再配置することでオクルージョンへの対応を試みる．オクルージョン領域内の検出手法においては提案特徴量である ECoHOG を適用する．

4.2.1 複数選手追跡手法

オクルージョンが無い場合の Particle Filter は前述の手法を適用するが，状態推定モデルにはノイズを考慮した等速直線運動を，尤度計算には色ヒストグラム比較を用いて追跡する．オクルージョンが発生した場合にはオクルージョン領域内で ECoHOG を用いた検出と速度情報を考慮した重心再配置を行う．

オクルージョン判定．色情報を特徴とした Particle Filter のみの追跡では，同じチームの選

図 4.23　選手間のユークリッド距離

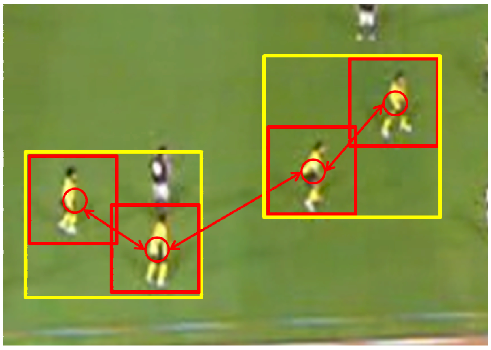

図 4.24　オクルージョン領域の判定：赤枠が追跡により獲得した選手領域，黄色矩形は距離判定によりオクルージョンと判定された領域．

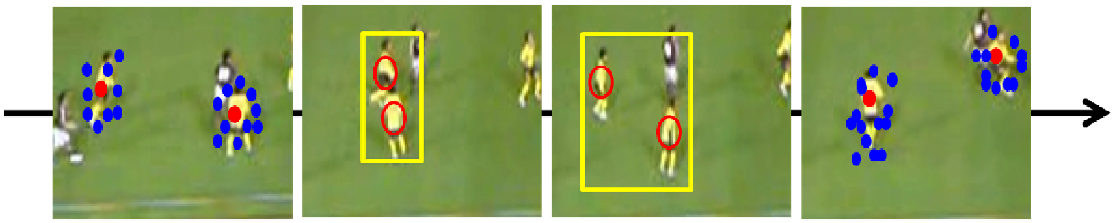

図 4.25　選手追跡における重心再配置の流れ：(左) 単一選手は Particle Filter により追跡する．青円はパーティクルの位置，赤円は重心．(中左) オクルージョンと判定された領域内 (黄色矩形) では識別器による選手検出と速度情報の考慮により重心を常に再配置する．(中右) 選手同士がすれ違っても誤追跡なく重心が配置されている．(右) オクルージョンから離れた際には再度 Particle Filter による追跡を行い，単一選手追跡に切り替える．

手のオクルージョンが発生した場合に誤追跡してしまう．そのため，ユニフォームの色が同じ選手同士に限りオクルージョン発生領域を判断する．オクルージョンかどうかの判断は，Particle Filter により得た重心の距離により判断する．Particle Filter の尤度評価により重心を得た後，以下により重心を算出する．距離はユークリッド距離により計測する．

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{33}$$

ここで，d はユークリッド距離，(x, y) は選手の座標を示す．距離の閾値は，Particle Filter がお互いのパーティクルの局所探索領域を侵害しないように設定する．たとえば，局所探索領域が 30×30 ピクセルの矩形だとすると，距離の閾値 d_{th} は $d_{th} = \sqrt{(15+15)^2 + (15+15)^2} = 30\sqrt{2}$ となる．選手間の距離を図 4.23 に，オクルージョンの判定については図 4.24 に示す．

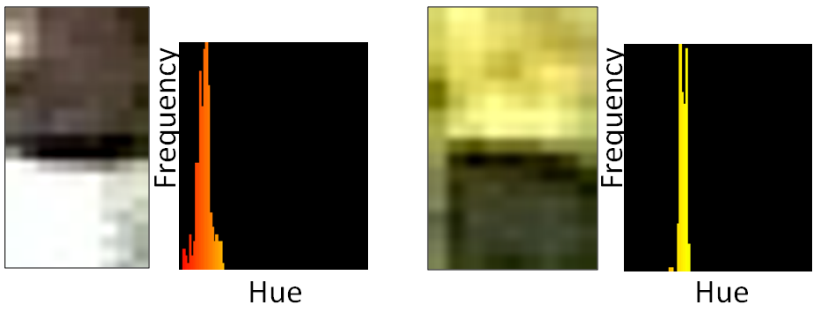

図 4.26　選手から取得するヒストグラム：両方のチームのユニフォーム画像と色相 (Hue) ヒストグラム画像．違う色の追跡は色ベースの Particle Filter により容易であるが，同じユニフォームの選手同士のオクルージョンは色ベースのみの追跡では困難である．

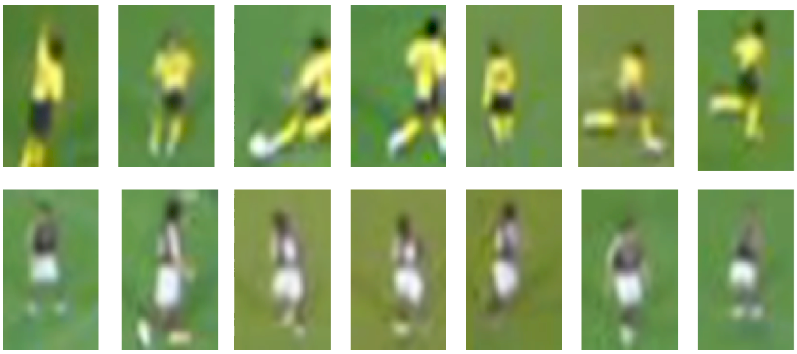

図 4.27　サッカー映像解析の正解画像例

選手検出と重心再配置．重心を再配置するために，オクルージョン領域内で選手を検出する．選手検出には ECoHOG+Real AdaBoost の識別器を適用する．Particle Filter は選手毎に割り当てるのでオクルージョン領域内の人数は既知である．よって，与えられた人数と領域内で検出した人数が一致したときのみ重心を再配置する．一人の選手を重複して検出しないように，一度検出されている周辺領域は特徴量を取得しないように設定している．追跡から重心再配置の様子を図 4.25 に示す．重心が離れている場合には Particle Filter の追跡を，選手が接近して Particle Filter の局所探索領域内に侵入した際には重心算出後に識別器による重心の再配置を行う．重心の距離が離れたら再び Particle Filter の追跡を行う．

図 4.28　サッカー映像解析の非正解画像例

4.2.2　選手追跡実験

選手追跡に適用した Particle Filter と生成した識別器の設定について説明する．Particle Filter について各選手に対して付加する．初期位置設定については手動により複数人の選手位置を指定する．モデルとなる選手のヒストグラムを図 4.26 に示す．また，識別器は Real AdaBoost により学習・生成する．特徴量は検出精度検証で示した通りである．正解画像と非正解画像の一例をそれぞれ図 4.27 と図 4.28 に示す．Particle Filter と Real AdaBoost の設定もそれぞれ表 4.5 と表 4.6 に示す．

表 4.5　Particle Filter の設定

パーティクル数	100 個/選手
局所探索領域	30 × 30 ピクセル (通常) 60 × 30 ピクセル (カメラ動作時)
位置ノイズ	−5〜＋5 ピクセル
速度ノイズ	−5〜＋5 ピクセル
尤度観測	色相ヒストグラム比較
重心算出	重み付き平均

次に，選手追跡実験での概要を説明する．ここではオクルージョン発生前後における追跡精度を検証する．また，同色のユニフォームを着た選手同士と異色のユニフォームを着た選手同士のオクルージョン時の追跡精度についても検証した．従来手法は Particle Filter の尤度観測方法を変更した手法としている．具体的には，色相ヒストグラムのみから尤度を観測した手法と，色相ヒストグラムとエッジを特徴量として取得した手法である．また，オクルージョン領域内にお

表 4.6　Real AdaBoost の設定

弱識別器数	50 個
正解画像	3,500 枚
非正解画像	7,600 枚
閾値	0.0
特徴量	ECoHOG

ける選手検出においても，HOG と ECoHOG による重心再配置を比較した．オクルージョンの前後で，指定された選手を見失わず，入れ替わりがない場合のみに追跡成功とした．提案手法と従来手法における，実験の結果を表 4.7 に示す．

表 4.7　オクルージョン発生時の追跡精度

	違チーム	同チーム
Color	98 / 100 (98%)	15 / 100 (15%)
Color + Edge	99 / 100 (99%)	26 / 100 (26%)
HOG を用いた重心再配置	98 / 100 (98%)	83 / 100 (83%)
CoHOG を用いた重心再配置	98 / 100 (98%)	87 / 100 (87%)
ECoHOG を用いた重心再配置	98 / 100 (98%)	90 / 100 (90%)

違うユニフォームの選手同士のオクルージョン時には，いずれの場合にも高い追跡精度を実現した (図 4.29 左)．これは，ユニフォームの違いにより，追跡対象の色相ヒストグラムの区別が容易にできたためである．カメラからの位置が遠い場所ではヒストグラム取得が困難なため追跡できなかったが，エッジを使った手法では追跡可能なシーンがあった．

同じチーム同士がオクルージョンした場合の考察を記述する．Particle Filter のみの追跡では，尤度の類似する物体がある場合に区別することができない (図 4.29 右)．そのため，ユニフォームが同色の選手が交差した場合には尤度が高く出る選手に重心が集まってしまう傾向が見られた (図 4.30)．ヒストグラムとエッジ特徴を組み合わせて尤度を計算する Particle Filter では，色だけでなく選手の形状を考慮して重心を動かすので，オクルージョン時の精度は向上している．背景に類似する色ヒストグラムが存在する環境でも追跡が出来るようになっているが，単一の選手を見ているに過ぎず，オクルージョンへの対応としては十分な精度が得られているとは言えない．提案手法では，色相ヒストグラムを特徴とした Particle Filter により重心を求

