IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 13, 2017, accepted October 10, 2017, date of publication October 20, 2017,
date of current version November 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2764998

Visual Discrimination and Large Area Mapping of Posidonia Oceanica Using a Lightweight AUV

FRANCISCO BONIN-FONT[1], ANTONI BURGUERA[1], AND JOSE-LUIS LISANI[2]
[1]Systems, Robotics and Vision Group, University of the Balearic Islands, 07122 Palma, Spain
[2]Mathematical Analysis and Processing of Images, University of the Balearic Islands, 07122 Palma, Spain

Corresponding author: Francisco Bonin-Font (francisco.bonin@uib.es)

This work was supported by the Ministry of Economy and Competitiveness under Contract TIN2014-58662-R (AEI,FEDER,UE), Contract DPI2014-57746-C3-2-R (AEI,FEDER,UE), and Contract TIN2014-53772 (AEI,FEDER,UE).

ABSTRACT Controlling and quantifying the presence of *Posidonia Oceanica* (P.O.) in the Mediterranean sea is crucial for the conservation of these endemic ecosystems and to underscore the negative impact of many anthropogenic activities. These activities, which include uncontrolled leisure anchoring or illegal drag fishing, directly affect the tourism and fishing industries. Nowadays, the control and quantification of P.O. is done by divers, in a slow and imprecise process achieved in successive missions of a duration limited by the capacity of the oxygen scuba tanks. This paper proposes the application of robotic and computer vision technologies to upgrade the current P.O. control methods, building large scale coverage maps using the imagery provided by an autonomous underwater vehicle endowed with a bottom-looking camera. The process includes four main steps: 1) training a classifier based on two different Gabor filter image patch descriptors and a *support vector machine*; 2) detecting P.O. autonomously, both *on-line* and *off-line*, in each individual image; 3) color photo-mosaicking the area explored by the vehicle to obtain a global view of the meadow structure; these mosaics are extremely useful to analyze the structure and extension of the meadow and to calculate some of the biological descriptors needed to diagnose its state; and 4) building a binary coverage map in which the classification results of areas with image overlap are fused according to four different strategies. The experiments, performed in coastal areas of Mallorca and Girona, evaluate and compare the proposed descriptors and fusion techniques, showing, in some cases, accuracies and precisions above 90% in the detection of different patterns of P.O., from video sequences at different locations, in different seasons and with different environmental conditions.

INDEX TERMS Posidonia Oceanica, Gabor filters, machine learning, photo-mosaicing, autonomous underwater vehicles.

I. INTRODUCTION

A. MOTIVATION

Posidonia Oceanica (P.O.) is an endemic low-growing seagrass of the Mediterranean that forms vast colonies with a great ecological value, playing a critical role in the equilibrium, maintenance, development and stability of coastal ecosystems and human livelihoods, because: a) they favour the deposition of new sediments on the seafloor and steady the unconsolidated ones, protecting the shoreline against erosion, b) they attenuate currents and wave energy, c) they are also related directly to the abundance of the biodiversity, being a source of food, a refuge for numerous species and a favorable substrate for many organisms, and d) meadows absorb great amounts of carbon and release oxygen to the water by means of photosynthesis, increasing its quality and transparency, mitigating the climate change [1]. Posidonia leaves are green when they are young and during spring and summer are active, but become brown in autumn and photosyntheticaly inactive. In winter, new leaves are generated.

Several biologists have studied the evolution of P.O. along and across the Mediterranean. Terrados and Medina-Pons [2] found a significant increase of density in two Posidonia meadows located in the Balearic Islands, monitored during 6 years. Contrarily, some other studies developed in different habitats, revealed the opposite pattern, showing evidences of decline on a global scale [3]. This decrement was in response to human impacting activities, such as boats petrol or diesel spill, which produce changes in water quality, and mechanical erosion due to uncontrolled leisure anchoring or dragging fishing.

However, other more optimistic points of view defend the idea that there is not a global and general decline but a decline due to an accumulation of local impacts, which can be overcome by acting upon these local causes [4].

The European Commission, in its directive DIR 92/43/CEE, identifies P.O. as a natural habitat of priority interest requiring the delimitation of special areas of conservation. Since, in general, the extension of P.O. meadows and its presence in areas not far from the coast is declining, monitoring and controlling these benthic habitats becomes a crucial task to preserve them and, as a consequence, preserve the benefits that they provide to the tourism and fishing industries, two strategic sectors in many Mediterranean resorts.

Nowadays, the control of P.O. is typically done by divers, who photograph and mark the perimeter of the meadows to see their extension, and install certain gauges inside for measuring their height. Sometimes, divers are tracked with acoustic localizers to build a georeferenced survey [5]. However, this process is slow, tedious, imprecise and limited by the autonomy of the scuba tanks.

Several approaches to map and control P.O. colonies are based on the exploitation of multispectral satellite imagery [6]. However, although the analysis of satellite imagery has revealed to be useful to detect the borders of meadows in shallow waters, it does not result effective in deeper areas where the water column complicates the perception of different blue tonalities.

Acoustic bathymetries performed with a *Side Scan Sonar* (SSS) attached to a vessel hull or to an underwater vehicle [7] can be also of great utility to detect and map the P.O. meadows.

Recently, lightweight *Autonomous Underwater Vehicles* (AUV) equipped with a variety of sensors such as SSS, GPS, *Doppler Velocity Logs* (DVL), *Inertial Measurement Units* (IMU) or cameras have been used to collect data in marine habitats colonized with P.O [8]. In these works the P.O. bottom coverage was estimated by segmenting dark regions corresponding to living P.O. from bright regions of dead matte or bottom sediment. However, the seagrass identification was not truly autonomous as it required the intervention of a human operator.

To the best of our knowledge, the first approach to automatic detection of P.O. using uniquely visual information was presented in [9], developed by the same research group and authors as this paper. In this contribution, images were characterized using Law's filters and automatically classified using a *Logistic Model Tree* (LMT) algorithm.

In the context of *Augmented Reality Subsea Exploration Assistant* (ARSEA), a Spanish national funded project (TIN2014-58662-R), we propose to upgrade the current methodology to visualize, map and control marine areas with P.O., using an AUV equipped with a bottom looking camera and programmed to navigate at a constant altitude. The use of an AUV makes it possible to extend the duration of missions while reducing costs and improving human security. Also, it increases both data resolution and accuracy, and allows highly accurate geo-references. Employing an AUV is combined with the application of several computer vision technologies to detect and map automatically the P.O.

The work presented next is an evolution of our previous work [9]. The main progress is reflected on several issues: a) the patch description has changed from Law's filters to Gabor filters, which are more suitable for the P.O. texture, b) the training dataset has been extended to images from additional environments which present P.O. in clear regression, c) the results of the P.O. binary discrimination are refined locally to smooth the borders, d) the construction of the coverage maps incorporates a pixel aggregation step for areas of image overlap. Unfortunately, there is no possible comparison with other visual algorithms used for the same purpose since, to the best of our knowledge, there is no comparable literature applied on similar purposes.

B. POSIDONIA INTERESTING BIOLOGICAL DESCRIPTORS

It is widely accepted [4], [10] that some of the most relevant biological descriptors needed to assess the ecological status of a P.O. meadow are:

(a) The density, measured as the number of leaves per square meter.
(b) Bottom coverage, expressed as the percentage of sea ground covered by live seagrass with respect to the percentage of ground surface covered by sand, rocks or dead matte. The conservation index Ci is defined as $Ci = P/(P + D)$, where P is the percentage of ground covered with live P.O. and D is the percentage of ground covered by dead P.O. matte, sand, other algae or rocks. This descriptor usually quantifies the dynamics of the meadow and the human impact.
(c) The lower and upper depth limits, which indicate the geographical location of the meadow and its boundaries. The upper depth limit refers to the closest part to the coast and it is easily detectable by means of aerial or satellite imaging, while the lower limits indicate the deepest boundary, been detectable only with remote sensing from vessels or underwater vehicles. These limits give information about the dynamics of the meadows and if they are progressing or regressing.

The majority of these ecological parameters can be computed by processing images of the meadow. Discriminating automatically the P.O. from the rest of elements of the sea ground can be of great utility to be more precise in the calculation of the bottom coverage and the conservation index, and the coverage maps can be useful to delimit the boundaries of the meadow.

C. OVERVIEW

Our goal is to autonomously build coverage maps of P.O. meadows using the images provided by a bottom-looking camera attached to an AUV or a ROV. These 2D maps are geo-referenced to absolute locations (in the basis of the AUV navigation data and a GPS) and can be of great utility to:

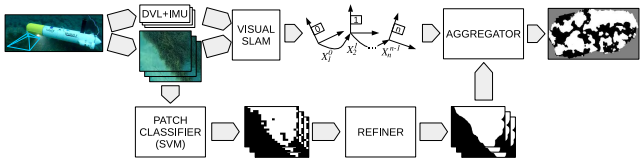

FIGURE 1. Summary of the P.O. coverage map building process.

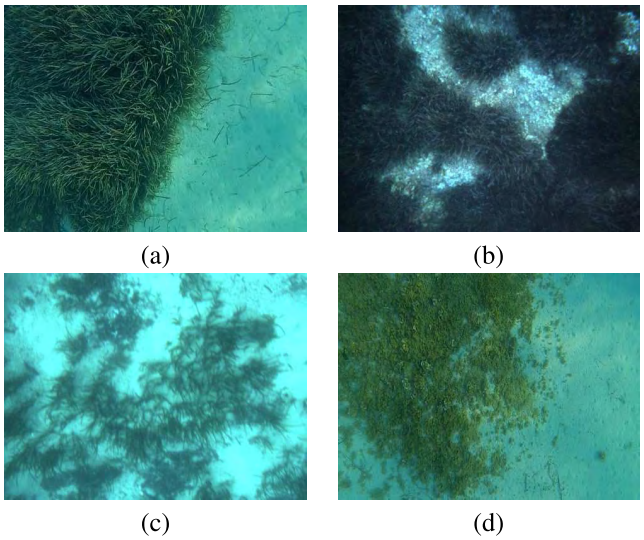

FIGURE 2. Examples of underwater images depicting (a) P.O., (b) P.O. observed under bad illumination conditions, (c) dying P.O. and (d) moss.

