# Semantic Mapping on Underwater Enviroment Using Sonar Data

M. Machado[1], P. Drews-Jr[1], P. Núñez[2] and S. Botelho[1]

***Abstract*— The use of robots in underwater exploration is increasing in the last years. The automation of the monitoring, inspection and underwater maintenance tasks requires the understanding of the environment. One of the key issues of these systems is to recognize in the objects in the scene. This paper proposes a method to provide a semantic mapping using acoustic images acquired by forward looking sonar (FLS). The method represents the environment using Gaussian probability density functions. Furthermore, we efficiently segment and classify the structures in the scene. Finally, we create a semantic map of the scene. We evaluate the method in a real dataset acquired by an underwater vehicle performing autonomous navigation and mapping tasks in a harbor area.**

## I. INTRODUCTION

The problem of building a map while robot is moving is an essential task in autonomous robotics, which has been extensively studied in the literature. A map of the environment allows robots to develop other important skills such as navigating, self-localizing or interacting, among others. How an autonomous robot builds a representation of its surrounding has been analyzed from different points of view in the scientific community in the last years (an interesting survey about mapping is found in [1]). Most of solutions proposed in the literature for this problem are addressed using representations of the spatial structure of the environment ($e.g.$, occupancy cells or geometric features like segment lines). However, using only a spatial representation of the environment is difficult to perform other tasks successfully. This tendency is now changing, and the scientific community is experiencing an increasing interest in so-called semantic solutions, which integrate semantic knowledge and geometrical information [2].

Recently, several advances in semantic mapping have been achieved. In fact, ground robots that incorporate capabilities for task planning and storing some semantic knowledge in their maps are commonly used ($e.g.$, classification of spaces, such as rooms, corridors or garden, and labels of places and/or objects) [2]. However, very few works have been achieved in underwater robotics where the semantic knowledge of the environment could be applied to predict changes and to take high-level decision. In fact, the mapping problem in underwater robots has been addressed typically by using geometric information and using sonar or RGB sensors [3]–[5].

In order to autonomously acquire semantic information from an underwater environment robots have to be equipped with sensors and the ability to extract high-level knowledge from the scene. This paper proposes an object detection and representation system by using a Forward Looking Sonar (FLS). The use of sonar offers the advantage to be invariant to the water turbidity, however, data suffer distortion and noise and thus, processing the acoustic signal still a challenge. Besides, the available data can be summarized to untextured range data and thus, only information about the shape of the detected objects is able to be acquired.

Several works proposed methods to identify objects on acoustic data as [6]–[10]. However, none of them recognize objects and create semantic maps. [6] proposed an underwater object classification on multi-beam sonar data by considering the specific domain knowledge with limited number of shapes.

The present work detects and recognize objects in the scene that allow us to create a semantic map. We evaluated our approach on real data acquired in a harbor area using a forward looking sonar (FLS). The images are segmented and described using an adaptation of the method previously proposed by the authors [11], [12]. Furthermore, the shapes are classified using a machine learning technique. A new tool is developed to annotate sonar data that allow us to train the supervised model. The proposed method is evaluate on the dataset ARACATI 2014 [13]. Figure 1 shows an example of the semantic map obtained using our approach.

## II. ACOUSTIC IMAGE FROM A FORWARD LOOKING SONAR

The forward looking sonars (FLS) are active devices which produce acoustic waves that propagate through the medium until they collide with an obstacle or be completely absorbed. When a wave collides with an obstacle, part of its energy is absorbed and the other part is reflected. The reflected portion which returns to the sensor is recorded using an array of hydrophones. The round trip of the wave is called *ping*.

The waves captured by the hydrophones are organized according to its return direction and their distance to the reflecting object. Acoustic returns from the same direction belong to the same beam. The returns recorded over time in a beam are called bin. A fan-shaped acoustic image $I(X, Y)$

Thanks to the partial supported of CNPq, CAPES, FAPERGS, Oil Brazilian Agency, PRH-27 FURG-ANP/MCT and IBP – Brazillian Petroleum, Gas and Biofuels Institute. This paper is a contribution of the INCT-Mar COI funded by CNPq Grant Number 610012/2011-8. This paper is also a contribution of the CAPES-DGPU project BS-NAVLOC (CAPES no 321/15, DGPU 7523/14-9, MEC project PHBP14/00083): Brazil-Spain cooperation on navigation and localization for autonomous robots on underwater and terrestrial environments.