図 4.29　オクルージョン：(左) 違うチームの選手同士のオクルージョン (右) 同じチームの選手同士のオクルージョン

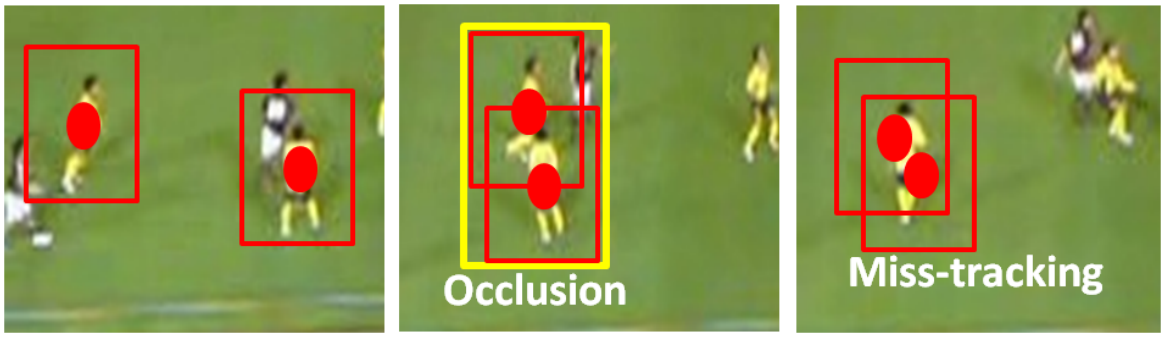

図 4.30　色相のみの追跡により生じた失敗例．左→右と時系列画像となっており，尤度がより高い選手へ 2 つの重心が移動している．

めた後，オクルージョン領域を探索し，Real AdaBoost により生成した検出器により選手を検出して重心を再配置する．Particle Filter による追跡後にさらに密に探索することにより，オクルージョン発生時にも追跡精度を高めることができた．また，運動方向を推定して重心を再配置しているので，オクルージョン後にも選手を見失うことが少なくなったと考えられる．HOG と CoHOG，ECoHOG による検出・重心再配置においても，検出精度の差がそのままオクルージョン時の追跡精度にも反映された．重心再配置はオクルージョン領域内の人数と検出数が一致する際に行っているが，HOG ではオクルージョン領域内において未検出や過検出が多数発生し，効果的な再配置が行えていなかったと考えられる．CoHOG では共起性の考慮によりオクルージョン下において対象となる選手とその周辺にいる選手，または背景との位置関係も含めて学習出来た．さらに，ECoHOG では強度の累積により CoHOG よりも精度を高めていることは前章までで実証済みであるが，外輪郭や上半身・下半身の境目を評価可能であるのでオクルージョン

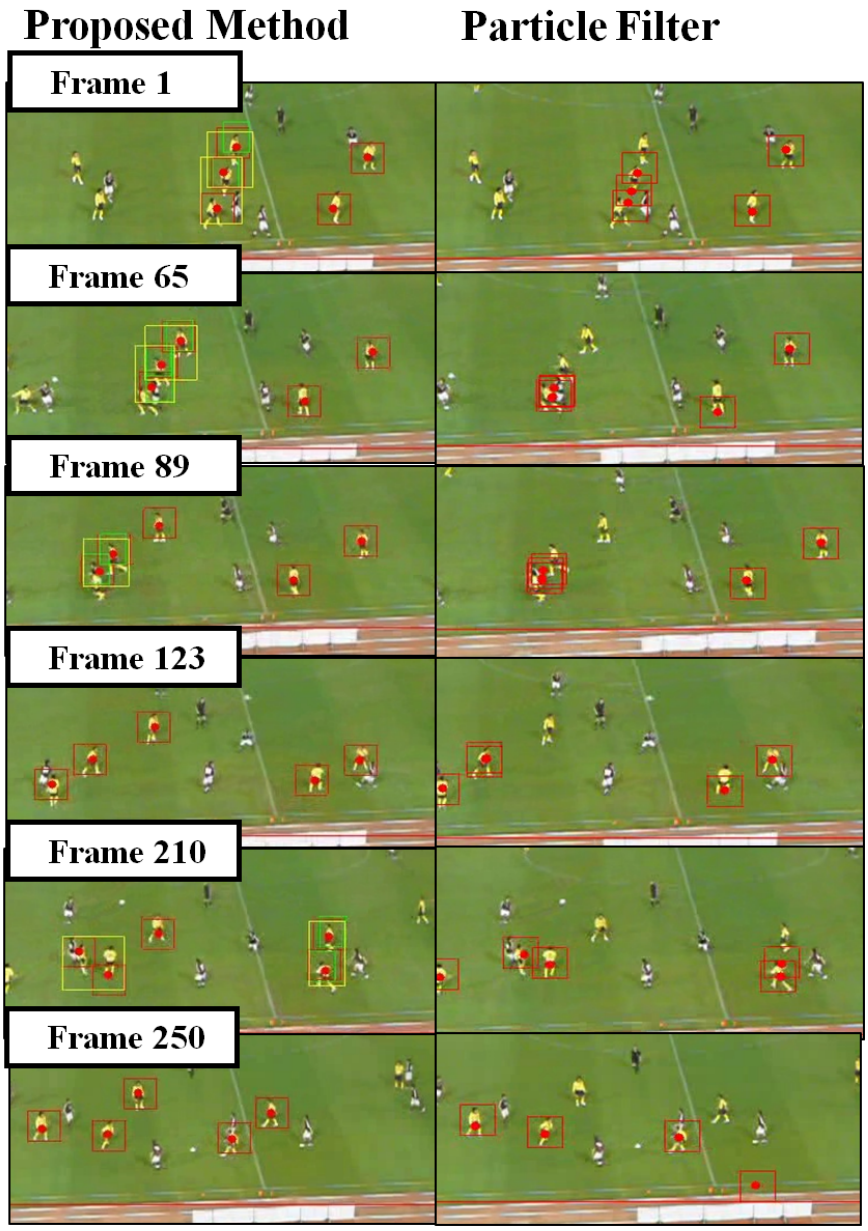

図 4.31　(左図) 提案手法による追跡と (右図) 色ベースの Particle Filter による追跡 (赤枠が Particle Filter の尤度計算適用範囲，赤円が重心位置，黄色枠がオクルージョン領域，緑枠は重心を再配置した位置)：従来手法では Frame1 から密集した場面において色相による尤度が高い選手側へ重心が映っている．Frame65 では 3 つのトラッカーが完全にひとつのピーク位置に移動している．乗り移った重心は最後の Frame250 まで復帰することなく誤追跡している．提案手法では Frame1 から重心再配置を実行している．その後，密集領域を抜ける Frame123 まで重心再配置が行われ，オクルージョンを脱しても追跡タグにミスがなく選手が追跡されている．複数人の追跡においても重心再配置が有効であることを示した．

している場面においても検出が可能である．オクルージョン環境下では特徴量の改善が直に検出率に直結し，重心再配置に適用可能であるので混雑環境における追跡精度自体も向上している．

ここで，色相のみによる追跡と提案手法による追跡の比較を図 4.31 に示す．実画像における追跡を図示しており，従来手法に対して提案手法の重心再配置により追跡がうまくいっている様子が分かる．従来手法ではトラッカーが尤度の高い方向に遷移してしまうが，提案手法では重心再配置によりオクルージョンを含む動画に対しても追跡タグにミスがなく選手が追跡される．複数人の追跡においても重心再配置が有効であることを示した．

課題として，選手が多数集まる場面やカメラから遠く離れた位置ではオクルージョン領域内の輝度勾配特徴の抽出が困難なため，HOG 特徴量の取得がうまくいかず，重心の再配置に失敗する場面があった．選手が多数集まる場面においても，選手の形状がうまく出ず，識別器による検出が困難であったことが考えられる．選手が多数集まる場面では領域内に何人の選手がいるかを判断し，その領域から選手が一人ずつ抜け出すときに重心を配置することで解決を図る．カメラからの距離がある場合は，画像上の選手が小さくなっているので，奥行きに応じて画像大きさの正規化をすることで，スケールの変換に対応可能と考えられる．

4.3 行動理解のための局所特徴量

コンピュータビジョン分野において，現在最も注目を集める技術のひとつが人物行動解析技術である．その中でも，近年特に注目されているのが行動理解に関する研究である [110] [111]．行動理解とはカメラに映り込んだ人物がその瞬間に「何をしているか」に着目し，映像解析により行動タグを付加することである．映像中で人物行動が取得できれば，人物の位置や時系列軌跡だけでなく「その位置で何をしているか」という要素が取得できるため人物解析技術もさらに拡張できる．

行動理解の研究はさまざまな学問分野において取り組みが報告されている．その一例として挙げられるのがウェアラブルセンサを用いた行動理解である [112]．位置情報を示すセンサや加速度センサなどの装着により位置情報や振動・周波数により行動が分かるようになった．装着したセンサの役割により，例えば $walk$，run や sit など日常で頻発する行動や $escalator$ や $driving$ などシステムを用いることにより変化する行動を捉えることに成功した．位置情報が動的に取得でき，しかも精度が良いという利点はあるものの，位置ベースの統計学習であるため，インタラ