1) control the state and evolution in time of the meadows, 2) study their spatial structure, and 3) to measure their global extension and some of the interesting descriptors suitable for assessing its ecological status, such as lower and upper depth limits and bottom coverage [4], [10].

This process involves several steps, which are summarized in Figure 1. First of all, as described in Section II, each of the gathered images is divided in patches, which are subsequently classified as depicting P.O. or not depicting it. Afterwards, as described in Section III, this rough classification is refined until a pixel-level accuracy is achieved. Finally, the refined classifications obtained for each of the gathered images are fused in order to build the global coverage map, as discussed in Section IV. To this end, accurate pose estimates are obtained by means of a visual SLAM and mosaicking approaches [11], [12], out of the scope of this paper. Finally, experimental results are exposed in Section V.

II. PATCH-LEVEL DETECTION

Each image obtained by the bottom-looking camera is divided into a set of $M \times N$ sub-images or patches. A descriptor is computed for each patch and used by a supervised learning approach to classify them as depicting P.O. or not.

Given the visual characteristics of P.O., texture descriptors appear to be distinctive enough to discriminate it from the seafloor and other seaweed species. As an example, Figures 2-(a) and 2-(c) show that even in very different states (young and dying), P.O. texture is clearly distinguishable. Moreover, the texture is significantly

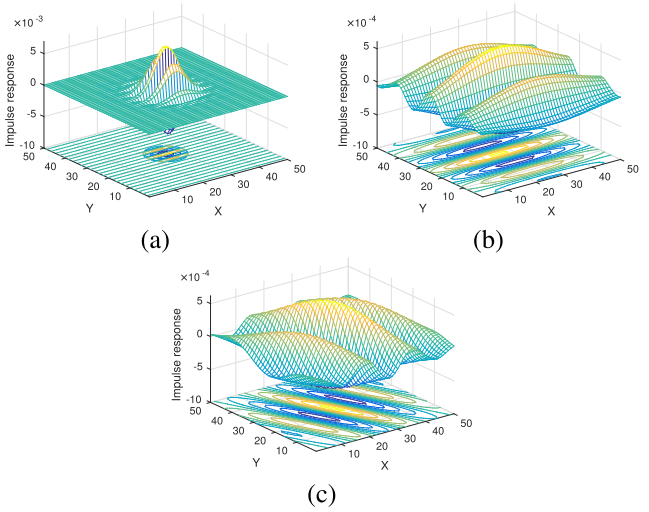

FIGURE 3. Examples of 2D Gabor filters with (a) small envelope scale, (b) large envelope scale and (c) large envelope scale and different carrier orientation.

different to non P.O. vegetation, such as the moss depicted in Figure 2-(d).

Color descriptors might be useful, although not sufficient. On the one hand, different plants may have colors very similar to P.O. On the other hand, the spectral components of light are differently absorbed by water and long wavelengths are usually lost first [13], giving the whole scene green tonalities. Finally, even small changes in illumination may lead to important changes in the perceived colors, as illustrated in Figure 2-(b). As a matter of fact, [14] shows that using color alone in underwater image segmentation leads to poor results and, in the particular case of P.O., tends to be overconfident.

A. PATCH DESCRIPTION

In order to assess the validity of texture analysis and quantify the importance of additional color information, this paper proposes two descriptors, one being based on texture alone and the other on texture and color.

The first descriptor, d_{GG} or *Gray-scale Gabor*, is solely based on texture and relies on 2D Gabor filters. Gabor filters [15] have two important features. On the one hand, they approximate the characteristics of the primary visual cortex of mammals. Thus, they are said to mimic certain parts of human visual perception [16]. On the other hand, they have been found to be particularly well suited for texture representation and discrimination [17]. More specifically, these filters have predominant orientations, similar to the P.O. leaves and, thus, they are likely to provide a strong response in front of P.O.

Roughly speaking, a 2D Gabor filter is the combination of a complex sinusoid, usually referred to as the *carrier*, and a 2D Gaussian-shaped function known as the *envelope*. A filter is characterized by the carrier *orientation* and the Gaussian dispersion or *scale*. The effect of the scale is exemplified in Figures 3-(a) and 3-(b). Also, Figure 3-(c) illustrates how the orientation affects the Gabor filter.

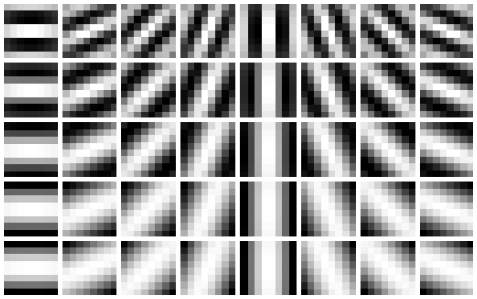

FIGURE 4. The proposed Gabor filter bank. Each row corresponds to a scale and each column to an orientation.

Our proposal is to generate a bank of 40 Gabor filters, involving 8 different orientations and 5 different scales and discretize each filter to an 8×8 matrix. This filter bank is shown in Figure 4.

In order to compute the descriptor d_{GG}, each patch is first converted to grayscale. To this end, each pixel value in the grayscale patch is computed as $0.2989 \cdot R + 0.5870 \cdot G + 0.1140 \cdot B$, being R, G and B the red, green and blue components of that pixel in the original color patch. That is, we compute the luminance, so that the perceived brightness of the pixel is quantified independently of its chromatic content [18].

Afterwards, the grayscale patch is convolved with the whole filter bank. From each convolution, two significant values are extracted: the *local energy* and the *amplitude*. The former is defined as

$$E = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} c(i,j)^2, \tag{1}$$

where m and n are the number of rows and columns, respectively, of c, which is the result of the convolution.

The amplitude is computed as follows:

$$A = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |c(i,j)|. \tag{2}$$

Being each patch convolved with 40 Gabor filters, a descriptor is composed of 80 values: 40 local energy values and 40 amplitude values.

The second descriptor, d_{CG} or *Color Gabor*, is computed in a similar way. In this case, the red, green and blue color channels are convolved separatedly with the Gabor filter bank and both local energy and amplitude are obtained. Thus, a descriptor is composed of 80 values per channel. That is, 240 values in total.

Overall, d_{GG} is solely based on texture information whilst d_{CG} considers both texture and color.

B. TRAINING AND CLASSIFICATION

Two classes, named 0 and 1, are defined for the patches. A patch belongs to class 1 or 0 if the majority of its pixels depict P.O. or not, respectively. Thus, detecting P.O. at the patch level is a binary classification problem.

Our proposal is to use a *Support Vector Machine* (SVM) [19] to perform the patch classification by means of a supervised learning schema and using the mentioned descriptors. Thus, a training is required prior to classification.

To this end, we first manually selected a set of 69 images of different resolutions gathered by an AUV with a bottom looking camera in several coastal areas of Mallorca. These images were taken under different illumination and environmental conditions. The purpose of choosing a variety of images from very different datasets was to take into consideration the diverse tonalities and textures of the P.O., depending on the environment, the depth and the life stage, to augment the classification range of the trained model. One third of the images in the dataset has only P.O. Another third has no P.O. at all, and the last third contains patches with P.O. and patches without it. A hand labeled ground truth was built for these images.

Afterwards, a Monte Carlo cross validation schema was used as follows. First, 14 of the 69 images (approximatedly a 20%) was randomly selected as the training set and the remaining 55 images (approximately an 80%) used as the test set.

Second, the aforementioned descriptors d_{CG} and d_{GG} were computed for the training set and used to train the SVM together with the ground truth.

Third, the descriptors were also computed for the test set and classified using the trained SVM. The quality of the classification was assessed thanks to the ground truth.

These steps were repeated 500 times, randomly building the training and test sets each time. In other words, each of the 500 tests involves classifying a random set containing 80% of the images using a SVM trained with the remaining 20%.

The results of these tests in terms of hit ratio are provided in [14]. However, in the context of this paper, the Monte Carlo cross validation served a different purpose. Among all 500 different training sets used during this cross-validation, the one leading to the best results was selected and used for further training. Being this training set composed of only 14 images (the abovementioned 20%), the training times in further experiments was substantially reduced. This small image set was later extended with additional images from two new environments (see Section V-A.2). Thanks to these additions, the resulting training set is improved and usable in a wider range of scenarios. Let this set of images be referred to as the *extended training set*.

Finally, a SVM with a *Radial Basis Function* (RBF) kernel [20] was trained with the *extended training set* and subsequently used to classify different image sets. The different parameters involved in this process will be described and experimentally assessed in Section V. As a result of this step, a rough classification in which patches are said to depict P.O. or not is achieved. In order to achieve a fine, pixel-level, classification, a refinement algorithm is proposed and described next.

III. PIXEL REFINEMENT

This section presents an algorithm for the post-processing of the initial results of the classifier presented in the previous sections. The algorithm refines the labels (0 or 1) of the classifier, which are initially assigned to patches of the original image. The refinement is obtained by comparing the color of the pixels in the boundaries between regions with different labels, with the average color of the pixels inside each region (i.e. the average color of the pixels labeled as 0 and the average color of the pixels labeled as 1).

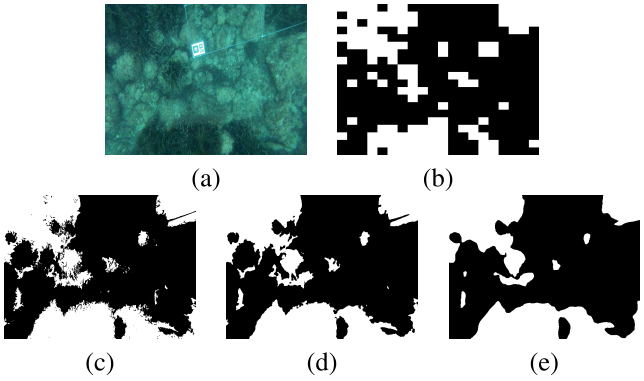

FIGURE 5. Examples of post-processing result. P.O. is depicted in white (label 1) and background in black (label 0). (a) Original image, (b) Initial result of the classifier. Post-processing results for increasing values of the parameter r: (c) 0, (d) 3 and (e) 10.

