Consumer Class Side Scanning Sonar Dataset for Human Detection

T. Aaltonen
Satakunta University of Applied Sciences, Pori, Finland
toni.aaltonen@samk.fi

***Abstract* - One of the leading problems for using modern deep neural networks with sonars is the lack of datasets, even rarer are datasets that are collected with consumer class sonars. This paper introduces a novel side scanning sonar dataset for humans under water. Data is collected with consumer class Garmin 8400 Xsv sonar with GT54UHD-TM transducer. Dataset is collected in shallow coastal water of the Baltic Sea, near Rauma Finland. The dataset contains 331 images of humans, and 364 images with other objects like tires and rocks. Dataset contains cropped images from objects, and full resolution images. Data is collected from two different locations, with different sonar settings. All images are from two rescue divers at the bottom of the sea. This Paper also introduces standard data split for collected dataset for training, validation, and test data for benchmarking different models with dataset, and the data collection system based on ROS.**

Keywords – side scanning sonar; dataset; rosbag; drowing; rescue diving

I. INTRODUCTION

Every year, more than 500,000 people drown in the world, most of them children and young adults [1]. It is of paramount importance for the survival of the victim of an underwater accident that the location is rapid, and that rescue, and possible resuscitation can begin immediately. On the other hand, in cold water, the victim may, under certain conditions, be able to recover more than half an hour after being submerged. Nowadays in Finland, the search continues for at least an hour after receiving an alarm.

In underwater rescue today, finding a victim is greatly influenced by the accuracy of the information about where the victim sank below the surface, if the place is well known and there are no significant currents in the place, there are a lot better chances of a successful rescue mission. The larger the area to be searched, the lower the chances of survival. It is especially difficult to find victim in murky waters, as conditions in Finland often are.

The process of underwater rescue in Finland has remained virtually unchanged through the decades. The safety diver is on the beach and handles a safety rope attached to the diver doing the dive. The searching diver starts from the side of the sector near the shore and begins to move along the bottom towards the other side of the sector, feeling the bottom with his hands until the safety diver detects that the other side of the sector has been reached, signaling the rescue diver to stop, and giving an additional half a meter of the rope. by which the diver moves away from the shore and begins to move toward the other side of the sector, while the bottom is felt by hand. This is continued until the victim is found, the sector has been searched, in which case the search area can be increased, or it is stated that the victim has been below the surface for at least an hour and there is virtually no hope of survival, and urgent search can be stopped.

Process has not changed much in recent years at least in Finland, even survey papers over 15 years ago were stating that development of autonomous underwater vehicles and cheaper consumer grade sonars where developed enough to start to deploy them in search and rescue mission [2].

Motivation behind this paper is to find the victim faster, with use relatively cheap consumer grade side scanning sonar with machine learning model to detect victim. These sonars targeted to consumers are priced from less than a thousand euros to couple of thousands of euros. Sonars used in research and in underwater robotics are usually lot more expensive from tens of thousands of euros hundreds of thousands of euros. Expensive sonars are usually more accurate, and in most cases, they are active scanning sonars, meaning that you get the full updated image even when sonar is still in place, opposite to cheaper side scanning sonars that will produce only one line of data, and sonar must move to get a full image.

The problem is that there are almost no public datasets about sonar images, and many studies are done with very limited amount of data, as stated also by other sonar data searchers [3]. For this work search about other sonar datasets was done in: kaggle databases, paperswithcode databases, datasetsearch provided by google.

kaggle datasets have 43 hits using sonar as keyword, most of these are replications of one old dataset that has sonar pings from rocks and mines [4]. Other search results are not related to underwater sonar but have for example twitter data where part of the user's name has sonar and similar nonrelated datasets, actual only sonar dataset is rock and mine dataset.

google over 100 hits with key word sonar, in this search there is also a lot of same rock or mine dataset replications. Most of the datasets found with this search was from National Oceanic and Atmospheric Administration, which part of United states of America government. In this NOAA data there is a lot of raw side scanning sonar survey data from different oceans and lakes, another big part of this data is water column sonar data. Other USA government agencies also contributed with different water and ice column datasets. Other nations also have some other sonar survey datasets in these searches, for example Raw side scanning sonar data from Submarine structures in the mid-Irish Sea Area of Search, done by Irish government.