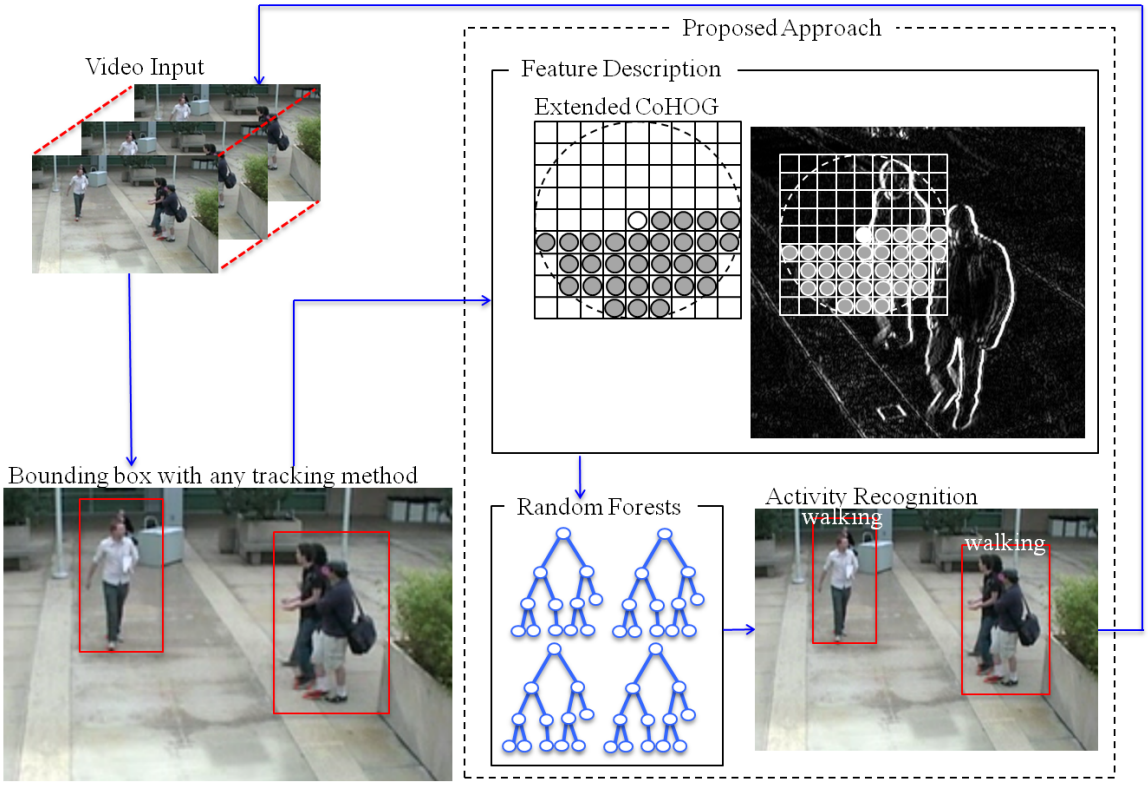

図 4.32　行動理解のフレームワーク：入力画像から人物を追跡した矩形を取得する．矩形領域内では局所特徴量を抽出して，Random Forests 識別器により行動を識別する．

クション認識や手先で変化する行動の変化を捉える事が困難である．この技術を補うために現在最も期待されているのがコンピュータビジョンによる行動理解である．コンピュータビジョンではウェアラブルセンサにより得られる位置情報も取得することができ，その上で頭部や人体の姿勢推定，身体から得られる特徴量などより詳細な行動を取得するための情報を取得できるというアドバンテージを持っている．上記のように監視・スポーツ・予防安全の場面においてウェアラブルセンサをつけているとは考えにくいので，万人に対して適用でき，位置や姿勢情報・局所特徴量を取得できるコンピュータビジョン技術が最有力とされている．

4.3.1　ECoHOG を用いた行動理解手法

前述の ECoHOG 特徴量を用いて，行動理解データセットにおいて評価を行う．前処理として追跡した矩形から特徴量を取得し，特徴量を取得，多クラス識別器である Random Forests を用いて行動を識別する．行動理解のフレームワークを図 4.32 に示す．

Random Forests．識別器には Random Forests を適用する．Random Forests は 2001 年に Breiman によって提案された識別や回帰を実現する手法のことで，過学習を防ぎ精度を向上させているという特徴を持つ [18].

Random Forests は複数のツリー型の弱識別器をつなぎ合わせて識別器を構成する仕組みであり，ツリーひとつひとつが弱識別器の役割を果たすバギングを採用している．ひとつの識別器は高い性能を持つわけではないが，組み合わせにより過学習を防ぐことやお互いを補いあって高い精度を実現可能である．

学習では決定木を構成していくが，基本的にはノード (node) をいかに分割関数を適用して分割するか，がキーになる．最終的に分割できない，もしくは木の深さが限界まできたなど決定された条件の際に決定木の学習を終了して葉ノードに最終的な出力を保存する．葉ノードにはそれぞれのクラスの確率分布が備えられており，識別や回帰の際に使用する出力値となる．特徴ベクトル $v = (v_1, v_2, ..., v_N)$ が入力として与えられたとき，根ノードに入力される．ここではランダムにベクトルの次元を選択して，分割関数に従ってノードの分割が行われる．この分割処理は再帰的に繰り返されて決められた条件に達したときに葉ノードを作成して学習を終了する．

あるノード j に特徴ベクトル v が到達したときの分割関数は以下のように示す．

$$h(v, \theta_j) \in 0, 1 \tag{34}$$

θ は分割関数のパラメータであり，$\theta = (\phi, \psi, \upsilon)$ で表せる．ϕ は n 次元ベクトル v からいくつかの次元をランダムで抽出するフィルタ，ψ は分割基準パラメータ，υ は分割の閾値を示す．識別の場合にはクラスのタグ c の事後確率 $p(c|v)$ が葉ノードに割り当てられる (図 4.33).

Random Forests におけるランダム性は，(i) 決定木を学習する際の学習データのサンプリングと (ii) 決定木の各ノードでの分割関数学習で採用されている．(i) は t 番目の決定木学習に利用する学習データセット $S_0^t \in S$ を選択する際に，S からランダムに抽出する．これにより，特徴ベクトル中から全ての次元を参照することがないので高速に，かつ決定木数を増やせば増やすほど精度が向上していく．(ii) ではランダム性の導入により分割関数の膨大なパラメータのうち，値の一部のみを利用できる．ランダムの度合いはパラメータにより決定できるため，学習データセットにより変更する必要があるが，相関性を保ちながらランダム性を導入する方が良い

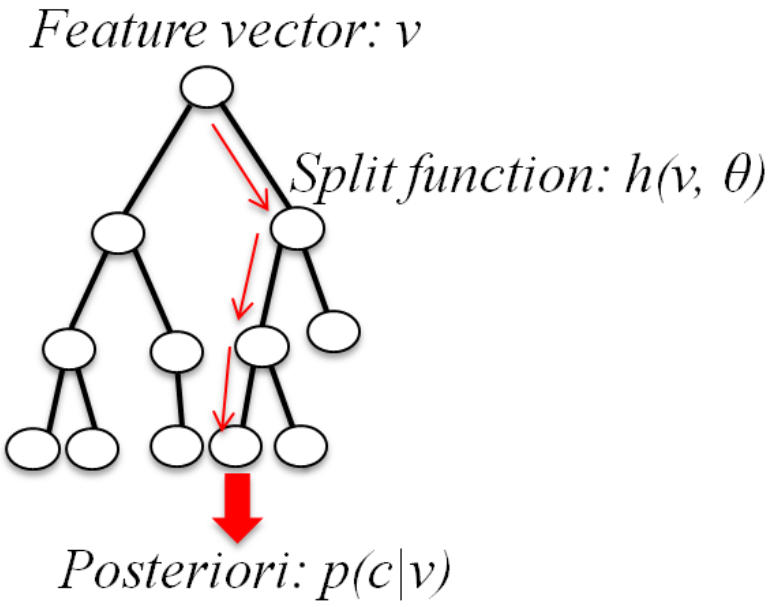

図 4.33　決定木と葉ノードの事後確率

とされている．ノードの分割にはエントロピーを基準にしており，

$$H(S) = -\Sigma_{c \in C} p(c) log(p(c)) \tag{35}$$

で表現される．全てのクラスが同じ確率で存在するときに $H(S)$ は最大となり，ひとつのクラスしか確率を持たない場合に最小となる．分割による情報利得を以下に示す．

$$I = H(S) - \Sigma_{i \in 1,2} \frac{|S^i|}{|S|} H(S^i) \tag{36}$$

また，学習を終了する条件については，予め設定した深さに達する，ノードに割り当てられた学習データの個数が一定値以下になった場合，もしくは分割による情報利得が一定値以下になった場合などがあり，通常それらすべてが同時に用いられる．

次に，Random Forests を用いた識別や回帰について説明する．ここでは学習により T 個の決定木が割り当てられている状態である．同じように根ノードに特徴ベクトルを入力するが，評価時には各決定木に割り当てられている分割関数により特徴次元を参照してノード間を移動する．葉ノードに達した際に格納されている事後確率を得る．全ての決定木から事後確率を得た場

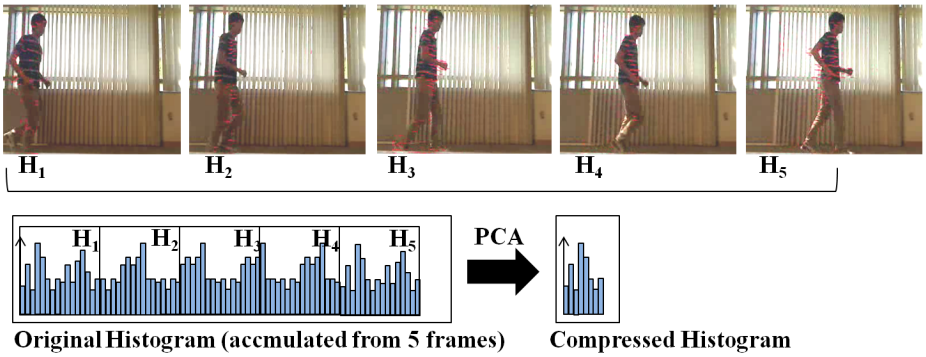

図 4.34　時系列での特徴累積と PCA による次元圧縮

合に確率の平均を以下のように取得する．

$$p(c|v) = \frac{1}{T}\Sigma_{t=1}^{T} p_t(c|v) \tag{37}$$

また，全てを掛け合わせた値を用いることもある．

$$p(c|v) = \frac{1}{T}\prod_{t=1}^{T} p_t(c|v) \tag{38}$$

識別クラスを決定する際には事後確率の平均値が最大のクラスを選択する．

$$C_i = argmax_{c_i} P(c_i|v) \tag{39}$$

行動理解のための時系列特徴累積．行動理解では時系列で生起する人物の行動を識別する必要があるため，連続する画像から取得する特徴量を蓄積して特徴ベクトルを形成する．ECoHOG は一フレームより取得する特徴量であるため，行動を識別するために時系列で連結させる．ここでは画像毎の特徴を時系列順に累積していくが，蓄積した後に次元を圧縮する．これは，時系列画像数によって累積する特徴次元数が高次元にならないようにするためである．図 4.34 に複数フレームからの特徴量取得や PCA による次元圧縮と最終的に得られるヒストグラムを示す．図では，5 フレームから ECoHOG を取得，時系列順に累積した後に次元圧縮を施した例である．

4.3.2　行動理解データセットにおける識別精度の比較

行動理解データセットとして一般的に知られる，Weizmann action dataset[100] や KTH dataset[101] を適用して，行動理解の場面における ECoHOG の精度を比較する．以下に，それ

図 4.35　Weizmann action dataset による行動理解の一例

表 4.8　Weizmann action dataset を用いた行動理解の精度比較

Framework	Accuracy(%)
Proposed method	97.1
Bregonzio *et al.* [106]	96.6
Satkin *et al.* [107]	95.8
Lin *et al.* [108]	95.4
Chaudhry *et al.* (HOOF)[105]	94.4
Jhuang *et al.* [109]	92.8
Klaser *et al.* (HOG3D)[103]	90.7

ぞれのデータセットにおける精度や考察を示す.

