

平成 25 年度
中部大学大学院工学研究科情報工学専攻

博士学位論文

**統計的学習法に基づく物体検出の高精度化と
効率化に関する研究**

土屋 成光

論文要旨

物体検出とは、対象とする物体を画像上で探索し、その位置や大きさを求める技術である。物体検出は機械に人の視覚や脳の機能を持たせるための前提技術であり、セキュリティやマーケティングなど多様な分野で研究が精力的に行われている。物体検出は、一般環境下で撮影された画像を入力としてその中に存在する物体領域を出力するため、物体に対して、その姿勢や向き、色や個体差に抛らず検出可能であることが求められる。カメラで撮影される画像において物体の姿勢や向き、色や個体差の要因は、その物体の見えを大きく変える。この変化を吸収するために、画像から特徴量を抽出し、その特徴量を統計的手法で学習する必要がある。しかし、様々な見えを持つ検出対象物体を1つのカテゴリとして識別することは複雑な問題であり学習が困難となる場合がある。これは、多様な見えを持つ検出対象と背景には類似する部分形状が存在し、その識別には非常に複雑な識別境界が必要とされるためである。複雑な識別境界を高精度に構築することは難しい。また、学習環境で高精度な検出システムを構築しても、実際にカメラを設置する設置環境ではその性能を十分に発揮できないという問題がある。これは、学習環境と設置環境の間では観測される対象物体の見えが大きく変化するためである。この性能低下に対しては、大きなコストを払い再学習などで識別器を設置環境に適応させる必要がある。

そこで、本研究では統計的学習法に基づく物体検出の高精度化と、設置環境への適応の効率化について取り組む。高精度化については、単一の識別器で対象物体の全ての変化を捉えることが困難であるため、まず類似性に基づき検出対象の集合を分割することで問題を簡易化し、その小集合それぞれに対し識別器を学習し統合する分割統治戦略の導入を提案する。これにより、多様な変化を持つ対象の高精度な検出が可能となる。

設置環境への適応の効率化のためには、学習後の識別器を簡易に調整して適応する仕組みが必要である。そこで、本研究では特徴量が識別能力にどの程度貢献しているかを求め、それに基づき識別器を調整する手法を提案する。これにより、構築した識別器を設置環境に合わせるための調整が簡便に可能となる。また、調整のみでは適応が困難な学習環境と設置環境の相違が大きい場合については、再学習のため学習用画像を再収集する必要がある。しかし、位置ずれなく大量の学習サンプルを作成するには、非常に大きな人的コストを要するという問題がある。これに対し、本研究では、CGを用いて設置環境における多様な学習用画像を自動生成する手法を提案する。これにより、人手のかからない効率的な適応を可能とする。更なる効率化として、本研究では、少数の学習画像から構築時間の削減を達成するために、転移学習と再学習を組み合わせたハイブリッド型転移学習を提案する。転移学習の際に低下する精度を全探索で補うことにより、大量の画像を用いた再学習に比べ高速な学習が可能となる。また、Random Forest に転移学習を導入し、マルチクラス識別に対しても効率的な適応が可能となる。

目次

第 1 章 序論	1
1.1 研究の背景	2
1.2 研究目的	3
1.3 本論文の構成	4
第 2 章 統計的学習に基づく物体検出とその関連研究	7
2.1 統計的学習を用いた物体検出システム	8
2.2 統計的学習に基づく識別器	8
2.2.1 Boosting	8
2.2.2 Support Vector Machine	13
2.2.3 Boosting と SVM	16
2.3 マルチクラス識別器	17
2.3.1 One vs Other 戦略	17
2.3.2 AdaBoost.MH	18
2.3.3 Joint Boosting	20
2.4 統計的学習による物体検出を困難とする要因	22
2.4.1 物体の多様性に起因する問題	24
2.4.2 カメラの設置環境に起因する問題	24
2.5 まとめ	26
第 3 章 分割統治戦略による識別器の高精度化	29
3.1 複雑な物体分類の問題と分割統治戦略の導入	30
3.1.1 複雑な変化を持つ物体の分類	30
3.1.2 分割統治戦略の導入	31
3.2 弱識別器の応答値に基づくサンプルの分割	32
3.2.1 Spectral Clustering	33
3.2.2 ノンパラメトリックなクラスタリングへの対応	33
3.2.3 サンプル分割結果	34
3.3 Divide and Conquer Boosting によるサブカテゴリに対する効率的な学習	37
3.3.1 Joint Boosting アルゴリズムの分割統治戦略への適用	37

3.3.2	Joint Boosting との相違点	39
3.4	評価実験	39
3.4.1	実験概要	40
3.4.2	弱識別器の応答値によるサンプル分割の有効性	40
3.4.3	DaCBoost によるサブカテゴリ学習の有効性	41
3.5	まとめ	43
第 4 章	特徴貢献度に基づく識別器の調整	45
4.1	識別器の調整と特徴量選択	46
4.1.1	Sequential Forward Selection	46
4.1.2	Sequential Backward Selection	46
4.1.3	従来法の問題点	46
4.2	Boosting による特徴量の貢献度評価	47
4.2.1	2 クラス問題に対する貢献度算出	47
4.2.2	マルチクラス問題における貢献度算出	47
4.2.3	貢献度に基づく特徴選択 (SBS-CRbB)	48
4.3	特徴貢献度に基づく特徴選択実験	49
4.3.1	特徴選択結果	50
4.3.2	特徴選択に要する計算時間の比較	52
4.4	貢献度に基づく識別器の調整	53
4.4.1	貢献度に基づく識別器の調整法	53
4.4.2	識別器の調整実験	54
4.4.3	移動体識別のための特徴量	54
4.4.4	貢献度に基づく特徴量の有効性の解析結果	57
4.4.5	特徴貢献度に基づく識別器の調整結果	57
4.5	まとめ	61
第 5 章	生成型学習に基づく学習の効率化	63
5.1	学習サンプルの自動生成による効率化	64
5.1.1	3 次元人体モデルによる人画像生成	65
5.1.2	ランダムサンプリングによる背景サンプルの自動収集	66
5.2	Multiple Instance Learning の導入によるノイズ低減	67
5.2.1	MILBoost	67
5.2.2	Negative-bag MILBoost	68
5.3	評価実験	73
5.3.1	実験概要	73
5.3.2	生成画像に基づく特定シーン学習の有効性	76
5.3.3	Negative-bag MILBoost によるノイズ低減の有効性	78

5.4	まとめ	80
第 6 章	実利用シーンに対する再学習の効率化	81
6.1	物体検出における転移学習	82
6.1.1	転移学習	82
6.1.2	Covariate-shift Boost(CovBoost) による転移学習	82
6.1.3	転移学習の問題点	84
6.2	ハイブリッド型転移学習による学習の効率化	86
6.2.1	本研究の問題設定	86
6.2.2	HOG 特徴量における Feature Shift	87
6.2.3	ハイブリッド型転移学習による学習	88
6.2.4	学習効率 ζ に基づく特徴空間選択	89
6.3	評価実験	91
6.3.1	実験概要	91
6.3.2	俯角変化に対するハイブリッド型転移学習の精度	91
6.3.3	考察	91
6.3.4	ハイブリッド型転移学習の高速性	92
6.4	まとめ	93
第 7 章	転移学習の導入による Random Forest の効率化	95
7.1	マルチクラス分類における転移学習	96
7.2	Random Forest	96
7.2.1	Random Forest の学習アルゴリズム	96
7.2.2	Random Forest の識別処理	99
7.3	Random Forest の転移学習	100
7.3.1	共変量シフトに基づく転移学習を導入した Random Forest	100
7.4	評価実験	104
7.4.1	2 クラス分類問題の転移	104
7.4.2	2 クラス分類問題への転移学習の導入による効率化	105
7.4.3	マルチクラス分類問題の転移	107
7.4.4	マルチクラス分類問題への転移学習の導入による効率化	108
7.5	まとめ	109
第 8 章	結論と展望	111
8.1	結論	111
8.2	展望	113
謝 辞		115

参考文献	117
研究業績一覧	123

目次

1.1	本論文の構成.	5
2.1	統計的学習法を用いた物体検出システム (例: 人検出).	8
2.2	AdaBoost による人の識別の流れ.	9
2.3	SVM によるマージン最大化.	13
2.4	カーネルトリックを用いた高次元への写像.	16
2.5	マルチクラス識別器.	18
2.6	弱識別器の共有.	22
2.7	物体の見えの変化 (INRIA person dataset).	24
2.8	視点の変化により発生する未検出.	25
2.9	標準的な学習サンプルの例 (INRIA Person Dataset).	26
3.1	分割統治的戦略による識別器の学習法 (概念図).	31
3.2	弱識別器応答値を用いたできる学習サンプルのカテゴリライズ.	32
3.3	Spectral Clustering の流れ.	33
3.4	medoid-shift と mean-shift の違い.	34
3.5	人手によるサンプル分割.	35
3.6	提案手法によるサンプル分割.	36
3.7	Joint Boosting と DaCBoost による境界の差.	39
3.8	INRIA Person Dataset での性能比較.	43
3.9	UIUC Image Database for Car Detection での性能比較.	43
4.1	Boosting に基づく特徴量の貢献度評価.	48
4.2	特徴削減に対する識別性能の比較.	51
4.3	計算時間の比較.	52
4.4	特徴貢献度に基づく識別器の調整.	53
4.5	モーメント特徴.	54
4.6	エッジ検出例.	55
4.7	オプティカルフローの抽出.	56
4.8	フローの分散の算出.	56

4.9	貢献度算出結果.	58
4.10	調整により改善された物体検出例.	60
5.1	生成型学習の流れ.	64
5.2	3次元人体モデル.	65
5.3	3次元人体モデル生成時のパラメータ.	66
5.4	実環境に対する3次元人体モデルの生成例.	67
5.5	提案手法におけるBagの構成.	68
5.6	Negative-Bag MILBoostによる学習サンプルの重みの更新.	71
5.7	学習データベース例.	75
5.8	各データベースで学習を行った際のDETカーブ.	76
5.9	人検出結果例 (緑: 正検出ウインドウ, 赤: 誤検出ウインドウ).	77
5.10	ノイズを混入した際の識別性能.	79
5.11	識別器の出力と重みの遷移.	79
6.1	俯角による学習サンプルの変化.	85
6.2	俯角の変化による転移学習の性能.	85
6.3	ハイブリッド型転移学習.	86
6.4	HOG特徴量の転移.	88
6.5	弱識別器のエラー率の推移 ϵ .	90
6.6	選択されたHOG特徴量.	92
7.1	Random Forestの構造.	97
7.2	Random Forestによる識別.	99
7.3	共変量シフトに基づく転移学習を導入したRandom Forest.	100
7.4	転移学習を導入したRandom Forestの構築.	101
7.5	各データベースのサンプル例.	104
7.6	転移学習を導入したRandom Forstsの識別性能.	106
7.7	マルチクラス分類問題のサンプル例.	107

表 目 次

3.1	サンプル分割手法の比較.	41
4.1	最大識別率 (RR) とその次元数 (dim) の比較.	52
4.2	調整前の識別性能.	57
4.3	識別器を再学習した際の識別性能.	59
4.4	識別器の調整 (HG).	59
4.5	識別器の調整 (SH,HG).	59
5.1	学習用データベースの内訳.	74
6.1	俯角による類似度の変化.	88
6.2	EER による性能評価.	91
6.3	目標識別器学習に必要な計算時間の比較.	92
7.1	Random Forest 構築のパラメータ.	105
7.2	マルチクラス分類問題に対する転移学習結果 (2,416 枚).	108
7.3	マルチクラス分類問題に対する転移学習結果 (800 枚).	108
7.4	マルチクラス分類問題に対する転移学習結果 (100 枚).	108

第1章

序論

本章では，本研究の背景及び目的，本論文の構成について述べる．

1.1 研究の背景

物体検出とは、対象とする物体を画像上で探索し、その位置や大きさを求める技術であり、高度道路交通システム (ITS) やビデオ監視システムなどのセキュリティ、顧客行動解析などのマーケティング、デジタルカメラにおける顔追跡など多岐にわたる分野への応用が可能であるため、1990年代から精力的に研究が行われている。物体検出は、一般環境下で撮影された画像を入力としてその中に存在する物体領域を出力するため、物体に対して一般性のある検出能力が求められている。ここでの一般性とは、物体の姿勢や向き、色や個体差に抛らず対象カテゴリに属する物体が検出可能であることを指す。カメラで撮影される画像において、物体の姿勢、向き、色、個体差等の複合的な要因は、その物体の見え自体を大きく変えてしまう。従って、同じ対象物体クラスの画像であってもクラス内の見えは大きく変動するため、画像中から物体を検出するのは難しい問題である。さらに、検出を難しくする要因は物体の多様性だけではなく、照明の変動や画像上での重なりによる物体領域の隠れなど環境に関する要因も物体検出を困難とする。

高精度な物体検出を実現するためには、画像から検出対象物体を表現する特徴を抽出し、多数の検出対象の画像から共通する要素を見つける必要がある。現在主流の物体検出法は、局所画像特徴量と統計的学習手法の組み合わせで実現されている。統計的学習手法による学習は、大量の学習サンプル画像に対して統計的に共通する要素を探索し、その要素を有効な特徴量として組み合わせることで識別器を構築することである。これにより、多くの物体画像に共通する要素に着目することで物体検出を実現している。しかしながら、画像局所特徴量と既存の統計的学習法を用いた物体検出法では実用化の条件を満たす検出性能に至らない場合がある。これは、多様な変化を持つ検出対象物体を同一のカテゴリとして学習することが困難なためである。例えば、画像上での物体の向きなどが変われば抽出される特徴量の値が大きく異なり、共通する情報が少なくなる。少ない情報を基に学習を行うため、未検出や誤検出も多く、長時間を費やしても目的とする性能に達しないことがある。

また、統計的学習手法を用いた場合、公開されている画像データベースによる評価が一般的であるが、必ずしも評価用データベースで高い性能を獲得した手法が実用化に適しているとは限らない。一般に、評価用データベースと実際の設置環境はカメラ位置などが異なるため、検出対象の見えも異なることが多い。統計的学習手法を用いた物体検出の性能は、学習サンプルに強い依存性がある。そのため、検出対象の見えが学習サンプルと異なると検出性能は低下し、実用的な精度に達しないという問題がある。設置環境に対する適応は、設置環境で新たに学習サンプルを収集し、そのサンプルに対して学習を行う再学習が一般的であり、再学習の際にも大量の学習サンプルの収集が不可欠である。しかし、学習サンプルの収集には大きな人的コストと時間的コストが必要となるため、設置環境毎に収集を行うのは困難である。そのため、人手の作業と、識別器の再学習に要する時間をできるだけ少なくし、かつ大量の学習画像を再収集した場合に比べて検出性能が劣ることのない、効率的な設置環境への適応を可能とするアプローチが望まれている。

ここまでをまとめると、物体検出の実用化という観点から、物体検出法は以下の条件を満たすことが重要である。

(1) 多様な見えのバリエーションを持つ検出対象の高精度な識別

人, 顔, 車両などの物体は多様な見えのバリエーションを持つことが一般的であり, 未検出や誤検出が発生しやすい. そのような検出対象を高精度に検出することは, 実用化の重要な条件である.

(2) システムを簡易に調整可能

実際にシステムを運用する環境は学習環境と異なる環境であり, 再学習が必要となる. 僅かな変化に対しては再学習を行うことなく, 簡易に識別器の調整が可能であることが望ましい.

(3) 特定環境に対してシステムを容易に再学習可能

物体の見えが大きく異なる場合, 識別器の調整では吸収できない. この場合は, カメラを設置する特定環境において学習サンプルを収集し, 再度識別器を学習することが必要となる. 学習サンプル収集のコストは膨大であり, 再学習におけるコストの低減は必須である.

(4) 少数の学習サンプルでシステムを再学習可能

条件 (3) に加え, 再学習の際に必要な学習サンプル数, 学習時間を可能なかぎり低減できることが望ましい.

1.2 研究目的

本研究の目的は, 1.1 で示した物体検出技術の実用化に向けて重要となる条件 (1) ~ (4) を満足するアプローチとして, 物体検出の高精度化と検出システムの設置環境への適応の効率化を実現することである.

条件 (1) に対しては, 分割統治戦略に基づく物体検出の高精度化を提案する. 従来の統計的学習法による識別器では, 多様に変化する物体の形状全てに対する高精度な識別が困難である. この問題は, 物体の多様な見えの変化全てを同一カテゴリとして学習することに起因する. これを解決するには, 物体カテゴリをその色, 形状, 姿勢など変化に応じて切り分け, 分割された複数の小問題に対して個別に最適な識別を行う必要がある. そこで, 本研究では学習サンプルを分割し, 分割された小問題に対して特徴を共有しながら学習する分割統治戦略による識別器を提案する. 本研究では, まず観測された学習用のサンプル集合を複数の類似した小集合に分割する手法として, 弱識別器の応答値を特徴空間としたサンプルのクラスタリングを提案する. 弱識別器の応答値は, 対象物体と背景領域の分類問題を学習した Boosting 識別器の解析により得られる特徴空間であり, 対象物体と背景領域の分類に有効な特徴量のみで構成される. この空間で複数の小問題にサンプルを分割することで, 物体の持つ多様な変化に対し個別に対処することが可能となり, 高精度な識別器の学習が可能となる. さらに, 小集合それぞれに対して各集合で共通する特徴を共有しながら学習を行

う Divide-and-Conquer Boosting(DaCBoost) を提案する。通常、複数の問題に対し各々識別器を構築する場合、それぞれの識別器が独立に同様の情報を学習するため、識別のための計算コストは大きく引き上がる。しかし、提案する DaCBoost では、共通する情報は個々の識別器において共有されるため、識別時の計算コストを大きく引き上げることはない。

条件 (2) に対しては、特徴量の貢献度に基づいて識別器の簡易な調整を可能にする手法を提案する。統計的学習法に基づく物体検出では、学習時の環境と実際の設置環境が一致しないと性能が低下するという問題がある。この問題を解決するために、本研究では Boosting 識別器より特徴量の貢献度を算出し、貢献度に基づいた識別器の調整や不要な特徴量の削減を行う。これにより、現場環境において簡易に環境変動への対処を行うことが可能となる。

カメラ俯角の変化等の要因により物体の見えが大きく変動する場合においては、再学習として、学習サンプルの再収集と識別器の再構築が必要である。しかし、学習サンプルは画像中の物体位置やサイズなどを正規化した良質なものが望ましく、そのような良質なサンプルを大量に収集するためには技術を持った人間が長時間を費やす必要があり、人的コストが高い。そこで、条件 (3) に対しては、CG を用いた生成型学習による再学習の効率化を提案する。3次元人体モデルを用いた CG により特定環境の学習用画像を大量に自動生成することで、設置環境への適応において人手の介在を大きく削減し、低コスト化が可能となる。

また、条件 (4) に対しては、転移学習と再学習を組み合わせたハイブリッド型転移学習による効率化を提案する。まず、転移学習に基づいて、より少ない学習用画像の追加と高速な学習時間での環境への適応を行う。さらに、部分的に再学習を取り入れることで、転移学習より高精度に、再学習より高速な環境適応が可能となる。

1.3 本論文の構成

本論文は、図 1.1 に示すように 8 つの章で構成されている。1 章では、本研究の背景と目的を述べた。2 章では、物体検出法と統計的学習法について述べる。3 章では、弱識別器の応答値に基づく学習サンプルの分割法と、DaCBoost による分割統治的戦略について提案する。4 章では、簡易に設置環境へ適応する方法として特徴貢献度に基づく特徴次元削減と識別器の調整法について提案する。5 章では、物体検出の実用化に向けて、学習環境と異なる設置環境への再学習に必要なコストを低減する方法として、3次元人体モデルによる学習用画像の自動生成と、その際発生する誤ラベル問題の解決方法として Negative-Bag MILBoost を提案する。6 章では、転移学習を導入した際に発生する負の転移による性能低下を解決する手段として、ハイブリッド型転移学習を提案する。7 章では、転移学習をマルチクラス識別器である Random Forest に展開する。8 章では、本論文の結論と展望について述べる。

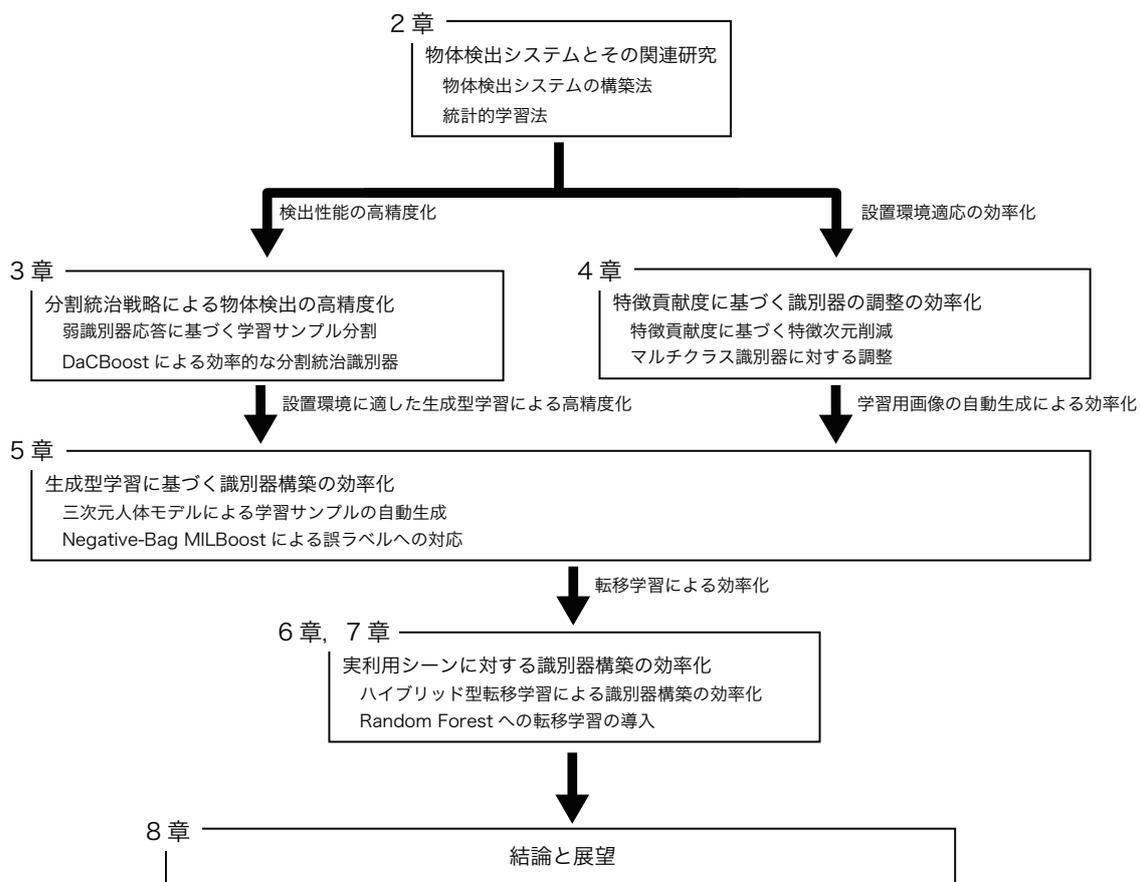


図 1.1: 本論文の構成.

第2章

統計的学習に基づく物体検出とその関連研究

物体検出は、画像から得られる特徴量を統計的学習法により学習することで構築した識別器を用いることで実現されている。本章では、統計的学習を用いた物体検出技術について述べ、特によく用いられる種々の統計的学習法について詳述する。また、これら統計的学習法を用いた物体検出を困難とする具体的な要因についても述べる。

2.1 統計的学習を用いた物体検出システム

本研究で対象とする物体検出システムについて述べる。物体検出問題は、画像中から物体の存在する領域の位置と大きさを求める問題である。物体検出は、図 2.1 に示すようにカメラで撮像した一般環境下での画像に対してウインドウスキャンを行い、各ウインドウが対象物体であるか否かを判定することで実現する。判定には、各ウインドウの画像特徴量と統計的学習により学習した識別器を用いる。統計的学習法とは、その特徴パターンが対象物であるか否かの判別基準を統計的に学習し、識別器を構築する手法である。顔検出 [1, 2] で用いられた Boosting, 人検出 [3] で用いられた SVM などが多く利用されている。本章ではこれらの統計的学習法について述べる。

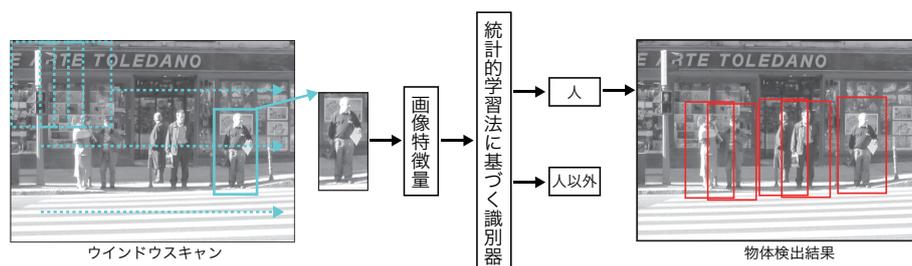


図 2.1: 統計的学習法を用いた物体検出システム (例: 人検出)。

2.2 統計的学習に基づく識別器

本節では、物体検出に用いられる統計的学習法として、Boosting と SVM について述べる。

2.2.1 Boosting

複数の識別器を組み合わせることで、汎化性能の高い識別器を構築する学習法として、アンサンブル学習がある。アンサンブル学習法には Bagging や Boosting 等が存在する。Bagging はランダムにサンプリングした学習サンプルに対して識別器を多数構築して確率的に汎化性能を向上するのに対し、Boosting では前段の識別器の性能不足を補うように識別器を逐次構築することで識別性能を向上する。この特性から、Bagging は回帰問題に、Boosting は識別問題に用いられているアンサンブル学習法である。

Boosting のアプローチは、1988 年に Kearns らの提起した「弱い識別器 (真の分布をそれほど高くない精度で分類する識別器) の集合は強い識別器 (真の分布を高い精度で識別する識別器) となり得るか?」という疑問に根ざしている [4]。この疑問に対する解答として提案された Boosting は、識別性能がそれほど高くない識別器を逐次的に複数組み合わせ、識別性能の高い識別器を実現する。組み合わせられる低性能な識別器を弱識別器 $h(x)$ 、複数の弱識別器 $h(x)$ から構成される高精度な識別器を強識別器 $H(x)$ と呼ぶ。この弱識別器をどのように選択、追加してより高精度な強識別器を構

築するかが Boosting の本質である。Boosting の代表的な学習アルゴリズムとして AdaBoost[5] がある。AdaBoost が提案されて以降、Real Adaboost[6]、Logit Boost[7]、Float Boost[8]、LP Boost[9] など AdaBoost の学習アルゴリズムが改良されてきた。ここでは、AdaBoost と AdaBoost を改良した Real AdaBoost の学習アルゴリズムについて述べる。

■ AdaBoost

Boosting の学習アルゴリズムでは、何らかの指標により弱識別器 $h(\mathbf{x})$ を逐次的に選択することで強識別器 $H(\mathbf{x})$ を学習する。この際の指標には重み付き誤差を用いるのが一般的である。AdaBoost の学習アルゴリズムを **アルゴリズム 2.1** に示す。AdaBoost では、学習サンプルに対する重み w という概念を取り入れ、弱識別器 $h(\mathbf{x})$ を選択後に学習サンプルに対して重み付けする。この学習サンプルに対する重みは、現在までに選択された弱識別器群、すなわち強識別器において識別が困難であるサンプルほど高く、次の弱識別器 $h(\mathbf{x})$ を選択する際に、その学習サンプルをどれほど考慮するかを表す。弱識別器を追加する度に、正しく識別したサンプルに対しては小さな重み、誤識別したサンプルに対しては大きな重みを与える。これを繰り返すことにより強識別器 $H(\mathbf{x})$ を学習する。

未知入力サンプルを識別する際には、図 2.2 に示すように各弱識別器 $h(\mathbf{x})$ の識別結果と重み α の積を計算し、これらの総和を強識別器 $H(\mathbf{x})$ の出力とする。この重み付き多数決といえる強識別器の出力がしきい値以上の値であれば対象物体、しきい値以下の値であれば非対象物体と識別される。

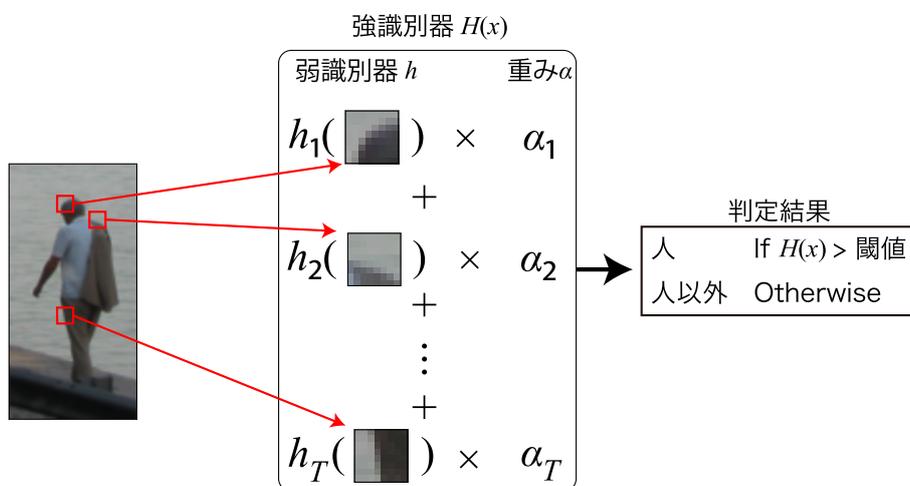


図 2.2: AdaBoost による人の識別の流れ。

AdaBoost の弱識別器 $h(\mathbf{x})$ は出力が 2 値であることから Discrete AdaBoost とも呼ばれる。また、AdaBoost の弱識別器 $h(\mathbf{x})$ はアルゴリズム的には明確に定義されておらず、識別率が 50% を超えるものであれば、どのように設計しても良いとされている。式 (2.3) で表わされる弱識別器 $h(\mathbf{x})$ は、文献 [2] の顔検出で用いられているしきい値処理を利用した識別器である。この弱識別器は、特徴量の値 $v(\mathbf{x})$ としきい値 θ を比較することにより特徴量を 2 値化する。 $p \in \{-1, +1\}$ はパリティ符号を表し、しきい値と特徴量の大小関係を変えるための符号である。

アルゴリズム 2.1: しきい値関数を弱識別器に用いた AdaBoost の学習アルゴリズム.

1. 入力: I 個の学習サンプル $\{\mathbf{x}_1, y_1\} \dots \{\mathbf{x}_I, y_I\}$ を用意する.

\mathbf{x} は学習サンプル, $y_i \in \{-1, 1\}$ は学習サンプルが属するクラスラベルを表す.

2. 初期化: 学習サンプルの重み D を初期化

$$D_1(i) = 1/I \quad (2.1)$$

3. 学習:

For $t = 1, \dots, T$ //学習ラウンド

- ・ 全ての弱識別器候補からエラー率 ϵ_t が最小の弱識別器候補 $h_t(\mathbf{x})$ を選択
- ・ エラー率 ϵ_t の算出

$$\epsilon_t = \sum_{i: h_t(\mathbf{x}_i) \neq y_i} D_t(i) \quad (2.2)$$

- ・ 弱識別器

$$h_t(\mathbf{x}) = \begin{cases} 1 & \text{if } p \cdot v(\mathbf{x}) > p \cdot \theta \\ 0 & \text{otherwise} \end{cases} \quad (2.3)$$

- ・ 弱識別器 h の重み α を算出

$$\alpha_t = \frac{1}{2} \log \left(\frac{1 - \epsilon_t}{\epsilon_t} \right) \quad (2.4)$$

- ・ 学習サンプルの重み D を更新・正規化

$$D_{t+1}(i) = \frac{D_t(i) \exp[-\alpha_t y_i h_t(\mathbf{x}_i)]}{\sum_{i=1}^I D_t(i) \exp[-\alpha_t y_i h_t(\mathbf{x}_i)]} \quad (2.5)$$

End for

4. 出力: 強識別器

$$H(\mathbf{x}) = \text{sign} \left[\sum_{t=1}^T \alpha_t h_t(\mathbf{x}) \right] \quad (2.6)$$

■ Real Adaboost

AdaBoost における弱識別器の出力 $h(\mathbf{x})$ は $\{+1, -1\}$ であり, \mathbf{x} の微小な変化によりしきい値処理の結果が反転すると, その重み α が反転して強識別器の出力に影響する. また, 多峰性を持つような複雑な学習サンプルの分布を表現するには多くの弱識別器を必要とする. Real AdaBoost では弱識別器の出力を実数として扱う拡張を取り入れている. そのため, 弱識別器の出力値は, 明確に対象クラスに属する場合大きくなり, 非対象クラスとの境界付近に位置する曖昧な場合では小さくなる.

これは、AdaBoost (Discrete) に比べ弱識別器 1 つあたりの分布の表現能力を大きく拡張し、少ない弱識別器数で高精度な識別を可能にする。Real AdaBoost の弱識別器も、AdaBoost と同様にどのように設計しても良いとされている。本論文では、文献 [10] で用いられた確率密度関数を利用した方法を採用する。確率密度関数は、学習サンプル全てを特徴量毎にヒストグラム化することで得られ、入力されたサンプルの特徴量の値から対象クラスらしさを確率分布で表現する。Real AdaBoost の学習アルゴリズムを **アルゴリズム 2.2** に示す。AdaBoost との大きな違いは、弱識別器の出力をしきい値による 2 値出力関数から確率密度関数 W による確率 (式 (2.11)) に基づく実数値へと変更したことである。式 (2.11) は入力されたパターンの観測値 k が $0 \sim K$ のどの区間に収まるかによって異なる出力を返す。学習ラウンド毎に式 (2.9), (2.10) で確率密度関数を求めることで、サンプルの重み $D_t(i)$ を考慮した重み付き誤差最小の弱識別器は式 (2.8) を最大とする弱識別器として求められる。弱識別器を選択する。サンプルの重み $D_t(i)$ は式 (2.12) により、選択された弱識別器で正解できなかったサンプルを重視するよう逐次更新される。

アルゴリズム 2.2: 確率密度関数を弱識別器に用いた Real AdaBoost の学習アルゴリズム.

1. 入力: I 個の学習サンプル $\{\mathbf{x}_1, y_1\} \dots \{\mathbf{x}_I, y_I\}$ を用意する.
 \mathbf{x} は学習サンプル, $y_i \in \{-1, 1\}$ は学習サンプルが属するクラスラベルを表す.

