Journal of Unmanned System Technology

Motion Planning for USV using Classification Capabilities

Oren Gal

Technion, Israel Institute of Technology and Haifa University

Abstract — This paper presents an automatic method to acquire, identify, and track obstacles from an Unmanned Surface Vehicle (USV) location in marine environments using 2D Commercial Of The Shelf (COTS) video sensors, and analyzing video streams as input. The guiding line of this research is to develop real-time automatic identification and tracking abilities in marine environment with COTS sensors. The output of this algorithm provides obstacle's location in x-y coordinates. The ability to recognize and identify obstacles becomes more essential for USV's autonomous capabilities, such as obstacle avoidance, decision modules, and other Artificial Intelligence (AI) abilities using low cost sensors. Our algorithm is not based on obstacles characterization. Algorithm performances were tested in various scenarios with real-time USV's video streams, indicating that the algorithm can be used for real-time applications with high success rate and fast time computation.

DOI: 10.21535/just.v8i1.1015

I. Introduction

PATH planning and obstacle avoidance are an important issues for Autonomous Underwater Vehicles (AUVs), and as can be noticed lately, these fields are extensively studied. Autonomy for AUVs challenges can be mainly described as combination of currents disturbances, limited perception capabilities, dealing with non-holonomic constrains by trying to face limited energy optimal trajectories with computational efficiency and no matter what –safety. Perception capability becomes crucial for USV's autonomous capabilities.

One of the most difficult challenges for USV navigation is to recognize and identify obstacles around the vehicle without human intervention. This task is known as Automatic Target Detection (ATD). An efficient ATD system should achieve high detection percentage for targets while maintaining a minimal false-alarm rate. This means that it must preserve an optimal balance between a high detection rate and a low error probability.

Although, ATD algorithms are very sensitive and unstable regarding clutter elements, *i.e.*, elements that are not targets but still part of the scenes with similar characteristics as the targets. Dealing with clutter in ATD algorithms was extensively studied[1,2,3].

One of the ATD algorithms methods is based on the target temperature. The contrast of the target were based on environment gradient of the target and the environment's contrast to recognize targets. These methods suffer from false alarms due to targets and environment similarity. Several methods were developed to characterize the targets and to distinguish between the target's and the environment's characteristic, avoiding false alarms[3,4].

Heuristic methods were introduced in the early 1980s based on threshold gradient in the image. The threshold was determined by the contrast of an object with its local background[3]. The segmentation part of such algorithms is based on standard edge operator using closing shape algorithms and filling steps[5,6,7,10].

Most of the previous works were studied in aerial and ground environments without considering special phenomena in marine environments, such as waves, clutter, vehicle's stability, etc. For the first time, our algorithm deals with ATD algorithms for USV's motion planning and automatic decision models using COTS sensors.

The algorithm based on common video format as an input, and computes targets locations around the vehicle. Our method is able to both detect objects of interests, and classify them with high probability.

The presented approach is based on the notion of Machine Learning, a well-known concept which has become the de-facto standard in many applications. The basic idea in this method, as opposed to the classical heuristics, such as gradient thresholds, is to provide a learning algorithm with a large data-set which contains samples of the problem along with their solution; the algorithm, after a process of "learning", will eventually provide a way to map between a sample and its solution. This paper aim at introducing the use of learning techniques in the domain of marine vessels detection. We show that well known algorithms perform well on the presented task.

Our contributions are:

- Presenting basic concepts in deep learning, and import them in the discipline of marine robotics.
- Training deep learning models on relevant data, as well as collecting and labeling it.
- Providing an algorithm which can detect and classify marine objects with high probability.

Corresponding author: Oren Gal (e-mail: orengal@technion.ac.il)
This paper was submitted on July 6 2018, and accepted on February 1, 2019.

II. Artificial Neural Networks

In this section, we present basic principles of deep learning; Artificial neural networks are defined, in addition to the training process.

Artificial neural network is a computational model which is extensively used in many complicated machine learning tasks, such as image classification, machine translation, automatic text generation and much more. The architecture of artificial neural network is inspired by the structure of the human brain: it consists of small computation units, called *neurons*, and connections between them. Each unit is capable of performing very simple task, but the combination of multiple neurons makes the model computationally strong.