One dataset was a website hosted by North Carolinas government agency NCDMF that has map with sonar scanned artificial reefs, where you can zoom in the map, and save the images of the reefs that you are interested. The entirety of this side scan sonar image dataset was obtained using towfish that were built by Chesapeake Technology and operated at frequencies of 1200 kHz and 600 kHz. The subsequent data analysis was carried out by NCDMF Staff using SonarWeb and SonarWiz software. NCDMF staff members were responsible for gathering all field data from 2000 to 2015 [5].

There is also a dataset of sonar images of shoals of fish, this dataset contains camera and sonar images of fishes in an aquarium. In total dataset contains 1334 sonar images, sonar used to record images is not specified [6].

In papers with code website there are only two datasets with sonar keyword in them, database contains over 7900 datasets at the moment of the writing this paper. There are datasets about common garbage in the ocean [7], [8]. Another one is datasets about fishes, the CFC dataset, which comprises more than 1,500 videos from seven diverse sonar sources, With over 500,000 annotations of different fish species [9].

Outside of these dataset search engines, some datasets were found by reading papers related to AI methods in sonar data, and following references from those papers. An underwater observation dataset for fish classification and fishery assessment, 524 datapoints taken with DIDSON sonar, published 2013[10]. The ARACATI 2017 dataset contains both optical aerial and acoustic underwater images, enabling researchers to compare the two perspectives. By utilizing these diverse image types, dataset allows researchers to examine how differences between image modalities can impact the accuracy of underwater localization [11].

Recently there has also been released a big dataset consisting of over 9000 Multibeam Forward-Looking Sonar images captured using Tritech Gemini 1200ik sonar. Dataset provides raw data of sonar images with annotation of 10 categories of target objects and contains images with human like doll [12]. This dataset is the closest that was found to ours, but it uses human like doll and not a real human, and it is captured using expensive sonar.

Most of the datasets containing sonar images are collected with expensive active scanning imaging sonars, and it is still common to not express the exact sonar model used in the metadata of the dataset. In our best efforts we didn’t find any public dataset that contains consumer class sonars with human images. Sonars that are used need to be cheap so that they can be deployed to small rescue departments, either used in boats or as part of an cheap underwater robot.

II. CATHERING THE DATA

The data was collected on a boat of the Satakunta rescue service, to which a Garmin 8400 Xsv sonar, and a GT54UHD-TM echo sounder were attached, as we can see in figure 1. The depth of the sensor could be adjusted by adjusting the height of the sensor mounting post, so data of different depths could be collected, even from the same places. When the boom was close to the maximum depth, it caused the sensor to wobble due to the water resistance, which could slightly reduce the accuracy of the data. However, this is likely to simulate the much faster movements of a much smaller underwater robot as it tries to maintain constant bearing and speed. The sonar image and other sensor data were collected in ROSBAG format on a laptop computer that acted as a ROS master. About 100 man-hours were used to collect the dataset, of which rescue divers accounted for about 30 hours.

Figure 1 Data collection devices installed into the boat of the rescue department. Power to the laptop and to the sonar is taken from the small gasoline generator. The system is designed to be easy to fit to most of the small boats without any holes or other permanent marks [6].

The data collection device shown in figure 2 had Garmin 8400 XSV. Main reason why this sonar plotter was selected was one of the few ones that had a HDMI output. So, it can be connected to laptop via HDMI to USB 3.0 capture card. 8400 XSV also has NMEA2000 data bus, that is also connected to same data stream with teensy 3.6 board. Teensy board was acting as ROS node with ROS-serial node connected with USB 2.0 to laptop. Data collected from NMEA2000 bus contained GPS position, Speed over ground, heading and depth. The device also had BNO055 IMU board connected with serial to USB 2.0 connection to laptop. Speed over water propeller sensor was used, but data was incorrect to point to be unusable.

