HOG(Histograms of Oriented Gradient)特徴量を用いた船舶 画像認識

松本 洋平¹

Ship Image Recognition using HOG

Yohei MATSUMOTO

Abstract

HOG (Histograms of Oriented Gradient) is a feature which is a basis of the modern object detection algorithms. I applied HOG-SVM (Support Vector Machine) detector to our ship detection system and evaluated it quantitatively. Firstly, I gathered large number of navigational images from ship mounted camera and manually annotated the area of ships in the images as the ground truth for training and evaluation. Then I carried out k -fold cross validation to evaluate HOG-SVM for various parameters. The result shows that HOG-SVM achieves 99% true positive and 0.0% false positive for ships with resolutions larger than 128x64.

Keywords : Artificial intelligence, Image recognition, HOG, SVM

キーワード: 人工知能, 画像認識, HOG, SVM

1. はじめに

船舶の衝突事故要因の多くは見張り不十分であるとされている⁽²³⁾。見張りにおいては船員自身の五感、レーダー・AIS (Automatic Identification System) 等の航海計器を駆使して情報収集が行われるが、目視によって得られる情報はとりわけ重要である。陸上に比して変化に乏しい海上に注意を持続することは難しく、人間にとっては過酷な労働である。

また、近年、我が国においては離島周辺の監視の重要性が増しており、我が国のような広大な領海を有する国家においては、監視のための船舶、人員を十分に確保することは困難である。

本研究はコンピュータビジョンによって船舶の自動的な監視を実現することを目標としており、達成されれば、上記の状況が改善され、船舶の運航の安全、領海の保安に役立つと期待する。

コンピュータビジョンを用いた船舶の監視タスクは、大きく2つの段階に分けられる。第一に撮影し

た画像中に船舶を発見すること、第二に発見した船舶を画像上で追尾し危険性の有無を判定することである。一般に、第一の部分は物体検出、第二の部分は物体追尾と呼ばれる。本研究は第一の部分である物体検出、すなわち、船舶検出に集中するものである。

本稿では船舶に搭載したカメラで撮影した画像を航海画像と呼ぶこととする。

航海画像からの船舶検出手法は、本学会の中でもいくつか発表されている⁽¹⁾⁻⁽⁹⁾。一方で、コンピュータビジョンの技術領域では船舶に限らない物体検出の技術が深く研究され非常に重厚な議論、評価がなされている⁽¹⁰⁾⁻⁽²²⁾。

物体検出の研究においては窓走査方式(Sliding Window Approach)が有力となっているが^{(13),(20)}、現在のところ、船舶検出に窓走査方式を適用した例は調査した文献の範囲では見つけることができなかった。

また、特定の撮影条件における特定の物体の検出

¹ 正会員 東京海洋大学大学院 海洋科学技術研究科 (〒135-8533 東京都江東区越中島 2-1-6)
matumoto@kaiyodai.ac.jp

を一般に特定物体検出と呼ぶ。特定物体検出手法においては、一般的な窓走査方式を踏襲しつつ、それぞれの状況に特化した改善を加える場合が多い。特定物体検出の例としては、車載カメラからの歩行者の検出が挙げられる。本研究の興味である航海画像からの船舶検出も特定物体検出に分類しうるが、その研究は上記の本学会に発表されている文献⁽¹⁾⁻⁽⁹⁾の他に例を見ない。

本研究では、特定物体検出として船舶検出手法を追求するが、それに先立って、一般的な手法として確立している HOG (Histograms of Oriented Gradient) と SVM (Support Vector Machine) を用いた窓走査方式の物体検出手法を船舶検出に適用し、基礎的な評価を行う。

物体検出アルゴリズムにとって重要であることは、検出率が高く、誤検出率が少ないことである。一般的な物体検出の研究においては、人が手作業で作成した正解データと比較して評価するのが普通である。この正解データをベンチマークデータと呼ぶ。従来の船舶検出手法の研究では、その意味において定量的な評価をしていない。本研究では多量の航海画像に対して船舶検出のベンチマークデータを作成し、船舶検出の精度を初めて定量的に評価した。