2. 初期化: 学習サンプルの重み D を初期化

$$D_1(i) = 1/I \quad (2.7)$$

3. 学習:

For $t = 1, \dots, T$ //学習ラウンド

· 評価値 Z_t が最大の弱識別器 $h(\mathbf{x})$ を選択

· 評価値 Z_t の算出

$$Z_t = 1 - 2 \sum_k \sqrt{W_+^k W_-^k} \quad (2.8)$$

· 確率密度関数 W_+, W_- の算出

$$W_+^k = \sum_{i: k \in K \wedge y_i = +1} D_t(i) \quad (2.9)$$

$$W_-^k = \sum_{i: k \in K \wedge y_i = -1} D_t(i) \quad (2.10)$$

· 弱識別器

$$h(\mathbf{x}) = \frac{1}{2} \ln \frac{W_+^k + \varepsilon}{W_-^k + \varepsilon}, \quad (\varepsilon = 1/I) \quad (2.11)$$

End for

· 学習サンプルの重み D を更新・正規化

$$D_{t+1}(i) = \frac{D_t(i) \exp[-y_i h_t(\mathbf{x}_i)]}{\sum_{i=1}^I D_t(i) \exp[-y_i h_t(\mathbf{x}_i)]} \quad (2.12)$$

End for

4. 出力: 強識別器

$$H(\mathbf{x}) = \text{sign} \left[\sum_{t=1}^T h_t(\mathbf{x}) \right] \quad (2.13)$$

2.2.2 Support Vector Machine

Support Vector Machine(SVM)は1960年代にVapnik等が考案したOptimal Separating Hyperplaneを起源とし、1990年代になってカーネル学習法と組み合わせた非線形の識別手法へと拡張された。カーネルトリックにより拡張したSVMは、現在知られている手法の中で最もパターン認識性能の優秀な学習モデルの一つである。その大きな特徴として、マージン最大化に基づく汎化性能の獲得とカーネルトリックによる非線形問題への対応が挙げられる。

マージン最大化 まず、図2.3(a)に示す線形分離可能な問題について考える。このような分布から線形判別関数を用いて識別境界を求める際、図2.3(a)のように複数の識別境界を考えることができる。これらの境界は、学習サンプルに対する識別率において等価である。しかし、実際に未知入力を識別させると、図2.3(a)中において境界線(1)や境界線(3)よりも境界線(2)の方がよい結果を示すことが多い。これは、境界線(2)はサンプルから識別境界への距離(マージン)が大きく、サンプルと実データの差を飲み込む見込みが最も高いためである。このような、実データに対する頑健性を汎化性能という。SVMは、識別境界と境界付近のサンプル群との距離を最大とすることで汎化性能の高い識別境界を構築する。

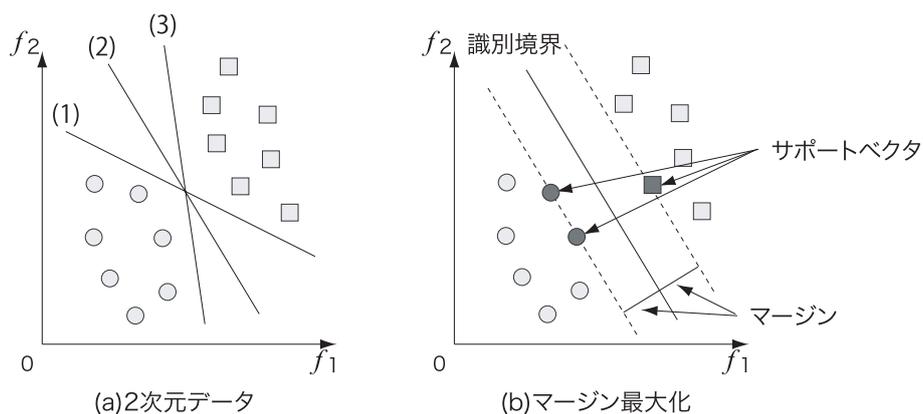


図 2.3: SVM によるマージン最大化.

各サンプルのマージンはサンプルから識別境界までの距離であるため、入力 \mathbf{x} に対する線形識別器を

$$f(\mathbf{x}) = \langle \mathbf{w} \cdot \mathbf{x} \rangle + b \quad (2.14)$$

$$= \sum_{i=1}^n w_i x_i + b \quad (2.15)$$

とすれば、その距離は

$$\begin{aligned}\gamma &= \frac{1}{2} \left(\frac{f(\mathbf{x}^+)}{\|\mathbf{w}\|_2} - \frac{f(\mathbf{x}^-)}{\|\mathbf{w}\|_2} \right) \\ &= \frac{1}{2\|\mathbf{w}\|_2} \{ (\langle \mathbf{w} \cdot \mathbf{x}^+ \rangle + b) - (\langle \mathbf{w} \cdot \mathbf{x}^- \rangle + b) \}\end{aligned}\quad (2.16)$$

として表現される。ここで \mathbf{x}^+ は正であるサンプル、即ちラベル $y = 1$ であるサンプルであり、 \mathbf{x}^- は負、即ち $y = 0$ であるサンプルである。black また、 I はサンプル数である。そこで、学習サンプル全てに正解する上で

$$\gamma = \frac{1}{2\|\mathbf{w}\|_2} \quad (2.17)$$

を最大とする境界を求めることで汎化性能を大きくすることができる。このマージン γ を最大化するには、ノルム $\|\mathbf{w}\|_2$ を最小化すればよい。マージンの最大化に関わるのは識別境界面に最も近いサンプルのみであるため、境界の構築に関わるか否かを重み付けして最適化する。最適化には、式 (2.19) に示す Lagrange の未定乗数法により得られる更新式を用いる。Lagrange の未定乗数法は束縛条件のもとで最適化を行う手法であり、本問題では識別性能とマージン双方の最大化を解く。また、この問題を二次凸最適化問題として解くために Karush-Kuhn-Tucker の相補条件 (KKT 条件) の制約が存在する。最急降下法を用いて解くと、SVM の学習は **アルゴリズム 2.3** となる。

式 (2.15) で用いる識別時の重みベクトル \mathbf{w} は

$$\mathbf{w} = \sum_{i=1}^I \alpha_i y_i \mathbf{x}_i \quad (2.21)$$

で求められる。ここで、識別境界付近以外のサンプルの持つ α は全て 0 であるため、重み $\alpha > 0$ となるサンプル \mathbf{x} のみの集合を \mathcal{S} とすると \mathbf{w} は以下となる。

$$\mathbf{w} = \sum_{i \in \mathcal{S}} \alpha_i y_i \mathbf{x}_i \quad (2.22)$$

\mathcal{S} に含まれるサンプルは、識別境界を支持することからサポートベクタと呼ばれる。ただし本アルゴリズムは線形分離可能な問題でない場合、線形分離が困難な少数の特異な点が大きな α を持ち、それらの点の境界が重視されるため、データ全体を正しく分類できない。そこで、実問題には誤差をどの程度許すかというパラメータを設置したソフトマージン SVM が用いられる。

カーネルトリック SVM を用いることで、線形分離可能な問題に対して高精度な識別境界を得ることができる。しかし、非線形で複雑な識別課題に対しては、必ずしも良い性能の識別器を構築できるとは限らない。非線形な問題に対応するための方法として、特徴ベクトルを非線形変換して、その空間で線形の識別を行う「カーネルトリック」と呼ばれる方法が知られている。この方法を用いることで複雑な識別境界を設定できるため、SVM の性能は劇的に向上する。

アルゴリズム 2.3: SVM の学習アルゴリズム (最急降下法) .

1. 入力: I 個の学習サンプル $\{\mathbf{x}_1, y_1\} \dots \{\mathbf{x}_I, y_I\}$ を用意する
 \mathbf{x} は学習サンプル, $y_i \in \{-1, 1\}$ は学習サンプルが属するクラスラベルを表す

2. 初期化: Lagrange 乗数 $\alpha_i = 0$ (2.18)

3. 学習:

For $t = 1, \dots, T$ //ここでは最適化回数を停止基準とする
For $m = 1, \dots, I$ //全サンプルについて最適化

· α を更新

$$\alpha_i = \alpha_i + \eta(1 - y_i \sum_{j=1}^I \alpha_j y_j \mathbf{x}_j^\dagger \mathbf{x}_i) \quad (2.19)$$

η は任意の学習係数

· KKT 条件 //識別境界付近のサンプル以外の重みを 0 にする

$$\alpha_i = 0 \text{ if } \alpha_i < 0 \quad (2.20)$$

End for

End for

4. 出力: 識別器

· 学習サンプルのうち $\alpha_i > 0$ でないものを抽出し, サポートベクタ \mathbf{S} とする
· サポートベクタ \mathbf{S} から重み係数ベクトル \mathbf{w} を算出

$$\mathbf{w} = \sum_{i \in \mathbf{S}} \alpha_i y_i \mathbf{x}_i$$

· 識別関数

$$f(\mathbf{x}) = \langle \mathbf{w} \cdot \mathbf{x} \rangle + b$$

一般に, 線形分離可能性はサンプル数が大きくなればなるほど難しくなり, 逆に, 特徴空間ベクトルの次元が大きくなるほど易くなる. 例えば, 特徴ベクトルの次元がサンプルの数よりも大きいなら, どんなラベル付けに対しても線形分離可能である. 図 2.4 に示すように, 2次元では分類不可能な問題をカーネル関数 $k(\mathbf{x}_i, \mathbf{x}_j)$ を用いて高次元に写像することで線形識別器である SVM を適用する.

カーネルとしては, 以下に示す多項式カーネル, RBF カーネル, シグモイドカーネル等が用いられる.

多項式カーネル

$$k(\mathbf{x}_i, \mathbf{x}_j) = (1 + \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle)^p \quad (2.23)$$

RBF カーネル

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|_2}{2\sigma^2}\right) \quad (2.24)$$

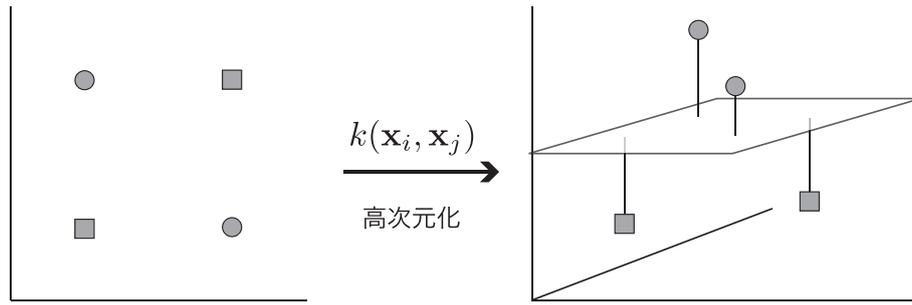


図 2.4: カーネルトリックを用いた高次元への写像.

シグモイドカーネル

$$k(\mathbf{x}_i, \mathbf{x}_j) = \tanh(c\mathbf{x}_i \cdot \mathbf{x}_j + \theta) \quad (2.25)$$

カーネルの選定は、サンプルの分布に対応したものを実験的に求めて設定することが多い。カーネル関数として利用できる関数の条件として、関数の対称性、Cauchy-Schwartz 不等式、Mercer の定理の 3 つを満たすことが必要である。関数の対称性とは

$$k(\mathbf{x}_i, \mathbf{x}_j) = k(\mathbf{x}_j, \mathbf{x}_i) \quad (2.26)$$

となることであり、Cauchy-Schwartz 不等式とは簡単に書けば

$$k(\mathbf{x}_i, \mathbf{x}_j)^2 = k(\mathbf{x}_i, \mathbf{x}_i)k(\mathbf{x}_j, \mathbf{x}_j) \quad (2.27)$$

を満足することである。そして、Mercer の定理は、対称性に加え

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sum_{l=1}^{\infty} \lambda_l \phi_l(\mathbf{x}_i) \phi_l(\mathbf{x}_j), \quad \lambda_l \geq 0 \quad (2.28)$$

が半正定を満たせば良い。これらのカーネルを用いることにより、非線形な識別境界を SVM によって設定することができる。

2.2.3 Boosting と SVM

Boosting は識別困難なサンプルを重視し、SVM は識別境界付近のサンプルを重視することで高精度な識別を可能とする。識別困難なサンプルとは、弱識別器で識別した際に別のクラスに間違えられるサンプルであり、つまり識別境界付近のサンプルである。そのため、両者は類似した性能を

持つことが多い。そこで、この2つの統計的学習法の違いは、どの弱識別器をどれだけ重視するか（AdaBoostにおける h, α ）と、どのサポートベクタとの類似をどれだけ重視するか（SVMにおける w ）という識別器の形の違いであるといえる。Boosting は、弱識別器を任意の特徴量1次元を用いたしきい値関数または確率密度関数とすることが可能である。この場合、Boosting では弱識別器がどの特徴量を捉えているかを解析することが容易である。本研究では、識別器の解析が容易であるため、統計的学習法の高精度化と効率化の基盤として、特徴量1次元に基づく判別関数を弱識別器とした Boosting を用いる。

2.3 マルチクラス識別器

AdaBoost, SVM は対象クラス, 背景クラスの2クラス分類を扱う手法である。複数種の検出対象を1つのシステムで同時に識別するマルチクラス分類も重要なタスクであり、盛んに研究が行われている。マルチクラス分類では、検出対象毎の2クラス分類問題として解く One vs Other 戦略が取られることが多いが、この戦略は独立した識別器を多数用意する必要があるため、識別結果の最適な統合が困難である。そこで、1999年に、検出対象毎の識別器を統合的に構築する手法として AdaBoost.MH[6] が提案されている。さらに2007年には、特徴の共有により効率的にマルチクラス識別器のを構築する Joint Boosting[11] が提案されている。

2.3.1 One vs Other 戦略

簡易にマルチクラス識別を実現する手法として、One vs Other 戦略がある。これは、対象とするクラス毎に自クラス (One) と他クラス全て (Other) の2クラス分類識別器を構築し、未知入力サンプルに対して、全てのクラスの出力を求める。そして、各クラスの出力を統合することでマルチクラスの判定結果を得る。しかし、複数の識別器を独立に学習するため識別器の出力値の範囲はクラスごとに大きく異なり、どのようにスケールリングして統合するかは経験的に決定する必要がある。One vs Other 戦略によるマルチクラス Boosting の強識別器は、図 2.5 のように構成される。識別は、未知入力サンプル \mathbf{x} に対して、各クラス c の2クラス識別器の出力 $H^c(\mathbf{x})$ を次式で求め、最大の出力を持つ識別器のクラスを最終的な出力とする。

$$H^c(\mathbf{x}) = \sum_{t=1}^T h_t^c(\mathbf{x}) \quad (2.29)$$

$$\hat{c} = \arg \max_c H^c(\mathbf{x}) \quad (2.30)$$

学習アルゴリズムは各クラスに対して**アルゴリズム 2.1**をそれぞれ適用する。

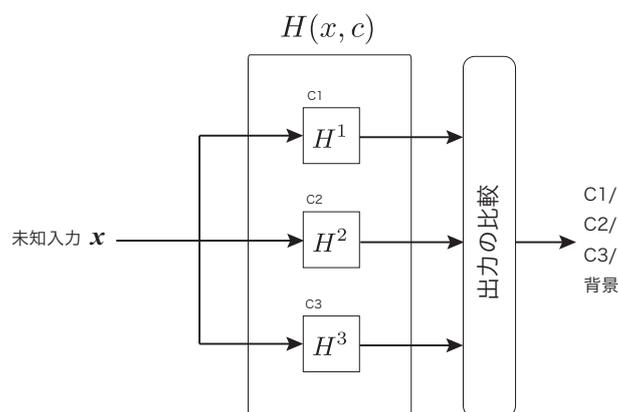


図 2.5: マルチクラス識別器.

2.3.2 AdaBoost.MH

One vs Other 戦略では、どのクラスに対してどの程度学習を行うかという点などを実験的に試行錯誤する必要がある。そこで、より効率的にマルチクラス問題を学習する方法として AdaBoost.MH[6] が提案されている。AdaBoost.MH の識別器は、One vs Other 戦略による識別器と同様に独立した各クラスに対する複数の 2 クラス識別器の集合であるが、学習過程において性能が低いクラスを重視して学習する点異なる。まず、クラスごとにサンプルの重みを用意する。次に、各クラスのうちどれか 1 クラスを対象とした弱識別器を選択する。この選択の際、クラス c の評価値算出は c に対応する重みを用いて重み付き誤差評価を行うため、現在エラーの大きいクラスについて正解可能な弱識別器の誤差評価値は小さくなり選択されやすくなる。これを繰り返すことで、クラスに大きく偏ることなくマルチクラス学習を行うことができる。AdaBoost.MH の強識別器の構成は、One vs Other 同様に各クラスに毎に独立した識別器の集合である。強識別器の学習は逐次的に、最も性能の低いクラスの精度を改善するよう、そのクラスの弱識別器を選択する。そのため識別器を、対象とするクラスを示す変数 c を入力に加え次式のように表現する。

$$H(\mathbf{x}, c) = \sum_{m=1}^M h_m(\mathbf{x}, c) \quad (2.31)$$

ここで、 \mathbf{x} は入力サンプル、 M は弱識別器 h_m の総数、 c はクラスラベルである。全クラスに対して $H(\mathbf{x}, c)$ を求め、最大の値を持つ $H(\mathbf{x}, c)$ が対象とする c が最終的な識別結果のクラスとなる。

アルゴリズム 2.4: AdaBoost.MH の学習アルゴリズム.

1. 入力: I 個の学習サンプル $\{\mathbf{x}_1, z_1\} \dots \{\mathbf{x}_I, z_I\}$ を用意する.
 \mathbf{x} は学習サンプル, $z_i \in \{-1, 1, \dots, C\}$ は学習サンプルが属するクラスラベルを表す.

2. 初期化: 学習サンプルの重み w を初期化

$$w_1^c(i) = 1/I \quad (2.32)$$

3. 学習:

For $t = 1, \dots, T$ //学習ラウンド

For $c = 1, \dots, C$ //すべてのクラス

・ 全ての c , 全ての特微量について評価値 J_{wse} を算出し,
 J_{wse} を最小とする $h_m(\mathbf{x}_i, c)$ を弱識別器として選択

$$J_{wse} = \sum_{c=1}^C \sum_{i=1}^N w_i^c (y_i^c - h_m(\mathbf{x}_i, c))^2 \quad (2.33)$$

End for

・ 学習サンプルの重み w を更新

$$w_{t+1}^c(i) = w_t^c(i) \exp^{(y_i^c - h_m(\mathbf{x}_i, c))^2} \quad (2.34)$$

End for

4. 出力: 強識別器

$$H(\mathbf{x}, c) = \sum_{m=1}^M h_m(\mathbf{x}, c) \quad (2.35)$$

2.3.3 Joint Boosting

AdaBoost を基に識別性能を向上した学習法として gentleboost[12] がある。Torralba らが提案した Joint Boosting[11] は、gentleboost をマルチクラス問題へ拡張している。マルチクラスへの拡張は、AdaBoost.MH のフレームワーク同様に、評価値を最小とする弱識別器を選択する形で実現する。AdaBoost.MH との違いは、弱識別器を共有し、弱識別器数の低減を可能とした点である。

■ gentleboost

gentleboost は 2 クラス問題に対する AdaBoost の発展形であり、AdaBoost に比べて、学習サンプルの分布を表現する能力が高い弱識別器を採用している。AdaBoost における弱識別器の出力は、しきい値判定を行った際の符号 $h(\mathbf{x}) \in \{1, -1\}$ と重み α の積で表現される。これは、しきい値判定結果によって出力がそのまま反転してしまい、繊細な出力が困難である。gentleboost の弱識別器は、 $h(\mathbf{x})$ の出力を実数 $h(\mathbf{x}) \in [-1, 1]$ とし、しきい値判定の結果に応じて正負それぞれ異なる大きさを出力できる。この点から、gentleboost はしきい値処理を用いながらも、2 区間のみで確率密度関数を構成した Real AdaBoost に近い性質を持つ。従って、Discreat 型と Real 型の間中間的な Boosting であるといえる。gentleboost の強識別器は以下となる。

$$H(\mathbf{x}) = \sum_{m=1}^M h_m(\mathbf{x}) \quad (2.36)$$

ここで、 \mathbf{x} は入力サンプル、 M は弱識別器 h_m の総数である。弱識別器 h_m は次式により計算される。

$$h_m(\mathbf{x}) = a\delta(x^f > \theta) + b \quad (2.37)$$

x^f は入力における特徴 f の値、 θ はしきい値、 δ は Kronecker のデルタ関数を表す。 a, b はしきい値判定された際の出力を決定するパラメータである。ここで、 a, b のパラメータは、それぞれしきい値以上の区間におけるサンプルの重み付き分布、しきい値以下の区間におけるサンプルの重み付き分布を式 (2.38)、式 (2.39) で表現する。

$$b = \frac{\sum_i w_i y_i \delta(\mathbf{x}_i^f \leq \theta)}{\sum_i w_i \delta(\mathbf{x}_i^f \leq \theta)} \quad (2.38)$$

$$a + b = \frac{\sum_i w_i y_i \delta(\mathbf{x}_i^f > \theta)}{\sum_i w_i \delta(\mathbf{x}_i^f > \theta)} \quad (2.39)$$

また、 y は \mathbf{x} が正例か負例かを示すラベルであり、 $\{1, -1\}$ をとる。式について指数評価関数 J を最も低くする弱識別器 h を逐次選択する。

$$J = E \left[e^{-yH(\mathbf{x})} \right] \quad (2.40)$$

その際、各サンプル i の重みが $w_i = e^{-z_i H(\mathbf{x}_i)}$ として更新される。gentleboost を AdaBoost.MH 同様マルチクラスに拡張した際の強識別器は次式である。

$$H(\mathbf{x}, c) = \sum_{m=1}^M h_m(\mathbf{x}, c) \quad (2.41)$$

このとき、弱識別器 h_m は以下となる。

$$h_m(\mathbf{x}, c) = a\delta(\mathbf{x}^f > \theta) + b \quad (2.42)$$

ここで、 \mathbf{x} は入力サンプル、 M は弱識別器 h_m の総数、 c はクラスラベルである。また、 \mathbf{x}^f はサンプル \mathbf{x} の f 番目の要素値を表す。

■ 特徴を共有するマルチクラス学習

Joint Boosting は、効率的なマルチクラス分類のための Boosting 手法であり各クラスに共通する特徴量（弱識別器）を共有しながら学習を行う。

AdaBoost.MH では、各クラスに対して弱識別器を設定して強識別器を逐次構築する。そのため、クラス間で共有可能な特徴量を共有できない。Joint Boosting では、ラベル c の対象を各クラスだけでなく、その和集合に対しても拡張し、各クラスとクラスの和集合をノード $S(n)$ で表現する。

例えば 3 クラス (1, 2, 3) の場合、ノード $S(n)$ は $S(1) : (1)$, $S(2) : (2)$, $S(3) : (3)$, $S(4) : (1, 2)$, $S(5) : (1, 3)$, $S(6) : (2, 3)$, $S(7) : (1, 2, 3)$ をそれぞれ表現する。これらノードを並列に置き、各ノードについて誤差最小の弱識別器 $h_m(v, S(n))$ を逐次探索する。

ここで、探索で得られた弱識別器を各ノードごとに格納する集合 $G^{S(n)}$ を図 2.6 のように定義する。弱識別器集合 $G^{S(n)}$ はそれぞれ $S(n)$ の表現するクラス集合に対し識別性能を持つため、図 2.6 中の上位であるほど複数のクラスで共有可能な識別器が格納されている。つまり、 G^{123} は検出対象 1 ~ 3 全てと背景を、 G^{12} は検出対象 1 と 2 全てと背景を分離するのに有効な弱識別器群である。この G^{123} と G^{12} は検出対象 1 に対する強識別器 $H(\mathbf{x}, 1)$ 、検出対象 2 に対する強識別器 $H(\mathbf{x}, 2)$ で共有される。この共有により、同様の弱識別器が各クラス毎に学習される無駄を防ぐことができる。Joint Boosting による各クラスに対する最終的な強識別器はそれぞれ以下となる。

$$\begin{aligned} H(\mathbf{x}, 1) &= G^{123}(\mathbf{x}) + G^{12}(\mathbf{x}) + G^{13}(\mathbf{x}) + G^1(\mathbf{x}) \\ H(\mathbf{x}, 2) &= G^{123}(\mathbf{x}) + G^{12}(\mathbf{x}) + G^{23}(\mathbf{x}) + G^2(\mathbf{x}) \\ H(\mathbf{x}, 3) &= G^{123}(\mathbf{x}) + G^{13}(\mathbf{x}) + G^{23}(\mathbf{x}) + G^3(\mathbf{x}) \end{aligned} \quad (2.43)$$

学習は、あるクラス集合 $S(n)$ に対応する弱識別器 $h(\mathbf{x})$ を毎ラウンド選択することで進む。 $h(\mathbf{x}, c)$ はノード $S(n)$ に含まれるクラス全てで共有される。その際、最良の弱識別器を選択するために以下の評価関数 J_{wse} という重み付き二次誤差を用いる。

$$J_{wse} = \sum_{c=1}^C \sum_{i=1}^N w_i^c (y_i^c - h_m(\mathbf{x}_i, c))^2 \quad (2.44)$$

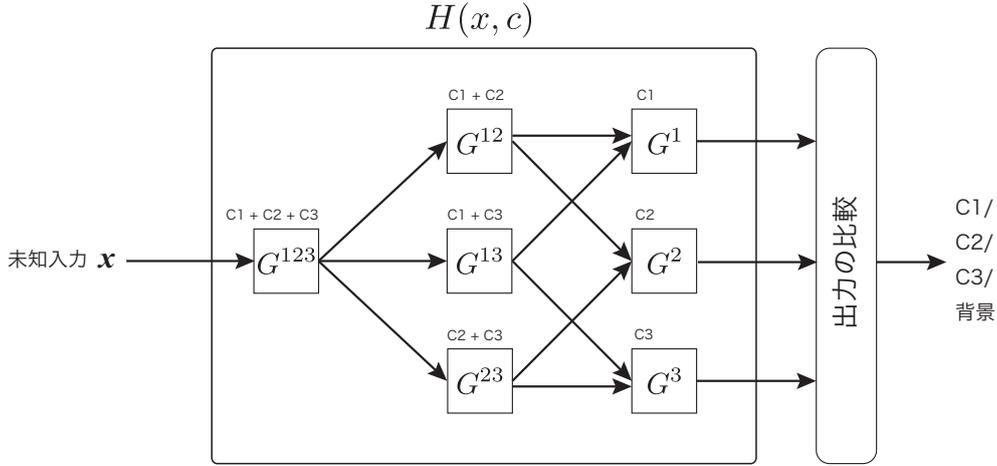


図 2.6: 弱識別器の共有.

C はクラス数, c は各クラスラベルを表し, 集合に対する弱識別器 $h_m(\mathbf{x}_i, S(n))$ を再度 $h_m(\mathbf{x}_i, c)$ として c 毎に表現している. また, y^c は \mathbf{x} がクラス c に含まれるかどうかを示すラベルであり, $\{1, -1\}$ をとる. この評価値を最小にするよう弱識別器を選択することで, 汎化性能の高い識別器の構築が期待できる. また, 以下の拡張により弱識別器をノードによるマルチクラス識別に適応させている.

$$h_m(\mathbf{x}, c) = \begin{cases} a\delta(\mathbf{x}^f > \theta) + b & \text{if } c \in S(n) \\ k^c & \text{otherwise} \end{cases} \quad (2.45)$$

ここで, a, b, k^c のパラメータはそれぞれ以下となる.

$$b = \frac{\sum_{c \in S(n)} \sum_i w_i^c z_i^c \delta(\mathbf{x}_i^f \leq \theta)}{\sum_{c \in S(n)} \sum_i w_i^c \delta(\mathbf{x}_i^f \leq \theta)} \quad (2.46)$$

$$a + b = \frac{\sum_{c \in S(n)} \sum_i w_i^c z_i^c \delta(\mathbf{x}_i^f > \theta)}{\sum_{c \in S(n)} \sum_i w_i^c \delta(\mathbf{x}_i^f > \theta)} \quad (2.47)$$

$$k^c = \sum_{c \notin S(n)} \frac{\sum_i w_i^c z_i}{\sum_i w_i^c} \quad (2.48)$$

b はしきい値以下の区間における確率密度を, a はしきい値以上の区間における確率密度をそれぞれ反映し, 負例が多い区間では負方向, 正例が多い区間では正方向の強度を示す. k^c は弱識別器 h_m が対象としないクラスについて識別器は k^c を応答として返し, 識別関数は介在しない.

2.4 統計的学習による物体検出を困難とする要因

本節では物体検出が困難となる要因について整理する.

現在主流である物体検出は、画像局所特徴量と統計的学習法の組み合わせにより実現され、統計的学習法による識別器には SVM や Boosting などの利用が提案されている。特に、Boosting は弱識別器として採用された特徴量のみを計算すればよいため計算コストが低く、特徴量の共起関係の表現が容易であるため、人、顔、車両等の物体検出に多く用いられている [13][14][15]。

検出対象となる物体の見えは、照明、視点によって大きく異なり、検出対象の向きや姿勢により、形状、テクスチャなどが全く異なる画像が得られる。人検出は、検出対象が非剛体であることと、その個体差の多様性から物体検出の中でも特に困難とされている。文献 [16] では、人検出を困難とする要因について、以下のように¹まとめられている。

1. 見えの個体差

「見えの個体差とは、人の衣服や体格など同一人クラス内における各個人の差である。衣服の色は人によって異なるため、人検出への有効な情報として利用しにくい。また、体格は大人と子供、性別によっても大きく異なるため、頭部や胴体、足などの見えや位置が人画像毎に異なる。」

2. 向きの変化

「人を正面から撮影する場合と横から撮影する場合では人の見えが異なる。」

3. 姿勢の変化

「人は非剛体な物体であり、自由な姿勢をとることができるため、姿勢により人の形状は大きく変化する。」

4. 視点の変化

「人の向きの違いと似ているが、ここではカメラの俯角の違いによる人の見えの違いを表わす。人を正面から撮影した人画像と斜め上から撮影した人画像では、人の見えは大きく異なる。」

5. 人領域の隠れ

「人とカメラの間に物体が存在する場合、画像上の人領域が部分的に観測できなくなる。そのため、部分的に欠損した人画像から人を検出することになる。」

6. 複雑な背景

「背景画像が複雑なテクスチャを含むことがある。特に、人画像のように縦エッジが連続しているような背景画像は、部分的な領域のみを見た場合には人画像に似ているため、誤識別する問題がある。」

これらの問題に対し、Dalal らは、Histograms of Oriented Gradients(HOG) 特徴量と SVM による人検出を提案している [3]。HOG 特徴量は、輝度勾配情報に基づき形状の人らしさを抽出する特徴であり、体型、服装等の個体差による問題を大きく低減することに寄与している。山内らは単純な HOG だけでなく、特徴間の共起表現を導入した Boosting による Joint 特徴量を提案している [13]。高木らも同様に、Geometric Context による共起表現を導入した Boosting により高精度な人と車両の検出を実現している [14]。しかし、人のように変動を多く含む物体を検出する際、変化の大きい学習サンプルを同一のクラスとして扱い識別器を構築するのは困難となる場合がある。文献 [13][14] の手法は特徴量の表現力を向上させることで精度向上しているが、学習自体は Real AdaBoost など

¹原文では番号付けされていないが、本論文では参照のため番号を付けた。

の既存の2クラス学習を用いており、統計的学習法の改良については検討されていない。

2.4.1 物体の多様性に起因する問題

図 2.7 は人検出用データベースとして公開されている INRIA person dataset[3] に含まれている画像である。全ての人画像は人が中心となるよう画像上で正規化されているが、体幹の傾き、顕著な年齢差、複雑な姿勢等の変化を含んでいる。このような変化は、人検出を困難とする要因のうち1, 2, 3に該当する。これらの変化(変動)を持つ複雑な対象物体を一つの識別器で学習すると、全体に共通する情報が減少し、識別性能の低下を招くことになる。



図 2.7: 物体の見えの変化 (INRIA person dataset).

2.4.2 カメラの設置環境に起因する問題

カメラの設置環境が異なると、以下に挙げる変化が生じる。

- 外乱等による特徴量の変化

学習サンプルと設置環境における光量の差や背景の違いなどのノイズにより、色やエッジの出力される値が変化する。これにより、しきい値や区間関数を用いた弱識別器が誤識別を起こす。

- 視点の変化

物体の向きの違いと似ているが、ここではカメラの俯角の違いによる物体の見えの違いを表わす。図 2.8(a)(b) に示す例のように、物体を正面から撮影した物体画像と斜め上から撮影した物体画像では、物体の見えは大きく異なり、未検出が発生する。

これらは人検出を困難とする要因の 4, 5, 6 に該当し、設置環境に依存した問題といえる。



図 2.8: 視点の変化により発生する未検出。

特に姿勢の変化は深刻な問題で、俯角の大きさやカメラの高さなどによって縦横比やパーツのサイズなどは全く異なってしまふ。しかし、実店舗や屋外環境への検出システム設置を考えた場合には、設置やメンテナンスのコスト面、また利用者の心理面からカメラの台数は可能な限り少なくすることが望ましいため、カメラをずらりと並べるのではなく俯角をつけて一台ですできるだけ広範囲の視野をカバーすることが求められる。このような場合、システムの設置箇所毎、あるいはカメラ毎に異なる俯角に合わせてシステムを構築し直すことが必要となる。しかし、各俯角や背景毎に大量のサンプルを収集することは非常にコストが大きく、困難であるため、設置環境において簡易な調整により十分な性能を得られる識別器の調整手法が必要である。

また、まるで見えが異なる場合、調整ではなく識別器の再構築を求められる。しかし、統計的学習法で必要とされる学習サンプルは、図 2.9 に示す例のように、切り出した画像中の人の位置やサイズなどが標準化されている必要があり、技術と人的コストを要する。もし学習サンプル中の人画像位置にズレがあれば、人を中心に検出することができず、十分に統計的な学習が得られない。サイズについても同様で、様々なサイズで映り込んでいる場合、学習や検出を困難とする要因となる。これらの要件を満たす学習サンプルを手で収集するためには、多くの専門的な人間が多大な時間を費やす必要があり、非常にコストが高い作業であるといえる。

これらの問題を解決するアプローチとして、少数の学習サンプルからスケール変化や回転、ノイズの付加などの実環境で測定されうる変動を含むように変形させた学習サンプルを生成し、生成したサンプルを用いて識別器を学習する生成型学習法 [17] が提案されている。文献 [18] では、認識対象である車載カメラから見える路面上の標識に対して、光学ぼけや動きぼけ等を考慮したサンプルを生成して学習に利用した。また、文献 [19] では道路標識に対して、位置ずれや回転などの形状の



図 2.9: 標準的な学習サンプルの例 (INRIA Person Dataset) .

