A Real-time Underwater Robotic Visual Tracking Strategy Based on Image Restoration and Kernelized Correlation Filters

Shihan Kong[1,2], Xi Fang[3], Xingyu Chen[1,2], Zhengxing Wu[2]*, Junzhi Yu[2]

1. University of Chinese Academy of Sciences, Beijing 100049, China

2. Key Laboratory of Management and Control for Complex Systems,
Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
E-mail: {kongshihan2016, chenxingyu2015, zhengxing.wu, junzhi.yu}@ia.ac.cn

3. School of Mechanical Engineering and Automation, Beihang university, Beijing 100191, China
E-mail: fangxi@buaa.edu.cn

Abstract: In this paper, a real-time underwater robotic visual tracking strategy (RUTS) based on underwater image restoration and Kernelized Correlation Filters (KCF) is developed for underwater robots. A real-time and unsupervised advancement scheme (RUAS), which is utilized in this strategy, performs robustly in restoring underwater images. The KCF, as a high-speed and accurate tracking method on land, is employed in this strategy. To handle the conflict between tracking speed and accuracy, we propose a tracking strategy based on KCF in video sequence restored by RUAS, comparing Histogram of Oriented Gradient (HOG) descriptors and raw pixels gray (RPG) descriptors. We define an index Ac to describe the tracking accuracy and regard the number of frames per second as computing speed. Results of contrast experiments show that the RPG, a much simpler descriptor, can achieve tracking accuracy as precise as HOG, accompanied by an increase of tracking speed up to 36%. Finally, experiments of the KCF-based tracker with RPG on different underwater objects demonstrate the feasibility of the formed RUTS.

Key Words: Underwater robotic vision, real-time underwater tracking, underwater image restoration, KCF

1 INTRODUCTION

The ocean, with numerous resources such as minerals, fuels, biological resources and so on, is full of unknown for mankind. Driven by this situation, an increasing number of researches on underwater robots are conducted all over the world. For example, Wu *et al*. proposed a novel robotic dolphin, which can be controlled robustly to monitor the water quality [1]. Underwater robotic vision allows underwater robots to achieve automatic control and operation. Particularly, technology of underwater visual tracking and image restoration are highly valuable for underwater robots to move towards destination and grab objects.

The underwater environment is greatly different from the atmospheric environment, where natural lighting is changed by refraction, abortion and scattering in the water [2]. As a result, the underwater image is of degraded quality, low contrast and blurred shape, which hampers the development of underwater visual object tracking. To resolve these problems, some algorithms of underwater image enhancement based on dark channel prior (DCP) [3] were provided. For instance, Mallik *et al*. introduced a method to enhance underwater image by DCP and contrast limited adaptive histogram equalization (CLAHE) [4], which has reliable results in non-real time process. Shortly before, Chen *et al*. adopted a real-time and unsupervised advancement scheme (RUAS) for underwater image restoration [5], and obtained a wonderful online testing consequence. A real-time tracker combined with methods of underwater image restoration was studied in depth by [6], but their underwater environment of experiments is widely different from the real marine environment.

Recently, many algorithms have been presented for visual object tracking on land. Bolme *et al*. employed an adaptive correlation filter, a Minimum Output Sum of Squared Error (MOSSE) filter, and a tracker based on MOSSE can operate at 669 frames per second, which is robust in various lighting and nonrigid deformation [7]. Based on correlation filters, Henriques *et al*. provided a Kernelized Correlation Filters (KCF) and demonstrated the high efficiency of the tracking method through experiments [8]. According to Henriques' work, the arithmetic speed and the tracking accuracy are connected with the feature description method of a target area. Histogram of Oriented Gradient (HOG) [9][10] descriptors are widely used in visual object tracking, which possess the characteristic of illumination invariance compared to using raw pixels gray (RPG) feature descriptors. Nevertheless, it will consume more time when HOG is used in a KCF based tracker [8].

Apparently, the primary problem of tracking object in real marine environment is the conflict between tracking accuracy and processing speed as is shown in Figure 1. To pursuit tracking accuracy, we choose HOG descriptors,

This work was supported in part by the National Natural Science Foundation of China (61633017, 61633004, 61603388, 61725305), in part by the Key Research and Development and Transformation Project of Qinghai Province (2017-GX-103), and in part by the Key Project of Frontier Science Research of Chinese Academy of Sciences (QYZDJ-SSW-JSC004).

