A mosaicking technique for object identification in underwater environments

Alexandra Pereira Nunes, Ana Rita Silva Gaspar, Andry M. Pinto and Aníbal Castilho Matos
CRAS, Instituto de Engenharia de Sistemas e Computadores Tecnologia e Ciencia, Porto, Portugal

Abstract

Purpose – This paper aims to present a mosaicking method for underwater robotic applications, whose result can be provided to other perceptual systems for scene understanding such as real-time object recognition.

Design/methodology/approach – This method is called robust and large-scale mosaicking (ROLAMOS) and presents an efficient frame-to-frame motion estimation with outlier removal and consistency checking that maps large visual areas in high resolution. The visual mosaic of the sea-floor is created on-the-fly by a robust registration procedure that composes monocular observations and manages the computational resources. Moreover, the registration process of ROLAMOS aligns the observation to the existing mosaic.

Findings – A comprehensive set of experiments compares the performance of ROLAMOS to other similar approaches, using both data sets (publicly available) and live data obtained by a ROV operating in real scenes. The results demonstrate that ROLAMOS is adequate for mapping of sea-floor scenarios as it provides accurate information from the seabed, which is of extreme importance for autonomous robots surveying the environment that does not rely on specialized computers.

Originality/value – The ROLAMOS is suitable for robotic applications that require an online, robust and effective technique to reconstruct the underwater environment from only visual information.

Keywords Perception, Computer vision, Mosaicking, ROLAMOS, Underwater

Paper type Research paper

1. Introduction

Underwater research is very important for several scientific applications, such as in geology, biology, archaeology, shipwreck recoveries, environmental assessments and ecological studies. Most of the time, activities related to survey usually require a large amount of operating time and technical and human resources with a strong financial impact on small- and medium-sized enterprises (SMEs) of the marine industry. Nowadays, underwater robots are tools for several scientific areas including, geology, biology, archaeology, shipwreck recoveries, environmental assessments and ecological studies. These studies usually demand specific information from the seabed, which requires a large amount of technical and human resources to conduct the field missions. As a consequence, these missions have a strong financial impact on SMEs of the marine industry. In the past years, the use of autonomous underwater vehicles (AUVs) is growing to support the activities of these SMEs to accomplish tasks that include mapping and visual observation of marine structures. Remotely operated vehicles (ROVs) and AUVs can be equipped with diverse sensors such as acoustic positioning sensors (long or ultra-short baseline), Doppler velocity log, gyroscopes and global positioning system, whose data have the ability to improve the navigation of such vehicles. Some of these sensors are not suitable when a vehicle navigates near to sea-floor (or other underwater structures) as it does not provide a reliable nor accurate information. In those situations, visual systems formed by high-resolution cameras represent obvious advantages; however, they are limited by sub-sea conditions and phenomena related to the light propagation in water: poor visibility, light absorption, the presence of suspensoids in water and the backscattering effect. Although being affected by such limitations, the visual systems provide relevant information from the scene that can be used for the navigation of the vehicle or even for scene understanding.

This research presents an improvement of a method that conducts the visual mapping of the sea-floor, and it is called robust and large-scale mosaicking (ROLAMOS), whose preliminary version was presented in ICARSC 2016. The main contribution of this work is the online characterization of the underwater environment based on monocular images captured and processed in real-time from a ROV (or AUV) without a specialized computer. This algorithm performs an efficient registration of a sequence of monocular images by following a frame-to-frame analysis divided into four distinct phases: pre-processing, estimation of the egomotion, creation of the mosaic and memory management. The first stage (the pre-processing) intends to increase the quality of the input image by removing spatial-temporal noise. The egomotion estimation is obtained based

The current issue and full text archive of this journal is available on Emerald Insight at: **www.emeraldinsight.com/0260-2288.htm**

Sensor Review
39/3 (2019) 387–396

[DOI 10.1108/SR-04-2018-0089]

Received 1 December 2017
Revised 16 April 2018
5 September 2018
Accepted 19 September 2018

on principles from the structure from motion (SfM) and, therefore, the three-dimensional displacement of consecutive frames is obtained by extracting features from observations and computing their correspondences between consecutive views (by assuming the epipolar constraint). This work proposes a more efficient mechanism for outlier removal because this mechanism is balanced in terms of computational demands versus the quality of the stitching procedure (considering requirements related to high-resolution and a large panoramic image). Additionally, the egomotion is verified during the entire stitching process through a consistency checking procedure that infers about the feasibility of motion that was estimated previously. This approach for consistency checking is conducted just before blending the current observation with the panoramic image. The blending process of ROLAMOS creates a mosaic without visible seams. Finally, a mosaic management phase organizes the information of the entire of ROLAMOS and makes it possible to continuously operate with high-resolution images during the large-scale reconstructions. This phase manages the process of mosaic building by taking into account the hardware's limitations which is relevant for representing large-scale environment (with full resolution) that exceed the memory of the conventional computers available on board.

The proposed architecture makes this method suitable for the interpretation of the scene as the visual mosaic is created online and provides a more general overview of the sea-bed. This is quite relevant for other methods (computer vision or artificial intelligence) responsible for the scene understanding such as, objects recognition. Moreover, the visual map for the entire underwater scene can be reconstructed in offline with higher quality from the information kept by ROLAMOS.

Therefore, the main contributions of this article are:

- an enhanced version of technique (ROLAMOS) (Pinto *et al.*, 2016) that composes the sea-floor from sequential and overlapping visual observations;
- promotion of more efficient and advanced perception systems for underwater applications;
- robustification of the stitching process that creates an accurate panoramic image through a mechanism that detects and removes the influence of outliers (whose motion profile does not fit in the egomotion);
- an efficient approach based on SfM principles, which is suitable for applications that do not have specialized devices;
- a management system that controls the growth of panoramic image based on the highest resolution allowed by the hardware configuration;
- the enhancement of an online perceptual system for underwater autonomous vehicles that provides a textured-scene which can be used for recognizing objects; and
- an extensive set of qualitative and quantitative evaluations of the results of this research in several scenarios: using underwater robots operating in real scenes.