[1] The authors are with NAUTEC, Intelligent Robotics and Automation Group, Univ. Federal do Rio Grande - FURG, Rio Grande - Brazil machado,paulodrews,silviacb@furg.br

[2]P. Núñez is with the Department of Computer and Communication Technology, Universidad de Extremadura, Cáceres, Extremadura, ES pnuntru@unex.es

978-1-5090-3656-1/16 $31.00 © 2016 IEEE
DOI 10.1109/LARS-SBR.2016.48

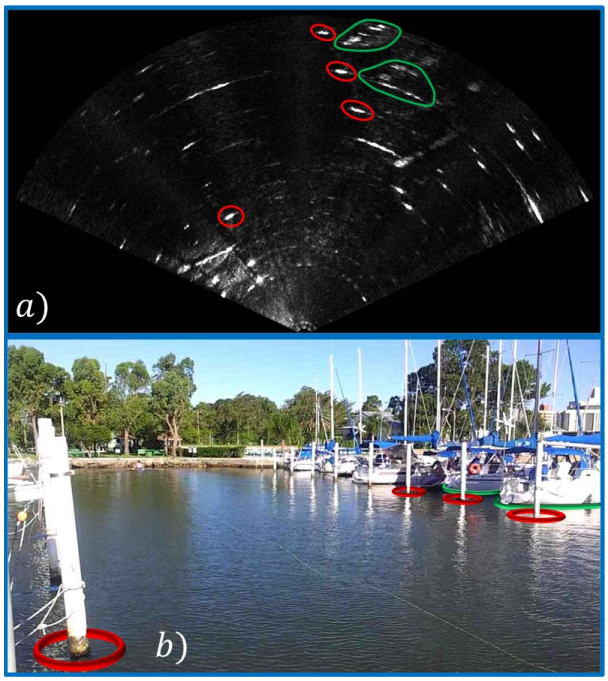



Fig. 1: Semantic map created using our approach. a) the sonar data acquired by a FLS and the segmented areas using colors. b) the water surface image with the associated objects. We show the poles in red and the hulls in green.

is one way to represent the beams and bins information recorded by the sonar for a certain period of time. In Figure 2 is shown how an acoustic image is described in function of their beams and their bins.

Figure 1a shows an example of acoustic image captured at harbor of Yacht Clube de Rio Grande, Brazil. In this image the pixels are associated with bins, and they are indexed according to their distance $r_{bin}$ and their azimuth direction $\theta_{bin}$ from the sonar, as show in Figure 2. Due to the FLS conception, the height information of a bin are not captured and, therefore, the acoustic image is a 2D representation of the observed environment.

Although the sonars are almost independent of the water turbidity conditions, they have some characteristics that make it difficult to handle and to extract information, such as:

- The inhomogeneous resolution. The amount of pixels used to represent a bin varies according to its distance to the sonar. In Figure 2 is shown two bins overlapped by a box. The orange box covers the farther one and the blue box covers the closer one. The covered area by orange box is bigger than blue box. This fact causes image distortion and objects flatness.
- The intensity variations of each bin. It is caused by water attenuation, changes in sonar tilt or sensitivity differences between the hydrophone.
- Acoustic reverberation caused when two or more acoustic returns from the same object are captured producing duplicated objects in the image.

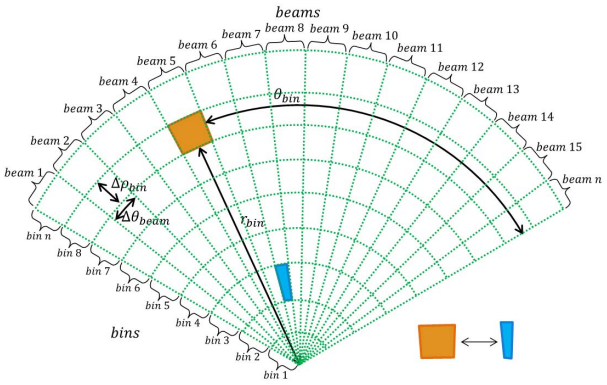



Fig. 2: Acoustic images description where the beams and the bins are depicted.

- The acoustic shadow effect produced by objects that block the path of the acoustic waves, producing a region without acoustic feedback after the blocking objects. These regions are characterized by a black spot in the image and hide a part of the scene causing occlusion of objects.
- The speckle noise due the low signal-to-noise ratio caused by mutual interference of the sampled acoustic returns.