978-1-5386-1243-9/18/$31.00

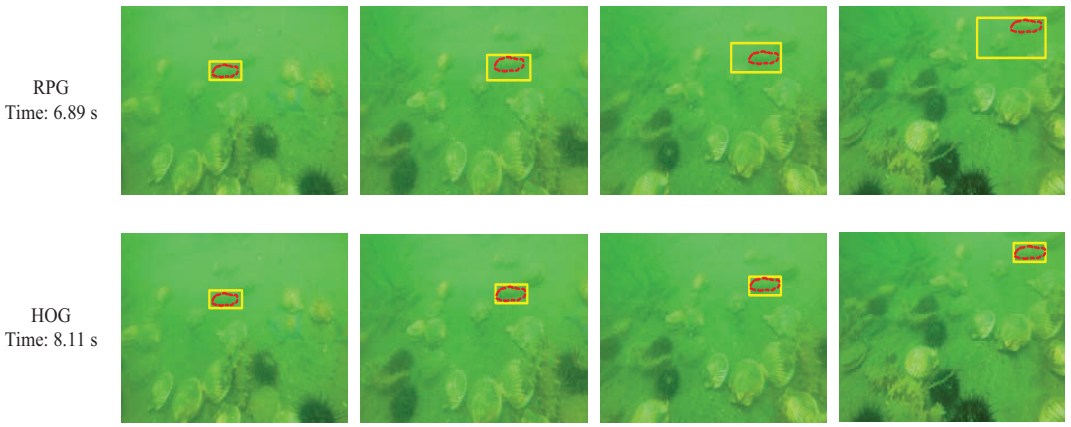

Figure 1: The conflict between HOG and RPG. The tracking object is a trepang, which is highlighted by a red dashed line in every frame. The time consumed is related to the computer hardware, for different CPU or GPU have different operational capabilities. From this figure, the conflict between HOG and RPG in tracking is noticeable.

by which the tracker performs precisely but consumes almost 1.2 times as much time as using RPG descriptors. However, results of a tracker with RPG descriptors are unsatisfactory and even preposterous sometimes. In practice, the size of the patch for objects gradually grows larger because of the low accuracy resulted by RPG descriptors, and affects the computing speed seriously. Our tracking strategy can improve the accuracy of RPG descriptors and it takes less time than using HOG. As a consequence, a compromise of tracking accuracy and processing speed is in need.

In this paper, we propose a KCF based real-time underwater robotic visual tracking strategy (RUTS). RUAS is used for underwater image restoration. Meanwhile, we attempt to compare RPG descriptors with HOG descriptors in tracking accuracy and computing speed. Through experiments on our underwater system in real marine environment, results demonstrate the satisfactory efficiency of RUTS with high computing speed and good enough tracking accuracy.

The rest of this paper is organized as follows. In Section II, the overview of RUTS, including the flowchart, comparison between HOG and RPG, the introduction of RUAS, and KCF, is described. In Section III, experiments and results in real underwater marine environment are evaluated and discussed. In Section IV, conclusions and future work are presented.

2 OVERVIEW OF THE RUTS

To realize the function of RUTS, feature descriptors for objects, image restoration and KCF are three important theoretical bases. In this section, the first subsection introduces the process of RUTS, and the last three subsections are organized to describe these theoretical basis and analyze the feasibility of RUTS in depth.

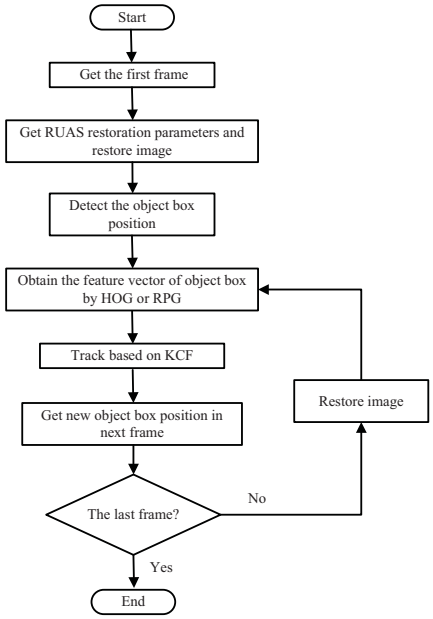

Figure 2: The flowchart of RUTS.