The article is organized as follows: Section 2 presents a state-of-art of the mosaic-based approaches that are used in robotic systems to perceive the environment. Section 3 shows the concept of the ROLAMOS method, where each phase is presented in details. Afterwards, Section 4 demonstrates the results of ROLAMOS that include experiments with sequences acquired from a ROV operating in a real scenario. The results indicate that ROLAMOS is suitable for online mapping of large-scale sea-floor because the visual information is registered and managed by taking into account the computational resources available onboard. Finally, Section 5 presents the most important conclusions of this research.

2. Related works

Dangerous and hazardous survey operations such as visual inspections, monitoring of sea structures and coral reefs and sea-floor mapping can be performed by underwater robotic vehicles. In some cases, these vehicles can reach places that are completely impossible for the human being (Gracias and Santos, 2000; Martin and Martin, 2002).

Nowadays, intelligent and autonomous robotic vehicles are developed with visual systems (Pinto *et al.*, 2014a, 2014b, 2014c) to obtain visual evidence of the sea-floor, and as a consequence, researchers have proposed some techniques to map large areas. Pizarro and Singh studied the impact of the reduced overlap between consecutive images (causing a radial distortion), unstructured motion and the poor lighting conditions. All these issues have a severe influence on the accuracy of the mapping process (Pizarro and Singh, 2003). Ferreira *et al.* (2013) proposed a real-time mosaicking based on a SLAM framework and applied to underwater image sequences. In addition to visual data, it uses other sensors such as lasers and acoustics. The authors claimed that this method is computationally efficient because it uses binary features (BRIEF) and local maps to provide scalability and a rapid construction of the mosaic. As noticed, the data from the navigation system of the robots can also be included in the mosaicking process to improve the quality of the final panoramic image (as well as to provide georeferenced data) (Ferrer *et al.*, 2007; Escartín *et al.*, 2008). A mechanism (Kim and Kweon, 2011) based on RANSAC was proposed to validate the consistency of the motion between consecutive observations. A tool for underwater large-area photo-mosaicking was presented by Marcon (2014). The work is focused on the quality, robustness and precision of the final map. It presents the possibility to intuitively create matches between unlinked images, several robust techniques to reject outliers and a global registration approach. Obviously, this tool was developed for offline image processing and cannot be embedded onboard of the vehicle. The Arnaubec *et al.* (2015) developed other software processing tool for semi-automatic, online or offline two-dimensional (2D) and three-dimensional (3D) optical mapping. To obtain the 2D mapping, the tool is optimized to process large-scale mosaics of many thousands of images. This is important to obtain a global representation of the seafloor without information about the vertical structures of the scene. In the other configuration, the 3D mapping is processed with diverse images taken from successive perspectives to obtain a highly precise bathymetry at small scale. The results of this technique indicate that both real-time and post-processing approaches are available. For the egomotion estimation, the work of Li *et al.* (2011) calculates the motion vector between consecutive frames from a simplified Lucas-Kanade method that is combined with an image registration technique (Lucas and Kanade, 1981). The estimation of the egomotion (affine model) is quite relevant for the warping and stitching stages. Bonny and Uddin (2016) presented a comparative study of four stitching techniques based on features (faster and more robust): SURF, FAST, Harris Corner Detector and MSER. The work emphasizes the importance of the amount of overlap between images and the minimization of the occurrence of the seam. From the results, the SURF achieved the best performance, and Harris Corners obtained a similar performance. The biggest disadvantage of this work is that it cannot be used online.

Currently, the limited visual perception capability in the underwater environment presents one of the greatest restrictions on the use of these type of vehicles. Therefore, it becomes crucial to develop perception methods that make underwater seafloor mapping possible whenever the vehicle moves closer to the bottom of the sea. These methods will allow a more efficient use of such systems in real environments and applications (Pinto *et al.*, 2014a, 2014b, 2014c). Maps generated by mosaicking methods can be used to object detection and identification, in particular, to discover sinked structures and to interpret the biological and physical characteristics of marine elements (Foresti and Gentili, 2002; Srividya and Shobha, 2014). A preliminary ROLAMOS version (Pinto *et al.*, 2016) proposed and tested with data sets publicly available. The preliminary algorithm is simple when compared to the enhanced version proposed by this research as it does not include alignment procedures and consistency checking. The proposed technique intends to decrease the gap between automatic processing of sensor data and missions that do not require the human intervention.

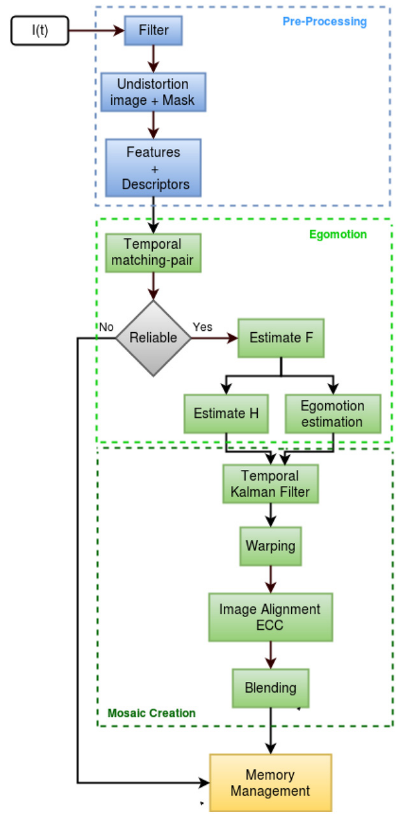
3. Robust and large-scale mosaicking

ROLAMOS is a mosaicking method formed by four main phases: a pre-processing stage, the egomotion estimation of the observer, a stitching stage and, simultaneously to the previous stages, a mosaic management mechanism ensures that the final panoramic image does not reach the full memory of the computer. The next sections present the concept and approaches followed in the architecture depicted in Figure 1.

3.1 Pre-Processing

In the pre-processing phase (a blue block on Figure 1) intends to remove spatial-temporal noise from the input image. This technique is named robust bilateral and temporal (Pinto *et al.*, 2014a, 2014b, 2014c) and uses the spatial and temporal evolution of the sequences to reduce the noise component and to make the images more appealing. It is quite common for the presence of suspensoids in the water column, which can be considered as noise component. Therefore, the quality of the texture information of moving observations is enhanced by applying this filtering method to rectified images (considering the calibration parameters of the camera) collected in underwater scenarios.