変化、背景などのテクスチャの変化、反射や影などの色の变化を考慮した生成モデルを用いて学習サンプルを生成した。しかし、これらの手法が認識対象とするのは比較的簡単な2次元パターンであり、人のような非剛体で複雑な形状や姿勢を持つ物体に対しては、同様のアプローチで学習サンプルを生成するのは困難である。これらのことから、特定環境における統計的学習法に用いるサンプル、特に非剛体の対象物体の学習サンプルを低コストで収集する手法が望まれている。さらに、近年では検出技術の向上に伴い特徴量の次元数は数千～数千万と膨大な量を用いることが多く、これら多くの次元から大量に得た多くの画像に共通する特徴量を探し学習を行うという、統計的学習による識別器構築自体に必要なコストも無視できなくなりつつある。これらの要因が物体検出の実用化を困難にしている。

2.5 まとめ

本章では物体検出技術において重要な要素である統計的学習法に基づく識別器について述べた後、それを実用化する際に困難となる要因について述べた。

統計的学習法は入力された学習サンプルについて正しい判別を行うよう学習し、非常に高い精度で正解する。しかし、入力された学習サンプルが多くのバリエーションを持つ複雑な問題であった場合、それを解決する識別器もまた複雑となり、精度は低下する。3章では、この問題を解決するために統計的学習法への分割統治戦略の導入を提案する。

物体検出を実用化する際に問題となるのは、設置環境への適応である。一般に学習サンプルと実際の設置環境で得られる画像は外乱などにより異なっており、その相違に起因して識別性能が低下する。そこで、4章では特徴量がどれだけ識別に貢献しているかに基づき、特徴量の削減や調整により簡易に設置環境へ適応する手法を提案する。

また、それでも適応不可能な相違がある場合は再学習が必要となるが、その際に学習サンプルを再収集する人的コストの大きさが問題である。基本的に、検出性能と良質なサンプルの量は不可分の関係にあり、設置環境に適したサンプルを大量に用意するための人的コストを低減する必要がある。

る。そこで、5章では学習サンプルの自動生成を用いて、再学習に使用するサンプルを大量かつ安価に用意し、設置環境に対して高精度な識別器を構築する手法を提案する。さらに、学習自体に必要なコストを削減するために、6章では転移学習を導入し、高速化に伴い発生する識別性能の低下を防ぐハイブリッド型転移学習による識別器を提案する。

第3章

分割統治戦略による識別器の高精度化

本章では、従来の物体検出に用いられる統計的学習法に分割統治戦略を導入することで高精度化するアプローチについて提案する。まず、3.1.1にて複雑な変化を持つ物体を分類する従来法について述べた後、3.1.2にて提案手法である分割統治戦略の導入について述べる。そして、3.2では分割過程における具体的なサンプル分割法として弱識別器の応答値に基づくサンプルの分割について述べた後、3.3で分割した学習サンプルに対する効率的な学習法として Divide-and-Conquer Boosting について述べる。

3.1 複雑な物体分類の問題と分割統治戦略の導入

3.1.1 複雑な変化を持つ物体の分類

2.4.1では、物体の複雑さにより識別性能が低下する問題を、図 2.7 を例として述べた。図 2.7 は人検出用データベースとして公開されている INRIA person dataset[3] に含まれている画像である。これらの画像は人が中心となるように配置されているが、体幹の傾き、顕著な年齢差、複雑な姿勢等変化が含まれている。これらの変化を吸収するように識別することは非常に困難な問題である。

従来、このような変化を伴う検出対象に対する解決策としては Viola らの顔検出 [15] に習うことが多い。この手法は Haar-like 特徴量と Boosting を用いて、顔の向きの変化に対する頑健性を得るために、学習画像を顔の傾きや方向毎に人手で分割して、サブカテゴリ毎に個別の識別器を構築するアプローチである。方向、傾き毎に構築した識別器のどれか一つでも反応すれば顔と判定する。また、Huang ら [20] は Viola らと同様に顔画像を右、正面、左の 3 クラスに手で分類し、各向きを Vector Boosting というマルチクラス識別器に入力することで同様に頑健性を担保している。しかし、これらの手法では、人手によりサブカテゴリに分類してラベル付けするためのコストが非常に大きい。さらに、構築される識別器の性能は、サブカテゴリの設定が的確であるかどうかにより大きく左右される。例えば図 2.7(h) はやや正面を向いていながらも全体としては左向きである。このように、どのラベルを付与すべきか判定が困難な例が存在する。また、図 2.7(f) は正面を向いているため頭、肩等においては図 2.7(a), (b) と同じ特徴を持つといえる。また、同時に、体幹が傾きにより脚部が左下方向に伸びている点で左向きの画像である図 2.7(c) と同じ特徴を持つといえる。識別においても、サブカテゴリ毎に分割して構築された識別器は「サブカテゴリ (右向き, 左向き等) らしさ」を学習する中で知識として「対象カテゴリ (顔, 人など) らしさ」について同様の情報をそれぞれ学習していると考えられる。これら重複した情報の学習は識別時において計算コストを引き上げる原因となる。

3.1.2 分割統治戦略の導入

3.1.1で述べたように、統計的学習法による2クラス分類では、同一検出対象内に多様性が存在すると問題が複雑化する。そこで本研究では、まず検出対象となる学習サンプルを自動的にサブカテゴリに分割する手法を提案する。次に、サブカテゴリ固有の情報は固有に、カテゴリ全体やカテゴリ群に共通する情報は共有して識別器を構築する手法を提案し、両手法の組み合わせによる分割統治アプローチを提案する。

本研究では、Joint Boosting をベースとして、弱識別器を共有しながら分割統治を行う Divide-and-Conquer Boosting (DaCBoost) を提案する。Joint Boosting はマルチクラスに対する学習であるため、サブカテゴリを持つ2クラス分類には不要な境界を学習することがある。そこで、提案する DaCBoost ではサブカテゴリ同士の分類を重視せず、各サブカテゴリと背景の境界のみを学習し、それらの結果を統合することで高精度な2クラス識別器を実現する。図3.1に提案手法の流れを示す。

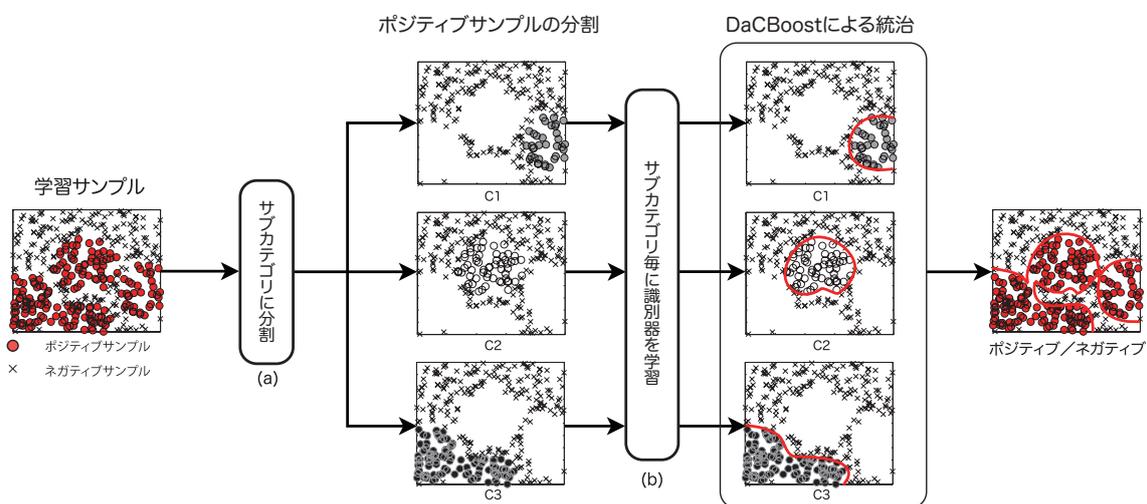


図 3.1: 分割統治的戦略による識別器の学習法 (概念図)。

提案手法の分割統治識別器を構築する手順は以下の2ステップとなる。

- サンプル分割

まず最初に、学習サンプルの全てのデータに対して1個の2クラス分類器を学習する。これは、通常の gentleboost による2クラス分類と同一の学習である。次に、gentleboost で得られた弱識別器の応答を用いて、図3.1(a)のようにサンプルを分割する。

- 共有学習

分割されたサンプル集合をサブカテゴリとし、図2.6で示した Joint Boosting と同様にサブカテゴリとその組み合わせをノードとして共有学習を行う。その結果、図3.1(b)のように各サブカテゴリに対し識別器が構築される。この時点で得られている2クラス識別器は全体の識別に

対して有効性を持つため、全サンプルを対象とするノード（例. 図 2.6 中の G^{123} ）として全サブカテゴリに利用し、学習を継続する。

以上の 2 つのステップにより、全体を大まかに識別する識別器、詳細な集合に対して識別する識別器をそれぞれ構築していく。こうして得られた各サブカテゴリへの識別器を統合することで、最終的な強識別器を構築する。

3.2 弱識別器の応答値に基づくサンプルの分割

従来、データを分割する場合には、特徴量空間における分布を利用して k-means などのクラスタリングを行う。しかし、特徴量空間は、識別対象のサンプルとそれ以外の背景サンプルの識別に寄与しない冗長な特徴次元を含んでいる。学習サンプルデータを分割するためには、識別対象のサンプルとそれ以外の背景サンプルを識別するために有効な特徴量のみを用いて分割することが必要である。村井らは、Boosting を用いた人領域のセグメンテーションにおいて、類似度の評価に Real AdaBoost による弱識別器の応答値を用いることで、識別器にマッチした類似度を算出している [21]。本手法は、Boosting の各弱識別器を識別に有効な 1 次元の特徴ベクトルと考えることができると述べている。弱識別器の応答値で構成される空間は、特徴空間とは異なり識別に有効でない特徴量は無視されるため、この空間においてクラスタリングすることで識別に有効な分割を得ることができる。そこで、提案手法では図 3.2 に示すように、まず学習サンプルに対して gentleboost で学習を行い、弱識別器群を得る。次に、弱識別器の応答値を属性としてサンプルのクラスタリングを行う。これにより、特徴空間での関係性だけでなく、学習を経た識別器の情報を用いることができるため、より識別に適したクラスタリングとなることが期待できる。

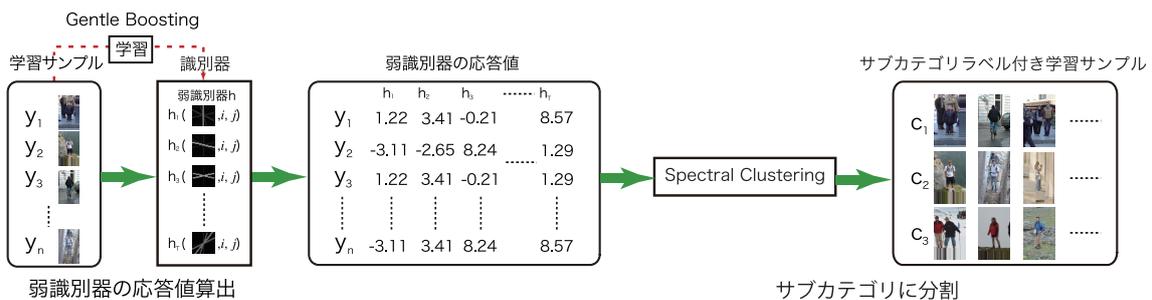


図 3.2: 弱識別器応答値を用いた学習サンプルのカテゴリ化。

3.2.1 Spectral Clustering

k-means に基づくクラスタリングは、以下の2つの前提に基づいている。1つは、各クラスタのデータ分布は超球形であること、2つ目は、各クラスタの半径は同じ長さであるということである。これらの性質に実データの分布が従うことは稀であり、問題となり得る。そこで、本研究では各サンプルの相関行列から次元圧縮を行い、圧縮された空間上でクラスタリングを行う Spectral Clustering[22] を用いる。Spectral Clustering は k-means に比べより多くの計算量を必要とするが、データの連結性などに着目したクラスタリング結果を得ることができる。図 3.3 に Spectral Clustering の流れを示す。

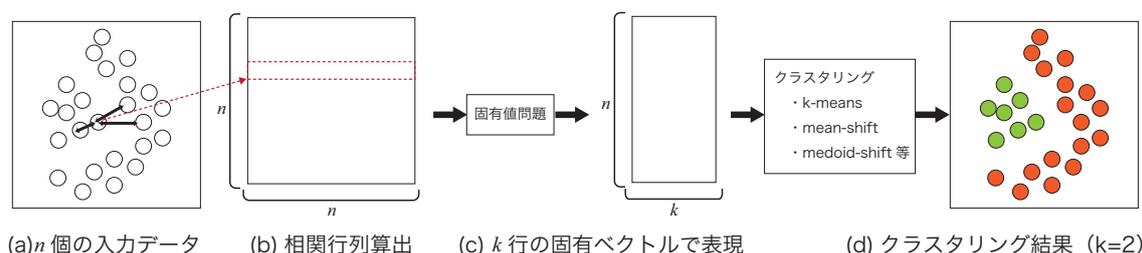


図 3.3: Spectral Clustering の流れ.

Spectral Clustering は、サンプル間の相関行列（類似度行列）を作成し、相関行列を固有値問題に帰結し解くことで次元圧縮効果を得ることができる。まず、図 3.3(a) のようなクラスタリング対象となるサンプルから、図 3.3(b) のように相関行列 A を求める。相関行列 A は式 (3.1) で表される。

$$A_{ij} = distance(i, j) \quad (3.1)$$

ここで、 i と j はクラスタリング対象となるサンプルである。距離関数 $distance()$ としては L2 ノルムや内積空間への写像などが用いられる。なお、本研究では L2 ノルムを用いた。この相関行列を、固有値問題に帰結して解く。まず、求められた A を用いて、距離行列 $L = D^{-1/2}AD^{-1/2}$ を求める。ここで、 D は $D_{ii} = \sum_{j=1}^N A_{ij}$ で求められる対角行列である。

次に、図 3.3(c) のように固有値 $E = [e_1, \dots, e_k]$ を計算する。 e_k は L の最小固有ベクトルから k 番目の固有ベクトルを抽出したものである。そして、 E を基に k-means クラスタリングすることで、図 3.3(d) のように任意のクラスタ数に分割することが出来る。

3.2.2 ノンパラメトリックなクラスタリングへの対応

提案手法では、クラスタ数をパラメータとして指定しないノンパラメトリックなクラスタリングを行う必要がある。そこで、Spectral Clustering の最後の処理に medoid-shift[23] を用いる。medoid-shift

は、同じくノンパラメトリックなクラスタリング手法である mean-shift[24] と類似しているが、図 3.4 に示すように、mean-shift がデータのセントロイド（平均値）を仮定して摂動し、クラスタリングするのに対し、入力されたデータの中央値を各クラスの代表として摂動し、各データ間の相互距離関係によりクラスタリングを行う点が異なる。medoid-shift は、データが存在しない特徴空間の点をクラスを中心とすることはない。そのため、mean-shift において困難となるファン状のような複雑なデータ分布に対して、比較的正しくクラスタリングできる手法である。

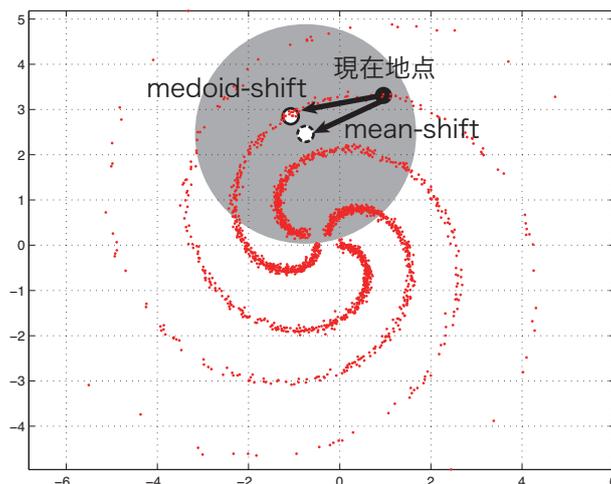


図 3.4: medoid-shift と mean-shift の違い。

3.2.3 サンプル分割結果

ここでは、人手による 3 クラス（正面／背面，左，右）へのラベリングと提案手法でのラベリングを比較する。データベースには、INRIA Person Dataset[3] を用いる。以下に人手によるサンプル分割と提案手法におけるサンプル分割について述べる。

人手によるラベリング

人手によるラベリングは、人の知識によりサンプルを分類するサブカテゴリ分類法である。しかし、特に多量のサンプルに対しては時間やコストの面に問題がある。また、サンプルにどのラベルを付与するかは作業者の主観に基づくため、作業者によりばらつきが出るという問題もある。

本実験では、INRIA Person Dataset の人画像 2,416 枚に対して、3つのサブカテゴリ（正面／背面，左，右）にラベリングを行った。各カテゴリの枚数は、正面／背面 1,308 枚，左向き 554 枚，右向き 554 枚である。各カテゴリに分類された画像から求めた平均勾配画像と各カテゴリのサンプル例を図 3.5 に示す。図 3.5 より、それぞれ方向別に分類されていることがわかる。

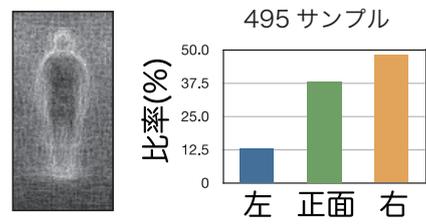
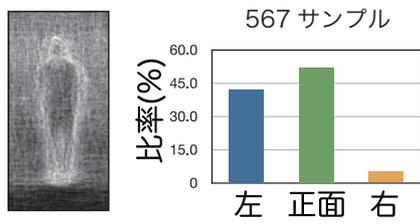


図 3.5: 人手によるサンプル分割.

弱識別器の応答値によるクラスタリング (提案手法)

提案手法によるクラスタリング結果の平均匀配画像とサンプル例を図 3.6 に示す. 弱識別器の応答値をクラスタリングした結果, 人画像は 4 つのカテゴリ C1~C4 に自動的に分割された. 図 3.6 のグラフはそれぞれ C1~C4 各カテゴリに対して, 方向別に人手でラベル付けしたサンプルがどの程度含まれていたかを比率で示している. 提案手法では, 弱識別器の応答値を入力ベクトルとして Spectral Clustering を行う際, 分割するカテゴリ数を自動的に決定する Medoide-shift を組み合わせている. 図 3.6 の平均匀配画像群に着目すると, カテゴリ C1 は頭部形状等にやや左に傾いた傾向が, C2 はやや右に傾いた傾向がある. また, C3 は最も多くのサンプルを含むため, 偏りの無い滑らかな平均画像が得られている. 一方, C4 は最もサンプル数が少ないため, 粗い平均画像である. C4 の特徴としては, 特に足下の部分にエッジが大きく出ている点と, 両足がやや開いている傾向がある.

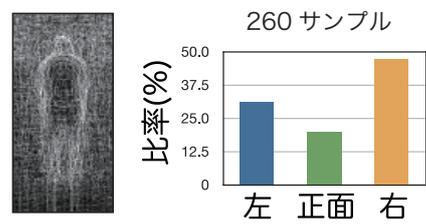
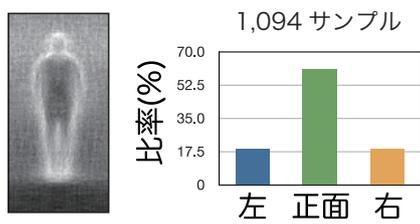
次に, 各カテゴリに含まれるサンプルの方向別の比率と, 各カテゴリのサンプル例に注目する. 図 3.6(a)(b) より, C1, C2 の比率はそれぞれ正面/背面サンプルと左, 右方向どちらかを中心としており, 人の傾きを中心としたカテゴリ分けを行っていることがわかる. また, 図 3.6(c) より, C3 は正面/背面サンプルを多く含み, 平均匀配画像からも明確に正面/背面を捉えるカテゴリであることがわかる. 図 3.6(d) より C4 はサンプル数自体が非常に少なく, 方向としても事前の知見である人の方向とは合致しない傾向を持つ. そこでサンプルを調査したところ, このカテゴリにクラスタリングされたサンプルの中には, 自転車に乗った人が含まれており, 脚部に複雑なエッジ構造を持つことが判明した. これは, 本実験における事前の知見には無かった「自転車に乗っている人」を中心としたカテゴリを自動的に発見したといえる.



(a)C1



(b)C2



(c)C3



(d)C4

図 3.6: 提案手法によるサンプル分割.

3.3 Divide and Conquer Boostingによるサブカテゴリに対する効率的な学習

3.2でサブカテゴリに分割されたデータに対し、各サブカテゴリ毎に識別器を構築する。本手法では、Joint Boostingをベースとしてサブカテゴリを含む2クラス問題向けに適応を行い、各サブカテゴリ間で共通して有効な弱識別器を共有して学習する..

3.3.1 Joint Boosting アルゴリズムの分割統治戦略への適用

Joint Boostingは各学習ラウンドにおいて全ての対象クラスの組み合わせを探索し、最良の識別器を学習することにより、全クラスで合計した特徴数（弱識別器数）を削減することが可能である。本研究では、全サブカテゴリを対象としたものが通常の2クラス問題を解く gentleboostと同様であることに着目し、段階的にサンプルを分割しながら学習を行うことでJoint Boostingと類似した識別器の構造を実現する。

まず、各サブカテゴリにおける弱識別器の評価を考える。Joint Boostingでは、カテゴリ間のエラーについても背景カテゴリとのエラー同様に評価するため、サブカテゴリ問題にはあまり寄与しない識別関数についても学習する。そこで、各サブカテゴリ毎に存在するサブカテゴリラベル $z = \{1, -1\}$ を、サブカテゴリに属するポジティブサンプル、属しないポジティブサンプル、ネガティブサンプルとして $z = \{1, 0, -1\}$ に拡張する。

次に、エラー評価を各ノードについて独立に求める。ここでは、他のノードに属するデータを自己と捉えてもエラーと判定せず、同様にそれらを他クラスと捉えても正解と判定しない。

次に、各ノードにおいて下記の評価関数を用いて弱識別器の評価を行う。

$$J_{wsc} = \sum_{c=1}^C \sum_{i=1}^N z_i^2 w_i^c (z_i^c - h_m(\mathbf{x}_i, c))^2 \quad (3.2)$$

上式により、対象ノード内に属しないデータは計算に関与しない。ここで、 a, b, k^c のパラメータはそれぞれJoint Boostingと同様の形で以下の式で計算される。

$$b = \frac{\sum_{c \in S(n)} \sum_i w_i^c z_i^c \delta(\mathbf{x}_i^f \leq \theta)}{\sum_{c \in S(n)} \sum_i z_i^2 w_i^c \delta(\mathbf{x}_i^f \leq \theta)} \quad (3.3)$$

$$a + b = \frac{\sum_{c \in S(n)} \sum_i w_i^c z_i^c \delta(\mathbf{x}_i^f > \theta)}{\sum_{c \in S(n)} \sum_i z_i^2 w_i^c \delta(\mathbf{x}_i^f > \theta)} \quad (3.4)$$

$$k^c = \frac{\sum_i w_i^c z_i}{\sum_i z_i^2 w_i^c} \quad c \notin S(n) \quad (3.5)$$

これらの操作は、ポジティブサンプルを各サブカテゴリで独立して保持することと同義である。これにより、各サブカテゴリ間の分離に着目した学習が行われなくなる。提案する DaCBoostの学習アルゴリズムを**アルゴリズム 3.1**に示す。

アルゴリズム 3.1: DaCBoost の学習アルゴリズム.

1. 入力: I 個の学習サンプル $\{\mathbf{x}_1, z_1\} \dots \{\mathbf{x}_I, z_I\}$ を用意する
 \mathbf{x} は学習サンプル, $z_i \in \{-1, 1, \dots, C\}$ は学習サンプルが属するクラスラベルを表す

2. 初期化: 学習サンプルの重み D を初期化

$$w_1^c(i) = 1/I \quad (3.6)$$

3. 学習:

· C 個のクラスの全ての組み合わせを示す N 個の要素を持つ集合をノード $S(n)$ とする

For $m = 1, \dots, M$ //学習ラウンド

For $n = 1, \dots, N$ //すべてのノード

· 評価値 J_{wse} が最小の弱識別器 $h_m(\mathbf{x}, c)$ を選択

$$h_m(\mathbf{x}, c) = \begin{cases} a\delta(\mathbf{x}^f > \theta) + b & \text{if } c \in S(n) \\ k^c & \text{otherwise} \end{cases} \quad (3.7)$$

· 評価値 J_{wse} の算出

$$J_{wse} = \sum_{c=1}^C \sum_{i=1}^N z_i^2 w_i^c (z_i^c - h_m(\mathbf{x}_i, c))^2 \quad (3.8)$$

· $h_m(\mathbf{x}, c)$ のパラメータ a, b, k^c の算出

$$b = \frac{\sum_{c \in S(n)} \sum_i w_i^c z_i^c \delta(\mathbf{x}_i^f \leq \theta)}{\sum_{c \in S(n)} \sum_i z_i^2 w_i^c \delta(\mathbf{x}_i^f \leq \theta)} \quad (3.9)$$

$$a + b = \frac{\sum_{c \in S(n)} \sum_i w_i^c z_i^c \delta(\mathbf{x}_i^f > \theta)}{\sum_{c \in S(n)} \sum_i z_i^2 w_i^c \delta(\mathbf{x}_i^f > \theta)} \quad (3.10)$$

$$k^c = \frac{\sum_i w_i^c z_i}{\sum_i z_i^2 w_i^c} \quad c \notin S(n) \quad (3.11)$$

End for

· 学習サンプルの重み w を更新

$$w_{t+1}^c(i) = w_t^c(i) \exp(z_i^c - h_m(\mathbf{x}_i, c))^2 \quad (3.12)$$

End for

4. 出力: 強識別器

$$H(\mathbf{x}, c) = \sum_{m=1}^M h_m(\mathbf{x}, c) \quad (3.13)$$

3.3.2 Joint Boosting との相違点

Joint Boosting では、あるノードについて識別器を評価する際に他のサブカテゴリを負例として用いる。このアルゴリズムにより決定される境界は、各サブカテゴリ間を分離するよう重点的に学習される。しかし、サブカテゴリを含む2クラス問題では、各サブカテゴリは本来同一クラスであるため、これらを分離する境界を学習するのは無駄となる。提案する DaCBoost では、あるノードについて計算するとき、そのノードに含まれないサブカテゴリについては考慮しないため無駄な境界を学習することがない。これにより、効率よくサブカテゴリ群とネガティブサンプル間の境界を学習可能である。図 3.7 に示すように、サブカテゴリ間での識別を考慮しないことで DaCBoost はより多くのポジティブ/ネガティブの分離が可能である。

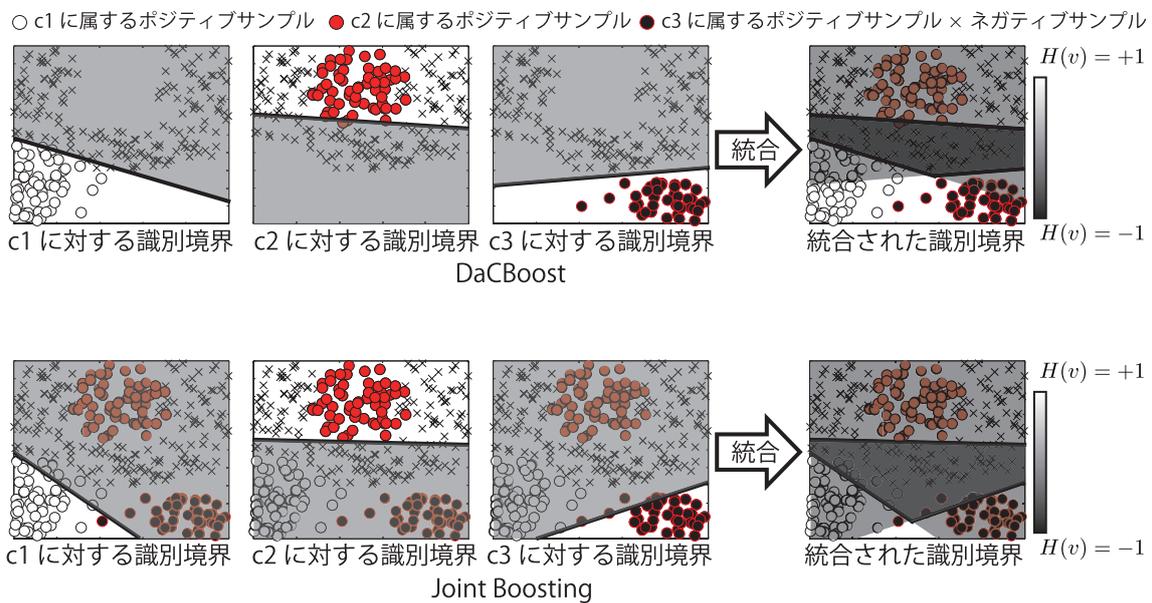


図 3.7: Joint Boosting と DaCBoost による境界の差。

3.4 評価実験

提案手法の有効性を示すために、以下の2種類の実験を行う。1つは、サンプル分割を行う際の分割手法についての比較である。本実験を通じて、弱識別器の応答値を用いてクラスタリングを行うことが提案する分割統治アプローチによる識別に有効であることを示す。2つ目は、2クラス問題に対する識別性能の比較である。ここでは、単体の2クラス識別器と、Viola らのように識別器を複数用いた際の識別性能との比較を行う。この実験を通じて、提案手法による分割統治アプローチが問題の識別に有益であることを示す。

3.4.1 実験概要

本実験では、人、車両それぞれ1種のデータセットを用いて識別精度を評価する。各データセット、特徴量、精度の評価基準について述べる。まず、実験に使用する各データセットについて説明する。

INRIA Person Dataset[3]

人検出問題として、INRIA から提供されているデータベースを用いる。学習用人画像 2,416 枚と学習用背景画像 1,218 枚、評価用人画像 1,135 枚と評価用背景画像 453 枚が用意されている。

UIUC Image Database for Car Detection[25]

車両検出問題として、UIUC の車両データベースを用いる。道路画像より切り出された車両画像 550 枚と同じく切り出された背景画像 550 枚を学習用に、評価用には道路画像 170 枚を持つデータセットである。

本実験では、人と車両共に特徴量として Histograms of Oriented Gradients (HOG) [3] を用いる。HOG 特徴量は、局所領域における輝度の勾配方向をヒストグラム化した特徴量として記述する。そのため、大まかな物体形状を表現することが可能であり、人検出や車検出等の物体検出に有効であることが報告されている [13][14]。

3.4.2 弱識別器の応答値によるサンプル分割の有効性

提案手法によるサンプル分割は、弱識別器の応答を用いることでより識別に有効な分離を実現する。その有効性を確認するために、各手法で分割したサンプルを用いて識別器を構築した際の識別性能を比較する。各比較手法の詳細について以下に示す。

(a) 人手によるラベリング

サンプルを分割する際の比較手法として、人手によるラベリングを行う。本実験では、INRIA Person Dataset 内の人画像に対して（正面／背面，左，右）へのラベリングを行った。3.1.3 で述べた通り、その平均勾配画像は図 3.5 のようにそれぞれ方向別に分類されていることがわかる。

(b) ランダムラベリング

サンプルに対するラベル付けをランダムで行う。カテゴリ数のみ手動による方向別ラベリングと同じ 3 として指定した。

(c) 特徴量によるクラスタリング

一般に、画像データの自動クラスタリングは画像より得られた特徴量を用いて行う。しかし、特徴量の距離を使ったカテゴリ分けは識別において有利に働くか否かは不明である。ここでは HOG 特徴量全ベクトルを用いたクラスタリングでラベリングする。

(d) 提案手法

弱識別器の応答値によるクラスタリングであり、3.1.3 で述べた通り結果は図 3.6 である。人画像を（左向き/正面，右向き/正面，正面，自転車等）として4カテゴリにラベリングを行っている。

■ 実験結果

本実験では、INRIA Person Dataset を用いる。INRIA Person Dataset を用いて、Equal-Precision-Recall Rate (EPR: 再現率と適合率を等しく設定) 時の検出率について比較を行った結果を表 3.4.2 に示す。表 3.4.2 より、提案手法である弱識別器の応答値による分割手法が最も検出性能が高いことがわかる。また、手動でラベリングしたものと特徴量によるクラスタリングはそれぞれ同程度である。これは、方向別などの人間の知識によるラベリングや単純な特徴量によるラベリングでは、必ずしも識別器に対して最適ではないことを示している。提案手法は人の向きという事前知識を用いてはいないが、右向き、左向きなどを考慮しつつ更に識別に有効なカテゴリとして自転車に搭乗した人というカテゴリを生成した。更に、人手では正面ラベルを付加されたサンプルの内相当数に、判別の観点からは右向きや左向きに非常に近いものが存在しているといえる。

表 3.1: サンプル分割手法の比較。

サンプル分割手法	検出率 (%)
(a) 人手によるラベリング	81.6
(b) ランダムラベリング	81.1
(c) 特徴量によるクラスタリング	81.5
(d) 提案手法	85.5

3.4.3 DaCBoost によるサブカテゴリ学習の有効性

提案手法の有効性を示すため、2クラス分類器である gentleboost、それをカテゴリ毎に複数構築した分割統治による識別器との比較を行う。取り扱う問題は2クラス識別として、サブカテゴリ相互の誤りは誤検出としない。本実験では、各手法の弱識別器の総数を500個と固定とする。各比較手法の詳細について以下に示す。

gentleboost

gentleboost は2クラス識別器であり、サンプル分割を行わない際の識別性能を示す。

複数の識別器

複数の識別器は One vs Other 戦略によりクラス毎に構築した gentleboost 識別器群であり、分割したサンプルに対して弱識別器の共有を行わず学習した際の識別性能を示す。

提案手法 (DaCBoost)

提案手法は DaCBoost により構築した識別器であり, 分割したサンプルに対して共通した特徴を共有しながら効率的に学習した際の識別性能を示す.