Weizmann Action Dataset [100]. Wezimann action dataset では 90 の映像があり，9 人の人物が 10 の行動 (bend, jack, jump, pjump, run, side, skip, walk, wave, and wave2) を実演している．Weizmann action dataset による行動理解の精度を表 4.8 に示す.

KTH Dataset [101]. KTH dataset は 600 の動画に 6 種類の行動 (boxing, handclapping,

図 4.36　KTH dataset による行動理解の一例

表 4.9　KTH dataset における検証実験

Framework	Accuracy(%)
Proposed method	83.3
CoHOG	77.2
Chaudhry *et al.* [105]	79.7

handwaving, jogging, running, and walking) が含まれている．KTH dataset では 160 × 120 ピクセルの動画にカメラモーションや人物の影領域などが含まれている．KTH dataset による行動理解の精度を表 4.8 に示す．

Weizmann action dataset では 97.1%，KTH dataset では 83.3% といずれも高い識別精度を達成した．行動理解の一例をそれぞれ図 4.35 と図 4.36 に示す．ECoHOG では共起性の考慮だけでなく，ペアとしてエッジ強度を累積しヒストグラムを作成しているため人物特有のエッジ強弱を識別できたと考えられる．また，エッジ強度を累積後，正規化を行っているため，KTH dataset の明るさの変動にも対応できた．表 4 に示すように CoHOG(77.2%) と ECoHOG(83.3%) の精度を比較してもエッジ強度の累積や正規化の有用性が分かる．ECoHOG は強度成分を累積することで，弱い成分よりも強い成分のエッジ強度を識別に重視することを前章の検出実験で述べた．検出だけでなく，行動認識においても記述性能を高めることは識別にとって有意義であるので，強度累積により人体の外輪郭など強い成分を持つエッジを重視することは有効であった．行動認識だけが時系列方向にも特徴を累積しているが弱いエッジ特徴に

引きずられずに記述ができている．強い特徴成分を重視して評価できるので，例えば “run”と “walk”など一見すると同じような姿勢の連続に見えても，外輪郭成分を見てみると変化のタイミングや一度に動く速度などの面において識別を向上させたと考える．

行動理解のように時系列で特徴を取得する場合には高次元になってしまう場合にも，PCA による次元圧縮が効果的であると考える．ここでは 5 フレームから取得した 15360 次元 (1 フレームあたり 3072 次元) の ECoHOG を圧縮して 100 次元の特徴量とした．

Weizmann action dataset における “jack”と “wave2”(どちらも両手を上下移動させる), KTH dataset においては “walking”, “jogging”や “running”で行動の特徴量が非常に類似していたために分類が困難であった．画像のブロック分割により左右の特徴が混同しない処理や時系列累積により変化する特徴を捉えているが，行動理解において更なる精度が必要な場合には複数の特徴量を組み合わせる必要があると考える．また，その他の要因として KTH dataset ではカメラモーションや影領域などを画像中に含んでいるため，形状における比較では精度が下がっている．

4.4 人物行動解析の応用へ向けたさらなる拡張

著者は先の人物行動を予測する「行動予測」に関する検討を重ねている [72]．コンピュータビジョンにおける人物行動解析の研究はその応用が広範に及び，実用化の側面からも非常に期待されているためさらなる発展が待たれる分野である．しかし一方でこれまでに提案されてきた技術は何れも行動の事後解析であり，「すでに起きている」事象の解析に重きが置かれていた (図 4.37)．これに対してカメラで撮影した行動を事前に阻止することや次の行動を推薦するといった高度な情報空間の構築など，多種多様なアプリケーションへの応用が期待できる (図 4.38)．カメラで撮影してから行動を解析し結果を出力する「事後解析 (Post-event Analysis)」だけでなく，未来に起こりうる行動を先読みする「事前予測 (Pre-event Prediction)」の概念は今後さらに必要になると考えられる．

著者は行動理解により得た連続する行動タグと，予め用意した行動履歴データベースの解析結果の照合により，先の行動を確率的に予測するフレームワークを確立した．行動理解により行動タグを獲得するが，行動理解は任意の手法により実現する．行動履歴データベースには「前の行動」「現在の行動」だけでなく，行動の並びが同じでもその次の行動は時間により変化すること

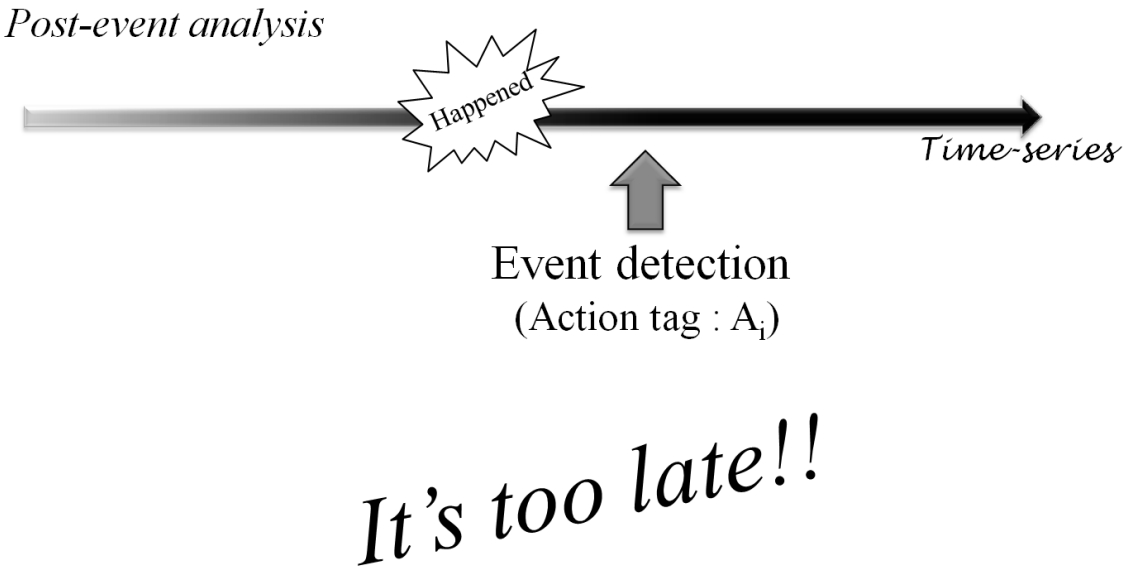

図 4.37　事後解析 (Post-event Analysis)：カメラで撮影/認識してから出力するため，事象が発生した後の行動認識になっている

を前提に「時間帯」を要素として含ませて「先の行動」を予測している．確率モデルには Naive Bayes 識別器を用いており高速な解析を実現している．ベイズの枠組みでの予測では，全ての行動に対してランク付けが可能であり，確率の高い順からソート可能である．実験では動画像解析とデータベース解析により平均して 5.0 秒，最大で 20.2 秒の行動予測を実現している．また，定点観測かつ行動理解の誤りがないという前提ではあるものの，最上位の行動において 81.0% の確率で行動を予測できることが分かった．また，時間帯により変化する行動の予測も実現しており，今後への行動予測への可能性を得ている．

「行動理解」と「行動解析」を用いることで「行動予測」を実現したが，この両者の統合により予測だけでない更なる可能性があると考える．見守りや監視以外の例として，人物動線の予測，人物間コミュニケーションの可視化，スポーツにおける戦術・選手軌跡予測が挙げられる．

人物動線の予測: 人物動線の予測は一部，Pellegrini らの “Local Trajectory Avoidance (LTA)” [73] や Kitani らの “Activity Forecasting” [74] により実現している．人物だけでなく環境や物体のモデリングにより，数秒後の人物の位置を確率的に予測している．ここで，実際の環境では一人の追跡だけでなく複数人のグループ追跡も必要になる．実際にグループの行動解析は研究段階にあり [75]，グループとしての行動解析も必要と考える．例えば，集団の動向解析と

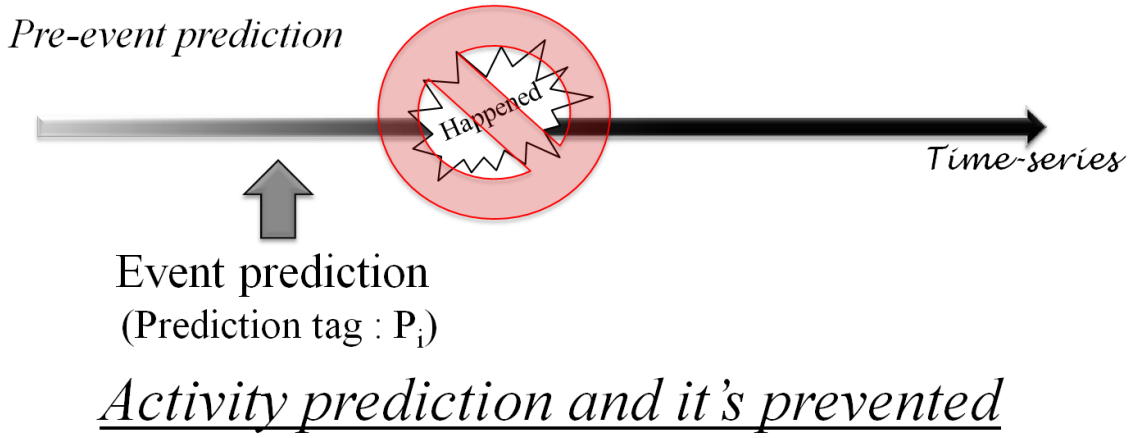

図 4.38　事前予測 (Pre-event Prediction)：データベースを予め解析して事前情報を持たせているため，事象が発生する前に察知することが可能

混雑予測である．事前に混雑する位置や度合いを把握できればサービスの向上に役立つ．著者はすでに動線の解析に着手しており [76]，大量のデータベースがある際にどのような特徴を捉え，環境認識の方式を検討し，統計モデルの設定についても改良を繰り返している．

