

Research Article

Optical Prior-Based Underwater Object Detection with Active Imaging

Jie Shen ¹, Zhenxin Xu,¹ Zhe Chen,^{1,2} Huibin Wang,¹ and Xiaotao Shi ²

¹College of Computer and Information, Hohai University, Nanjing 210098, China

²Hubei International Science and Technology Cooperation Base of Fish Passage, China Three Gorges University, Yichang 443002, China

Correspondence should be addressed to Xiaotao Shi; fishlab@163.com

Received 15 December 2020; Revised 29 March 2021; Accepted 17 April 2021; Published 27 April 2021

Academic Editor: Ning Cai

Copyright © 2021 Jie Shen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Underwater object detection plays an important role in research and practice, as it provides condensed and informative content that represents underwater objects. However, detecting objects from underwater images is challenging because underwater environments significantly degenerate image quality and distort the contrast between the object and background. To address this problem, this paper proposes an optical prior-based underwater object detection approach that takes advantage of optical principles to identify optical collimation over underwater images, providing valuable guidance for extracting object features. Unlike data-driven knowledge, the prior in our method is independent of training samples. The fundamental novelty of our approach lies in the integration of an image prior and the object detection task. This novelty is fundamental to the satisfying performance of our approach in underwater environments, which is demonstrated through comparisons with state-of-the-art object detection methods.

1. Introduction

Underwater object detection from images plays an important role in underwater scientific research and practical projects, providing condensed and informative content regarding underwater objects [1, 2]. However, underwater object detection is difficult owing to multiple factors. First, in contrast to objects on the ground, underwater objects are characterized by their streamlined contours and deceptive appearances. Commonly used image features (i.e., corners and color histograms) can hardly recognize them. Owing to the variety of underwater objects, there are only few carefully designed templates available. Second, challenging underwater conditions pose difficulties from two aspects: (i) the wavelength-selective attenuation reduces and distorts the contrast between the object and background [3] and (ii) the bias illumination and haze effects may generate false positive appearances [4]. These challenges together degrade the performance of underwater object detection significantly [5–7].

In general, there are two main approaches to underwater object detection. First, previous background models are updated for underwater tasks. However, this approach is problematic in theory, as the underlying mechanisms of previous models cannot possibly adapt to underwater scenes, specifically to noise distribution and object characteristics. Second, various learning methods are introduced for underwater object detection. A practical issue concerning these models is the small database available for underwater model training, which may result in errors in model convergence.

To solve the aforementioned problems, this paper proposes a novel foreground feature-based approach that detects objects by identifying and propagating features. Specifically, our approach develops a novel optical prior to locate the coarse region of objects, focusing on the feature seeds inside the object region. The complete object region can be extracted by propagating the seeds. The main contributions of this work are as follows. (i) Our approach integrates underwater active imaging principles and object

detection tasks, providing a novel strategy for underwater object detection. (ii) Our approach detects underwater objects by extracting and propagating feature seeds, demonstrating superior performance in terms of accuracy and generalization.

The remainder of this paper is organized as follows. Section 2 reviews the related literature and discusses the theoretical methodology of our approach. Section 3 presents the framework of our approach. Section 4 presents our novel optical prior model for underwater object detection. The seed selection process is described in Section 5. Section 6 presents the seed propagation process. The experimental results and analysis are discussed in Section 7. Finally, Section 8 concludes the paper.

2. Related Works

This section briefly reviews state-of-the-art research related to underwater object detection and theoretically discusses the novelty of our approach in comparison to these counterparts.

2.1. Object Detection from Underwater Images. Existing underwater object detection methods can be generally classified into two categories based on their fundamental strategies, that is, foreground feature extraction and background modeling.

In general, foreground feature extraction-based methods are applicable to predesigned objects with distinguishable appearances. For these objects of interest, carefully designed templates are available, and thus the detection can be realized by matching templates. For example, a color template was designed to detect light signs for underwater navigation [8]. An intensity template was established to detect the LED ring, guiding the docking position of the autonomous underwater vehicle (AUV) [9]. Moreover, different types of morphological templates have been developed to detect objects with distinguishable structures. For example, scale and rotationally invariant features were extracted for detecting man-made landmarks [10]. In terms of detecting underwater man-made landmarks, experimental results have shown that the shape feature is more stable than the color feature [11]. However, this benchmark only includes dock-mark samples while overlooking various natural objects under water. In addition to these man-made templates, many templates are automatically learned from training data. For example, the correspondence between raw images and their corrected results was modeled to restore the original radiation of underwater scenes, which can present the intrinsic contrast between an underwater object and its background [12]. As an extension to this method, an updated model was proposed by combining a number of low-complexity but moderately accurate color feature detectors [13, 14]. Moreover, saliency detection methods have been introduced to initially identify the region of underwater objects [15–19]. The problem is obvious, as common image features are possibly nonsalient owing to the light attenuation effect.

As mentioned above, underwater scenes are characterized by variability and dynamism. This degrades the background models significantly, and a few methods have been proposed for this purpose. For example, the average number of consecutive frames was calculated as the underwater background [20]. Moreover, the spatiocontextual Gaussian mixture model was used to estimate spatial and temporal correlations. A similar strategy was used in recent research wherein a probabilistic background was established for real-time fish detection [21].

In recent studies, various deep learning networks have been proven to be successful for object detection on the ground [22]. However, using this framework is relatively problematic for underwater tasks because of the small database under water. As a result, the performance of deep learning networks in underwater environments is insufficiently demonstrated [23].

2.2. Theoretical Discussion. In general, the main problems in current underwater detection result from challenging imaging environments. As a result, underwater images are significantly distorted and attenuated. The commonly used methods are significantly degraded in terms of their accuracy and generalization. This problem may be solved by learning strategies such as deep learning architectures. However, the small database and differences among underwater samples reduce the feasibility of learning strategies.

Current problems in underwater object detection have motivated our study. Our novel approach, which is different from other approaches, detects objects by using foreground features. In this study, foreground feature extraction is the key. To correctly identify object features, a novel underwater optical prior was modeled. This type of optical prior takes advantage of the optical principles in underwater imaging; specifically, it assumes that the optical collimation of active imaging necessarily overlaps with object bodies. Object features can be identified with the guidance of optical collimation. A complete object was extracted by propagating the object feature seeds.

In theory, this study develops a novel approach for underwater object detection, which extracts object features with the guidance of an optical prior [24]. This idea is beneficial to underwater image processing and requires further development in the future. Moreover, from the aspect of information fusion, the proposed approach can be understood as the fusion of external and internal features, such as optical and image features.

