https://doi.org/10.2352/ISSN.2470-1173.2019.11.IPAS-266

Multi-Class Detection and Orientation Recognition of Vessels in Maritime Surveillance

Amir Ghahremani, Yitian Kong, Egor Bondarev, Peter H.N. de With,
Eindhoven University of Technology, Video Coding and Architectures (VCA) Research Group; Eindhoven, The Netherlands.

Abstract

For maritime surveillance, collecting information about vessels and their behavior is of vital importance. This implies reliable vessel detection and determination of the viewing angle to a vessel, which can help in analyzing the vessel behavior and in re-identification. This paper presents a vessel classification and orientation recognition system for maritime surveillance. For this purpose, we have established two novel multi-class vessel detection and vessel orientation datasets, provided to open public access. Each dataset contains 10,000 training and 1,000 evaluation images with 31,078 vessel labels (10 vessel types and 5 orientation classes). We deploy VGG/SSD to train two separate CNN models for multi-class detection and for orientation recognition of vessels. Both trained models provide a reliable F1 score of 82% and 76%, respectively.

Introduction

Safety and security of harbors and waterways can be threatened by unknown pathless vessels. Additionally, harbor authorities need to control various operations in the harbors like cargo flows, analyze vessel behavior, passenger transportation, etc. For these reasons, it is crucial to have a surveillance system to automatically and continuously monitor the maritime/harbor environment. These systems process the visual data collected by cameras which are deployed along the shorelines and in harbor areas. However, the monitoring of vessels is a challenging task, due to changing weather conditions including wind, sunshine reflection, fog, water waves, etc., which deteriorate the recognition performance. Furthermore, since the vessels are not the only moving objects in a typical maritime scene, it is important to avoid false detection of irrelevant objects [1, 2].

Generally, a successful multi-class vessel detection is of vital importance for maritime surveillance. For instance, a surveillance system monitoring the cargo flow of a harbor needs to detect vessels that can transport cargo, while urban transportation systems require to identify and observe passenger ships, yachts, or taxi vessels. This implies that vessels need to be classified into multiple types.

In addition to multi-class vessel detection, recognizing the orientation (i.e. the camera viewing angle to a ship) of a detected vessel is also important in many maritime surveillance applications. For example, recognizing the orientation of a vessel can provide supplementary information for vessel behavior analysis and re-identification. In many surveillance domains (e.g. car surveillance), the orientation can be obtained by tracking the movement of the object. However, detecting the vessel orientation is more challenging, since the vessel moving direction is not always aligned with its orientation. Moreover, movement tracking is impossible on individual images.

Conventional object classification methods (e.g. SVM classifiers) [3–5] show an acceptable performance on benchmark datasets. However, these methods do not perform reliably on the real-world data collected from a maritime environment. With the emergence of convolutional neural networks (CNNs) [6–9] and efficient implementations on GPUs, a reliable outdoor multi-class vessel detection becomes applicable. Moreover, since CNNs can learn and detect various specifications of objects, the orientation of a vessel can also be recognized using these networks. However, to the best of our knowledge, a unified multi-class vessel detection and vessel orientation recognition system has not yet been reported in state-of-the-art surveillance systems.

This paper presents a real-time multi-class vessel detection approach for the maritime surveillance domain. The first objective is to reliably detect vessels and classify them in 10 major classes based on visual data collected in any possible maritime environment (i.e. river, lake, canal, harbor, or sea). The second objective is to enable accurate recognition of a vessel orientation into 5 orientation bins.

Finally, the supplementary objective is to capture, label, and publish a novel maritime dataset, containing 10,000 real-world training images and 1,000 validation images representing real-life traffic conditions in various maritime areas. This dataset includes both the classification and orientation labels. The images are captured from various locations in several cities and suburbs in the Netherlands (Amsterdam and Rotterdam) and Turkey (Istanbul). The images of this dataset are extracted from videos captured at different day/year-times from different camera positions and setups and include objects of various sizes, appearing occlusions and truncations.

This paper is organized as follows. The next section provides an overview of the related work and the succeeding section explains the proposed method. The last two sections present the experimental results with validation and conclusions of the paper.