If the color of the considered pixel is closest to the average color in the region with label 1, the label for this pixel is set to 1 in the processed image. If its color is closest to the average color in the region with label 0, then the label for this pixel is set to 0. The process is repeated iteratively until no more changes are produced. In order to increase the stability of the method, the average color of a given pixel is computed over a neighborhood of radius r (typically set to 3 in all the experiments), which is the only parameter of the algorithm. Algorithm 1 describes in detail the post-processing method and some results are displayed in Figure 5.

IV. AGGREGATION

At this point, the P.O. has been identified in each of the gathered images. Our goal is now to integrate each of these classified images to build the so called global coverage map. Although the P.O. meadow extension could be also evaluated using 3D reconstructions of the environment if the vehicle had a stereo rig, the same nature of the classification algorithm and the typology of the used images require the vehicle to move approximately at a constant height with the camera looking downwards, and forces the necessity to build 2D photo-mosaics to obtain coverage information. To achieve this goal, the overlapping regions between different images have to be identified. Afterwards, the identified overlapping regions have to be combined in order to consistently estimate the presence or absence of P.O.

Algorithm 1 Post-Processing

Input: Input (RGB) image $\boldsymbol{x}$, Classification result (labels, 1 for P.O., 0 otherwise) $\boldsymbol{b}$.
Parameter: Radius of neighborhood r.
Output: Post-processed result $\boldsymbol{bOut}$.

Compute average RGB value of pixels in $\boldsymbol{x}$ with label '1' and such that all of its four connected neighbors (top, down, left, right) have also label '1': $\overline{RGB}_1$;
Compute average RGB value of pixels in $\boldsymbol{x}$ with label '0' and such that all of its four connected neighbors (top, down, left, right) have also label '0': $\overline{RGB}_0$;
do
  for *all pixels* $p \in \boldsymbol{x}$ *with label '1' and such that some of its four connected neighbors (top, down, left, right) have label '0'* **do**
    Compute average RGB value in $\boldsymbol{x}$ of pixels in a neighborhood of $\boldsymbol{p}$ (radius of the neighborhood $= r$): $\overline{RGB}_{\boldsymbol{p},r}$;
    Compute Euclidean distances between the average color of pixel $\boldsymbol{p}$ and the average color inside regions with labels '1' and '0': $d_0 = ||\overline{RGB}_{\boldsymbol{p},r} - \overline{RGB}_0||$, $d_1 = ||\overline{RGB}_{\boldsymbol{p},r} - \overline{RGB}_1||$;
    If $d_0 < d_1$ assign label '0' to pixel $\boldsymbol{p}$;
  end
while *some pixel is assigned a label '0' in the* ***for*** *loop*;
do
  for *all pixels* $p \in \boldsymbol{x}$ *with label '0' and such that some of its four connected neighbors (top, down, left, right) have label '1'* **do**
    Compute average RGB value in $\boldsymbol{x}$ of pixels in a neighborhood of $\boldsymbol{p}$ (radius of the neighborhood $= r$): $\overline{RGB}_{\boldsymbol{p},r}$;
    Compute Euclidean distances between the average color of pixel $\boldsymbol{p}$ and the average color inside regions with labels '1' and '0': $d_0 = ||\overline{RGB}_{\boldsymbol{p},r} - \overline{RGB}_0||$, $d_1 = ||\overline{RGB}_{\boldsymbol{p},r} - \overline{RGB}_1||$;
    If $d_1 < d_0$ assign label '1' to pixel $\boldsymbol{p}$;
  end
while *some pixel is assigned a label '1' in the* ***for*** *loop*;

A. OVERLAP DETECTION

In order to properly detect the overlapping regions, the AUV motion has to be computed. Our proposal is to obtain *on-line* pose estimates by means of visual SLAM and, if necessary, improve these pose estimates *off-line* by means of a mosaicking algorithm. Being both pose estimates the result of a global optimization process, the drift is almost neglectable.

The specific visual SLAM and mosaicking approaches used in this study are described in [11] and [12] respectively, although other methods can be used.

Let X_{i+1}^{i} be the obtained motion estimate from the reference frame of image i to the reference frame of image $i+1$.

The pose of an arbitrary image j with respect to another image i can be computed as follows:

$$X_j^i = \begin{cases} \bigoplus_{k=i}^{j-1} X_{k+1}^k & j > i \\ 0 & j = i \\ \bigoplus_{k=1}^{i-j} \ominus X_{i-k+1}^{i-k} & j < i \end{cases} \quad (3)$$

where $\oplus$ and $\ominus$ denote the composition and the inversion of transformations [21]. Thus, a point p in the coordinate frame of an image j can be expressed in the frame of an image i as $q = X_j^i \oplus p$.

Let $B = [b0, b1, b2, b3]$ be the set of points defining the boundaries of an image with respect of its own coordinate frame expressed in pixels. For example, if the image coordinate frame is located at its center, as it happens with our visual SLAM approach, $B = [[-\frac{w}{2}, -\frac{h}{2}]^T, [\frac{w}{2}, -\frac{h}{2}]^T, [\frac{w}{2}, \frac{h}{2}]^T, [-\frac{w}{2}, \frac{h}{2}]^T]$, where w and h are the image width and height respectively. If the coordinate frames are located at the top-left corner, which is the case of the adopted mosaicking approach, $B = [[0, 0]^T, [w, 0]^T, [w, h]^T, [0, h]^T]$.

Having all the images the same resolution, the boundary polygon of image j with respect to image i can be computed as $B_j^i = X_j^i \oplus B$. Let bx_j^i and by_j^i denote the x and y coordinates, respectively, of the four points in B_j^i.

Our proposal is to select one of the gathered images, namely i, as a global reference frame and then computing B_j^i for each of the other images. In this way, the bounding box of the whole observed area with respect to i, $B_{all}^i = [x_{left}^i, y_{top}^i, x_{right}^i, y_{bottom}^i]$ can be easily computed from the coordinates of all the resulting image boundaries as follows:

$$x_{left}^i = \min_{\forall j} bx_j^i \quad (4)$$

$$x_{right}^i = \max_{\forall j} bx_j^i \quad (5)$$

$$y_{top}^i = \min_{\forall j} by_j^i \quad (6)$$

$$y_{bottom}^i = \max_{\forall j} by_j^i \quad (7)$$

The next step is to sample the whole bounding box at a desired sampling resolution δ and, for each sampling point, check if it lies within each of the individual boundary polygons. The coordinates of a sampled point p_s can be expressed with respect to the frame of an arbitrary image j as $\ominus X_j^i \oplus p_s$, as illustrated in Figure 6. Thus, if a sampled point lies within one or more images, the corresponding pixel intensities can be computed and stored.

After applying this process, a collection of pixel intensities $V_{x,y}$ is available for each sampled point (x, y). If the input images are the result of the P.O. classification, the value list $V_{x,y}$ holds information of the P.O. presence according to each image that observed the corresponding sampled point. Algorithm 2 summarizes the process.

It is important to emphasize that determining whether a sampled point lies within an image or not could also be achieved by simply checking if $\ominus X_j^i \oplus p_s$ is inside B. However, by pre-computing B_j^i and using a fast *Point In Polygon* (PIP) algorithm the computation time is significantly reduced.

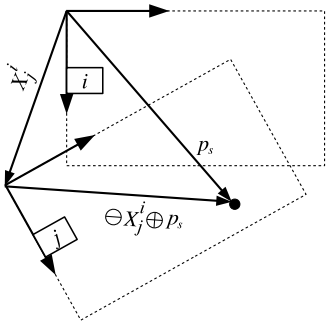

FIGURE 6. Coordinate transformation.

Algorithm 2 Building the Value List

Input: $[w, h]$: Image resolution
δ: Sampling resolution
i: Index of the reference image
$X_{0..n-1}^i$: Relative poses
$I_{0..n-1}$: P.O. classified images
Output: $V_{x,y}$: Set of P.O. detections per sampled point

for $j = 0$ **to** $n - 1$ **do**
  $B_j^i \leftarrow X_j^i \oplus [[0, 0]^T, [w, 0]^T, [w, h]^T, [0, h]^T]$
end
$[x_{min}, x_{max}] \leftarrow [\min_{j=0}^{n-1} bx_j^i, \max_{j=0}^{n-1} bx_j^i]$
$[y_{min}, y_{max}] \leftarrow [\min_{j=0}^{n-1} by_j^i, \max_{j=0}^{n-1} by_j^i]$
for $x = x_{min}$ **to** x_{max} **step** δ **do**
  for $y = y_{min}$ **to** y_{max} **step** δ **do**
    for $j = 0$ **to** $n - 1$ **do**
      if (x, y) **inside** B_j^i **then**
        $p = [p_x, p_y]^T \leftarrow \ominus X_j^i \oplus (x, y)$
        $V_{x,y} \leftarrow V_{x,y} \cup \{I_j([p_x], [p_y]\}$
      end
    end
  end
end

B. DATA FUSION

The goal of the data fusion is to properly aggregate the values in $V_{x,y} = [v_{x,y}^0, v_{x,y}^1, \ldots, v_{x,y}^{n-1}]^T$ in order to obtain a single value for each sampled point stating the likelihood of P.O. at these coordinates.

We propose four different aggregation strategies. These strategies, named $A_{x,y}^{mean}$, $A_{x,y}^{median}$, $A_{x,y}^{max}$ and $A_{x,y}^{min}$ consist on computing the mean, the median, the maximum and the minimum of $V_{x,y}$ respectively.

As, in our particular implementation the regions classified as P.O. are labeled as 1 and the regions not containing P.O. are labeled as 0, a single P.O. detection in $V_{x,y}$ leads to a P.O. result in $A_{x,y}^{max}$. Similarly, a single value in $V_{x,y}$ stating that no P.O. was present will result in not P.O. in $A_{x,y}^{min}$. These

are extremely conservative approaches that will be evaluated experimentally. Concerning $A^{mean}_{x,y}$, the result is a number between 0 and 1. Regarding $A^{median}_{x,y}$, a value of 0 or 1 is directly provided. Also, being ours a binary classifier, computing the median is equivalent to computing the majority label.