2.1 Process of RUTS

The flowchart of RUTS is shown in Figure 2. The first frame from a video sequence, is used to calculate the RUAS restoration parameters, by which we restore the image in every frame. At the same time, the computer obtains the location of the object box by intelligent detect algorithm or manual annotation. Then, the feature vector of the object box can be described by HOG or RPG. A tracker based on KCF is chosen to the position of new object box in the next frame. The feature vector of object is obtained over and over again after every image is restored, until the frame is the end of this video sequence.

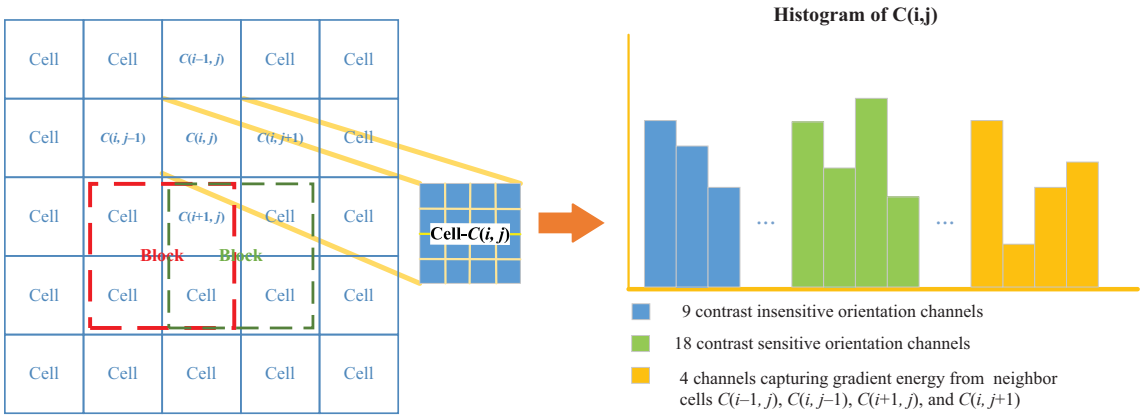

Figure 3: The process of obtaining HOG. The relationship between Cell and Block is described; the schematic diagram of the improved HOG used in this paper is visualized.

2.2 Comparison Between HOG and RPG

How to describe the object is a vital problem related to the tracking accuracy and computing speed for a tracker. HOG and RPG are two descriptors of different principle and complexity but all widely used in visual object tracking field.

Histograms of Oriented Gradient (HOG) was first presented by [9] for human detection. The window image is divided into cells, and a local 1-D histogram of gradient orientation, similar to a feature vector $h^i_{1\times n}$, is accumulated for every cell to represent its feature. Several cells constitute a larger and spatial connected block. Then histograms of cells regularly link, together forming a longer feature vector $\{h^1_{1\times n}, h^2_{1\times n}, h^3_{1\times n}, ...\}$. Contrast-normalization is applied in every overlapping block. Finally, feature vectors of overlapping blocks link together to form HOG feature. The relationship between Cell and Block is described in Figure 3. Briefly, HOG descriptors extract texture features of image meticulously and overcome the problem of variant illumination.

To realize better tracking performance and fit the circulant matrix used in KCF, a revised HOG put forward by [10] is chosen to describe object features in [8]. For each cell $C(i, j)$, the final feature has 31 channels, with 27 channels including 9 contrast-insensitive orientation channels and 18 contrast-sensitive orientation channels, and 4 channels capturing the overall gradient energy in square blocks of four cells around $C(i, j)$, which is visualized in Figure 3.

On the other hand, the most basic characteristic of an image is raw pixels gray (RPG), which can be converted from a colorful RGB image by the following equation [11].

$$Gray = R \times 0.333 + G \times 0.5 + B \times 0.1666 \quad (1)$$

where R, G, B represent the intensity of red channel, green channel, blue channel over each pixel. Contrast to HOG, RPG descriptor is quite simple sensitive to illumination, shadow and blur.Therefore, unless the underwater image is of great quality, RPG descriptor is accessible.

