

# Saliency-Based Adaptive Object Extraction for Color Underwater Images

Huibin Wang, Xin Dong, Jie Shen, Xuewen Wu, Zhe Chen  
 College of Computer and Information Engineering  
 Hohai University  
 Nanjing, China  
 Email: hbwang@hhu.edu.cn

**Abstract**—Because of the special optical underwater imaging environment, the contrast and quality of images are affected severely, causing it difficult to extract objects from underwater images. An adaptive underwater object extraction method based on the saliency maps is proposed in this paper. Firstly, preprocessing method is utilized to improve the color contrast and quality. Then multi-scale image features are combined into a single topographical saliency map. The most salient image location and primary object are directed by the saliency map. By calculating the Bhattacharyya distance of salient features between the primary object and background, the adaptive weights of conspicuity maps can be obtained, as well more accurate object can be extracted from the fuzzy images. Experiment results show that the proposed method can not only detect the objects effectively but also extract more accurate object areas. It has a better performance compared with other algorithms.

**Keywords**—underwater image; object extraction; saliency map; retinex

## I. INTRODUCTION

It is well known that object extraction from images is playing an important role in image project, and many methods were developed in the past decades, including edge detection [1], region growth [2], graph cut [3], etc. In the underwater environment, due to the major obstacles originating from the strong light absorption and scattering, underwater optical information acquisition and processing suffer from effects such as limited range, non-uniform lighting, low contrast, and diminished colors. It is hard to extract the objects accurately from the fuzzy underwater images [4]. Therefore, image enhancement and restoration process is essential to enhance the color and edge information in underwater images.

In the previous research, different features were introduced into object extraction, such as intensity, color and texture. For example, Zhang [5] utilized the texture as a main feature and presented a method based on the discrete fractional Brownian random field for underwater image segmentation. Liu [6] proposed a feature matching algorithm based on Hough transform and geometrical feature. However, because of the underwater noise, uneven illumination and low contrast, the majority of underwater imaging characteristics are difficult to extract. For example, due to the poor underwater light conditions, there are phenomenon of lots of false details and contours, self-shadowing in the image.

In our adaptive object extraction framework, preprocessing of images is firstly utilized to minimize the effects of underwater environment. Then multi-scale image features are combined into a single topographical saliency map, and the primary object is located. By calculating the Bhattacharyya distance of the features between the primary object and background, the adaptive weights of conspicuity maps can be obtained. The more precise object will be detected effectively overcoming the influence of environment, and an accurate extraction of the target is realized.

## II. ADAPTIVE UNDERWATER OBJECT EXTRACTION MODEL

### A. The Saliency Map

The classic saliency map model is proposed by Itti [7]. Generally, there are three steps in building a saliency map model. First, low-level vision features (color, intensity, orientations) are extracted from the original color image at several spatial scales (depending on the image size) [8]. The different spatial scales are obtained through the use of a dyadic Gaussian pyramid. Then the feature maps are combined into three conspicuity maps using across-scale combinations and normalization process. The final step is to combine the three different maps to obtain a saliency map.

### B. Adaptive Underwater Object Extraction Model

The main problem encountered in our underwater images is the low contrast, a problem which is closely related to changes in the imaging environment. We propose to use the saliency map building model with some modifications. Once the early saliency map  $S$  is extracted, we calculate the Bhattacharyya distance between the primary salient part and the background to modify the weights of conspicuity maps. And the revised weights are utilized to obtain an improved saliency map  $S'$ . This strategy can respond to changing environment rapidly and effectively, which is a new method for saliency map construction, shown in Fig. 1.

Firstly, we restore and enhance the fuzzy underwater images with Retinex algorithm [9] as a preprocessing step to reduce the interference of fuzzy and distortion on salient objects. The process of object extraction is divided into two parts: primary object area location and precise object extraction. In the primary object area location stage, we get color, intensity and orientation conspicuity maps according to the Itti model. After setting the same weights for the

