

A Novel Blending Technique for Underwater Giga-Mosaicing

Ricard Prados¹, Rafael Garcia¹, Nuno Gracias¹, Javier Escartín², László Neumann^{1,3}

¹Computer Vision and Robotics Group, University of Girona, 17001 Spain

²Equipe de Géosciences Marines, CNRS/IPGP UMR7154, 75238 Paris, France

³ICREA, Barcelona, 08010 Spain

Abstract—The fusion of several images of the same scene into a single and larger composite is known as *photo-mosaicing*. Unfortunately, the seams along image boundaries are often noticeable, due to photometrical and geometrical registration inaccuracies. *Image blending* is the merging step in which those artifacts are minimized.

Processing bottlenecks and the lack of medium-specific processing tools have restricted underwater photo-mosaics to small areas despite the hundreds of thousands of square meters that modern surveys can cover. Large underwater photo-mosaics are increasingly in demand for the characterization of the seafloor for scientific purposes. Producing these mosaics is difficult due to the challenging nature of the underwater environment and the image acquisition conditions, including extreme depth, scattering and light attenuation phenomena and difficulties in vehicle navigation and positioning.

This paper proposes strategies and solutions to tackle the problems of very large underwater optical surveys (*Giga-mosaics*), presenting contributions in the image pre-processing, enhancing and blending steps, resulting in an improved visual quality in the final photo-mosaic. A comprehensive review of the existing methods is also presented and discussed. Our approach is validated by a large optical survey of a deep-sea hydrothermal field, leading to a high-quality composite in excess of 5 Gigapixel.

Index Terms—Mosaicing, Blending, High-resolution Mapping, Underwater Optical Survey

I. INTRODUCTION

The underwater robots used in deep seafloor exploration, i.e. AUVs (Autonomous Underwater Vehicles) or ROVs (Remotely Operated Vehicles), are usually equipped with advanced sensor suites, including Ultra-Short Baseline (USBL) systems, accelerometers, inclinometers, acoustic imaging sensors, and optical cameras, among others. Among these, optical imaging provides short-range high-resolution visual information of the ocean floor.

Archeologists, geologists and biologists can benefit from these images as they provide the most precise and accurate representation of the selected areas, enabling detailed analysis of the structures of interest.

Nevertheless, the underwater medium adds particular challenges to the image acquisition task. When an underwater vehicle acquires images in deep waters, light attenuation impacts visibility range and color reproduction. Due to this attenuation phenomenon, image acquisition needs to be performed close to the seabed, considerably limiting the maximum area covered by a single photograph. Hence, optically mapping large seafloor areas can only be achieved by building image mosaics from a set of reduced-area pictures.

Over the last decade, the relevance of *photo-mosaicing* has grown significantly. As a clear example, numerous off-the-shelf still cameras now include built-in algorithms to fuse several pictures from a panoramic sequence into a single wide-angle view. Furthermore, gigapixel photo-mosaics [1] of the entire Earth are easily available through the Internet, using a limited bandwidth connection. In most cases, such large mosaics are created from terrestrial, aerial or space related imagery. The common photo-mosaicing problems for this kind of images, comprehending the compensation of different exposures and non-uniform illumination, are widely treated in the literature.

Unfortunately, performing underwater image surveys is a challenging task with a much higher level of complexity than conventional terrestrial or aerial image photo-mosaic generation. As stated in [2], and due to the constrained image acquisition conditions, both the navigation data and the images acquired should be used to recover an accurate estimate of the camera poses during the survey. This

information fusion is often performed by means of global alignment techniques [2], [3], [4], [5]. This is a mandatory step before generating precise visual maps from the seafloor. Usually, the short distance between the camera and the seafloor implies the presence of *parallax* effects (see Figure 1), which considerably affects the 2D mosaicing approach due to the violation of the planar assumption, *i.e.* the assumption of a flat scene, which allows the computation of 2D transformations between images. Furthermore, suspended particles causing the *scattering* phenomenon [6] are commonly present. Moving elements, such as fish and algae, are examples of other common issues in underwater photo-mosaicing. Consequently, once the global alignment problem has been solved, and an accurate vehicle trajectory has been obtained, photo-mosaicing still requires a final step. It is necessary to give to the heterogeneously appearing image dataset a continuous and uniform appearance in the form of a single large image. This is achieved by *image blending* techniques, and the processing procedure proposed here. Apart from the visual appearance, blending techniques are also important for proper interpretation and scientific exploitation of seafloor imagery (*e.g.* [7], [8]). The structures and objects of interest may cover a wide range of scales, from a few centimeters (macrofauna, rocks), and hence within individual images, to several tens or hundreds of meters (topographic scarps, fractures), therefore spanning several frames. To properly analyze such various features, and to understand the spatial relationships that may exist (*e.g.*, faunal assemblages associated with geological features) it is preferable to have a single, wide area photo-mosaic, where imaging artifacts are minimized, and where identified features of interest may be accurately represented regardless of their size and imaging conditions.

The goal of this work is to create a complete blending approach using state-of-the-art methods capable of generating and blending large-scale optical maps, such as that used for testing, obtained from the Mid-Atlantic Ridge seabed using a tele-operated robot. The developed blending technique focuses on two main ideas. Firstly, the richness of detail in the original images should not only be preserved but also enhanced when possible. Secondly, the algorithms should be able to deal with datasets of thousands of images covering large areas of the seafloor (on the order of several hundreds

of thousands of square meters). Consequently, the processing strategy needs to deal with underwater imaging while being well-suited for large input sequences.

There are various approaches to the problem of large-scale underwater photo-mosaicing in the literature. In the context of 2D mosaicing, Pizarro and Singh [9], [10] emphasized the global consistency of the result, obtaining very good results when dealing with large datasets (from hundreds to thousands of images). Richmond and Rock [11], [12], [13], instead, concentrated their efforts on reaching real-time performance in the mosaicing construction. The problem of obtaining a 3D reconstruction based on the fusion of images and navigation data was also addressed by Singh *et al.* [14], [15], [16]. Another 3D approach based on a stereo system has been proposed to recover a 3D estimate of the seabed morphology and to generate a navigable three-dimensional model for large areas [17]. This technique overcomes the well known problems of parallax on 2D mosaicing, although the detail of the 3D model is simplified to allow large sequence processing and fluid visualization. Unfortunately, the AUVs or ROVs performing the explorations are not always equipped with a stereo system, or the overlapping between consecutive images is not sufficient to perform a 3D reconstruction. Furthermore, the extraction of a mesh of triangles from the images acquired to generate the 3D model may lead to a loss in the detail of the information available from the original images. This is because when a texture, *i.e.* an image region, is applied to a triangle, even if it has been retrieved from the most appropriate view, it is typically stretched to fill the triangle's entire area. Consequently, the quality and sharpness of the texture information is reduced. The blending method, proposed in [17] and based on [18] (as used in [9]) for visualization efficiency reasons, may suffer from other known problems, such as ghosting and double contouring. A monocular 3D approach such as [19], [20] can be applied to perform Structure From Motion (SFM) when only information from a single camera is available. In that case, the small size of the overlapping area may also prevent the application of these kind of techniques. Furthermore, the complexity of the model will seriously constrain the size of the region to be mapped. Some approaches in the literature, such as [21] and [22], are able to deal with thousands of images, but with a

high computational cost and without achieving a full dense reconstruction. Concerning image blending in underwater imaging, only [9], [16], [23] have proposed methods to compensate the effects of light absorption and non-uniform illumination in the image preprocessing step.

Unfortunately, none of the techniques described above is suitable for our purposes due to several constraints. Our test dataset was acquired during the CNRS/IFREMER MOMAR08-Leg1 cruise by IFREMER's Victor-6000 ROV [24]. This vehicle is equipped with an OTUS 1-Mpixel still camera using artificial strobe lighting. Images were acquired at a height of about 8-10 m over the seafloor at intervals of between 10 and 17 seconds. Under these conditions, the reduced amount of overlap between successive frames is insufficient for the application of SFM techniques. The possibility of applying any 3D approach is therefore quite limited due to the lack of redundancy over time-consecutive images (*i.e.* multiple views of the "seam" area of the seafloor). The navigation information provided by an IXSEA[®] Octans sensor (heave, pitch, roll and heading) is used to estimate the initial topology of the photo-mosaic. The mapping payload of the Victor-6000 also features a RESON Seabat 7125 multi-beam echo-sounder with 512 beams. Nevertheless, this resolution, with respect to the field of view of the optical camera and the size of the scene elements, does not allow a 3D reconstruction at a resolution approaching that of the optical imagery.

The area surveyed, which extends $\approx 800 \times 700$ m, is located at a depth that ranges between 1600 and 1700 m. The area shows relatively moderate relief, with important slopes associated with fault scarps, in addition to three-dimensional structures, up to few tens of meters in height, corresponding to hydrothermal vents. Despite this topography, a 2D mosaicing approach that incorporates seam detection (reducing the disturbing effects of parallax) and blending presents itself as the most adequate for high quality underwater mapping. This provides a continuous and homogeneous view of large seabed areas. The obtained view is geo-referenced and that be readily exploited scientifically.

The main contributions of this paper are twofold. Firstly, it provides a comprehensive review of the current and most prominent state-of-the-art mosaicing and blending techniques, in order to evaluate their application in the underwater imaging context.



Fig. 1. Sequence corresponding to a straight trajectory of an AUV depicting the parallax problems. It shows the side and camera views of the robot's trajectory. One side of the chest disappears from the frame while the other arises due to the parallax effect.

Secondly, it proposes a new blending pipeline especially adapted to the underwater environment.

The rest of the paper is structured as follows. Section II presents a complete description of previous works in the field of image blending, pointing out the evolution of those algorithms and the context in which they were conceived. A classification criteria is proposed in Section III, followed by a summary table. A complete framework for underwater blending is presented in Section IV, along with results and discussion in Section V. Finally, conclusions derived from the results obtained and future work are presented in Section VI.

II. REVIEW ON STATE-OF-THE-ART BLENDING TECHNIQUES

Stitching two or more images together to create a photo-mosaic that enables the interpretation of the benthos by a scientist (biologist, geologist, archeologist, etc.) requires the use of a blending technique to obtain a seamless mosaic (see Figure 2).

There are three main concepts guiding image blending algorithms. Firstly, the effects of different illumination or exposure times between images should be minimized. Secondly, an adequate seam

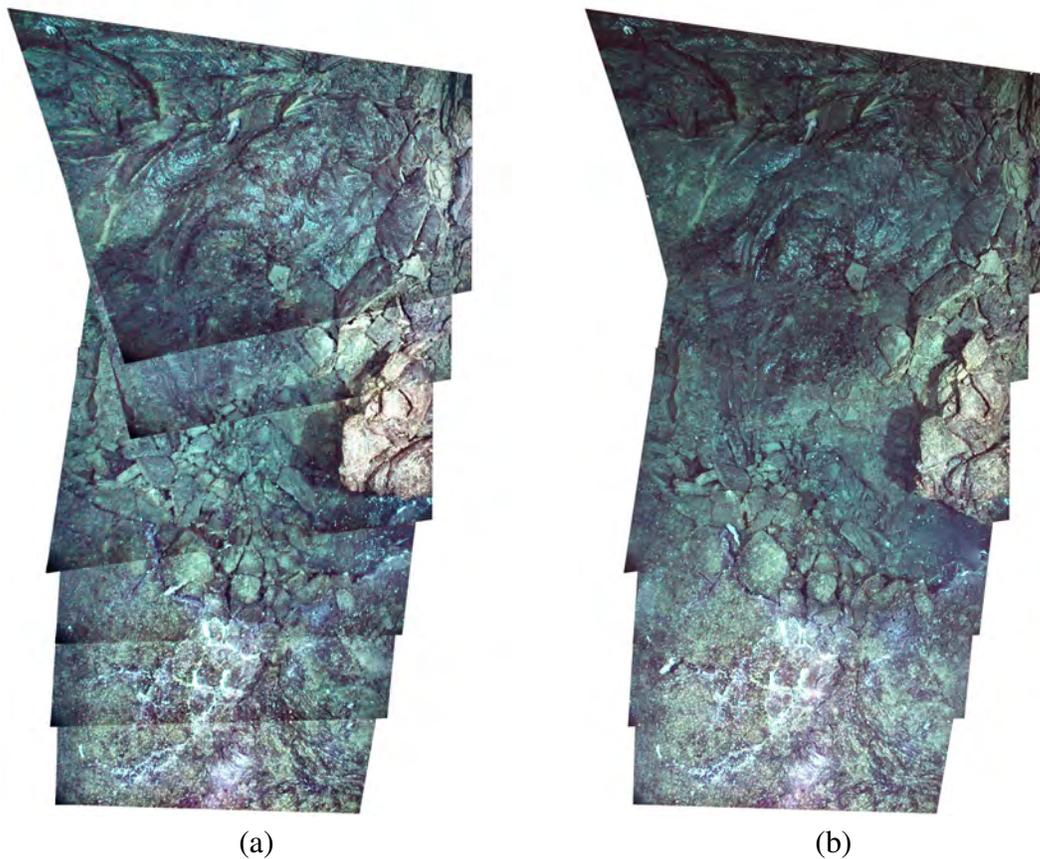


Fig. 2. Photo-mosaic built from six color images of two megapixels. The mosaic shows noticeable seams in (a), where the images have been only transformed and sequentially rendered on the final mosaic canvas, each one on top of the previous one. After applying a blending algorithm, the artifacts (image edges) disappear in the resulting mosaic (b). Images courtesy of Dan Fornari (Woods-Hole Oceanographic Institution).

should be found in order to reduce the visibility of micro-registration misalignments and moving objects. Lastly, a smooth transition through the selected seam must be applied to reduce the prominence of transitions between images.

In the early 70's, D. Milgram [25] addressed the problem of the seamless combination of two satellite images. The method searches the seam pixel offering the smoothest transition in a row-wise manner, and is intended to deal only with pairs of images horizontally registered. An arbitrary surround range is defined at each row around the seam pixel in order to smooth the transition using a weighted average. This algorithm established the basis for most of the methods that arose in the following decades. The author later proposed an improved approach, adding a pixel selection criterion in order to deal with shadows and moving objects [26]. A cost function permitted controlling the origin and the final pixel of the seam's path.

In the context of low-scale (order of mega-pixels) indoor photo-mosaicing, Peleg [27] introduced the

concept of *Seam-Eliminating Function (SEF)*. The *SEF* is based on a luminance smoothing function (*i.e.* a weighting map), computed by using a computationally expensive iterative relaxation algorithm, which is used to smooth the transition from an arbitrary number of overlapping images, setting the intensity differences along arbitrary seams at zero.

Burt and Adelson [18] introduced the concept of *Image spline* in 1983. The images to be fused are decomposed into a set of bandpass component images, and a separate spline with an appropriate transition width is applied to each band. The goal is to fuse the features with the same scale at each bandpass level. Finally, the splined bandpass components are recombined. The method suppresses the visibility of the seams and reduces the noticeability of any misalignments when registration is imperfect. However, it leads to double contouring and ghosting effects when the misalignment is significant (see Figure 3). In 1996, Hsu and Wu [28] extended the idea of Burt and Adelson [18] by applying the method in wavelet subspaces. The improvement

in the results obtained by Hsu and Wu [28] is negligible with respect to the method of Burt and Adelson [18], while being computationally more expensive.