As for calculated amount, an 100×100 pixels object box described by HOG with cell size 4 pixels, will generate a feature vector owning $25 \times 25 \times 31 = 19375$ elements. On the contrary, using RPG only produces a feature vector with $100 \times 100 = 10000$ elements. Furthermore, it is noticeable that RPG is much simpler than HOG when the size of object is relatively large. Using RPG instead of HOG under specific situation will improve tracking efficiency and benefit automatic control of underwater robots.

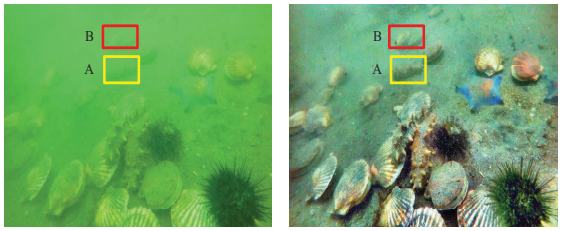

Figure 4: The performance of RUAS. The left degraded image is green and of poor quality due to underwater abortion and scattering, where Box A and Box B are almost disappeared. The right image is restored by RUAS, where the Box A is a trepang and the Box B is a scallop.

2.3 Analysis for Restoration Results of RUAS

Chen $et\ al.$ provided three parameters, which were related to underwater image degradation and color correction, by pre-searching in the the first frame of images sequences through artificial fish school algorithm [5]. The core of RUAS is a Wiener Filter in frequency domain as follows

$$V_{orig,C}(u, v) = [\frac{H(u, v)}{|H(u, v)|^2 + R}]V_{deg,C}(u, v) \quad (2)$$

where $V_{orig,C}$ represents one channel of original image, and $V_{deg,C}$ represents one channel of degraded image due to underwater scattering and abortion. R is the reciprocal of signal to noise ratio and will be implemented to restrict scattering. $H(u, v)$ is originated as a general image degradation model in turbulent media [12], expressed by

$$H(u, v) = e^{-k(u^2+v^2)^{5/6}} \quad (3)$$

where k is a crucial parameter related to the depth of water and the distance from the camera.
After Wiener Filter, color correction is implemented to the image by gamma factor as follows

$$I_{corrected,C} = I^{\gamma} \tag{4}$$

At this point, R, K, and γ have been introduced. To obtain a reliable combination of these three parameters, we employ a quality index of restored image expressed as

$$Q = \frac{\alpha\beta}{1+\eta} \tag{5}$$

α is a haze indicator, describing the level of haze by gradient computed by modified Tenengrad evaluation.

$$\alpha = \frac{1}{W}\sum_{i=0}^{M}\sum_{j=0}^{N}\sum_{k=0}^{7}|Gradient(V_g(i,j),k)|^2 \tag{6}$$

where $M \times N$ is the size of an input image; V_g is a gray-scale map, and orientations of gradient are regulated as $k \times 45°$. This indictor takes the textural feature and edge feature into consideration. Generally, the higher value of α reflects a clearer restored image.
β is a contrast indicator, which is calculated by histogram distribution in RGB channels, representing the image contrast.

$$\beta = \frac{1}{MN}\sum_{C\in\{R,G,B\}}\sqrt{\sum_{i=0}^{255}(h_C(i)\times i - \mu_C)^2} \tag{7}$$

where $h_C(i)$ stands for the data of histogram curves at gray level i for channel C; μ_C shows the average of histogram curves of channel C. Theoretically, objects can be distinguished more easily with a high value of β.
η is an imbalance indicator, which denotes the level of color correction.

$$\eta = |\mu_r - \mu_b| + |\mu_r - \mu_g| + |\mu_b - \mu_g| \tag{8}$$

Clearly, the η diminishes along with a better color correction.
The performance of RUAS shown in Figure 4. In addition, the amount of information in the restored image has been retained to a great degree, such as color information, texture and edge information, illumination information, etc., which is available for this paper to test HOG and RPG descriptors on underwater visual object tracking. In Section III, experiments will exhibit the tracking performance of HOG and RPG.