## III. METODOLOGY

The semantic mapping method is divided in steps that include image enhancement, segmentation, description of each segment and classification. A tool has been developed to perform all the steps of the proposed method and to create training data to the supervised classifier.

### A. Image Enhancement

We applied in this step an image correction method. The method is similar to proposed in [14]. First. we found the sonar insonification pattern by averaging a large group of acoustic images. After that, the sonar insonification pattern is applied to each image mitigating the effects of the nonuniform insonification and the overlapping problem of acoustic beams.

### B. Image Segmentation

Due to the low signal-noise ratio, the acoustic image segmentation step is one of the largest challenges faced by our methodology and its quality directly influence the final results. We segment images using an adaptation of the segmentation method proposed in our preview work [12]. The segments are formed by groups of 8-connected pixels extracted using an intensity threshold. A local parameters adjustment are performed by intensity peak analysis of each beam. The intensity threshold depend of the height of the local peak intensity. In Figure 3 is shown an example of this segmentation approach for a single beam, represented by the blue line. The extracted segments are marked by colored pixels.

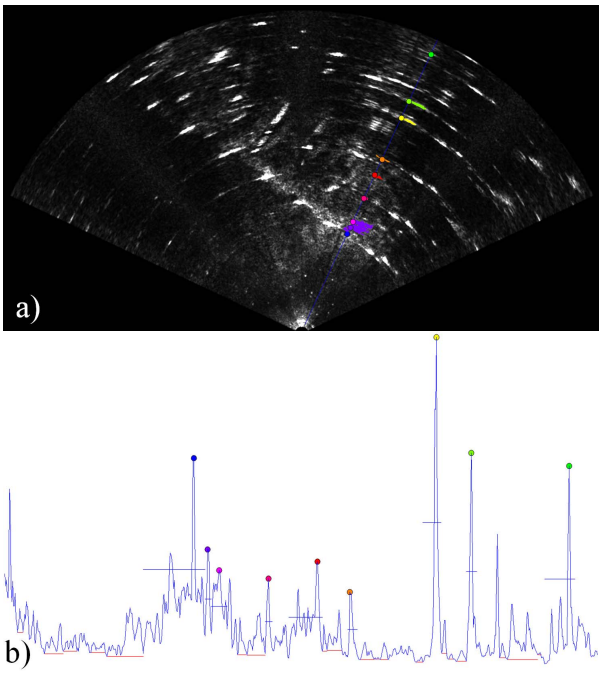



Fig. 3: Example of the segmentation step showing the peak analysis of a single beam. a) the beam intensity profile, where the horizontal axis represents the bins and the vertical axis represents the intensities; b) the acoustic image with the analyzed beam, represented by a blue line, and the extracted segments. The detected peaks are identified by colored circles on both images.

This segmentation approach presents three parameters named as $H_{min}$ , $\pi_{recursive}$ and $\pi_{end}$. $H_{min}$ defines the smallest intensity height of a peak to be used to extract one segment, $\pi_{recursive}$ is the intensity rate of the peak used to compute the intensity threshold for segmentation and $\pi_{end}$ defines the rate to close a peak analysis and start the next peak analyze. In a single beam intensity profile many peaks are detected and only the peaks with greater intensity variations are used to extract one segment, *i.e.* intensity larger than $H_{min}$.

### C. Describing segments

After the segmentation step, each segment is described using a Gaussian probabilistic function and the following information about each segment is computed. Initially, *width* and *length* are computed using a covariance matrix that relates the x and y position of each pixel of the segment. The eigenvalues and eigenvectors of the covariance matrix is computed using Singular Value Decomposition (SVD). The width is defined as the largest eigenvalue and height defined as the second largest eigenvalue. Furthermore, the segments *area* is computed using the Green's theorem that gives the relationship between a line integral around a simple closed curve. This area is computed using the implementation of the OpenCV library [15]. Finally, we determine the *convex hull area* of the segment, the *perimeter* of the segment, the mean and the standard deviation of the *acoustic intensity* of each segment.

These data are important to represent a segment. Almost all data are geometrical information, however the mean and the standard deviation of the intensities represents the acoustic data.

Based on these information, we defined two set of features to be used in the next step. The first **2D features** is only composed by *width* and *length*. In addiction to the *width* and *length*, we defined the **10D features**. They are composed of *Inertia Ratio*, *i.e.* width divided by the height, *mean* and *standard deviation* of the acoustic returns, *segmented area* and *convex hull area*. Furthermore, we compute the *convexity*, *i.e.* the segmented area divided by the convex hull area, the *perimeter* and the *amount of pixels* in the segment.