実験結果の比較には Detection Error Tradeoff(DET) カーブを用いる. DET カーブは横軸に False Positive Per Window(FPPW), 縦軸に miss rate を表わし, 左下の原点に近いほど検出性能が高いことを示す. FPPW は, 全背景画像に対し人のサイズに合わせたウインドウをラスタスキャンし, 人と誤検出したウインドウの割合を表す. miss rate は全ての切り出した人画像に対し検出を行い, 未検出であった割合を表す. これらはトレードオフの関係にあり, 識別器の出力 $H(\boldsymbol{x})$ を人と判定するしきい値によって上下する. FPPW, miss rate はそれぞれ以下の式で計算される.

$$FPPW = \frac{\text{False positive の数}}{\text{全てのテスト用背景画像}} \quad (3.14)$$

$$\text{miss rate} = \frac{\text{False negative の数}}{\text{全てのテスト用人画像}} \quad (3.15)$$

ここで, False positive は誤って人として検出された背景画像, False negative は人として検出されなかった人画像である.

■ INRIA Person Dataset

図 3.8 に, INRIA Person Dataset における比較結果を示す. 全体に, 提案手法が gentleboost に比べ低い miss rate を達成しており, FPPW= 0.01 時点で 0.08 向上 (検出率 8% 向上に相当) し, 最大で 0.1 以上の性能向上を見せた. 提案手法によるサンプルの分割が有効に働いたことがわかる. また, 同数の弱識別器でありながら, サブカテゴリ毎に独立して識別器を構築するよりも提案手法は平均して 0.05 miss rate が低く, DaCBoost で特徴を共有しながら, 分割されたデータを学習することによって効率的に学習が行われたといえる.

■ UIUC Image Database for Car Detection

図 3.9 に, UIUC Image Database for Car Detection での車両検出における比較結果を示す. 全体に, 提案手法が gentleboost を miss rate において下回っており, FPPW= 0.001 時点で 0.095 向上 (検出率 9.5% 向上に相当) し, 最大で 0.1 以上の向上を見せた. 車両のように方向などによって大きくエッジの構成が異なる対象に対して, 提案手法によるサンプルの分割が有効に働いたといえる. また, 本データセットにおいても提案手法の識別性能は, 各カテゴリ毎に識別器を構築する手法に比べて高い.

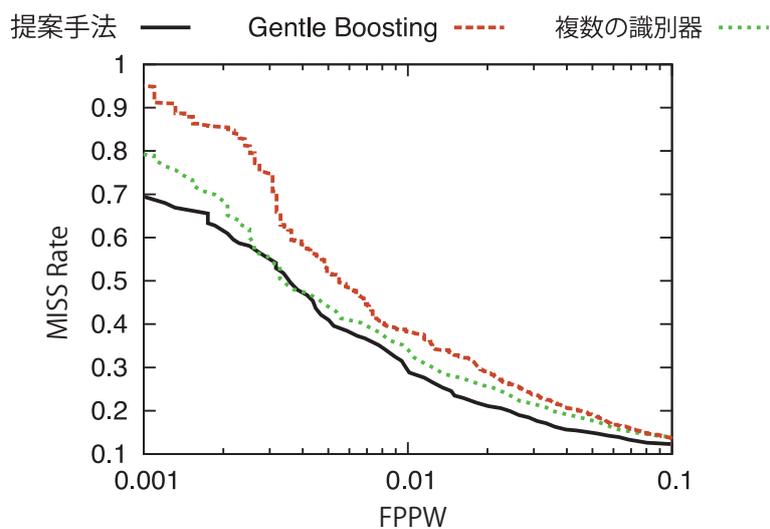


図 3.8: INRIA Person Dataset での性能比較.

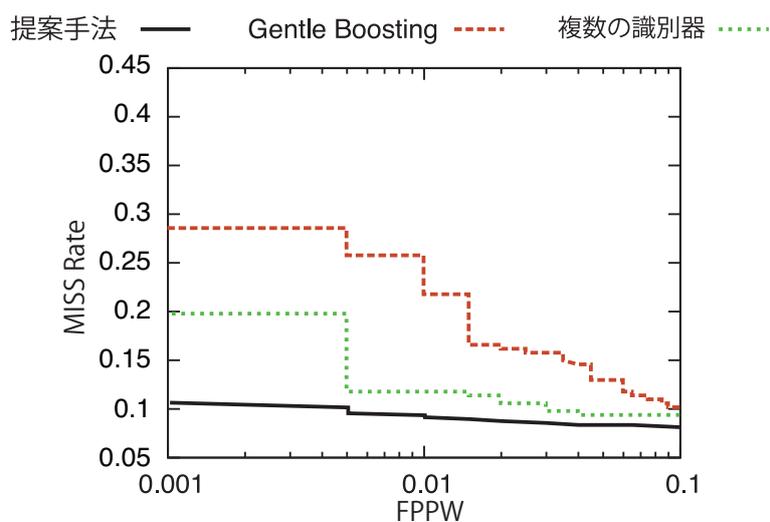


図 3.9: UIUC Image Database for Car Detection での性能比較.

3.5 まとめ

本章では、人や顔、車両などの向き、姿勢により様々な変動を含む物体を検出するため、サンプルデータ群の偏りをサブカテゴリと見なして分割し、それらに対して個別に問題を解き、その結果を統合する分割統治アプローチによる識別器の構築法を提案した。

本提案の貢献は、1つ目に偏りを持つサンプルを分割する教師無し的手法として弱識別器の応答

によるクラスタリング, 2つ目は DaCBoost により 2 クラス分類問題に対する分割統治の実現の 2 点である。人, 車両検出問題における評価実験の結果, 学習サンプルを分割しない従来の gentleboost と比較した結果, INRIA Person Dataset では約 8% (FPPW0.01 時), UIUC Image Database for Car Detection においては約 9.5% (FPPW0.001 時) 性能を向上させることができた。また, サンプル分割手法についても従来の特徴量や手動によるラベリングと比較して, 弱識別器の応答を用いる提案手法は 3%以上高精度であった。

本提案は 2 クラス問題における Boosting ベースの手法に容易に展開可能であり, 弱識別器数の低減, 識別性能の向上が期待できる。

第4章

特徴貢献度に基づく識別器の調整

3章では、識別問題を分割統治することで高精度な識別器を構築する手法について述べた。しかし、実際の設置環境において学習サンプルの収集環境から大きく変化すると検出性能は低下するという問題がある。そこで、本章ではまず4.1において環境変動に頑健な特徴の選択と設置環境での調整の必要性について述べた後、特徴選択の従来法について述べる。そして、4.2においてBoosting識別器に基づく特徴量の貢献度を提案し、貢献度に基づいた特徴選択や識別器の感度調整に用いる手法について述べる。

4.1 識別器の調整と特徴量選択

統計的学習法を用いた物体検出システムを実際に運用する際、カメラの設置環境と学習時の環境が異なると性能が低下する問題がある。特に、屋外環境下で撮影された人や自動車等の認識では、天候等の照明変動に影響を受けない特徴量を選択する必要がある。環境変動に対して頑健性を向上する手法として、識別に大きく寄与する特徴量のみを選択する特徴選択が挙げられる。識別に大きく寄与しない特徴量は学習サンプルの時点で対象の判別が困難な特徴量であり、このような特徴量は外乱によって容易に変動すると考えられる。そのため、識別器の構築段階で入力から排除してしまうことでシステムの頑健性を確保する。特徴選択は、任意の特徴量を棄却して識別器を構築し、その識別性能により決定するのが一般的である。特徴量の有効性評価は特徴量を排除または挿入した際の識別性能を基準とする。以下に、オーソドックスな二種類の特徴選択法について詳述し、その問題点について述べる。

4.1.1 Sequential Forward Selection

Whitney らにより提案された Sequential Forward Selection (SFS) [26] は forward 型と呼ばれ、特徴数が 0 の状態からスタートし特徴を増やしていく方法である。全ての特徴についてその特徴を追加した際の評価値を求め、最も評価の高い特徴を特徴集合に追加する。これを繰り返して特徴を順次追加し、評価値が低下する、または変化が無くなれば終了する。この際的评价値には、一般に識別率が用いられる。

4.1.2 Sequential Backward Selection

Marill らにより提案された Sequential Backward Selection (SBS) [27] は全ての特徴を持つ状態からスタートし特徴を削除していく手法であり、backward 型と呼ばれる。SFS 法とは逆に、SBS 法は不要と評価された特徴量を 1 つずつ特徴集合から削減することになる。全ての特徴についてそれを削減した際的评价値を求め、その結果を用いて特徴を順次、特徴集合から削除する。評価値が低下、または変化が無くなれば終了する。

4.1.3 従来法の問題点

SFS 法や SBS 法の特徴選択アルゴリズムは識別性能を評価とするため、追加または排除される特徴の組み合わせの数だけ識別器を構築する必要があり、探索の計算コストが大きい。具体的には n 次元の特徴量から m 個の特徴量を削減するために、その全ての組み合わせ ${}_nC_{n-m}$ 回の試行を必要とする。識別器構築に伴い必要な特徴探索時間は特徴次元数に比例する。SBS 法は特徴量をすべてサブセットに持つ状態から開始するため、特徴探索時間の増大が特に顕著である。また、SFS 法は順次特徴を挿入していくため SBS 法に比べ高速ではあるが、同時に組み合わせることで有効となる

特徴を考慮することができないため、精度が低いという問題がある。これらの理由により、特徴次元が多く、識別器の構築に時間を必要とする物体検出において、高速性と高精度を併せ持つ特徴選択アルゴリズムが望まれている。

4.2 Boosting による特徴量の貢献度評価

本節では、Boosting で構築された識別器を解析することにより簡易に特徴の有効性を評価する貢献度について述べる。また、その応用として、貢献度を用いた特徴選択について述べる。

4.2.1 2 クラス問題に対する貢献度算出

AdaBoost に代表される Boosting 手法は、2.2.1 で示したように、それぞれの弱識別器が最終的な判別に及ぼす度合いを値として持つ重み付き多数決である。本研究では弱識別器を 1 特徴次元に基づく判別関数として用いている。そこで、各弱識別器に用いられる特徴量と出力値の大きさから、各特徴量が最終判定にどの程度影響力を持つかを特徴量の貢献度として算出する。貢献度算出の流れを 4.1 に示す。まず、全ての特徴量と学習サンプルを入力として Boosting 識別器を学習する。図 4.1 の例では AdaBoost を用いている。次に、構築した強識別器から、各弱識別器の応答値を求めて、弱識別器に用いられている特徴量毎に総和し、各特徴量の貢献度とする。弱識別器から特徴量の貢献度へは、式 (4.1)、式 (4.2) により換算する。

$$CR_f = \sum_{t=1}^T h_t^{response} \cdot \delta[\mathcal{F}(h_t), f] \quad (4.1)$$

$$h_t^{response} = \sum_{r=1}^R h_{t,r}^2 \quad (4.2)$$

ここで、 δ は Kronecker のデルタ関数、 R は弱識別器の区間数であり、しきい値関数を用いているなら 2、確率密度関数を用いているならその分解能 k と同数となる。 $h_{t,r}$ は弱識別器 h_t の区間 r における応答値であり、 $\mathcal{F}(h_t)$ は、 h_t に採用された特徴量を求める関数である。つまり、提案手法による貢献度 CR_f は弱識別器の各区間の応答の二乗和を用いて算出する。二乗するのは応答値の正負関係なく大きさのみを見るためである。

4.2.2 マルチクラス問題における貢献度算出

マルチクラス問題では識別器に、Joint Boosting を用いる。Joint Boosting による学習では、2 章で述べたように、各クラスに対する弱識別器とその応答値（判別への重み）が共有されて与えられ

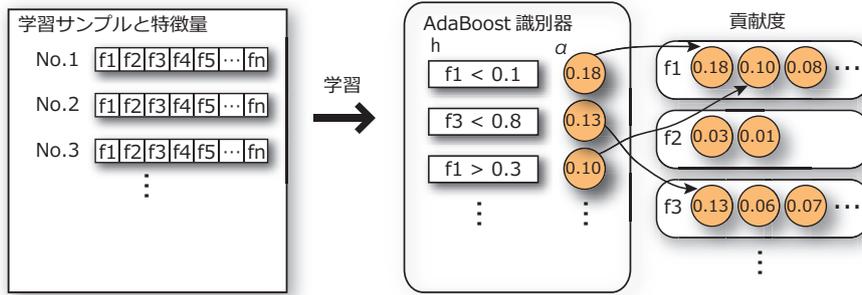


図 4.1: Boosting に基づく特徴量の貢献度評価.

る. そのため, クラス c に対する特徴量 f の持つ貢献度は式 (4.3) として算出する.

$$CR_{c,f} = \sum_{m=1}^M h_m^{response}(\mathbf{x}, c) \cdot \delta_K[F(h_m(\mathbf{x}, c)), f] \quad (4.3)$$

さらに, Joint Boosting が各クラス, クラス集合というノードについて弱識別器集合 G を構成しているため, 各ノード $S(n)$ に対する貢献度についても同様に, 弱識別器集合 $G^{S(n)}$ 内の弱識別器より式 (4.4) で求めることができる.

$$CR_{S(n),f} = \sum_{m=1, h_m(\mathbf{x}) \in G^{S(n)}}^M h_m^{response}(\mathbf{x}) \quad (4.4)$$

式 (4.4) は, 各弱識別器集合 $G(S(n))$ 毎に貢献度を収集したものであり, それぞれ複数クラスまたは個別クラスに対して特徴量がどれだけ有効であることを示す.

これにより, 複数のクラスで識別性能が低い際などにどの特徴量が問題であるかを客観的に評価できる.

4.2.3 貢献度に基づく特徴選択 (SBS-CRbB)

貢献度に基づく特徴選択は, 貢献度評価の値が最も低い特徴量を段階的に取り除くことで特徴を選択する. SBS 法と同じ backward 型であるが, SBS 法が 1 つの特徴量を削減するために削減候補と同数の識別器を構築する必要があるのに対し, 提案手法は 1 つの特徴量を削減するために 1 度の識別器の構築しか必要としない. これにより, 高速な特徴選択が可能となる. 貢献度による特徴選択アルゴリズムを **アルゴリズム 4.1** に示す. まず, n 次元の入力特徴セットと学習サンプルを入力する. 次に, 全特徴量を用いて識別器を構築し, 貢献度 CR_n を算出し, CR が最小であった特徴量を排除する.

アルゴリズム 4.1: 貢献度に基づく特徴選択アルゴリズム.

1. 入力: n 次元の特徴量セット $F = \{f_1, f_2, \dots, f_N\}$, I 個の識別器学習用データセット (\mathbf{x}_i, y_i) を用意

2. 初期化: 特徴集合 S に特徴量セット F の全ての特徴量を追加

・スコア $MaxScore=0$, $BestScore=0$

3. 特徴選択:

For Until $MaxScore < BestScore$ //スコア (識別性能) が低下するまで削除

・スコアの更新

$$BestScore = MaxScore$$

・ S に含まれる特徴量を用いて Boosting 識別器 $H(x)$ を構築

For $n = 1, \dots, N$ // 特徴量の貢献度, CR_1, CR_2, \dots, CR_N を算出

$$CR_n = \sum_{t=1}^T h_t^{response} \cdot \delta[\mathcal{F}(h_t), f_n]$$
$$h_t^{response} = \sum_{r=1}^R h_{t,r}^2$$

End for

・貢献度が最小の特徴 f_n を集合 S から削除

$$S = Exclusion(S, f_n)$$

・貢献度が最小の特徴量を削除した際の識別性能の算出

$$MaxScore = Score(S)$$

End for

4. 出力: 選択された特徴セット S

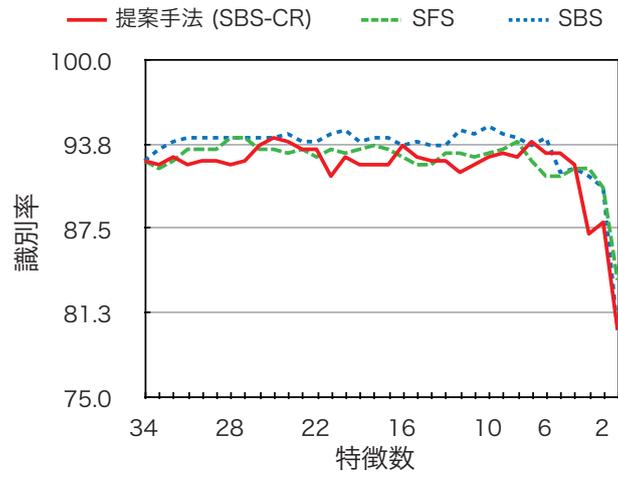
4.3 特徴貢献度に基づく特徴選択実験

貢献度に基づく特徴選択実験を行う。本実験では UCI-Machine Learning Repository[28] より, “ionosphere (34次元, 351 サンプル)”, “Heart (44次元, 80 サンプル)”, “Sonar (60次元, 208

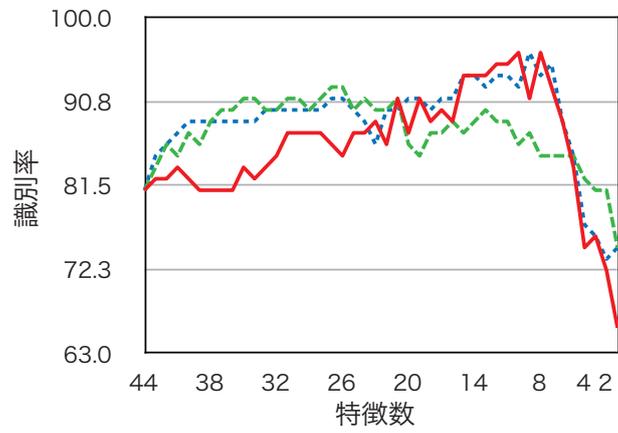
サンプル) ”の三種類のデータセットを実験に用いる。学習サンプルはそれぞれ総サンプル数の半数を無作為に選択して使用し、残りの半数を評価用サンプルとして用いる。これらのデータセットに対し、貢献度に基づく特徴選択 (SBS-CRbB) と SFS 法, SBS 法による特徴選択を行い, その際の識別性能と計算時間を比較する。

4.3.1 特徴選択結果

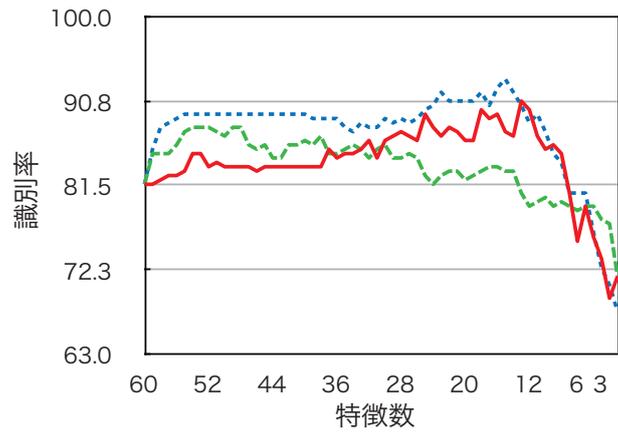
各手法の特徴選択結果を図 4.2 に示す。まず, すべてのデータセットにおいて, 提案手法は全特徴を使用したグラフ左端に比べ, 特徴選択により良好な識別性能を得ることができた。SBS 法が最も高精度を獲得しており, 提案手法は SBS 法の結果と同じ次元数にピークを持ち, SBS 法と傾向が似ていることがわかる。一方, SFS 法は特徴次元を少数にした際, 識別性能が大幅に低下している。



(a) Ionosphere



(b) Heart



(c) Sonar

図 4.2: 特徴削減に対する識別性能の比較.

表 4.1 に、各特徴選択法による最も高い識別率の値と、選択された特徴次元数を示す。提案手法は SBS 法とほぼ同数の次元で同程度の識別率を獲得していることがわかる。

表 4.1: 最大識別率 (RR) とその次元数 (dim) の比較。

	IonoSphere		Heart		Sonar	
	dim	best RR	dim	best RR	dim	best RR
SBS-CRbB	7	0.94	8	0.96	13	0.91
SFS	28	0.94	26	0.92	48	0.88
SBS	10	0.95	9	0.96	15	0.93

4.3.2 特徴選択に要する計算時間の比較

次に、各手法による特徴選択に必要な計算時間を図 4.3 に示す。横軸は削減対象の特徴数であり、縦軸はそのために必要な計算時間である。計算時間の計測は Intel Xeon processor 3.0GHz を搭載したコンピュータで行った。図より、計算時間において SBS 法は特徴数に比例して指数関数的に増大しており、SFS 法は緩やかではあるがこれも指数関数的に増大することがわかる。これは、特徴量セット中の 1 つの特徴量評価のために、1 つの識別器構築を行う必要があることに起因している。一方、提案手法は特徴量セットに対して 1 つの識別器を構築することで、特徴量セット中のすべての特徴量の貢献度を算出して特徴削減を行う。そのため、特徴数が増加しても計算時間の増大を抑えることができる。

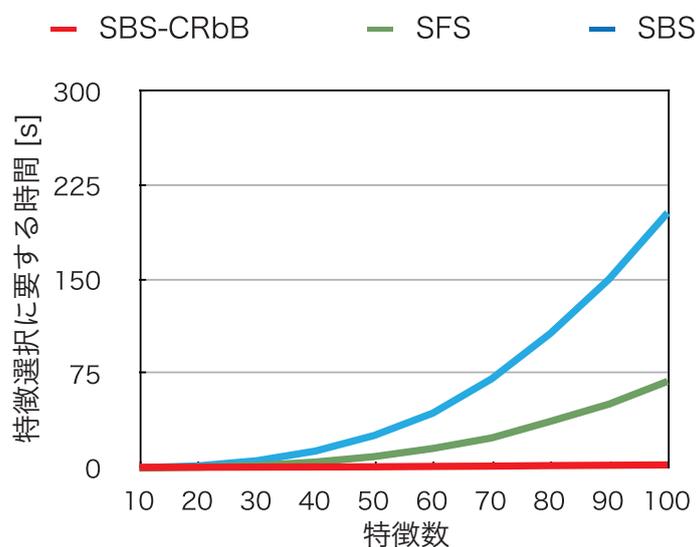


図 4.3: 計算時間の比較。

以上のことから、貢献度に基づく特徴選択は、SBS と同等の識別性能を獲得しながら、計算時間については SFS よりも高速で簡便な特徴選択法であるといえる。

4.4 貢献度に基づく識別器の調整

本節では、貢献度を用いて識別器を簡易に調整することで設置環境に対応させる手法について述べる。

4.4.1 貢献度に基づく識別器の調整法

屋外環境下における物体識別では、学習サンプルデータと設置現場で観測される物体の見えの違いから識別性能が低下する問題がある。これは、学習サンプル収集時の環境とシステム納入時における照明条件、天候などの違いにより発生する。このような問題への対処には、学習サンプルデータに設置環境で収集したデータを加え、識別器を再学習することが一般的である。また、環境変動に頑健な特徴量のみで構成されるよう特徴選択を行うことも有効である。しかし、システムを納入、設置する際の現場において、再学習や特徴選択による早急な対処は難しい。そこで、図 4.4 に示すような、貢献度を参考に、識別対象と各特徴量に対して識別時の感度が学習後に調整可能な識別器を考える。Boosting による識別器は、特徴次元のしきい値判別の結果を重みつき多数決することで識

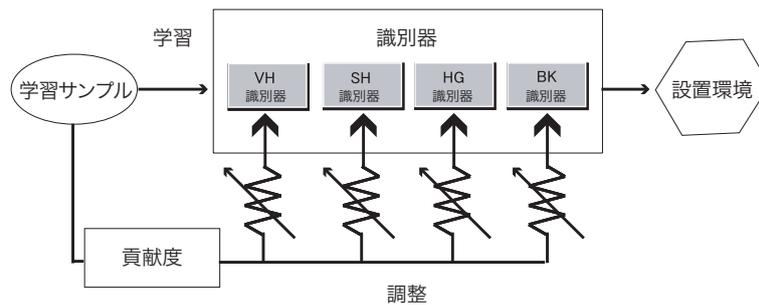


図 4.4: 特徴貢献度に基づく識別器の調整。

別を行う。そこで、弱識別器のしきい値を特徴量毎に調整することで、学習後の識別器を設置環境に対応させる。具体的には、式 (4.5) に示すように、学習後に得られた弱識別器におけるしきい値 θ を Δ_θ により調整することで、弱識別器の感度を変更する。

$$h_m(v, c) = a\delta(v^f > \theta + \Delta_\theta) + b \quad (4.5)$$

識別率の低い任意のクラスに対して貢献度の高い特徴量を優先的に調整することで、設置環境データに対して迅速な対応が可能となる。調整量 Δ_θ は自動決定することが望ましいが、本論文では実験的に求めた性能が低下しない値を与える。

4.4.2 識別器の調整実験

本実験では、貢献度に基づき識別器を評価用データに対し調整し、実際の設置環境へ簡易に対応可能であることを示す。まず、貢献度による特徴量の評価を行い、次に、算出した貢献度に基づいて識別器の調整を行う。これにより、学習サンプルと設置環境の相違で低下した識別性能が改善可能であることを示す。本実験では屋外で撮影されたマルチクラス問題を対象として取り扱う。

データベースとしては、屋外環境下での移動体検出法の評価用データを用いる。屋外に設置したカメラ映像からの移動体の検出・識別は、屋外自動ビデオ監視システムの研究分野で重要なタスクであり、多くの研究が行われている [29][30]。路上に設置したカメラにより収集した識別対象は、自動車 (VH) / 人 (SH) / 複数の人 (HG) / 自転車 (BK) の4クラスについて各10シーケンス、計40シーケンスの動画像を用いる。1シーケンスから40フレーム分の特徴量を抽出し、5シーケンス分200フレームの画像を学習サンプルに用いる。残りの5シーケンス分の200フレームは評価用の設置環境データに用いる。

4.4.3 移動体識別のための特徴量

識別器への入力特徴として、形状に着目した特徴量 (2種類)、テクスチャに着目した特徴量 (4種類)、時間情報に着目した特徴量 (1種類) を用いる。

● 形状に着目した特徴量

物体の全体的な形状を現す特徴として、移動体検出により検出した物体領域に対して、図4.5に示すモーメント特徴 (縦横比、傾き) を用いる。また、人と自動車などの複雑さを表す特徴として複雑度 (周囲長²/面積) を用いる。

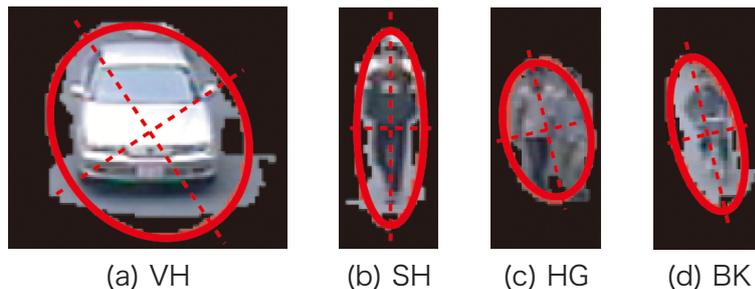


図 4.5: モーメント特徴.

- 縦横比と主軸の傾き (AS)
- 複雑度 (CS)

- **テクスチャに着目した特徴量**

物体の見えを現す特徴として、図 4.6 に示すようにエッジ強度とエッジ連結数をそれぞれ 45 度刻みの 4 方向について用いる。

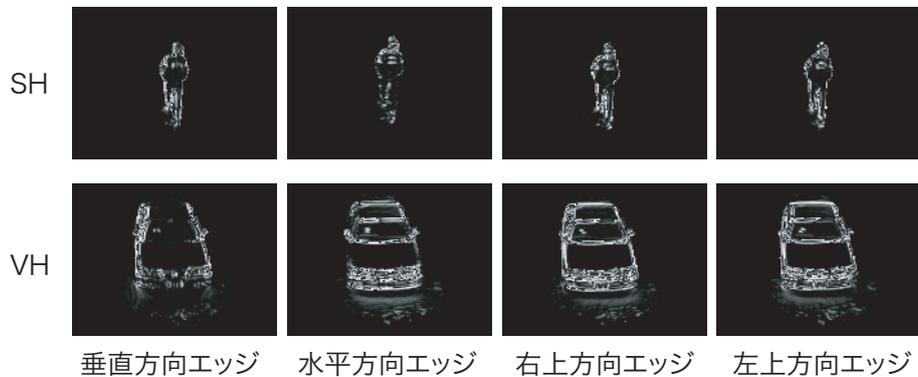


図 4.6: エッジ検出例。

- 垂直方向エッジ (V)
- 水平方向エッジ (H)
- 右上がり方向エッジ (R)
- 左上がり方向エッジ (L)

- **時間変化に着目した特徴量**

「人」と「自転車」のような、見えが似ている物体においては、形状やテクスチャ情報から種別を識別することは困難である。そこで、物体の動きを表す特徴として、オプティカルフローを用いる。図 4.7 に示すように、剛体である自動車や自転車のフローの向きは一定方向に観測され、非剛体である人のフローは足等に向きのばらつきが生じる傾向がある。本章では、フローの検出に特徴点の追跡手法である KLT 法 [31] を用いる。図 4.8(a) のように物体矩形領域からフローを抽出し、図 4.8(b) に示すように、物体矩形領域を 6 分割し、各分割領域内の各フローの向きの分散を算出する。最後に、図 4.8(b) に示すようにその傾きを求め、特徴量とする。

- オプティカルフローの分散 (OF)

これまでに述べた形状情報、テクスチャ情報、時間情報に基づく計 12 個の特徴量を、特徴量毎にその最小値が 0、その最大値が 1 となるようスケール変換し、識別器への入力特徴として用いる。

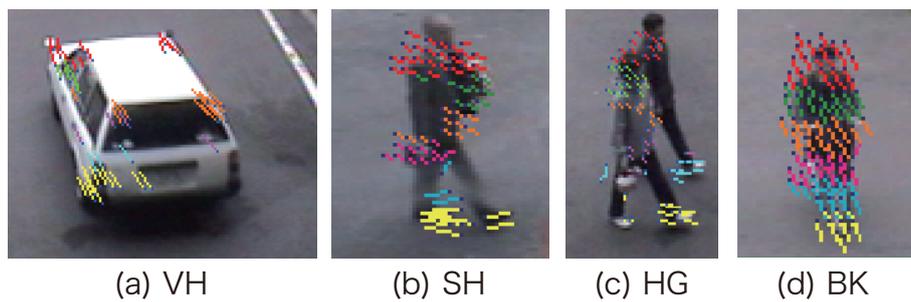


図 4.7: オプティカルフローの抽出.

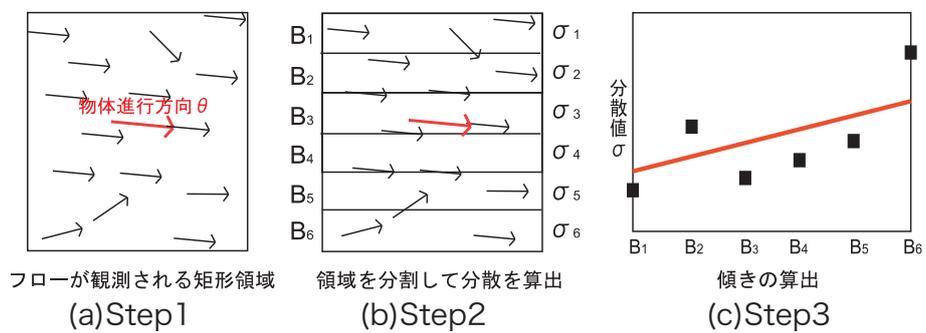


図 4.8: フローの分散の算出.

4.4.4 貢献度に基づく特徴量の有効性の解析結果

図 4.9 に、各ノードに対する特徴量の貢献度を示す。3 クラスで共有されるノードにおける貢献度を図 4.9(a) に、2 クラスで共有されるノードにおける貢献度を図 4.9(b) に、共有のない、自己と他すべてを分けるために有効な貢献度を図 4.9(c) に示している。なお、本問題では背景クラスが存在しないため、4 クラス全てと背景を分類する弱識別器集合 G^{1234} は選択されない。また、自動車 (VH)、複数の人 (HG)、自転車 (BK) と人 (SH) との分類に有効な弱識別器集合 G^{134} は、本実験では選択されなかった。

図 4.9(a-1) より、自動車 (VH)、人 (SH)、複数の人 (HG) の集合と自転車の分類には垂直方向のエッジ成分の貢献度が非常に高く、有効であることがわかる。また、図 4.9(a-2) や (a-3) より、それぞれエッジ情報が共有可能であることがわかる。さらに、図 4.9(b-2) では人、複数の人 (HG) と自動車 (VH)、自転車 (BK) という剛体、非剛体の分類に有効な特徴量として垂直方向のエッジ成分に加えて OF が非常に有効に働いていることがわかる。同様に、図 4.9(b-4) より自動車 (VH)、複数の人 (HG) と人 (SH)、自転車 (BK) の分類には、斜め方向のエッジに加え形状情報、中でも AS が非常に有効であることが観測できる。最後に、図 4.9(c) に着目すると、自動車 (VH) とその他の分類に対する水平エッジの有効性が非常に顕著に表れている。このように、提案手法は各特徴量の評価をクラスとその集合単位で詳細に表現することが可能である。

4.4.5 特徴貢献度に基づく識別器の調整結果

貢献度を用いた調整実験を行う。まず、未調整状態における設置環境データに対する識別結果を表 4.2 に示す。表より、調整前の識別器は、複数の人 (HG) クラスを自転車 (BK) クラスに誤識別することが多く、識別性能が低い。

表 4.2: 調整前の識別性能.