For the sake of simplicity, let us define $A^{mean}, A^{median}, A^{max}$ and A^{min} as the result of applying the mentioned aggregation criteria to all the sampled points (x, y). As A^{mean} provides values between 0 and 1, its output must be thresholded. Our proposal is to use the Otsu method [22] to this end.

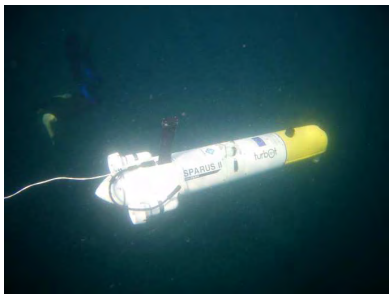

FIGURE 7. The Sparus II AUV.

V. EXPERIMENTAL VALIDATION

A. EXPERIMENTAL SETUP

1) THE AUV

The robot used for the experiments is a SPARUS II AUV [23] (see Figure 7). The vehicle is equipped with a DVL, a pressure sensor, an IMU, a GPS to be geo-referenced in the surface, an *Ultra Short Baseline* (USBL) acoustic link used for localization and data exchange between the robot and a ground station, and a stereo rig grabbing at 10 fps with its lens axis perpendicular to the seafloor. The vehicle has also two led bulbs facing downwards of 40W each one.

The vehicle estimates a first approximation of its displacement, global position and velocity from a two-layer *Extended Kalman Filter* approach fed with the DVL, the USBL, the pressure sensor and the IMU data [24]. This localization is refined each time the navigation module is able to detect, visually, a loop closing [11]. SPARUS works with the ROS middleware [25] to manage all the navigation, control and operation modules, which facilitates software integration and distribution.

2) THE TEST ENVIRONMENTS

The AUV was programmed to navigate in five different locations colonized with P.O. on the west and north-west coast of Mallorca and Girona, with different environmental conditions, such as: illumination (during the day and evening), turbidity (clear and turbid waters), density of P.O., and P.O. coloration. This permitted to get a wide range of different imagery containing P.O.. During each mission, several video sequences were recorded. From this imagery, the extended training set was built as described in Section II-B.

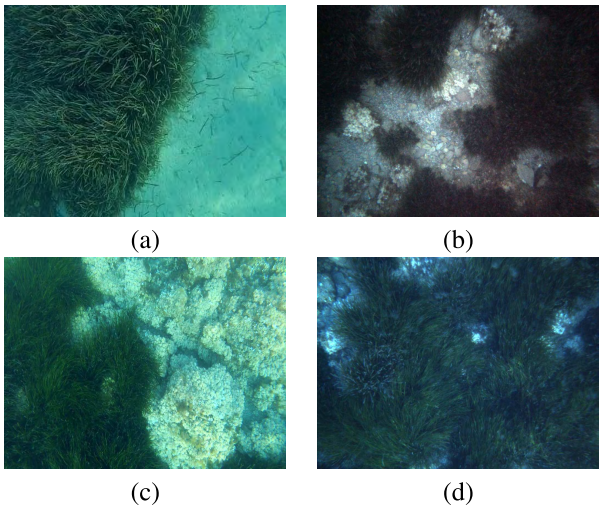

FIGURE 8. Some images from the original training dataset.

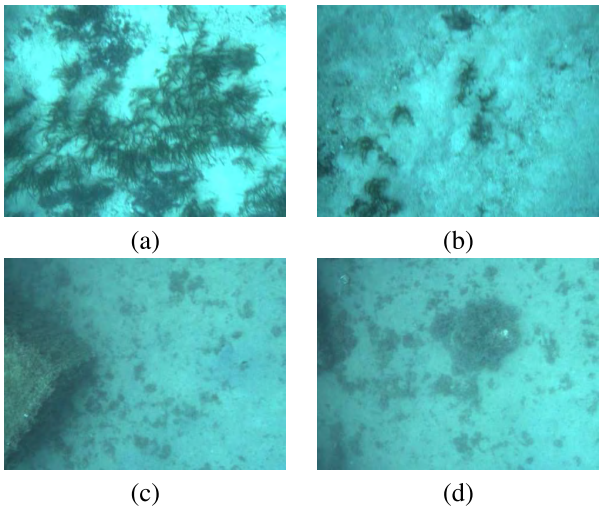

FIGURE 9. Some of the images used to enrich the original training dataset. (a) and (b) Palma Bay; (c) and (d) Sant Feliu harbor.

Figure 8 shows some of the original training set images. Figure 9 shows some of the images used to extend the original training set. The latter images came from two environments. The first is located in Palma Bay, near the sewage marine outfall, where the regressive state of the P.O. is clearly due to the spills coming from the sewage plants [26]. The second environment is located in Sant Feliu harbor (Girona), where the presence of any algae or seagrass is very scarce.

Two different types of assessments were performed: 1) *Off-line*, classifying images extracted from video sequences grabbed from the SPARUS, but processed after the mission, and 2) *On-line*, running the classifier with the best parameters, during the mission, feeding it with the images as they were captured by the camera.

Two different video sequences, one grabbed at Port de Valldemossa and another taken at Palma Bay were used to assess the different classifiers trained with the *extended training set*. None of the images of these two aforementioned sequences were included in the original training set. The

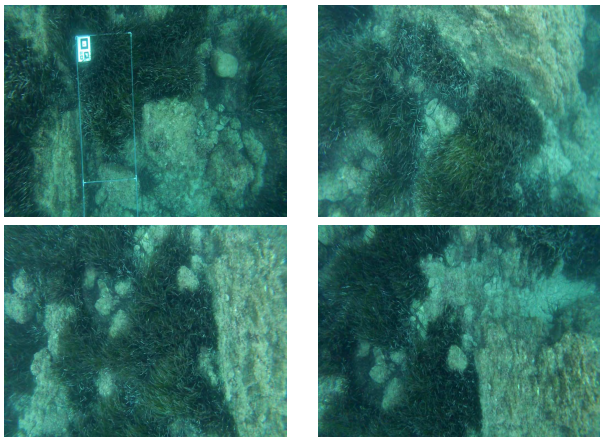

FIGURE 10. Some images extracted from the Valldemossa video sequence.

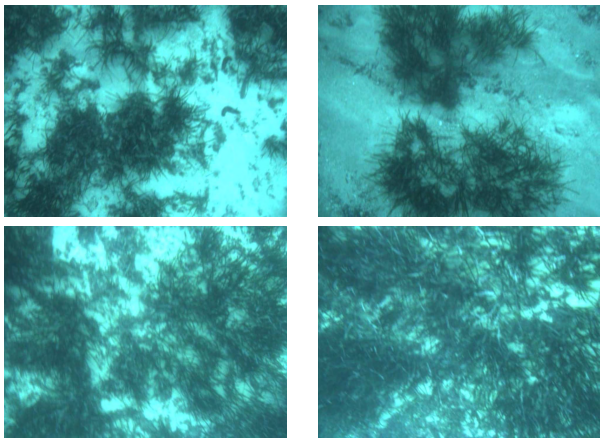

FIGURE 11. Some images extracted from the Palma Bay video sequence.

idea was to test the different classifiers with other video sequences non related with the training process, although some of the images can be of a similar style and appearance. Both sequences cover an area of 400 m^2 approximately. From the sequence of Port de Valldemossa, 333 key-images [12] were selected for the assessment. This sequence was grabbed at 4 meters depth, with low red color absorption, general good lighting conditions except a slight flickering in some parts of the route due to the sun light reflected on the water column, and over a P.O. meadow with excellent health state and dense bottom coverage. Figure 10 shows some frames from this video sequence. From the sequence of Palma Bay, 200 key-images were also selected for the assessment. This latter sequence was grabbed at 13 meters depth, with an important absorption of red frequencies, at 400 meters from the pipe mouth where the solid elements coming from the uncontrolled spill are deposited on the sea bottom. In this area the P.O. is in clear regression because of the organic and chemical pollution coming from the sewage and has a poor bottom coverage and density of leaves and bulbs. Figure 11 exemplifies this video sequence.

A hand labeled ground truth for all the images of both sequences was built and used to evaluate the results. Figure 12 shows the ground truth of sample images from Valldemossa and Palma Bay, in which P.O. is depicted in white and the background in black.

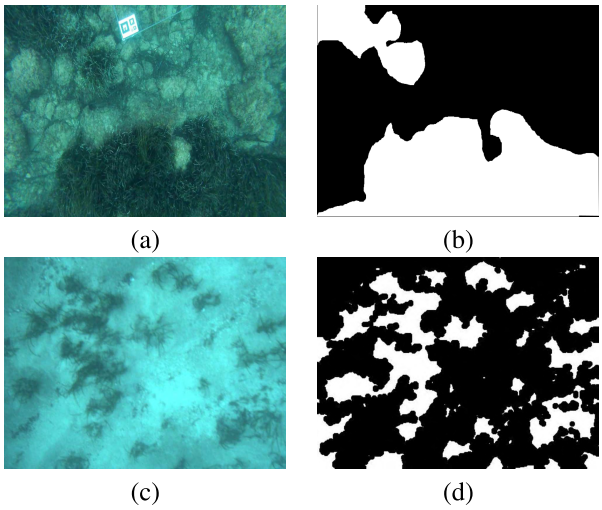

FIGURE 12. Frames from the video sequences (a),(c) and the corresponding ground truth (b),(d).

B. THE EXPERIMENTAL PLAN

Although preliminary results of the explained training and classification strategies were included in [14], [27], the results provided in the present paper extend these preliminary studies in several ways: first, a much wider range of test environments has been used and evaluated by means of a binary classifier extensive diagnostic based on *Receiver Operating Characteristic* (ROC) curves [28]; second, the process is now performed also in real time, with the P.O. classifier executed by the on-board AUV computer; third, different parameters, such as the patch size or the image resolution, are now experimentally evaluated; fourth, the assessment of the different pixel aggregation strategies applied on the coverage maps formation is also another novelty of this paper; fifth, all the validations of this work (classification and aggregation) were performed at a pixel level, while in the preceding references all the validations were done at a patch level, which always includes an additional level of uncertainty. The quality of the images is an important issue that can condition the quality of the results. In [27], the same authors evaluated the effect of applying several image enhancement techniques in the P.O. segmentation process. Furthermore, the training image set contains images of high quality taken with good illumination conditions at lower depths and images with lower quality at deeper environments. Of course, the results will vary depending on the quality of the classified images and the composition of the used training set.

