Classifying Vessels in Inland Waters Using Live Video Streams

Tomasz Hyla
Marine Technology Ltd.
Szczecin, Poland
e-mail: t.hyla@marinetechnology.pl

Natalia Wawrzyniak
Maritime University of Szczecin
Szczecin, Poland
e-mail: n.wawrzyniak@am.szczecin.pl

***Abstract*— Video surveillance systems used in vessels' traffic monitoring are being improved to eliminate the need for operator supervision. One desirable feature of such systems is the ability to count the number of passing vessels and determine their types. The paper presents straightforward algorithms that can determine the category of a vessel based on partial results of classification. The algorithms take as an input a series of vessel images acquired directly from the video stream by the detection module. Some of the images contain an inaccurately cropped ship or a part of it, and there are also occasions when ships or other objects in the background are erroneously cropped. The algorithms were implemented and experimental results from 100 video samples in four different qualities are presented. The results show that the method is suitable for a practical application.**

Keywords-vessel; classification; video surveillance; traffic monitoring; inland waters.

I. INTRODUCTION

Vessels' traffic is monitored on inland waters using video surveillance systems. These systems are being improved to eliminate the need for operator supervision. One desirable feature of such systems is the ability to count the number of passing vessels and determine their types. With the ability to assess the type of incoming vessel it is also possible to implement additional actions that can be undertaken depending on the detected type of a unit. For example, the movement of barges can be monitored to better plan the opening of bridges and locks, and recreational boats can be sent proper notifications.

This work is a part of the Ship Recognition (SHREC) project [1]. Its objective is to develop a system for vessel detection, identification, and tracking. Additionally, classifiers based on deep learning techniques have been developed that allow the vessels to be categorized into a series of types. These classifiers learned from images of vessels that were clipped out from previously recorded video streams. As the detection system returns a series of images of a ship with the same ID [2], the following question arises: what would be the quality of the classification if one tried to select a single class for the series of images? Additionally, the detection method sometimes inaccurately cuts the ship from the video frame and various artifacts appear in the background. There were no such images in the learning sets. Figure 1 shows how a picture with a vessel can be cut out of a video frame in such cases.

Together with the latest advances in computer vision for object detection and recognition, based closely on the development of deep learning techniques, massive progress was done in the field of ships identification and tracking in recent years. Traditional methods of classification focused on choosing key vessel's features, e.g., as in [3] or [4]. They obtained good results, however with small hull differences between vessels types and complex background of the environment, the practical use of these methods was limited. Then, with the development of Convolutional Neural Networks, such as Visual Geometry Group (VGG) [5], GoogLeNet [6], ResNet [7] and their wider accessibility, new methods were developed. The research works presented in [8] and [9] emphasize the latest focus on detection and classification methods using Convolutional Neural Networks (CNNs) for the purpose of ship recognition. One of the major problems in this field is building a large training sample set, because obtaining enough different images of different units is problematic. There is also no publicly available test dataset containing a wide variety of vessels, especially inland and recreational. This issue was raised in [10]-[12]. The latest research works use mainly end-to-end solution based on regression deep convolution networks, such as You Only Look Once (YOLO) [13][14].

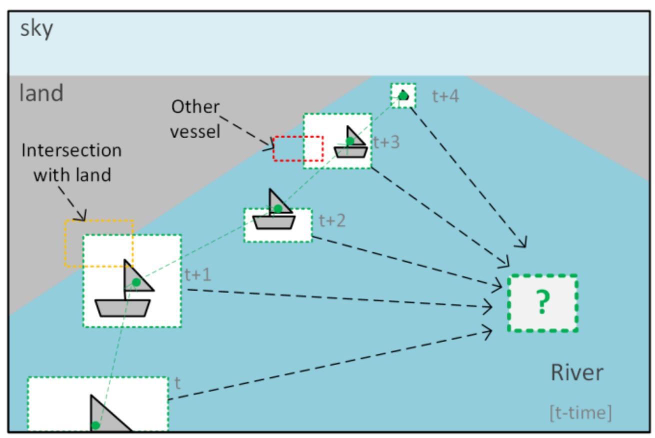

Figure 1. The problem of determining the class of a vessel on the basis of a series of images.

This paper presents straightforward algorithms that can determine the category of a vessel based on partial results. The algorithms take as an input a series of images of the vessel acquired directly by the detection module from the video stream. The category of a vessel for each image is determined by the classification module and, based on these results, the final category is computed. Some of the images contain an inaccurately cropped ship or part of it, and there are also occasions when ships or other objects in the

 ISBN: 978-1-61208-858-7

background are erroneously cropped. We present experimental results from 100 video samples in four different qualities, which show that the method is suitable for practical use.

