

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。

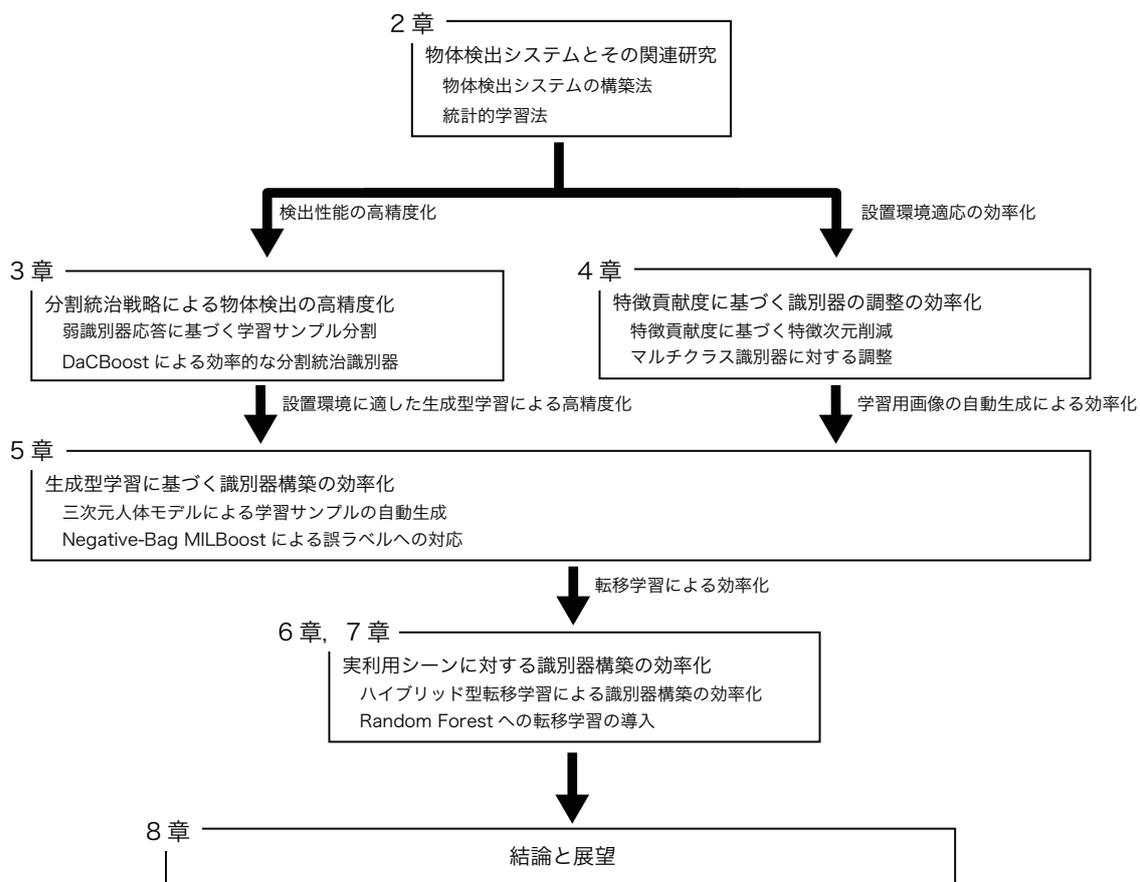


図 1.1: 本論文の構成.