Figure 1 Diagram of the feature-based image mosaicking approach



3.2 Egomotion estimation

Creating a 2D map from a sequence of monocular images collected in underwater scenarios is a challenging task because the positions of the camera can only be determined up to one parameter of ambiguity. As a consequence, the absolute position (and orientation) is only estimated when additional information from the navigation system is available, for instance, the metric distance between two points of the scene (or the distance from the camera to a point). This ambiguity imposes that temporal correspondences are usually characterized by a line that represents all the possible locations for each point in other points-of-view: the epipolar constraint. Thus, principles of SfM provide clues about the 3D structure of objects through the analysis of the 2D motion from distinct perspectives over time.

This paper considers the egomotion as a rigid transformation that characterizes the displacement (position and angles) occurred between consecutive observations capturing different points-of-view.

The estimation of egomotion of the observer encompasses a couple of approaches:

- *Direct methods*: Calculate the transformation by maximizing the photometric consistency over the whole overlapping image regions (Gutiérrez, 2013). This type of approach is suitable for image sequences having large overlapping regions.
- *Feature-based methods*: Compute the transformation between points-of-view using a set of points in images and a temporal association.

ROLAMOS follows a feature-based approach, which means that the extraction and the matching of features are performed between consecutive frames. The SURF feature description (Bay *et al.*, 2008) is used as it is invariant to scale and rotation (relevant characteristics for robotic applications operating in underwater scenes). After that, the temporal inconsistency of pairwise features is estimated by the epipolar geometry, which means that the feature representation in the second point-of-view must be located closely to the epipolar line.

The fundamental matrix provides a geometric relation between two images of the same scene acquired by different points-of-view. This matrix is estimated by using the epipolar geometry supported by a RANSAC-based approach. This approach makes it possible to detect wrong matches, each one defined as a pairwise feature with a large bilateral distance. These points are usually known as outliers and should not be considered for motion estimation, see equation (1). This process uses a cross-matching approach (with the k-nearest neighbor algorithm) for filtering pairwise features having a large matching distance.

$$err^{i}_{BL} = err^{i}_{f} + err^{i}_{p} \tag{1}$$

where err^{i}_{f} is the distance between the ith feature f to its epipolar line (defined by the p and the fundamental matrix). The sum of the bilateral error from all pairwise features gives an overall measure of the quality that can be expected for the egomotion estimation.

After consistency checking phase, the egomotion is obtained by decomposing the homography matrix (H) into rotation and translation (R and t) matrices (Nister, 2004). The homography is a transformation that maps points of the first image into the corresponding points on the second image. For the sea-floor scenario, the assumption that the scene is roughly flat can be considered without losing generalization. This assumption is necessary to estimate the homography (known as the co-planarity constraint). An singular-value decomposition (SVD) of the homography matrix retrieves some hypothesis for R and t, see equations (2)-(4):

$$R = UR_{\Lambda}V^{T} \tag{2}$$

$$t = Ut_{\Lambda} \tag{3}$$

$$n = Vn_{\Lambda} \tag{4}$$

where U and V are orthogonal matrices and Λ a diagonal matrix, which contains the singular values of matrix H. Up to eight different solutions are obtained as a result of the decomposition algorithm for the triplets: R_{Λ}, t_{Λ}, n_{Λ}. Impossible combinations of R and t are further eliminated by additional constraints related to the depth.

This egomotion estimation process (Figure 1) assumes that the dominant motion is expressed by a rigid transformation based on the set of pairwise features that are obtained during the computation of the fundamental matrix.

3.3 Stitching

The stitching process implies that the current image is projected by taking into account the homography matrix. This step is also known as warping, see Figure 1 (dark green blocks). Although ROLAMOS increases the robustness of the stitching process related to the quality of the egomotion estimation, by removing pairwise features that do not share a consistent motion profile, there are situations where the resulting pairwise features could be incoherent (due to the parallax effect, lack of good features and non-distribution of the features along the frame). In these cases, the warping process will produce unreliable results that can compromise the creation of the final panorama. To detect these situations, ROLAMOS incorporates an additional stage for outlier removal that is based on the empirical analysis of the egomotion and the homography estimative. A Kalman filter (with a model of constant acceleration) ensures that no sudden transitions are obtained between consecutive frames. Additionally, it is assumed that images are acquired sequentially within short periods of time (the aliasing constraint). As an example, sudden changes such as a turn of 180° or a large translation will not be admitted. If some of the previous conditions are detected, the stitching process of the current observation will not be concluded.

The alignment between two images with different viewpoints is estimated. This research uses the enhanced correlation coefficient (ECC) Maximization (Elibol *et al.*, 2013) is used to improve the quality of the warped image, which constitutes an additional alignment measure that is robust against geometric and photometric distortions. This additional registration layer (the ECC) uses the gradient information and achieves high accuracy (sub-pixel estimation) that is invariant to illumination changes due to the zero-mean normalized cross-correlation of an objective function. This layer refines the egomotion estimation as well as the registration process. Consequently, it increases the quality of the final panorama.

The final step of this process is the image blending, which combines two or more images and creates a panorama which enables the visual interpretation of the sea-floor and existing structures. Thus, this step takes into account other misalignments through the use of spatially-varying weights (feathering). A 2D Gaussian distribution centers the weights that create a gradient-like matrix for both images (new warped and panoramic image). This function determines how this new warped image will blend with the panoramic image in a computationally efficient management. More sophisticated blending functions are available in the literature; however, they are computational demanding for online processing. So, the blending stage makes it is possible to obtain a visually appealing panoramic image.

3.4 Memory management

The ROLAMOS is a technique mainly developed for AUV/ROV applications as it creates a robust and accurate panoramic image. When the AUV operates in large-scale environments, it has limited computation resources available onboard for processing the visual information.

