

RESEARCH PAPER

Open Access



Underwater visual feature matching based on attenuation invariance

Jinghao Sun¹, Hui Yu^{2*} , Yuan Rao¹ and Hong Zhang¹

Abstract

Underwater visual feature matching plays an important role in the localization and autonomous motion of underwater robots. However, feature descriptors are prone to change in degraded underwater images due to light attenuation, which leads to image color distortion, blur, and low illumination. Although the recent developments in deep learning have improved the performance of feature descriptors, the degradation of underwater images and the lack of underwater datasets have hindered the development of underwater visual feature matching and localization techniques. Herein, we propose an underwater visual feature matching method based on attenuation invariance. Our approach involves two main components. First, we study the attenuation differences between images to generate feature descriptors that are invariant to attenuation, thereby improving the accuracy of underwater visual feature matching. Second, we construct two datasets: the multiple water type (MWT) dataset, which contains over 30000 underwater images, and the underwater image feature descriptor (UIFD) evaluation dataset. These datasets are developed to compensate for the lack of evaluation datasets for underwater visual feature matching. The code and datasets are available at <https://github.com/Sun-Jinghao/UIFD>.

Keywords Feature matching, Attenuation invariance, Underwater datasets

1 Introduction

Feature matching is an important operation in computer vision and plays a crucial role in underwater localization techniques. This matching operation heavily relies on keypoint descriptors (Dai et al. 2019). There is a considerable difference in the pixel quality between underwater and in-air images (Lin and Liu 2014). Owing to refraction, absorption, and scattering of light by suspended particles in water and attenuation of light transmitted to the camera, raw underwater images are characterized by low contrast, blurred details, and color distortion (Liu et al. 2019a), as shown in Fig. 1.

In recent years, two approaches have been commonly adopted for underwater visual characterization. The first approach involves the direct use of descriptor methods for air images, and the other involves feature extraction and characterization after underwater image enhancement (UIE) and restoration. However, underwater image quality heavily relies on the inherent optical properties of the water body during the imaging process. Traditional methods can easily lead to color distortion when applied to underwater images, wherein the enhanced image appears under-saturated and over-saturated in disparate regions because optical properties specific to underwater images are neglected (Liu et al. 2019a). Compared to the feature matching of air images, feature matching of underwater images involves not only underlying point-of-view variations and light variations but also variations in light attenuation, such as chromatic aberration and blurring (Qi et al. 2021). Existing studies on feature descriptors considered the invariances regarding viewpoint, scale, and illumination, which are variable in

*Correspondence:

Hui Yu
hui.yu@port.ac.uk

¹ Faculty of Information Science and Engineering, Ocean University of China, Qingdao 266100, China

² School of Creative Technologies, University of Portsmouth, Portsmouth PO1 2DJ, UK



Fig. 1 Real underwater images suffering from severe chromatic aberration, blurred fog, and low contrast. These characteristics make feature descriptors in degraded underwater images prone to change

underwater images that are blurred and suffer from low-quality chromatic aberration.

Further, traditional manual descriptors are limited by researchers' prior knowledge and specific application scenarios and are gradually replaced by learning-based descriptors. Although deep learning-based feature descriptors are data-driven, developing underwater datasets for research is considerably challenging. Moreover, there is a lack of existing underwater training datasets that contain pairs of underwater images for matching training. The current research focus primarily involves feature descriptors of air images, overlooking the need for attenuation invariance in underwater image feature descriptors. Consequently, matching features between two degraded underwater images remains highly challenging (Oliver et al. 2010), and there is a dearth of evaluation metrics for underwater visual feature matching in current research.

Motivated by the above observations, we propose a novel approach toward underwater image feature descriptors based on attenuation invariance. Our proposed approach can maintain attenuation invariance in underwater images that suffer from light scattering and absorption and improve the matching accuracy of degraded underwater images. We provide a foundation for underwater vision for ocean exploration and small underwater image descriptor evaluation datasets to compensate for the lack of underwater image datasets in current research.

