Seafloor-Invariant Caustics Removal From Underwater Imagery

Panagiotis Agrafiotis ⓘ, Konstantinos Karantzalos ⓘ, *Senior Member, IEEE*, and Andreas Georgopoulos

Abstract—Mapping the seafloor with underwater imaging cameras is of significant importance for various applications including marine engineering, geology, geomorphology, archaeology, and biology. For shallow waters, among the underwater imaging challenges, caustics, i.e., the complex physical phenomena resulting from the projection of light rays being refracted by the wavy surface, is likely the most crucial one. Caustics is the main factor during underwater imaging campaigns that massively degrades image quality and affects severely any 2-D mosaicing or 3-D reconstruction of the seabed. In this article, we propose a novel method for correcting the radiometric effects of caustics on shallow underwater imagery. Contrary to the state-of-the-art, the developed method can handle the seabed and riverbed of any anaglyph, correcting the images using real pixel information, thus, improving image matching and 3-D reconstruction processes. In particular, the developed method employs deep learning architectures to classify image pixels to "noncaustics" and "caustics." Then, it exploits the 3-D geometry of the scene to achieve a pixelwise correction, by transferring appropriate color values between the overlapping underwater images. Moreover, to fill the current gap, we have collected, annotated, and structured a real-world caustic data set, namely, R-CAUSTIC, which is openly available. Overall, based on the experimental results and validation, the developed methodology is quite promising in both detecting caustics and reconstructing their intensity.

Index Terms—Caustics, data set, fully convolutional network, image segmentation, sun flickering, underwater 3-D reconstruction.

I. INTRODUCTION

RADIOMETRIC effects of water refraction include the chromatic aberration, which can be handled using high-quality lens and the rippling caustics or sun flickering which are of really high importance [1], [2]. These complex physical phenomena result from the light rays being refracted by a curved surface, such as the wavy interface between air and water. Caustics effects (see Fig. 1) are apparent both in overwater and
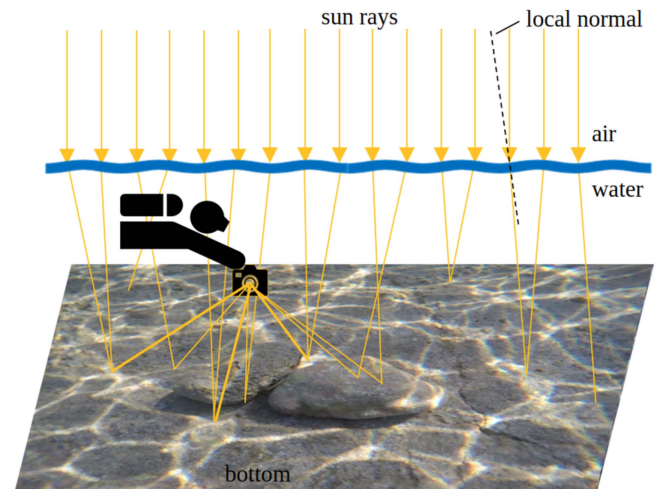
Fig. 1. Due to waves on the water surface, the refraction of natural sky illumination is spatially varying. This creates 3-D patterns of variable light flux, caustics, and 2-D illumination patterns. These patterns vary in time due to dynamic surface waves.

underwater imagery depicting the bottom of optically clear water bodies (seabed, lakebed, etc.). In the overwater cases, where large-scale mapping operations are mostly involved, they can be avoided by increasing the flying height, thus, increasing the ground sampling distance (GSD) or by acquiring imagery with the sun angle less than 30° over the horizon [3], [4], [5], [6]. However, in the shallow underwater cases, in most of which high detailed 3-D reconstructions and textures are required, this is not possible since the camera will exit the water, introducing also the severe geometric errors caused by the refraction on the water surface [4]. Moreover, rippling caustics' generation precedes wavy water surface, a water state in which overwater mapping is not suggested since additional errors would be introduced due to the waves [7], [8]. Till now, to avoid these intense lighting artefacts on the bottom, image acquisition is performed under overcast conditions, or with the sun low on the horizon [1], [2].

Hence, caustics seem to be the main factor degrading the underwater image quality and affecting the image-based 3-D reconstruction process in very shallow waters [1]. These effects are adversely affecting image matching algorithms by throwing off most of them, leading to less accurate matches [1] and causing issues in the simultaneous localization and mapping (SLAM)-based navigation of the remotely operated vehicles (ROV) and autonomous underwater vehicles (AUV) on shallow waters [9]. Also, they are the main cause for dissimilarities in the generated 3-D models' textures and orthoimages [1].

Thus far, the rippling caustics effect has not attracted enough attention from the scientific communities. However, during the last years, where automated SfM and MVS techniques are widely used by nonexperts, especially for mapping submerged cultural heritage or benthic ecosystems in shallow waters, caustics have been getting a lot of attention since, in many cases, they prevent the 3-D reconstruction [1], [10]. At the same time, only a few techniques have been proposed for the removal of caustics from images and videos in the context of image enhancement. However, recent literature [1], [10] indicates that the more successful the caustics removal is, the more valid matches are appearing in the stereo pairs and further processing with structure-from-motion and multiview stereo (SfM-MVS) techniques becomes possible for a number of applications.

The technical contributions of this work are as follows.

1) A new method for correcting the radiometric effects of rippling caustics on the underwater imagery in shallow areas. The method, contrarywise to the state-of-the-art, can handle seabed of any anaglyph, correcting the images using real pixel information. It first relies on state-of-the-art deep learning tools which can accurately classify the pixels of the image and then exploits the 3-D geometry of the scene to achieve a pixelwise correction, by transferring color values between the overlapping images. The proposed pixelwise image correction method improves 3-D reconstruction results or in many cases enables the 3-D reconstruction in areas that was previously impossible.

2) A new benchmark data set: The R-CAUSTIC: Rippling CAUSTICs underwater Image data set, is the first real-world annotated data set for caustics detection and correction. The data set includes seven areas of different characteristics and ground truth images without caustics, images with caustics and a binary annotation of caustics for each pose as well as camera calibration parameters. The specific data set provides an opportunity to evaluate, at least to some extent, the performance of different caustics detection and image segmentation approaches.

The rest of this article is organized as follows. Section II presents the physics and the effects of reppling caustics in key point detection and matching, and discusses the related work and limitations. Section III presents the real world data set on underwater rippling caustics while Section IV presents the proposed method for pixelwise image radiometric correction. In Section V, the steps of the proposed method for pixelwise image correction are tested over real-world data sets. Finally, Section VI concludes this article.

II. CONTEXT AND RELATED WORK

A. Rippling Caustics

In optics, a caustic or caustic network is the envelope of light rays resulting from the projection of light rays being reflected or refracted by a curved surface [11], such as the wavy interface between air and water. Therefore, caustics can be the patches of light or their bright edges, often having cusp or spinode singularities [12].

Rippling caustics are commonly formed when light shines through waves on a body of water [12] (see Fig. 1). Kinematically, wavy fringes can be clearly recognized from specks due to their relative smooth motion in contrast to the quick twinkling behavior of the scattering. So, sunlight rippling caustics have an optical flow field, while specks do not. Also, morphologically, sunlight rippling caustics have characteristic strip and ringlike outlines (see Fig. 1) in contrast with the randomly dispersive location of the specks. On the contrary, the morphological characteristics of the caustic waves, like the connectivity existing among many brilliant points (see Figs. 1 and 2), is much more discriminating than the spectra of the brightness gradient. Connected bright pixels differentiate from the underlying background scene by the fact that the scene generally looks rather matte textured, but not dominantly brilliant. Furthermore, the brilliant points of the scattering are relatively small, well spread, and generally not connected to each other [13].

1) Rippling Caustics Impede Key Point Detection and Image Matching: To demonstrate the severe effects of rippling caustics in the underwater imagery in shallow waters, and consequently, their effects in key point detection and matching processes, real-world data are presented and processed below. In Fig. 2, three consecutive images with caustics of the same seabed area, captured from exactly the same camera position and orientation and with an interval of 5 s (from left to right) are presented (more details on the data set used can be found in Section III). It is obvious that rippling caustics, being dynamic phenomena, cause a differentiation of the pixel values of the same area of the bottom. Key points detected using the SIFT algorithm [14] are also mapped on these images with blue dots. As can be seen in the zoomed areas of the images depicted in the third row of Fig. 2, the detected key points on and around the areas affected by the phenomenon are totally different for the images acquired at time t, t + 5 seconds, and t + 10 seconds. Indeed, the boundaries of the rippling caustics on the seabed, appear to be a dominant area in the feature detection step, due to the dominant gradients between the bright and darker areas of the image. On the contrary, the interior area of rippling caustics, which is very bright or sometimes burnt, is inappropriate for feature detection since it is characterized by the absence of texture. Together with the variation of the phenomenon through time, these are exactly the reasons why these effects are adversely affecting image matching algorithms by throwing off most of them, leading to less accurate matches [1] and causing issues even in SLAM navigation in shallow waters [9].

To demonstrate these negative effects of rippling caustics on the key point detection and matching processes, experiments were performed on the images presented in Fig. 2 and their respective caustics free image, using SIFT [14] algorithm and brute-force matching (see Fig. 3).

It is of really high importance to state here that the attempted matching process was performed on images having exactly the same camera orientation, thus, depicting exactly the same area of the seabed in a different time. The matching of the above three consecutive frames is attempted and compared with the matching results over the caustics-free images, where the caustics-free image is matched with itself. This matching was
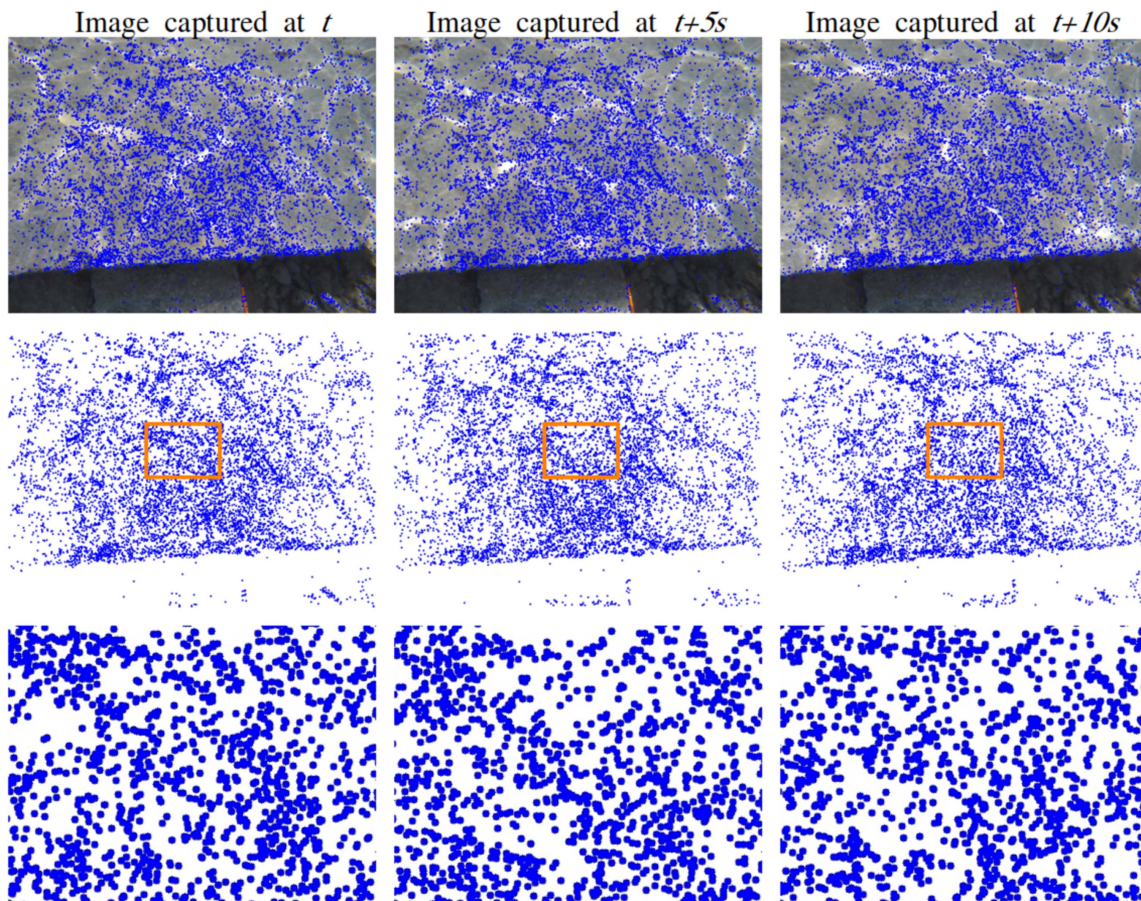
Fig. 2. Images with caustics of the same seabed area from the same camera position, captured with an interval of 5 s (from left to right) and the detected key points using SIFT [14]. In the first row the RGB images are demonstrated, in the second row the detected points on a white background while in the third row zoomed areas are depicted.

based on n-space Euclidean distance and performed both from left-to-right and right-to-left for redundancy. To filter these matches, the RANSAC algorithm was utilized to identify the inliers of the obtained point correspondences [15]. At the end of this step, a set of matched points is found in the given scenes, which are presented in Fig. 3 with lines of blue color. The matched points between the images captured at t and t+5 seconds are presented in Fig. 3(a) while the matched points between the images captured at t+5 and t+10 seconds are presented in Fig. 3(b). In Fig. 3(c), the matches between the image itself, without rippling caustics, are presented.

As also performed in [1], during the performed tests, it was decided not to evaluate the number of the total and valid matches only, but also the geometry of the matches, since some valid results of the RANSAC filtering are still matching the wrong points. Quantitative results suggested that in the first two matching examples, the matched points between the first two images were 47, between the second and the third were 43, while between the same caustic-free image were 1044. The above results suggest that the rippling caustics effect indeed affects key point detection and matching process in the underwater imagery in shallow waters, thus affecting the image-based 3-D reconstruction in these areas.