2.4 Kernelized Correlation Filters

Visual object tracking can be regarded as a non linear regression, to classify the object and background in [8]. Henriques $et\ al.$ utilized the circulant matrix, the non linear region regression and the correlation filter to achieve a high speed and accuracy tracker. In this paper, the principle will not be elaborated in details, but the process of KCF is discussed as follows.

An $m \times n \times c$ size training image patch i, an $m \times n$ size Gaussian-shaped regression target r, and an $m \times n \times c$ size training image patch t are given. First of all, the Gaussian kernel is derived as

$$k^{ii} = exp\{-\frac{1}{\sigma^2}(||i||^2 + ||i||^2 - 2\mathcal{F}^{-1}(\sum_c \hat{i}_c^* \odot \hat{i}_c)\} \tag{9}$$

where the hat $\hat{}$ denotes a DFT of a vector, the star * denotes a conjugated vector and the $\odot$ represents element-wise product in this subsection from now on; the c stands for different channels. Next, the vector of ridge regression coefficients A is estimated by

$$\hat{A} = \frac{\hat{r}}{\hat{k^{ii}} + \lambda} \tag{10}$$

where λ is a regularization parameter in case of over fitting. Similar to equation (9), k^{ti} and $\hat{k^{ti}}$ can be obtained. Finally, the response is expressed by the equation below.

$$\hat{f(t)} = \hat{k^{ti}} \odot \hat{A} \tag{11}$$

When $\hat{f(t)}$ is IDFT to spatial domain, the location of max value of response is the new object location of the patch. In our experiments, it is proven that employing KCF in RUTS is feasible and dependable.

3 EXPERIMENTS AND RESULTS

In this section, contrast experiments on RUTS between RPG and HOG are conducted, followed by experiments on RUTS with RPG on different objects to verify the feasibility. After experiments, results are obtained and discussed in the end.

3.1 Contrast Experiments on RUTS and Results

The video sequences of real marine environment were captured by our underwater robot platform diving in the sea area at Zhangzi Island, Dalian, China. Objects are three kinds of marine organisms, i.e., trepang, scallop, and urchin. We pre-processed the video sequences by marking object patches manually every frame. HOG and RPG descriptors are tested separately to evaluate their tracking performance. OpenCV 3.0 in the programming environment Microsoft Visual Studio 2012 is used for the experiment on the computer with the Intel Core i7-6700 3.40 GHz CPU.
In the experiment, the size of tracking patch is variable to fit the object more closely. The accuracy in every frame is estimated by

$$Ac = \frac{\sqrt{||P_{tl} - P_{tl,truth}||^2 + ||P_{br} - P_{br,truth}||^2}}{||P_{tl,truth} - P_{br,truth}||} \tag{12}$$

where P_{tl} and $P_{tl,turth}$ are experimental and true top-left coordinates of the object patch, and P_{br} and $P_{br,turth}$ are bottom-right coordinates analogously. Smaller value of Ac represents a higher fitting degree between tracking box and the actual object box, which denotes a better tracking performance. In practice, $Ac \leq 0.13$ can be considered as an acceptable tracking accuracy.

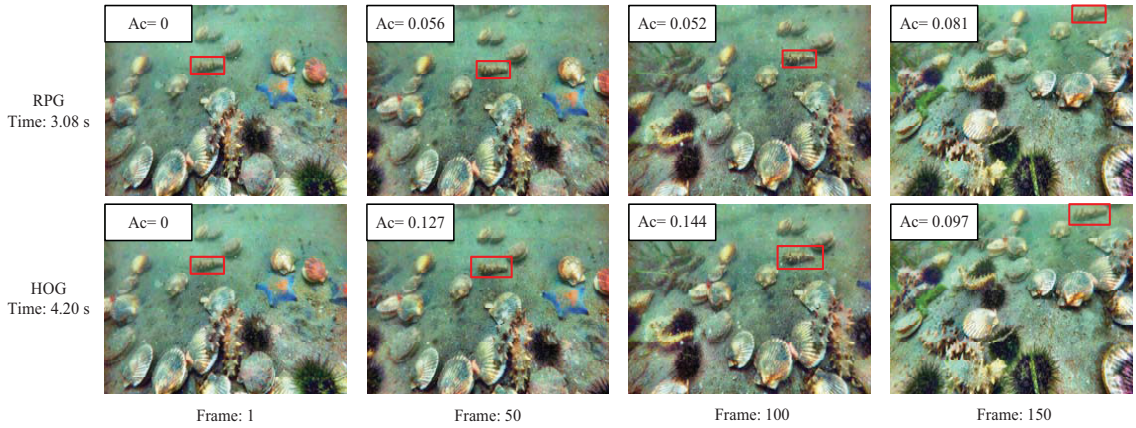

Figure 5: The results of contrast experiment on trepang, the time consumed and tracking accuracy are displayed on.