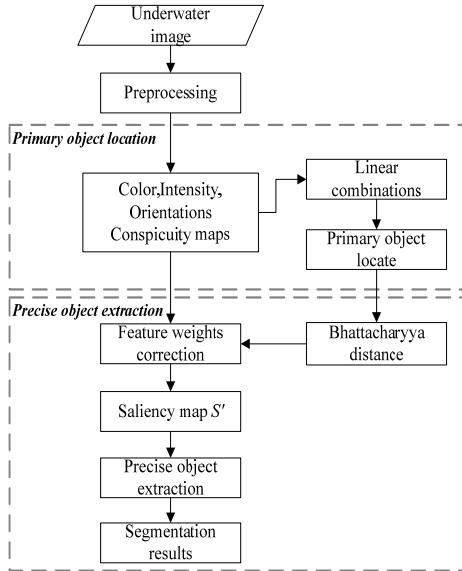


Figure 1. Adaptive Underwater Object Extraction Model

conspicuity maps, a saliency map is combined, which represents the conspicuity at every location in the visual field and guides the selection of attended locations. We use outer rectangular to represent the bounds of primary object location. In the precise object extraction stage, we calculate the Bhattacharyya distances between the primary object and the background to update the weights of conspicuity maps, using the color, intensity and orientation histograms. And an improved saliency map can be obtained with the revised weights. Finally, the precise object will be extracted from the new saliency map.

### III. PRIMARY OBJECT AREA LOCATION

#### A. Extraction of Color Conspicuity Map

Four broadly-tuned color channels are created:

$$R = r - \frac{(g+b)}{2} \quad (1)$$

$$G = g - \frac{(r+b)}{2} \quad (2)$$

$$B = b - \frac{(g+r)}{2} \quad (3)$$

$$Y = \frac{(r+g)}{2} - \frac{|r-g|}{2} - b \quad (4)$$

Four Gaussian pyramids  $R(\sigma)$ ,  $G(\sigma)$ ,  $B(\sigma)$  and  $Y(\sigma)$  are created from these color channels, where  $\sigma \in [0, \dots, 8]$  is the scale, and 12 feature maps for color are computed as follow:

$$RG(c, s) = |(R(c) - G(c)) \ominus (G(s) - R(s))| \quad (5)$$

$$BY(c, s) = |(B(c) - Y(c)) \ominus (Y(s) - B(s))| \quad (6)$$

$$I_1 = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} [N(RG(c, s)) + N(BY(c, s))] \quad (7)$$

The across-scale difference between two maps, denoted as ' $\ominus$ ', is obtained by interpolation to the finer scale and

point-by-point subtraction.  $N(\bullet)$  is a normalization operator, and ' $\oplus$ ' is an across-scale addition, which consists of reduction of each map to scale four and point-by-point addition [7].  $I_1$  is the primary color conspicuity map.

#### B. Extraction of Intensity Conspicuity Map

The intensity of an image is obtained as  $I = (r+g+b)/3$  where  $r$ ,  $g$ , and  $b$  represent the red, green, and blue channels.  $I$  is used to create a Gaussian pyramid  $I(\sigma)$ . The  $r$ ,  $g$  and  $b$  channels are normalized by  $I$  in order to decouple hue from intensity. Here, both types of sensitivities are simultaneously computed in a set of six maps  $I(c, s)$  with  $c \in \{2, 3, 4\}$  and  $s = c + \delta$ ,  $\delta \in \{3, 4\}$ .

$$I(c, s) = |I(c) \ominus I(s)| \quad (8)$$

$$I_2 = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} N(I(c, s)) \quad (9)$$

Where  $I(c)$  is the scale of  $c$ ,  $I(s)$  is the scale of  $s$  and 6 feature maps for intensity are computed.  $I_2$  is the intensity conspicuity map.

#### C. Extraction of Orientations Conspicuity Map

The orientation features is achieved using oriented Gabor filter for image smoothing and down-sampling, and the orientation in the Gabor filter are  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  and  $135^\circ$ . The formula for calculating the difference between the center and surround scales:

$$o(c, s, \theta) = |o(c, \theta) \ominus o(s, \theta)| \quad (10)$$

$$I_3 = \sum_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} N \left( \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} N(o(c, s, \theta)) \right) \quad (11)$$

And 24 feature maps for orientation are computed.  $I_3$  is the orientation conspicuity map.

#### D. Primary Object Area Location

For the enhanced image, respectively,  $I_1$ ,  $I_2$  and  $I_3$  indicates the color, intensity and orientation conspicuity map, the primary salient region as shown:

$$S = \frac{I_1 + I_2 + I_3}{3} \quad (12)$$

The primary object area could be located by its maximum enclosing rectangular after segmentation. This area is denoted as  $V$ .

