J Intell Robot Syst
https://doi.org/10.1007/s10846-017-0721-4

Object Recognition and Semantic Mapping for Underwater Vehicles Using Sonar Data

Matheus dos Santos[1] · **Paulo Drews Jr.**[1] · **Pedro Núñez**[2] · **Silvia Botelho**[1]

Received: 19 December 2016 / Accepted: 29 September 2017

Abstract The application of robots as a tool to explore underwater environments has increased in the last decade. Underwater tasks such as inspection, maintenance, and monitoring can be automatized by robots. The understanding of the underwater environments and the object recognition are required features that are becoming a critical issue for these systems. On this work, a method to provide a semantic mapping on the underwater environment is provided. This novel system is independent of the water turbidity and uses acoustic images acquired by Forward-Looking Sonar (FLS). The proposed method efficiently segments and classifies the structures in the scene using geometric information of the recognized objects. Therefore, a semantic map of the scene is created, which allows the robot to describe its environment according to high-level semantic features. Finally, the proposal is evaluated in a real dataset acquired by an underwater vehicle in a marina area. Experimental results demonstrate the robustness and accuracy of the method described in this paper.

Keywords Robot vision · Underwater robot · Semantic mapping · Object recognition · Forward looking sonar

✉ Matheus dos Santos
matheusmachado@furg.br

Paulo Drews Jr.
paulodrews@furg.br

Pedro Núñez
pnuntru@unex.es

Silvia Botelho
silviacb@furg.br

[1] NAUTEC, Intelligent Robotics and Automation Group - Center of Computational Science, Univ. Federal do Rio Grande - FURG, Rio Grande, Brazil

[2] ROBOLAB, Robotics Laboratory - Computer and Communication Technology Department, Universidad de Extremadura, Cáceres, Spain

1 Introduction

The ability to construct a map while the robot moves is essential for performing autonomous tasks and has been extensively studied in the literature. Map building allows the robot to develop autonomous skills such as navigation, interaction with environment and self-localization, among others. The scientific community has been studying new ways of representing the map of the environment in the last decades (an interesting review about mapping can be found in [18]). Most of the approaches proposed in the literature to solve this problem explore the spatial information of the environment (e.g., geometric features like segment lines or occupancy cells). However, only with the spatial representation of the environment is difficult to perform other tasks successfully. Now, this tendency is changing and the scientific community is experiencing an increasing interest in so-called semantic solutions, which integrate geometrical information and semantic knowledge [10].

Recently, several advances were made in the semantic mapping. Generally, ground robots that are able to perform tasks planning usually combines semantic knowledge in their maps (e.g., places classification, such as rooms, passageways or garden, and labels of objects) [10]. However, there are very few works in underwater robotics that consider the semantic map to predict changes in the environment and make high-level decisions. In fact, the problem of underwater mapping has typically been treated with

Published online: 23 October 2017

geometric information extracted from acoustic or optical sensors like sonar and RGB cameras [1, 7, 15].

In order to semantically describe and recognize an underwater environment, a robot needs a system able to extract high-level knowledge from the scene. Typically, the RGB sensors have been used in the literature to extract and characterize the robot's environment. However, in underwater scenarios, these RGB images provide little information due to water turbidity.

The sonar offers the benefit to be invariant to the water turbidity, however, its images are noisy and have distortion problems that make the processing a challenge. The data captured by a sonar can be summarized in a picture with an untextured set of ranges whose the most notable characteristic is the shape of the objects.

Some works propose strategies to identify objects on acoustic images as [4–6, 11, 14]. However, none of them recognize objects and create semantic maps in these scenarios.

In the work presented in this paper, a method for semantic mapping is provided. The proposal is able to detect and recognize objects in the scene allowing the robot to build a semantic map. The acoustic images are segmented, and the shape of each cluster is described geometrically. Each shape is then classified into six different classes (Pole, Boat, Hull, Stone, Fish and Swimmer) using the well-known Support Vector Machine (SVM) algorithm. Besides, a tool was developed to annotate the sonar data, allowing the training during the supervised model.

This approach was developed to integrate with the topological graph proposed in a previous work [12], making it possible to construct more reliable maps for the localization problem. Since it would be possible to establish a reliability relation for each object detected based on its behavior in the environment. For example, static objects such as stones and poles have more confidence than dynamic objects such as fish, boats, and swimmers for the localization problem.

This work also extends our previous contributions [16], bringing a new statistical results and a new segmentation stage. A local adjustment of the segmentation parameters is performed automatically based on the average intensity of the acoustic bins. Besides, this paper describes with details the experiments that validate the proposal: new results were generated evaluating the solution on real data acquired by FLS in a marina. The Fig. 1 demonstrates the kind of information that can be obtained by the approach

2 Acoustic Image from a Forward Looking Sonar

The Forward Looking sonars (FLS) are active devices that create acoustic waves. The waves spread through the underwater environment in the forwarding direction until striking an object or being completely assimilated by the medium.

According to the object composition, a portion of the waves that striked the object are reflected back to the sonar. The reflected waves that achieve the sonar are recorded by an array of hydrophones. The signal is processed and discretized in intensity values called bins. The bins are indexed in an image according to its return direction θ_{bin} and traveled distance r_{bin} as show in Fig. 2. An acoustic image acquired in a marina area of the Yacht Clube of Rio Grande, Brazil, is shown in Fig. 1b.