A. Preliminaries

Let us begin with feedforward neural network. A more detailed explanation may be found it[19]. The goal of feedforward neural net is to approximate some function f. For example, a classifier $f_{(x)}$ is a function that maps input to category; in the context of the paper $f_{(x)}$ can be a function which on given a photo of a marine vessel returns its class.

Feedforward neural network is composed of multiple *layers* (hence *depth*); a layer is a function of a certain form which will be explained soon. The entire network is the composition of the layers: assume we have two layers: $l_1(x)$, $l_2(x)$, then the network will output $f(x) = l_2(l_1(x))$. l_1 is called the first layer (or **input layer**), l_2 is the second layer, and so on. The last layer is called the **output layer**, and all the layers between the input and output are called **hidden layers**. The depth of the network is defined as the number of layers.

Layer l has a *linear transformation* and an *activation function*: $f(W_l u + b_l)$ where u is the previous layer's output, b_l is the bias and f is the activation function (which operates coordinate-wise on the vector). The activation function depends on the task. For example, in binary classification task one may use the sigmoid function: $S(x) = \frac{e^x}{1+e^x}$ or rectangular function. The matrices W and biases b are called the *weights* of the network; these are the variables which we optimize during the learning process.

In practical problems usually there are dozens of thousands variables –what makes the learning process so difficult.

The Universal Approximation Theorem states that every continues function f on compact subset of R^n can be approximated up to any desired accuracy using feed-forward neural network with a single hidden layer and sigmoid function[14]. However, the size of the hidden layer may be exponential in the desired accuracy. Thus, variety of architectures were proposed, for many different applications.

B. Training process

Given a set of points $\{x, f^*(x) + \varepsilon\}$ where $x \in R$, f^* is an arbitrary function and ε is noise, the training process aims to adjust the functions of the neural net so that for all x, $f(x) \approx f^*(x)$.

The training process is essentially an optimization problem: the loss function, the function that measures the performance of the model, is being minimized, subject to the variables of the network. Since the network is usually very large, it is impossible to minimize the loss function using simple differentiation, an approximation method is needed.

One of the most simple and intuitive method is computing the gradient and "follow" its direction. There are no optimality guarantees for this method; we are likely not complete the process at a real minimum of the function. Computing the gradient is done using back-propagation algorithm[20].

C. Convolution neural network

Convolution neural networks are special type of neural networks. They composed, in addition to the layers type which were previously discussed, with convolution layers. For further explanation, we refer the reader to [21].

Convolution neural networks performs extremely well on many image recognition tasks, and they form the basis of the network used in this paper.

III. Experimental results

We compared a trained version of MobileNetSSD[17] and YOLO[18]. These trained networks are capable to detect wide range of object, however, the focus of our work is to detect objects in the sea. Thus, a further training process is needed.

We have collected 5 hours of video, with resolution of 3840×2160 pixels and 30 frames per second. This video contained moving merchants, tugboats, kites and kayaks. After collecting it, the video was labeled. In order to synthetically create more training data, we used a few methods of data augmentation, including: adding Gaussian noise, rotation and scaling.

On each epoch, the data were divided to training set (90 percent of the data) and validation set (10 percent of the data).

Two measures were brought into account:

- Confidence Loss: this measures how confident the network is of the objectness of the computed bounding box. Categorical cross-entropy is used to compute this loss.
- Location Loss: this measures how far away the networks predicted bounding boxes are from the ground truth ones from the training set. L2-Norm is used here.

It is possible to observe the results in Figure 1 and Figure 2. Additionally, results of the algorithms can be seen in Figure 3, Figure 4 and Figure 5.

The trained model is used as the detector and classifier.

Move videos describing our results can be found in the following links:

- https://youtu.be/6qUF8Iu1fY4
- https://youtu.be/3bIw0FOa5AM
- https://youtu.be/HUTjE4zdIVU