B. Related Work on Caustics Detection and Removal

For many years, the computer graphics research community has focused on the generation of caustics and as a result, many techniques have been proposed that generate photorealistic results. At the same time, only a few techniques have been proposed for the removal of caustics from images and videos in the context of image enhancement. A brief overview of the most relevant work in caustics removal is provided next.

Trabes et al. [13] proposed a technique that involved tuning a filter for sunlight-deflickering of dynamically changing underwater scenes. They employed a continuous parameter optimization inside a basic filter, which provided feedback for further improving the performance of the filter. Being an optimization, the filter's performance is highly sensitive to suboptimal parameters, and in particular, the segmentation parameter which is part of the objective function in the optimization. A different approach was proposed in [16], where a mathematical solution was presented involving the calculation of the temporal median between images within a sequence. A strong assumption of this work, is the fact that feature matching (Harris corner detection variant in [17]) is employed for the formation of the sequence which makes this approach very susceptible to the light
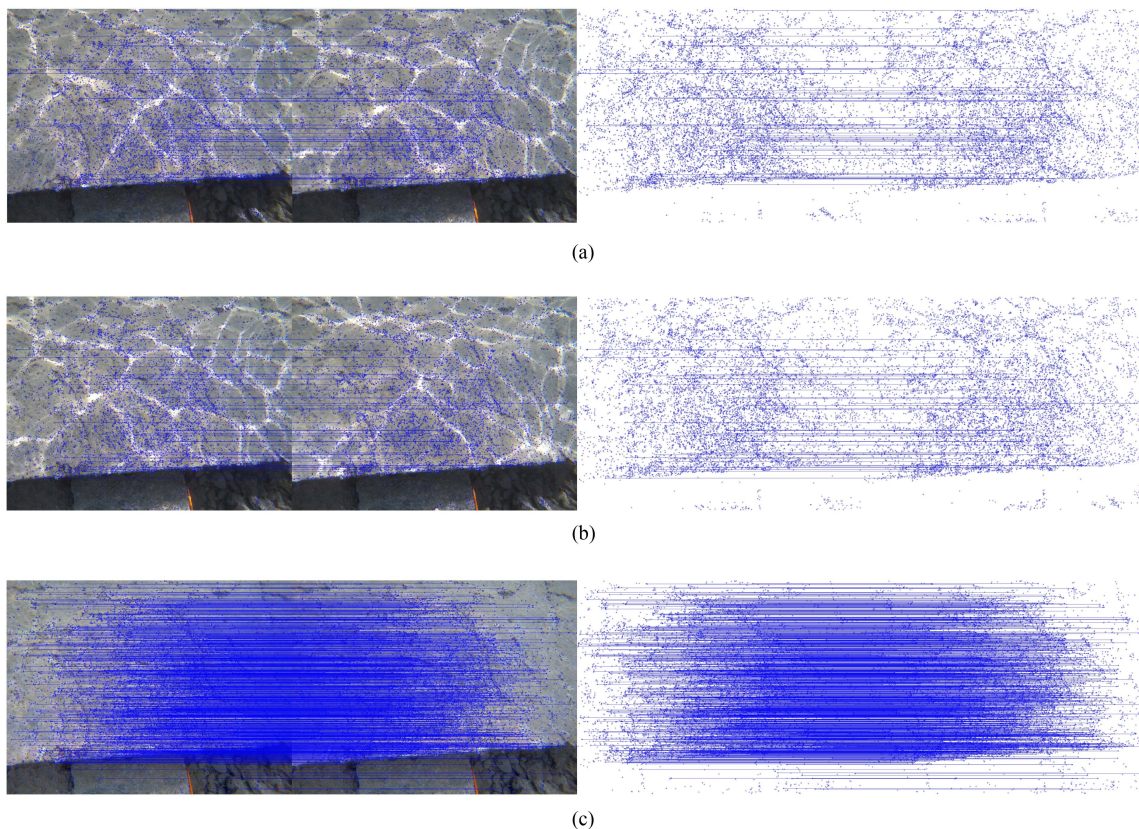
Fig. 3. Matching results between the image captured at t and t+5 seconds (a), between the image captured at t+5 and t+10 seconds (b), and between the image without caustics and itself (c). Left column demonstrated matches over the RGB images while right only matches on a white background.

variations in the images and, in particular, caustics effects. The same authors later extended their work in [18] and proposed an online sunflicker removal method that treated caustics as a dynamic texture. As reported in the paper, this only works if the seabed or bottom surface is flat. Similar approaches have also been proposed for general cases of dehazing and descattering of images such as [19], [20], and [21]. Schechner and Karpel [22] proposed a method based on processing a number of consecutive frames. These frames were analyzed by a nonlinear algorithm which preserved consistent image components while filtering out fluctuations. Their proposed method, however, does not take into account the camera motion which almost always leads to registration inaccuracies. Forbes et al. [10] proposed a solution based on two small and easily trainable convolutional neural networks (CNNs). To detect caustics, a small CNN was trained over synthetic data to overcome the obstacle of not having ground truth data available for real world underwater caustics. Being the first deep-learning based solution for caustics removal, this method is very important for a variety of applications, however, when it comes to reliable and accurate underwater 3-D reconstruction and mapping, especially of submerged cultural heritage or benthic community, artificial parts on the imagery should rather be avoided. This proposed solution was extensively evaluated in terms of keypoint detection, image matching, and 3-D reconstruction performance in [1]. Tripathy et al. [23] presented a clustering-based approach for underwater sunlight

flicker removal. As in some other works, their method is based on temporal median calculation and thresholding to be used as an online algorithm for AUVs to minimize the sunlight flicker distortions in the frames during the video survey and was tested in a controlled pool environment.

Other methods for correcting the underwater imagery, i.e., a piecewise linear transformation between the overlapping areas of the images, exploiting the matched keypoints as control points for the transformation, fail to deliver accurate and reliable 3-D reconstruction, since the x-parallax of the 3-D objects is not taken into account, delivering point clouds of different depths on the corrected areas.

Despite the innovative and complex techniques developed, addressing caustics removal with current procedural methods requires that strong assumptions are made on the many varying parameters involved, e.g., scene rigidity, camera motion, seabed flatness, etc. As demonstrated in [1], replacing the caustics affected area with synthetic pixel information, delivers quite appealing results, albeit slightly increases the matched points by 2%–3%, since caustics are replaced with artificial information that cannot be cross-matched between overlapping images. Moreover, real ground truth for caustics is not easily available, preventing recent advances in machine learning to jump in. Contrary to the state-of-the-art, the developed method can handle seabed and riverbed of any anaglyph, correcting the images using real pixel information, thus, improving image matching and 3-D
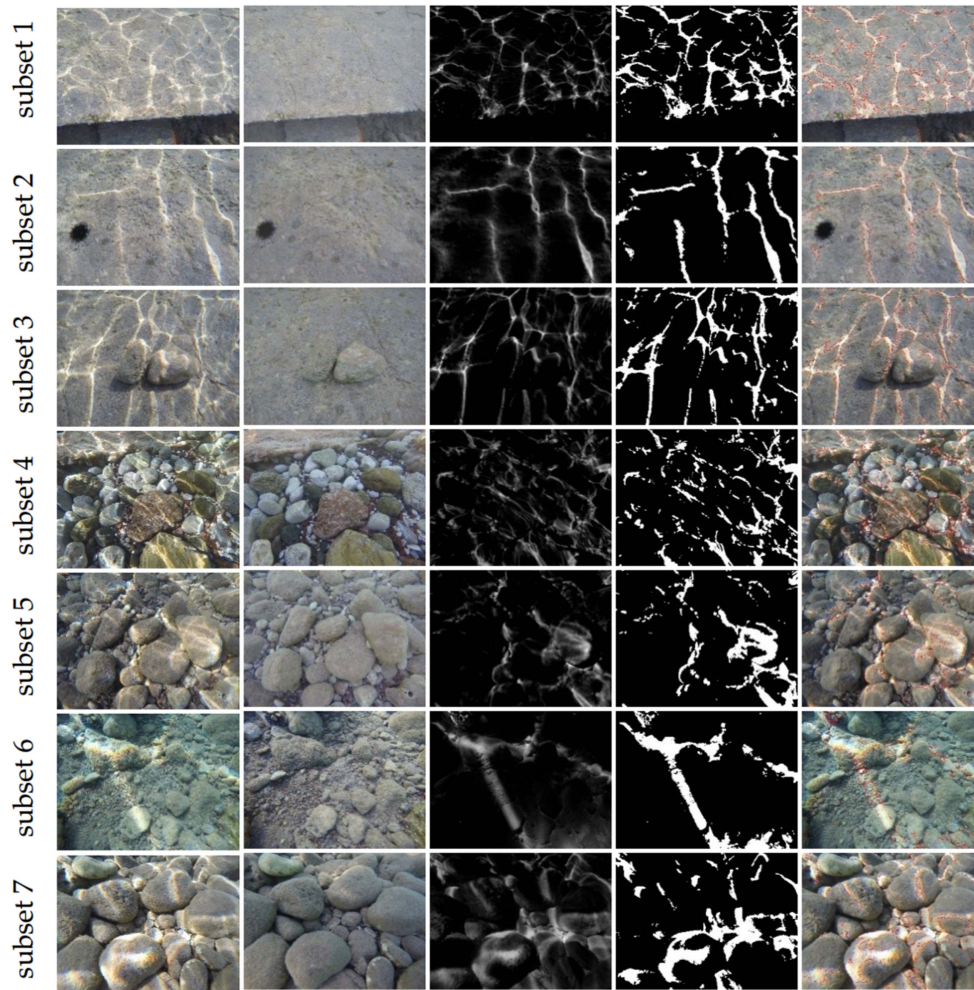
Fig. 4. Sample images from the R-CAUSTIC data set. First column: RGB images. Second column: The reference images. Third column: The difference images between the images with caustics and the reference images. Fourth column: The thresholded difference image. Fifth column: The RGB images with the detected contours. Each row depicts images for the respective subset of the data set.

reconstruction processes. In particular, the developed method employs deep learning architectures to classify image pixels to "noncaustics" and "caustics." Then, exploits the 3-D geometry of the scene to achieve a pixelwise correction, by transferring appropriate color values between the overlapping underwater images. Moreover, to fill the current gap, we have collected, annotated, and structured a real-world caustic data set, namely, R-CAUSTIC, which is openly available. Overall, based on the experimental results and validation, the developed methodology is quite promising in both detecting caustics and reconstructing their intensity.

III. R-CAUSTIC: RIPPLING CAUSTICS UNDERWATER IMAGE DATA SET

In the literature, the data sets presented in [24] are the only ones available related to rippling caustics. These data sets contain raw material and results of the underwater experiments performed in a pool, in the Red Sea and in the Mediterranean Sea. However, the images are of low resolution and the data set is

not created with the purpose of being used with machine learning and deep learning frameworks due to the lack of ground truth images. Also, the scenes of this data set are not representing a realistic set up for underwater image-based 3-D reconstruction of the seabed or lakebed.

In this work, to fill this gap, R-CAUSTIC, which is a real-world underwater caustics benchmark data set containing 1465 underwater images is presented (see Fig. 4). Together with the RGB imagery, the corresponding generated ground truth images are provided for facilitating the training and testing of machine and deep learning methods for image classification. The data set contains 7345 images in total. The specific data set also provides the necessary data to evaluate, at least to some extent, the performance of 3-D reconstruction approaches.

A. Data Collection

Data were acquired using a GoPro Hero 4 Black action camera with image dimensions of 4000×3000 pixels, focal length of 2.77 mm, and pixel size of 1.55 μm, and a tripod. Action cameras
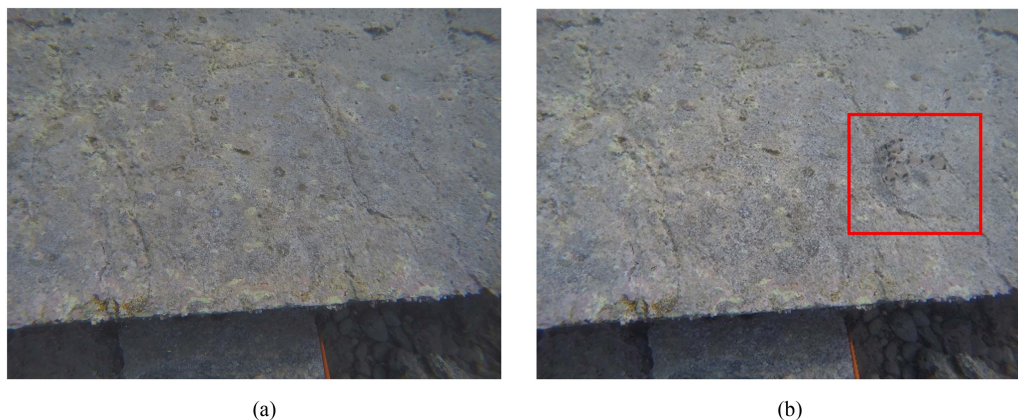
Fig. 5. Reference image created using the artificial shadow in (a) and the reference image created using the pixels having the lowest values compared with the pixels of the same position in the rest of the imagery in (b). In the red rectangle, the multiple fish appearing in the reference image are highlighted.

are widely used for underwater image acquisition [2]. The data set was captured in near-shore underwater sites at depths varying from 0.5 to 2 m. No artificial light sources were used. Due to the wind, the turbulent surface of the water created dynamic rippling caustics on the seabed. In total, 1465 images were collected using the same camera, separated into seven different subsets according to the seabed type of the depicting area or their overlap; subsets 1, 4, 5, 6, and 7 contain stereo images, subset 3 tri-stereo images, and subset 2 consists of multistereo imagery acquired in seven consecutive camera poses. As can be seen in Fig. 4 where each row stands for a typical image from each subset, the collected imagery presents a large variability in terms of scene complexity, color, caustics complexity, frequency, and scale. Images were collected with five seconds interval to describe as much different instances of the caustics as possible. The 7 subsets of the data set capture flat and 3-D seabed surfaces.

