Ship Detection in Optical Satellite Imagery

Saad Rasheed Abbasi[1] Souvik Roy[1] and Gaurav Shalin[1]

Abstract— **This project proposes a processing pipeline for automated detection of ships in optical satellite imagery. Dataset from Kaggle's Airbus Ship Detection Challenge is used for the project. This dataset has 190,000, 768 x 768 pixel images with complex backgrounds of clouds, shore lines, waves and ship-wakes. The large field of view images in the dataset makes saliency detection a necessary first step. The saliency map provides several candidates that can be tested with a classifier. Several feature descriptors are explored, compared and combined to determine the best method for ship detection. Principle Component Analysis (PCA) is employed to reduce the feature size and its results are also investigated. This project achieves a 82% classification score with classical pattern recognition methods.**

I. INTRODUCTION

Automated ship detection is an essential task with applications in maritime security, vessel traffic services and fishing management. Currently, many private organizations operate a Vessel Monitoring System that uses ship-borne equipment to track the movement and location of vessels in the high seas. However, such technology is expensive and not all ships are able carry this technology. This makes the tracking of such ships difficult, if not impossible.

Synthetic Aperture Radar remains the sensor of choice as it is largely invariant to weather changes and there is large amount of data available. In the past decade, widespread use of commercial satellites has dramatically increased the size and resolution of optical satellite imagery. This has enabled a large number of researchers to solve this problem via optical imagery. While the task may seem trivial as it involves detecting a foreground object against a fairly uniform background, the presence of waves, clouds and differing resolutions make the task challenging [1].

II. RELATED WORK

The methods reported in literature can be broadly split into three approaches: threshold based, where the researchers attempt to take advantage of the difference in pixel intensity between the sea and the vessel, statistical methods, which make the assumption about the pixel distributions of the sea and finally classification based methods which attempt to classify the ships based on patterns they exhibit, after extracting some feature descriptors. This paper discusses published work based on the classifier approach.

Among the papers used for literature survey, the state of the art performance is by Tang et al.,2015 [2]. The paper reports an accuracy of 97.58 % on optical spaceborne images (4000 5-m-resolution SPOT 5 panchromatic images with the size of 2000 2000 (in pixels)), where the accuracy is defined as the percentage of correctly detected ships. In the paper, wavelet features are extracted to increase the detection efficiency. After the 2D-DWT (Discrete Wavelet Transform), the original image is decomposed into a low-frequency subband (denoted as LL) and several horizontal/vertical/diagonal high-frequency subbands (denoted as LH, HL, and HH). The singularities of LL are extracted to train the first deep neural network (DNN);as LH, HL, and HH describe the image details in different orientations (horizontal, vertical, and diagonal), these subbands are combined before training the second DNN. Then, the outputs of the two DNNs are pooled for final decision making by an extreme learning machine (ELM).

The focus of this paper, however, is on classical machine learning approaches. Among various classifiers, SVM is the most common classifier used for vessel detection, most prominently by Li and Itti in [3]. The authors have cut broad-area satellite images into small image chips and have analyzed them in two complimentary ways: attention/saliency analysis uses local features and their interactions across space, while gist analysis focuses on global non-spatial features and their statistics. Both feature sets are then used to classify each chip as containing target(s) or not, using a support vector machine. For the saliency feature computation, an image was analyzed along multiple low-level feature channels to give rise to multi-scale feature maps. Ten feature channels were adopted which are as -: intensity, orientation (0 , 45 , 90 and 135 , combined into one orientation channel), local variance, entropy, spatial correlation, T-junctions, L-junctions, X-junctions, endpoints, and surprise. The authors used 18 different feature channels for the gist feature computation.These features included intensity, four orientations (0 , 45 , 90 , and 135), and four L-junctions (0 , 45 , 90 , and 135), four T-junctions (0 , 45 , 90 , and 135), four endpoints (0 , 45 , 90 , and 135), and X-junction. The saliency features and gist features were combined together to form the final saliency-gist feature vector which were then fed to RBF (radial basis function) based SVM to complete the classification task. A ROC area of 0.952 was reported on 300 training samples for the experiment on detecting boats.

