Marine Snow Detection for Real Time Feature Detection

Alexandre Cardaillac* and Martin Ludvigsen*†‡

*Department of Marine Technology, Norwegian University of Science and Technology, Trondheim, Norway
†Department of Arctic Technology, University Centre in Svalbard, Longyearbyen, Norway
‡ Department for Arctic and Marine Biology, The Arctic University of Norway, Tromsø, Norway
{alexandre.cardaillac, martin.ludvigsen}@ntnu.no

***Abstract*—Underwater images are often degraded due to backscatter, light attenuation and light artifacts. One important aspect of it is marine snow, which are particles of varying shape and size. Computer vision technologies can be strongly affected by them and may therefore provide incorrect and biased results. In robotic applications, there is limited computational power for online processing. A method for real time marine snow detection is proposed in this paper based on a multi-step process of spatial-temporal data. The RGB colored images are converted to the YCbCr color space before they are decomposed to isolate the high frequency information using a guided filter for a first selection of candidates. Convolution with an uniform kernel is then applied for further analysis of the candidates. The method is demonstrated in two use cases, underwater feature detection and image enhancement.**

***Index Terms*—Marine Snow Detection, Image Processing, Real Time Processing**

I. Introduction

Modern underwater drones operations often require visual feedback to complete the missions. However, in the marine environment, the captured images are often degraded. It can be caused by three main reasons [1]–[3]:

- Light absorption: provokes in water a reduction of the light energy and therefore light attenuation and color distortions.
- Scattering: caused by light reflection, it largely contributes to the degradation of the image. Backward scattering results in haziness and low contrast whereas forward scattering in blurriness.
- Marine snow: Biological debrief and other macroscopic floating particles that increase the previously mentioned effects.

The third degradation problem, marine snow, it is often considered as a source of noise and bias in computer vision tasks such as scene perception, understanding or mapping. Marine snow can be of variable size and shape and quantity. In Figure 1, examples of images containing marine snow are shown. It is particularly important to filter it out when considering computer vision methods such as feature detection and matching.

This work was supported by the BugWright2 EU H2020-Project under the Grant agreement No. 871260.

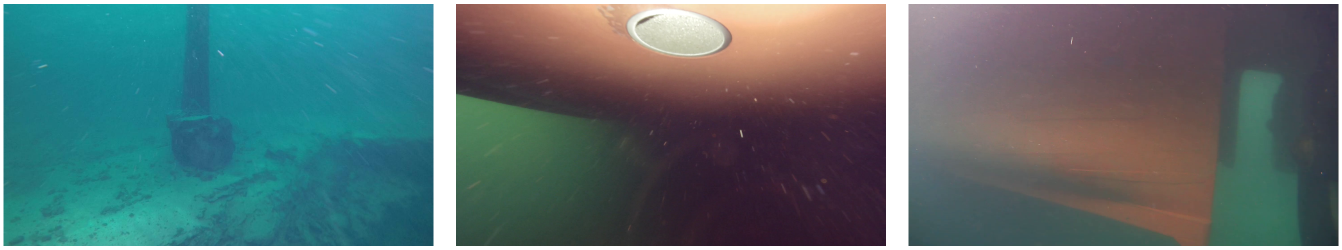
When the scene needs to be perceived as it is, with the level of accuracy provided by the available sensors, only detection of the marine snow should be done. Removing marine snow may cause the loss of image information as pixels are replaced in the process. For complex scenes this may cause reconstruction errors, and ultimately lead to observation and navigation biases for missions.

Hence this paper describes a method focusing on detection instead of removal, although it can be done easily with the addition of median filtering as presented as an example at the end of the paper.

In [4], a adaptive probabilistic approach is proposed by considering the probability of marine snow existence based on pixel intensities in local patches. In [5], in a supervised manner, the marine snow is detected based on pixel dissimilarity with its their neighbors - in patches. The median filtering concept is then used to filter out the candidates. However, the two mentioned methods perform on single image and are not able to detect all the marine snow, especially those of bigger size. Methods using a sequence of images have proven to be more reliable. In [6], a three steps approach is proposed. A Gaussian background modeling algorithm is first employed to extract the scene information of a video. Then, marine snow detection based on [5] is performed in each frame. Finally, an inpainting algorithm using prior data is used to restore the scene behind the detected particles. In [7], a real time method is proposed to remove marine snow from video sequences. It is based on spatiotemporal patch analysis and 3D median filtering.