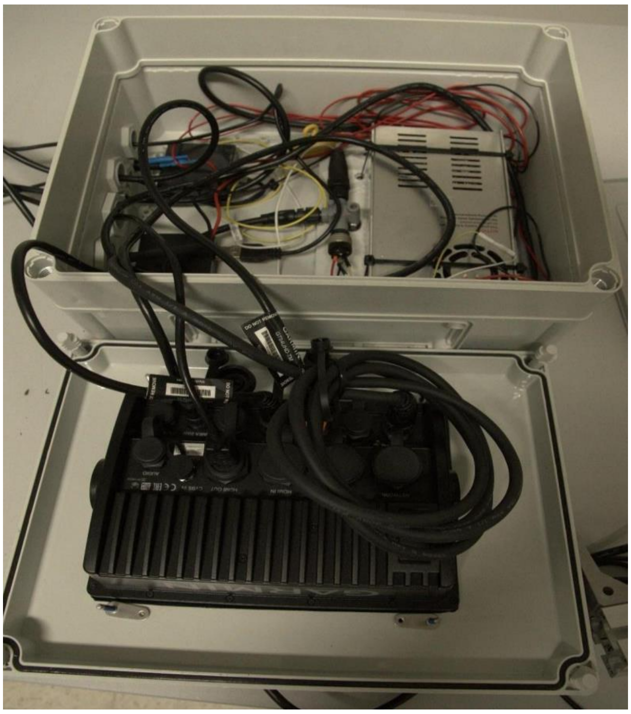

Figure 2 Data collection box, sonar plotter with touch screen is fitted to cover of the box. Meanwell power supply is fitted to the bottom of the box. IMU box, and NMEA2000 to ROS-serial microcontroller is clued to bottom in the left [7].

Sensor boom has 3 different sensors connected to it, as we can see from figure 4. First in the left is scanning sounder LVS12, next is GT54UHD-TM side scanning sonar, last sensor is a speed over water sensor. Sensor boom is made from aluminum. In figure 3 there is the connection diagram, where different connections, sensors and devices can be seen, all different dataflows are changed to USB signals before connecting to computer.

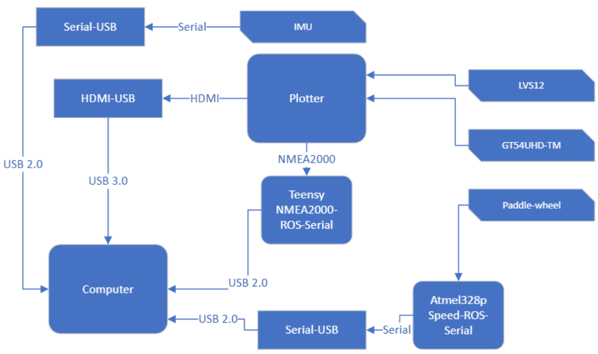

Figure 3 a connection scheme of the data gathering setup. Lines are connecting cables, and boxes are devices.

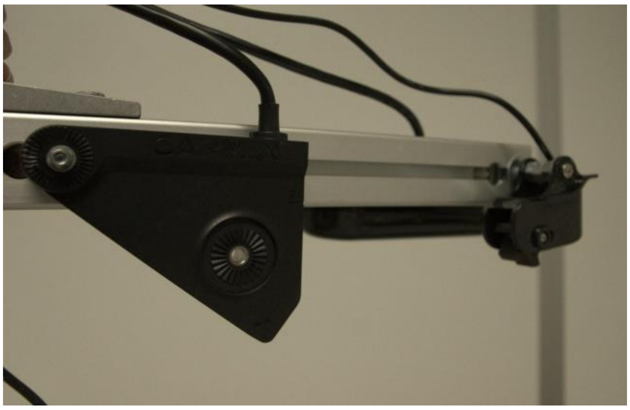

Figure 4 sensors that were used in the data collection. from the left first is real time sensor LVS12, next in the back side is GT54UHD-TM sonar, and the third sensor is speed over water paddle sensor [7].