3. Framework of the Proposed Approach

The framework of the proposed approach is illustrated in Figure 1. In general, the framework consists of two main phases: feature seed extraction and seed propagation. In the first phase, the optical prior is combined with various image features, such as color, texture, and contrast, to extract the most distinguishable seeds of underwater objects. In the second phase, the selected seeds are propagated by the

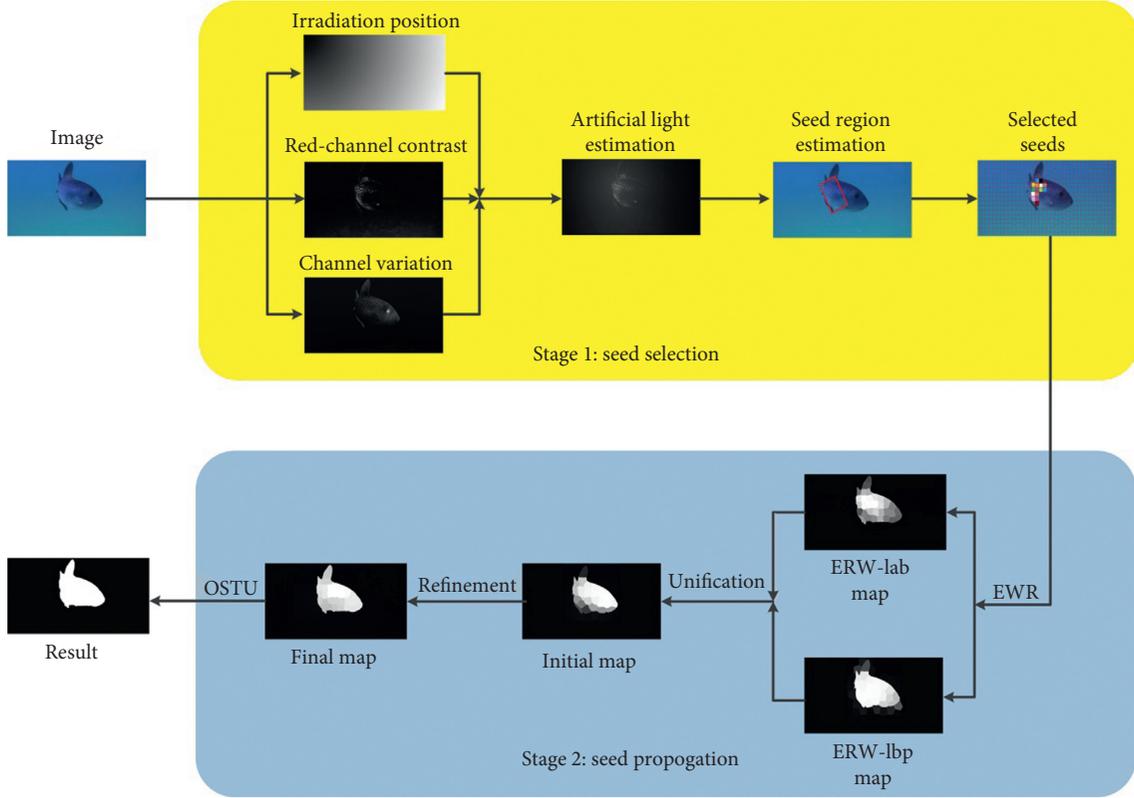


FIGURE 1: Framework of our proposed approach.

extended random walk (ERW) algorithm to identify the complete region of the underwater objects [25].

4. Underwater Optical Prior

Active imaging technologies have been widely used to compensate for light attenuation in underwater environments. However, artificial power causes an inhomogeneous distribution of the scene light. This effect was considered a drawback in previous studies; however, this optical inhomogeneity is regarded as a novel valuable cue for identifying underwater objects in this study. According to underwater active imaging principles, an artificial source is necessarily collimated at the object of interest to enhance visibility. As a result, this type of optical collimation can help identify the region of underwater objects.

Regarding underwater active imaging (see Figure 2), the light irradiance along the optical axis (line AB in Figure 2) decreases with transmission distance [26].

$$E(z, \lambda) = E_0 \frac{1}{(z+a)^2} e^{-\alpha(\lambda)z}, \quad (1)$$

where z is the distance from a point on the optical axis to the light source, λ is the light wavelength, and E_0 is the initial irradiance. The quadratic term $1/(z+a)^2$ measures the change in irradiance. The exponential term $e^{-\alpha(\lambda)z}$ accounts for the light attenuation with the attenuation coefficient $\alpha(\lambda)$.

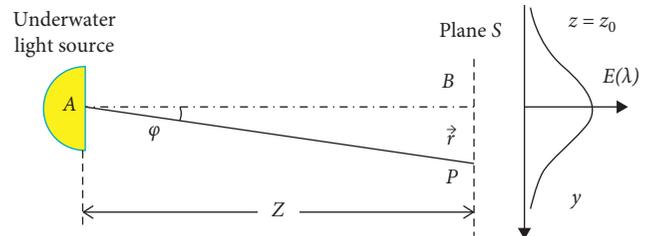


FIGURE 2: Underwater light field distribution. Plane S is perpendicular to the optical axis of the underwater light source, the distance between them being z . The distance between points P and B is denoted as \vec{r} . For any perpendicular plane located at a distance $z = z_0$ from the underwater light source, the irradiance $E(\lambda)$ at any point generally decreases with its distance to the optical axis.

Consider a point P in an underwater object, and the irradiance at P (denoted as $E_P(\lambda)$) can be represented as

$$E_P(\lambda) = E(z, \lambda) f(\vec{r}), \quad (2)$$

where $f(\vec{r})$ models the distribution of irradiance in plane S perpendicular to the optical axis and the vector \vec{r} denotes the coordinates of point P in the plane. Considering that the light field is generally symmetric around the optical axis, the function $f(\vec{r})$ can be approximated using a Gaussian model. According to the optical model (Equations (1)–(2)), the power of the artificial source is maximized at the object and then gradually decreases. Thus, we have more opportunities to identify the object features along the optical collimation. Generally, the optical collimation of a light

source can be identified by the intensity feature. This is the case for natural scenes covered by parallel ambient light; however, this is not true for most underwater conditions because of the complex light components. As a result, the region with the largest intensity may be distant from the optical collimation. To solve this problem, a series of optical principles is proposed in this study, to comprehensively model the optical prior.

- (i) According to the underwater active imaging model (Equation (2)), the irradiation of any point in the optical collimation region depends on the radius of the collimation center.
- (ii) According to the Beer–Lambert law, the wavelength-selective attenuation is reduced in the optical collimation region for a short sight distance, such that the color-channel variation is low, and the red channel is more obvious in the optical collimation region.

These two optical principles can be analyzed with three mathematical computations.

- (i) Irradiation-position relation in the local region: this measures the relationship between the irradiation of any points and their distance from the point with the maximum intensity in the local region.

$$D_x^d = \exp_{m \in \Omega_x}(D(x, m)) = \exp\left(\sqrt{(\xi_1 - \xi_2)^2 + (\gamma_1 - \gamma_2)^2}\right), \quad (3)$$

where $D(x, m)$ is the Euclidean distance from point x to point m that has the maximum intensity in the local region, Ω_x , centered at point x and (ξ_1, γ_1) and (ξ_2, γ_2) are the coordinates of points x and m , respectively.

