A Mosaicking Approach for Visual Mapping of Large-Scale Environments

Andry Maykol Pinto, Henrique Pinto and Anibal C. Matos INESC TEC and

Faculty of Engineering, University of Porto, Portugal Email: {andry.pinto, ee09143, anibal}@fe.up.pt

Abstract—Nowadays, the technological and scientific research related to underwater perception is focused in developing more cost-effective tools to support activities related with the inspection, search and rescue of wreckages and site exploration: devices with higher autonomy, endurance and capabilities. Currently, specific tasks are already carried out by remotely-operated vehicles (ROV) and autonomous underwater vehicles (AUV) that can be equipped with multiple sensors, including optical cameras which are extremely valuable for perceiving marine environments; however, the current perceptual capability of these vehicles is still limited.

In this context, the paper presents a novel mosaicking method that composes the sea-floor from a set of visual observations. This method is called RObust and Large-scale MOSaicking (ROLAMOS) and it enables an efficient frame-to-frame motion estimation with outliers removal and consistence checking, a robust registration of monocular images and, finally, a mosaic management methodology that makes it possible to map large visual areas with a high resolution. The experiments conducted with realistic images have proven that the ROLAMOS is suitable for mapping large-scale sea-floor scenarios because the visual information is registered by managing the computational resources that are available onboard, which makes it appropriate for applications that do not have specialized computers. Further, this is a major advantage for automatic mosaic creation based on robotic applications, that require the location of objects or even structures with high detail and precision.

I. Introduction

Underwater surveys are extremely valuable for inumerous scientific applications, for instance, in archeology, geology and biology and even for shipwreck recoveries, ecological studies and environmental damage assessments. Conducting those surveys usually require a large amount of operating-time, human and technical resources which often culminates in a strong financing impact for small and medium enterprises (SME) of the marine and offshore industry.

In the last decade, the robotic community has developed applications to support the activity of SMEs, which includes mapping and perceiving marine structures through autonomous and remotely-operated vehicles. Such vehicles are usually equipped with several sensors like acoustic positioning sensors (long or ultra-short baseline), DVL (doppler velocity log), gyroscopes and GPS (global positioning system). These sensors are relevant for assuring the control and navigation of the vehicle during medium and large endurance missions however, they are not suitable when the vehicle stays at small distances from the bottom of the sea since they are unable to provide reliable and accurate information which is

required for almost every close range operations. Therefore, high resolution cameras are being installed in AUV. Obviously, visual systems are limited by the conditions and phenomenons of the environment like the poor visibility, the presence of suspensoids in water, light absorption and the backscattering effect; however, they provide relevant and helpful clues from the environment that can be used during the navigation or even for visual inspection.

This research presents a novel approach for visual mapping of the sea-floor. This method is called RObust and Large-scale MOSaicking (ROLAMOS) and conducts an efficient registration of monocular images based on frame-to-frame analysis. The algorithm is formed by three phases: estimation of the egomotion, mosaic creation and management. The estimation of the egomotion is obtained based on SfM (structure from motion) principles and, thus, the three-dimensional displacement of consecutive frames based on a motion model is obtained by extracting a set of correspondences between both views and the epipolar constraint. The phase that creates the mosaic uses overlapping fields of view to produce a high resolution image, often called by panorama. This phase conducts the stitching of consecutive frames into the panorama reference based on the egomotion. In addition, the egomotion is validated during the stitching process through a consistence checking procedure (infers about the feasibility of motion estimated within the context of the existing panorama) which is performed right after updating the panorama with the current observation. Finally, the mosaic management phase of the ROLAMOS technique gives the opportunity for stitching consecutive images at high resolution during the large-scale reconstructions (without scaling the resolution down to accommodate new observations). Therefore, it manages the process of mosaic creation based on the hardware's limitations.