More recently, methods using deep learning approach have been proposed [8]–[11]. Despite the fact that these methods perform well, they are computationally expensive and can have difficulties to generalize as their performance is dependant on the data they trained on. Therefore, in dynamic scenes with a moving camera and complex structures, they can fail to reliably detect the marine snow and remove it.

In this paper, a real time method for marine snow detection is proposed and applied to detect potential outliers in applications such as feature detection. It consists of multiple steps: first, a sequence of three images centered in the image that needs to be processed is selected. A domain transformation of these images is performed, from RGB to YCbCr color space to



(a) (b) (c)

Fig. 1: Examples of images containing marine snow in different situations. (a, taken from [7], includes marine snow with relatively few background features. (b) and (c), taken from an underwater inspection of a ship hull. In these cases, marine snow can bias automatic inspection pipelines based on vision.

reduce the resolution by keeping only the Y component which can be considered as a representation of luminance of the image and highlight potential relevant data in the images. Since marine snow particles are predominantly seen as bright spots in the images, this will help to isolate them. An optional step is then proposed to pre-select candidates based on a guided filter to retain the high frequency information of the images. To decide on the presence of marine snow, a convolution with uniform kernel is performed on the previous and next frames and then compared to the current frame and if conditions are met, the concerned pixels are marked as marine snow. A final step for result refinement is applied, it consists of mathematical morphology operations to modify or remove potential noise and unwanted results such as malformed marked regions. All the steps present in the algorithm are represented in Figure 2.

II. Proposed Method

The entirety of the algorithm uses a sequence of three images defined as $\{I_{n-1}, I_n, I_{n+1}\}$ with I_n, the n^{th} frame in the video sequence that will be processed for marine snow detection based on the previous frame in time I_{n-1} and the following one I_{n+1}. It is important to note that the proposed methods works best when the captured frames are close in time.

A. Image domain conversion

Transforming the image color domain can have multiple benefits: reducing the resolution of the image, which can greatly impact the processing time, and also highlight information that would not have been possible to observe with the original image. For both of these reasons, the color space of the selected images are converted from RGB to YCbCr, which represents color as brightness together with two color differences, also sometimes referred as luminance. In this paper, the JPEG modified version of the Rec. 601 YCbCr is employed[1]. It supports the components Y, CB and CR with full 8-bit range of $[0...255]$, they are defined as follows:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.1687 & -0.3313 & 0.5 \\ 0.5 & -0.4187 & -0.0813 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \quad (1)$$

with Y the luma component and Cb and Cr, the blue-difference and red-difference chroma components. All the operations to detect marine snow are performed on the Y component, so the other two can be discarded.

B. Guided filter decomposition for candidate selection

The guided filter [12] has recently found application in marine snow detection [10]. Originally used for filtering applications like dehazing, it was then used for rain detection and removal [13], [14]. The filter can be used as a low-pass filter to decompose the image I into a low frequency layer I_{low} and a high frequency layer I_{high} such that:

$$I = I_{low} + I_{high} \quad (2)$$

For speed performance purpose, the fast guided filter [15] is employed. By using the subsampling principles, the computation complexity is reduced from $O(N)$ to $O(N/s^2)$ for a subsampling ratio s.

It is applied to each selected frame with exaggerated parameters which results in isolation of pixels associated to marine snow. However some of the pixels part of the background are still included. To filter them out, the time parameter is included

[1] JPEG File Interchange Format: https://www.w3.org/Graphics/JPEG/jfif3.pdf

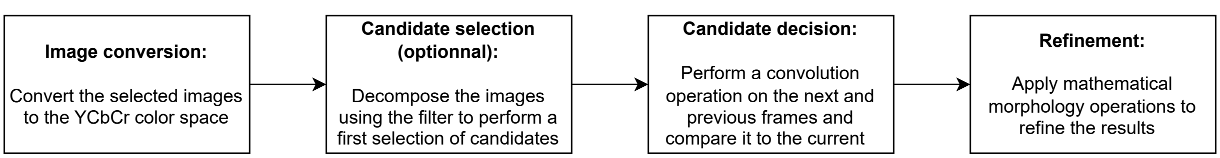


Fig. 2: The four-steps process to detect marine snow. Note the second step is optional because marine snow can be detected without it but necessary to achieve better results.

