A Performance Analysis of Feature Extraction Algorithms for Acoustic Image-Based Underwater Navigation

António José Oliveira *[ID], Bruno Miguel Ferreira [ID] and Nuno Alexandre Cruz [ID]

INESC TEC and Faculty of Engineering, University of Porto, 4200-465 Porto, Portugal; bmmf@fe.up.pt (B.M.F.); nacruz@fe.up.pt (N.A.C.)
* Correspondence: antonio.ventura@fe.up.pt

Abstract: In underwater navigation, sonars are useful sensing devices for operation in confined or structured environments, enabling the detection and identification of underwater environmental features through the acquisition of acoustic images. Nonetheless, in these environments, several problems affect their performance, such as background noise and multiple secondary echoes. In recent years, research has been conducted regarding the application of feature extraction algorithms to underwater acoustic images, with the purpose of achieving a robust solution for the detection and matching of environmental features. However, since these algorithms were originally developed for optical image analysis, conclusions in the literature diverge regarding their suitability to acoustic imaging. This article presents a detailed comparison between the SURF (Speeded-Up Robust Features), ORB (Oriented FAST and Rotated BRIEF), BRISK (Binary Robust Invariant Scalable Keypoints), and SURF-Harris algorithms, based on the performance of their feature detection and description procedures, when applied to acoustic data collected by an autonomous underwater vehicle. Several characteristics of the studied algorithms were taken into account, such as feature point distribution, feature detection accuracy, and feature description robustness. A possible adaptation of feature extraction procedures to acoustic imaging is further explored through the implementation of a feature selection module. The performed comparison has also provided evidence that further development of the current feature description methodologies might be required for underwater acoustic image analysis.

Keywords: acoustic image analysis; feature detection; feature matching; sonar data; underwater navigation

1. Introduction

Accurate navigation is fundamental for an Autonomous Underwater Vehicle (AUV) to be able to safely explore unknown environments. For their overall robustness, sonars emerge as one of the most preferable technologies for this purpose [1], being able to generate acoustic images of the AUVs' surroundings. Detecting and recognizing reliable reference landmarks throughout sequences of these images is crucial, since this information can then be used to improve pose estimation, as detailed in Figure 1.

In this work, special focus is given to feature detection and matching towards AUV localization in confined and unknown environments, through the application of feature extraction algorithms commonly employed for visual odometry purposes. The characteristics of such environments and the lack of previous knowledge about its layout accentuate the need for accurate localization. In recent years there has been increased research interest on the application of these algorithms towards acoustic image analysis [2–4], due to the recorded performance for optical imaging [5–8]. Furthermore, for localization purposes, the goal of employing these algorithms is not to identify particular sets of features, unlike more specific solutions proposed in the literature [9], but rather to achieve a generic solution, enabling its application in unknown environments. Isolated objects (e.g., moorings,

mines, pillars) in the insonified volume of water normally appear as blobs in an acoustic image. In confined environments, planar walls result in parabolic curves. Salient features on walls may originate easily detected features but they strongly differ from a given point of view to another, causing additional difficulties in matching. Corners on walls are usually represented as two intersecting curves on the polar image. Due the variety of features and their deformation as the viewpoint changes, it is difficult to anticipate the type of feature to look for. The challenge is amplified especially if generic features in open waters and confined environments are expected. We argue that a comparison, based on state-of-the-art feature extractors, has the potential to better understand the strengths and weaknesses of such algorithms and to lead to their adaptation for acoustic images. Due to the differences between acoustic and optical images, it is not clear that such an adaptation is possible.
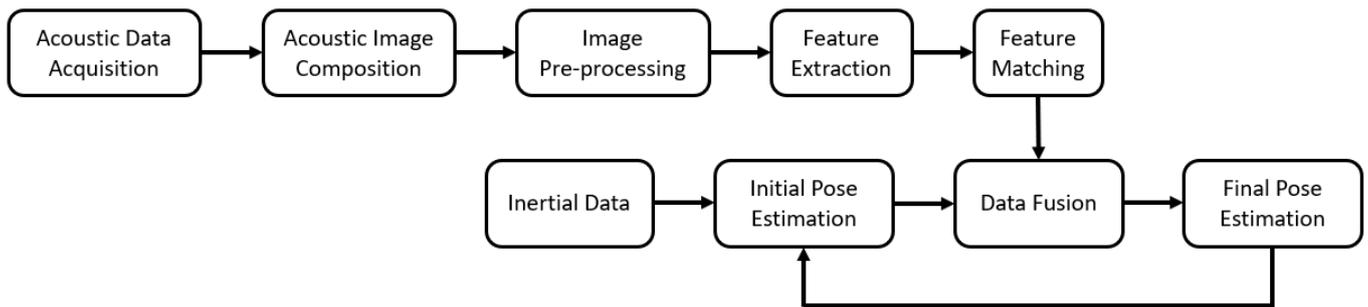

Figure 1. Principle diagram for acoustic image-based underwater navigation.

Feature extraction algorithms developed for optical image analysis take into account specific characteristics of such images. Acoustic images greatly differ from optical images. Firstly, structure representation varies from optical imaging to acoustic imaging. This comes as a consequence of the nature of the image acquisition process, based on the emission and reception of acoustic waves. Such is evident in the acoustic representation of a square water tank, shown in Figure 2. Secondly, acoustic images are affected by the presence of acoustic shadows, low resolution, distortion, and range-varying attenuation [3]. In more confined areas, the occurrence of multiple echoes makes it even harder to extract useful information for navigation. Moreover, due to relatively low scanning speeds of this type of sensors, the motion of the AUV will result in a deformation of environmental features' acoustic representation, posing additional challenges to the identification and matching throughout a sequence of images. Figure 2 also illustrates the impact of the multipath problem. These problems further hamper the application of feature extraction algorithms, since they may result in incorrect landmark detection.
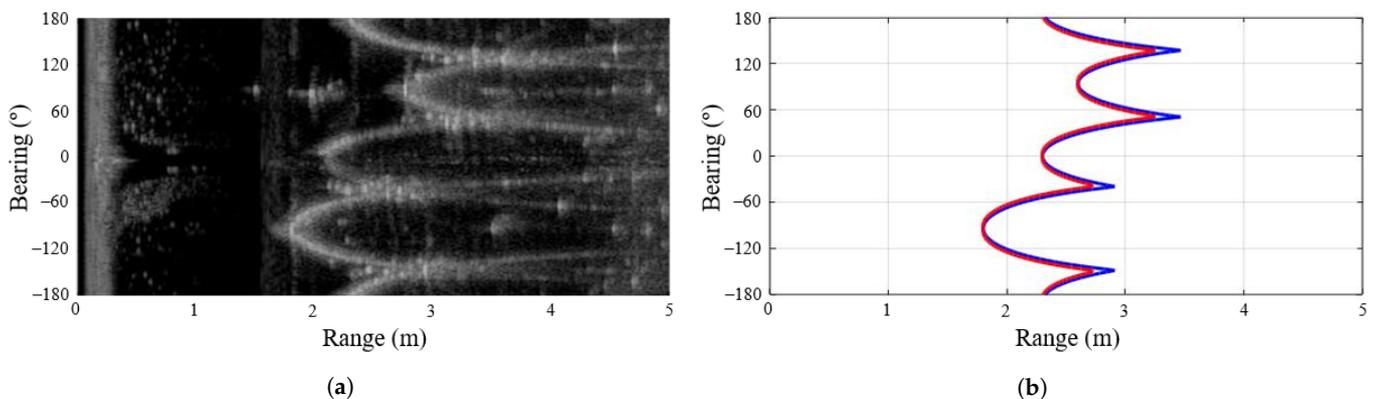

Figure 2. Illustration of challenges portrayed by acoustic image analysis. (**a**) Acoustic image obtained through mechanical scanning imaging sonar from a water tank and (**b**) ideal polar representation of the walls of the water tank (blue) and ideal sonar representation (red).

This article presents a performance analysis and comparison of the SURF (Speeded-Up Robust Features), ORB (Oriented FAST (Features from Accelerated Segment Test) and Rotated BRIEF (Binary Robust Independent Elementary Features)), BRISK (Binary Robust Invariant Scalable Keypoints), and SURF-Harris algorithms. Our goal is to assess the feasibility of employing these tools to acoustic images generated by a Mechanical Scanning Imaging Sonar (MSIS), in a structured environment, for landmark detection and recognition. These algorithms were selected as the most suitable for the proposed task based on previous analysis available in the literature, for both optical and acoustic imaging. This comparison was based on the collection of typical statistics in image processing evaluation but also in a more qualitative evaluation. To the best of our knowledge, this paper contributes to the literature as the first addressing this topic so exhaustively and especially focused on MSIS acoustic images and AUV operation in confined environments. We highlight the following contributions:

- We present a detailed analysis on the performance of each algorithm under the challenges initially described and outline the more appropriate characteristics for localization and detected limitations. This analysis was performed in light of several metrics, such as the number of features detected, the total number of associations performed, and the number of incorrect associations observed;
- We propose and implement a feature selection procedure focused on removing features resulting from a multipath effect or secondary echoes.

In what follows, Section 2 is focused on a comparison of acoustic images and optical images, while Section 3 introduces a detailed literature review of the algorithms under analysis. The methodology adopted for this study is introduced throughout Section 4, along with further details regarding the datasets and software used. The results obtained and further discussion are presented in Section 5. Section 6 summarizes the main conclusions of this study.

2. Optical Images and Acoustic Images

Optical image acquisition is based on the capture of electromagnetic waves in the visible spectrum, while acoustic images are generated through insonification with multiple echoes. These different processes result in distinct image characteristics. Moreover, when image acquisition is performed in motion, the resulting image transformation is clearly distinct. In this section, we further highlight major differences and possible problems that may affect the performance of the algorithms considered, in an effort to provide the reader further insight into the challenges faced.

In an initial assessment, we take into consideration the example provided by Figure 3. Contrary to the characteristic sharpness of optical images, acoustic images evidence significant blur. This property may impact the performance of feature extractors, since these are focused on the detection of salient features. The characteristics of confined environments make acquired images more prone to multipath problems. Unlike optical images that accurately replicate a scene, these problems result in the introduction of image artifacts. This can be observed in the previous figure, where the water tank walls are distinctively represented and, beyond their limits, a number of artifacts appear, resulting from secondary echoes. In terms of feature extraction, this is a similar situation to trying to retrieve features from an optical image portraying an object placed in front of a set of mirrors: It will be possible to extract several features but only a few would correspond to the actual object. Furthermore, due to the properties of the acoustic beams emitted by a MSIS, the occurrence of artifacts resulting from reflections on the AUV's own body, as highlighted in Figure 3, are also common. The described phenomena are a source of ambiguity for the detection of landmarks, since the image elements introduced are very similar to the acoustic representation of actual landmarks.

As it is also possible to verify, the considered acoustic images make use of a polar coordinate system and thus do not portray simple two-dimensional projections of environment scenes, as in the case of optical images. This property has a strong effect on landmark

representation, which may affect the performance of feature extractors, since these aim at identifying image features with well-defined characteristics. The acoustic portrayal of corner-like structures presents a good example of such a problem. A number of feature extractors are focused on the detection of these. However, in acoustic images, corner-like features are portrayed as the intersections of two curves, as illustrated in Figure 4, which may invalidate some of the assumptions made by feature extraction algorithms about corner-features' properties.
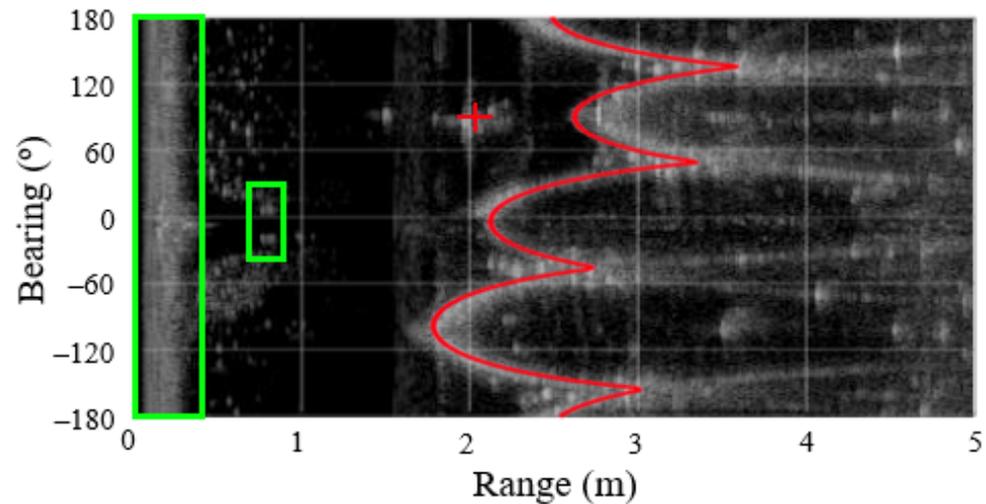

Figure 3. Acoustic image obtained through mechanical scanning imaging sonar from a water tank. The representation of the tank walls and a floater are marked in red. Reflections on the Autonomous Underwater Vehicle's (AUV) body are highlighted in green.
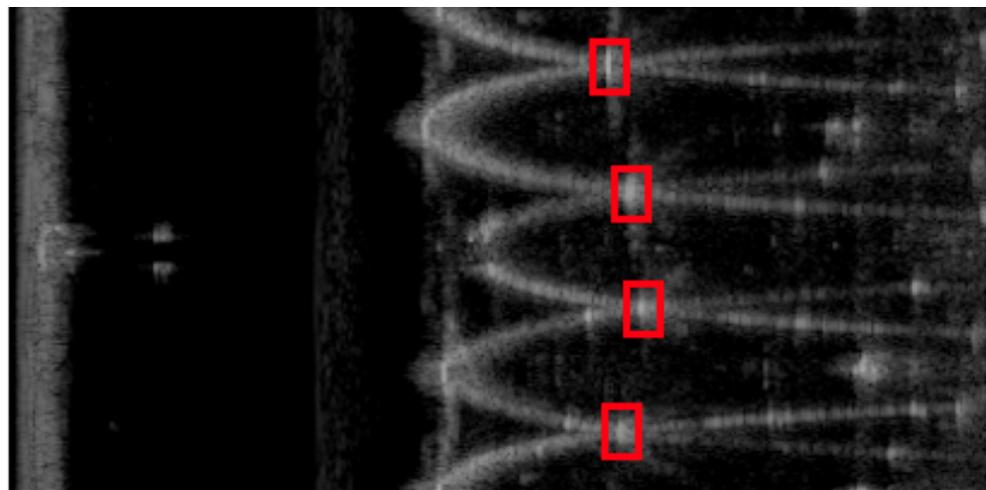

Figure 4. An example of corner-like structures' acoustic image representation (highlighted in red).