### D. Segment Classification

In this step, each segment are classified using a supervised classifier. The Support Vector Machine (SVM) technique is adopted. The SVM technique is a classifier that models the data as a k-dimensional vector and defines an optimal hyperplane that best separates the vectors depending on your class. The hyperplane is defined by an optimization algorithm in the training step.

The training data of SVM is generated using the developed tool that allows the manual annotation of each segment label. The Figure 4 shows a screenshot of the tool with some annotated segments. The tool was developed using the OpenCV library [15].

The classification using SVM is based on the libSM library [16]. Its implementation presents several type of kernels that allow us to deal with non-linear classification. The available kernels are: polynomial, radial basis function (RBF) and sigmoidal kernels. Among these kernels, we empirically chose the radial basis function (RBF). As described in [16], the two parameters must to be defined: $\gamma$ and $C$.

These parameters are optimally set through an auto training function that build a grid with the classifier performance by varying the two parameters ($\gamma$, $C$). The performance of the classifier is calculated by cross validation. The training data are divided into $k$ groups, one of them is used for cross-validation and the others train the classifier. The minimum and maximum value of the grid must be defined to use auto training function. In this work was defined a grid starting with 0.1 and ending with 60 for both parameters $\gamma$ and $C$.

## IV. Experimental Results

The experimental results are performed using the acoustic images of a FLS from dataset ARACATI 2014. The training dataset of the SVM classifier was created using the developed tool. Resuls are performed using **2D features** and **10D features** as described in Sec. III-C.

### A. Dataset ARACATI 2014

The dataset ARACATI 2014 provided by [13] was created using a mini Remote Operated Vehicle (ROV) LBV300-5 from Seabotix equipped with a Forward Looking Sonar

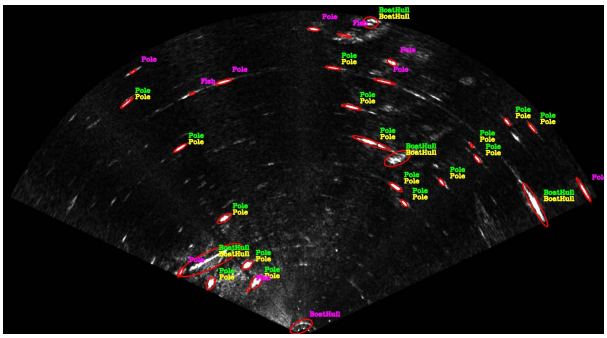


Fig. 4: A sample acoustic image of the training set generated by the developed tool. A demonstration video is available [17].

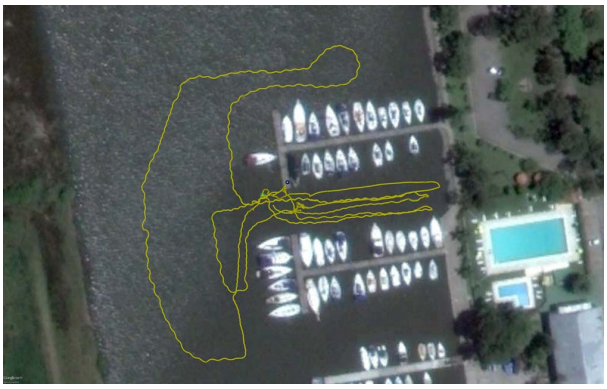


Fig. 5: Satellite image of the harbor with the trajectory traveled by the ROV during the acquisition of the Dataset ARACATI 2014 [13].

BueView P900-130 (900kHz) and a Differential Global Positioning System (DGPS). Throughout entire path the ROV remained closer to the water surface because the use of DGPS. The harbor structures such as poles, piers and boat hulls and also stones are visible in the acoustic images. Some of these objects are highlighted in Figure 1. The Figure 5 shows a satellite image of the harbor with the trajectory traveled by the ROV.

### B. The SVM training dataset

The SVM training dataset generated with the developed tool consists a total of 548 segments over 148 acoustic images which were manually classifieds in one of the five different classes: Pole, Boat Hull, Stone, Fish and Swimmer. In Table I is shown the total amount of segments in each class. Table II is shown the parameters used by the segmentation algorithm.

