DETECTING OBSTACLES FROM CAMERA IMAGE AT OPEN SEA

MIKLÓS KNÉBEL

School of Electrical Engineering

Master's thesis
Espoo, 2020

Supervisor

Prof. Quan Zhou
Associate Professor

Advisor

Kalevi Tervo, D.Sc.
Corporate Executive Engineer

Aalto University
School of Electrical Engineering

Aalto University
School of Electrical
Engineering

Aalto University, P.O. Box 11000, 00076 AALTO
www.aalto.fi
Abstract of the Master's thesis

Author Miklós Knébel

Title Detecting obstacles from camera image at open sea

Degree Programme ICT Innovation

Major Autonomous Systems **Code of major** ELEC3055

Professor in charge Prof. Quan Zhou

Advisor Kalevi Tervo D.Sc. (Tech.)

Date 30.7.2020 **Number of pages** 54 **Language** English

Abstract

While self-driving cars are a hot topic in these days, fewer people know that the same level of automation is being developed in the maritime industry. To enhance safety on board and to ensure the optimal utilization of crew members, automated assistant solutions are implemented on cargo ships and vessels.

This thesis deals with a monocular camera-based system, that is capable of detection obstacles in open sea scenarios, and to estimate surrounding vehicles' distance and bearing. After a solid research of existing methods and literature, an algorithm has been developed, containing three main parts. First, the real-world measurement data and camera images are being processed. Secondly, object detection is achieved with the YOLO deep learning methods, that achieves at a high framerate and can be used for real-time applications. Lastly, distance and bearing values of detected obstacles are estimated based on geometrical calculations and mathematical equations, that are validated with ground truth measurement data.

Having multiple weeks of recorded measurement data from a RoPax vessel operating from Helsinki, allowed testing and validation already during the development phase. Results have shown that the systems' detection capability is highly affected by the image resolution, and that distance estimation performance is reliable until 2-3 kilometers, but estimation errors rise at farther distances, due to physical limitations of cameras. In addition, as an interesting evaluation method, a survey has been conducted with industry professionals, to compare human distance estimation capability with the developed system. As a conclusion it can be stated that a significant need and huge potential can be found in automated safety solution in the maritime industry.

Keywords Obstacle detection, Maritime, Computer Vision, Deep Learning

PREFACE

I want to thank Prof. Quan Zhou for his clear guidance in these challenging times, like the COVID-19 pandemic has been in the past month. Our lives have changed drastically, a quick adaptation was necessary to keep up with our daily routines. Special thanks for all the help and expert knowledge I received from Kalevi Tervo from ABB, who has been my advisor from the beginning on and provided me a real-life, industrial challenge. Last but not least, I would like to thank Jukka Peltola and Stefano Marano from ABB, who were always available when practical help was needed.

* * *

Espoo, 30.7.2020

Miklós Knébel

TABLE OF CONTENTS

SYMBOLS AND ABBREVIATIONS

Abbreviations

Notation	Interpretation
AIS	Automatic Identification System
CNN	Convolutional Neural Network
COCO	Common Objects in Context (Dataset)
CPA	Closest Point of Approach
DL	Deep Learning
FD	Fourier Descriptors
FOV	Field of View
HD	High Definition
Lidar	Light Detection and Ranging
MMSI	Maritime Mobile Service Identity
OOW	Officer of the Watch
Radar	Radio Detection and Ranging
RCS	Radar Cross Section
SD	Standard Definition
SMD	Singapore Maritime Dataset
SoTA	State-of-the-Art
TCPA	Time to Closest Point of Approach
YOLO	You Only Look Once

1. INTRODUCTION

1.1. Background of research problem

Robotics, automation, and autonomous systems are currently leading technological fields. Sensor technology, data analytics and computing power has gone through significant improvement in the last years, that enables and increasing level of automation [1].

Although the industry of self-driving cars has the largest attention, since it has impact on most households' future mobility, well-being, and safety, nearly all other fields are part of a rushing development phase in the transportation industry. In case of cargo ships, large vessels and ocean liners the trend is the same, due to newest technological innovations they are going through a dynamic improvement like never before. Automatic applications can help navigation on opens sea, docking in harbors or detecting obstacles on the route that could lead to hazardous events. With a stepwise implementation of these safety and comfort functions on watercrafts, single assistant functionalities are turning into highly automated, unmanned vessels.

This thesis investigates the possibility and feasibility of a monocular camera-based object detection systems, that is capable of recognizing and localizing relevant obstacles at open sea scenarios and estimate their distance and bearing. This kind of computer vision-based solution can significantly contribute to the safety and reliability of automated vessels, help in navigation, and could lead to unmanned bridge conditions in the industry.

1.2. Motivation

The future of maritime industry is going in the direction of a so called B0 conditionally and periodically unmanned ships. Advanced technological solutions have led to a decreasing number of persons on the bridge already in the past years. Although cargo vessels are equipped with a large variety of sensors, the regulatory frameworks are not yet ready to allow the absence of the Officer of the Watch (OOW). Regardless of weather conditions, visibility or easily manageable traffic situations, to ensure safe operations regulations require personal to look out the bridge window at all times [1].

The operating crew on the bridge often spends an entire work shift with looking at radar screens, monitoring the environment through the window without intervention or touching any equipment, even in secure scenarios with clear visibility conditions. Same and similar monotonous and actionless work often leads to frustration, mental fatigue and lack of alertness. All these effects can result in human failure, lower reaction times and incorrect decision-making.

To meet the requirements of current regulations, human eyes might be replaced by intelligent camera systems, that enable the optical scanning of the environment at all times. If it could be proved by supportive measurements, that computer vision-based solutions perform similarly or even better than human lookout persons, the level of automation in the maritime industry could increase further. The implementation of such an optical monitoring system would result in more efficient utilization of vessel crew, increasing mental health and a safer journey.

Furthermore, the motivation behind utilizing a monocular camera system is to be able to reach significant cost reduction. Although most distance estimation algorithms are based on stereo vision, and a single camera solution is full of technical challenges, a functionally working estimation system would bring many benefits and competitive advantages for future applications.

1.3. Objectives and scope

The objective of the thesis is to develop a monocular camera-based distance estimation system as a combination of object detection and distance measurement, based on the research of feasible methods of state-of-the-art solutions. Although computer vision and machine learning methods provide multiple tools to create innovative solutions, each use-case requires a different approach. Combining the right object detection technique with solid mathematical models, would allow to return radial distance and bearing values of vessels from given input images, within a certain accuracy.

Camera based solutions have always some level of uncertainty, since they rely on visual attributes and optical phenomenas that are highly affected by weather conditions and visibility. The scope of the thesis, as an experimentation in the field, therefore focuses on a solution for optimal conditions, exaggerated scenarios such as nights and fog are not guaranteed. Furthermore, as the assistive functionality is planned to facilitate the crew's monotonous work, the relevant setting is at open sea. As a matter of course, the purpose of the thesis is to create a working solution, where the main functionalities are working properly, in order to build a solid ground for future developments.

1.3.1. Goal of the thesis

The overall goal of the thesis is to confirm or refute the fact, whether a monocular camera-based assistance system could significantly improve the safety and establish new features in the maritime industry. To compare the developed system's results to the existing human performance, a numerical comparison should be made.

In addition, based on validated data, performance measures should evaluate the object detection efficiency and distance estimation accuracy of the solution, with the aim to give substantial advice on the feasibility and emerged limitations that may have effect on the installation in commercial use.

1.3.2. Scope of the technical realization

The technical realization consists of two main steps, first the relevant objects must be detected on given input images, then an estimation has to be given on distance and bearing. Although a general solution, that is compatible with multiple vessels, would be an ideal scenario, computer vision-based systems often require prior knowledge of the camera mounting, setting and camera calibration.

The thesis is based on the data of a cruise ship, operating from Helsinki. Since the camera has a fixed mounting position, the system's estimation is optimized on the given vessel's technical parameters, meaning that a stable operation can be only guaranteed on the mentioned vessel or ship, with really similar properties.

Furthermore, the objective of the thesis is to detect obstacles in open sea scenarios, therefore it's not intended to work properly in crowded situations, where a city's landscape or group of islands can be seen in the background, instead of a clear horizon. Finally, vessels and cruise ships have to deal with a variety of weather conditions and have to operate also under limited visibility. As mentioned before, most camera systems are highly affected by visibility, and since the thesis was developed on a predefined setting with hardware limitations, it cannot be assured that foggy, dark periods and similar challenging scenarios are also handled properly.

1.4. Overview

The thesis is divided into three main parts, theoretical introduction to provide proper background knowledge, a detailed explanation of the technical solution and finally, a demonstration of results and drawing conclusion.

1.4.1. Theoretical background

The theoretical introduction is laying a foundation to have proper background knowledge on automation in maritime industry, trends in computer vision and possibilities with state-of-the-art solutions of deep learning. To understand the chosen methodology and the details of the technical realization, it is necessary to introduce the technology briefly.

1.4.2. Methodology of technical solution

The main part of the thesis deals with a thorough explanation of the developed technical solution. Details of data labelling, object detection and distance estimation will be described in this section.

1.4.3. RESULTS AND CONCLUSION

Finally, the systems performance will be evaluated based on multiple metrics, and based on the outcome, a sufficient conclusion will be drawn to answer the questions mentioned as the goals of the thesis. As an outlook, appropriate suggestions will be made based on the results and arisen limitations.

2. THEORETICAL AND CONCEPTUAL BACKGROUND

2.1. Automation of vessels

2.1.1. History and future outlook

The operation of cargo vessels and container ships has always been a mixture of complex tasks. Controlling the engines, docking in harbors, navigation between continents and steering in crowded traffic situations are just a few essential tasks that captains, and the crew must handle on a regular basis.

The rapid growth of automation has reshaped the working principles of these crafts, while the reliability of robotic applications have taken over many tasks in the last decades. The effect can be seen clearly on Figure 1, due to assistive and automatic functions, the minimally required crew size has been diminished significantly.

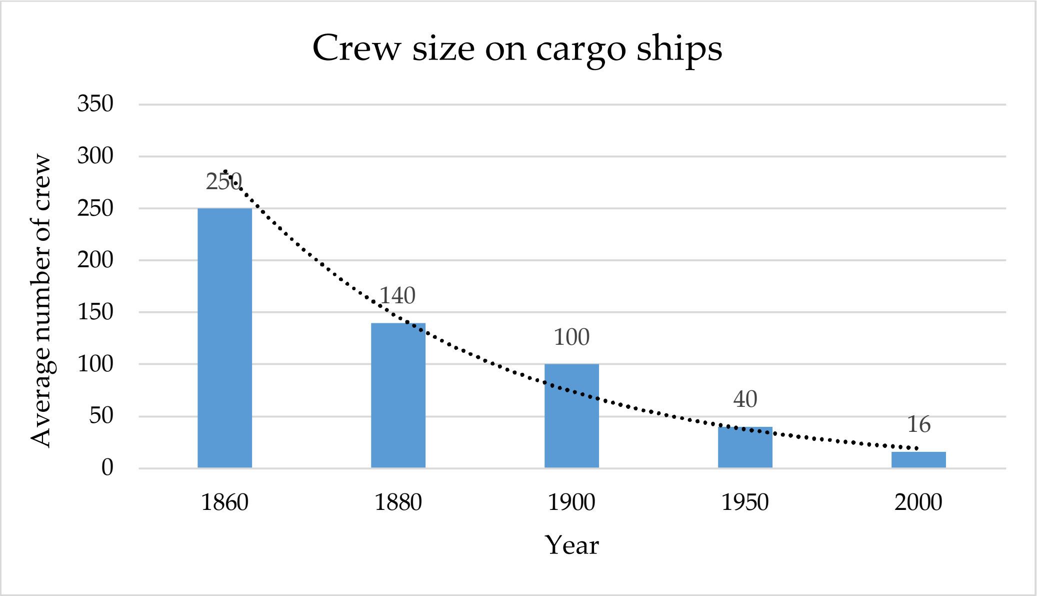


Figure 1 - The changing number of crew size on cargo ships [3]

The rising need for safety, simplification of operations and reduction of cost are all key indicators why developments are accelerated, and the digitalization of shipping is more dynamic as ever before. The application of game changer technologies in marine solutions are fundamental steps to keep a competitive advantage in the industry [3].

Although the tendency shows that an unmanned operation could be reached soon, autonomous solutions still require the presence of humans on board. There is a large step between partly automated, remote controlled or fully automated systems [4].

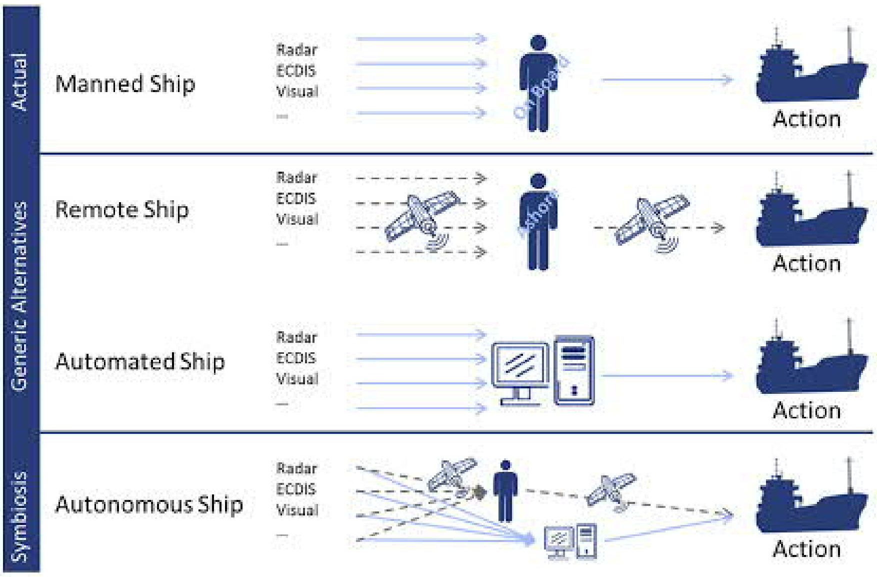


Figure 2 - High level working principle of automated solutions [6]

At the current stage, automated vessels have assistive functionalities, such as obstacle detection and obstacle avoidance, that rely on sensor data. The forming smart harbors and the increasing communication between collaborative vessels establish the way for self-docking solutions soon. In long term, all these developments tend to reach a fully automated vessel, where small ferry boats, but even large cargo ships could navigate from one harbor to another autonomously, without human intervention.