3. Feature-Based Image Analysis

The present section provides a brief overview of the algorithms considered and of some fundamental concepts.

3.1. Fundamental Concepts

The main goal of feature extraction algorithms is to retrieve key features from images that remain locally invariant, so that it is possible to repeatedly detect these despite viewpoint changes or in the presence of image rotation or changes on brightness and scale [2,10]. They encompass two main procedures: *Feature detection* and *feature description*.

Feature detectors enable the detection of feature-points (also called interest points or keypoints) in an image. These features typically take the form of corners, blobs, edges,

junctions, or lines. The form of the features depends on the feature detector employed [5]. The corresponding *feature descriptor* enables the feature effective recognition, which is of utmost importance for posterior matching.

In order to achieve the desired robustness to image transformation, additional image processing techniques are commonly implemented. To achieve invariance to scale (feature size), scale-space analysis procedures are implemented. These allow the application of feature detectors across the image and at different scale dimensions [10]. A common solution is based on the construction of a scale-space pyramid, whose layers consist of n octaves and n intra-octaves, formed by progressively half-sampling the original image. The feature detector is then applied on each octave and intra-octave. The occurrence of image noise or brightness changes is another common problem, resulting from image acquisition conditions. Image smoothing techniques are applied in order to reduce feature descriptor sensitivity to such problems, enhancing stability and repeatability. Furthermore, the application of orientation-normalized descriptors is a common solution used to attain rotation invariance, which is also key for overall robustness. Initially, a characteristic interest point direction is calculated and, subsequently, the feature descriptor is built taking into account such information.

Some feature extractors are made available with their proper feature detectors and descriptors, such as SURF, BRISK, or ORB. In other cases, feature detectors and feature descriptors are designed individually and are possible to be paired in several different combinations.

3.2. Algorithm Selection

A few works [2,3] have applied and compared the application of feature extraction algorithms for acoustic image analysis but the conclusions diverge, especially if their performances on optical images [5,6,8] are taken into account. As reported in the literature, for optical images, ORB (Oriented FAST and Rotated BRIEF) and BRISK (Binary Robust Invariant Scalable Keypoints) are better in repeatability and computational efficiency than SURF (Speeded-Up Robust Features) or SIFT (Scale Invariant Feature Transform) [5]. SIFT is the most accurate algorithm [3] while SURF is the most robust [2]. Nonetheless, when it comes to acoustic imaging, some of these premises do not hold. In this scenario, Harris and FAST (Features from Accelerated Segment Test) are able to extract the most number of keypoints [3] but present low matching ratios. SURF evidences the highest matching ratio and overall robustness to changes in rotation, scale and brightness [2]. BRISK returns the lesser number of keypoints [2]. Moreover, it has been noticed that, for acoustic images, the association between the same feature in different images is troublesome, due to the overall feature ambiguity.

Taking these initial considerations into account, the SURF, ORB, BRISK, and SURF-Harris feature extractors have been selected for the comparison carried out.

The SURF algorithm [11] is composed of a feature detector and a descriptor and it is inspired on the SIFT algorithm [12]. However, SURF procedures are simpler, resulting in a similar performance as SIFT but at significantly lower computation costs [11]. In [11], the authors have registered a decrease of 66% in the computation time required for the application of the SURF algorithm to a single image, when compared to the application of SIFT, while attaining similar performances for the number of matches performed. Furthermore, [2] has highlighted this algorithm as one of the most preferable for underwater feature based navigation. For these reasons, the SIFT algorithm, a reference algorithm for feature extraction, has not been considered for the purpose of this study.

The SURF-Harris feature extractor, proposed in [2], combines one of the most popular interest point detectors, the Harris Corner Detector [13], with SURF's descriptor, in an effort to conjugate the higher number of features detected by Harris with the robustness of SURF's descriptor. When applied to acoustic images retrieved from a sidescan sonar in [2], it has shown improvements of 63% on the number of features detected and of 33% on the number of features matched, comparatively to SURF.

The ORB algorithm [14] makes use of a combination of the FAST detector [15] and the BRIEF (Binary Robust Independent Elementary Features) descriptor [16], whilst proposing solutions to deal with the limitations of these methods, such as the lack of an orientation component in FAST or the impossibility of computing an oriented BRIEF descriptor [14]. The obtained algorithm shows lower computational costs than algorithms such as SURF or SIFT [5], a valuable characteristic for real time applications, such as underwater navigation. In [5], feature extraction conducted with the ORB algorithm is shown to require less than half the computation time required for the application of the SURF algorithm. Despite the good performance presented by ORB for optical image analysis, to the extent of our knowledge, the application of this algorithm to acoustic imaging has not been extensively addressed.

The BRISK algorithm [10] was developed with the objective of achieving high quality performance at lower computational costs than reference algorithms such as SURF or SIFT [10]. For that purpose, it makes use of the FAST detector and a binary descriptor. Both ORB and BRISK are identified in the literature as two of the most computationally efficient algorithms, a key characteristic for navigation purposes. Nevertheless, the BRISK algorithm is able to achieve similar computation costs to those characteristic of the ORB algorithm, while providing better accuracy for image scale and rotation variations than the latter, according to [5].

3.3. Feature Detection

In this section further detail on the feature detection procedure implemented by each of the selected algorithms is introduced. Additional methodologies employed for increased robustness are also presented. It is important to highlight that feature detection depends on the computation of a detection score, a metric resulting from the application of a feature detector, which enables evaluating if a candidate point corresponds to an interest point.

Regarding SURF's detector, it is based on the computation of the Hessian matrix. It allows improved performance in computation time and accuracy, since the computation of the determinant of the Hessian provides a simple metric for the determination of interest points. Despite being able to detect features of different forms, interest points are more commonly found at blob-like structures [11]. To improve the required computation time for the detection procedure, Gaussian second-order derivatives used in the computation of the Hessian matrix are approximated with box filters, as depicted in Figure 5. These are evaluated even faster by making use of integral images. An integral image $I_\Sigma(\mathbf{x})$ at location $\mathbf{x}$ is defined as the sum of all pixel values of the input image I in a rectangular region formed between the point $\mathbf{x}$ and image origin. The use of box filters and integral images allows a faster scale-space analysis, guaranteeing feature scale invariance without constructing a scale-space pyramid. Box filters are also employed for image smoothing, to achieve feature invariance to noise and brightness changes.

The SURF-Harris algorithm makes use of the Harris Corner Detector. As the name indicates, it aims at detecting corner-like features. This detection procedure is based on the local auto-correlation function of a signal. When applied to an image pixel $I(x, y)$, the auto-correlation function allows the analysis of the intensity structure of its local neighborhood. If the function is flat, that indicates that the region is of approximately constant intensity. If it is ridge shaped, that is indicative of an edge while a corner results in a sharply peaked auto-correlation function. The simplicity of this detection procedure results in a faster interest point evaluation, as initial tests performed have pointed out. During interest point analysis, image regions are successively smoothed with Gaussian filters. Nonetheless, the points retrieved are not scale invariant, since scale-space analysis is not performed.

The ORB interest point detection routine is focused on corner-like features. It is based on an initial computation of FAST interest points. This feature detection procedure is based on the analysis of the circular neighborhood of 16 pixels around each corner candidate. The procedure searches for sets of n contiguous pixels, with all intensities higher than that of the candidate pixel, I_p, plus a threshold t, or lower than I_p minus a threshold.

If any of these sets are found, the point is considered a corner point. The FAST detector is applied to each layer of the scale-space pyramid, in order to achieve invariance to scale. A posterior removal of interest points along edges is performed using a Harris corner measure. The application of this procedure makes ORB an order of magnitude faster than SURF for feature detection, according to [14]. Before the feature descriptor building, the original image is smoothed using Gaussian filters.
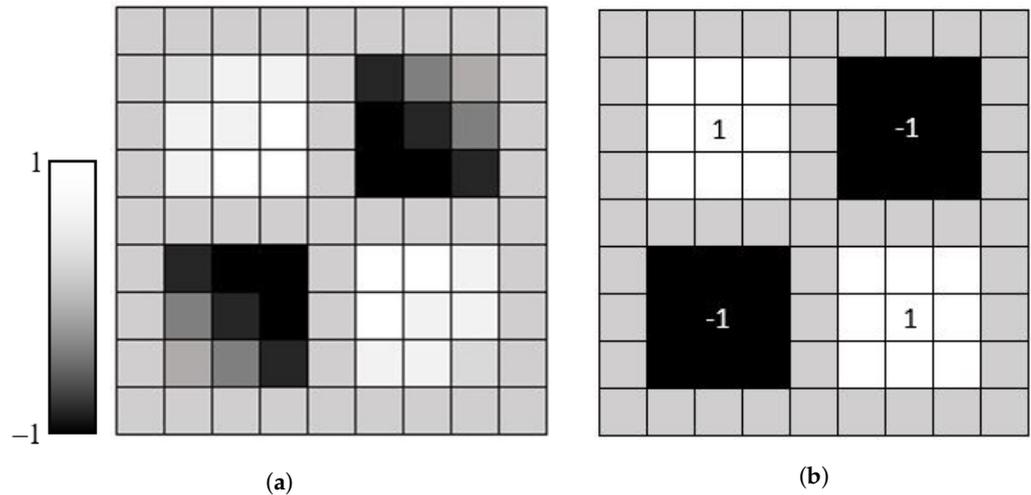

Figure 5. Illustration of box filter approximation. (**a**) Gaussian second order partial derivatives in the *xy*-direction and (**b**) approximation using box filters.

Similarly to the ORB detector, the BRISK feature detector employs an extension of the FAST detector, focusing on computation efficiency. With the aim of achieving invariance to scale, FAST is applied to each layer of the scale-space pyramid. The detected points are subjected to non-maximum suppression in scale-space. Gaussian filters are employed for image smoothing.

The previous details on feature detection are summarized in Table 1, hereafter.

Table 1. Characteristics of the selected feature detection procedures.

	Pre-Processing				Feature Detector		
	Image Smoothing		Space-Scale Analysis		Blobs	Corner	
Algorithm	Gaussian Filters	Box Filters	Scale-Space Pyramid	Integral Images and Box Filters	Hessian Matrix	FAST	Harris
SURF		✓		✓	✓		
SURF-Harris	✓						✓
ORB	✓		✓			✓	
BRISK	✓		✓			✓	

3.4. Feature Description

The present section introduces the feature description procedures implemented by each of the selected algorithms, along with additional methodologies employed for increased robustness.

The SURF descriptor implies two main operations, orientation calculation and descriptor construction. The orientation calculation is performed based on the horizontal and vertical intensity variations for each pixel within a circular neighborhood of the interest point. These variations are measured through the computation of the Haar Wavelet [17] responses in the *x* and *y* directions. The overall orientation is obtained through the sum of all individual responses. The descriptor is then built by firstly constructing a square region around the interest point, oriented according to the calculated orientation, as illustrated

in Figure 6. Haar Wavelet responses are calculated for several sub regions of this region, which will compose the feature descriptor. Note that this procedure is also employed by the SURF-Harris algorithm.
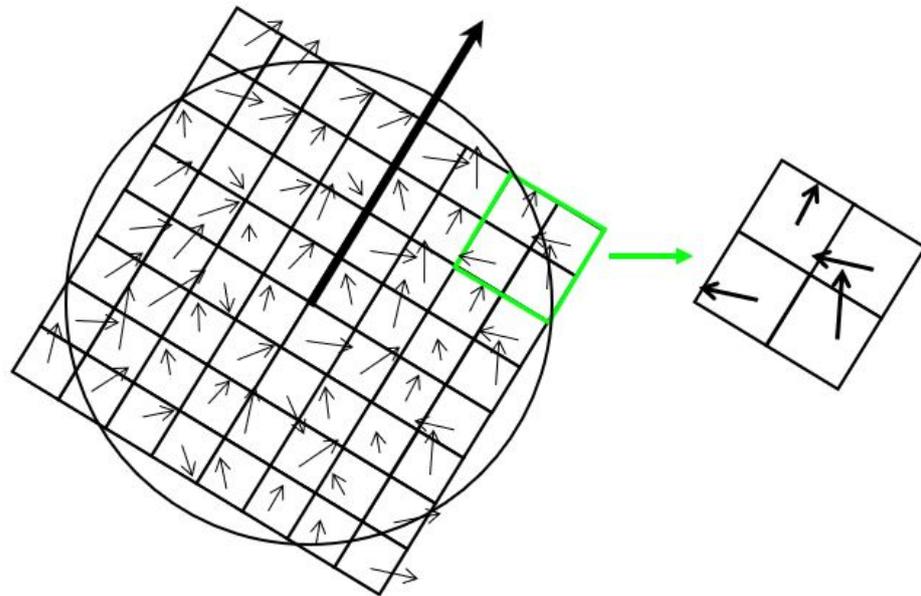

Figure 6. SURF (Speeded-Up Robust Features) interest point orientation calculation and sub region analysis [2]. Arrows indicate each sub region Haar Wavelet response.