Apart from SVM other classical classifiers has been used, as by Johansson in [4], who used random forests to detect small vessels in high resolution optical satellite images from Google Maps. This authors work has the performance of the vessel detection ranging from 85-99% depending on the tolerance for the number of false detections. Shi et.al. in [5] uses Adaptive Boosting for classifying ships. The authors

[1]Department of System Design Engineering, Faculty of Engineering, University of Waterloo, 200, University Avenue, Waterloo, Ontario, Canada, N2L 3G1

first convert the panchromatic image into a fake hyperspectral form to enhance the separability between ships and background. Considering the ship candidates as anomaly, the abnormal and fluctuating spectral vectors are projected while the normal and flat ones are suppressed. Afterward a new extension of histograms of oriented gradients (HoGs) is extracted by combining it with features generated on basis of a Circle Frequency (CF) filter. Finally the hypotheses for ships are generated using AdaBoost algorithm.

III. PROJECT FOCUS

This project proposes a pipeline to detect and localize ships in wide field of view images of the sea acquired with optical satellites. The dataset is explored in Section V. This project compares different feature descriptors and evaluates the results of a combined feature descriptor. The feature descriptors explored are i) histogram of oriented gradients (HoG) ii) local binary patterns (LBP) and iii) color histograms. Pattern recognition is performed by a random forest classifier. This classifier was chosen due to its simplicity, immunity to outliers and robustness to noise [12]. Each feature descriptors best parameters are selected by measuring the classifiers performance. Finally, the performance of the pipeline is evaluated after reduction of the feature vector using PCA.

IV. BACKGROUND

A. Local Binary Patterns

The key in achieving a good generalized classifier is selecting a feature descriptor that provides a fairly unique pattern or signature for a ships image. Texture is an important characteristic of various images varying from microscopic images to multispectral remotely sensed images [6]. Image texture provides information about the physical properties of object such as smoothness or roughness, or difference in surface reflectance [7]. One of the best local texture descriptor is the Local Binary Pattern (LBP). It has been popular because of its resistance to lighting changes and low computational cost. It is a simple and efficient operator that labels each pixel based on the properties of the pixels surrounding this pixel [6]. LBP considers an image to be composed of micropatterns. Binary gradient directions are concatenated to form patterns and finding the first-order circular derivative of these patterns gives the LBP [6].To compute the canonical LBP operator [8] at each pixel location, the values of a small circular neighbourhood around this central pixel (Ic) are considered. It is calculated as follows:

$$LBP(N,R) = \sum_{n=0}^{N-1} s(\bar{I}_n - I_c)2^n$$

Where N is the number of pixels in the neighborhood, R is the radius, and s(x)=1 if x is greater than or equal to 0, otherwise s(x)=0. To describe the texture of the image, the histogram of these binary numbers are used [6]. LBP have been used for ship detection in satellite images as by Haigang, S. et al. in [9], by Song, Z. et al. in [10] and by Yang, F. et al. in [11].

B. Color Histogram

A color histogram represents the distribution of color in an image. The color range or span of an image is divided into multiple ranges and the color histogram gives the number of pixels that have colors in each of these ranges. Unlike intensity histogram which are used for monochromatic images, color histograms are used for RGB or HSV images. As the color histogram can have very compact representation and is computationally simple, it is a popular technique for extracting color features from an image.

C. Random Forest

Random Forest [12] is a supervised learning algorithm. The forest it builds is an ensemble of Decision trees to get better and stable predictions. The algorithm brings extra randomness into the model when the tree grows. Instead of searching for the best feature while splitting a node, it searches for the best feature among a random subset of features. This process creates a wide diversity, which generally results in a better model. Random forest makes it very easy to measure the relative importance of each feature. It measure the feature importance by looking at how much the tree nodes, that use that feature, reduce impurity on average across all trees in the forest. It computes this score automatically for each feature after training and scales the result so that sum of all importance is equal to 1.

Random Forest run-times are quite fast and they are able to deal with unbalanced and missing data. The steps below will show how a Random Forest is trained for number of trees.

- In the first step, we need to sample N cases at random with replacement to create a subset of the data.
- At each node:
  - For a number m, predictor variables are selected at random from all the predictor variables.
  - The predictor variable that provides the best split, according to an objective function, is used to do a binary split on that node.
  - At the next node, another m variables are chosen at random from all predictor variables and the same steps are repeated.