Although the sonars have the benefit of being independent of turbidity, their data have some characteristics that make it difficult to process and extract information. These characteristics can be summarized in:

- Non-homogeneous resolution: The bin resolution in a number of pixels changes according to its range r_{bin} to the sonar. An illustration is shown in Fig. 2, where two bins are overlapped by a box. The orange box is farther than the blue box, then, the orange box cover a bigger area. Hence, the resolution of acoustics images decreases according to the bin distance r_{bin}. This fact causes image distortion and objects stretching making their recognition harder.
- Non-uniform intensity: It is not guaranteed that an object will always be represented with the same pixel intensities on the acoustic images. Because of the signal attenuation caused by the water, distant objects tend to have a lower intensity than near objects. Typically this problem is mitigated with a mechanism that compensates the signal loss according to the traveled distance. However, the intensity variations can also be

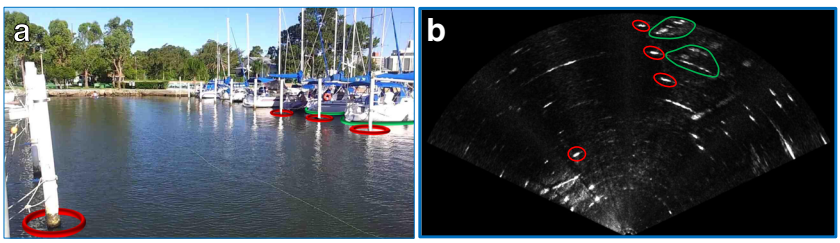

Fig. 1 An example of the semantic map created from an acoustic image collected in a marina. Both images have a visual intersection. In **a** an RGB image captured on the surface (underwater RGB cameras can not obtain data because of the high-level turbidity conditions). In **b** an underwater image captured by a Forward Looking Sonar. The highlighted objects in red are poles and in green are boat hulls

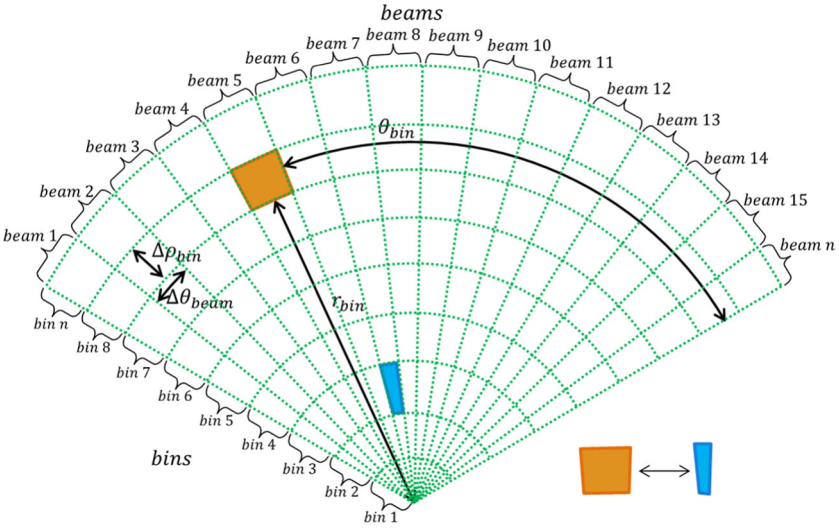

Fig. 2 A representative scheme of image formation of an FLS. Each bin can be identified on the polar coordinate system (θ_{bin}, r_{bin}) and has an angular resolution $\Delta\theta_{beam}$ and a range resolution $\Delta\rho_{bin}$. For this reason, the most distant bins have a lower resolution than the nearest bins. This effect can be visualized on the blue and orange highlight polygons

caused by changing the sonar tilt angle or by sensitivity differences between its transducers.

- The speckle noise: The FLS has a low signal-to-noise ratio and the speckle noise in the acoustic image are caused by mutual interference of the sampled acoustic returns.
- Acoustic shadow: The Acoustic shadow is caused by objects that block the passage of acoustic waves generating a region of occlusion in the image. Because it is an active device, the sonar displacement moves the acoustic shadows and the occlusion areas significantly changing the scene.
- Acoustic reverberation and multipath problem: A transmitted wave may travel through indirect paths due to secondary reflections. Depending on the environment it can generate different effects that include the creation of "ghost" objects and thus change the quality and interpretation of the acoustic image.
- The FLS construction concept: Because of the construction concept of an FLS, it is not possible to determine the vertical direction of an acoustic return. Therefore the acoustic images of an FLS are 2D horizontal projections of the observed environment. This fact generates an ambiguity problem because equidistant objects at different heights are mapped to the same position in the acoustic image.

Because of these problems, techniques for enhancing, segmenting and describing of acoustic images, specifically developed for FLS, are required.

3 Metodology

The proposed method has four steps which include image enhancement, segmentation, segment description and classification as indicated in the illustration of Fig. 3.

3.1 Image Enhancement

In this step, an image correction process is applied on the image to mitigate the non-uniform intensity problem. First, the sonar insonification pattern is computed by averaging a significant number of images captured by the same sonar. The averaged image shows the regions where the pixels have almost the same intensity values in all images. These regions represent constant problems associated with sensitivity difference between the sonar transducers, the overlapping of acoustic beams and the loss of signal intensity. The insonification pattern is applied in each acoustic image in order to normalized theses constant problems. This approach is similar to the proposed in [9] and [8].

The Fig. 4c shows the insonification pattern found by averaging 3675 acoustic images. The same insonification pattern is applied to all images regardless of the FLS position. Figure 4a shows an acoustic image without correction, Fig. 4b shows the image after correction. Although the method reduces non-uniform insonification problems by removing constant effects on the image, this normalization involves low-intensity values and affects only the background of the images.

The FLS BlueView P900-130 generates 16-bit images, i.e. the pixels intensities cover a range of 0-65535. The nonuniform insonification problems usually occur in an intensity range of 0-335. The high-intensity regions, which represent the objects present in the scene, usually are not affected by this method.

Since the grayscale acoustic images are truncated in 8-bits (all pixels are saturated in 255) for visualization propose, it is not possible to see the intensities differences between the light and dark regions of the image. For this reason, the Fig. 4d, e and f show a surface plot of the image intensities. It is possible to visualize the acoustic intensity peaks, these peaks are not affected by the correction method.