The problem of non-static objects in the overlapping regions was addressed by Davis [29] in 1998, who found an optimal seam using Dijkstra's algorithm [30] through the photometric differences computed between two registered images. The obtained path tends to cut around a moving object, leaving it either totally in or out of the final mosaic image. In 2001 Uyttendaele *et al.* [31] proposed a method to suppress the ghosting effect in mosaic images due to moving objects, along with a block-based procedure to adjust the exposure across multiple images. The authors proposed a search for regions of difference (RODs) on the overlapping areas. The goal is to use information from only one image for each ROD, keeping information from one of the images, and ignore corresponding information in the other images. The RODs are used to build a graph, where the minimum weight vertex cover [32] must be computed. However, this method is not entirely robust, and situations can appear where a wrong elimination of RODs causes gaps in the overall mosaic.

In 2003, Pérez *et al.* [33] proposed a generic interpolation machinery based on solving Poisson equations for seamless editing and cloning of selected regions, being the first important approach in the gradient domain. Through suitably mixing the gradient of a given image with that of another one, it becomes possible to fuse image regions (namely objects) convincingly. As an extension of the technique presented by Bertalmio in [34], Pérez *et al.* proposed modifying the problem of image interpolation through the Poisson equation by introducing further constraints in the form of a guidance field. In the same context, Levin *et al.* [35] proposed a method based on several cost functions for the evaluation of the quality of the stitching. The authors named GIST (Gradient-domain Image STitching) the developed framework based on this method, providing two main approaches to image stitching. First, images are combined in the gradient domain. Second, the mosaic image is inferred by optimization over the image gradient. The drawback of the methods working exclusively in the gradient domain are the computational resources required to deal with large datasets, such as the examples

provided in this paper.

In 2004 Agarwala *et al.* proposed a technique which combined the two main classes of blending algorithms [36]. Firstly, *graph-cut optimization* [37] was used to find the optimal seam. Secondly, *gradient-domain fusion* [33] was applied to reduce or remove any remaining visible artifacts. The developed framework required user guidance to select the image interest regions, thus being unsuitable for the automatic generation of photo-mosaics. In 2007, Agarwala [38] presented a hierarchical approach to improve the efficiency of gradient-domain compositing. The efficiency increase was achieved by observing that the pattern of the differences between a simple color composite and its associated gradient-domain composite can be predicted *a priori*. These differences are solved by adaptively subdividing the domain using a *quadtree* hierarchical structure [39]. Unfortunately, when the number of overlapping images increases and the overlapping regions become smaller, the performance of the method decreases. Su *et al.* [40] proposed a method based on the minimization of a blending energy function with two variation terms. First, the *image value variation* measures the difference between corresponding pixel values of the images to be combined and the photo-mosaic itself. Second, the *first derivative variations* measure the difference between the blended values of each respective first derivative and the first derivative of the mosaic. The method does not address the seam finding problem, and its computational cost makes it unsuitable for large image datasets.

Regarding computational and memory cost reduction of Dijkstra's based optimal seam finding, Gracias *et al.* [23] proposed a method using watersheds and graph-cuts. The use of watershed segmentation to find possible cuts over areas of low photometric differences allowed their algorithm to reduce the search to a smaller set of watershed segments, at the cost of sacrificing a certain degree of precision in the computed path. Furthermore, the use of graph-cuts over image pairs guarantees a globally optimal solution for each overlapping region.

Eden *et al.* [41] presented in 2006 a blending approach including a two-step graph-cut procedure, working on a global radiance space. Firstly, the positions of moving objects in the scene are defined (manually or automatically). Secondly, the entire available dynamic range is used to render the photo-

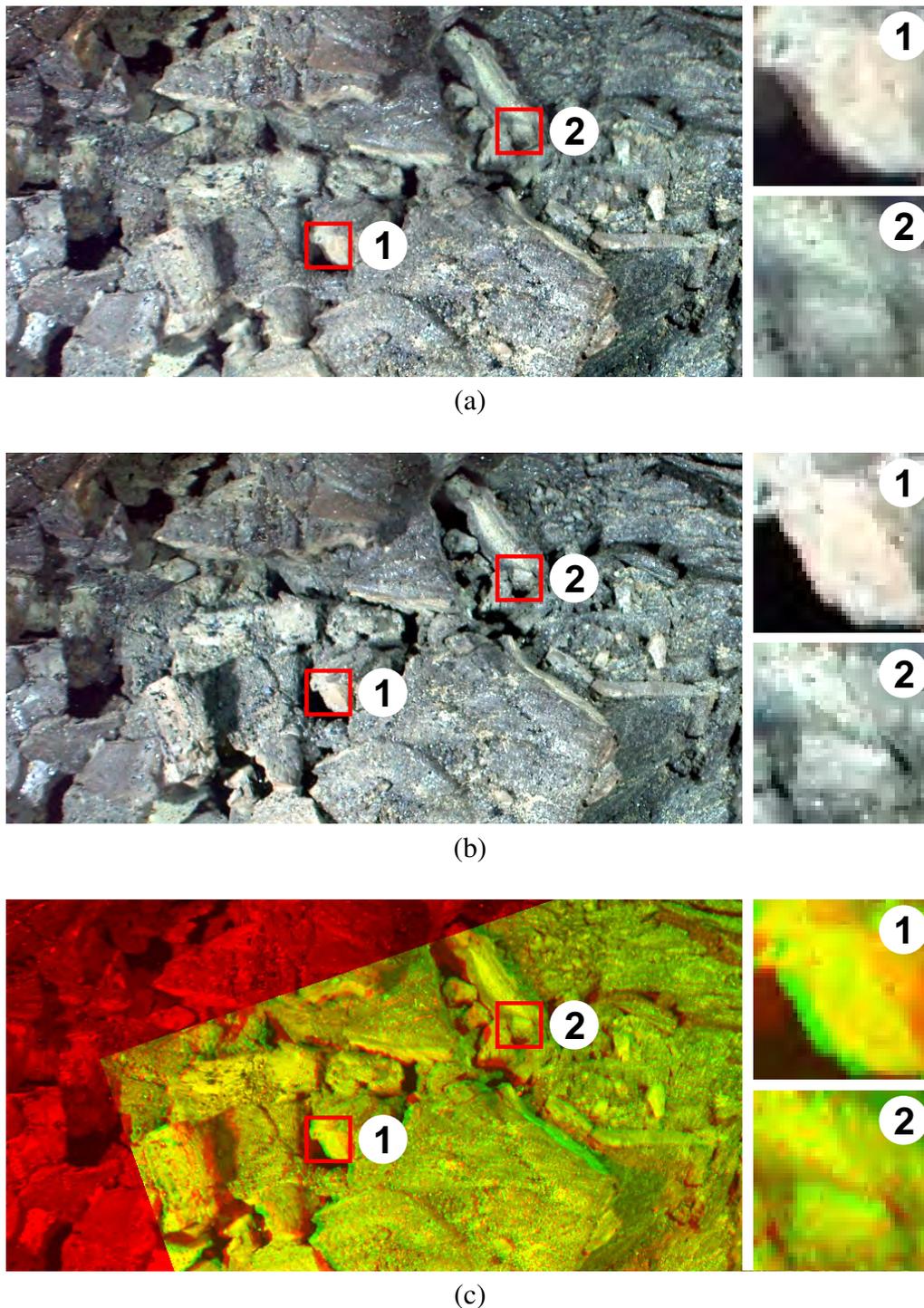


Fig. 3. Sample photo-mosaic region with (a) and without (b) ghosting and double contouring in the transition region due to registration inaccuracies. Seabed structures 1 and 2 are noticeably blurry in (a) while having a sharp appearance in (b). (c) shows two overlapping images of a given photo-mosaic (I_1 and I_2) represented in the red (I_1) and green (I_2) channels. Consequently, perfectly registered regions should appear in yellow, while regions showing misalignments, with a reddish or greenish appearance, are small and they appear as easily noticeable visual artifacts in the first image of the mosaic. The image without ghosting and double contouring has been obtained using the blending approach proposed in this study. Images courtesy of Dan Fornari (Woods-Hole Oceanographic Institution).

mosaic. Therefore, a High Dynamic Range (HDR) image can be obtained from the photo-mosaicing process. The gradient blending step is performed as in [36].

The problem of stitching images in real-time for online photo-mosaicing was addressed by Zhao [42] in 2006. There are three main advantages in the proposed flexible blending technique: a) good results and embedded implementation, b) comprehensive treatment of geometry, time and user control and c) exposure imbalance handling. Flexible blending has its basis in the sequential implementation of image blending features. Unfortunately, some drawbacks prevent its application to large-scale underwater mosaicing. Firstly, the selected strategy of the blending step [18] does not offer good enough results when registration problems occur. Secondly, the exposure correction mechanism selected may lead to a global exposure degeneration. Lastly, the behavior of the method when faced with large input image sequences is unknown.

Few approaches in the literature have specifically faced the problem of underwater imagery mosaicing. Gu and Rzhano [43], similarly to [36], proposed a graph-cut technique to select the optimal seam between two images, and the application around this boundary of a pure gradient domain fusion. The method claims to overcome the short comings of pure graph-cut techniques, which show noticeable seams in the case of changing illumination conditions, and gradient domain fusion, which produce blurring in case of misalignment. The authors do not define a criteria for selecting the contributing image in the case of multiple images overlapping in the same region. Thus, [43] is limited to “panoramic” mosaics where only two images overlap in the same area. Furthermore, a conventional graph-cut approach, in the case of gigapixel mosaics, may lead to non-optimal seams in the presence of different exposures or illuminations.

III. CLASSIFICATION OF TECHNIQUES

The list of papers that conform the state-of-the-art on image blending is long, and the main requirements for conventional panorama image generation have been satisfyingly addressed by several of them. Unfortunately, blending in underwater photo-mosaicing is a specific application, which has not been treated thoroughly in the literature. Consequently, despite the numerous image mosaicing

methods, not all of which are adequate to deal with large-scale underwater photo-mosaicing. In order to highlight the properties, benefits and drawbacks of the current methods, and with the aim of evaluating their application to underwater mosaicing, a classification is proposed.

There are several criteria that determine the behavior and performance of a given blending algorithm, including their capability of dealing with high resolution underwater photo-mosaics. Table I provides a comprehensive comparison of the most relevant blending techniques proposed in the literature. The specially important categories for underwater applications (mostly working with monochrome images) are *exposure correction* and elimination of *ghosting* and *double contouring*, concerning image quality and *scalability* in large-scale photo-mosaicing.

A. Basic Principle

Two main groups of algorithms can be found in the literature in the context of image blending [35]: *Transition Smoothing* (TS) and *optimal Seam Finding* (OS) techniques. *Transition smoothing* methods, also known as *feathering* [31] or *alpha blending* methods [46], attempt to minimize the visibility of seams by smoothing the common overlapping regions of the combined images. TS methods often suffer from *Ghosting*, a blurring of the finest details (*i.e.* low frequency image components), and *Double Contouring*, consisting in practice on a partial duplication of certain scene structures (*i.e.* high frequency image components), if registration is not accurate enough or the scenario considerably violates the planar scene assumption for 2D mosaicing. *Optimal seam finding* methods place the seam between images where photometric differences in their joining boundaries are minimal [47], [29]. OS methods are not able to deal with images with different *Exposures*, as it is often the case in underwater imagery due to 3D relief, oblique terrain, variations in the altitude of the vehicle, etc. Finally, *Hybrid* (TS/OS) methods combine the benefits of both algorithm groups (*e.g.* [25], [36]).

B. Domain

The *Domain* in which the process is carried out (*Luminance / Radiance*, *Wavelet* or *Gradient*) strongly influences the properties of the blending.

TABLE I
BLENDING TECHNIQUES: COMPARISON TABLE

Method	Principle*	Domain*	Scalability	Color*	Multiresolution	Ghosting Removal	Double Contouring	Exposure Correction	Moving Objects	Real-Time	Sequential Processing	Locality	Dynamic Range	Parallax Robustness
(1975) D. L. Milgram [25]	TS / OS	L	Low	GS	No	No	Yes	Yes (*)	Indirect	No	Yes	Local	Low	Yes
(1977) D. L. Milgram [26]	TS / OS	L	Low	GS	No	No	Yes	Yes (*)	Indirect	No	Yes	Local	Low	Yes
(1981) S. Peleg [27]	TS	L	Yes	GS	No	Yes	No	No	No	No	No	Global	Low	No
(1983) P. Burt & E. Adelson [18]	TS	L	Low	CW	Yes	No	Yes	No	No	No	Yes	Local	Low	No
(1996) C. T. Hsu & J. L. Wu [28]	TS	W	Low	CW	Yes	No	Yes	No	No	No	Yes	Local	Low	No
(1998) J. Davis [29]	OS	L	Yes	CW	No	Yes	No	No	Yes	No	No	Global	Low	Yes
(2001) M. Uyttendaele et al. [31]	OS	L	Yes	CW	No	Yes	No	Yes	Yes	No	No	Global	Low	Yes
(2003) P. Pérez et al. [33]	TS	G	Low	CW	No	No	Yes	No	No	No	No	Local	Low	No
(2004) A. Levin et al. [35]	TS	G	Low	CW	No	No	Yes	No	No	No	Yes	Local	Low	No
(2004) A. Agarwala et al. [36]	TS / OS	G	Yes	CW	No	Yes	No	No	Yes	No	No	Global	Low	Yes
(2004) M. Su et al. [40]	TS	L	Low	CW	Yes	No	Yes	No	No	No	Yes	Local	Low	No
(2006) A. Eden et al. [41]	TS / OS	R / G	Yes	CW	No	Yes	No	No	Yes	No	Yes	Global	High	Yes
(2006) W. Zhao [42]	TS	L	Yes	CW	No	No	Yes	Yes	No	Yes	Yes	Local	Low	No
(2006) F. Gu & Y. Rzhanov [43]	TS / OS	G	Low	CW	No	Yes	No	No	Yes	No	Yes	Local	Low	Yes
(2007) A. Agarwala [38]	TS / OS	G	Yes	CW	No	Yes	No	No	Yes	No	No	Local	Low	Yes
(2008) R. Szeliski et. al. [44]	TS / OS	G	Yes	CW	No	Yes	No	No	Yes	No	No	Local	Low	Yes
(2009) N. Gracias et al. [23]	OS	L	Yes	CW	No	Yes	No	No	Yes	No	Yes	Local	Low	Yes
(2009) A. Mills & G. Dudek [45]	TS / OS	L	Yes	CW	No	Yes	No	Yes	Yes	No	Yes	Local	Low	Yes
(2010) Johnson-Roberson et al. [17]	TS	L	High	CW	Yes	No	Yes	No	No	No	Yes	Local	Low	Yes
(2012) Proposed Approach	TS / OS	L / G	Yes	SC	No	Yes	No	Yes	Yes	No	No	Local	High	Yes

* TS: Transition Smoothing, OS: Optimal Seam, L: Luminance, W: Wavelet, G: Gradient, R: Radiance, GS: Gray Scale, CW: Color Wise, SC: Single Channel.

Gradient blending methods, widely used in the latest publications [44], [48], [45], are able to unify different *Exposures* seamlessly and can lead implicitly to high dynamic range from a set of low dynamic range images. However, they require solving large sparse systems of equations to recover the *Luminance* from the gradient vectors. In contrast, *Luminance* based methods typically have lower computational requirements.

