

博士論文

色コントラストに基づく勾配特徴量による
人物検出に関する研究

九州工業大学

大学院生命体工学研究科 脳情報専攻

松村 遼

2019 年

目次

第1章	序論	1
1.1	研究背景	1
1.2	先行研究	2
1.3	研究目的	6
1.4	論文構成	8
第2章	関連手法	9
2.1	はじめに	9
2.2	Histograms of Oriented Gradients	9
2.3	Color Self Similarity	12
2.4	性能評価手法	13
2.5	おわりに	15
第3章	色類似度算出による色コントラストの定量化手法	17
3.1	はじめに	17
3.2	色空間	17
3.3	色類似度算出手法	19
3.4	おわりに	21
第4章	局所色相関分布に基づく Histograms of Oriented Gradients	22
4.1	はじめに	22
4.2	特徴計算法	22
4.3	実験	25
4.3.1	実験概要	25

4.3.2	実験結果	25
4.4	考察	32
4.5	おわりに	36
第 5 章	ピクセル間色コントラストに基づく Histograms of Oriented Gradi-	
	ents	37
5.1	はじめに	37
5.2	特徴計算方法	37
5.3	実験	38
5.3.1	実験概要	38
5.3.2	実験結果	40
5.4	フレーム画像での検出実験	48
5.4.1	実験概要	48
5.4.2	実験結果	48
5.5	考察	50
5.6	おわりに	54
第 6 章	全体を通しての考察と今後の課題	55
6.1	全体を通しての考察	55
6.2	今後の課題	59
第 7 章	結論	61
	謝辞	63
	参考文献	64

第1章 序論

1.1 研究背景

物体検出 [1] は、画像認識タスクの内のひとつであり、入力画像の中から検出対象の物体の位置とクラスを特定するものである。物体検出の中でも、特に人物検出に関する研究が近年盛んに行われている。人物検出技術はロボットビジョンや監視システム [2]、自動車運転補助システムに代表される、Intelligent Transport System (ITS) [3–5] などの様々な分野で応用されており、非常に重要な基盤技術となっている。

人物検出技術の応用例としては、上述した監視システムの他に、マーケティング利用を目的とした人物行動流解析システムにおける人物追跡処理の前処理として用いる事例も挙げられる [6]。ITS 分野における応用事例では、各自動車メーカーから人物検出技術を用いた衝突被害軽減ブレーキシステム [7–12] を搭載した車両が販売され、人物検出技術が実用化されるにまで至っている。

物体検出では、背景差分法 [13–17] に代表されるような背景の見え方を基準に背景ではない前景を検出対象物体として検出する方法と、検出対象物体の見え方を基準に検出対象物体を検出する方法が存在する。これらはそれぞれ背景モデル、物体モデルと呼ばれている [1]。背景モデルに基づく手法では、カメラ自身の位置や視点が変化しない、つまり大きな背景の変動が起こらないことを前提としているものが多く、旋回を行うパンチルトカメラや車載カメラ、移動ロボット搭載カメラなどのような移動観測系においては適用が困難である。一方で、物体モデルに基づく手法では物体自身の見え方を基準に検出を行うものであるため、前述した観測系や様々な環境に適した手法であると言える。

しかしながら、実際の環境下において観測される検出対象は、姿勢・形状の変化、照明変動や個体差による色、模様の変化、輝度コントラストの低下といった見え方の変動が起こりうる。人物検出技術の高精度化、高信頼性の確保のためには、このような検出対象の見え方の変動に対しても頑健である手法の確立が望まれている。

1.2 先行研究

近年の物体モデルに基づく人物検出手法では、対象物体の局所的な情報である局所特徴量と、Adaboost [18] や Real Adaboost [19] に代表されるブースティング、サポートベクタマシン (Support Vector Machine : SVM) [20]、ランダムフォレスト (Random Forest) [21] などの統計的学習法を組み合わせた手法が数多く提案されている。このような局所特徴量と統計的学習法を組み合わせた手法の代表的な例として、Viola らの顔検出法 [22] が挙げられる。Viola らは、局所領域内の輝度差を算出し、これを特徴量とする Haar-like 特徴量を提案している。Haar-like 特徴量は顔検出におけるデファクトスタンダードな特徴量となっている。

人物検出に有効な局所特徴量として、輝度に基づく勾配特徴量が提案されている。Levi らによって提案された Edge Orientation Histograms (EOH) 特徴量 [23] は、ソーベルフィルタにより得られたエッジ画像からエッジ強度とエッジ方向を算出し、局所領域内における、任意の方向同士の累積エッジ強度の累計の比を特徴量としている。また、定義された形状パターンに基づき、エッジの繋がりを捉える Edgelet 特徴量 [24] も提案されている。

これらの中で最も成功を収めている局所特徴量として、Histograms of Oriented Gradients (HOG) 特徴量 [25] が挙げられる。HOG 特徴量は、局所領域内における輝度の勾配方向と強度をヒストグラム化した特徴量であり、検出性能の高さと実装のしやすさにより、多数の拡張がなされている。EHOG 特徴量 [26] は、HOG 特徴量の特徴である、複数の隣接する局所領域による正規化処理を単純化することによって、特徴量の次元削減と計算量削減を行っている。Co-occurrence Histograms of Oriented Gradients (CoHOG) 特徴量 [27] は、勾配方向の組み合わせを利用して、ヒストグラムを作成することにより共起特徴量を算出し、より高精度な検出を可能にした。S-HOG 特徴量 [28] は、入力画像を3つの領域に分割し、それらの領域で勾配方向ヒストグラムを作成することにより、物体の対称性を捉える特徴量を算出している。また、HOG 特徴量や HOG ベースの特徴量は、船舶検出 [28]、歩容認識 [29]、地雷検出 [30] など様々なアプリケーションに応用されている。

上記のように、人物検出では輝度情報に基づいた特徴量が多数提案されている。物体検出においては、輝度情報以外に色情報も一般的に使用されている [31–35]。しか

しながら、人物検出の場合では、人物が着用している衣服の色が個人で異なっており、色情報を特徴量として検出に利用するのは困難である。

この問題を解決し、色情報を人物検出に適用した例として Color Self-Similarity (CSS) 特徴量 [36] が挙げられる。CSS 特徴量は、局所領域間の色の類似度を算出し、特徴量とする。これにより、同一物体らしさや非同一物体らしさを表現することが可能で、人物検出に有効な色情報を捉えることができる。

これまでに説明した HOG 特徴量と CSS 特徴量は、ともに人物検出に有効な特徴量である。しかしながら、HOG 特徴量は輝度コントラストによるエッジ情報を捉えることが可能だが、色情報は捉えることができない。CSS 特徴量は色の相関情報を捉えることが可能だが、色コントラストによるエッジ情報は捉えることができない。つまり、両者はともに色コントラストによる色エッジ情報を捉えることができない。

図 1.1(b),(c) に示すように、輝度コントラストが低下した画像では、赤枠で囲まれた部分のように輝度差による勾配が消失する部位も発生し、人物検出にとって重要なエッジ情報を検出することが不可能になる。対照的に輝度コントラストが低下しても、図 1.1(a) のように色の境界、つまり色コントラストによる勾配は消失せず、色差によって生まれる色エッジ情報を検出することが可能である。後述するように、色エッジ情報は心理学的観点、工学的観点からにおいても重要な情報であることが指摘されている。

人間の視覚システムに関する心理学的研究においては、色は境界の認識に重要な役割を果たしていると報告されている [37, 38]。また、色エッジはより信頼性の高い物質境界の指標になるとも報告されている [39]。

工学分野における画像処理研究では、エッジの 90% がグレースケール画像とカラー画像で一致しており、輝度勾配で検出できない色エッジが 10% 存在すると報告されている [40]。これらのエッジには、図 1.1(a) からわかるように人体部分のエッジも含まれており、人物検出において有効な特徴となるものも含まれているため、これらを検出可能にすることで、人物検出精度が向上すると予想される。

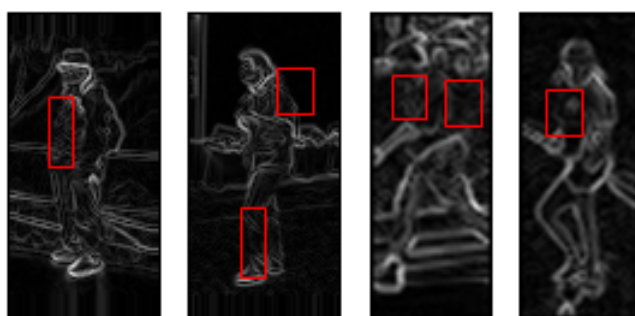
これまで、画像処理分野では、カラー画像の各色チャネルでのエッジ検出結果を統合する、シングルチャネルベースの手法、色情報をベクトルとして扱い、偏微分や順序統計によりエッジ検出する、ベクトルベースの手法といった多数の色エッジ検出手



(a) 入力画像



(b) グレースケール画像



(c) 勾配画像

図 1.1 輝度コントラストが低下した画像

法が提案されてきた [40–43]。しかしながら、これらの研究は色エッジ検出法の提案のみに留まっており、人物検出を含む、物体検出に適用した例は存在しない。Wang らは、色の情報を利用した HOG 拡張の特徴量である color HOG 特徴量 [44] を用いた交通標識検出手法 [45] を提案している。color HOG 特徴量は画像の RGB 各チャンネルで HOG を算出している。上記で挙げた色エッジの検出法の中では、各色チャンネルで得られた結果を統合することで色エッジの検出を行う手法が存在する。color HOG は画像の RGB 各チャンネルに対して HOG を算出するのみで、これらを統合する処理は含まれておらず、色エッジを検出しているとは言い難い。また、明るさの要素が独立していない RGB 色空間を使用しているため、照明変動に対して頑健でないと考えられる。

色エッジを人物検出に適用した例として、後藤ら [46] の取り組みが挙げられる。後藤らは色の類似度に基づく HOG 特徴量である CS-HOG 特徴量 [46] を提案している。CS-HOG 特徴量は、注目領域と画像中の全てのピクセル間での色類似度を算出し、注目領域ごとに類似度画像を作成する。そして、それらに対して HOG 特徴量の算出を行う。これは、注目領域の色に基づいた、前景と背景を分離する領域分割画像を作成していると考えられ、この画像に対して勾配計算を行うことにより、色エッジの検出を行っているものである。人物検出は HOG 特徴量よりも高精度に行えるが、 64×128 ピクセルの画像に対して、特徴計算を行うと特徴量は約 97 万次元という膨大な次元数になってしまう。倍精度実数で特徴量を保持することを考えると、1 枚の画像につき、約 8M バイトの RAM 使用量となる。

Adaboost、Real Adaboost のようなブースティングや SVM、ランダムフォレストは、学習時に全入力画像の特徴量を保持しなければならない、バッチ学習である。INRIA データセット [25] を用いた学習では、解像度 64×128 ピクセルの画像を 7,428 枚使用する。これを考慮すると、学習時に特徴量のみで約 60G バイトの RAM を使用することになる。実際に、後藤らも RAM を 256G バイト搭載したコンピュータで学習を行っており [46]、汎用的な性能のコンピュータでは学習が行えないことが事実として挙げられる。

この他に color-CoHOG 特徴量 [47] が提案されている。color-CoHOG は CoHOG を拡張した共起特徴量であり、上記と同様に色エッジを人物検出に適用した例となって

いる。本特徴量は、RGB 色空間を用いて色エッジの検出を行っている。RGB の各チャンネルで勾配計算を行い、それらを平均することで勾配方向を決定している。color-CoHOG も color HOG と同様に RGB 色空間を用いているため、照明変動に対して頑健でないと考えられる。

また、近年では Convolutional Neural Network (CNN) に代表される、ディープラーニングを用いた高精度な物体検出が実現されている。人物検出においても、ディープラーニングを用いた手法が提案されており [48, 49]、高精度な検出を実現している。文献 [48, 49] はともに、Region proposal ベースの検出手法であり、フレーム画像中から物体候補領域を検出した後に、それらをネットワークへ入力することで人物か否かの識別を行っている。[48] では、Single Shot MultiBox Detector (SSD) [50] で検出された候補領域を入力とするネットワークによる検出結果と、Fully Convolutional Network によるセグメンテーション結果を統合することにより、人物検出を行う。[49] は、Selective Search [51] を候補領域検出に用いた Fast R-CNN [52] による人物検出を行っている。

しかしながらディープラーニングは、学習された特徴はネットワークの重みとして表現されるため、どのような特徴が学習されているのか解析するのが困難だという問題がある。1 節で述べたように、人物検出技術は監視システムや自動車運転補助システム等で用いられている。上記のシステムにおいてディープラーニングを用いた場合、誤認識によって重大な事故・事件が発生した際に、誤認識要因の対策をとることやシステムを利用するユーザに対して判断根拠の説明を行うことが困難となる。

1.3 研究目的

2 節で述べたように、人間の視覚システムにおいて色エッジは重要な役割を果たしており、工学的な観点からも重要な情報であることがわかる。これまでに色エッジを人物検出に適用している例は少なく、CS-HOG 特徴量 [46] や color-CoHOG 特徴量 [47] が挙げられるのみである。

しかしながら、CS-HOG 特徴量は特徴量の次元数が膨大になってしまう問題を抱えている。実際に、後藤らも RAM を 256G バイト搭載したコンピュータで学習を行っており [46]、汎用的な性能のコンピュータでは学習が行えないことが事実として挙げ

られる。color-CoHOG 特徴量は、前節で述べたように色エッジの検出に、明るさの要素が独立していない RGB 色空間を用いるため、照明変動に対して頑健でないと考えられる。また、RGB 各チャネルで得られた結果を平均化して統合しているため、エッジの方向が正確に検出できない可能性も考えられる。

人物検出技術を、より実用的、汎用的な技術にするには、検出精度を向上させるが重要である。局所特徴量を用いた物体検出手法が提案されて以降、局所特徴量同士の関連を捉える共起特徴量 [53–55] が提案されている。2 節で挙げた CoHOG 特徴量及び、上記した color-CoHOG 特徴量も共起特徴量であり、人物検出においても共起特徴量が精度向上に有効であると報告されている。これらの共起特徴量は、局所特徴量の拡張手法であり、基となる局所特徴量を高精度化することで、更なる精度向上が見込める。

上記に加えて、人物検出を利用するシステムのコストを削減することが実用化にとって重要であると考えられる。また、組み込み機器などでは、搭載 RAM が少ないことも考えられるため、特徴量次元を削減することも重要である。

本研究では、色コントラストに基づく勾配特徴量の提案を行い、色エッジの人物検出への適用を行う。提案手法は、色相 (Hue)、彩度 (Saturation)、明度 (Value) で表される HSV 色空間上での色差により算出される色の類似度を色コントラストと捉え、これに対して勾配計算を行うことで色エッジ検出を可能にする。加えて、提案手法における特徴計算法により、CS-HOG と比較して、特徴次元数の削減を可能とする。人物検出実験により、提案手法が低輝度コントラスト条件下でのエッジ検出を可能とし、色エッジ及び色エッジを用いた提案手法が検出精度向上に有効であること、輝度エッジに基づいた既存手法よりも高精度な検出が可能なことを示す。加えて、複数のデータセット及び学習法を用いて提案手法の精度検証を行い、提案手法による精度向上がデータセット依存、学習法依存ではないことを示す。

また、本研究で提案する特徴量は、ブースティング、SVM、ランダムフォレストなどの学習法において、どのような特徴が学習されたか、どのような特徴が識別に寄与したかを解析することが容易であるため、この点がディープラーニングに対して優位な点になると言える。

1.4 論文構成

本論文の構成を以下に示す。本論文は全7章からなる。第1章は序論であり、研究背景と先行研究、研究目的、ならびに論文の構成について述べた。第2章では、提案手法の基となった HOG 特徴量と CSS 特徴量、及び物体検出における性能評価手法について述べる。第3章では、提案手法で採用した色コントラストの定量化手法について述べる。第4章では、色コントラストに基づく勾配特徴量である、局所色相関分布に基づく Histograms of Oriented Gradients を提案し、提案手法と既存手法の性能比較実験を行い、提案手法に対する考察を述べる。第5章では、色コントラストに基づく勾配特徴量である、ピクセル間色コントラストに基づく Histograms of Oriented Gradients を提案し、提案手法と既存手法の性能比較実験を行い、提案手法に対する考察を述べる。第6章では、全体を通しての考察として、他手法との比較やディープラーニングとの比較、提案手法の今後の発展性を述べ、今後の課題を述べる。第7章は結論であり、本研究の統括を述べる。

第2章 関連手法

2.1 はじめに

本章では、本研究の提案手法の基となった特徴量について述べる。また、併せて物体検出における性能評価手法について、その際に必要となる関連手法も含めて述べる。1節では、Histograms of Oriented Gradients (HOG) 特徴量について、2節では、Color Self Similarity (CSS) 特徴量について述べる。3節にて、性能評価に用いる Detection Error Tradeoff (DET) カーブと False Positive Per Window (FPPW)、False Positive Per Image (FPPI) という2つの評価指標について述べる。

2.2 Histograms of Oriented Gradients

Histograms of Oriented Gradients (HOG) 特徴量 [25] とは、セル (図 2.1 : 図では 5×5 ピクセルのサイズ) と呼ばれる局所領域内における輝度の勾配方向と強度をヒストグラム化した特徴量であり、後述する正規化処理により、照明変動や局所的な幾何学変化に頑健という特徴を持つ。以下に HOG 特徴量の計算手順を示す。

まず、各ピクセルの輝度値 L から勾配強度 m と勾配方向 θ を式 (2.1) 及び (2.2) により算出する。次に算出した勾配強度と m と勾配方向 θ を用い、セル内において勾配方向ヒストグラムを作成する。ここで、 $f_x(x, y)$ と $f_y(x, y)$ は、注目ピクセル位置 (x, y) における隣接ピクセル (図 2.2) の輝度値の差分であり、式 (2.3) で算出される。

$$m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \quad (2.1)$$

$$\theta = \tan^{-1} \frac{f_y(x, y)}{f_x(x, y)} \quad (2.2)$$

$$\begin{cases} f_x(x, y) = L_{(x+1, y)} - L_{(x-1, y)} \\ f_y(x, y) = L_{(x, y+1)} - L_{(x, y-1)} \end{cases} \quad (2.3)$$