	VH	SH	HG	BK	識別率 [%]
VH	120	0	0	0	100.0
SH	4	100	0	16	83.3
HG	28	5	25	62	20.8
BK	7	1	10	112	93.3

そこで、図 4.9 の G^{34} に着目し、複数の人 (HG)、自転車 (BK) 共に貢献度が高い特徴である複雑度 (CS) と左上がり方向エッジ (L) を選択し、HG クラスの識別感度を調整する。ここで問題となるのが、どの弱識別器に対してどの程度変更を加えるかという点である。本研究では、任意のクラス固有のノードの弱識別器 (複数の人 (HG) クラスであれば G^3) に対して調整を加える。調整量については調整/再学習用設置環境データを用いて高性能であった $\Delta th = 0.3$ を採用した。表 4.3 に設置環境データを加えて再学習した結果を示す。表 4.4 に貢献度に基づいた調整を行った際の

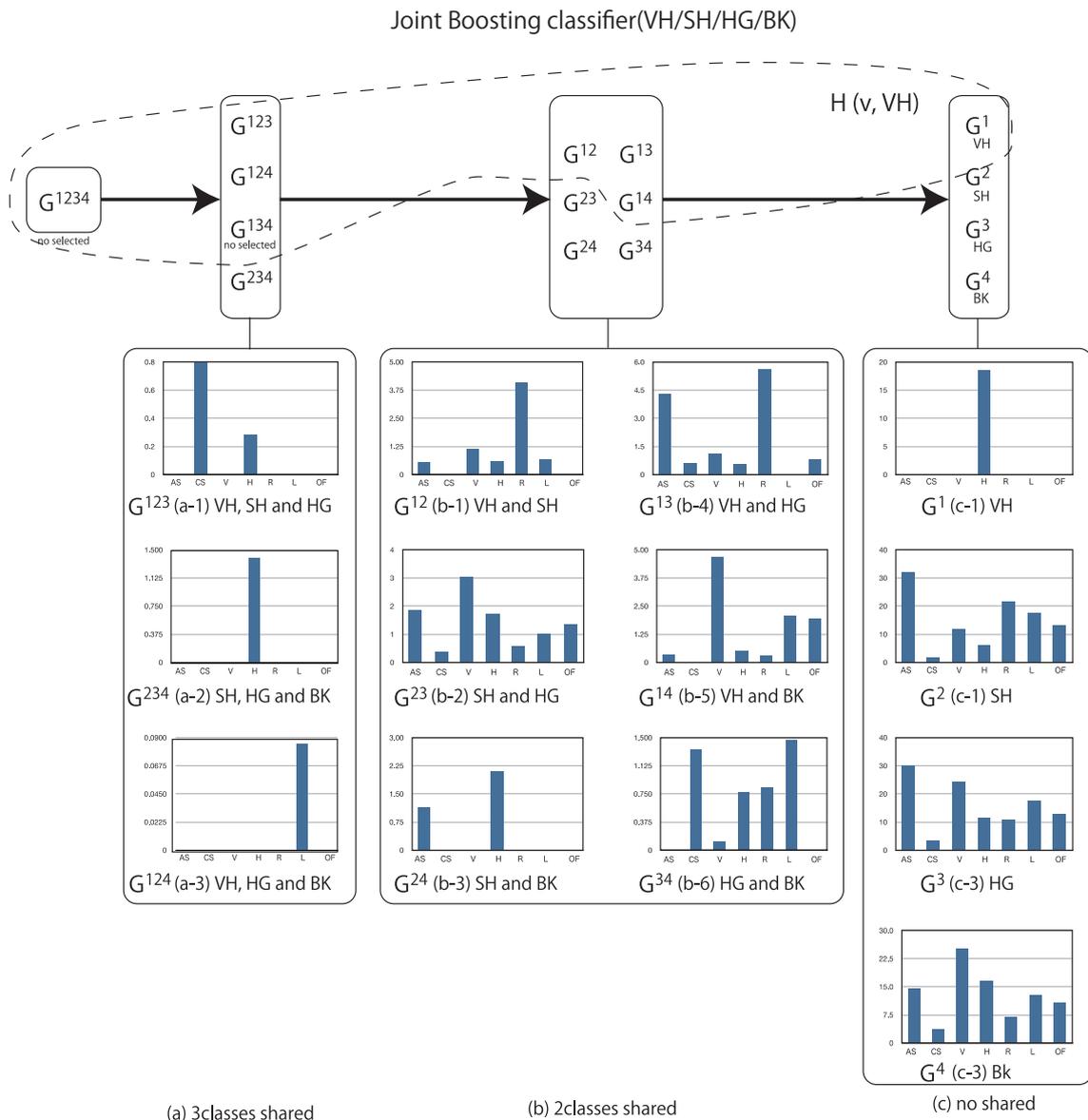


図 4.9: 貢献度算出結果.

結果を示す.

再学習と調整のどちらも複数の人 (HG) クラスの識別性能が大きく改善され、平均で7%以上識別性能を向上していることがわかる。しかし、再学習の表 4.3 では、再学習前では存在しなかった自転車 (BK) クラスから人 (SH) クラスへの誤識別が27%増加している。これは、再学習前に問題であった複数の人 (HG) クラスについては改善されたものの、別の問題が発生したといえる。一方、提案手法は貢献度の高い特徴を問題のクラスについてピンポイントで調整することで、問題であった複数の人 (HG) クラスを自転車 (BK) クラスへ誤識別する点を解決しながら、他クラス間の誤

表 4.3: 識別器を再学習した際の識別性能.

	VH	SH	HG	BK	識別率 [%]
VH	120	0	0	0	100.0
SH	0	119	0	1	99.1
HG	0	44	71	5	59.1
BK	3	33	13	87	72.5

表 4.4: 識別器の調整 (HG).

	VH	SH	HG	BK	識別率 [%]
VH	120	0	0	0	100.0
SH	3	86	17	14	71.6
HG	9	2	77	32	64.1
BK	6	0	19	111	92.5

識別は大きく変化していない。これは、設置環境下において特定の問題が発生した場合、その問題に対して直接アプローチできる提案手法の利点を表している。さらに、提案手法は再学習と異なり、この上に微調整を加えることができる。複数の人 (HG) クラスに対する調整により識別率が低下した人 (SH) クラスについて、図 4.9(c-1) より人 (SH) クラスと他の分類に最も有効である AS について行った結果を表 4.5 に示す。表 4.5 より、再度の微調整を行うことで、平均性能で再学習を上回ることも可能であることがわかる。

表 4.5: 識別器の調整 (SH,HG).

	VH	SH	HG	BK	識別率 [%]
VH	0	120	0	0	100.0
SH	2	96	11	11	80.0
HG	9	8	75	28	62.5
BK	5	3	19	109	90.8

本実験の最後に、人 (SH) クラス、自転車 (BK) クラスと誤識別されていたが調整により複数の人 (HG) クラスと正解できたデータ例を図 4.10(a), (b), (c) に示す。

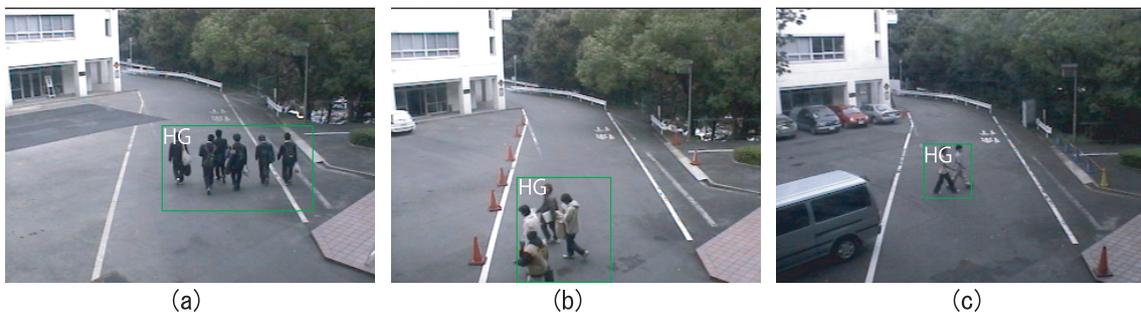


図 4.10: 調整により改善された物体検出例.

4.5 まとめ

本章では、まず、識別器における特徴削減や調整の必要性と従来法の持つ計算コストの問題を述べた。次に、Boosting 識別器より特徴量の識別への貢献度を算出することを提案し、貢献度を用いた特徴削減法と、識別器の調整法について述べた。特徴削減の対象として UCI-ML Repository から 3 種類の問題について評価実験を行い、貢献度に基づいた特徴削減を行うことで、SFS 法以上の高速性で SBS 法と同程度の高精度であることを確認した。

また、設置環境において Boosting 識別器を簡易に直接調整するフレームワークを提案した。評価実験では、屋外環境下で撮影された映像における移動体を自動車／人／複数の人／自転車の各クラスに識別するための入力特徴として、形状に着目した特徴量 (2 種類)、テクスチャに着目した特徴量 (4 種類)、時間情報に着目した特徴量 (1 種類) を使用した。評価実験の結果、貢献度に基づいて調整を行うことで、調整前と比べて 7%以上性能が向上することを確認した。これにより、識別器を設置現場で簡易に調整し、短時間で設置環境へ適応することが可能となる。

第5章

生成型学習に基づく学習の効率化

4章では、特徴量の貢献度を用いて、再学習することなく簡易に識別器を調整するアプローチを提案した。しかし、カメラの設置角度が大きく変化する場合、画像上の見えは、物体のパーツの位置やサイズも大きく変化するため、調整での対応が困難となる。このような場合、再学習を必要とするが、再学習には大量の学習サンプル収集等の大きな人的コストが必要となるという問題がある。

本章では、特定シーンに対する再学習を効率化、高精度化する手法を提案する。5.1では、人体モデルのCGを用いて特定シーンに特化した学習サンプルを生成する方法について述べる。5.2では、誤ラベルが付与された学習サンプルの混在を考慮した Negative-Bag MILBoost による学習について述べる。これにより、特定シーンから自動的に大量の学習サンプルを収集し、人的コストを大きく削減可能となることが期待できる。

5.1 学習サンプルの自動生成による効率化

カメラの設置視点の変化等の大きな変動があった場合、4章で述べた調整での対応は困難となり、再学習を必要とする。再学習の際に問題となるのが、画像中の人の位置やサイズなどが標準化された学習サンプルを大量に収集するために必要な膨大な人的、時間的コストである。

そこで、本章では3次元人体モデルを用いた学習サンプルの自動生成と Negative-Bag MILBoost による生成型学習法を提案する。学習サンプルの自動生成法として、Mar ら [32] はCGを用いた人体画像生成法を提案している。本手法は、低コストで大量に学習サンプルを得ることを目的としており、ゲームで用いられる街のシーンのCGから、人画像の切出しを行う手法である。そのため、設置環境の特定シーンを反映した学習サンプルの収集は不可能である。そこで、本研究では、3次元人体モデルを用いて特定シーンに特化した学習用ポジティブサンプルを自動的に生成し、ネガティブサンプルも同様に特定シーンから収集する。

このように、特定シーンに特化した学習サンプルを学習に用いることで、特定シーンにおいては、一般的なデータベースで学習した識別器と比べて精度を向上することができる。また、学習用ネガティブサンプルは、映像からランダムに自動的に切り出して収集するため、人画像が含まれることがある。このように誤ったラベルを付与された学習サンプルを用いて学習すると、識別能力が低下するという問題がある。そこで、本研究では誤ラベルのサンプルが混在しても悪影響を受けない Negative-Bag MILBoost を用いる。

提案手法の流れを図5.1に示す。提案手法は、特定シーンに特化した学習サンプルを自動生成するために、ポジティブサンプルを、3次元の人体モデルのCGを用いて生成する(図5.1(a))。ネガティブサンプルは映像中から切り出した画像を用いる(図5.1(b))。このように生成した学習サンプルを Negative-Bag MILBoost に入力することで識別器を学習する。

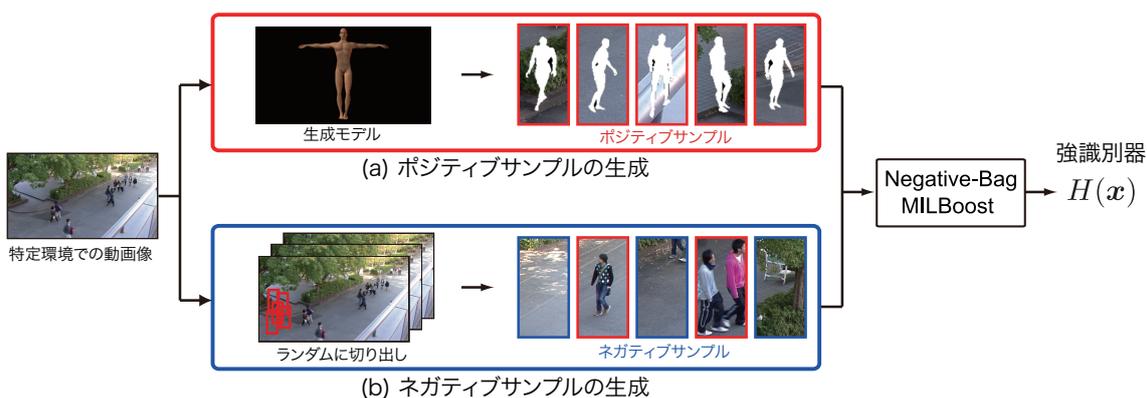


図 5.1: 生成型学習の流れ。

学習用ポジティブサンプルの生成には、図5.2に示すような人体モデルと仮想カメラを用いる。これにより、様々な視点からの自由な姿勢の人体シルエット画像を生成することが可能となる。

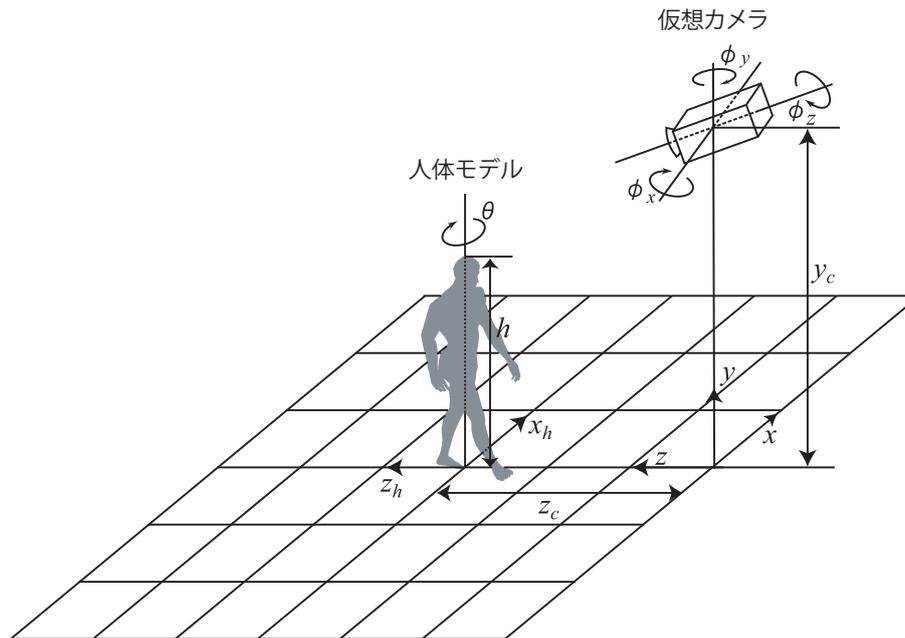


図 5.2: 3次元人体モデル.

5.1.1 3次元人体モデルによる人画像生成

提案手法で使用する人体モデルには、形状モデルやモデルの各パーツの階層構造、動作データなどが含まれている。人体の形状モデルは、19のパーツが存在し、これらのパーツは階層的な構造で表現される。そのため、例えば右肩を動かした場合、右腕や右手が連動して動く。本研究では、19のパーツに歩行動作のパラメータを与えることで、歩行姿勢の人体モデルを表現する。また、下記の各パラメータを与えることで、図5.3に示すように任意の視点から撮影した人体モデルの姿勢を生成することができる。

- カメラパラメータ
カメラ位置 x_c, y_c, z_c , カメラ角度 ϕ_x, ϕ_y, ϕ_z
- 人体形状パラメータ
身長 h , 人の向き θ , 人の位置 x_h, y_h, z_h
- テクスチャ
背景のテクスチャ T_{bg} , 人体のテクスチャ T_{in}

特定シーンに特化した人体シルエット画像を得るために、実環境に設置したカメラのパラメータを3次元人体モデルに入力する。本研究では固定カメラを想定しており、事前に得たカメラパラメータを用いる。人形状パラメータにおける人の向きと位置は事前に決定できないため、一様にランダム

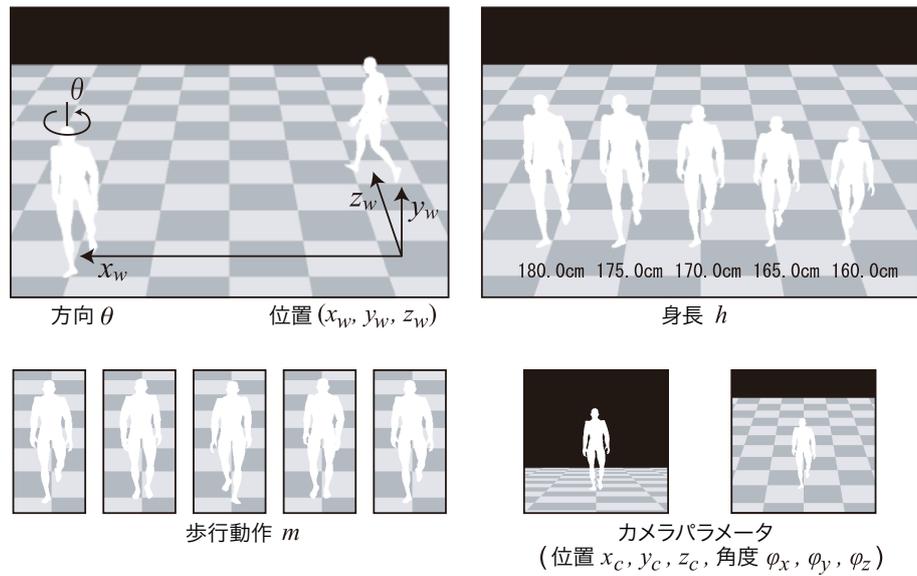


図 5.3: 3次元人体モデル生成時のパラメータ。

設定とする。人の身長は、文部科学省の統計調査 [33] より平均身長 171.9cm とする。人体のテクスチャは、衣服などを考慮することも考えられるが、多種多様な衣服を用意することが難しい。また、人体のテクスチャを張り付けないことも考えられるが、このようなサンプルを用いて学習した場合、人の内部はテクスチャがないものとして学習される。そこで、本研究では事前に用意した自然画像のテクスチャをランダムで貼り付ける。これにより、衣服によるテクスチャ変化の影響を受けずに、人の輪郭を学習する。また、設置したカメラから得られる画像を背景のテクスチャとして用いる。

図 5.4 はカメラの高さ $y_c = 6.2m$ 、カメラ角度 $\phi_x = 21^\circ$ とした際の人体シルエット画像を生成した例である。このように特定シーン画像において人体シルエットを生成し、人体領域を中心に切り出した画像を学習用ポジティブサンプルとして用いる。

5.1.2 ランダムサンプリングによる背景サンプルの自動収集

ネガティブサンプルは、撮影した映像中からランダムで収集する。これにより、特定シーンにおいては、一般的なデータベースで学習した識別器と比べ、特定シーンに特化したネガティブサンプルを収集できるため、識別器の高精度化が可能となる。しかし、ランダムにサンプルを収集した場合、図 5.1(b) に示すように、ネガティブサンプルとして人画像が収集される問題がある。この問題を解決するために、本研究では、誤って付与されたラベルを持つサンプルの混在を考慮した識別器の学習法として、Negative-Bag MILBoost を提案する。

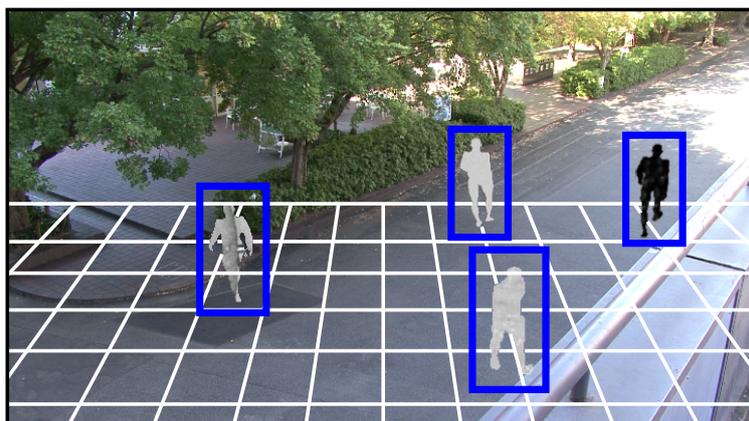


図 5.4: 実環境に対する 3 次元人体モデルの生成例.

(camera position $(x_c, y_c, z_c) = (0m, 6.2m, 0m)$, camera angle $(\phi_x, \phi_y, \phi_z) = (21^\circ, 0^\circ, 0^\circ)$)

5.2 Multiple Instance Learning の導入によるノイズ低減

本章では誤ラベルが付与されたサンプルが混在してしまう問題を解決するために、Negative-Bag MILBoost による識別器の学習法について述べる。

5.2.1 MILBoost

Multiple Instance Learning (MIL) [34] は、複数のサンプルから構成される Bag に対してラベルを付与する。Bag は画像座標上の位置が近傍関係にある複数のサンプルにより構成され、Bag 中のサンプルに一つでもポジティブが含まれていればポジティブ Bag とラベル付けされ、それ以外はネガティブ Bag とラベル付けされる。MIL では、Bag に与えられたラベルに基づき識別器を学習する。そのため、ラベルが付与されていない未知のサンプルを含むデータに対しても学習が可能なアプローチとなる。本研究では、MIL を Boosting に導入した MILBoost [35] を拡張した Negative-Bag MILBoost を提案する。

MILBoost は MIL の学習モデルを Boosting に導入した学習アルゴリズムであり、Viola らは顔検出器の学習に適用している。顔検出器の学習サンプルは顔画像を大量に必要とするため、顔画像の収集を対象領域とその周辺領域を切り出し集めた Bag として扱い、ポジティブサンプルの収集を容易にしている。このポジティブサンプルには、画像中に背景の比率が高い顔画像等が含まれるが、MIL を導入することによって、それらの影響を低減し、高精度な顔検出を実現している。MILBoost は、Boosting により学習する際、各 Bag と各サンプルに対するクラス尤度を求め、サンプルに対する重みの更新時に、クラス尤度を用いて誤って付与されたラベルのサンプルの重みを小さくする。これにより、誤って付与されたラベルの影響を抑制することが可能となる。

5.2.2 Negative-bag MILBoost

MILBoost[35]では、ポジティブサンプルのラベルに誤ラベルが混在する問題を対象としている。しかし、提案手法ではネガティブサンプルのラベルに誤ラベルが混在する問題である。そこで、本研究の問題設定にあわせて、ネガティブサンプルにMILBoostのBagを適用する。ここでは、まずBagの作成方法について述べ、次にNegative-Bag MILBoostの学習アルゴリズムについて述べる。

■ Bagの作成方法

本研究の問題設定では、ネガティブサンプルに誤ラベルを含むため、文献[35]のBagの構成を図5.5に示すように変更する。図5.5(a)に示す従来法では、ポジティブサンプルが誤ラベルを含むため、複数のポジティブサンプルをBagとし、ラベル付けが確実なネガティブサンプルは単体の画像を1つのBagとして学習に利用する。一方、提案手法においては図5.5(b)で示すように、人体モデルから生成した人画像を一つのBagとして扱う。ネガティブBagは、映像中から切り出した複数のサンプル一つのBagとして扱う。このネガティブBagの中に人画像が含まれる可能性がある。

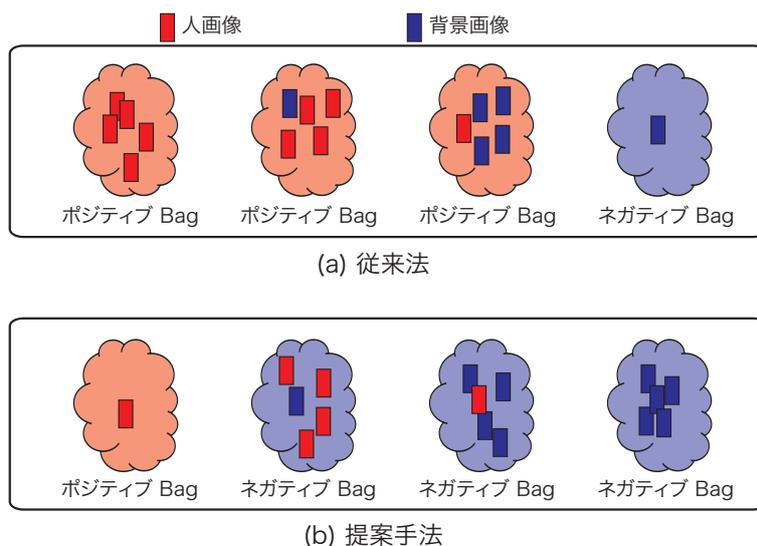


図 5.5: 提案手法における Bag の構成。

■ Negative-Bag MILBoost の学習

MILBoost の学習は、通常の Boosting と共通のアルゴリズムであるが、学習サンプルの重みの更新法が異なる。本研究では、学習サンプルに対する重みの更新処理を本研究の問題設定に適應する

ように拡張する。Real AdaBoost [6] をベースとした Negative-Bag MILBoost の学習アルゴリズムを**アルゴリズム 5.1** に示す。

まず、学習の事前準備として J 個の学習サンプルが所属する I 個の Bag を用意する。これらの Bag は、ネガティブサンプルを 1 つでも含んでいればネガティブ Bag、それ以外はポジティブ Bag としてラベルが与えられている。次に、 i 番目の Bag に所属する j 番目のサンプルの重み $w_t(i, j)$ を式 (5.1) により初期化する。

次に、学習サンプルから抽出した特徴ベクトル \mathbf{x} を入力とした確率密度関数 W_+ , W_- を式 (5.2), (5.3) により作成する。特徴量には、Histogram of Oriented Gradients(HOG) 特徴量 [3] を用いる。確率密度関数 W_{\pm} は、HOG 特徴量で得られる特徴次元毎にヒストグラムとして表現される。学習サンプルの重み $w_t(i, j)$ は、そのサンプルの特徴量の値を K 段階で量子化した値 k とラベル y に基づき、確率密度関数 W_+^k または W_-^k に累積される。ここで t は学習回数を表す。確率密度関数 W_{\pm} を作成した後、式 (5.4) より弱識別器 $h(\mathbf{x})$ の評価値 Z を求める。評価値 Z は弱識別器の識別性能を表し、小さいほどポジティブサンプルとネガティブサンプルを分離する能力があることを表す。 L 個の弱識別器候補について評価値を求め、最も識別性能が高い弱識別器候補を t 個目の弱識別器 $h_t(\mathbf{x})$ (式 (5.5)) とする。1 つの弱識別器で 1 つの特徴量を観測し、その全てについて評価を行うため、 L は特徴量の次元数と同数である。

弱識別器を選択後、式 (5.7), (5.8) から得られる Bag のクラス尤度 p_i とサンプルのクラス尤度 p_{ij} を用いて学習サンプルの重みを更新する。以上の処理を T 回繰り返すことで、式 (5.9) より最終識別器 $H(\mathbf{x})$ を得る。

アルゴリズム 5.1: Negative-Bag MILBoost の学習アルゴリズム.

1. 入力: J 個の学習サンプルを含む I 個の Bag とラベル $y \in \{1, 0\}$ を用意

2. 初期化: 学習サンプルの重み $w_t(i, j)$ を初期化

$$w_1(i, j) = \frac{\text{Bag of class}}{\text{Bag of all}} \quad (5.1)$$

3. 学習:

For $t = 1, \dots, T$ // 学習ラウンド

For $l = 1, \dots, L$ // 弱識別器の候補数

• 弱識別器の確率密度関数 W_{\pm} の算出

$$W_+^k = \sum_{i,j:k \in K \wedge y_i=1} w_t(i, j) \quad (5.2)$$

$$W_-^k = \sum_{i,j:k \in K \wedge y_i=0} w_t(i, j) \quad (5.3)$$

• Z_l の評価

$$Z_l = 2 \sum_{k=1}^K \sqrt{W_+^k W_-^k} \quad (5.4)$$

End for

• Z_l が最小の $h_t(\mathbf{x})$ を弱識別器として選択

• 弱識別器 $h(\mathbf{x})$

$$h(\mathbf{x}) = \frac{1}{2} \ln \frac{W_+^k + \varepsilon}{W_-^k + \varepsilon} \quad (\varepsilon = 1/J) \quad (5.5)$$

• 重み $w_t(i, j)$ の更新

$$w_t(i, j) = \begin{cases} -p_{ij} & \text{if } y_i = 1 \\ \frac{p_{ij} \times (p_i)}{1 - p_i} & \text{if } y_i = 0 \end{cases} \quad (5.6)$$

$$p_i = \prod_{j \in \text{Bag}_i} p_{ij} \quad (5.7)$$

$$p_{ij} = \frac{1}{1 + \exp(-H_t(\mathbf{x}))} \quad (5.8)$$

End for

4. 出力

• 強識別器 $H(\mathbf{x})$

$$H(\mathbf{x}) = \text{sign} \left(\sum_{t=1}^T h_t(\mathbf{x}) \right) \quad (5.9)$$

■ Negative-Bag を用いた誤ラベルへの対応

図 5.6 に Negative-Bag MILBoost による学習サンプルの重みの更新の概念図を示す.

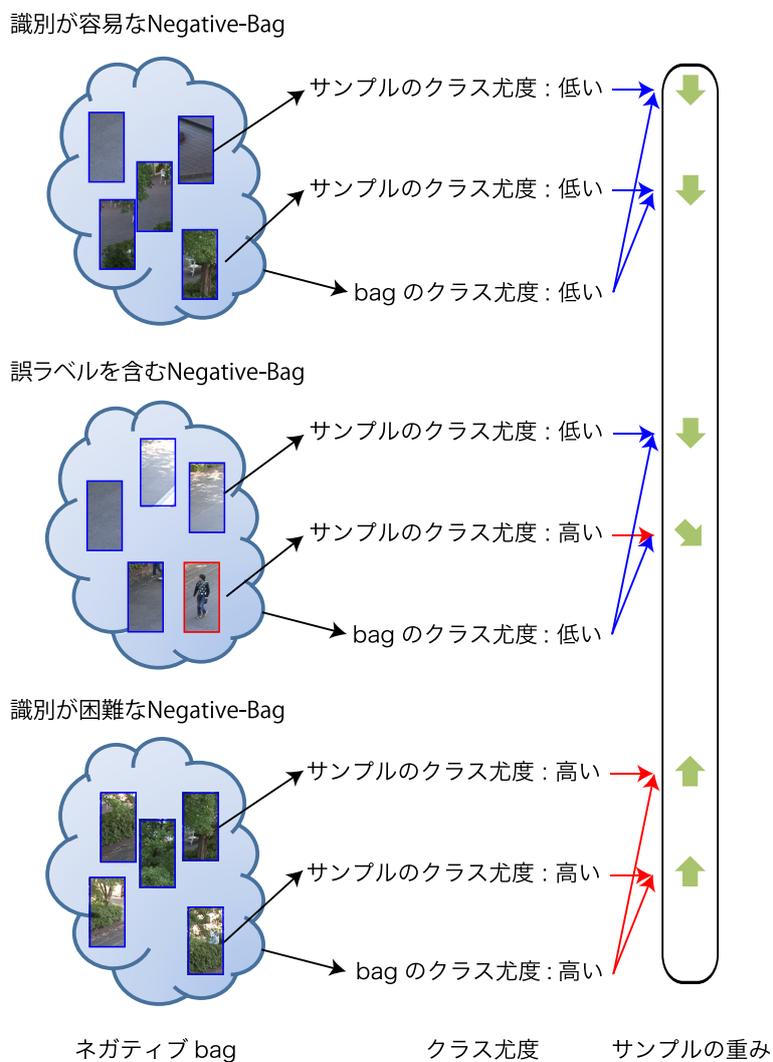


図 5.6: Negative-Bag MILBoost による学習サンプルの重みの更新.

MILBoost では, サンプルに対するクラスラベルが付与されていないため, i 番目の Bag に含まれる j 番目の学習サンプルの重み $w_i(i, j)$ は式 (5.6) に従い Bag のラベルに基づき更新される. ポジティブ Bag に含まれるサンプルに対しては, サンプルのクラス尤度 p_{ij} により重みを更新する. クラス尤度は, 高い値であるほど人画像, 低い値であるほど背景画像の可能性が高いことを表している. ネガティブ Bag に含まれているサンプルは, サンプルのクラス尤度 p_{ij} と Bag のクラス尤度 p_i により重みを更新する. サンプルのクラス尤度 p_{ij} の値が総じて低いと Bag のクラス尤度 p_i の値も

低くなる。クラス尤度の低いサンプルは識別が容易なネガティブサンプルであるため、Bag 中のサンプルの重み $w_t(i, j)$ は低下する。もし、その中に誤ラベルされたポジティブサンプルが含まれていても、Bag のクラス尤度 p_i が十分に低ければ、そのサンプルをノイズであると捉え、サンプルの重み $w_t(i, j)$ は低下する。一方、サンプルのクラス尤度 p_{ij} の値が総じて高いと Bag のクラス尤度 p_i の値が高くなる。この各サンプルは識別が困難なネガティブサンプルであるため、Bag 中のサンプルの重み $w_t(i, j)$ を向上させ、重点的に学習を行う。このように、Bag 単位での尤度を用いることでノイズの影響を低減することができる。

■ 識別

識別時は、従来法である Real AdaBoost と同様に計算する。式 (5.9) に示すように複数の弱識別器の線形和を求め、しきい値により判定する。しきい値より大きい場合は人、そうでない場合は背景として識別結果を出力する。

5.3 評価実験

提案手法の有効性を示すために2つの評価実験を行う。1つ目の実験では、特定シーンに特化した学習サンプルの自動生成の有効性を示す。2つ目の実験では、誤ラベルに対する評価実験により、Negative-Bag MILBoost による学習法の有効性を示す。

5.3.1 実験概要

特定シーンに特化した学習サンプルの自動生成による有効性を評価する。4種類のデータベースを用いて識別器を学習し、識別性能を比較する。Database 1はINRIA Person Dataset[3]に含まれている人画像をポジティブサンプルとして使用し、提案手法により自動収集した特定シーンの背景画像をネガティブサンプルとして使用する。Database 2は実環境で収集した人画像をポジティブサンプル、提案手法により実環境から自動収集した背景画像をネガティブサンプルとして使用する。Database 3は提案手法により生成した人画像をポジティブサンプル、INRIA person datasetの背景画像をネガティブサンプルとして使用する。Database 4は提案手法により生成した人画像をポジティブサンプル、提案手法により実環境から自動収集した背景画像をネガティブサンプルとして使用する。一般的なデータベースの人画像を用いた Database1、実環境の人画像を用いた Database2 と提案手法である Database 4 との比較により、提案手法で生成する人画像を用いて、実環境でサンプルを収集したものと同等に特定シーンの学習が可能であることを示す。また、提案するネガティブサンプルの自動収集についても評価するため、一般的なデータベースである INRIA Person Dataset のネガティブサンプルと生成画像の組み合わせである Database3 と提案手法である Database 4 を比較する。データベース毎に使用するポジティブサンプル、ネガティブサンプルをまとめると以下のようになる。

- Database 1 : INRIA Pos. + 生成 (Generated) Neg.
- Database 2 : 実環境 (Real environment) Pos. + 生成 Neg.
- Database 3 : 生成 Pos. + INRIA Neg.
- Database 4 : 生成 Pos. + 生成 Neg.