In all the images, the diver is wearing a wet suit and a face mask. In most of the pictures, the diver's oxygen bottles are so that they are not visible in the picture. Divers are in various positions in different pictures. Data was collected over two days at sea off Rauma. The first day's data is from the Syväraumanlahti area, and the second day's data is from the Maanpäännnokka area. In the data of both days, there are pictures from different points of the bottom in different positions of both divers. Divers have a considerable height difference. In the photos from the first day, both divers have flippers. There are no flippers in the photos from the second day.

The dataset has also left all the worst quality images, which humans have great difficulty in identifying. The Ground Truth information was obtained from these images because it was known that there is a person in it based on the time stamp and location. Otherwise, some would certainly have gone unrecognized. Including such low-quality images in the dataset will certainly worsen the detection accuracy but gives a more realistic picture of the actual detection percentage when using inexpensive consumer grade echo sounders.

During the data preparation phase, the ROSBAG files were manually reviewed and checked, and at the same time stamps were recorded for the points where a person can be seen. After this, a python program was prepared that goes through the ROSBAG files and saves the images from the timestamp points where there is a person. In addition to this, the data was reviewed a second time. In this round, we searched for all those timestamp points where some other objects can be found, which can be used in teaching the neural network for teaching negative cases. Of course, many more such images could have been captured from the data, but we wanted to keep the ratio of positive and negative cases close to one. As we can see from picture figure 5, there are usually more than one object, a pile of rocks or another pattern in the bottom, that can be used. In figure 6 we can see an example of the image where there is a human in the image, some images are easier for humans to recognize than others, this one is between easy and average.

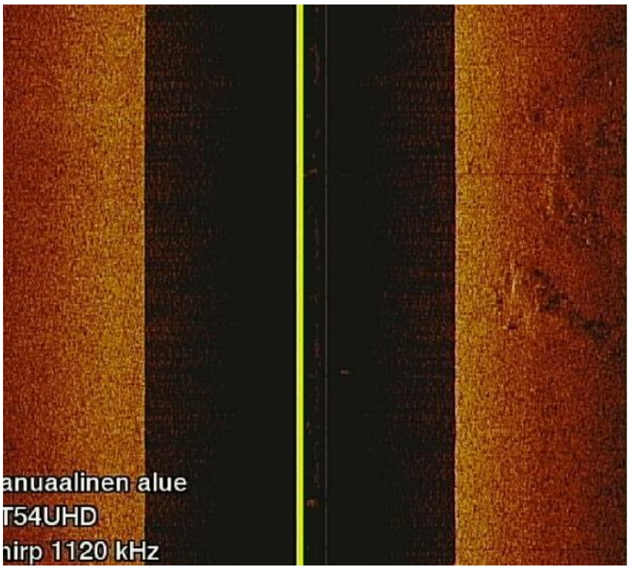

Figure 5 Full size picture from the dataset, boat moves from the bottom to the top, black area is the height of the water, and then finally the sea bead in left and right.

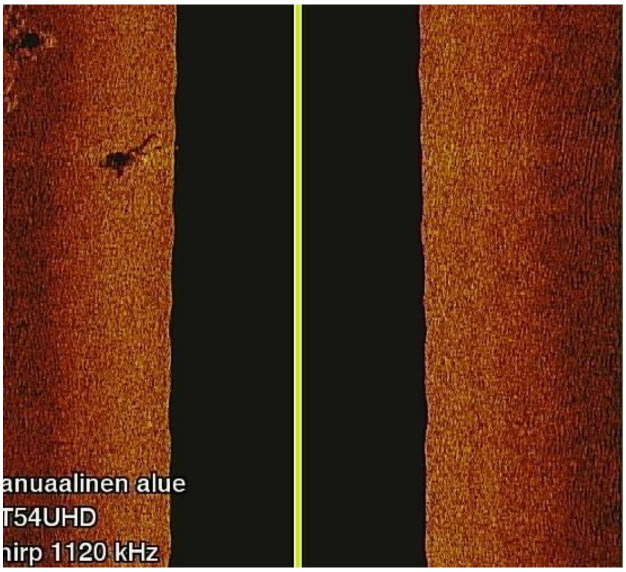

Figure 6 We can see the human in side scanning image, diver is located upper left side of the image, not all images in dataset are as easy find for human eye as this one, legs of the diver are easy to detect in this image [13].