ただし、算出された勾配方向は0から180度へ変換する。得られた勾配方向は、20度ずつに分割し、9方向のヒストグラムを作成する。最後に、式(2.4)により、任意の数のセルで構成されるブロック(図2.1: 図では 3×3 セルのサイズ)という領域ごとにヒストグラムを正規化する。ここで、 $v(\cdot)$ はヒストグラムのビンの値、つまり特徴量であり、 b はブロック内の任意のセルのビン番号である。 k は、ブロック内全てのヒストグラムのビンの総数である。 ε は計算不能になることを防ぐ係数である。1セルずつブロックを移動させながら、正規化を行っていく。

$$v(b) = \frac{v(b)}{\sqrt{(\sum_{i=1}^k v(i)^2) + \varepsilon}} \quad (2.4)$$

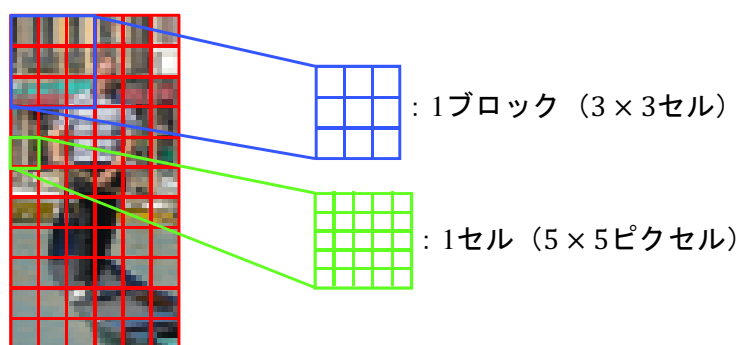


図 2.1 セルとブロックの例

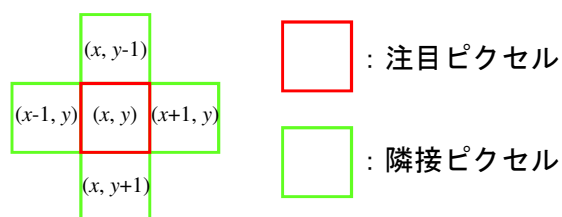
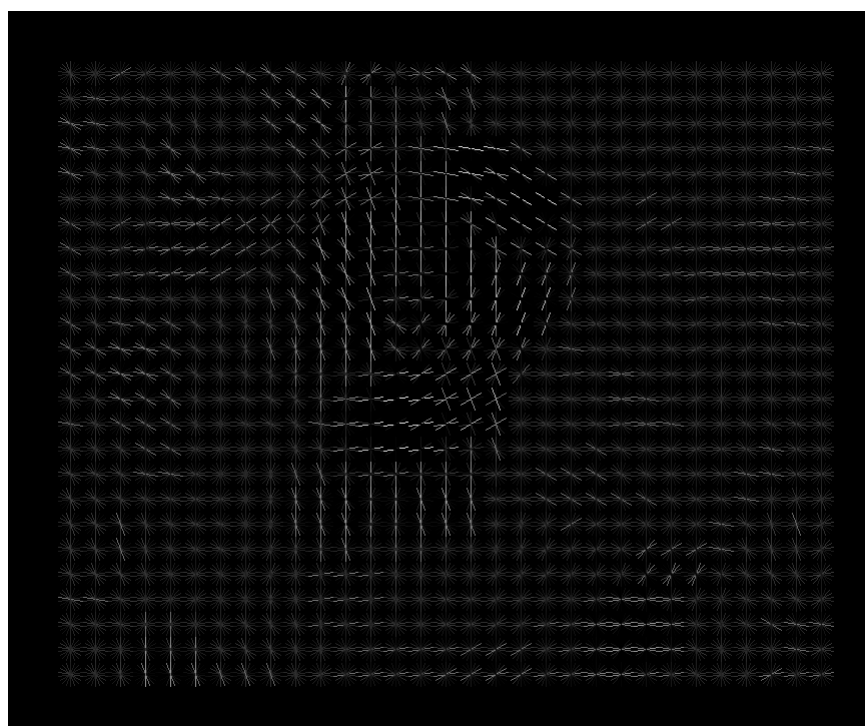


図 2.2 注目ピクセルと隣接ピクセル

図 2.3 に HOG 特徴量を可視化した例を示す。図 2.3(b) 中に示される白い線分が、その局所領域内のエッジであり、明るさにより強度を表している。このように、HOG 特徴量では局所領域内のエッジを捉えることができる。



(a) 入力画像



(b) HOG 可視化画像

図 2.3 HOG 特徴量可視化の例

2.3 Color Self Similarity

Color Self-Similarity (CSS) 特徴量 [36] は、2つのセル間の色の類似度を特徴量とする。図2.4のように、セル領域内で色ヒストグラムを作成し、これを類似度計算に用いている。2つの領域間の色類似度を捉えることにより、同一物体らしさや非同一物体らしさを表現する特徴量となる (図2.5)。

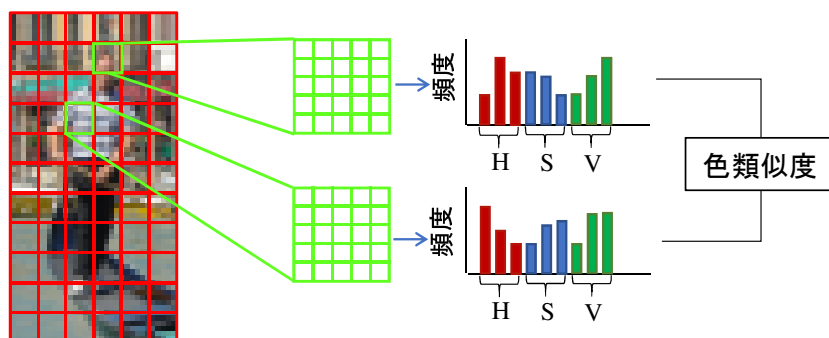
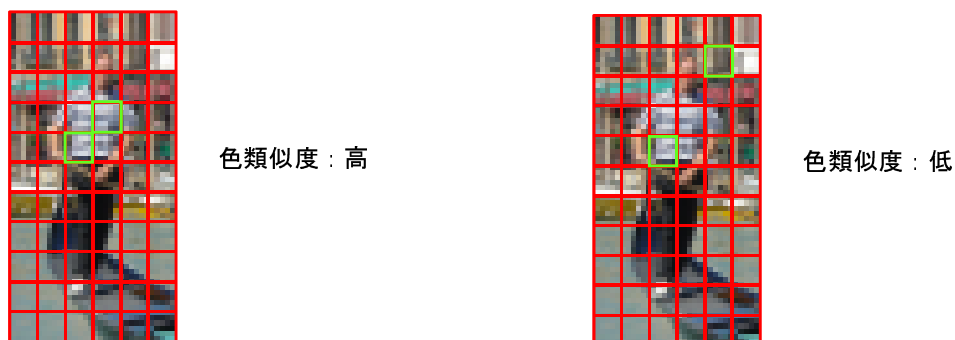


図 2.4 CSS 特徴量



(a) 同一物体らしさを表す領域の組合せ

(b) 非同一物体らしさを表す領域の組合せ

図 2.5 同一物体らしさと非同一物体らしさを表す例

CSS特徴量の算出手順を以下に示す。まず、画像中の各セルで色ヒストグラム $C_{hist} = \{C(H, 1), C(H, 2), \dots, C(H, N), C(S, 1), C(S, 2), \dots, C(S, N), C(V, 1), C(V, 2), \dots, C(V, N)\}$ を作成する。ここで、 $\{H, S, V\}$ は、色相 (Hue)、彩度 (Saturation)、明度 (Value) で表される HSV 色空間での各要素であり、 N は量子化数である。

$C(r \in \{H, S, V\}, n)$ は、以下の式 (2.5) で算出される。このように $C(r, n)$ はセル内の各ピクセルの色 $P_c(x, y, H)$ 、 $P_c(x, y, S)$ 、 $P_c(x, y, V)$ を累積する。

$$C(r, n) = \sum_{x=1}^M \sum_{y=1}^M \delta[f(P_c(x, y, r), n)] \quad (2.5)$$

ここで、 M はセルの大きさ (ピクセル)、 n は量子化の番号 $n = (1, 2, \dots, N)$ 、 $f(P_c(x, y, r))$ は $P_c(x, y, r)$ を量子化した値である。

上記のようにして作成した色ヒストグラムを用いて、セル i, j からヒストグラムインターセクション式 (2.6) により、類似度 $S(i, j)$ を算出する。 B は色ヒストグラムのビン数である。

序論で述べたように、人物検出においては人物が着用している衣服の色が個人で異なっているため、色情報を特徴量として検出に利用するのは困難であった。CSS 特徴量は、色類似度を特徴量とすることで、色情報を有効な特徴量として扱うことに成功している。

$$S(i, j) = \frac{\sum_{b=1}^B \min(C_{hist}^{(i)}[b], C_{hist}^{(j)}[b])}{\sum_{b=1}^B C_{hist}^{(i)}[b]} \quad (2.6)$$

2.4 性能評価手法

物体検出などの検出タスクの評価手法として、Detection Error Tradeoff (DET) カーブ [56] が挙げられる。DET カーブとは、横軸に誤検出率 (False positive rate)、縦軸に未検出率 (False negative rate : Miss rate と呼ばれる) をプロットした両対数グラフである。

物体検出においては、統計的学習法により生成された識別器から出力される値を、閾値によってそれが検出対象か否かを判別する。つまり、閾値で識別器の動作特性が変化するため、各閾値でのエラー値をプロットした性能曲線によって性能を評価する。

DET カーブは、横軸に誤検出率 (False positive rate)、縦軸に感度 (True positive rate) をプロットした Receiver Operating Characteristic (ROC) カーブの変形である。Martin らは、ROC カーブよりも DET カーブの方が、比較対象同士の性能差を容易に観察可能だと述べている [56]。この理由により、物体検出では DET カーブを性能評価手法として用いる場合が多い。

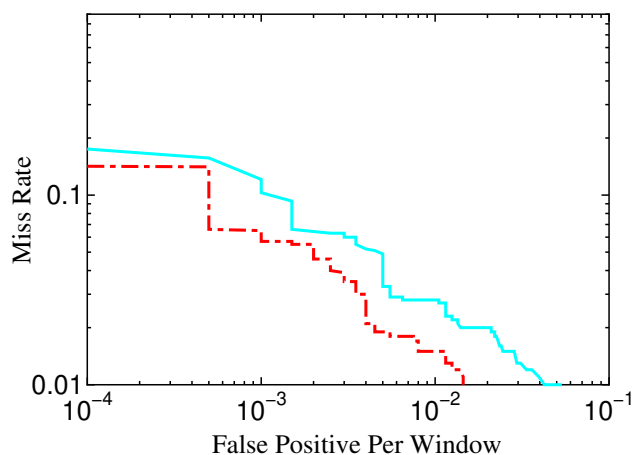


図 2.6 DET カーブの例

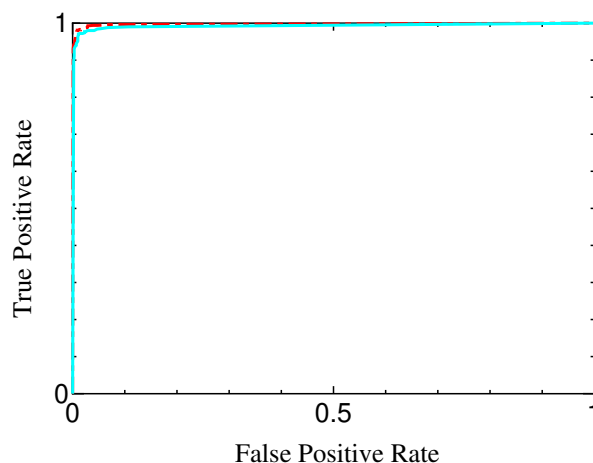


図 2.7 ROC カーブの例

図 2.6 に DET カーブの例を示す。DET カーブでは、曲線が左下の原点に向かうほど性能が良いことを表す。また、図 2.7 に図 2.6 と同様のデータの ROC カーブを示す。図より、DET カーブの方が性能差を容易に把握できることがわかる。

また、物体検出では False Positive Per Window (FPPW) と False Positive Per Image (FPPI) という 2 つの指標が評価に使用される。FPPW とは、検出ウィンドウあたりの誤検出数を表し、FPPI とはフレーム画像あたりの誤検出数を表す。FPPW は、学習用画像と同様の検出対象、非検出対象が映る切り出し画像を入力として与え、1 検出ウィンドウあたりの誤検出数を算出する。FPPI は、検出ウィンドウをフレーム画像中でラスタスキャンした場合のフレーム画像あたりの誤検出数を算出する。そし

て、識別閾値を変化させた場合の FPPW、FPPI とそれぞれに対する未検出率を DET カーブにプロットする。

フレーム画像から物体を検出する場合、画像中に映る物体のスケール変化に対応するため、フレーム画像のサイズを変化させながら検出ウインドウをラスタスキャンする (図 2.8(a))。この場合、検出対象物体付近でその物体が何度も検出される (図 2.8(b))。そのため、検出結果の領域を 1 つに統合する必要がある (図 2.8(c))。先行研究においては、Non Maximum Suppression [57] という領域統合手法が提案されている。Non Maximum Suppression は、領域の重なり度合いを算出し、これが閾値を超えた領域を削除する処理を行う。これ以外に、MeanShift クラスタリングを統合処理に用いることもあるが、Non Maximum Suppression は計算量が少ないことから、物体検出においてよく用いられる。本研究においても、FPPI 評価において本手法を検出結果統合処理に採用する。また、検出結果統合処理の性能は、検出精度に大きく影響するため、高速で精度の良い手法が求められている。

2.5 おわりに

本章では、Histograms of Oriented Gradients (HOG) 特徴量と Color Self Similarity (CSS) 特徴量、性能評価手法である Detection Error Tradeoff (DET) カーブと False Positive Per Window (FPPW)、False Positive Per Image (FPPI) について述べた。

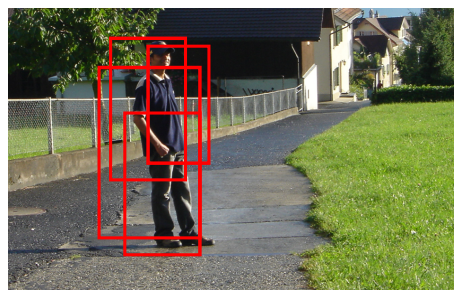
2 節において、特徴量の算出方法を示した。HOG 特徴量は、局所領域内の輝度の勾配方向と強度をヒストグラム化した特徴量であり、物体の大まかな形状を表現することができる。また、HOG 特徴量の可視化画像により、HOG 特徴量が局所領域内のエッジを捉えることが可能なことを示した。

3 節において、CSS 特徴量の算出方法を示した。CSS 特徴量は、局所領域内において色ヒストグラムを作成し、ヒストグラムインターセクションで領域同士の類似度を算出する。これにより、同一物体らしさや非同一物らしさを表現することができる。また、同一物体らしさ、非同一物体らしさを表す領域の組合せの実例を示した。

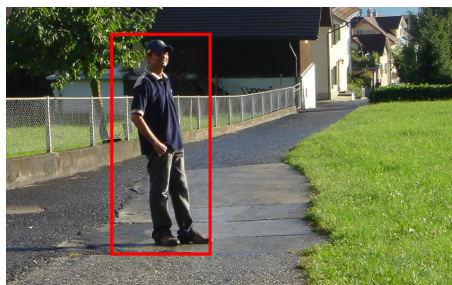
4 節において、DET カーブの例と利点を示した。FPPW と FPPI それぞれの算出方法を示し、FPPI での評価の際に必要な検出ウインドウ統合処理についても述べた。



(a) 検出ウィンドウの走査



(b) 検出結果統合前の例



(c) 検出結果統合の例

図 2.8 フレーム画像での検出例

第3章 色類似度算出による色コントラストの定量化手法

3.1 はじめに

本章では、提案手法で採用する色コントラストの定量化手法について述べる。色コントラストによりエッジを検出するには、色コントラストを定量化し、これに対して勾配計算を行う必要がある。本章2節で、画像処理で一般的に扱われる色空間と提案手法で採用した HSV 色空間について述べる。3節にて、提案手法で採用した色類似度計算手法について述べる。

3.2 色空間

表色系とは、色彩を体系的に表現する方法であり、色を定量的に扱うことができる。表色系は、知覚色に基づく顕色系と、光の混合理論に基づく混色系の2種類に分類することができる。顕色系の代表例としてマンセル表色系が、混色系の代表例として CIE 表色系が挙げられる。マンセル表色系は、色を色相 (Hue)、彩度 (Saturation)、明度 (Value) の三属性によって、色を表現する表色系である。CIE 表色系は、Commission Internationale de l'Éclairage (CIE: 国際照明委員会) により体系化された表色系であり、CIE-RGB、CIE-XYZ、CIE-L*u*v*、CIE-L*a*b*など、様々な表色系が定義されている。

CIE-RGB は、赤 (Red)、緑 (Green)、青 (Blue) を原色とする表色系である。CIE-XYZ は、平均的な観測者によって知覚可能な色を表現できる三刺激値を X、Y、Z として定義し、これによって表現される表色系である。RGB 表色系では正確に再現できない色が存在するという問題を、数学的に回避するために定義された。CIE-L*u*v*と CIE-L*a*b*は、CIE-XYZ の色知覚の不均一性を改善した表色系である。また、CIE-XYZ は基準表色系として定義され、CIE-XYZ から RGB や CIE-L*u*v*、CIE-L*a*b*、それぞれの表色系へと変換できる [58]。

色空間とは、それぞれの表色系で扱われる色要素を座標軸と考え、これにより構成される3次元空間で色を表現するものである。色空間において、色は空間上の位置、座標として決定される。

CIE-RGB 表色系に基づき、RGB を要素とした RGB 色空間や、CIE- $L^*u^*v^*$ 表色系、CIE- $L^*a^*b^*$ 表色系に基づいた、CIE- $L^*u^*v^*$ 色空間、CIE- $L^*a^*b^*$ 色空間、また、マンセル表色系のように、色相 (Hue)、彩度 (Saturation)、明度 (Value) の三属性によって、構成される HSV 色空間といった様々な色空間が存在する。CIE- $L^*u^*v^*$ 色空間において、 L^* は明るさを、 u^* は赤-緑の軸を、 v^* は黄-青の軸を表している。CIE- $L^*a^*b^*$ 色空間も同様に、 L^* は明るさを、 a^* は赤-緑の軸を、 b^* は黄-青の軸を表している。2つの色空間の主な違いは、実装されている色順応モデルにある [58]。RGB、CIE- $L^*u^*v^*$ 、CIE- $L^*a^*b^*$ 、HSV、それぞれの色空間を図 3.1 に示す。