本研究の学術的貢献は以下の2つにまとめられる。

- (1) 航海画像からの船舶検出に、HOG・SVM を適用しその精度を定量的に評価した。
- (2) 航海画像からの船舶検出のためのベンチマークデータを作成し、船舶検出の定量的な評価を可能とした。

2 章では従来研究として、本学会に見られる船舶検出手法、窓走査方式に基づく物体検出手法、物体検出の評価において重大な役割を果たすベンチマークデータの作成について従来の研究を紹介する。3 章では本研究におけるベンチマークデータの作成について述べ、4 章で実験手順と評価結果を示し、その結果を 5 章で考察する。そして、最後に、6 章でまとめる。

2. 従来研究

従来の船舶検出・物体検出の研究について紹介する。その上で、物体検出の評価に必要なベンチマークデータの作成の取り組みについて解説する。なお、本稿における議論は単一のカメラを用いた船舶の検出に集中する。ステレオカメラや他のセンサ(赤外線カメラ、レーダー、ライダー等)の活用も考

えられるが、今回は扱わない。

2.1 船舶検出の研究

船舶を対象とした物体検出の研究としては時系列画像から船舶の運動情報を抽出し利用したものが多い⁽¹⁾⁻⁽⁸⁾。

文献^{(1)-(4),(6)-(8)}は背景領域のオプティカルフローが類似していることや、船舶領域のオプティカルフローが基本的に時間的に直進するという洞察に基づいて、背景領域から船舶領域を分離し抽出する手法である。オプティカルフローの計算にはテンプレートマッチング、Horn-Schunk 法、Lucas-Kanade 法が用いられている。

文献⁽⁵⁾は背景差分法を用いており、船体動揺による背景領域の運動は、天空領域のテンプレートマッチングによって検出し除去することで対応している。

上記の研究のように、移動体の運動情報を用いることの問題としては、見合い関係によっては運動を抽出できない場合があること、遠方にある移動体は動きが少なく運動の抽出が困難であることが挙げられる。

ところで、後に紹介する窓走査方式の物体検出においても運動情報を補助的に活用する動きはあって、例えば、文献⁽¹⁰⁾はオプティカルフローを時間微分することで、動きの境界(Motion Boundary)を抽出し特徴量とする手法を提案している。これは非常に膨大な計算量を要するが、歩行者の検出においては一定の効果を挙げている^{(11),(12)}。

運動情報を用いない船舶検出の研究としては文献⁽⁹⁾があり、船舶領域の色の分散が背景領域に比して大きくなるという性質を利用している。

2.2 窓走査方式による物体検出

一般的に物体検出においては窓走査方式が成功をしている^{(12),(13)}。ここで窓走査方式と呼んでいるのは、撮影した画像のあらゆる部分画像に対して画像認識アルゴリズムによって対象物の有無を判定する方法である。以下に窓走査方式の物体検出アルゴリズムの擬似コードを示す。

```
detectObject(I, O){
    O = {null};
    for all possible window s (x, y, w, h) in I
        if (classifyObject(I, (x, y, w, h)) == true)
            push(O, (x, y, w, h));
}
```

}

ここで I は入力画像を表し、 O は検出した物体を囲む矩形領域のリストを表す。 O は最初空(null)に初期化され、 I 中に検出対象が存在すれば、最終的に検出対象を含む矩形領域 (x, y, w, h) が格納される。 (x, y, w, h) は矩形を表す 4 値の組であり、左上の座標 (x, y) 、矩形の幅・高さ (w, h) をつなげたものである。ループでは画像 I の内部をあらゆる可能な位置・大きさの矩形領域が検索される。classifyObject は I 中の矩形 (x, y, w, h) の領域の目標物の有無を調べる処理(以下、画像認識)、push はリスト O に検出領域の矩形 (x, y, w, h) を追加する処理を表している。