The main contributions of this study are summarized below:

- (1) We propose an attenuation invariance-based training strategy for underwater image matching that enables descriptor networks to be trained for matching noisy images from multiple water bodies. By learning the attenuation differences between the images, we generate feature descriptors that are invariant to attenuation, thus improving the accuracy of underwater visual feature matching.
- (2) We construct an underwater dataset of over 30000 images, namely the multiple water type (MWT) dataset, and an underwater image feature descriptor (UIFD) evaluation dataset to compensate for the

lack of evaluation datasets for underwater visual feature matching.

- (3) We propose a novel underwater image feature descriptor network based on attenuation invariance. The proposed network can generate reliable descriptors even under underwater light scattering and absorption.

The remainder of this paper is organized as follows: Section 2 presents related works on underwater image feature matching. Section 3 introduces the proposed dataset and method. Section 4 discusses the experiments. Finally, Section 5 presents the conclusions.

2 Related work

Underwater Image Enhancement. Underwater images often suffer from degradation due to chromatic aberration, blurring, low contrast, and uneven illumination caused by the complex underwater environment during imaging (Li et al. 2020b). Existing methods for UIE can be broadly categorized into physical model-based and data-driven methods (Peng et al. 2023). AHE (Adaptive Histogram Equalization) (Pizer et al. 1987) focuses on local histogram computation to redistribute image brightness and achieve local adaptive enhancement. Because the UIE model is highly similar to the haze model (McCartney 1976), DCP (Dark Channel Prior) (He et al. 2010) has been commonly used for UIE.

Recently, deep learning-based methods have shown remarkable improvements in UIE (Li et al. 2017; Fabbri et al. 2018; Uplavikar et al. 2019; Islam et al. 2020; Li et al. 2020a). UWCNN (Li et al. 2020a) is a UIE method based on convolutional neural networks, which learns the difference between the synthesized underwater image and the corresponding clear image through the network structure, and then reconstructs the degraded underwater image to obtain an enhanced image. Deep SESR (Islam et al. 2020) consists of two main parts: the feature extraction network (FENet) and the auxiliary attention network (AAN), which realizes UIE and improves the resolution of the image. WaterGAN (Li et al. 2017) employs a generative adversarial network (GAN) to synthesize degraded/clear image pairs and proposes a two-stage refinement network. To address the problem of lacking paired underwater image datasets, UWCNN (Li et al. 2020a) simulates the underwater images based on a physical model and ten different water types.

Traditional Feature Descriptor. Visual feature matching is an important technique in computer vision applications, which is valuable in image retrieval (Rui et al. 1999) and robot navigation and localization (Leng et al. 2018). A good feature descriptor should remain unique and robust in various image transformations, such as changes

in scale and illumination. Among the traditional feature descriptors, SIFT, SURF, and Daisy are notable gradient-based descriptors. SIFT descriptors (Lindeberg 2012) have been widely used for image matching and recognition because of their stability. SIFT descriptor is invariant in rotation and scaling transformations and is a robust scale invariant descriptor. Lin and Liu (2014) applied the SIFT descriptor to the bi-directional matching localization of underwater images. Bay et al. (2006) proposed the SURF descriptor, which is faster to compute than SIFT. Tola et al. (2009) proposed the Daisy local feature descriptor that can estimate gradients in different directions for key points, demonstrating high-density computational efficiency.

Furthermore, there are binary descriptors, such as Brief, ORB, and BRIEF, which offer advantages in terms of memory consumption and computational speed compared to gradient-based descriptors. Brief (Calonder et al. 2010) constructs feature description vectors based on the size relationship of pixel pairs surrounding key points. The similarity of descriptors is then evaluated by calculating the distances between feature vectors using the Hamming distance. This approach requires less storage and enables fast matching. Rublee et al. (2011) proposed ORB descriptors based on Brief. The key points are extracted using FAST (Rosten and Drummond 2006), and the descriptors are generated in the same way as BRIEF; thus, the ORB descriptors are computed quickly. It is worth noting that traditional descriptors have been developed based on researchers' prior knowledge and have mostly been limited in their application to specific scenarios, lacking extensive use in underwater image analysis.