D. Saliency Detection

A saliency map is a topographic representation of saliency which refers to visually dominant locations. The basic idea of salient maps is to plot salient pixels that have unique features than most of the other pixels in the image. Saliency map is generated with the dark pixels representing the salient regions.

In this project we implement the saliency model by Itti [13]. Our saliency maps are based on three features: intensity, color and orientation. The resulting saliency model is created by hierarchical decomposition of the features and their combination to the single map. As the first step to find the saliency map, intensity of image is obtained by converting the image to grayscale. Then for color extraction, the image is converted to red-green-blue-yellow color space and the information about local orientation is extracted using Gabor filter in four angles. The next step is to create gaussian pyramids from which feature maps are created as difference of Gaussian between finer and coarser scales of pyramids. The model creates three conspicuous maps for intensity, color and orientation by combining the created feature maps. The final saliency map is a mean of the conspicuous maps.

E. *Histogram of Oriented Gradients*

The Histogram of Oriented Gradients (HoG) [14] is an efficient way to extract features out of the pixel colors for building an object classification system. In HoG, the distribution of direction of gradient are used as features. Gradient is the change in x and y derivative of an image. The gradients are useful as the magnitude of gradient is very high around edges and corners. This provides information about shape of an object.

The steps below shows how HoG is computed.

- The patch size of an image have to be in fixed aspect ratio.
- The horizontal and vertical gradients are calculated with direction. The magnitude of gradient appears when there is a sharp change in intensity. So it highlights the outlines of an object. And for color image, we choose the pixel with maximum of the magnitude of gradient of the three color channels and corresponding angle.
- The image is divided into cells (8x8) and histogram of gradient is calculated. The 8x8 patch makes the representation of histogram of gradient less sensitive to noise. Since, the ships in our case are smaller, therefore we choose 8x8 for our 768x768 image size. The histogram contains 9 bins corresponding to angles 0, 20, 40 160.
- Normalizing the histogram so they are not affected by lightning variation.
- Calculating the HoG feature vector for the entire image patch concatenating horizontal and vertical image patches.

F. *Principle Component Analysis*

Principle Component Analysis (PCA) is employed to reduce the size of the feature vector. This is done by representing the data accurately in a lower dimensional space while preserving largest variances in data. Since this project aims to combine feature descriptors to generate a more discriminator vector, many of the elements could possibly be redundant or lower the classification accuracy on the test set.

PCA operates by computing the eigenvalues and eigenvectors of a given dataset. The eigenvalues correspond to each feature's variance. The higher the variance, the more that particular feature contributes to distinguishing between different classes. A mapping function W is found out which transforms the original data X_D to a lower dimensional data Y_d such that:

$$Y = W^T X$$

The mapping function W consists of eigenvectors $(w_1, w_2 ...)$ which are called principal components. Each of these eigenvectors correspond to the largest eigenvalues, i.e to maximize the variance, w_1 is the eigenvector corresponding to the largest eigenvalue λ_1, w_2 is the eigenvector corresponding to the second largest eigenvalue λ_2 and so on. These eigenvectors correspond to each row of W.

V. Implementation

A. *Dataset Description*

The dataset chosen for this work consists of 190,000 images of the sea taken from optical satellites. It is unbalanced with approximately 75% of the images with no ship whereas the remaining 25% include one or more ships. In addition, the images can contain ship wakes, clouds, docks, land or shore lines and sea waves which often induce a repeating pattern in the images.

Fig. 1: An image consisting of ships

Figure 1 demonstrates the most clear images that are available in the dataset with one ship. Figure 2 demonstrates the images obscured with clouds. Figure 3 and 4 demonstrates the difficult aspects of the dataset containing images with sea, land, shore-lines, waves and multiple ships.

Fig. 2: Images of clouds obscuring ships - potentially making classification challenging

Fig. 3: Images with Waves, land, dockyards

Fig. 4: Images of land water with high intensity reflections

In order to ensure that the classifier did not misclassify these patterns as ships, it was pertinent to choose a good feature descriptor. Since classical pattern recognition algorithms do not typically require hundreds of thousands of images, the dataset was reduced to 10,000 randomly chosen images. The images were chosen at random, without replacement, ensuring the data was still unbalanced. The dataset labels the pixels which contain the ship via run length encoding, implying that most modern algorithms treat ship detection as a segmentation problem rather than a object detection problem.