C. Scalability

We define the *Scalability* of a method as the ability to deal with more than two overlapping images. This property might be constrained by two main factors. The first is the nature of the method itself, as in [25], [26] or [28], which cannot work with more than two overlapping images. The second is related to computational requirements: non-optimized *Gradient* algorithms suffer from poor computational scalability when the input dataset is extremely large, as in the case of Giga-Mosaics.

D. Color and Dynamic Range

The acquired *Color* changes significantly as a function of the robot's altitude due to the

wavelength-dependent spectral absorption of the media. Mosaic blending techniques generally use a *Channel Wise* approach, where three color channels are processed independently and later reunified into a single color image. Those methods have no control over perceptual color attributes. Several approaches in the literature address the color balancing problem in the image photo-mosaicing pipeline, based on exposure compensation in single [31], [49] or multiple channels [50], [51] and on color transfer techniques [52], [53]. Unfortunately, when dealing with extremely large datasets, keeping the consistency of the image's global appearance is a difficult task when using methods available in the literature.

The *Dynamic Range* of the image and the quantization of the data provided by the camera sensor strongly influence the accuracy of the final scene representation. Any *High Dynamic Range* blending method will require a *Tone Mapping* algorithm, in order to display the *High Dynamic Range* mosaic image onto a *Low Dynamic Range* device, like conventional screens or printers.

E. Multiresolution

The main advantage of the *Multiresolution* approach (Burt and Adelson [18]) is the significant

reduction of the noticeability of *Double Contours* due to registration inaccuracies. An important shortcoming is that the method requires keeping several representations of the same image in memory. The price of the seamless appearance comes at the loss of high frequency details. The *multiresolution* approach is applied by Su *et al.* [40] to the wavelet domain.

F. Local / Global and Real-Time Operation

Global methods require knowing all the final mosaic information beforehand, while *Local* methods can work on small parts of the final photo-mosaic, joining them together upon completion. Obviously, *Global* methods often require high computational resources. *Local* methods may not be able to solve some problems such as exposure compensation during a pair-wise sequential processing.

Methods able to work in *Real-Time* [42] are optimized towards a high performance for large sequences. The results obtained are not as accurate as those from off-line approaches, but are acceptable when on-line feedback is required. *Real-time* techniques are typically based on the *Sequential Processing* of the input data. Methods like [25] or [28] can process images pair-wise and add the result to a final mosaic canvas. Nevertheless, this strategy is not appropriate for long sequences where the drift accumulated due to the sequential registration and changes in the exposure times may result in inconsistent overlapping regions.

G. Relevant Visual Performance Criteria

Different *Exposures* between images are especially common in underwater imaging, mainly due to the difficulty of the AUV or ROV in keeping a perfectly constant altitude during the survey. This fluctuation in altitude also causes changes in the scene's illumination conditions, because the artificial light sources are part of the vehicle. Consequently, the camera requires a constant readjustment of the automatic exposure. The *Exposure Correction* might be performed *actively*, by preprocessing the image sequence, but may also be corrected by means of a *Gradient* domain techniques, given that this domain is not sensitive to time exposure changes. Unfortunately, wrong automatic exposure adjustments may lead to loss of information on overexposed regions, which cannot be corrected. *Ghosting* and *Double Contouring* are mainly

due to geometrical registration inaccuracies. When two overlapping images are not properly aligned, non-coincident features are smoothed, and thereby *Ghosted*, when fused using a transition smoothing method, while strong contours appear twice in the blended photo-mosaic. Underwater, the *forward scattering* phenomenon is responsible for loss of contrast [54] and, thereby, ghosting appears when merging images from significantly different depths (see Figure 4). *Moving Objects* often appear in underwater imaging, *e.g.* fish, algae, crustacea and other life forms, as well as floating objects. Most *Optimal Seam Finding* algorithms are able to deal with *Moving Objects*, keeping a single representation of each object in the final map. The *Parallax Robustness* determines the ability of a given blending algorithm to deal with a sequence where the 2D assumptions were violated, *i.e.* when the relief of the scene is not negligible compared to the distance from the camera to the scene. Underwater scenarios are characterized by frequent seabed depth-changes, as well as the direction of shadows produced by the artificial lighting systems of the AUV or ROV. *Optimal Seam Finding* techniques are typically the most indicated methods to deal with this problem. The *Parallax Robustness* is also strongly related to tolerance to moving objects, since methods able to deal with moving objects are often able to handle parallax issues.

IV. PROPOSED FAMEWORK

A photo-mosaicing pipeline has been conceived to address the most relevant problems particular to underwater imaging. Nevertheless, the application field of the presented approach can be extended to the generation of conventional panoramas of terrestrial or aerial images. Figure 5 shows the sequence of steps that are performed in our approach and applied to the construction of a high resolution blended photo-mosaic from the deep-seafloor.

A. Input Sequence Preprocessing

Apart from *exposure variations*, which are a common issue in terrestrial images, the rest of the problems are not directly addressed by conventional panorama generation software. To deal with inherent underwater imaging problems (*non-uniform illumination*, *light attenuation*, *scattering*,

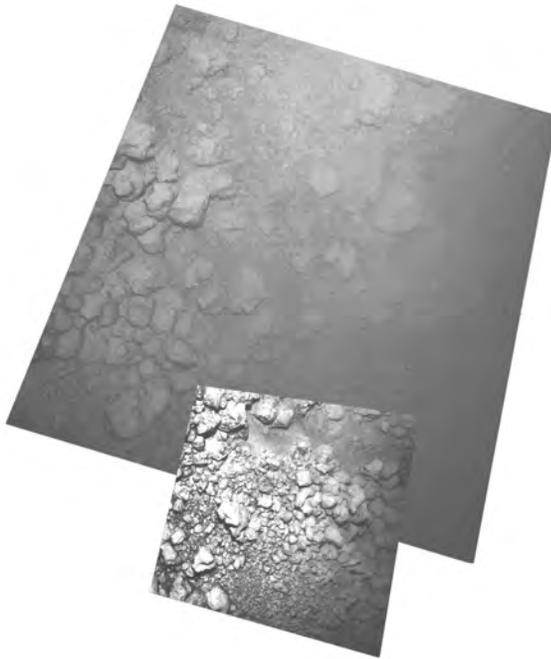


Fig. 4. Registration of two images acquired at significantly different altitudes. The image acquired at the higher altitude shows strong light attenuation and scattering. These effects cause a noticeably different appearance between the two images.

exposure variations, etc.), we perform image pre-processing, which in our experience is a key step, strongly impacting the quality of the final photo-mosaic rendering.

1) *Inhomogeneous Lighting Compensation*: The *lighting inhomogeneity* problem in deep water is due to the lack of natural global (or ambient) lighting, and to the necessity of using artificial light sources with limited power. The illumination systems are rigidly attached to the AUVs or ROVs, and typically concentrate the light rays in the area on which the camera is focused. The acquired image borders suffer from darkening, due to *light attenuation*, principally induced by *light absorption* of the water. The effect is similar to *vignetting*, although the phenomenon is not produced by the camera lens but by the medium itself. All images from a given sequence are affected to some degree by this issue. The illumination distribution from artificial light sources changes with the distance between the camera and the seafloor. Colors are also affected due to *light absorption*, resulting in *depth-dependent* color profiles in the images acquired, *i.e.* different numerical representations corresponding to the same real scene colors.

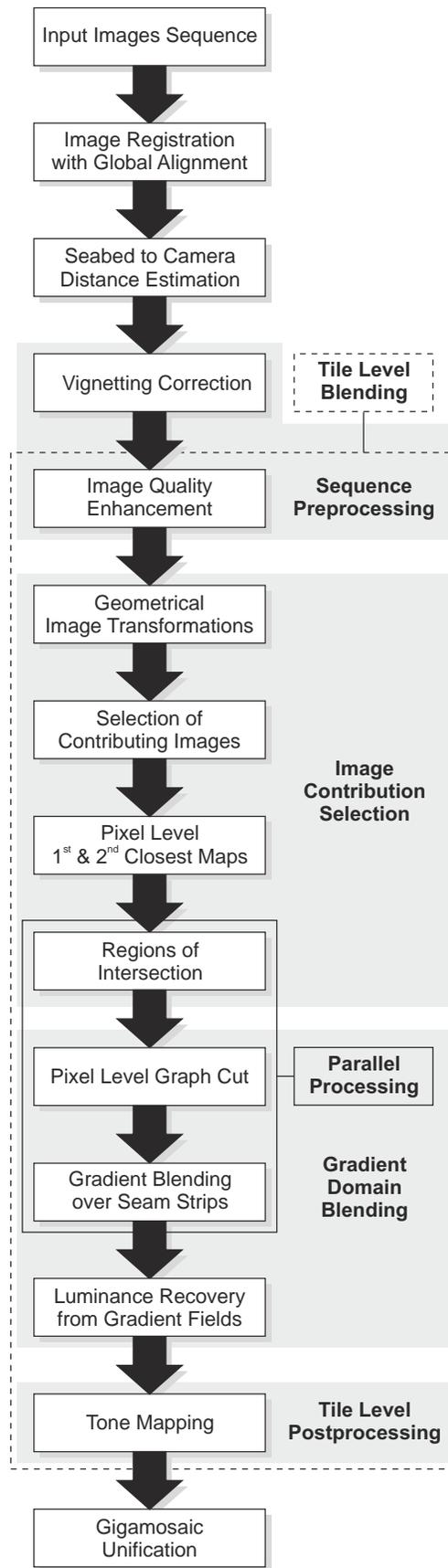


Fig. 5. Full processing pipeline of the proposed underwater photo-mosaicing approach. Some of the processing steps can be executed using parallel computing techniques to increase the performance of the algorithm.

Imaging conditions hinder the application of a unique compensation function on all the acquired images in absence of precise information on the placement and nature of the light sources, the distance between the camera and the seabed, and the 3D structure of the scene. These circumstances result in the loss of a global terrain perception, which is a cognitive sensation factor highly dependent on lighting consistency.

A feasible correction of *lighting inhomogeneity* and *vignetting-like* artifacts in a single step consists of the application of a 2D “inverse illumination distribution” to the original input images [55], [56], [57], [58]. The main aim of this operation is to enhance the luminance of the darkened image borders in order to obtain uniform illumination throughout the image. If a high sensitivity camera with a high pixel depth ($> 8bpp$) is available, not only the luminance but also the detail richness can be enhanced in the region affected by *light absorption*.

The illumination pattern describing the “inverse illumination distribution” function can be estimated from a subset of pictures showing low texture and reduced 3D structure (i.e., flat, sedimented terrain). As this function changes with the distance from the light source to the seabed, a three-step approach is proposed (Figure 6) to correct the lighting artifacts. It is based on two main ideas: (1) the application of a depth dependent inverse illumination distribution, and (2) the automatic selection of images to compute this pattern in a given depth-range based on the Total Variation (TV) metrics [59] described in the next section.

Quasi-altitude Estimation

Underwater image acquisition platforms usually record not only image sequences but also other synchronized data like surface GPS positioning, acoustic positioning, heading and altitude, among others. Unfortunately, image altitude is not always available for every data set. Consequently, as a first step, the images of a given sequence should be classified by altitude in order to apply a different lighting correction function to each one, assuming that precise information on the distance from the camera to the seafloor is not available. In order to solve this issue, a quasi-altitude estimation is proposed instead.

Given a sequence of images and their corresponding registration parameters in the photo-mosaic

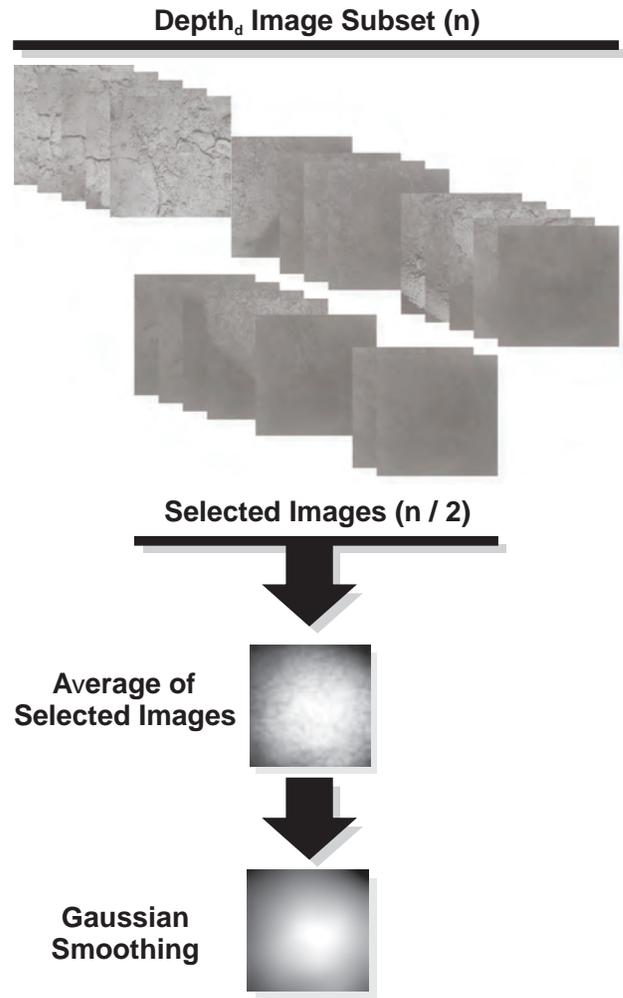


Fig. 6. Lighting Pattern Compensation procedure. The images of a sequence are classified into depth subsets, and a different lighting pattern compensation function is computed for each one. The figure shows a given set of n images from which the $n/2$ images having the lower Total Variation (TV) value have been selected. Next, the images are averaged and the result normalized and smoothed using a Gaussian filter with an adaptively selected σ .

frame, it is possible to determine which ones were acquired closer to the seabed and which ones farther, by computing the size of a footprint or a scale in the image once it is registered in the 2D photo-mosaic coordinate system. Specifically, it is possible to consider only the diameter of the transformed pictures (i.e. the size of the longest diagonal) since this scale and the altitude are highly correlated when the focal length is assumed constant. Once an image list has been built and sorted according to their diagonal length, images can be classified into subsets of similar altitudes.

Image Selection

For each image subset, a distinct compensation

function for the light distribution should be computed from images with low texture content and a homogeneous appearance. Low textured images are best suited for this estimation due to their low average gradient length. An adequate ranking metric for the selection of these images is the (TV).

$$TV = \frac{1}{W \cdot H} \sum_{x=1}^{W-1} \sum_{y=1}^{H-1} \|g(x, y)\| \quad (1)$$

Equation 1 shows the computation of the normalized TV for a given image, where W and H are the width and height sizes and $\|g\|$ denotes the L_1 or L_2 norm of the g gradient vector. The gradient vector components for the last row and column in a given image are set to 0.