After that, the pictures were in the form shown in figures 5 and 6. After that, the picture of the desired object was still cut from the larger picture. The final images have a resolution of 100x70 pixels and have three color channels. This cropped image size was used, because almost all images fit nicely on that. As we can see from figures 7, many patterns of the bottom can look a little like a human. In figure 8 we can see one example of the final cropped human image.


Figure 7 Cropped image taken from Figure 4, this is example of what kind of features the are in the negative dataset, from pictures where the are some features in the bottom what are not humans.


Figure 8 This image is the cropped version of figure 6. In the image, the diver is lying on his stomach at the bottom of the sea, legs are in upper right part of the picture, body of the diver is casting small shadow behind him in the upper body [13].

The reliability of the collected sonar data is good, and the internal optimal error is zero in light of the best information, but its external reliability is more difficult to define. This refers to how well the data represents all cases that can occur in transom echoes with a person on the waterbed. The only practical way to measure this is to use data from two different days as a separate data set in neural network training, one for training and validation and the other day's data for testing. If the results of the neural network trained with the training data generalize to the test data, it can be said that the dataset is comprehensive enough to generalize at least to a certain extent to the general situation with the sonar used. This can be said because there are many differences in first and second day data, depth is different, bottom of the sea is different, brightness settings are different, and the divers don’t wear flippers in second day, what makes the shape of the foot different. With small dataset like this, division to different days we can test the generalization of the model a lot better, and this is recommended way to use this dataset.

The final training/validation data set collected on the first day consists of 205 images of a person at the bottom of the sea, and 249 images with an object other than a person. Images of, for example, stones, plants, and car tires belong to this category. In addition to this, the test data set made from the data set recorded in the second data collection session contains a total of 125 images of people at the bottom of the ocean and 84 images where there are other objects or plain bottom of the ocean. All in all, there are a total of 695 pictures, of which there are 331 pictures of people, and 364 pictures without people.

Default dataset division is done in the way that would make train and validation dataset as separated from the test data as possible, and the second dataset is assigned to test date. Train and validation are made from the first dataset, by splitting it with random function. Full size images and cropped images of human are all with one-to-one naming, starting from 000000.jpg. Nonhuman images are not always one to one mapped from full size to cropped, because there are some pictures with multiple cropped pictures.

Default division for train and validation datasets, is provided with the dataset. Dataset is ready to use with neural networks without any preprocessing, with categorical detection networks like densenet, renesnets, and object detectors based on that kind of networks. This is main intended use of dataset, because in general that type of detectors needs less data to train, and the nature of the image produced by side scanning sonar is that is updated one horizontal line at the time, so sliding window type of object detection fits well to the problem. Extensive data augmentation is recommended, because of the small dataset size. If used with Yolo type object detection algorithms then use of full-size images are recommended, and annotation is needed, and is not currently provided with the dataset.

III. Conclusion and discussion

There are many well-known limitations in this published dataset. The size of the dataset is still very limited, and it contains only data from the two different people, and all pictures are from adults, also in the images divers have wet suit, and that might give a different echo than human skin or other clothing. Still, it gives a novel dataset to understand what kind of quality can be obtained from relatively cheap side scanning sonars, and there is enough data to do preliminary tests with neural networks.

One important observation was that with consumer class sonars, the depth of the water needs to be shallow, best range seems to be from 1,5 – 3 meters and after 6 meters the human size object starts to be too small to detect for trained human eye. That means that sonar connected directly on the bottom of the boat limits the detection range to quite shallow water. Use of the sensor boom can extend this with some meters. Other options are to use depth controlled towed sonar, or underwater robot.