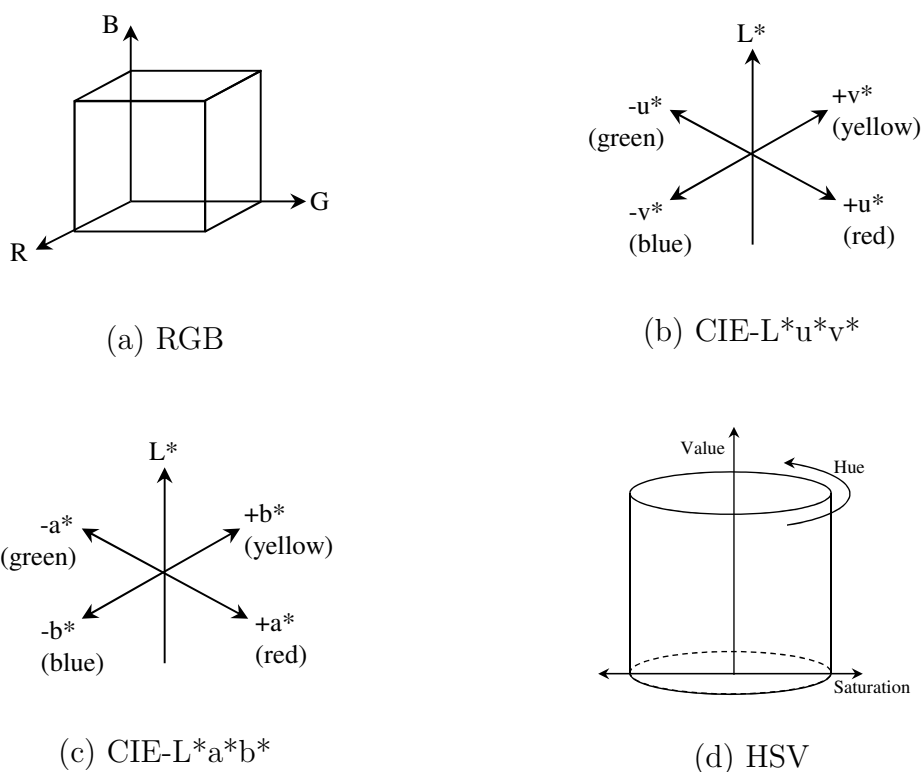


図 3.1 色空間

デジタル画像は、RGB 色空間を用いて色の情報を保存している。つまり、画像の各ピクセルごとに RGB のチャンネルが存在し、ピクセル値を構成している。このこ

とより、画像処理では色情報を扱う際、一般的に RGB 色空間が使用されることが多い [41,42]。しかしながら、RGB 色空間は、その他の色空間のように明るさの要素が独立しておらず、照明変動に対して頑健ではない。そのため、CIE-L*u*v*、CIE-L*a*b*、HSV などの色空間を使用する例も存在する。

RGB 色空間から、CIE-L*u*v*、CIE-L*a*b*色空間へ変換するためには、CIE-XYZ 色空間を経由して変換しなければならない (RGB → XYZ → L*u*v*もしくはL*a*b*)。一方で、HSV 色空間は RGB 色空間からの直接の変換が可能である。RGB 色空間から HSV 色空間への変換を式 (3.1, 3.2, 3.3) で示す。ここで、 R, G, B は RGB それぞれの値であり、MAX と MIN は RGB の値の内、最大と最小のものを表す。

$$H = \begin{cases} \text{undefined}, & \text{if } MIN = MAX \\ 60 \times \frac{G-R}{MAX-MIN} + 60, & \text{if } MIN = B \\ 60 \times \frac{B-G}{MAX-MIN} + 180, & \text{if } MIN = R \\ 60 \times \frac{R-B}{MAX-MIN} + 300, & \text{if } MIN = G \end{cases} \quad (3.1)$$

$$S = \frac{MAX - MIN}{MAX} \quad (3.2)$$

$$V = MAX \quad (3.3)$$

計算コストを考慮すると、HSV 色空間は、CIE-L*u*v*、CIE-L*a*b*色空間よりも少ない計算量で RGB 色空間から変換可能である。また、先行研究では、次節で詳細を述べる色類似度算出において、HSV 色空間が最も有効であったと報告している [36]。これらの点を踏まえ、本研究では色空間に HSV 色空間を採用した。

3.3 色類似度算出手法

本節では、色空間上での色類似度算出手法について説明する。色コントラストは、色差により色類似度として算出する。前節で説明した CIE-L*u*v*色空間や CIE-L*a*b*色空間のような 3 次元色空間上での 2 色間の差は、図 3.2 のようにユークリッド距離として算出される [58]。CIE-L*u*v*色空間上での色差計算式を式 (3.4)、CIE-L*a*b*色空間上での色差計算式を式 (3.5) に示す。

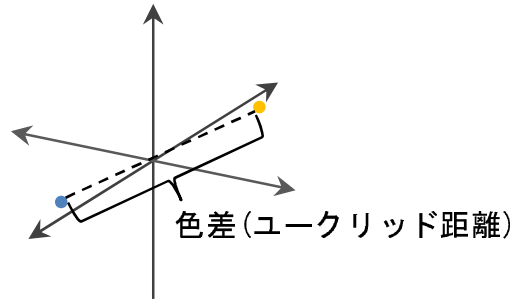


図 3.2 色空間上での色差

$$D_{uv} = \sqrt{(\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2} \quad (3.4)$$

$$D_{ab} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \quad (3.5)$$

色相、彩度、明度の3軸で表現される HSV 色空間上においても、色差は上記のようにユークリッド距離として算出することができる [59]。しかしながら、HSV 色空間は円柱座標系で表現されているため、色相、彩度、明度の各要素値を直接距離計算に使用することはできない。そこで、以下の式 (3.6) を用いて、色相、彩度、明度の各要素を $\{u, r, v\}$ で表現される直交座標系へと変換する。

$$\begin{cases} u = \cos H \times S \\ r = \sin H \times S \\ v = V \end{cases} \quad (3.6)$$

座標系変換後、以下の式 (3.7) を用いてユークリッド距離を算出し、式 (3.8) により、類似度を算出する。

$$D_{ur} = \sqrt{(\Delta u)^2 + (\Delta r)^2 + (\Delta v)^2} \quad (3.7)$$

$$S = \frac{1}{D_{ur} + 1} \quad (3.8)$$

提案手法では、このようにして得られた色コントラストを勾配計算に用いる。

3.4 おわりに

本章では、提案手法で用いる色コントラスト定量化手法について述べた。2節において、画像処理で一般的に扱われる色空間と提案手法で採用した色空間について述べた。また、3節で色類似度算出手法について述べた。色コントラストは、色差により色類似度として算出すること及び提案手法で採用する HSV 空間上での色差がユークリッド距離で算出できることを示した。

本研究では、上記のように色コントラストを色類似度として、ユークリッド距離に基づいて算出し、提案手法での勾配計算に用いる。このように提案手法では、ベクトルとして表される色情報を、スカラーとして扱っている。

提案手法の狙いは、輝度差が大きくないが色差が存在するような状況下での、エッジ検出を可能にすることである。人物検出の場合、前景となる人物（主に人物が着用する衣服）と背景との間の輝度差や色差に規則性は存在しないと考えられる。そのため、輝度差や色差といったベクトル情報を、ユークリッド距離のようなスカラー情報に単純化しても問題ないとする。

文献 [40, 41, 43] における色エッジ検出法は、各色チャンネル上で得られた結果を統合する必要があったり、色情報をベクトルとして扱い、ベクトル演算によりエッジ検出を行うため、色エッジ検出のための計算コストが高い。また、多くの手法が RGB 色空間を使用しているため [41–43]、照明変動に頑健ではない。前述したように、提案手法は輝度差や色差のようなベクトル情報を、スカラー情報に単純化している。そのため、先行研究の色エッジ検出手法に比べて計算コストは低い。加えて、色空間として HSV 色空間を採用しているため、照明変動に対して頑健であると考えられる。

第4章 局所色相関分布に基づく Histograms of Oriented Gradients

4.1 はじめに

本章では、色コントラストに基づく勾配特徴量である、局所色相関分布に基づく Histograms of Oriented Gradients (HOG) [60–62] を提案する。提案手法は、局所領域内において、色コントラストによる色相関分布を作成する。これに対して勾配計算を行い、特徴量を算出する。色コントラストは、注目ピクセルと近傍ピクセル間で算出される。色相関分布の作成は、注目ピクセルの色を基準とした領域分割と考えることができる。これにより、局所領域内における物体の形状を抽出することが可能で、色コントラストに基づいた色エッジを検出することができる。

本章2節で提案手法の詳細について述べ、3節にて実験を行い、既存手法との性能比較を行う。4節にて考察を述べる。

4.2 特徴計算法

局所色相関分布に基づく HOG の特徴計算法について述べる。まず、局所領域 (セル) 内で注目ピクセルと近傍ピクセル (図 4.1) との色類似度を計算する。3章で述べた手順と同様にして色コントラスト算出を行う。また、注目ピクセルは、セルの中心ピクセルである。注目ピクセルと近傍ピクセルそれぞれの HSV 各要素を直交座標 $\{u, r, v\}$ へ変換した後、式 (4.1) でユークリッド距離 $D_{i:n}$ を算出し、式 (4.2) で類似度 S_n を算出する。

$$D_{i:n} = \sqrt{(u_i - u_n)^2 + (r_i - r_n)^2 + (v_i - v_n)^2} \quad (4.1)$$

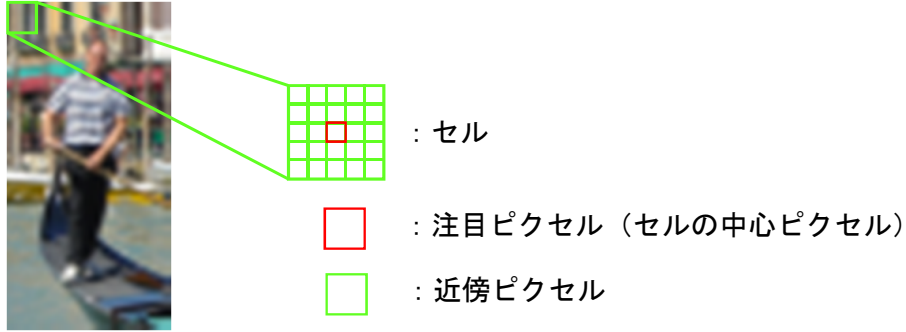


図 4.1 セル及びセル内における注目・近傍ピクセル

$$S_n = \frac{1}{D_{i:n} + 1} \quad (4.2)$$

ここで、 i は注目ピクセルの位置 (x, y) を表す。 n は近傍ピクセルの位置 $(x+m, y+m)$ を表す。 m は $-\frac{M}{2} \sim \frac{M}{2}$ の範囲であり、 M はセルサイズである。上記により、局所色相関分布 $CCorrelation = S_{(x+m, y+m)}$ が作成される。

次に、作成した色相関分布に対して Histograms of Oriented Gradients (HOG) [25] と同様の勾配計算を行う。式 (4.3) 及び (4.4) で勾配強度と方向を算出し、勾配方向ヒストグラムを作成する。ただし、算出された勾配方向は 0 から 180 度へ変換する。得られた勾配方向は、20 度ずつに分割し、9 方向のヒストグラムとなる。ここで、式 (4.3)～(4.5) の (x, y) は勾配計算における注目ピクセル位置である。

セルは 1 ピクセルずつオーバーラップして移動させながら、勾配方向ヒストグラムを作成する。また、各セルで得られたヒストグラムのビン値をセルごとに順に並べていき、特徴ベクトル $[v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, \dots, v_N]$ を構成し、特徴量とする。 $v_1 \sim v_9$ は、1 番目のセルの 1～9 ビンの強度値であり、以降、2 番目のセルの 1～9 ビンの強度値、3 番目のセルの 1～9 ビンの強度値と並んでいく。このため、提案手法において、特徴ベクトルの次元はエッジの方向、空間的位置についての情報を表すことになる。図 4.2 に本提案手法での特徴計算の概要を示す。

$$m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \quad (4.3)$$

$$\theta = \tan^{-1} \frac{f_y(x, y)^2}{f_x(x, y)^2} \quad (4.4)$$

$$\begin{cases} f_x(x, y) = S_{(x+1, y)} - S_{(x-1, y)} \\ f_y(x, y) = S_{(x, y+1)} - S_{(x, y-1)} \end{cases} \quad (4.5)$$

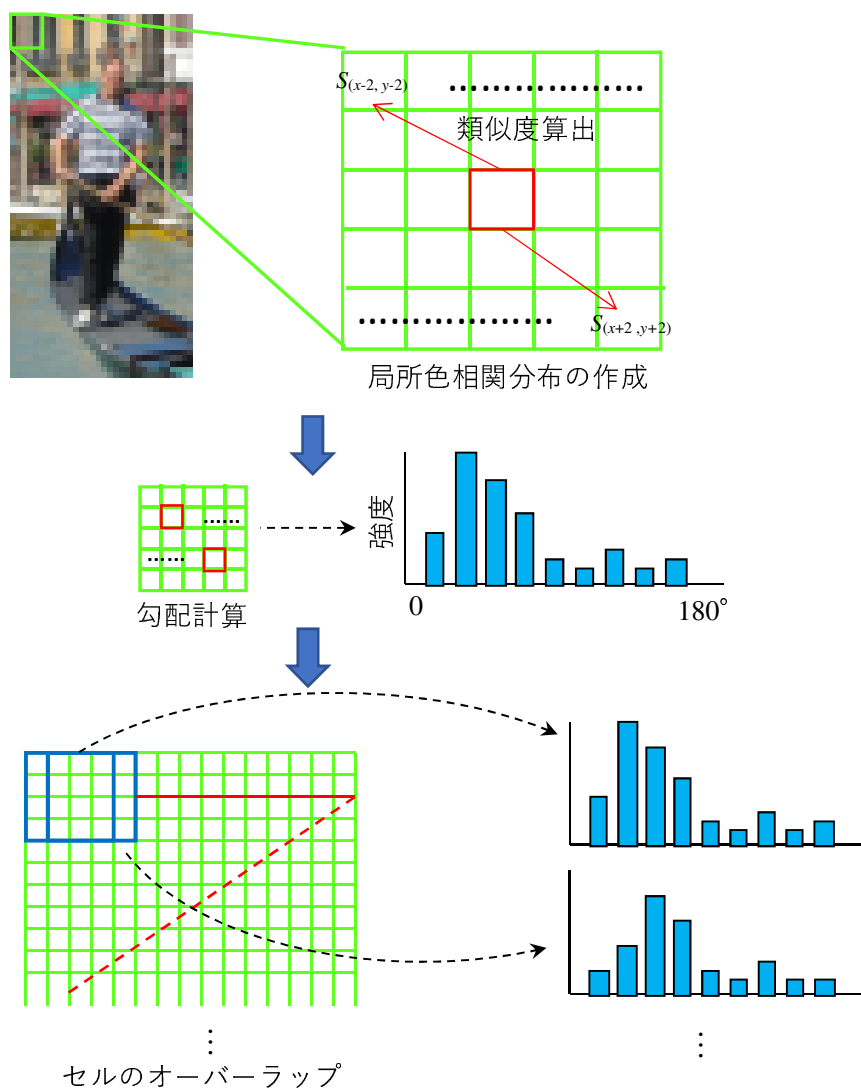


図 4.2 特徴計算概要

4.3 実験

4.3.1 実験概要

本章で提案する局所色相関分布に基づく HOG の人物検出性能評価実験を行う。本実験において、既存手法である HOG [25] との精度比較を行い、提案手法の有効性を確認する。提案手法と HOG は、Real Adaboost [19]、サポートベクタマシン (Support Vector Machine : SVM) [20]、ランダムフォレスト (Random Forest) [21] により学習を行い、識別器を構築する。Real Adaboost は筆者が実装したものを、SVM とランダムフォレストは、それぞれ LIBSVM [63] と Scikit-learn [64] を用いた。Real Adaboost では弱識別器の数を 500 として、ランダムフォレストでは木の数を 300、深さを 4 として学習を行っている。

性能評価手法として、2 章で述べた Detection Error Tradeoff (DET) カーブ [56] を用いる。また、False Positive Per Window (FPPW) により、提案手法と HOG の比較を行う。本実験で用いる人物画像データセットは、INRIA Person Dataset [25] と NICTA Pedestrian Dataset [65] である。学習やテストに用いた画像数内訳を表 4.1 に示す。また、INRIA、NICTA データセットの例を図 4.3 に示す。

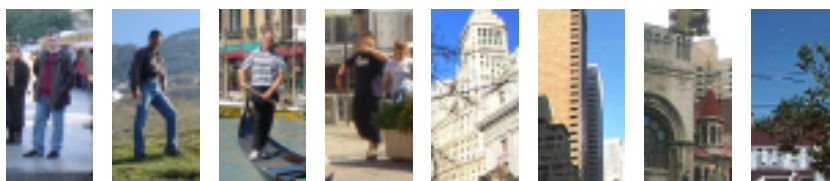
表 4.1 INRIA, NICTA データセットにおける学習・テスト画像数内訳

データセット	学習/テスト	ポジティブサンプル	ネガティブサンプル	計
INRIA	学習	2,416	4,832	7,248
	テスト	1,132	1,132	2,264
NICTA	学習	3,000	6,500	9,500
	テスト	1,000	1,000	2,000

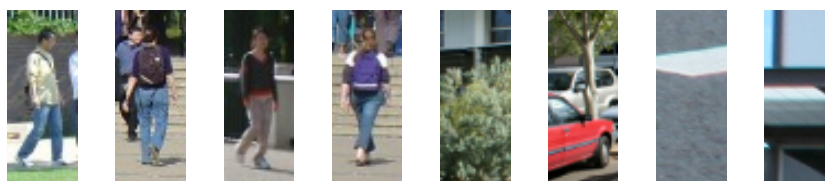
4.3.2 実験結果

図 4.4 および 4.5 に INRIA データセット、NICTA データセットでの実験結果 DET カーブを示す。それぞれの図はともに、(a) が Real Adaboost、(b) が SVM、(c) がランダムフォレストでの結果を示している。

また、実験結果比較の参照点として 10^{-3} FPPW 以下の値を用いる。人物検出などの物体検出においては、誤検出を抑えつつ、未検出を少なくすることが重要である。そのため、実験では低 FPPW での Miss rate 比較を行い、誤検出を抑えた際にどの程



(a) INRIA Person Dataset



(b) NICTA Pedestrian Dataset

図 4.3 本研究で用いた人物画像データセット

度未検出を減少させたかに注目する。本研究においては、提案手法によって誤検出を減少させ、かつ未検出を減少させる方針であるため、このように低 FPPW での評価を行う。INRIA データセットにおいて、 9×10^{-4} FPPW における Miss rate を比較した結果を表 4.2 に示す。NICTA データセットにおいて、 10^{-3} FPPW における Miss rate を比較した結果を表 4.3 に示す。