実験結果の比較には DET カーブを用いる。

■ データベース

実験に用いるデータベースの詳細について述べる。INRIA Pos. と INRIA Neg. は、INRIA Person Dataset に含まれている人画像と背景画像である。このデータベースは、多様な人の姿勢、向き、視点、照明の変動や背景テクスチャを含んでおり、非常に汎用性の高いデータベースである。実環境 Pos. は、実環境下で撮影した映像から人手により切り出した人画像である。生成 Pos. は、5.1にて述べた3次元人体モデルを用いて生成した人画像である。生成 Neg. は、実環境画像からランダ

ムに生成した背景画像である。本実験に使用する実環境下の映像は、人の通行量が多い屋外の通路を撮影したものである。カメラの高さ 6.2m, 俯角 21° としてカメラを設置し、約 1 時間の映像を撮影した。提案手法により生成するサンプルは、人体モデルに上記のカメラの高さと俯角、背景テクスチャを与えることで生成した。表 5.1 に学習に使用する画像データベースの枚数と種類を示す。図 5.7 に学習に使用した各データセットのサンプル例を示す。評価用データベースには、実環境下で撮影した動画画像からランダムで選択した 450 フレームを評価用データベースとして用いる。

表 5.1: 学習用データベースの内訳

	ポジティブ	ネガティブ
Database 1	INRIA Pos.(2,416)	生成 Neg.(12,180)
Database 2	実環境 Pos.(2,416)	生成 Neg.(12,180)
Database 3	生成 Pos.(2,416)	INRIA Neg.(12,180)
Database 4	生成 Pos.(2,416)	生成 Neg.(12,180)



(a) INRIA positive samples (INRIA Pos.)



(b) 生成 positive samples (生成 Pos.)



(c) 実環境 positive samples (実環境 Pos.)



(d) 生成 Negative samples (生成 Neg.)



(e) INRIA Negative samples (INRIA Neg.)

図 5.7: 学習データベース例.

5.3.2 生成画像に基づく特定シーン学習の有効性

各データベースを用いて学習した際の DET カーブを図 5.8 に示す。

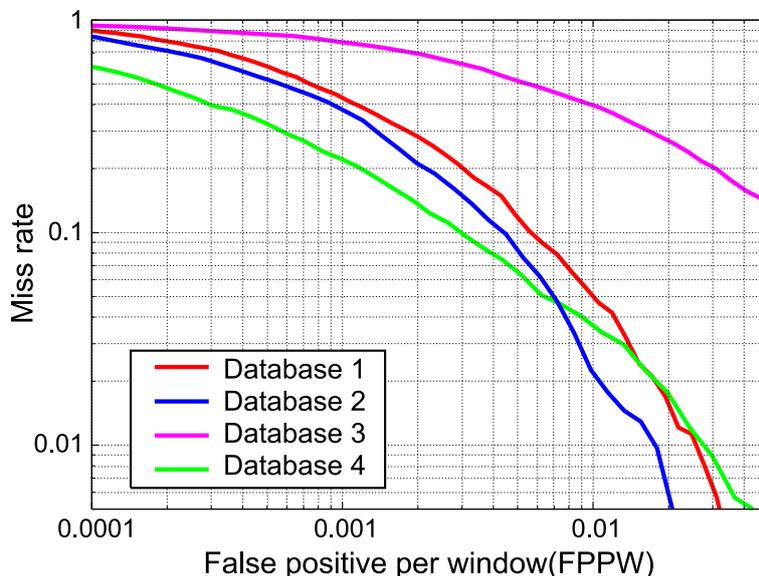


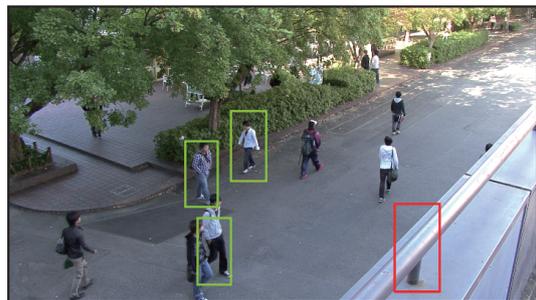
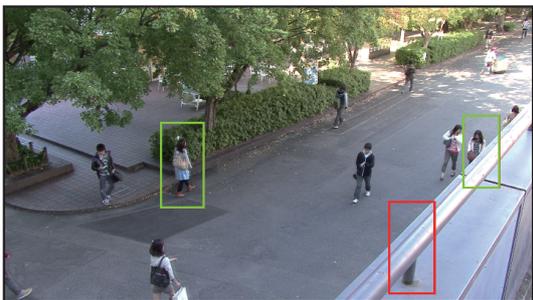
図 5.8: 各データベースで学習を行った際の DET カーブ。

まず、ネガティブサンプルが同一の Database 1, Database 2, 提案手法である Database 4 を比較すると、検出性能が最も高いのは人体モデルから生成したサンプルを用いた提案手法であった。これは、実環境下で撮影した映像に対応した人の見えを生成できたからといえる。実環境下の映像から人手で切り出したサンプルを用いた Database 2 は、自動生成よりも低い結果となった。これは、人画像を人手で大量に切り出す際に、切り出し基準が曖昧になることがあり、これが識別器に悪影響を及ぼしたと考えられる。汎用性のあるデータベースを用いた Database 1 の結果が最も低い検出率となった。これは、学習用データベースの INRIA Person Dataset は実験環境とカメラ位置が異なるため、サンプル中の人の見えも大きく異なるからといえる。

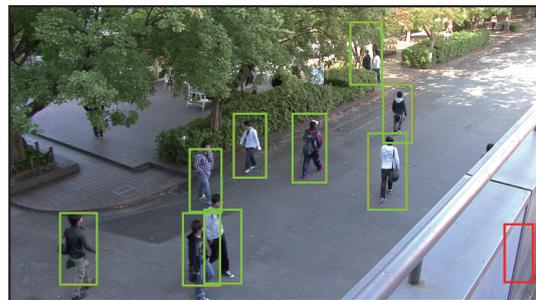
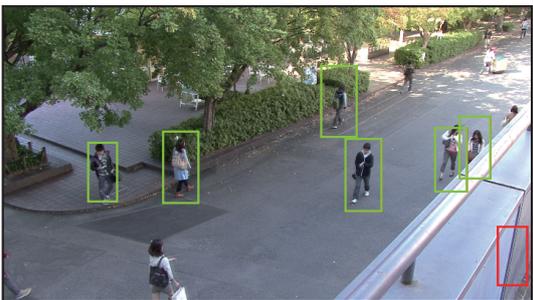
次に、Database 3 と提案手法を比較すると、実環境下で撮影した映像の背景を用いた提案手法の方が良い結果が得られた。これは、提案手法では実環境の画像から切り出した背景画像を学習用ネガティブサンプルとして用いているため、実環境のシーンに特化した識別器となり検出性能が大きく向上したといえる。

提案手法と従来法それぞれの人検出結果例を図 5.9 に示す。

図 5.9 より、同一画像において FPPW を等しくした時、提案手法はより多くの人を正しく検出できていることがわかる。



(a) 従来法



(b) 提案手法

図 5.9: 人検出結果例 (緑：正検出ウインドウ, 赤：誤検出ウインドウ).

5.3.3 Negative-bag MILBoost によるノイズ低減の有効性

ネガティブの誤サンプルに対応した Negative-Bag MILBoost による学習法の有効性を評価する。提案手法と Real AdaBoost を比較する。

- 従来法：Real AdaBoost
- 提案手法：Negative-Bag MILBoost

Negative-Bag MILBoost の有効性を確認するために、学習用のネガティブサンプルへ人画像を混在させて識別器を学習する。ネガティブサンプルにおける人画像の割合を 0% ~ 30% まで変動させ、その際の識別性能を比較する。実験に使用するデータベースは、ポジティブサンプルには INRIA Person Dataset の人画像 1,200 枚を用いる。ネガティブサンプルには INRIA Person Dataset の背景画像 4,000 枚を用いる。ネガティブサンプルのノイズとして用いる人画像は、INRIA Person Dataset のポジティブサンプルに使用していない 1,200 枚を 0%~30% の割合で混在させる。評価に使用するデータベースは、INRIA データベースにおける評価データセットと同じものを使用する。本実験の比較には Equal Error Rate(EER) を用いる。EER は、DET において miss rate と FPPW が等しい時の値であり、この値が低いほど性能が高いことを表す。DET との違いは、1つの手法の評価が1つの値で表現できることであり、本実験のように、複数のバリエーションをもつ学習サンプルを用いて手法を比較する際に用いられる。

■ 実験結果

実験結果を図 5.10 に示す。実験結果より、従来法はネガティブサンプル中に人画像の含有率が高くなるに従って EER が高くなる。一方、提案手法 (NB MILBoost) では EER の増大を抑制していることがわかる。人画像の含有率が 15% の場合を比較すると、提案手法は従来法よりも EER が 6.1% 低い。提案手法はネガティブサンプル中に人画像が含まれていても、識別器の学習に及ぼす悪影響を低減することができたといえる。また、混入率 0% 時点においても提案手法の EER がわずかに低いのは、ノイズ低減効果による差であると考えられる。

同じネガティブ Bag に含まれる人画像と背景画像の強識別器の出力と、サンプルのクラス尤度の遷移を図 5.11 に示す。横軸は学習回数、左縦軸はサンプルの重み、右縦軸は強識別器の出力値を表す。図 5.11 より、ネガティブ Bag の人画像の識別結果は人らしい高い出力値を持つことがわかる。通常の Boosting の場合、ラベルと異なる識別結果を出力すると、次の弱識別器の選択では識別困難なサンプルとして重みが上昇する。しかし、図 5.11 より、提案手法では逆に重みが低下していることがわかる。これは、Bag 単位での尤度を用いて本サンプルをノイズと判断して、学習に利用しないことを意味する。その結果、ネガティブサンプルに混入されているにもかかわらず、識別の際は人らしい出力値を保つことができた。一方、背景画像は背景と識別され、出力、重み共に低下している。ネガティブ Bag においては、サンプルのクラス尤度が低いほど、学習サンプルの重みが低下する。そのため、提案する Negative-Bag MILBoost の学習アルゴリズムは、誤って付与されたラベルを持つ人画像の悪影響を低減することができた。

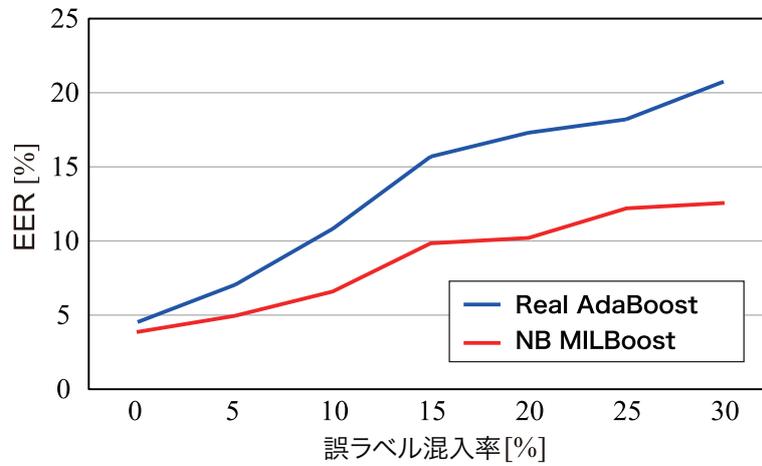


図 5.10: ノイズを混入した際の識別性能.

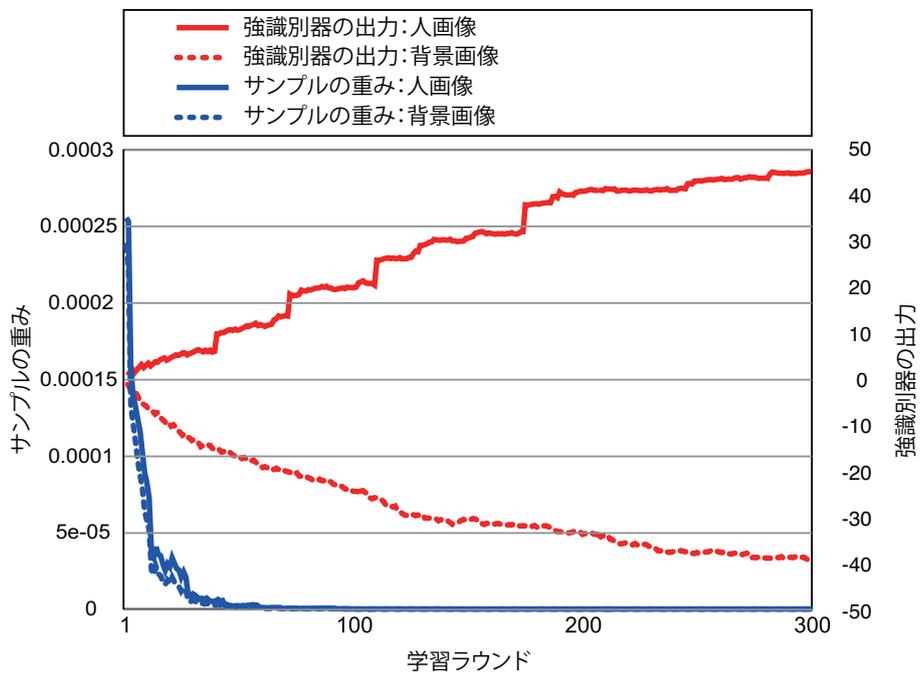


図 5.11: 識別器の出力と重みの遷移.

5.4 まとめ

本章では、3次元人体モデルから学習サンプルの自動生成と Negative-Bag MILBoost を用いた生成型学習法について述べた。特定シーンにおいて、3次元人体モデルから生成した学習サンプルを用いることにより、実環境に特化した識別器を学習することが可能である。さらに、Negative-Bag MILBoost を用いることにより、誤ラベルを付与されたサンプルに対して悪影響を受けない学習を実現可能であることを述べた。しかしながら、現場環境への適応の際に再度識別器を学習する計算コストは学習用画像の枚数と特徴量の次元数に比例して大きくなり、この計算コストについても実用化の障害であるといえる。そこで、6章では転移学習による高速な再学習を導入して計算コストを効率化する。

第6章

実利用シーンに対する再学習の効率化

5章では、統計的学習法に基づく人検出において、学習サンプル収集に伴う人的コストを削減し、特定環境への適応を自動化する手法について述べた。しかしながら、再学習を行う際に学習に要する時間コストが大きな問題として残る。学習サンプルの追加と特徴空間を少数に抑えて省力化する手法として Boosting に基づく転移学習が提案されている。しかし、事前学習シーンと特定シーンの間に大きな変化がある場合、転移学習を用いて適応するのは困難である。そこで本章では、転移学習の高速性と再学習の高精度を両立するハイブリッド型転移学習に取り組む。6.1.1において転移学習について述べた後、6.1.3においてその問題点について述べる。6.2では学習効率に基づいて特徴空間を選択的に切り替えるハイブリッド型転移学習を提案する。これにより、追加する学習サンプル数の削減と学習の高速化による効率化が期待できる。

6.1 物体検出における転移学習

5章で述べた生成型学習では、CGを用いて学習サンプルの自動収集が可能である。これにより学習サンプル収集に要する人的コストの削減を実現したが、再学習を行う場合、識別器の学習に時間を要することが問題として残されている。識別器の学習時間は学習サンプル数と特徴次元の数に比例する。特に、生成型学習で様々な視点の学習サンプルを大量に生成した場合、学習時間を要することになる。設置環境などの特定シーンに対して短時間で学習を行う手法として、事前に類似した問題に対して構築された識別器を応用する転移学習が提案されている。本節では、転移学習の従来法とその問題点について述べる。

6.1.1 転移学習

転移学習とは機械学習の分野で利用される学習手法の一つである。転移学習という言葉には様々な解釈があるが、機械学習の分野においてはNIPS 2005 workshop - inductive transfer: 10 years laterの論文募集要項で、「新規タスクの効果的な仮説を効率的に見つけ出すために、一つ以上の別のタスクで学習された知識を得て、それを適用する問題」と定義されている。この定義に従う転移学習は、事前ドメインとよばれる事前に学習した識別器や学習サンプルを、目的ドメインと呼ばれる新たな環境に対する識別器の構築に利用することで効率化を目指している。転移学習による効率化として、人検出における俯角の変化への対応 [36] や新規サンプルの追加 [37] などが提案されている。転移学習を用いた研究は2010年代以降増加しており、SOINN[37] や TRAdaBoost[38], [39], CovBoost[36] などの手法が提案されている。以下では、共変量シフトに基づく転移学習を Boosting に導入した Covariate-shift Boost (CovBoost) を転移学習について述べる。

6.1.2 Covariate-shift Boost(CovBoost) による転移学習

Covariate-shift Boost(CovBoost) とは、事前学習によって学習した弱識別器の情報と事前学習に用いた学習サンプルを目的とする特定シーンの学習に用いることで、特定シーンにおける学習を効率化する Boosting 手法である。特定シーンでの学習サンプルが少量でも、事前シーンの大量の学習サンプルを用い、大量のデータを用いて全特徴量を探索した時と同程度の検出精度を維持することが可能となる。これにより、事前シーンの学習サンプルと検出対象である特定シーンの学習サンプルに見えの変化がある場合の再学習の省力化を実現した [36]。その後、半教師型オンライン学習による人検出に展開されている [40]。CovBoost の大きな特徴は特徴の転移と、共変量を用いた転移学習アルゴリズムの2点である。

特徴の転移

CovBoost は、学習の前準備として一度学習し、識別器として採用された有効な特徴量から、目標学習にも有効な特徴量を再生成する。これを特徴の転移という。特徴の転移は、事前学習によって選択

された特徴量を目的ドメインに転移させることで行われる。文献 [36] では、Haar-like 特徴量の転移を対象としている。学習により選択された Haar-like 特徴量の領域を示す座標 4 点を乱数を用いて変動し、特徴の転移候補として生成する。これを再度学習することで有効に転移した領域が選択される。有効である可能性が高い領域から生成するため、通常の学習と比べて高速となる。

共変量を用いた転移学習

Boosting 手法は、一般的に式 (6.1) に示す損失関数 L を最小化する強識別器 $H(\mathbf{x})$ を求めることが目標となる。ここで、 Ω は学習サンプルの総数、 \mathbf{x} と y はそれぞれ学習サンプルとクラスラベルを示す。

$$L = \sum_{\Omega} e^{-yH(\mathbf{x})} \quad (6.1)$$

CovBoost では入力する全学習サンプル集合 Ω に、事前学習で用いた学習サンプル (事前ドメイン) と再学習に用いる学習サンプル (目標ドメイン) の二種類が存在する。特徴量により観測される事前ドメインの確率密度分布 $p_a(\mathbf{x})$ は、目標ドメインの確率密度分布 $p_t(\mathbf{x})$ と異なる分布を持つ可能性がある。すなわち、 $p_a(y|\mathbf{x}) \neq p_t(y|\mathbf{x})$ であると考えられる。ここで、事前ドメインの学習サンプルのうち、目標ドメインのネガティブサンプルに類似したポジティブサンプルなど目標ドメインの学習を阻害するサンプルを、学習に利用しないことが必要となる。そこで、 $\frac{p_t(y|\mathbf{x})}{p_a(y|\mathbf{x})}$ として事前ドメインのサンプルに重みを与えることで目標ドメインの学習に有効なサンプルを取捨選択する。ここで $\frac{p_t(y|\mathbf{x})}{p_a(y|\mathbf{x})}$ は共変量と呼ばれ、共変量を用いて事前ドメインのサンプルを目標ドメインに挿入することを共変量シフトによる転移学習と呼ぶ。また、共変量 $\frac{p_t(y|\mathbf{x})}{p_a(y|\mathbf{x})}$ は記号 λ で表す。共変量シフトを用いた転移学習の目的関数を式 (6.2) に示す。ここで、 $(\mathbf{x}_i, y_i) \in T_t$ と $(\mathbf{x}_j, y_j) \in T_a$ はそれぞれ目標ドメイン t と事前ドメイン a の学習サンプルの特徴ベクトル \mathbf{x} とそのクラスラベル y を示す。

$$\tilde{L} = \sum_{(\mathbf{x}_i, y_i) \in T_t} e^{-y_i H_t(\mathbf{x}_i)} + \sum_{(\mathbf{x}_j, y_j) \in T_a} \lambda_j e^{-y_j H_t(\mathbf{x}_j)} \quad (6.2)$$

共変量 λ は、ベイズ推定を用いて次のように変形することができる。

$$\begin{aligned} \lambda &= \frac{p_t(\mathbf{x}, y)}{p_a(\mathbf{x}, y)} \\ &= \frac{p(\mathbf{x}, y|t)}{p(\mathbf{x}, y|a)} \\ &= \frac{p(t|\mathbf{x}, y)p(\mathbf{x}, y)}{p(t)} \frac{1}{\frac{p(a|\mathbf{x}, y)p(\mathbf{x}, y)}{p(a)}} \\ &= \frac{p(t|\mathbf{x}, y)p(a)}{p(a|\mathbf{x}, y)p(t)} \end{aligned} \quad (6.3)$$

$p(a) = p(t)$ と仮定すると、式 (6.3) は目標ドメインの条件付き確率と事前ドメインの条件付き確率の比となる。ここで、条件付き確率を次式のようなロジスティック関数に変形する。

$$p(t|\mathbf{x}, y) = \frac{1}{1 + e^{yH_t(\mathbf{x})}} \quad (6.4)$$

$$p(a|\mathbf{x}, y) = \frac{1}{1 + e^{yH_a(\mathbf{x})}} \quad (6.5)$$

ここで $H_a(\mathbf{x})$ とは、事前学習によって構築された強識別器を示し、 H_t は事前ドメインと目標ドメインの両方を使って学習された強識別器を示す。この式 (6.4) と式 (6.5) から λ は式 (6.6) のように変形できる。

$$\lambda = \frac{1 + e^{-yH_a(\mathbf{x})}}{1 + e^{-yH_t(\mathbf{x})}} \quad (6.6)$$

λ は事前ドメインのサンプルが目標ドメインにどれだけ適合しているかを識別器 H_a , H_t により表現し、目標ドメインに適合しているほど大きな値となる。

6.1.3 転移学習の問題点

Pang 等は、視点の異なる環境において CovBoost により識別器を転移することで、新規に収集する学習サンプルを 3 分の 1 に削減しても同等の性能を実現した。しかし、カメラの俯角が大きく変化する場合、図 6.1 に示すように画像上の見えが大きく変わるため、特徴の転移が困難となり、識別性能の低下を招く。図 6.2 は、カメラ俯角を $30^\circ \sim 50^\circ$ に変化した際の転移学習による識別性能を DET により示したものである。事前学習には俯角 20° の学習サンプルを、特徴量には HOG を用いた。図 6.2 より、俯角を大きく変化すると性能が低下していることがわかる。これは図 6.1 の 50° のようにサンプルの見えが大きく変化すると、転移した特徴のみでは目標ドメインにうまく適合できないことを示している。

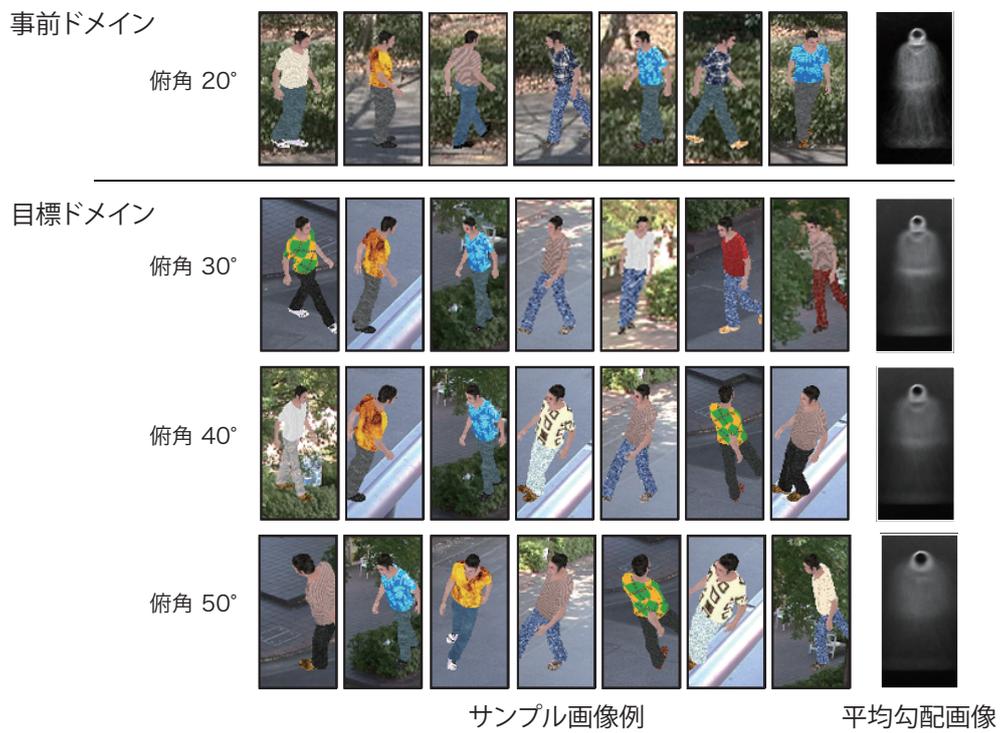


図 6.1: 俯角による学習サンプルの変化.

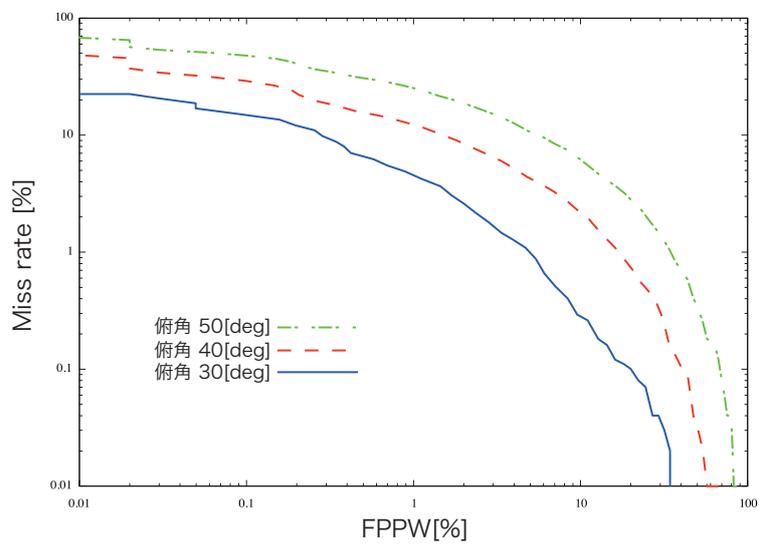


図 6.2: 俯角の変化による転移学習の性能.

6.2 ハイブリッド型転移学習による学習の効率化

転移学習を用いることで少量の目標学習サンプル収集で高精度な識別が可能となるが、事前学習と目標学習でのシーンが大きく異なる場合は適応できない。これは、シーンの大きな変化により特徴量の転移が困難となるためである。そこで本研究では、図 6.3 に示すように転移により得られる高速に探索可能な特徴量と、探索に時間を要する全ての特徴量をそれぞれ特徴空間として用意し、学習効率に基づいて転移特徴空間と全特徴空間を選択的に切り替えるハイブリッド型転移学習を提案する。事前に用意した大規模なデータセットを用いて事前学習を行い、事前識別器 H_a を得る。この H_a から特徴の転移によって得られる特徴空間を転移特徴空間 F_{Tr} とし、共変量を用いて目標ドメインと事前ドメインの両学習サンプルから学習を行い、目標識別器 H_t を構築する。これを転移学習ステップとする (図 6.3(a))。この際、学習サンプルに対するエラー率から学習効率を算出する。そして、その効率が低ければ、転移特徴のみで十分な成果が得られないとして、全探索特徴空間 F_{Re} より学習を行う再学習ステップ (図 6.3(b)) へ移行する。このようにハイブリッド型転移学習を定義し、再学習より高速かつ、従来の転移学習より高精度な識別器を構築することを目的とする。

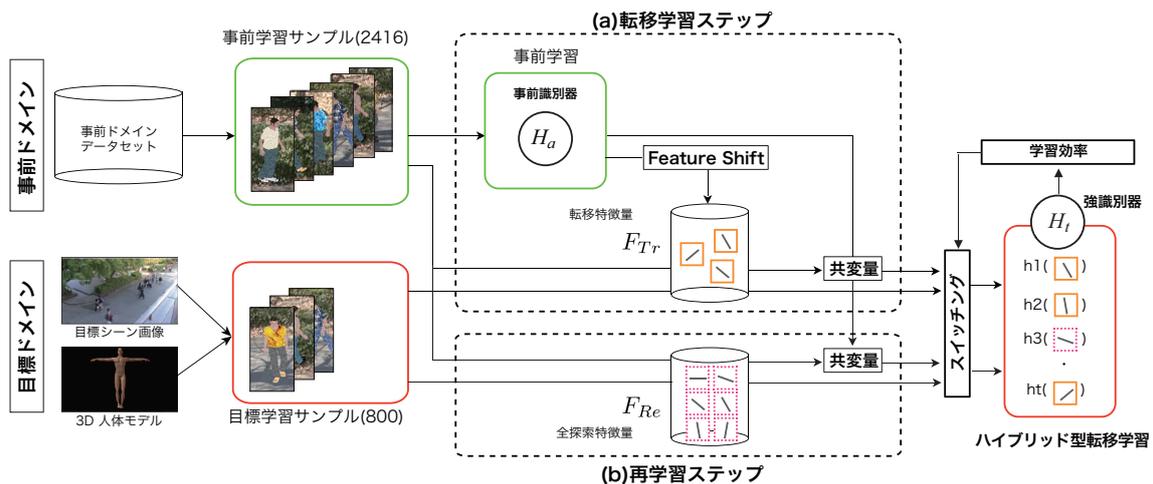


図 6.3: ハイブリッド型転移学習.

6.2.1 本研究の問題設定

本研究では、俯角の少ない人画像データの学習を事前ドメインとし、実際の設置環境である俯角の大きい特定シーンのデータを目標ドメインと定義する。

事前ドメイン

事前ドメインの学習はオフライン処理が可能であるため、大量のデータを使用する。学習サンプル

のポジティブサンプルは5章で述べた, CGで生成した人画像を用いるため, サンプル収集のコストはないものとする. 事前ドメインとしては俯角 20° の人画像 2,416 枚を使用し, HOG 特徴量による AdaBoost で学習した識別器を使用する.

- ポジティブサンプル数: 2,416 枚
- ネガティブサンプル数: 12,180 枚
- シーン: 俯角 20°

目標ドメイン

目標ドメインは, カメラの俯角の異なる特定シーンを使用する. 事前ドメインのデータと異なり, 新規に収集する必要がある点から, サンプル数が少ないことが望ましい. 目標ドメインとして, カメラ俯角を 30° , 40° , 50° として各俯角を 800 枚生成した 3 種類を用意する.

- ポジティブサンプル数: 800 枚
- ネガティブサンプル数: 12,180 枚
- シーン: 俯角 30° , 40° , 50°

本研究では, 上記に定義した問題に対して転移学習を行う. 再学習に比べ, 少数 (文献 [36] より事前ドメインの $1/3$) の目標ドメインサンプルを用いて, 精度を維持しながら高速に識別器を学習することを本研究の問題設定とする.

6.2.2 HOG 特徴量における Feature Shift

文献 [36] の CovBoost では Haar-like 特徴量の転移を行う. 本研究では, 人検出に用いられる HOG 特徴量に対して特徴の転移を行う. まず, 図 6.4(1) のように, 事前学習で選択された弱識別器の局所特徴量の中心座標を求める. この座標を中心に, 図 6.4(2) のように正規乱数に従い K 個の候補領域を発生させる. 文献 [36] では, $K = 50$ を妥当としている. ここで, 提案手法では, 候補領域から局所特徴量のヒストグラムを求め, 図 6.4(3) のように事前学習で選択された弱識別器の局所特徴量のヒストグラムと比較して両ヒストグラムの類似度を求める. ヒストグラムの類似度算出には次式に示すバタチャリヤ係数を用いる.

$$Bhattacharyya = \sum_{i=1}^n \sqrt{p(\mathbf{x})q(\mathbf{x})} \quad (6.7)$$

ここで $p(\mathbf{x})$ と $q(\mathbf{x})$ はそれぞれ異なるドメインの確率密度関数である. 最後に, 事前学習で選択された弱識別器と最も高い類似度を持つ転移候補を転移先の特徴量とし, その集合を転移特徴空間 F_{Tr} と定義する. これに対し, 再学習と同様に画像から全特徴量を抽出したものを全特徴空間 F_{Re} と定義する. 表 6.1 に各俯角ごとの類似度の平均値を示す. 6.1.3 で述べたように, 事前ドメインの学習データから離れるほど類似度は低くなる. それらの差が図 6.2 のように識別性能を低下させる一因であるといえる.

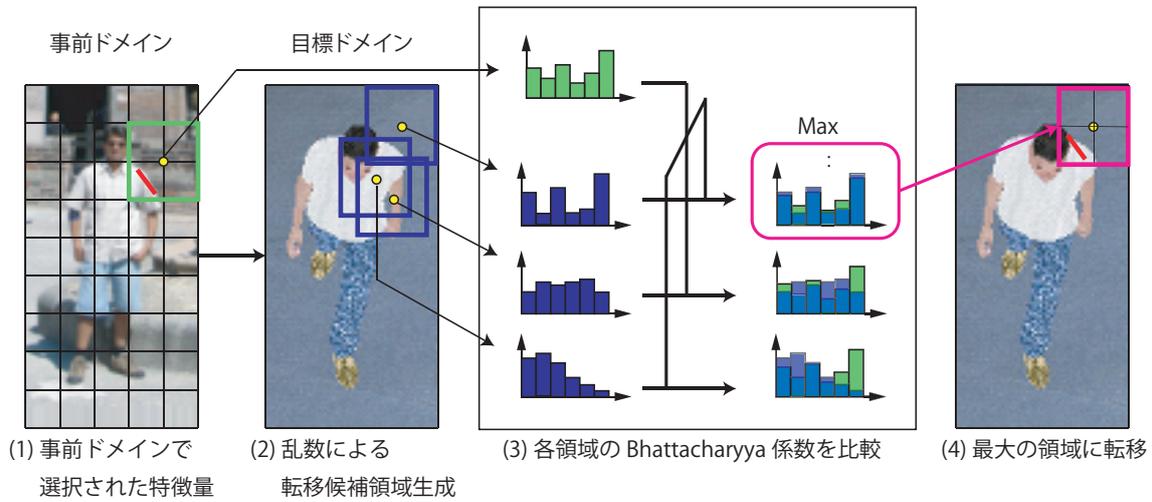


図 6.4: HOG 特徴量の転移.