To mitigate the limitation related to the large amount of flash memory (RAM) that is needed to create mosaics, ROLAMOS proposes a hierarchical structure to manage the computational resources available onboard, see the yellow blocks in Figure 1. The proposed memory management stage guarantees that the ROLAMOS does not jeopardizes the situations awareness module that is also running inside the AUV (e.g. for object recognition). Therefore, the first goal of the memory management

Sensor Review
Volume 39 · Number 3 · 2019 · 387–396

is to control the size of the panoramic image depending on the RAM (random-access memory) available over time. The mosaic image is limited in terms of the maximum size and the stitching pipeline halts when the egomotion of the new observation increases the size of the mosaic to beyond the safe threshold. In those situations, the current mosaic image and complementary data are saved in the disk as a "patch" and, a new mosaic can be started safely. A "patch" is formed by a local portion of the panoramic image (composition of sequentially blended images) that does not exceed the maximum number of pixels and also by additional information that supports the reference frame, see Table I.

Table I summarizes the additional information: egomotion, the corner's locations and homographies of images. Based on this information, a hierarchical scheme can be created where each "patch" is referred to a "global" frame for further reconstruction.

The next step is restarting the process of mosaic creation, more specifically, starting a new patch that will comprise a set of new warped images. For an offline analysis (after the mission to be concluded): the global mosaic of large-scale environments cannot be generated from the patches at full resolution as it probably will exceed the maximum number of pixels allowed by hardware (for conventional systems). Thus, this method proposes a solution to scale down the global mosaic due to a scaling factor that is calculated to fit all patches into the global mosaic frame without exceeding the maximum pixel limit. In this way, it is possible to visualize the entire panoramic image at lower resolution however, the resolution of a specific area of the panoramic could be extended to the maximum resolution allowed by the hardware configuration.

The seafloor is usually characterized by areas that are featureless or have low contrast. In those regions of the seafloor, it might not exist the necessary number of temporal correspondences of pairwise features belonging to points-of-view and, therefore, the stitching process cannot continue – the ROLAMOS will save the local panoramic image ("Patch") – and will restart the again.

4. Results

A set of experiments were conducted as part of this research to analyze the performance of the proposed technique – ROLAMOS. For that it was used an i7-4700 @ 2.6GHz × 8 processor computer, without graphics processing unit. These experiments can be divided into two main groups: sequential images obtained by data sets and sequential images obtained using an ROV operating in a real scenario. The first experiment uses data set images from the Scott Reef (Bryson *et al.*, 2013) and Florida Reef Tract (Lirman *et al.*, 2006). The Scott Reef data set comprises a set of underwater images acquired by an AUV which covered an area of 75 × 50 m of sea-floor and it was equipped with a high-resolution camera (1,360 × 1,024). As can be seen from the image sequence, there are three dominating substrates: reef, sand and transition area, which cause several challenges in computing the egomotion of the vehicle. On the other hand, a sequence of images was obtained during a survey to a reef in the Florida Reef Tract, which is composed by several vertical and horizontal paths obtained by Phantam XTL ROV. This vehicle was equipped with a Flea digital camera and the images have a resolution of 1,024 × 768 pixels. An experiment was conducted using a ROV belonging to INESC TEC (FEUP) that is equipped with an underwater camera. This acquisition system is composed by a Mako G-125 with a 6 mm lens, that allows a resolution of 1,292 × 964 at 30 frames per second (fps). It is important to highlight that some features of this camera, namely: the auto gain and the auto exposure, as well as the color correction, hue and saturation, were kept fixed for the entire experience. The visual system was calibrated (Wang *et al.*, 2010) using a chessboard to estimate the intrinsic and extrinsic parameters from corner extraction. After that, the images can be rectified. The trajectory made by the ROV during the experience follows closely a free camera motion as the movement was not restricted in 3D. The frame rate of data acquisition is high to ensure that the consecutive frames present overlapping regions. To verify the quality of the final map, some objects were placed at the bottom of the underwater environment. These objects make it possible to verify the texture information that can be expected from the panoramic images created by ROLAMOS, which is relevant to a scene understanding algorithms (e.g. for object recognition).

4.1 Visual accuracy

The ROLAMOS is suitable for realistic robotic applications for monitoring structures of the sea-floor, as well as to support mission that involves the detection of objects. In both situations, the visual accuracy of the panoramic image is extremely important.

Table I Brief description of the variables to the global and patch modules

Short Description			
Global		**Patch**	
Created after image cycle is finished		Created when the pixel count of the current panorama exceeds a defined	
size_x	Total size of the global mosaic	size_x	Total size of the global mosaic
size_y		size_y	
N_images	Sum of all images	n_images	Sum of images in the patch
N_patches	Sum of all patches	offset_x	Offset of the global mosaic from the first image
offset_x	Offset of the global mosaic from the first	offset_y	in the whole mosaic
ofsset_y	image in the whole mosaic	first_homography	Homography of the first image in the patch
scale	Scale factor for blending the patches	last_homography	Homography of the last image in the patch
	into the full resolution global mosaic		
Egomotion – the motion estimation from visual observations			

An example of the visual accuracy of the ROLAMOS technique can be verified in Figure 2, which represents a partial representation of a mosaic image that encompasses eight underwater images, obtained by Phantam XTL ROV. This result demonstrates the high definition of the reconstruction process that can be performed by ROLAMOS: areas of the mosaic that represent the frontiers from different observations do not include visible artifacts that are usually caused by the inaccurate combination of several observation for the same spatial location. These visual artifacts are the main evidence of an imprecise egomotion estimation or an inappropriate blending function. These factors have a strong impact on the visual accuracy of panoramic images. Figure 2 demonstrates that the texture information of the panoramic image created by ROLAMOS has high quality, which is caused by the innovations presented by ROLAMOS.

Figure 3 presents the result of a sequence of 22 images whose trajectory includes one direction change. This scenario has a good texture information as a total of (approximately) 30,000 features was obtained during the registration/warping process described above. Moreover, the minimum number of good matches (after the computation of both fundamental and homography matrices) was always higher than 12, and consequently, the egomotion estimative was reliable and no observation was discarded.

The final step of the mosaic creation process is the blending, which combines two sequential observations having overlapping regions into a single image. The blending is responsible for an uniformization of the panoramic image, avoiding visible seams between observations.

