MI-Hough Forest による類似物の影響を受けない特定物体検出

村井 佑輔 † 三品 陽平 † 小関 亮介 ‡ 松浦 康寿 ‡ 山内 悠嗣 † 山下 隆義 † 藤吉 弘亘 † † 中部大学 ‡(株) 豊田自動織機

E-mail: ym@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

Abstract

Hough Forest は、入力画像から切り出した局所パッチ 画像を Random Forest に入力し、辿り着いた末端ノー ドのクラス確率を物体中心へのオフセット量を用いて 投票する物体検出法である.検出の際に、学習画像に類 似した領域が存在すると、その領域にも投票するため 誤検出が発生するという問題がある.本研究では、学習 サンプルの重みを導入し、ネガティブサンプルに類似 するポジティブサンプルの重みを低くすることで、学 習画像と類似した領域への投票を抑制する MI-Hough Forest を提案する。類似した領域への投票を抑制する ために, Multiple Instance Learning における Bag の 概念を導入する. Hough Forest の学習において, Bag 内にネガティブサンプルとの類似度が高いポジティブ サンプルがある場合,決定木の階層毎に Bag 内の全て のサンプルの重みが小さくなるように更新する. これ により、ネガティブサンプルと類似したポジティブサン プルからの投票を抑制することが可能となる.評価実 験により、従来の Hough Forest と比べて位置に対する 検出精度が8%、回転変化に対する検出精度が7%向 上することを確認した.提案手法は、回転変化した学 習サンプルを1つの Hough Forest で表現することがで きるため、1回のラスタスキャンで全方向の対象物の検 出が可能であり、テンプレートマッチング等の従来手 法と比べて高速であるというメリットもある.

1 はじめに

統計的学習法を用いた画像からの物体検出は,領域 ベース手法と局所パッチベース手法の二つに分類でき る.領域ベース手法は,検出対象の大きさに合わせて 切り出した画像全体から,特徴抽出を行い,識別器に より対象物であるか否かを判定する.歩行者や車両の 検出では,輝度勾配から計算した特徴量[1]をSVM[2] や Adaboost[3] などの統計的学習法を用いて識別器を 学習する.学習した識別器を用いて,入力画像を網羅 的にラスタスキャンして対象物を検出する.このとき, 物体の一部に隠れや変形が発生すると,検出精度が低 下するという問題がある.

一方, 局所パッチベース手法は, パッチを用いた投票 処理を行うため、部分的に隠れた物体に対して頑健な検 出が可能である.局所パッチベース手法は一般化ハフ変 換による投票処理を基としている。2008年に Leibe ら はコードブックを用いた Implicit Shape Model(ISM)[4] を提案した. ISM では、入力画像から抽出した特徴点 を中心としたパッチ画像をコードブックを参照してベク トル量子化し、同一のコードブックに含まれる学習サン プルの情報を用いて、物体の重心位置を投票する. ISM の性能は、抽出したパッチが属するコードブックに依存 しており、コードブックの作成が重要である。2009年 に Gall らにより, Random Forest [5] をパッチの識別に 利用した Hough Forest[6] が提案された. Hough Forest は, Random Forest を用いて, パッチ画像を2クラス に分類する決定木を構築する。検出時は、 ラスタスキャ ンの際に Hough Forest をトラバーサルし、対象物体と 識別されたパッチから重心位置へ投票する.

2012年には、Hough Forest における物体検出のため の学習サンプルの設計法が報告されている [8].文献 [8] では、検出対象のシルエット画像を作成し、ポジティブ 領域を含む割合を考慮した学習サンプルを生成する.ポ ジティブ領域の含有率が低い (ネガティブ領域を多く含 む) パッチが学習サンプルに含まれている場合、検出精 度が低下するという知見が得られている.これは、ネガ ティブサンプルと類似しない識別に有効なポジティブ サンプルの選定の重要性を示している.また、類似物 への誤検出を抑制する手法として、対象物と類似物の 分離性を最大化するような画素を事前に選択して照合 する手法が提案されている [9].このように、ポジティ ブサンプル内のパッチの選定は重要な課題である.