Deep Learning-based Feature Descriptors. Deep learning techniques have also been gradually applied to local image characterization, with their powerful learning and data-driven capabilities, which are not limited by prior knowledge. This technique also overcomes the limitations of scale and illumination variation change. L2Net (Tian et al. 2017) learns high-performance descriptors in Euclidean space through convolutional neural networks (Albawi et al. 2017). Szkocka et al. (2017) proposed a method for training convolutional neural networks and applied it to L2Net to obtain HardNet descriptors through lossy metric learning, which shows superior performance in image retrieval and view matching under extreme conditions. SuperPoint (DeTone et al. 2018) proposes a superior-performing SuperPoint keypoint detector and presents robust descriptors in multiview geometry. SOSNet (Tian et al. 2019) proposes second-order similarity regularization based on second-order similarity. The resulting SOSNet descriptors demonstrate excellent performance in the benchmark test set, and the

second-order similarity regularization optimization can significantly improve the matching performance of the learned descriptors. R2D2 (Revaud et al. 2019) learns keypoint descriptors only in high-confidence regions and trains keypoint detectors and descriptors jointly. The LISRD (Pautrat et al. 2020) descriptor enables the learned descriptor to remain invariant while obtaining more local feature information. LISRD demonstrates remarkable performance in public datasets. The GIFT descriptor (Liu et al. 2019b) uses a group of convolutional networks to extract features, thus ensuring the specificity of the descriptor and improving the performance of the relative pose estimation in practical applications.

3 Proposed dataset and method

3.1 MWT and UIFD dataset

In the current research, there is a lack of paired underwater image data for various types of water bodies, and they are not well categorized. Additionally, acquiring actual underwater image data is extremely difficult and time-consuming, especially for multiple water body types. To address this issue, researchers commonly use underwater image generation methods to integrate underwater data. Among these, two methods are commonly used. The first method utilizes the depth data of air images to generate underwater images using underwater image synthesis models based on the underwater imaging principle, such as the IPO-based underwater image synthesis method. The second method involves learning the physical prior through the style transfer method to convert the underwater image type. Therefore, herein, we contribute to the field by constructing an underwater image dataset (i.e., MWT) using the aforementioned methods, which includes over 30000 underwater images (Fig. 2).

Currently, publicly available measurement datasets of local feature descriptors for underwater images are limited. An evaluation dataset that can assess the impact of color distortion and blur on descriptors in degraded underwater images while incorporating fundamental geometric and illumination variations is crucial. In this section, we discuss the development of a dataset (i.e., UIFD) specifically designed to evaluate local feature descriptors in underwater images.

3.1.1 IPOs-based underwater image synthesis

We employ an IPO-based underwater image synthesis method (Ding et al. 2019) to generate underwater datasets for various water bodies. This method incorporates the inherent optical properties of different water bodies into the synthesis process. Jerlov (1977) classified the marine environment into ten types: I, IA, IB, II, III, 1C, 3C, 5C, 7C, and 9C. Further, Solonenko and Mobley (2015) provided the absorption and scattering parameters

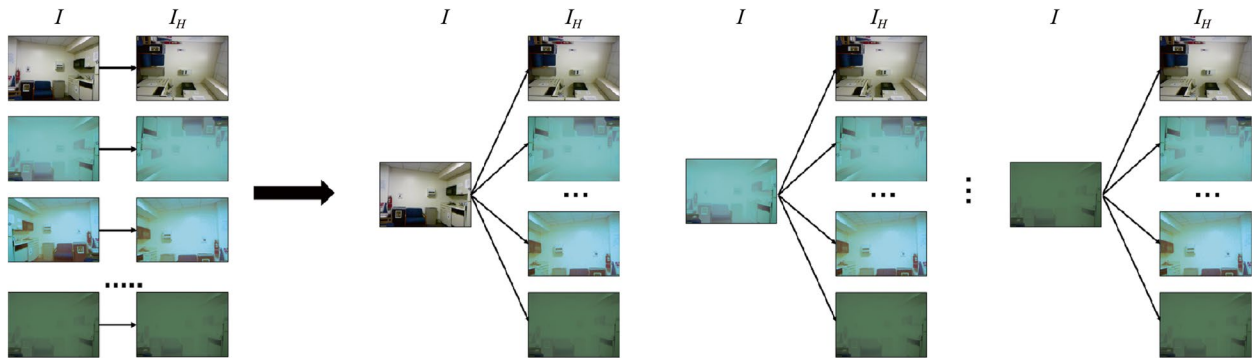


Fig. 2 Anchor image (I) obtained under various water body conditions and the matching image (I_H). I_H is a prior image with varying degrees of chromatic aberration and blurring added using an underwater image generation model based on image I

at wavelengths of 650 nm (red), 525 nm (green), and 450 nm (blue) for all Jerlov water body types.