Our approach uses L_2 norm, *i.e.* Euclidean metrics, to evaluate the homogeneity of the images. Once the TV measure has been computed for all the images of a given altitude subset, an image subset of low TV is used to estimate the light distribution. The aim of this measure is to identify images containing structures rich in detail. The presence of high frequency noise, mainly due to scattering on macroscopical particles in suspension of scattering (see Figure 7), may skew the image quality evaluation. The TV magnitude of the image may be increased inappropriately leading to scenarios in which the dominant part of the metrics comes from high frequency noise. Nevertheless, the unwanted effects of the high frequency components can be avoided by building lower resolution images from the originals with $N \times N$ super-pixels. This simple approach significantly reduces the effects of the high frequency components in both the image and the TV measure. In our experiments, 8×8 linearly averaged super-pixels have been used. Thereby, the original images of 1024×1024 pixels have been reduced to 128×128 pixels. The images obtained preserve the most prominent seabed structures, but strongly reduce the effects of the scattering phenomena, allowing the usage of the TV as an image quality evaluation metric. For each depth-range, the images with a TV value below the median have been used to compute the illumination correction function. To obtain this function, the selected images are averaged and the result is smoothed by a low-pass Gaussian filter to reduce the remaining high frequency components, as explained next.

Gaussian Filtering

In order to compensate for the light attenuation problems and obtain an image with a homogeneous illumination l_H , the acquired luminance values are divided by a given l_G compensation mask as shown in Equation 2.

$$l_H(x, y) = \frac{l(x, y)}{l_G(x, y)} \quad (2)$$

where l is the image luminance value and l_G is the illumination pattern.

$$l_G(x, y) = \frac{1}{N} \sum_{k=1}^N l_k(x, y) \quad (3)$$

The lighting compensation pattern l_C before the Gaussian smoothing is obtained using Equation 3, which computes the average value for every pixel position given a stack of N images. Finally, the obtained compensation mask l_C is smoothed with a low-pass Gaussian filter to obtain the illumination distribution l_G function. This distribution is then used for the *lighting inhomogeneity* compensation, as per Equation 4, where $\langle \rangle$ denotes Gaussian smoothing.

$$l_G(x, y) = \langle l_C \rangle \quad (4)$$

The value for σ , used in the Gaussian convolution, is selected adaptively for each altitude subset. Starting from the average image l_C in Equation 3, a set of increasing values $\sigma_1, \sigma_2, \dots, \sigma_k$ is sequentially applied, until the resultant smoothed TV reaches a value under a threshold $TV(l_{G(\sigma)}) < \varepsilon$. In our experiments we used values in the $\frac{d}{256}, \frac{d}{128}, \dots, \frac{d}{32}$ range, where d is the shortest dimension of a given image. With this threshold condition, the appropriate smoothness and uniformity of the blurred image are ensured.

2) *Gradient based Image Enhancement*: As the altitude of the robot increases, new problems appear. One of them is *backward scattering* (see Figure 7a), which is an additive noise in the form of “marine snow” patterns due to light reflection from suspended particles. Another is *forward scattering* produced by the local light inter-reflections from the particles (see Figure 7b), including water molecules, which causes the loss of contrast and sharpness in the images. When artificial light is

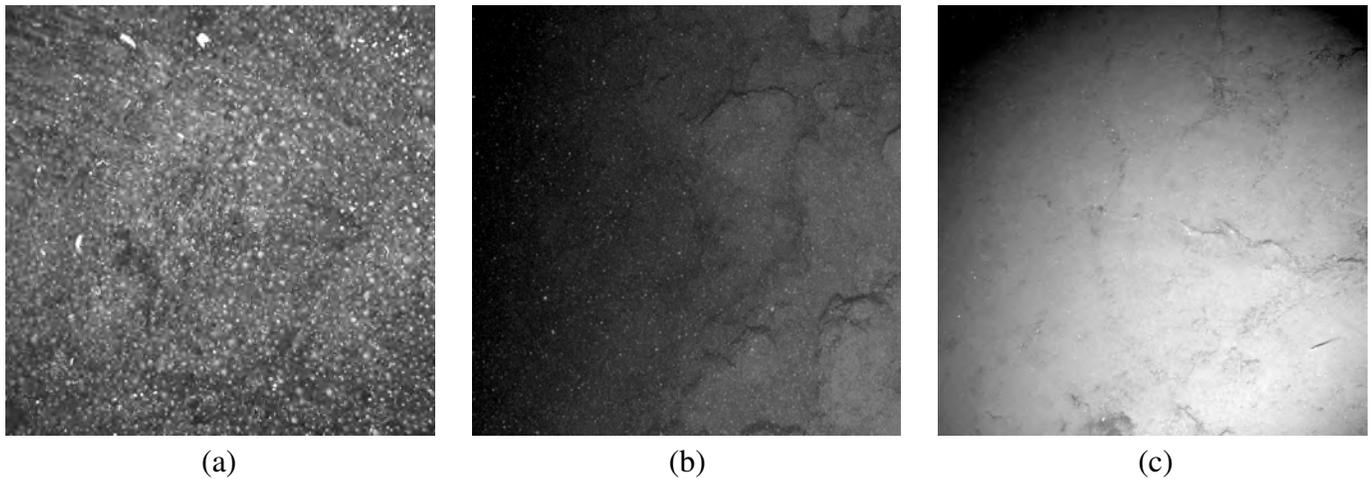


Fig. 7. (a) Example of *backward scattering* due to the reflection of rays from the light source on particles in suspension, hindering the identification of the seafloor texture. (b) Example of *forward scattering* caused by the local inter-reflection of the light suspended particles, hiding the terrain behind them. (c) Effects produced by *light absorption* of the water resulting in an evident loss of luminance in the regions farthest from the focus of the artificial lighting.

shed over particles in suspension, part of the light is reflected back due to *backward scattering* and part is diffused in the medium, contributing to image blurriness. This altitude-dependent blurring due to *forward scattering* is a most troublesome source of image quality degradation, and is particularly disruptive when registering images affected by this phenomenon at different levels. Finally, another problem is *light absorption* (see Figure 7c), which leads to an exponential luminance loss, resulting in increasing image noise after a luminance correction.

The strategy to enhance the high frequency details affected by the phenomena described above is a simple and global approach, selecting the highest quality image in a given surrounding region from the whole set (as is explained in the *Image Quality Estimation* section), and using it as a contrast or gradient reference. The aim is to obtain a visually consistent area by providing all images with a similar appearance to the reference one. To avoid unpredictable visual effects, the non-global approaches of homomorphic filtering [60], CLAHE [61] (Figure 8) and histogram specification [52] are not used, due to the following reasons. On one hand, homomorphic filtering may lead to an excessively homogeneous appearance in the filtered image, and to a loss of global consistency in the appearance of the photo-mosaic. The suppression of low frequencies performed by this kind of filter may provide some advantages in the visibility of local details, but in gigamosaicing, depending on the zoom factor, every

spatial frequency can be important to recognize and understand the nature and morphological attributes of the seabed structures. On the other hand, the histogram specification is highly dependent on the reference image, therefore the modified image may often lose its realistic appearance. Thus a simple but robust local contrast stretching has been applied to equalize a given sequence of images.

Image Quality Estimation

There is not a single and objective criterion to identify the image with the highest visual quality from a given set because the concept of “quality” involves different cognitive aspects. However, phenomena that affect image detail richness and sharpness, like *scattering* and *light absorption*, are known to grow with the distance from the camera to the seabed.

This simple and fast approach may lead to poor results when the selected image presents an over-exposed region due to being acquired too close to the seabed under strong illumination, for example. A more robust selection of a reference image is to use the *TV* to rank image quality as well. Thus, the image with highest *TV* may be selected as the reference image while ensuring that over-exposed regions will not be included in this selection. According to our experimental validation, the image with the highest *TV* coincides, in most of cases, with the closest one taken to the seabed, and with the second or the third closest images in the few

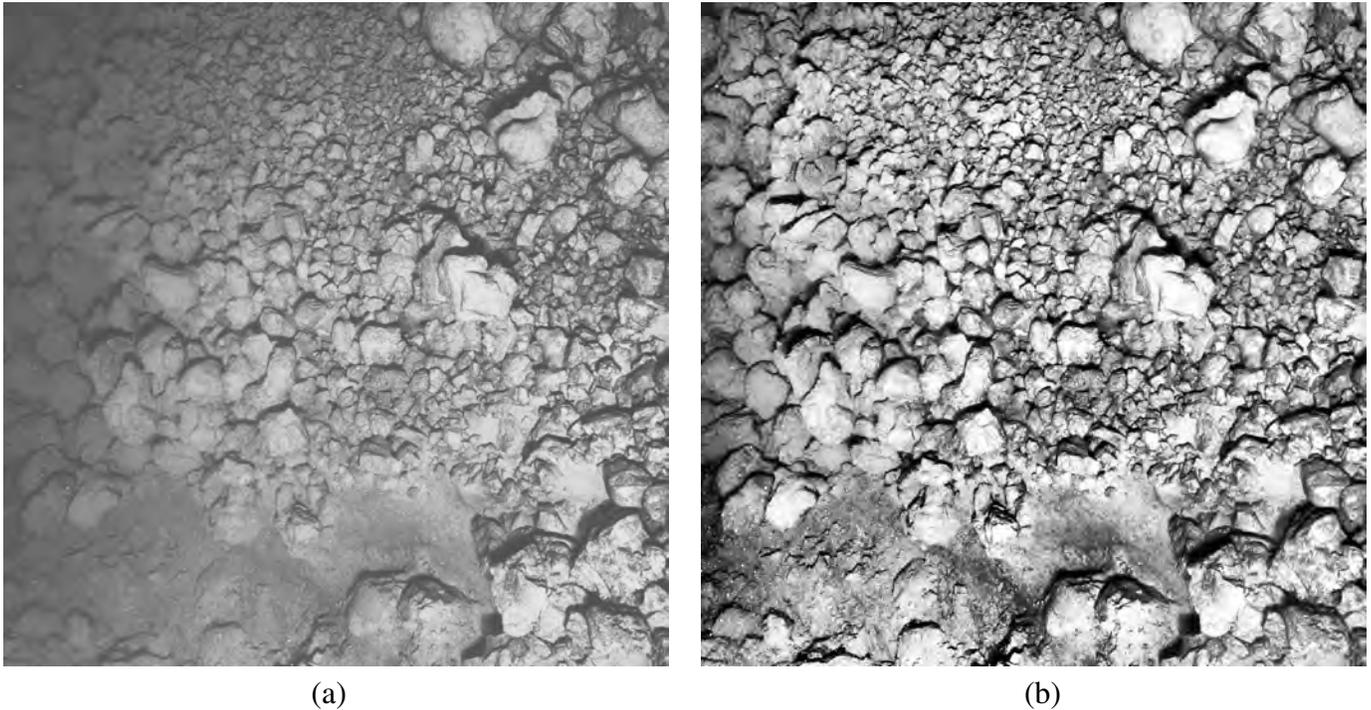


Fig. 8. (a) Image lacking contrast on its left side. (b) Image processed with a Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm, showing enhanced details in the originally lower-contrast regions. The appearance of the processed image is less realistic than the original due to an aggressive level of local filtering.

remaining cases.

Global Contrast Stretching

The TV value of the selected reference image is used to compute the stretching factors that will be applied for a global contrast (or gamma amplification) to all the other images. This stretching factor should be below a given threshold T_s to avoid overamplification of areas of poor contrast, *e.g.* textureless sediment-covered regions. T_s depends on the Signal to Noise Ratio (SNR) in the image, which can vary highly depending on water quality, lighting intensity, and/or the camera sensor. An experimentally selected value of $T_s = 5$ has been used in the experiments performed. Despite the application of these gradient corrections, the merging of images from highly different depth categories will produce noticeable seams due to their distinct blurring levels. This problem is, in itself, a challenging field of research in underwater imaging.

The stretching factor $\frac{TV_{reference}}{TV(k)}$ is applied to enhance the x and y gradient components of the k -th image.

B. Image Registration with Global Alignment

While image registration is not directly related to the blending procedure, and therefore is not at the

core of the work presented here, the quality of image registration significantly affects the final quality of the rendered photo-mosaic.

Even when navigation data (such as USBL positioning, heading, depth, etc.) are available, pair-wise image registration is still required to ensure a precise camera motion estimation. Pair-wise registration has been performed in our experiments using a feature-based approach, mainly involving the well known image feature detectors and descriptors of SIFT [62] and SURF [63]. When building a 2D photo-mosaic from a set of images acquired by a camera close to the seabed, the *planar assumption* of the scene can be violated due to the microbathymetry of the seafloor. The 3D geometry of the scene, in addition to the short camera distance, results in *parallax*. This problem increases the difficulty of estimating the 2D planar transformation between consecutive images, often leading to *misregistrations*.

A global alignment strategy [2], [3] is required to reduce the inaccuracies in a simple sequential pairwise registration. The strength of the global alignment arises from closing-loops, because they allow us to improve the camera trajectory estimate significantly when re-visiting an already mapped area. In absence of loop-closings, and considering

input sequences of thousands of images, the drift accumulated by the pairwise transformations leads to significantly inconsistent (misaligned) photo-mosaics.

C. Image Contribution Selection

The *parallax* effect influences both the image registration and image blending procedures. On one hand, image panorama software often fails to register sequences with strong *parallax* since they only assume camera rotation. On the other hand, and even using the best possible registration, the double contouring problem will appear when merging two or more images, if the vehicle (and the camera) translates and the scene is not perfectly planar.

The solution to avoiding ghosting artifacts is the use of information from a single image for each pixel in the final photo-mosaic whenever possible. Blending is performed in a narrow region around the optimally computed seams, and consequently, information on more than one image is fused only in a small fraction of the final photo-mosaic. Ghosting may occur in these regions, but is significantly constrained to specific areas, and is dependent on the width of the transition region.

1) *Image Discarding*: First, the frames of the original images are mapped into the global photo-mosaic frame using the image registration parameters in order to know their shape and coverage area in the final photo-mosaic coordinate system. The altitude estimation is computed, assuming that depth information is not available in the navigation data. It is possible to discard low quality images covering a region in the scene if higher quality images are available for this same region. The discarding procedure is performed using logical operations on the polygons describing the images, which is an efficient and low resource usage approach.

Each image is defined as a quadrilateral described by four vertices corresponding to the four corners of the image once registered in the photo-mosaic frame. Additionally, the polygons are sorted decreasingly according to their corresponding image *TV* value. At each step of the iterative process, a new image trapezoid of the sorted list is added to the final photo-mosaic polygon using simple binary operators. If the area covered by the new trapezoid has already been fully covered by the photo-mosaic polygon (*i.e.* the quadrilateral does not intersect the

photo-mosaic polygon and lies inside this one), the image is discarded, because this same region has already been covered by higher quality images. Otherwise, the photo-mosaic polygon is updated and the image is accepted. With this strategy, information on lower quality images will not be used in the final photo-mosaic image. A consequence of this image selection is a speeding-up of the later steps in the process.

2) *Pixel-Level First and Second Closest Maps*:

The proposed blending methodology determines the first and second closest maps at pixel level. The *first closest map* contains, for each pixel coordinate in the photo-mosaic, the index of the image whose center is the closest (see Figure 9). The *second closest map* does the same, but with the second closest image indices. Similarly to [23], the overlapping of these two maps will use a graph-cut algorithm to compute the seam-strips for blending. Two image indices are selected for every *seam pixel*. Therefore, every pixel outside the seams (most of the photo-mosaic) is associated to a single image.

The Euclidean distance between pixel $I^M(x, y)$ in the photo-mosaic frame and the center of a given n -th image $I^n(x, y)$ is weighted by a factor $w_n(s)$, as shown in Equation 5:

$$d_M^n(x, y) = w_n(s) \cdot \sqrt{(x_M - x_n)^2 + (y_M - y_n)^2} \quad (5)$$

where the scalar factor $w_n(s)$ is a size-ratio between the n -th image and the image having the smallest area once registered. For time efficiency reasons, the ratio is not computed based on the area of the warped images, but on the length of their diameters, as explained in the *Quasi-depth estimation* section, to obtain a rough and fast approximation, as shown in Equation 6:

$$w_n(s) = s_{min}/s_n \quad (6)$$

where s_{min} is the diameter of the smallest image for a given set and s_n is the diameter of a given n -th image.