画像認識(classifyObject)の内容は、通常、(1)部分画像からの特徴量の抽出、(2)機械学習アルゴリズムによる識別の 2 ステップで行われる。

窓走査方式の先駆的研究である文献⁽¹⁴⁾は画像認識のための特徴量として Haar-Wavelet を用い、機械学習アルゴリズムとして SVM を用いた。以来、様々な特徴量と機械学習アルゴリズムの組み合わせが検討され膨大な数の論文が出版されている。

特徴量としては Haar-like 特徴⁽¹⁵⁾や、SIFT (Scale Invariant Feature Transform)⁽¹⁶⁾、SURF (Speed Up Robust Feature)⁽¹⁷⁾、本研究で用いる HOG⁽¹⁸⁾が有名である。HOG は識別対象の画像領域を空間微分し、複数のセルに分割し、セルごとに計算した勾配ベクトルのヒストグラムを特徴量とするものである。

機械学習アルゴリズムとしては SVM の人気が高い。SVM は基本的には線形識別関数を決定する手法であるが、マージン最大化を学習の目的とするところに特徴がある。SVM の他には Boosting とそのバリエーションがしばしば用いられる。

とりわけ HOG と SVM の組み合わせは最新の物体検出アルゴリズムの基礎となっており、これに基づいた多くの研究成果が報告されている。例えば、近年、成功した DPM (Deformable Part Model)⁽¹⁹⁾はパーツ毎の小検出窓を設け、それぞれに HOG 特徴量を計算し、つなぎあわせ、対象物の変形や回転に対応している。

この他にも多くの物体検出の最新の研究は HOG・SVM に基づいており、本研究もまた船舶を対象にした場合の HOG・SVM の応答を調べ、次の発展の基礎を作る。

2.3 学習データ

物体検出アルゴリズムを提案する場合には、その

アルゴリズムの精度を評価する必要があるが、そのためには膨大な数の正解データ(ベンチマークデータ)を用意し、アルゴリズムの検出結果と比較する必要がある。検出アルゴリズムの能力とは、具体的には、対象物の存在する領域を囲む矩形が正しく対象物であると検出される率 TPR (True Positive ratio) と対象物の存在しない領域が誤って対象物として検出される率 FPR (False Positive ratio) のことである。言うまでもなく、TPR / FPR の比が大きい方が優れたアルゴリズムである。これまでの航海画像認識の研究報告ではいくつかの成功例を示すにとどまっており、TPR / FPR を評価していない。

一方、前述の窓走査方式では高次元の特徴量を用い、機械学習によって検出窓の画像を判定することから、機械学習に与える学習データが多量に必要である。ここで、学習データとは既知の正解データのことであり、つまり、評価のために必要となるベンチマークデータと同じデータである。

すなわち、物体検出アルゴリズムの研究の要はベンチマークデータにある。実際、画像認識のコミュニティに置いてはベンチマークデータが数多く作られ、それぞれのアルゴリズムの学習と客観的な評価を可能としている。

PASCAL VOC⁽¹³⁾は最も一般的な物体検出を対象にしたベンチマークデータセットとして知られている。PASCAL VOC のデータは、例えば、画像共有サイトにおいて様々な対象物をキーワードとして集めた画像に、対象物の存在する矩形領域を指定したデータを付したものである。収集した画像の撮影条件は全く固定されておらず、最も難しい物体検出の課題を提示している。

一方で、例えば、車載カメラから撮影した画像からの歩行者や車などの特定物体検出を目的とするベンチマークセットが多数存在している⁽¹⁸⁾⁻⁽²¹⁾。カメラの位置が固定されていること、検出対象物体が決まっていること、などから、それらは PASCAL VOC に比べると容易である。

ところで、陸上交通や人間を対象にした物体検出のためのベンチマークセットは数多く存在するが、海上交通を対象としたものは現在のところ存在していない。これを提供することは本研究の主要な目的のひとつである。