TABLE I: Dataset information

| *Classname* | *Amount* |
|---|---|
| *Pole* | 282 |
| *Boat Hull* | 86 |
| *Stone* | 122 |
| *Fish* | 46 |
| *Swimmer* | 12 |
| *Total* | 548 |

TABLE II: Segmentation parameters

| *Parameter* | *Value* |
|---|---|
| $H_{seach}$ | 250 |
| $\pi_{end}$ | 0.1 |
| $\pi_{recursive}$ | 0.94 |
| $searchdistance$ | 10 |
| $minsegsize$ | $10 pixels$ |
| $maxsegsize$ | $1200 pixels$ |

### C. Results using 2D features

Firstly, we performed experiments using 2D features. All the 548 segments were used for training the SVM classifier using cross-validation and auto training function that estimate the better parameters $\gamma$ and $C$ using k-subsets of the training data.

Six tests were performed. Three of them consider all the five classes. The other three tests exclude the class stone. The class stone is critical because represents the segments with large size of width and height. This class hinders the training of other classes whose predominance of small segments is greater. The tests were carried out varying amount of sub groups created by the auto training function, the parameter $k$ . The results can be seen in Table III.

TABLE III: Results using 2D features

| *Five class results* | | | | |
|---|---|---|---|---|
| *Parameters* | | | *Result* | |
| $\gamma$ | $C$ | $k$ | $Hit(\%)$ | $Figure$ |
| 14.204 | 10.671 | 2 | 85.94 | $6d$ |
| 3.740 | 18.905 | 5 | 85.58 | $6e$ |
| 3.400 | 20.796 | 10 | 83.02 | $6f$ |
| *Result without stone class* | | | | |
| *Parameters* | | | *Result* | |
| $\gamma$ | $C$ | $k$ | $Hit(\%)$ | $Figure$ |
| 49.030 | 1.311 | 2 | 88.26 | $6a$ |
| 3.740 | 18.905 | 5 | 85.58 | $6b$ |
| 57.665 | 0.500 | 10 | 88.02 | $6c$ |

For the case of 2D features, interesting images can be generated to show the classifier hyperspace and its hyperplanes that separate each class. These images are shown in Figure 6. In this image each circle represents a segment used on the training step. Each class is represented by a color such as fish is yellow, pole is green, boat hull is red, swimmer is blue and stone is cyan. The origin of the classifier space is in the upper left corner. The horizontal axis grows to right and represents the segment width, and the horizontal axis grows to down and represents the segments height.

The classifier space must be normalized before training to achieve a good results. This normalization reduces the scale problem and makes all the dimensions have the same importance to the classifier. The maximum and minimum values adopted in the normalization are obtained using the analysis of the extracted segments (Table IV).

TABLE IV: Maximum and minimum values to normalize the dataset.

| $DimensionName$ | $WithStones$ | | $WithoutStones$ | |
|---|---|---|---|---|
| | $Min$ | $Max$ | $Min$ | $Max$ |
| $Width$ | 1.49 | 94.42 | 1.49 | 51.13 |
| $Height$ | 3.45 | 394.92 | 3.45 | 186.92 |

Using only the width and height of the segments, we correctly classified 85.94 % of the segments. One of the difficulties of this approach is that there are segments with similar width and height and does not belong to the same class. For this reason, we consider more information about the segment to achieve better results.

### *D. Results using 10D features*

We performed experimental evaluation using 10D features to achieve better results. As previously described, we normalized every dimension using the minimum and maximum values shown in Table V.

TABLE V: Maximum and minimum values to normalize the dataset.

| $DimensionName$ | $Min$ | $Max$ |
|---|---|---|
| $Width$ | 1.49 | 94.42 |
| $Height$ | 3.45 | 394.92 |
| $Inertia\ Ratio$ | 0.072 | 0.77 |
| $Std.\ Intensity$ | 67.6105 | 2538.62 |
| $Mean\ Intensity$ | 264.727 | 1777.71 |
| $Area$ | 0 | 43805.5 |
| $Hull\ Area$ | 5 | 20032 |
| $Convexity$ | 0 | 3.00408 |
| $Perimeter$ | 15.3417 | 21463.3 |
| $Pixel\ count$ | 10 | 1200 |

TABLE VI: Ten dimensional experiment results

| $Parameters$ | | | $Result$ |
|---|---|---|---|
| $\gamma$ | $C$ | $k$ | $Hit(\%)$ |
| 1.919 | 44.579 | 2 | 85.58 |
| 1.442 | 22.876 | 5 | 87.95 |
| 8.819 | 8.819 | 10 | 93.97 |

We obtained the results shown in Table VI using 10D features. These results surpass those obtained using 2D features. The classifier is able to correctly classify up to $93\%$ of the segments using the training data divided into 10 subgroups. Another test was conducted without using the auto training function. In this test were manually defined the parameters $\gamma = 28.45$ and $C = 41.55$. The classifier could correctly classify $99.81\%$ of the segments. Despite the possible overfitting, this result shows that the classifier is able to distinguish the five classes of the problem.