The ORB algorithm makes use of the BRIEF feature descriptor. Nonetheless, this algorithm does not perform any interest point orientation calculation. The absence of the orientation component is solved by ORB through the calculation of the intensity centroid [18], which assumes that a corner's intensity is displaced from its center and allows the calculation of a characteristic orientation for each interest point. Regarding descriptor construction, the BRIEF descriptor is based on performing a series of binary intensity comparisons on the previously smoothed version of the original image. Such tests are performed between pixels in the neighborhood of the interest point. In order to achieve rotation invariance, a steered version of the BRIEF algorithm is employed by ORB. Here, each set S of n binary intensity tests performed at locations (x_i, y_i) is defined as:

$$S = \begin{pmatrix} x_1, ..., x_n \\ y_1, ..., y_n \end{pmatrix}. \tag{1}$$

Making use of the interest point orientation θ and the corresponding rotation matrix R_θ, a steered version of S, S_θ is defined as follows:

$$S_\theta = R_\theta S, \tag{2}$$

which defines the new rotation invariant test locations.

BRISK also makes use of a binary descriptor resulting from the concatenation of the results of several brightness comparison tests, a solution proven to be computationally efficient [10]. For the descriptor building, the algorithm starts off by sampling the neighborhood of each interest point based on an unique pattern (N locations defined equally spaced on circles concentric with the interest point). These points are used to compute local intensity gradients. With this information, a characteristic orientation of each interest point is calculated. The binary descriptor is afterwards built by re-sampling the interest point neighborhood with the initial sampling pattern oriented accordingly to the previously calculated orientation. Each bit of the descriptor is determined through intensity comparisons between pixels in the neighborhood of the interest point.

The information presented regarding feature description is hereafter summarized in Table 2.

Table 2. Characteristics of the selected feature description procedures. ORB: Oriented Features from Accelerated Segment Test and Rotated Binary Robust Independent Elementary Features; BRISK: Binary Robust Invariant Scalable Keypoints.

Algorithm	Descriptor Nature	Orientation Calculation Procedure
SURF	String based	Haar Wavelet response
SURF-Harris	String based	Haar Wavelet response
ORB	Binary	Intensity centroid
BRISK	Binary	Intensity gradient

3.5. Feature Matching

Feature extraction algorithms are a valuable tool for applications related to image analysis. However, in a number of them, such as pose estimation or object tracking, the detection of key feature points, by itself, does not provide sufficient information for accomplishing the intended goal. Therefore, this procedure must be complemented with a mechanism that enables the association of the same feature point in different images, taking advantage of the corresponding feature descriptor. Thus, usually, a feature matching procedure is associated to the application of feature extraction techniques.

Here, feature descriptors are compared and the points that evidence minimum deviations between their descriptors are taken as matches [2]. The matching procedure can be carried out according to different strategies, such as nearest neighbor, threshold based matching, or nearest neighbor distance ratio [5]. The comparison between a given feature point and match candidates feature points is performed based on the corresponding matching score. This metric rests on the computation of the L1 or L2 (Euclidean) norms, in the case of string based descriptors, or the Hamming distance, in the case of binary descriptors, between their corresponding descriptors. If a matching threshold is employed, a match is only retained if the corresponding matching score is lower than this threshold, as described in Figure 7. Otherwise, it is rejected.
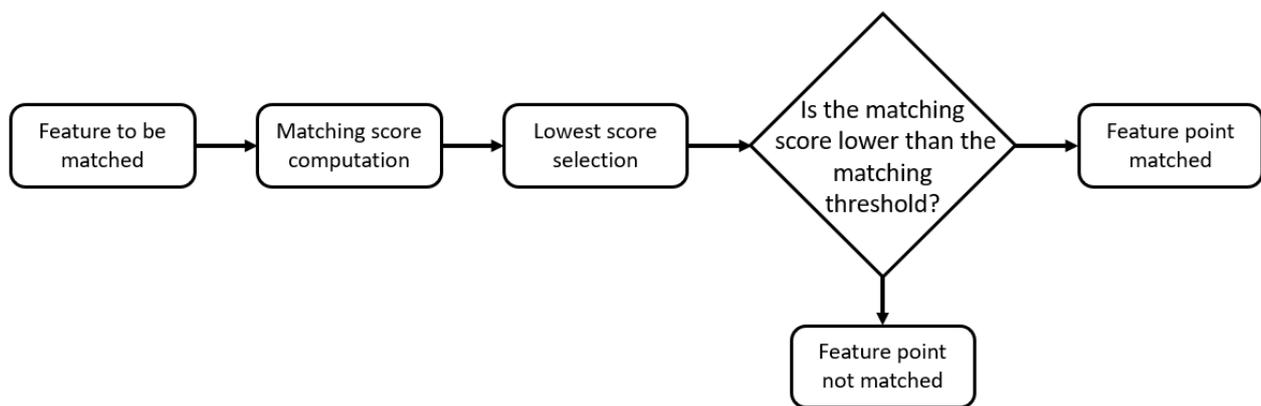

Figure 7. The feature matching procedure.

4. Comparison Methodology

This study was performed making use of MATLAB R2019a software, along with functions and tools available in the Image Processing Toolbox, for image preparation and feature extraction and matching. The specifications of the device where the required tests were performed on are presented in Table 3.

For the preparation of the datasets used, the SHAD AUV [19] was deployed, equipped with a Tritech Micron Sonar [20]. The sonar was mounted in the forward hull of the vehicle, oriented according to its longitudinal axis. The comparison performed comprised two

stages. An initial analysis was carried out taking into account the standard pipeline for visual odometry illustrated in the principle diagram in Figure 1. Afterwards, we focused our analysis on an adaptation of such architecture to the characteristics of the environments under study, through the incorporation of a feature selection module, as detailed in Figure 8, and consequent impact on the algorithms' performances. The remainder of this section details the employed methodology.

Table 3. Hardware specifications of the testing device.

Hardware Component	Specifications
CPU	Intel® Core™ i7-8565U Quad-core 1.80 GHz
GPU	NVIDIA GeForce MX250
RAM	8 GB

Acoustic Data Acquisition → Acoustic Image Composition → Image Pre-processing → Feature Extraction → Feature Selection → Feature Matching
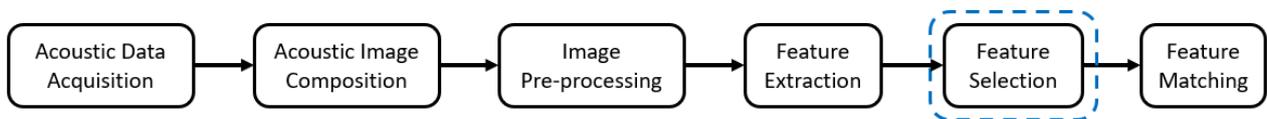
Figure 8. Inclusion of the filtering layer (highlighted in blue) into the image analysis procedure.

4.1. Datasets

Data surveys were carried out in a rectangular tank with the following dimensions: $4.6 \times 4.4 \times 1.8$ m (length × width × depth), using the SHAD AUV, both depicted in Figure 9. A floater was also deployed and moored close to one of the corners of the tank, so as to provide for a more distinguishable obstacle.
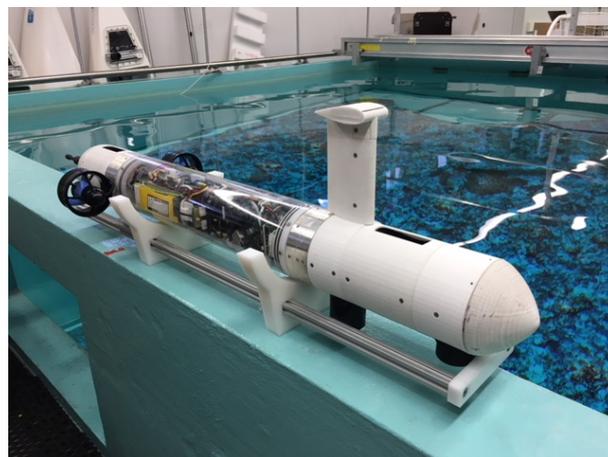

Figure 9. The SHAD AUV and the water tank structure used for testing procedures.

The presented testing scenario was chosen due to the associated complexity. Acoustic images retrieved from confined environments, such as a water tank, are generally more affected by acoustic echoes and multipath transmission, allowing a more thorough analysis of the algorithms in study. The diversity of features depicted in the acquired datasets is found to be representative of the class of environments under study, allowing to generalize conclusions drawn to other operation scenarios. Furthermore, working in a structured environment enables ground-truth localization measurements that are important for the proposed comparison.

Two datasets were acquired during this process, each one composed of a series of 360° acoustic scans of the AUV's surroundings. For *dataset*1, comprising a total of 8 acoustic scans, the AUV was kept immobile at surface, at position $(0.1, 0.4)$ of the tank, as illustrated in Figure 10a. For *dataset*2, comprising 11 acoustic scans, data collection was carried out with the vehicle moving horizontally at approximately constant velocity from position

$(0.1, 0.4)$ of the tank to position $(-1.7, 0.6)$, as detailed in Figure 10b. These scenarios were designed in order to better portray real mission situations and perform an analysis in light of the challenges described. Ground-truth measurements were performed regarding the AUV's initial and final positions and the floater's position, for both datasets.
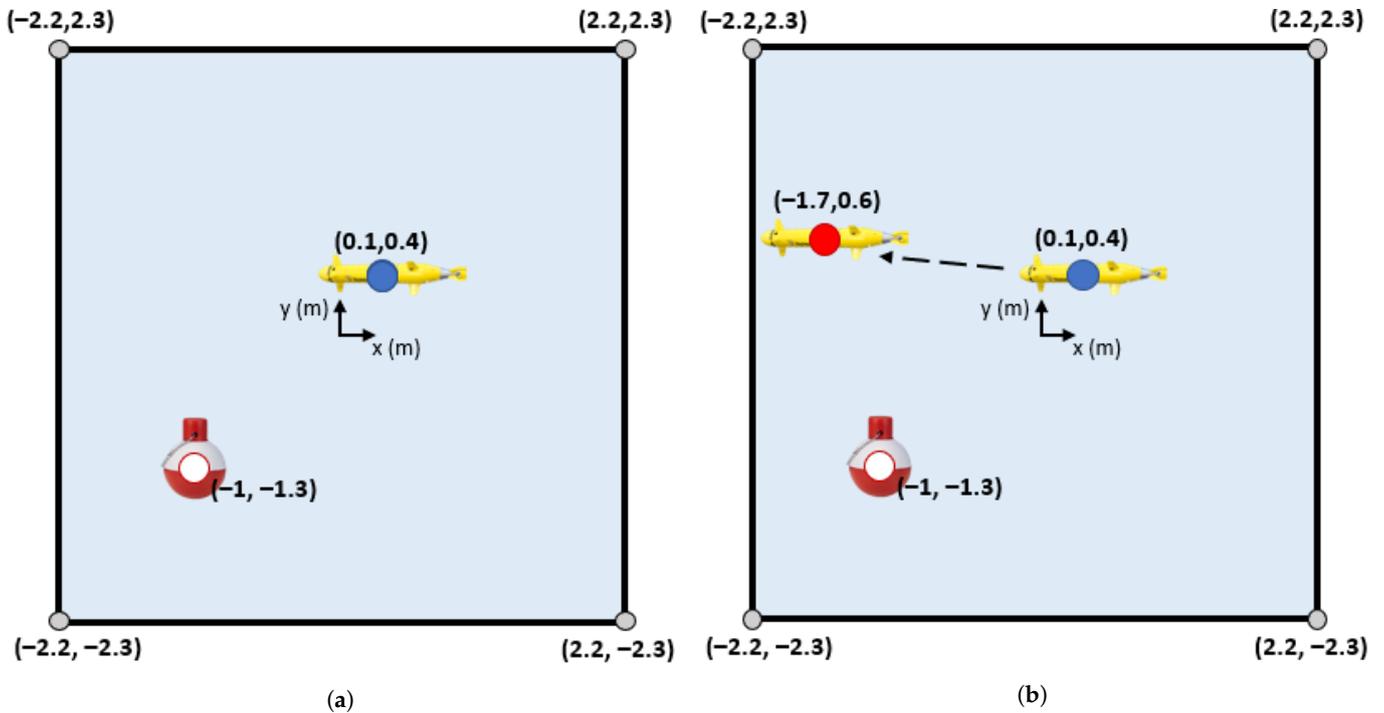

Figure 10. Illustration of the scenarios portrayed in each dataset. (**a**) Scenario for *dataset*1 and (**b**) scenario for *dataset*2, with initial (blue) and final (red) AUV positions marked.