### IV. WEIGHT DESIGN OF THE CONSPICUITY MAP

In order to extract the more precise object, we adjust the proportions of the conspicuity maps as shown:

$$S' = \alpha_1 \cdot I_1 + \alpha_2 \cdot I_2 + \alpha_3 \cdot I_3 \quad (13)$$

Where  $I_1$ ,  $I_2$  and  $I_3$  are the three conspicuity maps for color, intensity and orientation features and  $\alpha_1 + \alpha_2 + \alpha_3 = 1$ .

#### A. Factors for Comparison

In order to measure the saliency between the primary object areas and the background, we calculate their color, intensity and orientation histograms and compare the differences between them using Bhattacharyya distance.

For the HSV color space is much in line with human visual system (HVS), we convert the images to HSV space and carry out the non-uniform quantization of each component. The hue component of the image is quantized to 16 levels, and the saturation and value components are quantified to 4 levels. Then, the H-S histogram can be transformed to a one-dimensional vector, which represents 64 kinds of colors. Let the image's color histogram is His:

$$His = [c_1, c_2, \dots, c_{64}] \sum_{i=1}^{64} c_i = 1, 0 \leq c_i \leq 1 \quad (14)$$

Where,  $c$  represents 64 levels of colors, such as those in Fig. 2(b).

The image's intensity information is denoted as  $I=(r+g+b)/3$ . Let  $n_i$  is the number of times that gray level  $i$  occurs, the probability of gray level  $i$  in the image is shown as follow:

$$P_x(i) = \frac{n_i}{n}, i \in 0, \dots, L-1 \quad (15)$$

Where,  $L$  is the gray level of the image,  $n$  is the number of all pixels in the image,  $P$  is the intensity distribution normalized to 0~1. The intensity histogram of Fig. 2(a) is shown in Fig. 2(c).

Four orientation-selective pyramids are created using Gabor filtering at  $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ . We use the following family of Gabor functions:

$$h(x, y, \theta, \lambda, \sigma_x, \sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{-\pi\left[\left(\frac{x-x_0}{\sigma_x}\right)^2 + \left(\frac{y-y_0}{\sigma_y}\right)^2\right]\right\} \cdot \exp\left(\frac{2\pi i x y \theta}{\lambda}\right) \quad (16)$$

Where  $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ . The probability of threshold-filtered orientation feature map is shown as:

$$D_x(i) = \frac{n_i}{n}, i \in 0^\circ, 45^\circ, 90^\circ, 135^\circ \quad (17)$$

Where,  $n$  is the total number of pixels and  $D$  is the orientation histogram of the image. The orientation histogram of an example is shown in Fig. 2 (d).

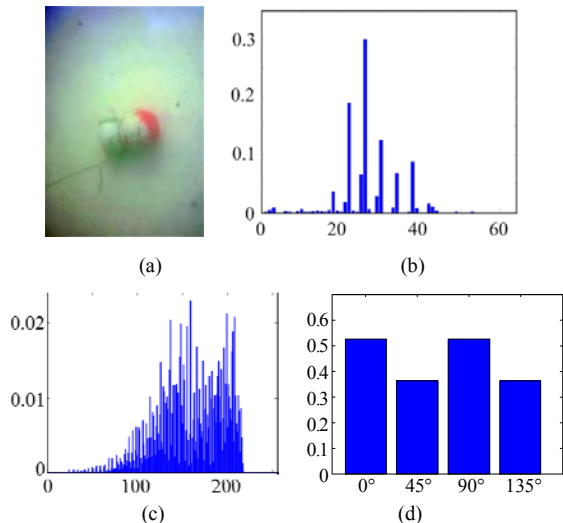


Figure 2. (a) Example of Enhanced Underwater Image. (b) Color Histogram. (c) Intensity Histogram. (d) Orientation Histogram.

### B. Bhattacharyya Distance Calculations for Weights Design

In order to measure the discrete probabilities between the primary object and the background, we use Bhattacharyya distance computation of the histograms for description. The Bhattacharyya distance between  $p$  and  $q$  is given by [10]:

$$d(p, q) = \sqrt{1 - \rho[p, q]} \quad (18)$$

Where  $\rho[p, q] = \sum_{u=1}^m \sqrt{p_u q_u}$  is the Bhattacharyya coefficient between  $p$  and  $q$ .

We have analyzed the Bhattacharyya distances between the primary object location area and the background, found that the distances were significantly correlated with the salient features. For more apparent features, increasing the weights can produce a better saliency map. But for less apparent features, the weights should be decreased.