1) TESTED PARAMETERS

In order to fully evaluate the proposed approach, experiments have been conducted with different values for the following parameters:

TABLE 1. Parameter values used in the experiments. The value within each cell denotes the corresponding patch size, in pixels. The number in parentheses is the code assigned to each combination.

		Number of patches		
Descriptor	**Resolution**	**20×15**	**32×24**	**40×30**
d_{GG}	160×120	8×8 (1)	5×5 (2)	4×4 (3)
	320×240	16×16 (4)	10×10 (5)	8×8 (6)
	640×480	32×32 (7)	20×20 (8)	16×16 (9)
d_{CG}	160×120	8×8 (10)	5×5 (11)	4×4 (12)
	320×240	16×16 (13)	10×10 (14)	8×8 (15)
	640×480	32×32 (16)	20×20 (17)	16×16 (18)

Image resolution By testing different input image resolutions, the effects on the time consumption will be quantified as well as the changes in P.O. detection quality. To this end, the input images have been downsampled to 160×120, 320×240 and 640×480 pixels, prior to any of the described processes. It is important to emphasize that, in all cases, the original aspect ratio is preserved.

Number of patches The number of patches affect the detection quality and the granularity of the patch level classification. In this way, small patches may not hold enough information to achieve a proper classification, whilst large patches may difficult the refinement step. To quantify these effects, the resized images have been divided in 20×15 patches, 32×24 patches and 40×30 patches. These divisions lead in all cases to square patches.

Finally, as stated in Section II-A, two descriptors have been defined: d_{GG} (gray scale) and d_{CG} (color). Each of these two descriptors will be tested with all the combinations of the aforedescribed parameters.

Combining three different resolutions, three different number of patches and two descriptors leads to the 18 situations summarized in Table 1. The table shows the resulting patch size for each possible combination. For example, if the image is downsampled to 160×120 and divided in 20×15 patches, the patch size will be 8×8 pixels. Additionally, a number is provided within parentheses for each combination. This number is a code that will be used in further explanations to refer to each particular combination of parameters.

2) QUALITY MEASURES

In order to assess the application of the trained classifiers on different image sets, one experiment has been conducted for each of the aforementioned 18 combinations. Each experiment involved the following steps:

1. Training the SVM with the extended training image set, the hand labeled ground truth and the corresponding parameters of image resolution, number of channels and patches. The trained patches were labeled as P.O. if the majority of pixels in the ground truth were labeled as P.O. and as non P.O. otherwise.
2. Classifying each frame of the two video sequences, the one of Valldemossa and the other grabbed in Palma Bay.
3. Measuring the total classification time for each image in each sequence and for each experiment. That is, we measured the time spent to classify all the patches in each image. Then, the mean and the standard deviation of the classification time were computed. A low variance is a good indicator that the corresponding mean time reflects the achievable frame rate when the algorithm is executed *on-line*.
4. Calculating the total number of *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP) and *False Negatives* (FN), for each image of each video sequence at a pixel level. A positive appears when a pixel is classified as P.O. and a negative refers to pixels classified as non P.O. True and false indicate whether the corresponding pixel in the ground truth coincides with the obtained classification or not, respectively.
5. For every one of the 18 experiments, the mean values of TP, TN, FP and FN were calculated from all the individual values of TP, TN, FP and FN obtained for each image of the sequence. These 18 mean values of TP, TN, FP and FN, were used to calculate 18 values of *Accuracy*, *Precision*, *Recall* and *Fall-out*, one for each experiment, defined as:

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$Fall - out = \frac{FP}{FP + TN} \tag{11}$$

The *Accuracy* defines the hit rate of our classifier with respect to the whole population of classified elements, that is, how many elements have been correctly classified with respect to the total of treated elements. The *Precision* denotes the percentage of TP with respect to all classified as positives.

The *Recall* is the percentage of TP with respect to the number of all elements really positive, and the *Fall-out* represents the number of FP with respect to the number of all elements really negative.

6. The 18 different values of Recall and Fall-out form the ROC curve for the pixel comparison criteria. ROC curves are a classical tool used in a variety of disciplines, from medicine [29] to robotics [30], to analyze and diagnose a binary classifier as certain parameters are varied. The ROC curves plot the Fall-out in the horizontal axis vs the Recall in the vertical axis. ROC curves analysis permit to obtain the optimal classification model including the optimal parameters, which are those that provide a trade-off between a minimum *Fall-out* with a maximum *Recall*. ROC curves close to the diagonal line (so called *line of no discrimination*) indicate a random classifier while ROC curves close to the lines $y = 1$

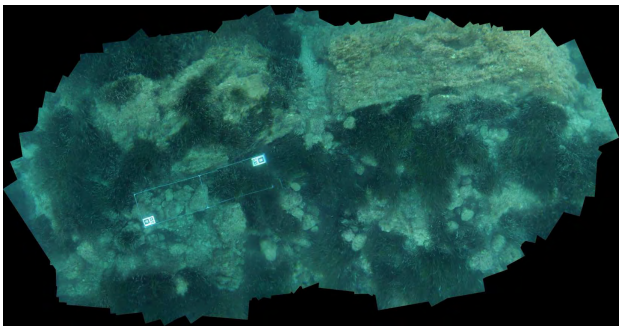

FIGURE 13. Original photo-mosaic of the Valldemossa video sequence.

and $x = 0$ indicate a classifier with high performance. Points above the *line of no discrimination* indicate a classification better than random. For each ROC curve, its *Area Under the Curve* (AUC) was calculated as a quantitative measurement of the classifier performance. Area values range from 0.5 (no apparent accuracy) to 1.0 (perfect accuracy) as the ROC curve moves towards the left/top boundaries [31]. A common and accepted approximation to a diagnostic test is [32]: areas between 0.90 and 1 correspond to an excellent (A) classifier, areas between 0.80 and 0.90 reflect good (B) classifiers, areas between 0.70 and 0.80 denote fair (C) systems and areas under a 0.6 correspond to poor (D) or fail (F) classifiers.

In order to evaluate the different pixel aggregation strategies in the formation of the coverage maps, the following actions were performed:

1. For every one of the 18 different experiments on the Valldemossa dataset, the coverage map was computed according to Section IV, using the 4 different aggregation strategies: A^{mean}, A^{median}, A^{max} and A^{min}. That makes a total of 18×4 different resulting coverage maps.
2. For every different coverage map, the number of TP, TN, FP and FN were computed at a pixel level, comparing all these maps with a ground truth coverage map obtained hand labeling the original color mosaic obtained using [12] with the original color key frames of the Valldemossa dataset (see the original photo-mosaic in Figure 13). This photo-mosaic is essential to see the structure and state of the meadow. The resulting color photo-mosaic from the images of Palma Bay was not reliable, since it was extremely difficult to obtain visual loop closings from sandy and dead matte bottoms, which generated evident misalignments in the resulting mosaic.
3. The *accuracy*, *precision*, *recall* and *fall-out* of each experiment/aggregation pair were calculated.

C. EXPERIMENTAL RESULTS

1) INITIAL CLASSIFICATION

Figure 14 shows the mean and standard deviation of the classification time per image, calculated for the Valldemossa and Palma Bay datasets, and for each one of the 18 different experiments. These time data has been obtained *off-line*, from a laptop with very similar characteristics as the vehicle computer [23], that is, an *Intel* i7 processor working at 2.5 GHz, 4 cores, 8GB of RAM and an Ubuntu 16.04 O.S. Data is grouped in increasing image resolutions and number of color channels (1 channel for gray scale and 3 channels for color). In each group, data is sorted in ascending order by the number of patches in the x and y directions. The mean values are depicted as bars and the standard deviation intervals are depicted as vertical lines over each bar top.

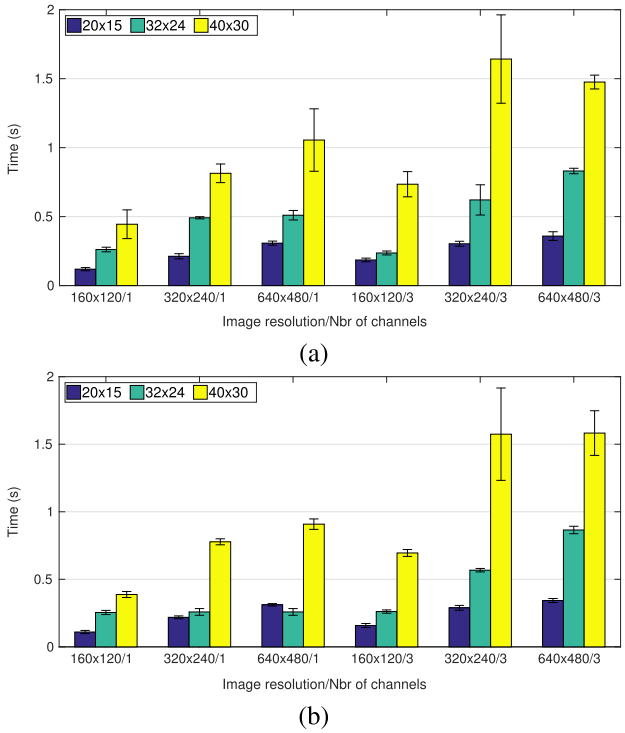

FIGURE 14. Mean and standard deviation of the classification time per image, for the Valldemossa (a) and Palma Bay (b) datasets. Data is grouped in image resolutions, and sorted by number of patches.

Both graphics reflect a similar pattern: a) in each group, the classification time increases as the number of patches increases, which suggest that, although the patch size is smaller, more time is needed to process the larger amount of them, b) the classification time is prone to increase as the image resolution increases, and c) the time is globally higher for color images than for gray scale images, comparing groups with the same image resolution. Standard deviations are, in general, small, except for those settings that imply highest execution times, including 3 channels, resolutions of 320×240 and 640×480, and the highest number of patches (40×30).