3. ベンチマークデータの作成

ベンチマークデータの作成は、船舶検出アルゴリ

ズムの研究に重要な役割を担っている。以下では本研究で行った船舶検出のためのベンチマークデータの作成について説明する。

3.1 画像の収集

ベンチマークデータを作成する以前に航海画像を収集する必要がある。PASCAL VOC では画像共有ウェブサイトを使って画像を収集しているが、今回のベンチマークデータは航海見張りへの対応を想定しているため、航海中に船橋から撮影した画像を利用した。また、原則として航海見張りにおいては自船の全周囲を監視する必要があるが、今回は多数のカメラを用意することができなかったため、前方を撮影した画像のみ利用した。

カメラは東京海洋大学練習船・汐路丸のフライングブリッジの前方の手すり(高さ約 8.4m)の中央、正船首方向に、水平線が画像の上部 1/3 の位置に映るように設置した。

カメラは 4912×3264 画素、 $23.6\text{mm} \times 15.6\text{mm}$ のセンサを備え、レンズの焦点距離は 18mm である。なお、この設定における水平画角は約 66 度、垂直画角は約 47 度である。

撮影は汐路丸の実験航海の往路(勝どき・館山)にて行い、2012 年 7 月 24 日に 146 枚、2013 年 1 月 30 日に 60 枚、2013 年 3 月 6 日に 371 枚、2013 年 3 月 7 日に 221 枚、2013 年 3 月 8 日に 180 枚、合計 978 枚の画像を得た。2013 年 3 月 7 日の視界がしばしば 1~2 マイル程度までに制限されたことを除けば、概ね明快な画像が撮影された。今回はこの 978 枚の画像を用いてベンチマークセットを作成した。

なお、現実的なベンチマークセットを作成するという観点からは、より様々な気象条件で撮影する必要がある。全周囲の撮影も含めて、ベンチマークセットの元となる画像の充実は今後も継続する課題である。

3.2 ラベリング

画像に対象物の位置データを付していく作業をラベリング、あるいは、アノテーションと呼ぶ。いうまでもなく、これは多くの人手と時間を要する作業である。LabelMe⁽²²⁾は誰もがラベリング作業に参加できるウェブサイトを提供し、ベンチマークセットの充実を図っているが、質を維持できない問題がある。

今回、収集した航海画像へのラベリングに際して

はこの作業を補助するツールを作成した。ツールは、航海画像を順に読み込んで表示し、ユーザが描いた矩形の座標(x, y, w, h)を保存する機能を有している。ラベリングはこのツールを使用して学生 3 名と著者自身が行った。

なお、HOG・SVM は傾いた画像には適用できない。見張りのシステムとして完成させる場合には文献⁽²³⁾で示したような画像の安定化方法を用いることを想定している。ただし、今回のベンチマークデータでは、高解像度の画像を必要とした都合上、撮影機材がビデオカメラではなく、デジタルスチルカメラである。この画像に対してはフレームレートが少ないため文献⁽²³⁾の方法が使えない。したがって、ベンチマークデータの作成において、傾いた画像は手作業で除去した。

3.3 サンプリング

HOG・SVM の学習および評価を行うに当たって、前記ラベルを付された画像から、船舶を含む部分画像と含まない部分画像をサンプリングした。前者をポジティブデータ、後者をネガティブデータと呼ぶ。

今回はすべての船舶に対して単一の学習モデルを構築することを想定しており、入力画像の解像度(以下、窓サイズと呼ぶ)は実験毎に統一した。今回の研究では 16×8 、 32×16 、 64×32 、 128×64 の 4 通りの窓サイズで評価を行なっている。窓のアスペクト比(横/縦)を 2 にしたのは、全船舶の矩形の統計をとったところ、平均アスペクト比が約 2.1 であったためである。