using the previous and next frames in time. By performing spatial comparison over time it is possible to differentiate the background from the marine snow. To do so, the pixel wise minimum value is extracted from the previous and next images and then subtracted to the current one:

$$P = I_{high,n} - \min(I_{high,n-1}, I_{high,n+1}) \tag{3}$$

With P the resulting image that can be considered as a potential map in which high pixels values are likely to be marine snow. A threshold value λ_p must be chosen to do the selection of potential candidates. It is based on the following condition:

$$P_{f,n} = \exists p \in P : p > \lambda_p \tag{4}$$

With P_f, the filtered set containing all the existing pixels p in P that satisfy the condition. This step is optional as it is not required to find marine snow but it can improve the results by providing additional inputs on the likelihood that a pixel belongs to marine snow in the image. When time performance is very important or the most robust results are not needed, this step can be skip.

C. Convolution operations for candidate decision

Using again the Y component of each selected frame, operations based on convolutions are defined to make a decision on the presence of marine snow. To achieve this, convolution with an uniform kernel is first applied to the previous and next frames, Y_{n-1} and Y_{n+1}. This corresponds to a moving average filter. Considering a two dimensional square sliding window of size $2k+1$, therefore containing $(2k+1)^2$ pixels, it can be mathematically defined as:

$$Y_f(y,x) = \frac{1}{(2k+1)^2} \sum_{i=-k}^{k} \sum_{j=-k}^{k} Y(y+i, x+j) \tag{5}$$

With Y_f the filtered output and (y,x), the coordinates within the frame of the window's center. Using the average provides smoothing in the images, thus reducing the noise. For this filter to appropriately affect marine snow, the kernel size needs to be chosen carefully. Indeed, a kernel too small will propagate the marine snow of bigger size while too large kernels, will remove the marine snow but will also have a too large smoothing effect on the background which results in important loss of local information.

All the tools to make the decision on the existence of marine snow in the images are now available. All that remains to be done is to determine for each pixel whether it is marine snow or part of the underwater scenery. To achieve this, the resulting smoothed images are separately subtracted to the current image and a comparison to a threshold for each pixel value is applied:

$$S(Y, Y_f) = \exists (y,x) \in C : Y(y,x) - Y_f(y,x) > \lambda_s \tag{6}$$

With S the new set of candidates using the raw Y component and a filtered one Y_f from a different frame, i.e. previous or next frame. C corresponds to the set of possible combinations of 2D pixel coordinates within the image. When processing with a total of three frames, the mask M_n, final set of candidates for the n^{th} frame, is defined as the intersection between the two set S based on the previous and next frames. It is defined as follows:

$$M_n = S(Y_n, Y_{f,n-1}) \cap S(Y_n, Y_{f,n+1}) \tag{7}$$

When the optional step explained in Section II-B is included, the set M_n becomes instead:

$$M_n = S(Y_n, Y_{f,n-1}) \cap S(Y_n, Y_{f,n+1}) \cap P_{f,n} \tag{8}$$

D. Result refinement

In some conditions, very small regions are categorized as marine snow. They are too small to have any use. For this reason, they need to be removed. Also, sometimes, some of the regions are not well formed and need to be reconstructed. Both can be done at the same time using mathematical morphology operations.

These are well known techniques when working with binary images that may contain imperfections due to thresholding operations. Morphological image processing assists in the removal of theses imperfections.

In the context of marine snow detection, the mask M_n is used as the binary image as it contains for each pixel the binary information of whether it is good or not. A structuring element, which is a template of small shape is then used to probe the image. It also contains binary values which are compared to the corresponding local neighbourhood of pixels in the image during the operation to test them and decide if a pixel should be added or removed.

The closing operation, defined as a dilation operation followed by an erosion operation is applied to the mask. Considering an image A and structuring element B, subsets of $\mathcal{Z}^2$, the dilation of A by B is defined as follows:

$$A \oplus B = \{z : \hat{B}_z \cap A \neq \emptyset\} \tag{9}$$

The dilation operation has the effect of expanding the shapes and filling in the small holes contained in the input image. It therefore also makes these shapes more visible and globally larger.