The robustness of this step increases the quality of the final panoramic image and facilitates the upcoming interpretation of the scenario. This step is considered very expensive in terms of computational costs; thus, the ROLAMOS technique proposes a simple but effective implementation of a blending function. Figure 4 shows the influence of different blending approaches. More specifically, Figure 4(a) depicts the mosaic image obtained with a simple average function, while Figure 4(b)-(d) shows the results obtained by applying 2D Gaussian function with a standard deviation values of 0.2, 0.1 and 0.01, respectively. As expected, the average blending function leads to a blur effect along the transition of images (frontiers between observations) and, in addition, it smooths edges and regions with high texture. In contrary to the previous result, the 2D Gaussian function preserves the texture information of panoramic image (specially, regions of transition between observations). The best result was obtained for a 2D Gaussian function with a small standard deviation.

Figure 5 presents a patch created by an ROV in the INESC TEC laboratory at the Faculty of Engineering of University of Porto. This demonstrates the performance of ROLAMOS during a mission executed by a small ROV in a relevant scenario. An anchor was placed in the sea-floor to make possible to validate the quality of the mosaic, as well as to verify if it possible to identify and recognize objects (Section 4.5). The visual observations acquired by the ROV were completely

Figure 2 A mosaic image created by a sequence of eight images of the Florida Reef Tract data set with ROLAMOS

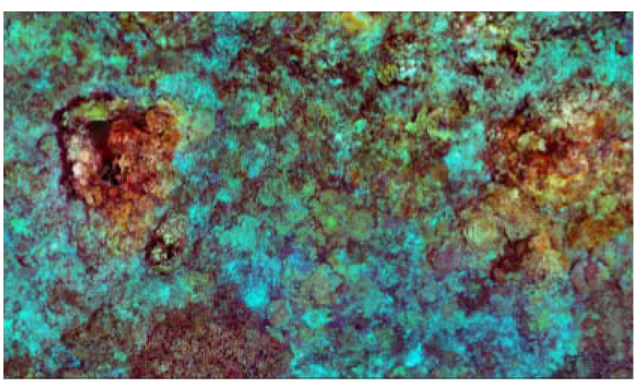
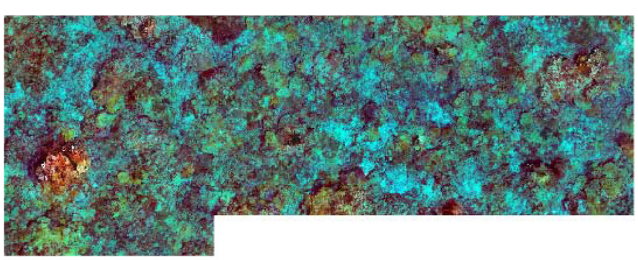
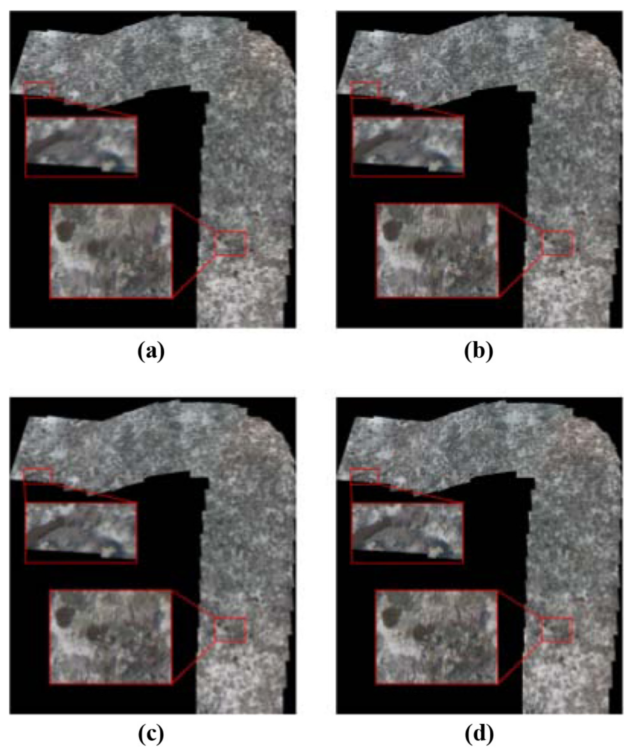


Figure 3 A mosaic image of the Florida Reef Tract data set, with one direction change.



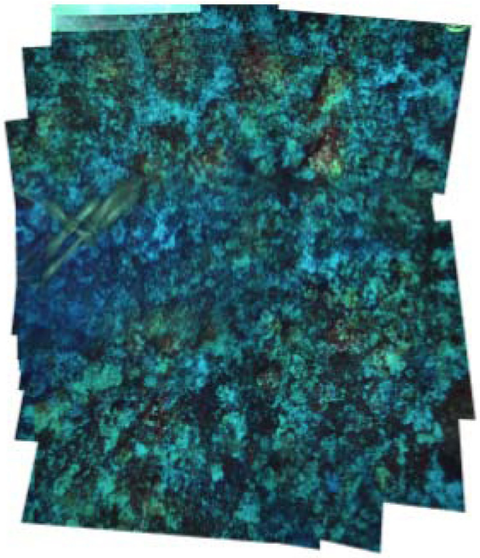
Note: After 22 images, each cycle time took about 3.3 s, which is not considerable, taking into account the intended application with the proposed method

Figure 4 Panoramic result of a sequence of 22 real images with different blending methods



Notes: (a)Simple average; (b) the blending technique (2D Gaussian function) with a standard deviation of 0.2; (c) 0.1; (d) 0.01

Figure 5 A patch created by stitching of 670 images acquired by a ROV in a real scenario



processed and stitched by ROLAMOS. As can be noticed, the final panoramic image has high resolution, all objects are well represented (as the gradient-information was kept intact) and they can be identified easily by humans. It is also important to refer that the estimation of the egomotion was correctly determined as it captured the trajectory made by the ROV: with several direction changes and revisited areas, otherwise, it would be impossible to characterize and reconstruct this environment at full resolution.