サンプリングに関しては 2 通りの方法を実施した。ここではそれぞれをアスペクト比固定サンプリング、アスペクト比可変サンプリングと呼ぶ。以下、窓サイズを $w \times h$ 、ラベル付けされた船舶領域の矩形を(x, y, w_b, h_b)としてそれぞれについて説明する。

(1) アスペクト比固定サンプリング

$s = \max(w_b / w, h_b / h)$ とする。ポジティブデータはサイズ $s w \times s h$ の矩形の中央に船舶領域が収まるようにサンプルし $w \times h$ に縮小する。ただし、 $w_b < w$ かつ $h_b < h$ の場合は破棄する。ネガティブデータは、サプルしたポジティブデータの近くで、サイズ $s w \times s h$ で 2 回サンプルし $w \times h$ に縮小する。ポジティブデータの近傍に船舶を含まずにサンプル可能な領域がない場合は、ネガティブデータはサンプルしない。

(2) アスペクト比可変サンプリング

Table 1 Number of sampling data

Window Size	Fixed Aspect Ratio		Variable Aspect Ratio	
	Positive	Negative	Positive	Negative
16x8	5598	11100	5287	10498
32x16	4690	9280	3814	7588
64x32	2786	5532	1806	3588
128x64	1128	2216	588	1146

ポジティブデータは矩形(x, y, w_l, h_l)そのままサンプルし、 $w \times h$ に縮小する(アスペクト比が変わる)。ただし、 $w_l < w$ または $h_l < h$ の場合は破棄する。ネガティブデータはポジティブデータの近くで、サイズ $w_l \times h_l$ で2回サンプルし $w \times h$ に縮小する。ポジティブデータの近傍に船舶を含まずにサンプル可能な領域がない場合は、ネガティブデータはサンプルしない。

いずれのサンプリング方法でもネガティブデータは最大2倍の数サンプリングすることになっている。これは、実際に検出窓が走査するとき、窓の中にはネガティブデータが入ることが多く、より多くのネガティブデータを学習しておく方が良い結果につながることに配慮したためである。

それぞれの窓サイズ、サンプリング方法でのサンプル数を Table 1 に示す。解像度が小さいデータ程サンプル数が多くなっている。また、アスペクト比固

定の方が可変の場合に比してサンプル数が大きくなっているが、これは、アスペクト比固定の方がサンプルを破棄する条件が緩いためである。128×64でサンプリングした場合のポジティブ・ネガティブ双方のデータの例を Fig.1 に示す。

今回はサンプル済みの画像に対して HOG・SVM の訓練と評価を行うのでサンプリング方法の違いによって実験時間は変化しないが、実際、窓走査方式で画像全体を走査する場合、アスペクト比固定の窓よりもアスペクト比可変の窓の方がサンプリング対象となる矩形のバリエーションが増えるため処理時間は非常に長くなる。

一方で、アスペクト比可変の場合は窓全体に船舶が満たされるため、アスペクト比固定の場合に比して、HOG・SVM に入力される船舶の情報量が多い。Fig.1 の水上バスが比較になっていてわかりやすいかと思う。アスペクト比固定(a)のデータの水上バスは窓の中で縦方向に 1/3 程度しか情報を持っていない。一方で、アスペクト比可変(c)のデータは窓サイズに合うように横方向が縮んでいる。後に見るが、このことは識別精度を向上させるという点で有利である。

Fig.1 Positive and negative samples of fixed and Variable aspect ratio.

Table 2 Experimental setup

Bench set	Window Size	k	Training Data	Evaluation Data
Fixed AR	16x8	10	B_i	$B \setminus B_i$
	32x16	8	B_i	$B \setminus B_i$
	64x32	2	$B \setminus B_i$	B_i
	128x64	10	$B \setminus B_i$	B_i
Variable AR	16x8	10	B_i	$B \setminus B_i$
	32x16	8	B_i	$B \setminus B_i$
	64x32	2	$B \setminus B_i$	B_i
	128x64	10	$B \setminus B_i$	B_i