人物間コミュニケーションの可視化: 人間同士のコミュニケーションは普遍的であるが，現在まで可視化するツールが数少ないことも事実である．近年のコンピュータビジョン分野では人物間のインタラクションを認識する研究も発生しており [77]，インタラクション認識と蓄積，データベース解析によりコミュニケーションの可視化が実現できる．例として，一人の人物行動を蓄積した性格の分析と，複数人物間のインタラクション認識を分析した人間関係の構築が考えられる．性格分析により，個の解析とその間に起こる人間関係の相互作用解析を可能とする．なお，コミュニケーションの可視化においては心理学や社会学など様々な要素を含む研究テーマになると予想される．

スポーツにおける戦術・選手軌跡予測: スポーツ映像解析では主に選手追跡の研究がおこなわれてきた．特徴量の改良と追跡モデルの高度化により，精度が著しく向上してきた．追跡精度が向上したことで，今後は戦術の解析に移行すると考える．そのための技術として，行動解析の技術は戦術を解析しさらにパターン化するための重要なフレームワークである．また，人物動線の予測と同じく，選手軌跡の予測も可能であるが，スポーツ映像解析ではボール，選手間の位置関係，フォーメーションなど実際の環境とは異なるためスポーツに合わせたモデルを構築する必要

図 4.39　KITTI Vision Benchmark Suite の一例

図 4.40　Caltech Pedestrian Benchmark Dataset の一例

がある．

その中でも，交通安全においては最も効果が大きいとされている．現在までの交通安全分野における予防安全技術において，ステレオカメラやレーザを用いた前方車両や人物など障害物の検知，周囲環境の 3 次元再構成などが研究されてきた．著者はステレオカメラではなく，単眼カメラを用いた歩行者検出と追跡を提案した [78]．しかし，歩行者予防安全技術が持つセーフティシステムは時速約 30km/h で走行する自動車の停止が限界である．今後更なる安全を提供するためにも自動車走行環境に潜む危険を事前に顕在化させる「危険予測」へ取り組む必要がある．近年

そのためのデータセットの充実もあり，益々実現可能性が高まっているといえる．例えば，カールスルーエ工科大学 (KIT: Karlsruhe Institute of Technology) と豊田工業大学シカゴ校 (TTI: Toyota Technological Institute at Chicago) が共同で作成した KITTI vision benchmark suite [79] やカリフォルニア工科大学 (Caltech: California Institute of Technology) が配布している Caltech pedestrian detection benchmark [15] は歩行者の映像を多数含んでいる (図 4.39，図 4.40)．特に，KITTI vision benchmark suite ではステレオカメラの映像だけでなく，オプティカルフローやオドメトリ，物体情報，車両の位置，道路位置などのデータが同時に配布されている．Caltech pedestrian detection benchmark においても約 2300 人の歩行者が含まれているだけでなく，オクルージョンや歩行者の大きさなどの詳細な解析も Dollar らによって付加されている [15]．これらのデータセットを用いることで歩行者だけでなく，自車両や周囲環境のモデリングを可能とする．一方で自動車技術会より交通事故のヒヤリハットデータベースを配布している [80]．自動車技術会では 100 数十台のドライブレコーダをタクシーに搭載し，2006 年から 2008 年までの間に 33000 件のヒヤリハット事例を集積することに成功している．ヒヤリハットデータベースには自動車同士の接触だけでなく，歩行者や自転車など衝突に至れば重大事故になるケースも収録されている．危険予測を実現するためには歩行者・自車両・周囲環境のモデリングをする一方で事故のケースを解析して，危険の定義をする必要もある．実際の場面において危険を予測するためには定義した危険と実環境から取得したそれぞれの要素がどれほど類似しているかを判断し，何秒後にどの程度の危険になるかを予測できるフレームワークを構築する．数秒前に危険を予測できるとすれば，現在よりも速い速度で走行する自動車を制御して停止できる．車載カメラにより事後に解析して歩行者を検知するだけでなく，事前に危険を予測して自動車を制御する「危険予測」の概念が未来の交通安全技術には必要不可欠である．危険予測の概念を確立することが予防安全システムを拡張するための糸口であり，著者は今後も実現可能性を探る．

4.5 本章のまとめ

本章では提案した特徴量 ECoHOG を，様々なシーンにおける人物検出事例へと適用し，その応用可能性を検討した．歩行者予防安全では車載カメラから撮影した前景映像中から歩行者を捉える技術について，歩行者の左右対称性に着目した前処理や Tracking-by-detection を用いた追跡を実装した．前処理に関しては歩行者と背景の学習画像から閾値を決定し，追跡に関しては

ECoHOG の適用や車両運動モデルの適用により従来手法よりも精度を高めている．サッカー映像解析では，単一の選手追跡に色情報を用いた Particle Filter を適用しているが，オクルージョン発生領域にて選手を検出して速度を考慮して重心を再配置する．オクルージョン領域では色情報を用いた追跡はより重心の高い選手にトラッカーが追跡してしまうが，より接近した状態でも重心を検出できる形状ベースの手法に切り替えて追跡する．実験では ECoHOG を適用した識別器が HOG よりも混雑した状況での検出を可能にし，オクルージョンした状態でも追跡精度を向上させている．実験では 2-3 人のオクルージョン発生時の追跡において Particle Filter 追跡と ECoHOG 識別器の統合による手法は 90% の追跡精度を実現した．人物の行動理解では予めトラッカーにより抽出した人物矩形から特徴量を取得して行動を判別する方式を採用している．人物矩形からは ECoHOG 特徴量を適用しているが，行動理解においては時系列の情報が必須であるため，複数フレーム特徴量を蓄積することにより行動理解においてより強力なフレームワークを構築している．実験では Weizmann action dataset や KTH dataset に対して適用しており，従来法と比べて高い性能を実現している．

人物行動解析の応用へ向けた拡張として，「行動予測」の概念を提唱した．現在までのコンピュータビジョンで提案された人物行動解析の技術は何れも行動の事後解析であり，「すでに起きている」事象の解析に重きが置かれていた．一方で事前に人物の行動を予測することができれば，今後人物行動解析への応用が拡がると考える．

局所特徴量の改善と人物行動解析への応用を実現した．「歩行者予防安全」では複雑背景，移動するカメラからの歩行者検出，「サッカー映像解析」では素早く運動し，選手によるオクルージョンが多数発生する場面での複数選手追跡を，「行動理解」では姿勢変動を含み時系列に生起する動作からの人物行動識別を解決した．ECoHOG ではエッジ強度を累積することにより，重要度の高い成分を識別に重視する．対象物体のテクスチャの相違や複雑背景でも安定した特徴ベクトル取得を可能としているため複雑背景やオクルージョンの多数発生する場面においての検出を高精度化した．また，人物の外輪郭の形状等，より重要度の高い成分を特徴として抽出する傾向にあるので時系列に生起する人物行動からも詳細に特徴を記述できたと言える．主成分分析による特徴次元の圧縮については，特徴空間での分離を容易にして少ない学習サンプルでも効果的に歩行者のモデリングができているため精度が向上している．

5 結論

本章の概要

本章では，結論として本論文を総括し，今後の課題と将来展望を示す．

本論文では，人物検出を対象とした局所特徴量の改良に取り組んだ．局所特徴量 CoHOG を対象としてエッジ強度の累積・ヒストグラム正規化・次元圧縮を施した．実験では局所特徴量の設定や従来手法との比較を行い，人物行動解析の様々な場面において局所特徴量を応用することでコンピュータビジョンの様々な問題を解決した．

人物検出のための局所特徴量の改善として，共起特徴を対象として CoHOG をさらに改良した特徴量である ECoHOG を提案した．ECoHOG では人物のエッジ強度を特徴量として累積し，人物の強度成分の分布や曲率度合を表現可能にした．明度状況に依存してエッジの強度は変化してしまうため，ヒストグラムを正規化した．さらには主成分分析により効果的な次元圧縮方法を適用した．

実験では，INRIA person dataset や Daimler pedestrian benchmark dataset，実環境で撮影したデータセットを適用して ECoHOG の設定方法や従来手法との比較について検証した．ECoHOG の設定について，エッジ強度の累積には和算による累積が特徴の情報量を損なわずに特徴記述でき，特に過検出を低減する手法であることが分かった．一方で積算による累積では強い特徴量のみを残す特徴量であり，未検出を低減する手法であることを求めた．次元圧縮では情報量だけでなく，特徴空間サイズという指標も考慮して，双方のバランスが非常に重要であることが分かった．実験的ではあるものの，4608 次元の ECoHOG に対しては 100 次元が最も精度が良かった．従来手法との比較について，ECoHOG が最も高い性能を示した．エッジ強度の累積やヒストグラムの正規化，次元圧縮により精度が向上している．CoHOG からは特に，姿勢変動，部分的な遮蔽，複雑背景等において検出精度が向上している．

提案した局所特徴量 ECoHOG を歩行者予防安全，サッカー映像解析，行動理解の各場面で適用した応用例についても述べた．歩行者予防安全では車載カメラから撮影した前景映像中から歩行者を検出・追跡する手法について歩行者の左右対称性に着目した前処理や時系列で検出結果を対応付ける Tracking-by-detection を用いた追跡を用いた．サッカー映像解析では単一の選手追

跡に色ベースの Particle Filter を適用しているが，オクルージョン発生領域にて選手を検出して速度を考慮して重心を再配置した．オクルージョン領域での検出と重心再配置においては特徴量の精度がそのまま追跡精度に直結することを実証した．また，行動理解においては時系列の情報が必須であるため，複数フレーム特徴量を蓄積することや，Random Forests を適用することにより複数クラスの識別を可能にした．

これらの取り組みから，ECoHOG を適用し従来の局所特徴量の課題を一部克服したと言える．共起性の考慮や人物特有のエッジ強度の累積，正規化により複雑背景や照明変動や姿勢の変化などの場面において精度の向上が見られた．また，歩行者予防安全の場面においては速度のある車両から撮影し，ぼけ等不鮮明なエッジを含む画像中でもエッジ強度ペアの総合的な評価により検出可能としたり，サッカー映像解析の場面においてもオクルージョン領域内における選手同士の遮蔽に対してもロバストな検出が出来た．行動理解においても局所特徴量自体の記述性能を高めているため，時系列特徴にした際にも詳細な形状まで表現でき，複雑な行動の識別にも役立てることに成功した．