本稿では、類似物への投票を抑制する Hough Forest ベースの物体検出法の実現を目的とする.我々は、学習 サンプルの重みを導入し、ネガティブサンプルに類似す るポジティブサンプルの重みを決定木の階層毎に低くす ることで、学習画像と類似した領域への投票を抑制する MI-Hough Forest を提案する.類似した領域に対する 投票を抑制するために、Multiple Instance Learning[7] における Bag の概念を利用して重みの更新を行う.検 出の際に、重みが低いポジティブサンプルからの投票 値は小さくなるため, 誤検出を抑制することが可能と なる.

2 Hough Forest

Hough Forest[6] は、大量の学習サンプルを用いて、 パッチ画像を識別する Random Forest(決定木群[5])を 学習する。検出時には、入力画像から切り出したパッチ 画像を Random Forest に入力し、辿り着いた末端ノー ドのポジティブサンプルの割合とオフセット量のリスト を用いて投票することで物体を検出する。以下に Hough Forest の学習と物体検出法について述べ、Hough Forest の問題点について整理する。

2.1 Hough Forest の学習と検出

学習画像より切り出したパッチを学習サンプルとし て用いる.検出対象物体からグリッドサンプリングに より切り出したパッチ画像群をポジティブサンプル,非 対象物体から切り出したパッチ画像群をネガティブサ ンプルとする.ポジティブサンプルは、物体中心まで のオフセット量を保持している.このように作成した 学習サンプルを用いて図1に示すように、各サブセッ ト毎に決定木を学習する.



図 1 Hough Forest の学習過程

決定木の分岐ノードでは、2点の画素値の差をしきい 値処理により左右の子ノードに分岐する.末端ノード は、ポジティブサンプルの割合と物体中心へのオフセッ ト量のリストを保存する.

Hough Forest を用いた物体検出では、入力画像から 切り出したパッチ画像を各決定木に入力し、辿り着い た末端ノードを求める.末端ノードに保存されている ポジティブサンプルの割合をオフセット量を用いて物 体中心に投票することで尤度マップを作成する.尤度 マップの極大値を MeanShift[10]を用いて位置を求め、 検出対象物体として検出する.

2.2 Hough Forest の問題点

Hough Forest による物体検出時の尤度マップを図 2 に示す.検出対象から切り出したパッチ A と非検出対 象から切り出したパッチ B は,形状が類似しているた め,同じ末端ノードに辿り着き,検出対象と同様に投 票処理が行われる.そのため,検出対象と類似した形 状領域を誤検出をすることになる.これは検出対象か ら切り出したパッチ画像全てを学習に用いるからであ る.この問題を解決するには,Hough Forest に用いる 学習用ポジティブサンプルの取捨選択を自動に行う必 要がある.



図 2 Hough Forest の問題点

3 MI-Hough Forest

本章では,提案手法である MI-Hough Forest につい て述べる. MI-Hough Forest では類似物への投票を抑 制するために, Multiple Instance Learning[7] における Bag の概念を利用して学習サンプルに重みを導入する.

3.1 学習サンプルの作成と Bag の構成

本研究では、360 度方向に回転したパーツの検出を 対象とする.学習サンプルとして、検出対象の距離画 像を 360 度方向に 1 度ずつ回転した画像を生成して用 いる.各回転角における生成した学習画像からグリッ ドサンプリングにより切り出したパッチ画像 \mathcal{I}_{ij} を用 いて Bag $B_i = \{(\mathcal{I}_{ij}, k_{ij}, \mathbf{d}_{ij}, o_{ij})\}(i = 1, 2, \cdots, I, j = 1, 2, \cdots, J)$ を作成する.ここで、 \mathcal{I}_{ij} は *i* 番目の Bag に所属する *j* 個目のパッチ、 k_{ij} はクラスラベル、 \mathbf{d}_{ij} は パッチ画像位置からの物体中心へのオフセット量、 o_{ij} はパッチの回転角である.あらかじめ教師信号がある画 像から Bag を作成するため、ポジティブサンプルとネ ガティブサンプルが混同するような Bag は作成しない.

ポジティブサンプルは、検出対象の同一角度におけ る画像の近傍領域から切り出したパッチ画像群を1つ の Bag とする. 識別に有効でないポジティブサンプル の重み付けを目的としているため、ネガティブサンプ ルは、非検出対象の画像から切り出した1枚のパッチ 画像を1つの Bag とする.