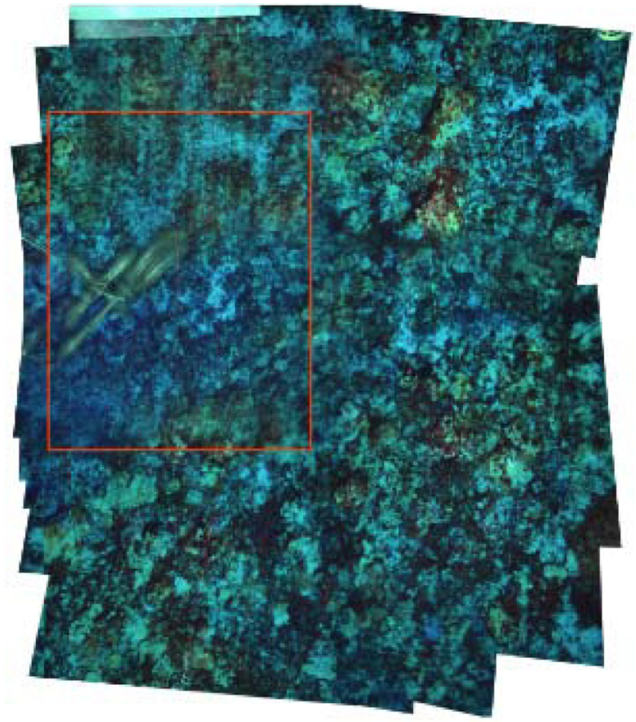
4.2 Image alignment

The impact of the image alignment mechanism incorporated in ROLAMOS can be qualitatively analyzed in Figure 5. This figure shows the panoramic image obtained with the ECC alignment technique. This technique maximizes the ECC function which constitutes a robust measure against geometric and photometric distortions. The result of ROLAMOS without this image alignment phase is depicted in Figure 6. As can be noticed in the area highlighted in this figure, the ECC method leads to a better egomotion estimation and preserves the texture information of the individual observations in the final panoramic image. Although the object placed at the bottom of the sea-floor is visible in both cases, the object was represented with better details in the panoramic image resulted from the image alignment phase (Figure 5). Thus, the previous results demonstrate that ROLAMOS is able to aggregate all observations in a suitable manner for underwater robotic applications.

4.3 Mosaic management

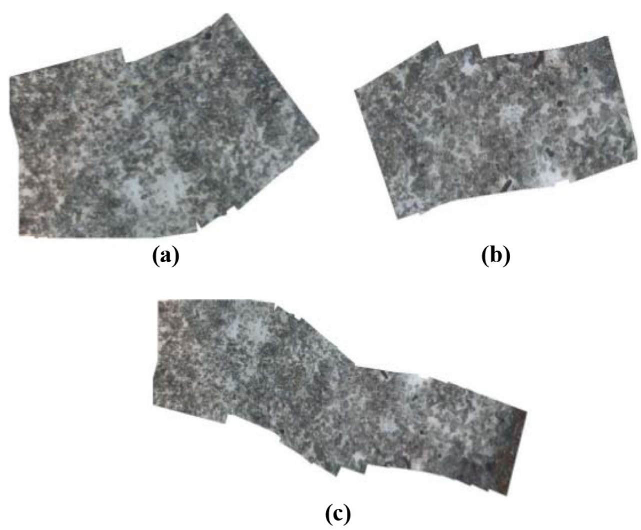
The full Scott Reef data set comprises 9,831 images, on a roughly square area, which means that it approximately a 100×100 image grid. A rough estimation of the memory size of the resulting global mosaic will be of about 6,845 million pixels by assuming that each image overlaps the previous by 50 per cent. As such, the representation of this large-scale environment with full resolution will be impossible for conventional computers

Figure 6 A patch created by stitching of 670 images acquired by a ROV in a real scenario, without ECC



because the number of pixels exceeds the memory installed in the majority of computers used in AUV or ROV. In this sense, the ROLAMOS implements a memory management mechanism that splits the global panoramic image into local patches that are created online, see Figure 7(a) and (b). The global mosaic can be

Figure 7 Two examples of the patch creation in a horizontal representation



Notes: (a) and (b) are the sequential patches created by the mosaic management approach of ROLAMOS, with a maximum number of pixels of five million; (c) is the final mosaic obtained from the patches

Sensor Review
Volume 39 · Number 3 · 2019 · 387–396

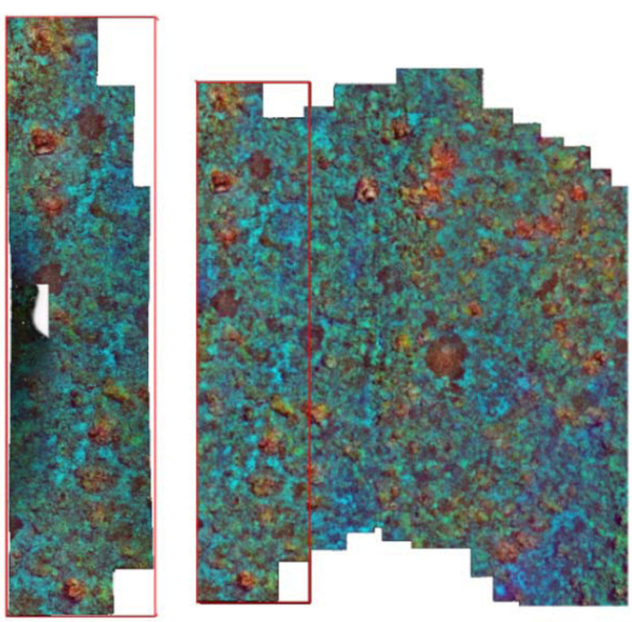
derived from these patches by using the proposed large-scale image mosaicking algorithm, see Figure 7(c). More specifically, this example consists of a panoramic image obtained by 22 sequential observations that represent 30 million pixels. It should be noticed that each patch saves complementary information (e.g. coordinates, homographies and egomotions) that is relevant for the creation of the global panoramic image – usually done at offline. Although the global mosaic of the current example does not exceed the memory capacity of the current system installed onboard of the ROV, it can be used as a valid proof-of-concept. Therefore, ROLAMOS manages the creation of a mosaic by building different sub-representations of the environment. ROLAMOS creates each patch during online operations, and this patch will be used by object recognition algorithms as it represents a more contextualized information of the scene (when compared to isolated observations that usually represent a small portion of the underwater scene).

4.4 Comparison with state-of-art technologies

This section compares the results obtained by ROLAMOS with other techniques that could be found in the literature. Some techniques (online and offline) were selected to determine and compare the impact of different approaches in the overall performance of the final panoramic image.