- (ii) Red-channel contrast: this measures the irradiation difference in the red channel between the optical collimation region and outside regions.

$$C_x^r = \sum_{\forall y \in N} C(I_x^r, I_y^r) = \sum_{\forall y \in N} I_x^r - I_y^r, \quad (4)$$

where $C(I_x^r, I_y^r)$ is the absolute intensity difference between points x and y in the red channel and the superscript r denotes the red channel.

- (iii) Channel variation: this measures the variation between color channels.

$$V_x^c = \sqrt{\sum V(I_x^c, I_x^i)} = \sqrt{(I_x^r - I_x^i)^2 + (I_x^g - I_x^i)^2 + (I_x^b - I_x^i)^2}, \quad (5)$$

where $V(I_x^c, I_x^i)$ is the channel variance of point x in the $r * g * b$ color space. Herein, the Euclidean distance is used to measure the difference.

According to these three measurements, the optical collimation can be identified at points such that their irradiation distribution is of a Gaussian model, they must have a large contrast against other points in the red channel, and they should be homogeneous among color channels.

This can be mathematically modeled using three types of relationships.

- (i) Positive relationship between the irradiation and position relation and channel variation: Inside the optical collimation region, the points with a lower channel variation must be spatially closer to the point with the maximum irradiation in local regions, and vice versa.

$$\text{corr2}(D_x^d, V_x^c), \quad (6)$$

where corr2 is the two-dimensional correlation operator between two matrices.

- (ii) Inverse relationship between the red channel contrast and irradiation–position relation: Inside the optical collimation region, the points with a higher red contrast must be spatially closer to the point with the maximum irradiation in local regions, and vice versa.

$$\text{corr2}(D_x^d, 1 - C_x^r). \quad (7)$$

- (iii) Inverse relationship between the contrast of the red channel and channel variation: Inside the optical collimation region, the points with a higher red contrast must be lower in the channel variation, and vice versa.

$$\text{corr2}(V_x^c, 1 - C_x^r). \quad (8)$$

Inside the optical collimation region, all of the above three two-dimensional correlations should be maintained at a high level.

$$S_x = \text{corr2}(D_x^d, V_x^c) \text{corr2}(D_x^d, 1 - C_x^r) \text{corr2}(V_x^c, 1 - C_x^r). \quad (9)$$

The results for S are shown in Figure 3. From Figure 3, we can observe that the S values inside the optical collimation are significantly larger than the values outside. Hence, using S values initially can help identify underwater object regions. This can be deemed the external optical prior for underwater object detection, as it does not consider any object-specific features.

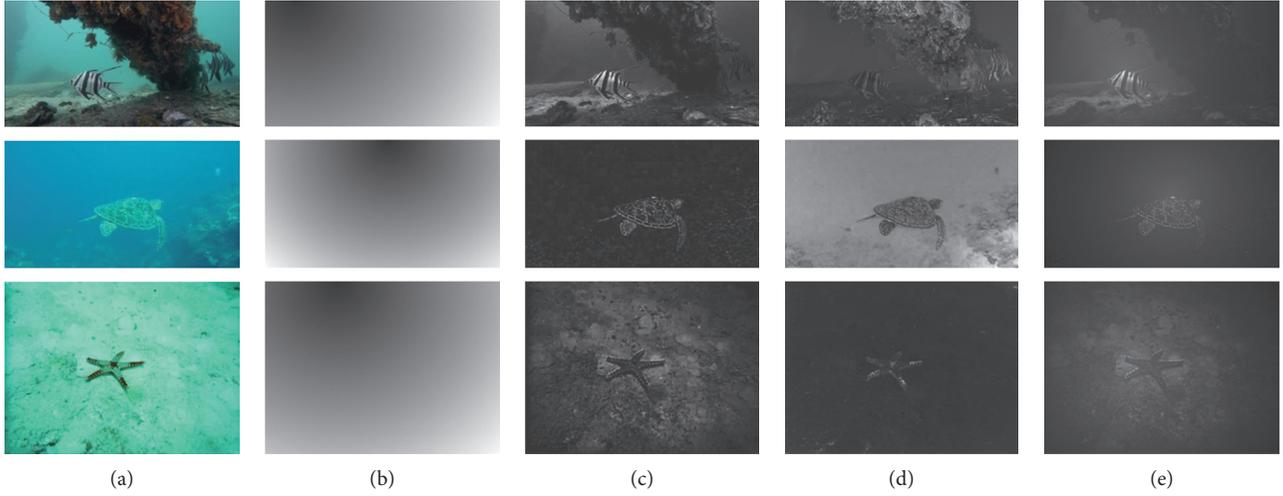


FIGURE 3: Optical collimation. (a) Original image. (b) Irradiation-position relation. (c) Red-channel contrast. (d) Channel variation. (e) Map of S .

5. Seed Selection Refining by Optical Prior and Image Features

After modeling the optical prior, feature seeds are refined using image features, that is, textural and contrast features. Accordingly, the underwater image is transferred from the pixel level to the superpixel level. The edge density in each superpixel was calculated as a textual feature [27].

$$\xi_i = \frac{|\ell_i|}{|N_i|}, \quad (10)$$

where ℓ_i is the total length of the edges in the i -th superpixel and N_i is the size of the i -th superpixel.

$$|\ell_i| = C\{OE, TG\},$$

$$OE_{\theta,s} = (I * f_{\theta,s}^e)^2 + (I * f_{\theta,s}^o)^2, \quad (11)$$

$$TG = \chi^2(g, h) = \frac{1}{2} \sum_i \frac{(g(i) - h(i))^2}{g(i) + h(i)},$$

where OE is the oriented energy for detecting and localizing the composite edges, TG denotes the texture gradient, and C is a classifier for combining cues. $f_{\theta,s}^e$ and $f_{\theta,s}^o$ are quadrature pairs of even- and odd-symmetric filters at the orientation θ and scale s . g and h are two half-disc histograms.

The contrast feature of any point denotes their difference from the average of the background points where $S_x < T$.

$$\lambda_x^c = \sum \gamma(I_x^c, I_{ave}^c) = \sum ((I_x^c(r) - I_{ave}^c(r))^2 + (I_x^c(g) - I_{ave}^c(g))^2 + (I_x^c(b) - I_{ave}^c(b))^2)^{1/2}, \quad (12)$$

where $\gamma(I_x^c, I_{ave}^c)$ is the difference between point x and the average background.

$$\begin{aligned} I_{ave}^c &= [I_{ave}^c(r), I_{ave}^c(g), I_{ave}^c(b)] \\ &= \frac{\sum_x [I_x^c(r), I_x^c(g), I_x^c(b)]}{\zeta} \text{ such that } S_x < T, \end{aligned} \quad (13)$$

where ζ is the number of the background points with $S_x < T$.

The comprehensive evidence for identifying seeds can be formulated from multiple cues.

$$\phi_i = S_i \xi_i \lambda_i, \quad (14)$$

where

$$\begin{aligned} S_i &= \sum_{x \in SP_i} S_x, \\ \lambda_i &= \sum_{x \in SP_i} \lambda_x, \end{aligned} \quad (15)$$

where SP_i is the i -th superpixel. The superpixels with the largest k value of ϕ_i were selected as the seeds.