2.1.2. OPERATIONAL BENEFITS

Naturally, there are solid reasons and a variety of benefits that require the advancement and digitalization of vessels.

One of the main arguments is safer operation. In 96% of accidents in marine scenarios, human errors are shown to be the root cause [6]. With reducing human-caused errors, many accidents and collisions could be prevented, and a safer operation could be ensured [7]. The motivation is not to replace and dismiss human labor, but to use the power of engineering to eliminate risks and to assist decision making.

Furthermore, increased efficiency is needed for a sustainable business and to reduce costs. With the help of newest technologies, a more intelligent navigation can be reached that results in a more economical service. The optimal utilization of vessels does not only provide an economic business model for technology provider companies, but allows a more environmental friendly operation of such systems.

2.1.3. Technological challenges

The robustness, reliability and the future acknowledgement of automatic functionalities rely on technology. If the machine will be the number one decision maker in the coming decades, the systems must detect and measure all surrounding vehicles and obstacles to understand traffic scenarios.

Changing weather conditions complicate the perception of the environment, therefore proper sensor technology must be used, that will be introduced in later sections. Although many sensors exist and are implemented on vessels, sensor fusion makes it possible to process reliable measurement values and to create robust systems. Another technological challenge is having adequate computing power. Multiple high-quality sensors produce a large amount of data that has to be stored with a high framerate. In addition, state-of-the-art solutions are often based on machine- and deep-learning methods that solve complex equations and require compelling computing power.

Besides the fact, that latest technology is the fundamental building brick of such solutions, technical challenges are not the withdrawing factors of these systems. Several successful pilot projects globally have proven, that technology could be ready and solid enough to proceed with autonomous functions.

2.1.4. Legal barriers

One of the main restraining forces for the implementation of automatic solutions is the absence of a comprehensive legal framework. Regulations lag behind technology and cannot hold the dynamic speed of technological developments.

There is a major uncertainty in legislation since autonomous ships have never existed before. Similar use-cases, such as self-driving cars, have also a poorly established framework, therefore legislators are limited in relying on pioneering regulations. A strong collaboration is needed between tech companies and law-makers, since a transparent development roadmap and technical capabilities of systems define the base for new legal rights and licenses.

Moreover, the compliance of varying national and international regulations are challenging, due to the fact that countries stand at different levels of legislation. Thus, shipping has always played a key role in global transportation, regulations on autonomous functions must hold on international levels, to avoid intermittent operation opportunities. In practice, all affected countries will not be able to change their laws in the same dynamic, due to disharmonic innovativeness and technical, economical and political interests. Therefore, most innovative countries are the leaders of reformatory technologies, where successful pilot projects can set trustfulness, so countries from the second wave can adapt, based on the reliable experience of pioneers.

Since most of the automatic functions will likely require connectivity with surrounding vessels and cloud-based solutions, cyber security will play a key role in enhancing safe operations. From one hand, they are exposed to a high risk of hacker attacks that might be able to take over control of systems or disturb sensors technology. On the other hand, telecommunication coverage is often poor on open seas where even a blind operation must be guaranteed. Cyber security methodologies today are already at a high technological level, but missing regulations and exact requirements are again delaying factors for applying innovations of the future.

2.2. Commonly used sensors in maritime situations

A key element of automatic solutions is the perception of the environment with the help of sensors. Marine vehicles run under both normal an extreme weather conditions, such as low visibility at night, storms, foggy days or even snowing. Since most situations have devious drawbacks, vessels are equipped with a wide variety of sensors to have a proper overview of the environment, that allows a safe navigation. In this section, the most common maritime equipment is introduced to have a better understanding behind the motivation of applying a camera-based system.

2.2.1. RADAR

Radars, that transmit and receive electromagnetic waves were applied from old times onward. Already in the middle of the 20th century, they were counted as irreplaceable navigational instruments [8]. This fact has not changed since then, radars as still one of the primary sensors that provide safe navigation on seas. In case a larger vessel or cargo ship is in the range of detection, the transmitted electromagnetic waves get reflected, while the receiver unit processes the signals. Based on the returning energy the distance, bearing and velocity of an existing object can be calculated.

The widespread utilization of radars is due to its many advantageous properties. A robust working principle and resistance against weather conditions, darkness and fog allow a long range of detection under various conditions. In the maritime industry, a cleared sea and the low number of objects foster the distance measurement accuracy, compared to the applications on land.

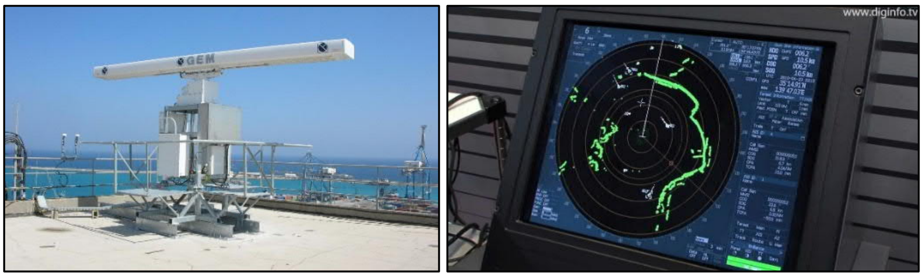


Figure 3 - Marine radar and radar image [10] [11]

Despite the fact, that radars are robust sensors with high measurement accuracy, the measurements lack many informative properties of the objects. Based on the received energy, the radar cross section (RCS) is the only key indicator in defining the size and type of the detected obstacles [11]. Unfortunately, the RCS values are highly influenced by the material and distance of the objects, therefore the radar measurements observe rather the existence and position of sea vehicles. Lastly, smaller vessels often reflect a small amount of electromagnetic waves, thus cannot be detected properly by the radar.

2.2.2. LIDAR

Light Detection and Ranging (LIDAR) is a laser-based distance measurement technology commonly used in autonomous systems. The sensor emits pulsed laser beams in certain directions, usually in 360 degrees, that get reflected by objects inside the detected range. Based on the returning signals and elapsed time, the distance of points can be estimated. The rotating sensor then generates a 3D point cloud of the environment, based on the single measurement points [12].

Compared to radar sensors, one of the main advantages of lidars is that they represent detected objects by many measurement points, that can more precisely determine the exact shape which is the basis of a profound object classification. Furthermore, the working principle allows an independence on the quality of natural lighting conditions, it can be used both in daylight and at night under normal circumstances [13].

On the other hand, the reachable measurement range is limited to a few kilometers, and the resolution deteriorates significantly in the far. In addition, snowing, heavy rainfall, and dense fog, all of them induce measurement errors since the laser beams may get reflected by the particles of the residual.

2.2.3. AIS

The Automatic Identification System (AIS) is the most informative platform in the maritime industry, an indispensable tool for collision avoidance. Although the equipment is not a sensor, the automatic tracking system transmits valuable details of surrounding vessels and sea vehicles. The monitoring system provides vital signals such as a unique identification number, position, speed and course of the vessel, the anticipated destination, and the estimated time of arrival. Its undoubtful, that no other sensor could elaborate such a wide variety of information [14].

Despite the fact, that the AIS system appears like a complex and reliable source in general, it has many downsides as well. The broadcasting frequency can sometimes alter significantly, second based signaling can attenuate to even three minutes. The ego sea vehicle is exposed to the vessels in the intermediate environment, since the transmission of AIS signals is their duty, the ego vessel itself is just a passive listener. In the unlikely event, when the transmission drops out, the AIS is turned off or a small and medium size vessel is not registered in the system at all, it is not possible to receive any information of surrounding vehicles. For the above-mentioned reasons, one cannot rely truly on the AIS data, there is certainly a need for an additional active monitoring solution.

2.2.4. Camera and Thermal camera

Vision based sensors, such as cameras and thermal cameras combine many advantages of previously mentioned sensors, and have the most similar properties in functionality, compared to human eyes. Based on regulations, the human vision and the sight-based environment monitoring is a mandatory prescription on the bridge.

Cameras have a wide viewing angle and can observe the environment on a large range. Theoretically, the limitation of the viewing distance is the horizon, or objects that are above the horizon even further, assuming that a modern industrial camera is used. Computer vision-based solutions are a commonly used and efficient tool in detection and classification of objects, such as vessels, islands, reefs or mainland. Based on lens properties and geometrical rules, even distances, bearings and sizes can be estimated.

Unfortunately, bad weather conditions, fog and darkness, that result low visibility affect the observation capability of cameras pretty much. If the lighting conditions are not suitable, the physical requirements for the proper operation are not met. Thermal cameras can be a great suit for applications facing multiple environmental conditions, however a high resolution and clear thermal contours are necessary. Since thermal cameras do not see colors, they can only differentiate between objects and background elements if a significant difference in temperature is visible. In the maritime use-case, the hull of vessels has often similar thermal reflection as the sea nearby, but the engine compartment and the chimney can be detected due to their higher temperature.

2.2.5. Comparison of Sensors

There is no straightforward answer whether one sensor is better than the other. All of them have many attributes than can be beneficial or detrimental in certain use-cases. As a summary, the following Table compares the most relevant properties of sensor types introduced earlier.

	Radar	Lidar	Camera	Thermal camera
Range of detection	🟢	🔴	🟡	🟡
Dependency on weather	🟢	🔴	🟡	🟡
Availability in darkness	🟢	🟢	🔴	🟢
Object classification	🔴	🟡	🟢	🟢
Data density	🟡	🟡	🟢	🟢
Cost of equipment	🟡	🔴	🟢	🟡

Good: 🟢 Fair: 🟡 Poor: 🔴

Table 1 – Comparison of radar, lidar, camera and thermal camera

The main motivation of the thesis is to evaluate the possibility of applying monocular cameras for distance estimation of objects. As the table also states, camera systems could be financially sustainable solutions for detecting vessels and estimating their distance and bearing.

2.3. *Object detection methodologies and deep learning*

Computer vision as a tool can be applied in many use-cases. Object recognition is used to identify a specific object on an image, while classification aims to select the class or category of an object. The relevant field of computer vision regarding the thesis is object detection, that not only identifies the specific category of the object, but also finds its location on the image. The output is usually a bounding box, that shows the external boundaries of the target. Object detection methods are applied in various fields of technology, such as surveillance and tracking persons on security cameras, face detection or recognizing cars and pedestrians for driver assistant systems [15].

2.3.1. Traditional computer vision methods for object detection

The human brain can easily detect and classify objects, but what attributes characterize a different type of vessels specifically? Is it the shape or color? While the human brain processes the information automatically, the characterization and definition of marine vehicles have to be taught for computers and camera systems.

The usage of Fourier Descriptors is a traditional, template-based computer vision method to detect objects. The approach uses shape information, the contours of the objects are represented by vectors and the outline itself is described as a mathematical function. In order to detect vessels, the mathematical function of a reference object's contour has to be set and compared with other edges on the image. If the function of the target's edge is similar to the template, the object of the category is recognized on the image. Fourier descriptors are efficient methods in some cases, but the dynamically changing environment in maritime scenarios are a way too complex task to solve. On a horizontal image, multiple types of vessels would all require a template shape to be compared, and changing orientation and alignment also lead to multiple outlines of vessels. If satellite images would be used, the FD process may perceive the cigar-shaped boats from a top-view, but that is not the case in this research. Furthermore, vessels in the far are represented by only a low number of pixels, where the mathematical functions of the object's outline would inaccurate, making the comparison results uninterpretable [16].

Another possibility could be the saliency method, that tries to identify image regions which stand out relatively to their neighboring pixels and grab attention on the picture [17]. Based on a research paper submitted by members of the Nanyang Technological University in Singapore, a method called global sparsity potential has been developed to find maritime obstacles on the sea. The basic idea of the feature is to find texture of objects that are not similar to other areas on the sea surface, meaning the presence of distinct objects [18]. Although the research paper states great results, based on an analysis of a large image dataset utilized in the thesis, many cases cannot be detected with the method. A starting point of the algorithm is the segmentation of the sky and the sea surface to select the region of interest. It turned out, that the approach is limited to vessels, which's shape is located below the horizon, since that is the region where the comparison takes place. In many general traffic scenarios, a significant part of the vessels are above the horizon. Moreover, the dataset has shown, that there is a wide variety of light conditions, where humidity or fog blurs the image in a way, that the container ships' or ocean liner's texture and color appears the same as the water surrounding them. Last but not least, the surface of oceans is really noise from a computer vision perspective, since the waves, sunshine and clouds can create changing colors and strong edges, that could lead to false detections.

2.3.2. Object detection using the power of deep learning

Besides traditional computer vision methods, deep learning (DL) gained ground as one of the top techniques in object detection for its performance and adaptability. These State-of-the-Art algorithms are based on convolutional neural networks with many layers, that can learn object classes on thousands of training images recognize new input images with a high accuracy and efficiency.