This weighting prioritizes pixels from images acquired at low altitude, close to the seabed, and consequently less affected by the imagery artifacts described above. This weighting also maximizes the contribution of “higher-quality” images to the final photo-mosaic image. Therefore, in cases like that

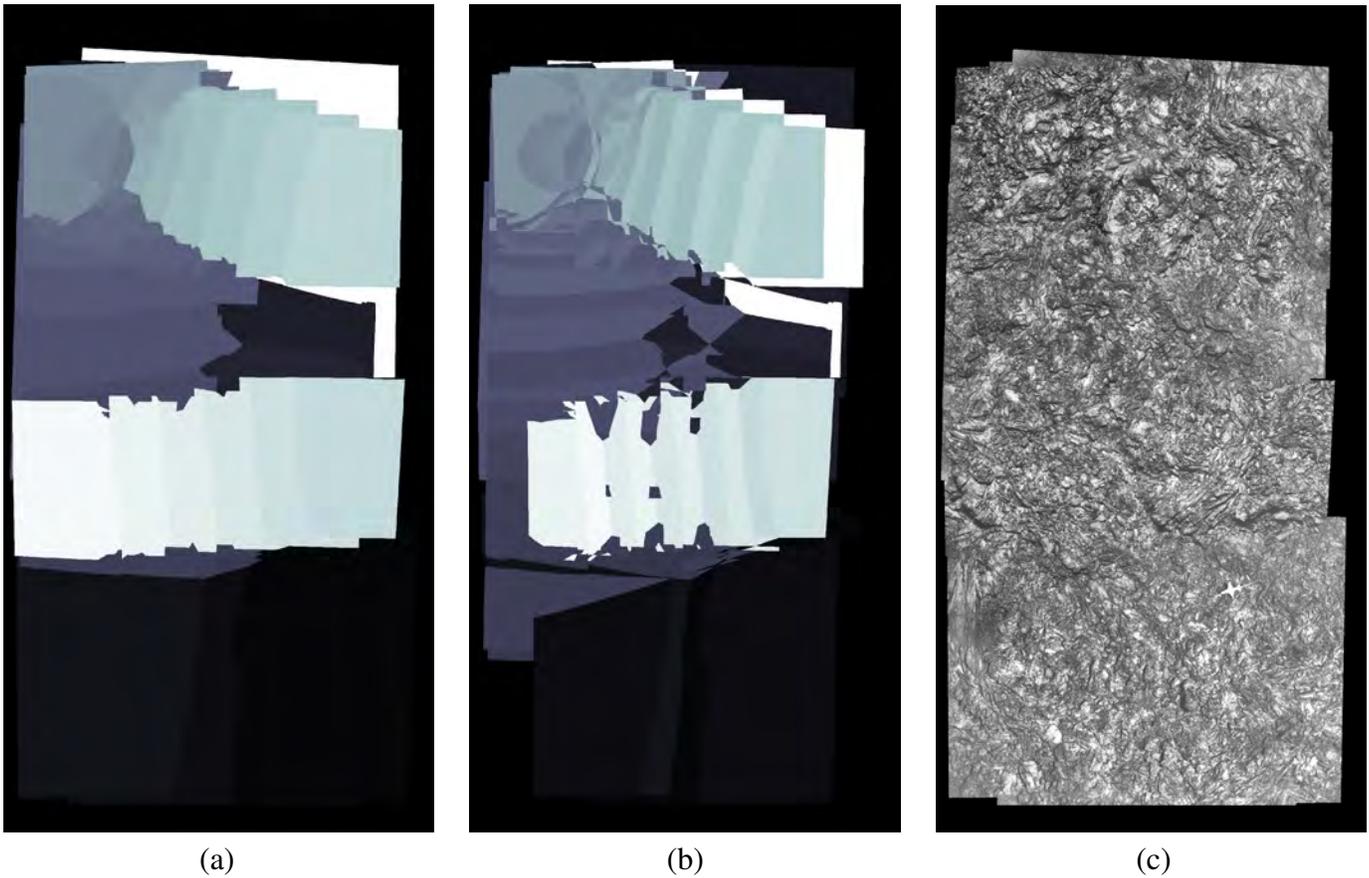


Fig. 9. (a) First closest map and (b) second closest map corresponding to the registered images finally blended into the (c) photo-mosaic. Each pixel's blue level in the closest maps represents the index of the image having the closest and second closest image centers. The distance measure gives more priority to pixels belonging to images which have been acquired at a lower altitude, consequently showing a higher level of detail richness.

shown in Figure 4, only a small percentage of the pixels from the smaller overlapping image are lost while computing the smooth transition, while the most significant percentage of the original image is preserved.

3) *Regions of Intersection*: The overlap between the first and second closest maps determines the regions in which the pixel level graph-cut should be performed. Therefore, for each overlapping patch the texture from the two best-quality images is available, and the graph-cut is used to find the optimal boundary seam, determining the contribution of each one in the final photo-mosaic. Each region of intersection $ROI_{i,j}$ between the two images i and j , where a given i is the closest image, j is the second closest image, and $R_{i,j}$ denotes the photo-mosaic region where i and j coincide simultaneously, is defined as $ROI_{i,j} = R_{i,j} \cup R_{j,i}$.

D. Gradient Domain Blending

1) *Pixel-Level Graph-Cut*: The proposed blending strategy uses an optimal seam finding algorithm to compute the best boundaries of the overlapping image areas. A pixel level graph-cut is performed on the regions of intersection determined by the first and second closest maps. In contrast to [23], the graph-cut is performed at the pixel level in order to guarantee the maximum accuracy of the cut, given that the main aim of the algorithm is reaching a high image quality. The algorithm searches for the boundary that minimizes the cost of the transition from one side to the other of the border line for every pair of pixels. The function has three weighted terms controlling the behavior of the cut:

$$C = \mu_1 \cdot f(I_1, I_2) + \mu_2 \cdot s(g_1, g_2) + \mu_3 \cdot L \quad (7)$$

The first term, $\mu_1 \cdot f(I_1, I_2)$, measures the intensity differences between overlapping pixels. The second

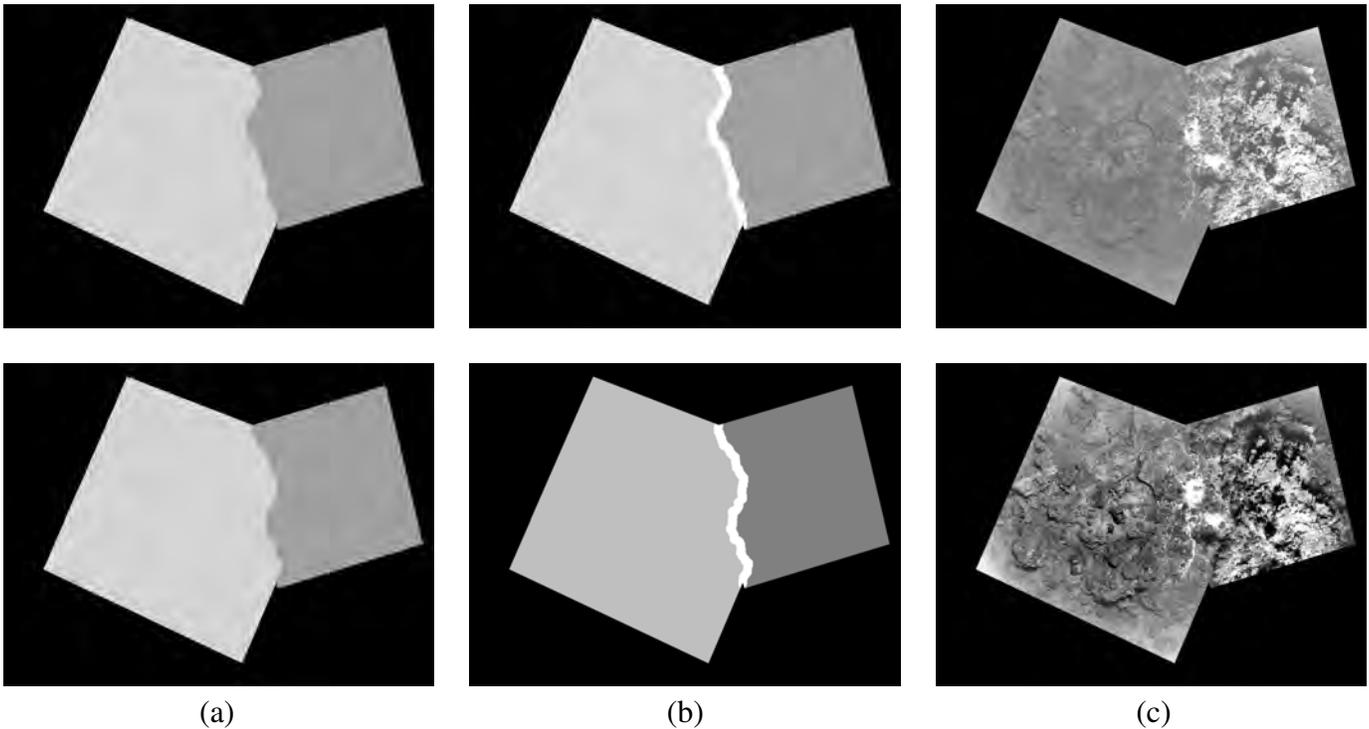


Fig. 10. Example of a pixel level graph-cut performed between two overlapping images acquired at different altitudes, and consequently evidencing differences in appearance. (a, top) Result of the graph-cut performed on the images without enhancement, (b, top) depicts, in white, the narrow strip (20 pixels on each side of the cut) where the gradient domain blending is performed and (c, top) shows the blended image pair. (a, bottom) is the result of the graph-cut performed on the images after being enhanced using the proposed neighboring based enhancement approach, (b, bottom) depicts, in white, the narrow strip where the gradient domain is performed and (c, bottom) shows the blended image pair. Notice that the results of the pixel-level graph-cuts are different before and after the application of the image enhancements.

term, $\mu_2 \cdot s(g_1, g_2)$, measures the gradient vector differences along the boundary B seam. Finally, the third term, $\mu_3 \cdot L$, measures the length L of the seam. The three weighting factors μ_1 , μ_2 and μ_3 control the behavior of the cut. The gradient term, which is not used in this way in the literature [23], allows us to deal with differently exposed overlapping regions. Here an intensity-based graph-cut considers the differences between neighboring pixels, being large even if the registration is accurate, and thereby avoids those regions where the cut is performed. Instead, if the difference between the gradient vectors along the seam path is used, the optimal seam is found independently of the different image exposures. In the case of misregistration of any moving elements in the scene, the term $\mu_2 \cdot s(g_1, g_2)$ avoids bisecting those elements by having the seam line by-passing them. This is due to the fact that even a large L value in the by-pass has a lower cost than crossing a double contour, with large gradients, of a given structure. The gradients are also less sensitive to other illumination issues, such as those caused by artificial lighting and non-uniform light-

ing. Furthermore, working in the gradient domain compensates the exposures when recovering the luminance images from the gradient vectors. Despite the benefits of the gradient term, the intensity term is kept in order to favor low photometric differences when registration is highly accurate. Therefore, a weighted addition between both intensity and gradient domain terms is proposed.

The effects of *parallax* and registration inaccuracies are minimized, since the graph-cut tends to place the seam in textureless regions where morphological differences are low. For the same reason, cuts over moving objects tend to be avoided, thus benefiting the visual consistency of the blended results.

Performing a graph-cut, especially at a pixel level, can be a computationally expensive operation when the size of the region to process is significantly large. The regions in which a graph-cut is applied are determined by the intersection between the first and second closest maps. This division of the mosaic into smaller regions allows using only one binary graph-cut per region [23]. Compared with

other approaches using multi-label graph-cuts, this approach has the benefit of being parallelizable, taking advantage of recent multi-core processors, to overcome one of the main bottlenecks in the processing pipeline. Furthermore, it is well suited for large-area underwater mosaics where all overlapping regions typically contain only two images at most. In this case, multi-label graph-cuts do not provide any advantage.

2) *Gradient Blending over Seam Strips*: Once an optimal seam has been found, a smooth transition between neighboring regions needs to be performed. Even for sequences in which the images have been preprocessed to solve non-uniform illumination problems, exposure artifacts and contrast level equalization, the graph-cut result may lead to an image with noticeable seams. Therefore, smoothing the transition between the image patches is required. The image fusion around the computed seams should be performed in a limited region, being both wide enough to ensure a smooth transition and narrow enough to reduce the noticeability of ghosting and double contouring.

A new transition smoothing approach is proposed in this paper. The applied method is a weighted average around the seams in the gradient domain, as shown in Equation 8, where g_x^1 , g_y^1 , g_x^2 and g_y^2 are the x and y gradient fields for the two involved images, \hat{g}_x and \hat{g}_y are the x and y gradient fields after the blending and μ is the smoothing transition function. Specifically, a 3rd order Hermite function is applied. The advantage of performing the weighted average in the gradient domain is the automatic compensation of different exposures between neighboring images when the luminance image is integrated from the gradients as a final step.

$$\begin{aligned}\hat{g}_x(x, y) &= \mu \cdot g_x^1(x, y) + (1 - \mu) \cdot g_x^2(x, y) \\ \hat{g}_y(x, y) &= \mu \cdot g_y^1(x, y) + (1 - \mu) \cdot g_y^2(x, y)\end{aligned}\quad (8)$$

Figure 12 shows the different results obtained by some gradient domain blending approaches, stressing the benefits of the proposed approach.

E. Luminance Recovery from Gradient Fields

After independently processing each overlapping strip region around the seams, the resulting patches need to be unified into a single larger image. Each

processed patch should be updated in the final photo-mosaic image, while information belonging to regions without an overlap should be recovered from the corresponding original images.

Once the final gradient domain photo-mosaic has been composed, after the “strip-blending”, an inconsistent gradient field is obtained. In order to recover the luminance values from the gradient fields \hat{g}_x and \hat{g}_y , the well-known Poisson equation $\nabla^2 I = \tilde{G}$ should be solved [64]. ∇^2 is the Laplacian operator $\nabla^2 = \frac{\delta^2 I}{\delta x^2} + \frac{\delta^2 I}{\delta y^2}$ and \tilde{G} is the divergence of the vector field G , defined as $\tilde{G} = \frac{\delta G_x}{\delta x} + \frac{\delta G_y}{\delta y}$. Solving the differential equation requires specifying the boundary conditions. As frequently assumed in the literature, Neumann boundary conditions are defined as $\nabla I \cdot \mathbf{n} = 0$, *i.e.* the derivative in the direction \mathbf{n} normal to the boundary is zero. The differential equation is solved using a multigrid Poisson solver [65].

The solution obtained from the gradient solver is defined up to a free additive term for the recovered intensity value. Consequently, a mapping algorithm, such as Minimum Information Loss [66], should be applied to determine this factor. The main goal of the mapping algorithm is to appropriately manipulate the dynamic range of the computed image in order to make it fit the limited range of a display device while keeping the maximum amount of detail information.