6. Feature Seed Propagation

In our approach, underwater objects were detected by propagating the selected seeds. Specifically, the ERW algorithm is opted as it can propagate seeds to more distant areas, thereby preventing false negatives in object regions [25].

6.1. Graph Construction. We initially measured the correlation between superpixels prior to seed propagation. In this study, the superpixel-to-superpixel correlation is modeled by an undirected graph of the input image, $G = (V, E)$. Here, V is the node set including all superpixels over the image, $V = \{SP_1, SP_2, \dots, SP_N\}$, and E is the connection between any two nodes. The similarity between the nodes is measured by the weight matrix $W = H^{N \times N}$, and each element of W is defined as follows:

$$w_{ij} = \exp\left(\frac{k(SP_i) - k(SP_j)_2}{2\sigma^2}\right), \quad (16)$$

$$\arg \min_f \left\{ \omega \sum_{i=1}^m \left(d_{SP_i} f_i - \sum_{j \in C_i} w_{ij} f_j \right) + \frac{1}{2} \sum_{i,j} w_{ij} (f_i - f_j)^2 + \lambda \sum_{i=1}^m (f_i - y_i) \right\}, \quad (17)$$

where the first term is a second Laplacian term aiming to propagate saliency to distant nodes and f_i denotes the label of the node SP_i ; therefore, when $f_i = 1$, it denotes the seed, and in contrast, when $f_i = 0$, it denotes other places, and C_i is the node set centered at the superpixel SP_i . The second term is the standard random walk formula, the third term ensures the correctness of the seeds, and y_i denotes the output of an external seed classifier. Parameters ω and λ are trade-off parameters.

6.3. Combination. The features of any nodes $k(SP_i)$ are characterized by their color and texture information. Complementation between these two features can enhance the reliability of object detection. For example, the color feature can better identify parts with distinguishable colors (see Figure 4(a)). Alternatively, the texture feature plays a more important role in identifying textural parts (see Figure 4(b)). The mean CIElab color and local binary pattern (LBP) features were utilized and separately propagated within their own feature spaces. Therefore, two maps can be obtained, and their intersection is regarded as the result of underwater object detection.

$$R = R_{lab} \times R_{lbp}, \quad (18)$$

where R_{lab} and R_{lbp} denote the maps generated by the CIElab and LBP features, respectively, and R represents the final detection result (see Figure 4(c)).

It is important to note that the above calculations were all processed at the superpixel level, which consequently reduces the accuracy of the object detection results. A diffusion process was added to refine the detection results. The superpixels over the detection results are clustered into k partitions, and r superpixels are included in each cluster. In each partition, the value of the i -th superpixel is smoothed by the average of the remaining superpixels.

where $k(SP_i)$ denotes the feature extracted in superpixel SP_i and σ is the controlling constant. The degree of any node is defined as the sum of edges connected to them: $d_{SP_i} = \sum_{j=1}^N w_{ij}$. The degree matrix is defined as $M = \text{diag}\{d_{SP_1}, d_{SP_2}, \dots, d_{SP_N}\}$. The Laplace matrix was generated by $L = M - W$.

6.2. Extended Random Walking. According to ERW [25], seed propagation is realized by minimizing the following energy function:

$$\overline{\mathfrak{R}}_i^{(M)} = \beta_1 \sum_{j \neq i, j=1}^r \theta_{ij} \mathfrak{R}_i^{(M)} + \beta_2 \mathfrak{R}_i^{(M)}, \quad (19)$$

$$\theta_{ij} = \frac{1 - (1/\psi(l_i, l_j))}{\sum_{j=1}^r (1 - (1/\psi(l_i, l_j)))},$$

where $\overline{\mathfrak{R}}_i^{(M)}$ and $\mathfrak{R}_i^{(M)}$ are the refined and original values at the i -th superpixel, respectively, θ_{ij} denotes the color feature, and β_1 and β_2 are the balance parameters. A comparison between the original and refined results is shown in Figure 4(d). As shown in Figure 4(d), the effect of the diffusion process can be clearly demonstrated; after diffusion, the detection results become more homogeneous in the underwater object region, while the contrast between the object and background is enlarged. The binary result can be obtained using the thresholding method (see Figure 4(e)) [28].

6.4. Pseudocode for Underwater Object Detection. In general, our approach detects underwater objects through three successive phases: feature seed selection, propagation, and combination. The step-by-step pseudocode details are listed in Algorithm 1.

7. Experimental Evaluation and Analysis

Both qualitative and quantitative evaluations are proposed in this section to demonstrate the performance of our underwater object detection method. We first present the results of our novel seed selection for diverse samples, as it is the key and fundamental novelty of our approach. Then, our underwater object detection results are presented and compared to other typical object detection methods. Our approach can efficiently detect underwater objects in underwater images. This aspect is shared by comparing methods that include two classic image segmentation methods (PCNN [29] and graph-based segmentation methods [30]) and eight saliency-based

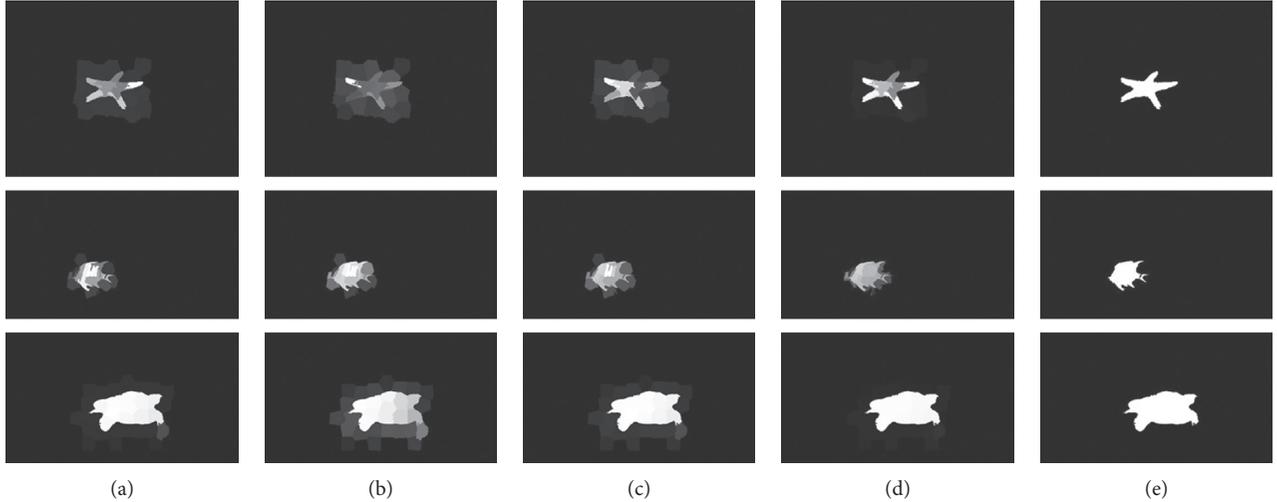


FIGURE 4: Seed propagation for the samples in Figure 3. (a-b) Detection results propagated in Lab and LBP spaces. (c) Initial detection result. (d) Refined result. (e) Binary result.