With the same subsets A and B, the erosion of A by B is expressed as follows:

$$A \ominus B = \{z : B_z \subseteq A\} \tag{10}$$

Morphological erosion has the opposite effect of dilation. It results in shape reduction but also removes very small objects that can be considered as floating pixels.

Based on these two operations, morphological closing can be formulated as follows:

$$A \bullet B = (A \oplus B) \ominus B \tag{11}$$

The closing operation tends to eliminate small holes and fill inner gaps in the objects, but also smooth them out and merges narrow breaks and long thin gaps.

By applying this closing operation on the marine snow mask, $M_n \bullet B$, a cleaner mask is obtained that corresponds now to the final set of pixels associated with marine snow.

III. Results and discussion

To test the proposed method, three scenarios are considered. First, the detection performance together with time consumption will be evaluated. The solution will then be tested for two real time applications: image enhancement and feature detection.

The implementation of the methods is done in Python and executed on a laptop containing an Intel Core i7 vPRO CPU and 16GB RAM. Video sequences were extracted from ROV video recordings of ship inspection missions with an image resolution of 1920 × 1080 and 25 images per second.

A. Evaluation

To evaluate the detection performance, it is necessary to possess the ground truth data. For this reason, the frames of a sequence were manually labeled for binary classification.

To evaluate the correctness of the pixel-wise classification, the supervised metric, F1 Score is employed. It provides contour accuracy $\mathcal{F}$ (eq. 14) defined as the harmonic mean of the precision $\mathcal{P}$ (eq. 12) and recall $\mathcal{R}$ (eq. 13) of the model.

$$\mathcal{P} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{12}$$

$$\mathcal{R} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{13}$$

$$\mathcal{F} = \frac{2 \times \mathcal{P} \times \mathcal{R}}{\mathcal{P} + \mathcal{R}} \tag{14}$$

The results of the classification evaluation are displayed in Table I. The results of the proposed method are compared with the method proposed in [7] which was re-implemented in Python.

TABLE I: Classification performance evaluation of different solutions for marine snow detection on video sequences

Method	F Score (%)
proposed method: full	74.18
proposed method: light	69.95
(B. Cyganek and K. Gongola) [7]	61.32

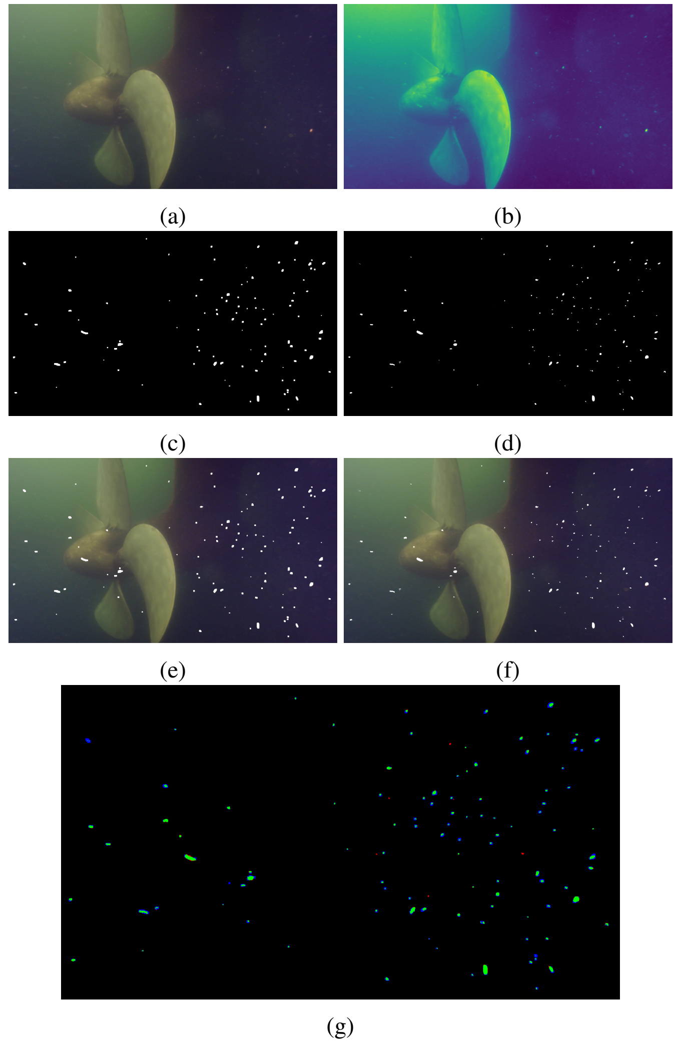


Fig. 3: Marine snow detection on a frame containing a ship's propeller. (a) corresponds to the original frame with (b) its Y component after conversion to the YCbC color space. (c) and (d) are respectively the ground truth and detected masks. In (e) and (f) the masks are superposed to the original frame. (g) presents the classification results compared to the ground truth.