1) Reference and Ground Truth Image Generation: To capture the reference images of the data set (see Fig. 4, second column), an artificial shadow was introduced in the scene, above the water, to protect the water surface from the sun rays [see Fig. 5(a)]. Since the images of the same branch of the data set (i.e., left or right) are captured by the same camera position, an alternative approach was also implemented for the reference images generation. This was to create an image using the pixels having the lowest values compared with the pixels of the same position in the rest of the imagery. To perform that, images were transferred to the lab color space to choose the pixels with the lower luminosity.

Although this approach is generating images with great quality, in some cases, in the created reference image objects appear that are apparent in one of the scenes, having lower values in the lab space compared with the rest of the pixels of this position. This led, in many cases, to the appearance of multiple fish in the reference image, even in the branch of the data set there is only one fish moving between the image instances [see Fig. 5(b)]. Having captured the reference images for each part of the data set, the per-element differences $\Delta(x, y)$ between the reference and each image of the data set containing caustics is calculated and saved in a new image, the difference image (see

Fig. 4, third column). To generate a more accurate difference image, containing only caustics and not other differences in the pixels' colors, a color transferring approach between the images containing caustics and the reference images is performed, as described in Section IV-B1. Color transferring is performed from the images with caustics to the reference image. Also, since the difference image contains noise due to scattering and passing particles, the image is smoothed with a spatial Gaussian pre-filter with a kernel size of 3 x 3 to 7 x 7 pixels, depending on the scene.

To compensate for changes in illumination conditions during the acquisition phase and prepare the imagery for the ground truth image generation, the difference images' pixel values are scaled and shifted so their minimum value is 0 and the maximum value is 255. Finally, the thresholded difference images are created (see Fig. 4, fourth column). To ensure that the ground truth images do not include any false positives, a Canny edge detector [25] is employed to facilitate the projection of the detected contours on the original images containing caustics for visual inspection (see Fig. 4, fifth column).

Although the above procedure is very consistent for generating the ground truth images, it must be noticed that caustics are very complex physical phenomena and by their nature, they do not have a very specific boundary. When the camera to object distance is quite small (i.e., 0.5 m) and the incidence angle of the light rays to the surface of the water is also large, their boundary is characterized by very intense chromatic aberration effects. On the contrary, when the incidence angle is closer to the perpendicular, their boundary is of the same color values as the core of the caustics bright white. The chromatic aberration intensity is also depending on the direction of the ripples compared with the lighting rays' direction. As the angle of the ripples with the lighting direction increases, the aberration increases too.

2) Underwater Camera Calibration: Since the collected data set enables also the 3-D reconstruction of the imaged areas of the seabed, underwater camera self-calibration [26] was performed to deliver the retrieved parameters together with the data set. Self-calibration techniques do not use any calibration object. If images are taken by the same camera with fixed internal parameters, correspondences between three
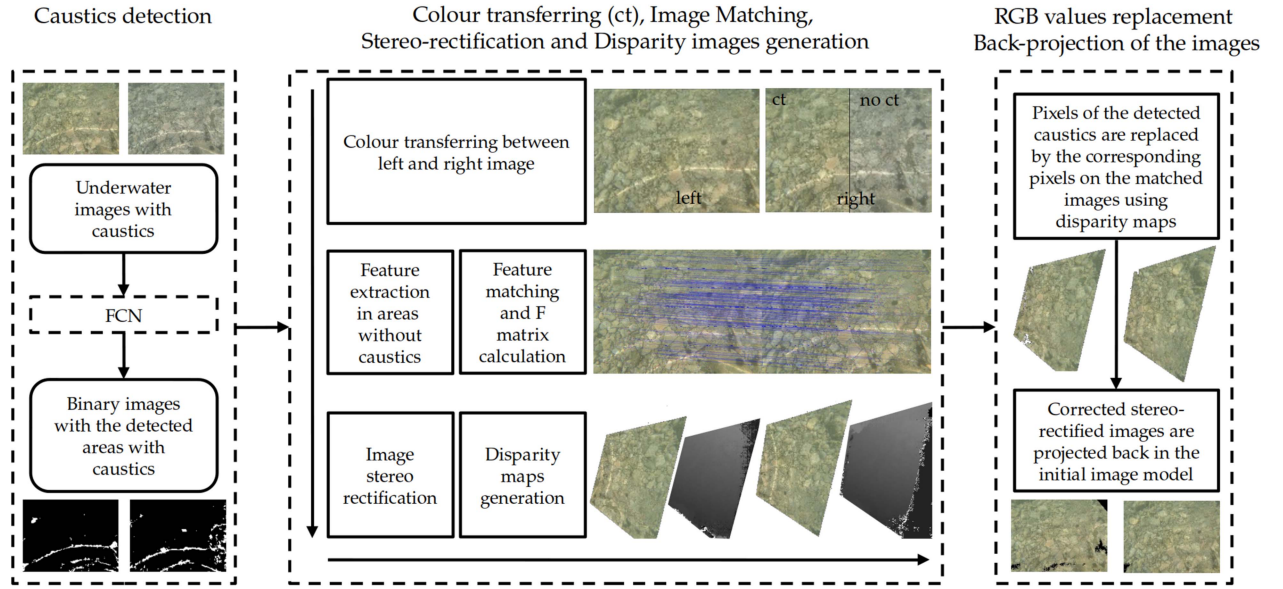
Fig. 6.   Proposed method for the pixelwise correction of the rippling caustics on the underwater imagery.

images are sufficient to recover both the internal and external parameters [27]. Nowadays, self-calibration is widely applied for the camera-housing system in underwater cases, as it is assumed that refraction effects of the housing are compensated by the interior orientation parameters [7], [28]. In the performed approach, since the intrinsics were shared between all the images of the data set, the intrinsic camera parameters were estimated through COLMAP [29] SfM pipeline which exploits a RANSAC variant called P3P(f)-RANSAC proposed by [30].

Results are included in the data set as a separate file. The camera models used are of OpenCV [31]. For more information on the models, readers may refer to [26], [31], and [32]. Below, the estimated intrinsics are given with unrefined principal point and with refined principal point, respectively

$$[f_x, f_y, c_x, c_y, k_1, k_2, p_1, p_2]$$
$$= [2416.65, 2424.26, 2000, 1500,$$
$$\quad - 0.092734, 0.108091, 0.002016, -0.000773]$$
$$[f_x, f_y, c_x, c_y, k_1, k_2, p_1, p_2]$$
$$= [2343.20, 2346.41, 1930.11, 1481.97,$$
$$\quad - 0.101428, 0.098774, -0.000378, 0.000014].$$

IV. PROPOSED METHOD FOR PIXELWISE IMAGE RADIOMETRIC CORRECTION

To deal with the feature detection and matching problem in images with caustics, the very accurate detection of the not affected areas of the images is proposed in this work, enabling feature detection only on these areas. For the approach presented here, it is of utmost importance to maintain as much of the original RGB information of the images as possible. Next, to correct the affected imagery, a pixelwise method based on the stereo or the multiview geometry is proposed. This method works

on the overlapping area of the imagery and takes advantage of the pixel correspondences in this area. The overall workflow of the proposed method is illustrated in Fig. 6 while the different modules of the method are described in detail in the following sections.

Initially, to detect the unaffected areas with very high accuracy, reliability, and repeatability over different caustics patterns, different types of seabed, luminosity, and visibility conditions, a fully convolutional neural network (FCN) classifier is exploited. The classifier is being trained using the R-CAUSTIC data set.

Having detected the unaffected and consequently the affected areas on the initial imagery since it is a binary classification problem, a color transferring approach is performed between the consecutive or overlapping images. These images are then matched using only the key points detected on the not-affected areas of the images classified as "noncaustics," are stereo-rectified and their respective disparity maps are generated. In this step, the binary images resulted by the FCN are also reprojected using the homography matrices computed for the initial imagery rectification. Having all the necessary data available, the pixels classified as "noncaustics" are not processed while the rest of the pixels are replaced by their corresponding pixels on the matched images using the disparity maps. Finally, the corrected stereo-rectified images are projected back onto the initial camera model to facilitate further SfM-MVS processing and texturing.

A. Pixel Classification

Although the unaffected areas are appearing on the images as pixels having lower exposure than the caustics' pixels, for the method presented here, a sophisticated approach was chosen to be implemented. This was decided because of the very intense effects of chromatic aberration appearing on the boundaries of caustics, not being able to be detected by a simpler method such

as thresholding. Moreover, it was also important to classify as not-affected areas white colored areas and reflective materials of the seabed such as, e.g., marbles and ground control points (GCP) markers placed for photogrammetric campaigns in the shallower waters (see Section V). Since underwater imagery can be captured by very different distances from the seabed, different imaging sensors, different luminosity, and visibility conditions and caustics size might vary depending on the water surface state, it was considered important for the deep fully convolutional network to be able to deal with multimodal and multiscale image data for semantic labeling. To that direction, several state-of-the-art architectures were implemented and used for training and testing, providing insights for their performance over the specific problem of caustics. As such, a similar approach to the one presented in [33], which is based on SegNet [34], a U-Net [35], two FCN models with ResNet50 and ResNet101 backbones [36], respectively, and two DeepLabV3 [37] models with a ResNet50 and ResNet101 backbones, respectively, were trained, evaluated, and tested using the R-CAUSTIC data set as well as real world images. For more information on these results (see Section V).

B. Stereo-Rectification and Disparity Maps Generation

The most important steps of the proposed method are the accurate matching and stereo-rectification of the imagery and, consequently, the accurate disparity maps generation and filtering. However, before these steps, a color transferring approach is performed between the images of the stereopair or even the whole block.

1) Color Transferring in the CIE-$l\alpha\beta$ Color Space: Color transferring is an important step of the proposed method. It will prevent the appearance of visible patches of pixels in the corrected imagery and will deliver seamless corrected imagery. Indeed, due to different exposure and white color balancing of the images, it is very likely that the replaced regions are clearly recognizable by color differences creating visual artifacts in the image that are not part of the object itself. The main goal of this step is to obtain a visual consistent and pleasing corrected image, by equalizing the color profile of all images of the data set while fine details are preserved

To apply this color transferring approach, the well-established methodology presented in [38] is exploited. It is based on the $l\alpha\beta$ color spaced developed by [39] which minimizes the correlation between the three image channels. This facilitates different operations in different color channels with the confidence that undesirable cross-channel artifacts will not occur [38]. In addition, this color space is logarithmic, which means, to a first approximation, that uniform changes in channel intensity tend to be equally detectable [38]. In the implemented method, first, the RGB images are converted to $l\alpha\beta$ color space. There, the l axis represents an achromatic channel, while the α and β channels are chromatic yellow–blue and red–green opponent channels. To achieve the color transferring, the mean and standard deviations along each of the three axes are computed for both the source and target images. These mean values are subtracted from the image data points [38], i.e., the values of each channel. Consequently,

the resulted data points are scaled comprising the new image by factors determined by the respective standard deviations. After this transformation, the new data points (of the new image) have standard deviations that conform to the image. Next, instead of adding the averages that were previously subtracted, the averages computed for the image are added. Finally, the result is converted back to RGB via log-LMS (long, medium, short), LMS, and XYZ color spaces and the new image is ready for further processing. Color transferring was performed sequentially, starting from an image selected by the user or the first image of a data set.

2) Feature Detection, Matching, and Epipolar Geometry Retrieval: Next, feature detection is performed using BRISK [40] detector. BRISK is adaptive, offering high quality performance as in other state-of-the-art algorithms, albeit at a dramatically lower computational cost (an order of magnitude faster than SURF in some cases) [40]. To detect matchable features, the generated binary images from the previous step are exploited here. To that direction, for detecting the necessary features, BRISK [40] is working only in areas that are classified as "noncaustics" in the binary mask. Even for this specific approach BRISK detector was chosen due to its lower computational cost. SIFT, SURF, ORB, or other state-of-the-art detectors can also be used to detect the necessary features.

The detected features on an image are then matched to the corresponding features on the overlapping images and the mapping of these features between these images is stored in a vector. This matching is based on n-space Euclidean distance and it is performed both from left-to-right and right-to-left for redundancy. However, since in feature matching, several blunders might occur, the RANSAC [15] algorithm is utilized to identify the inliers of the obtained point correspondences. The algorithm takes all the matched points as input, formulates a mathematical model that incorporates the majority of the points, and filters out the remaining points which are considered as outliers. To accomplish that, the fundamental matrix is computed and the measure for thresholding inlier points is the distance from the epipolar line. At the end of this step, a set of matched points is found in the given scenes.

3) Stereo-Rectification and Disparity Maps Generation: Knowing the epipolar geometry of the overlapping images, the initial imagery is then projected to form the stereo-rectified imagery in pairs. By using exactly the same matrices which describe the epipolar geometry of the two images, the predicted binary images resulting from the FCN are also projected in their stereo-rectified form. These projections are necessary to reduce the complexity of the problem across horizontal epipolar lines and facilitate the pixelwise image radiometric correction. Since for the vast majority of the divers acquired underwater imagery, the optical axes of the cameras are not parallel to each other, to have horizontal epipolar lines parallel to the baseline, the reprojection of both image planes onto a common plane parallel to the baseline needs to be performed. The rectification approach followed was first proposed in [41] and involves the decomposition of each rectifying homography into a projective and an affine component. Then the projective component that minimizes a well-defined projective distortion criterion is found. The affine

component of each homography is further decomposed into a pair of simpler transforms; one designed to satisfy the constraints for rectification, the other is used to further reduce the distortion introduced by the projective component.

The stereo-rectified imagery is then used for the disparity maps generation. This is achieved by stereo-processing the imagery by semiglobal matching (SGM) [42], [43], evaluating in eight line directions instead of 16, to reduce the processing time. The exact approach followed is described in [42] where the consistent semiglobal matching (CSGM) is first presented. Compared with SGM, CSGM uses the same steps but also the intensity consistent disparity selection. It uses a pixelwise, mutual information-based matching cost for compensating radiometric differences of input images, a feature necessary when processing images with caustics. Pixelwise matching is supported by a smoothness constraint that is usually expressed as a global cost function. CSGM performs a fast approximation by pathwise optimizations from all the eight directions. In addition, postprocessing steps for removing outliers, recovering from specific problems caused due to the caustics' effects and the interpolation of gaps are also applied.