Input: source image X (size n), T , ω , λ , σ , β_1 , β_2 .
Output: the object region
For n iterations **do**
 Feature seed extraction
 $S_x \text{corr2}(D_x^d, V_x^c) \text{corr2}(D_x^d, 1 - C_x^r) \text{corr2}(V_x^c, 1 - C_x^r), S_i \sum_{x \in \text{SP}_i} S_x$
 $\lambda_x^c \sum \gamma(I_x^c, I_{\text{ave}}^c) = \sum ((I_x^c(r) - I_{\text{ave}}^c(r))^2 + (I_x^c(g) - I_{\text{ave}}^c(g))^2 + (I_x^c(b) - I_{\text{ave}}^c(b))^2)^{1/2}, \lambda_i \sum_{x \in \text{SP}_i} \lambda_x$
 $\xi_i (|\ell_i|/|N_i|)$
 $\phi_i S_i \xi_i \lambda_i$
 Find k largest ϕ_i to identify the feature seeds
 Feature seed propagation
 Generate detection results R_{lab} and R_{lbp} in CIELab and LBP feature spaces by graph optimization
 $w_{ij} \exp((k(\text{SP}_i) - k(\text{SP}_j)^2)/2\sigma^2)$
 $\text{argmin}_f \{ \omega \sum_{i=1}^m d_{\text{SP}_i} f_i - \sum_{j \in C_i} w_{ij} f_j + 1/2 \sum_{i,j} w_{ij} (f_i - f_j)^2 + \lambda \sum_{i=1}^m (f_i - y_i) \}$
 End for
 Result combination
 $RR_{\text{lab}} \times R_{\text{lbp}}$
Return The binary map R for object detection

ALGORITHM 1: Optical prior-based underwater object detection.

methods (Itti [31], HFT [32], MR [33], GBVS [34], PBS [35], SF [36], wCtr* [37], and RRWR [38]). The code for the baseline methods was downloaded from the websites provided by the authors, and default parameters were used. Here, we did not attempt to adapt the model parameters to special underwater data. There are three possible reasons for this. First, all compared methods can work in a fully automatic manner, while the parameter setting plays a minor role in the detection performance [39]. Moreover, parameter setting is unnecessary for most compared methods, which is only required for PCNN and graph-based methods. Second, unlike the images on the ground, underwater images are characterized by large intervariability. Taking the images in Figure 3 as examples, we can find that not only the objects but also the background hues are significantly changed between

scenes. Therefore, it is unreasonable to adjust the parameters to all possible underwater scenes, as well as special samples. Hence, we used unified parameters in all experiments when detecting underwater objects. Finally, experimental comparisons in this section are intended to verify the underlying mechanisms of different methods, such as feature modeling and classification. The theoretical mechanism difference between the methods determines their performance variation in underwater object detection. Further, carefully moderating each parameter is partially problematic, which can be deemed as an overfitting issue for underwater object detection. To ensure a fair performance evaluation, the parameters in our approach are fixed for different samples without any adaptation process. In contrast to the model learning strategy, our approach and the compared methods can

detect underwater objects without additional learning processes.

7.1. Dataset and Experimental Setup. Underwater images available on YouTube were collected to establish a benchmark for experimental evaluations [40–48]. More than 200 images and 50 scenes were included in our underwater benchmarks. The object region of each image was labeled by 10 volunteers, and the average results were deemed the ground truth. Images in this dataset differ in several aspects, such as participants, context, and ambient light. The diversity of these images can ensure a comprehensive and fair evaluation of different aspects, such as background removal, noise inhibition, and object identification. The quantitative performance of the object detection is evaluated with respect to six criteria [49]: precision (Pr), similarity (Sim), true positive rate (TPR), F-score (Fs), false positive rate (FPR), and percentage of wrong classifications (PWC).

$$\begin{aligned}
 \text{Pr} &= \frac{\text{tp}}{\text{tp} + \text{fp}}, \\
 \text{TPR} &= \frac{\text{tp}}{\text{tp} + \text{fn}}, \\
 \text{Fs} &= 2 \times \frac{\text{Pr} \times \text{TPR}}{\text{Pr} + \text{TPR}}, \\
 \text{Sim} &= \frac{\text{tp}}{\text{tp} + \text{fp} + \text{fn}}, \\
 \text{FPR} &= \frac{\text{fp}}{\text{fp} + \text{tn}}, \\
 \text{PWC} &= 100 \times \frac{\text{fn} + \text{fp}}{\text{tp} + \text{tn} + \text{fp} + \text{fn}}.
 \end{aligned} \tag{20}$$

where tp, tn, fp, and fn denote the scores of the true positive, true negative, false positive, and false negative, respectively. The parameter tp is evaluated by the number of pixels that belong to the object in both the detection results and the ground truth. The parameter tn is the number of pixels that belong to the background in both the detection results and the ground truth. The parameter fp counts the number of pixels that belong to the background in the ground truth but are mistaken as the object in the detection result. The parameter fn is the number of pixels that are the object in the ground truth but are mistaken as the background in the detection results. In each of these experiments, we maintained the resolution of all the inputs as the original resolution. Additionally, we kept the tunable parameters in our approach fixed in different experiments as follows: $T = 0.25$, $\omega = 0.1$, $\lambda = 0.01$, $\sigma = 0.05$, $\beta_1 = 0.5$, and $\beta_2 = 0.5$, which can ensure the automation of our approach.

7.2. Seed Selection. This section evaluates the performance of our proposed seed selection method, which is the fundamental novelty of our approach, and also aids in validating

our detection results. A qualitative presentation is depicted in Figure 5. From the results in Figure 5, most of the selected seeds were correctly located on the object regions. Given the ground truth for experimental samples, seed selection results can be quantitatively evaluated by the TP, FP, and Pr scores, which comprehensively calculate the ratio of the seeds that were located inside and outside the objects. The quantitative evaluation of our seed selection method with 30 samples is given as follows: TP=0.4486, FP=0.0010, and Pr=0.9416. From these quantitative scores, we can conclude that almost all of the selected seeds were correctly located inside the object regions. This is a prerequisite for the good performance of underwater object detection.

7.3. Underwater Object Detection. Experimental evaluations are realized by comparing our approach to state-of-the-art object and saliency detection methods. Figure 6 presents the experimental results for twelve scenes with diverse background and object categories. The first column in Figure 6 presents the original images, the second column shows the ground truth, the third to fourth columns show the results of PCNN and graph-based segmentation methods [29, 30], the fifth to twelfth columns present the saliency-based segmentation methods [31–38], and the last column shows the results obtained by our approach.

According to the results in Figure 6, different methods demonstrate discrepant performance in various scenes. In contrast to the compared methods, our approach presents the most consistent performance that our approach can robustly segment different underwater objects under all conditions. Overall, the saliency-based methods performed better than the image segmentation methods in the underwater object detection experiments. Other segmentation methods can only segment images into homogeneous partitions but cannot identify the object regions. The most comparable performance to ours is obtained by SF, which identifies objects using the contrast feature that is also considered in our approach. This similarity between our approach and SF proves the importance of the contrast feature for object identification in underwater environments.