For an enhanced image  $I$ , the primary object area is denoted as  $V$  and the background area is denoted as  $V'$ . According to equations (14~17), the histograms of color, intensity and orientation features of  $V$  are calculated and denoted respectively as  $His_1, P_1$  and  $D_1$ . And the histograms of  $V'$  are  $His_2, P_2$  and  $D_2$ . Their Bhattacharyya distances can be calculated from the Equation (18) as:  $d_{His} = d(His_1, His_2)$ ,  $d_P = d(P_1, P_2)$ ,  $d_D = d(D_1, D_2)$ . Therefore, the optimum weights can be obtained by normalization of  $d_{His}$ ,  $d_P$ , and  $d_D$ .

$$\alpha_1 = \frac{d_{His}}{d_{His} + d_P + d_D} \quad (19)$$

$$\alpha_2 = \frac{d_P}{d_{His} + d_P + d_D} \quad (20)$$

$$\alpha_3 = \frac{d_D}{d_{His} + d_P + d_D} \quad (21)$$

### C. Underwater Object Extraction

An improved saliency map can be obtained as below:

$$S' = \alpha_1 \cdot I_1 + \alpha_2 \cdot I_2 + \alpha_3 \cdot I_3 \quad (22)$$

Where  $\alpha_1 = d_{His} / (d_{His} + d_P + d_D)$ ,  $\alpha_2 = d_P / (d_{His} + d_P + d_D)$  and  $\alpha_3 = d_D / (d_{His} + d_P + d_D)$ . And the precise object areas can be extracted by segmenting the improved saliency map of the image.

## V. RESULTS AND DISCUSSION

### A. Underwater Image Enhancement Results

Retinex is a human perception-based image processing algorithm which provides color constancy and dynamic range compression [9]. The image enhancement results are shown in Fig. 3.

The contrast parameters and standard deviations can reflect the effects of enhancement. And the results of the above experiment are shown in Table I, which shows that the Retinex algorithm for underwater image enhancement is effective.

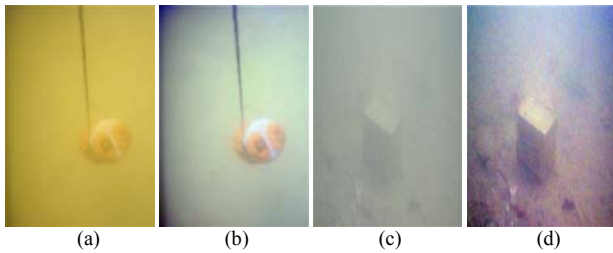


Figure 3. Examples of Underwater Image Enhancement with Retinex Algorithm. (a) Original Image 2. (b) Enhancement of Image 2. (c) Original Image 3. (d) Enhancement of Image 3.

TABLE I. Compare of the Images in Fig. 4

	Contrasts	Standard Deviations
Original Image 1	0.0032	0.0567
Enhanced Image 1	0.0260	0.1611
Original Image 2	0.0013	0.0256
Enhanced Image 2	0.0245	0.1512

*B. Primary Object Area Location Results*

In the primary object area location stage, feature extraction is achieved through linear filtering for a given feature type (e.g., intensity, color or orientation), followed by a center-surround operation which extracts local spatial

discontinuities for each feature type. And 6 for intensity computed at different spatial scales, 12 for color and 24 for orientation [8]. The set of feature maps is generated from center-surround differences between center fine scales and surround coarser scales. And three conspicuity maps are constructed from feature maps through across-scale combinations and normalization process. The primary object area location results of underwater images are shown in Fig. 4.

*C. Obtainment of Weight Values*

Using the methods proposed in Section 4, we calculated the Bhattacharyya distances of color, intensity and orientation histograms between primary object regions and the background regions respectively. After this, the optimum weights can be obtained by Equations (19-21) comparing the original weights  $\alpha_1=\alpha_2=\alpha_3=1/3$ .

For the image1 in Fig. 5(a), the distances are  $d_1=0.6879$ ,  $d_2=0.6176$  and  $d_3=0.6115$ , and the optimum weights are  $\alpha_1=0.3588$ ,  $\alpha_2=0.3222$  and  $\alpha_3=0.3190$ . For the image2 in Fig. 5(a), the distances are  $d_1=0.4923$ ,  $d_2=0.6059$  and  $d_3=0.7827$ , and the optimum weights are  $\alpha_1=0.3221$ ,  $\alpha_2=0.2617$  and  $\alpha_3=0.4162$ . For the image3 in Fig. 5(a), the distances are  $d_1=0.4221$ ,  $d_2=0.6257$  and  $d_3=0.6121$ , and the optimum weights are  $\alpha_1=0.2543$ ,  $\alpha_2=0.3770$  and  $\alpha_3=0.3687$ . Meanwhile, the optimum weights will change for different images.