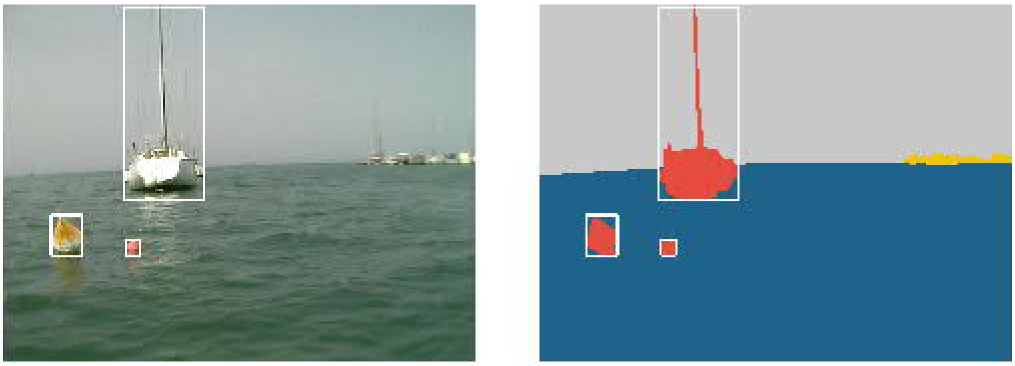


Figure 4 – Object detection and image segmentation in maritime surveillance [20]

Image segmentation is the pixel-wises semantic annotation of the whole picture, where image parts get divided into regions with homogeneous class labels [19]. A deep neural network can be trained to distinguish between the sea, background, sky, land or a specific object seen on the image. During the process, each pixel is assigned to an image part category. Although the detected obstacles' form can be described more precisely than in other methods, this level of accuracy is not intimately relevant for distance estimation, moreover, the increased need for computing power could be detrimental for the intended system.

The topmost technological method required for the thesis is object detections using deep convolutional networks. Given an input image, the network can decide the existence of vessels and also find the location represented by a bounding box. In order to learn such features, a large amount of datapoints need to be labelled. On one hand, some datasets, such as the Singapore Maritime Dataset, exists already that contain more than 200 thousand images that can be used for custom training [20]. On the other hand, some high level, open source detectors are already available online and have the capability of detecting required objects. As an example, an easy-to-use, highly efficient network is YOLO, that can detect vessels and performs on an industrial level. Due to its individual approach, it achieves a high framerate that allows a close to real-time perception.

2.4. Computer vision for distance estimation

The first challenge of the thesis task is to detect the obstacles on the sea. That is an essential part, since the position of the vessels set the base for the distance estimation algorithms. This section introduces conceptual possibilities for measuring distances based on visual information.

2.4.1. Stereo cameras

Stereo cameras are a fundamental solution for depth perception on images. Using two cameras with a special alignment can create a pixel-wise depth map of the visible environment. Stereo vision is similar to the human eyes sight and the brains capability to understand two dimensional images in space. The method is based on geometrical calculations, where the recorded images are shifted with a known offset, and based on these prior adjustments, the two pictures can be confronted, and distance can be calculated [21].

The main downside of the approach is that all the infrastructure, hardware and computation is doubled. First of all, two camera have to be installed, a need for larger computational power is essential for real-time image processing and also, double of the normal storage space is required to process the images taken at the same time [22] [23].

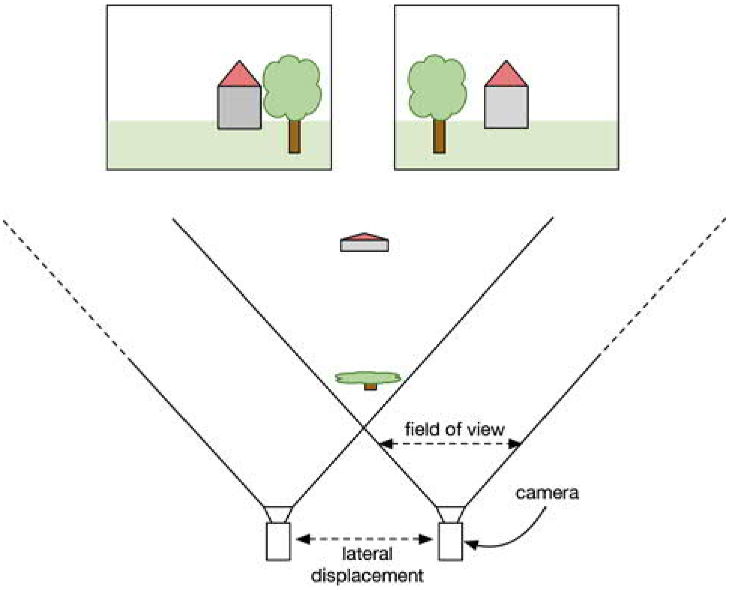


Figure 5 - Stereo camera model illustration [23]

2.4.2. Monocular cameras

Monocular camera-based systems for distance estimation are still under research, there is no established method yet. As mentioned earlier, a significant cost reduction can be achieved with solutions using single cameras, they have a huge potential, but of course, the developments are more challenging compared to other technologies.

First of all, one of the basic but theoretically correct and implementable solution is the usage of the pinhole camera model. Based on prior knowledge of the detected surface, mounting position and lens properties, geometric equations and mathematical models can be set up and distance calculations can be made. In general, if the size of an object is known, a conversion between pixels and distances can be made and a proportional technique can be applied. After a successful calibration, the distance can be easily estimated by the number of pixels that the edge of an object contains. However, in maritime scenarios, the orientation and type of vessel are continuously changing, taking any prior assumption on the size of vessels could be misleading. Open sea scenarios can be approximated as a flat surface, where with the help of known Earth curvature properties and camera mounting height, distances can be estimated based on the relative alignment on the surface.

In some cases, the relative position of the marine objects is inspected between the ego vessel and the horizon, which is detected in the early phase of algorithms [24]. Since the horizon is a straight line in optimal case, with the help of open source computer vision libraries, it can be detected with Hough Transform or Canny Edge Detection. From one hand, the horizon can be used as a reference point with known distance, but it can also be used to define region of interests and a separator between sky and sea.

Even if the utilization of the horizon seems logical, many doubts appear in real world applications. First of all, if an island, mainland or a mountain appears in the background, the horizon detection may fail. The clear contour that can be seen on open sea will disappear. Secondly, foggy, and cloudy weather often obscure the horizon, it changes to a blurred line or cannot be seen at all. Using the horizon as part of an algorithm might create to many dependencies that lead to failed operation in complicated weather conditions. Finally, the roll-pitch-yaw angles have to be considered in case a camera is mounted on a buoy, but large ocean liners that are used in the thesis are significantly less affected.

Additionally, a more advanced solutions called Optical Flow also exist, that uses a single monocular camera for distance estimation. The idea behind the method is that it analysis the pixelwise change and the displacement of certain pixel regions between two consecutive images [25]. In robotic applications or ground vehicles, they can be used on small distances with reasonable alteration of images. Unfortunately, at open sea the scenario is quite static, there is lack of changing reference points since the only objects are the targets themselves, that are moving with unknown velocity and orientation as well.

2.4.3. Distance estimation with machine learning

Finally, distance estimation on images could be also achieved with monocular cameras and machine learning techniques. The Institute of Automation in Bremen has developed DisNet, a Multi Hidden-Layer Neural Network for railway operations, applied on smaller distances compared to the maritime use-case [26]. The inputs of the supervised learning method are bounding box coordinates of various objects together, with ground truth measurement values coming from a Lidar sensor.

More than 2000 datapoints have been used to train the model, which is actually an image-based regression problem of coordinates and distances. Unfortunately, the maritime scenario requires not just objects with distance in a few hundred meters, but datapoints from multiple kilometers and with widely spread vessels on the side of images as well. Based on rules of thumb, a training of such a model would require nearly 10.000 training images for the maritime application, which is not possible given the available ground truth measurements and the scope of the thesis. Moreover, the trained model highly depends on the mounting height and mounting angle, it's hard to parametrize as a general solution, therefore in case a similar solution is implemented on a new vessel, another dataset has to be recorded and the model must be retrained.

As a summary, traditional geometrical solutions with single cameras tend to be the best practices for distance estimation in the current use-case. Besides their easy implementation and acceptable computing power needs, they can be parameterized well with just a few variables, that allow a scalable and sustainable solution for the future.

3. TECHNICAL DESIGN AND REALIZATION

3.1. Research plan

Many solutions exist in the field of object detection and measuring values on images. Even though, the challenges of the current marine use-case require a grounded choice of technologies to achieve prominent results. This section explains in detail, how the advantageous properties of existing methods have been combined to build a complex, multifunctional system.

3.1.1. INTRODUCTION OF THE METHODOLOGIES

The development is built on three main components. First of all, data collection was needed to pair up real-world camera images with ground truth measurement data. Secondly, the object detection method had to be developed, to recognize and localize vessels on images. Finally, based on location inputs of the object detection, a distance estimation approach had to be established.

The data serving as a base for the thesis was a large set of measurements, recorded on a cruise ferry that operates from Helsinki. Camera images, GPS positions, internal sensors and many more were recorded in various traffic and weather conditions. RGB cameras are the first input for the system, were object detection is applied. Based on the analysis of the literature review and existing solutions, a deep learning-based method has been chosen. In case a cargo ship, ocean liner or other type of vessel gets selected by the detector, a representative point must be chosen for further calculations on distance end bearing. During the development phase, GPS coordinates served as a comparison, to validate the estimation capability of sensors. While current ships are equipped with a large number of sensors, the final system developed as part of the thesis are not based on sensor fusion. In contrary, a blind method is used, where the input information relies only on one single RGB color image.

3.1.2. THEORETICAL CAPABILITIES AND LIMITATIONS

Besides having the power of newest technology, some theoretical capabilities and limitations, the system is affected by, must be clarified in advance. Two main facts have the most impact on the systems performance, one of the is the curvature of the Earth, the other is the image resolution and quality the input images are recorded with.

As a fact, the Earth is spherical, which means that a human or a camera sees the horizon as the end of the ocean surface. In theory, but also in practice, object farther than the horizon can be also seen, however, the current use-case limits the system for objects before the horizon. It is necessary to see the bottom of objects since their location is a crucial step in distance estimation. Unfortunately, only the upper part of objects can be seen behind the horizon, which means that their real height is unknown and could be arbitrary high in theory. In addition, the observation height plays a key role in determining how far the horizon can be seen.

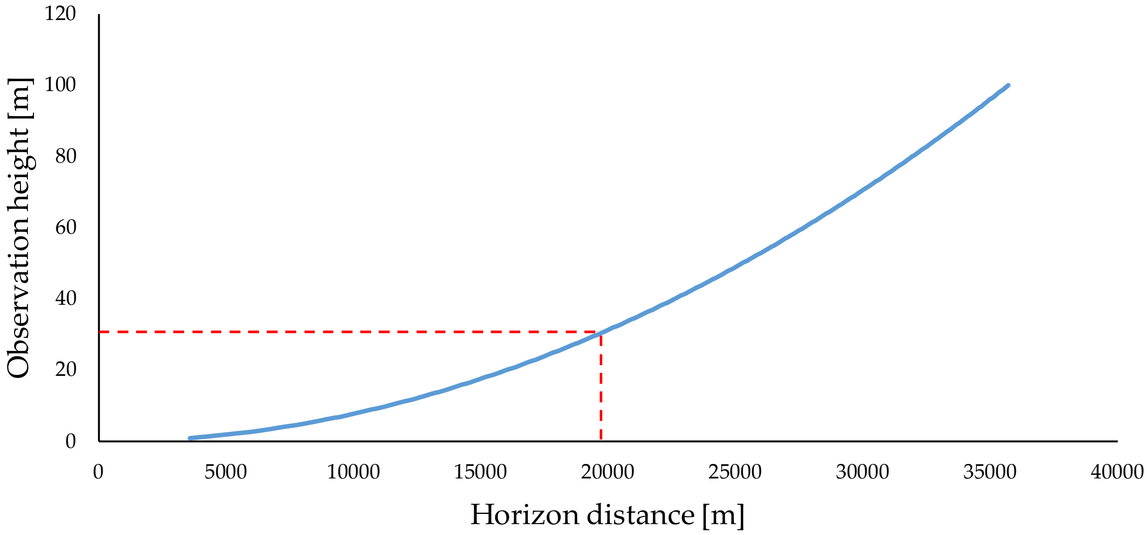


Figure 6 – Distance of the horizon based on observation height

As *Figure 6*. also states, the higher the observer is placed, or a camera is mounted, the farther the horizon is seen. The video cameras used in the thesis are set around a height of 30 meters, meaning that the real distance of the horizon is around 20 kilometers, assuming the globe as a perfect sphere with radius of 6371 kilometers. Therefore, the upper theoretical limit of detecting an obstacle on the sea, using the currently equipped vessels, is around 20 kilometers.

Another key factor is the resolution of the images, that highly affects both object detection, and the distance observation as well. An important question, but difficult to be answered, is the minimum observable size of vessels in term of pixels. At a shorter distance, vessels appear larger and are represented by many pixels. From one hand, they can be detected easier due to their size, but the real advantage is, that a more detailed shape and color pattern can be identified, due to the large number of representing pixels. In contrary, at a farther distance, the shape of marine obstacles only rely on a few pixels, meaning that only a low-resolution, discrete contour is visible. The vessels' most straightforward characteristic, for deep learning-based object detection algorithms is the connection of edges and the outline. Unfortunately, due to the above-mentioned reasons and basic physical parameters, it is expected to see deteriorating performances at higher distances.