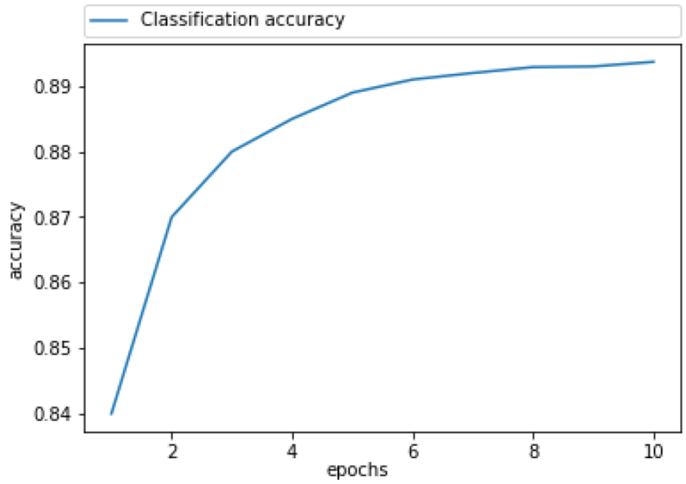

Figure 1 Classifier Accuracy vs. Number of Epochs

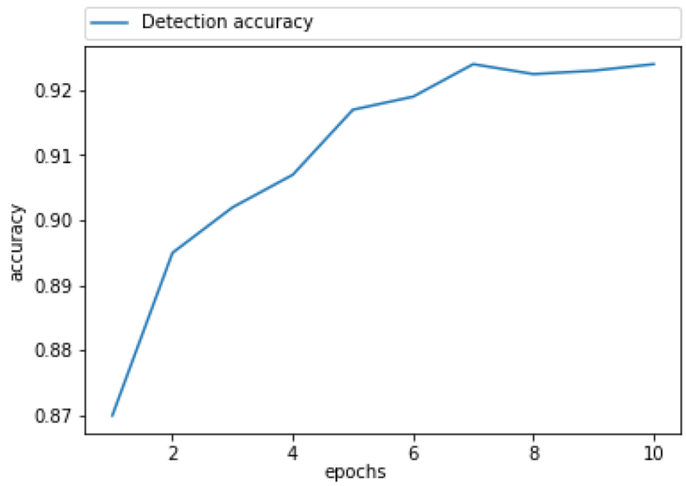

Figure 2 Detector Accuracy vs. Number of Epochs

Figure 3 Small Boat Detection and Classification

Figure 4 Boat Detection and Classification With Hard Conditions

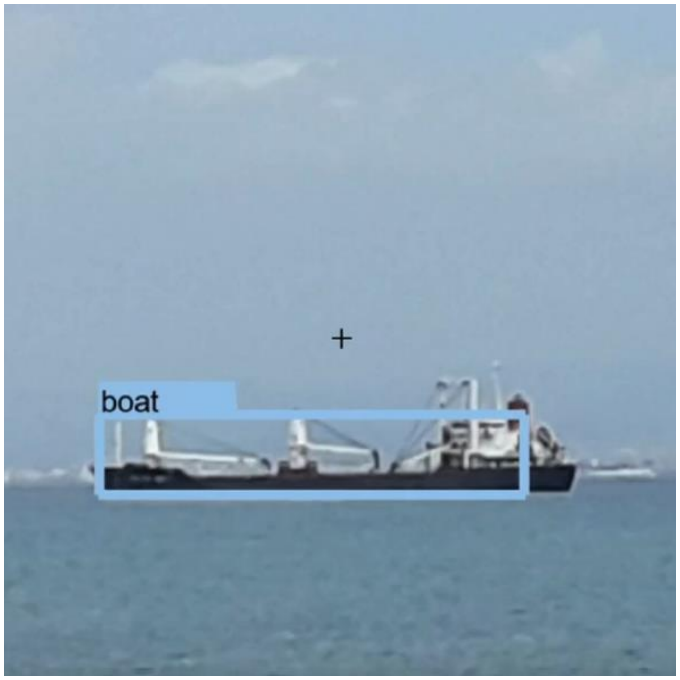

Figure 5 Complicated Boat Objects Detection and Classification

IV. CONCLUSION

This paper presents a basic and very efficient algorithm for marine environments that was tested on different scenes. Yet, there are some limitations and untested cases that should be treated in future research. Our algorithm is limited to sensor noises and false alarm dealing with very cluttered environments. The algorithm has been tested with stabilize sensor on the platform, which simplify horizon line recognition. In case of not stabilized sensor or panoramic image integrating several video streams, we expect a very limited success rate.

Another well know problem in marine environments is related to sun effects. The algorithm might classify sun light effects as targets, which will cause false alarms. This challenge should be treated in our further research, for more accurate ATD abilities for accurate USV trajectory planning.