When the optional modified guided filter is included, explained in Section II-B, the method is referred as full, and light otherwise.

It is possible to observe that it can achieve a F Score of up to 74.18%. Overall, most of the labeled particles are detected, but not exactly covered as the ground truth, especially the borders which are often further inside the ground truth objects. An example of classification is presented in Figures 3. It is a frame extracted from the sequence containing a large variety of marine snow with a ship's propeller in the background.

It appears clearly that the classification results are close to the ground truth but also that the size and shape of particles differ. In Figure 3g, the classification results are compared to the ground truth and a color is associated to each test result category. Green pixels are correctly detected marine snow, true positives, blue pixels are wrongly classified as background, false negatives, red pixels are wrongly classified marine snow, false positives, and black pixels, true negatives, correspond to correctly classified pixels that does not contain marine snow, i.e. background.

For real time applications, it is important to consider the processing time. The frame rate is measured and averaged over a whole video sequence with two different popular resolutions: 1920×1080 and 1280×720. The results are displayed in Table II with the frame rate expressed in Frames per second (FPS) and once more compared to the method proposed in [7].

TABLE II: Speed performance evaluation of different solutions for marine snow detection on video sequences of different resolutions

Method	Image resolution	Measured FPS
proposed method: full	1920 × 1080	9.54
proposed method: full	1280 × 720	13.98
proposed method: light	1920 × 1080	12.26
proposed method: light	1280 × 720	19.83
(B. Cyganek and K. Gongola) [7]	1920 × 1080	3.36
(B. Cyganek and K. Gongola) [7]	1280 × 720	8.13

The proposed solution can reach up to approximately 20 FPS in its minimal configuration, i.e. without the modified guided filter step and with lower resolution. The time performance can be affected in multiple ways: the video settings, processing platform but is also largely impacted by the method parameters, e.g. convolution kernel size, thresholds etc. For the evaluations, convolution kernels of size $k = 8$ were used together with thresholds between 0.0 and 0.05.

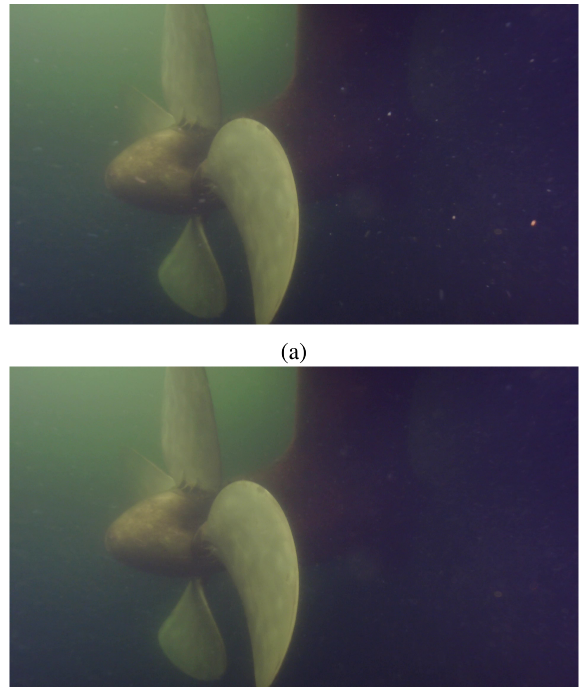
B. Application to image enhancement

To further test the proposed method, two use cases are considered. The first one is image enhancement. The objective is to perform marine snow removal based on detection results from the method presented here. To achieve this, a 3D median filter inspired from [7] is employed. For each value in the mask M_n, the median value of a 3D patch for each color channel is calculated and used to replace the pixel in the original image I_n. The 3D patches are based on the previous and next frames, for a pixel located in (y, x) in I_n, the 3D patch is composed of a set of pixels in two 2D patches centered in (y, x) with both next and previous frames, assuming the patch length q is odd. The 3D patch is therefore of size $q \times q \times 2$.