また、DET カーブで注目した FPPW 区間以外の精度を評価するため、Receiver Operating Characteristic (ROC) カーブと Area Under the Curve (AUC) を用いる。図 4.6 及び 4.7 に INRIA データセット、NICTA データセットにおける ROC カーブを示す。それぞれの図はともに、(a) が Real Adaboost、(b) が SVM、(c) がランダムフォレストでの結果を示している。また、表 4.4 及び 4.5 に INRIA データセット、NICTA データセットにおける AUC を示す。

図 4.4 に示す INRIA データセットにおける実験結果より、提案手法は、Real Adaboost、SVM、ランダムフォレストによる、どの識別器においても、 10^{-3} FPPW 以下の Miss rate の削減が確認できた。表 4.2 より、提案手法は HOG の Miss rate を、Real Adaboost で 3.71 ポイント、SVM では 29.42 ポイント、ランダムフォレストでは 20.05 ポイント削減した。また、表 4.4 より、提案手法は HOG の AUC を Real Adaboost で 0.0047、SVM では 0.0086、ランダムフォレストでは 0.0077 増加させた。

図 4.4 に示す NICTA データセットにおける実験結果でも、提案手法は、Real Adaboost、SVM、ランダムフォレストによる、どの識別器においても、 10^{-3} FPPW 以下の Miss rate の削減が確認できた。表 4.3 より、提案手法は HOG の Miss rate を、Real Adaboost で 2.50 ポイント、SVM では 16.40 ポイント、ランダムフォレストでは 7.90 ポイント削減した。また、表 4.5 より、提案手法は HOG の AUC を Real Adaboost で 0.0014、SVM では 0.0047、ランダムフォレストでは 0.0040 増加させた。

表 4.2 INRIA データセットにおける 9×10^{-4} FPPW での Miss rate 比較

(a) Real Adaboost		(b) SVM		(c) ランダムフォレスト	
	Miss rate		Miss rate		Miss rate
HOG	0.2402	HOG	0.4010	HOG	0.5556
提案手法	0.2031	提案手法	0.1068	提案手法	0.3551

表 4.3 NICTA データセットにおける 10^{-3} FPPW での Miss rate 比較

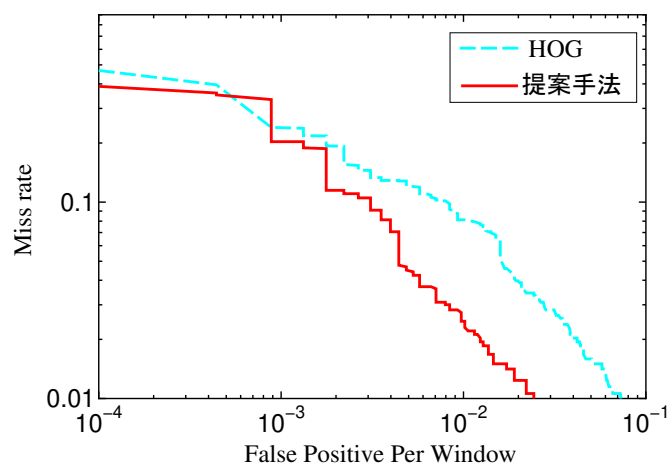
(a) Real Adaboost		(b) SVM		(c) ランダムフォレスト	
	Miss rate		Miss rate		Miss rate
HOG	0.1380	HOG	0.2210	HOG	0.3290
提案手法	0.1130	提案手法	0.0570	提案手法	0.2500

表 4.4 INRIA データセットにおける AUC 比較

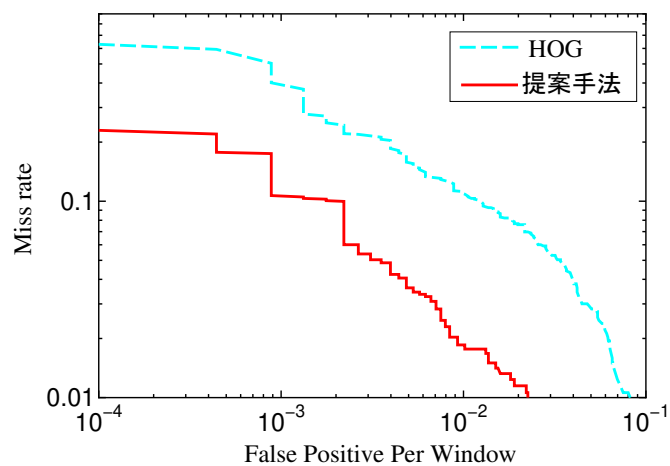
(a) Real Adaboost		(b) SVM		(c) ランダムフォレスト	
	AUC		AUC		AUC
HOG	0.9917	HOG	0.9864	HOG	0.9870
提案手法	0.9964	提案手法	0.9950	提案手法	0.9947

表 4.5 NICTA データセットにおける AUC 比較

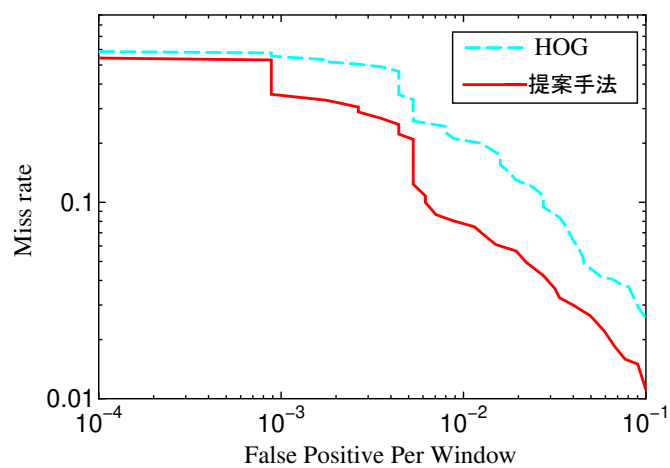
(a) Real Adaboost		(b) SVM		(c) ランダムフォレスト	
	AUC		AUC		AUC
HOG	0.9958	HOG	0.9920	HOG	0.9898
提案手法	0.9972	提案手法	0.9967	提案手法	0.9938



(a) Real Adaboost

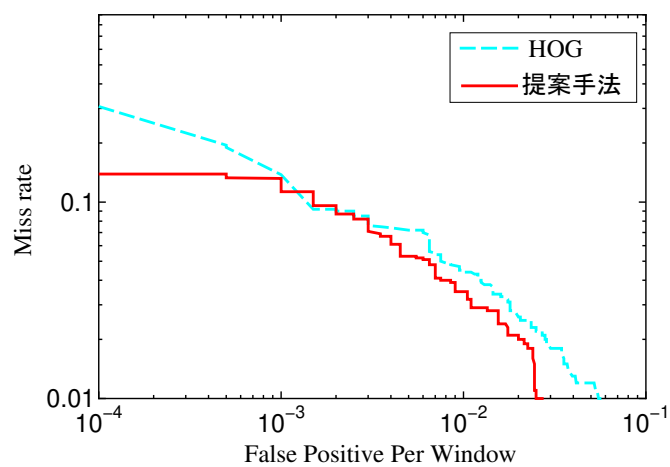


(b) SVM

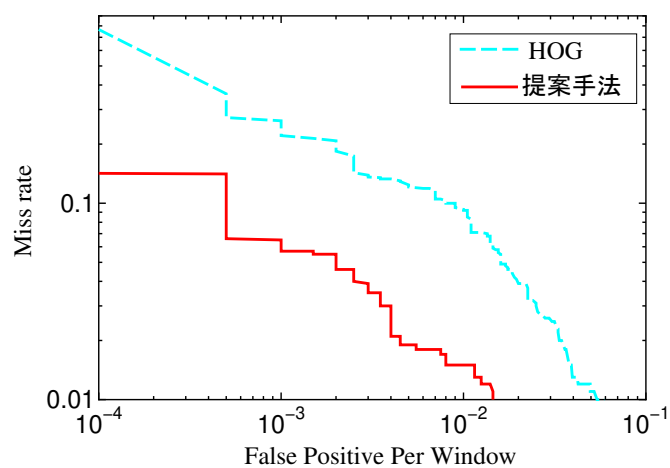


(c) ランダムフォレスト

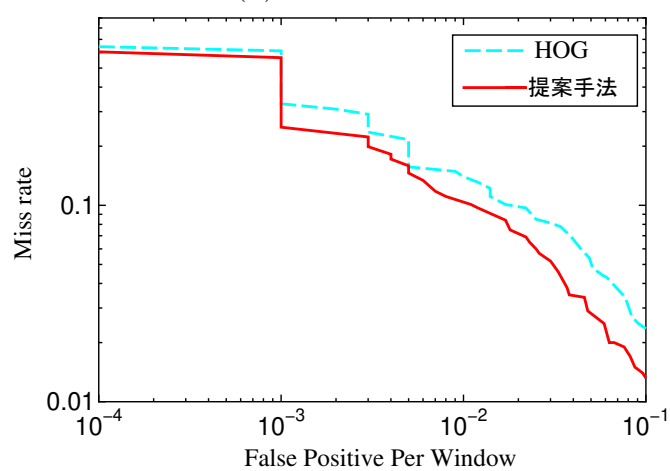
図 4.4 INRIA データセットでの実験結果 DET カーブ



(a) Real Adaboost

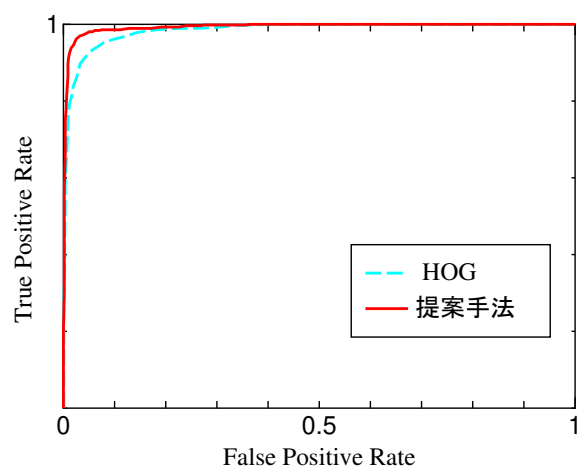


(b) SVM

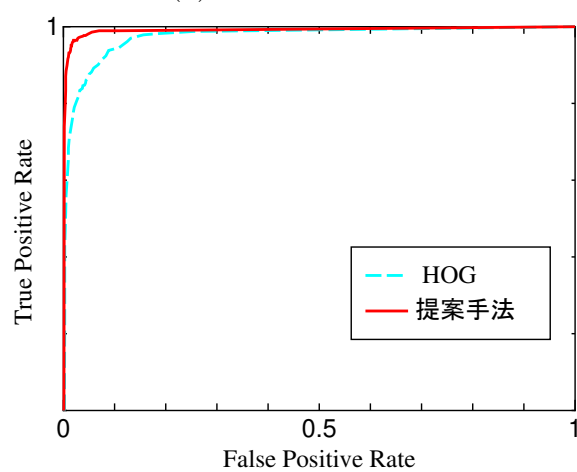


(c) ランダムフォレスト

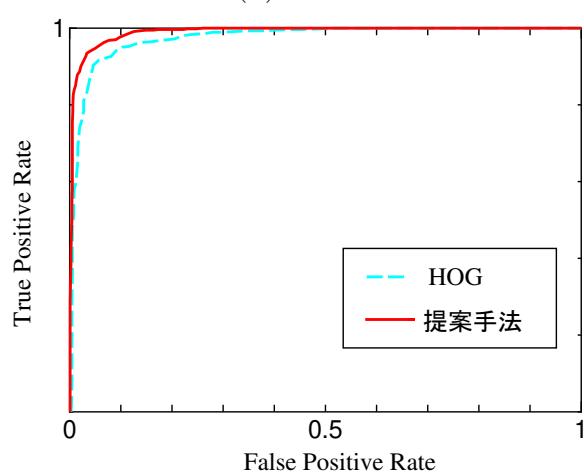
図 4.5 NICTA データセットでの実験結果 DET カurve



(a) Real Adaboost

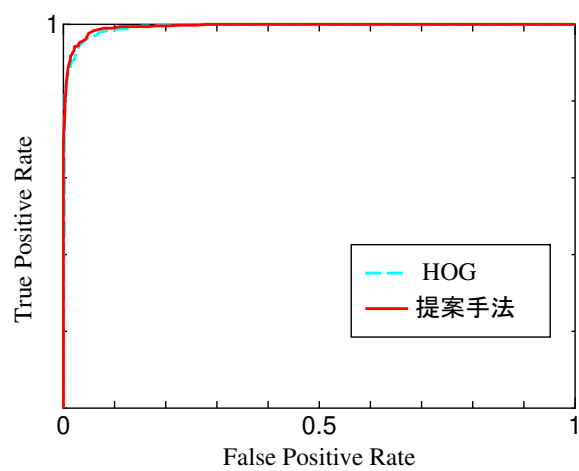


(b) SVM

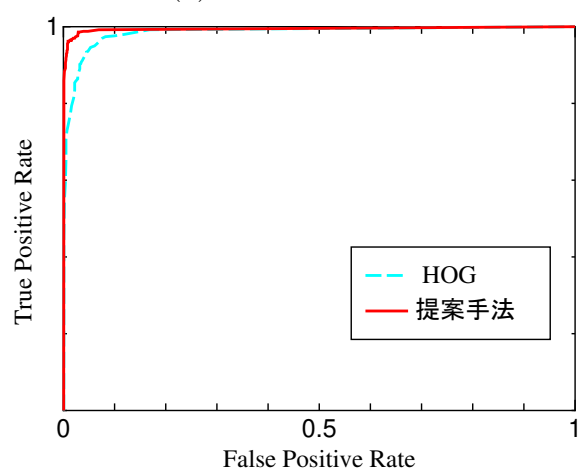


(c) ランダムフォレスト

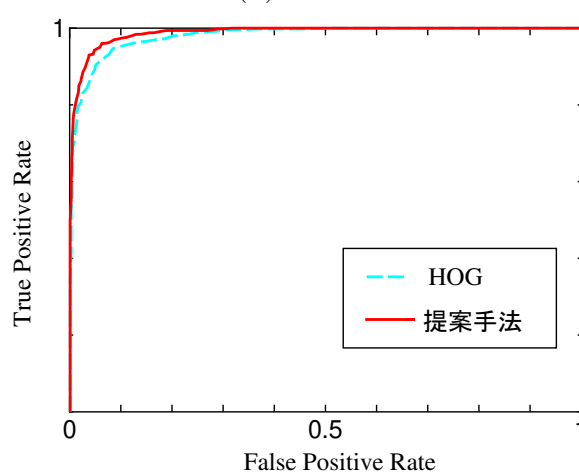
図 4.6 INRIA データセットでの実験結果 ROC カーブ



(a) Real Adaboost



(b) SVM



(c) ランダムフォレスト

図 4.7 NICTA データセットでの実験結果 ROC カurve

4.4 考察

前節の実験により、提案手法である、局所色相関分布に基づく HOG の有効性が確認された。3章4節で述べたように、提案手法の狙いは輝度差はないが色差が存在する状況でのエッジ検出を可能とするものであり、これにより検出精度向上を図っている。

図4.8に局所色相関分布を可視化した例を示す。図において、左列が入力画像であり、赤枠で示される注目セルの輝度分布と局所色相関分布の可視化画像を、右列の上段と下段に示している。局所色相関分布の可視化画像においては、明るいほど類似度が高いことを表す。また、このときのセルの注目ピクセルは、人物が着用している衣服部分となっている。

輝度分布から確認できるように、このセルにおいては輝度コントラストの低下により、人物の上腕部分の輝度差が消失しており、エッジの検出が困難となっている。これに対して、局所色相関分布では、色コントラストに基づく色相関分布によって、前景である人物と背景の分離を行うことで、上腕部分の形状を抽出しており、この部分のエッジ検出が可能となっていることがわかる。このように、提案手法では、輝度コントラストの低下により輝度差が消失した状況でも、色コントラストによる色エッジを検出することが可能であり、これによって人物検出の精度が向上したと考えられる。

図4.9に HOG では識別できなかったが、提案手法により識別可能となった人物画像とそれらの輝度分布及び局所色相関分布の例を示す。図中の1列目、及び3列目が入力画像であり、赤枠で示される注目セルの輝度分布と局所色相関分布の可視化画像を、2、3列目の上段と下段に示している。それぞれの画像中に赤枠で示される注目セルにおいて、輝度コントラスト低下が確認できる。上段1列目の人物は腹部、2列目の人物は脚部、下段1列目の人物は胸部、2列目の人物は背部の輝度差が消失している。輝度情報に基づく HOG においては、これらの部分のエッジ検出ができなかったことが誤識別となっていた原因として挙げられる。

一方で、それぞれのセルにおける局所色相関分布を確認すると、これによって人物形状の抽出が行えており、輝度差では検出できていなかった人物のエッジが検出可能となっていることがわかる。提案手法では色エッジにより人物のエッジを検出するため、輝度エッジよりも、より人物の形状を捉えやすくなっている。これによって、HOG では誤識別となっていた画像も提案手法で正しく識別することが可能となり、

検出精度の向上へと繋がった。

しかしながら、局所色相関分布では、セル内において注目ピクセルの色を基準に前景と背景の分離を行い、人物の形状抽出・エッジ検出を行うため、セルによっては人物形状が抽出できず、人物のエッジを検出することが困難になる場合が存在する。

上記のことが発生する場合のシミュレーション例を図 4.10 に示す。図において、左列に示す画像中の黄色と薄い水色で示される領域が背景、濃い水色で示される領域が人物の領域であると仮定する。このとき、図中に赤枠で示される注目セルの局所色相関分布の可視化画像を右列に示す。注目セルにおいては、黄色で示される背景領域のピクセルが注目ピクセルとなっている。

この例における局所色相関分布を見ると、背景領域の黄色が前景として抽出され、それ以外の色を持つ背景領域と人物領域がともに背景として抽出されている。図 4.11 に実際の人物画像での例を示す。図中に赤枠で示される注目セルにおいて、背景の赤い物体のピクセルが注目ピクセルとなっている。この例では、人物の腕部分が背景に映り込む他の物体とともに背景として抽出されている。このように背景に複数の物体が映り込むような複雑な領域では、人物の形状が抽出されずエッジ検出も行えないため、検出に影響を及ぼす可能性がある。この問題については、学習用画像データセットに上記した複雑な背景を持つ画像を多く含めることで、解決可能だと考えられる。これにより、人物の形状が正しく抽出されない、背景に複数物体が映り込む複雑な領域が、学習時に特徴として選択される可能性が減少すると考えられる。