B. Pipeline

A random forests classifier is trained by selecting a 64 x 64 region of interest (ROI) around the ships location. While the exact center of each ship is not provided by the dataset, the run-length encoded pixels are converted to rectangular vertices and their center is computed. For each ROI three feature descriptors are computed. The random forest classifier is trained on 9,000 such ROIs. The classifiers performance is evaluated on the full 768 x 768 pixel images. The performance is evaluated in two major steps. First all the interest points in an image are selected via a saliency map. These points are then subsequently classified by the random forests classifier into ‘ship‘ or ‘no-ship‘ categories. The remainder of this section provides details on the pipeline and individual blocks.

1) Saliency Map Extraction: A saliency map emphasizes the interesting regions of an image. The saliency map is computed by first transforming a RGB image into HSV colorspace. The HSV colorspace is much suited to such as tasks as it separates the luminance and chromo components of an image, allowing for independent processing of either of the components. The HSV image is then used to compute a back-projection. The back-projected image is then passed through a mean-shift filter which clusters similar intensity values as one. A mean shift filter convolves the nearby pixels and a spatial mean is computed along with a mean value of the intensities. The resultant image is converted to black & white and its histogram is equalized. This image is then thresholded, with the black regions representing the salient regions.

2) Candidate Selection:: Since the image at this stage is a mask consisting of white and black pixels, contours are detected via OpenCVs contour detection algorithm. A rectangular approximation of each contour is computed. All pixels that are encompassed by this rectangle are extracted and rescaled to a 64 x 64 image.

3) Ship Detection:: For each extracted ROI, the feature descriptors are computed and fed to the random forest classifier. If the classifier predicts the label as ship, a green box is drawn over the region. If the classifier predicts the label as no-ship, a red rectangle is drawn.

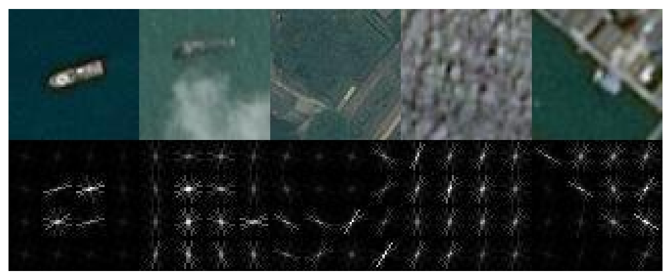

Fig. 5: Visualization of Histogram of Oriented Gradients of Various Candidates Selected by Saliency Map

VI. RESULTS

The experiments were performed with each feature descriptor independently and then combined. As mentioned earlier, 10,000 images were randomly selected from the dataset which also ensured that the original unbalance of the dataset was reflected in the new subset. The smaller subset enabled the authors to collect results at a reasonable rate via 10-times 10-fold cross-validation. Here the results for each component of the pipeline are presented, followed by the final result of classifier. Both correctly and incorrectly classified candidates are shown.

A. Saliency

Figure 6 shows the results of the saliency maps. The black squares are the candidates for further prediction by the classifier. In Figure 7 the saliency map outlines the land and the lone ship in the image.Figure 9 consists of waves with no salient features.