Moreover, the distance resolution, that defines the estimation capability of systems, is dependent on the number of vertical pixel points. Images describe the visual world in a discrete way by pixels. A pixel does not only mean a single point in the real world, but cover a larger sea surface area, in the current case. Since the camera is mounted in an angle of nearly 15 degrees compared to horizonal, as a result, a changing resolution will be seen in relation to the distances.

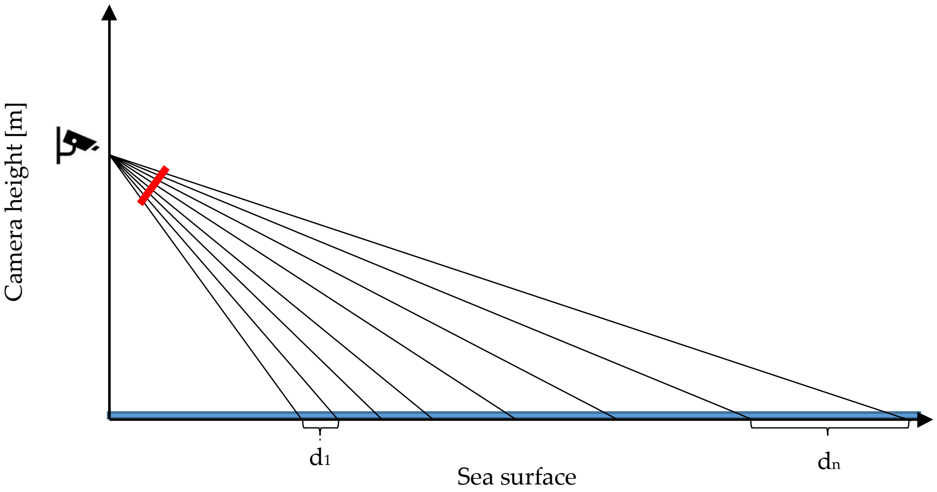


Figure 7 – Changing distance resolution based on pixel steps in the image plane (red)

The exemplary drawing on *Figure 7.* demonstrates that there is a significant difference in observed areas by pixels at the bottom and the top of images. At the beginning, at shorter distances, a single step in pixels means 10-15 centimeters change in reality. The d_1 and d_n values indicate, as further the pixels are determined, the larger the covered distances get. Based on mathematical calculations, the distance steps reach even a few kilometers difference near the horizon, meaning that a detection error of only one pixel will immediately lead to an estimation error, of kilometers order of magnitude. Such physical limitations must be considered when designing estimation systems and evaluating their performance.

Last but not least, illumination properties, visibility conditions and changing environment will play a key role in the system's performance. The water surface is a highly textured plane that can change rapidly among different faces, such as an orange color at sunset, blue when clear sky or grey in stormy weather. The detection of small objects therefore results in a challenging task, since the fore and background do not have a homogeneous character, larger waves and clouds can eventuate strong environmental contours, while the critical obstacle, such as a grey ship might be blurred by fog. Given the above, camera-based observation systems have not just physical limitations, but are affected by a variate of complication circumstances and challenges.

3.2. Data collection and labelling

The method developed in the thesis is a concept solution for industrial application purposes. A standardized camera with fixed mounting height and positions had to be set to conduct further research on existing methodologies. The primary source information for the computer vision-based system is the camera images itself. An industrial camera equipped on a RoPax vessel cruising from Helsinki served as the base for the developments. Although single camera images would be sufficient for object detection methods, distance estimation requires ground truth measurement data from additional sensors, to have a profound basis for comparison and validation purposes.

3.2.1. Description of data

The proprietary communication protocol used in the thesis is developed specifically for real-time system's applications. The protocol is programming language independent and is suitable for applications where low latency and high bandwidth is crucial. Messages contain both internal and vessel specific sensors measurements, which can be processed online or offline by using the protocol-specific API.

One of the main features of the protocol in the maritime use-cases is the Automatic Identification System (AIS), where vital details are sent across vessels inside a given region. AIS encompasses a unique identificatory called Maritime Mobile Service Identity (MMSI), GPS position coordinates, vessel bearing and speed and many more describing the current status of surrounding vessels. Once a ship is equipped with AIS, an ego vehicle can track and localize traffic companions within a certain radius. From the perspective of the research work, the GPS coordinates are the most crucial information to generate ground truth training points. Once the camera images, that are received in the communication files, contain vessels, the timestamp can be compared with recorded AIS information in the same moment, meaning that based on GPS location, the vessels can be recognized and identified instantly. Basic mathematical calculations can later on determine distances and bearings with a high accuracy.

3.2.1.1. Internal sources

In this thesis the recorded internal measurements of a maritime technology company were used, that contain cruises operating from Helsinki in various weather and lightning conditions. The recordings are coded with a communication protocol and had to be processed beforehand, to reach camera images and AIS data in appropriate format. In total, 200 files were recorded, each of them containing a timeframe of 20 minutes, resulting an overall set of more than 65 hours of operation time. Naturally, not all of them contain applicable scenarios or visible vessels, but a large dataset was available for development. Detailed data processing methods will be described in later sections.

3.2.1.2. External sources

While the internal dataset contained both images and distance measurements, some external data sources can be found as well. However, it is important to highlight, that accurate distance information was only available for the used dataset, but object detection could be assisted with external data.

One of the most expanded data is collected in the Singapore Maritime Dataset (SMD). It contains both onshore and onboard camera images and videos of crowded traffic situations around the Singapore harbor. Many researches focusing on object detection and classification of vessels are based on SMD. In this thesis, the object detection method is a more sophisticated and commonly applied method, where the usage of external sources would not result in significant performance improvements or would be out of the scope. Unfortunately, SMD does not contain distance information, therefore the performance could not have been validated. Besides SMD, other maritime datasets exist as well, but none of them provide additional measurements. Usually internal, industry-heavy research has been conducted in the field, were open-sourced data would endanger competitive advantage.

3.2.2. Labelling method

The labelling method, to combine camera images with GPS coordinates of visible vessel were divided into two main subtasks. First of all, the raw measurements had to be processed, meaning that the camera images and the AIS data had to be subtracted from a large set of information, that was encrypted in the binary files. After gaining the right format of images and the GPS information surrounding obstacles, a timestamp-wise comparison was made. The next sections explain the two subtasks in detail.

3.2.2.1. Processing raw data files

As mentioned before, the measurement files contain all the necessary information that serves as a base for the thesis. The raw data file stores the AIS information, that was transmitted by marine vehicles in a range of 40-50 kilometers around the ego vehicle. The vessel specific information arrives in the AIS data structure, containing the MMSI identifier, latitude and longitude information paired with a global timestamp. Although many more details are transmitted and recorded, in this case GPS coordinates are essential, but also sufficient values for further calculation.

AIS information is received as individual packages, each vessel in transmitting its information every few seconds. In many cases, multiple packages arrive at the same timestamp, but still individually. The regularity of messages varies and is depending on the velocity as well, since vessels moving at a higher speed transmit their values every few seconds, but docked ships send their updated AIS only after a couple of minutes. Besides arising challenges with the transmission frequency, it is also a validation, that camera-based systems have a market need and potential, since objects could be tracked at a much higher framerate compared to the AIS.

Based on the incoming data from the files, the information gets processed immediately, distance and bearing are calculated while running. In a real-world application, this would mean real-time calculation time. The distance of two distinct GPS locations have been calculated with the help of the python geopy library, using a formula invented by Thaddeus Vincenty. Built upon two iterative steps, geodesic distances can be calculated using latitude and longitude coordinates as inputs. The accuracy of the approach outperforms other methods, where instead of a perfect spherical, a precise ellipsoidal modeling of the Earth is undertaken [27]. For bearing calculation, some initial statements have to be clarified. A few general approaches exist, that are used in maritime situations and geodesy, where given two different points with coordinates, the absolute bearing can be defined [28]. However, it is important to mention that the camera system will be only able to estimate a relative bearing value, since the heading of the vessel will not be used as an input parameter, to ensure a standalone functionality. After the calculation of the absolute bearing, the heading needs to be extracted to provide comparable results for validation purposes later. Given two points with coordinates in radian ***A*** *(lat₁, long₁)* and ***B*** *(lat₂, long₂)*, the following equations can be applied:

$$diff_{Long} = long_2 - long_1 \quad (1)$$

$$x = \sin\,(diff_{Long}) \cdot \cos\,(diff_{Long}) \quad (2)$$

$$y = \cos(lat_1) \cdot \sin(lat_2) - \sin\,(lat_1) \cdot \cos\,(lat_2) \cdot \cos\,(diff_{Long}) \quad (3)$$

$$Bearing_{initial} = atan\left(\frac{x}{y}\right) \quad (4)$$

Note, that the *atan()* function returns values between the range of -180° and 180°, but in the current use-case, the compass bearing must be calculated. Hence, the initial bearing value needs to be normalized, to receive results between 0° and 360°. After the conversion of the initial bearing from radians to degrees, the final step is:

$$Bearing_{compass} = (Bearing_{initial} + 360)\ \%\ 360 \quad (5)$$

When the calculations are done for a single message, the information package, containing MMSI identifier, latitude, longitude, distance and bearing values, get stored in a pandas data frame [29]. In addition, a simple filter is added to each line, that checks whether a vessel is possibly visible or not. If a vessel is less than 20 kilometers far, and the coordinate is within a certain viewing angle compared to the ego ship, the visible flag is set to true, otherwise to false.

	MMSI	Latitude	Longitude	Radial distance	Bearing	Visible flag
0	230986940	60,1542	24,8895	0,7899	-11,3493	True
1	230046990	59,7365	24,6029	4,2768	4,8112	True
2	636014356	60,2057	25,6238	1,1189	36,4791	True
3	230125940	60,0902	25,9856	2,8719	-48,0089	False
4	230184000	59,9436	24,9259	2,6544	-22,5734	True
5	255805884	59,9074	25,558	7,1415	55,9423	False
	…	…	…	…	…	…

Table 2 – Stored AIS values in Pandas Dataframe

Two important design considerations have to be explained. First of all, a real-time refreshing table could be made, that always shows the last known position of each vessel. This method could be used on vessels in real application, where the exact position of surrounding vessels is required always at the current time. In this case, the data is only needed to be comparable with the image stream to identify detected objects. Unfortunately, the image stream and AIS messages are not always synchronized, that could lead to miscalculations and the processing and comparison of both information at the same time were not efficient. More details are explained in the next section. The other design concept that has to be mentioned is the event-based refreshing of the database. Logically, one would only consider the incoming AIS messages, where distances and bearing get calculated and stored after no further update is received. Nevertheless, since the camera is mounted on a moving vessel, and sometimes the transmission of signal is delayed by 10-15 seconds, the moving state of the ego vessel must be acknowledged in addition. For this reason, the database must be refreshed when an event of incoming AIS message is happening, or the ego vehicle transmits a new position. In practice, at a given timestamp all newly received positions have to be processed and all known positions from the last timestamp have to be updated. Since the method is calculation heavy, the data is not utilized in real time, but for each measurement file, a large data frame gets stored to a CSV file, containing all incoming information, ego position updates and necessary calculations for each timestamp within a recorded time window. The stored CSV file is later on serving as a base for the next step, where GPS coordinates, distances and bearing get compared with the camera images based on the recorded timestamp.

3.2.2.2. Comparison of camera images and relevant obstacles

Camera images arrive as an image stream with a high framerate. However, with lack of vessels, many images are not relevant and even when some are visible, it is not necessary to process images more frequently than a couple of seconds. Each image is paired with a timestamp, that serves as a base for comparison. Using the generated CSV file described in the earlier section, theoretically visible objects will be selected for each image and the visible object in practice can be assigned.

This method allows to determine ground truth distance and bearing values to vessels appearing in the images. Unfortunately, it is a manual task. Although the detection of objects is possible by a computer, the pairing with measurement values require accurate decision making. First of all, a false comparison can undermine the performance of the system, as the whole distance estimation method is based on the generated validation data. Therefore, a manual supervision is needed to avoid possible errors. Secondly, measurement based, theoretically visible objects cannot always be localized on the images, due to visibility conditions, covering of geographical objects or overlapping obstacles. Moreover, the MMSI identifier has to be checked to assure a correct detection and pairing. Last but not least, the timestamp of the image stream and the refreshing of AIS information is not necessarily fully aligned. The two systems work independently, therefore the most accurate pairing can be done only, when the closest timestamp is chosen. As mentioned in the earlier section, the comparison is not made while running and processing the files, because the shifted timestamps could results situations, when the closest AIS update is after the image stream, which leads to more accurate values, which could not have been produced if only earlier updates would have been taken into consideration.

Moreover, an important design choice is the analysis of all surrounding vehicles even in the past. The flag, that states the potential visibility of objects, is helpful when choosing the seen object from the measurements. However, in many cases a standing or moving object outside of the viewing area can fall into the visible region of the vessels change position and rotate accordingly. Therefore, the incoming AIS message stream always needs to be processed and values from the past have to be updated, without filtering out implausible obstacles. This case is mostly relevant for vessels transmitting their information only every few minutes, since the ego vessel's trajectory can change drastically within that time.

This section has introduced a parallel method that helps to generate validation data to determine the performance of distance estimation methods later on. Although, an adaptive, real-time solution would be preferred on the bridge, the current use-case differs, and different requirement are needed. In this research, the only input for the real-life application is a single camera image, where object detection and estimation has to be applied. Therefore, the separation of processes was only needed to create training data and will not affect or delay calculation times for the final system.