Related Work

In general, CNNs [10–12] provide a high detection rate, robustness, and timeliness. As an important advantage, CNNs can learn any arbitrary attribute of an object. Hence, it is possible to train a CNN to detect an object of interest and even further to recognize a specific property of the detected object. Since we propose a unified multi-class vessel detection and vessel orientation recognition method, this section provides a brief overview on related state-of-the-art work.

Multi-Class Vessel Detection

Recent multi-class object detection methods dominantly use CNNs. Generally, these works focus on either proposing a new CNN or improving existing CNNs, such as Faster RCNN [13] or SSD [14]. For multi-class vessel detection, there are several methods [15–18] focusing on detecting vessels in the synthetic aperture radar (SAR) images. In [15], the authors improve the Faster RCNN by combining the traditional constant false alarm rate method to select better region proposals generated by Faster RCNN. With this, the proposed method improves the accuracy of the predicted vessel locations. The work in [18] proposes SVD-Net, which jointly utilizes the CNN and the singular value decomposition algorithm to learn more discriminative features from the SAR images with clouds and different vessel sizes. Moreover, [17] proposes a new model called S-CNN, embedding an improved saliency detection method. This method results in more accurate detection, especially for small vessels located offshore.

Vessel Orientation Recognition

We have not found any work on vessel orientation detection. However, there are a few vehicle orientation estimation approaches. In [19], the authors introduce a Faster RCNN-based architecture to simultaneously detect the location of vehicles and estimate their orientations. This method adds another classification layer with softmax loss function to predict the orientation on the ground plane.

Single Shot Detector (SSD)

SSD is a feedforward ConvNet that explores the presence of an object instance in the pre-defined default bounding boxes, followed by a non-maximum suppression stage to produce the final detection. This detector allows to omit the region-proposal generation stage, encapsulating all the computations in a single network. As stated in [14], SSD achieves 76.9% mean Average Precision (mAP) on the PASCAL dataset.

Compared to Faster-RCNN based on a region proposal network, the SSD network has a multi-scale and multi-aspect-ratio architecture. This makes SSD a faster network, which is essential for real-time surveillance applications. Additionally, this architecture offers a better performance on smaller objects [20], which is essential for outdoor surveillance. Therefore, in this paper, we extend the SSD network to address the multi-class vessel detection and vessel orientation recognition problems.

Classification and Orientation Recognition

In this paper, we propose a unified vessel classification and orientation analysis method for real-world maritime surveillance applications. The method uses images extracted from the videos as input. It detects vessels in these images and provides the class and orientation labels. This section presents the proposed methods. First, the multi-class detection approach is explained. Then, the vessel orientation recognition is presented.

In this work, we deploy a Single Shot Detector (SSD) to perform multi-class vessel detection and vessel orientation recognition. For the model generation, we train a VGG [21] multi-class vessel model on our new real-world vessel dataset, to learn the following 10 major classes: container ship, river cargo, tanker, passenger vessel, yacht, sailing vessel, fishing vessel, tug vessel, taxi vessel, and small boat. The annotated dataset is captured at various locations in several cities and suburbs in the Netherlands (Amsterdam and Rotterdam) and Turkey (Istanbul). The images are extracted from videos captured in different day/year-periods from different camera positions and setups. The dataset includes vessels of various sizes, appearing occlusions and truncations. The dataset includes images from all possible water region types, i.e. lakes, rivers, canals, harbors, and seas. The original video streams have a resolution of 1080×1920 pixels and are downscaled to 512×512 pixels to match the detector input. The resulting public dataset includes 10,000 outdoor images and 28,260 bounding boxes for vessels.

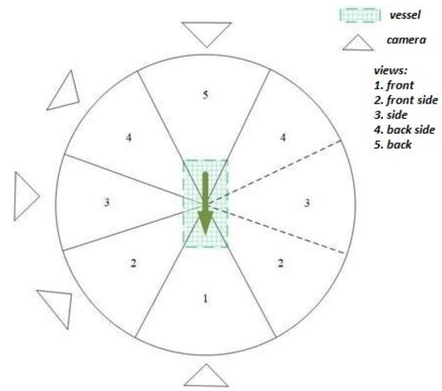

Figure 1: Five defined orientation bins.