These datasets consist of a series of intensity measurements, which are organized in the form of intensity arrays of the received acoustic signals. For each scanning angle α_i, an ordered set of m intensities values, represented by 8-bit integers, is generated [21]. The total number of measurements bins, m, returned by the sonar for each scanning angle is a function of the maximum range and the bin length for which the sonar is configured. Each scan is composed of a complete revolution of the sonar's head. For the purpose of this study, the sonar was configured as detailed in Table 4.

Table 4. Sonar parameter configuration for dataset construction.

Sonar Configuration Parameter	Assigned Value
Maximum Range	5.0 m
Bin Length	1.25×10^{-2} m
Angle Step	$1.8°$
Number of Bins (per intensity array)	399 bins
Number of Intensity Arrays (per acoustic image)	200 arrays
Scan Acquisition Time (per image)	$\sim$8 s

4.2. Acoustic Image Composition

By default, each intensity value is associated with a detection position, represented in polar coordinates through a scanning angle of α_i and a bin position of ρ_j. Therefore, an acoustic image I_{m_p}, as the one in Figure 11a, can be constructed by concatenating all the intensity arrays of a single scan as follows:

$$
I_{m_p} = \begin{bmatrix}
i(\alpha_1,\rho_1) & i(\alpha_1,\rho_2) & \cdots & i(\alpha_1,\rho_{399}) \\
i(\alpha_2,\rho_1) & i(\alpha_2,\rho_2) & \cdots & i(\alpha_2,\rho_{399}) \\
\vdots & \vdots & \cdots & \vdots \\
i(\alpha_{200},\rho_1) & i(\alpha_{200},\rho_2) & \cdots & i(\alpha_{200},\rho_{399})
\end{bmatrix}, \tag{3}
$$

where $i(\alpha_i,\rho_j)$ represents the intensity value at bin j of the intensity array for the α_i scanning angle.

Since the datasets used for the analysis performed are composed of raw acoustic data, the construction of the corresponding acoustic images can also be performed according to a cartesian coordinate system. An example of both image composition procedures is depicted in Figure 11.
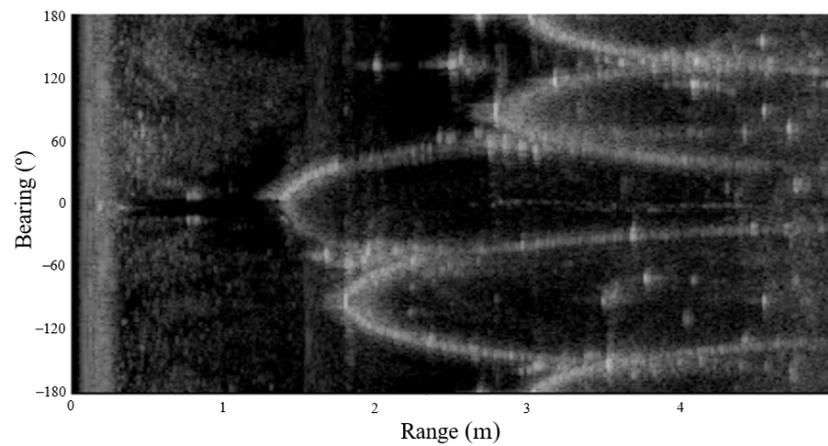

(a)
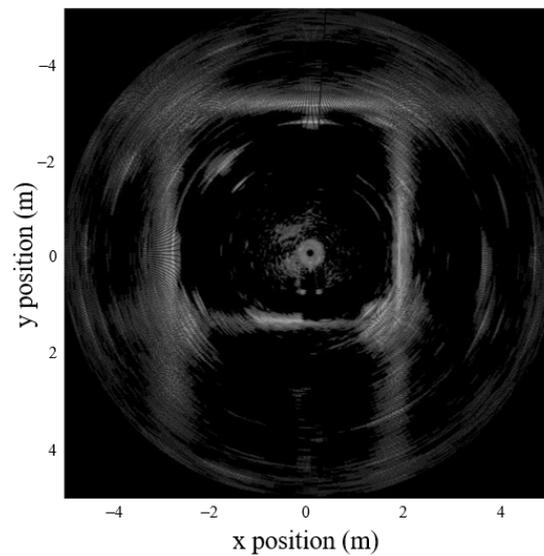
(b)

Figure 11. Results of different image composition techniques. (**a**) Image composed based on polar coordinate system and (**b**) image composed based on cartesian coordinate system.

This comparison was performed only taking into account the polar coordinate based image composition methodology. The reason for this decision comes from a brief analysis of both processes. The cartesian image composition algorithm is computationally more demanding and requires increased memory storage, mainly due to the resulting image size. Furthermore, as shown in Figure 11, the position from which data is acquired has a significant impact on the portrayed features in the resulting image, posing additional

challenges to the efficient performance of feature extraction algorithms. In the case of the cartesian image composition algorithm, this problem is more severe, since the loss of detection resolution with distance leads to increased image blur. Moreover, converting from polar to cartesian coordinates would introduce artifacts to interpolate bins in order to guarantee fixed-sized images. Such phenomena make these images even more affected by viewpoint changes and thus require further pre-processing.

4.3. Image Pre-Processing

The impact of image pre-processing was explored in [2], where it was found that the number of correct matches increases with the application of an image pre-processing routine based on thresholding and image smoothing operations. Thus, a similar image preparation step has been implemented.

The performed threshold operation targets lower intensity points. These are a common cause for the occurrence of false-positives. Note that, in the context of this study, a false-positive refers to any point that is classified as a feature-point but is in fact the result of noise or multipath effect. The value of the threshold applied was manually adjusted so as to promote a balance between the amount of background noise removed and features' structural integrity. The result of this operation is exemplified on Figure 12b.
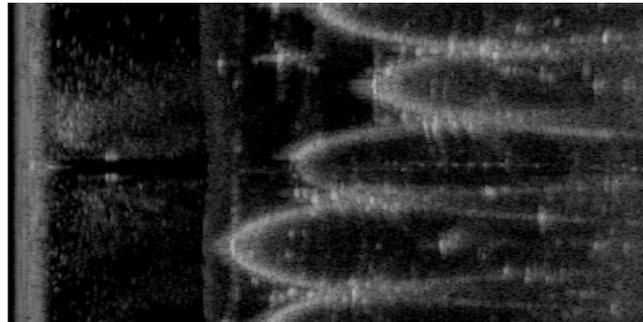

(**a**)
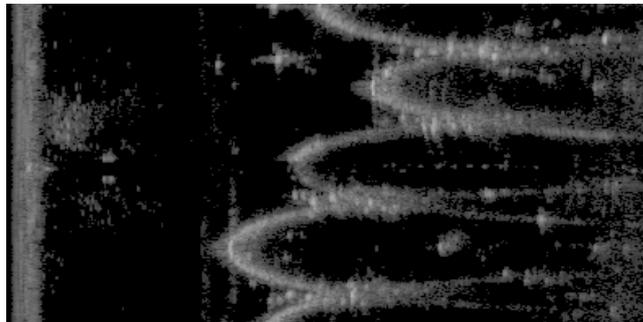
(**b**)
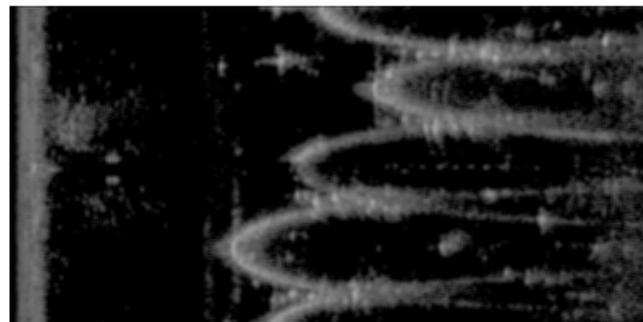
(**c**)

Figure 12. Intermediate results of image pre-processing routine applied to image 6 of *dataset*2. (**a**) Original image, (**b**) image after threshold operation, and (**c**) image after filtering operation.

The image is then passed through an average filter to remove the "salt and pepper" noise generated in the previous stage, as shown in Figure 12c. This procedure provides satisfying balance between noise removal capabilities and image blurring. Furthermore, the insertion of some degree of smoothing in the image contributes to an increasing stability and repeatability of feature descriptors [16].

4.4. Feature Extraction

After being pre-processed, the same image is passed through the feature extraction algorithms selected, in order to extract interest points and their respective descriptors.

Each of the studied feature detectors makes use of its own metric for classifying a point as an interest point or not, the detection score. However, contrary to what would be expected, there is not a clear relation between the detection threshold used and the number of false-positives that occur. This is due to the characteristics of acoustic images, since, despite being generated by secondary acoustic echoes or noise, false-positive points are very similar to interest points generated by environmental features, fulfilling the detection criteria. For this reason, feature detection accuracy, that is, the ability of the algorithm to distinguish an actual feature point from a false-positive, is crucial.

In order to achieve uniformity across all the methods compared, the detection threshold of each one was set so as to allow the detection of at least 50 interest points in each image of both datasets. The corresponding descriptors are calculated, stored, and passed to the next stage. For the purpose of comparison, the distinction between true interest points and false-positives was possible to be manually performed by using ground-truth positions and the dimensions of the tank, as depicted in Figure 13.
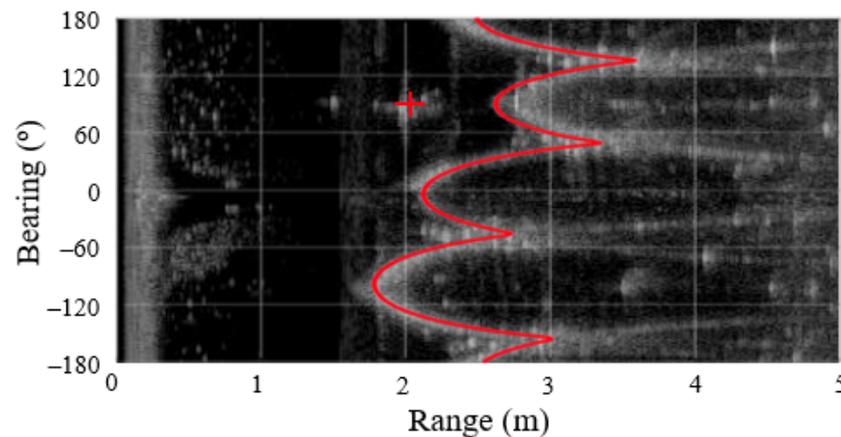

Figure 13. Overlap of acoustic image with ground-truth information, to define the boundary between regions of true-positives (left side) and false-positives (right side). The expected positions of the tank walls and floater are marked in red.

4.5. Feature Matching

For the purpose of feature-based navigation, the matching step is of key importance for the overall performance. Firstly, it is through this step that data which enables the application of visual odometry techniques is extrapolated. Secondly, since false-positive points are volatile, especially if the vehicle moves, it is expected that a vast majority of these are not possible to be matched throughout a collection of scans. Thus, the matching step acts as a filtering step, improving the confidence on the used data.

In this step, only the interest points detected in each pair of consecutive images are taken into account. For SURF and SURF-Harris, the sum of squared differences is used to perform the required comparison, since their respective descriptors are string based. In the case of BRISK and ORB, that make use of binary descriptors, feature comparison is performed through Hamming distance computation. These metrics compose the matching score and allow the application of a nearest neighbor search algorithm as a matching

strategy. A feature is only matched if its closest neighbor matching score is lower than this threshold. Note that, for the purpose of feature matching, the concept of neighborhood is related to the employed metric for descriptor comparison and not, as usually, to the feature's position in image coordinates. A match performed between two interest points referent to different features is considered an outlier, as portrayed in Figure 14.
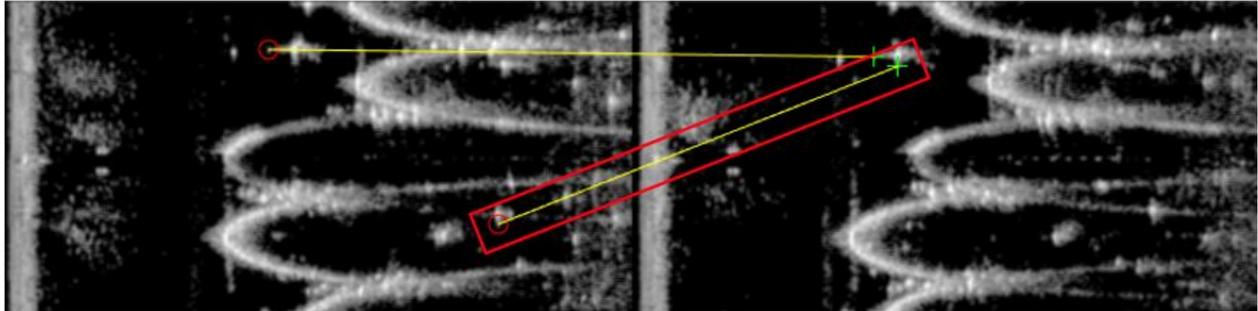

Figure 14. An example of a matching outlier, highlighted in red.