3.3. Object detection

After the receiving and processing of data has been discussed in the previous section, the second phase of the algorithm deals with object detection on camera images, which will build the ground for the distance estimation methods later. Object detection is a complex task, where relevant marine obstacles have to be identified as a classification problem, and as an additional step, an image-based, pixel-wise localization has to be made. Based on a detailed experimentation of State-of-the-Art methodologies, a deep learning-based approach has been chosen as the primary object detection method.

3.3.1. Using YOLO object detection library for experimentation

Many convolutional neural network-based architectures deal with object detection, such as RetinaNet-50, RetinaNet-101, R-CNN, Fast R-CNN or YOLO. Region proposal based convolutional networks such as R-CNN are not designed for real time implementation, since they can only perform at a low framerate. Fast and Faster R-CNN are already sped up solutions that are close to real-time, but region proposal methods are still bottlenecks of the system.

You Only Look Once (YOLO) object detection library, is a SoTA solution that outperforms many other architectures due to an outside of the box principle, interpreting the challenge as a regression problem. Instead of defining regions, YOLO uses a single convolutional network that predicts bounding boxes and class probabilities for each of them, only after one single look at an image [30].

An image is split into a certain number of grids, typically with a 19x19 pixel size, each of them responsible for predicting a number of bounding boxes. In the next step, a class probability and offset value is assigned to each bounding box, meaning that the probability is determined, whether a cell contains a certain object. Furthermore, the class with the maximal probability is chosen as the type of the object, while a particular grid cell with the highest probability also localizes the object within an image. YOLO is a magnitude faster compared to other object detection algorithms and can reach even up to 45 frames per second.

Although 45 frames are not required in the maritime use-case, a system capable of real-time detection is necessary for safety critical industrial applications. Nevertheless, the YOLO network struggles with identifying small objects, which would be an important feature for detecting obstacles at higher distances. In the current solution, the detection speed and detection capability hat to be compromised, where the framerate benefits provided by YOLO were significantly higher, as the detection capability of other networks [31].

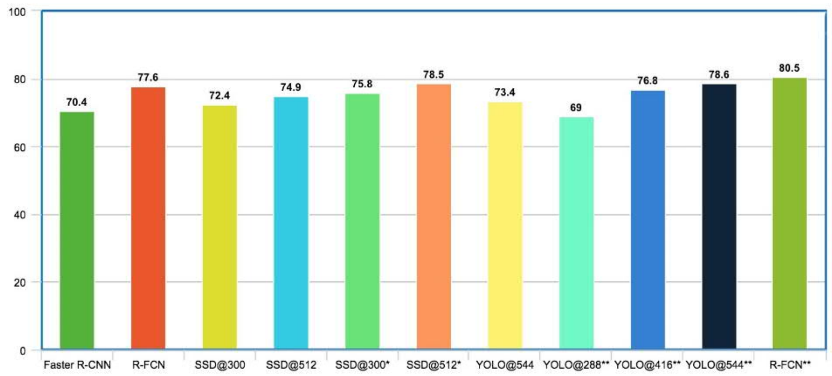


Figure 8 – Accuracy comparison of deep learning architectures [31]

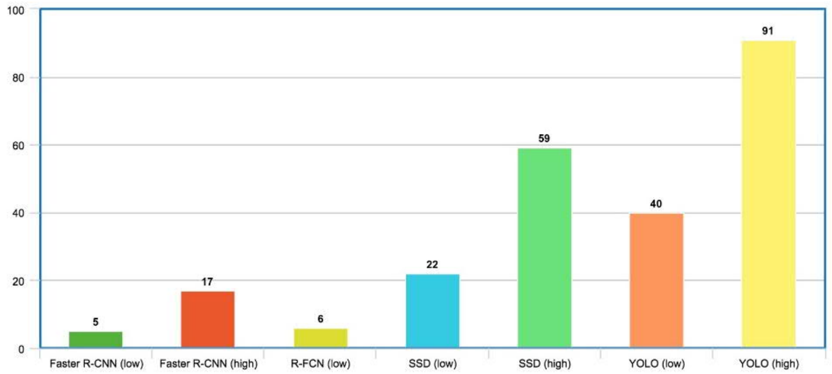


Figure 9 – Detected frames per seconds of deep learning architectures [31]

The Darknet YOLO v3 approach, which was chosen as the primary object detection method in the thesis, was trained on the COCO dataset on hundred thousands of images. In total, around 80 object categories are pretrained on the networks with a large set of data, which also allowed the detection of ships and vessels. Interestingly, the original YOLO v3. Implementation based on the work of J. Redmon and A. Farhadi [32] had slightly lower performance as the high-level, easy to use, open source computer vision library called cvlib, which is based on exactly the same networks [33]. After both methods were implemented, the cvlib has been chosen for its user friendliness. Based on further analysis, the color coding and processing method of saturation values seem to lead to dissimilar results in detection performance.

YOLO has shown compelling results during the development phase. While larger vessels in the front of the image are easily detectable, smaller ships at farther distances, near the horizon were often not detected. As mentioned in earlier section, on low quality images, the shape of the vessels in the far are not well-sophisticated, since only a low number of pixels define the outline of the hull. A more detailed performance evaluation can be found later in the Results and Discussion section.

3.4. Distance estimation with traditional computer vision methods

In the *Theoretical overview* section multiple methodologies for distance estimation have been introduced, strengths, and weaknesses of have been analyzed and a combination of methods has been selected for further steps. A basic requirement for the developed concepts solution was the usage of a monocular camera from the beginning on. In this section, two methods are introduced, a tradition bottom-up geometric solution based on the pinhole camera model and lens properties, and a sampling-based top-down method, that builds a distance mapping on processed measurement values as a regression problem.

3.4.1. PRIOR ASSUMPTIONS AND KNOWN UNCERTAINTIES

As it was partly discussed in the Theoretical capabilities and limitations section, some prior assumptions and uncertainties have to be given attention to and be explained. Due to the fact, that the thesis discusses a concept solution using a monocular camera, limitations can be applied, and the performance can be analyzed, but the effect of prior assumptions have to be dealt with.

First of all, in the oncoming calculation models, the ocean in the observable range is considered as a flat surface. Based on calculations with a horizon distance at 20 kilometers, the maximal offset, between a perfect spherical shape and the flat representation, is below 10 meters which can be negligible at the current stage. From one hand, it does not have a significant effect on distance estimation, and sometimes, at the open sea also waves can reach to that height in normal circumstances. Secondly, the perspective representation of the image plane causes a distortion in distance steps of pixels. As mentioned earlier, at lower pixels, that show closer parts of the ocean in reality the pixelwise step in distance is low and precise. However, as farther points are covered, a pixel step can lead to kilometers change in the distance plane. As *Figure 10* states, the larger distances are observed, the less pixels are representing a certain distance region meaning a lower resolution quality.

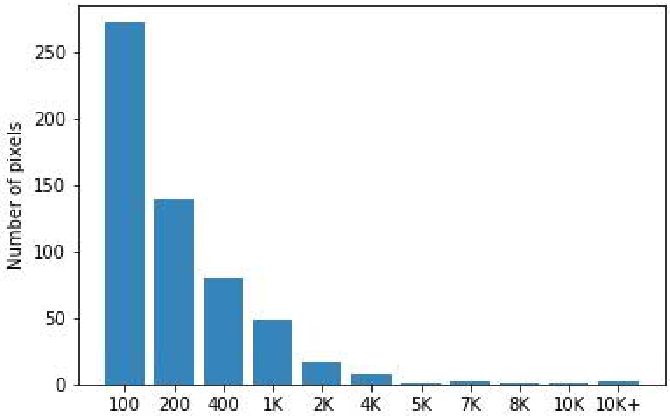


Figure 10 - Number of pixels for distance categories

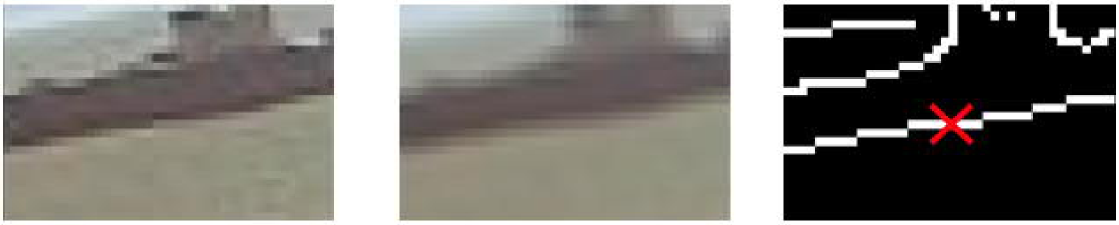
Furthermore, there are three main factors that create uncertainty in the estimation methodology. In context to the previously mentioned change in distance resolution, the pixel selection of vessels must be as precise as possible. For estimation of parameters, a representative point has to be selected after detection of marine obstacles. The representative point is on the bottom part of the vessel, that has the highest correlation with the measured distances in real life scenarios. Due to some uncertainty in the YOLO based object detection method, bounding boxes might be shifted by a few pixels around or inside a given object. While it does not cause problems at shorter distances, issues arise when estimating vessels in the far. If even one pixel difference takes place, the observed value will lead to an error of kilometers magnitude or more. Therefore, some correction steps need to be undertaken, to prevent large pixel errors and minimize estimation digressions.

The next uncertainty is caused by the asynchronous sampling of AIS messages and camera images. Camera images used as inputs of the system can be accessed with a frequency depending on the framerate. In contrary, the AIS and ago motion information messages, that are the main values for ground truth distance calculation, arrive in unpredictable time intervals and shifted compared to the images. For this reason, the surrounding vehicles seen on the image cannot be compared and combined with AIS data from measured at the same timestamp. Based on an image, the algorithm chooses the values from the closest possible timestamp, whether it was recorded before or after the image was taken, since the closest timestamps will achieve the smallest prediction error. In conclusion, although the best possible AIS and image combination is selected, there is still an error by default that can lead to incorrectness in the distance and bearing calculation.

Lastly, the vessel size itself generates uncertainty to the system. The detected objects are a few hundred meters long, and it is not explicitly defined, which point of it is or has to be taken into consideration. In marine traffic, the Closest Point of Approach (CPA) is usually selected, since that defines the most critical position and the shortest acting time, the Time to Closest Point of Approach (TCPA). Although it is the most logical reference point, it is not necessarily correct when considering measurement files as ground truth data for validation. In most cases, the AIS transmission point of a vessel varies between the front and the back of the boat. If always the closest points would be considered, the system would be affected by a large, unpredictable distance error from the beginning on. Therefore, to average the error and to create a generic solution, always a point from the middle of the ships have to be selected. Hence, the AIS transmission points error, that cannot be dealt with beforehand, is minimized by this design decision.

3.4.2. Removal of bounding box error

All three uncertainties as a combination deteriorate the distance estimation capability of the system. In order to minimize the risk of known error sources, the bounding box error can be reduced by a short set of functions.



(1) (2) (3)

Figure 11 – Removal of the bounding box error. (1) Range of interest (2) Gaussian blur (3) Canny edge detection

In case an object is detected properly, a bounding box is created to define the region the obstacle is in. Sometimes the bounding box is larger than the object, in other cases a few sides or parts are cut through. Both cases are possible error sources that have to be prevented. When a bounding box is set, the algorithm extends it to a range of interest, practically meaning, that the bottom part of the box is expanded. Hereby it is assured, that the bottom of the detected vessel is inside of the analyzed region. As a next step, a Gaussian blur is applied with a kernel size of 3x3. Based on tests of multiple combinations, 3x3 based smoothing provided the best results as a preprocessing step. Finally, the Canny edge detection method was implemented, to identify the outline, especially the bottom part of the vessel. With the help of this step-by-step method, the middle point of the vessel's bottom can be easily selected, meaning also, that the bounding box error could be minimized significantly.

3.4.3. Detailed explanation of a bottom-up method

Now that objects are detected and some elementary error sources have been eliminated, a solid ground is set to apply distance estimation methods. The first bottom-up method builds up a mathematical model from geometrical principles, into a high-level distance mapping of image pixels. The approach is based on perspective geometry of image points on the ocean surface, approximated as a plane.

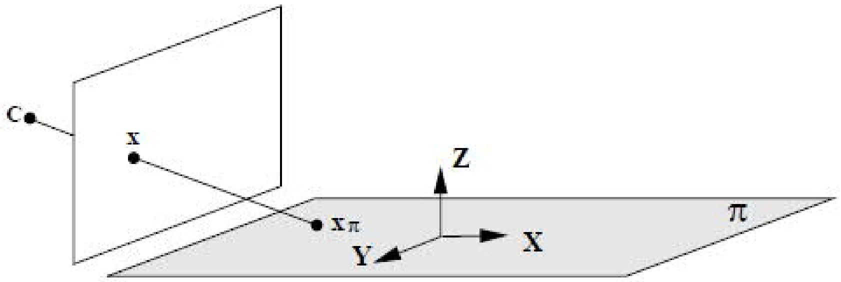


Figure 12 - Perspective image of points on a plane [34]

Given the mounting height and mounting angle of the camera, the image resolution, and the camera field of view the geometric mapping can be applied. This geometrical solution has been selected to be able to parametrize the model with initial values, to have an adaptable solution for vessels with different setups.

Parameter type	Parameter value
Mounting height	30 m
Mounting angle	-15°
Image resolution	704 x 576
Camera FOV	82°x 60°

The principle of the estimation method is, that a fixed camera setting defines a clear part on the ocean surface for each pixel, that concludes the whole visible region as one. As a first step, to each pixel with X-Y coordinates a distance value gets assigned, that is calculated based on geometrical properties and equations, since points on the image and scene planes are related by a projective transformation. Defining the values for the whole image, result in a complex, pixelwise distance mapping. When an object gets detected and localized, a representative point is selected and based on their X-Y coordinates, a distance value can be looked up and returned from the predefined mapping.