The rest of this paper is organized as follows. In Section 2, the algorithms for determining vessel category are described. Section 3 describes the experiments and their results. Section 4 concludes by presenting the results in the context of their practical application.

II. METHOD

The SHREC system contains a detection and tracking method that returns a series of images for each vessel passing in front of the camera. To classify a vessel into one of the categories, we have used a pretrained, 22-layer convolutional neural network GoogLeNet that can classify images into 1000 object categories (with none related to vessel classification). During two years of video registration, we obtained thousands of images of different vessels (both conventional and recreational) and initially divided the set into 21 categories of ships. As much as possible, attempts were made to collect photos representing each vessel category from different angles. However, it appeared that the agreed categories were too detailed, as we could not cover each category with enough image representation and some categories were almost impossible to distinguish from each other. Finally, the categories were reduced first to 11 and then to 7 (inland barge, port services ship, kayak, motor yacht, passenger ship, sail yacht, other). The reduction was made in consultation with monitoring centers operators. The transfer training was used to retrain the GoogLeNet CNN for chosen set of the images. The full research on this matter was presented in [15]. The quality of the obtained classification varied between 67% and 99%, depending on the similarities between classes. Nevertheless, in practice we do not need a correct classification for each image of a unit, but an accurate recognition from a series of images representing the detected vessel while it passes in from of camera. Two straightforward algorithms to determine one category for a series of images were designed:

1. Algorithm v1: compute the category for each image, then return the one that occurs most often. If two categories occur the same number of times, return the one that was computed first.
2. Algorithm v2: as above, but rejects small images, i.e., with width or height below 224 pixels. If all images are reject, proceed as in algorithm v1.

Such algorithms should allow to reject erroneous classifications based on vessel images that have a lot of objects in their background or contain incorrectly clipped vessels from a frame. A simple metric was defined: correct vessel classification ratio that is the ratio of the number of vessels correctly categorized over the sum of all vessels categorized. The metric is used to determine the quality of the proposed algorithms.

III. EXPERIMENTAL RESULTS

The classification based on a series of vessel images was tested using the prototype version of the SHREC system. The classification module was implemented using C# and OpenCvSharp3-AnyCPU (OpenCV wrapper for C#). In our experiments, the following computers were used:

1. Detection service and system core module: Intel Core i7-8750H, 16 GB RAM, SSD 256 GB, NVIDIA GeForce GTX 1050Ti, Windows 10 Pro (laptop).
2. Classification service: Intel Core i7-8700K, 32 GB RAM, SSD 1 TB, NVIDIA Quadro P4000, Windows 10 Pro (workstation).

We have used batch processing to play in real time all video files in each data set. Each data set contains 100 video samples. Each video sample shows one passing vessel in front of the camera. Figure 2 shows a series of images of the vessel that were captured from one of the video samples and the partial classification results. The data sets do not include video samples used to train the CNN classification network. The video samples were recorded in high 4K quality and then transcoded to lower quality and full high definition resolution, which resulted in four data sets:

1. Set A: 4K 3840 × 2160, 30 frames/s, bitrate 66 Mb/s, H.264 Advanced Video Coding (AVC) High@L5.1
2. Set B: 4K 3840 × 2160, 30 frames/s, bitrate 10 Mb/s, H.264 AVC High@L5.1
3. Set C: Full High Definition (FHD) 1920x1080, 30 frames/s, bitrate 8 Mb/s, H.264 AVC High@L4.2
4. Set D: FHD 1920x1080, 30 frames/s, bitrate 3 Mb/s, H.264 AVC High@L4.2

TABLE I. TEST RESULTS

Algorithm	Set A	Set B	Set C	Set D
Version 1	81%	80%	85%	81%
Version 2	80%	75%	81%	80%

Table 1 contains classification results for algorithm v1 and algorithm v2 for each data set. Algorithm v1 provided slightly better results. The correct vessel classification ratio for algorithm v1 and set C was 85% and it was the best result. Algorithm v2 for set B returned the worst result of 75%. The correct classification ratio for other tests was 80% or 81%. On average, a series of images for one vessel contains 20 elements.

IV. CONCLUSION

The classification quality depends on the detection quality. The detection method returns a series of images containing vessel's images that are clipped from a video frame based on motion detection. The background of the images is not cleared from other objects. It would be best to use camera views with scenes, where the ships background is only water. However, the system is intended for inland waters where cameras are located mainly on bridges and in narrow passages. Therefore, it is practically impossible to provide a homogeneous background. Additionally, some of the images in the series are clipped (when a vessel enters or leaves a camera view).