図 3 Bag の構築とサブセットの作成

3.2 MI-Hough Forest 学習の流れ

MI-Hough Forest の学習過程を図4に示し,各処理 過程を以下に述べる.



図 4 MI-Hough Forest の学習過程

Step1:重みの初期化 学習サンプルの重み $w_{ij}^{(d)}$ を $w_{ij}^{(0)} = 1/N$ に初期化する。学習サンプル群から $T(t = 1, 2, \dots, T)$ 個のサブセットを作成する。サブセットは、図3に示すように学習サンプル集合からランダムサンプリングすることにより作成する。作成したサブセットを用いてT本の決定木を構築する。決定木は、並列処理により構築する。

Step2:分岐関数の候補 階層 d における分岐関数の候

補 $h^{(d)}$ には、学習パッチ画像Tとテンプレートパッチ Tとの類似度S(T,T)としきい値 τ により式(1)のよう に定義する.

$$h_{T,\tau}^{(d)}(\mathcal{I}) = \begin{cases} 0 & \text{if } S(\mathcal{I},T) < \tau \\ 1 & \text{otherwise} \end{cases}$$
(1)

テンプレート T は分岐ノードにおけるポジティブサン プル群から選択される.図5に示すように学習パッチ Tは、テンプレート T との類似度がしきい値 τ 以下の 場合は左、それ以外は右に分岐される.

Step3:分岐ノードの決定 分岐関数候補の中から最適 なものを決定する.式(1)におけるパラメータTはポ ジティブサンプル群から選択されたテンプレートパッ チを表す.パラメータ τ は、学習パッチ画像Tとテン プレートTの類似度と比較するためのしきい値である. 以上の2パラメータ(T, τ)をランダムで選択し、最適 なパラメータを選択する.評価関数 U_* を式(2)に示す.

$$\underset{T,\tau}{\arg\min(U_{\star}(\{p_i|h(\mathcal{I}_i)=0\})+U_{\star}(\{p_i|h(\mathcal{I}_i)=1\}))}$$
(2)

 ${p_i | h(\mathcal{I}_i) = 0}$ は左の子ノードに分割されたサンプル集合, ${p_i | h(\mathcal{I}_i) = 1}$ は右の子ノードに分割されたサンプル集合である.評価関数 U_* には,以下の2つの基準を階層毎に切り替えて用いる.一つ目は,エントロピーを用いて,各子ノードに分岐したサンプル集合Aにつ



図5 分岐ノードにおけるサンプルの分割方法

いて式(3)のように定義する.

 $U_1(A) = |A|(-c \cdot \log c - (1 - c) \cdot \log(1 - c))$ (3)

ここでcはサンプル集合 Aに含まれるポジティブサン プルの割合であり、サンプルの重み w_{ij} を用いて式 (4) から算出する.

$$c = \frac{\sum_{i \in S_n j \in S_n \land k_{ij} = 1} w_{ij}^{(d)}}{\sum_{i \in S_n} \sum_{j \in S_n} w_{ij}^{(d)}}$$
(4)

二つ目はオフセット量 \mathbf{d}_{ij} のばらつきを評価する関数 であり式 (5) のように定義する.

$$U_2(A) = \sum_{i:k_{ij}=1} (\mathbf{d}_{ij} - \mathbf{d}_A)^2$$
(5)

 \mathbf{d}_A はオフセット量 \mathbf{d}_{ij} の平均値であるので U_2 は オフセット量の分散に相当する。各決定木において、 Step3~Step4 を階層 d における全てのノードが作成 されるまで繰り返す。

Step4:サンプルの重み更新 各決定木における階層 dの全ノードを作成後,各子ノードにおいて学習サンプ ルの重みを更新する.重みの更新は,ポジティブサン プルのみを対象とする.ポジティブサンプルの重み w_{ij} は,ポジティブサンプルのクラス尤度 p_{ij} と Bag のク ラス尤度 p_i の積より更新する.ポジティブサンプルの クラス尤度 p_{ij} を式 (6)により求める.

$$p_{ij} = \frac{1}{1 + \exp(1 - \mathcal{F}(\mathcal{I}_{ij}))}$$
 if $k_{ij} = 1$ (6)