In terms of running time saving for *on-line* applications, settings which imply gray scale frames and lower number of patches are the fastest (less than 0.5 seconds/frame). However, additional statistical values must be analyzed to decide an option with a good trade-off between speed and good performance.

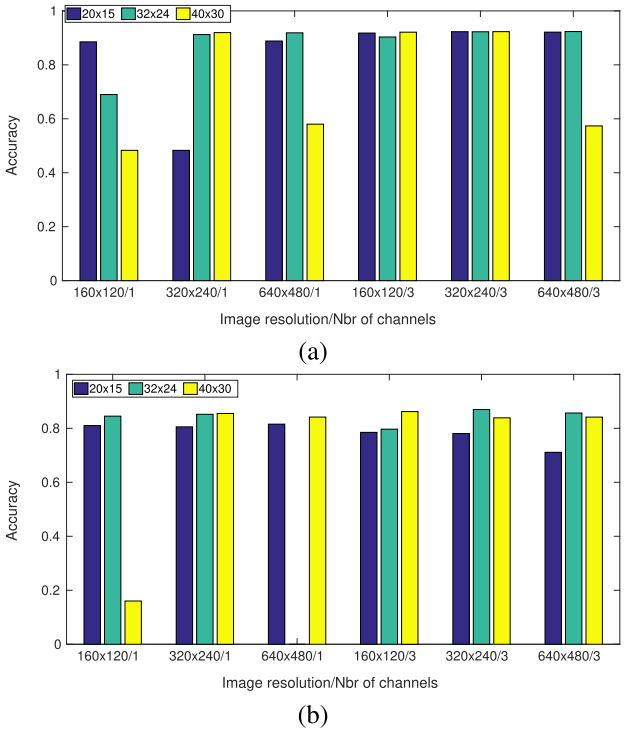

FIGURE 15. Accuracy of the 18 experiments for Valldemossa (a) and Palma Bay (b).

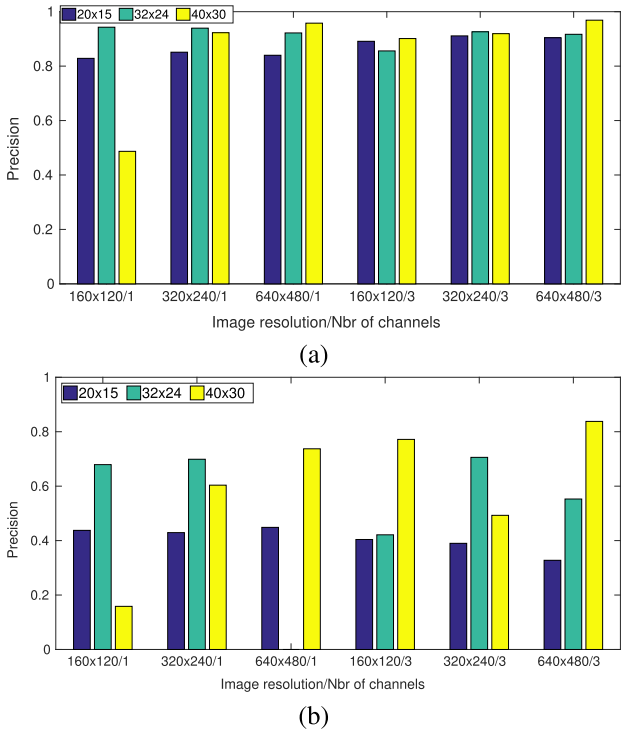

FIGURE 16. Precision of the 18 experiments for Valldemossa (a) and Palma Bay (b).

Figures 15 and 16 show the mean accuracy and precision, as defined by equations 9 and 11, for the 18 experiments. The data is grouped in the same way as in Figure 14. Accuracy and precision data corresponding to experiment number 8 for the Palma Bay dataset is missing since all the classified frames indicated a complete absence of Posidonia. In consequence, the aforementioned ratios did not make much sense.

The analysis of plots 15 and 16 permits to infer several conclusions:

- Although results for the Palma Bay dataset are slightly worse than for the Valldemossa dataset, for both datasets, the accuracy is prone to get or to exceed the 80% in the majority of the experiments which involve a three channel classification; this means that the number of (FP+FN) tends to be small.
- Accuracy results involving one channel classification do not follow any clear trend and are not conclusive since some setting combinations generate a good result in one dataset and worse for the other. However, the combinations involving one channel and a resolution of 320×240 with either 32×24 or 40×30 patches seem to be stable and to present a good performance in both datasets, in terms of accuracy.
- Precision for almost all combinations applied on the Valldemossa dataset exceeds 80%, meaning low levels of FP. However, for the Palma Bay sequence, the number of FP seems to be higher since only 4 combinations reach or clearly exceed 70%. One of these combinations includes a single channel, a resolution of 320×240 and 32×24 patches, and the other two involve 3 channels and 40×30 patches.

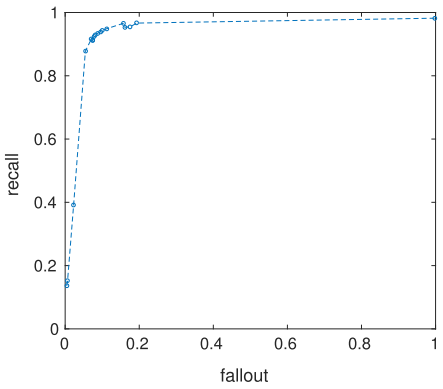

FIGURE 17. Classification ROC curve for the Valldemossa video sequence.

In summary, for *off-line* evaluations, where the classification time is not as critical as in real-time applications, settings including a 3 channel description can be used to obtain a better classification performance. However, for *on-line* applications where the data discrimination must be fast and in real time, a combination with a single channel, a resolution of 320×240 pixels and with 32×24 patches can be used to get an acceptable performance.

Figures 17 and 18 show the ROC curves build from the 18 *Fall-out* and *Recall* values, computed following the pixel level comparison criteria, for the Valldemossa and Palma Bay

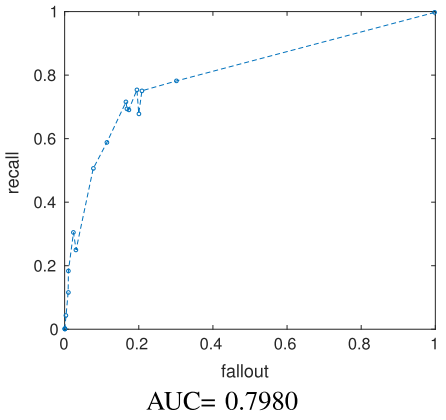

FIGURE 18. Classification ROC curve for the Palma Bay dataset.

datasets, respectively. Likewise, the AUC is indicated below each curve to provide a numerical diagnose of the classifier. The AUC for the classifier applied on the Valldemossa dataset defines it as an excellent classifier while the AUC for the classifier applied on the Palma Bay dataset reflects a fair-good classifier. Although the results of the classifier for the Palma Bay dataset are worse than the ones obtained for the Valldemossa dataset, the global performance on Palma Bay is still highly reliable, given the quality of the images, the difficulty to discriminate dead and alive P.O., to describe their texture, and the poor density of seagrass in this environment.

For both datasets, the classification diagnose performed at a pixel level gives a high AUC, which means a high reliability for the classifier. The *best* points of the ROC curve are those that represent the best trade-off between a high *Recall* and a low *Fall-out*. In the case of Figure 17: points (0.09603, 0.9392), corresponding to experiment 16, (0.07836, 0.9246), corresponding to experiment 15, (0.08118, 0.9283), corresponding to experiment 17, (0.08795, 0.9345), corresponding to experiment 13 and (0.09965, 0.9427) and corresponding to experiment 12. For the same curve, the points that present the worse relations *Fall-out* vs *Recall* are (0.9985, 0.982), for the experiment 3, and (0.00423, 0.136), for experiment 18. The *best* points of this curve correspond to those obtained with the settings (image resolution, patch size and number of channels) that most likely provide the best results on the classifier. The *worse* points on the curves indicate those combinations to avoid since they provide the worse classification results. In this case, the experiments that involve the description and classification of color images report much more reliable results than the ones with gray-scale images.

2) CLASSIFICATION AFTER PIXEL AGGREGATION AND MOSAICKING

Figures 19-(a) and 19-(b) show the *accuracy* and *precision*, respectively, of each pixel aggregation strategy, for each of the 18 experiments on the Valldemossa dataset.

Observing both figures, the general conclusions would be: a) in terms of accuracy, for the majority of the combinations

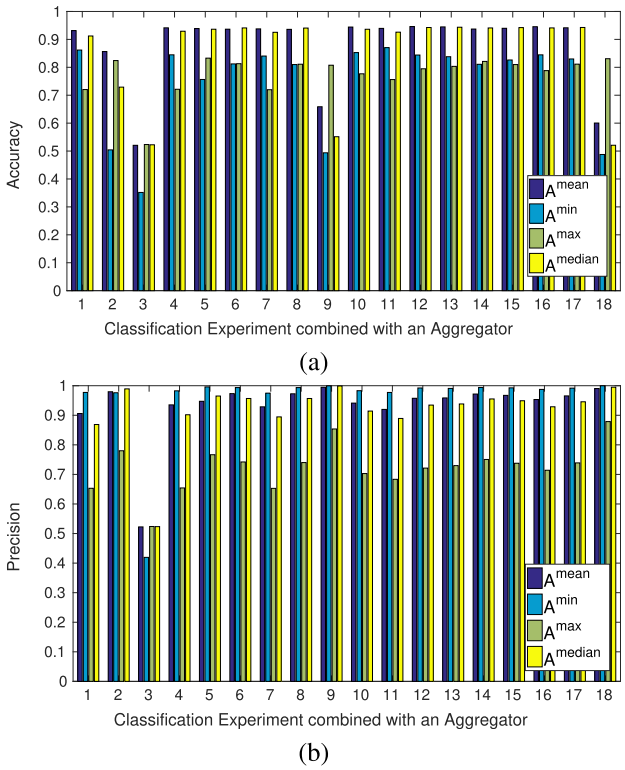

FIGURE 19. Accuracy (a) and Precision (b) of the aggregation strategies, for the Valldemossa dataset.

that involve a 3 channel Gabor description, the A^{mean} and the A^{median} pixel aggregation strategies have a superior performance than A^{min} and A^{max}; high accuracies mean low levels of falsely classified pixels, b) the precision is higher than 0.8 for all aggregation strategies, except for all experiments with the A^{max} aggregator and the experiment number 3 (d_{GG} with the minimum image resolution and 40×30 patches); this means a general lack of FP after the pixel aggregation.