F. Giga-mosaic Unification

The described photo-mosaicing pipeline is currently implemented in Matlab™, using MEX files and parallel computing when possible. This allows efficiently blending of photo-mosaics up to 60 Mpixels in a standard personal computer with 4 GB of RAM in less than 5 minutes. Nevertheless, this mosaic size (*i.e.* <0.1 Gpixels) is small in the gigapixel scale in which this work is interested, and a solution should be used to reach the desired 5-15 Gpixels required to process the currently available data sets.

The amount of RAM may become a limitation when dealing with gigapixel images, especially if the images have more than 8 bpp (*e.g.*, 16 bpp gray scale images or 24..48 bpp color images). The strategy proposed to reduce the computer requirements consists of decomposing the problem into sub-problems (*i.e.* rectangular *tiles*), in order

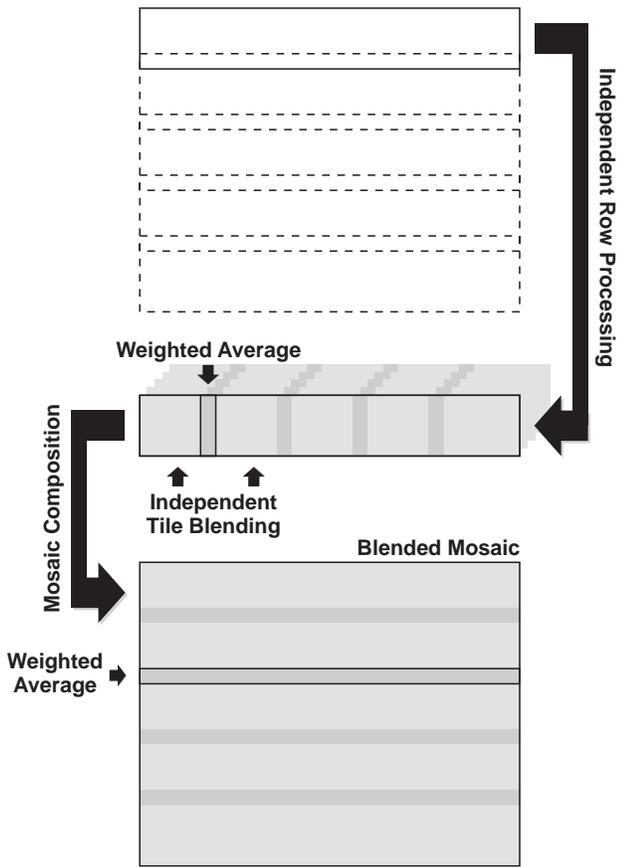


Fig. 11. Tiling scheme for the Giga-photo-mosaic blending. Each *tile* is processed as an independent photo-mosaic and blended with the previously processed neighboring ones in a given *global-strip* (i.e., a row of blended tiles), using a weighted average in the luminance domain. Next, each two neighboring *rows* are blended using the same approach. The Giga-photo-mosaic is the result of joining all the *global-strips*.

to solve them sequentially and finally unify them into the mosaic image.

The price of this decomposition is the need of a second level blending of the tiles. This is similar to “strip-blending” (presented in Section IV-D2) applied in the optimal seams, but is performed in the intensity domain for computational reasons. When compared with gradient domain operations, intensity blending is inexpensive and can deal with large amounts of data. Given a tile blended in the gradient domain, it cannot be aligned with a neighboring one without applying a blending method for several reasons. Every tile has a different free factor after the luminance recovery using the Poisson solver, which is multiplicative when working with $\log I$ values, and should be equalized throughout all the tiles. But even after multiplying the pixel intensities of every tile with the corresponding constant factor, the tiles will not match perfectly. The nature of the

Poisson solver leads to a result in which the gradient inconsistencies in the seam regions between images penetrates the whole tile, leading to gradual and subtle but perceptible luminance shifts. These are the reasons why a *tile-overlap* intensity blending has to be performed on the tile overlapping region. This blending does not have to face issues such as misregistration or moving objects, in which the gradient blending shows its benefits, because the tiles are perfectly aligned. Consequently, this kind of blending will compensate the gradient differences of overlapping tiles coming from different Poisson solutions, hiding the small luminance discrepancies between neighboring tiles. The decomposition differs necessarily from the theoretically exact solution, but the differences are not perceptible.

Although the *tile-level* pipeline described above is straightforward, its technical implementation deserves further clarification owing to the need to manage available computational resources with such large problems (i.e. gigapixel photo-mosaics).

The rectangular “canvas” of the full photo-mosaic is divided into a regular grid of overlapping *tiles*. The size of the *tiles* depends on the available RAM. For time efficiency reasons, the space required to store the final photo-mosaic image and a single *global-strip* (i.e., a row of tiles) are allocated in memory, avoiding an excessive amount of slow hard drive sequential accesses.

A weighted average smoothing in the intensity domain is used to join neighboring *tiles* in a given rectangular overlapping region. In our experiments, the size of the overlapping regions varied between 15% - 25% of the *tile* size, depending on the initial spatial image arrangement. Once a *tile* has been processed, it is stored in the current *global-strip*, performing a blending with the previously processed tile (when available). When an entire *global-strip* has been processed, it is stored in the hard drive to save RAM space and the same procedure is repeated on the next strip. The strategy used to blend two neighboring *tiles* is also used to blend two neighboring *global-strips*. The proposed blending strategy avoids the problem of simultaneously fusing more than two images of a given region using the intensity domain blending. In this way, it is not necessary to compute complex transition functions on areas with multiple overlaps. Figure 11 shows the giga-mosaic unification strategy described above. Moreover, the proposed approach can effectively be

applied to color images, as illustrated in Figure 2. In this case, each channel is processed independently. When recovering the intensities from the gradient domain, the gain of every channel in the final composite color image should take into account the original correlations among red (R), green (G) and blue (B) components for a set of sampled points in different images in the sequence. However, it should be noted that, depending on the sampling (especially for images taken at different altitudes) and the different absorption characteristics of the medium for R, G and B, this simple technique may result in color shifting. Therefore, further work is required for accurate color blending in underwater imagery.

V. RESULTS AND DISCUSSION

The datasets presented below were collected by the Victor-6000 ROV [24] during the MOMAR08-Leg1 cruise (IFREMER/CNRS, France) at the Mid-Atlantic Ridge. A 3-day survey of the Lucky Strike hydrothermal field was performed with a black and white, high sensibility camera system (OTUS) installed in the geophysical mapping payload (Module Route). The payload also included 4 Flashes (1200 joules), a RESON Seabat 7125 multi-beam echo sounder and a SIMRAD EK60 echo sounder altimeter. The onboard navigation system also included an RDI Doppler Velocity Log (DVL), an iXSEA OCTANS fiber-optic gyrocompass and a Parascientific depth sensor.

The robot was teleoperated from the N/O “Pourquoi pas?” research vessel, and was programmed to follow lawn-mower trajectories at a constant altitude of 6 meters over the seafloor. The area surveyed was located at a depth of $\sim 1,700$ m, and positioning estimates were obtained by means of a POSIDONIA ultra-short baseline system (USBL).

A rough photo-mosaic was built using the navigation data as a first estimate of the camera position for all the images in the sequence. From this initial estimation, a pair-wise feature detection and matching processing was performed in order to improve the quality of the sequential pair-wise registration. Finally, a global alignment algorithm [2], [67] was applied to the whole set of features detected to maximize the global consistency of the registration.

The resulting globally aligned photo-mosaic is the input data of the presented blending approach,

which allows the processing of the whole data set on a 16-core computer with 128 GB of RAM in less than two days. Nevertheless, this implementation also allows the processing of regions of the photo-mosaic of arbitrary size on smaller computers with less memory and processing power by adjusting the tile sizes.

Figure 13 shows an example comparing the results of the approach proposed in this paper to that of Szeliski [68], implemented in the Microsoft ICE (MICE) software, which also applies a graph-cut technique to determine the optimal seams around which a gradient blending is performed. The absence of an adaptive image enhancement step leads to differences in appearance that are noticeable between neighboring images acquired at significantly different depths. Furthermore, the proposed approach results in enhanced sharpness due to the usage of a very narrow region of transition around the optimal seams. It is important to mention that Szeliski’s method was not able to register some of our test sequences, and therefore it has not always been possible to perform a direct comparison.

Again comparing the results of the MICE software with those provided by the proposed approach, Figure 14 illustrates the difference in appearance of a given region in the photo-mosaic where different criteria to determine the contributing images have been used. Beyond the automatic image contrast enhancement, our approach selects the image with the greatest richness of detail for each region in the photo-mosaic, while discarding the poorer ones. This method results in a sharper mosaic image.

Figure 15 illustrates the advantages of our approach, based on graph-cuts using a combination of both gradient and luminance domain information, over the simpler approach of using only the intensity domain. Furthermore, these examples demonstrate the benefits that the proposed 2D approach may offer even in the presence of strong 3D on the scene. Nevertheless, when the 3D of the scene is significant and the information required to reconstruct it is available, the best approach is to recover the terrain structure using SFM [19], [20], and to perform an orthogonal projection of the triangle mesh obtained.

The photo-mosaic built from the complete dataset presented in Figure 16 covers an area of roughly 0.6 square km with a resolution of 10 mm per pixel. This includes 21,262 images, with a seafloor coverage of $\approx 56\%$. The processing time required to

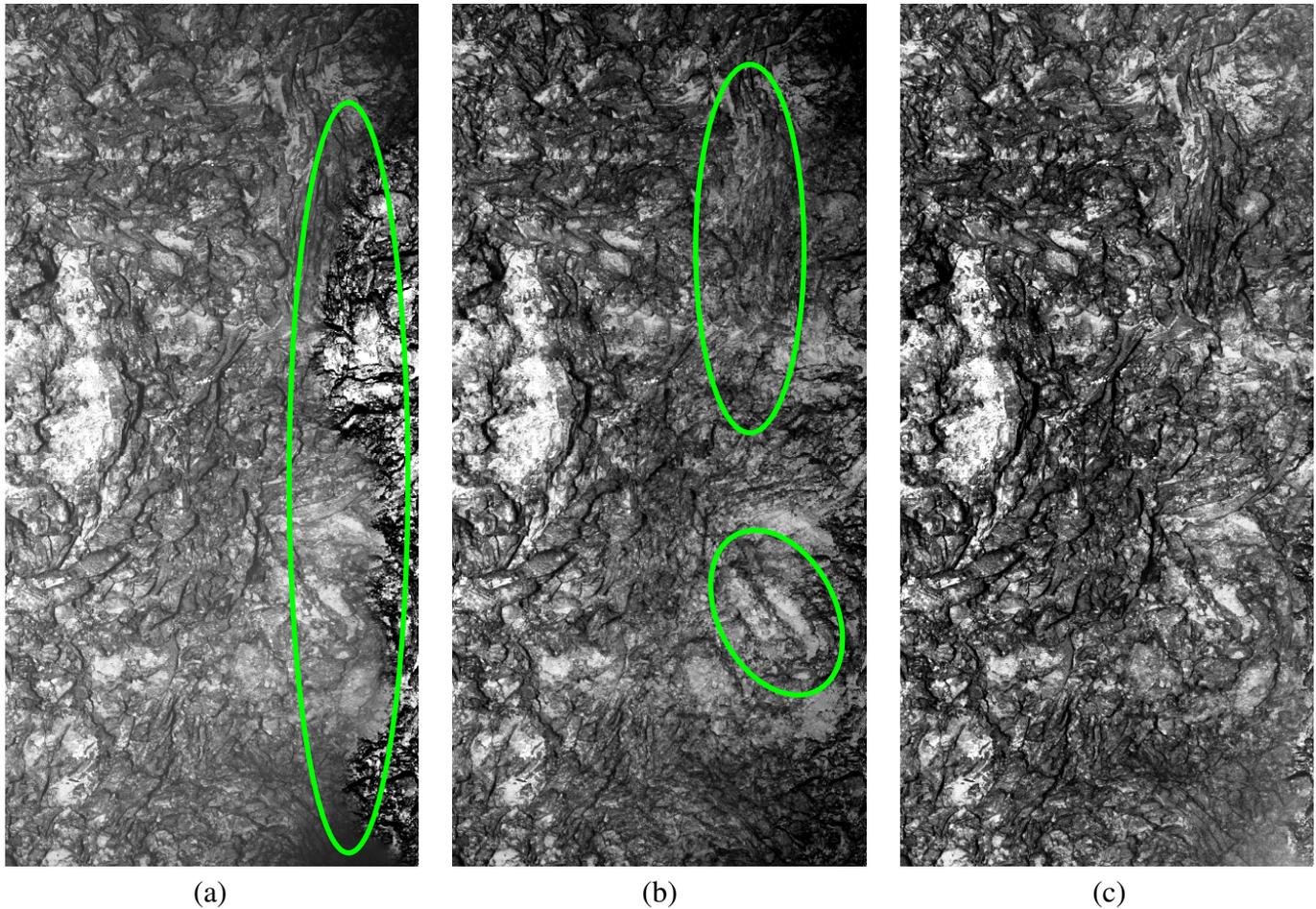


Fig. 12. Detail of a given overlapping region blended with three different gradient approaches. (a) shows the result of setting the gradient values around the optimal seam boundary at zero [53], in order to enforce continuity through the join. (b) shows the result of a weighted average gradient blending performed on all the available overlapping pixels [33]. (c) shows the result of the proposed approach, which only performs a gradient blending on a narrow region around the optimal computed seam. The transition is sufficiently smooth to provide a sensation of consistency in the image while avoiding ghosting and double contours in the overlapping areas.

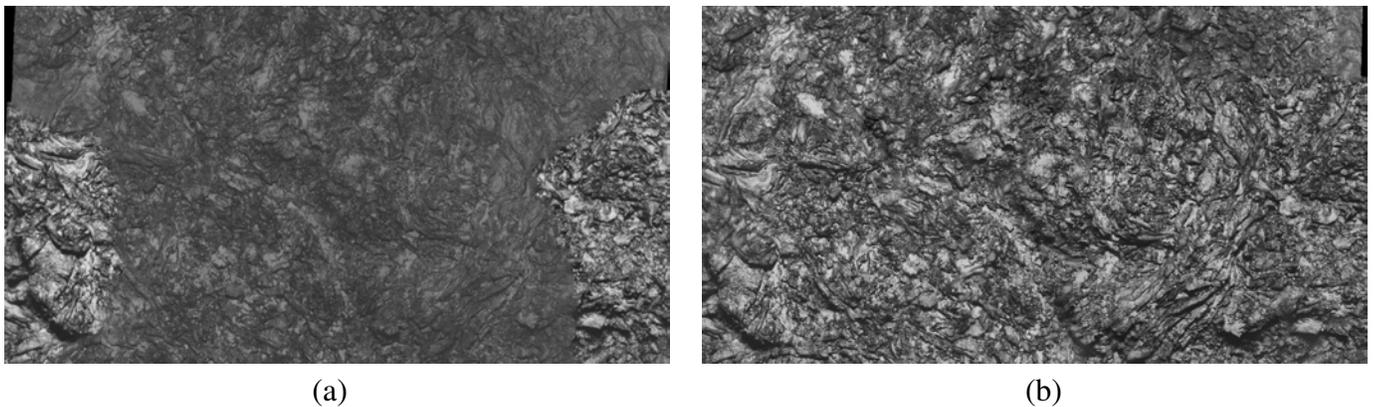


Fig. 13. (a) Detail of an underwater photo-mosaic region generated with Szeliski's method [68] (direct result of Microsoft ICE software) without automatic image enhancement, and (b) the result obtained by our approach with adaptive contrast enhancement. The global appearance of the image is uniform and the central part of the photo-mosaic is perceptually more informative after the contrast improvement.