Occlusions and mismatches can be distinguished as part of the left/right consistency check. Regarding the interpolation for the gaps caused by the mismatched pixel areas on the caustics, it is performed by propagating valid disparities through neighboring invalid disparity areas. To achieve the best possible interpolation, this is done similarly to SGM along paths from eight directions. According to [42], where the implementing postprocessing was first presented, this approach emphasizes the use of all information without a preference to foreground or background. Moreover, instead of the mean, the median is used for maintaining discontinuities in cases where the mismatched area is at an object border. The implemented interpolation method has the advantage that it is independent of the used stereo matching method. Finally, median filtering can be useful for removing remaining irregularities and additionally smooths the resulting disparity image.

C. RGB Values Replacement and Back-Projection of the Images

Considering that for each stereo-rectified image, the binary image of "noncaustics" and "caustics" classes and the disparity maps are available in the same projection, the pixelwise image correction can be performed. To that direction, for each pixel of an image that is not classified as "noncaustics" in the binary image, the RGB values of the corresponding pixel in the overlapping image are found, using the disparity map and vice versa. The relation between the source and the target pixels is expressed in the following:

$$x_{target} = x_{source} - disparity_{source} \qquad (1)$$

$$x_{source} = disparity_{target} + x_{target} \qquad (2)$$

where x_{target} is the horizontal coordinate of the pixel in the n+1 image of the stereo pair and x_{source} the horizontal coordinate of the pixel in the n image of the block. This process is performed only if the target pixel is classified as "noncaustics."

That way, the areas that are not classified as "noncaustics," are pixel-wisely corrected by exploiting the rigorous geometry of the stereo-pair. By using this approach, the disparity (x-parallax) of each pixel is taken into account, facilitating a more accurate and reliable correction approach that is not adversely affecting the later SfM and MVS steps, but it improves them. This is achieved as this horizontal displacement between rectified feature points is related to the depth of the feature. This way the method can be used to recover the images over 3-D structures without affecting the 3-D position of the SfM and MVS processing calculated subsequent.

V. EXPERIMENTAL EVALUATION

In this section, the proposed method for pixelwise image correction is tested end evaluated over the R-CAUSTIC data set and other real-world images collected, proving its robustness, accuracy, and reliability.

A. Pixel Classification

To form a reliable image pixel classification model, the data set presented in Section III was used to train and cross-validate SegNet [34], U-Net [35], two FCNs with ResNet50 and ResNet101 backbones [36], respectively, and two DeepLabV3 [37] models with a ResNet50 and ResNet101 backbones, respectively. To that direction and to prove and evaluate their generalization and domain adaptation potential over different types of scenes and subsets, seven training-testing approaches were performed for each of the FCNs.

To achieve that, the seven subsets were categorized based on the similarity of the scene in terms of caustics frequency, patterns and seabed uniformity and characteristics. Subsets 1, 2, and 3 are considered to have similar background, although caustics are of very different scale and complexity, subsets 5 and 7 are also considered similar. The rest of the subsets are not similar to each other. They represent totally different types of seabed and caustics.

B. Choosing a Shallow or a Deep ML Architecture?

Concerning the selection of a deep architecture for addressing the image classification problem, before the decision that an FCN would be more appropriate for the proposed solution, several tests were performed using shallower ML architectures. Training and testing were performed for classifying the underwater imagery using AdaBoost [44] and specifically AdaBoost-SAMME [45] having decision trees [46] as the base estimator, decision trees [46], K-nearest neighbors (KNN) [47], quadratic discriminant analysis (QDA), random forests [48] and linear support vector machines [49]. Also, a simple thresholding approach was performed. Regarding the thresholding, its value was changed over the different images, while for the rest of the methods, the default values as set in [50] were used. Resulting metrics from training over subsets 4, 5, 6, and 7 and testing over subsets 1, 2, and 3 are presented in Table I.

As can be seen in Table I, the metrics achieved by all the FCNs outperform the rest of the methods used for image classification.

TABLE I
TESTING PERFORMANCE OF THE VARIOUS METHODS

Method	F1 (%)	Accuracy (%)
AdaBoost	89.02	80.48
Decision Trees	87.37	77.93
KNN	87.61	78.30
QDA	84.34	73.49
Random Forests	87.56	78.21
Linear SVM	89.61	81.40
SegNet	96.46	93.89
U-Net	96.32	93.64
FCN+ResNet50	96.48	93.96
FCN+ResNet101	96.50	94.00
DeepLabV3+ResNet50	95.79	92.96
DeepLabV3+ResNet101	95.98	92.96

Number in bold are the best metrics achieved. F1 stands for "noncaustics" F1 score.

The linear SVM follows with almost 10% less accuracy and after that, AdaBoost comes with almost 13% less accuracy. By comparing the precision score of the rest of the classifiers, they are achieving quite similar percentages, however, this is not the case for the recall and the F1 scores where only the linear SVM and the AdaBoost classifiers are quite close. It is also evident that AdaBoost is outperforming the results of the decision trees, however, this was expected. For the majority of the performed tests, the shallower classifiers as well as the thresholding were not able to distinguish between caustics and other bright artifacts on the scene, i.e., white rocks with high accuracy. Moreover, they did not succeed in reliably detecting the boundary of the caustics when it is characterized by the intense chromatic aberration effect. This is also reflected in the achieved metrics in Table I. Two typical examples are presented in Figs. 7 and 8.

There, the input RGB image is given, along with the predictions of the tested shallow architectures. In the red rectangles, some typical cases of false negatives, i.e., pixels that are wrongly classified as "caustics" are highlighted. On the other hand, orange rectangles indicate areas of false positive, i.e., pixels that are wrongly classified as "noncaustics." This, together with the experiments performed over the different subsets, highlights the difficulty of those methods to generalize over the different types of seabed and caustics.

The metric and visual results of the above experiments, justify the use of a deeper architecture for classifying the underwater imagery affected by caustics, as also used in [10]. Even if the resulted metrics of most of the tests performed using the shallower architectures are lower than those achieved by the exploited FCN architecture, they do not prohibit their use. However, one has to consider the specific goal of the application. As such, for a more close to real time application, where only key point detection and matching is needed, i.e., visual odometry, etc., AdaBoost, decision trees, KNNs, QDA, or even simple thresholding could

be used due to their minimal prediction times, compared to the SVMs and the FCN architectures.

However, when it comes to the maximum improvements on the sparse and the dense 3-D reconstruction using the less necessary images, a deep architecture seems to be the best solution. By classifying the pixels of the underwater imagery with the highest accuracy, only the pixels that it is really necessary to be replaced are undergoing the correction process, keeping as much as possible of the original imagery unprocessed. This way, there is more area on the image available for the keypoint detectors, ensuring a more robust SfM process. In addition, as already proved, the FCN, achieves to detect in an accurate and reliable degree the intense chromatic aberration effects on the boundaries of the caustics, something really important for facilitating a more realistic corrected imagery and as will be shown in Section V-D, enabling the generation of a more complete dense 3-D point cloud.

1) Training and Cross-Validating Deep ML Models: Taking into account the categorization of the subsets based on the type of seabed and caustics they are depicting, eight (8) different training and testing approaches were formed (see Tables II and III) to check the generalization and domain adaptation potential and the trained networks. For each case, six different models where trained and tested, forming the respective subcase: (a) for SegNet [34], (b) for U-Net [35], (c) and (d) for two FCNs with ResNet50 and ResNet101 backbones [36], respectively, and (e) and (f) for the two DeepLabV3 [37] models with a ResNet50 and ResNet101, respectively. In the first case, the models are trained over subsets 4, 5, 6, and 7 and testing is performed over subsets 1, 2, and 3. In the second case, the models are trained over subsets 3, 4, 5, 6, and 7 and testing is performed over subsets 1 and 2. The third case involves training on 1, 2, 3, 4, and 6 and testing on subsets 5 and 7. For the fourth approach, the models are trained over subsets 1, 2, 3, 5, and 7 and tested on subsets 4 and 6. In the fifth, sixth and seventh approach, the models are trained over only subsets 1, 3, and 6, respectively, and tested over the same subsets; 2, 4, 5, and 7. Finally, in the eighth case, the models were trained using all the available data to form a model able to predict over various unseen data. The number of the images used for training these three models (three for each network) are exactly the same with the previous training approaches. Cases 5–7 aimed to prove the necessity of the variability of the scenes of the data set and to reply to the question whether it is better to train the model using N images coming from one subset or using N images coming from many subsets.

For training the FCNs, a sliding window approach to extract 128×128 patches was used. Due to the size of the images being too large to fit all at once to the GPU used, they were broken into those smaller patches for training the deep learning models. The stride of the sliding window defines the size of the overlapping regions between two consecutive patches. At training time, a smaller stride allows us to extract more training samples and acts as data augmentation. At testing time, a smaller stride allows us to average predictions on the overlapping regions, which reduces border effects and improves the overall accuracy. During training, cross-validation and testing, a 32 pixels stride was used. Models are implemented using the PyTorch framework. Torch is a scientific computing framework with wide support for
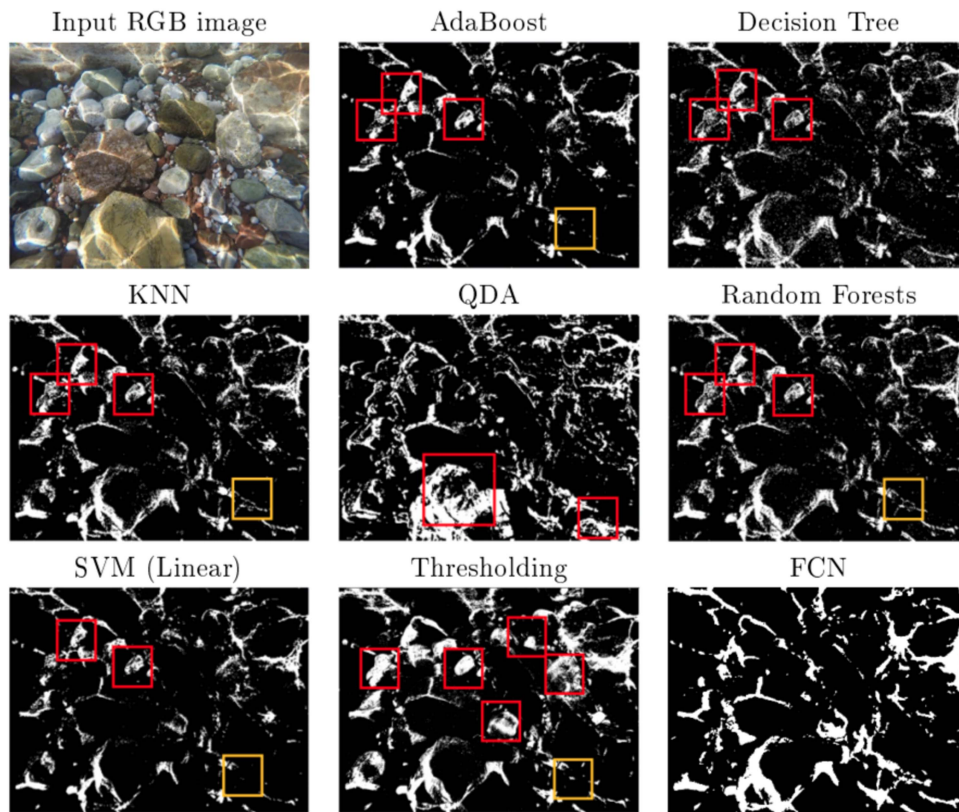
Fig. 7.   Input RGB image together with the predictions of the tested architectures for an image from subset 4.
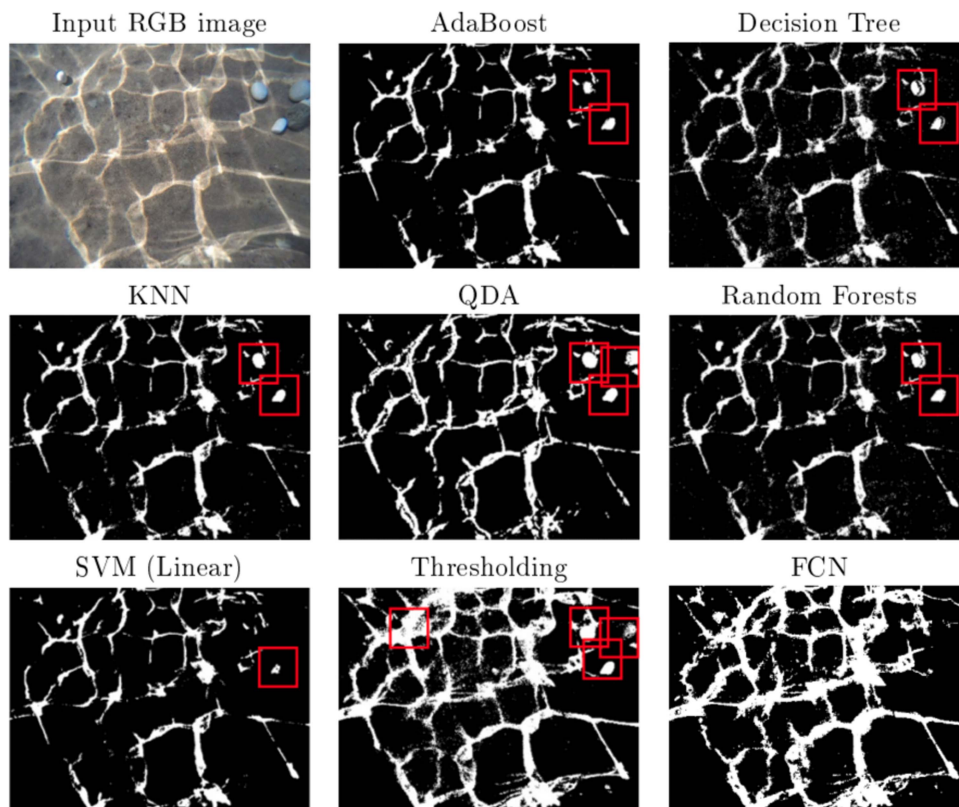

Fig. 8.   Input RGB image together with the predictions of the tested architectures for an image of a real world application, not contained in the data set.