図 4.12 に、HOG では識別可能であったが、提案手法では識別できなかった画像の例を示す。これらの画像は人物以外にも様々な物体が画像中に映り込んでいる。このように複雑な背景を持つ画像では人物を誤識別してしまうため、画像データセットによっては検出精度の低下が発生する可能性も考えられる。しかしながら、INRIA データセット及び NICTA データセットといった複数のデータセットを用いた本実験において、提案手法は既存手法の HOG の精度を超えることが確認できた。上述した要因により、HOG のような輝度エッジを用いた特徴では識別可能であった画像が、本手法で識別できなくなることがあるが、それ以上に識別可能となった画像のほうが多いといえるため、色エッジによる人物検出は有効だと考えることができる。また、これらの例の画像も上記した対策により、識別可能になると考えられる。

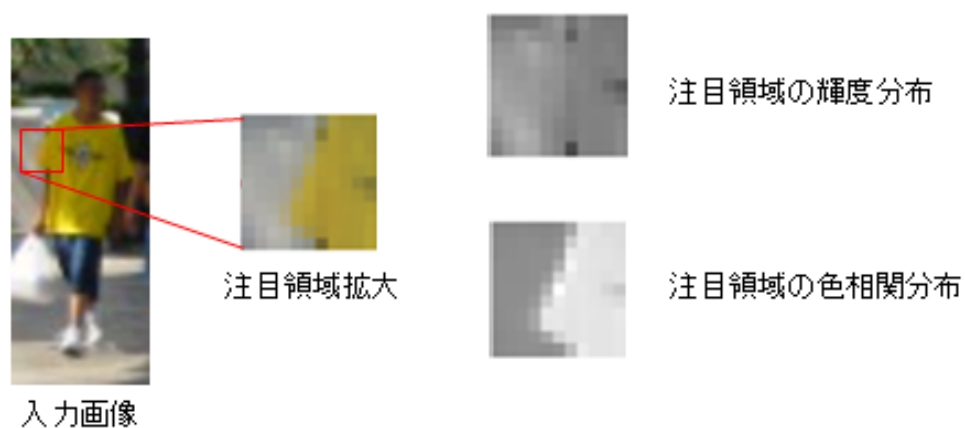


図 4.8 局所色相関分布可視化の例



図 4.9 HOG で識別できなかったが、提案手法で識別可能になった画像例

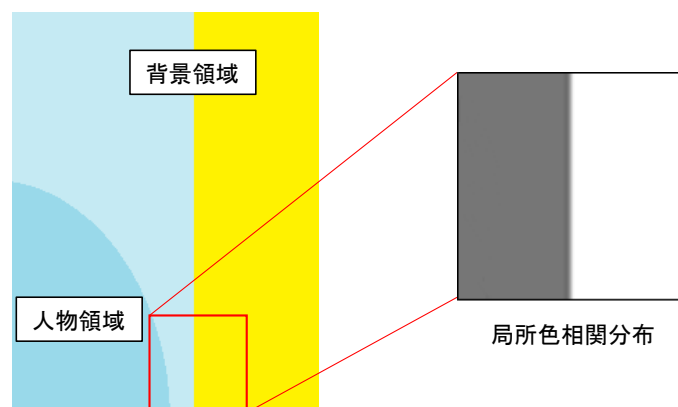


図 4.10 人物形状抽出が困難となる状況のシミュレーション例



図 4.11 実際の画像において人物形状抽出が困難となる例



図 4.12 HOG では識別可能だったが、提案手法では識別できなかった画像例

4.5 おわりに

本章では、色コントラストに基づく勾配特徴量である、局所色相関分布に基づく Histograms of Oriented Gradients (HOG) [60–62] を提案した。2 節で、提案手法の特徴計算法を述べた。また、3 節で実験結果について、4 節で考察について述べた。

提案手法は、局所領域内の注目ピクセルと近傍ピクセルの色コントラストを算出し、色相関分布を作成する。これに対して勾配計算を行うことで、色コントラストに基づく色エッジ検出を可能とし、色エッジを特徴量とすることができる。

人物画像データセットを用いて、既存の輝度情報に基づいた Histograms of Oriented Gradients (HOG) 特徴量 [25] との性能比較実験を行った。DET カーブによる比較、AUC による比較を行い、提案手法により人物検出精度が向上することを示した。

また、局所色相関分布の可視化例や、提案手法により識別が可能となった画像例や識別ができなかった画像例を示し、これらについて考察を行った。これにより、提案手法が低輝度コントラスト状況下でのエッジ検出を可能にし、色エッジが人物検出に有効であることを示した。提案手法により、HOG で識別可能だった画像が誤識別となった例もあったが、それ以上に識別可能となった画像が多かったため、本手法は人物検出に有効な特徴量であると考えることができる。

第5章 ピクセル間色コントラストに基づ く Histograms of Oriented Gradients

5.1 はじめに

本章では、色コントラストに基づく勾配特徴量である、ピクセル間色コントラストに基づく Histograms of Oriented Gradients (HOG) を提案する。前章の局所色相関分布に基づく HOG は、局所領域内において、背景が複雑であった場合に人物の形状を抽出できないことが考えられる。また、色相関分布を作成しなければならないため計算コストが高い。

本章で提案する、ピクセル間色コントラストに基づく HOG は、注目ピクセルと勾配計算の対象となる上下左右の隣接ピクセル間の色コントラストを求め、これに対して勾配計算を行う。前章の手法が、局所領域内の注目ピクセルの色を基準として、領域分割的に物体形状を抽出し色エッジ検出を行うのに対して、本手法は一般的なエッジ検出手法と同様の演算に基づいて、色エッジ検出を行うものである。

本章2節で提案手法の詳細について述べ、3節にて実験を行い、既存手法との性能比較を行う。4節にて考察を述べる。

5.2 特徴計算方法

ピクセル間色コントラストに基づく HOG の特徴計算方法について述べる。まず、注目ピクセルと隣接ピクセル間の色コントラストを、3章で述べた手順と同様にして算出する。注目ピクセルと隣接ピクセルそれぞれの HSV 各要素を直交座標 $\{u, r, v\}$ へ変換した後、式 (5.1) でユークリッド距離 $D_{i:n}$ を算出し、式 (5.2) で類似度 S_n を算出する。ここで、 i は注目ピクセルの位置 (x, y) を表す。 n は隣接ピクセル位置を表し、 $n \in \{(x-1, y), (x+1, y), (x, y-1), (x, y+1)\}$ である。

$$D_{i:n \in \{(x-1,y), (x+1,y), (x,y-1), (x,y+1)\}} = \sqrt{(u_i - u_n)^2 + (r_i - r_n)^2 + (v_i - v_n)^2} \quad (5.1)$$

$$S_{n \in \{(x-1,y), (x+1,y), (x,y-1), (x,y+1)\}} = \frac{1}{D_{i:n} + 1} \quad (5.2)$$

算出した類似度を用いて Histograms of Oriented Gradients (HOG) [25] と同様の勾配計算を行う。式 (5.3) 及び (5.4) で勾配強度と方向を算出し、勾配方向ヒストグラムを作成する。算出された勾配方向は 0 から 180 度へ変換し、20 度ずつに分割して 9 方向のヒストグラムとする。

セルは 1 ピクセルずつオーバーラップして移動させながら、勾配方向ヒストグラムを作成する。各セルで得られたヒストグラムのビン値をセルごとに順に並べていき、特徴ベクトル $[v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, \dots, v_N]$ を構成し、特徴量とする。 $v_1 \sim v_9$ は、1 番目のセルの 1~9 ビンの強度値であり、以降、2 番目のセルの 1~9 ビンの強度値、3 番目のセルの 1~9 ビンの強度値と並んでいく。本提案手法においても、局所色相関分布に基づく HOG と同様に、特徴ベクトルの次元はエッジの方向、空間的位置についての情報を表すことになる。図 5.1 に提案手法での特徴計算の概要を示す。

$$m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \quad (5.3)$$

$$\theta = \tan^{-1} \frac{f_y(x, y)^2}{f_x(x, y)^2} \quad (5.4)$$

$$\begin{cases} f_x(x, y) = S_{(x+1,y)} - S_{(x-1,y)} \\ f_y(x, y) = S_{(x,y+1)} - S_{(x,y-1)} \end{cases} \quad (5.5)$$

5.3 実験

5.3.1 実験概要

本章で提案するピクセル間色コントラストに基づく HOG の人物検出性能評価実験を行う。本実験において、既存手法である HOG [25] との精度比較を行い、提案手法

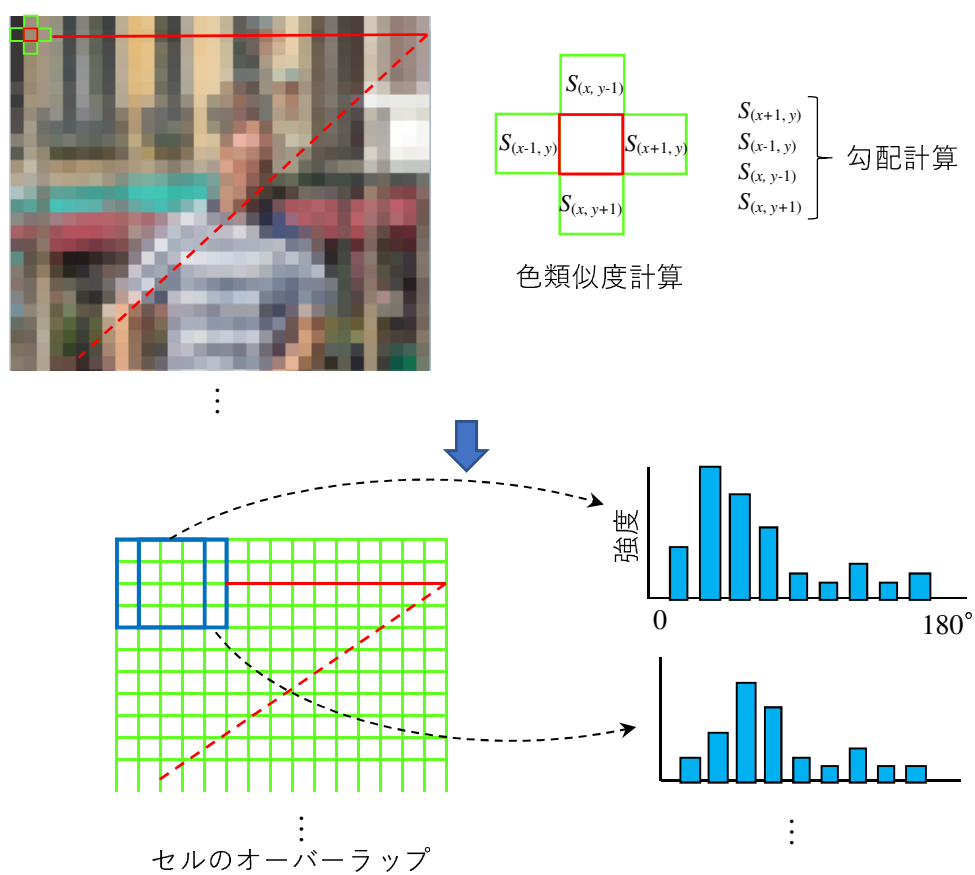


図 5.1 特徴計算概要

の有効性を確認する。また、前章で提案した局所色相関分布に基づく HOG との精度比較も行う。

前章と同様に、Real Adaboost [19]、SVM [20]、ランダムフォレスト [21] により学習を行って識別器を構築し、実験を行う。Real Adaboost とランダムフォレストの弱識別器数、木の数、深さは前章の実験と同様の値としている。

性能評価手法として、Detection Error Tradeoff (DET) カーブ [56] を用いる。また、FPPW により、提案手法と局所色相関分布に基づく HOG 及び HOG の比較を行う。本実験で用いる人物画像データセットは、前章の実験で用いたデータセットと同様の INRIA Person Dataset [25] と NICTA Pedestrian Dataset [65] である。学習やテストに用いた画像数内訳を表 5.1 に示す。

表 5.1 INRIA, NICTA データセットにおける学習・テスト画像数内訳

データセット	学習/テスト	ポジティブサンプル	ネガティブサンプル	計
INRIA	学習	2,416	4,832	7,248
	テスト	1,132	1,132	2,264
NICTA	学習	3,000	6,500	9,500
	テスト	1,000	1,000	2,000

5.3.2 実験結果

図 5.2 及び 5.3 に INRIA データセット、NICTA データセットでの実験結果 DET カーブを示す。それぞれの図はともに、(a) が Real Adaboost、(b) が SVM、(c) がランダムフォレストでの結果を示している。また、局所色相関分布に基づく HOG は提案手法 I、ピクセル間色コントラストに基づく HOG は提案手法 II と示した。

前章での実験と同様に、結果比較の参照点として 10^{-3} FPPW 以下の値を用いる。INRIA データセットにおいて、 9×10^{-4} FPPW における Miss rate を比較した結果を表 5.2 に示す。NICTA データセットにおいて、 10^{-3} FPPW における Miss rate を比較した結果を表 5.3 に示す。

また、本実験においても DET カーブで注目した FPPW 区間以外の精度を評価するため、Receiver Operating Characteristic (ROC) カーブと Area Under the Curve (AUC) を用いる。図 5.4 及び 5.5 に INRIA データセット、NICTA データセットにおけ

る ROC カーブを示す。それぞれの図はともに、(a) が Real Adaboost、(b) が SVM、(c) がランダムフォレストでの結果を示す。表 5.4 及び 5.5 に INRIA データセット、NICTA データセットにおける AUC を示す。

図 5.2 に示す INRIA データセットにおける実験結果より、Real Adaboost、SVM、ランダムフォレストそれぞれによる識別器において、提案手法 II は従来法の HOG と比較して、 10^{-3} FPPW 以下での Miss rate 削減が確認できた。表 5.2 より、提案手法 II は、HOG の Miss rate を Real Adaboost で 1.76 ポイント、SVM では 25.18 ポイント、ランダムフォレストでは 23.32 ポイント削減した。また、表 5.4 より、提案手法 II は、HOG の AUC を Real Adaboost で 0.0043、SVM では 0.0080、ランダムフォレストでは 0.0071 増加させた。

提案手法 I の Miss rate と比較すると、Real Adaboost で 1.95 ポイント増加、SVM では 4.24 ポイント増加している。ランダムフォレストにおいては、3.27 ポイント削減となった。AUC においては、Real Adaboost で 0.0004、SVM では 0.0006、ランダムフォレストでは 0.0006 減少となった。

図 5.3 に示す NICTA データセットにおける実験結果でも、提案手法 II は、Real Adaboost、SVM、ランダムフォレストによる識別器において、 10^{-3} FPPW 以下での Miss rate 削減が確認できた。表 5.3 より、提案手法 II は、HOG の Miss rate を Real Adaboost では 2.70 ポイント、SVM では 11.80 ポイント、ランダムフォレストでは 13.50 ポイント削減した。表 5.5 より、提案手法 II は、HOG の AUC を SVM では 0.0003 増加させた。Real Adaboost、ランダムフォレストでは、それぞれ、0.0001、0.0016 減少している。提案手法 I の Miss rate と比較すると、Real Adaboost で 0.20 ポイント削減、SVM では 4.60 ポイント増加、ランダムフォレストにおいては、5.60 ポイント削減となった。AUC においては、Real Adaboost で 0.0015、SVM では 0.0044、ランダムフォレストでは 0.0056 減少した。

実験結果より、提案手法 II は、従来法の HOG よりも検出性能を向上させたといえる。NICTA データセットにおいて、Real Adaboost とランダムフォレストでの AUC が HOG よりも減少しており、低 FPPW のみでの精度向上となったが、4 章 3 節で述べたように、人物検出などの物体検出においては、誤検出を抑えつつ未検出を少なくすることが重要であるため、低 FPPW における性能改善は意味があると考えられる。

また、提案手法IIと提案手法Iの比較において、提案手法Iの方が、精度が高いと判断できる結果となった。しかしながら、学習法によっては、低 FPPW において、提案手法であるピクセル間色コントラストに基づく HOG の精度を下回ることもあり、学習法による精度差が発生することを確認した。

また、1節で述べたように提案手法Iでは、局所色相関分布を作成しなければならず、色エッジ検出のための計算コストが高い。提案手法IIは、注目ピクセルの上下左右の隣接ピクセル間の色コントラスト計算を行えばよいだけなので、提案手法Iと比較して、計算量の削減も可能である。画像サイズ 64×128 ピクセルの画像において、提案手法Iの計算時間は 520.26[ms]、提案手法IIの計算時間は 346.16ms であり、色エッジ検出のための計算時間を短縮することができた。なお、上記の計算時間は特徴計算処理の最適化や並列化を行っていない状態であり、さらなる高速化が可能である。

表 5.2 INRIA データセットにおける 9×10^{-4} FPPW での Miss rate 比較

(a) Real Adaboost		(b) SVM		(c) ランダムフォレスト	
	Miss rate		Miss rate		Miss rate
HOG	0.2402	HOG	0.4010	HOG	0.5556
提案手法 I	0.2031	提案手法 I	0.1068	提案手法 I	0.3551
提案手法 II	0.2226	提案手法 II	0.1492	提案手法 II	0.3224

表 5.3 NICTA データセットにおける 10^{-3} FPPW での Miss rate 比較

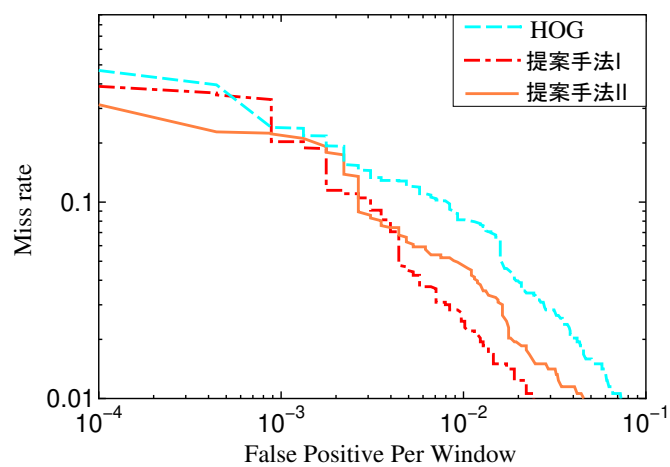
(a) Real Adaboost		(b) SVM		(c) ランダムフォレスト	
	Miss rate		Miss rate		Miss rate
HOG	0.1380	HOG	0.2210	HOG	0.3290
提案手法 I	0.1130	提案手法 I	0.0570	提案手法 I	0.2500
提案手法 II	0.1110	提案手法 II	0.1030	提案手法 II	0.1940