Table 1 summarizes the average performance of the different methods. The scores in Table 1 present the average results for the 30 samples. For Pr, TPR, Fs, Sim, and C, larger values indicate better performance, while lower values for the FPR and PWC show better results. In general, our approach won the best in six criteria and ranked fourth in the TPR criterion. This is because the coarse regions given by the GBVS, Itti, and graph-based methods increase the tp value while reducing the fn value, resulting in a significant increase in the TPR value. By comprehensively understanding all criteria, our method provides the best results in contrast to the other methods. This functional advantage can theoretically be explained from two perspectives. First, image degeneration and distortion will reduce the feature contrast between the object and the

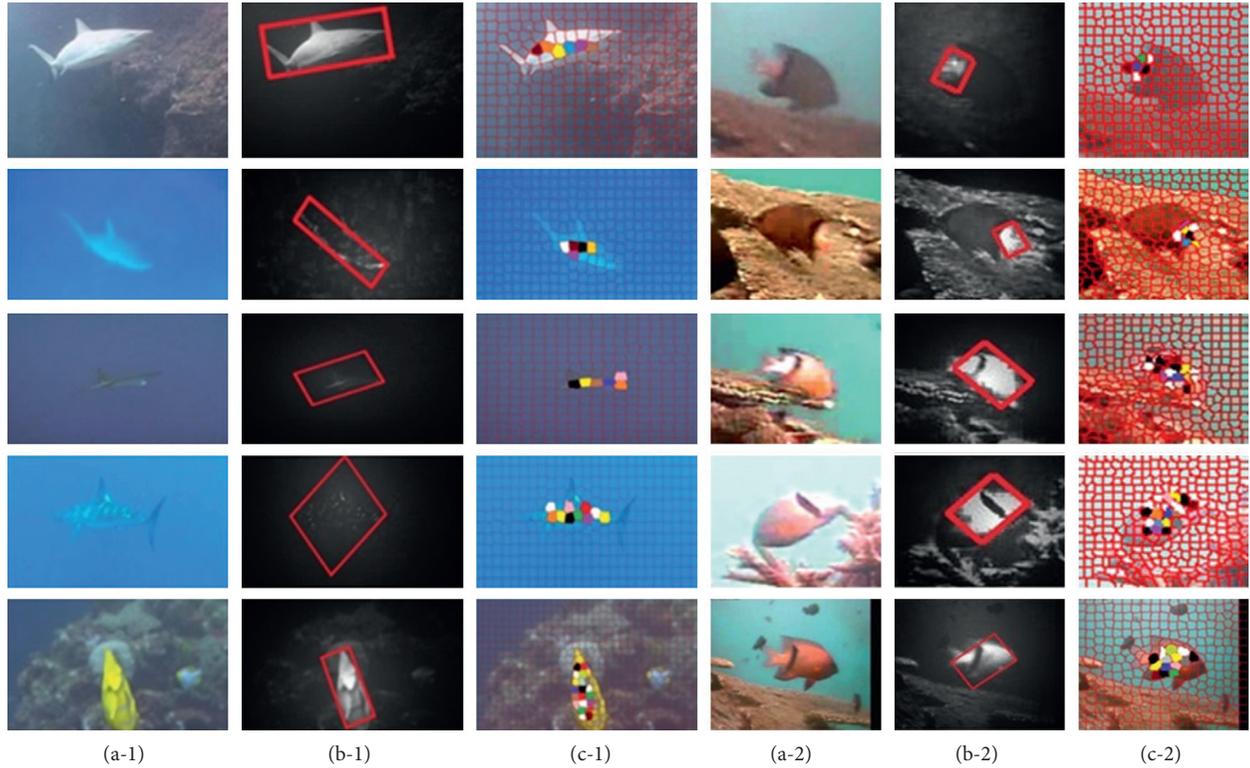


FIGURE 5: Seed selection. (a) Original samples. (b) Optical collimation region. (c) Selected seed (chromatic superpixel).

background, such that the clustering-based method cannot identify the object region, which can be observed by the results generated by the PCNN and graph methods (Figures 6(c) and 6(d) and the first and second rows in Table 1). In contrast, the saliency detection methods have better ability in pop-out region detection, which has more opportunities to identify the objects of interest (Figures 6(e)–6(l) and the third to tenth rows in Table 1). However, the regions given by the saliency detection methods were considerably coarse. This is caused by light scattering in water, where a haze effect is generated around

the objects. This is the reason for the irregular edge expansion of the object detection results. Second, the functional advantages of our approach are presented in Figure 6(m) and the last row in Table 1. These advantages demonstrate the contributions of our novel approach to underwater object detection. The most representative features can be refined using our feature seed selection process. However, these feature seeds are limited and cannot identify the complete region of the objects. This problem is solved by our seed propagation phase, which can extract the complete regions of underwater objects.