The mean *precision*, *accuracy* and *recall* were: 0.8825, 0.8148 and 0.8079, respectively.

Four ROC curves were formed from the *recall* and *fall-out* values, computed from the data obtained after the comparison of the resulting coverage maps formed with the four different aggregation strategies over the 18 different experiments, and the ground truth coverage map.

In consequence, every ROC curve represents the performance of each pixel aggregation process applied on the 18 different experiments, performed with different detector settings. Figures 20-(a) and 20-(b) show the aforementioned ROC curves with their respective AUC obtained from the A^{mean} and A^{median} aggregators. Figures 21-(a) and 21-(b) show the ROC curves with their AUC, from the A^{max} and A^{min} aggregators. According to the curve shape and the AUC values, the A^{mean} and A^{median} aggregators are the ones that present an excellent performance, thus the ones recommended to be used in this kind of applications. On the other side, the A^{max} aggregator presents a good performance and the A^{min} aggregator a poor performance.

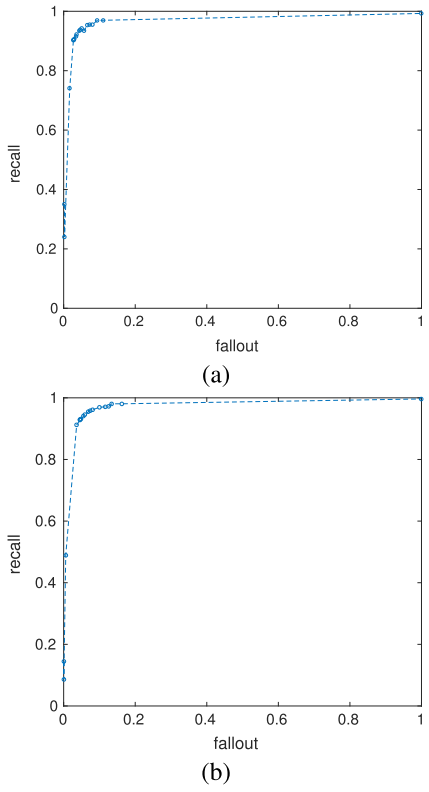

FIGURE 20. ROC curves of the aggregation results for the Valldemossa dataset. (a) A^{mean} and (b) A^{median}.

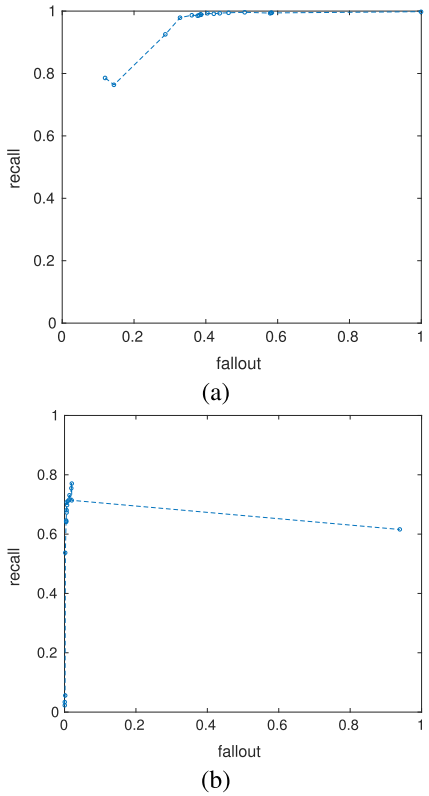

FIGURE 21. ROC curves of the aggregation results for the Valldemossa dataset. (a) A^{max} and (b) A^{min}.

Some samples of the best combinations could be, for instance, A) two points of the A^{median} aggregator curve: (0.05549, 0.94), corresponding to the experiment 15 (d_{CG}, resolution: 320×240; 40×30 patches) and (0.04624, 0.9295) corresponding to experiment 6 (dGG, resolution: 320×240; 40×30 patches), and B) two points of the A^{mean} curve: (0.0459, 0.9384), corresponding to the experiment 12 (d_{CG}, resolution: 160×120; 40×30 patches) and (0.06565, 0.9531) corresponding to experiment 10 (d_{CG}, resolution: 160×120; 20×15 patches).

Contrarily, some of bad combinations would be in the points (0.002375, 0.5376), corresponding to experiment 5 (dGG, resolution: 320×240; 32×24 patches) with an A^{min} aggregation, or (0.1443, 0.7637), from experiment 9 (dGG, resolution: 640×480; 40×30 patches) and aggregation A^{max}.

In consequence, an adequate combination involves, for instance, a 3-channel description with a A^{mean} or A^{median} aggregation, an image resolution of 320×240 and 40×30 patches. On the contrary, a bad combination would be any 1-channel description with a A^{min} aggregation.

Figure 22-(a) shows the hand labeled ground truth of the P.O. coverage map, corresponding to the mosaic of Figure 13. Figures 22-(b) and 22-(c) show, respectively, the resulting coverage map from experiments 15 and 6, using the pixel aggregation A^{median}. Figures 22-(d) and 22-(e) show the resulting coverage map from experiment 12 and 10, respectively, using the pixel aggregation A^{mean} and before the Otsu thresholding. Using these highly realistic coverage maps, the automatic calculation of the P.O. bio-parameters of bottom coverage, conservation index, and upper and lower borders is very easy, just counting black and white pixels and delimiting the transitions between black and white parts.

On the other side, Figure 23 shows two coverage maps obtained with a *bad* combination classifier-settings/pixel-aggregation: (a) experiment 5, aggregation A^{min} and (b) experiment 9, aggregation A^{max}.

As an important conclusion, Figure 20 suggests that the aggregation process using A^{median} or A^{mean} improves the initial classification results for the Valldemossa dataset (compare with Figure 17).

D. *Executing* on-line *THE P.O. CLASSIFIER*

The software package corresponding to the classifier was wrapped into a ROS node and installed in the vehicle to be run *on-line*, during the mission, as the robot moved around the environment and grabbed the video sequence. The robot image processing pipeline includes a conversion from raw frames to RGB encoding, a rectification according to the camera calibration parameters stored in the vehicle and a down-sampling by 2 of the original resolution.

The classification node was trained in gray-scale, first, and then with color images, using patches of 8×8 pixels and resolutions reduced to 320×240 pixels.

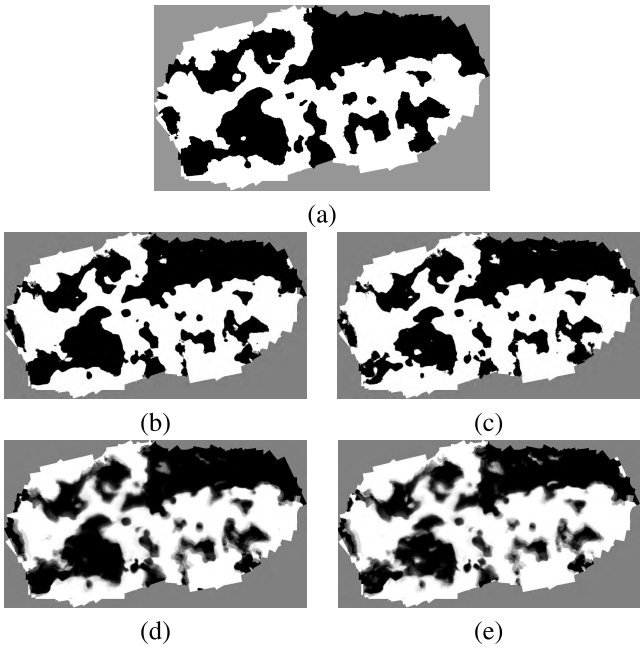

FIGURE 22. (a) Hand labeled ground truth of the coverage map for mosaic of Figure 13. Four coverage maps, obtained with an *excellent* combination classifier-settings/pixel-aggregation: (b) experiment 15, aggregation with A^{median}, (c) experiment 6, aggregation with A^{median}, (d) experiment 12, aggregation with A^{mean} and (e) experiment 10, aggregation with A^{mean}. Results corresponding to (d) and (e) are shown prior to the Otsu thresholding.

FIGURE 23. Two coverage maps, with a *bad* combination classifier-settings/pixel-aggregation. (a) Experiment 5, aggregation with A^{min}, (b) experiment 9, aggregation with A^{max}. P.O. is marked in white and the background in black.

The node returned frames classified in gray-scale at 5.8 fps and at 2.6 fps using the 3 channels, being the original frame rate of 7.5 fps. The reduction of the frame rate for the classified images with 3 channels is evident, but, as it has been seen in previous sections, the classification results are significantly better. The purpose of detecting P.O. *on-line* is to get this environmental information in real time, as fast as possible, as the vehicle moves.

An illustrative video showing the execution of the classifier *on-line* can be seen in https://www.youtube.com/watch?v=lG9szHFPnjk&t=17s. Images show different marine environments located in Mallorca, colonized with P.O. with different textures. The video shows, at the left of the screen, the original images captured from the camera. The results of the classification are superimposed to the original frames and shown at the right of the screen, in green. This is a proof of concept that it is possible to train the system, *off-line*, with a variety of P.O. images with diverse characteristics, and then apply the trained model *on-line* on different environments, getting promising results.

VI. CONCLUSIONS AND FORTHCOMING WORK

In this paper we propose the use of several image processing and machine learning techniques to automatically detect and quantify the presence of Posidonia Oceanica in video sequences of vast areas of sea floor grabbed with an AUV.