To calculate the values of the distance map pixelwise, first the middle column in the vertical direction has to be determined. Knowing the camera height, the mounting angle from the horizontal direction and the vertical FoV of the camera, distances can be calculated with the following formula using the pinhole camera model:

$$d = \frac{h}{\tan{(\beta_i)}} \tag{6}$$

where β_i is the angle of the viewing axis and the horizontal direction. *Figure 13* indicates two essential angles, β_{mid} that is equal with the mounting angle and defines the optical axis, which is the middle point of the image. Furthermore, β_0 indicates the lowest visible part of the image, that is equal with the border of a blind spot region.

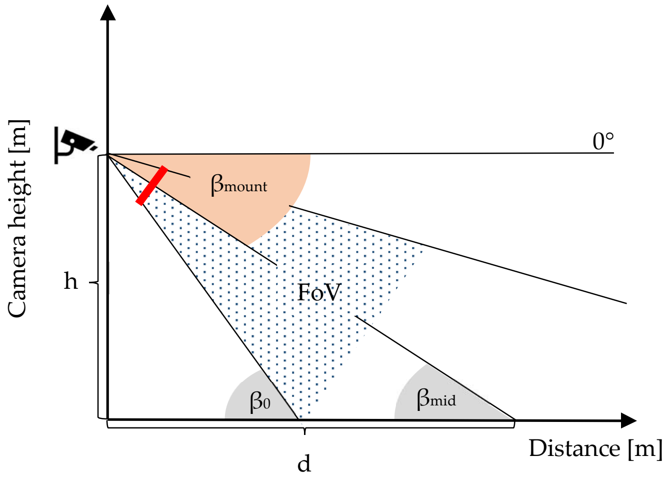


Figure 13 – Field of View, Mounting Angle, Blindspot angle, Optical axis angle

While the calculation of remarkable points is trivial, the challenge relies in the division of viewing axes and angles between the endpoints. At first, a solution would be to split the vertical Field of View in the number of the vertical image resolution to equal parts. Unfortunately, that assumption is mathematically not correct, since the angles determined by the viewing ranges are constant, but in reality, the pixels, that are a unit step on the image are constant and the viewing angle is changing.

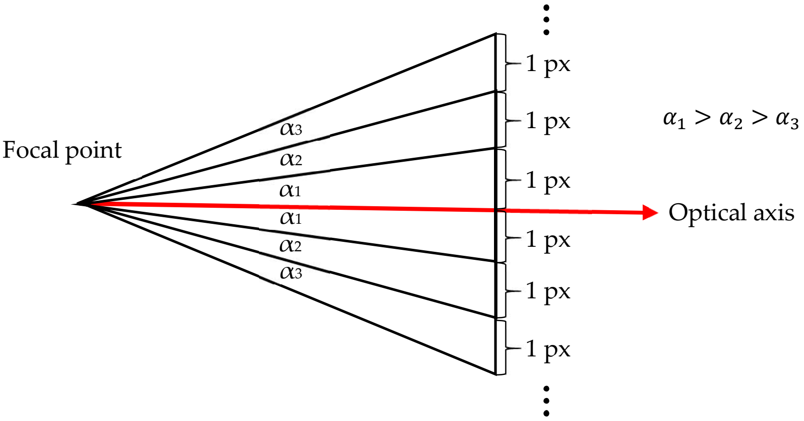


Figure 14 – Approximated vertical alignment of pixels and angles

To utilize the β_i angles on *Figure 14*, that are necessary for distance estimation, the α_i angles have to be calculated and added to the mounting angle, by that the optical axis of the image plane is rotated. Based on the sketch seen on Figure 13, the following equations can be determined:

$$\tan(\alpha_1) = \frac{1\ px}{dist_{pinhole}} \tag{7}$$

Where $dist_{pinhole}$ is the distance between the focal point and the image plane. With reshaping the equation, the angles shown in *Figure 14* can be calculated as follows:

$$\alpha_1 = \text{atan}\left(\frac{1\,px}{dist_{pinhole}}\right) \tag{8}$$

$$\alpha_1 + \alpha_2 = \text{atan}\left(\frac{2\,px}{dist_{pinhole}}\right) \tag{9}$$

$$\alpha_1 + \alpha_2 + \alpha_3 = \text{atan}\left(\frac{3\,px}{dist_{pinhole}}\right) \tag{10}$$

Following these steps until 288, that is the half of the image resolution and defines the half of the image plane, brings us to a general equation:

$$\sum_{i=1}^{n=288} \alpha_1 = \text{atan}\left(\frac{pixelstep_{vertical} \cdot i}{dist_{pinhole}}\right) \tag{11}$$

Where:

$$n = \frac{image\ resolution\ height}{2} = \frac{576}{2} = 288 \tag{12}$$

$$pixelstep_{vertical} = \frac{FoV\ vertical}{image\ resolution\ height} = \frac{60}{576} = 0.10417 \tag{13}$$

$$dist_{pinhole} = \frac{\frac{image\ resolution\ height}{2}}{tan\left(\frac{FoV\ vertical}{2}\right)} = \frac{288}{\tan(30°)} \tag{14}$$

With the equation *(11)*, all the changing angles can be calculated and added step by step to the direction of the optical axis, thereby all the vertical distance can be calculated with the equation *(6)*.

Now that the vertical angles and distances are assigned correctly, the horizontal division of pixel values has to be completed. The camera used in the application has a wide viewing range, meaning that the distance is highly affected at the side of the images. As a reference, the values of the middle column on the image are used for each row to generate the mapping sideways. Similarly, to the vertical calculations, the angle steps are calculated based on the horizontal FOV of the camera and the horizontal resolution of the image. The parceling is completed, when each X-Y points on the image were assigned an estimated distance value, which can be approximately displayed as heatmap of distances and coordinates. *Figure 15* is based on calculated values but serves only as a visualization to show the extension of distance categories. In the developed algorithm, each single pixel has a unique distance value assigned.

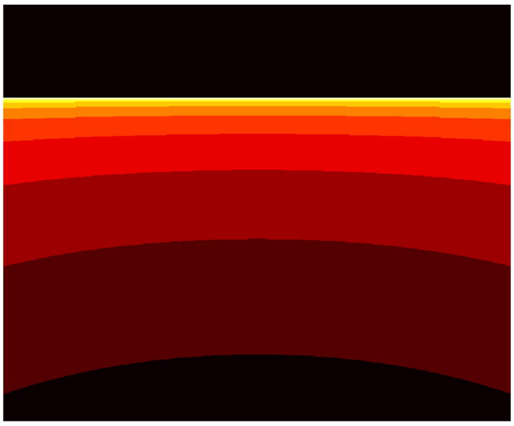


Figure 15 - Generated distance heatmap

As a final step, the distance heatmap has to be aligned with the real camera image. The camera used in the thesis was set up with a wide-angle lens that creates a barrel distortion on the image. To avoid a bent horizon and amorphous shape of vessels, an image rectification process has been applied with the help of camera calibration values. Interestingly, the rectified camera images could still not be perfectly overlaid, some misalignment has occurred near the horizon. The geometrical calculations are theoretically correct, but the issue might rely on approximated mounting height and angle from beforehand. A wrong mapping between the image and the distance map result in wrong estimation, therefore the issue needed to be corrected. To achieve a correct re-mapping, a recalculation has been done that takes the pixel value of the horizon as an additional parameter. Knowing the exact location of the farthest point, helps to recalculate the mapping between the horizon and the bottom of the image, that is the first visible part after a blind spot region.

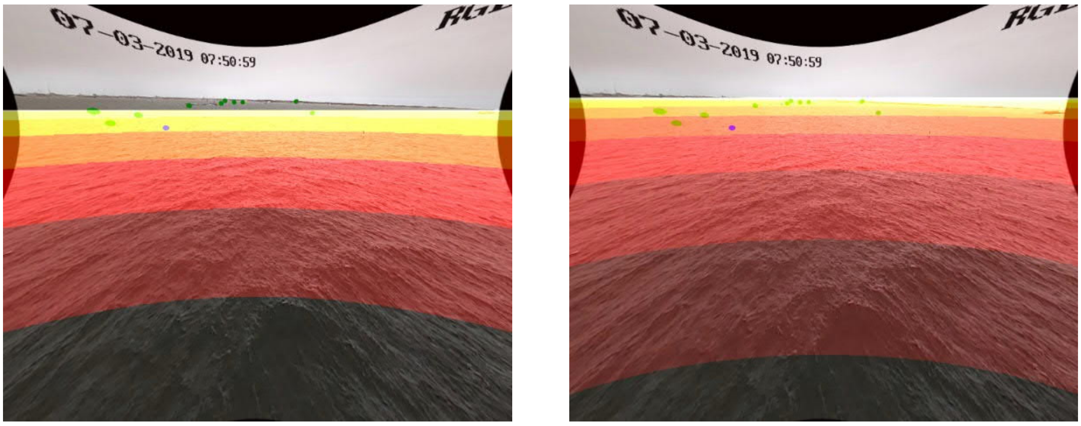


Figure 16 – Shifted mapping of distance heatmap and camera image (left), corrected alignment (right)

With the completion of the final correction step, a fully functional concept solution has been developed. Giving an input image, the obstacle gets detected using the YOLO network, a reference point gets selected and the adequate distance values get returned based on X-Y coordinates. The developed solution was validated on over 100 datapoints with known real-world distances and image location. Detailed results and evaluation are explained in later sections.

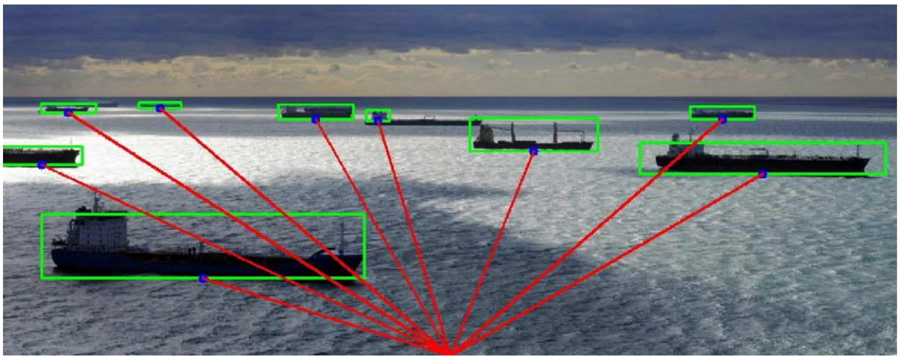


Figure 17 - Visualization of detected vessels

A main advantage of the solution is adaptability, due to the fact that changing a few initial parameters can recalculate the distance heatmap, if the method would be applied with a camera at a higher mounting point or a lower vessel. Furthermore, the basic geometrical calculations are mathematically proven and do not require special computational power, meaning that a real-time application is possible. On the other hand, some drawbacks are due to the discrete representation of the world through pixels. Each pixel defines a smaller or larger region on the ocean surface, where the whole region is concluded with on distance value. In reality, the distances vary within the observed range, but that cannot be analyzed in more detail by cameras. In the current implementation, the shortest distance within a region was selected for safety purposes, but in the future, the averaged middle-point could be considered. Furthermore, theoretical knowledge with prior assumptions and methods in optimal scenarios cannot always be applied in practice, where many more factors are affecting a system. Therefore, another method has been developed as well, that relies only on recorded data points.

3.4.4. Detailed explanation of a sampling based top-down method

The idea behind a sampling based, top-down method was to eliminate prior assumptions or geometric calculations and have a focus only on sampled data points. Similarly, to the solution of the Institute of Automation in Bremen, a set of 100 points have been selected by their X-Y values with known distance. The question is whether any correlation can be found among X-Y points and distances based on a low number of training points.

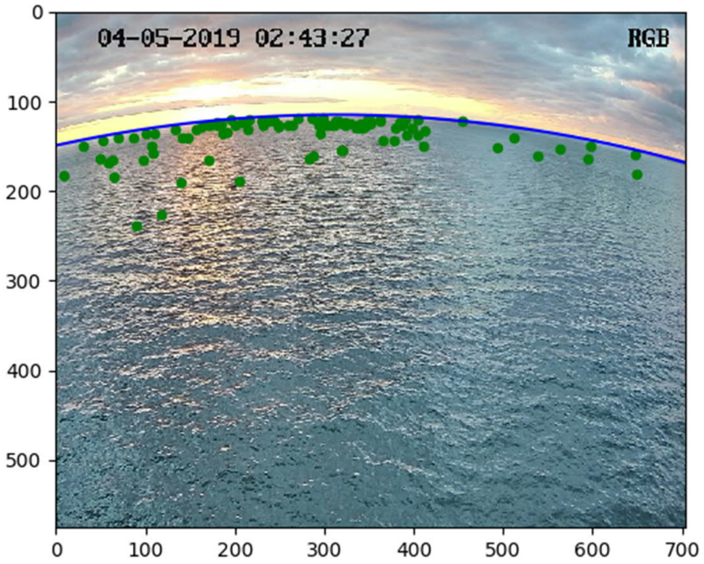


Figure 18 - Sampled datapoints and quadratic function of horizon

The only input for the approach, besides the set of datapoints, is the mathematical function of the horizon, that can be calculated based on a few reference points. The horizon is appearing on the images as a quadratic function with a certain offset from the top of the image. This particular quadratic function contains all the points that are approximately at an equal distance around the visible horizon. Similarly, all equally far points, from the base point of the ego vessel, are located on a curved line, that is shifted with a certain value from the top. When interpreting the challenge as a regression problem, preliminary results have shown that level of 0,8211 correlation can be found, up to four kilometers, between the distance of datapoints and the shifting value of the function. At regions between 2-3 kilometers, the method was able to predict distances with a relative error of less than 10%. However, at smaller distances and ranges above 4 kilometers, the method failed and could not provide acceptable results.