Therefore, contributions of this article include:

- A novel technique that composes the bottom of the sea from sequential and overlapping visual observations. This method is called RObust and Large-scale MOSaicking (ROLAMOS);
- An efficient methodology that estimates the egomotion based on the SfM principles which is suitable for applications that do not have specialized computer devices:
- A mechanism that increases the robustness of the process that creates the panorama (an accurate stitching of frames by removing outliers and by evaluating the egomotion);

- A mosaic management approach to create a panorama image of large-scale underwater environment at the highest resolution allowed by the hardware configuration:
- Promotion of efficient vision perception methods for underwater robotic applications;
- Qualitative evaluations of the ROLAMOS, considering realistic scenarios.

The article is organized as follows: section II presents a brief review of mosaic-based approaches that are used in robotic systems to perceive the environment. Section III shows the concept of the ROLAMOS technique, where each phase is explained with more detail. Afterwards, section IV presents the practical results. The experiments were conducted using realistic sequences and they have proven that the ROLAMOS is suitable for mapping the large-scale sea-floor scenarios because the visual information is registered by managing the computational resources, which makes it suitable for applications that do not have specialized computers. Finally, section V presents the most important conclusions of this research.

II. RELATED WORKS

Several intelligent and autonomous robotic vehicles have been developed with visual systems [12] capable of operating in the underwater scenario. They are able to perform tasks that would, otherwise, be too dangerous, expensive or even impossible to achieve. Namely, tasks such as visual inspections and monitoring sea structures, like dams, dikes or other infrastructures, as well as coral reefs and sea-floor mapping [5], [10]. Aryo Ibrahim et al. (2010) [6] present a mosaic technique for a robotic application. The technique maps the areas and detects moving objects. The authors match invariant features SIFT (scale invariant feature transform) between frames in order to compute the matrix that describes the geometrical transformation (projective model). The matrix is used by the warping process which aligns the current frame with previous frames. Jing Li et al. (2011) [8] use a simplified Lucas-Kanade method combined with an image registration technique to obtain the motion vector between consecutive frames. This defines the egomotion (affine model) of the vehicle. This intends to detect the moving objects which are retrieved by a temporal differencing approach. The researching work presented in [7] proposes the detection of moving objects using the Kanade-Lucas-Tomasi feature tracking [9]. The homography is calculated using the 5-point RANSAC (RANdom SAmple Consensus) RANSAC (random sample consensus) with all features of the two consecutive frames. The RANSAC analyses inconsistent features during the computation of the homography matrix and an onlineboosting algorithm is used to follow moving objects. The Elibol et al. (2014) [4] presented a image mosaicking algorithm for visual mapping in underwater environments using multiple underwater robots that identifies overlapping image pairs in the trajectories carried out by the robot formation. Currently, the limited visual perception capability in the underwater environment presents one of the greatest restrictions to the use of these type of vehicles. Therefore, it becomes crucial to develop perception methods that make underwater sea-floor mapping possible whenever the vehicle moves close to the bottom of the sea. These methods will allow a more efficient use of such systems in real environments and applications [11].

III. ROBUST AND LARGE-SCALE MOSAICKING (ROLAMOS)

The RObust and Large-scale MOSaicking (ROLAMOS) has three main phases: the estimation of the egomotion, stitching and mosaic management. The next sections will explore the concepts and approaches involved in each phase of the ROLAMOS architecture, see Figure 1.

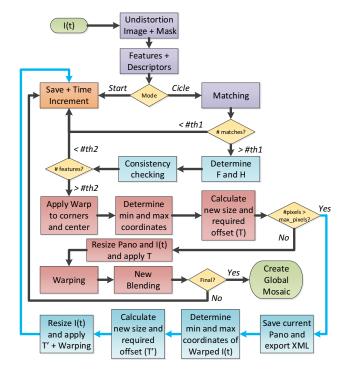


Fig. 1. Overall architecture proposed by the ROLAMOS technique to reconstruct large-scale environments.

A. Estimating the egomotion and stitching

Figure 2 depicts the main elements of the two starting phases of ROLAMOS.