Besides this, we train a vessel orientation model that is able to classify the orientation of a vessel into the 5 following classes: back view, back-side view, side view, front-side view, and front view. Fig. 1 illustrates these orientation bins. We also annotate the published dataset according to these 5 orientation classes.

Furthermore, we annotate and generate two separate evaluation datasets, to test the 10-class vessel detection and the 5-class vessel orientation recognition accuracy. Both evaluation datasets comprise 1,000 real-world vessel images with 2,818 labels. As part of the contribution, we provide these four annotated datasets to open public access [22].

Empirical Validation

In this work, we deploy two VGG/SSD networks for multi-class and orientation detection, named as SSD-classification and SSD-orientation, respectively. We have trained these two models using our annotated datasets for 240,000 iterations. Both models are evaluated with a confidence score threshold of 0.5.

Table 1 illustrates the experimental results. The SSD-classification has categorized 2,030 vessels correctly while allowing only 124 miss classifications. The computed mAP, recall, and F1-measurement of the vessel classification model are 0.94, 0.73, and 0.82, respectively. The SSD-orientation model has resulted in 1,801 correct orientations, at the cost of 318 false classifications. The mAP, recall, and F1-measurement obtained by the orientation detection model are 0.85, 0.68, and 0.76, respectively. Fig. 2 illustrates several resulting examples of the classifiers.