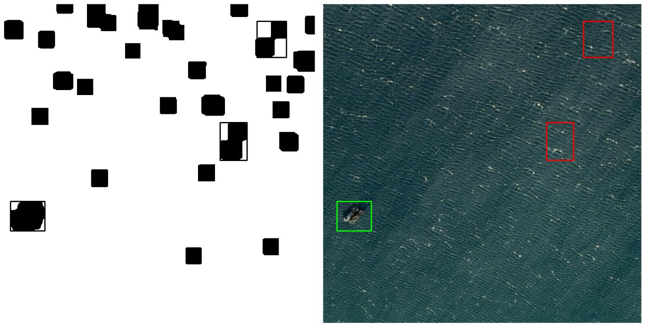

Fig. 6: Saliency map and Classification

Fig. 9: No Salient region and Classification

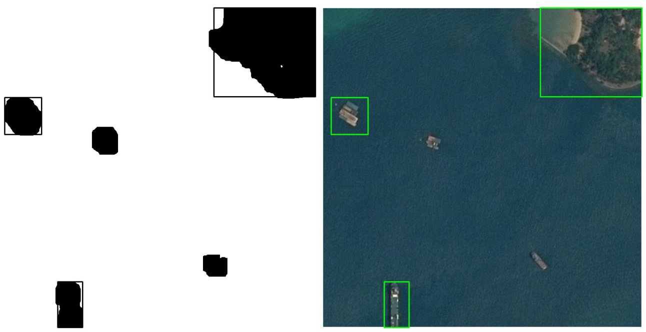

Fig. 7: Saliency map and Classification

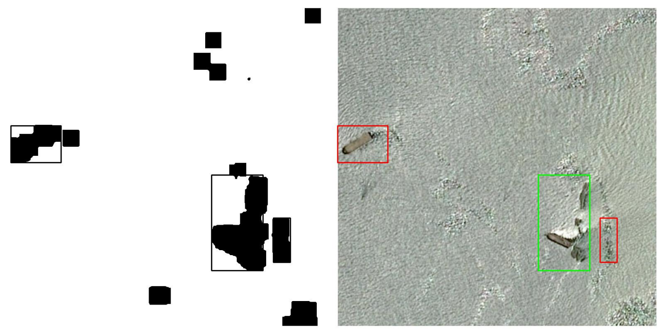

Fig. 10: Misclassification of ship - Salient region and Classification

B. Feature Descriptors

TABLE I: Results with Random Forest classifier with various feature descriptors. Results acquired via 10-times 10-fold cross validation.

No.	Feature Vector	Score	Variance
1.	HoG	66.8	285×10^{-05}
2.	LBP	48.9	402×10^{-05}
3.	Color Histogram	82.7	113×10^{-05}
4.	HoG, LBP	68.9	240×10^{-05}
5.	HoG, Color Histogram	79.0	300×10^{-05}
6.	LBP, Color Histogram	75.0	250×10^{-05}
7.	HoG, LBP, Color Histograms	85.3	99×10^{-05}

Table 1 demonstrates the results of feature engineering. Three feature descriptors are compared and combined. Histogram of oriented gradients, local binary patterns and color histograms were employed because three components dominate the dataset: repetitive patterns in images due to waves, gradients in ROI due to ship edges and color distribution. Thus, this implementation utilizes local binary patterns to discern texture information and histogram of oriented gradients to use information grained from the ship-sea edges. The color histogram allows discernment when there is not enough textual or gradient information, such as clouds. The above results were validated using 10-times 10-fold cross validation.

C. PCA

The combination of all three feature descriptors resulted in feature vector of size 690. To make the feature more manageable in size, PCA was performed. It was determined empirically that a PCA transformation with vector size of 200 performed the best. Table II shows the result.

TABLE II: Classification Accuracy with PCA.

No.	Number of Principle Components	Accuracy
1.	100	77.0
2.	200	81.6
3.	150	81.2
4.	180	79.9
5.	210	79.4