Complex regression problems require a large number of datapoints, to find a generalized solution with reasonable results and with avoiding overfitting. The principle of the approach was an interesting experiment, but a dataset of only 100 values were with orders of magnitudes smaller as would be needed. As the paper, created by the researchers and developers of DisNet also states, machine learning based distance estimation requires at least thousands of training datapoints to achieve significant results [26]. Unfortunately, collecting that amount of data was not possible during the thesis and was out of scope. Moreover, linear regression-based solutions, applied for the current use-case, try to approximate mathematically provable geometrical theorems, meaning that a higher uncertainty and accuracy is expected from the beginning on.

3.5. Bearing estimation based on geometry and sampling points

As the parameters estimated earlier, the calculation of the bearing is also affected by a number of uncertainties. While the physical limitation that result in distance estimations are not relevant for bearing, camera distortion and lens properties and mappings between image plane and ocean surface do influence the results. The method used for bearing estimation is solved as a regression problem with over 100 datapoints.

Datapoints on an image have clear X-Y values, which can serve as a base for multiple calculations. A basic approach is to calculate an angle from the ego vessels reference point on the image, which is the middle pixel of the bottom row.

$$\gamma = -\operatorname{atan}\left(\frac{\frac{X_{imageSize}}{2} - X_{dataPoint}}{Y_{imageSize} - Y_{dataPoint}}\right) = -\operatorname{atan}\left(\frac{352 - X_{dataPoint}}{576 - Y_{dataPoint}}\right) \qquad (15)$$

Naturally, the angle on the image plane is not equal with the bearing value on the water surface, but with known values, the calculated angle for each datapoint can be plotted in relation to the real-world bearing value.

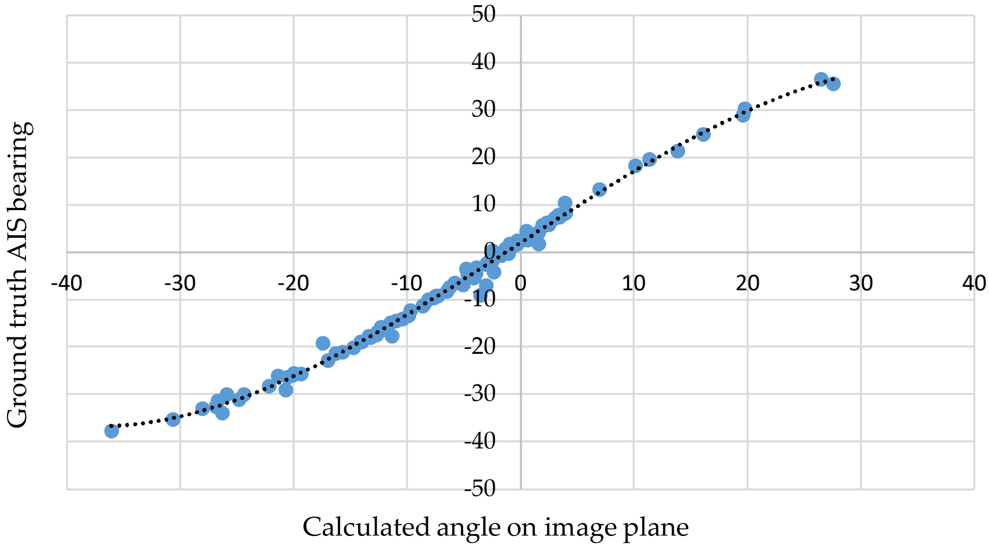


Figure 19 - Bearing estimation based on regression

As it can be seen on Figure 19, a pattern can be recognized and a with help of the Least-Squared-Error method, a regressive trendline with a known mathematical function can be set. Although linear regression has also achieved outstanding results with over 0.98 correlation, a third order polynomial line has been fitted to the datapoints. Due to the tangent functions hyperbolic attribute, the values follow the regression line without overfitting.

$$Bearing = -0.0004 \cdot \gamma^3 - 0.0004 \cdot \gamma^2 + 1.5529 \cdot \gamma + 2.00449 \qquad (16)$$

Given an X-Y value on the cameras input, the angle on the image can be calculated from the reference point and with the polynomial equation, the bearing can be estimated. Based on a test set of 100 datapoints, a correlation of 0.9943 has been achieved, resulting in an overall performance of less than 0.59° average absolute error.

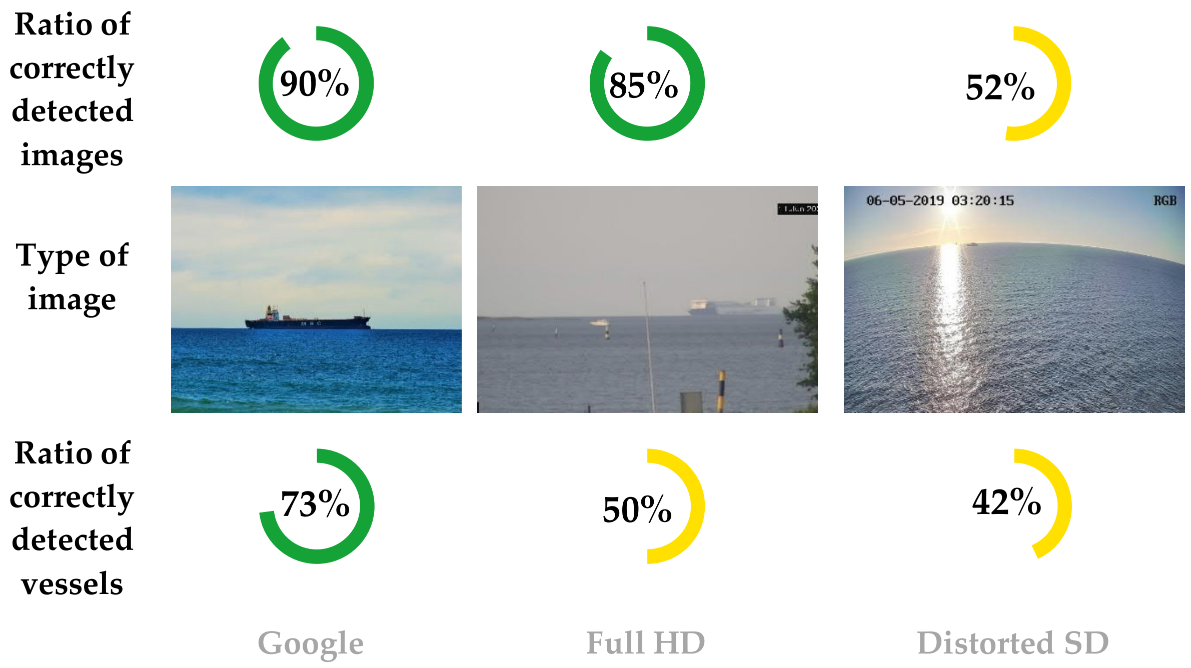
4. RESULTS AND DISCUSSIONS

4.1. *Performance analysis*

During the development phase, many methods have been tested in each research area. The analysis in this section deals with the most promising methods that either provide good results or have beneficial properties, such as adaptability or usability. Similarly, how the development phase was divided into object detection and distance estimation, the performance analysis presents both areas independently.

4.1.1. Performance of object detection

After a solid research on methodologies, the YOLO deep learning frameworks has been chosen as the primary object detection method for the thesis. As a general evaluation for the deep learning method, only a few false positive cases have occurred, but most issues were true negative cases, when existing vessels could not be identified and localized on camera images. Close-by, colorful shapes were detected correctly, but ships at farther distances and distorted forms on the side of images were challenging. Presumably, the explanation is that the training dataset for the framework has been prepared on large quality, colorful images with clear shapes and adequate contrasts. In contrary, images taken in real-life conditions show low-stimulus scenarios with less outstanding features and blurred colors, not even talking about foggy weather. Moreover, due to a low image quality, smaller appearing vessels did not contain enough pixels to define a clear outline, that could be detected as a feature for the system.



As it was predicted in detail in earlier sections and proved with test measurements, the image quality and resolution play a key role in detection performance. The evaluation has measured in how many cases at least one vessel was detected on an image and how many vessels were recognized from all. Three different image classes have been tested, Google images, Full HD camera images and distorted SD images that were used in the thesis work. Results has shown, that while Google like images have detected vessels on images correctly in 90% of the cases, only half of the validation images were recognized as containing a vessel for Distorted SD images. In addition, on images with multiple vessels 73% of all objects have been detected, but the real-world full HD images have already dropped back to 50%. While at least one vessel was recognizable in most cases, smaller vessels were already a difficulty for the system. Although it needs to be highlighted, that the distorted SD resolution images contained some highly challenging scenarios, only 42% of the ships were recognized. Besides the fact of being detected, the confidence of an object class does also play an important role. In can be stated, that most objects in the ship category were classified with an average confidence score between 70-90%.

Identifying different performances across image quality, the re-training of the network with custom, low-quality images and shapes was considered. However, some benchmark measurements have shown, that using transfer learning for a network does not necessarily provide better results until a large number of images are not reached. The YOLO network's ship category was trained with ten thousands of images, while re-trained solutions with a few thousand had a lower performance compared to the original network. Considering the scope of the thesis and the available datasets, the retraining of the network, to achieve better performances, was not reasonable and possible at the time.

4.1.2. Performance of distance estimation

After experimentation with multiple methods, the bottom-up approach has provided the best result, that builds up a distance map for each image pixel based on geometrical calculations. As it was predicted in the beginning, it has been proven that the estimation error rises on larger distances, due to physical limitations of the world's representation by cameras.

Results have been evaluated based on a test set of 100 datapoints, that had known distances in form of AIS information. At ranges lower than three kilometers, a relative error of only 11-12% has been achieved. As the distances get larger and the image resolution deteriorates, since the number of representative pixel points decreases, the relative error starts to raise significantly as it can be seen on *Figure 20*.

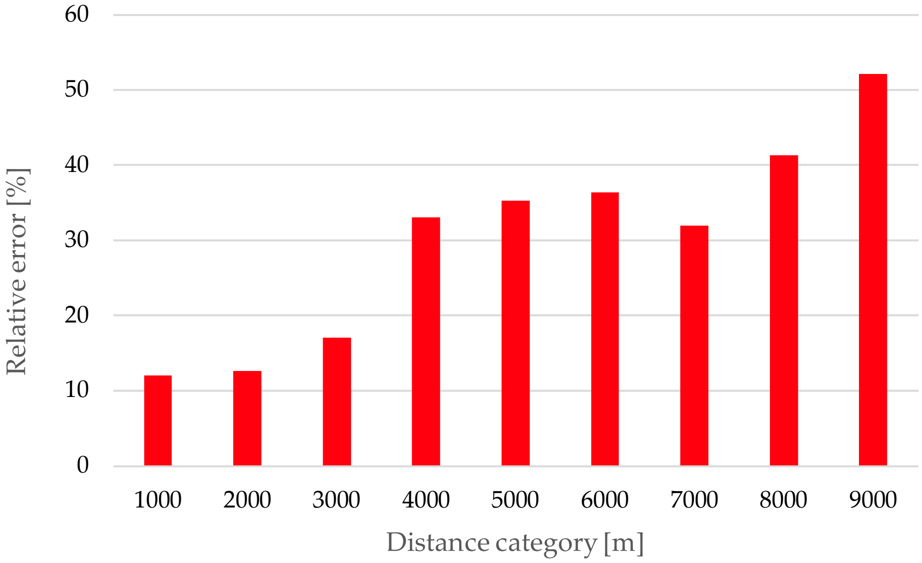
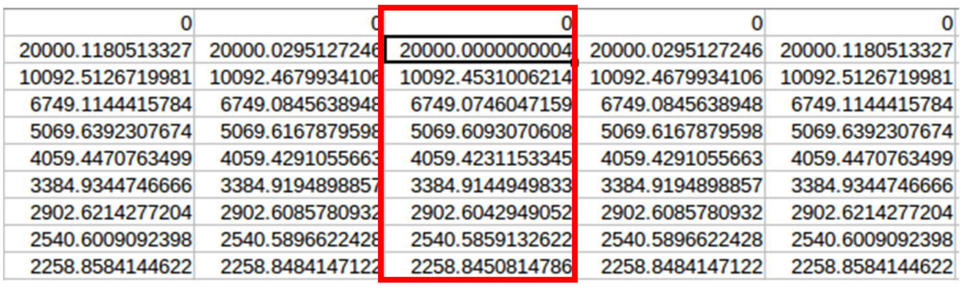


Figure 20 - Relative error of distance estimation based on range categories

Unfortunately, when using the camera as a discrete representation of the world, some physical limitations have to be taken care of. *Figure 21* shows explicitly, how drastically the distance values of pixels near the horizon can change. In the thesis, most error sources were identified in the beginning on, and possible improvement factors have been applied. Even though, a higher resolution was physically not possible to achieve on larger distances.