 $\mathcal{F}(\mathcal{I}_{ij}) = 2c - 1$ と定義し、ノード内のポジティブサン プルの割合 *c* が高いほど p_{ij} は高くなり、逆にポジティ ブサンプルの割合が低いほど p_{ij} は低くなる. Bag のク ラス尤度 p_i は Bag に属するサンプル集合のクラス尤度 を用いて式 (7) より計算する.

$$p_i = \frac{1}{J} \sum_{j \in B_i} p_{ij} \quad \text{if} \qquad k_{ij} = 1 \tag{7}$$

ポジティブサンプルの重み w_{ij}^{d+1} は, p_{ij} , p_i を用いて式 (8) より計算する.

$$w_{ij}^{(d+1)} = \begin{cases} p_i \cdot p_{ij} & \text{if } k_{ij} = 1\\ w_{ij}^{(d)} & \text{otherwise} \end{cases}$$
(8)

最後に,ノード内のポジティブサンプルの重みを式(9) により正規化する.

$$w_{ij}^{(d+1)} = \frac{w_{ij}^{(d+1)}}{\sum_{i \in S_n j \in S_n \land k_{ij} = 1} w_{ij}^{(d+1)}}$$
(9)

ポジティブサンプルの重み $w_{ij}^{(d+1)}$ は, Bag 内にクラス 尤度 p_{ij} が低いポジティブサンプルが多く含まれている と低くなり,逆にクラス尤度 p_{ij} が高いポジティブサン プルが多く含まれていると高くなる.

Step5:Step2~Step4の繰り返し Step2~Step4の ノードの作成と重み更新を,サンプル数が一定数未満 になるか,もしくは指定した深さDに達するまで繰り 返す.階層が深くなるにつれ,重みが低いサンプルの 影響を低減した分岐構造を持つ決定木が構築される.

Step6:末端ノードの作成 末端ノード*L*には,ポジ ティブサンプルの重み付き割合 C_L と物体中心へのオ フセット量のリスト D_L ,回転角 o_{ij} を保存する.

以上の処理過程により決定木を構築することで, 階 層が深くなる度に非対象物体と類似しているポジティ ブサンプルの重みを自動的に低くすることができる.

3.3 特徴量と類似度計算

本研究では、式(1)の分岐関数において、Dominant Orientation Templates(DOT)[11] による類似度を用い る.DOTは、セルと呼ばれる局所領域ごとに作成した 勾配方向ヒストグラムと、セルの注目画素の勾配方向 により求める.DOT 特徴の抽出例を図6に示す.セル の特徴では、勾配情報を0~360度を22.5度間隔に分け たビンに投票し、しきい値処理することにより2値化 する.セルの注目画素では、0~360度の方向を45度間 隔で分けたビンを用意し、セルの注目画素の勾配方向 に対応するビンに1、それ以外は0にすることで2値化 する.



図7 末端ノードの情報



図 6 DOT 特徴

DOT 特徴による類似度は式 (10) より計算する.

$$S(\mathcal{I},T) = \sum_{\substack{P_m^{\mathcal{I}} \in \mathcal{I} \\ P_m^{\mathcal{I}} \in T}} \delta(P_m^{\mathcal{I}} \otimes P_m^{\mathcal{I}} \neq 0), m = 1, 2, \cdots, M (10)$$

 $P_m^{\mathcal{I}} \ge P_m^{\mathcal{T}}$ はそれぞれ $\mathcal{I} \ge T$ のセル mにおける DOT 特 徴を表している. \otimes は AND 演算子, δ はクロネッカー のデルタ関数であり, $\mathcal{I} \ge T$ の AND 演算をした結果, 1 つでも 1 となるビットがある場合に 1 を返す.

3.4 投票処理

本節では、MI-Hough Forest による投票処理につい て説明する.投票の際には、回転角 θ ごとに投票平面 を用意して3次元尤度マップを構成する.そして、入 力画像の位置yから切り出されたパッチ画像 $\mathcal{I}(y)$ を各 決定木に入力し、各決定木ごとのポジティブサンプル 割合 $P(c|\mathcal{I}(y))$ を得る.そして、各決定木の $P(c|\mathcal{I}(y))$ を辿り着いた末端ノードに保存されている回転角 θ に 対応する投票空間 $V_{\theta}(y)$ に投票する.

$$V_{\theta}(\mathbf{y}) = \sum_{\mathbf{y} \in I(\mathbf{x})} P(c | \mathcal{I}(\mathbf{y}))$$
(11)

作成した尤度マップに対して,以下の手順により物 体の位置と回転角を求める.