TABLE II
AVERAGE TESTING PERFORMANCE AFTER 30 EPOCHS OF TRAINING

Case	Training subsets	Testing subsets	Method	Caustics F1 (%)	Non caustics F1 (%)	Total accuracy (%)
1a	4,5,6,7	1,2,3	SegNet	77.64 ±0.06	96.46 ±0.01	93.89 ±0.02
1b	4,5,6,7	1,2,3	UNet	76.53 ±0.24	96.32 ±0.02	93.64 ±0.05
1c	4,5,6,7	1,2,3	FCN-ResNet50	78.61 ±0.45	96.48 ±0.06	93.96 ±0.11
1d	4,5,6,7	1,2,3	FCN-ResNet101	**79.16** ±0.33	**96.50** ±0.07	**94.00** ±0.17
1e	4,5,6,7	1,2,3	DeepLabV3+ResNet50	78.62 ±0.41	95.79 ±0.07	92.96 ±0.13
1f	4,5,6,7	1,2,3	DeepLabV3+ResNet101	78.98 ±0.36	95.98 ±0.06	93.25 ±0.11
2a	3,4,5,6,7	1,2	SegNet	**83.59** ±0.01	**97.20** ±0.01	**95.22** ±0.01
2b	3,4,5,6,7	1,2	UNet	81.88 ±0.19	97.02 ±0.02	94.88 ±0.03
2c	3,4,5,6,7	1,2	FCN-ResNet50	83.22 ±1.31	96.91 ±0.40	94.82 ±0.59
2d	3,4,5,6,7	1,2	FCN-ResNet101	83.39 ±0.34	97.13 ±0.06	95.10 ±0.11
2e	3,4,5,6,7	1,2	DeepLabV3+ResNet50	81.25 ±0.05	96.30 ±0.02	93.82 ±0.04
2f	3,4,5,6,7	1,2	DeepLabV3+ResNet101	82.94 ±0.46	96.83 ±0.11	94.65 ±0.19
3a	1,2,3,4,6	5,7	SegNet	78.45 ±0.45	97.44 ±0.06	95.42 ±0.11
3b	1,2,3,4,6	5,7	UNet	79.08 ±0.51	97.44 ±0.06	95.44 ±0.11
3c	1,2,3,4,6	5,7	FCN-ResNet50	79.72 ±0.76	97.47 ±0.10	95.50 ±0.18
3d	1,2,3,4,6	5,7	FCN-ResNet101	**80.56** ±0.61	**97.61** ±0.08	**95.74** ±0.14
3e	1,2,3,4,6	5,7	DeepLabV3+ResNet50	79.72 ±0.61	97.16 ±0.14	95.16 ±0.16
3f	1,2,3,4,6	5,7	DeepLabV3+ResNet101	80.18 ±0.51	97.31 ±0.07	95.27 ±0.13
4a	1,2,3,5,7	4,6	SegNet	73.03 ±0.89	96.58 ±0.14	93.93 ±0.24
4b	1,2,3,5,7	4,6	UNet	73.19 ±0.13	96.57 ±0.14	93.92 ±0.24
4c	1,2,3,5,7	4,6	FCN-ResNet50	77.45 ±1.01	97.05 ±0.19	94.78 ±0.14
4d	1,2,3,5,7	4,6	FCN-ResNet101	77.87 ±0.97	**97.06** ±0.15	**94.81** ±0.26
4e	1,2,3,5,7	4,6	DeepLabV3+ResNet50	77.97 ±0.77	96.77 ±0.11	94.36 ±0.16
4f	1,2,3,5,7	4,6	DeepLabV3+ResNet101	**78.19** ±0.84	96.80 ±0.15	94.42 ±0.25
5a	1	2,4,5,7	SegNet	**72.99** ±0.09	**96.30** ±0.03	**93.49** ±0.05
5b	1	2,4,5,7	UNet	73.26 ±0.14	96.27 ±0.04	93.45 ±0.07
5c	1	2,4,5,7	FCN-ResNet50	71.64 ±0.16	95.84 ±0.04	92.74 ±0.07
5d	1	2,4,5,7	FCN-ResNet101	72.20 ±0.17	95.97 ±0.04	92.96 ±0.08
5e	1	2,4,5,7	DeepLabV3+ResNet50	70.91 ±0.13	95.56 ±0.04	92.30 ±0.06
5f	1	2,4,5,7	DeepLabV3+ResNet101	71.57 ±0.13	95.74 ±0.13	92.58 ±0.13
6a	3	2,4,5,7	SegNet	68.51 ±0.14	95.08 ±0.04	91.49 ±0.07
6b	3	2,4,5,7	UNet	67.62 ±0.19	94.90 ±0.07	91.19 ±0.11
6c	3	2,4,5,7	FCN-ResNet50	68.60 ±0.07	95.10 ±0.07	91.52 ±0.07
6d	3	2,4,5,7	FCN-ResNet101	**69.17** ±0.47	**95.22** ±0.09	**91.72** ±0.15
6e	3	2,4,5,7	DeepLabV3+ResNet50	65.42 ±0.53	93.97 ±0.14	89.74 ±0.23
6f	3	2,4,5,7	DeepLabV3+ResNet101	66.72 ±0.31	94.35 ±0.11	90.35 ±0.16
7a	6	2,4,5,7	SegNet	67.55 ±0.12	96.45 ±0.01	93.61 ±0.01
7b	6	2,4,5,7	UNet	63.01 ±0.19	96.22 ±0.01	93.14 ±0.0.01
7c	6	2,4,5,7	FCN-ResNet50	66.52 ±0.07	96.56 ±0.02	93.77 ±0.02
7d	6	2,4,5,7	FCN-ResNet101	64.71 ±0.27	96.58 ±0.02	93.77 ±0.02
7e	6	2,4,5,7	DeepLabV3+ResNet50	69.84 ±0.06	96.42 ±0.01	93.60 ±0.02
7f	6	2,4,5,7	DeepLabV3+ResNet101	**70.81** ±0.09	**96.84** ±0.01	**94.30** ±0.01

Entities in bold indicate the best score.

machine learning algorithms that puts GPUs first. All the models were trained over 30 epochs using stochastic gradient descent (SGD) with a base learning rate of 0.01, a momentum of 0.9, a weight decay of 0.0005, and a batch size of 60. The learning rate was divided by 10 after 20 epochs. Regarding the number of the epochs of training, various applied approaches indicated that more than 30 epochs offer nothing more to the network's performance. To this direction, the network was trained on 30 epochs where for each epoch were used 10.000 samples acquired

from ten images of each subset over 167 iterations. For the training approaches on which only one subset was used, the same number of samples was retrieved from 50 images.

Fig. 9 presents the training loss and cross-validation accuracy for training FCN-ResNet101 on subset 1 (case 5 d). Lines in green color represent the running-window cross-validation accuracy while lines in orange the cross-validation accuracy on all the validation data for each 100 iterations. Lines in red represent the training loss for the same running-window while

TABLE III
AVERAGE TESTING PERFORMANCE AFTER 30 EPOCHS OF TRAINING

Case	Training subsets	Testing subsets	Method	Caustics F1 (%)	Non-caustics F1 (%)	Total accuracy (%)
8a	70% of 1-7	30% of 1-7	SegNet	83.41 ±0.04	97.64 ±0.02	95.87 ±0.02
8b	70% of 1-7	30% of 1-7	UNet	83.54 ±0.20	97.62 ±0.03	95.83 ±0.05
8c	70% of 1-7	30% of 1-7	FCN-ResNet50	85.13 ±0.46	97.80 ±0.06	96.17 ±0.41
8d	70% of 1-7	30% of 1-7	FCN-ResNet101	**85.26** ±0.51	**97.83** ±0.08	**96.22** ±0.14
8e	70% of 1-7	30% of 1-7	DeepLabV3+ResNet50	84.16 ±1.24	97.42 ±0.28	95.56 ±0.47
8f	70% of 1-7	30% of 1-7	DeepLabV3+ResNet101	84.63 ±0.39	97.50 ±0.09	95.69 ±0.16

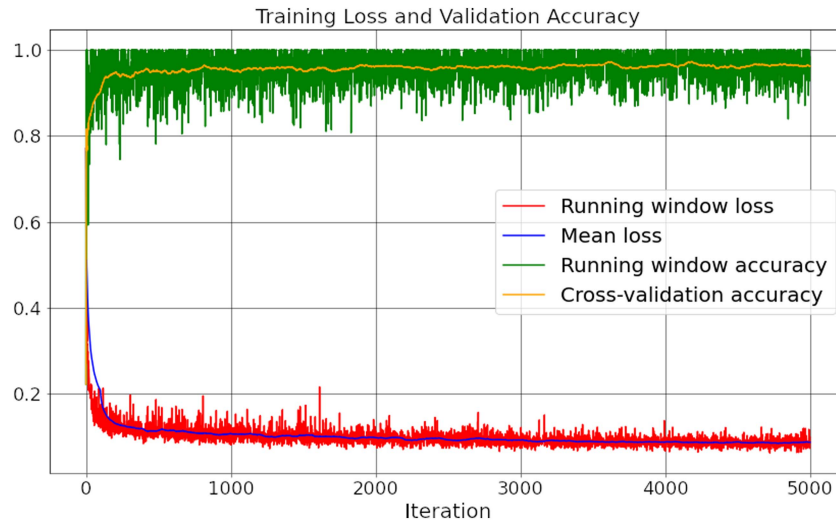Entities in bold indicate the best score.

Fig. 9. Training loss and cross-validation accuracy for training the FCN with the ResNet101 backbone over subset 1.

lines in blue represent the mean loss for each iteration. In Fig. 9, which is representative for the majority of the performed training approaches, one can notice that the cross-validation accuracy (orange line) is over 92%–94% and the mean training loss (blue line) is less than 0.15 even from the first 2500 iterations (15 epochs). However, moderate oscillation in the running window cross-validation accuracy and training loss is present in most of the multisubset training approaches.

The oscillation observed in the same figure is directly related to the complexity and the clarity of the caustics on the subsets used for training. To investigate the source of this oscillation, initially the base learning rate is altered. By the tests performed it is concluded that the reduction of the learning rate did not decreased the oscillation of the accuracy but on the contrary it increased it while the increase of the learning rate did not affect it. This implies that the oscillation of the running-window values is caused by the variance of the cross-validation data and especially the use of a random sample. Increasing the window size or reducing batch size will effect in reduction of the oscillation, however, this is not the case for the performed experiments. Regarding the accuracy of all the cross-validation performed every 100 iterations for all the validation data and the mean loss, no remarkable oscillation is noticed, proving the reliability of the models.

2) Testing the Models: To test the different models trained over unseen subsets of the data set, two metrics are used; *F1 score* which is a single metric that combines *recall* and *precision* using the harmonic mean and *accuracy* which is the ratio of the correctly labeled subjects to the whole pool of subjects. Tables II and III present the various metrics calculated after testing the several trained models over 30 epochs. It is obvious that for all the trained models, the achieved metrics for "noncaustics" pixel classification are quite high, indicating their great generalization. It is also evident that the models trained only on subsets 3 or 6 (cases 6 and 7) are achieving lower metrics compared to the rest, proving the necessity of the R-CAUSTIC data set as a whole. It is reminded that the main goal here was the very accurate detection of the nonaffected areas of the images, enabling feature detection and color values transferring only from those areas.

Overall, the results delivered by all the trained models are very satisfying, classifying with very high accuracy and reliability the pixels of the underwater images. Differences on the sensitivity of the models are expected, however this highlights the need to train a final model using all the available subsets. However, the slight prevalence of the models trained on the FCN-ResNet101 should be mentioned, while SegNet follows. The small variations observed in the metrics were expected due to the diverse characteristics of the subsets used for training and
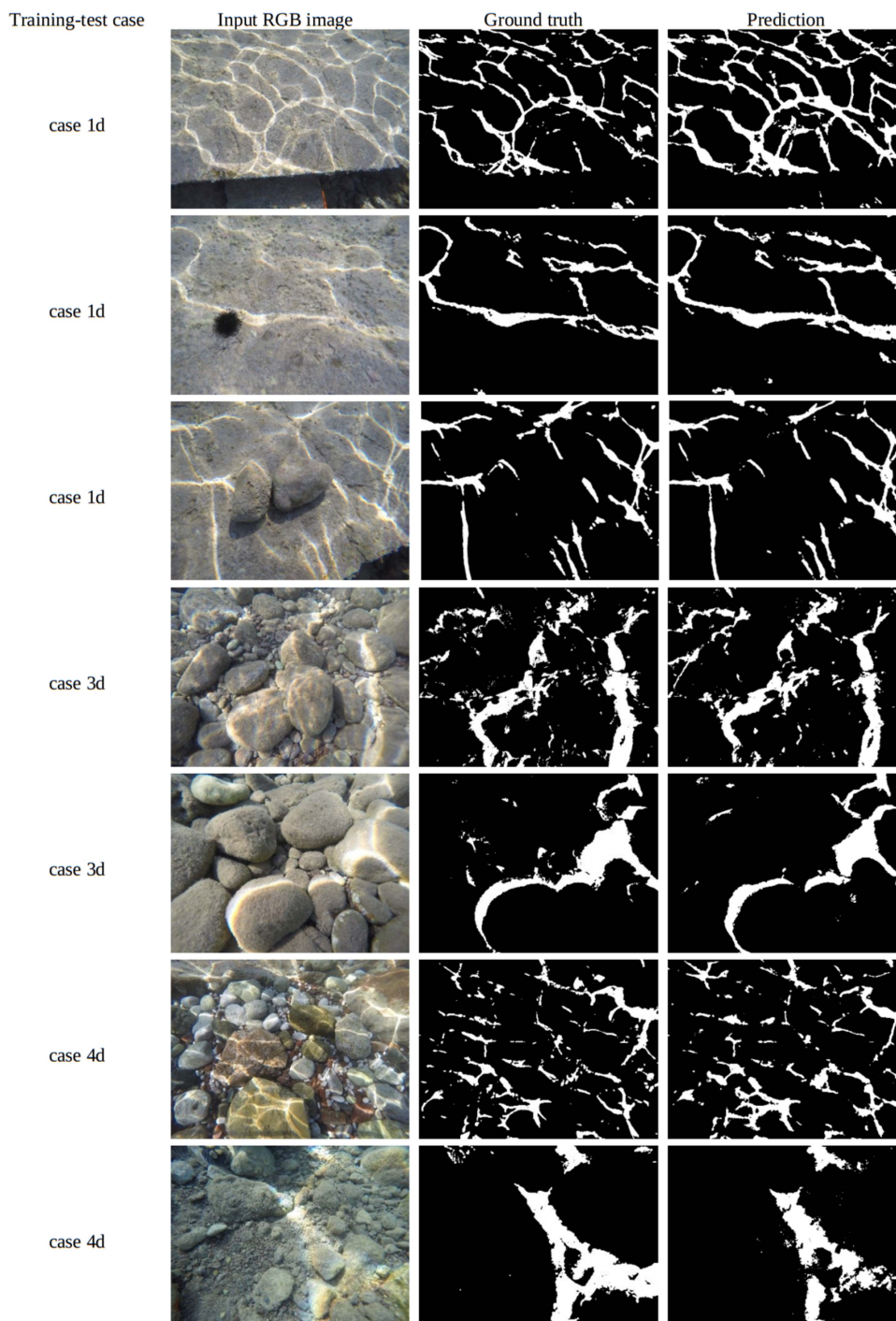
Fig. 10. Typical examples of the RGB images (second column), the ground truth binary images (third column), and the predicted binary images (fourth column) using the models trained in the respective cases (first column).