4.6. Feature Selection

In the second stage of the evaluation carried out, we have concentrated on adapting the above described methodology to the characteristics of acoustic imaging, in an effort to enhance the performance of the algorithms studied. As detailed in Section 2, in confined and structured environments not all the information reproduced in an image is of value, due to the presence of several artifacts resulting from multipath interference and secondary echoes. These are a common source of false-positives. To address such a challenge, a selection module aimed at removing these points has been developed and included into the architecture considered for the first stage of the study.

Taking into account the properties of acoustic waves, the information portrayed by acoustic echoes posterior to reflections on the closest obstacles is more likely to be affected by the problems initially described. Thus, to identify the corresponding image regions it is necessary to retrieve the acoustic echoes associated to the closest obstacles. The developed procedure is based on this premise and is further detailed hereafter:

1. For each intensity array composing each image, a five-level threshold procedure [22], based on the Otsu's method, is applied to segment the acoustic data into six different classes, minimizing intra-class variance;
2. The intensity array is analyzed again using the highest threshold previously defined. The goal is to identify the first intensity value higher than this threshold, which is expected to be associated to a reflection in the closest obstacle. The corresponding bin position is stored;
3. Using the retrieved bin position, plus a margin term, interest points resulting from acoustic information depicted in subsequent bins are removed. These are likely to be false-positive points.

The margin term was introduced in order to prevent relevant features from being erroneously disregarded. Its value was defined as the size of the larger detection kernel among the algorithms selected, in this case, the SURF algorithm. The application of this procedure to each intensity array of an acoustic image allows to define a feature rejection region, as illustrated in Figure 15. Interest points detected within this region are removed and feature matching proceedings are then applied to the filtered set of interest points.
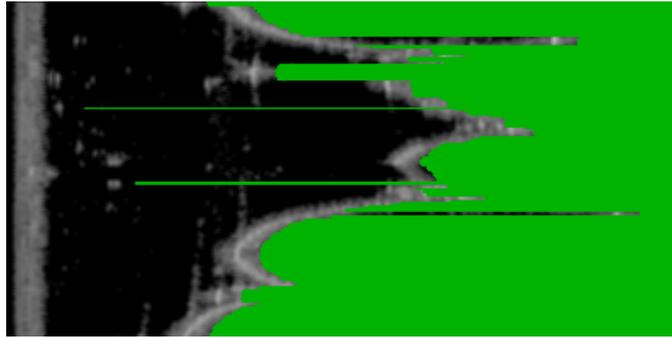
Figure 15. Illustration of the feature rejection region (in green) employed in image 9 of *dataset*2.

5. Results and Discussion

The four feature extraction algorithms are compared in light of extraction and matching. The impact of the feature selection algorithm is evaluated, considering that the incorrect matching of features or the correct matching of false-positives impact negatively. The analysis is conducted keeping in mind that the outcome of these algorithms should serve localization purposes.

5.1. Feature Detection Results

To assess the performance of the feature detection component, the total number of detected interest points and the number of false-positives were collected. To guarantee uniformity across the performed comparison, the measurement of the number of false-positives was only performed for the 50 interest points with higher detection scores extracted by each algorithm. The chosen value was determined experimentally and found to be sufficient for a sound representation of the algorithms' performance.

Figure 16 portrays an example of the obtained results for each of these algorithms. In an initial qualitative assessment it can be perceived that the distribution of the detected interest points varies according to the employed algorithm. SURF points are typically more scattered throughout the image. This is mainly due to the greater variety of feature shapes that SURF may be able to detect, along with its scale-space analysis methodology. In the case of ORB, BRISK, and SURF-Harris, the propensity to generating point clusters is notorious in the presence of strong features. For the purpose of localization, SURF presents a more attractive feature distribution for posterior pose estimation.

Regarding the total number of features detected through each method in each image, the results are provided in Figures 17 and 18, where it is possible to observe that ORB and BRISK are able to extract the most features in every single image, while SURF and SURF-Harris retrieve significantly less interest points. Figures 19 and 20 display the number of false-positives observed amongst the 50 interest points selected in each image. As it can be observed, ORB and SURF perform similarly, producing less false-positives than BRISK or SURF-Harris. This comes as a direct consequence of these algorithms tendency to detect a greater density of interest points in the same region.

The application of the feature selection module was focused on improving feature detection accuracy. Detection accuracy, besides having a direct impact on localization, also affects the matching procedure. More false-positives may lead to more outliers. As it is possible to observe in Figure 21, the module is successful in removing interest points detected beyond the limits of the water tank's walls representation.

Figures 22 and 23 helps in better understanding the significance of false-positives on the application of feature extraction to acoustic imaging. In comparison to the results collected during the first stage of the evaluation, the total number of detected interest points in each image of both datasets decreases to less than half the initial values. Nevertheless, the ORB and BRISK algorithms are still able to detect significantly more interest points than both the SURF and SURF-Harris algorithms.

The actual impact on the number of false-positives detected is portrayed in Figures 24 and 25. A decrease in the number of false-positives detected is evident, particularly in the case of the ORB algorithm. It was also observed that ORB provides enhanced robustness to background noise, a valuable characteristic for operation in confined environments. The motion of the AUV has a more relevant impact on the number of false-positives detected. This is owed to the occurrence of higher intensity secondary echoes whose contributions are wrongly preserved.


(**a**)

(**b**)

(**c**)