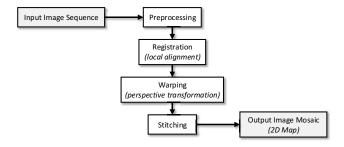


Fig. 2. Diagram of the feature-based image mosaicking.

The first step is the preprocessing where the image is rectified according to the camera calibration parameters. This step ensures that no misalignments occur due to distorted images and reduces the errors for the estimation of the egomotion.

The image registration stage determines the best alignment between two images taken from different viewpoints [3]. Building a two-dimensional map from a sequence of images poses an extremely difficult task. Nevertheless, the camera poses can only be determined up to a one parameter of ambiguity - a scale factor. The real relative camera poses could be estimated only if additional information is provided like the distance from the camera to a point or the distance between two points in the scene. In other words, all the possible locations of a point that is seen in one image is a line that goes through the corresponding point and the epipolar point on the other image. This is known as epipolar constraint and it is the basis of the Structure from Motion (refers to the process of extracting the three-dimensional structure of objects by analyzing the projected two-dimensional motion over time).

The ROLAMOS is a feature-based technique which means that keypoints are extracted and matched between consecutive frames. After that, the inconsistence level of pairwise keypoints is estimated by using the epipolar geometry: a keypoint of the second view should be located close to the epipolar line generated by the matched keypoint of the first view. The fundamental matrix (a relationship between any two images of the same scene) is required to estimate the epipolar lines and is computed using a RANSAC approach with all keypoints from both frames. In this way, a pairwise keypoint having a large bilateral distance (sum of distances of each keypoint to its corresponding epipolar line) represents a wrong match that should not be considered during the estimation of the egomotion. Figure 3 provides a visual illustration of the bilateral error (err_{BL}^i), see also equation 1.

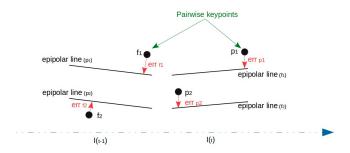


Fig. 3. Detecting wrong matches between keypoints of two consecutive images based on the epipolar geometry. The keypoints notated by f and p represent features obtained in image I(t-1) and I(t), respectively. The superscript represents the number of the pairwise keypoint match.

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

where err_f^i is the distance between the keypoint fi to its epipolar line (defined by the pi and the fundamental matrix). Please notice that, the sum of the bilateral error from all pairwise keypoints gives an overall measure of the quality that will be expected for the egomotion estimation.

In this paper, the model that captures the egomotion follows closely the planar transformation, also known as homography. Homography is a transformation that maps points in one image to the corresponding points in the other image. This is only possible due to the co-planarity constraint, which means in practice that the scene is roughly flat (a realistic assumption for sea-floor scenarios). The homography transformation is calculated by analyzing the resulting pairwise keypoints (the RANSAC algorithm fits the motion model into the remaining inliers). This egomotion allows overlaying the current observation into a common reference frame and, therefore, the image can be stitched into the panorama. The stitching process implies that the current image is projected (projective transformation) based on the homography matrix, because it has been acquired from a different point of view. This is also known as warping.

The final step of the stitching process is the image blending: combining two or more partially overlapping images together to create a photo-mosaic that enables the visual interpretation of the sea-floor. This intends to compensate for exposure differences and other mis-alignments and is accomplished by a spatially-varying weighting (feathering). Then, a blending function determines how this new projected (warped) image will blend with the panoramic image. The algorithm uses a 2D Gaussian distribution to center the weights. This creates a gradient-like matrix for both the warped image and for the panoramic image. This blending function makes it possible to create a visually pleasant panorama.