A straightforward comparison between ROLAMOS and the OpenCV (stitching module) is available on Figure 8. Figure 8(a) depicts a partial mosaic with visual artifacts created by the OpenCV method. As can be noticed, only some observations obtained from the entire trajectory were stitched correctly – see the region of the figure highlighted by a red rectangle – which indicates the existence of a small black region without any texture information. Therefore, the OpenCV method was unable to reconstruct the entire underwater scenario. Figure 8(b) depicts the same underwater sequence reconstructed by the ROLAMOS.

Figure 8 Panorama obtained by OpenCV Stitching function (a) and by ROLAMOS method (b) with Florida Reef Tract data set

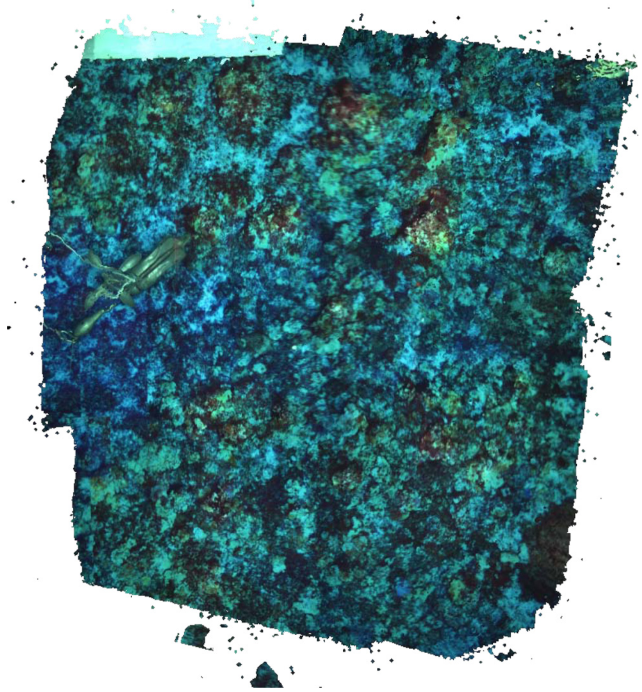


(a) (b)

In practice, Figure 8(b) represents a panoramic image that was obtained by stitching a set of 1,208 visual observations. The panoramic image depicts the trajectory performed by the ROV during its close-range operation (by a factor of scale due to the unknown depth) as it is an accurate representation of the underwater environment. Moreover, the complete panoramic image of Figure 8(a) was created in approximately 3 h, while the panoramic image of Figure 8(b) was obtained by ROLAMOS in less than 2 h.

ROLAMOS was conceived to be an online method that can be incorporated onboard of a ROV or AUV. There are techniques available in the literature that create panoramic images offline with a high degree of perceptual accuracy, for instance, the Pix4DMapper. This software has the ability to create a high-quality map but the observations must be considered as data sets and the path made by the observer should be well defined. The Pix4DMapper configured with two main steps namely, pre-processing and photomosaic creation. The pre-processing verifies the number of features collected from each image (it should be above the minimum number required for stitching). In the next step, images are stitched and the photomosaic is created. Figure 9 shows the result obtained by Pix4DMapper with the same images used by ROLAMOS of Figure 5). As can be noticed, the results obtained by the ROLAMOS and Pix4DMapper are similar in aspect. ROLAMOS represented the underwater scenario without distorting the objects (e.g. the anchors) which can be important when one of its goals is to allow the identification of objects. The panoramic image of ROLAMOS, formed by 670 images, was computed less than 1 h; however, the Pix4DMapper required more than 2.5 h. These results demonstrate that the quality of the panoramic images from both approaches is

Figure 9 A patch created by stitching of 670 images acquired by a ROV in a real scenario, by the Pix4DMapper software



comparable, indicating that ROLAMOS represents a good balance between performance (online calculations) and the perceptual quality of the panoramic image.

4.5 Object identification

This section presents the application of one advantage of ROLAMOS – path generation – that can be used by other scene understanding algorithms to retrieve additional information from the environment being observed by the ROV or AUV. As example, this research used an object identification approach based on bag of words (BOW) to recognize the anchor at the bottom of the sea-floor during the patch creation process. A classifier based on support vector machine was trained using the dictionary of features obtained by the BOW-based approach that requires images that represent the object, see Figure 10. The training procedure of this classifier followed the one-vs-all approach that used 100 samples of images representing the object. Using this simple but effective process, the classifier was able to detect the presence of anchors in 83 per cent of the underwater images (30 testing samples were considered).

5. Conclusion

This research describes a novel technique called ROLAMOS, which composes sequences of the sea-floor from visual observations. This method supports recent perception algorithms for recognizing objects and for scene understanding. This technique presents a robust registration of monocular images though an efficient frame-to-frame motion estimation process (with outlier removal and consistency checking stages) and a mosaic management mechanism. Overall, the method makes it possible to reconstruct large underwater areas with a high resolution. Moreover, ROLAMOS incorporates a mosaic management approach to overcome hardware limitation which is usual for underwater vehicles.

An extensive set of experiments were performed in this research to validate the performance of ROLAMOS. An ROV from the INESC TEC was used in a real scenario to collect realistic data. Results obtained in multiple conditions, scenarios and missions demonstrate that the technique presented in this work has the ability to accurately reconstruct underwater scenes. ROLAMOS was compared to other state-of-the-art implementations, and results from this analysis show that ROLAMOS is a reliable method as a high number of visual observations were stitched together, creating complete panoramic images with high resolution. Therefore, ROLAMOS is suitable for robotic applications that require for an online, robust and effective technique to reconstruct the underwater environment from visual information.

Figure 10 Example of images used for training of the classifier

Acknowledgements

This work is financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation – COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT – Fundação para a Ciência e a Tecnologia within project POCI-01-0145-FEDER-030010 and by STRONGMAR Project – STRengthening MARritime Technology Research Center. The authors would like to acknowledge the Australian Center for Field Robotics' marine robotics group for providing provide real sea-floor imagery data of the Scott Reef 25 (Bryson *et al.*, 2013) and the Research Centre in Underwater Robotics of the University of Girona for providing real sea-floor visual data of the Florida Reef Tract (Lirman *et al.*, 2006).

The author(s) of this article have not made their research data set openly available. Any enquiries regarding the data set can be directed to the corresponding author.