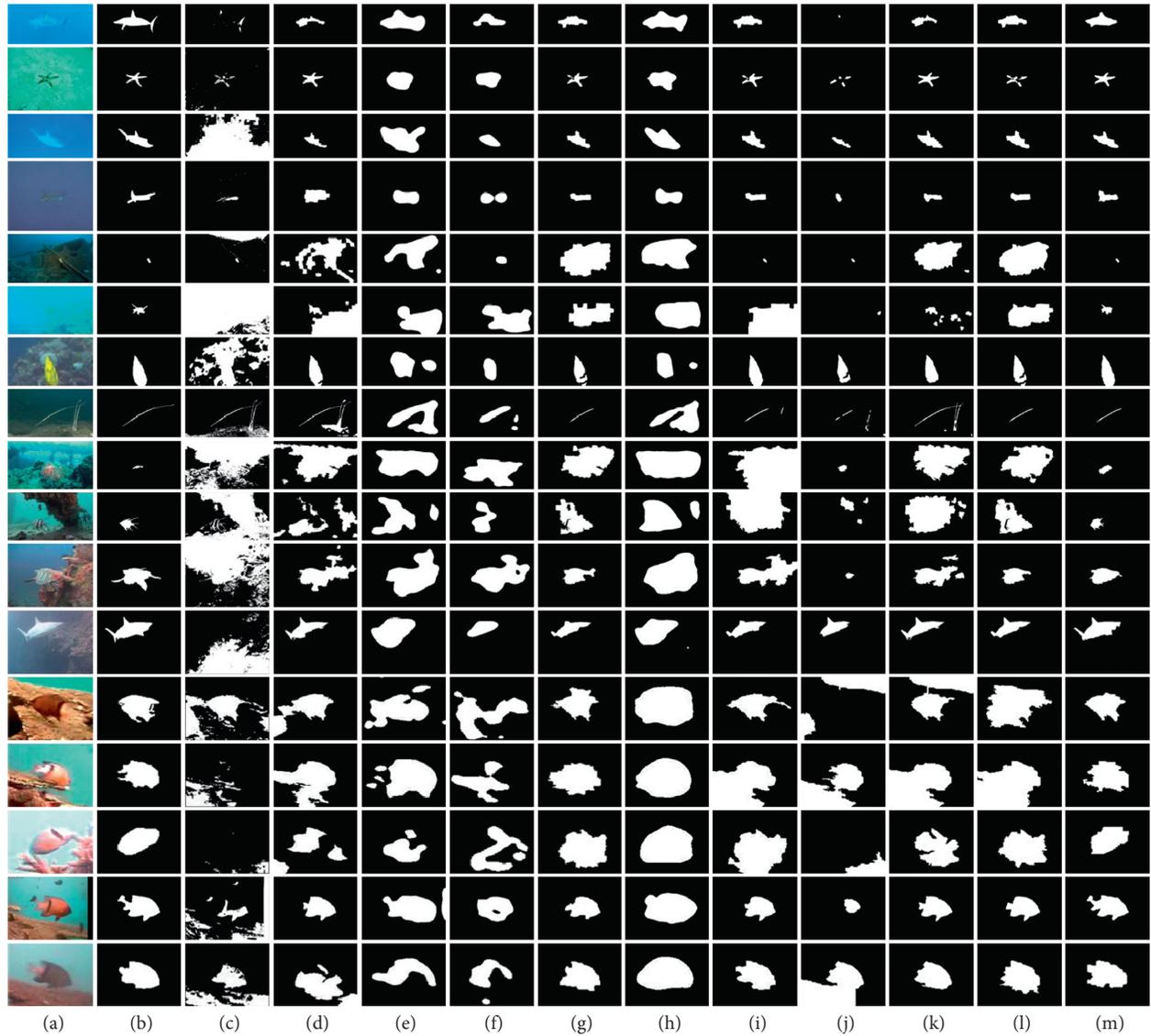


FIGURE 6: Result comparison. (a) Underwater sample. (b) Ground truth. (c) PCNN. (d) Graph. (e) Itti. (f) HFT. (g) MR. (h) GBVS. (i) PBS. (j) SF. (k) wCtr*. (l) RRWR. (m) OURS.

TABLE 1: Quantitative average performance comparison.

Method	Pr	TPR	Fs	Sim	FPR	PWC	C
PCNN	0.3780	0.5969	0.3337	0.2201	0.1465	18.1596	0.2516
Graph	0.4727	0.9122	0.4967	0.3987	0.0936	9.7105	0.4124
Itti	0.2360	0.9239	0.3723	0.2402	0.1037	12.1497	0.2312
HFT	0.5692	0.8073	0.5976	0.4620	0.0578	7.4780	0.4938
MR	0.7534	0.8159	0.7099	0.6073	0.0437	5.7969	0.6475
GBVS	0.3065	0.9127	0.3957	0.3018	0.1076	11.2413	0.3104
PBS	0.6724	0.7912	0.6745	0.5134	0.0805	8.6571	0.5192
SF	0.6893	0.4914	0.5080	0.3988	0.0344	6.7280	0.4277
wCtr*	0.6792	0.8238	0.6653	0.5473	0.0601	7.3151	0.5787
RRWR	0.7305	0.8104	0.6840	0.5766	0.0559	6.8902	0.6142
OURS	0.8885	0.8595	0.8624	0.7725	0.0080	1.7897	0.8135

8. Conclusions

Underwater object detection is difficult owing to the challenging optical environments. Commonly used detection methods can hardly distinguish an object from the scene background, often resulting in a high false positive ratio. To solve this problem, we propose a novel optical prior-based approach for detecting underwater objects. Theoretically, our work develops a novel interdisciplinary idea for underwater object detection that models and introduces an external prior into the detection task. Unlike data-driven strategies, our prior method is independent of the training data and can be directly obtained by optical disciplines. Using this prior method can correctly guide foreground object feature extraction. The object features provide sufficient cues for identifying the object regions. The advantages of the proposed idea have been demonstrated by a variety of underwater samples, where the scene background and objects are different. Moreover, we believe that this new approach will provide valuable inspiration for future computer vision research in water. However, we acknowledge that our approach may fail in some cases owing to the lack of feature seeds. This problem likely occurs when objects have a high-contrast appearance. This problem will be discussed in future work.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported partly by the National Natural Science Foundation of China (nos. 61903124 and 62073120), the Fundamental Research Funds for the Central Universities (nos. B200202186 and B200202181), the National Natural Science Foundation of China (no. 51922065), the Natural Science Foundation of Jiangsu Province (no. BK20201311), the Hubei International Science and Technology Cooperation Base of Fish Passage (no. HIBF2020010), and the Open Research Fund of Jiangxi Provincial Key Laboratory of Water Resources and Environment of Poyang Lake (no. 2020GPSYS01).

References

- [1] M. W. Westneat, "Skull biomechanics and suction feeding in fishes," *Fish Physiology*, vol. 23, pp. 29–75, 2005.
- [2] W. Zhang, X. Pan, X. Xie, L. Li, Z. Wang, and C. Han, "Color correction and adaptive contrast enhancement for underwater image enhancement," *Computers & Electrical Engineering*, vol. 91, p. 106981, 2021.
- [3] J.-C. Zhou, D.-H. Zhang, and W.-S. Zhang, "Classical and state-of-the-art approaches for underwater image defogging: a comprehensive survey," *Frontiers of Information Technology & Electronic Engineering*, vol. 21, no. 12, pp. 1745–1769, 2020.
- [4] S. Kuanar, D. Mahapatra, M. Bilas et al., "Multi-path dilated convolution network for haze and glow removal in nighttime images," *The Visual Computer*, vol. 37, pp. 1–14, 2021.
- [5] J. Zhou, Z. Liu, W. Zhang, D. Zhang, and W. Zhang, "Underwater image restoration based on secondary guided transmission map," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 7771–7788, 2021.
- [6] J. Aguzzi, C. Costa, Y. Fujiwara, R. Iwase, E. Ramirez-Llorda, and P. Menesatti, "A novel morphometry-based protocol of automated video-image analysis for species recognition and activity rhythms monitoring in deep-sea fauna," *Sensors*, vol. 9, no. 11, pp. 8438–8455, 2009.
- [7] M. Moniruzzaman, S. M. S. Islam, M. Bennamoun, and P. Lavery, "Deep learning on underwater marine object detection: a survey," in *Advanced Concepts for Intelligent Vision Systems*, vol. 10617, pp. 150–160, Springer, Cham, Switzerland, 2017.
- [8] S. C. Yu, T. Ura, T. Fujii et al., "Navigation of autonomous underwater vehicles based on artificial underwater landmarks," in *Proceedings of the Conference on An Ocean Odyssey MTS/IEEE Oceans*, vol. 1, pp. 409–416, Honolulu, HI, USA, 2001.
- [9] P. M. Lee, B. H. Jeon, and S. M. Kim, "Visual servoing for underwater docking of an autonomous underwater vehicle with one camera," in *Proceedings of the Celebrating the Past ... Teaming Toward the Future*, vol. 2, pp. 677–682, San Diego, CA, USA, 2003.
- [10] A. Negre, C. Pradalier, and M. Dunbabin, "Robust vision-based underwater target identification and homing using self-similar landmarks," in *Field and Service Robotics*, vol. 42, pp. 51–60, Springer, Berlin, Germany, 2008.
- [11] F. D. Maire, D. Prasser, M. Dunbabin et al., "A vision based target detection system for docking of an autonomous underwater vehicle," in *Proceedings of the 2009 Australasian Conference on Robotics and Automation*, University of Sydney, Sydney, Australia, 2009.
- [12] G. Dudek, M. Jenkin, C. Prahacs et al., "A visually guided swimming robot," in *Proceedings of the 2005 IEEE/RSJ International Conference On Intelligent Robots And Systems*, pp. 3604–3609, Edmonton, Canada, 2005.
- [13] J. Sattar and G. Dudek, "Robust servo-control for underwater robots using banks of visual filters," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 3583–3588, Kobe, Japan, 2009.
- [14] D. Kim, D. Lee, H. Myung, and H.-T. Choi, "Artificial landmark-based underwater localization for AUVs using weighted template matching," *Intelligent Service Robotics*, vol. 7, no. 3, pp. 175–184, 2014.
- [15] D. L. Rizzini, F. Kallasi, F. Oleari et al., "Investigation of vision-based underwater object detection with multiple datasets," *International Journal of Advanced Robotic Systems*, vol. 12, pp. 1–13, 2015.
- [16] D. R. Edgington, K. A. Salamy, M. Risi et al., "Automated event detection in underwater video," in *Proceedings of the Celebrating the Past ... Teaming Toward the Future*, vol. 5, pp. 2749–2753, San Diego, CA, USA, 2003.
- [17] M. C. Chuang, J. N. Hwang, and K. Williams, "A feature learning and object recognition framework for underwater fish images," *IEEE Transactions on Image Processing*, vol. 25, pp. 1862–1872, 2016.
- [18] Y. Zhu, L. Chang, J. Dai et al., "Automatic object detection and segmentation from underwater images via saliency-based