Besides all the mechanisms regarding the robustness of the estimation of the egomotion, the ROLAMOS improves the robustness of the stitching process related to the quality of the egomotion estimation. There are still situations were the egomotion is wrongly estimated due to the parallax effect or even to the lack of good features (non-distributed along the frame) and, in those situations, the warping process will produce undesirable results which may compromise the creation of the full panorama. Detecting these situation is crucial for a reliable and robust creation of the panorama and, therefore, the ROLAMOS also implements an additional outliers removal method based on the empirical analysis of the resulting homographies. The homography filtering guarantees that no sudden transitions are applied between consecutive images, which could lead to an inconsistent mosaic. It is assumed that the images are taken sequentially within short periods of time (for example, once every second): no sudden changes such as a full 180 degrees turn or an exaggerated x or y translation or even a (large) change in scale will not be admitted. If some of the previous issues are detected, the stitching process will not be conducted, see the blue blocks from the center of Figure 1.

B. Mosaic Management

The ROLAMOS is a technique that being designed for AUV applications, it has the capability to create a robust and accurate panorama based on underwater sequences. A new problem rises when the AUV operates in large scale environments because the mosaicking process (that was presented in previous sections) conducted with high definition has large memory requirements which could compromise the operation of such vehicles, specially, if the panorama is used

¹The SURF [1] feature description was used in this research since it is invariant to scale and rotation, which are relevant characteristics for underwater mosaicking.

for navigation tasks. As the mosaic grows with an increasing number of processed images, it can reach sizes which the processing system may not be able to manage. Therefore, an hierarchical structure is proposed to handle with hardware's (memory) issues.

TABLE I. Brief description of the variables and files associated to the global and patch modules.

Global		Patch	
		Short description	
Created after image cycle is finished		Created when the pixel count of the current panorama exceeds a defined	
size_x size_y	total size of the global mosaic	size_x size_y	total size of the patch
N_images	sum of all images	n_images	sum of images in the patch
N_patches offset_x	sum of all patches offset of the global mosaic from the	offset_x offset_y	offset of the patch from the first image in the whole mosaic (global reference)
offset_y	first image in the whole mosaic	first_homography	homography of the first image in the patch
scale	scale factor for blending the patches into the full resolution global mosaic	last_homography	homography of the last image in the patch
		Associated files	
global.xml	stores all the variables above	patchxml	stores all the variables above
global.png	mosaic at maximum resolution (according to maximum pixel count)	patchpng patch mask.png	full resolution image of the patch full resolution mask of the patch
globalsmall.png	small scale of the global mosaic, resized to a standard resolution	patch_small.png patch_smallmask.png	small scale image of the patch small scale mask of the patch

Note: the '__' in the patch's file names are to be replaced with the patch number = 0, 1, ..., N_patches - 1

The major objective is to prevent the panoramic image from growing to a size too big for the system to manage, mainly limited by the available RAM (random-access memory) and, for that purpose the size of the mosaic image is limited to a maximum number of pixels (threshold), see the blue blocks from the bottom of Figure 1. The mosaicking pipeline is halted if the homography of the new input image causes the resulting panoramic image to grow beyond that threshold. The current mosaic image has to be saved into the hard drive disk as a 'patch' and a new mosaic has to be started. Then, a 'patch' is formed by a portion of the panoramic image (composition of sequentially blended images) that does not exceed the maximum number of pixels and by additional information that supports the reference of this sub-mosaic in the global panorama frame. In this context, the 'global' panorama is the complete 2D map and is composed by all the patches.

Images are sequentially added to a patch until the maximum allowed number of pixels is reached. At this point, the current mosaic image and the relevant information are saved into the hard drive. Table I summarizes the additional information: the offset (translation required to view the full image), the corner locations and homographies of the first and last warped images. This creates an hierarchical scheme where each 'patch' can be referred to a 'global' mosaic for further reconstruction. The next step is restarting the mosaic image, this means, starting a new patch that should be comprised of the new warped image. This process is repeated until the sequence of the input images ends. The global mosaic of large scale environments cannot be directly generated from the patches at full resolution because it would exceed the maximum number of pixels. The proposed solution is to scale down the global mosaic, where a scaling factor is calculated so that all patches can fit into the global mosaic frame without exceeding the maximum pixel limit. Then, the patches are loaded one by one and resized according to the global scaling factor and, afterwards, they are blended. This makes it possible to visualize the entire panorama image but at lower resolution however, the resolution of a specific area of the panorama could be extended to the maximum resolution allowed by the hardware configuration whenever it must be analyzed with detail.