4. 評価

4.1 実験方法

3章で説明したベンチマークデータを用いて HOG・SVM を用いた船舶画像認識の評価を行った。

評価において、HOG のパラメータとしては、セルサイズを 4×4 画素、ブロックサイズを 2×2 セル、ブロックストライドを 8 ピクセル、ヒストグラムビン数を 9 とした。この HOG の設定は典型的な設定⁽¹⁰⁾である。SVM はソフトマージン型の線形 SVM とし C パラメータは学習毎に自動最適化を行い、0.1~0.5 の間の値となっている。HOG の計算は C++ でコーディングし、SVM は OpenCV の実装を用いた。

ところで、一般に機械学習の評価においては、汎化性能を調べる必要から、学習に用いるベンチマークデータと、評価に用いるベンチマークデータを分けなくてはならない。そのために、今回は k -fold CV で評価を行った。 k -fold CV とは、ベンチマークデータセット B を k 分割して得た各 $B_1, \dots, B_k$ に関して、 $B \setminus B_i$ を用いて学習を行い、 B_i を用いて評価を行うことを $i=1 \sim k$ で k 回繰り返す、それぞれの評価値の平均を最終的な評価値とする方法である。

本研究における評価では 4 種類の窓サイズ (16x8、32x16、64x32、128x64) とアスペクト比固定とアスペクト比可変の 8 通りの組み合わせでそれぞれに k -fold CV を実施し、TPR と FPR を評価するが、Table 1 に示す通り、窓サイズによってベンチマークデータの数が大きく違う。 k -fold CV は基本的に少ないベンチマークデータを有効利用する方法であるが、学習に用いるデータが多すぎると SVM の学習が収束しない場合がある。

k は実験毎に SVM の学習が収束する限界の値に選ぶこととし、学習データが多い場合は、通常の k -fold CV とは逆に B_i を用いて学習を行い、 $B \setminus B_i$ を用いて評価を行うこととした。Table.2 に評価の設定をまとめる。

Fig.2 TPR vs. FPR (fixed aspect ratio)

Fig.3 TPR vs. FPR (variable aspect ratio).

4.2 結果

Fig.2 と Fig.3 にアスペクト比固定の場合とアスペクト比可変の場合のそれぞれの実験結果を示す。グラフの横軸は FPR、縦軸は TPR で、SVM の決定関数の値の判定しきい値を $-1 \sim 1$ の間で 0.1 ずつ変化させながら 4 種類の窓サイズに関して TPR と FPR の点をプロットしている。(決定関数とは線形識別関数であり、SVM は線形識別関数の値が判定しきい値を上回るか否かでポジティブ・ネガティブの別をつける。)

グラフにおいて、曲線が左上隅に近いほど SVM の識別精度が良いことを表している。

基本的な結論として窓サイズが大きいと識別精度が良いといえる。具体的な数値として TPR/FPR の比が最大であったのは、固定アスペクト比の場合、128x64 で 0.91/0.0、64x32 で 0.61/0.01、32x16 で 0.57/0.02、16x8 で 0.48/0.01、可変アスペクト比の

場合、 128×64 で 0.99/0.0、 64×32 で 0.68/0.01、 32×16 で 0.67/0.01、 16×8 で 0.59/0.02 であった。このことから、可変アスペクト比の場合がいずれの場合も固定アスペクト比の場合に比較して良い結果を出しているといえる。

5. 考察

一般に検出窓で一枚の画像内を走査し判定する場合、判定回数は極めて多い。今回撮影した航海画像に対しては、固定アスペクト比の窓でも画像あたり約 1000 万箇所を判定を行う。すると、例えば FPR が 0.01 であると、誤検出が 10 万程度出現することになる。通常、誤検出は画像あたり 1 回以下にしたい。つまり、FPR は 10^{-7} 以下に収めたい。TPR を 0.8 以上に維持したいとすれば、実用に耐えられる窓のサイズは 128×64 以上ということになる。