We choose a video sequences with 150 frames, tracking a trepang by RUTS and considering the tracking accuracy and time consumed, with results shown in Figure 5. The average Ac of the tracker with the simpler RPG descriptors is 0.063, which is much smaller than the average $\bar{Ac} = 0.123$ of the tracker with HOG descriptors, as distinctly perceived. The processing speed of the tracker with RPG is up to about 48.7 frames per second, and the speed of the tracker with HOG is merely 35.7 frames per second. In short, results of the contrast experiment show that the RPG, a much simpler descriptor, can achieve tracking accuracy as precise as HOG, accompanied by an increase of tracking speed up to 36%.

3.2 Experiments on RUTS with RPG

After the contrast experiments on RUTS, results show that describing objects with RPG performs even better than describing with HOG in underwater object tracking. To verify the feasibility of RUTS further, we conducted three experiments on tracking different objects in other underwater videos captured from the real marine environment. The objects include sea urchin, scallop, and starfish, which have different sizes and features. Results are expressed in the Table 1, and the tracking process is visualized in Figure 6.

Experiments on tracking different underwater objects indexes that the tracking accuracy and computing speed is satisfactory, so RUTS can be a practical tracking method for underwater visual tracking. However, the RPG is out of rotation invariance, as exposed in the process of tracking scallop, for which an improved descriptors must be proposed in the future. Additionally, when the size of object box becomes larger, the computing speed will be lower. The stable tracking accuracy of RUTS is research motivation for us.

3.3 Discussion

Through these experiments on RUTS, the results indicate that RUTS, a compromise strategy, is feasible and realize the accurate and real-time visual object tracking. The image is restored by RUAS. In this situation, the gray-scale feature can contain enough information of objects, even better than HOG. In brief, the quality of underwater image is adequate, which is the core reason of succeeding in precise tracking by RUTS. Note that on account of the computing capability of our computer, the experimental results can indicate the different tracking accuracy and speed between RPG and HOG descriptors, but cannot stand for the best computing speed of the strategy.

Compared with other similar researches on underwater visual object tracking, we use a novel method of underwater image restoration, and the experimental environment in this paper is the real ocean. Images in the marine environment is full of uncertain geometric deformation including rotation, affine transformation, and so on. RPG based on gray-level, is hardly able to describe these features of a object, which reduces the tracking accuracy.

4 CONCLUSION AND FUTURE WORK

The fundamental purpose of this paper is to propose a real-time visual tracking strategy for underwater objects, which can track object using a simpler descriptors. HOG-based tracking is better in terms of accuracy but time-consuming, and using RPG can track at a high speed but with low accuracy. In this paper, we compare the tracking performance of HOG and RPG descriptors with RUAS-processed underwater video, proposing a compromise strategy to visually track objects underwater. According to the analysis and discussion of HOG and RPG descriptors, the underwater image restoration method RUAS, and the experimental results, the feasibility of RUTS is demonstrated.

In future research, the underwater tracking accuracy will be further enhanced, since RUTS is only a compromise method for tracking without strong stability. An improved robust RUTS will be proposed in the future work. Meanwhile, the combination of underwater vision tracking and autonomous underwater robot operation will be investigated. With such capabilities, the underwater robot can move towards to the objects and grab them through underwater vision tracking.