IV. RESULTS

A set of comprehensive experiments were conducted as part of this research. These experiments intend to analyze the behavior of the proposed technique called ROLAMOS during the creation of large-scale mosaics and, therefore, factors such as the visual accuracy and the computational performance are further evaluated and discussed in this section. A sequence of images from the Scott Reef [2] are used for evaluating ROLAMOS and, in this manner, the results can be compared by future researches. The Scott Reef comprises a set of underwater images that were obtained from a AUV which covered an area of 75m×50m of sea-floor and it was equipped with a high-resolution camera (1360×1024)². As can be seen from the image sequences, there are three dominating substrates: reef, sand and transition area which cause several challenges in computing the egomotion of the vehicle.

A. Visual accuracy

The visual accuracy of the panorama obtained by the ROLAMOS is of extremely importance since it will be used by a realistic robotic application for monitoring man-made structures of the see-floor as well as for navigation tasks in the vehicle (in future versions). An example of the visual accuracy of the ROLAMOS technique can be confirmed in figure 5 and it represents a mosaic image resulting from stitching five underwater images. As can be noticed, the reconstruction process achieves a visual perception of the sea-floor with high definition: areas of the mosaic representing the frontiers from different observations do not evidence visible artifacts that are caused by an inaccurate combination of several observation values for the same spatial location. These visual artifacts are the fundamental cause of an imprecise estimation of the egomotion or even an inappropriate blending function: both issues have a strong impact in the warping and stitching steps depicted in figure 2.

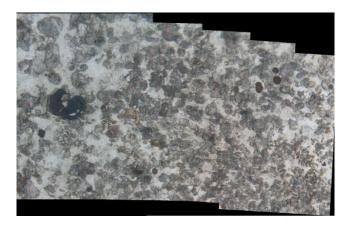


Fig. 5. A mosaic image created by a sequence of 5 images with ROLAMOS.

Figure 6 presents four results for an exactly sequence of images but with a different blending technique. Figure 6 a) depicts the mosaic image obtained with a simple average

²The authors would like to acknowledge the Australian Center for Field Robotics' marine robotics group for providing the real sea-floor imagery data of the Scott Reef 25 [2].



Fig. 4. A mosaic image created by a sequence of 95 images with ROLAMOS.

blending, while figure 6 b), c) and d) are the 2D Gaussian function with a standard deviation value of 0.2, 0.1 and 0.01 respectively. As expected, the average blending function causes a blur effect during the transition of images (frontiers between observations) and, moreover, it smooths the edges and regions with high texture. On the contrary, the spatially-varying weighting (2D Gaussian function) of the current image into the mosaic lead to better results since regions of the panorama with transitions or texture are preserved with small loss of information, specially for a small standard deviation.

B. Mosaic Management

The full Scott Reef dataset is comprised by 9831 images, on a roughly square area, which means that it approximately a 100 by 100 image grid, and assuming that each image overlaps the previous by 50%, a rough estimation of the resulting global mosaic would be of about 6,845 million pixels. Therefore, the representation of this large scale environment with full resolution is impossible for conventional computers because it exceeds the memory of the majority of such devices.

Figures 7 a), b) and c) are the results of patches created by the proposed large scale image mosaicking algorithm. The final mosaic obtained from these patches are presented in Figure 8. This panorama consists of 22 sequential images and represents 30 million of pixels. Although the global mosaic does not exceeds the real system capacity, it is a valid proofof-concept because it demonstrates the ability of the mosaic management approach to create different sub-representations of the mosaic. In addition, the ROLAMOS aggregate these representations in a suitable manner that makes it possible to reconstruct the panorama at full resolution (in predefined zones due to hardware's limitations). Considering a more relevant experiment, the figure 4 represents a panorama that is obtained by stitching 95 visual observations and, as can be noticed, it presents the navigation path that was executed by the AUV during its close range operation (obviously, by a factor of scale due to the unknown depth). These experiments showed that the threshold value for the maximum pixel count can be dependent on the hardware's limitation, for instance, how many RAM memory is available. The ROLAMOS uses a few intermediary matrices to perform all computations and, for reference, a realistic approximation of the amount of RAM used for an input sequence of images with a resolution of 1280 by 960 pixels, resulting in a global mosaic with a total resolution of