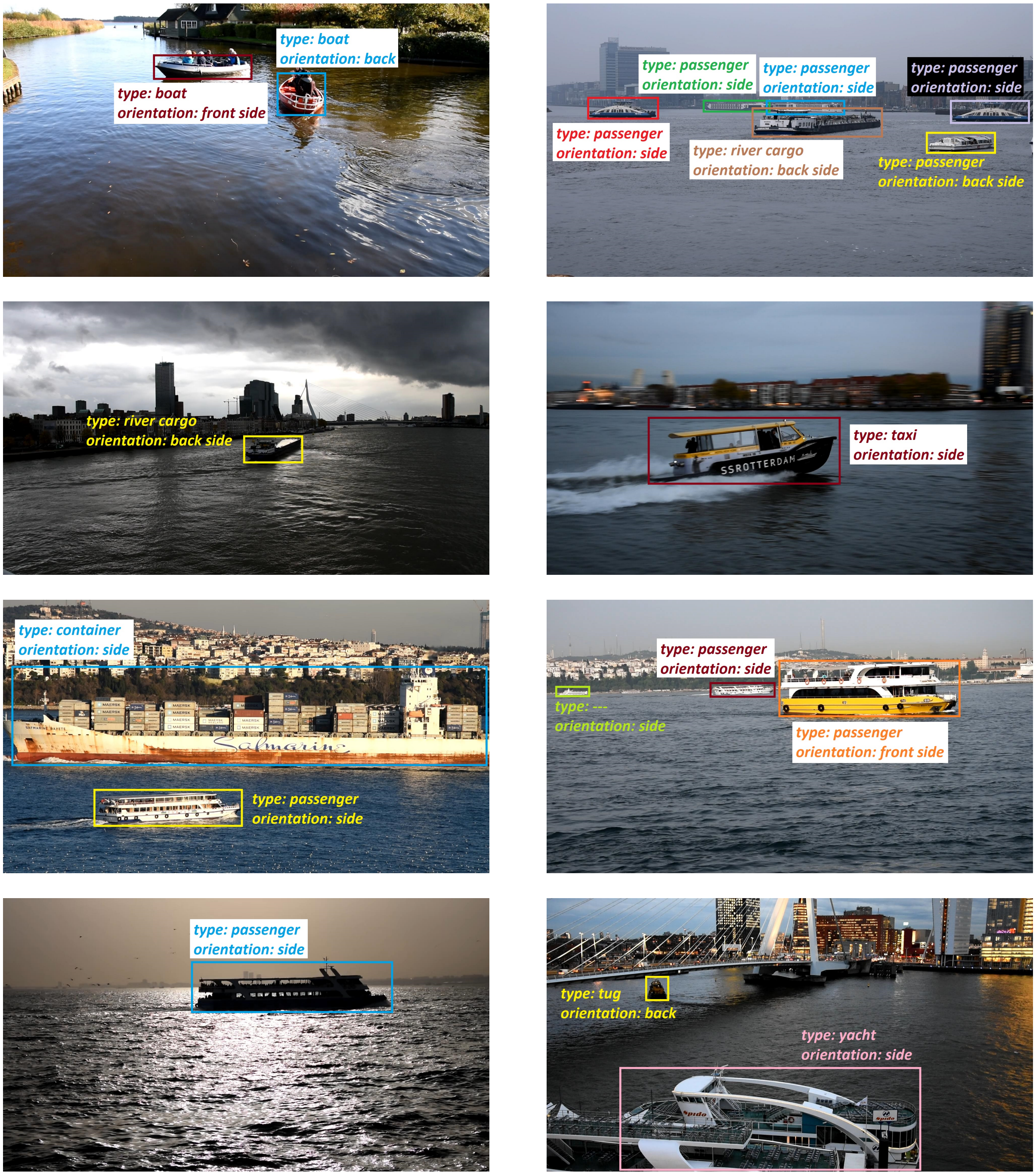

Figure 2: Eight multi-class vessel detection and vessel orientation recognition examples provided by our approach.

Conclusions

This paper presents experiments on applying a VGG/SSD network for both multi-class vessel detection and orientation recognition. The parallel network models were trained and validated using four novel annotated datasets with images obtained in several types of maritime regions (harbors, rivers, canals, sea, etc.), encompassing a vast variety of camera settings and weather conditions. The obtained vessel classification and orientation recognition models perform reliably with an F1 score of 82% and 76%, respectively. The orientation detection is instrumental for later vessel behavior analysis. These results reveal the feasibility of the SSD/VGG networks for training robust CNN models to detect multiple vessel classes and orientations. The obtained detectors based on these networks are fast and reliable and therefore suited for real-time implementations which are typically desired in surveillance applications.

Table 1: Multi-class vessel detection and vessel orientation recognition performance.

	SSD-classification	SSD-orientation
TP	2,030	1,801
FP	124	318
FN	754	847
mAP	0.94	0.85
Recall	0.73	0.68
F1	0.82	0.76

Acknowledgement

This work is partly funded by the European ITEA PS-Crimson and the H2020 Interreg PASSANT projects.

References

[1] X. Bao, Zinger, S., de With, P. H. N., Wijnhoven, R. G. J., and Han, J., "Water Region and Multiple Ship Detection for Port Surveillance", in Proc. the 33rd WIC Symposium on Information Theory in the Benelux, pp. 20-27, 2012.

[2] A. Ghahremani, E. Bondarev, P.H.N. de With, "Water Region Extraction in Thermal and RGB Sequences Using Spatiotemporally-Oriented Energy Features," IS&T Electronic Imaging - Algorithms and Systems, 2017.

[3] A. Andreopoulos and J. Tsotsos, "50 Years of Object Recognition: Directions Forward," Computer Vision and Image Understanding, vol. 117, pp. 827—891, August 2013.

[4] R. Muralidharan, and C. Chandrasekar. "Object recognition using support vector machine augmented by RST invariants." International Journal of Computer Science Issues (IJCSI) 8, no. 5, 2011.

[5] Uijlings, Jasper RR, Koen EA Van De Sande, Theo Gevers, and Arnold WM Smeulders. "Selective search for object recognition." International journal of computer vision 104, no. 2, 154-171, 2013.

[6] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788. 2016.

[7] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Region-based convolutional networks for accurate object detection and segmentation." IEEE transactions on pattern analysis and machine intelligence 38, no. 1, 142-158, 2016.

[8] Lin, Tsung-Yi, Priyal Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. "Focal loss for dense object detection." IEEE transactions on pattern analysis and machine intelligence, 2018.

[9] Herrmann, Christian, Miriam Ruf, and Jürgen Beyerer. "CNN-based thermal infrared person detection by domain adaptation." In Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything, vol. 10643, p. 1064308. International Society for Optics and Photonics, 2018.