- 1. 回転角 θ の平面において局所領域を走査
- 2. 局所領域ごとに総和を求める
- 3. θ平面における局所領域の総和の最大値を探索

- 4.1~3を回転角分(360回)繰り返し処理
- 5. 総和が最大となる平面における θ を回転角,局所 領域の注目点を位置として検出

上記の処理を行うことで,入力画像に対して1回のラ スタスキャンのみで全方向を向いた対象物を検出する ことが可能となる.

4 評価実験

提案手法の有効性を評価するために,従来法との精 度比較を行う.まず,シミュレーション実験により基本 性能の評価し,次に実画像に対する性能検証を行う.

4.1 実験概要

従来法には ZNCC, Chamfer Matching, DOT, Hough Forest を用いる.以下に,従来法の詳細を述べる. **Zero-mean Normalized Cross-Correlation[12]** ZNCC は,入力画像およびテンプレートの輝度値の平 均値とそれぞれの輝度値の差を計算することで,明る さが変動する場合でも安定した検出が可能である.

Chamfer Matching[13] Chamfer Matching は,入 力画像とテンプレート画像間のエッジの相違度に基づ いてマッチングを行う手法であり,距離変換画像を用 いてエッジの相違度が小さくなる方向へ効率的に探索 することが可能である.

Dominant Orientation Templates[11] DOT は, 入力画像とテンプレート画像間の勾配情報に基づく類 似度によりマッチングを行う手法であり,勾配情報を バイナリで表現することにより高速なマッチングが可 能である.

4.2 シミュレーション実験による性能評価

シミュレーション実験では,図8に示すような CAD モデル (図8(a)) から生成した距離画像 (図8(b)) を用 いる.評価には,類似した形状を含む3種類の物体 (物



図8 CADモデルと生成した距離画像

体 A,物体 B,物体 C)を 0~360 度に回転させた距離



図 9 シミュレーションデータにおける実験結果

画像 360 × 3 枚を用いた. 位置に対する真値は対象物 体を画像中心に配置して,画像の中心位置を真値とす る。検出した位置と回転角の誤差がしきい値以内であ れば成功とする. 図9に ZNCC, Chamfer Matching, DOT, Hough Forest, MI-Hough Forest の検出率を示 す. ノイズなしの場合において,図9(a)から物体Bに おいて提案手法による精度向上が最も顕著であった。位 置に対しては、従来法と比べて誤差を許容するしきい 値が 3pixel の場合,平均で7%精度を向上させること ができた. 回転角に対しては, 誤差を許容するしきい 値が3°の場合,平均で8%向上させることができた. ノイズありの場合においても、図 9(b) から物体 B にお ける精度向上が顕著であった. 位置に対しては, 従来法 と比べて誤差を許容するしきい値が 3pixel の場合,平 均で6%精度を向上させることができた。回転角に対 しては、誤差を許容するしきい値が3°の場合、平均で 7%向上させることができた.

以上の結果から,提案手法は学習サンプルへの重み 付けを導入することで高精度な検出が可能であること を確認した.

4.3 実画像に対する性能検証

次に,実画像における性能検証を行う.実画像には, レンジファインダより撮影した点群データを変換した 距離画像を用いる.テストデータは手動のターンテー ブルを約10°間隔で回転させることで撮影する.この とき,レンジファインダから対象物体までの距離は約 50cmである.