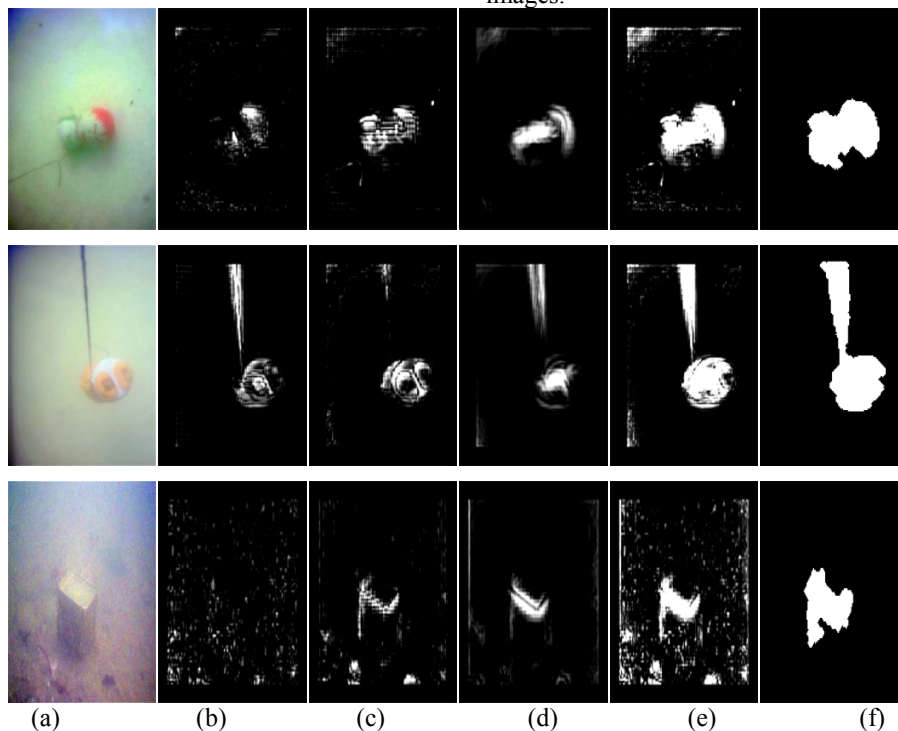


Figure 4. The Achievement of The Silency Map for image 1, image 2 and image 3. (a) Enhanced Underwater Images. (b) Color Feature Maps. (c) Intensity Feature Maps. (d) Orientation Feature Maps. (e) The Combined Saliency Map. (f) Primary Object Location.

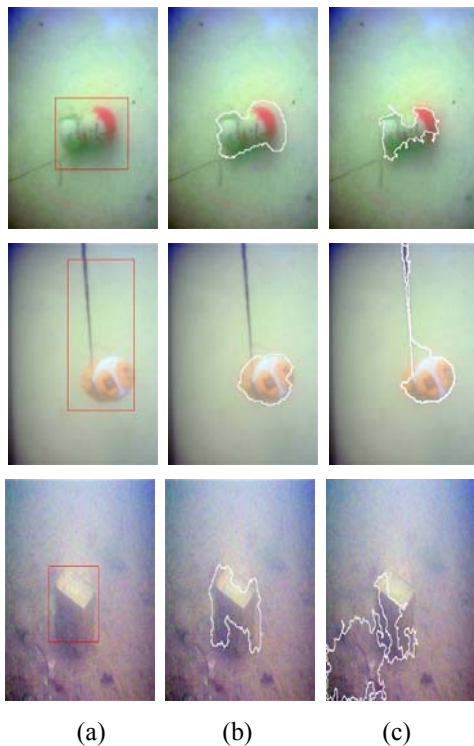


Figure 5. Results of the target extraction. (a) Primary Object Regions. (b) Object Extraction Results by Our Method (c) Edge Detection and Morphological Processed Results of Enhanced Underwater Images.

Comparing with our method, shown in Fig.5(b), the edge detection (with Sobel operator) and morphological processing method is used to process these images and the results are shown in Fig. 5(c).