$$I_{n,c}(y, x) = \mathrm{Median}\left(\mathrm{Patch_{n-1,c}}(y, x) \cup \mathrm{Patch_{n+1,c}}(y, x)\right) \tag{15}$$

with $I_{n,c}(y, x)$ the focused pixel that will be replaced in the n^{th} frame and c^{th} color channel, located in (y, x). $Patch_n$ is the 2D patch of size q centered in (y, x) in the n^{th} frame.

An example result after removing the marine snow is presented in Figure 4. Even if the floating particles have been correctly detected, in some cases there may remain residues in the reconstructed image. This is often due to the speed of these particles in consecutive frames which can make the median filter retain pixel values corresponding to the marine snow. Overall, it is possible to observe in the figure small and large particles are correctly removed.



(a)

(b)

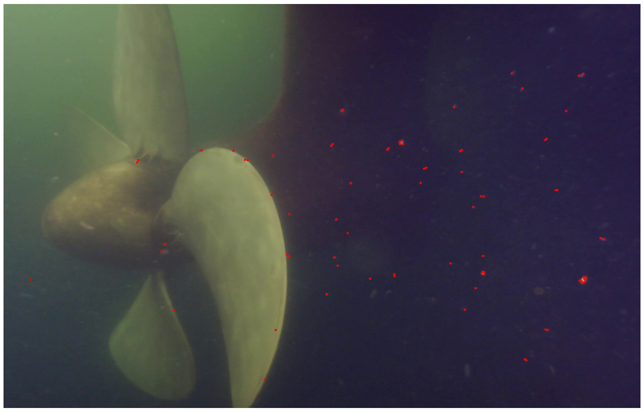
Fig. 4: Marine snow removal on a frame containing a ship's propeller. (a) is the original frame and (b) is the frame after marine snow removal with 3D median filter of size $9 \times 9 \times 2$.

C. Application to feature detection

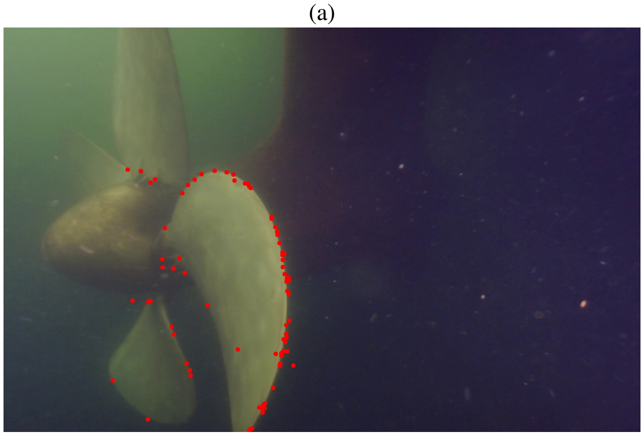
The second use case is feature detection. It is particularly important to make sure good features are detected in order to have relevant results with feature matching to be able to perform computer vision tasks such as navigation, object tracking, 3D reconstruction. To test the proposed method with feature detection, an unprocessed frame and the keypoints location are compared to the detected marine snow location. To achieve this, the Shi-Tomasi corner detector [16] is employed to find the most relevant features in the frame. The OpenCV[2] implementation of the detector is used in this test.

The results are presented in Figure 5. The best 75 feature points are marked with red circles in each image. In 5a the detector is used on the raw image without any additional processing. As a result, all the best detected features are located on the marine snow which would cause very biased results for computer vision tasks. In 5b, unlike the previous one, the proposed solution is applied to filter the set of detected keypoints. Those that are located on marine snow are removed and replaced with new ones at a different location. This results in significant improvements as all the best found keypoints are now located on the propeller. When applied over a whole video sequence, this will inherently lead to improved feature detection. When used together with keypoint tracking or matching, it is not necessary to apply the filter for every frame as the keypoints are not computed every time.

[2]OpenCV (Open Source Computer Vision Library): https://opencv.org



(a)



(b)

Fig. 5: Feature detection on a frame containing marine snow and a ship's propeller. In (a) feature detection is applied on the original frame without any additional processing whereas in (b), keypoints that are located on detected marine snow are removed and replaced by new ones.