We utilize a clear image E_λ^{in} from NYU-Depth V2 (Silberman et al. 2012) and its depth value $d(x)$, combined with the absorption coefficient $\alpha(\lambda)$, scene depth $d(x)$ and scattering coefficient $\beta(\lambda)$, to obtain the transmittance $T_\lambda(x)$. Next, we randomly select a background light value B_λ of size 0.5 to 1. Based on the principle of light propagation in water, we use the image in air to simulate the attenuation and scattering of light in water.

$$I_\lambda(x) = (E_\lambda^{\text{in}} \cdot C_\lambda) + (1 - T_\lambda(x)) \cdot B_\lambda \quad (1)$$

where C_λ is the wavelength attenuation coefficient. According to the underwater image formation model formula, the degraded underwater image is obtained for different water body types. To compensate for the lack of real underwater datasets, ten simulated underwater images are generated based on scattering and attenuation parameters covering different water body types. The synthetic dataset includes over 30000 images and encompasses water depths ranging from 3 to 10 m.

3.1.2 Underwater image based on style transfer

In addition to using imaging model-based methods for underwater image generation, another existing approach involves the use of the style transfer method of color domain conversion. To enhance dataset diversity, prevent overfitting of the descriptor network, and expand the number of datasets, this paper incorporates an improved CycleGAN (Zhu et al. 2017) network for synthesizing underwater images pertaining to multiple water body types. The CycleGAN network employs cyclic consistency and adversarial networks to learn the mapping between the source and target domains. This ensures that the generated images in the source domain are indistinguishable from those in the target domain while also

enforcing cyclic consistency to enhance the learning process.

Due to the ambiguous classification of water body types in existing underwater datasets, we have categorized underwater images with similar color domains into a group. Specifically, underwater images are divided into blue, green, and turquoise groups. To achieve this, we employed the CycleGAN network. The network is trained using the blue water and green water datasets as the source and target domains, respectively. This training process ensures that images from the source domain can generate images with the color domain of the target domain while preserving the content. Similarly, images from the target domain can generate underwater images with a consistent style and content as the source domain. To incorporate the physical prior of the underwater image without altering its content, we replace the original generator and residual block with U-Net (Ronneberger et al. 2015) in the generator, as shown in Fig. 3.

CycleGAN has two mappings: G and F . The mapping function G generates the clear image $G(x)$ from $x \in X$, whereas the mapping function F reconstructs the degraded image x by $G(x) \rightarrow F(G(x))$. Discriminator D_Y examines whether the generated image $G(x)$ and clear image y are derived from the same distribution. Additionally, $y \rightarrow F(y) \rightarrow G(F(y))$ and D_X are similar inverse processes.

Real underwater images are also utilized in addition to the abovementioned two methods to further expand the number of datasets. This approach enhances the authenticity of data and helps to prevent network overfitting.

3.1.3 UIFD dataset

The evaluation dataset constructed herein is mainly used to evaluate the impact of underwater light attenuation on descriptor performance. According to the characteristics of underwater light attenuation, the dataset is