(**d**)
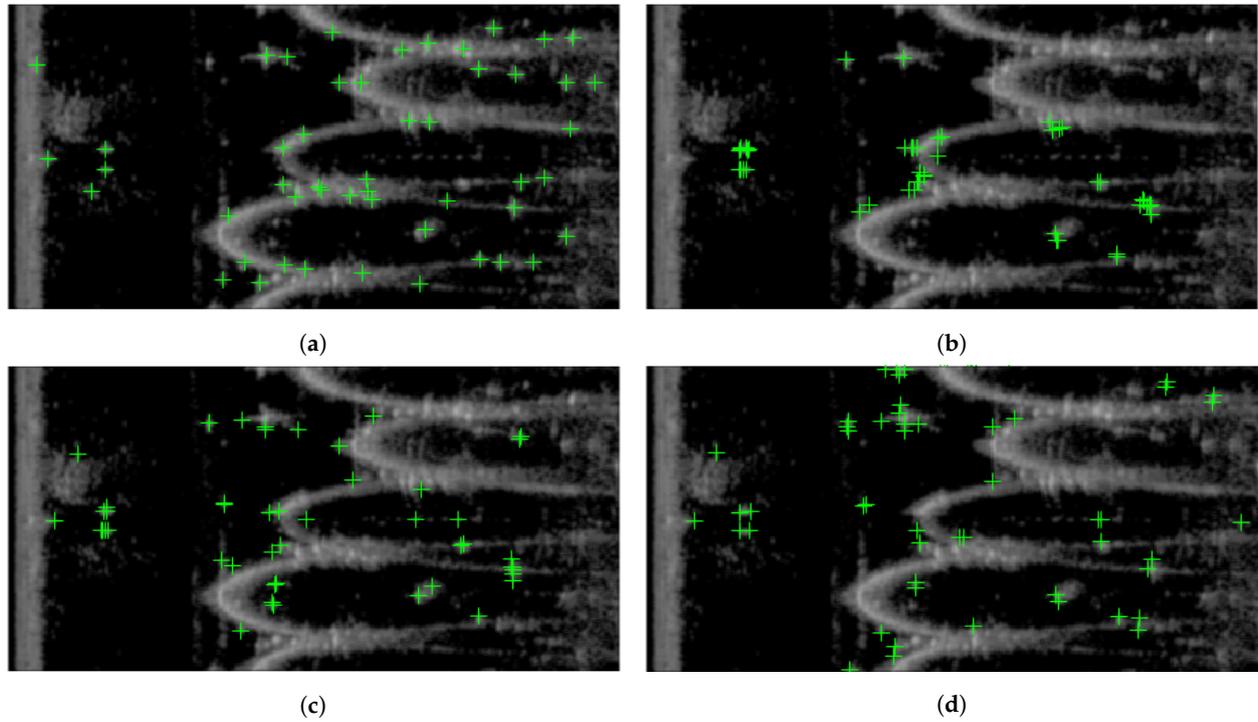
Figure 16. Illustration of the 50 best interest point detected in image 6 of dataset 2 with each algorithm under analysis. (**a**) Interest point detection with SURF algorithm, (**b**) interest point detection with ORB algorithm, (**c**) interest point detection with BRISK algorithm, and (**d**) interest point detection with SURF-Harris algorithm.
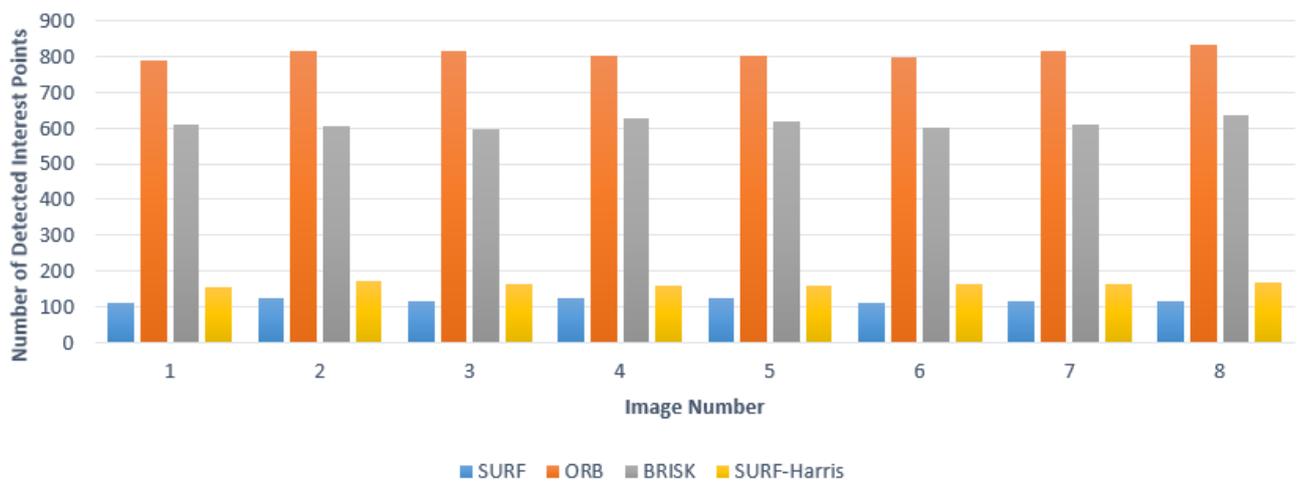
Figure 17. Total number of detected interest points by each algorithm in each image of *dataset*1.
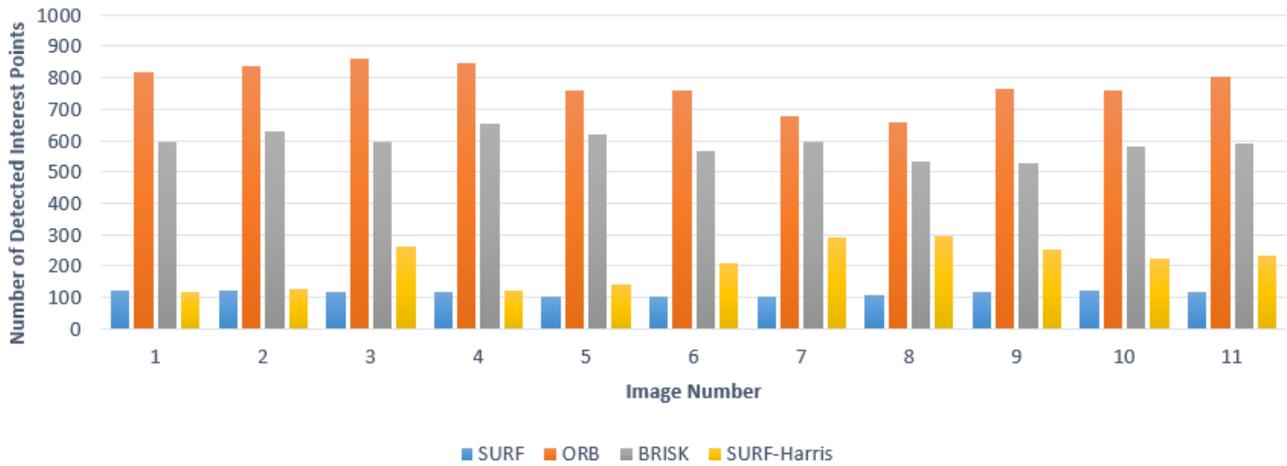
Figure 18. Total number of detected interest points by each algorithm in each image of *dataset*2.
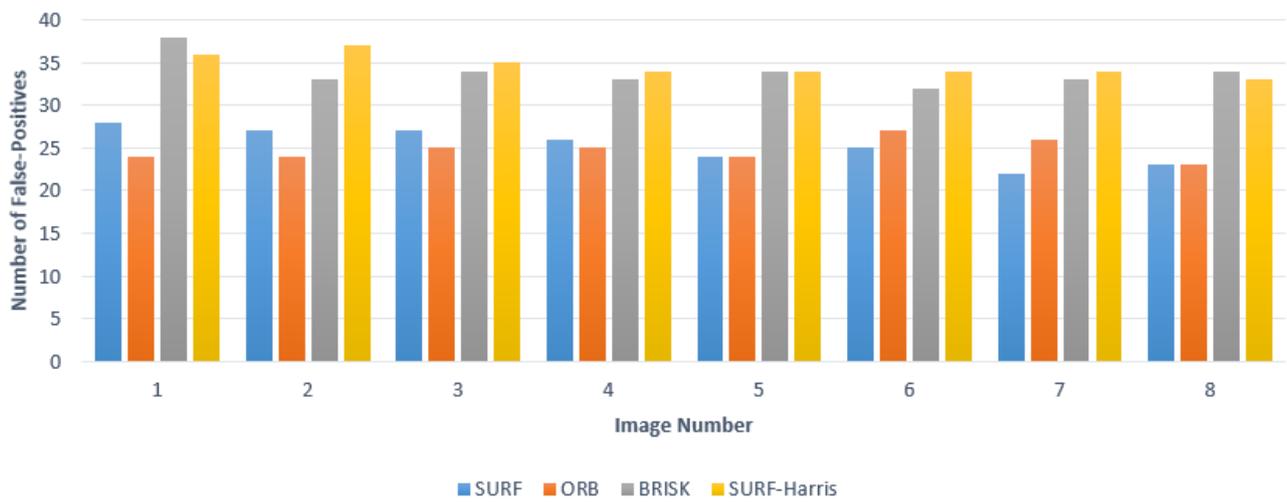
Figure 19. Number of false-positives detected amongst the 50 best interest points detected in each image of *dataset*1.
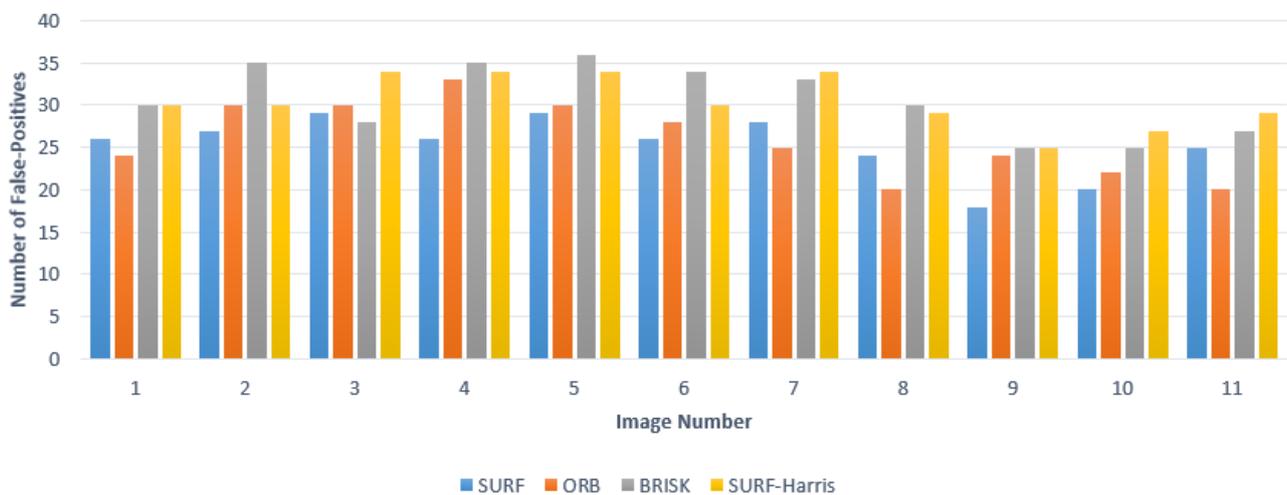
Figure 20. Number of false-positives detected amongst the 50 best interest points detected in each image of *dataset*2.
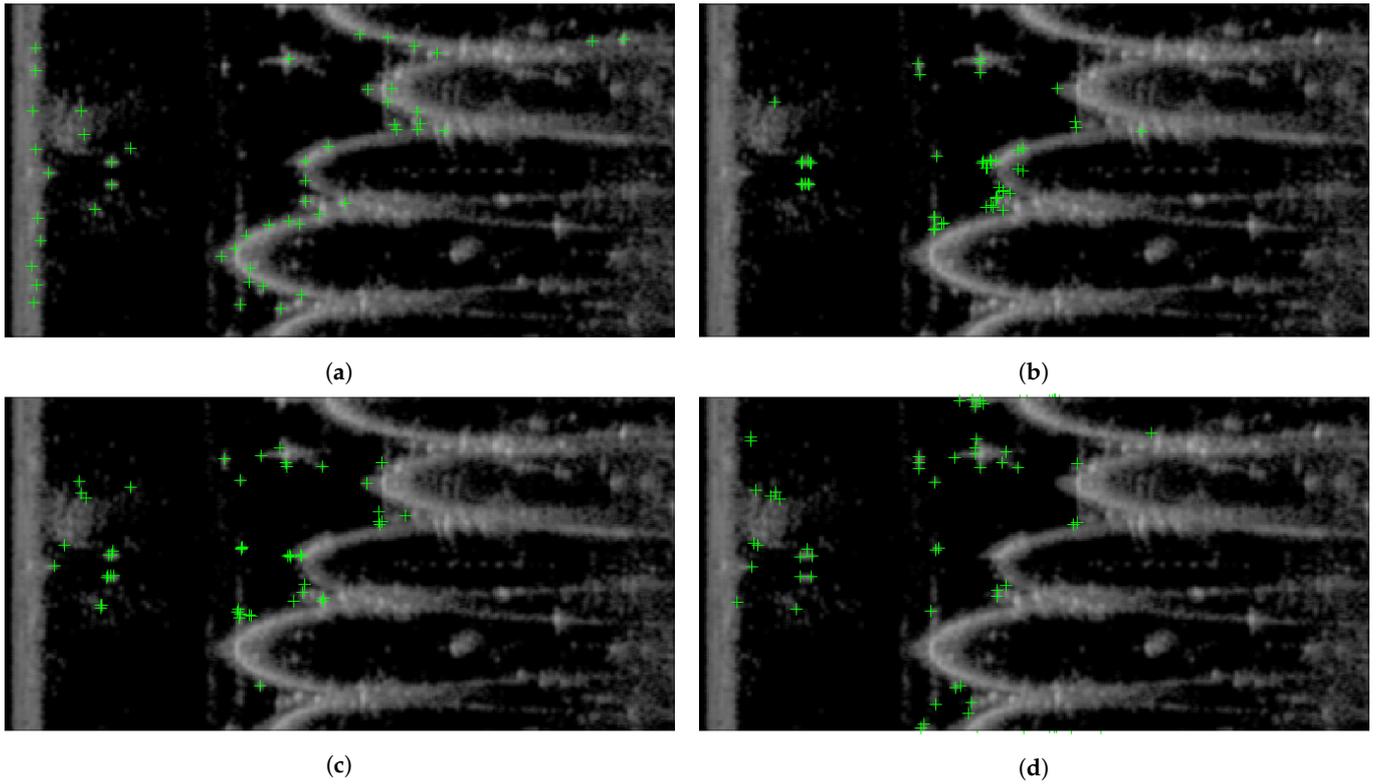
Figure 21. Illustration of the 50 best interest point detected in image 6 of dataset 2 with each algorithm under analysis employing the selection module. (**a**) Interest point detection with SURF algorithm, (**b**) interest point detection with ORB algorithm, (**c**) interest point detection with BRISK algorithm, and (**d**) interest point detection with SURF-Harris algorithm.
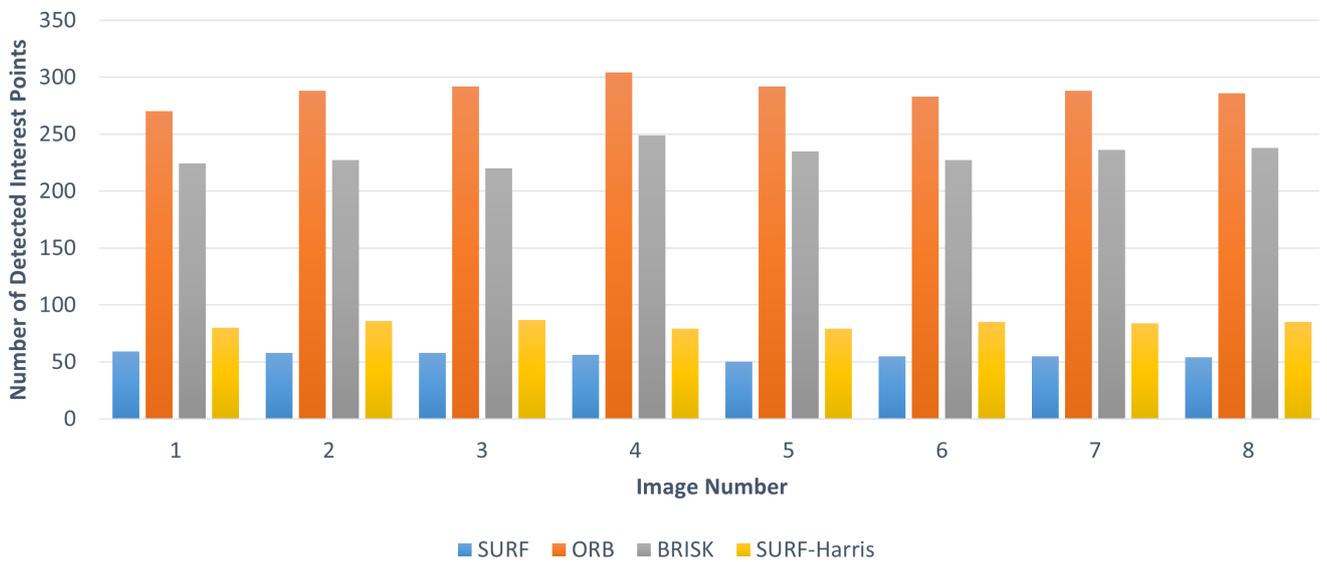

Figure 22. Total number of detected interest points in *dataset*1 employing the selection module.
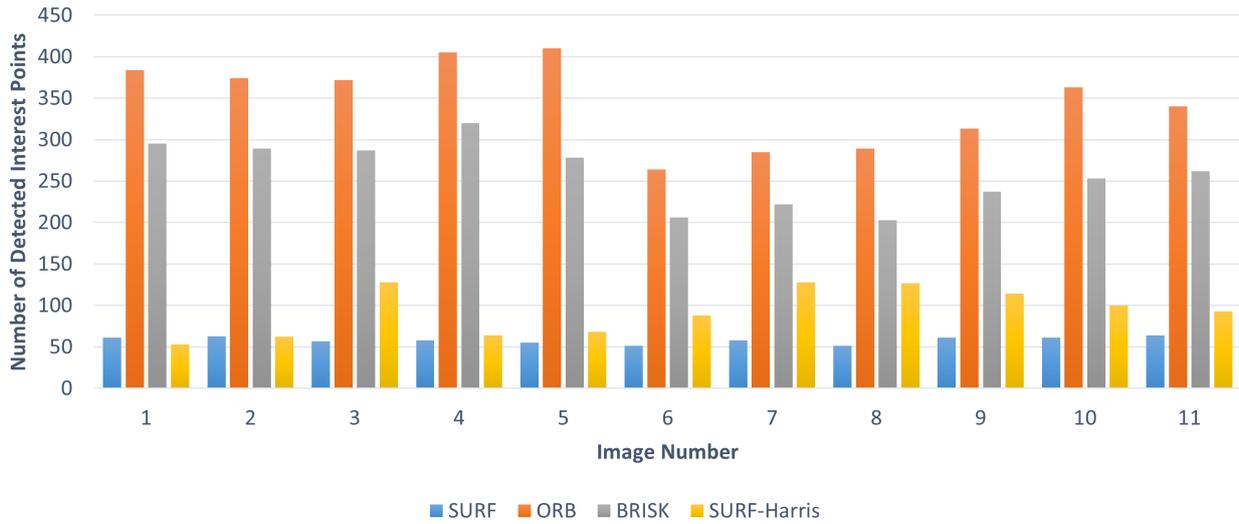
**Figure 23.** Total number of detected interest points in *dataset*2 employing the selection module.
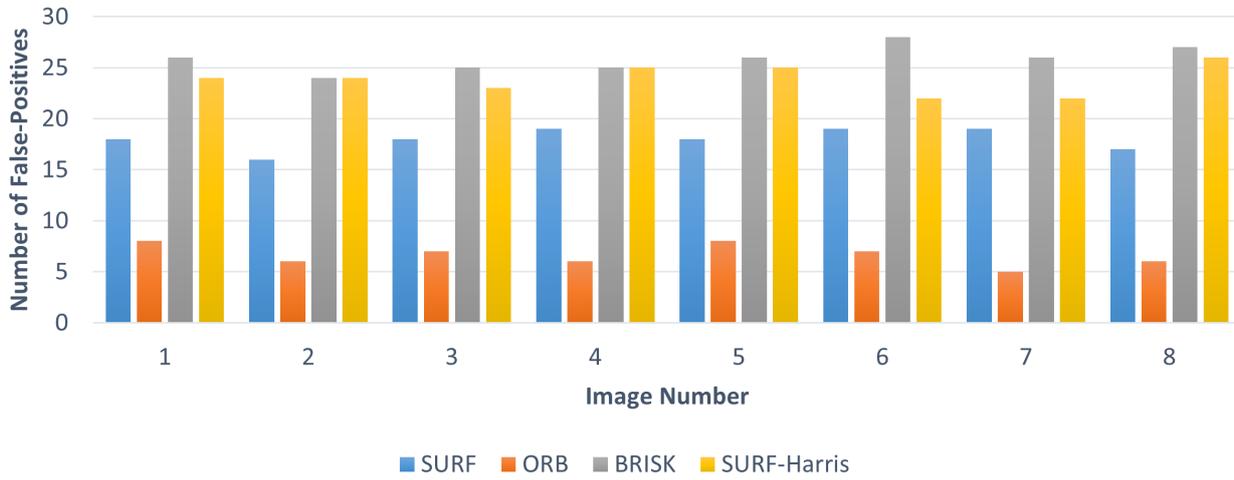
Figure 24. Number of false-positives detected amongst the 50 best interest points detected in each image of *dataset*1 employing the selection module.
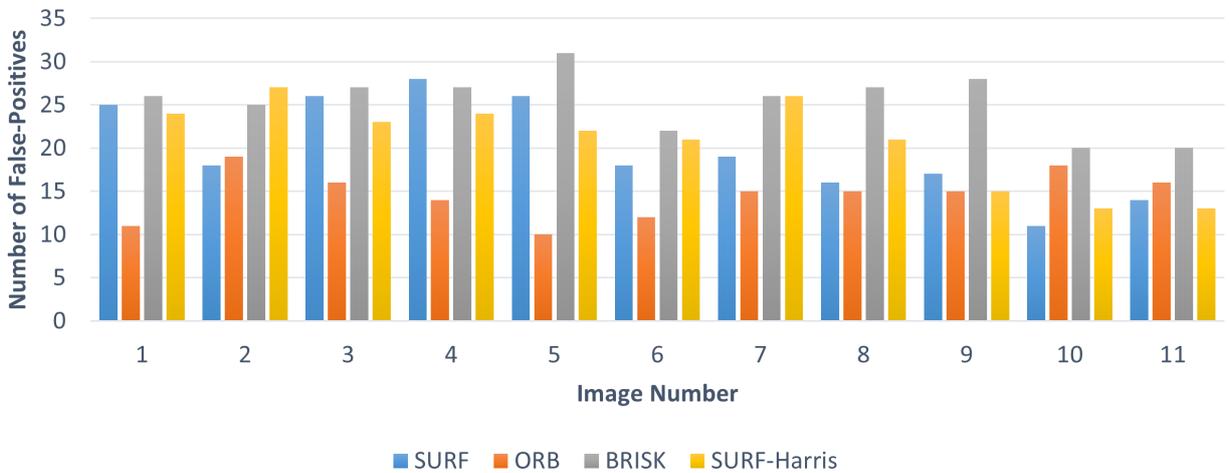
Figure 25. Number of false-positives detected amongst the 50 best interest points detected in each image of *dataset*2 employing the selection module.