IV. Conclusion

In this paper, a solution for marine snow detection in image sequences is proposed. It is designed for real time performance without requiring specific hardware setups. The method relies on the Y component after conversion to the YCbCr color space. According to the needs and computational capacities, an optional step for preselection of candidates is proposed based on a guided filter. A moving average filter is applied to the previous and next frames in time and compared to the current one for candidate selection. Finally, morphological image processing is performed to remove residual noise and have better shaped objects.

Based on the previous evaluations, the solution has shown that it is capable of working properly in presence of structures. It can reliably detect small and large floating particles, with a high but not absolute detecion rate. The tuning parameters included in the method makes it adaptable and suitable for most situations. The configuration of the processing system impact the detection speed, and a good trade-off in the tuning process can be found according to the user's setup. The performance can easily be improved with code optimization, mathematical simplification and assumptions which will be studied in future work. The proposed method could be further developed by adaptive estimation threshold values given the conditions in a given underwater scene.

References

[1] A. Arnold-Bos, J.-P. Malkasse, and G. Kervern, "A preprocessing framework for automatic underwater images denoising," *Proceedings of the European Conference on Propagation and Systems 2005*, 03 2005.

[2] C. Mobley, *Light and Water: Radiative Transfer in Natural Waters*, 01 1994.

[3] S. Bazeille, I. Quidu, L. Jaulin, and J.-P. Malkasse, "Automatic underwater image pre-processing," *Proceedings of CMM'06*, 10 2006.

[4] S. Banerjee, G. Sanyal, S. Ghosh, R. Ray, and S. N. Shome, "Elimination of marine snow effect from underwater image - an adaptive probabilistic approach," in *2014 IEEE Students' Conference on Electrical, Electronics and Computer Science*, 2014, pp. 1–4.

[5] F. Farhadifard., M. Radolko., and U. Freiherr von Lukas., "Single image marine snow removal based on a supervised median filtering scheme," in *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP, (VISIGRAPP 2017)*, INSTICC. SciTePress, 2017, pp. 280–287.

[6] F. Farhadifard, M. Radolko, and U. Freiherr von Lukas, "Marine snow detection and removal: underwater image restoration using background modeling," in *WSCG 2017: full papers proceedings: 25th International Conference in Central Europe on Computer Graphics, Visualization and Computer Visionin co-operation with EUROGRAPHICS Association*. Václav Skala - UNION Agency, 2017, pp. 81–89.

[7] B. Cyganek and K. Gongola, "Real-time marine snow noise removal from underwater video sequences," *Journal of Electronic Imaging*, vol. 27, no. 4, pp. 1 – 10, 2018.

[8] M. Koziarski and B. Cyganek, "Marine snow removal using a fully convolutional 3d neural network combined with an adaptive median filter," in *International Conference on Pattern Recognition*. Springer, 2018, pp. 16–25.

[9] L. Li, T. Komuro, K. Enomoto, and M. Toda, "Removal of floating particles from underwater images using image transformation networks," in *Pattern Recognition. ICPR International Workshops and Challenges*, A. Del Bimbo, R. Cucchiara, S. Sclaroff, G. M. Farinella, T. Mei, M. Bertini, H. J. Escalante, and R. Vezzani, Eds. Cham: Springer International Publishing, 2021, pp. 414–421.

[10] Y. Wang, X. Yu, D. An, and Y. Wei, "Underwater image enhancement and marine snow removal for fishery based on integrated dual-channel neural network," *Computers and Electronics in Agriculture*, vol. 186, p. 106182, 2021.

[11] Q. Jiang, Y. Chen, G. Wang, and T. Ji, "A novel deep neural network for noise removal from underwater image," *Signal Processing: Image Communication*, vol. 87, p. 115921, 2020.

[12] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013.

[13] R. Li, L.-F. Cheong, and R. T. Tan, "Heavy rain image restoration: Integrating physics model and conditional adversarial learning," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 1633–1642.

[14] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1715–1723.

[15] K. He and J. Sun, "Fast guided filter," 2015. [Online]. Available: https://arxiv.org/abs/1505.00996

[16] J. Shi and Tomasi, "Good features to track," in *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1994, pp. 593–600.