ところで、撮影される距離 d にある長さ L の対象物の画像上での大きさ l は焦点距離 f と画素ピッチ p を用いて次の式で表される。

$$l = \frac{fL}{dp}$$

本研究で用いたカメラは $p=4.8\mu\text{m}$ であり、この場合の $L=2.5 \sim 320\text{m}$ の物体の距離と画像上でのサイズの関係を図 4 に示す。図 4 の横軸は距離、縦軸は画像上でのサイズを表している。例えば、幅 2.5m の物体は 100m の距離で 100 ピクセル程度になる。漁船等の幅は数 m であり、 128×64 の窓サイズで検出する場合には、100m 以下に接近しなくては認識できないことになる。目視で識別できなくてはならないのは、主にレーダーや AIS で識別できない小型船舶になるであろうことを考えると、現状の識別精度では、運動性能の高い小型の船舶からの見張りの役にしか立たないと考えられる。

また、今回用いた航海画像はビデオカメラでは得られない高解像度である。見張りへの実用にはビデオレートでの撮影を可能とする必要があり、スーパーハイビジョン(4K, 8K)クラスのビデオカメラが必要になる。

一方で、解像度とフレームレートが向上すると、それだけ計算量も増大する問題がある。現状の航海画像であっても、固定アスペクト比の場合でさえ、 128×64 の窓サイズでは 1 枚あたり 30 秒程度の処理時間を要する。

課題をまとめると、(1)カメラの解像度を向上する、

Fig.4 Relationship between object size and distance.

(2)フレームレートを向上する、(3)アルゴリズムを高速化する、(4)低解像度での認識精度を向上する、の 4 点が挙げられる。ところで、処理時間は窓サイズに比例して決まるため、低解像度での認識精度を向上させることができれば、(1)～(3)のいずれも改善すると期待できる。この点から、(4)の課題は最優先課題ということになる。

窓サイズと識別精度の関係は複雑であり、本研究の結果から低解像度での検出精度を向上する余地があるのか否か論じることは難しい。

窓サイズが大きい場合の識別精度が高いのは、主には対象を識別する情報(=画素数)が多く与えられているためであると解釈できる。しかしながら、窓サイズが大きい場合のベンチマークデータは近景を写したものになる場合が多く、すると、窓サイズが小さいデータに比してコントラストが高い明瞭な画像である場合が多いと想像される。つまり、窓サイズを小さくする場合の難しさには、画素の減少に加えて、コントラストの低下という難しさも加わっている。

ただ、 16×8 の船舶であっても人の目で見て認識できていることから、船舶であることを示す情報が十分含まれており、処理方法次第で、低解像度での識別精度を向上することは可能であると考えられる。

6. まとめ

船舶画像認識に HOG・SVM を適用し、評価を行った。また、評価に先立って、膨大な量の船舶画像のベンチマークデータを作成した。

ベンチマークデータは機械学習のための学習データとしてもその評価のためのデータとしても、船舶検出の研究に欠かせないものである。ベンチマークデータの一般性を確保するため、さらなる充実をは

かりたい。

HOG・SVMの評価においては、高解像度のアスペクト比可変のサンプルに対してのみ実用的な識別精度($TPR/FPR = 0.99/0.0$)を達成している。これは約100mの距離にある漁船を識別できる程度である。

より遠方の船舶に対しても高い識別精度を達成するためには、より低解像度での識別精度を高める必要がある。幸い物体検出に関して極めて多くの研究報告がなされており、今後はそれらの応用を検討しながらよりよい方法を模索して行きたい。