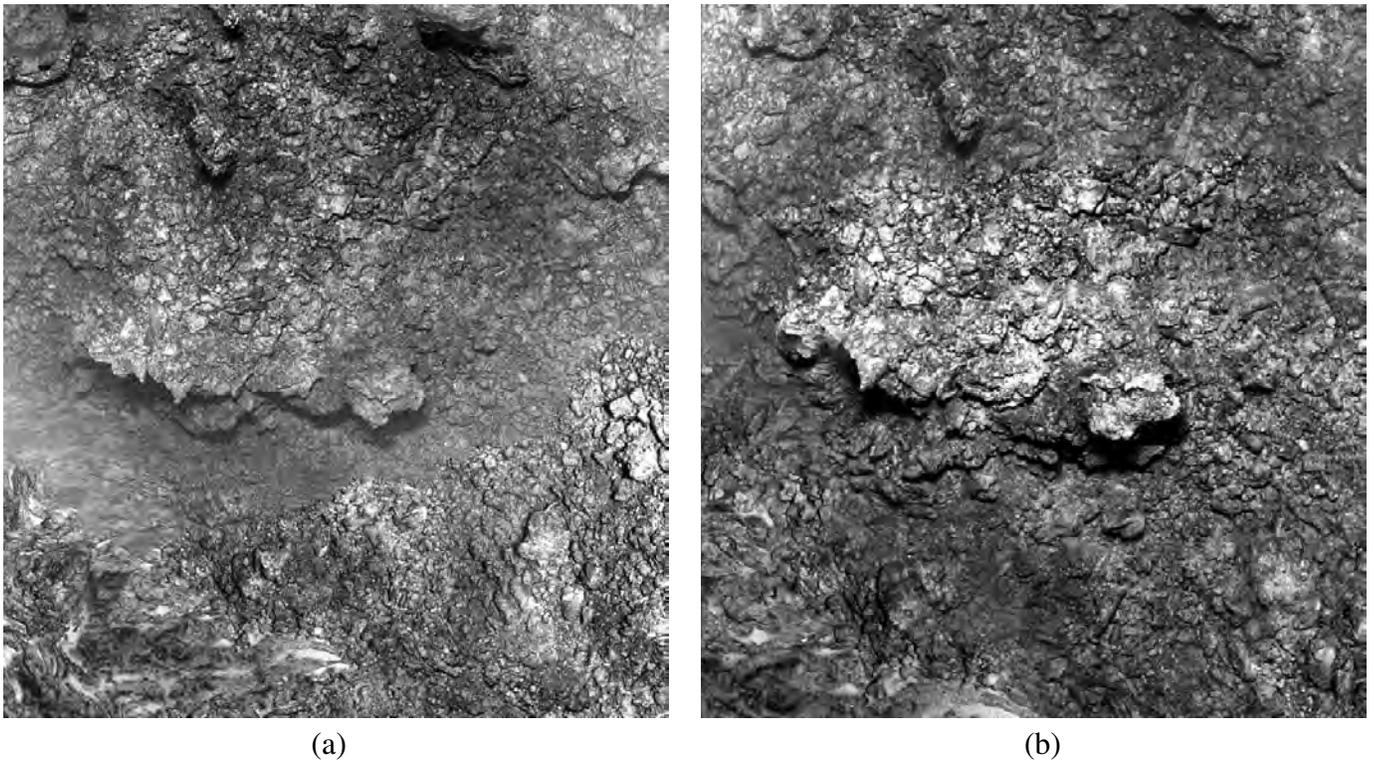


Fig. 14. (a) Detail of an underwater photo-mosaic region using Szeliski's method and (b) the result obtained by our approach with contributing image selection based on a quality rank estimation. The proposed approach leads to an image richer in details and with higher contrast.

blend the photo-mosaic was 36 hours, taking into account that the sparseness of the image distribution allows speeding the process up by skipping the non-populated areas. The visualization of the photo-mosaic image, with a size of $78,651 \times 62,722$ pixels (*i.e.* ~ 5 Gpixels), was performed using our mosaicViewer [69] software. This application generates a multi-resolution pyramid of the whole photo-mosaic and uses dynamic caching techniques to speed up the visualization at different zoom levels, thus avoiding loading the complete mosaic to memory. Furthermore, this software allows displaying the original images of the data set on top of the blended photo-mosaic. It is thus possible for the user to access global and consistent visualization of the whole data set (through the blended image) simultaneously, while inspecting the original image information from selected interest regions.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we present a new approach for large-scale underwater image blending. The presented pipeline extends the typical photo-mosaicing techniques to the more complex and challenging underwater medium. We can now build Giga-photo-

mosaics of large areas (in the km^2 range), allowing us to image large seafloor expanses for scientific explorations (*e.g.*, geology, biology, ecology), applied purposes (*e.g.*, civil engineering), and environmental monitoring (with repeated surveys).

Deep-ocean underwater imaging suffers from specific problems that require the application of particular solutions. The contributions of this paper concern all the photo-mosaicing steps (image preprocessing, enhancing and blending) that can significantly improve the final image quality and visual consistency.

Firstly, in the image preprocessing step, a *depth dependent illumination compensation function* is computed and applied to the original images in order to solve the non-uniform illumination appearance due to light attenuation.

Additionally, if precise depth information is not available, an *altitude estimation* based on the size of the image projection (once registered) has been exploited in different steps of the pipeline.

Concerning image enhancement, the image contrast variability due to different acquisition altitudes has been compensated using an *adaptive contrast enhancement* based on an image quality reference

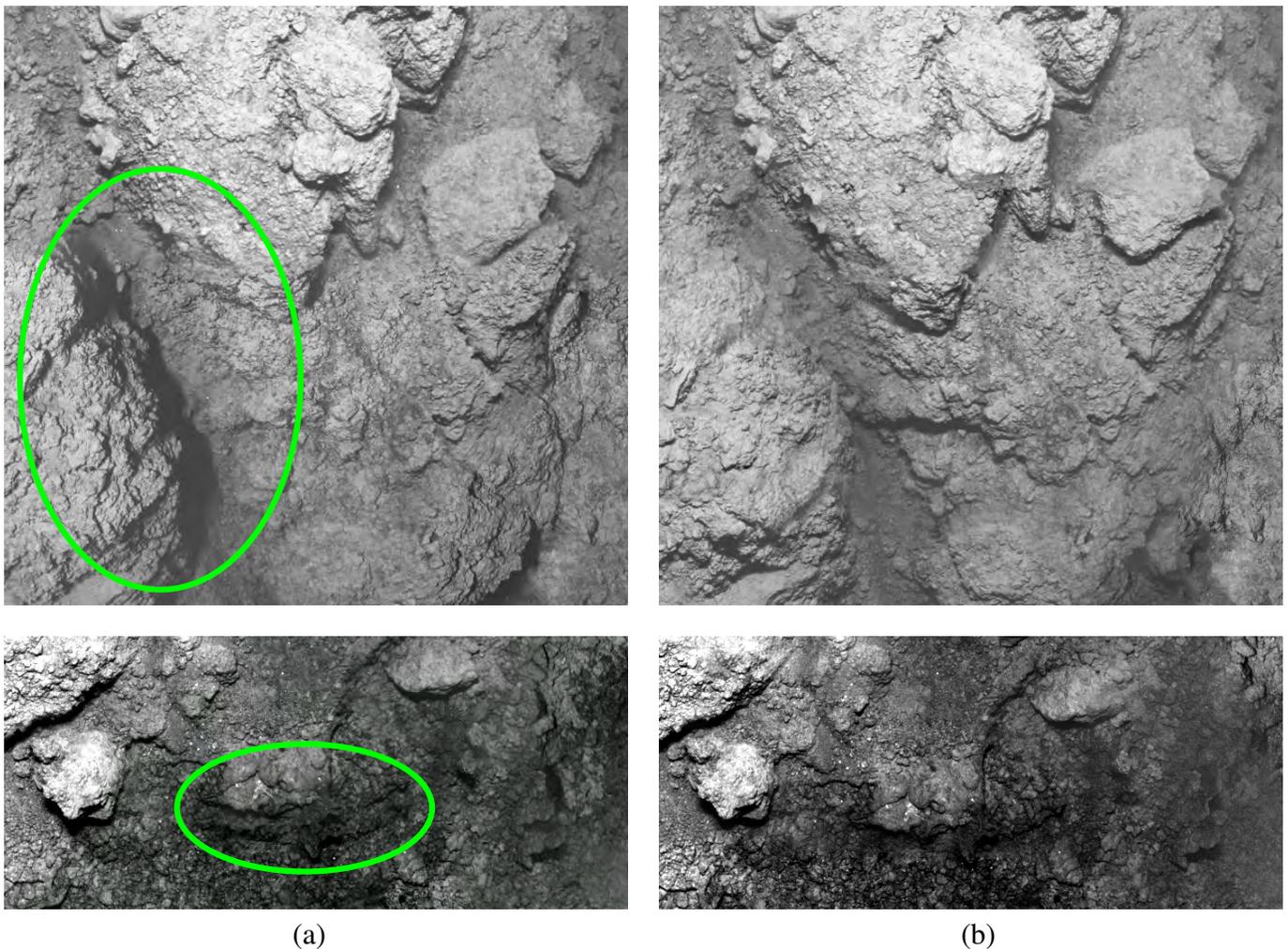


Fig. 15. Comparison of our approach with common state-of-the-art methods in terms of graph-cut performance. (a-top and bottom) Details of an underwater mosaic region blended using an intensity based graph-cut algorithm and (b-top and bottom) the result obtained by the proposed gradient based graph-cut strategy. (a-top) The shadow in the top left corner region has been interpreted differently by both approaches, leading, in the case of our graph-cut (b-top), to an unshadowed valley. The highlighted regions in image (a-bottom) present object doubling that has been suppressed in image (b-bottom). The graph-cut shows a different behavior in both cases, leading to a noticeable difference in the image contribution selection.

selected through a TV criteria. This criteria has also been applied to give a higher priority to the information coming from the *higher quality images* when building the first and second closest maps which allow us to perform the graph-cut on the overlapping regions. Consequently, the contribution from sharper and informative images is higher than from contrastless or poorly detailed ones.

In the blending step, the proposed graph-cut strategy operates in the *image gradient domain* over the overlapping regions, in contrast with several state-of-the-art methods [70], [23], [71] working on the intensity domain image differences. This approach allows finding an adequate seam even if the overlapping images have been acquired with different exposures. For a given image region acquired with

two different exposures, even after the non-uniform illumination compensation, an intensity domain approach will find photometric differences between pixels which do not correspond to real scene structure misalignments. A gradient domain method is unaffected by this problem since gradient values are not exposure dependent, assuming that exposure differences are purely modeled as shifts.

The smooth transition around the optimally selected seams is performed in a *narrow strip*, ensuring the maximum possible sharpness and avoiding double contouring in that region. This smoothing is performed also in the gradient domain, as it also compensates the possible different exposures between images. Finally, a *blending strategy for very large-scale photo-mosaics*, i.e. Giga-photo-

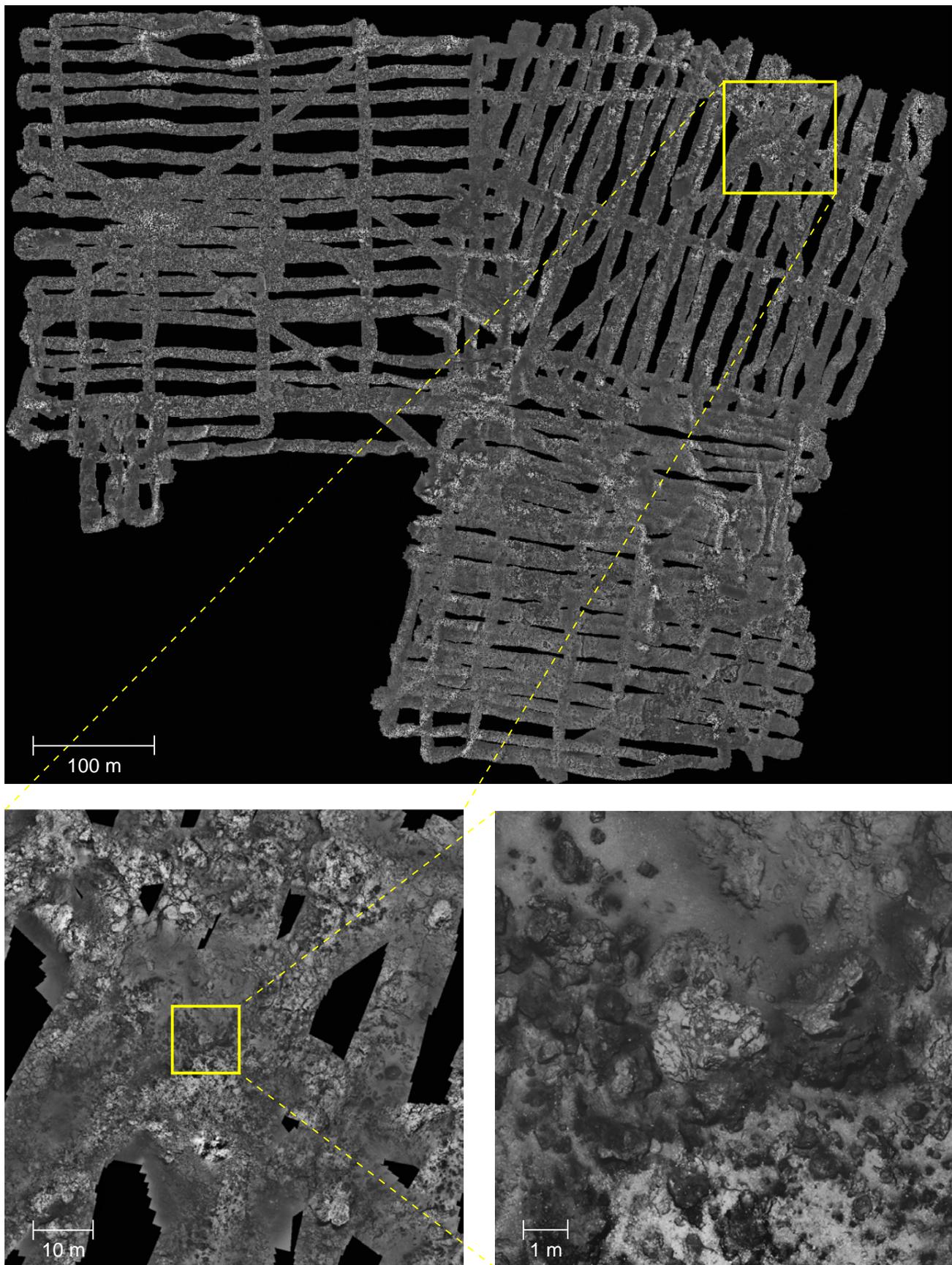


Fig. 16. Three levels of zoom for a given region of the giga-mosaic. The first zoom level shows the global appearance of the whole photo-mosaic image. The second zoom level shows a selected region depicting the local consistency appearance obtained by the proposed blending approach. Finally, the third zoom level shows a small seamless region at the maximum resolution of 10 mm per pixel. The bounding box of this photo-mosaic is 78,651 by 62,722 pixels (total ~ 5 GPixels), and the covered area corresponds to 493,315 m^2 .

mosaics, has been developed and tested on real data, generating images in excess of 5 GPixel, and having, as its only limitation, the maximum size of the tile that can be processed in a given amount of RAM.