testing and the different FCN architectures used for training. It is also an indication of the performance of the trained models over different types of caustics. However, these variations are not affecting the overall accuracy of the models in a severe degree. Actually, the differences in the F1 score for caustics and "noncaustics" are mostly attributed to the indefinite boundary of caustics due to the chromatic aberration effect, and as such the differences between the predicted and the ground truth values, especially in subsets 3 and 6. This is also the reason for the

lower accuracy achieved in "caustics" class, since the chromatic aberration affects a larger percentage of this class, compared to the "noncaustics" class.

In Fig. 10, the input RGB images with the respective ground truth and SegNet's predictions are illustrated. For the first and the second row of the figure which is showing the predictions over an image from 1 and 2 subset, respectively, the binary image is predicted by the model trained on 3, 4, 5, 6, and 7 subsets.
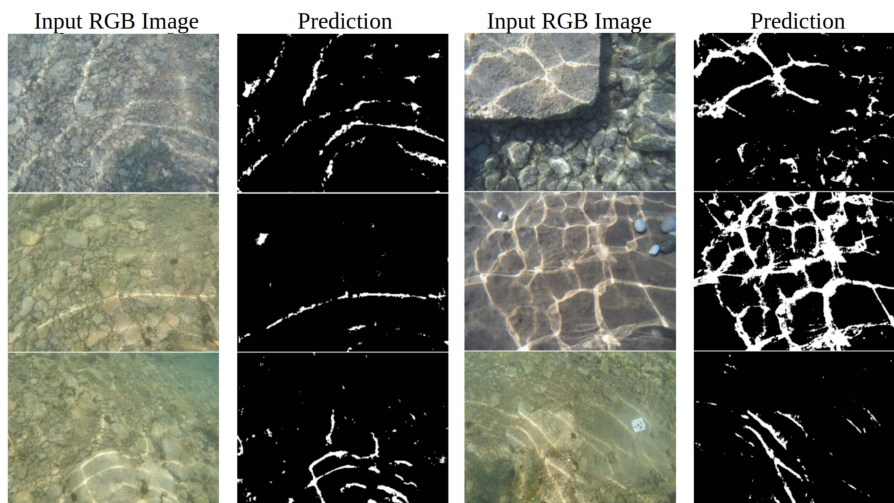
Fig. 11. Various real-world RGB images (first and third column) and the respective predictions using the model trained on all the available subsets (case 8 d) (second and fourth column).

Especially in the first row, it can be noticed that the pixels classified as "caustics" exceed those in the ground truth. These results come from the training of the model in more close-range imagery, giving it the ability to detect the intense chromatic aberration on the boundaries of the caustics. However, this cannot be achieved in the ground truth image. Following, the third and the fourth row present the predicted binary masks using the model trained over the 1, 2, 3, 4, and 6 subsets for images belonging to the fifth and the seventh subsets, respectively. Finally, the fifth and the sixth rows are illustrating the predicted binary masks using the model trained on the 1, 2, 3, 5, and 7 subsets for images belonging to the fourth and the sixth subsets, respectively.

In addition, an FCN-ResNet101 model was trained exploiting all the available subsets, to be used for classifying unseen underwater imagery. For training this model, the same parameters as before are used. Results over images captured for real-world underwater photogrammetric applications in shallow waters are presented in Fig. 11.

There, it is obvious that the trained model can generalize over different types of seabed and caustics with high reliability. A very important outcome is also that the model achieves to classify really bright areas of the image that are not caustics as "noncaustics" correctly. This can be seen on the last two images of Fig. 11 where the white stones on the seabed and a white ground control point (GCP) are classified as "noncaustics." This example with the GCP is of really high importance for the underwater photogrammetric applications.

C. Pixelwise Image Radiometric Correction

In this section, results regarding the pixelwise image correction are presented and evaluated. Fig. 12 presents typical examples of corrected images from the presented data set while Fig. 13 presents examples of real-world corrected imagery.

The first column depicts the original images with caustics, the second column depicts the binary images predicted by the trained FCN model and the third column demonstrates the corrected imagery. By observing the corrected imagery, it is clear that the proposed method achieves impressive visual results, since in the vast majority of the corrected images the rippling caustics are mitigated enough for certain image processing applications, like 3-D reconstruction, and the replaced pixels are not clearly obvious. However, in some cases like the one presented in the second row of Fig. 12, the corrected areas are obvious. This is a result of an inadequate color transferring approach, indicating great differences in luminosity between the overlapping images used for the correction. In fact, in some cases, these differences cannot be compensated by the color transferring only. However, as will be shown by the experiments presented in Section V-D, this is not affecting the 3-D reconstruction performance in a measurable degree. To solve this issue, Poisson blending [51] could be implemented on the boundary of the corrected with the uncorrected areas, however, this would affect the pixel values in an uncontrollable degree, affecting even more the SfM-MVS processes.

To demonstrate the importance of the color transferring step proposed in this methodology, in cases where no extreme differences between the source and the target images are apparent, a typical example is given in Fig. 14.

There, a corrected image created by the proposed methodology with the color transferring step is presented in (a) while the same image corrected by the proposed methodology without applying the color transferring step is presented in (b). It is obvious that in the case where the color transferring is not applied, the replaced pixels are obviously enough, since they are characterized by lower RGB values.

Although it is not clearly obvious in the scale at which the images are presented in Fig. 12, the proposed approach achieved really high accuracy in the pixel replacement process. When having a closer look to the corrected imagery, no offsets and other pixel displacements are obvious, especially in formations on the scene, being continuous between a corrected and a noncorrected area of the image.
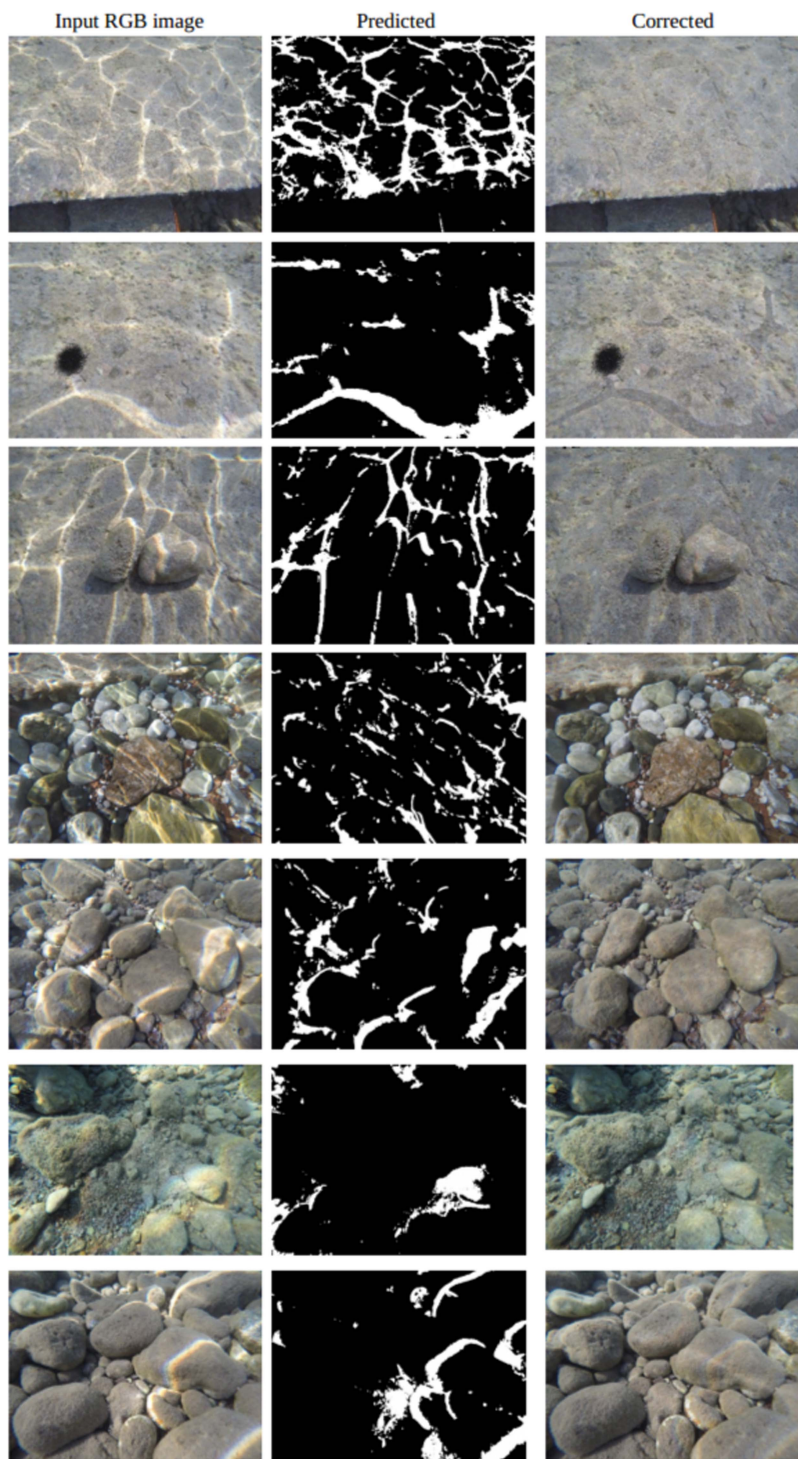
Fig. 12. Examples of corrected images for the seven subsets. The first column depicts the original images with caustics, the second column depicts the binary images predicted by the trained FCN model and the third column demonstrates the corrected imagery.

D. Improving 3-D Reconstruction

Following the pixelwise correction of the underwater imagery, the corrected data were processed with a commercial SfM-MVS software for evaluating the improvements on the 3-D reconstruction, which was the initial objective of the effort. To that direction, six different test cases were processed (see Figs. 15

and 16). Typical images for these six test cases can be seen in Fig. 12.

The first three test cases are using images of the already presented benchmark data set; the first test case is using tri-stereo imagery from subset 3, the second testcase is using stereo imagery from subset 1 while for the third test case multiple-view stereo imagery (seven consecutive images) from subset 2 was
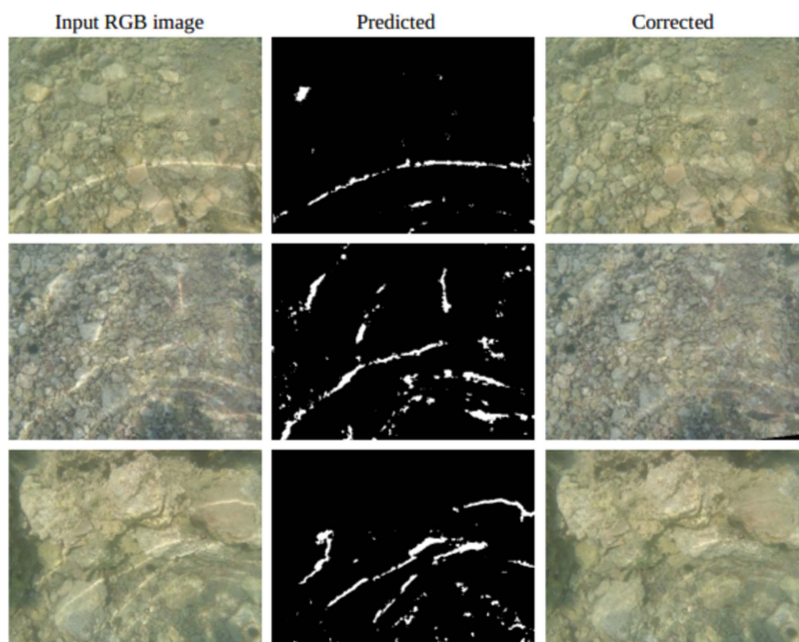
Fig. 13. Examples of corrected images for real-world imagery. The first column depicts the original images with caustics, the second column depicts the binary images predicted by the trained FCN model, and the third column demonstrates the corrected imagery.
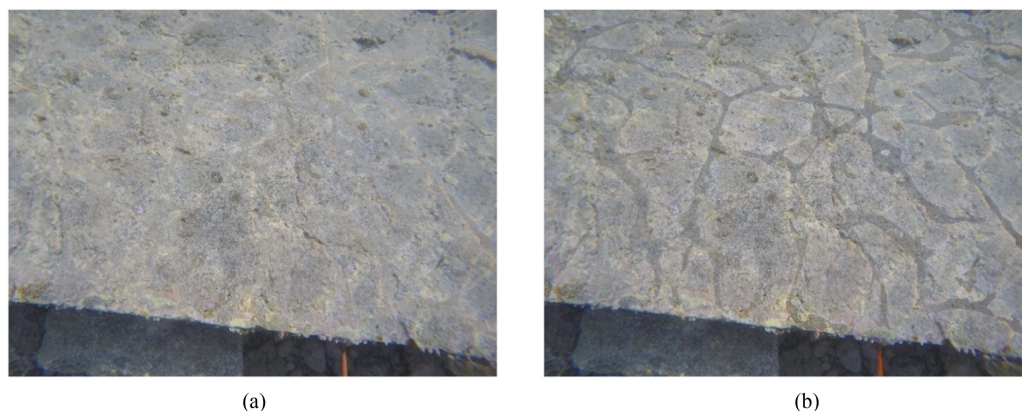

Fig. 14. Corrected image by the proposed methodology with the color transferring step (a) and the same image corrected by the proposed methodology without applying the color transferring step (b).

used. These three test cases will facilitate the evaluation of the proposed correction methodology in terms of 3-D reconstruction improvement, compared also to the generated 3-D point clouds using the already available reference imagery. The 3-D reconstruction results are presented in Fig. 15. Results of the rest of the four subsets presented in Section III are not presented here since due to the complexity of the scene and the poor texture of the smooth and glossy rocks, the improvements are not easy to be highlighted, since the 3-D point clouds of the reference imagery are incomplete too. As such, the subsets presented in Fig. 15 are those that are generating the most complete 3-D reconstructions, facilitating a detailed and direct comparison of the results.