提案した局所特徴量により，高い精度で検出や認識を実現しているが現状ではさらに性能を高める必要がある．例えば，遠方に見える人物を検出することは，高速で走行する自動車の動きを止める上では非常に重要な技術となる．しかし，現在の検知技術では約 30m を境に性能が落ち始める．これを，さらに遠方に人物が離れていても検出できる技術があれば，さらに危険を回避できる能力が高まる．また，スケール変化やカメラ位置による見えの変化にも対応できるような特徴構成にするのが望ましい．よって，局所特徴量はさらに精緻な特徴を捉えるように改善する必要がある．現状では 2 つの画素の共起を捉える特徴量により形状を記述しているが，これからは 3 - 4 画素の共起も検討する必要がある．3 - 4 画素の共起では次元数が膨大に増えてしまうため超高次元特徴になってしまうことが課題になる．

主成分分析において，汎用的なモデルを作成することも課題の一つである．主成分分析においては分散を最大にするような固有ベクトルを生成するためにモデルとなるサンプル画像を入力する必要がある．場面が変化しても恒常的に人物のモデルを変換できるような固有ベクトルのモデリングが課題となる．

また，学習画像のバリエーションを拡張させる必要もあると考える．現在では人物画像が数千オーダの枚数であるため，必ずしもジェネラルな人物モデルを形成出来ているとは言えない．多

数の場面における多数の人物の画像を収集することが検出や尤度計算の精度にも直結する．人物の角度，照明，姿勢などアピアランス変化に着目して学習画像を集める必要がある．その他は各パラメータの最適化が挙げられる．多数の場面においてシステムを適用して最適なパラメータをそれぞれ決定していく．

実利用化への課題としては，いかなる場面においてもリアルタイム性を持たせることが挙げられる．実際に利用可能なシステムとして運用するためには GPU 等高速なハードウェアでの実装を行う．システムとして最適化することにより，格段に処理時間の短縮が期待できる．入力から検出までをハードで実装し，各工程において処理時間を大幅に高速にする．前処理に関しても問題によって最適化する必要がある．

研究課題として，3 - 4 画素の共起が必要と述べた．しかし，3 - 4 画素の共起では特徴次元数が膨大に増えてしまうため，真に必要な画素の組み合わせや取得する位置を制限する必要がある．また，共起特徴に限らず，現在の局所特徴量の研究においては複数の特徴量を統合する傾向にある．複数の特徴量を統合する試みでは例えば「形状」や「動き」など取得方法の異なる複数の特徴量を組み合わせることで相補的な特徴記述ができることからもきたいされているが，特徴空間が膨大になってしまい，統計的な機械学習に非常に不利であった．

そこで必要とされる技術がデータマイニング (Data Mining) を適用して特徴空間内を探索する特徴選択 (Feature Mining) と呼ばれる概念である．まずはデータマイニングについて概説した後で特徴選択の研究について説明する．

データマイニングの概説．データマイニングは大量のデータの中から明らかにされていない法則を発見する技術のことであり，ビッグデータから有益な情報を導き出すための技術としても注目を集めている．データマイニングでは統計学やパターン認識，人工知能の知見を適用することで網羅的にデータベースを解析可能となる．最近のビッグデータ解析ではデータの数を集めることが重要なのではなく，全ての要素のパターンを網羅的にそろえることが大事であるという実験結果が得られている．データマイニングによる解析技法としては，

1. データクレンジング：データベースのノイズの除去や欠損を補てんする．取得したデータはそのままでは使用できないことが多いので背景知識やルールを用いることでデータベースを修正する．はずれ値の除去やカテゴリ化などもデータクレンジングに含む．

2. データ結合：複数のデータリソースをひとつに統合する．その際にデータの種類が異なる場合もあるので変換も含めてデータ結合とする．複数のデータを結合可能なのでデータ数を拡張することが可能.
3. データ選択：解析するデータをデータベースから選択する．
4. データ分析：データベース中から対象となるデータ抽出や知識を取得する．
5. データ変換：数値を量子化してテキストにする，や形式を統一するなどがデータ変換として挙げられる．
6. 知識表現：得られた知識を可視化したり，テキストとして表示するためのデータ分析

などが挙げられる．しかし，データマイニングの難しさとしてはデータが決まったフォーマットを持たないため，目的を決めないと所望の結果が出ないことが多い．解析した数値に関する考察にも注意が必要である．また，データベースへ追加する要素や場面に合わせたデータの考案も非常に重要となる．また，統計手法も多様化しているが，データマイニングで適用されるトップ 10 アルゴリズムには以下の手法が挙げられている [35].

1. C4.5
2. K-means
3. SVM
4. Apriori
5. EM algorithm
6. Page Rank
7. AdaBoost
8. k-Nearest Neighbor
9. Naive Bayes
10. CART

特徴選択．Feature Mining の分野ではいくつかの研究がおこなわれている．Dollar らは多数の特徴データベースから識別に必要な特徴量を見つけ出して次元数の削減や学習の高速化を図った [66]．Quack らは Bag-of-features による特徴量表現を用いて画像中の同時出現確率を求めた

[67]．特徴取得ウインドウを用意して，物体が持つ Bag-of-featues をアソシエーションルール [68] を用いて同時に出現しやすい特徴を解析している．特定の物体に現れやすい組み合わせの発見により主成分分析など多変量解析手法に依存せず識別率を保ったまま次元数の削減に成功している．しかし，アピアランスから取得する特徴量は見え方が変わった場合に特徴ベクトルが変化してしまうため，マイニングの結果が反映されないという問題点が発生する．Dollar らの手法について，特徴ベクトルが変わると識別空間内において目的とする識別クラスとは離れてしまい識別率が低下することも考えられる．また，Quack らの手法についても見え方が変わると特徴ベクトルを量子化した Bag-of-features が変わってしまい，あらかじめ作成したモデルとは異なる特徴が現れてしまうという問題点が挙げられる．Gilbert らは，XYT の二次元空間 (XY，XT，YT) に対して特徴点を抽出し，1 フレームから 1,500 点以上の特徴点を抽出，マッチングにより行動を理解した [70]．各特徴点は XYT 空間の特徴を持ち，アソシエーションルールに基づくデータマイニング [68] により行動クラスに特有の頻出パターンを抽出して特徴取得の効率化を図った．マイニングされた特徴量は他のクラスの行動と分類しやすいという指標にも基づいており，非常に効果的な特徴量となっている． STIP は疎な特徴点抽出であると指摘した上で，XY，XT，YT 空間に対して複数の階層を設けた Harris corner で特徴点を抽出し密な特徴点探索を実現している．抽出した特徴点の 1. 取得したスケール，2. XY，XT，YT のチャネル，3. 方向を基にして特徴量を構成する．XYT の $3 \times 3 \times 3$ のグリッドから取得した位置関係と 1 - 3 の特徴量から 5 つの digit をアソシエーションルールに従って頻出かつ分離がしやすい特徴量を選択，行動理解のための特徴量にしている．この結果，KTH dataset においても 93.6% の精度を実現しているだけでなく，24fps と高速な処理時間を実現した．特徴選択では膨大になった特徴次元数を絞るだけでなく，いかに効率的に特徴量を構成するか，という特性も持ち合わせる．主成分分析では一度全ての特徴量を取得した後，固有ベクトルにより特徴ベクトルを変換しているが，特徴選択では必要な特徴のみを取得可能なため，特徴を取得する処理時間が大幅に削減できる．また，2 画素の共起を 2 つ組み合わせて 4 画素の共起にしているという研究例も存在する [71].

特徴選択は取得した特徴ベクトルを，データマイニングすることにより有効な特徴量を知識として抽出する研究である．主成分分析のように一度全てのベクトルを取得してから別の空間に射影するのではなく，有効な特徴量のみを取得するので次元の削減と高速化を同時に満たすことが

できる．また，有効な特徴量のみを学習に使用可能であるので精度向上ができることでも知られている．

謝　辞

本研究は著者が慶應義塾大学大学院理工学研究科在学中の 2009 年 10 月より同大学理工学部電子工学科の青木義満准教授のもと研究致しました．青木准教授には著者が芝浦工業大学在学中から 6 年以上もの間御指導を頂きました．研究を一貫して見て頂いただけでなく，大学院受験や就職活動，博士課程進学へ向けて背中を押して頂いたこともあり，ここまで来れたと強く実感しております．心より感謝の念を申し上げます．

そして，本研究の副査を快く了承頂きました池原雅章教授，岡田英史教授，斎藤英雄教授，満倉靖恵准教授には深く感謝いたします．池原教授や岡田教授には修士課程から研究の御審査を頂きまして数多くのコメントを頂きました．斎藤教授には同分野として学内外での活動を支えて頂けただけでなく，訪問研究員として海外に滞在する際に御推薦頂きまして誠に感謝しております．満倉准教授には学内外の研究御指導だけでなく，生活面においても大変多くの御指導を頂きました．

著者が博士論文の研究を遂行するにあたり，独立行政法人交通安全環境研究所の松井靖浩博士，高橋国夫氏には大変多くの御助言を頂きました．特に松井博士からは交通安全の見解だけでなく，研究者としての在り方について学ぶことができ，大変貴重な機会を頂きました．交通安全分野に研究が少しでも役立てられて光栄に思います．

また，博士課程在学中には複数の研究機関に滞在させて頂きました．カリフォルニア大学リバーサイド校の Prof. Bir Bhanu，Dr. Mehran Kafai(現 Hewlett Packard Labs)，独立行政法人産業技術総合研究所の佐藤雄隆博士，岩田健司博士，大西正輝博士，ミュンヘン工科大学の Prof. Nassir Navab，Dr. Slobodan Ilic には約 2 年の間御指導を頂きました．著者が研究を遂行するための貴重なご意見を頂けただけでなく，世界で活躍する基準を示して頂けたことは研究者として今後活動する上での礎となりました．

現在までの共同研究を支えてくれた皆さんにも感謝しております．著者が学部在学中には東芝プラントシステムの山田義浩氏，シスプロの田波厚氏，サイバー大学の宮地恵美氏にお世話になりました．修士課程在学中からはパナソニック株式会社の田靡雅基氏，里雄二氏，大島京子氏，藤田光子氏，丸谷健介氏，加賀屋智之氏と密に監視カメラにおける研究開発を重ねてきました．商用の技術を積み上げていく様子を直に感じ取ることができました．

著者の研究室生活を支えて頂きました慶應義塾大学/芝浦工業大学青木研究室の皆様に感謝の

意を表します．博士課程の先輩として御指導頂きました現大阪大学の中澤満博士，同研究室の博士課程として互いに刺激し合いながら成長できた林昌希氏，松尾清史氏，齊藤俊太氏，橋本潔氏，そして研究室で多くの時間を共に過ごせた仲間に深く感謝致します．

末筆になりますが，厳しくも温かく見守り生活を支えてくれた父・片岡光男氏，努力や仲間の大切さを教え自分の思う方向に導いてくれた母・故片岡清美氏，そして今の自分を形成するために関わった全ての親友にこの場を借りて感謝し，研究者としての新しい物語を進めます．

2013 年 12 月

片岡裕雄

参考文献

[1] T. B. Moeslund, A. Hilton, V. Kruger, L. Sigal, “Visual Analysis of Humans: Looking at People”, ISBN 978-0-85729-997-0, Springer, 2011.