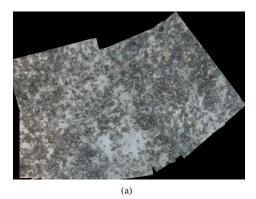
10,000 by 10,000 pixels is estimated to be slightly over 2 GB of required memory.

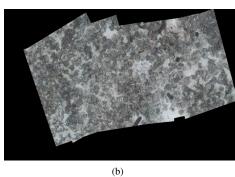
V. CONCLUSION

This research presented a novel technique that composes the bottom of the sea from sequential and overlapping visual observations. This method is called RObust and Large-scale MOSaicking (ROLAMOS). The ROLAMOS enables a robust registration from close range and monocular images, an efficient frame-to-frame motion estimation with outliers removal and consistence checking and, finally, a mosaic management methodology that makes it possible to map large visual areas with a high resolution. Usually, the composition of large scale image mosaics can escalate to sizes that are not manageable by the operating system, leading to memory problems. The mosaic management approach overcomes this system limitation of memory by composing the mosaic in large patches with a manageable size which can be integrated after by selecting the area of the scenario that should be recreated at full resolution. Results obtained from realistic underwater sequences proved the robustness of the process that creates the panorama (an accurate stitching of frames by removing outliers and by evaluating the egomotion). Moreover, the mosaic management approach proposed by ROLAMOS enables the creation of a panorama image that represents a large-scale underwater environment at the highest resolution allowed by the hardware configuration. Therefore, the proposed method is suitable for robotic-based applications that require a robust and effective technique to reconstruct the underwater environment from visual information. For future activities, the ROLAMOS technique will be tested in other realistic scenarios. In addition, a navigation methodology will be developed to provide helpful and complementarity information for close range operations of AUVs.

ACKNOWLEDGMENT

This work is financed by the ERDF European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme within project POCI-01-0145-FEDER-006961, and by National Funds through the FCT Fundacao para a Ciencia e a Tecnologia (Portuguese Foundation for Science and Technology) as part of project UID/EEA/50014/2013. The research leading to these results has received funding from





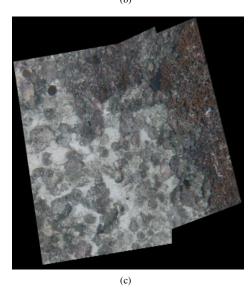


Fig. 7. Result of the proposed Large Scale Image Mosaicking algorithm from a sequence of 22 images: (in a horizontal representation). (a), (b) and (c) are the sequential patches created by the mosaic management approach of ROLAMOS. Patches were set to a maximum number of pixels of 5 million.

the European Union's Horizon 2020 - The EU Framework Programme for Research and Innovation 2014-2020, under grant agreement No.692427. And, Project "NORTE-01-0145-FEDER-000036" is financed by the NORTE 2020, under the PORTUGAL 2020 Partnership Agreement, and through the ERDF.



Fig. 8. The resulting global mosaic of the proposed Large Scale Image Mosaicking algorithm from a sequence of 22 images (in a horizontal representation).