By comparing the 3-D point clouds generated using the reference imagery (left column of Fig. 15) to the 3-D point clouds

generated using the uncorrected and the corrected imagery, significant differences can be noticed. It is clear that caustics are preventing a proper 3-D reconstruction of the scene, since when images affected by them are used, the resulting 3-D point clouds are incomplete for all the tested cases. More specifically, dense image matching (DIM) algorithms are failing to match the corresponding points of the affected areas and only the unaffected areas are appearing in the 3-D point cloud. It is also important to highlight that as reported also in Table IV, when using the uncorrected imagery, not all of the images were aligned for the test case 1 and test case 3. In the latter, this resulted in a much less covered area by the 3-D point cloud.

Coming to the third column, it is obvious that when the imagery corrected by the proposed methodology is used, the
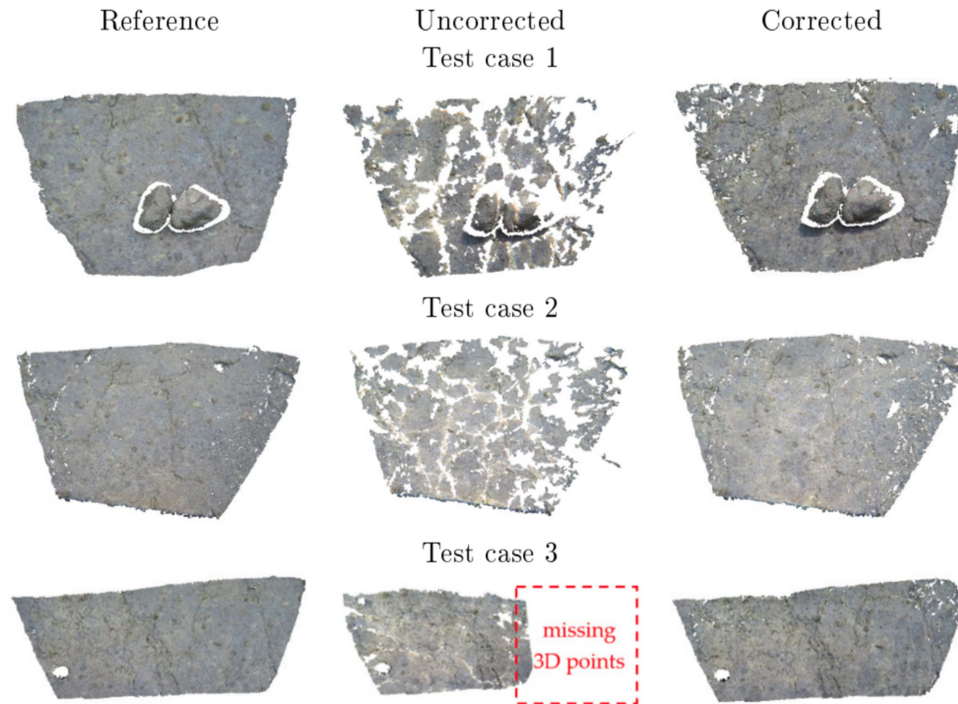
Fig. 15. Three-dimensional reconstructions using the reference imagery (left column), the uncorrected imagery (middle column) and the corrected imagery by the proposed methodology (right column). Each row refers to a different test case.

TABLE IV
EVALUATION METRICS BETWEEN THE UNCORRECTED AND THE CORRECTED IMAGES

Test case	Images aligned/ total	Caustics in pixels [%]	Matched points	Matched points change [%]	Dense points	Dense points change [%]	D	D change [%]	R
1 uncorrected	2/3	28.1	1327		37505		7764.99		0.003
1 corrected	3/3		2210	+66.5	72261	+92.7	8140.25	+4.87	0.003
2 uncorrected	2/2	15	606		1783027		106908.18		0.001
2 corrected	2/2		820	+35.3	2118934	+18.8	106718.97	-0.17	0.001
3 uncorrected	5/7	15.4	2562		556926		12364.76		0.001
3 corrected	7/7		4173	+62.9	6137345	+1102	78761.6	+637	0.001
4 uncorrected	2/2	8	3920		188203		1590.53		0.007
4 corrected	2/2		4916	+25.4	190442	+1.19	1620.51	+1.88	0.007
5 uncorrected	2/2	2.7	4315		144435		3146.87		0.004
5 corrected	2/2		5257	+21.8	147948	+2.43	3135.25	-0.37	0.004
6 uncorrected	2/2	4.5	2433		253778		3996.95		0.003
6 corrected	2/2		3075	+26.4	259839	+2.39	4010.57	+0.34	0.003

D is the surface density and R is the roughness. Numbers for these two metrics are relative.

completeness of the 3-D point cloud is improved to a great degree, delivering point clouds very similar to the ones generated by the reference imagery. Some insignificant differences can be observed on the perimeter of the reconstructed area. These differences are not attributed to some defect of the corrected imagery but on the extremely intense chromatic aberration that is apparent at the areas of the image having large radial distance when the luminosity of the scene is increased. This effect is not that intense on the reference images since they are characterized by lower illumination. In these cases, chromatic aberration is

Uncorrected   Corrected

Test case 4
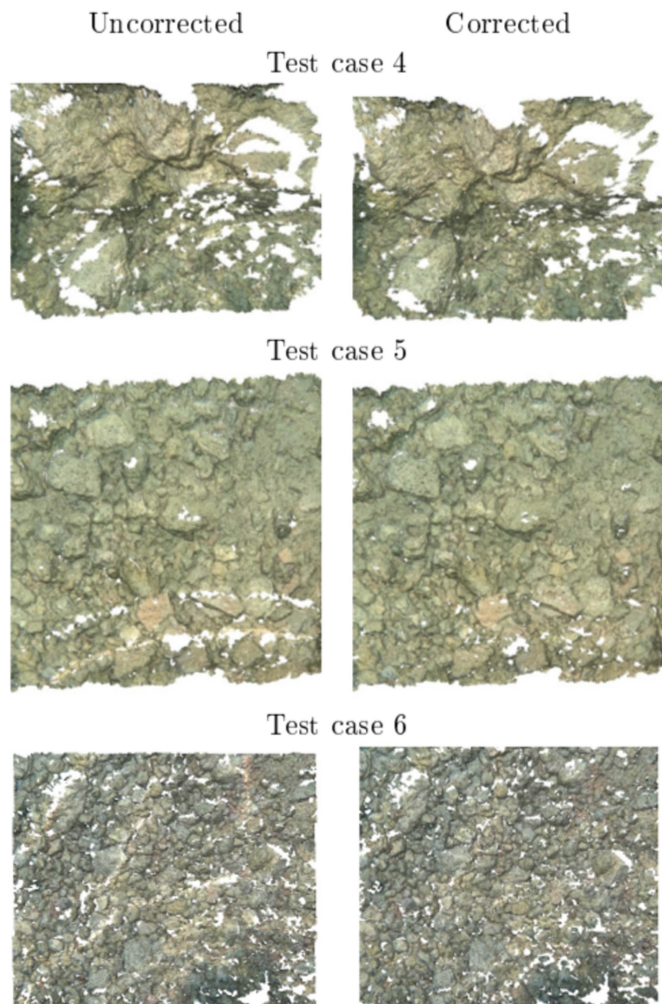
Test case 5

Test case 6

Fig. 16. Resulting 3-D point clouds for the rest three tests performed over real-world imagery. In the left column the 3-D point cloud generated using the original uncorrected imagery is presented while in the right column, the respective 3-D point cloud of the corrected imagery is presented.

caused by the lens of the camera, as a result of the different refractive indices of the lens for each wavelength of light.

In Fig. 16, three different tests performed over imagery used for real underwater 3-D reconstruction tasks are presented. Again, as in the first three test cases, when using real world imagery, the imagery resulted by the proposed methodology achieved to deliver a more complete 3-D point cloud in the areas covered by caustics.

To facilitate a deeper evaluation of the improvements on the 3-D reconstruction, all the 3-D point clouds presented above were imported into Cloud Compare freeware [52] for further investigation. In particular, the following parameters and statistics that are widely used also in the literature for evaluating 3-D point clouds [8], [53], [54], were computed for each point cloud.

Total number of sparse and dense points: All the 3-D points of the point cloud were considered for this metric, including any outliers and noise [52]. For the purposes of the work presented here, the total number of 3-D points reveals the effect of the correction methodology on the matchable pixels among the

images. The more corresponding pixels are found in the DIM step on the images, the more points are generated. Higher values of total number of points are considered better in these cases; however, this should be crosschecked with the surface density metric, since it might be an indication of noise on the point cloud.

Surface density: The density is estimated by counting the number of neighbors N (inside a sphere of radius R) for each point [52]. The surface density used for this evaluation is defined as $N/(Pi \times R^2)$, i.e., the number of neighbors divided by the neighborhood surface. Cloud Compare [52] estimates the surface density for all the points of the cloud and then it calculates the average value for an area of 1m2 in a proportional way. Surface density is considered to be a positive metric, since it defines the number of the points on a potential generated surface, excluding the noise being present as points out of this surface. This is also the reason of using the surface density metric instead of the volume density metric.

Roughness: For each point, the "roughness" value is equal to the distance between this point and the best fitting plane computed on its nearest neighbor [52], which are the points within a sphere centred on the point. Roughness is considered to be a negative metric since it is an indication of noise on the point cloud, assuming an overall smooth surface.

Table IV presents the above metrics, together with the ratio of the aligned images by the total images, the average percentage of the pixels of the images containing caustics based on the predicted binary images, the number of the matched points (the number of the points of the sparse point cloud) and the respective percentage of the change achieved between the uncorrected and the corrected imagery, the number of the dense points and the respective percentage of change, the density D and the respective percentage of change, and finally the roughness R.

By observing the metrics presented in Table IV, a first important outcome is that when the corrected imagery is used, more images are aligned in the image alignment step. Moreover, it can be noticed that for all the performed tests, when the corrected imagery is used, more points are matched, delivering a denser sparse point cloud and a more robust 3-D geometry of the scene. It is noted that in [1], a state-of-the-art method increased matched points only by 2%–3%, ten times less than the smallest increase achieved here. These first two outcomes were expected, and the background is already reported in Section II-A1. The increase of the matched points, is proportional to the number of the images covered with caustics. However, there is not a strict relation between those two. This is not satisfied for test case 1 and test case 3, since not the same number of images are aligned. An increase in the number of the dense points is also observed. However, this is not of the same magnitude as that for the matched points. Again, here test case 1 and test case 3 are excluded for the same reasons as before. Coming now to the density of the point clouds and their roughness, no significant differences can be observed.

Regarding the density, this was expected since the unaffected areas of the images are remaining the same, so there is no reason for generating more 3-D points there. However, the fact that the roughness of the point clouds is remaining the same is of great importance. In the literature [54], it is reported that most

of the underwater image enhancement methods are increasing the roughness of the generated 3-D point clouds. Nevertheless, this does not apply for the proposed methodology, highlighting the accuracy and the quality of the performed pixelwise image correction.

VI. Conclusion

In this article, a novel method for correcting the radiometric effects of caustics on the shallow underwater imagery was presented. In addition, R-CAUSTIC, the first real-world data set on underwater rippling caustics was delivered and documented.

The proposed method first relies on state-of-the-art deep learning tools which can accurately classify the pixels of the image as "noncaustics" or "caustics" and then exploits the 3-D geometry of the scene to achieve a pixelwise correction, by transferring color values between the overlapping images. The method depends on the good matches among the images, since an accurate fundamental matrix calculation is a prerequisite. However, if no good matches can be achieved, even when the detected masks are exploited in the key point detection step, images are not even appropriate for image-based 3-D reconstruction, which this method intends to improve.

Results suggest that the tested FCNs architectures achieve very high accuracy in this binary classification problem even from the first 15–20 epochs. Differences on the sensitivity of the models were also expected, however, this highlighted the need to train a final model using all the available subsets. Pixel classification results over images captured for real world underwater photogrammetric applications in shallow waters suggested that the trained models can generalize over different types of seabed and caustics with high reliability. Caustics correction experiments performed illustrated the robustness and the reliability of the method over different types of seabed, different types of caustics and different anaglyph of the scene. The need of the color transferring step that is proposed was also highlighted.