There are two parameters used to measure the extraction results of images [11], shown in following equations:

$$C_{good} = \frac{card\{\Omega_{in} \cap \Omega_o\}}{card\{\Omega_{in}\}} \quad (23)$$

$$C_{false} = \frac{card\{\Omega_{in} \cap \Omega_b\}}{card\{\Omega_b\}} \quad (24)$$

Where  $\Omega_{in}$  is the extraction object region,  $\Omega_o$  is the real object and  $\Omega_b$  is the background region.  $C_{good}$  is the ratio of extraction region to real object region and  $C_{false}$  is the ratio of false extraction region to background regions. So for the former, a higher value is better; for the later, lower is better. The compare results of these methods are shown in Table II.

From the results shown in Table II, our method have better performance in object extraction than primary object detect and edge detectoin methods. The more precise object regions can be obtained with our method, when the  $C_{good}$  are improved significantly and the  $C_{false}$  are reduced to a very small value. It is proved that our method is feasible, effective and sufficiently accurate in color object extractions.

TABLE II. Compare of Extraction Results with Different Methods

	Image 1		Image 2		Image 3	
	$C_{good}$	$C_{false}$	$C_{good}$	$C_{false}$	$C_{good}$	$C_{false}$
<b>Primary Object Extraction</b>	0.8480	0.0127	0.8786	0.0608	0.6515	0.0213
<b>Edge Detection and Morphological Processing</b>	0.8551	0.0112	0.8477	0.0715	0.6413	0.5212
<b>Our Method</b>	0.9246	0.0675	0.9163	0.0217	0.8574	0.0251

## VI. CONCLUSION

The theoretical and experimental results show that the proposed method can extract underwater objects effectively, and the more accurate object regoins can be obtained with weights redesigned and adaptive saliency map rebuilt, which is very useful for further object recognition. In this paper, the single object detection with low visibility in underwater environment of color images is studied, and multi-object extraction under complex background will be our study subjects in the future.

## REFERENCES

- [1] P. Meer and B. Georgescu. "Edge Detection with Embedded Confidence," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, Dec. 2001, pp. 1351-1365, doi: 10.1109/34.977560.
- [2] L. Garcia Ugarriza, E. Saber, et al. "Automatic Image Segmentation by Dynamic Region Growth and Multiresolution Merging," IEEE Transactions on Image Processing, vol. 18, Oct. 2009, pp. 2275-2288, doi: 10.1109/TIP.2025555.
- [3] Y. Boykov, G. Funka-lea. "Graph Cuts and Efficient N-D Image Segmentation," International Journal of Computer Vision, vol. 70, Nov. 2006, pp. 109-131, doi: 10.1007/s11263-006-7934-5.
- [4] R. Schettini, S. Corchs. "Underwater Image Processing: State of the Art of Restoration and Image Enhancement Methods," EURASIP Journal on Advances in Signal Processing. 2010, doi: 10.1155/2010/746052.
- [5] Tiedong Z, Lei W, Zaibai Q, Yue M. "Underwater Image Detection Based on the Discrete Fractional Brownian Random Field," Opto-Electronic Engineering, vol. 38, no. 8, pp. 41-46, Aug. 2008.
- [6] Lixian L, Shuhong J, Tao C, "Detection and Recognition of Underwater Target Based on Feature Matching," Modern Electronics Technique, vol. 34, no. 4, pp. 73-76, Feb. 2011.
- [7] L. Itti, C. Koch, E. Niebur. "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, pp. 1254-1259, Nov. 1988, doi:10.1109/34.730558.
- [8] D. Walther, D. Edgington, C. Koch. "Detection and Tracking of Objects in Underwater Video," In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 544-549, 2004.
- [9] Z. Rahman, D. Jobson, G. Woodell. "Resiliency of the Multiscale Retinex Image Enhancement Algorithm," In Proceedings of the IS&T Sixth Annual Color Imaging Conference: Color Science, Systems and Applications, pp. 129-134, Nov. 1998.
- [10] D. Comaniciu, V. Ramesh, P. Meer. "Real-time Tracking of Non-rigid Objects Using Mean Shift," IEEE Conference on Computer Vision and Pattern Recognition, pp. 142-149, Jun. 2000.
- [11] C. Barat, R. Phlypo. "A Fully Automated Method to Detect and Segment a Manufactured Object in an Underwater Color Image," EURASIP Journal on Advances in Signal Processing, 2010, doi:10.1155/2010/568092.