Table 1: Results of Tests on RUTS with RPG

Object type	Original size of Object Box (Pixels × Pixels)	Tracking accuracy (Average of Ac)	Tracking speed (Frames per second)
Sea urchin	120 × 90	0.116	27.34
Scallop	110 × 70	0.084	48.07
Starfish	100 × 75	0.120	43.71

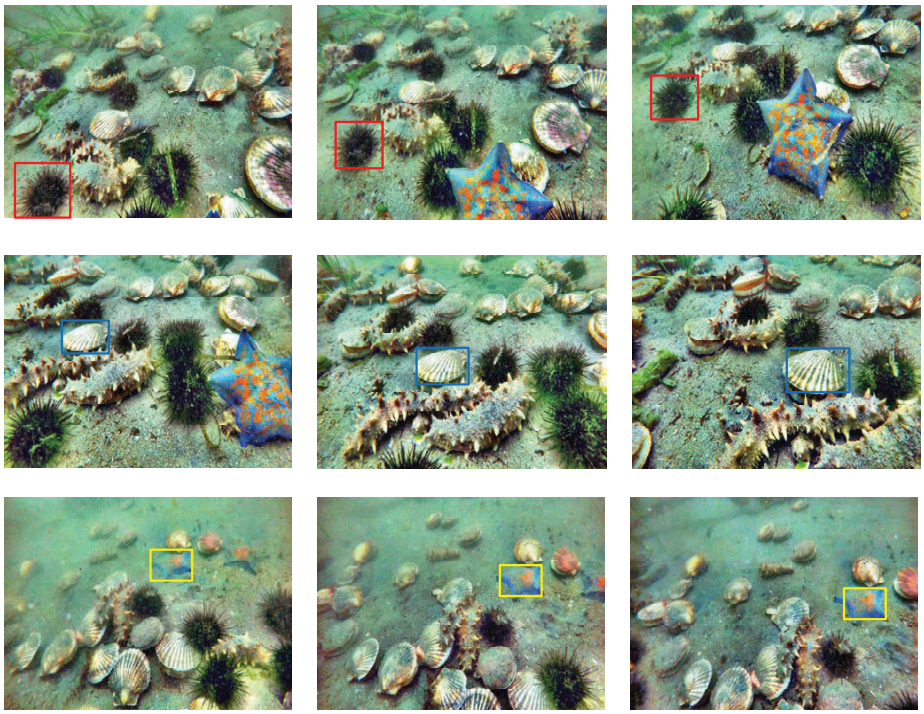

Figure 6: The process of tests on sea urchin, scallop, and starfish.

REFERENCES

[1] Z. Wu, J. Liu, J. Yu, and H. Fang, "Development of a novel robotic dolphin and its application to water quality monitoring," *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 5, pp. 2130-2140, 2017.

[2] Y. Y. Schechner and N. Karpel, "Clear underwater vision," in *Procedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Washington, USA, Jul. 2004, pp. 536-543.

[3] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341-2353, 2011.

[4] S. Mallik, S. S. Khan, and U. C. Pati, "Underwater image enhancement based on dark channel Prior and Histogram Equalization," in *IEEE International Conference on Innovations in Information Embedded and Communication Systems*, Coimbatore, Tamilnadu, India, Mar. 2016, pp. 139-144.

[5] X. Chen, Z. Wu, J. Yu, and L. Wen, "A real-time and unsupervised advancement scheme for underwater machine vision," in *IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems*, Hawaii, USA, Aug. 2017, pp. 271-276.

[6] D. Lee, G. Kim, D. Kim, H. Myung, H. T. Choi, "Vision-based object detection and tracking for autonomous navigation of underwater robots," *Ocean Engineering*, vol. 48, no. 7, pp. 59-68, 2012.

[7] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Procedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, USA, Jun. 2010, pp. 2544-2550.

[8] J. F. Henriques, C. Rui, P. Martins, and J. Batista, "High-speed tracking with Kernelized Correlation Filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583-596, 2015.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Procedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, USA, Jun. 2005, pp. 886-893.

[10] P. F. Felzenszwalb, R. B. Girshick, and D. Mcallester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627-1645, 2010.

[11] T. Kumar and K. Verma, "A theory based on conversion of RGB image to gray image," *International Journal of Computer Applications*, vol. 7, no. 2, pp. 7-10, 2010.

[12] N. R. Stanley and R. E. Hufnagel, "Modulation transfer function associated with image transmission through turbulent media," *Journal of the Optical Society of America*, vol. 54, no. 1, pp. 50-60, 1964.