## V. Conclusion

We presented a method to detect and classify objects using a FLS images. The method allows us to build semantic maps of underwater environments. The method uses SVM classifier to detected objects by using its geometrical and acoustic information.

A tool has been developed to create the training data and perform the objects classification. It was shown that it is possible to identify and classify objects automatically in a real environment at a harbor using a forward looking sonar images.

The semantic map can be adopted to assist in mapping and localization of an autonomous robot. For example the information of static objects such as classes pole and stones, and dynamic objects such as swimmer, boat hull and fish can be used to build a more accurate environment map for autonomous navigation.

Future works will be focused in integrate the proposed approach in SLAM method. Furthermore, the development of autonomous navigation using semantic information is also on going.

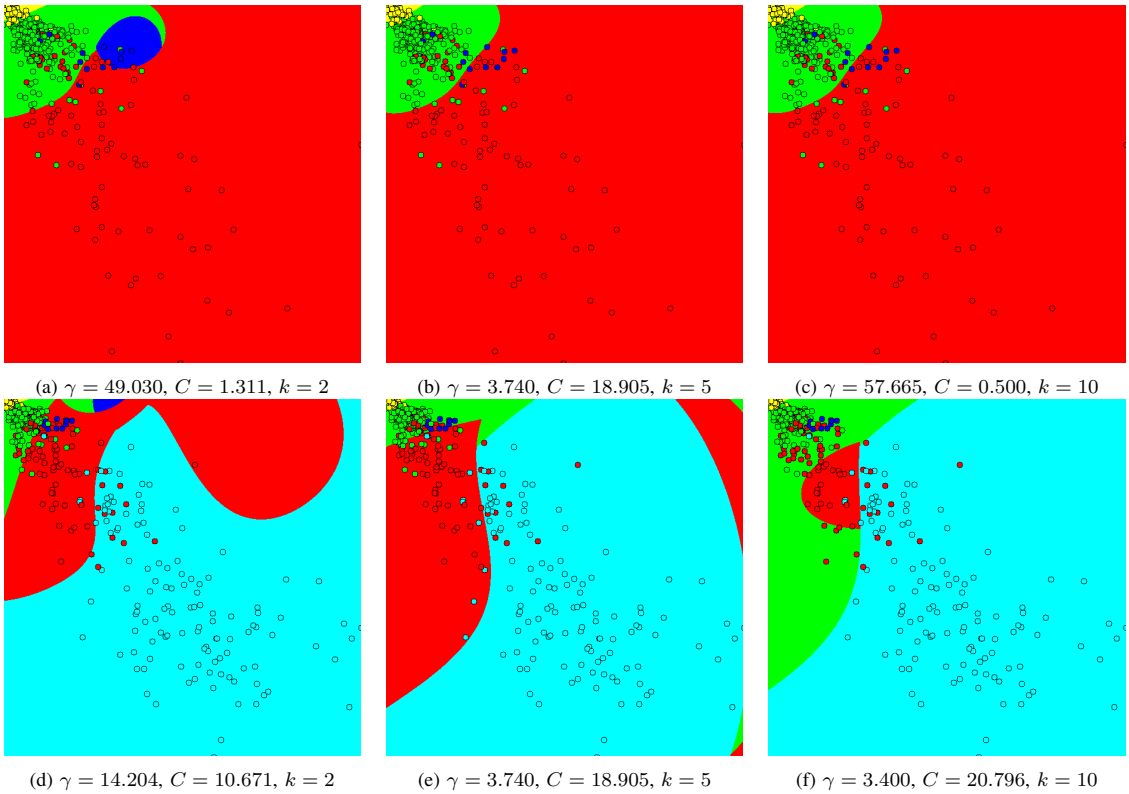
## References


[1] S. Thrun, "Exploring artificial intelligence in the new millennium," G. Lakemeyer and B. Nebel, Eds., 2003, ch. Robotic Mapping: A Survey, pp. 1–35.