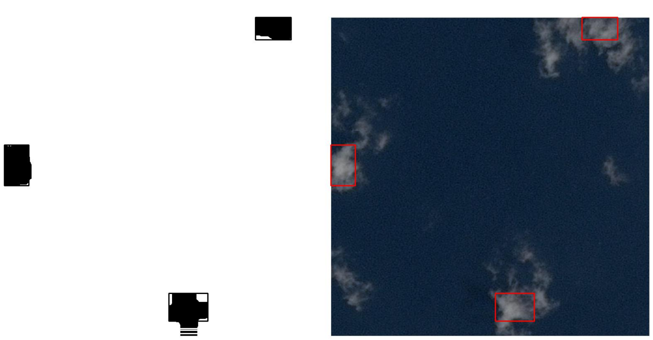

Fig. 8: Image with Clouds - Salient map and Classification

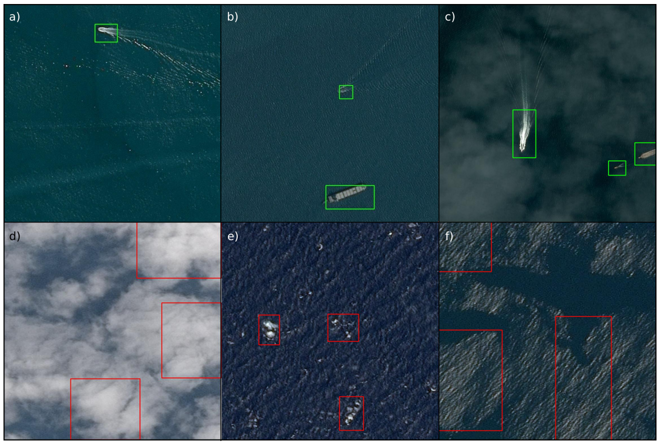

Fig. 11: Demonstration of localization and ship detection via the proposed pipeline. Green rectangles indicate that the classifier considers the region to consist of a ship. Red rectangles indicate that the classifier considers the region to not have a ship.

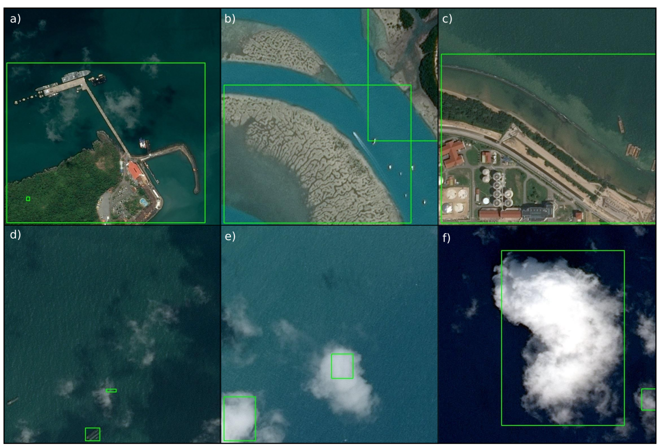

Fig. 12: Misclassified candidates. Green rectangles indicate that the classifier considers the region to contain a ship.

VII. DISCUSSION

Figure 11 and Figure 12 illustrate some typical examples of correctly and incorrectly classified candidates. The pipeline is able to classify small and large ships in the dataset. This is primarily due to the sensitivity of the saliency map and the feature descriptors. Fig. 7a) demonstrations the robustness of the propose framework to intensity changes in the background due to ship wakes as it is able to detect a small ship. Fig. 7b) illustrates the detection of multiple ships in a repeating background, whereas fig. 7c) shows how ships obscured by clouds are still detected. Fig. 7d), e) and f) demonstrate the classifier correctly predicting the regions as 'not-ship' even though the saliency map considers these regions to be important.

The misclassification on sea occurs due to two primary types of pictures: high sharpness images of ocean-waves and clouds. Color histograms showed significant improvement during development of the algorithm however there is still approximately 10% false positives due to clouds and ship wakes. Most of the misclassified examples occur on land and this particular type of error can be easily improved on by incorporating a land-sea segmentation process in the pipeline. The segmentation will ensure that any candidates detected on land can be safely ignored.

From the results it is seen that the combination of feature descriptors is most effective. Utilizing just the histogram of colors is also extremely effective. This is due to the fact that ships are chromatically very different from the sea and clouds. In general, the distribution of colors in images of the sea and clouds consist of accumulation around single or a few peaks whereas the distribution of colors in ships would much spread out. Thus, this feature descriptor performs exceptionally well for most cases. Local binary patterns and histogram of oriented gradients on their own do not perform as well. Local binary patterns, however, in combination with histogram of gradients perform fairly well, posting a score of 70%. This illustrates that textual information and gradient information, on their own, is not sufficient to distinguish between ships and other targets in the images. The primary weakness of HoG and LBP were ocean waves with significant sharpness. Such regions were often highlighted by the saliency map and the classifier often misfired on such targets. This is due to similarity in the high frequency components of ocean waves and ship images. HoG captures the gradients between ship and the surrounding sea, with the straight edges being considerably strong gradients. Similarly, sharp images of ocean waves provided straight lines which are similar to those by ship. The high frequency components of sharp images appear as strong gradients leading the classifier to misclassify such images. Hence, HoG alone is not a good feature descriptor. The parameters for each of the feature descriptors were hyper tuned and validated via 10-times 10-fold cross validation.

