SIGNAL2021

Classifying Vessels in Inland Waters Using Live Video Streams

Tomasz Hyla, Natalia Wawrzyniak

PRESENTER: TOMASZ, MARINE TECHNOLOGY LTD., THYLA@MARINETECHNOLOGY.PL

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) in 2017-2021

Presenter

- Tomasz Hyla, PhD
  - Lead Scientist in SHREC project, Marine Technology Ltd., Poland
  - Associate Professor, West Pomeranian University of Technology, Szczecin, Poland
- Research interest:
  - applied computer science, designing new solutions for commercial applications
  - inland navigation, vessel identification
  - cybersecurity, blockchain, digital signatures

- https://www.researchgate.net/profile/Tomasz-Hyla
- https://www.linkedin.com/in/tomaszhyla/

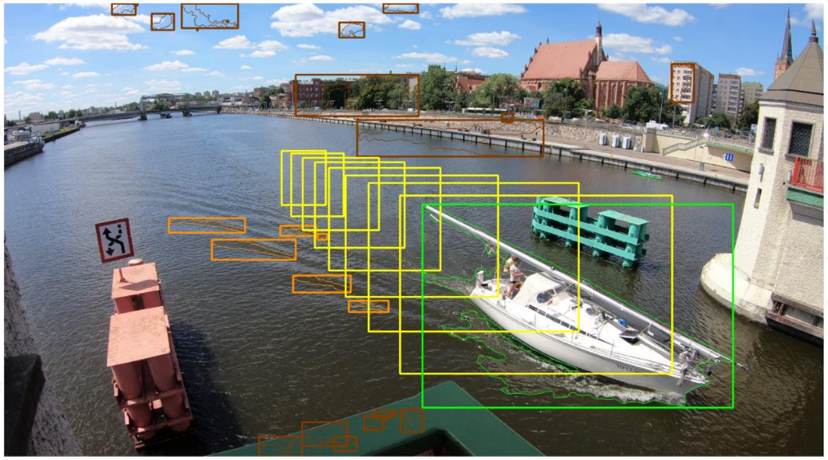
SHREC Project

My paper is a part of the project in which we have built the complete solution for vessel:

- detection
- identification
- classification

Our approach:

- designed for inland waters
- modular approach
- use AI and analytical techniques
- can use existing video monitoring systems