表 5.4 INRIA データセットにおける AUC 比較

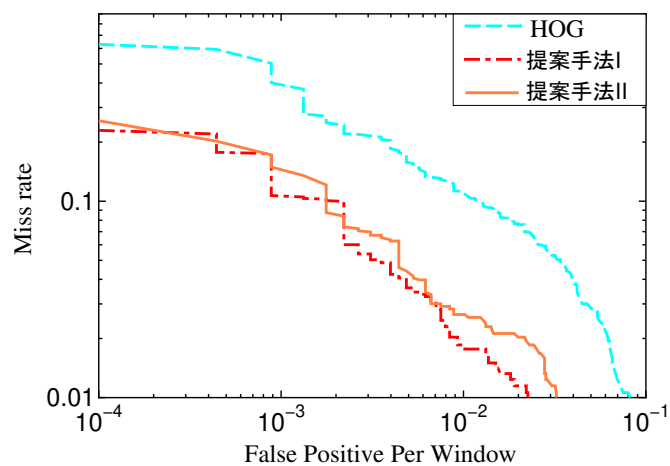
(a) Real Adaboost		(b) SVM		(c) ランダムフォレスト	
	AUC		AUC		AUC
HOG	0.9917	HOG	0.9864	HOG	0.9870
提案手法 I	0.9964	提案手法 I	0.9950	提案手法 I	0.9947
提案手法 II	0.9960	提案手法 II	0.9944	提案手法 II	0.9941

表 5.5 NICTA データセットにおける AUC 比較

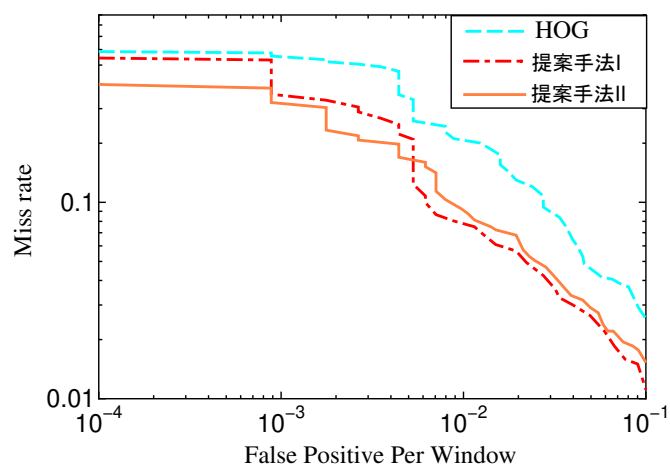
(a) Real Adaboost		(b) SVM		(c) ランダムフォレスト	
	AUC		AUC		AUC
HOG	0.9958	HOG	0.9920	HOG	0.9898
提案手法 I	0.9972	提案手法 I	0.9967	提案手法 I	0.9938
提案手法 II	0.9957	提案手法 II	0.9923	提案手法 II	0.9882



(a) Real Adaboost

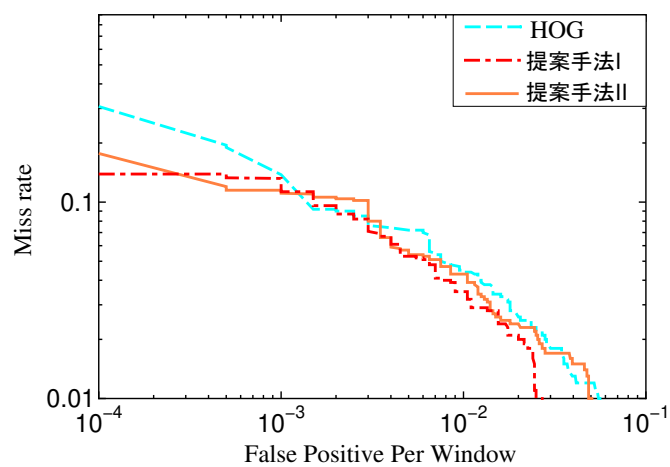


(b) SVM

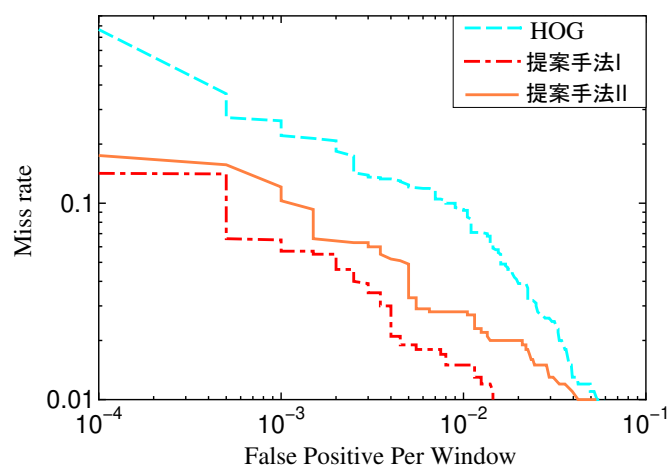


(c) ランダムフォレスト

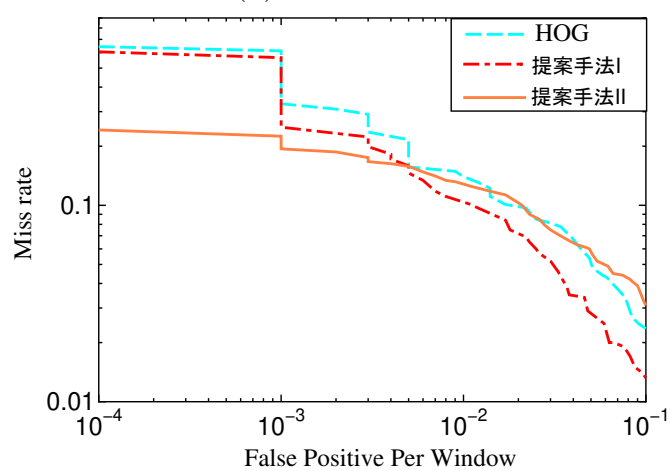
図 5.2 INRIA データセットでの実験結果 : DET カーブ



(a) Real Adaboost

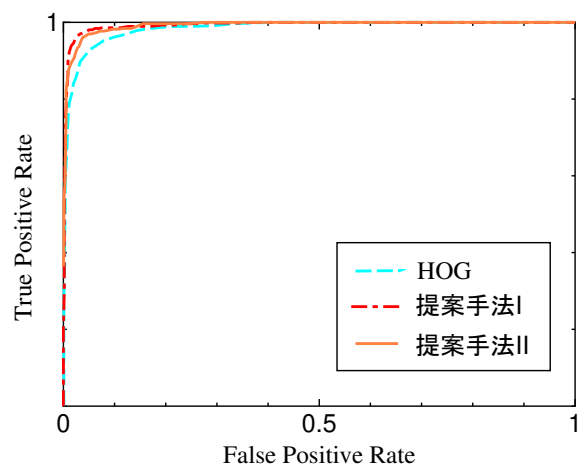


(b) SVM

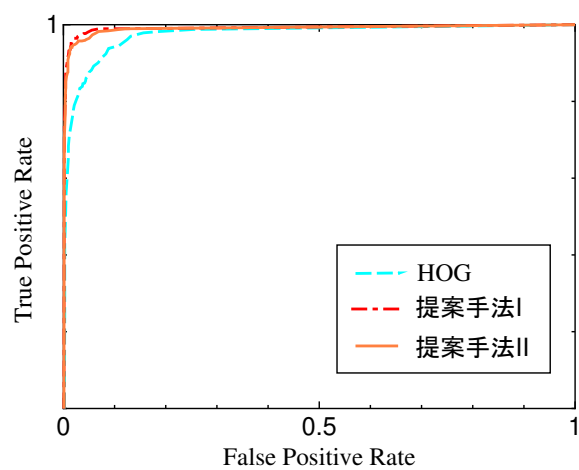


(c) ランダムフォレスト

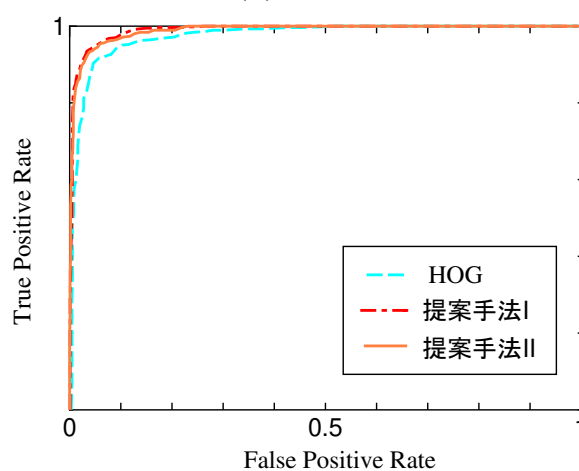
図 5.3 NICTA データセットでの実験結果 : DET カーブ



(a) Real Adaboost

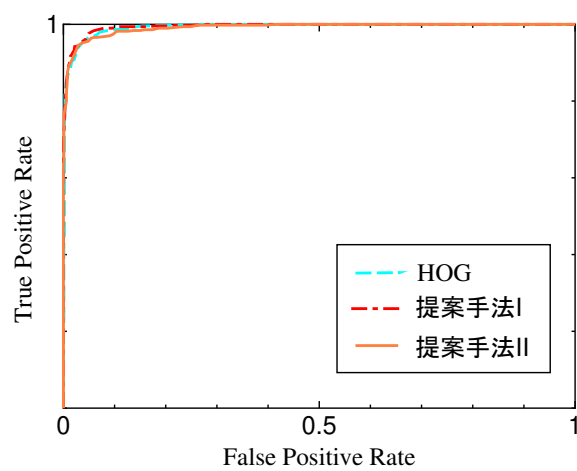


(b) SVM

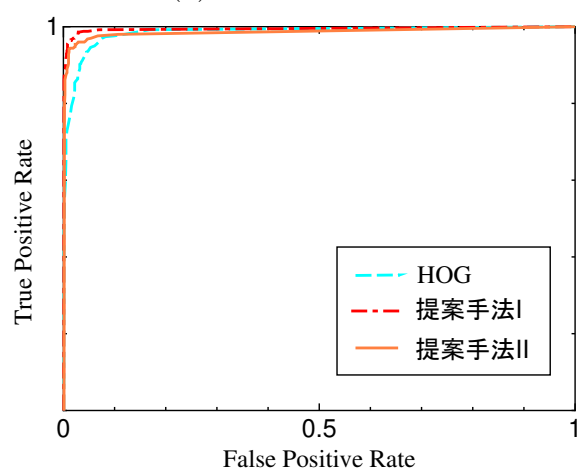


(c) ランダムフォレスト

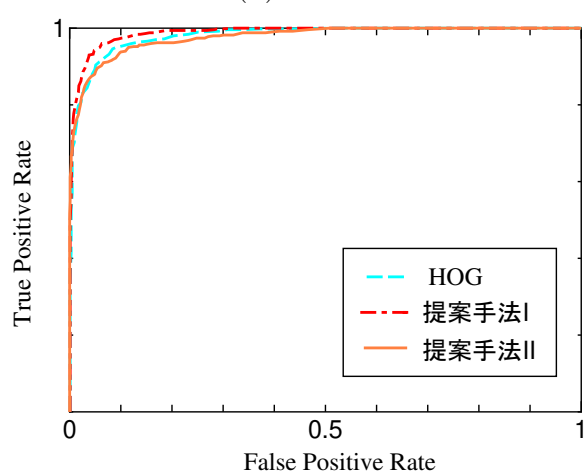
図 5.4 INRIA データセットでの実験結果 : ROC カーブ



(a) Real Adaboost



(b) SVM



(c) ランダムフォレスト

図 5.5 NICTA データセットでの実験結果：ROC カーブ

5.4 フレーム画像での検出実験

5.4.1 実験概要

本実験では、ピクセル間色コントラストに基づく HOG を用いて、フレーム画像に対して検出実験を行い、HOG との精度を比較する。3 節の実験より、最も精度が高かった SVM による識別器を本実験で用いる。画像データは INRIA Person Dataset [25] のものを用いる。INRIA データセットには、検出実験用のフレーム画像とそれに対する人物の正解位置を記録したアノテーションデータが存在する。3 節で用いた NICTA データセットではこれらのデータが用意されていないため、本実験は INRIA データセットのみを用いることになる。評価には、これまでの実験と同様 DET カーブを使用し、False Positive Per Image (FPPI) により比較を行う。また、2 章 3 節で述べたように検出結果の統合処理には、Non Maximum Suppression [57] を用いる。フレーム画像を対象とした実験では、人物の正解位置以外の領域がネガティブデータでありネガティブ数が膨大となるため、誤検出率の算出が困難である。そのため、本実験では ROC カーブや AUC を用いての評価は実施せず、DET カーブでの評価のみとする。

5.4.2 実験結果

図 5.6 に、実験結果の DET カーブを示す。図中の凡例において、ピクセル間色コントラストに基づく HOG は提案手法 II と示している。結果比較の参照点として 0.25FPPI の値を用いる。実験結果より、ピクセル間色コントラストに基づく HOG は、HOG と比較して 0.9FPPI 以下での Miss rate 削減が確認できた。0.25FPPI において、HOG の Miss rate と比較してみると 8.2 ポイントの削減となった。0.9FPPI 以上では、HOG の精度が上回るが、物体検出では未検出を抑えつつ、誤検出を減少させることが重要である。そのため、低 FPPI における Miss rate 改善は人物検出にとって重要な意味を持つと言える。また、本実験における検出結果の例を 5.7 に示す。図中央の左列の画像では、3 人並ぶ子供のうち、中央の子供が未検出となっている。これは、Non Maximum Suppression の統合閾値を調整することで対応可能である。図下段の画像はそれぞれ背景と自転車の前輪部分が誤検出されている。今後は、このように誤検出となった領域を調査し、その原因を解明していく必要がある。

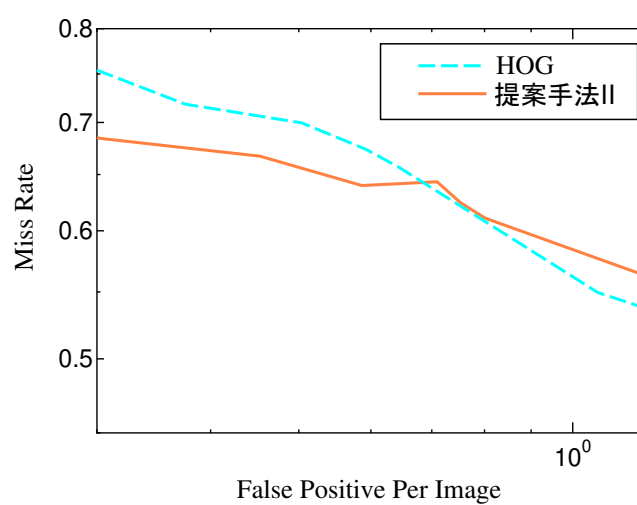


図 5.6 INRIA データセットでの FPPI 評価

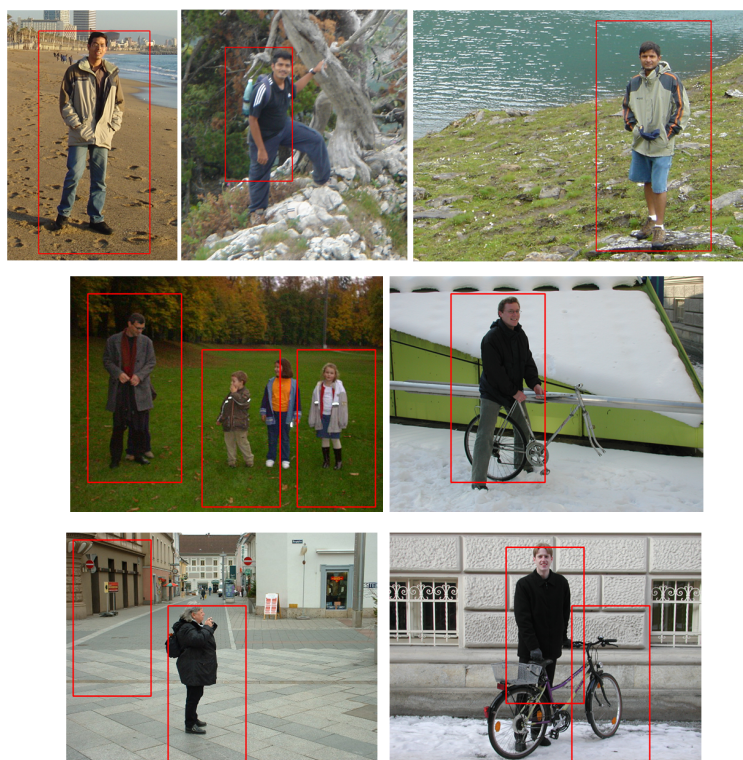


図 5.7 本実験における検出結果例

5.5 考察

前節の実験により、提案手法である、ピクセル間色コントラストに基づく HOG の有効性が確認された。本手法も、局所色相関分布に基づく HOG と同様に、色エッジを検出することによって検出精度向上を図っている。

図 5.8 に HOG 及び提案手法における勾配計算による勾配画像の例を示す。図中に赤枠で示される領域に注目すると、輝度差では検出できなかった人物の肩部分のエッジが検出されていることが確認できる。このように、提案手法では前章で提案した局所色相関分布に基づく HOG と同様に、色コントラストに基づく色エッジの検出が可能であることがわかる。

また、図 5.9 に HOG では識別できなかったが、提案手法により識別可能となった人物画像と、それらの HOG 及び提案手法における勾配画像の例を示す。(a) が入力画像であり、(b) が HOG における勾配画像、(c) が提案手法における勾配画像である。例として挙げているのは、局所色相関分布に基づく HOG と同様のものである。

1 列目の人物は腹部、2 列目の人物は脚部、下段 1 列目の人物は胸部、2 列目の人物は背部の輝度差が消失しており、エッジが検出されていないことがわかる。前章でも述べたように HOG においては、これらの部分のエッジ検出ができず人物の形状を捉えられなかったことが誤識別の原因として挙げられる。局所色相関分布に基づく HOG と同様に、本手法においても輝度コントラストが低下した状況で、色コントラストによるエッジの検出を行うことができる。これによって人物検出の精度向上が可能となったと言える。

図 5.10 に HOG で識別可能であったが、提案手法では識別できなかった人物画像と、それらの HOG 及び提案手法における勾配画像の例を示す。(a) が入力画像であり、(b) が HOG における勾配画像、(c) が提案手法における勾配画像である。右列の画像では、人物が着用しているコートとインナーの境界のエッジが検出されている。中央列と右列の画像では、背景に映る構造物のエッジが多数検出されていることがわかる。このような人物以外のエッジによって、これらの画像は誤識別したと考えられる。

色エッジは輝度によって検出できないエッジも検出可能であるため、これらのように、人物以外のエッジも多数検出してしまう。ここで挙げた例のように色エッジが人物検出精度の低下を招く可能性も考えられる。しかしながら、INRIA データセット及