Dataset can also be used for training for rescue people for use side scanning sonars without any computer vision, with sonars mounted in to the boat. Observation after doing the work with this dataset is that after looking at sonar data with humans, detection comes a lot easier. Typical features start to clearer to see. In shallow murky coastal waters this could be helpful to detect victim faster than current methods.

Data is published in github repository, with following link: https://github.com/tonaalt/sonar_human_dataset

IV. Future studies

Next work is to publish results of the trained neural networks based on this data, using different neural network architectures. This study has been mostly done, but still no peer review papers have been published.

More data is needed to gather, to get more certain results, and it would be better to collect some images without wetsuit. An interesting option that we have been starting to investigate is the creation of synthetic or semi-synthetic data using different AI and traditional methods.

Acknowledgment

This work has been supported by Finnish Ministry of education project AI Forum and Satakunta Rescue Department Unit of Rauma.

References

[1] D. Szpilman, J. L. M. Bierens, A. J. Handley, and J. P. Orlowski, “Current Concepts Drowning,” 2012.

[2] A. J. Murphy, M. J. Landamore, and R. W. Birmingham, “The role of autonomous underwater vehicles for marine search and rescue operations,” *Underwater Technology*, vol. 27, no. 4, pp. 195–205, 2008, doi: 10.3723/ut.27.195.

[3] M. Jegorova, “Generative Neural Data Synthesis for Autonomous Systems,” no. July, 2021, doi: 10.13140/RG.2.2.19876.60801.

[4] R. P. Gorman and T. J. Sejnowski, “Analysis of hidden units in a layered network trained to classify sonar targets,” *Neural Networks*, vol. 1, no. 1, pp. 75–89, Jan. 1988, doi: 10.1016/0893-6080(88)90023-8.

[5] “Artificial Reef Sonar Scans | Artificial Reef Sonar Scans | NC OneMap.”https://data-nconemap.opendata.arcgis.com/datasets/artificial-reef-sonar-scans/explore (accessed Mar. 29, 2023).

[6] T. Kei, K. Shin, M. Katsunori, and T. Koji, “A dataset of camera and sonar images.” 2018.

[7] M. Valdenegro-Toro, “Deep Neural Networks for Marine Debris Detection in Sonar Images,” no. April, 2019, [Online]. Available: http://arxiv.org/abs/1905.05241

[8] M. Valdenegro-Toro, A. Preciado-Grijalva, and B. Wehbe, “Pre-trained Models for Sonar Images,” 2021, [Online]. Available: http://arxiv.org/abs/2108.01111

[9] J. Kay *et al.*, “The Caltech Fish Counting Dataset: A Benchmark for Multiple-Object Tracking and Counting,” Jul. 2022, [Online]. Available: http://arxiv.org/abs/2207.09295

[10] E. McCann, L. Li, K. Pangle, N. Johnson, and J. Eickholt, “Data descriptor: An underwater observation dataset for fish classification and fishery assessment,” *Sci Data*, vol. 5, Oct. 2018, doi: 10.1038/sdata.2018.190.

[11] M. M. D. Santos, G. G. De Giacomo, P. L. J. Drews-Jr, and S. S. C. Botelho, "Cross-View and Cross-Domain Underwater Localization Based on Optical Aerial and Acoustic Underwater Images," *IEEE Robot Autom Lett*, vol. 7, no. 2, pp. 4969–4974, 2022, doi: 10.1109/LRA.2022.3154482.

[12] K. Xie, J. Yang, and K. Qiu, "A Dataset with Multibeam Forward-Looking Sonar for Underwater Object Detection," *Sci Data*, vol. 9, no. 1, p. 739, Dec. 2022, doi: 10.1038/s41597-022-01854-w.

[13] "tonaalt/sonar_human_dataset: Public sonar dataset dataset, two categories humans and other. Gathered with garmin consumer side scanning sonar." https://github.com/tonaalt/sonar_human_dataset (accessed Mar. 30, 2023).