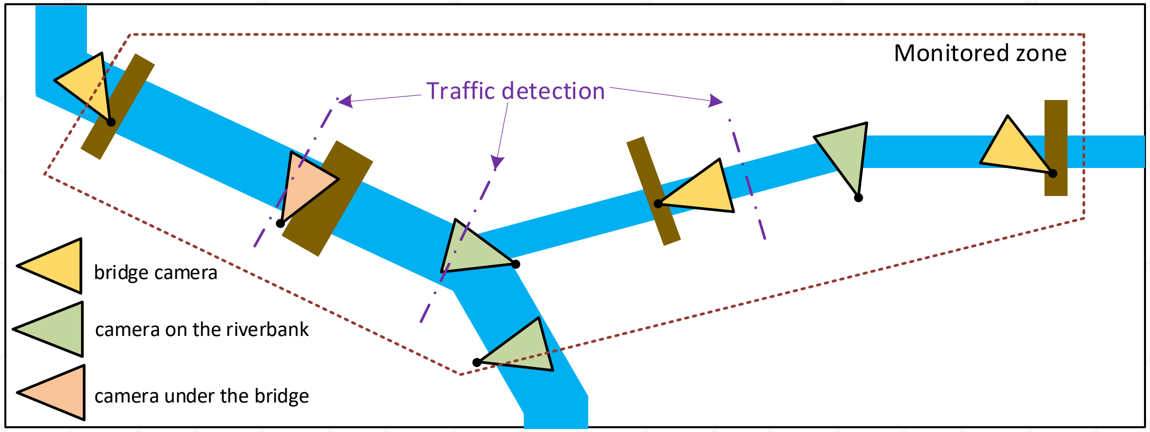


Outline

1. Introduction
2. Classification method
3. Experimental results
4. Conclusion

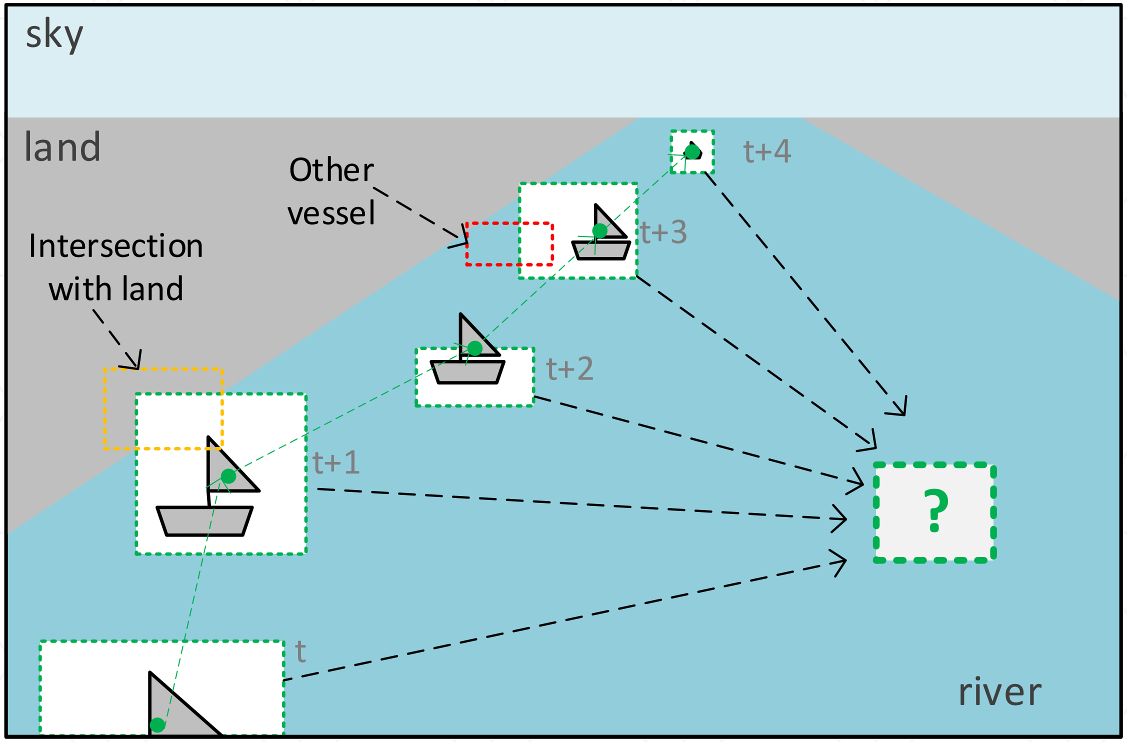
1. Introduction

- The 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 system is the ability to count the number of passing vessels and determine their types.
- This work is a part of the Ship Recognition (SHREC) project. Its objective is to develop the system for vessel detection, identification, and tracking