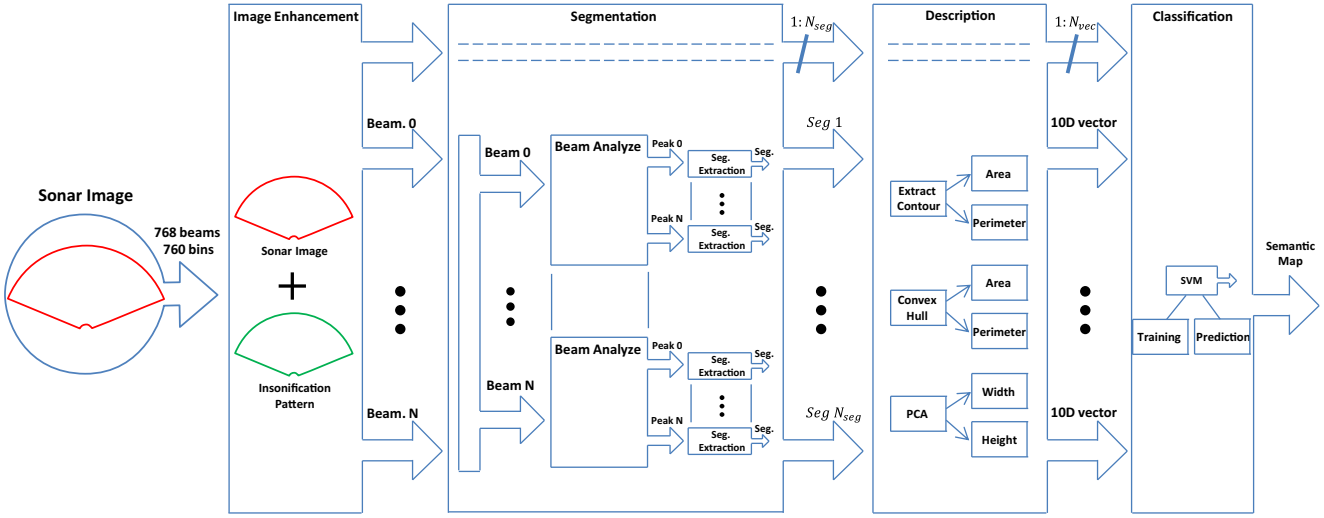

Fig. 3 The proposed pipeline of the semantic mapping system. First, the images are enhanced applying a normalization with the sonar insonification pattern. On the second stage, the images are segmented by an intensity peak analyze approach. The third stage describes each segment extracting geometric and pixel intensities information. On the last stage, the supervised classifier Support Vector Machine is trained to recognize the segment descriptions. At the end, the semantic information of each segment is outputted

This method provides a significant improvement for the image visualization and mosaic construction problem as in [8, 9, 13] and for the image segmentation problem.

3.2 Image Segmentation

Because of low signal to noise ratio and the phenomena described in Section 2, the acoustic images are very noisy and represent a significant challenge faced by our methodology and its quality directly influence the final results.

The main idea of this segmentation approach is to separate the objects of interest from the background (seabed). As objects are more efficient than the seabed to reflect acoustic waves. They are characterized by high-intensity spots on the images. For this reason, we adopted an approach based on the acoustic image formation to detect peaks of intensity. Each acoustic beam B is analyzed individually, bin by bin.

The average intensity $I_{mean}(b, B)$ is calculated for each bin b of a given beam B through Eq. 1.

$$I_{mean}(b, B) = \frac{1}{win_{sz}} \sum_{i=b-win_{sz}}^{b} I(i, B), \tag{1}$$

where win_{sz} is the window size, in number of bins, admitted in the averaging; b and i are bin identifiers; B is a beam identifier; $I(i, B)$ is the intensity of i^{th}-bin of B^{th}-beam. The intensity $I_{peak}(b, B)$ is an offset of $I_{mean}(b, B)$ as shown in Eq. 2.

$$I_{peak}(b, B) = I_{mean}(b, B) + h_{peak}. \tag{2}$$

Where h_{peak} is a constant that determines the minimum height of a peak of intensity. A sequence of bins with an intensity $I(b, B)$ greater than $I_{peak}(b, B)$ are considered part of a peak and are not considered on the $I_{mean}(b, B)$ computation.

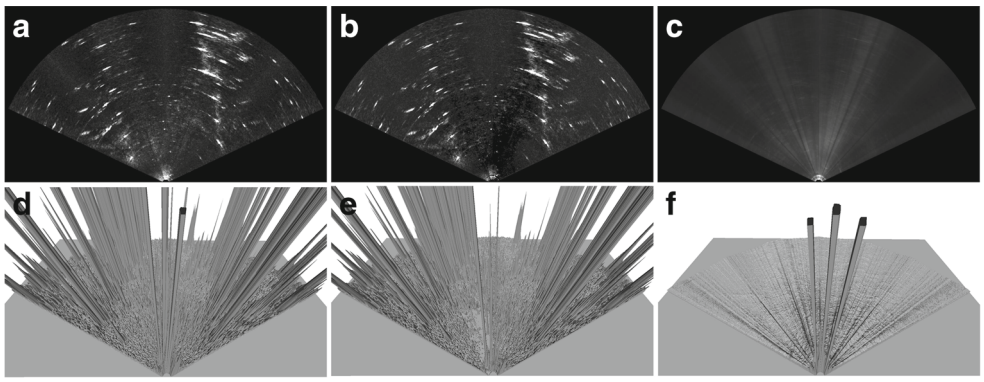

Fig. 4 Application of image correction method. In **a** the image before correction, in **b** the corrected image, in **c** the insonification pattern obtained by averaging images. The images **d**, **e** and **f** show a 3D surface plot of the respective (**a**), (**b**) and (**c**) images

In this sequence, the bin b_{peak} with the greater intensity $I(b_{peak}, B)$ is adopted to adjust the segmentation parameters.

Figure 5 shows in red values of $I_{mean}(b, B)$, in blue values of $I(b, B)$ and in green values of $I_{peak}(b, B)$ of all bins of a single beam B. The peaks detected b_{peak} are represented by colored circles.

The detected b_{peak} peaks are defined by quadruple $\{x, y, I(b_{peak}, B), I_{mean}(b_{peak}, B)\}$, where x, y is the bin b_{peak} position in the image. After the detection of all peaks, a search for connected pixels is performed for each peak, starting at the peak of lower intensity $I(b_{peak}, B)$ until the highest intensity peak.

The 8-way connection is adopted as the neighborhood criterion by the breadth-first search algorithm. In this search, all the connected pixels are visited if $I(i, j) > I_{mean}(b_{peak}, B)$ or its relative distance to the segment border is lower than the parameter D_{seg} in pixels, where $I(i, j)$ is the pixel intensity.

The distance criterion is adopted to reduce the multi-segmentation issue of a single object caused when a group of high-intensity pixels is divided by low-intensity pixels. This effect is generated by noise or by acoustic shadows. Figure 6 shows the behavior of the segmentation algorithm by changing the D_{seg} parameter.