 ISBN: 978-1-61208-858-7

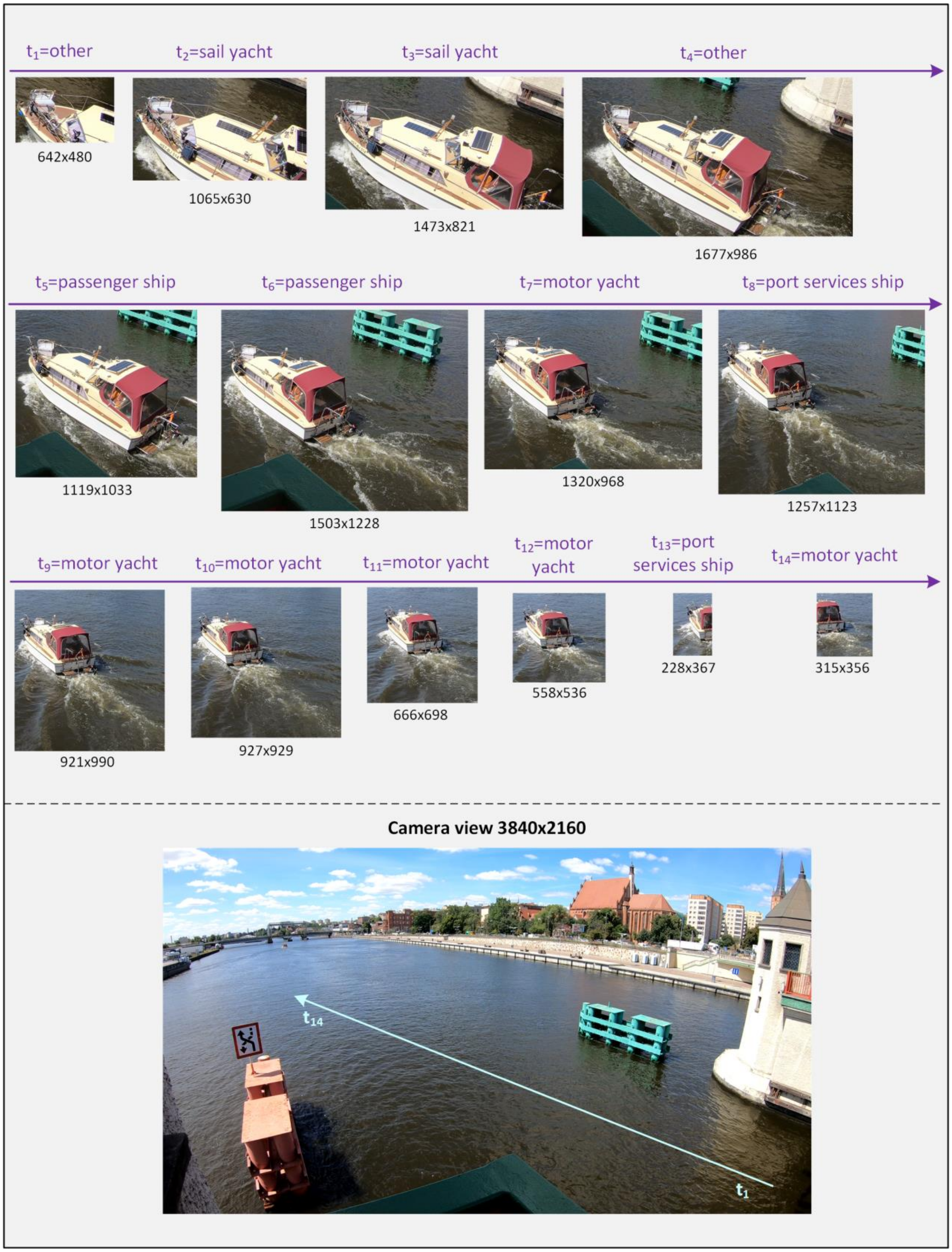

Figure 2. Example series of images with partial classification results.

In the case of heavy vessel traffic, there may be a situation where an additional vessel or vessels appear in the image. However, there will be one or several such images in a series, and in most cases the algorithms will reject the faulty classification results caused by them, as there will be more correct results remaining. In addition, the system requires that the cameras are located in narrow passages or on bridges, which limits the cases of passing ships.

During the preliminary works, three pre-trained neural networks were tested: AlexNet, SqueezNet, and GoogLeNet. Those networks are available in Matlab 2020b with Deep Learning Toolbox. The networks were trained with the same data sets and the best results were returned by GoogLeNet. In the future, our image database (the training dataset) will be periodically updated. The available pre-trained networks will be tested again. Replacing the classification module in our system is straightforward. If some new network returns better results, it can be used instead of the current one.