References

Arnaubec, A. Opderbecke, J. Allais, A.G. and Brignone, L. (2015), "Optical mapping with the ARIANE HROV at IFREMER: the MATISSE processing tool", OCEANS 2015 Genova, pp. 1-6, doi: 10.1109/OCEANS-Genova.2015.7271713.

Bay, H., Tuytelaars, T. and Van, L. (2008), "Surf: speeded up robust features", *Computer Vision and Image Understanding*, Vol. 110 No. 3, pp. p433-p466.

Bonny, M.Z. and Uddin, M.S. (2016), "Feature-Based image stitching algorithms", *2016 International Workshop on Computational Intelligence*, pp. 198-203, doi: 10.1109/IWCI.2016.7860365.

Bryson, M., Johnson-Roberson, M., Pizarro, O. and Williams, S. (2013), "Automated registration for multi-year robotic surveys of marine benthic habitats", *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 205-211, doi: 10.1109/IROS.2013.6696832.

Elibol, A. Gracias, N. and Garcia, R. (2013), "Efficient topology estimation for large scale optical mapping", Springer Tracts in Advanced Robotics, pp. 1-76, doi: 10.1007/978-3-642-30313-5_1.

Escartín, J., García, R., Delaunoy, O., Ferrer, J., Gracias, N., Elibol, A., Cufi, X., Neumann, L., Fornari, D.J., Humphris, S.E. and Renard, J. (2008), "Globally aligned photomosaic of the lucky strike hydrothermal vent field (Mid-Atlantic Ridge, 37°18.5'N): release of georeferenced data, mosaic construction, and viewing software", *Geochemistry Geophysics Geosystems*, Vol. 9 No. 12, doi: 10.1029/2008GC002204.

Ferreira, F., Veruggio, G., Caccia, M., Zereik, E. and Bruzzone, G. (2013), "A Real-Time mosaicking algorithm using binary features for ROVs", *2013 21st Mediterranean Conference on Control & Automation*, pp. 1267-1273, doi: 10.1109/MED.2013.6608882.

Ferrer, J. Elibol, A. Delaunoy, O. Gracias, N. and Garcia, R. (2007), "Large-Area Photo- Mosaics using global alignment and navigation data", OCEANS 2007, pp. 1-9, doi: 10.1109/OCEANS.2007.4449367.

Foresti, G. and Gentili, S. (2002), "A hierarchical classification system for object recognition in underwater environments", *IEEE Journal of Oceanic Engineering*, Vol. 27 No. 1, pp. 66-78.

Gracias, N. and Santos, V. (2000), "Underwater video mosaics as visual navigation maps", *Computer Vision and Image Understanding*, Vol. 79 No. 1, pp. 66-91.

Gutiérrez, R.P. (2013), "Image blending techniques and their application in underwater mosaicing", Universitat de Girona.

Kim, W.J. and Kweon, I. (2011), "Moving object detection and tracking from moving camera", *8th International Conference on Ubiquitous Robots and Ambient Intelligence*, pp. 23-26, doi: 10.1109/URAI.2011.6146005.

Li, J., Yuan, H. and Hong, A. (2011), "Moving vehicle detection in dynamic background from airborne monocular camera", *Energy Procedia*, Vol. 13, pp. 3955-3961, doi: 10.1016/j.egypro.2011.11.568.

Lirman, D., Gracias, N., Gintert, B., Gleason, A., Reid, R., Negahdaripour, S. and Kramer, P. (2006), "Development and application of a video-mosaic survey technology to document the status of coral reef communities", *Environmental Monitoring and Assessment*, Springer, Vol. 125 Nos 1/3, pp. 59-73.

Lucas, B.D. and Kanade, T. (1981), "An iterative image registration technique with an application to stereo vision", *International Joint Conference on Artificial Intelligence*, pp. 674-679.

Marcon, Y. (2014), "LAPMv2: an improved tool for underwater Large-Area Photo-Mosaicking", Oceans – St. John's, pp. 1-10, doi: 10.1109/OCEANS.2014.7003185.

Martin, C.J.M. and Martin, E.A. (2002), "An underwater photomosaic technique using adobe photoshop", *The International Journal of Nautical Archaeology*, Vol. 31 No. 1, pp. 137-147.

Nister, D. (2004), "An efficient solution to the five-point relative pose problem", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26 No. 6, pp. 756-770, doi: 10.1109/TPAMI.2004.17.

Pinto, A.M., Moreira, A.P. and Costa, P.G. (2014a), "An architecture for visual motion perception of a surveillance-based autonomous robot", *IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pp. 205-211.

Pinto, A.M., Pinto, H. and Matos, A.C. (2016), "A mosaicking approach for visual mapping of large-scale environments", *2016 International Conference on Autonomous Robot Systems and Competitions*, pp. 87-93.

Pinto, A.M., Correia, M.V., Moreira, A.P. and Costa, P.G. (2014b), "Unsupervised flow-based motion analysis for an autonomous moving system", *Image and Vision Computing*, Vol. 32 Nos 6/7, pp. 391-404, doi: 10.1016/j.imavis.2014.04.003.

Pinto, A.M., Costa, P.Q., Correia, M.V. and Moreira, A.P. (2014c), "Enhancing dynamic videos for surveillance and robotic applications: the robust bilateral and temporal filter", *Signal Processing: Image Communication*, Vol. 29 No. 1, pp. 80-95, doi: 10.1016/j.image.2013.11.003.

Pizarro, O. and Singh, H. (2003), "Toward large-area mosaicing for underwater scientific applications", *IEEE Journal of Oceanic Engineering*, Vol. 28 No. 4, pp. 651-672, doi: 10.1109/JOE.2003.819154.

Srividya, S. and Shobha, G. (2014), "A survey on: underwater video processing for detecting and tracking moving objects", *International Journal of Engineering and Computer Science*, Vol. 3 No. 1, pp. 3768-3771.

Wang, Y.M., Li, Y. and Zheng, J.B. (2010), "A camera calibration technique based on OpenCV", *International Conference on Information Sciences and Interaction Sciences*, pp. 403-406, doi: 10.1109/ICICIS.2010.5534797.

Corresponding author

Alexandra Pereira Nunes can be contacted at: apn@inesctec.pt