1. The Problem

The problem of determining the class of a vessel based on a series of images



2. Our method – 1 step

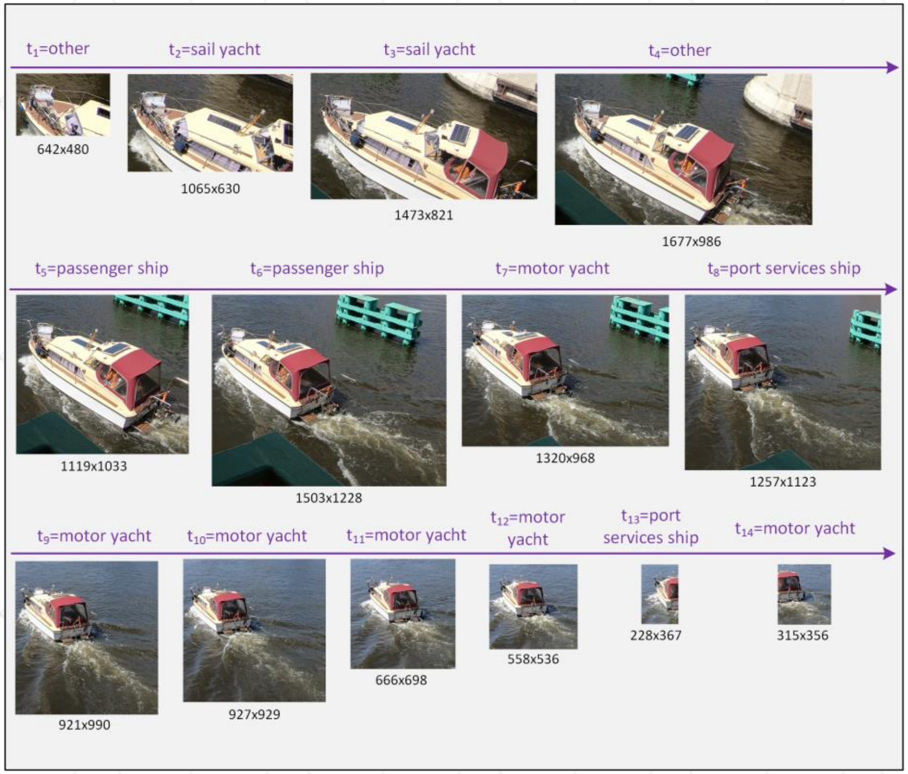
- The input is a series of images for each vessel passing in front of the camera.
- To classify a vessel into one of the categories, we have used pretrained, 22-layer convolutional neural network GoogLeNet that can classify images into 1000 objects categories (with none related to vessel classification).
- During two years of video registration, we obtained thousands of images of different vessels and divided the set into 7 categories of ships:
    1. inland barge
    2. port services ship
    3. kayak
    4. motor yacht
    5. passenger ship
    6. sailing yacht
    7. other
- The transfer training was used to retrain the GoogLeNet CNN.
- The quality of the obtained classification varied between 67 and 99% depending on each class similarities with each other.

2. Our method – 2 step

- 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 detected vessel while it passes in from of camera.
- Two algorithms to determine one category were designed:
  - Algorithm v1: compute category for each image, return the one that occurs most often. If two categories occur the same number of times, return the one that was computed first.
  - Algorithm v2: as above, but rejects small images, i.e., with width or height below 224 pixels. If all images are rejected proceed as in algorithm v1.
- A simple metric was defined:
  - **correct vessel classification ratio** – a ratio of vessels correctly categorized to the sum of vessels categorized correctly and incorrectly.

3. Experiments

Input data visualization



Camera view 3840x2160

3. Experiments

- The classification based on series of vessel's images was tested using the prototype version of the SHREC system.
- The classification module was implemented using C# and OpenCvSharp3-AnyCPU
- The following computers were used:
  - Detection service and system core module: Intel Core i7-8750H, 16 GB RAM, SSD 256 GB, Windows 10 Pro (laptop).
  - **Classification service:** Intel Core i7-8700K, 32 GB RAM, SSD 1 TB, Windows 10 Pro (workstation).
- We have used batch processing to play in real time all video files in each data set.
- The data set contain 100 video samples. Each video sample shows one passing vessel in front of the camera.
- The data set do not include video samples used to train the CNN classification network.
- The video samples were recorded in high 4K quality and then transcoded:
  - **Set A**: 4K 3840 × 2160, 30 frames/s, bitrate 66 Mb/s, H.264 Advanced Video Coding (AVC) High@L5.1
  - **Set B**: 4K 3840 × 2160, 30 frames/s, bitrate 10 Mb/s, H.264 AVC High@L5.1
  - **Set C**: FHD 1920x1080, 30 frames/s, bitrate 8 Mb/s, H.264 AVC High@L4.2
  - **Set D**: FHD 1920x1080, 30 frames/s, bitrate 3 Mb/s, H.264 AVC High@L4.2

3. Experiments

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

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

4. Conclusion

1) The classification quality depends on the detection quality.
2) It would be best to use camera views with scenes, where the ships' background is only water. However, it is practically impossible to provide a homogeneous background.
3) 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.
4) During the preliminary works, three pre-trained neural networks were tested: AlexNet, SqueezNet, and GoogLeNet. The networks were trained with the same data sets and the best results were returned by GoogLeNet.
5) 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.

4. Conclusion

6) Discarding small images did not improve the quality of the classification and worsened it by a few percentage points.
7) 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.
8) 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.
9) In future works, we are planning to test algorithms for background removal to increase classification accuracy.

Thank you for your attention!

Questions?