3.3 Describing Segments

After the segmentation step, each segment is described using a Gaussian probabilistic function and the following information about each segment is computed.

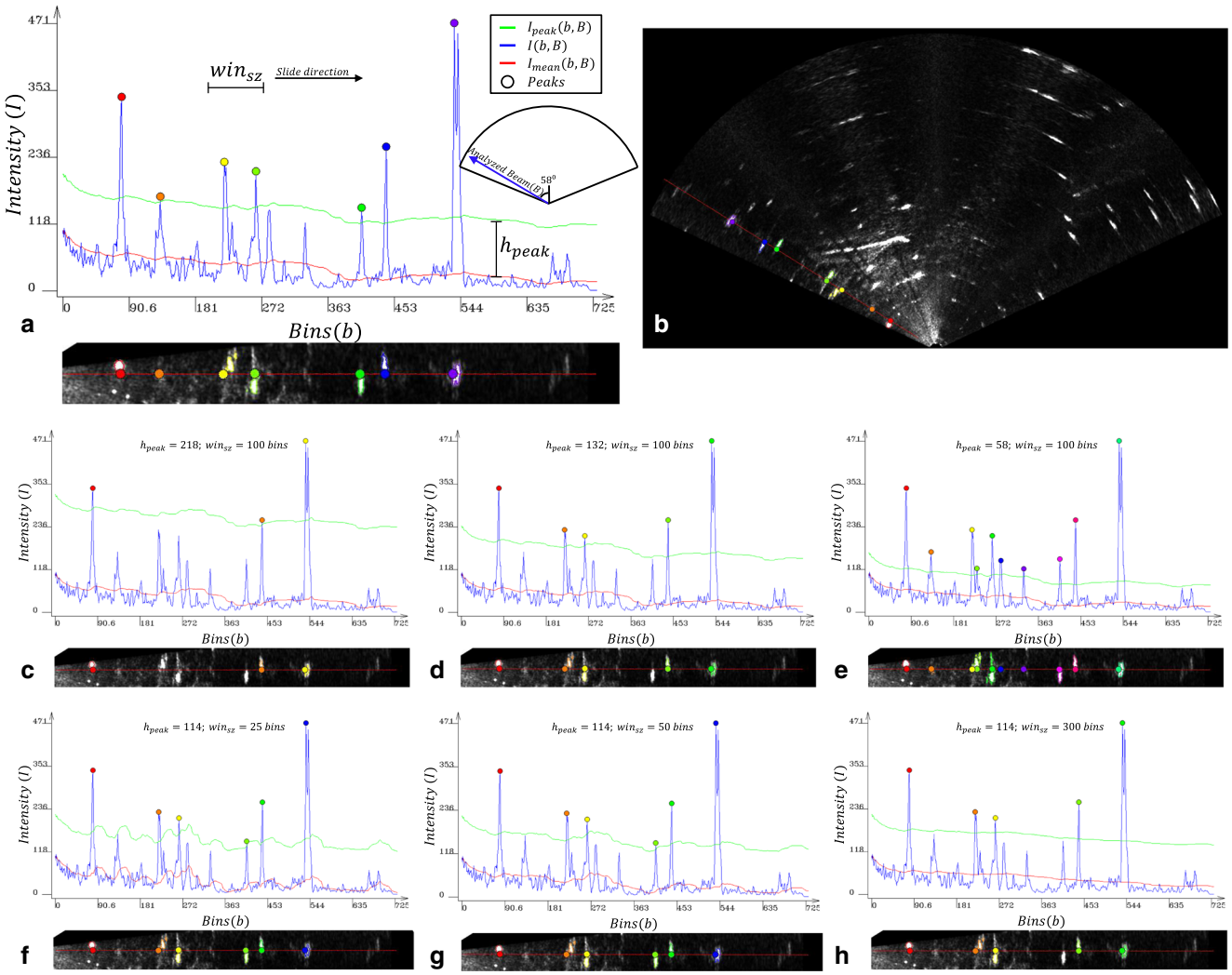

Fig. 5 Local tuning parameters for segmentation. The graph represents the analysis of one acoustic beam B ($\theta_{bin} = 123^{circ}$). In this analysis, the peaks of intensity are detected and used to locally adjust the segmentation parameters. The blue line represents the bins intensities $I(b, B)$; the red line represents the mean intensities $I_{mean}(b, B)$, and the green line represents the minimum intensity for peak detection $I_{peak}(b, B)$. The colored circles represent the detected peaks. As can be seen in Figure **b**, each segment is extracted based on the intensity and position of the detected peaks in Figure (**a**). The behavior of I_{peak} calculated by Eqs. 1 and 2 can be observed in the figures (**c**), (**d**), (**e**) when the parameter h_{peak} is changed and in the figures (**k**), (**l**), (**m**) when the parameter win_{sz} is changed

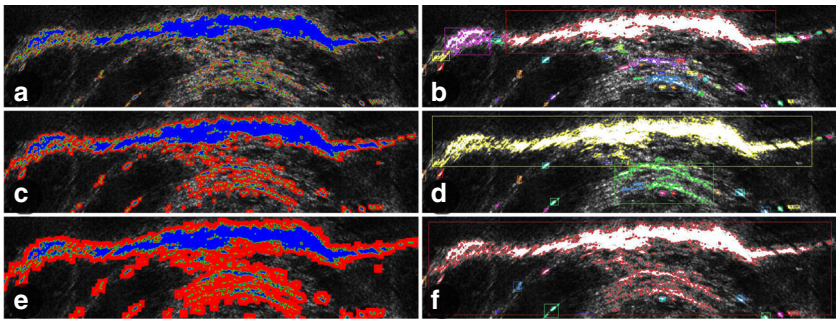

Fig. 6 Segment extraction step, after detecting the intensity peaks a search for connected pixels is performed. This images shown the segment extraction of the same image change the parameter D_{seg}. The images on the left show the pixel search process; the visited pixels (included on the segment) are shown in blue, the segment contour pixels are shown in green and the visited pixels by the distance criteria is shown in red. The right images show the extracted segments. On figures **a** and **b** were used $D_{seg} = 1$; on figures **c** and **d** were used $D_{seg} = 4$ and on figures **e** and **f** were used $D_{seg} = 10$