Our approach divides every image of each sequence in patches of the same size and describes each patch using color and a bank of Gabor filters. The process includes three phases:

a) A training phase using a heterogeneous group of images and a SVM to obtain a classification model. b) A classification phase using the trained model, in which the P.O. is discriminated from the background in all the images of a video sequence. This phase also involves a refinement step in which the rough SVM classification is iteratively improved until reaching a pixel level classification. The overall process can be run off-line, but also on-line, from the robot computer, during the mission. c) The construction of a coverage map with all the classified key frames of the video sequence, which turns out to be a photo-mosaic of the surveyed area, but with the P.O. highlighted in white and the rest in black; the challenging point in the construction of these maps lies in the pixel aggregation strategy used to determine the value of a map pixel shared with different images. These coverage maps are an excellent tool to evaluate the state, extension and several biological parameters of the P.O. meadows visualized in the inspected area.

Experiments include an extensive combination of different parameter settings, such as several patch sizes, different image resolutions or the use of one or three color channels on two different video sequences. Also, four different pixel aggregation strategies have been tested for the coverage map formation.

The quality of the P.O. classification and pixel aggregation for coverage map formation has been evaluated using ROC curves and compared with hand-made ground truth images and coverage maps. Results of extensive tests reveal that, although a couple of combinations including gray-scale images have an stable and adequate performance, including the color in the patch description increases notably the classifier success rate. On the other side, an additional and important conclusion is that the use of pixel aggregation techniques for the formation of the coverage map improves the initial classification results.

In summary, for *off-line* applications, using 3 channel Gabor descriptors is preferable than using one channel Gabor descriptors, but, if the classifier must be run *on-line*, the combination with only one channel descriptor, an image resolution of 320 $\times$ 240 pixels and 40 $\times$ 30 patches can also be used, guaranteeing a good trade-off between classified frame rate and classification accuracy.

The ongoing work (out of the scope of this paper) includes the use of the on-line P.O. detection for the vehicle mission re-planning, in a dynamic path-planning context. The idea is to apply a navigation schema able to re-direct, on-line, the vehicle towards certain areas of interest, depending on the environmental data captured in real time, and the programmed criteria. For instance, drive, automatically, the AUV towards zones densely colonized with P.O. or follow the border of the meadow. Both implementations, the stand alone and the ROS versions, are available for the scientific community, in two public *GitHub* repositories, [33] and [34].

REFERENCES

[1] E. Diaz-Almela and C. Duarte, "Management of natura 2000 habitats 1120, *Posidonia beds (Posidonia oceanicae)," European Commission, Brussels, Belgium, Tech. Rep. 2008 01/24, 2008.

[2] J. Terrados and F. J. Medina-Pons, "Inter-annual variation of shoot density and biomass, nitrogen and phosphorus content of the leaves, and epiphyte load of the seagrass Posidonia oceanica (L.) Delile off Mallorca, western mediterranean," *Sci. Marina*, vol. 75, no. 1, pp. 61–70, 2011.

[3] G. Jordà, N. Marbà, and C. M. Duarte, "Mediterranean seagrass vulnerable to regional climate warming," *Nature Climate Change*, vol. 2, no. 11, pp. 821–824, 2012.

[4] D. Moreno, P. A. Aguilera, and H. Castro, "Assessment of the conservation status of seagrass *(Posidonia oceanica)* meadows: Implications for monitoring strategy and the decision-making process," *Biol. Conservation*, vol. 102, no. 3, pp. 325–332, 2001.

[5] D. Scaradozzi *et al.*, "Innovative technology for studying growth areas of Posidonia oceanica," in *Proc. IEEE WorkShop Environ., Energy Struct. Monitor. Syst.*, Sep. 2009, pp. 71–75.

[6] R. Matarrese, M. Acquaro, A. Morea, K. Tijani, and M. T. Chiaradia, "Applications of remote sensing techniques for mapping Posidonia oceanica meadows," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2008, pp. 906–909.

[7] G. Di Maida *et al.*, "Discriminating between Posidonia oceanica meadows and sand substratum using multibeam sonar," *ICES J. Marine Sci.*, vol. 68, no. 1, pp. 12–19, 2011.

[8] A. Vasilijevic, N. Miskovic, Z. Vukic, and F. Mandic, "Monitoring of seagrass by lightweight AUV: A Posidonia oceanica case study surrounding Murter island of Croatia," in *Proc. Medit. Conf. Control Autom. (MED)*, Jun. 2014, pp. 758–763.

[9] F. Bonin-Font, M. M. Massot, and G. O. Codina, "Towards visual detection, mapping and quantification of Posidonia oceanica using a lightweight AUV," *IFAC-PapersOnLine*, vol. 49, no. 23, pp. 500–505, Jun. 2016.

[10] C. Pergent-Martini *et al.*, "Descriptors of *Posidonia oceanica* meadows: Use and application," *Ecol. Indicators*, vol. 5, no. 3, pp. 213–230, 2005.

[11] P. L. Negre, F. Bonin-Font, and G. Oliver, "Cluster-based loop closing detection for underwater slam in feature-poor regions," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2016, pp. 2589–2595.

[12] E. Garcia-Fidalgo, A. Ortiz, F. Bonnin-Pascual, and J. P. Company, "Fast image mosaicing using incremental bags of binary words," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2016, pp. 1174–1180.

[13] F. Bonin-Font, A. Burguera, and G. Oliver, "New solutions in underwater imaging and vision systems," in *Imaging Marine Life: Macrophotography and Microscopy Approaches for Marine Biology*. 2013, pp. 23–47.

[14] A. Burguera, F. Bonin-Font, and E. García-Fidalgo, "Building large-scale coverage maps of Posidonia oceanica using an autonomous underwater vehicle," in *Proc. MTS/IEEE Oceans*, Aberdeen, U.K., Jun. 2017, pp. 1–6.

[15] D. Gabor, "Theory of communication. Part 1: The analysis of information," *J. Inst. Elect. Eng. III, Radio Commun. Eng.*, vol. 93, no. 26, pp. 429–441, 1946.

[16] S. Marcelja, "Mathematical description of the responses of simple cortical cells," *J. Opt. Soc. Amer.*, vol. 70, no. 11, pp. 1297–1300, Nov. 1980.

[17] M. A. Hoang, J.-M. Geusebroek, and A. W. M. Smeulders, "Color texture measurement and segmentation," *Signal Process.*, vol. 85, no. 2, pp. 265–275, Feb. 2005.

[18] A. K. Jain, *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1989.

[19] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[20] T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel methods in machine learning," *Ann. Stat.*, vol. 36, no. 3, pp. 1171–1220, 2008.

[21] R. Smith, M. Self, and P. Cheeseman, "A stochastic map for uncertain spatial relationships," in *Proc. 4th Int. Symp. Robot. Res.*, 1988, pp. 467–474.

[22] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[23] M. Carreras *et al.*, "SPARUS II, design of a lightweight hovering AUV," in *Proc. 5th Int. Workshop Marine Technol. (MARTECH)*, 2013, pp. 1–2.

[24] E. Guerrero-Font, M. Guerrero-Font, P. L. Negre, F. Bonin-Font, and G. O. Codina, "An USBL-aided multisensor navigation system for field AUVs," in *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst. (MFI)*, Sep. 2016, pp. 430–435.

[25] M. Quigley *et al.*, "ROS: An open-source robot operating system," in *Proc. ICRA Workshop Open Source Softw.*, 2009, pp. 1–5.

[26] F. Bonin-Font, M. M. Campos, P.-L. N. Carrasco, G. O. Codina, E. G. Font, and E. G. Fidalgo, "Towards a new methodology to evaluate the environmental impact of a marine outfall using a lightweight AUV," in *Proc. IEEE Oceans*, Aberdeen, U.K., Jun. 2017, pp. 1–8.

[27] A. Burguera, F. Bonin-Font, J. Lisani, A. B. Petro, and G. Oliver, "Towards automatic visual sea grass detection in underwater areas of ecological interest," in *Proc. IEEE Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, Berlin, Germany, Sep. 2016, pp. 1–4.

[28] T. Fawcet, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.

[29] A. R. Henderson, "Assessing test accuracy and its clinical consequences: A primer for receiver operating characteristic curve analysis," *Ann. Clin. Biochem.*, vol. 30, no. 6, pp. 521–539, 1993.

[30] D. Seita *et al.*, "Large-scale supervised learning of the grasp robustness of surface patch pairs," in *Proc. IEEE Int. Conf. Simulation, Modeling, Program. Auto. Robots (SIMPAR)*, San Francisco, CA, USA, Dec. 2016, pp. 216–223.

[31] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 521–539, 1982.

[32] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, 2011.

[33] A. Burguera, F. Bonin-Font, and J. Lisani. (2017). *POD Source Code: STAND ALONE Application*. [Online]. Available: https://github.com/aburguera/POD_STANDALONE

[34] A. Burguera, F. Bonin-Font, and J. Lisani. (2017). *POD Source Code: ROS Node*. [Online]. Available: https://github.com/aburguera/POD_ROS

FRANCISCO BONIN-FONT received the degree in telecommunications engineering from the Polytechnical University of Catalonia, Barcelona, in 1996, and the Ph.D. degree in computer engineering from the University of the Balearic Islands in 2012. He has been ten years with the industry of information technology services addressed to bank business, before he initiated his academic activities. He has participated as a Technician and a Researcher in nine projects funded by the Spanish Scientific Council and the European Commission. He is also an Assistance Senior Lecturer with the Department of Mathematics and Computer Science, University of the Balearic Islands. He has authored or co-authored over 40 papers, among journals, book chapters, and conference proceedings in the field of image processing and underwater robotics, during his research activities in the Systems, Robotics and Vision Group, University of the Balearic Islands.

ANTONI BURGUERA received the Ph.D. degree in computer engineering from the Universitat de les Illes Balears in 2009. He has been involved as participant and as leading researcher, in several projects granted by the local administration, the Spanish Scientific Council and the European Commission. He is currently an Associate Professor with the Department of Mathematics and Computer Science, University of the Balearic Islands and member of the Systems, Robotics and Vision group. He has authored over 50 articles in mobile robotics, mostly focusing on acoustic and visual sensor data filtering and processing.

JOSE-LUIS LISANI received the Ph.D. degree in computer science and applied mathematics from the Universities of Illes Balears, Spain and Paris-Dauphine, France, in 2001.

He is currently an Assistant Professor with the University of the Balearic Islands, Spain. His research interests include the analysis and processing of color images and video sequences. He has co-authored the book *A Theory of shape identification* (Springer LNM, 2008).