び NICTA データセットといった複数のデータセットを用いた実験において、提案手法は既存手法の HOG の精度を超え、誤識別以上に識別可能となった画像のほうが多いことが確認できるため、本提案手法も局所色相関分布に基づく HOG と同様に人物検出の精度向上に有効であると言える。

上記した人物以外のエッジが検出されるという問題に対しては、輝度エッジ特徴量との併用により解決可能だと考えられる。色エッジを用いた提案手法と輝度エッジ特徴量が相互補完的な役割を果たすことによって、人物以外のエッジが学習に影響を与えることを減少させ、より人物形状を捉えやすくなると考えられる。

また、局所色相関分布に基づく HOG と本提案手法との比較結果より、局所色相関分布に基づく HOG は、低 FPPW において学習法による性能の差が発生することがわかった。4 章 4 節で述べたように、局所色相関分布では注目セルによっては、人物形状が抽出できずに人物のエッジが検出できない場合がある。これについては逆に、注目セルによっては背景に映り込む物体の形状を抽出せず、人物形状のみを抽出することが可能であるとも言える。これらの領域で検出されたエッジをうまく学習することができれば、人物検出精度が向上すると考えられるが、学習法によっては、この特徴を学習できなかったことが可能性として挙げられる。局所色相関分布に基づく HOG と本提案手法はともに、色エッジを検出し特徴量とするものだが、上述したことによって学習法による性能の差が発生したと考えられる。

実験結果より、各データセットにおいて局所色相関分布に基づく HOG を特徴量とした SVM による識別器が、最も性能が良いことがわかった。それ以外の Real Adaboost、ランダムフォレストにおいては、局所色相関分布に基づく HOG よりも本提案手法の精度が上回ることが多かった。実用化を考えると、特徴計算がより高速に行えることは重要である。これらの結果と特徴計算時間を考慮すると、本提案手法であるピクセル間色コントラストに基づく HOG の方が、局所色相関分布に基づく HOG よりも有利であると考えられることができる。

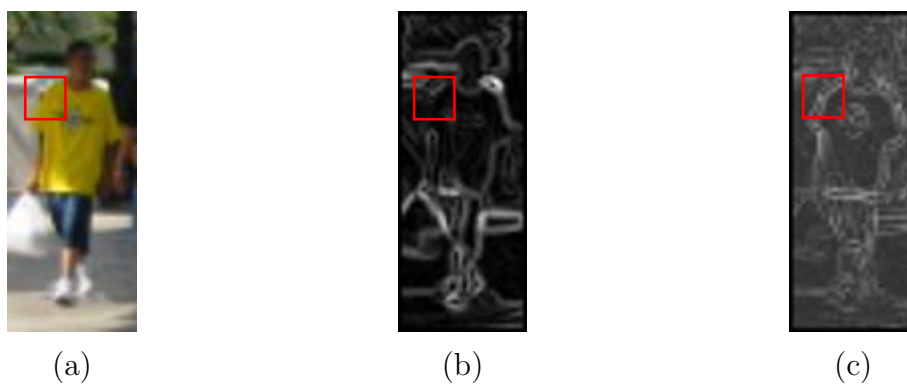


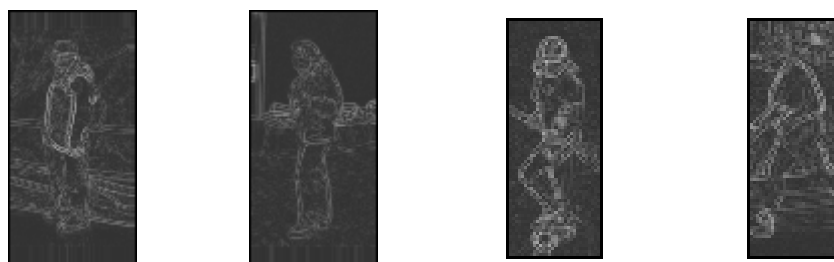
図 5.8 HOG 及び提案手法における勾配計算による勾配画像の例



(a) 入力画像



(b) HOG における勾配計算による勾配画像



(c) 提案手法における勾配計算による勾配画像

図 5.9 HOG で識別できなかったが、提案手法で識別可能になった画像例

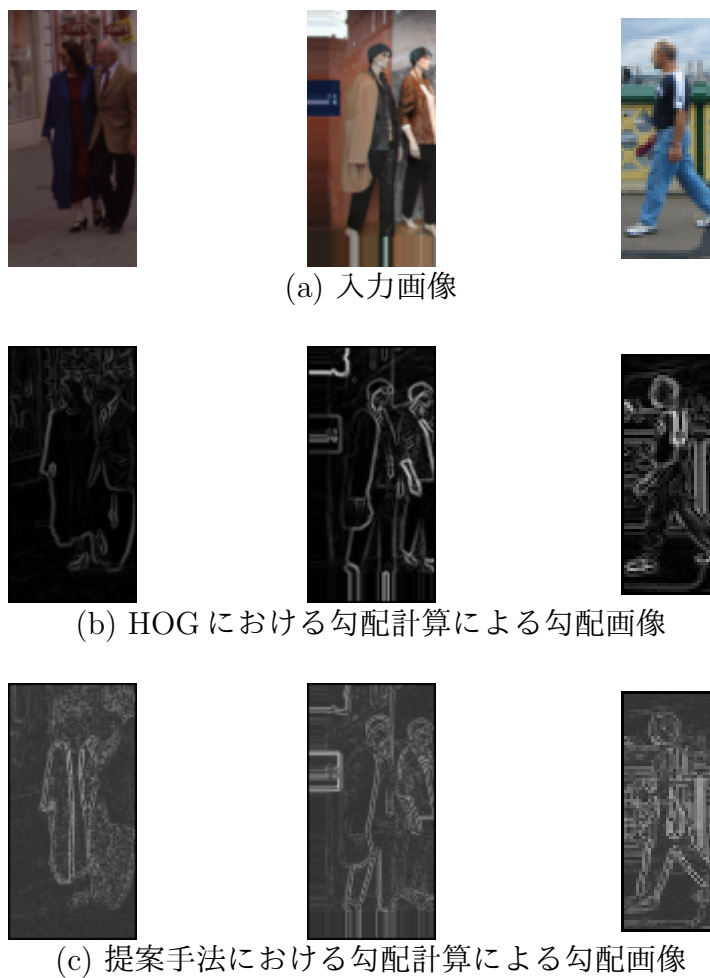


図 5.10 HOG で識別可能だったが、提案手法で識別できなかった画像例

5.6 おわりに

本章では、色コントラストに基づく勾配特徴量である、ピクセル間色コントラストに基づく Histograms of Oriented Gradients (HOG) を提案した。2節で、提案手法の特徴計算法を述べた。また、3節及び4節で実験結果について、5節で考察について述べた。

提案手法は、注目ピクセルと勾配計算の対象となる上下左右の隣接ピクセル間の色コントラストを求め、これを用いて勾配計算を行うため、4章で提案した局所色相関分布に基づく HOG のようにセル内で色相関分布を作成することなく、色エッジ検出を行うことが可能である。

人物画像データセットを用いて、既存の輝度情報に基づいた Histograms of Oriented Gradients (HOG) 特徴量 [25] 及び局所色相関分布に基づく HOG との性能比較実験を行った。DET カーブによる比較、AUC による比較を行い、提案手法が既存手法である HOG の精度を上回ることを示した。さらに、本提案手法が局所色相関分布に基づく HOG よりも、高速に特徴計算が行えることを示した。

また、提案手法により識別が可能となった画像例や識別ができなかった画像例を示し、これらについて考察を行った。これにより、本提案手法も局所色相関分布に基づく HOG と同様に、低輝度コントラスト状況下での人物検出を可能にしたことを示した。提案手法によって、HOG で識別可能だった画像が誤識別となった例もあるが、誤識別以上に識別可能となった画像のほうが多く、人物検出精度も向上したため、本手法も人物検出に有効な特徴量だと考えることができる。

第6章 全体を通しての考察と今後の課題

6.1 全体を通しての考察

4章と5章にて、局所色相関分布に基づく Histograms of Oriented Gradients (HOG) とピクセル間色コントラストに基づく Histograms of Oriented Gradients (HOG) を提案した。またこれらの性能評価実験において、Real Adaboost [19]、SVM [20]、ランダムフォレスト [21] といった複数の学習法で識別器を構築し、INRIA Person Dataset [25] 及び NICTA Pedestrian Dataset [65] を用いて評価を行った。実験結果より、各識別器、各データセットにおいて、提案手法は、Histograms of Oriented Gradients (HOG) の性能を上回るか同程度の性能を見せた。このことから提案手法は学習法やデータセットに依存することなく、従来手法である HOG よりも精度を向上させることができたと言える。4章と5章の考察で述べた通り、提案手法が HOG で識別可能だった画像を誤識別することがあるが、INRIA データセット、NICTA データセットにおいても性能向上が見られるため、精度がデータセットに依存することはないと考えられる。また、局所色相関分布に基づく HOG とピクセル間色コントラストに基づく HOG の間において、学習法による精度の差が発生しているが、従来手法である HOG からの精度向上が学習法に依存するものではないことが実験結果より確認できる。

後藤らが提案した CS-HOG 特徴量 [46] では、序論で指摘したように特徴次元数が膨大となることが問題として挙げられる。 64×128 ピクセルの画像に対して、CS-HOG 特徴量を算出すると、967,680 次元となる。局所色相関分布に基づく HOG とピクセル間色コントラストに基づく HOG は、ともに CS-HOG よりも少ない特徴次元とすることが可能である。上記と同様に 64×128 ピクセルの画像に対して、 10×10 ピクセルのセルサイズで特徴量を算出した場合、特徴次元数は局所色相関分布に基づく HOG とピクセル間色コントラストに基づく HOG とともに、62,073 次元となり、CS-HOG と比較して 905,607 次元の特徴を削減したことになる。特徴次元数の削減は実用化に向けても非常に有利に働くと考えられる。また、HOG の次元数は上記サイズの画像に

対して、 10×10 ピクセルのセルサイズで特徴量を算出した場合、次元となり、提案手法の方が次元数が多い。4章2節及び5章2節の特徴計算法で述べたように、提案手法ではセルを1ピクセルずつ、オーバーラップさせながら勾配方向ヒストグラムを作成している。そのため、HOGよりも次元数が多くなる。しかしながら、提案手法は汎用的な性能のコンピュータで学習可能であるため、実用化の際に特徴次元数が障壁になることはないと考えられる。

上記でCS-HOG特徴量と提案手法の比較を行ったが、他の類似手法との比較も述べる。序論で挙げたように、色の情報を利用したcolor HOG特徴量 [44,45] が、色エッジを利用したcolor-CoHOG特徴量 [47] が提案されている。

color HOG特徴量は画像のRGB各チャネルでHOGを算出しているのみで、これらを統合する処理は含まれていないため、色エッジを検出しているものではないと考えられる。また、RGB色空間を使用しているため、照明変動に頑健ではない。提案手法は色エッジの検出により、これを特徴量として人物の検出を行っている。加えてHSV色空間を使用しているため、照明変動に対しても頑健であると考えられる。この点でcolor HOGと提案手法の違いがある。

color-CoHOGはCoHOG特徴量の拡張であり、提案手法と同様に色エッジに注目した特徴量である。本手法は、色エッジ検出にRGB色空間を用いており、RGBの各チャネルで勾配計算を行い、それらを平均して勾配方向を決定している。また、その後、局所領域内において、YCbCr色空間に基づくピクセル間カラーマッチングを行い、同じ色のピクセル同士、異なる色のピクセル同士それぞれで、勾配方向ペアを求め、共起特徴量を算出している。色エッジ検出のみに注目すると、本手法もRGB色空間で色エッジの検出を行っており、照明変動に対して頑健でないと考えられる。また、RGB各チャネルで得られた結果を平均化して統合しているため、エッジの方向が正確に検出できない可能性も考えられる。

RGBから直接特徴量を算出することができるため、計算コスト的にはcolor-CoHOGの方が有利かもしれないが、提案手法は上記のようにHSV色空間を用いており、照明変動に対する頑健性を持つと考えられる。この点において、提案手法とcolor-CoHOGの差異が存在する。

色エッジを人物検出へと適用した先行研究例は、上記の CS-HOG 特徴量と color-CoHOG 特徴量の 2 例が挙げられる。色情報を人物検出へと適用した例が少ないのは、序論で述べたように、人物が着用している衣服の色が個人で異なっているため、色を特徴量として利用するのが困難だからである。一方で、色エッジを人物検出へと適用した例が少ないのは、これまで提案されてきたシングルチャネルベース、ベクトルベースの色エッジ検出手法 [40–43] の計算コストの高さ、計算の複雑性に起因していると考えられる。3 章 4 節で述べたように、提案手法は輝度差や色差のようなベクトル情報をスカラー情報に単純化し、色エッジ検出の計算コスト低下を図ると同時に色エッジ検出のための計算も単純化している。このため、提案手法は容易に人物検出に適用することが可能となっている。

序論で述べたように、Convolutional Neural Network (CNN) に代表されるディープラーニングによって、画像認識の精度が飛躍的に向上している。本研究で提案しているようなハンドクラフト特徴量と比較しても、さらに高精度な検出や認識が行える。しかしながら、CNN などのディープラーニングは、識別に至る判断根拠を理解することが困難である。ディープラーニングは、学習においてデータの特徴を学習しており、従来の画像認識のように特徴量の設計を行う必要がない。どのような特徴が学習されるのかはネットワークに依存し、学習された特徴はネットワークの重みとして表現されるため、ブラックボックスのような状態となる。このようにディープラーニングでは、学習された特徴を解析するのが困難なため、誤認識をした場合に何が原因となったのかが分からないという問題がある。

近年では、Adversarial Attack という、機械学習に対する攻撃手法が報告されており、これによって機械学習による識別に影響を与えることが可能である。また、ディープラーニングによる交通標識認識を誤認識させる手法も報告されている [66]。このような手法によって、ディープラーニングを用いた自動車の運転補助システムや自動運転システムが攻撃された場合、重大な事故が発生する可能性が考えられる。上述したように、ディープラーニングでは、どのような特徴が学習されているかわからないため、識別に寄与した特徴を解析することも困難である。このため、例に挙げた運転補助システム、自動運転システムにおいては、誤認識が要因となって事故が発生した際に、ユーザに対して事故原因、システムの判断根拠を説明することが不可能であると

いう問題も発生する。guided backpropagation [67] や Grad-CAM [68] など、入力画像中のどの部分が識別に重要であったかを可視化する手法も提案されているが、どのような特徴が識別に寄与したのかを解析することは未だに困難である。

一方で、本研究のようなハンドクラフト特徴量を用いた手法では、識別に寄与する重要な特徴の解析が容易に行える。4章2節及び5章2節で述べたように、提案手法では特徴量の次元がエッジの方向、空間的位置を表すため、どのような特徴が学習されたか、識別に寄与したか容易に解析できる。

人物検出技術を実用化するにあたり、誤認識の要因を解析できることは、その後の精度向上にとって非常に重要であると考えられる。また、運転補助システム、自動運転システム以外でも、機械学習の判断を基に人間が業務を行う場面において、判断の根拠を理解できることは重要な意味を持つ。本研究で提案した、局所色相関分布に基づく HOG とピクセル間色コントラストに基づく HOG は従来型のハンドクラフト特徴量であるが、上記のように識別の要因を解析できることが、ディープラーニングに対して有利な点となる。

最後に本研究の発展性について述べる。提案手法はともに low level な局所特徴量である。これまで、局所特徴量同士の関連性を捉える mid level な特徴量である共起特徴量が提案されている。共起特徴量には、Adaboost や Real Adaboost を用いた特徴の共起手法 [53–55] が提案されており、局所特徴量の共起により、特徴の関連性、例えばエッジの連続性、対称性の表現を可能とすることで、精度向上を図っている。提案手法においても、これらの共起手法を導入することで容易に共起特徴量へと拡張可能で、更なる精度向上が見込める。加えて、color-CoHOG で用いられている共起手法を導入することも容易であり、提案手法を color-CoHOG のような共起特徴量へと拡張することも可能である。また、複数の特徴量の併用により、人物検出精度が向上することが報告されている [69, 70]。提案手法においても、異なる特徴量との併用を考えることができる。5章5節で述べたように輝度エッジ特徴量と併用すれば、色エッジと輝度エッジが相互補完的な役割を果たすことが期待できる。

CNN では、畳み込み層によってエッジなどの特徴を検出し、プーリング層によって特徴の位置ずれに対するロバスト性を高めている。提案手法では、このプーリング層に該当するような特徴の位置ずれに対する対処は特になされていない。そのため、

CNNにおけるプーリング層のような処理を提案手法に導入し、特徴の位置ずれに対するロバスト性を高めることで、更なる精度向上を見込むことができる。

また、人間の視覚システムは、色と輝度が同時に変化する境界は物体表面色の変化として知覚し、色の変化がなく輝度のみが変化する境界を影による物体表面色の変化として知覚していることがわかっている [71]。このことに基づき、提案手法と輝度エッジ特徴量の相関情報を特徴量とすれば、陰影の影響を除去可能なエッジ特徴量を設計することができる。このように、心理学的知見を導入することでさらなる発展を見込むことができる。

6.2 今後の課題

本研究における今後の課題として、次の2点が挙げられる。

- Real Adaboost 及びランダムフォレストの学習パラメータの最適化
- 提案手法における精度向上要因の更なる解析

Real Adaboost 及びランダムフォレストの学習パラメータは4章3節1項の実験概要で述べたとおりである。4章と5章の各章における実験では、ランダムフォレストのみ他の学習法よりも精度が低い結果となっているが、これは学習パラメータの影響であると考えられる。本研究におけるランダムフォレストの学習パラメータは、Scikit-learn [64] で実装されているグリッドサーチを用いて決定している。グリッドサーチとは、学習パラメータを一定の範囲内で変化させながら、学習データに対する識別精度が最も良かったパラメータを採用し、識別器を構築するパラメータ調整法である。グリッドサーチにおいて、木の数は100から500、深さは1から5までの範囲を探索して、パラメータを決定している。また、Real Adaboost においては、弱識別器数が1000までの範囲でテストデータに対する識別精度を検証し、精度がほぼ頭打ちとなったパラメータを採用している。

ランダムフォレストにおいては、木の深さや数を増加し、選択される特徴数を増加させることで、さらなる精度向上を図ることができる。今後は木の数、深さの上限値を上げて探索を行うことで、Real Adaboost や SVM の性能に迫ることができると考えられる。Real Adaboost においても、弱識別器数をさらに増加させ、より詳細なパ