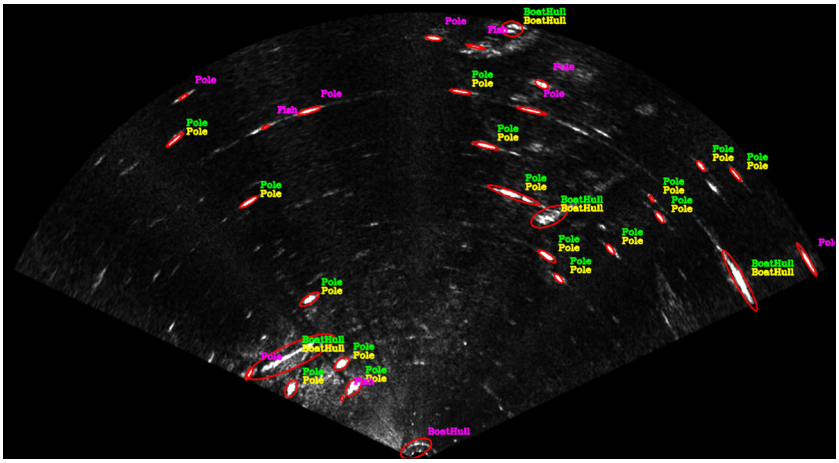

Fig. 7 A sample acoustic image of the training set generated by the developed tool. A demonstration video is available on https://youtu.be/G6c1pBVKI1E

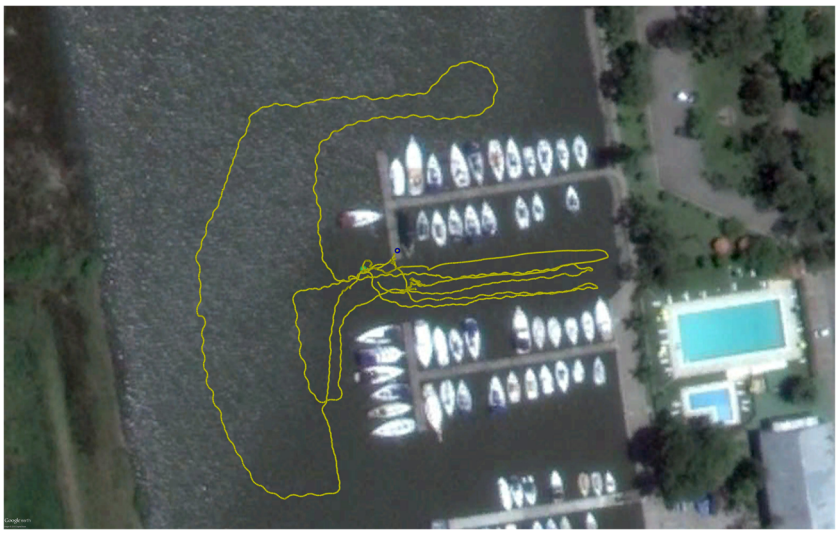

Fig. 8 Satellite image of the marina with the trajectory traveled by the ROV during the acquisition of the Dataset ARACATI 2014 [17]. Map data: Google, DigitalGlobe 2016

Table 1 Dataset information

Class name	Number	Training	Validation
Pole	187	121	66
Boat Hull	128	83	45
Stone	91	59	32
Fish	90	58	32
Swimmer	14	9	5
Total	510	330	180

Initially, *width* and *height* 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 is 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 [2]. 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.

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 *height*. In addiction to the *width* and *height*, 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 *number of pixels* in the segment.

3.4 Segment Classification

In this stage, the supervised classifier Support Vector Machine (SVM) is adopted to classifier each segment. The SVM classifier models the data as a k-dimensional vector and defines an optimal hyperplane that best classifies the vectors depending on its data. The hyperplane definition is an iterative optimization process executed on the training step of the classifier.

A tool was developed to automatically segment the acoustic image, to allow the manual annotation of the segments by a graphical interface and to automatically classify the segments. The tool was developed with the OpenCV library [2] and a screenshot with some annotated segments is shown in Fig. 7.

Table 2 Segmentation parameters

Parameter	Value
$Bearing$	130 degrees
$nBeams$	768 beams
H_{peak}	132
Win_{sz}	100 bins
D_{seg}	4 pixels
$MinSegSize$	20 pixels
$MaxSegSize$	9000 pixels

The SVM implementation is based on the libSVM library [3] which presents several types of kernels that allow us to deal with nonlinear classification. The available kernels are Polynomial, Radial Basis Function (RBF) and Sigmoidal kernels. Since the RBF kernel is know as the best choice in most cases because of its capability to handle with nonlinear classification, it was the adopted kernel for the segment classification. As documented in [3], the kernel parameter function γ and C must be defined.

To define theses parameters an auto training function builds a grid with the classification performances by altering the two parameters (γ, C). The performance of the classifier is calculated by cross validation. The training data are divided into k groups, one of them is adopted for cross-validation and the others train the classifier. The combination which results in the best performance is chosen as the optimal parameter value. The range and variation step of the parameters to build the grid must be defined. In this work we adopted a grid starting in 0.1, ending in 60 with a step of 0.01 for both parameters γ and C.

4 Experimental Results

The experimental results were obtained using real acoustic images from the dataset ARACATI 2014 in which a Forward

Table 3 Ranges values for normalization

Dimension name	Min	Max
Width	1.84	170.09
Height	5.86	817.85
Inertia Ratio	0.07	0.91
Std. Intensity	44.92	1293.59
Mean Intensity	164.76	403.15
Area	0	35205.5
Hull Area	13	107622
Convexity	0	0.83
Perimeter	0	6482.33
Pixel Count	20	9000