5.2. Feature Matching Results

Feature description and matching are evaluated by analyzing the number of matches performed per image pair (including outliers) and the number of resulting outliers. Again only the 50 interest points with higher detection scores extracted in each image were taken into account.

Tables 5 and 6 display the number of matches performed per image pair, as well as the number of outliers detected. Sample results are shown in Figure 26. It is possible to conclude that the motion of the AUV has a great impact on the number of matches performed. A significant decrease in the number of matches is recorded, as well as a rise in the number of outliers, from the *dataset*1 results to *dataset*2 results. This conclusion is moreover supported by the strong decrease in matched features observable from image pair 2–3 to image pair 3–4. The reason for this is associated with the fact that image 2 is acquired while the AUV is just starting its motion, so feature deformation affecting this image is not so intense.

Table 5. Number of matches performed and corresponding number of outliers for *dataset*1.

	SURF		ORB		BRISK		SURF-Harris	
Image Pair	Matches	Outliers	Matches	Outliers	Matches	Outliers	Matches	Outliers
1–2	41	1	38	1	32	0	29	0
2–3	43	0	45	0	40	0	35	0
3–4	43	0	42	0	44	0	34	1
4–5	41	0	43	0	43	0	38	0
5–6	47	0	43	0	43	0	37	1
6–7	43	0	41	0	44	1	33	0
7–8	44	0	42	0	38	0	38	0

Table 6. Number of matches performed and corresponding number of outliers for *dataset*2.

	SURF		ORB		BRISK		SURF-Harris	
Image Pair	Matches	Outliers	Matches	Outliers	Matches	Outliers	Matches	Outliers
1–2	29	3	23	1	13	1	20	3
2–3	24	2	20	1	16	1	21	3
3–4	18	8	9	4	2	1	21	11
4–5	16	9	4	2	2	0	23	13
5–6	14	5	7	2	2	1	15	8
6–7	15	6	2	1	2	2	11	8
7–8	11	4	4	0	4	1	15	9
8–9	17	7	2	1	4	1	14	13
9–10	17	5	5	1	3	2	14	9
10–11	16	5	8	2	7	1	15	14

It is clear that both SURF and SURF-Harris are able to match far more features in every image than both BRISK and ORB. Yet, it is also possible to observe that both SURF and SURF-Harris produce higher numbers of outliers than BRISK and ORB. This fact may indicate that binary descriptors may be less adequate to the scenario in study but are able to provide information less affected by incorrect matches. String based descriptors may generate more matches but these are prone to incorrect associations.

The results obtained after application of the feature selection method are provided in Tables 7 and 8. By comparing this information, it can be observed that the decay in the number of matches performed and the rise in the number of outliers recorded from one scenario to the other is still notorious. So, it is possible to argue that the number of false-positives generated by the motion of the AUV does not severely limit the matching

performance. This issue highlights the need for further adaptation of the feature extraction algorithms to the characteristics of acoustic images.
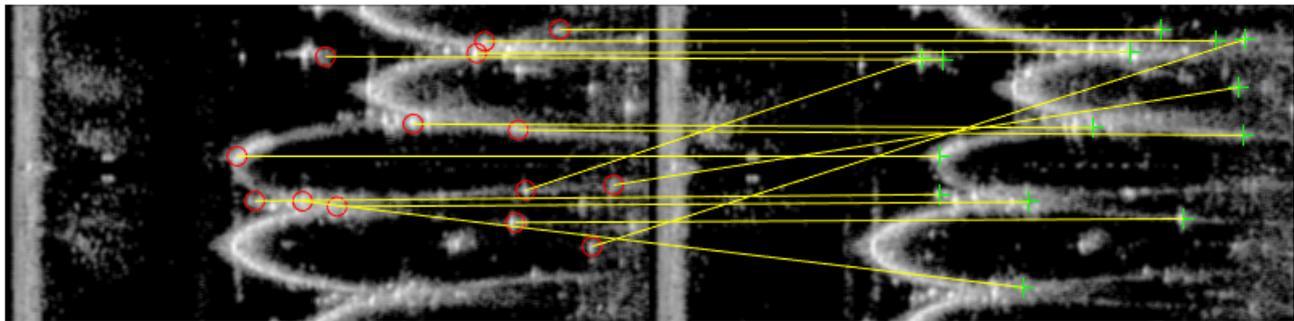
(a)
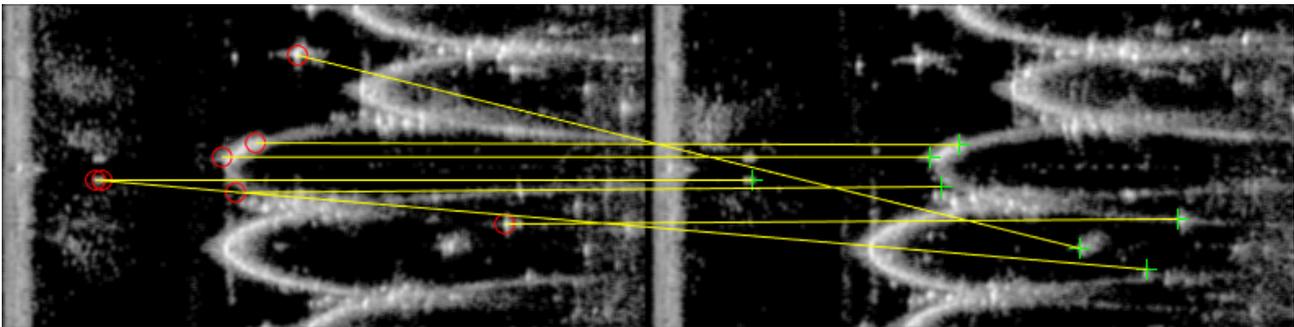
(b)
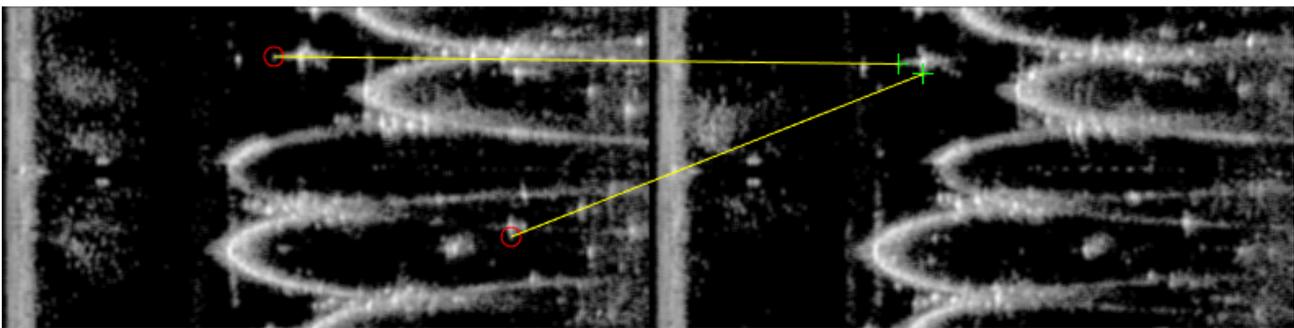
(c)
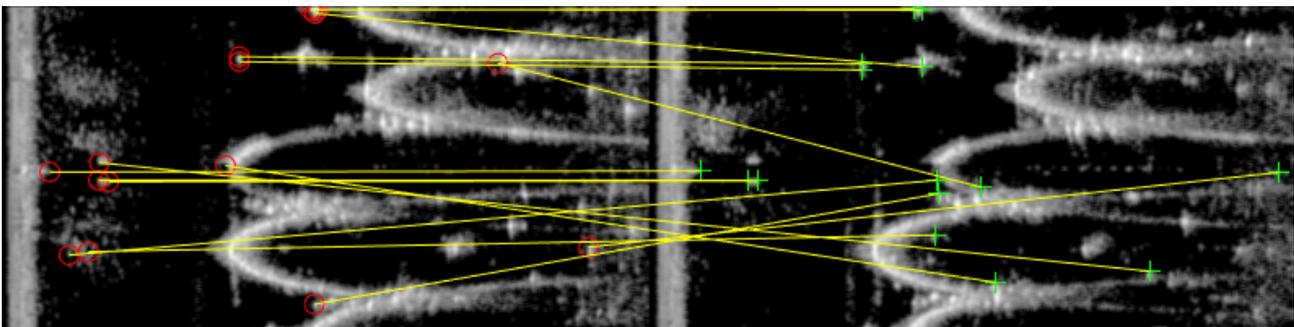
(d)

Figure 26. Feature points matched between images 5 and 6 of *dataset*2 with the algorithms under analysis. This illustration corresponds to the information portrayed on the third entry of Table 6. (**a**) Feature matching with SURF algorithm, (**b**) feature matching with ORB algorithm, (**c**) feature matching with BRISK algorithm, and (**d**) feature matching with SURF-Harris algorithm.

In the case of *dataset*1, it is observable that a reduction in the number of matches is performed by each algorithm. In contrast, for *dataset*2, ORB, BRISK, and SURF-Harris

algorithms evidence an increase in the number of matches performed. Furthermore, every algorithm exhibits an overall decrease in the number of outliers registered, especially ORB and BRISK. Figure 27 summarizes the effect of the feature selection module on the matching procedure. The improvement on the ORB algorithm's performance is notorious and strengthens the accuracy of its matching procedure. BRISK and SURF-Harris also show an overall improvement. SURF's performance does not benefit from false-positive removal due to a verified sensitivity to elements resulting from reflections on the AUV's body.

Table 7. Number of matches performed and corresponding number of outliers for *dataset*1 employing the selection module.