ラメータ探索を行っていくことで、さらに精度が向上する可能性がある。このように学習パラメータによる精度向上が見込めるため、今後、学習パラメータについて検証し、最適なパラメータを探索することが課題となる。その上で、どの学習法が最も精度が良くなるのかを検証することで、提案手法の実用化の可能性を高められると考えられる。

また、提案手法の精度向上要因については、色エッジを特徴量として用いているため、輝度では検出できないエッジが検出可能となり、それが性能向上に繋がったと4章5節と5章5節で述べている。さらに色エッジによってエッジが検出可能となった実際の画像例も示している。このように提案手法は、色エッジを人物検出に用いているという点が、従来のHOGのような輝度エッジ特徴量による人物検出に関する先行研究に対する有利な点となっている。しかしながら、提案手法においては、Real Adaboostやランダムフォレストで、どのような特徴が選択されているか、SVMでどの特徴の重みが高くなっているのかの解析は行っておらず、選択された特徴や重みが高くなっている特徴が、従来のHOG特徴量と比較してどのような違いがあるのかについてはわかっていない。そのため、今後この点について解析する必要がある。4章5節と5章5節において、提案手法によって誤識別した例、HOGで人物だと識別されたが、提案手法では背景だと識別された例を示した。提案手法における誤識別例では、HOGで背景と識別されたが、提案手法では人物だと識別された例はなかったが、このような誤識別例も、上記の解析を行うことによって、より有効な解決策を見出せる可能性があると考えられる。

第7章 結論

本論文では、色エッジの人物検出への適用と人物検出の精度向上を目的として、色コントラストに基づく勾配特徴量を提案した。色コントラストを捉えるために色類似度に着目し、これを勾配計算へ用いることを考え、特徴量の設計を行った。色コントラストに基づく勾配特徴量として、局所色相関分布に基づく Histograms of Oriented Gradients (HOG)、ピクセル間色コントラストに基づく Histograms of Oriented Gradients (HOG) の2つの特徴量を提案した。

第2章では、提案手法の基となった Histograms of Oriented Gradients (HOG) 特徴量と Color Self Similarity (CSS) 特徴量、性能評価手法である Detection Error Trade-off (DET) カーブと False Positive Per Window (FPPW)、False Positive Per Image (FPPI) という2つの評価指標について述べた。また、FPPIでの評価の際に必要な検出ウィンドウ統合処理についても述べた。

第3章では、提案手法で採用する色コントラストの定量化手法について述べた。まず、画像処理で一般的に扱われる色空間と提案手法で採用した HSV 色空間についてその採用理由について述べ、色類似度算出手法について述べた。色コントラストは、色差により色類似度として算出すること及び HSV 空間上での色差がユークリッド距離で算出できることを示した。

第4章では、局所色相関分布に基づく Histograms of Oriented Gradients (HOG) を提案した。提案手法は、画像の局所領域内において、注目ピクセルと近傍ピクセル間の色コントラストに基づく局所色相関分布を作成し、これに対して勾配計算を行うものである。提案手法における局所色相関分布の作成は、注目ピクセルの色を基準として、局所領域内で前景背景を分離する領域分割として考えることができる。これによって物体形状を抽出すること及び色コントラストに基づいた色エッジの検出が可能となる。人物画像データセットを用いた性能比較実験により、提案手法が輝度情報に基づいた既存手法の性能を上回ることを確認した。考察にて、本手法が色コントラストに基づく色エッジ検出が可能なことを示した。これらより、本手法が低輝度コント

ラスト状況下でのエッジ検出を可能にし、色エッジが人物検出の精度向上に有効であることを示した。

第5章では、ピクセル間色コントラストに基づく Histograms of Oriented Gradients (HOG) を提案した。提案手法は、注目ピクセルと勾配計算の対象となる隣接ピクセル間の色コントラストを算出し、これらを勾配計算に利用するものである。これによって、一般的なエッジ検出法と同様の演算で色エッジの検出を行うこと、色エッジ検出のための計算量を削減することが可能となる。人物画像データセットを用いた性能比較実験により、提案手法が輝度情報に基づいた既存手法の性能を上回ることを確認した。考察にて、本提案手法が局所色相関分布に基づく HOG と同様に、色コントラストに基づく色エッジ検出が可能なこと、色エッジにより低輝度コントラスト状況下でのエッジ検出を可能にしたことを示した。

第6章では、全体を通しての考察について述べた。提案手法が学習法や画像データセットに依存せずに人物検出精度を向上させること、既存の色エッジ特徴量との比較、Convolutional Neural Network との比較について述べ、今後の展望として、提案手法の改良案を示した。また、本研究における今後の課題を述べた。

以上のように、本論文では色コントラストに基づいた勾配特徴量を提案した。人物画像データセットを用いた性能比較実験により、提案手法が色エッジにより人物検出精度を向上させることを確認し、検出精度向上が学習法、データセットに依存するものではないことを示した。これらによって、色エッジ及び色エッジを用いた提案手法が精度向上に有効であることを示した。

謝辞

本論文は、筆者が九州工業大学大学院生命体工学研究科 脳情報専攻博士後期課程在籍時の研究成果をまとめたものである。本研究を遂行するにあたり、大学院工学研究院 基礎科学研究系 准教授 花沢明俊先生には終始懇切なる御指導と御鞭撻を賜った。また、花沢先生の異動に伴い、指導教員を引き継いでくださった、大学院生命体工学研究科 人間知能システム工学専攻 教授 夏目季代久先生にも懇切なる御指導、御鞭撻を賜った。ここに深謝の意を表する。

本論文執筆にあたり、有益な御討論、御助言を頂いた大学院工学研究院 基礎科学研究系 教授 浅海賢一先生、大学院生命体工学研究科 人間知能システム工学専攻 准教授 堀尾恵一先生、田向権先生に感謝の意を表する。

また、前勤務先を退職する際にその決断を尊重してくれ、様々な援助や励ましをくれた修士時代の同期の学友である、永田尚人君、片山智弘君、松田充史君、筒井瑛士君、田丸哲也君と、前勤務先同期の諸氏に感謝の意を表する。

大島商船高等専門学校 情報工学科 5 年生の領家直哉君には第 5 章のフレーム画像での検出実験の一部を補助してもらった。ここに感謝の意を表する。

最後に、今日に至るまで暖かく支えてくださった両親に深謝の意を表し謝辞とする。

参考文献

- [1] 鷺見和彦, 関真規人, 波部齊, ”物体検出—背景と検出対象のモデリング—.” 情報処理学会研究報告コンピュータビジョンとイメージメディア (CVIM), 2005(88 (2005-CVIM-150)), pp.79-98, 2005.
- [2] 原田典明, 石寺永記, 大網亮磨, 中尾敏康, ”人物行動を把握する画像解析技術と適応例.” NEC 技報, 63(3), pp.39-43, 2010.
- [3] Papageorgiou, C., and Poggio, T., Trainable pedestrian detection. *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, Vol. 4, pp.35-39, 1999.
- [4] Gavrila, D. M., and Munder, S., Multi-cue pedestrian detection and tracking from a moving vehicle. *International journal of computer vision*, 73(1), pp.41-59, 2007.
- [5] Cho, H., Seo, Y. W., Kumar, B. V., and Rajkumar, R. R., A multi-sensor fusion system for moving object detection and tracking in urban driving environments. *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pp.1836-1843, 2014.
- [6] 大澤達哉, 若林佳織, ”映像モニタリングのための人物追跡技術.” NTT 技術ジャーナル, 2007.
- [7] マツダ株式会社, i-ACTIVSENSE,
<http://www.mazda.co.jp/cars/axela/feature/safety/i-activsense/>

- [8] トヨタ自動車株式会社, Toyota Safety Sense,
<https://www.toyota.co.jp/jpn/tech/safety/toyota-safety-sense/>
- [9] 日産自動車株式会社, インテリジェント エマージェンシーブレーキ,
https://www.nissan-global.com/JP/TECHNOLOGY/OVERVIEW/emergency_brake.html
- [10] 株式会社 SUBARU, アイサイト, <https://www.subaru.jp/safety/eyesight/>
- [11] 本田技研工業株式会社, 衝突軽減ブレーキ 〈CMBS〉 ,
<https://www.honda.co.jp/hondasensing/feature/cmbs/>
- [12] ダイハツ工業株式会社, スマートアシスト III,
https://dport.daihatsu.co.jp/information/sumaashi/?_ga=2.133362252.357306717.1547455116-1078961026.1547455116
- [13] Cucchiara, R., Grana, C., Piccardi, M., and Prati, A., Detecting moving objects, ghosts, and shadows in video streams. *IEEE transactions on pattern analysis and machine intelligence*, 2003.
- [14] Mittal, A., Monnet, A., and Paragios, N., Scene modeling and change detection in dynamic scenes: A subspace approach. *Computer vision and image understanding*, 113(1), pp.63-79, 2009.
- [15] Xin, B., Tian, Y., Wang, Y., and Gao, W., Background subtraction via generalized fused lasso foreground modeling. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.4676-4684, 2015.
- [16] 岡村健史郎, 仲貴大, 宗藤萌子, 河合宏紀, 松村遼., ”湾岸領域監視を目的とした固有空間法の精度向上.” 独立行政法人国立高等専門学校機構大島商船高等専門学校紀要 国立高等専門学校機構大島商船高等専門学校編, (47), pp.61-72, 2015.
- [17] 松山隆司, 和田俊和, 波部齐, 棚橋和也., ”照明変化に頑健な背景差分.” 電子情報通信学会論文誌 D, 84(10), pp.2201-2211, 2001.

- [18] Freund, Yoav, and Robert E. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting." *Journal of computer and system sciences* 55.1 (1997): 119-139.
- [19] Schapire, R. E., and Singer, Y., Improved boosting algorithms using confidence-rated predictions. *Machine learning*, 37(3), pp.297-336, 1999.
- [20] Boser, B. E., Guyon, I. M., and Vapnik, V. N., A training algorithm for optimal margin classifiers. *In Proceedings of the fifth annual workshop on Computational learning theory*, ACM, 1992.
- [21] Breiman, L., Random forests. *Machine learning*, 45(1), pp.5-32, 2001.
- [22] Viola, P., and Jones, M. J., Robust real-time face detection. *International Journal of Computer Vision*, 57(2), pp.137-154, 2004.
- [23] Levi, K., and Weiss, Y., Learning object detection from a small number of examples: The importance of good features. *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, Vol. 2. IEEE, 2004.
- [24] Wu, B. and Nevatia, R., Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors. *IEEE International Conference of Computer Vision*, Vol. 1, 90-97, 2005.
- [25] Dalal, N., and Triggs, B., Histograms of oriented gradients for human detection. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 1. IEEE, pp.886-893, 2005.
- [26] Hou, C., Ai, H., and Lao, S., Multiview Pedestrian Detection Based on Vector Boosting. *Asian Conference on Computer Vision*, Springer, Berlin, Heidelberg, 2007.

- [27] Watanabe, T., Ito, S., and Yokoi, K. , Co-occurrence histograms of oriented gradients for pedestrian detection. *Pacific-Rim Symposium on Image and Video Technology*, Springer, Berlin, Heidelberg, 2009.
- [28] Qi, S., Ma, J., Lin, J., Li, Y., and Tian, J., Unsupervised ship detection based on saliency and S-HOG descriptor from optical satellite images. *IEEE Geoscience and Remote Sensing Letters*, 12(7), pp.1451-1455, 2015.
- [29] Liu, Y., Zhang, J., Wang, C., and Wang, L., Multiple HOG templates for gait recognition. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, IEEE, 2012.
- [30] Torrione, P. A., Morton, K. D., Sakaguchi, R., and Collins, L. M., Histograms of oriented gradients for landmine detection in ground-penetrating radar data. *IEEE Transactions on Geoscience and Remote Sensing*, 52(3), pp.1539-1550, 2014.
- [31] Gevers, Theo, and Arnold WM Smeulders., "Color-based object recognition." *Pattern recognition* 32.3, pp.453-464, 1999.
- [32] Phung, Son Lam, Abdesselam Bouzerdoum, and Douglas Chai., "A novel skin color model in ycbcr color space and its application to human face detection." *Image Processing. 2002. Proceedings. 2002 International Conference on*. Vol. 1. IEEE, 2002.
- [33] Bahlmann, C., Zhu, Y., Ramesh, V., Pellkofer, M., and Koehler., T. "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information." *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*. IEEE, 2005.
- [34] Shinichiro Omachi and Masako Omachi, "Traffic light detection with color and edge information." *Computer Science and Information Technology, International Conference on(ICCST)*, pp. 284-287, 2009.

- [35] 山田寛, 松尾直志, 島田伸敬, 白井良明., ”手話認識のための見えの学習による手領域検出と形状識別.” 画像の認識・理解シンポジウム (MIRU2009), pp.635-642, 2009.
- [36] Walk, S., Majer, N., Schindler, K. and Schiele, B., New features and insights for pedestrian detection. *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*, IEEE, 2010.
- [37] Treisman, A. M., and Gelade, G., A feature-integration theory of attention. *Cognitive psychology*, 12(1), pp. 97-136, 1980.
- [38] Treisman, A., Features and objects in visual processing. *Scientific American*, 255(5), pp. 114B-125, 1986.
- [39] Moutoussis, K., The physiology and psychophysics of the color-form relationship: A review. *Frontiers in psychology*, 6, 1407, 2015.
- [40] A. Koschan, A comparative study on color edge detection, in *Proc. 2nd Asian Conf. Comput. Vis.*, vol. III, pp. 574-578, 1995.
- [41] Koschan, Andreas, and Mongi Abidi., ”Detection and classification of edges in color images.” *IEEE Signal Processing Magazine* 22.1, pp.64-73, 2005.
- [42] Dutta, Soumya, and Bidyut B., Chaudhuri., ”A color edge detection algorithm in RGB color space.” *Advances in Recent Technologies in Communication and Computing, 2009. ARTCom'09. International Conference on*. IEEE, 2009.
- [43] Akinlar, C., and Topal, C., ColorED: Color edge and segment detection by Edge Drawing (ED). *Journal of Visual Communication and Image Representation*, 44, pp.82-94, 2017.
- [44] Wang, G., Ren, G., Wu, Z., Zhao, Y., and Jiang, L., A robust, coarse-to-fine traffic sign detection method. *Neural Networks (IJCNN), The 2013 International Joint Conference on*. IEEE, pp.1-5, 2013.

- [45] Wang, D., Hou, X., Xu, J., Yue, S., and Liu, C. L., Traffic sign detection using a cascade method with fast feature extraction and saliency test. *IEEE Transactions on Intelligent Transportation Systems*, 18(12), pp.3290-3302, 2017.
- [46] 後藤雄飛, 山内悠嗣, 藤吉弘亘., "CS-HOG: 色の類似性に基づいた形状特徴量." 電子情報通信学会論文誌 D 96.7, pp.1618-1626, 2013.
- [47] Ito, S., and Kubota, S., "Object classification using heterogeneous co-occurrence features." *European Conference on Computer Vision*, Springer, Berlin, Heidelberg, 2010.
- [48] Du, X., El-Khamy, M., Lee, J., and Davis, L., Fused DNN: A deep neural network fusion approach to fast and robust pedestrian detection. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 953-961, 2017.
- [49] Li, J., Liang, X., Shen, S., Xu, T., Feng, J., and Yan, S., Scale-aware fast R-CNN for pedestrian detection. *IEEE transactions on Multimedia*, 20(4), pp.985-996, 2018.
- [50] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., and Berg, A. C., Ssd: Single shot multibox detector. In *European conference on computer vision*, pp. 21-37, Springer, Cham, 2016.
- [51] Uijlings, J. R., Van De Sande, K. E., Gevers, T., and Smeulders, A. W., *Selective search for object recognition. International journal of computer vision*, 104(2), pp.154-171, 2013.
- [52] Girshick, R., Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pp. 1440-1448, 2015.
- [53] Mita, Takeshi, Toshimitsu Kaneko, and Osamu Hori., "Joint haar-like features for face detection." *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. Vol. 2. IEEE, 2005.

- [54] 尾崎貴洋, 山内悠嗣, 藤吉弘亘., "Joint HOG 特徴を用いた 2 段階 AdaBoost による車両検出." 動的画像処理実利用化ワークショップ (DIA2008), I1-13, 2008.
- [55] 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 D 92.8, pp.1125-1134, 2009.
- [56] Martin, A., Doddington, G., Kamm, T., Ordowski, M., and Przybocki, M., "The DET Curve in Assessment of Detection Task Performance." *Fifth European Conference on Speech Communication and Technology*. 1997.
- [57] Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D., Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 32(9), pp.1627-1645, 2010.
- [58] Tkalcic, M., and Tasic, J. F., "Colour spaces: perceptual, historical and applicational background." *The IEEE Region 8 EUROCON 2003. Computer as a Tool., Ljubljana, Slovenia, 2003*, pp. 304-308 vol.1.
- [59] J. R. Smith., "Integrated spatial and feature image systems: Retrieval, compression and analysis." *Ph.D. dissertation, Columbia Univ., New York*, 1997.
- [60] Matsumura, R., and Hanazawa A., Human Detection by Gradient Orientation Histogram applied on Local Color Correlation Distributions, *2013 International Workshop on Advanced Image Technology Proceedings*, pp.210-215, 2013.
- [61] Matsumura, R., and Hanazawa A., Gradient Orientation Histogram applied on Local Color Correlation Distributions improves Human Detection. *2014 International Workshop on Advanced Image Technology Proceedings*, pp.94-97, 2014.
- [62] 松村遼, 花沢明俊., "局所的色相関分布に対する勾配方向ヒストグラムによる人物検出." 信学技報, vol. 112, no. 227, NC2012-38, pp.11-16, 2012.
- [63] Chang, C. C., and Lin, C. J., LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3), 27, 2011.

- [64] Pedregosa, F., et al., Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct), pp.2825-2830, 2011.
- [65] G. Overett, L. Petersson, N. Brewer, L. Andersson and N. Pettersson., A new pedestrian dataset for supervised learning. *Intelligent Vehicles Symposium, 2008 IEEE*, IEEE, 2008.
- [66] Evtimov, Ivan, et al. "Robust physical-world attacks on deep learning models." *arXiv preprint arXiv:1707.08945 1*, 2017.
- [67] Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M., Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*, 2014.
- [68] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D., "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE International Conference on Computer Vision*. 2017.
- [69] Viola, Paul, Michael J. Jones, and Daniel Snow., "Detecting pedestrians using patterns of motion and appearance." *International Journal of Computer Vision*, 63.2, pp.153-161, 2005.
- [70] Yamauchi, Y., Fujiyoshi, H., Hwang, B. W., and Kanade, T., "People detection based on co-occurrence of appearance and spatiotemporal features." *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008.
- [71] Kingdom F. A. A., Color brings relief to human vision. *Nature Neuroscience*, 6 (6), pp.641–644, 2003.