学習用画像には、3 種類の物体(物体 A、物体 B、物 体 C)を撮影した1枚の画像を0~360度まで回転させ た画像360×3枚を用意した。画像サイズは500×500 であり、位置に対する真値は検出対象物体にマーカを 張り、マーカ位置を目視により求めた。評価実験では、 位置ずれ誤差が3pixel以内かつ回転角に対する誤差が 3°以内のときに成功として検出率を計算する。図11



図 10 各物体の距離画像

に ZNCC, Chamfer Matching, DOT, Hough Forest, MI-Hough Forest の検出率と検出時間を示す.実画像 においても、実験1と同様に物体 B における提案手法 による精度向上が最も顕著であった.図12に各手法に おける検出例と尤度マップを示す.検出対象が物体 B のとき,従来法では非検出対象の領域に高い値が存在 するが,提案手法では尤度が低く,誤検出を抑制する ことができた.図13に物体 B における階層毎のパッチ の重みの可視化例を示す.図13から,物体 B 以外には 存在していない曲線形状のパッチの重みが高く,それ 以外は重みが低いことがわかる.このことからも提案 手法における学習サンプルの重みが検出に有効に働い ていることがわかる.



図 11 検出結果と処理時間

次に,処理時間については,ZNCC, Chamfer Matching, DOT は回転角ごとに用意された 360 枚のテンプ レート全てと照合を行う必要があるため多大な時間を 要している.しかし, Hough Forest, MI-Hough Forest は決定木により 360 枚のテンプレートを表現している ため,1回のラスタスキャンで各回転の対象物の検出が 可能であり,1.1[sec] で検出することができた.以上の 結果より,実画像実験においても提案手法では高精度 かつ高速な検出ができることを確認した.

5 おわりに

非検出対象と類似した学習サンプルの重みを低くす ることで類似物との誤検出を抑制することが可能な MI- Hough Forest を提案した.提案手法は,従来法に対し て類似物の影響を抑えることができ,従来法よりも対 象物を安定して検出できることを示した.さらに,1つ の Hough Forest で回転変化した学習サンプルを表現す ることができるため,1回のラスタスキャンで検出可能 であり,高速な処理を実現することができた.今後は, 実時間処理に向けてさらなる高速化と視点変化への対 応について検討する予定である.

参考文献

- N. Dalal, and B. Triggs, "Histograms of Oriented Gradients for Human Detection", Computer Vision and Pattern Recognition, vol.1, pp.886-893, 2005.
- [2] C. Cortes, and V. Vapnik, "Support-vector networks", Machine Learning, pp.273-297,1995.
- [3] Y. Freund, and R. E. Schapire, "Experiments with a new boosting algorithm", International Conference on Machine Learning, pp.148-156, 1996.
- [4] B. Leibe, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation ", International Journal of Computer Vision, vol.77, no.1-3, pp.259-289, 2008.
- [5] L. B. Statistics, and L. Breiman, "Random forests", Machine Learning, pp.5-32, 2001.
- [6] J. Gall, and V. Lempitsky, "Class-specific hough forests for object detection", Computer Vision and Pattern Recognition, 2009.
- [7] T. G. Dietterich, and R. H. Lathrop, "Solving the multipleinstance problem with axis-parallel rectangles", Artificial Intelligence, vol.89, pp.31-71,1997.
- [8] 清水彰一,渡邉信太郎,三輪祥太郎,関真規人,平位隆史,藤吉弘 亘,"Hough Forests を用いた物体検出のための学習サンプル 設計法", Vision Engineering Workshop, 2012.
- [9] 櫻本泰憲,斎藤正孝,橋本学,"類似物との識別に有効な画素 群を用いた高速画像照合アルゴリズム", Vision Engineering Workshop, 2012.
- [10] D. Comaniciu, and P. Meer, "Mean Shift Analysis and Applications", in IEEE International Conference on Computer Vision, pp.1197-1203, 1999.
- [11] S. Hinterstoisser, V. Lepetit, S. Ilic, P. Fua, and N. Navab, "Dominant orientation templates for real-time detection of texture-less objects", Conference Computer Vision and Pattern Recognition, pp.2257-2264,2010.
- [12] "テンプレートマッチングの魅力", http://isl.sist.chukyou.ac.jp/Archives/SSII2013TS-Hashimoto.pdf.
- [13] H. G.Barrow, J. M.Tenenbaum, R. C.Bolles, and H. C. Wolf, "Parametric correspondence and chamfer matching:Two new techniques for image matching", Proceedings of the 5th International Joint Conference on Artificial Intelligence, pp.659-663, 1977.



図 13 階層ごとの重み