The highlight of the algorithm is the simple concept that can be use on COTS sensor without special hardware, and as far as we know this specific issue was not yet studied extensively in marine environments. The algorithm is based on simple basic algorithms from the image processing world, which are suitable for real-time application.

We presented a new algorithm to acquire identify and track obstacles from USV systems using COTS sensor for autonomic navigation. The algorithm is based on previous image processing filters and algorithms, and been been adapted to the marine environment challenges.

The algorithm was successfully tested on real-time video from USV systems, and can be apply in real-time applications dealing with CPU time constraints. Further research directions include light effects and not stabilized systems.

Acknowledgement

The author would like to thank Gal Yehuda for simulations and remarks on this paper.

References

[1] Bhanu, C. and Holben, R. D.: Model-based segmentation of FLIR images. IEEE Trans. Aerosp. Electron. Syst. 26, 2–10 (1990).

[2] Ben-Yosef, N. Bahat, B. and FeiginBhanu, G. and Holben, C.: Simulation of IR images of natural backgrounds. Appl. Opt. 22, 190–193 (1983).

[3] Ratches, J. A. Walters, C. P. Buser, R. G. and Guenther, B. D.:Aided and automatic target recognition based upon sensor inputs from image forming systems. IEEE Trans. Pattern Anal. Mach. Intell. 19, 1004–1019 (1997).

[4] Casasent, D. P. and Neiberg, L. M.:Classifier and shift-invariant automatic target recognition neural networks. Neural Networks. 8, 1117–1129 (1995).

[5] Schachter, B. J. Lev, A. Zucker, S.W. and Rosenfeld, A.:An application of relaxation methods to edge reinforcement. IEEE Trans. Syst. Man Cybern. 7, 813–816 (1997).

[6] Walters, D. K. W.:Computer vision model based on psychophysical experiments.Pattern Recognition by Humans and Machines, 2 (1986).

[7] Broy, M.: Early vision. In: Rosenfeld,A.. (eds.) Perspectives in Computing, pp. 190-206. Academic Press, New York (1986).

[8] Marham, K. C.:Comparison of segmentation processes for object acquisition in infrared scenes. IEEE Radar Signal Process. 136, 13–21 (1989).

[9] Davies, E.R.:Machine Vision: Theory, Algorithms, Practicalities, 2nd ed., Academic Press, New York. (1997).

[10] Marham, K. C.:Clutter metrics for target detection systems. IEEE Trans. Aerosp. Electron. Syst. 30, 81–91 (1994).

[11] Vijaya, K.:A tutorial survey of composite filter designs for optical correlators. Appl. Opt. 31, 4773–4798 (1992).

[12] Flannery, D. Loomis, J. and Milkovich, M.:Transform-ratio ternary phase-amplitude filter formation for improved correlation discrimination. Appl. Opt. 27, 4079–4083 (1988).

[13] http://sites.google.com/site/orenusv/home/research-1/usvvision

[14] Cybenko, George. : Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals, and Systems. 2, 303–314, (1989).

[15] Gundogdu, Erhan and Solmaz, Berkan : MARVEL: A Large-Scale Image Dataset for Maritime Vessels. Computer Vision – ACCV 2016. 4, 165–180, (2016).

[16] Zhang, Mabel M and Choi, Jean and Daniilidis, Kostas and Wolf, Michael T and Kanan, Christopher. : Vais: A dataset for recognizing maritime imagery in the visible and infrared spectrums. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2, 10–16 (2015).

[17] Howard, Andrew G and Zhu, Menglong and Chen, Bo and Kalenichenko, Dmitry and Wang, Weijun and Weyand, Tobias and Andreetto, Marco and Adam, Hartwig : Mobilenets: Efficient convolutional neural networks for mobile vision applications. Mathematics of Control, Signals, and Systems. (2017).

[18] Redmon, Joseph and Divvala, Santosh and Girshick, Ross and Farhadi, Ali : You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition. 2, 779–788, (2016).

[19] Goodfellow, Ian and Bengio, Yoshua and Courville, Aaron and Bengio, Yoshua : Deep learning. MIT press Cambridge(2016).

[20] Hinton, G and Sejnowski, T : A Theoretical Framework for Back-Propagation.

[21] Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E : Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems. 2, 1097–1105, (2012).