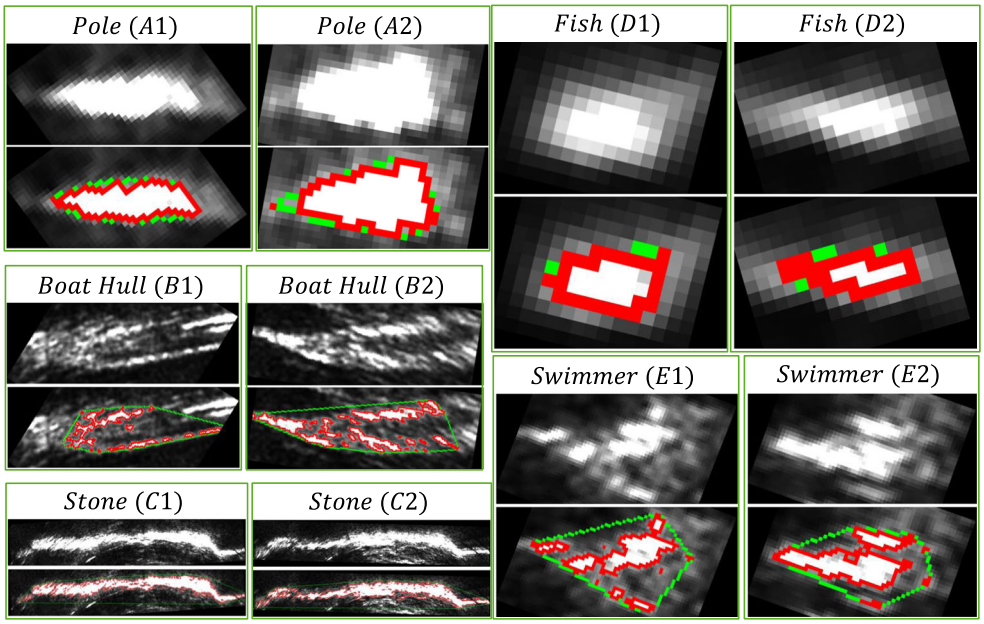

Fig. 9 Segmentation results using the parameters of Table 2. Pixeis in red represent the segment contour, pixeis in green represent the convex hull. The information extracted from each segment is shown on Tables 4 and 5

Looking Sonar (FLS) was used. The images were processed and the segments were manually annotated using the developed tool. The results were obtained using **2D features** and **10D features** as described in Section 3.3.

4.1 Dataset ARACATI 2014

The dataset ARACATI 2014 was made available in [17]. The dataset is composed by 9659 acoustic images recorded in a marina by a 2D Forward Looking Sonar (FLS) Teledyne BlueView P900-130 mounted in mini Remote Operated Vehicle (ROV) LBV300-5 manufactured by Seabotix.

The FLS Teledyne BlueView P900-130 has an angular resolution (beam width) $\Delta\rho_{bin} = 1^{circ}$ and a range resolution $\Delta\theta_{beam} = 0.0254$ m. The generated images are 16-bits gray scale with a resolution of 1429×781.

During the entire path, the ROV remained close to the water surface to keep the Differential Global Positioning System (DGPS) working. The marina structures such as poles, piers and boat hulls and stones are visible in the acoustic images. Some of these objects are highlighted in Fig. 1. The Fig. 8 shows the entire path traveled by the ROV at the marina.

4.2 The SVM Training Dataset

The SVM training dataset was regenerated on this extended version using the developed tool. The training data consists a total of 510 segments over 257 acoustic images

Table 4 Feature information

	Pole		Boat		Stone	
Dim.	A1	A2	B1	B2	C1	C2
Width	8.3	10.7	31.1	34.2	**78.0**	74.7
Height	26.3	19.7	106.1	130.8	**755.9**	772.5
Inertia Ratio	0.31	0.54	0.29	0.26	0.10	**0.09**
Std. Intensity	395.2	**647.4**	115.9	145.5	157.1	147.3
Mean Intensity	291.9	**346.5**	**189.9**	195.8	201.9	193.3
Area	24	10.5	680.5	1293.5	**31339.5**	28367.5
Hull Area	195	171.5	3457.5	5181.5	**82339.5**	80409.5
Convexity	0.123	**0.061**	0.196	0.249	0.380	0.352
Perimeter	112.8	47.3	307.5	253.8	**3138.1**	**25332.4**
Pixel count	85	66	650	1020	6679	**7192**

Table 5 Feature information continuation

	Fish		Swimmer	
Dim.	D1	D2	E1	E2
Width	4.8	**3.1**	22.1	14.7
Height	**8.2**	9.7	38.3	31.6
Inertia Ratio	**0.58**	0.31	0.57	0.46
Std. Intensity	**73.2**	112.4	134.2	136.7
Mean Intensity	195.4	204.7	203.0	202.5
Area	**1.5**	4	276.5	251.5
Hull Area	30.5	**26**	902.5	555.5
Convexity	0.049	0.153	0.306	**0.452**
Perimeter	**15.2**	21.4	115.8	118.2
Pixel Count	**24**	27	218	166

Table 6 2D feature results

Parameters			Result Hit(%)						
γ	C	k	*All*	Pole	Boat	Stone	Fish	Swimmer	Figure
59.334	27.680	2	85	86.3	82.2	87.5	96.8	0	10a
59.334	53.940	5	84.4	86.3	80	90.6	93.7	0	10b
6.024	44.579	10	86.6	90.9	80	93.7	93.7	0	10c
Overfitting test									
41.55	28.45	–	90	94.6	92.1	85.7	95.5	35.7	10d

which were manually classifieds in one of the five different classes: Pole, Boat Hull, Stone, Fish and Swimmer. The data was separated into two group, the validation data (35%) and the training data (65%). To avoid the overfitting problem, the validation set never is used in the training step, and the training set never is used to evaluate the SVM classifier.

The number of segments in each class is shown on Table 1. The adopted parameters of the segmentation algorithm are shown on Table 2, where $Bearing$ is the opening of the sonar field of view, $nBeams$ is the number of bins and $MinSegSize$ and $MaxSegSize$ define respectively the minimum and maximum size of a segment in pixels. The parameters H_{peak}, Win_{sz} and D_{seg} were previously defined in Section 3.2. This parameter was empirically defined doing several tests.

To avoid problems with the scale between the dimensions of the vectors so that one dimension becomes more important than others because it is on a larger scale. All data are normalized using the maximum and minimum values of each dimension found the training dataset, theses values are shown on Table 3.