[10] Kang, Kai, Hongsheng Li, Junjie Yan, Xingyu Zeng, Bin Yang, Tong Xiao, Cong Zhang et al. "T-cnn: Tubelets with convolutional neural networks for object detection from videos." IEEE Trans. Circuits and Systems Video Technol. vol. 28, no. 10, 2896-2907, 2018.

[11] Gao, Hongbo, Bo Cheng, Jianqiang Wang, Keqiang Li, Jianhui Zhao, and Deyi Li. "Object Classification using CNN-Based Fusion of Vision and LIDAR in Autonomous Vehicle Environment." IEEE Transactions on Industrial Informatics, 2018.

[12] Li, Junliang, Hon-Cheng Wong, Sio-Long Lo, and Yuchen Xin. "Multiple Object Detection by a Deformable Part-Based Model and an R-CNN." IEEE Signal Processing Letters 25, no. 2, 288-292, 2018.

[13] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances Neural inform. proc. syst., pp. 91-99. 2015.

[14] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. "SSD: Single shot multibox detector." In European conference on computer vision, pp. 21-37. Springer, Cham, 2016.

[15] Miao Kang, Xiangguang Leng, Zhao Lin, and Kefeng Ji. A modified faster r-cnn based on cfar algorithm for sar ship detection. In Remote Sensing with Intelligent Processing (RSIP), 2017 International Workshop on, pages 1–4. IEEE, 2017.

[16] Guang Yang, Bo Li, Shufan Ji, Feng Gao, and Qizhi Xu. Ship detection from optical satellite images based on sea surface analysis. IEEE Geosci. Remote Sensing Lett., 11(3):641–645, 2014.

[17] Ruiqian Zhang, Jian Yao, Kao Zhang, Chen Feng, and Jiadong Zhang. S-cnn-based ship detection from high-resolution remote sensing images. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 41, 2016.

[18] Zhengxia Zou and Zhenwei Shi. Ship detection in spaceborne optical image with svd networks. IEEE Transactions on Geoscience and Remote Sensing, 54(10):5832–5845, 2016.

[19] Carlos Guindel, David Martin, and Jose Maria Armingol. Joint object detection and viewpoint estimation using cnn features. In Vehicular electronics and safety (ICVES), 2017 IEEE international conference on, pages 145–150. IEEE, 2017.

[20] A. Ghahremani, E. Bondarev, P.H.N. de With, "Cascaded CNN Method for Far Object Detection in Outdoor Surveillance", 14th International Conference on Signal Image Technology and Internet Based Systems, Spain, 2018.

[21] K. Simonyan, and A. Zisserman, Very deep convolutional networks for large-scale image recognition, NIPS, 2015.

[22] http://vca.ele.tue.nl/

Author Biography

Amir Ghahremani received his BSc. degree in Electrical Engineering with emphasis on telecommunication from Azad University, Urmia, Iran, in 2009. This was followed by a MSc. degree in Electrical Engineering with emphasis on Electronics at the Khajeh Nasir Toosi University of Technology, Tehran, Iran, in 2014. Since 2015, he is working as a PhD at the Eindhoven University of Technology (TU/e), the Netherlands. His research interests include computer vision, semantic content analysis, and machine learning.

Egor Bondarev obtained his PhD degree in the Computer Science Department at TU/e, in research on performance predictions of real-time component-based systems on multiprocessor architectures. He is an Assistant Professor at the Video Coding and Architectures group, TU/e, focusing on sensor fusion, smart surveillance and 3D reconstruction. He has written and co-authored over 50 publications on real-time computer vision and image/3D processing algorithms. He is involved in large international surveillance projects like APPS and PS-CRIMSON.

Peter H.N. de With is Full Professor of the Video Coding and Architectures group in the Department of Electrical Engineering at Eindhoven University of Technology. He worked at various companies and was active as senior system architect, VP video technology, and business consultant. He is an IEEE Fellow, has (co-)authored over 400 papers on video coding, analysis, architectures, and 3D processing and has received multiple papers awards. He is a program committee member of the IEEE CES and ICIP and holds some 30 patents.