0	0	0	0	0
20000.1180513327	20000.0295127246	20000.0000000004	20000.0295127246	20000.1180513327
10092.5126719981	10092.4679934106	10092.4531006214	10092.4679934106	10092.5126719981
6749.1144415784	6749.0845638948	6749.0746047159	6749.0845638948	6749.1144415784
5069.6392307674	5069.6167879598	5069.6093070608	5069.6167879598	5069.6392307674
4059.4470763499	4059.4291055663	4059.4231153345	4059.4291055663	4059.4470763499
3384.9344746666	3384.9194898857	3384.9144949833	3384.9194898857	3384.9344746666
2902.6214277204	2902.6085780932	2902.6042949052	2902.6085780932	2902.6214277204
2540.6009092398	2540.5896622428	2540.5859132622	2540.5896622428	2540.6009092398
2258.8584144622	2258.8484147122	2258.8450814786	2258.8484147122	2258.8584144622

Figure 21 - Distance values near the horizon in the middle of the image

Giving the circumstances, the available data sources and the condition that a monocular camera should be used, a well performing distance estimation method has been developed. The successfulness of the method relies in fact, that early recognition of error sources and limitations have been made and proven by validated results. Naturally, the acceptable error of such system is hard to determine, since the ground of comparison is not settled and expectations are dependent on sensor types and application goals. To make a relevant validation for the estimation method in the current use-case a human comparison has been made.

4.1.3. Human comparison

During the thesis, an online research survey has been conducted to measure the human distance estimation capability. The survey, that has been filled out by 55 industry professionals, contains 10 images of open sea scenarios with vessels, where the volunteers had to estimate distances with limited prior knowledge. A part of the survey assessed the background of volunteers, thereby a clear line could be drawn between captain and OOWs working on a ships bridge, crew members on vessels or even engineers working in maritime related companies. Thereby, the performance could be also separated and compared with people, having different background knowledge and skills.

At first, the average of human estimates have shown distinguished results compared to the ground truth values. However, detailed investigation has indicated, that a large variance can be seen, sometimes even a range from 200 meters up to 20 kilometers. Since crew members from the bridge are the most relevant category, on comparison is based on their filtered answers. In 80% of the cases, the system developed in the thesis has estimated a better value than the absolute average estimates of captains. When observing individual cases across all cases, only in 17% of the volunteers could perform equally good or better than the developed algorithm. Interestingly, when all participants are considered, already 23% could perform equally or better, meaning that based on the results, a more general composition had achieved better estimations.

As a conclusion it can be stated that the developed system in the thesis could outperform human distance estimation capabilities on many levels. Due the high variance and error of the manual monitoring of the environment, such technical system could contribute to enhance safer operations in the maritime industry.

4.2. *Adaptability*

A key question of the developed system is in what form it can be adapted to safety critical functionalities of future autonomous solutions. First of all, the research of SoTA solutions serve as a solid guideline, which methodologies can have the potential to be implemented. Knowing their advantages and drawbacks, the optimal approach can be selected for given use-cases. Secondly, the YOLO based object detection method, that was selected in the thesis, is an adaptive solution, independent from camera settings. As it was stated in the *Results* section, possible improvement options have to be considered, to be able to detect smaller objects on larger distances. Finally, the geometrical distance estimation method has been designed in a way, to be only dependent on a minimal number of input parameters. With known mounting settings, camera properties and just a few camera images, the method recalculates the distance mapping for the whole visible area. Although, a more detailed validation would be required for further steps, the fully functional algorithms developed in the thesis can serve as a stable ground for future developments.

5. CONCLUSION

5.1. Retrospection to motivation and technological solutions

The overall goal of the thesis was to confirm or refute the fact, whether a monocular camera-based assistance system could improve the safety and establish new features in the maritime industry. However, one of the main challenges was to define, how to measure improvements, what is the base of the comparison and how to decide what is considered as an acceptable solution. In addition, an initial goal was to give substantial advice on feasibility and emerged limitations, that might have an effect on the implementation in commercial use.

At the beginning of the thesis, literature review has been conducted to explore existing solutions and methodologies, to understand strength and weaknesses that might affect the system. Later on, a framework has been built, to generate validation data of existing measurements, such has RGB images, thermal images and processing of GPS coordinates. Besides using state-of-the-art solutions as a general guideline, multiple methodologies have been experimented as a next step. Initial results have served as a base to move towards YOLO based object detection and the geometrical bottom-up distance estimation. Moreover, the developed approaches have been validated on ground truth data, where results have shown an outstanding evaluation for bearing estimation and good results for distance estimation. To answer the question how well the system's performance is, a manual visual monitoring has been used as a base of comparison. In summary, the developed concept solution outperformed the human distance estimation capability, that was collected in form of an online survey, with more than 50 industry professionals.

Emerged limitations have been discovered in detail already at an early stage, that affect the performance and accuracy of the system. Furthermore, eventual physical limitations have been defined that might set boundaries for industrial applications. With lessons learned from existing solutions and having knowledge on occurring limitations, a tangible result is that future suggestions can be made to foster decision makers on feasibility of future applications.

As an overall conclusion it can be stated, that if certain infrastructural requirements can be achieved in a cost effective way, and some physical limitations, such as observable distance, are admissible, then a similar solution can definitely enhance safer operation, which leads to an optimal and efficient utilization of crew members on the long run.

5.2. Future suggestions

The review of existing solutions, experimentation with various methods and validation on real-world data have led to a clear conception of future improvements possibilities and realistic expectations.

First of all, there is some possible space of improvement in the object detection method. The thresholding and saliency method, that was experimented but not used in the end, has shown promising results, but a general approach could not be developed. Changing lighting conditions have impacted the image composition too heavily, and parameters had to be finetuned for each individual case. The method might bring false positive cases, but a more detailed experimentation might lead to new findings. The YOLO object detection has performed well on Google-like images, but a significant deterioration has been found on real-life images, with less characteristic shapes and low-contrast colors. Applying a transfer learning method and retraining the deep learning network should be considered with a large dataset of real-life, lower quality images. Unfortunately, lack of training images and the scope, limited the experimentation possibilities in the thesis, but there is significant potential in improving the detection rate. However, in the current use-case, training the framework with labeled data points of small vessels might lead to an increasing number of false positives. Regardless of the outcome, the first step for the approach needs to be a generation of a large dataset with thousands of training images.

The distance estimation system faced many challenges. As a first suggestion, the exact use-case and application area needs to be reviewed, considering the physical limitations of camera-based systems. At shorter distances, the evaluation has shown acceptable and applicable results, but the working principle of cameras are a restraining force for the performance at larger distances. Due to the changing resolution of distance estimates, when a step in pixels is taken, a new mapping method should be considered. Each pixel represents a certain area on the water surface, which can grow up to kilometers size in the far. Instead of assigning an exact numerical value to the whole area, the distance range covered by the pixel could be added. Thereby, unnecessary estimation errors could be eliminated, since a higher sampling accuracy cannot be achieved anyway. Last but not least, a high mounting height and a large negative mounting angle is suggested. In that case, a long range can be observed, but the majority of pixels is facing the water surface, meaning that a higher resolution could be achieved for estimation. As a matter of course, the highest possible image quality needs to be implemented, to improve detection performance and estimation accuracy.

In addition, novel method has not been covered in the thesis, namely the usage of thermal cameras. Based on some available test images, the object detection could be implemented in open sea scenarios. In these cases, clear vessel shapes have been identified, but crowded situations deteriorate visible contours. Unfortunately, the YOLO network was unable to detect vessels on thermal images, but if the network could be trained with objects on a large set of thermal images, a stable detection might be achieved. Having a working solution as suggested could allow a utilization of camera systems in all visibility conditions such as night or fog.

As a summary, the research work in the thesis has already reached valuable results that serve as a guidance for future continuance. With the help of proven and experienced suggestions, further improvements can be applied and a development direction could be settled, that could raise detection rate and estimation accuracy to an industrially applicable level.

6. REFERENCES

[1] E. Lehtovaata and K. Tervo, "B0 – a conditionally and periodically unmanned bridge," *Abb*, p. 12, 2018.

[2] V. Bertram, "Towards Unmanned Ships 1," pp. 1–52, 2013.

[3] A. N. Update, A. For, T. H. E. Finnish, and M. Cluster, "Smart Maritime Technology Solutions," 2017.

[4] "Automation of Ships in Ports and Harbours," vol. 44, no. 0, pp. 1–7.

[5] MUNIN, "The Autonomous Ship," 2016. [Online]. Available: http://www.unmanned-ship.org/munin/about/the-autonomus-ship/.

[6] A. Kirchner, "Rise of the Machines – Legal analysis of Seaworthiness in the context of autonomous shipping," 2019.

[7] M. Blanke, M. Henrique, and J. Bang, "DTU Management Engineering A pre-analysis on autonomous ships," pp. 1–27, 2017.

[8] K. M. Kaspersen, "Marine Radar Properties, Analysis and Applications," 2017.

[9] G. Elettronica, "Vessel Traffic Service (VTS)." [Online]. Available: http://www.gemrad.com/vessel-traffic-service-vts/.

[10] T. Keiki, "Marine Radar BR-3200 and ECDIS EC-8000/8500 : DigInfo." [Online]. Available: https://www.youtube.com/watch?v=yS6Fp-ifUgQ.

[11] M. Knébel, "Speciális környezet felismerésének fejlesztése radar alapú vezetéstámogató rendszerekben," 2017.

[12] N. Haf and O.-A. Pantazis, "LIDARs Usage in Maritime Operations and ECO – Autonomous Shipping , for Protection , Safety and Navigation for NATO allies Awareness."

[13] D. Thompson, "Maritime Object Detection, Tracking, and Classification Using Lidar and Vision-Based Sensor Fusion," *Diss. Theses*, 2017.

[14] B. J. Tetreault, "Use of the Automatic Identification System (AIS) for maritime domain awareness (MDA)," *Proc. Ocean. 2005 MTS/IEEE*, pp. 1590–1594, 2005.

[15] B. U. of T. (TUB) C. V. and R. S. Group, "Automatic Image Analysis - Typical Tasks," 2019.

[16] B. U. of T. (TUB) C. V. and R. S. Group, "Automatic Image Analysis - Fourier Descriptors," 2019.

[17] R. Cong, J. Lei, H. Fu, M. M. Cheng, W. Lin, and Q. Huang, "Review of visual saliency detection with comprehensive information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 10, pp. 2941–2959, 2019.

[18] X. Mou and H. Wang, "Image-Based Maritime Obstacle Detection Using Global Sparsity Potentials," *J. Inf. Commun. Converg. Eng.*, vol. 14, no. 2, pp. 129–135, 2016.

[19] T. Cane and J. Ferryman, "Evaluating deep semantic segmentation networks for object detection in maritime surveillance," *Proc. AVSS 2018 - 2018 15th IEEE Int. Conf. Adv. Video Signal-Based Surveill.*, 2019.

[20] S. Moosbauer, D. Konig, J. Jakel, and M. Teutsch, "A benchmark for deep learning based object detection in maritime environments," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2019-June, pp. 916–925, 2019.

[21] P. Alizadeh, "Object Distance Measurement Using a Single Camera for Robotic Applications by Peyman Alizadeh A thesis Submitted in partial fulfillment of the requirements for the degree of Master of Applied Sciences (M A Sc) in Natural Resources Engineering The Facult," *Object Distance Meas. Using a Single Camera Robot. Appl.*, p. 126, 2015.

[22] M. H. Clemens Holzmann, "Single-Camera Stereo Vision for Mobile Devices," 2012. [Online]. Available: https://mint.fh-hagenberg.at/?p=1149.

[23] T. Huntsberger, H. Aghazarian, A. Howard, and D. C. Trotz, "Stereo vision-based navigation for autonomous surface vessels," *J. F. Robot.*, vol. 28, no. 1, pp. 3–18, 2011.

[24] R. Gladstone, Y. Moshe, A. Barel, and E. Shenhav, "Distance estimation for marine vehicles using a monocular video camera," *Eur. Signal Process. Conf.*, vol. 2016-Novem, pp. 2405–2409, 2016.

[25] R. Ranftl, V. Vineet, Q. Chen, and V. Koltun, "Dense Monocular Depth Estimation in Complex Dynamic Scenes," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 4058–4066, 2016.

[26] M. A. Haseeb, J. Guan, D. Ristić, and A. Gräser, "DisNet : A novel method for distance estimation from monocular camera," *10th Planning, Percept. Navig. Intell. Veh.*, 2018.

[27] C. M. Thomas and W. E. Featherstone, "Validation of Vincenty's formulas for the geodesic using a new fourth-order extension of Kivioja's formula," *J. Surv. Eng.*, vol. 131, no. 1, pp. 20–26, 2005.

[28] A. Upadhyay, "Formula to Find Bearing or Heading angle between two points: Latitude Longitude," 2015. [Online]. Available: https://www.igismap.com/formula-to-find-bearing-or-heading-angle-between-two-points-latitude-longitude/.

[29] Pandas, "Pandas DataFrame," *Pandas*, 2014. [Online]. Available: https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.html.

[30] R. Gandhi, "R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms," 2018. [Online]. Available: https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e.

[31] J. Hui, "Object detection: speed and accuracy comparison (Faster R-CNN, R-FCN, SSD, FPN, RetinaNet and YOLOv3)," 2018. [Online]. Available:

https://medium.com/@jonathan_hui/object-detection-speed-and-accuracy-comparison-faster-r-cnn-r-fcn-ssd-and-yolo-5425656ae359.

[32] J. Redmon and A. Farhadi, “YOLO v.3,” *Tech Rep.*, pp. 1–6, 2018.

[33] A. Ponnusamy, “cvlib - high level Computer Vision library for Python,” 2018. [Online]. Available: https://github.com/arunponnusamy/cvlib.

[34] A. Z. Richard Hartley, *Multiple View Geometry in computer vision*. .