- region merging,” in *Proceedings of the OCEANS 2016*, pp. 1–4, Shanghai, China, 2016.
- [19] J. Li, R. M. Eustice, and M. Johnson-Roberson, “High-level visual features for underwater place recognition,” in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 3652–3659, Seattle, WA, USA, 2015.
- [20] H. Liu, J. Dai, and R. Wang, “Combining background subtraction and three-frame difference to detect moving object from underwater video,” in *Proceedings of the OCEANS*, pp. 1–5, Shanghai, China, 2016.
- [21] A. Salman, S. Maqbool, A. H. Khan, A. Jalal, and F. Shafait, “Real-time fish detection in complex backgrounds using probabilistic background modelling,” *Ecological Informatics*, vol. 51, pp. 44–51, 2019.
- [22] H. Qin, X. Li, J. Liang, Y. Peng, and C. Zhang, “DeepFish: accurate underwater live fish recognition with a deep architecture,” *Neurocomputing*, vol. 187, pp. 49–58, 2016.
- [23] X. Sun, J. Shi, L. Liu et al., “Transferring deep knowledge for object recognition in low-quality underwater videos,” *Neurocomputing*, vol. 275, pp. 897–908, 2018.
- [24] S. Kuanar, V. Athitsos, D. Mahapatra et al., “Low dose abdominal CT image reconstruction: an unsupervised learning based approach,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 1351–1355, Taipei, Taiwan, 2019.
- [25] M. Jian, R. Zhao, X. Sun et al., “Saliency detection based on background seeds by object proposals and extended random walk,” *Journal of Visual Communication and Image Representation*, vol. 57, pp. 202–211, 2018.
- [26] Y. Guo, H. Song, H. Liu et al., “Model-based restoration of underwater spectral images captured with narrowband filters,” *Optics Express*, vol. 24, no. 12, pp. 13101–13120, 2016.
- [27] D. R. Martin, C. C. Fowlkes, J. Malik et al., “Learning to detect natural image boundaries using local brightness, color, and texture cues,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 530–549, 2004.
- [28] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [29] M. M. Subashini and S. K. Sahoo, “Pulse coupled neural networks and its applications,” *Expert Systems with Applications*, vol. 41, pp. 3965–3974, 2014.
- [30] X. Wang, Y. Tang, S. Masnou et al., “A global/local affinity graph for image segmentation,” *IEEE Transactions on Image Processing*, vol. 24, pp. 1399–1411, 2015.
- [31] L. Itti and C. Koch, “Computational modelling of visual attention,” *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.
- [32] J. Li, M. D. Levine, X. An, X. Xu, and H. He, “Visual saliency based on scale-space analysis in the frequency domain,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 996–1010, 2013.
- [33] C. Yang, L. Zhang, H. Lu et al., “Saliency detection via graph-based manifold ranking,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3166–3173, Portland, OR, USA, 2013.
- [34] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” *Advances in Neural Information Processing Systems*, vol. 19, pp. 545–552, 2007.
- [35] C. Yang, L. Zhang, and H. Lu, “Graph-regularized saliency detection with convex-hull-based center prior,” *IEEE Signal Processing Letters*, vol. 20, pp. 637–640, 2013.
- [36] F. Perazzi, P. Krähenbühl, Y. Pritch et al., “Saliency filters: contrast based filtering for salient region detection,” in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 733–740, Providence, RI, USA, 2012.
- [37] W. Zhu, S. Liang, Y. Wei et al., “Saliency optimization from robust background detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2814–2821, Columbus, OH, USA, 2014.
- [38] C. Li, Y. Yuan, W. Cai et al., “Robust saliency detection via regularized random walks ranking,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2710–2717, Boston, MA, USA, 2015.
- [39] H. Chang, F. Zhang, S. Ma, G. Gao, H. Zheng, and Y. Chen, “Unsupervised domain adaptation based on cluster matching and Fisher criterion for image classification,” *Computers & Electrical Engineering*, vol. 91, Article ID 107041, 2021.
- [40] Underwater Vision, 2016, <http://www.youtube.com/user/bubblevision>.
- [41] Underwater Vision, 2008, <https://www.youtube.com/watch?v=P7257ozFHkI>.
- [42] Underwater Vision, 2009, <https://www.youtube.com/watch?v=i1T70Ev2AYs>.
- [43] Underwater Vision, 2010, <https://www.youtube.com/watch?v=Pt4ib8VIFVA>.
- [44] Underwater Vision, 2008, <https://www.youtube.com/watch?v=2AzEeh87Z38>.
- [45] Underwater Vision, 2012, https://www.youtube.com/watch?v=a9_iVF4EA-o.
- [46] Underwater Vision, 2010, <https://www.youtube.com/watch?v=BNq1v6KCANo>.
- [47] Underwater Vision, 2012, https://www.youtube.com/watch?v=kK_hJZo-7-k.
- [48] Underwater Vision, 2010, <https://www.youtube.com/watch?v=NKmc5dlVSRk>.
- [49] A. W. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, “Visual tracking: an experimental survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1442–1468, 2014.