Image Pair	SURF		ORB		BRISK		SURF-Harris	
	Matches	Outliers	Matches	Outliers	Matches	Outliers	Matches	Outliers
1–2	38	1	41	1	31	1	30	0
2–3	41	1	44	0	38	0	33	1
3–4	40	0	44	0	43	0	36	1
4–5	41	2	41	1	39	0	36	0
5–6	39	1	40	0	38	0	36	1
6–7	41	1	44	0	43	0	32	2
7–8	43	1	39	0	39	0	35	0

Table 8. Number of matches performed and corresponding number of outliers for *dataset*2 employing the selection module.

| Image Pair	SURF		ORB		BRISK		SURF-Harris	
	Matches	Outliers	Matches	Outliers	Matches	Outliers	Matches	Outliers
1–2	24	2	25	2	13	0	22	6
2–3	18	2	23	0	18	0	26	7
3–4	20	5	11	1	5	2	16	6
4–5	14	10	6	1	4	1	17	10
5–6	15	9	5	1	4	1	16	9
6–7	12	2	6	1	4	0	16	11
7–8	12	3	6	1	1	0	14	8
8–9	19	8	1	1	4	1	15	10
9–10	14	2	8	1	2	0	17	9
10–11	14	3	9	3	5	0	14	11
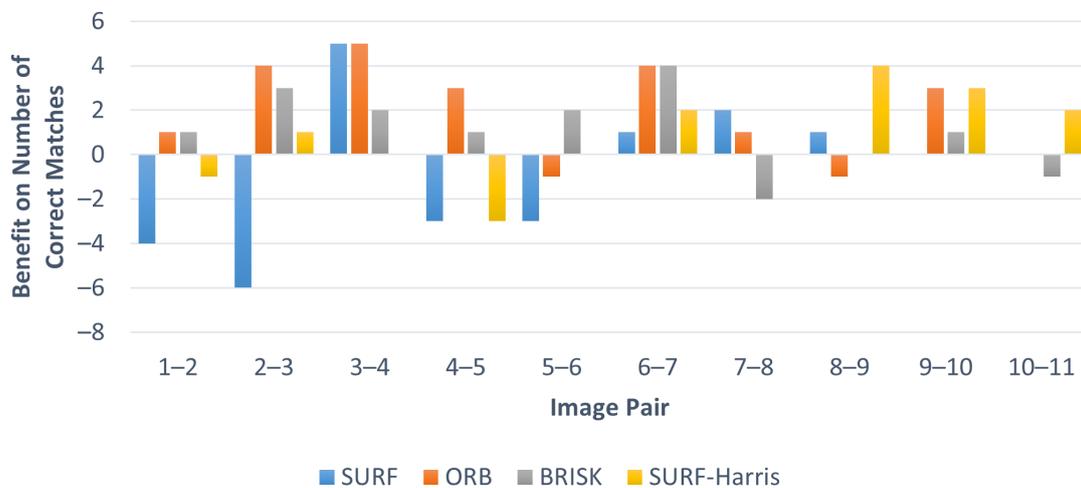
Figure 27. Improvement on the number of correct matches performed for the 50 best interest points for each image pair of *dataset*2 through feature selection.

5.3. Computation Time Results

The measurement of the computation time required for the feature extraction and feature matching routines was performed in an effort to better assess the real time applicability of the studied algorithms for localization purposes. The collected results are displayed in Tables 9 and 10. Note that, since these time measurements are presented for each consecutive image pair, the tabulated feature extraction computation time refers to the sum of the feature extraction computation time required for each image of the pair.

Table 9. Computation time measurements results performed for each algorithm for *dataset*1.

Image Pair	SURF			ORB			BRISK			SURF-Harris		
	Detection Time (s)	Matching Time (s)	Total Time (s)	Detection Time (s)	Matching Time (s)	Total Time (s)	Detection Time (s)	Matching Time (s)	Total Time (s)	Detection Time (s)	Matching Time (s)	Total Time (s)
1–2	0.141	0.005	0.146	0.020	0.004	0.024	0.571	0.004	0.574	0.033	0.004	0.037
2–3	0.143	0.003	0.146	0.019	0.004	0.022	0.581	0.003	0.584	0.045	0.054	0.100
3–4	0.156	0.004	0.160	0.032	0.006	0.037	0.590	0.002	0.592	0.082	0.007	0.090
4–5	0.167	0.003	0.1707	0.034	0.003	0.037	0.687	0.002	0.689	0.103	0.006	0.110
5–6	0.172	0.004	0.176	0.020	0.006	0.025	0.801	0.002	0.803	0.132	0.044	0.176
6–7	0.163	0.003	0.166	0.017	0.003	0.021	0.847	0.003	0.849	0.113	0.057	0.170
7–8	0.155	0.004	0.158	0.021	0.003	0.024	0.812	0.002	0.815	0.058	0.056	0.114
Average Total Time			0.160			0.027			0.701			0.114
Standard Deviation			0.011			0.007			0.120			0.048

Table 10. Computation time measurements results performed for each algorithm for *dataset*2.

Image Pair	SURF			ORB			BRISK			SURF-Harris		
	Detection Time (s)	Matching Time (s)	Total Time (s)	Detection Time (s)	Matching Time (s)	Total Time (s)	Detection Time (s)	Matching Time (s)	Total Time (s)	Detection Time (s)	Matching Time (s)	Total Time (s)
1–2	0.141	0.003	0.144	0.015	0.003	0.019	0.543	0.019	0.562	0.094	0.054	0.148
2–3	0.143	0.003	0.146	0.012	0.004	0.016	0.555	0.003	0.557	0.120	0.005	0.126
3–4	0.149	0.003	0.152	0.012	0.003	0.015	0.576	0.003	0.579	0.073	0.052	0.125
4–5	0.156	0.003	0.158	0.017	0.004	0.020	0.584	0.029	0.613	0.078	0.052	0.130
5–6	0.160	0.005	0.165	0.020	0.003	0.022	0.582	0.011	0.594	0.079	0.006	0.085
6–7	0.147	0.003	0.150	0.020	0.002	0.022	0.574	0.002	0.576	0.068	0.005	0.073
7–8	0.153	0.003	0.156	0.018	0.002	0.020	0.561	0.003	0.564	0.068	0.004	0.072
8–9	0.167	0.004	0.170	0.015	0.003	0.018	0.566	0.003	0.569	0.051	0.004	0.054
9–10	0.197	0.004	0.201	0.018	0.003	0.021	0.576	0.003	0.579	0.060	0.004	0.063
10–11	0.209	0.003	0.212	0.019	0.004	0.023	0.639	0.002	0.641	0.073	0.008	0.080
Average Total Time			0.165			0.020			0.583			0.096
Standard Deviation			0.023			0.003			0.026			0.033

It should be taken into account that the time measurements obtained are heavily dependent on the tools and computational capacity available, whose consequences are supposed to be evenly distributed over the results. A fairer comparison can be established on average times. These results reveal that ORB is the fastest algorithm, whereas BRISK is the slowest. SURF-Harris shows shorter computation times than SURF. These differences are mainly due to each algorithm's feature extraction computation speed. Furthermore, it is important to stress that both BRISK and SURF-Harris present a high variability on the required detection time, which can be inadequate for real-time applications.

6. Conclusions

In this article, a comprehensive performance evaluation of the SURF, ORB, SURF-Harris, and BRISK feature extraction algorithms applied to acoustic images was presented. The results obtained indicate that the characteristics of acoustic images put significant challenges to all these algorithms. The adaptation of feature extraction proceedings to acoustic imaging was explored and the proposed feature selection module produced performance improvements, especially in the case of the ORB algorithm. The feature extraction algorithms analyzed prove that these methodologies allow the identification and

matching of acoustic image features. Therefore, underwater vehicle localization through such information is possible.

Taking into account all the different aspects analyzed, we highlighted the characteristics of the SURF and ORB algorithms. These detected the lowest numbers of false-positives, valuable for localization purposes. SURF benefited from a better interest point distribution and was able to generate more matches. ORB presented lower computation requirements and a more robust matching procedure. The application of the feature selection module enabled an enhancement of the ORB's performance, making it a more interesting solution for confined environments. Nevertheless, SURF should not be disregarded as a possible solution, especially for open water environments. For such scenarios, SURF's detection procedure will benefit from the existence of scattered and isolated landmarks.

Future Work

The application of image pre-processing procedures have proven to be effective in reducing the number of incorrect matches performed. The literature presents several other methodologies for image preparation, which can be of great interest for this goal. We highlight the potential of contrast enhancement techniques, since the intensity difference between a candidate point and its neighborhood points is a key factor for classifying it as an interest point.

As it was possible to observe, the feature matching procedure presents itself as a troublesome operation in the context of acoustic imaging, especially when acoustic images are acquired with an underwater vehicle in motion, which leads to strong image deformation due to relatively slow scanning speeds. It is then possible to assume that the current information portrayed by the studied feature descriptors may be insufficient for acoustic image matching. Therefore, further research could be conducted in order to assess the possibility of extending existing feature descriptors with additional information. This purpose may be achieved more readily with string based descriptors, which further emphasizes SURF's fitness to acoustic image analysis.

Throughout this article we emphasized the impact of reflections on the AUV's body on detection accuracy. The resulting image elements are typically static and easily identifiable. The application of background subtraction techniques may provide an interesting tool to remove such artifacts.

Abbreviations

The following abbreviations are used in this manuscript:

AUV	Autonomous Underwater Vehicle
BRIEF	Binary Robust Independent Elementary Features
BRISK	Binary Robust Invariant Scalable Keypoints
FAST	Features from Accelerated Segment Test

MSIS	Mechanical Scanning Imaging Sonar
ORB	Oriented FAST and Rotated BRIEF
SIFT	Scale-Invariant Feature Transform
SURF	Speeded Up Robust Features

References

1. Sahoo, A.; Dwivedy, S.K.; Robi, P.S. Advancements in the field of autonomous underwater vehicle. *Ocean. Eng.* **2019**, *181*, 145–160. [CrossRef]
2. Dinuka, H.M.; Bandara, D. Sonar Image Based Advanced Feature Extraction and Feature Mapping Algorithm for Under-Ice AUV Navigation. Ph.D. Thesis, University of Tasmania, Hobart, Australia, 2017.
3. Tueller, P.; Kastner, R.; Diamant, R. A Comparison of Feature Detectors for Underwater Sonar Imagery. In Proceedings of the OCEANS 2018 MTS/IEEE Charleston, Charleston, SC, USA, 22–25 October 2018; pp. 1–6. [CrossRef]
4. Song, S.; Yu, S. Navigation Method of AUV to Target Point by Optical and Acoustic Sensor Switching Strategy. In Proceedings of the 2018 15th International Conference on Ubiquitous Robots (UR), Honolulu, HI, USA, 26–30 June 2018; pp. 769–774. [CrossRef]
5. Tareen, S.A.K.; Saleem, Z. A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. In Proceedings of the 2018 International Conference on Computing, Mathematics and Engineering Technologies: Invent, Innovate and Integrate for Socioeconomic Development, iCoMET 2018, Sukkur, Pakistan, 3–4 March 2018. [CrossRef]
6. Hidalgo, F.; Bräunl, T. Evaluation of Several Feature Detectors/Extractors on Underwater Images towards vSLAM. *Sensors* **2020**, *20*, 4343. [CrossRef] [PubMed]
7. Ferrera, M.; Moras, J.; Trouvé-Peloux, P.; Creuze, V. Real-Time Monocular Visual Odometry for Turbid and Dynamic Underwater Environments. *Sensors* **2019**, *19*, 687. [CrossRef] [PubMed]
8. Fraundorfer, F.; Scaramuzza, D. Visual odometry: Part II: Matching, robustness, optimization, and applications. *IEEE Robot. Autom. Mag.* **2012**, *19*, 78–90. [CrossRef]
9. Xiangyang, Z.; Jiaruo, H.; Lixiang, M. Image Representation of Acoustic Features for the Automatic Recognition of Underwater Noise Targets. In Proceedings of the 2012 Third Global Congress on Intelligent Systems, Wuhan, China, 6–8 November 2012; pp. 144–147. [CrossRef]
10. Leutenegger, S.; Chli, M.; Siegwart, R.Y. BRISK: Binary Robust Invariant Scalable Keypoints. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2548–2555. [CrossRef]
11. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [CrossRef]
12. Low, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
13. Derpanis, K. *The Harris Corner Detector*; York University: Toronto, ON, Canada, 2004; pp. 2–3.
14. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2564–2571. [CrossRef]
15. Rosten, E.; Porter, R.; Drummond, T. Faster and better: A machine learning approach to corner detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 105–119. [CrossRef] [PubMed]
16. Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P. BRIEF: Binary Robust Independent Elementary Features. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; 6314 LNCS; Springer: Berlin/Heidelberg, Germany, 2010; pp. 778–792.
17. Porwik, P.; Lisowska, A. The Haar-wavelet transform in digital image processing: Its status and achievements. *Mach. Graph. Vis.* **2005**, *13*, 79–98.
18. Rosin, P.L. Measuring Corner Properties. *Comput. Vis. Image Underst.* **1999**, *73*, 291–307. [CrossRef]
19. Goncalves, C.S.; Ferreira, B.M.; Matos, A.C. Design and development of SHAD—A Small Hovering AUV with Differential actuation. In Proceedings of the OCEANS 2016 MTS/IEEE Monterey, OCE 2016, Monterey, CA, USA, 19–23 September 2016; pp. 1–4. [CrossRef]
20. Tritech International. *Micron Sonar Product Hardware Manual*, 2nd ed.; Tritech International: Westhill, UK, 2020. Available online: https://www.tritech.co.uk/product/small-rov-mechanical-sector-scanning-sonar-tritech-microns (accessed on 13 May 2020).
21. Fula, J.P.; Ferreira, B.M.; Oliveira, A.J. AUV Self-localization in Structured Environments Using a Scanning Sonar and an Extended Kalman Filter. In Proceedings of the OCEANS 2018 MTS/IEEE Charleston, Charleston, SC, USA, 22–25 October 2018. [CrossRef]
22. Akay, B. A study on particle swarm optimization and artificial bee colony algorithms for multilevel thresholding. *Appl. Soft Comput. J.* **2013**, *13*, 3066–3091. [CrossRef]