REFERENCES

- [1] H Bay, T. Tuytelaars, and L. Van. Surf: Speeded up robust features. Computer Vision and Image Understanding, 110(3):433–466, 2008.
- [2] M. Bryson, M. Johnson-Roberson, O. Pizarro, and S. Williams. Automated Registration for Multi-year Robotic Surveys of Marine Benthic Habitats. In IEEE, editor, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 205–211, 2013.
- [3] Armagan Elibol, Nuno Gracias, and Rafael Garcia. Efficient Topology Estimation for Large Scale Optical Mapping. Springer Tracts in Advanced Robotics, Springer, first ed. edition, 2013.
- [4] Armagan Elibol, Jinwhan Kim, Nuno Gracias, and Rafael Garcia. Efficient image mosaicing for multi-robot visual underwater mapping. Pattern Recognition Letters, 46(1):20–26, 2014.
- [5] N. Gracias and Santos-Victor. Underwater Video Mosaics as Visual Navigation Maps. Computer Vision and Image Understanding, 110(3):66–91, 2000.
- [6] Aryo Wiman Nur Ibrahim, Pang Wee Ching, G.L. Gerald Seet, W.S. Michael Lau, and Witold Czajewski. Moving Objects Detection and Tracking Framework for UAV-based Surveillance. In 2010 Fourth Pacific-Rim Symposium on Image and Video Technology, pages 456– 461. IEEE, November 2010.
- [7] Won Jin Kim and In-So Kweon. Moving Object Detection and Tracking from Moving Camera. In 8th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), pages 23–26, November 2011.
- [8] Jing Li, Hong Ai, and Jianzhu Cui. Moving Vehicle Detection in Dynamic Background from Airborne Monocular Camera. Energy Procedia, 13:3955–3961, Jan 2011.
- [9] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *In International Joint Conference on Artificial Intelligence (IJCAI)*, pages 674–679, 1981.
- [10] Colin J.M. Martin and Edward A. Martin. An underwater photomosaic technique using adobe photoshop. *The International Journal of Nautical Archaeology*, 31(1):137 – 147, 2002.
- [11] Andry Maykol Pinto, Miguel V. Correia, A. Paulo Moreira, and Paulo G. Costa. Unsupervised flow-based motion analysis for an autonomous moving system. *Image and Vision Computing*, 32(6-7):391–404, 2014.
- [12] Andry Maykol Pinto, A. Paulo Moreira, and Paulo G. Costa. An Architecture for Visual Motion Perception of a Surveillance-based Autonomous Robot. In IEEE, editor, IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), pages 205– 211, 2014.

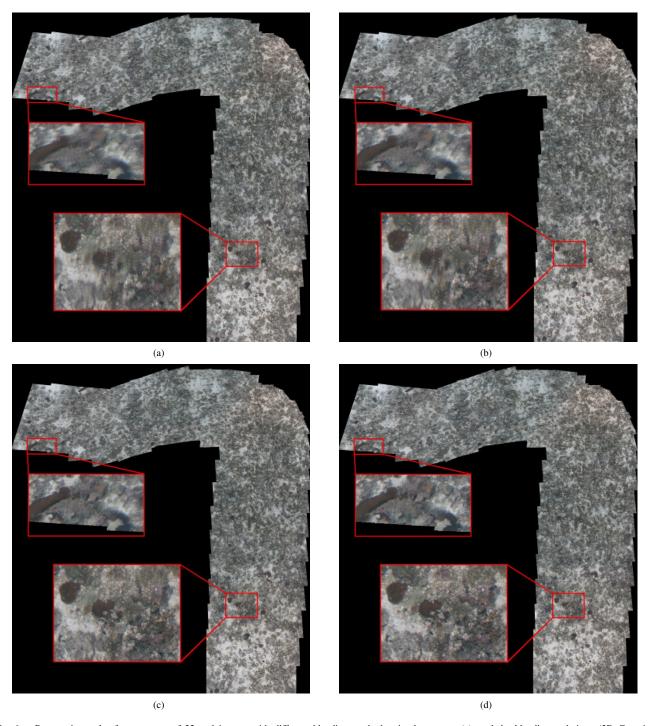


Fig. 6. Panoramic result of a sequence of 22 real images with different blending methods: simple average (a); and the blending technique (2D Gaussian function) with a standard deviation of 0.2 (b), 0.1 (c) and 0.01 (d)