The object recognition on acoustic images is not a trivial task, as shown on Fig. 9, the shape of the segment is the most distinctive feature for recognition. A quantitative information extracted from the segments is shown on Tables 4 and 5 where the highest and lowest values are highlighted in bold. It is possible to see that the stones are the

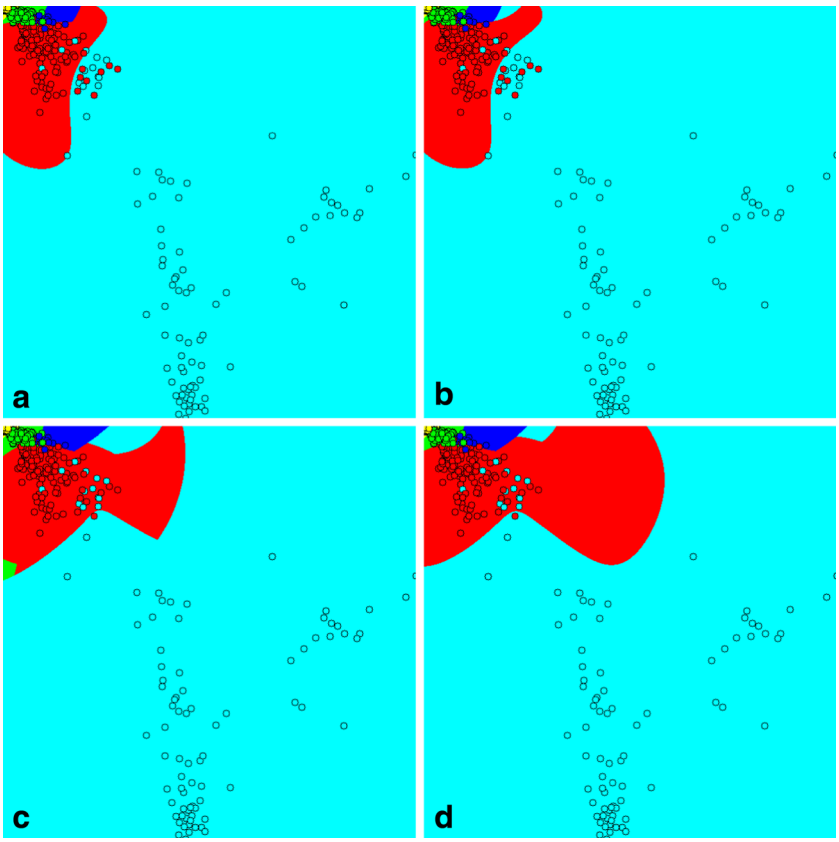

Fig. 10 Images generated to show the classifier hyperspace and its hyperplanes that separate each class for each one of the six tests performed using 2D features. Each class is represented by a color such as fish is yellow, pole is green, boat hull is red, swimmer is blue and stone is cyan. The background colors represent the hyperplanes and each circle represents one object of the ground truth data set

Table 7 10D feature results

Parameters			Result Hit(%)					
γ	C	k	*All*	Pole	Boat	Stone	Fish	Swimmer
1.311	53.940	2	71.6	78.7	77.7	93.7	34.3	20
0.235	49.037	5	82.7	74.2	86.6	90.6	96.8	20
1.442	11.739	10	86.1	81.8	91.1	90.6	93.7	20
Overfitting test								
41.55	28.45	–	99.6	99.4	100	100	98.8	100

largest segments; the fish are the smallest segments and the poles is the most convex segments.

4.3 Results Using 2D Features

Firstly, we performed experiments using 2D features. The first 3 tests were carried unsing 330 segments of the tranning set and the auto training function, that estimate the better parameters γ and C using cross-validation of k-subsets of the training data.

The classifier performance was evaluated comparing the obetened classification with the annotated classification on the ground truth for the 180 segments of the validation data. The results are shown in Table 6.

Using only the width and height of the segments, we correctly classified 86.1% of the validation data using 10 subgroups for cross validation. The classes stones and fish had the highest hit rate (93.7%) and the swimmer class had no hit. The main reason for the swimmer class has no hit is the dataset limitation that has few swimmer images.

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 on Fig. 10, where each circle represents a segment and its position represent the extracted values, e.g. width and height.

The horizontal axis represents the segment width and grows to right, and the vertical axis represents the segments height and grows to down.

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 color represents the classifier hyperplanes and each circle color represent one segment classification annotated on ground truth.

Another 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 did a test forcing the overfitting of the classifier using all available data, e.g. the 510 segments. As shown in the last line of Table 6, 90% of the segments were correctly classified.

For this reason we consider more information about the segment to achieve better results on 10D feature.

4.4 Results Using 10D Features

The same tests performed with 2D features was made to 10D feature. The results are shown on Table 7 revealing a similar behavior for both approaches. The 2D feature got a slightly higher hit rate (86.6% against 86.1% of 10D feature). The main difference was found in the forced overfitting test. The classifier hit 99.6% of the segments using 10D feature against 90% of 2D feature.

Despite the possible overfitting, this result shows that the classifier is able to distinguish the five classes at least of the problem.

5 Conclusion

A method to automatically detect and classify objects in acoustic images of a 2D Forward Looking Sonar (FLS) is proposed. The object segmentation is performed by an algorithm specifically developed for acoustic images. A segment description approach was suggested using geometric and acoustic intensities reflected by the objects. The object classification is performed by the Support Vector Machine (SVM) classifier using the Radial Basis Function (RBF) kernel.

A tool was developed to annotate the image segments and perform automatic object classification. The results showed that it possible to identify and classify objects in real environments such as a marina allowing the creation of semantic maps.

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 on performing new tests in larger and differents environments, produce new public datasets, explore and make comparisons with new classification approaches like Convolutional Neural Network (CNN) or traditional classifiers such as Random Trees (RT) and K-Nearest Neighborhood (KNN).

Finally, we are interested to integrate the proposed method in a Simultaneous Localization and Mapping (SLAM) approach and perform autonomous navigation using semantic information.

Acknowledgments We thank to CNPq, CAPES, FAPERGS, Oil Brazilian Agency, PRH-27 FURG-ANP/MCT and IBP -- Brazillian Petroleum, Gas and Biofuels Institute to support this research. This paper is a contribution of the INCT-Mar COI funded by CNPq Grant Number 610012/2011-8 and 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 terrestrial and underwater environments.