[2] A. Yilmaz, O. Javed, M. Shah, “Object Tracking: A Survey”, ACM Computing Survey, Vol.38, Issue 4, No.13, 2006.

[3] T. B. Moeslund, A. Hilton, V. Kruger, “A survey of advances in vision-based human motion capture and analysis”, Computer Vision and Image Understanding (CVIU), vol.104, pp.90-126, 2006.

[4] Z. Zeng, M. Pantic, G. I. Roisman, T. S. Huang, “A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions”, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol.31, No.1, pp.39-58, 2009.

[5] P. Viola, M. Jones, “Robust Real-time Object Detection”, International Journal of Computer Vision, Vol.57, Issue 2, pp.137-154, 2001.

[6] H. Ling, S. Soatto, N. Ramanathan, D. W. Jacobs, “A Study of Face Recognition as People Age”, International Conference on Computer Vision (ICCV), pp.1-8, 2007.

[7] H. Lu, Y. Huang, Y. Chen, D. Yang, “Automatic Gender Recognition based on Pixel-pattern-based Texture Feature”, Journal of Real-Time Image Processing, Vol.3, Issue1-2, pp.109-116, 2008.

[8] C. Shan, “Smile Detection by Boosting Pixel Differences”, IEEE Transactions on Image Processing, Vol.21, No.1, pp.431-435, 2012.

[9] I. S. Kim, H. S. Choi, K. M. Yi, J. Y. Choi, S. G. Kong, ”Intelligent Visual Surveillance - A Survey”, International Journal of Control, Automation, and Systems, Vol.8, No.5, pp.926-939, 2010.

[10] J. R. Wang, N. Parameswaran, ”Survey of Sports Video Analysis: Research Issues and Applications”, Pan-Sydney area workshop on visual information processing, pp.87-90, 2005.

[11] D. Geronimo, A. M. Lopez, “Vision-based Pedestrian Protection Systems for Intelligent

Vehicles", ISBN 978-1-4614-7987-1, SpringerBriefs in Computer Science, 2014.

[12] R. E. Schapire, Y. Singer, "Improved Boosting Algorithms Using Confidence-rated Predictions", Machine Learning, No.37, pp.297-336, 1999.

[13] T. Gandhi, M. M. Trivedi, "Pedestrian Collision Avoidance Systems: A Survey of Computer Vision Based Recent Studies", IEEE Intelligent Transportation Systems Conference (ITSC), pp.976-981, 2006.

[14] D. Geronimo, A. M. Lopez, A. D. Sappa, T. Graf, "Survey of Pedestrian Detection for Advanced Driver Assistance Systems", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol.32, No.7, pp.1239-1258, 2010.

[15] P. Dollar, C. Wojek, B. Schiele, P. Perona, "Pedestrian Detection: An Evaluation of the State of the Art", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol.34, No.4, pp.743-761, 2012.

[16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision (IJCV), Vol.60, pp.91-110, 2004.

[17] Y. Ke, R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2004.

[18] L. Breiman, "Random Forests", Machine Learning, Vol.45, No.1, pp.5-32, 2001.

[19] V. Lepetit, P. Fua, "Keypoint recognition using randomized trees", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol.28, No.9, pp.1465-1479, 2006.

[20] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Van Gool, "A comparison of affine region detectors", International Journal of Computer Vision (IJCV), Vol.65, No.1, pp.43-72, 2005.

[21] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, "Speeded-Up Robust Features (SURF)", Computer Vision and Image Understanding (CVIU), Vol.110, No.3, pp.346-359, 2008.

[22] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, C. Bray, "Visual categorization with bags of keypoints", Workshop on Statistical Learning in Computer Vision, European Conference on Computer Vision (ECCV), pp.1-22, 2004.

[23] J. B. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations", Symposium on Mathematical Statistics and Probability, University of California Press, pp.281-297, 1967.

[24] B. D. Lucas, T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision", International Joint Conference on Artificial Intelligence, pp.674-679, 1981.

[25] N. Dalal, B. Triggs, C. Schmid, "Human Detection using Oriented Histograms of Flow and Appearance", European Conference on Computer Vision (ECCV), 2006.

[26] D. Gavrila, "A Bayesian, exemplar-based approach to hierarchical shape matching", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol.29, No.8, pp.1408-1421, 2007.

[27] G. Borgefors, "Distance transformations in digital images", Computer Vision Graphics and Image Processing, Vol.34, No.3, pp.344-371, 1986.

[28] A. Broggi, M. Bertozzi, A. Fascioli, M. Sechi, "Shape-based pedestrian detection", IEEE Intelligent Vehicles Symposium, 2000.

[29] M. Bertozzi, A. Broggi, M. Carletti, A. Fascioli, T. Graf, P. Grisleri, M. Meinecke, "IR Pedestrian Detection for Advanced Assistance Systems", DAGM Symposium, 2003.

[30] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.886-893, 2005.

[31] T. Ojala, M. Pietikainen, T. Maenpaa, "Multiresolution grayscale and rotation invariant texture classification with local binary patterns", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, No.7, pp971-987, 2002.

[32] S. Liao, X. Zhu, Z. Lei, L. Zhang, S. Z. Li, "Learning Multi-scale Block Local Binary Patterns for Face Recognition", Advances in Biometrics, Vol.4642, pp.828-837, 2007.

[33] J. Trefny, J. Matas, "Extended Set of Local Binary Patterns for Rapid Object Detection", Computer Vision Winter Workshop, 2010.

[34] K. Levi, Y. Weiss, "Learning Object Detection from a Small Number of Examples: the Importance of Good Features", Computer Vision and Pattern Recognition (CVPR),

pp.53-60, 2004.

[35] B. Wu, R. Nevatia, “Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet Based Part Detectors”, International Journal of Computer Vision (IJCV), Vol.75, pp.247-266, 2007.

[36] P. Sabzmeydani, G. Mori, “Detecting Pedestrians by Learning Shapelet Features”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1-8, 2007.

[37] T. Mita, T. Kaneko, B. Stenger, O. Hori, “Discriminative Feature Co-occurrence Selection for Object Detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol.30, No.7, pp.1257-1269, 2008.

[38] 三井相和，山内悠嗣，藤吉弘亘，“Joint HOG 特徴を用いた 2 段階 AdaBoost による人検出”，画像センシングシンポジウム (SSII)，2008.

[39] 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”，電子情報通信学会論文誌，Vol.92-D-II，No.8, pp.1125-1134，2009.

[40] Y. Yamauchi, M. Takagi, T. Yamashita, H. Fujiyoshi, “Feature Co-occurrence Representation based on Boosting for Object Detection”, Computer Vision and Pattern Recognition Workshop (CVPRW), pp.31-38, 2010.

[41] S. Walk, N. Majer, “New Features and Insights for Pedestrian Detection”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1030-1037, 2010.

[42] 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 CS-HOG の提案 CS-HOG: Color Similarity-based HOG feature”，第 18 回画像センシングシンポジウム (SSII)，IS3-04，2012.

[43] T. Watanabe, S. Ito, K. Yokoi, “Co-occurrence Histograms of Oriented Gradinets for Pedestrian Detection”, Pacific-Rim Symposium on Image and Video Technology (PSIVT), pp.37-47, 2009.

[44] M. Nishiyama, A. Seki, T. Watanabe, “Stereo-based Pedestrian Detection Using Two-stage Classifiers”, IAPR Conference on Machine Vision Applications, pp.520-523, 2011.

[45] M. Hiromoto, R. Miyamoto, “Cascade Classifier Using Divided CoHOG Features for Rapid Pedestrian Detection”, International Conference on Computer Vision Systems

(ICVS), pp.53-62, 2009.

[46] B. Leibe, A. Leonardis, B. Schiele, "Robust Object Detection with Interleaved Categorization and Segmentation", International Journal of Computer Vision (IJCV), Vol.77, pp.259-289, 2008.

[47] M. Andriluka, S. Roth, B. Schiele, "Pictorial Structured Revisited: People Detection and Articulated Pose Estimation", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[48] P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, "Object Detection with Discriminatively Trained Part based Models", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol.32, No.9, pp.1627-1645, 2010.

[49] C. Vondrick, A. Khosla, T. Malisiewicz, A. Torralba, "HOGgles: Visualizing Object Detection Features", International Conference on Computer Vision (ICCV), 2013.

[50] M. Isard, A. Blake, "CONDENSATION - conditional density propagation for visual tracking", International Journal of Computer Vision (IJCV), Vol.29, No.1, pp.5-28, 1998.

[51] D. Ziou, S. Tabbone, "Edge Detection Techniques An Overview", International Journal of Pattern Recognition and Image Analysis, Vol.8, No.4, pp.537-559, 1998.

[52] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms", IEEE Transactions on Systems, Man and Cybernetics (SMC), Vol.9, No.1, pp.62-79, 1979.

[53] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions", Bulletin of the Calcutta Mathematical Society, Vol.35, pp.99-109, 1943.

[54] J. H. Ward, "Hierarchical Grouping to Optimize an Objective Function", Journal of the American Statistical Association, No.58, Vol.301, pp236-244, 1963.