謝辞

この論文はJSPS 科研費 25820419 の助成を受けたものです。データ収集及に協力してくださった汐路丸の乗員の皆様、研究室の学生に感謝の意を表します。

参考文献

- (1) 鈴木充雄・大島正毅：航海環境画像からの航行船舶検出に関する研究，日本航海学会論文集，No.100, pp.13-19, 1999.
- (2) 新保雅俊・大島正毅：フレーム間の照合性を手がかりとする時系列画像からの航行船舶の抽出：日本航海学会論文集，No.109, pp.113-118, 2003.
- (3) 新保雅俊・平澤雅人・大島正毅：時系列画像の複雑背景化における抽出領域のフレーム間対応を用いる航行船舶の抽出，日本航海学会論文集，No. 111, pp.181-191, 2004.
- (4) 新保雅俊・平澤雅人・大島正毅：航海環境時系列画像からの航行船舶の抽出と追跡，日本航海学会論文集，No.113, pp.115-125, 2005.
- (5) 桐谷伸夫・福戸淳司：海上観測画像を用いたフレーム間差分による動体認識，日本航海学会論文集，No. 113, pp.107-113, 2005.
- (6) 新保雅俊・平澤雅人・大島正毅：マルチセンサーにより取得した航海環境時系列画像からの航行船舶の抽出，日本航海学会論文集，No.114, pp.65-72, 2005.
- (7) 周立輝・新保雅俊・大島正毅：航海環境時系列画像からのリアルタイム船舶抽出手法，日本航海学会論文集，No. 117, pp.175-182, 2007.
- (8) 新保雅俊・周立輝・山本茂広：航海見張りを支援する画像処理手法の検討，No.119, pp.91-97, 2008.
- (9) 新保雅俊・周立輝・平澤雅人・大島正毅：HSV分析による航海環境時系列画像からの船舶抽出，日本航海学会論文集，No.116, pp.69-76, 2007.
- (10) N. Dalal, B.Triggs, and C.Schmid: Human Detection using Oriented Histogram of Flow and Appearance, Proc. ECCV, 2006.
- (11) S. Walk, N. Majer, K. Schindler, and B. Schiele: New Features and Insights for Pedestrian Detection, Proc. CVPR, 2010.
- (12) P.Dollor, C. Wojek, B. Schiele, and P.Perona: Pedestrian Detection: An Evaluation of the State of the Art, IEEE Trans. PAMI, vol. 34, no.4, pp.743-761, 2012.
- (13) M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman: The PASCAL Visual Object Classes (VoC) Challenge, J. Computer Vision, Vol.88, pp.303-338, 2010.
- (14) C. Papageorgious and T. Poggio: A Trainable System for Object Detection, J. Computer Vision, Vol. 38, no. 1, pp. 15-33, 2000.
- (15) P.A. Viola and M.J. Jones: Robust real-time face detection, J. Computer Vision, Vol. 57, No. 2, pp.137-154, 2004.
- (16) D.G. Lowe: Distinctive Image Features from Scale-invariant Keypoints, J. Computer Vision, Vol. 60, No. 2, pp.91-110, 2004.
- (17) H. Bay, T. Tuytelaars, and L. V. Gool: SURF: Speeded Up Robust Features, Proc. ECCV, 2004.
- (18) N. Dalal, B. Triggs, and C. Schmid: Human Detection using Oriented Histogram of Flow and Appearance, ECCV, 2006.
- (19) P. Felzenszwalb, D. McAllester, and D. Ramanan: A discriminatively trained multiscale deformable part model, IEEE Trans. PAMI, vol. 32, No. 9, pp.1627-1645, 2010.
- (20) P. Dollar, C. Wojek, B. Schele, and P. Perona: Pedestrian Detection: A Benchmark, ICCV, 2007.
- (21) M. Enzweiler and D. M. Gavrila: Monocular Pedestrian Detection: Survey and Experiments, IEEE Trans. PAMI, 2009.
- (22) B. Russel, A. Torralba, K. Murphy, and W. T. Freeman: LabelMe: A Database and Web-based Tool for Image Annotation, J. Computer Vision, Vol. 77, No.1-3, pp157-173.
- (23) 松本洋平：逆合成法を用いた航海画像の安定化，日本航海学会論文集，No.127, pp.205-214, 2012.
- (24) 海上保安庁：海難事故の現況と対策について（平成24年版），統計資料，2012.