表 6.1: 俯角による類似度の変化.

カメラ俯角 [°]	30	40	50
Bhattacharyya 係数	0.981	0.972	0.968

6.2.3 ハイブリッド型転移学習による学習

提案するハイブリッド型学習では、事前ドメイン T_a と目標ドメイン T_t の異なるシーンから切り出したサンプル群を用いる。目標ドメインの各サンプル \mathbf{x}_i と事前ドメインの各サンプル \mathbf{x}_j はクラスラベル y_i または y_j を持ち、ポジティブサンプルには +1, ネガティブサンプルには -1 を設定する。次に、サンプルの重みを初期化する。サンプルの重みは目標ドメインと事前ドメインのそれぞれで正規化したものを初期値とし、それぞれの重みを $D_t(\mathbf{x}_i)$ と $D_a(\mathbf{x}_j)$ と表現する。弱識別器の選択は、式 (6.2) を最小化するような $h(\mathbf{x})$ を求めることである。ここで、式 (6.2) を $h(\mathbf{x}) = 0$ の一次のテイラー展開により式 (6.8) の近似式を求める。

$$h_t(\mathbf{x}) = \arg \min_{h_t} \left(\sum_{(\mathbf{x}_i, y_i) \in T_t} e^{-2y_i D_t(\mathbf{x}_i)} y_i h_t(\mathbf{x}_i) + \sum_{(\mathbf{x}_j, y_j) \in T_a} \lambda_j e^{-2y_j D_a(\mathbf{x}_j)} y_j h_t(\mathbf{x}_j) \right) \quad (6.8)$$

この式を最小にする $h(\mathbf{x})$ を求めることで弱識別器 $h_t(\mathbf{x})$ を選択する。各弱識別器 $h(\mathbf{x})$ は、転移特徴空間 F_{Tr} と全特徴空間 F_{Re} を探索して求める。まず転移特徴空間 F_{Tr} を探索し、得られた弱識別器のエラー率から学習効率 ζ を算出する。その値がしきい値以下のとき、全特徴空間 F_{Re} において弱識別器の再選択が行われる。 ζ の算出法については次節で示す。次に、式 (6.9) でエラー率 ϵ_m を

算出する.

$$\epsilon_m = \frac{\sum_{h(\mathbf{x}_i) \neq y_i} e^{-2y_i D_t(\mathbf{x}_i)} + \sum_{h(\mathbf{x}_j) \neq y_j} \lambda_j e^{-2y_j D_a(\mathbf{x}_j)}}{\sum_i e^{-2y_i D_t(\mathbf{x}_i)} + \sum_j \lambda_j e^{-2y_j D_a(\mathbf{x}_j)}} \quad (6.9)$$

算出したエラー率から, 選択した弱識別器に対する重み α_m を次式により算出する.

$$\alpha_m = \frac{1}{4} \ln \frac{1 - \epsilon_m}{\epsilon_m} \quad (6.10)$$

最後に, 目標ドメイン及び事前ドメインの学習サンプルの重み $D_t()$ と $D_a()$ を次式により更新する.

$$D_t(\mathbf{x}_i) = D_t(\mathbf{x}_i) e^{-2y_i \alpha_t h_m(\mathbf{x}_i)} \quad (6.11)$$

$$D_a(\mathbf{x}_j) = D_a(\mathbf{x}_j) e^{-2y_j \alpha_t h_m(\mathbf{x}_j)} \quad (6.12)$$

以上の処理を, 事前学習の学習回数と同数繰り返して, 新たな弱識別器 $h_t(\mathbf{x})$ を求める. 識別時は, 通常の AdaBoost と同様に, 次式に示すような弱識別器の重み付きの多数決を求めることにより識別を行う. 強識別器を式 (6.13) に示す.

$$H_t(\mathbf{x}) = \text{sign} \left[\sum_{t=1}^M \alpha_t h_t(\mathbf{x}) - th \right] \quad (6.13)$$

ここで th はしきい値であり, M は事前学習の学習回数である.

6.2.4 学習効率 ζ に基づく特徴空間選択

前節で述べた特徴転移では, 転移尤度の高い特徴量の転移特徴空間 F_{Tr} において転移学習を行う. これにより学習時間の削減 (探索コストの低下) が可能となるが, 事前学習データと目標学習データに大きな相違がある場合, 共通な特徴が観測されないことが考えられる. そこで, 提案手法では転移特徴空間と, 再学習同様の全特徴空間を選択的に切り替える. 転移が有効な場合には高速な転移学習を行い, 転移が困難な場合には全特徴空間を用いて学習を行う. 本研究では, それぞれの特徴空間を以下のように定義する.

転移特徴空間

- 特徴次元: 100 (事前学習により選択)
- 弱識別器のしきい値探索: 必要なし
- 計算コスト: 低

- ドメイン間の差が大きいと性能低下

全特徴空間

- 特徴次元：3,780
- 弱識別器のしきい値探索：各次元ごとに 100 段階
- 計算コスト：高
- 目標ドメインに最適化

ここで必要となる切り替え指標は、転移学習で補えるか否かを判定する値である。図 6.5(a) に転移学習時と再学習時の Boosting における弱識別器のエラー率 ϵ (式 (6.9)) の推移を示す。

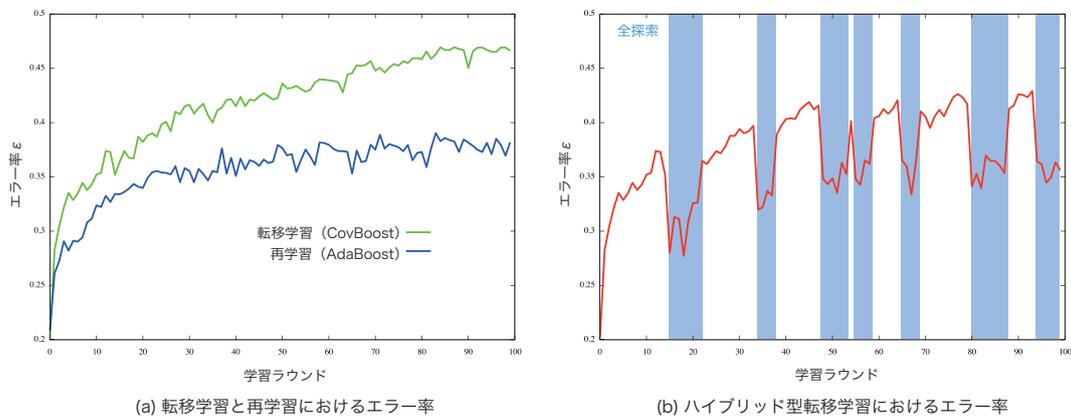


図 6.5: 弱識別器のエラー率の推移 ϵ .

弱識別器のエラー率は転移学習，再学習共に学習が進むにつれ高くなる。これは，学習が進行するにつれ適応困難なサンプルの重みの値が高くなるためである。この値の変化量が小さくなると，ある程度学習は収束し，大きな改善はないと考える。そこで，本研究ではエラー率 ϵ の勾配を学習効率 ζ として定義し，切り替えの指標として利用する。このエラー率の傾きを観測し，転移学習が進行して傾きが緩やかになり，その絶対値がしきい値を下回る際に全探索を適用する。なお，勾配の算出には直近 5 点の最小二乗近似を用いる。全探索により有効な特徴を発見できれば大幅に ϵ が下がるため勾配が拡大し，再度転移学習へと移行する。学習効率を用いた切り替えを行った際のエラー率の推移を 6.5(b) に示す。図より，全探索へ移行する度に大幅にエラー率が改善され，十分に ϵ が下がると転移学習に再度移行することがわかる。また，学習の終了条件は学習サンプルに対する誤識別率の収束を用いる。一般には誤識別率に加えエラー率の収束も用いられるが，本手法ではエラー率が大きく振動的な振る舞いをするため終了条件としては用いない。

6.3 評価実験

提案手法の有効性を示すために、識別精度と学習速度の観点から検証を行う。

6.3.1 実験概要

HOG 特徴量と AdaBoost による識別器を事前学習とする。HOG 特徴量のセルサイズは 8、ブロックサイズは 2 とすると計 3,780 次元となる。転移時に生成する特徴候補数は 1 特徴量あたり 50 とする。事前学習サンプルのポジティブサンプルとしては、図 6.1 で示した俯角 20° の CG を用いて生成した人画像を 2,416 枚使用し、ネガティブサンプルとしては、ポジティブサンプルを生成したシーンの背景画像から 12,180 枚切り出して使用する。目標学習のポジティブサンプルとしては、図 6.1 でそれぞれ示した俯角 30~50° の CG を用いて生成した人画像を 800 枚用い、ネガティブサンプルとして、事前ドメイン同様に切り出した背景画像 12,180 枚を用いる。ただし、比較手法である再学習については CG を用いて生成した人画像を 2,416 枚使用し、事前学習を行わない。評価用サンプルはそれぞれ、ポジティブサンプルには特定シーンの俯角に設定した 3D 人体モデルによる CG 生成画像 10,000 枚を、ネガティブサンプルには生成に用いた背景画像から 10,000 枚切り出したものを使用する。性能評価は EER (Equal Error Rate) により行う。

6.3.2 俯角変化に対するハイブリッド型転移学習の精度

表 6.2 に、カメラ俯角を変化させた 3 種の目標シーンに対する各手法の識別性能を示す。

表 6.2: EER による性能評価。

カメラ俯角 [°]	提案手法 [%]	転移学習 [%]	再学習 [%]
30	2.24	2.61	2.10
40	2.52	4.78	2.39
50	2.36	4.23	2.35

表より、従来の転移学習では、シーンの変化とともに性能が大きく低下することがわかる。一方、提案手法は大きなシーン変化が生じてても、再学習と同等であることがわかる。これは、転移特徴量のみでは対応できなかった点を、再学習と同等の全探索特徴により改善できたといえる。

6.3.3 考察

図 6.6 は提案手法で選択された特徴量のうち、転移学習ステップで選ばれたものを (a) 転移特徴量、再学習ステップで選ばれたものを (b) 全探索として可視化したものである。図 6.6(a) より、転移特

微では標準的な肩のエッジや脚部の縦方向のエッジなどが転移できていることがわかる。一方、図 6.6(b) の全探索では横エッジが目立ち、俯角の変化により発生した上半身の見えの変化に適応した特徴が選択されている。図 6.6(c) に提案手法全体として (a) と (b) を重ねたものを、(d) に再学習で選択された特徴量を表示する。両者を比較すると、特徴の位置関係や勾配方向が類似していることから、(c) の提案手法は転移特徴と全探索の組み合わせにより (d) の再学習に近い特徴の構成を獲得することで高精度な転移学習を実現したといえる。

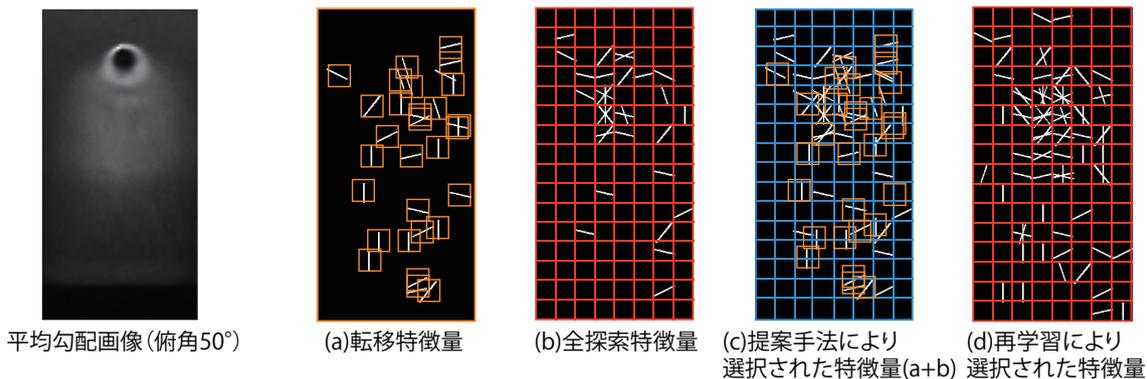


図 6.6: 選択された HOG 特徴量.

6.3.4 ハイブリッド型転移学習の高速性

提案手法は転移学習と同様の少サンプルで、再学習に近い精度を発揮する。しかし、目標シーンへの適応のための再学習にはサンプルの収集コストの他に学習に要する時間コストの問題も存在する。そこで、提案手法と従来の転移学習、再学習に要する時間の比較を表 6.3 に示す。

表 6.3: 目標識別器学習に必要な計算時間の比較.

カメラ俯角 [°]	提案手法 [min]	転移学習 [min]	再学習 [min]
30	12.1	0.58	60
40	15.2	0.58	60
50	13.6	0.58	60

表 6.3 より提案手法は、従来の転移学習に比べ 20 倍以上の時間を必要とする。これは、全探索ステップ 1 回が転移学習ステップ 1 回に比べ大きく処理時間を必要とするためである。しかし、表 6.2 で示した通り提案手法は従来法より高精度である。一方、再学習に対しては、提案手法では転移学習法における特徴の転移により、同等精度のまま 4~5 倍の高速化を実現している。これにより、さ

らなるデータ追加やカスケード構造など、長時間の学習を必要とする問題に、本手法を適用することで効率化が期待できる。

6.4 まとめ

本章では、まず、効率的な再学習として転移学習の導入とその有効性を述べた後、転移学習には、事前ドメインと目標ドメインの相違が大きい場合に深刻な性能の低下を招くという問題が存在することを示した。

次に、この問題に対し、転移学習において特徴を転移させた転移特徴空間と、再学習と同様の全特徴空間の2つの特徴空間を設定し、適応的に特徴空間を選択するハイブリッド型転移学習法を提案した。弱識別器のエラー率 ϵ の勾配に基づく学習効率に従った空間選択により、従来の転移学習より同等 $\sim 2.26\%$ 性能を向上しつつ、従来の再学習に比べ約3.9倍高速な学習が可能となる。これにより、高速性と高精度を両立した効率的な学習を実現した。また、転移学習で得られた特徴量と提案手法の全探索で得られた特徴量を組み合わせて可視化したところ、再学習により得られる特徴量と非常に類似した勾配が得られた。これは、転移した特徴量のみでは大きな変化に対応できないことと、その問題を全探索で選択した特徴量で補うことが可能であることを示している。この点から、転移学習では異なるドメイン間の転移に失敗するような困難な問題に対して、提案手法は全探索とのハイブリッド戦略を導入することで再学習と同様の改善を得ることができたといえる。

第7章

転移学習の導入による Random Forest の効率化

本章では、転移学習をマルチクラス分類問題へ展開する。ここでは、マルチクラス識別器として様々な応用に利用されている Random Forest を対象とし、転移学習の導入法について述べる。まず、Random Forest の概要について述べた後、転移学習の Random Forest への導入について述べる。Random Forest において転移学習を実現することで、マルチクラスの物体検出問題においても効率的な識別器の構築が期待できる。

7.1 マルチクラス分類における転移学習

5章と6章で示した学習サンプルの生成や転移学習を用いることで、2クラス識別器の再学習を大きく効率化することができた。しかし、2000年代以降の物体検出では1クラスのみでなく、画像中に存在する複数の物体クラスを認識するマルチクラス分類問題も重要なタスクとされている [41][42]。マルチクラス分類問題で注目されている識別器に Random Forest がある。Random Forest は高い汎化性能を持つ上に、決定木の集合体であることから並列処理による高速な識別が可能である。しかし、Random Forest もまた統計的学習法であるため、その識別性能は学習サンプルに依存し、設置環境の変化に対する再学習は避けられない。そこで、Random Forest に転移学習を導入することで、マルチクラス物体検出に対しても識別器の構築における効率化が可能となる。本章では、マルチクラス分類問題における設置環境への対応の効率化について述べる。

7.2 Random Forest

Random Forest [43] とは、2001年に Breiman によって提案された複数の決定木構造を持ったマルチクラス識別器を構築するアンサンブル学習アルゴリズムである。図 7.1 に Random Forest の構造を示す。このアルゴリズムの特徴は、Bagging [44] と同様にブートストラップを取り入れ過学習を防ぐことが可能な点や、Random Feature Selection を取り入れることで特徴ベクトルの次元数が大きくても高速に学習が可能である点である。このようなメリットを持ち、クラス識別、クラスタリング、回帰分析に利用することができるため、様々な分野に応用されている。

コンピュータビジョンの分野でも、文字認識 [45]、物体認識 [46]、物体セグメンテーション [47]、人体姿勢推定 [48]、医用画像処理 [49] で用いられており、それぞれの研究課題のブレイクスルーを実現するフレキシブルな学習アルゴリズムとして利用されている。

7.2.1 Random Forest の学習アルゴリズム

Random Forest は、学習サンプルを用いて図 7.1 に示すような複数の決定木構造を持つ識別器を構築する。各決定木は、分岐ノード (Split Node) と末端ノード (Leaf Node) により構成され、分岐ノードを繰り返し作成し、一定の基準により分岐が不可能になった際に、末端ノードを作成することにより構築される。分岐ノードは、ある 1次元の特徴量としきい値の大小関係により左右に分岐する。また、末端ノードは辿り着いた学習サンプルの各クラス確率を保存する。これらの構造を持つ識別器の構築アルゴリズムについて以下に述べる。

Random Forest の学習アルゴリズムを **アルゴリズム 7.1** に示す。まず、学習サンプルを I 、木の数 T 、木の深さ D 、サブセット $I = (I_1, I_2, \dots, I_T)$ 特徴選択回数 F 、しきい値選択回数 TH を定義する。 F, TH は分岐ノードを作成するためのパラメータである。サブセットは学習サンプルからランダムで選択した小さなデータ集合であり、サブセット間のデータのオーバーラップを許容して作成される。これらのサブセットを利用することにより決定木を構築する。

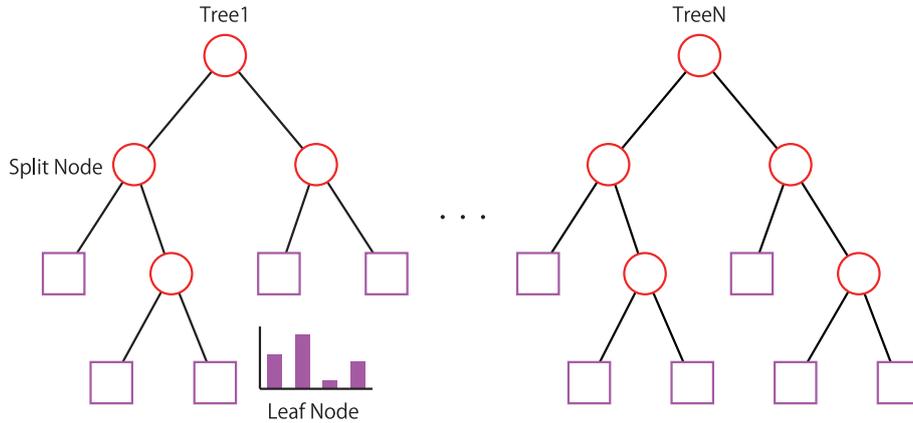


図 7.1: Random Forest の構造.

決定木の構築は、分岐が不可能になるまで分岐ノードを繰り返し作成する。分岐ノードはある特徴量 f としきい値 t を用いて左もしくは右への分岐を決めるものである。そのため、 f と t は様々な候補の中から厳選する必要がある。しかし、これらの候補を全探索すると特徴次元が大きくなるに従い学習に時間を要するという問題が発生する。この問題を解決し、高速な学習を実現するために Random Feature Selection を導入する。予め定めた特徴選択回数 F としきい値選択回数 TH を用いて、分岐ノードの候補を $F \times TH$ の組み合わせだけ用意する。この候補の中のみから最適なものを探すことにより、高速に学習を行う。最適な分岐ノードを探す基準に情報利得 ΔE を利用し、 ΔE が最も大きくなる組み合わせを分岐ノードとして用いる。ある分岐ノード n に分かれてきたデータ集合 I_n 、分岐ノード組み合わせ f と t を用いて式 (7.1)、式 (7.2) により I_l と I_r を算出する。

$$I_l = \{i \in I_n | f(\mathbf{v}_i) < t\} \quad (7.1)$$

$$I_r = I_n \setminus I_l \quad (7.2)$$

この I_l と I_r を用いて、式 (7.3) により情報利得 ΔE を算出する。

$$\Delta E = E(I) - \frac{|I_l|}{|I_n|} E(I_l) - \frac{|I_r|}{|I_n|} E(I_r) \quad (7.3)$$

ここで、関数 $E(I)$ は式 (7.4)(情報エントロピー)、または式 (7.5)(ジニ係数) である。

$$E(I) = - \sum_{i=1}^n p(c_i) \log p(c_i) \quad (7.4)$$

$$E(I) = \sum_{i=1}^n p(c_i)(1 - p(c_i)) \quad (7.5)$$

ここで、 $p(c_i)$ はクラス c_i の確率 (相対頻度) を表しており、学習サンプルにつけられている教師信号により求められる。これらの処理を繰り返すことにより分岐ノードを作成してゆき、情報利得が 0

アルゴリズム 7.1 : Random Forest の学習アルゴリズム

1. 入力: 学習サンプル : I を用意する.

パラメータ: 木の数 : T , 木の深さ: D

2. サブセット作成: T 個のサブセット : $I = (I_1, I_2, \dots, I_T)$

For $k = 1, \dots$, 木の数 T まで

サブセット I_k を用いて決定木を作成

For $l = 1, \dots$, 特徴量選択数 F まで

分岐関数 f をランダムに選択

For $m = 1, \dots$, しきい値選択数 TH まで

・ しきい値 t をランダムに選択

・ f と t を用いてデータを I_l と I_r に分岐

$$I_l = \{i \in I_n | f(\mathbf{v}_i) < t\}$$

$$I_r = I_n \setminus I_l$$

・ 教師信号を利用して情報利得 ΔE を算出

$$\Delta E = -\frac{|I_l|}{|I_n|} E(I_l) - \frac{|I_r|}{|I_n|} E(I_r)$$

・ if $\Delta E > \Delta E_{old}$ f, t, I_l, I_r を保存

End for

End for

If $gain = 0$ または D に達した \rightarrow 末端ノードを作成

教師信号を用いて $P(c|l)$ を計算

・ Else I_l, I_r を用いて再帰処理

End for

3. 出力: 識別器

未知サンプル \mathbf{x} が到達したノード l よりクラス c の確率を算出

$$P(c|\mathbf{x}) = \frac{1}{T} \sum_{t=1}^T P_t(c|l)$$

クラス確率が最大のクラスに属すると推定

$$C_i^* = \arg \max_{c_i} P(c_i|\mathbf{x})$$

になった場合に末端ノード l を作成し, 分岐されてきたデータを用いて確率 $P(c|l)$ を計算する. これにより, 各決定木の学習が可能となる.

7.2.2 Random Forest の識別処理

Random Forest の識別処理の流れを、図 7.2 に示す。未知データを \boldsymbol{x} とした時に各決定木に入力し、たどり着いた末端ノードに保存されている $P(c|l)$ を出力する。各決定木から出力された $(P_1(c|l), P_2(c|l), \dots, P_T(c|l))$ を用いて以下の式により未知データの事後確率 $P(c|\boldsymbol{x})$ を算出する。

$$P(c|\boldsymbol{x}) = \frac{1}{T} \sum_{t=1}^T P_t(c|l) \quad (7.6)$$

これを用いて次式の単純ベイズ識別器により未知データのクラスを推定する。

$$C_i^* = \arg \max_{c_i} P(c_i|\boldsymbol{x}) \quad (7.7)$$

以上の処理によりマルチクラス識別が可能となる。

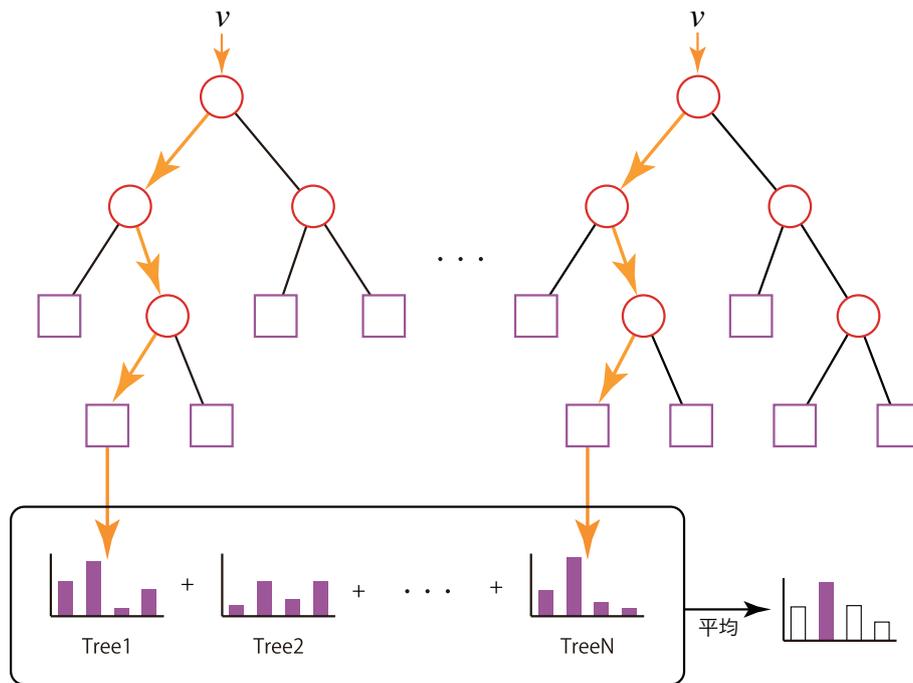


図 7.2: Random Forest による識別。

7.3 Random Forest の転移学習

Random Forest はアンサンブル学習を容易に実現できることから、様々な問題に展開されている。しかしながら、統計的学習法である以上、学習環境と設置環境が大きく異なる際の性能低下を避けられず、再学習を必要とする。この再学習に転移学習を導入することで識別器の構築における効率化を実現する。本節では共変量シフトに基づく転移学習を導入した Random Forest のアルゴリズムについて述べる。

7.3.1 共変量シフトに基づく転移学習を導入した Random Forest

共変量シフトに基づく転移学習を導入した Random Forest の学習の流れを図 7.3 に示す。

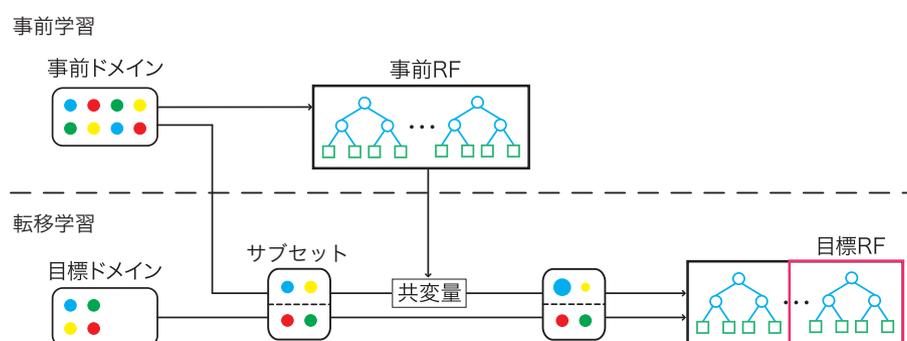


図 7.3: 共変量シフトに基づく転移学習を導入した Random Forest.

転移学習の前提として、十分な事前ドメインの学習サンプルが存在し、その学習サンプルを用いた Random Forest は構築済みであるとする。この事前ドメイン $X_a = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_j\}$ と事前に構築した Random Forest (事前 RF) を、少数用意した目的ドメインの学習サンプル $X_t = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i\}$ の学習に利用する。この際、事前ドメインサンプルは共変量シフトにより目標ドメインの学習への有効性で重み付けされ、有効でない事前ドメインサンプルは、転移学習ではその影響を失う。このとき共変量は、事前 RF と現在構築している目標 Random Forest (目標 RF) の出力の比により決定する。

決定木の転移学習は、図 7.4 に示すように以下の 3 つのステップで行われる。

- 1) 事前ドメイン X_a 、目標ドメイン X_t から同数のサンプルを抽出し、サブセットを作成する。
- 2) サブセットに対し決定木を学習する。ここで、サブセット内の X_a に属するサンプルは共変量 λ で重み付けされている。
- 3) 構築した決定木を目標 Random Forest に加え更新する。更新された目標 Random Forest と事前学習した Random Forest により、共変量 λ を更新する。

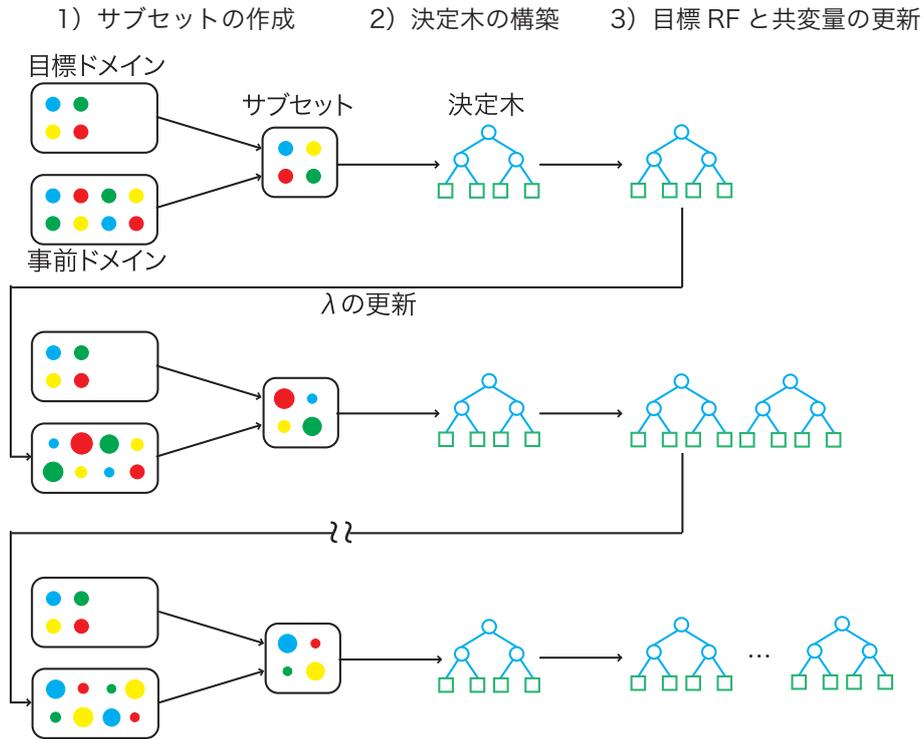


図 7.4: 転移学習を導入した Random Forest の構築.

以上のステップを繰り返し、十分な数の決定木となるまで目標ドメインに対する Random Forest を構築する。通常の Random Forest との違いは、入力サンプルが部分的に重み付けられている点と、その重みを更新する点である。ここで、各決定木に与えられるサブセット I は、 X_a と X_t から同数サンプルを採用することで作成される。これにより、事前ドメインと目標ドメインの数量差を無視できる。共変量 λ は 6 章で述べた式 (6.6) の、Boosting による識別器の出力 H を、式 (7.6) で示した Random Forest の出力 $P(c|\mathbf{x})$ に置き換えた次式で求める。

$$\lambda_j = \frac{1 + e^{P_a(c_k|\mathcal{X}_j)}}{1 + e^{P_t(c_k|\mathcal{X}_j)}} \quad (7.8)$$

ここで $P_a(c_k|\mathcal{X}_j)$ は j 番目の事前ドメインサンプル \mathcal{X}_j を事前 Random Forest に入力した際の正解クラスラベル c_k の尤度、 $P_t(c_k|\mathcal{X}_j)$ は同様に構築中の目標 Random Forest に入力した際の尤度である。一本目の決定木を構築する際は、 $P_t(c_k|\mathcal{X}_j)$ は 0 とし、 λ_j を $1 + e^{P_a(c_k|\mathcal{X}_j)}$ で初期化する。

分岐関数の決定は、通常の Random Forest 同様にランダムに選択された特徴量としきい値に対し

て情報利得を算出し、情報利得が最大となる候補を分岐関数として採用する。

$$\Delta E = \frac{|I_l|}{|I|} E(I_l) - \frac{|I_r|}{|I|} E(I_r) \quad (7.9)$$

$$E(I) = \sum_{k=1}^C P(c_k) \log P(c_k) \quad (7.10)$$

提案手法は共変量を用いて、目的ドメインの学習に有効でない事前ドメインサンプルの影響が小さくなるように事前ドメインサンプルを重み付ける。そのため、次式のようにドメインごとに異なる計算でクラス確率を算出して統合することで、両ドメインのクラス確率 $P(c_k)$ を算出する。

$$P(c_k) = \frac{|I^{t,c_k}| + |\Lambda^{a,c_k}|}{|I^t| + |\Lambda^a|} \quad (7.11)$$

$$|\Lambda^{a,c_k}| = \sum_{j:y_j=c_k} \lambda_j \quad (7.12)$$

ここで、 $|I^{t,c_k}|$ はサブセット内の目標ドメインにおいてクラスラベルが c_k であるサンプルの数、 $|\Lambda^{a,c_k}|$ は事前ドメインにおいてクラスラベルが c_k であるサンプルに付与された λ の総和である。また、 $|I^t|$ は全てのクラスについて $|I^{t,c_k}|$ を総和したものであり、 $|\Lambda^a|$ は全てのクラスについて $|\Lambda^{a,c_k}|$ を総和したものである。

任意の深さに達するまで分岐を繰り返し決定木を構築した後、 λ を式 (7.8) に基づき更新し、次の新しい決定木を作成していく。決定木の作成は、決定木の本数が任意の数に達することで終了する。以上の処理により、任意の本数の決定木で構成される Random Forest を構築することができる。ここで、転移学習における事前ドメインサンプルの選択基準である λ は、初期状態では 1 が与えられ、決定木が追加される度に目標 RF と事前 RF の違いを正しく表現できるよう更新される。そのため、学習の序盤に構築された決定木における λ の信頼性は低く、目標ドメインの学習に適合していない可能性がある。Boosting に基づく転移学習の 1 つである TrAdaBoost[38] は、同様の理由で T ラウンド学習した際の前半部分は切り捨て、後半 $T/2$ の弱識別器のみで識別を行っている。本研究では文献 [38] と同様に、前半 $T/2$ の決定木を切り捨てることで、学習初期の悪影響を低減する。

提案手法の学習アルゴリズムを **アルゴリズム 7.2** に示す。

アルゴリズム 7.2: 共変量シフトによる転移学習を導入した Random Forest の学習

1. 入力: I 個の目標ドメイン学習サンプル X_t と J 個の事前学習サンプル X_a とそれらのクラスラベル c_k を用意する.