[2] I. Kostavelis and A. Gasteratos, "Semantic mapping for mobile robotics tasks: A survey," *Robotics and Autonomous Systems*, vol. 66, pp. 86 – 103, 2015.

[3] D. Ribas, P. Ridao, J. D. Tards, and J. Neira, "Underwater slam in man-made structured environments," *Journal of Field Robotics*, vol. 25, no. 11-12, pp. 898–921, 2008. [Online]. Available: http://dx.doi.org/10.1002/rob.20249

[4] F. Guth, L. Silveira, S. S. Botelho, P. Drews-Jr, and P. Ballester, "Underwater slam: Challenges, state of the art, algorithms and a new biologically-inspired approach," in *IEEE 5th RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*, 2014, pp. 1–6.



(a) $\gamma = 49.030$, $C = 1.311$, $k = 2$ (b) $\gamma = 3.740$, $C = 18.905$, $k = 5$ (c) $\gamma = 57.665$, $C = 0.500$, $k = 10$

(d) $\gamma = 14.204$, $C = 10.671$, $k = 2$ (e) $\gamma = 3.740$, $C = 18.905$, $k = 5$ (f) $\gamma = 3.400$, $C = 20.796$, $k = 10$

Fig. 6: Images generated to show the classifier hyperspace and its hyperplanes that separate each class for each one of the six tests performed using 2D features. Each class is represented by a color such as fish is yellow, pole is green, boat hull is red, swimmer is blue and stone is cyan. The background colors represent the hyperplanes and each circle represents one object of the ground truth data set.


[5] S. Botelho, P. Drews-Jr, M. S. Figueiredo, C. Rocha, and G. L. Oliveira, "Appearance-based odometry and mapping with feature descriptors for underwater robots," *Journal of the Brazilian Computer Society*, vol. 15, pp. 47 – 54, 09 2009.

[6] E. Galceran, V. Djapic, M. Carreras, and D. P. Williams, "A real-time underwater object detection algorithm for multi-beam forward looking sonar," *Navigation, Guidance and Control of Underwater Vehicles (NGCUV)*, vol. 3, pp. 306–311, 2012.

[7] H. Cho, J. Pyo, J. Gu, H. Jeo, and S. C. Yu, "Experimental results of rapid underwater object search based on forward-looking imaging sonar," in *Underwater Technology (UT), 2015 IEEE*, Feb 2015, pp. 1–5.

[8] S. Reed, Y. Petillot, and J. Bell, "An automatic approach to the detection and extraction of mine features in sidescan sonar," *Oceanic Engineering, IEEE Journal of*, vol. 28, no. 1, pp. 90–105, 2003.

[9] J. Guo, S.-W. Cheng, and T.-C. Liu, "Auv obstacle avoidance and navigation using image sequences of a sector scanning sonar," in *Underwater Technology, 1998. Proceedings of the 1998 International Symposium on*, Apr 1998, pp. 223–227.

[10] Y. Lu and E. Sang, "Underwater target's size/shape dynamic analysis for fast target recognition using sonar images," in *Underwater Technology, 1998. Proceedings of the 1998 International Symposium on*, Apr 1998, pp. 172–175.

[11] M. Machado, P. Ballester, G. Zaffari, P. Drews-Jr, and S. S. C. Botelho, "A topological descriptor of acoustic images for navigation and mapping," in *IEEE 12th Latin American Robotics Symposium LARS*, 2015, pp. 1–6.

[12] M. Machado, G. B. Zaffari, A. C. Duarte, D. A. Fernandes, P. Drews-Jr, and S. Botelho, "A modified topological descriptor for forward looking sonar images," in *MTS/IEEE Oceans - Shanghai*, 2016, accepted for publication.

[13] L. Silveira, F. Guth, P. Drews, P. Ballester, M. Machado, F. Codevilla, N. Duarte, and S. Botelho, "An open-source bio-inspired solution to underwater SLAM," in *IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles NGCUV*, april 2015.

[14] K. Kim, N. Neretti, and N. Intrator, "Mosaicing of acoustic camera images," *Radar, Sonar and Navigation, IEE Proceedings -*, vol. 152, no. 4, pp. 263–270, Aug 2005.

[15] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[16] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[17] "Sonar image classification," https://youtu.be/G6c1pBVKI1E, accessed: 2016-05-24.