Discarding small images did not improve the quality of the classification and worsened it by a few percentage points. That was a result different from what was expected. On the other hand, as expected, increasing the compression ratio within a given resolution has caused a loss of qualification quality as compressed images have less details. However, the difference is not significant, therefore more compressed video streams can be used.

One of the most surprising results is the better performance for test set C (FHD resolution) than for set A (4K resolution). This is probably influenced by the detection method. The detection algorithm internally compresses frames to 1280x720 resolution. However, the result for both sets will not be exactly the same, as the compression for FHD is done differently (4k->FHD->720p instead of 4K->720p). We will investigate the exact cause of this in further works. Additionally, we are planning to test algorithms for background removal to increase the classification accuracy.

ACKNOWLEDGEMENT

This scientific research work was supported by National Centre for Research and Development (NCBR) of Poland under grant No. LIDER/17/0098/L-8/16/NCBR/2017) and under grant No. 1/S/IG/16 financed from a subsidy of the Ministry of Science and Higher Education for statutory activities.

REFERENCES

[1] N. Wawrzyniak and A. Stateczny, “Automatic watercraft recognition and identification on water areas covered by video monitoring as extension for sea and river traffic supervision systems,” Polish Marit. Res., vol. 25, pp. 5–13, 2018, doi: 10.2478/pomr-2018-0016.

[2] N. Wawrzyniak, T. Hyla, and A. Popik, “Vessel detection and tracking method based on video surveillance,” Sensors (Switzerland), vol. 19 (23), 5230, pp. 1-14, 2019, doi: 10.3390/s19235230.

[3] G. K. Yüksel et al., "Ship recognition and classification using silhouettes extracted from optical images," 2016 24th Signal Processing and Communication Application Conference (SIU), Zonguldak, Turkey, 2016, pp. 1617-1620, doi: 10.1109/SIU.2016.7496065.

[4] S. Li, Z. Zhou, B. Wang, and F. Wu, “A novel inshore ship detection via ship head classification and body boundary determination,” IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 12, pp. 1920–1924, 2016.

[5] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2015, https://arxiv.org/abs/1409.1556.

[6] C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[7] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE Conferenceon Computer Vision and Pattern Recognition, pp. 770–778, IEEE Computer Society, Las Vegas, NV, USA, June 2016.

[8] T. Akiyama, Y. Kobayashi, J. Kishigami, and K. Muto, “CNN-based boat detection model for alert system using surveillance video vamera,” In 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE), 2018, pp. 669–670, doi: 10.1109/GCCE.2018.8574704.

[9] B. Solmaz, E. Gundogdu, V. Yucesoy, A. Koç, and A. A. Alatan, “Fine-grained recognition of maritime vessels and land vehicles by deep feature embedding,” IET Comput. Vis. vol. 12, pp. 1121–1132, 2018, doi: 10.1049/iet-cvi.2018.5187.

[10] M. Wlodarczyk-Sielicka and D. Polap, “Automatic Classification Using Machine Learning for Non-Conventional Vessels on Inland Waters,” Sensors, vol. 19 (14), 3051, pp. 1-17, 2019, https://doi.org/10.3390/s19143051.

[11] S. Liu and W. Deng, “Very deep convolutional neural network based image classification using small training sample size,” Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia, 3–6 November 2015; pp. 730–734, 2015.

[12] M. Milicevic, I. Obradovic, K. Zubrinic, and T. Sjekavica, “Data Augmentation and Transfer Learning for Limited Dataset Ship Classification,” Wseas Trans. Syst. Control, vol. 13, pp. 460–465, 2018.

[13] Z. Huang, B. Sui, J. Wen, and G. Jiang, “An Intelligent Ship Image/Video Detection and Classification Method with ImprovedRegressive Deep Convolutional Neural Network,” Complexity, art id 1520872, pp. 1-11, 2020, doi.org/10.1155/2020/ 1520872.

[14] Y. Jie, L. Leonidas, F. Mumtaz, and M. Ali, “Ship Detection and Tracking in Inland Waterways Using Improved YOLOv3 and Deep SORT,” Symmetry, vol. 13, no. 2, p. 308, Feb. 2021, doi: 10.3390/sym13020308.

[15] K. Bobkowska and I. Bodus-Olkowska, “Potential and use of the GogleNet Ann for the purposes of Inland Water Ship Classification,” Polish Maritime Research, vol. 4 (108), vol. 27, pp. 170-178, 2020, doi:10.2478/pomr-2020–0077.