[55] http://pascal.inrialpes.fr/data/human/

[56] S. Munder, D. M. Gavrila, "An Experimental Study on Pedestrian Classification", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol.28, No.11, pp.1863-1868, 2006.

[57] J. Begard, N. Allezard and P. Sayd, "Real-Time Human Detection in Urban Scenes:

Local Descriptors and Classifiers Selection with AdaBoost-like Algorithms", Proc. IEEE CVPR Workshops 2008, pp.1-8, 2008

[58] Yibo Cui, Lifeng Sun, Shiqiang Yang, "Pedestrian Detection Using Improved Histogram of Oriented Gradients", Visual Information Engineering (VIE2008), pp.388-392, 2008

[59] C.F.Olson, "Parallel Algorithms for Hierarchical Clustering", Parallel Computing, Vol.21, pp.1313-1325, 1995

[60] 財団法人交通事故総合分析センター, "交通統計 平成 21 年版", 2010.

[61] 内閣府, "平成 21 年交通安全白書", 2009.

[62] 国土交通省, "道路運送車両の保安基準", 2012.

[63] 柴田英司, "新開発ステレオカメラによる運転支援システム「EyeSight」の開発", 自動車技術, Vol.63, No.2, pp.93-98, 2009

[64] 葛巻清吾, "安全への取り組み", 自動車技術, Vol.63, No.12, pp.11-16, 2009

[65] 財団法人交通事故総合分析センター, ITARDA インフォメーション, No.53, 2004

[66] P. Dollar, Z. Tu, H. Tao, S. Belongie, "Feature Mining for Image Classification", IEEE Conference on Computer Vision and Pattern Recognition, 2007.

[67] T. Quack, V. Ferrari, B. Leibe, L. V. Gool, "Efficient Mining of Frequent and Distinctive Feature Configurations", International Conference on Computer Vision, 2007.

[68] R. Agrawal, T. Imielinski, A. N. Swami, "Mining association rules between sets of items in large databases", ACM SIGMOD International Conference on Management of Data, pp.207-216, 1993.

[69] C. Cortes, V. Vapnik, "Support-Vector Networks", Machine Learning, vol.20-3, pp.273-297, 1995.

[70] A. Gilbert, J. Illingworth, R. Bowden, "Action recognition using mined hierarchical compound features", IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI), Vol.33, No.5, pp.883-897, 2011.

[71] T. Kobayashi, "BoF meets HOG: Feature Extraction based on Histograms of Oriented p.d.f. Gradients for Image Classification", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.747-754, 2013.

[72] 片岡裕雄，青木義満，“行動理解とデータマイニングを適用した人物意図推定・行動予測”，画像の認識・理解シンポジウム (MIRU2012)，2012.

[73] S. Pellegrini, A. Ess, K. Schindler, L. V. Gool, “You’ll Never Walk Alone: Modeling Social Behavior for Multi-target Tracking”, International Conference on Computer Vision (ICCV), pp.261-268, 2009.

[74] K. M. Kitani, B. D. Ziebart, D. Bangnell, M. Herbert, “Activity Forecasting”, European Conference on Computer Vision (ECCV), pp.201-214, 2012.

[75] B. Zhou, X. Tang, X. Wang, “Measuring Crowd Collectiveness”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.

[76] H. Kataoka, K. Iwata, Y. Satoh, I. Yoda, M. Onishi, Y. Aoki, “Big Trajectory Data Analysis for Clustering and Anomaly Detection”, IAPR Conference on Machine Vision Applications (MVA), 2013.

[77] M. S. Ryoo, J. K. Aggarwal, “Spatio-temporal Relationship Match: Video Structure Comparison for Recognition of Complex Human Activities”, IEEE International Conference on Computer Vision, pp.1593-1600, 2009.

[78] H. Kataoka, K. Tamura, K. Iwata, Y. Satoh, Y. Matsui, Y. Aoki, “Extended Feature Descriptor and Vehicle Motion Model with Tracking-by-detection for Pedestrian Active Safety”, IEICE Transactions on Information and Systems, 2014.

[79] A. Geiger, P. Lenz, R. Urtasun, “Are We Ready for Autonomous Driving? The KITTI Vision Behchmark Suite”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.

[80] 永井正夫，鎌田実，“交通事故ゼロを目指して新たな取り組みへ -ヒヤリハットデータベースの紹介-”，社団法人自動車技術会 Press Information，2009.

[81] I. Kitahara, Y. Ohta, H. Saito, S. Akimichi, T. Ono, T. Kanade, “Recording of Multiple Videos in Large-scale Space for Large-scale Virtualized Reality ”, Proceedings of International Display Workshops (AD/IDW’01), pp. 1377-1380, 2001.

[82] T. Koyama, I. Kitahara, Yuichi Ohta, “Live Mixed-Reality 3D Video in Soccer Stadium”, ISMAR, pp.178-187, 2003.

[83] N. Kasuya, I. Kitahara, Y. Kameda, Y. Ohta, “Watching a Player’s View in Real Soccer Scenes by Using Player Trajectories”,The Second Korea-Japan Workshop on Mixed Reality (KJMR), 2009.

[84] N. Inamoto, H. Saito, “Free Viewpoint Video Synthesis and Presentation from Multiple Sporting Videos”, ICME, 2005.

[85] T. Bebie, H. Bieri, “SoccerMan - reconstructing soccer games from video Sequences”, ICIP1998, pp898-902, 1998.

[86] 島脇巧，三浦純，白井良明，“シーン検索システムのための長時間サッカー中継映像の解析”，情報処理学会 CVIM 研究会, 2004.

[87] D. Yow, B. L. Yeo, M. Yeung, B. Liu, “ANALYSIS AND PRESENTAITON OF SOCCER HIGHLIGHTS FROM DIGITAL VIDEO”, ACCV95, pp.499-503, 1995

[88] J. Assfalg, M. Bertini, C. Colombo, A. D. Bimbo, W. Nunziati, “Automatic Interpretation of Soccer Video for Highlights Extraction and Annotation”, SAC, pp.769-773, 2003.

[89] 熊野雅仁，岩本健，有木康雄，“ボールと選手に着目したディジタルカメラワークの実現法 -ディジタルシューティングによるサッカー解説映像生成システムに向けて-”，画像の認識・理解シンポジウム (MIRU2004)，2004.

[90] 加藤大一郎，“新しい番組制作支援技術知的ロボットカメラと放送番組への応用”，NHK 技研 R&D，No.48，pp.34-47，1998.

[91] 井口泰典，土居元紀，真鍋佳嗣，千原國宏，“スポーツ映像放送のための実時間映像解析によるマルチカメラの自動制御と自動スイッチング”，映像情報メディア学会誌，Vol.56，No.2，pp.271-279 (2002).

[92] Y. Zhang, H. Lu, C. Xu, “Collaborate ball and player trajectory extraction in broadcast soccer video”, ICPR2008, pp.1-4, 2008.

[93] K. Kim, M. Grundmann, A. Shamir, I. Matthews, J. Hodgins, I. Essa, “Motion Fields to Predict Play Evolution in Dynamic Sport Scenes ”, 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR2010), pp.840-847, 2010.

[94] R. Hamid, R. Kumar, M. Grundmann, K. Kim, I. Essa, J. Hodgins, “Player Localization

Using Multiple Static Cameras for Sports Visualization", 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR2010), June 2010.

[95] http://ias.cs.tum.edu/projects/aspogamo/

[96] M. Beetz, S. Gedikli, J. Bandouch, B. Kirchlechner, N. Hoyningen-Huene, A. Perzylo, "Visually tracking football games based on TV broadcasts", Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI), 2007.

[97] S. Gedikli, J. Bandouch , N. Hoyningen-Huene, B. Kirchlechner , M. Beetz, "An Adaptive Vision System for Tracking Soccer Players from Variable Camera Settings", Proceedings of the 5th International Conference on Computer Vision Systems (ICVS), 2007.

[98] M. Beetz, J. Bandouch, S. Gedikli, N. Hoyningen-Huene, B. Kirchlechner, A. Maldonado, "Camera-based Observation of Football Games for Analyzing Multi-agent Activities", Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2006.

[99] M. Beetz, B. Kirchlechner, M. Lames, "Computerized Real-Time Analysis of Football Games", IEEE Pervasive Computing, 2005.

[100] L. Gorelick, M. Blank, E. Shechtman, M. Irani, R. Basri, "Actions as space-time shapes", ICCV2005, 2005.

[101] C. Schuldt, I. Laptev, B. Caputo, "Recognizing human actions: A local SVM approach", ICPR2004, 2004.

[102] Y. Freund, R. E. Schapire, "A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting", EuroCOLT1995, pp.23-37, 1995.

[103] A. Klaser, M. Marszalek, C. Schmid, "A spatio-temporal descriptor based on 3D-gradients",BMVC2008.

[104] B. Wu, R. Nevatia, "Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors", International Conference on Computer Vision, pp.90-97, 2005.

[105] R. Chaudhry, A. Ravichandran, G. Hager, R. Vidal, "Histograms of oriented optical flow and binet cauchy kernels on nonlinear dynamical systems for the recognition of

human actions", CVPR2009, pp.1932-1939, 2009.

[106] M. Bregonzio, S. Gong, T. Xiang, "Recognising action as clouds of space-time interest points", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[107] S. Satkin, M. Herbert, "Modeling the temporal extent of actions", ECCV2010, pp.536-548, 2010.

[108] Z. Lin, Z. Jiang, L. Davis, "Recognizing Actions by Shape-Motion Prototype Trees", International Conference on Computer Vision (ICCV), pp.444-451, 2009.

[109] H. Jhuang, T. Serre, L. Wolf, T. Poggio, "A Biologically Inspired System for Action Recognition", International Conference on Computer Vision (ICCV), 2007.

[110] J. K. Aggarwal, M. S. Ryoo, "Human Activity Analysis: A Review", ACM Computing Survey, Vol.43, Issue 3, No.16, 2011.

[111] J. K. Aggarwal, Q. Cai, "Human motion analysis: A review", Computer Vision and Image Understanding (CVIU), Vol.73, 3, pp.428-440, 1999.

[112] O. D. Lara, M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors", IEEE Communications Surveys & Tutorials, Vol.15, No.3, pp.1192-1209, 2013.