Regarding future work, there are still several open problems that will require the development of new techniques. On one hand, it is worth noting that the illumination in underwater imagery provides strong visual clues for human perception of the seafloor relief. This effect becomes obvious when visualizing the whole photo-mosaic image, where it becomes more difficult to perceive the real relief of the underwater terrain. This problem can be addressed using an image based computer graphics solution, if a DEM (Digital Elevation Model) is available. The global lighting can be synthetically applied to the photo-mosaic image taking into account a rough estimate of the 3D structure of the scene. This solution is independent of the 3D recovery using SFM techniques, because the aim is not only to determine the terrain relief but also to be able to synthetically apply a global illumination. On the other hand, the *parallax* is one of the most difficult issues to solve when building seamless 2D photo-mosaics. The only way to address the blending on a strong 3D seabed consists of recovering its 3D structure using, for instance, SFM techniques. This is not possible for sparse low-overlapping image sets, thus making the blending of a 3D structure extremely challenging.

Furthermore, the different levels of sharpness for neighboring images acquired at very different altitudes are a problem, which can be partially addressed using costly image processing techniques. Image sharpening algorithms can be applied to enhance the visual quality of the poorer images, obviously up to a given level. This image enhancement, like the contrast enhancement proposed in this work, would be performed in an adaptive way, depending on the neighboring images context.

Finally, despite the fact that the available image sequences are in gray scale, dealing with colors is another open issue, due to the properties of the underwater medium, especially concerning light spectral absorption. Generating a blended color photo-mosaic with a realistic global appearance is a challenge for the near future.

VII. ACKNOWLEDGMENTS

This work has been partially supported by the Spanish Ministry of Science and Innovation under grant CTM2010-15216 (MUMAP), the European Union under grant ICT-2011-288704 (MORPH) and the CIDEM from the Catalan Government under grant VALTEC09-2-0122. L. Neumann was supported by ICREA from the Catalan Government. N. Gracias was supported by the Spanish MCINN under the Ramon y Cajal program. Additional support was provided by ANR (France) Mothseim (project NT05-3_42213) and CNRS to J. Escartín. CNRS and IFREMER (France) financed the MoMAR'08 cruise where the data processed here were acquired.

REFERENCES

- [1] J. Kopf, M. Uyttendaele, O. Deussen, and M. Cohen. Capturing and viewing gigapixel images. *ACM Transactions on Graphics (SIGGRAPH)*, 26(3), 2007.
- [2] J. Ferrer, A. Elibol, O. Delaunoy, N. Gracias, and R. Garcia. Large-area photo-mosaics using global alignment and navigation data. In *MTS/IEEE OCEANS*, 2007.
- [3] A. Elibol, R. Garcia, O. Delaunoy, and N. Gracias. A new global alignment method for feature based image mosaicing. In *ISVC '08: Proceedings of the 4th International Symposium on Advances in Visual Computing, Part II*, pages 257–266, Berlin, Heidelberg, 2008. Springer-Verlag.
- [4] A. Elibol, N. Gracias, and R. Garcia. Augmented state-extended Kalman filter combined framework for topology estimation in large-area underwater mapping. *Journal of Field Robotics*, 27(5):656–674, 2010.
- [5] A. Elibol, R. Garcia, and N. Gracias. A new global alignment approach for underwater optical mapping. *Ocean Engineering*, 38(10):1207–1219, 2011.
- [6] B. Shao, J.S. Jaffe, M. Chachisvilis, and S.C. Esener. Angular resolved light scattering for discriminating among marine picoplankton: modeling and experimental measurements. *Opt. Express*, 14(25):12473–12484, Dec 2006.
- [7] T. Barreyre, J. Escartín, R. Garcia, M. Cannat, E. Mittelstaedt, and R. Prados. Structure, temporal evolution, and heat flux estimates from the lucky strike deep-sea hydrothermal field derived from seafloor image mosaics. *Geochem. Geophys. Geosyst.*, 13:Q04007–, April 2012.
- [8] E. Mittelstaedt, J. Escartín, N. Gracias, J.A. Olive, T. Barreyre, A. Davaille, M. Cannat, and R. Garcia. Quantifying diffuse and discrete venting at the tour eiffel vent site, lucky strike hydrothermal field. *Geochem. Geophys. Geosyst.*, 13:Q04008–, April 2012.
- [9] O. Pizarro and H. Singh. Toward large-area mosaicing for underwater scientific applications. *Oceanic Engineering, IEEE Journal of*, 28(4):651 – 672, oct. 2003.
- [10] H. Singh, J. Howland, and O. Pizarro. Advances in large-area photomosaicking underwater. *Oceanic Engineering, IEEE Journal of*, 29(3):872 – 886, july 2004.
- [11] K. Richmond and S. Rock. A real-time visual mosaicking and navigation system. In AUSI, editor, *Proceedings of the Unmanned Untethered Submersible Technology Conference (UUST)*, August 2005.

- [12] K. Richmond and S. Rock. An operational real-time large-scale visual mosaicking and navigation system. *Sea Technology*, 48:10–13, March 2007.
- [13] K. Richmond and S. Rock. Real-time visual mosaicking and navigation of the USS Macon. In *Proceedings of the Unmanned Underwater Technology Conference (UUST)*. AUSI, August 2007.
- [14] O. Pizarro, R. Eustice, and H. Singh. Large area 3d reconstructions from underwater surveys. In *OCEANS '04. MTS/IEEE TECHNO-OCEAN '04*, volume 2, pages 678 – 687 Vol.2, nov. 2004.
- [15] R. Eustice, H. Singh, J.J. Leonard, and M.R. Walter. Visually mapping the RMS Titanic: Conservative covariance estimates for SLAM information filters. *The International Journal of Robotics Research*, 25(12):1223–1242, 2006.
- [16] H. Singh, C. Roman, O. Pizarro, R. Eustice, and A. Can. Towards high-resolution imaging from underwater vehicles. *Int. J. Rob. Res.*, 26:55–74, January 2007.
- [17] M. Johnson-Roberson, O. Pizarro, S.B. Williams, and I. Mahon. Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys. *J. Field Robotics*, 27(1):21–51, 2010.
- [18] P. Burt and E. Adelson. A multiresolution spline with application to image mosaics. *ACM Trans. Graph.*, 2(4):217–236, 1983.
- [19] T. Nicosevici and R. Garcia. On-line visual vocabularies for robot navigation and mapping. In *IROS'09: Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems*, pages 205–212, Piscataway, NJ, USA, 2009. IEEE Press.
- [20] T. Nicosevici, N. Gracias, S. Negahdaripour, and R. Garcia. Efficient three-dimensional scene modeling and mosaicing. *J. Field Robot.*, 26(10):759–788, 2009.
- [21] S. Agarwal, N. Snavely, I. Simon, S.M. Seitz, and R. Szeliski. Building rome in a day. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 72 –79, 29 2009-oct. 2 2009.
- [22] Y. Furukawa, B. Curless, S.M. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1434 – 1441, june 2010.
- [23] N. Gracias, M. Mahoor, S. Negahdaripour, and A. Gleason. Fast image blending using watersheds and graph cuts. *Image and Vision Computing*, 27:597–607, April 2009.
- [24] P. Simeoni, J. Sarrazin, H. Nouze, P.M. Sarradin, H. Ondreas, C. Scalabrin, and J.-M. Siquin. Victor 6000: New high resolution tools for deep sea research. module de mesures en route. In *OCEANS 2007 - Europe*, pages 1 –6, June 2007.
- [25] D.L. Milgram. Computer methods for creating photomosaics. *IEEE Transactions on Computers*, 24(11):1113–1119, 1975.
- [26] D.L. Milgram. Adaptive techniques for photomosaicking. *IEEE Transactions on Computers*, C-26(11):1175–1180, Nov. 1977.
- [27] S. Peleg. Elimination of seams from photomosaics. In *Pattern Recognition and Image Processing*, volume 16, pages 90–94, 1981.
- [28] C.T. Hsu and J.L. Wu. Multiresolution mosaic. *Consumer*, 42(4):981–990, November 1996.
- [29] J. Davis. Mosaics of scenes with moving objects. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA, USA, June 1998.
- [30] E.W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.
- [31] M. Uyttendaele, A. Eden, and R. Szeliski. Eliminating ghosting and exposure artifacts in image mosaics. In *Proc. Int. Conf. on Comp. Vision and Patt. Recog. (CVPR2001)*, pages 509–516, 2001.
- [32] J. Chen, I. A. Kanj, and W. Jia. Vertex cover: Further observations and further improvements. *Journal of Algorithms*, 41(2):280 – 301, 2001.
- [33] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. *ACM Transactions on Graphics*, 22(3):313–318, 2003.
- [34] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proceedings of SIGGRAPH. New Orleans, USA*, pages 417–424, jul 2000.
- [35] A. Levin, A. Zomet, S. Peleg, and Y. Weiss. Seamless image stitching in the gradient domain. In *Proc. of the European Conference on Computer Vision (ECCV04)*, Prague, Czech Republic, May 2004.
- [36] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. In *ACM Transactions on Graphics (SIGGRAPH)*, August 2004.
- [37] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Proc. IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, Nov. 2001.
- [38] A. Agarwala. Efficient gradient-domain compositing using quadrees. *ACM Trans. Graph.*, 26(3):94, 2007.
- [39] S. Hanan. *Applications of spatial data structures: Computer graphics, image processing, and GIS*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1990.
- [40] M. Su, W. Hwang, and K. Cheng. Analysis on multiresolution mosaic images. *IEEE Transactions on Image Processing*, 13(7):952–959, July 2004.
- [41] A. Eden, M. Uyttendaele, and R. Szeliski. Seamless image stitching of scenes with large motions and exposure differences. In *CVPR '06: Proc. of the 2006 IEEE Computer Vision and Pattern Recognition*, pages 2498–2505, Washington, DC, USA, 2006. IEEE Computer Society.
- [42] W. Zhao. Flexible image blending for image mosaicing with reduced artifacts. *International Journal of Pattern Recognition and Artificial Intelligence*, 20(4):609–628, 2006.
- [43] F. Gu and Y. Rzhonov. Optimal image blending for underwater mosaics. In *OCEANS 2006*, pages 1–5, Sept. 2006.
- [44] R. Szeliski, Uyttendaele, M., and D. Steedly. Fast poisson blending using multi-splines. Technical report, Interactive Visual Media, April 2008.
- [45] A. Mills and G. Dudek. Image stitching with dynamic elements. *Image and Vision Computing*, 27(10):1593–1602, 2009.
- [46] T. Porter and T. Duff. Compositing digital images. In *Proceedings of SIGGRAPH*, pages 253–259, New York, NY, USA, 1984. ACM.
- [47] A. Efros and W. Freeman. Image quilting for texture synthesis and transfer. *Proceedings of SIGGRAPH*, pages 341–346, August 2001.
- [48] Y. Xiong and K. Pulli. Color correction based image blending for creating high resolution panoramic images on mobile devices. In *ACM SIGGRAPH Asia 2009 Posters*, SIGGRAPH Asia '09, pages 47:1–47:1, New York, NY, USA, 2009.
- [49] M. Brown and D. G. Lowe. Recognising panoramas. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, page 1218, Washington, DC, USA, 2003. IEEE Computer Society.
- [50] G.Y. Tian, D. Gledhill, D. Taylor, and D. Clarke. Colour correction for panoramic imaging. In *Information Visualisation, 2002. Proceedings. Sixth International Conference on*, pages 483 – 488, 2002.
- [51] J. Jia and C.K. Tang. Image registration with global and local luminance alignment. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, pages 156 –163 vol.1, oct. 2003.
- [52] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley. Color

- transfer between images. *IEEE Computer Graphics and Applications*, 21(5):34–41, sep./oct. 2001.
- [53] Y. Xiang, B. Zou, and H. Li. Selective color transfer with multi-source images. *Pattern Recognition Letters*, 30(7):682–689, 2009.
- [54] J.S. Jaffe, K.D. Moore, J. McLean, and M.P. Strand. Underwater optical imaging: Status and prospects. *Oceanography*, 14:66–77, 2001.
- [55] R. Garcia, T. Nicosevici, and X. Cufi. On the way to solve lighting problems in underwater imaging. In *OCEANS '02 MTS/IEEE*, volume 2, pages 1018–1024 vol.2, oct. 2002.
- [56] D. P. Capel. *Image Mosaicing and Super-resolution*. PhD thesis, University of Oxford, Oxford, UK, 2001.
- [57] Y. Rzhanov and F. Gu. Enhancement of underwater videomosaics for post-processing. In *OCEANS 2007*, pages 1–6, oct. 2007.
- [58] N. Gracias, S. Negahdaripour, L. Neumann, R. Prados, and R. Garcia. A motion compensated filtering approach to remove sunlight flicker in shallow water images. In *OCEANS 2008*, pages 1–7, sept. 2008.
- [59] T. Chan and S. Jianhong. *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2005.
- [60] R. Guillemaud. Uniformity correction with homomorphic filtering on region of interest. In *ICIP (2)*, pages 872–875, 1998.
- [61] S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J.B. Zimmerman, and K. Zuiderveld. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3):355–368, 1987.
- [62] D. Lowe. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*, volume 20, pages 91–110, 2003.
- [63] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *European Conference on Computer Vision*, pages 404–417, 2006.
- [64] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. *ACM Trans. Graph.*, 21(3):249–256, 2002.
- [65] M. Kazhdan and H. Hoppe. Streaming multigrid for gradient-domain operations on large images. *ACM Trans. Graph.*, 27(3):1–10, 2008.
- [66] L. Neumann, K. Matkovic, and W. Purgathofer. Automatic exposure in computer graphics based on the minimum information loss principle. *Computer Graphics International Conference*, 0:666–667, 1998.
- [67] J. Escartin, R. Garcia, O. Delaunoy, J. Ferrer, N. Gracias, A. Elibol, X. Cufi, L. Neumann, D.J. Fornari, S.E. Humphris, and J. Renard. Globally aligned photomosaic of the lucky strike hydrothermal vent field (Mid-Atlantic ridge, 37deg18.5'n): Release of georeferenced data, mosaic construction, and viewing software. *Geochemistry, Geophysics, Geosystems*, 9:Q12009, dec. 2008.
- [68] R. Szeliski. Image alignment and stitching: a tutorial. *Found. Trends. Comput. Graph. Vis.*, 2:1–104, January 2006.
- [69] Mosaicviewer, <http://eia.udg.edu/~rafa/mosaicviewer.html>.
- [70] V. Kwatra, A. Schodl, I. Essa, G. Turk, and A. Bobick. Graphcut textures: Image and video synthesis using graph cuts. In *ACM Transactions on Graphics (SIGGRAPH)*, July 2003.
- [71] Y. Xiong and K. Pulli. Color correction based image blending for creating high resolution panoramic images on mobile devices. In *ACM SIGGRAPH Asia 2009 Posters, SIGGRAPH Asia '09*, pages 47:1–47:1, New York, NY, USA, 2009.