References

1. Botelho, S., Drews, P. Jr., Figueiredo, M.S., Rocha, C., Oliveira, G.L.: Appearance-based odometry and mapping with feature descriptors for underwater robots. J. Braz. Comput. Soc. **15**, 47–54 (2009)
2. Bradski, G.: The OpenCV library. Dr. Dobb's J. Softw. Tools Prof. Prog. **25**(11), 120–123 (2000). ISSN: 1044–789X, http://www.drdobbs.com/open-source/the-opencv-library/184404319
3. Chang, C.C., Lin, C.J.: Libsvm: a library for support vector machines. ACM Trans. Intell. Syst. Technol. (TIST) **2**(3), 27 (2011)
4. Cho, H., Pyo, J., Gu, J., Jeo, H., Yu, S.C.: Experimental results of rapid underwater object search based on forward-looking imaging sonar. In: Underwater Technology (UT), 2015 IEEE, pp. 1–5 (2015). https://doi.org/10.1109/UT.2015.7108235
5. Galceran, E., Djapic, V., Carreras, M., Williams, D.P.: A real-time underwater object detection algorithm for multi-beam forward looking sonar. Navigation, Guidance and Control of Underwater Vehicles (NGCUV) **3**, 306–311 (2012)
6. Guo, J., Cheng, S.W., Liu, T.C.: Auv obstacle avoidance and navigation using image sequences of a sector scanning sonar. In: Proceedings of the 1998 International Symposium on Underwater Technology, 1998, pp. 223–227 (1998). https://doi.org/10.1109/UT.1998.670096
7. Guth, F., Silveira, L., Botelho, S.S., Drews, P. Jr., Ballester, P.: 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, pp. 1–6 (2014)
8. Hurtós, N.V.: Forward-Looking Sonar Mosaicing for Underwater Environments. Ph.D. Thesis, Universitat de Girona (2014)
9. Kim, K., Neretti, N., Intrator, N.: Mosaicing of acoustic camera images. IEE Proceedings Radar, Sonar and Navigation **152**(4), 263–270 (2005). https://doi.org/10.1049/ip-rsn:20045015
10. Kostavelis, I., Gasteratos, A.: Semantic mapping for mobile robotics tasks: a survey. Robot. Auton. Syst. **66**, 86–103 (2015)
11. Lu, Y., Sang, E.: Underwater target's size/shape dynamic analysis for fast target recognition using sonar images. In: Proceedings of the 1998 International Symposium on Underwater Technology, 1998, pp. 172–175 (1998). https://doi.org/10.1109/UT.1998.670085
12. Machado, M., Zaffari, G., Ballester, P., Drews, P. Jr., Botelho, S.: A Topological Descriptor of Forward Looking Sonar Images for Navigation and Mapping, pp. 120–134. Springer International Publishing, Cham (2016). https://doi.org/10.1007/978-3-319-47247-8_8
13. Negahdaripour, S., Firoozfam, P., Sabzmeydani, P.: On processing and registration of forward-scan acoustic video imagery. In: Proceedings the 2nd Canadian Conference on Computer and Robot Vision, 2005, pp. 452–459 (2005). https://doi.org/10.1109/CRV.2005.57
14. Reed, S., Petillot, Y., Bell, J.: An automatic approach to the detection and extraction of mine features in sidescan sonar. IEEE J. Ocean. Eng. **28**(1), 90–105 (2003)
15. Ribas, D., Ridao, P., Tardós, J.D., Neira, J.: Underwater slam in man-made structured environments. J. Field Rob. **25**(11-12), 898–921 (2008). https://doi.org/10.1002/rob.20249
16. Santos, M.M., Drews, P. Jr., Nunez, P., Botelho, S.S.C.: Semantic mapping on underwater environment using sonar data. In: IEEE 13th Latin American Robotics Symposium LARS, pp. 1–6 (2016)
17. Silveira, L., Guth, F., Drews, P., Ballester, P., Machado, M., Codevilla, F., Duarte, N., Botelho, S.: An open-source bio-inspired solution to underwater SLAM. In: IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles NGCUV (2015)
18. Thrun, S.: Robotic mapping: a survey. In: Exploring Artificial Intelligence in the New Millennium, pp. 1–35 (2003)

Matheus dos Santos is Graduated in Computer Engineering at Federal University of Rio Grande (2013) and Master in Computer Engineering at the Federal University of Rio Grande (2015). Nowadays he is a doctoral student in Computer Modeling at the Federal University of Rio Grande acting on Computer Vision, Acoustic Images Processing, Optical Acoustical Systems, Simultaneous Localisation and Mapping, Mobile Robotics.

Paulo Drews Jr. is a D.Sc. and M.Sc. in Computer Science at Federal University of Minas Gerais(UFMG) under supervisor of Prof. Mario Campos, Belo Horizonte, Brazil. His main research interests are robotics, computer vision, image processing, pattern recognition and machine learning. He was a researcher at the ISR Coimbra. He was also a visiting researcher in the ASL at QCAT-CSIRO, Australia. Currently, he is Assistant Professor at Federal University of Rio Grande.

Pedro Núñez is Assistant Professor associated to Tecnología de los Computadores y las Comunicaciones department, at University of Extremadura (Spain). He is Telecommunications Engineer since 2003, and PhD in 2008, at the University of Malaga (Spain). During these years he developed his research as member of the group of Ingeniería de Sistemas Integrados (ISIS) at University of Malaga. In 2007 he starts as professor at University of Extremadura, whitin the Robolab (Laboratorio de Robótica y Visión Artificial) group. He has published more than 40 international publications, some of them in the most important international conferences or journal on the robotics topic, and currently he is leader of different research projects related to social robots.

Silvia Botelho is graduated in Electric Engineering (1991) and received Master degrees (1994) in Computer Science at Federal University of Rio Grande do Sul – UFRGS. She has Ph.D. in Robotics, Informatics, and Telecommunications at Centre National de la Recherche Scientifique (2000). Nowadays she is Full Professor at Universidade Federal do Rio Grande - FURG and director of the Center of Computational Science - C3 and director of the Intelligent Robotics and Automation Group - NAUTEC. Her research is mainly focused in Intelligent Robotic and Automation, applied to Oil and Gas industry and Education.