These score are achieved for the dataset where there is huge variations in terms of the lighting conditions, or shape, size and orientation of ships, or presence of clouds and land areas which gives a lot of misclassifications. Further, the number of ships to be detected in an image for this project were kept limited to a maximum of two which has also led to decrease in accuracy as some images have more than two ships. Since there are lot of redundant and uninformative features, feeding only the most important features in the classifier after extracting them using PCA has improved the detection accuracy.

VIII. CONCLUSIONS

The results demonstrate that it is possible to achieve considerably accurate classification rate with simple pattern recognition methods up to 85%. The key is to utilize feature engineering and utilize the right classification method.

The feature descriptor of color histogram was found to be better than HoG and Local Binary Patterns for the dataset. Combining all these features, however, gave the best classification score. Since there were many uninformative features, extracting 200 best features (out of 690) using PCA gave the best detection score. Among the classifiers used for this project, Random forest with 200 trees outperformed other classical methods especially SVM with a significant margin, owing to the characteristics of the dataset. To improve the score, further work can be done on segmenting land part in images before finding saliency map. This would allow the classifier not to give false positives on the land areas. This can be a challenge, however, in case the whole image is a land area.

IX. FUTURE WORK

The proposed work illustrates that it is possible to reach significant classification using basic feature engineering and utilizing the right classifier. However, this pipeline fails often due to the saliency map highlighting targets such as waves, clouds and land. As mentioned before, a significant improvement can be made if land-sea segmentation is incorporated into this pipeline. It will allow the framework to exclude any object that are on land from any further considerations.

Another simple method to improve the classification rate would be to utilize ship geometry. Ships are typically long and and narrow and this attribute can be used as a feature vector (i.e. the ratio of a ship's length to width). Such features would aid in distinguishing between a ship and random shapes that the classifer may consider to be a ship otherwise. The challenge in this approach would be segmenting the ship in the ROI.

REFERENCES

[1] Kanjir, Urka, Harm Greidanus, Kritof Otir.,2018. Vessel detection and classification from spaceborne optical images: A literature survey. Remote sensing of environment 207.

[2] Tang, J., Deng, C., Huang, G.-B., Zhao, B., 2015. Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine. IEEE Trans. Geosci. Remote Sens. 53, 11741185.

[3] Li, Z., Itti, L., 2011.Saliency and gist features for target detection in satellite images. IEEE Trans. Image Process. 20, 20172029.

[4] Johansson, P., 2011. Small vessel detection in high quality optical satellite imagery. In: Tech. Report Chalmers University of Technology Sweden.

[5] Shi, Z., Yu, X., Jiang, Z., Li, B., 2014. Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature. IEEE Trans. Geosci. Remote Sens. 52, 45114523.

[6] Brahnam, Sheryl, et al., eds. Local binary patterns: new variants and applications. Springer Berlin Heidelberg, 2014.

[7] Tuceryan, M., Jain, A.K.: Texture analysis. C.H. Chen, L.F. Pau, P.S.P. Wang, (eds.) The handbook of pattern recognition and computer vision, pp. 207248: World Scientific Publishing Co., Singapore (1998)

[8] Ojala, T., Pietikainen, M., Maeenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell.24(7), 971987 (2002)

[9] Haigang, S., Zhina, S., 2016. A novel ship detection method for large-scale optical satellite images based on visual LBP feature and visual attention model. ISPRS. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 917921.

[10] Song, Z., Sui, H., Wang, Y., 2014. Automatic ship detection for optical satellite images based on visual attention model and LBP. In: 2014 IEEE Workshop on Electronics, Computer and Applications, pp. 722725.

[11] Yang, F., Xu, Q., Li, B., 2017. Ship detection from optical satellite images based on saliency segmentation and structure-LBP feature. IEEE Geosci. Remote Sens. Lett. 14, 602606.

[12] Breiman, Leo. Random forests. Machine learning 45.1 (2001): 5-32.

[13] Itti, Laurent, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis Machine Intelligence 11 (1998): 1254-1259.

[14] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1.