Concerning the improvements on the 3-D reconstruction of the scenes, the effectiveness of the proposed method was clearly obvious, since complete 3-D point clouds were delivered, leaving no doubts about the achieved results. It was proven that when the corrected imagery is used for performing 3-D reconstruction tasks, more images are aligned, and more points are matched. This delivers a more robust, complete, and reliable 3-D reconstruction. Moreover, it was also considered very important that the proposed method did not increase the roughness of the generated dense point clouds for all the testing cases. The proposed method will enable the users to capture less images and deliver more complete results, covering also larger areas. This will reduce the SfM-MVS processing time and the revisit of the underwater site for extra data will be avoided.

By using the proposed method, a more chromatically consistent and realistic representation of the seabed is achieved. From the experiments performed, some remaining artifacts noticed on the images, especially in the areas having large radial distance, are attributed to the really intense chromatic aberration effect. However, this was not to be dealt with within the context of this work. It was found that this effect is negatively affecting mainly the SfM-MVS process, compared with the skipping of the color transferring step. To overcome these issues, a channel-based correction of the refraction effect on the water-lens-air interface has to be performed. However, it can be easily avoided by using dome ports instead of flat ports and by avoiding fish-eye lenses and generally very small focal lengths. Although this approach was developed intending to correct the radiometrically affected areas on the underwater imagery, it can also be exploited to correct overwater, aerial, and satellite imagery for specularities, shadows, and occlusions caused by illumination conditions, objects, or even clouds.

References

[1] P. Agrafiotis, D. Skarlatos, T. Forbes, C. Poullis, M. Skamantzari, and A. Georgopoulos, "Underwater photogrammetry in very shallow waters: Main challenges and caustics effect removal," *ISPRS - Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XLII-2, pp. 15–22, 2018. [Online]. Available: https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2/15/2018/

[2] F. Menna, P. Agrafiotis, and A. Georgopoulos, "State of the art and applications in archaeological underwater 3 D recording and mapping," *J. Cultural Heritage*, vol. 33, pp. 231–248, 2018.

[3] P. Agrafiotis, K. Karantzalos, A. Georgopoulos, and D. Skarlatos, "Learning from synthetic data: Enhancing refraction correction accuracy for airborne image-based bathymetric mapping of shallow coastal waters," *PFG–J. Photogrammetry, Remote Sens. Geoinformation Sci.*, vol. 89, no. 2, pp. 91–109, 2021.

[4] P. Agrafiotis, K. Karantzalos, A. Georgopoulos, and D. Skarlatos, "Correcting image refraction: Towards accurate aerial image-based bathymetry mapping in shallow waters," *Remote Sens.*, vol. 12, no. 2, 2020, Art. no. 322.

[5] P. Agrafiotis, D. Skarlatos, A. Georgopoulos, and K. Karantzalos, "Depthlearn: Learning to correct the refraction on point clouds derived from aerial imagery for accurate dense shallow water bathymetry based on SVMs-fusion with LiDAR point clouds," *Remote Sens.*, vol. 11, no. 19, 2019, Art. no. 2225.

[6] P. Agrafiotis, "Image-based bathymetry mapping for shallow waters," Ph.D. dissertation, School of Rural, Surveying and Geoinformatics Engineering, Nat. Tech. Univ. Athens, Athens, Greece, 2020.

[7] D. Skarlatos, F. Menna, E. Nocerino, and P. Agrafiotis, "Precision potential of underwater networks for archaeological excavation through trilateration and photogrammetry," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 42, no. 2/W10, pp. 175–180, 2019.

[8] P. Agrafiotis, G. I. Drakonakis, D. Skarlatos, and A. Georgopoulos, "Underwater image enhancement before three-dimensional (3 D) reconstruction and orthoimage production steps: Is it worth," in *Latest Developments in Reality-Based 3D Surveying and Modelling*, F. Remondino, A. Georgopoulos, D. González-Aguilera, and P. Agrafiotis, Eds., MDPI: Basel, Switzerland, 2018.

[9] E. Trabes and M. A. Jordan, "A node-based method for slam navigation in self-similar underwater environments: A case study," *Robotics*, vol. 6, no. 4, 2017, Art. no. 29.

[10] T. Forbes, M. Goldsmith, S. Mudur, and C. Poullis, "Deepcaustics: Classification and removal of caustics from underwater imagery," *IEEE J. Ocean. Eng.*, vol. 44, no. 3, pp. 728–738, Jul. 2019.

[11] D. K. Lynch, W. C. Livingston, and W. Livingston, *Color and Light in Nature*. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[12] Wikipedia, "Caustic (optics)." Accessed: Apr. 2023.. [Online]. Available: https://en.wikipedia.org/wiki/Caustic_(optics)

[13] E. Trabes and M. A. Jordán, "Employing feedback to filter caustic waves in underwater scenes in motion," in *Proc. Argentine Symp. Artif. Intell.*, 2015, pp. 240–247.

[14] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE 7th Int. Conf. Comput. Vis.*, 1999, pp. 1150–1157.

[15] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[16] 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 *Proc. IEEE OCEANS Conf.*, 2008, pp. 1–7.

[17] N. Gracias and J. Santos-Victor, "Underwater video mosaics as visual navigation maps," *Comput. Vis. Image Understanding*, vol. 79, no. 1, pp. 66–91, 2000.

[18] A. Shihavuddin, N. Gracias, and R. Garcia, "Online sunflicker removal using dynamic texture prediction," in *Proc. VISAPP*, 2012, pp. 161–167.

[19] N. Joshi, M. Cohen, and S. Rainier, "Lucky imaging for multi-image denoising, sharpening, and haze removal, computational photography (ICCP)," in *Proc. IEEE Int. Conf.*, Cambridge, MA, USA, 2010, pp. 1–8.

[20] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[21] R. Fattal, "Single image dehazing," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–9, 2008.

[22] Y. Y. Schechner and N. Karpel, "Attenuating natural flicker patterns," in *Proc. IEEE/MTS Oceans Techno-Ocean*, 2004, pp. 1262–1268.

[23] A. Tripathy, M. Singh, and H. Roy, "Clustering based approach for underwater sunlight flicker removal," in *Proc. Glob. Oceans: Singap.–US Gulf Coast*, 2020, pp. 1–7.

[24] Y. Swirski and Y. Y. Schechner, "3deflicker from motion," in *Proc. IEEE Int. Conf. Comput. Photogr.*, 2013, pp. 1–9.

[25] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986, doi: 10.1109/TPAMI.1986.4767851.

[26] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[27] R. I. Hartley, "An algorithm for self calibration from several views," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1994, pp. 908–912.

[28] D. Skarlatos and P. Agrafiotis, "Image-based underwater 3 D reconstruction for cultural heritage: From image collection to 3 d. critical steps and considerations," in *Visual Computing for Cultural Heritage*. Berlin, Germany: Springer, 2020, pp. 141–158.

[29] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4104–4113.

[30] T. Sattler, C. Sweeney, and M. Pollefeys, "On sampling focal length values to solve the absolute pose problem," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 828–843.

[31] OpenCV, "Camera calibration and 3 D reconstruction." Accessed: Apr. 2023. [Online]. Available: https://docs.opencv.org/4.x/d9/db7/tutorial_py_table_of_contents_calib3d.html

[32] J.-Y. Bouguet, "Camera calibration toolbox for Matlab." Accessed: Apr. 2023. [Online]. Available: http://robots.stanford.edu/cs223b04/JeanYvesCalib/htmls/links.html

[33] N. Audebert, B. Le Saux, and S. Lefèvre, "Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks," *ISPRS J. Photogrammetry Remote Sens.*, vol. 140, pp. 20–32, 2018.

[34] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2015, pp. 234–241.

[36] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[37] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[38] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 34–41, Jul./Aug. 2001.

[39] D. L. Ruderman, T. W. Cronin, and C.-C. Chiao, "Statistics of cone responses to natural images: Implications for visual coding," *JOSA A*, vol. 15, no. 8, pp. 2036–2045, 1998.

[40] S. Leutenegger, M. Chli, and R. Y. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 2548–2555.

[41] C. Loop and Z. Zhang, "Computing rectifying homographies for stereo vision," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 1999, pp. 125–131.

[42] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

[43] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 807–814.

[44] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Proc. Eur. Conf. Comput. Learn. Theory*, 1995, pp. 23–37.

[45] T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class adaboost," *Statist. Interface*, vol. 2, no. 3, pp. 349–360, 2009.

[46] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 3, pp. 660–674, May/Jun. 1991.

[47] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Amer. Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[48] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[49] B. Schölkopf et al., *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT press, 2002.

[50] F. Pedregosa et al., "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[51] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," in *Proc. ACM SIGGRAPH Papers*, 2003, pp. 313–318.

[52] D. Girardeau-Montaut, "Cloudcompare (version 2.11 alpha) [GPL software]." Accessed: Apr. 2023. [Online]. Available: http://www.cloudcompare.org/

[53] P. Agrafiotis, G. Drakonakis, A. Georgopoulos, and D. Skarlatos, "The effect of underwater imagery radiometry on 3D reconstruction and orthoimagery," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 42, no. 2/W3, pp. 25–31, 2017.

[54] M. Mangeruga, F. Bruno, M. Cozza, P. Agrafiotis, and D. Skarlatos, "Guidelines for underwater image enhancement based on benchmarking of different methods," *Remote Sens.*, vol. 10, no. 10, 2018, Art. no. 1652.

Panagiotis Agrafiotis received the M.Eng. diploma in rural and surveying engineering in 2012, the M.Sc. degree in geoinformatics in 2016, and the doctorate degree in engineering from the National Technical University of Athens (NTUA), Athens, Greece, in 2020.

During the last nine years, he lived in Germany, Greece, and Cyprus, working in more than 17 research projects as a Researcher with the Technical University of Berlin, the Photogrammetric Vision Lab. of CUT, the Lab. of Photogrammetry, NTUA, laboratories with which he still closely collaborates, the ATHENA Research and Innovation Center and the Institute of Communication and Computer Systems (ICCS) of the School of Electrical and Computer Engineering (ECE) of the NTUA, with the responsibility of applied computer vision algorithms design, overwater, and underwater 3-D reconstruction, machine learning, laser scanning, and person detection and tracking for survivor localization on optical and thermal imagery. He is currently a Postdoctoral Researcher doing research in the areas of 3-D computer vision, remote sensing and machine learning, focusing on image-based bathymetric mapping and seabed image analysis. He has authored 36 scientific publications. His current research interests include image-based bathymetry mapping for shallow waters using machine learning, image processing, and image analysis.

Dr. Agrafiotis has participated in the organizing committee of three international conferences and in the scientific committee of seven more. During his studies, he earned various awards and travel grants for conferences and a full Ph.D. Scholarship from NTUA. In December 2020, his team 3[Deep]Vision, won the 1st Prize of the integrated Geomatics solution awarded by the European GNSS Agency for our proposal "Bathymetry from UAV Imagery and Machine Learning." He has served as a reviewer in 24 journals of high impact factor.

Konstantinos Karantzalos (Senior Member, IEEE) received the engineering diploma in surveying engineering from the National Technical University of Athens (NTUA), Athens, Greece, in 2000, and the Ph.D. degree in remote sensing from SRSE, NTUA, in collaboration with Ecole Nationale de Ponts et Chaussees (CERTIS/Imagine, ENPC, France), in 2007.

In 2007, he joined the Department of Applied Mathematics, Ecole Centrale de Paris, Gif-sur-Yvette, France, as a Postdoctoral Researcher. He is currently a Professor with the NTUA, joining the Remote Sensing Laboratory (RSLab) at the School of Rural, Surveying and Geo-Informatics Engineering. He is also occasionally affiliated with the Institute of Communications and Computer Systems (ICCS) and the Athena Research Center. He has numerous publications in international journals and conferences. He has 19 years of research experience, while being involved with more than 28 EU and national excellence/competitive research projects as a Principal Investigator and as a Researcher toward the design, development, and validation of state-of-the-art methodologies and cutting-edge technology in geomatics, earth observation, and computer vision. His teaching and research interests include earth observation and remote sensing, geospatial big data and analytics, computer vision and machine learning, environmental monitoring, and precision agriculture.

Dr. Karantzalos was the recipient of a number of awards and honors for his research contributions. He also serves as a Scientific/Program Committee Member at international conferences and as a reviewer in top-rank international journals.

Andreas Georgopoulos received the diploma of surveying engineering from the National Technical University of Athens (NTUA), Athens, Greece in 1976, and the M.Sc. and Ph.D. degrees in photogrammetry from the University College London, London, U.K., in 1977 and 1981, respectively.

Since 1980 he has been teaching photogrammetry and documentation of monuments at an undergraduate and postgraduate level with UCL, NTUA and as a Visiting Professor with KULeuven (RLICC - ArchDOC), Leuven, Belgium, Department of Civil and Geomatics Engineering, Cyprus University of Technology (CUT), Limassol, Cyprus, and the University of the Aegean, Mytilene, Greece. Since 1985, he has been leading and participating in numerous research projects with the Lab oratory of Photogrammetry, NTUA, concerning digital photogrammetry and monument documentation. He is currently an Em. Professor of Photogrammetry with the School of Rural, Surveying and Geoinformatics Engineering, NTUA. He has published approximately 270 scientific papers in international journals and conference proceedings. His research interests include 3-D modeling of cultural heritage, photogrammetric automation, and digital contemporary techniques.

Dr. Georgopoulos has served as the Vice-Head (1998–2002) and the Head (2002–2006) of the School of Rural and Surveying Engineering, and a Member of the Research Committee of NTUA (1999–2020). Since 2006, he has been a Member of the Executive Board of the ISC of ICOMOSCIPA-Heritage Documentation and has served as the Secretary General (2010–2014), President (2015–2020), while currently he is the Vice President. He is also a Board Member of ICOMOS.