事前に X_a を Random Forest により学習し, 識別器を構築しておく.

2. 初期化: 事前ドメインの重み付けとなる共変量 λ_j を初期化

$$\lambda_j = 1 + e^{P_a(c_k|\mathcal{X}_j)}$$

3. 学習:

For $t = 1, \dots, T$ //決定木の本数

· X_t, X_a から同数のサンプルを無作為に選択し, サブセットを作成

· 決定木の作成 (アルゴリズム 7.1)

クラス確率

$$P(c_k) = \frac{|I^{t,c_k}| + |\Lambda^{a,c_k}|}{|I^t| + |\Lambda^a|}$$

$$|\Lambda^{a,c_k}| = \sum_{j:y_j=c_k} \lambda_j$$

· λ_j の更新

$$\lambda_j = \frac{1 + e^{P_a(c_k|\mathcal{X}_j)}}{1 + e^{P_t(c_k|\mathcal{X}_j)}}$$

End for

4. 出力: 目標ドメインに対する Random Forest

後半部分 ($t = T/2$) の決定木のみを抽出し, Random Forest とする.

$$P(c_k|\mathbf{x}) = \frac{2}{T} \sum_{t=T/2}^T P(c_k|l)$$

7.4 評価実験

転移学習を導入した Random Forest の有効性を評価するため、目標ドメインに事前ドメインを加えたデータ集合を用いて再学習した Random Forest と目標ドメインのみで再学習した Random Forest と比較する。

7.4.1 2クラス分類問題の転移

2クラス分類問題での評価実験は、事前ドメイン、目標ドメインの評価データとして、異なる人検出データセットを用いる。本実験では、事前ドメインとして DaimlerChrysler dataset を、目標ドメインとして INRIA person dataset を用いる。各ドメインの例を図 7.5 に示す。



(a)DaimlerChrysler Mono Pedestrian Detection Benchmark Dataset



(b)INRIA Person Dataset

図 7.5: 各データベースのサンプル例。

事前ドメイン：DaimlerChrysler Mono Pedestrian Detection Benchmark Dataset[50]

15,600 枚の人画像と 6,700 枚の背景画像が学習サンプルとして、56,500 の人領域を含む 21,800 枚の画像が評価サンプルとして用意されている。

目標ドメイン：INRIA Person Dataset[3]

2,416 枚の人画像と 1,218 枚の背景画像が学習サンプルとして、1,135 の人画像と 453 枚の背景画像が評価サンプルとして用意されている。

入力特徴量は HOG 特徴量 [3] とする。Random Forest 構築の際のパラメータを 7.1 に示す。

表 7.1: Random Forest 構築のパラメータ.

木の本数	50
木の最大深さ	5
特徴次元数	3,780
分岐ノード候補のランダム生成数	$63(= \sqrt{3,780})$

事前ドメインに目標ドメインを転移する提案手法と以下の2つの手法を比較する.

“事前ドメイン+目標ドメイン”

事前ドメインに目標ドメインを転移を用いず加えたサンプル集合を入力として Random Forest を学習する.

“目標ドメインのみ”

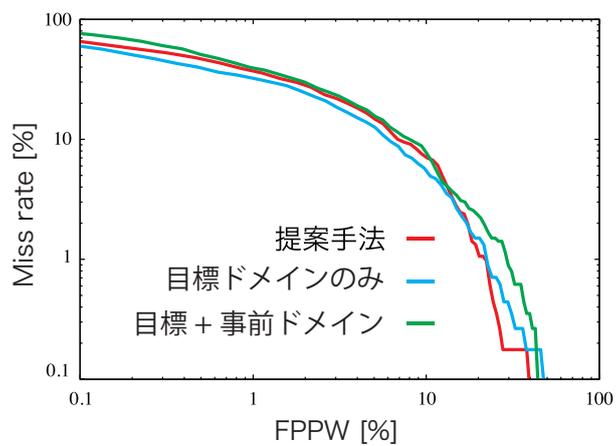
目標ドメインのサンプルのみを入力として Random Forest を学習する.

7.4.2 2クラス分類問題への転移学習の導入による効率化

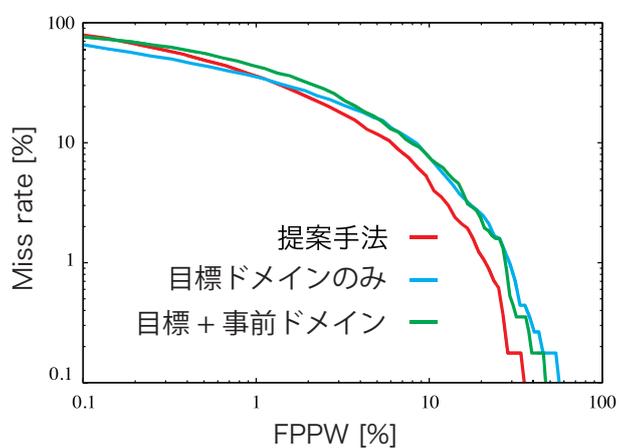
目標ドメインのポジティブ学習サンプルを 2,416 枚から 800 枚, 100 枚と順に削減した際の DET カーブを図 7.6 に示す. まず, 図 7.6(a) より, 目標ドメインの学習サンプルを 2,416 枚全て使用した場合, 提案手法と“事前ドメイン+目標ドメイン”, “目標ドメインのみ” は同程度の性能となることがわかる.

800 枚に学習サンプルを削減した場合, 図 7.6(b) のように“目標ドメインのみ”, “事前ドメイン+目標ドメイン” は性能を低下させる. これは, 目標ドメインの学習サンプル数が削減されることで目標ドメインに対する学習が十分でなくなるためである. 一方, 提案手法はその性能を維持していることがわかる.

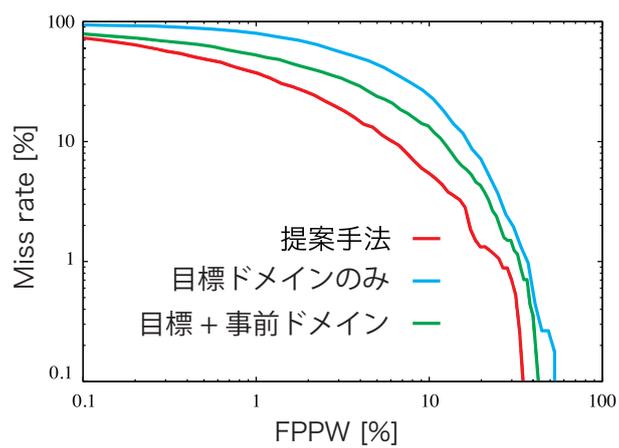
さらに, 目標ドメインの学習サンプルを 100 枚まで削減した場合, 図 7.6(c) より, “目標ドメインのみ” の識別性能は大きく低下する. これは, 100 枚の人画像の学習ではデータセット全体の分布を捉えることはできないことを示している. また, “事前ドメイン+目標ドメイン” は性能の低下が緩やかである. これは, 事前ドメイン中に存在する目標ドメインの識別に有効なサンプルを学習できたためであると考えられる. しかし, 事前ドメイン中には目標ドメインの学習に適さないサンプルも多く含まれるため, 2,416 枚や 800 枚での結果に比べ識別性能は低下している. 一方, 提案手法は高い識別性能を獲得しており, その識別性能は 2,416 枚, 800 枚時と比較しても大きく低下しない. 以上より, 転移学習により識別に有効な事前ドメインサンプルのみを重視しながら, 目標ドメインの学習に取り込むことで, 目標ドメインの学習サンプルの不足を補うことができたと考えられる.



(a) 入力サンプル 2,416 枚



(b) 入力サンプル 800 枚



(c) 入力サンプル 100 枚

図 7.6: 転移学習を導入した Random Forsts の識別性能.

7.4.3 マルチクラス分類問題の転移

マルチクラス分類問題での評価実験では、人体モデルの向きを示すパラメータ θ を正面、右向き、背面、左向きと 90° ずつ 4 分割した生成人体画像データを用いる。ここで、正面と背面を同一クラスとし、正面／背面、右向き、左向きの 3 クラスに背景クラスを加えた計 4 クラスをカテゴリとする。

本実験では、事前ドメインとして俯角 20° の生成人体画像データセットを、目標ドメインとして俯角 30° の生成人体画像データセットをそれぞれ用いる。各データセットは 5 章で述べた学習サンプル生成手法により生成した。図 7.7 に、各ドメイン、各クラスのサンプル画像例を示す。



図 7.7: マルチクラス分類問題のサンプル例。

事前ドメイン：俯角 20°

俯角を 20° に設定し、2,416 枚の人画像と 12,180 枚の背景切出し画像を学習サンプルとして生成する。

目標ドメイン：俯角 30°

俯角を 30° に設定し、2,416 枚の人画像と 12,180 枚の背景切出し画像を学習サンプルとして、同数の画像を評価サンプルとして用意する。俯角の変化により、各カテゴリの見えは大きく変化する。図 7.7 の俯角 20° と俯角 30° の平均匀配画像を比較すると、同じ左向きでも体の角度がより広く拡散

し、ぼやけていることがわかる。

7.4.4 マルチクラス分類問題への転移学習の導入による効率化

目標ドメインのポジティブ学習サンプルを 2,416 枚、800 枚、100 枚と削減した際の識別性能を、それぞれ表 7.2、表 7.3、表 7.4 に示す。

表 7.2: マルチクラス分類問題に対する転移学習結果 (2,416 枚)。

	正面／背面 [%]	右向き [%]	左向き [%]	背景 [%]	平均 [%]
提案手法	55.8	91.0	88.1	97.1	83.0
目標ドメインのみ	60.0	90.8	88.7	98.5	84.5
事前ドメイン+目標ドメイン	53.5	87.6	84.6	98.8	81.1

表 7.3: マルチクラス分類問題に対する転移学習結果 (800 枚)。

	正面／背面 [%]	右向き [%]	左向き [%]	背景 [%]	平均 [%]
提案手法	55.0	77.5	73.7	98.9	76.3
目標ドメインのみ	60.4	81.5	77.8	98.8	79.6
事前ドメイン+目標ドメイン	42.9	77.8	73.2	94.6	72.1

表 7.4: マルチクラス分類問題に対する転移学習結果 (100 枚)。

	正面／背面 [%]	右向き [%]	左向き [%]	背景 [%]	平均 [%]
提案手法	44.4	70.9	71.9	98.8	71.5
目標ドメインのみ	47.6	28.8	43.7	98.9	54.8
事前ドメイン+目標ドメイン	36.1	65.3	67.1	94.7	65.8

表 7.2 より、目標ドメインの学習サンプルを 2,416 枚全てを使用した場合、提案手法、“目標ドメインのみ”、“事前ドメイン+目標ドメイン”の3手法ともほぼ同程度の平均性能となることがわかる。

次に、800 枚に学習サンプルを削減した場合、表 7.3 より、各手法とも精度が低下しており、サンプル数が不足していることがわかる。特に、“事前ドメイン+目標ドメイン”は 72.1%と大きく性能を低下させている。これは、マルチクラス分類問題において俯角が異なると、事前ドメインと目標ドメインの差が大きくなり、適応できない事前ドメインサンプルが多く発生するためであると考えられる。提案手法は転移学習により、適応できない事前ドメインサンプルを学習に用いないことで、“事前ドメイン+目標ドメイン”に比べ性能低下を抑制できたと考えられる。

さらに、目標ドメインの学習サンプルを 100 枚まで削減した場合、表 7.4 より、“目標ドメインのみ”の識別性能は大きく低下する。特に、右向き、左向きについての識別率が顕著に低下しており、

両カテゴリについて学習サンプルが十分でないことがわかる。また，“事前ドメイン+目標ドメイン”と提案手法は性能の低下が緩やかである。これは、目標ドメインで不足している右向きカテゴリと左向きカテゴリに有効なサンプルを、事前ドメインから補い学習できたためであると考えられる。しかし、事前ドメイン中には目標ドメインの学習に適さないサンプルも多く含まれるため、共変量を用いた転移学習を行う提案手法が高精度である。

7.5 まとめ

本章では、Random Forest に転移学習を導入することによる再学習の効率化について述べた。決定木構築時に共変量シフトを用いることで、目標ドメインの学習に適した事前ドメインサンプルのみを利用して学習できる。このため、少数の目標学習サンプル追加で十分な学習が可能となる。評価実験において、目標ドメイン数を 100 枚と少量にした際に、目標ドメインのみの再学習では著しく性能が低下するのに対して、提案手法は学習サンプルを 800~2,416 枚使用した際と同程度の識別性能を得ることができた。これは、多くのカテゴリについて大量にサンプルを得ることが困難なマルチクラス分類問題において有効である。これにより、提案手法は Random Forest をベースとした様々なアプリケーションにおける学習の効率化に利用が可能である。

第8章

結論と展望

本論文では、物体検出技術の実用化のために、問題の分割統治による検出技術の高精度化と、特定環境への適応の効率化を行った。

8.1 結論

各章のまとめは次の通りである。

2章では、統計的学習法に基づく検出技術について述べた後、検出に用いられる統計的学習法各種について詳述した。統計的学習法は多量の学習サンプルの統計に基づき、対象とする物体を高精度に判別する。しかしながら、対象クラスが多様なバリエーションを含む場合については学習が困難となる問題、多量に用意した学習サンプルと実際の設置環境が相違する場合には性能が低下するという問題が存在することを述べた。

3章では、多様な見えのバリエーションを含む物体を高精度に検出するため、問題を弱識別器応答値に基づきサンプルを分割する手法と、分割された問題を弱識別器を共有しながら解決する DaCBoost について述べた。問題の分割は弱識別器の応答値空間において Spectral Clustering を行うことで、サンプル集合の分割として行われる。これにより、類似したバリエーションの見えをもつ問題は自動的に集合し、個別に解決することで高精度な識別が可能となる。さらに、そのまま個別に識別器を構築して解決すると弱識別器数が膨大になるため、DaCBoost により弱識別器の情報を共有しながら学習を行うことで、同じ弱識別器数において高精度な識別器を構築することが可能となる。

4章では、特徴量の貢献度評価を提案した後、その貢献度に基づき特徴削減や識別器を調整することで、現場環境に対して適応が可能となることを述べた。また、Boosting に基づき算出される貢献度は、特徴の有効性に妥当な評価を与えることができ、検出システム構築の一助となる。

5章では、設置対象とする特定環境に対してさらなる効率化、高精度化を実現するために、特定環境に特化した学習サンプルの自動生成手法を提案した。学習サンプルは実環境の背景画像と3次元人体モデル CG の合成で生成し、特定環境に特化した大量の学習サンプルを用いて学習することで高精度な識別が可能であることを示した。さらに、学習サンプル生成の自動化に伴い、ネガティブサンプルにノイズが混在し得る問題については、Multiple Instance Learning をネガティブサンプルに適用した Negative-Bag MILBoost により解決することを提案した。これにより、簡易に大量のサンプルを用意して、識別器を構築することが可能となる。

6章では、さらに別側面からの効率化として、新規収集するサンプルの数と学習に要する時間を削

減する手法について述べた。転移学習を導入することで識別器の学習は高速となる。しかしながら、ドメイン間の差が大きすぎるなど転移が困難な場合に、その識別性能は大きく低下する。この問題に対し、高速な転移特徴空間と低速であるが高精度な全探索特徴空間を切り替えながら学習するハイブリッド型転移学習を提案した。これにより、再学習より非常に高速で、転移学習より高精度な識別を実現することが可能となる。

7章では、より多様な問題に対応できる手法として、マルチクラス識別器である Random Forest に転移学習による効率化を導入した。転移学習を Random Forest に導入することで、マルチクラス分類問題においても、学習サンプルが少数でありながらも十分な学習を可能とする。

1章では、物体検出の実用化に向けた重要な要件として以下の条件を列挙した。以下に、各条件に対して、本研究での貢献についてまとめる。

- (1) 多様なバリエーションを持つ検出対象の高精度な識別
- (2) 再学習なくシステムを簡易に調整可能
- (3) 特定環境へシステムを容易に再学習可能
- (4) 少数の学習サンプルでシステムを再学習可能

条件 (1) については、3章で述べた分割統治戦略により多様なバリエーションを持つ物体の検出という複雑な問題を、弱識別器の応答値と Spectral Clustering により簡易な小問題の集合に分割して、個別に識別器を構築することで高精度な識別が可能となる。さらに、分割した各問題に対して、DaCBoost で学習を行うことで互いに共通する部分は情報を共有しながら効率的に識別器構築が可能となる。従って、多様な見えの変化を持つ物体に対する精度の低下や識別器の肥大化という問題を解決する高精度化となる。

条件 (2) については、4章で述べた貢献度に基づいた識別器の調整により、識別に有効な特徴量を選別し、再学習なく現場環境に適応可能となる。この問題に対して、Joint Boosting より貢献度を算出することで、各検出対象クラスの関連や共通して利用できる特徴量などを解析することが可能となる。また、貢献度は学習時に、環境変動に弱い特徴を除去する特徴選択にも用いることが可能である。そのため、従来用いられる SFS 法より高速に、SBS 法と同程度の精度で特徴選択が可能となる。従って、様々な環境にシステムを設置する際、識別器の再学習などで問題となる多大な人的コストを解消する。

条件 (3) については、5章で述べた特定環境に適した学習サンプルの自動生成により、人手のコストを最小限に、大量の学習サンプルを用いて再学習することが可能となる。これにより、特定の環境に適した識別器を安価に構築することが可能となるため、高精度な識別システムの実用が可能となる。さらに、サンプルの自動生成に起因するミスラベルについても、ノイズの影響を低減できる Negative-bag MILBoost を用いた学習により、ミスラベルを含む学習サンプルから問題なく高精度

な識別器を構築できる。従って、再学習が必要な、学習環境とシステム設置環境との差異が大きい場合において多大な人的コストが必要となる問題を解決することが期待できる。

条件 (4) については、6 章で述べたハイブリッド型転移学習により、更に少数のサンプル追加で高速に再学習することが可能となる。転移学習では高速性と引き換えに失われる識別性能を維持したまま、通常の再学習に比べ非常に高速な学習法となる。また、7 章で述べた Random Forests への転移学習の導入は、マルチクラス識別を中心とした多様な問題の効率化に寄与する。5 章で述べた特定環境に適した学習サンプルの自動生成にこれらの提案を加え、調整などでは対応できない差異を持つ設置環境に対しても識別性能を低下させることなく非常に効率的に適応可能となる。

8.2 展望

本論文では、物体検出技術の実用化のために統計的学習法に基づく物体検出の高精度化と効率化について述べた。そして、評価実験により、複雑な問題に対する精度の低下、特定環境への適応、より少数の学習サンプルでの学習という各問題に対して従来法に比べ効率的かつ高精度な成果を得た。今後、より実用的かつ汎用的な物体検出を実現するためには以下の 2 つの課題がある。

一つ目の課題は、同一環境内で時々刻々と変化する環境変動への対応である。本論文では異なる環境に適応する手法について種々述べたが、時間による光量の違い、一時的な影、遮蔽などの変化については対応していない。これらの変化は同一環境内でもシステム稼働時に周期的または突発的に発生するため、オンラインで学習して適応するアプローチが望ましい。

二つ目の課題は、より効率的な学習として、積極的な事前学習結果の利用が挙げられる。現在、本研究の手法も含む転移学習の多くは、事前ドメインから目標ドメインへのサンプルの転移を実現する手法が主軸であり、事前識別器はそのサンプルの取捨選択を補助するに過ぎない。しかし、効率的な転移という意味ではまさにその事前識別器で学習された識別ルールを転移可能とすべきである。これにより、再学習の無駄を大幅に低減し、すでにある識別器の調整で多様な環境に対応可能となることが期待できる。

また、本研究では物体検出の実用化にむけた条件に対して、分割統治による高精度化、貢献度に基づく識別器の調整法、学習サンプルの自動生成による効率化、ハイブリッド型転移学習による効率化を個々に提案した。物体検出の実用化のためには、本研究の成果を統合して扱うフレームワークの実現が重要である。

謝 辞

本研究は、著者が中部大学大学院工学研究科情報工学専攻後期博士課程在学中に、同大学工学部情報工学科 藤吉弘亘教授の指導のもとに行ったものである。研究の遂行にあたり、常日頃ご指導を賜りました中部大学工学部情報工学科 藤吉弘亘教授に深く感謝の意を表します。本論文をまとめるにあたり、有益なご討論、ご助言を賜りました和歌山大学院システム工学研究科 和田俊和教授、中部大学工学部情報工学科 岩堀祐之教授、平田豊教授に謹んで感謝いたします。本研究において、多くのご助言を頂きました中部大学情報工学科 山内悠嗣研究員、オムロン株式会社 山下隆義氏、株式会社日立製作所 日立研究所 弓場竜氏に心より感謝申し上げます。最後に、本研究にご協力して頂いた藤吉研究室の皆様にご感謝致します。

参考文献

- [1] P. Viola and M. Jones: “Rapid Object Detection Using a Boosted Cascade of Simple Features”, in proceedings of *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 511–518, 2001.
- [2] P. Viola and M. Jones: “Robust Real-Time Object Detection”, *International Journal of Computer Vision*, Vol. 57, No. 2, pp. 137–154, 2004.
- [3] N. Dalal and B. Triggs: “Histograms of Oriented Gradients for Human Detection”, in proceedings of *IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 886–893, 2005.
- [4] M. Kearns: “Thoughts on Hypothesis Boosting”, Unpublished, Project for Ron Rivest’s machine learning course at MIT: 1988.
- [5] Y. Freund and R. E. Schapire: “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting”, *Journal of Computer and System Sciences*, Vol. 55, No. 1, pp. 119–139, August 1997.
- [6] R. E. Schapire and Y. Singer: “Improved Boosting Algorithms Using Confidence-Rated Predictions”, *Machine Learning*, Vol. 37, No. 3, pp. 297–336, 1999.
- [7] J. Friedman, T. Hastie, and R. Tibshirani: “Additive Logistic Regression : a Statistical View of Boosting”, *Annals of Statistics*, 1998.
- [8] S. Li, Z. Zhang, H. Shum, and H. Zhang: “FloatBoost Learning for Classification”, in proceedings of *MIT Press*, pp. 993–1000, 2003.
- [9] A. Demiriz and K. P. Bennett: “Linear Programming Boosting via Column Generation”, *Machine Learning*, pp. 225–254, 2001.
- [10] B. Wu, H. Ai, C. Huang, and S. Lao: “Fast Rotation Invariant Multi-View Face Detection Based on Real Adaboost”, in proceedings of *International Conference on Automatic Face and Gesture Recognition*, pp. 79–84, 2004.

- [11] A. Torralba, K. P. Murphy, and W. T. Freeman: “Sharing Visual Features for Multiclass and Multiview Object Detection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 5, pp. 854–869, 2007.
- [12] J. Friedman, T. Hastie, and R. Tibshirani: “Additive Logistic Regression: a Statistical View of Boosting”, *Annals of Statistics*, Vol. 28, pp. 337–407, 1998.
- [13] Y. Yamauchi, H. Fujiyoshi, Y. Iwahori, and T. Kanade: “People Detection Based on Co-occurrence of Appearance and Spatio-temporal Features”, *National Institute of Informatics Transactions on Progress in Informatics*, No. 7, pp. 33–42, 2010.
- [14] 高木雅成, 山内悠嗣, 藤吉弘亘: “Geometric Context を用いた特徴量間の共起による物体検出の高精度化”, 電子情報通信学会論文誌, Vol. 93-D, No. 8, pp. 1429–1438, 2010.
- [15] P. Viola, O. M. Way, and M. J. Jones: “Robust Real-Time Face Detection”, *International Journal of Computer Vision*, Vol. 57, No. 2, pp. 137–154, 2004.
- [16] 山内悠嗣: 画像局所特徴量の共起性に基づいた人検出の高精度化に関する研究. PhD thesis: 中部大学院工学研究科情報工学専攻博士論文: 2012. 未公開.
- [17] 村瀬洋: “画像認識のための生成型学習”, 情報処理学会論文誌. コンピュータビジョンとイメージメディア, Vol. 46, No. 15, pp. 35–42, 2005.
- [18] M. Noda, T. Takahashi, I. Ide, H. Murase, Y. Kojima, and T. Naito: “Recognition of Road Markings from In-Vehicle Camera Images by a Generative Learning Method”, in proceedings of *International Conference on Machine Vision Applications*, pp. 514–517, 2009.
- [19] K. Doman, D. Deguchi, T. Takahashi, Y. Mekada, I. Ide, and H. Murase: “Construction of cascaded traffic sign detector using generative learning”, in proceedings of *International Conference on Innovative Computing Information and Control*, pp. 889–892, 2009.
- [20] C. Huang, H. Ai, Y. Li, and S. Lao: “Vector boosting for rotation invariant multi-view face detection”, in proceedings of *IEEE International Conference on Computer Vision*, Vol. 1, pp. 446–453, 2005.
- [21] 村井陽介, 藤吉弘亘, 伊藤誠也, 数井誠人: “弱識別器の応答に基づく類似シルエット画像選択による Chamfer Matching を用いた人領域のセグメンテーション”, 第 12 回画像の認識・理解シンポジウム, pp. 1468–1475, 2009.
- [22] A. Y. Ng, M. I. Jordan, and Y. Weiss: “On Spectral Clustering: Analysis and an algorithm”, in proceedings of *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, pp. 849–856. MIT Press, 2001.

- [23] Y. A. Sheikh, E. A. Khan, and T. Kanade: “Mode-seeking by Medoidshifts”, *Computer Vision, IEEE International Conference on Computer Vision*, Vol. 0, pp. 1–8, 2007.
- [24] C. Dorin and M. Peter: “Mean Shift: A Robust Approach Toward Feature Space Analysis”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 5, pp. 603–619, 2002.
- [25] S. Agarwal, A. Awan, D. Roth, and I. C. Society: “Learning to detect objects in images via a sparse, part-based representation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, pp. 1475–1490, 2004.
- [26] A. W. Whitney: “A Direct Method of Nonparametric Measurement Selection”, *IEEE Transactions on Computers*, Vol. 20, No. 9, pp. 1100–1103, September 1971.
- [27] T. Marill and D. Green: “On the Effectiveness of Receptors in Recognition Systems”, *IEEE Transactions on Information Theory*, Vol. 9, No. 1, pp. 11–17, 2006.
- [28] K. Bache and M. Lichman, “UCI Machine Learning Repository”, 2013.
- [29] R. T. Collins, A. J. Lipton, H. Fujiyoshi, and T. Kanade: “Algorithms for Cooperative Multisensor Surveillance”, in proceedings of *Surveillance, Proceedings of the IEEE*, 2001.
- [30] 藤吉弘亘, 小村剛史, 矢入(江口) 郁子, 香山健太郎, 吉水宏: “歩行者 ITS のためのフレーム間差分による移動体検出法とその評価”, *情報処理学会論文誌. コンピュータビジョンとイメージメディア*, Vol. 45, No. 13, pp. 11–20, 2004.
- [31] B. D. Lucas and T. Kanade: “An Iterative Image Registration Technique with an Application to Stereo Vision”, in proceedings of *DARPA Image Understanding Workshop*, pp. 674–679, 1981.
- [32] J. Mar, V. David, D. Ger, and M. L. Antonio: “Learning Appearance in Virtual Scenarios for Pedestrian Detection”, in proceedings of *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 137–144, 2010.
- [33] “平成 20 年度体力・運動能力調査調査結果統計表:文部科学省”, 2009.
http://www.mext.go.jp/b_menu/houdou/21/10/attach/1285568.htm.
- [34] T. Dietterich, R. Lathrop, and T. Lozano-Pérez: “Solving the Multiple Instance Problem with Axis-Parallel Rectangles”, *Artificial Intelligence Journal*, Vol. 89, pp31–71, 1997.
- [35] P. Viola, J. Platt, and C. Zhang: “Multiple Instance Boosting for Object Detection”, in proceedings of *Advances in neural information processing systems*, Vol. 18, 2006.

- [36] J. Pang, Q. Huang, S. Yan, S. Jiang, and L. Qin: “Transferring Boosted Detectors Towards Viewpoint and Scene Adaptiveness.”, *IEEE Transactions on Image Processing*, Vol. 20, No. 5, pp. 1388–1400, 2011.
- [37] F. Shen, H. Yu, K. Sakurai, and O. Hasegawa: “An Incremental Online Semi-supervised Active Learning Algorithm Based on Self-organizing Incremental Neural Network”, *Neural Computing and Applications*, Vol. 20, No. 7, pp. 1061–1074, 2011.
- [38] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu: “Boosting for transfer learning”, in proceedings of *International Conference on Machine Learning*, pp. 193–200, 2007.
- [39] X. Wang, W. Li, and M. Wang: “Transferring a generic pedestrian detector towards specific scenes”, *2013 IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 0, pp. 3274–3281, 2012.
- [40] G. Li, L. Qin, Q. Huang, J. Pang, and S. Jiang: “Treat samples differently: Object tracking with semi-supervised online CovBoost”, in proceedings of *International Conference on Computer Vision*, pp. 627–634, 2011.
- [41] A. Oliva and A. Torralba: “Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope”, *Int. J. Comput. Vision*, Vol. 42, No. 3, pp. 145–175, 2001.
- [42] A. Torralba, B. C. Russell, and J. Yuen: “LabelMe: Online Image Annotation and Applications”, *Proceedings of the IEEE*, Vol. 98, No. 8, pp. 1467–1484, 2010.
- [43] L. Breiman: “Random Forests”, *Machine Learning*, Vol. 45, No. 1, pp. 5–32, 2001.
- [44] L. Breiman: “Bagging Predictors”, *Machine Learning*, Vol. 24, No. 2, pp. 123–140, 1996.
- [45] Y. Amit and D. G. Y: “Shape quantization and recognition with randomized trees”, *Neural Computation*, Vol. 9, pp. 1545–1588, 1997.
- [46] J. Gall, A. Yao, N. Razavi, L. V. Gool, and V. Lempitsky: “Hough Forests for Object Detection, Tracking, and Action Recognition”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 11, pp. 2188–2202, 2011.
- [47] J. Shotton, M. Johnson, and R. Cipolla: “Semantic texton forests for image categorization and segmentation”, in proceedings of *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.
- [48] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake: “Real-time human pose recognition in parts from single depth images”, in proceedings of *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

- [49] J. Margeta, K. McLeod, A. Criminisi, and N. Ayache: “Decision Forests for Segmentation of Left Atrium From 3D MRI”, in MICCAI Workshop on Statistical Atlases and Computational Models of the Heart, 2013

- [50] M. Enzweiler and D. M. Gavrilu: “Monocular Pedestrian Detection: Survey and Experiments”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 12, pp. 2179–2195, 2009.

研究業績一覧

学術論文

- [1] 土屋成光, 藤吉弘亘, “Boosting に基づく分割統治的戦略による高精度な識別器構築手法”, 電子情報通信学会論文誌, Vol.J95-D, No.3, pp.656-665, 2012.
- [2] 後藤雄飛, 土屋成光, 山内悠嗣, 藤吉弘亘, “近似計算を導入した線形識別器の早期判定による高速な識別”, 電子情報通信学会論文誌, Vol.J97-D, No.2, pp.294-302, 2014.
- [3] 土屋成光, 山内悠嗣, 藤吉弘亘, “人検出のための生成型学習と Negative-Bag MILBoost による学習の効率化”, 電気学会論文誌 C, Vol.134, No.3, pp.450-458, 2014.

国際会議発表論文 (査読あり)

- [1] Masamitsu Tsuchiya, Hironobu Fujiyoshi, “Evaluating Feature Importance for Object Classification in Visual Surveillance”, in Proceedings of International Conference on Pattern Recognition, pp. 978-981, 2006.
- [2] Masamitsu Tsuchiya, Hironobu Fujiyoshi, “A Method of Feature Selection Using Contribution Ratio Based on Boosting”, in Proceedings of International Conference on Pattern Recognition, 2008.
- [3] Masamitsu Tsuchiya, Hironobu Fujiyoshi, “Hybrid Transfer Learning for Efficient Learning in Object Detection ”, in Proceedings of Asian Conference on Pattern Recognition, 2013.
- [4] Yohei Mishina, Masamitsu Tsuchiya, Hironobu Fujiyoshi, “Boosted Random Forest ”, in Proceedings of International Conference on Computer Vision Theory and Applications, pp. 594-598, 2014.

学会口頭発表 (査読あり)

- [1] 土屋成光, 藤吉弘亘, “屋外環境下における移動体識別に用いる入力特徴の AdaBoost による評価”, 画像センシングシンポジウム, pp. 277–283, 2006.
- [2] 土屋成光, 藤吉弘亘, “Boost 学習に基づく特徴量の貢献度の評価”, 情報科学技術フォーラム, pp. 71–74, 2007.

学会口頭発表 (査読なし)

- [1] 土屋成光, 藤吉弘亘 “調整可能な AdaBoost による識別器”, 電気関係学会東海支部連合大会, 2006
- [2] 土屋成光, 藤吉弘亘. “Boosting に基づく分割統治的戦略による高精度な識別器構築手法の提案”, 電子情報通信学会 パターン認識・メディア理解研究会, vol. 109, no. 182, pp81–86, 2009.
- [3] 土屋成光, 山内悠嗣, 山下隆義, 藤吉弘亘, “ハイブリッド型転移学習による物体検出における学習の効率化”, 電子情報通信学会 パターン認識・メディア理解研究会, vol. 112, no. 385, pp. 329–334, 2013.
- [4] 藤吉弘亘, 山内悠嗣, 土屋成光, 山下隆義, “統計的学習法による物体検出の高精度化と効率化-人検出の実用化に向けて-”, 情報処理学会 コンピュータビジョンとイメージメディア研究会, vol. 2013-CVIM-186, No.26, 2013.
- [5] 後藤祐斗, 土屋成光, 藤吉弘亘, “Inside-Out カメラからの 3 次元 Scan path とシーン構造の復元”, 電子情報通信学会 パターン認識・メディア理解研究会, vol. 113, no. 75, pp. 59–64, 2013.