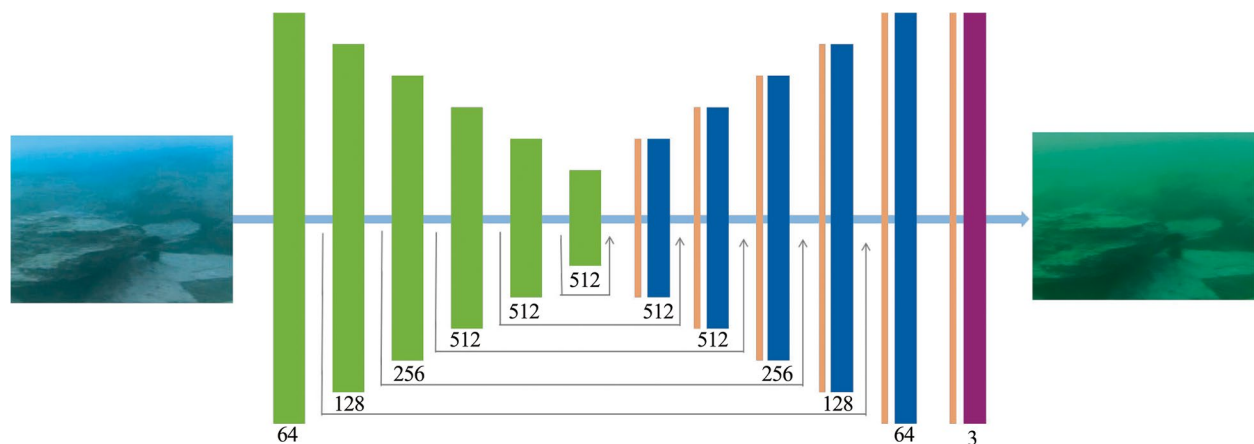


Fig. 3 Overview of the improved CycleGAN generator network architecture. We used U-Net instead of the original generator and residual block to address the issue of jaggedness in the generated image

divided into three subcategories: chromatic aberration, blur, and clean. Chromatic aberration in underwater images is primarily caused by the absorption of underwater light, while blurring is mainly caused by scattering. To provide a basis for comparison, we include clear underwater images. The image samples used herein are consistent with the classification description of the dataset. The chromatic aberration images predominantly exhibit chromatic aberration as the main cause, with minimal blurring effects. Similarly, the blurry category is less affected by chromatic aberration. The evaluation dataset is shown in Fig. 4.

An anchor image meeting the classification criteria is selected in each image sequence or video-converted image sequence. Another reference image is selected at an interval of 5–10 frames to ensure accurate polar motion. In evaluating the impact of water body type on the descriptors, it is challenging to match the image pairs in the datasets with chromatic aberration and blurring. Hence, the SIFT algorithm is used to extract key points and generate SIFT descriptors for matching. Since image matching is difficult, we manually select paired key points as the correct match to obtain the homography matrix between the image pairs. First, a key point coordinate is selected from the key points of the anchor image, and then, the corresponding key point is searched in the

reference image. The correctness of the matching points is determined using the `findHomography` function in OpenCV to find the homography matrix H for the coordinates of the keypoint pairs, which serves as the ground truth for matching the image pairs.

Due to the lack of underwater data, it is difficult to obtain multiple images in continuous scenes. To address this issue, each evaluation dataset sequence consists of anchor and reference images. The anchor image is randomly rotated and scaled using data augmentation techniques, and the homography transformation matrix serves as the ground truth between the anchor and augmented images. Consequently, each sequence comprises six images, including two real images and four augmented images. To assess the performance of the image local feature descriptor after UIE, we apply AHE (Pizer et al. 1987), DCP (He et al. 2010), UWCNN (McCartney 1976), and Deep SESR (Islam et al. 2020) for UIE using the evaluation dataset.

3.2 Network architecture

As shown in Fig. 5, we propose an underwater feature descriptor network based on attenuation invariance. The network takes RGB images as input and generates 128-dimensional feature descriptors as output. The network architecture comprises a five-layer convolutional module that operates at three simultaneous scales in parallel. Each module layer comprises a convolutional layer, a ReLU layer, a convolutional layer, and an IN layer.

After two convolution layers, the IN layer is used to normalize each feature map to obtain a high-resolution feature map F of size $H \times W \times 32$. Inspired by U-Net (Ronneberger et al. 2015) and HRNet (Sun et al. 2019), herein, we introduce parallel connectivity into the network to perform feature fusion at multiple scales. Through multibranch feature map fusion,

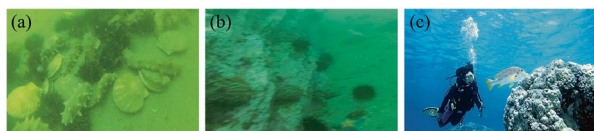


Fig. 4 **a** Underwater degraded image mainly affected by chromatic aberration; **b** underwater degraded image mainly affected by blurring; **c** clear underwater image less affected by chromatic aberration and blurring

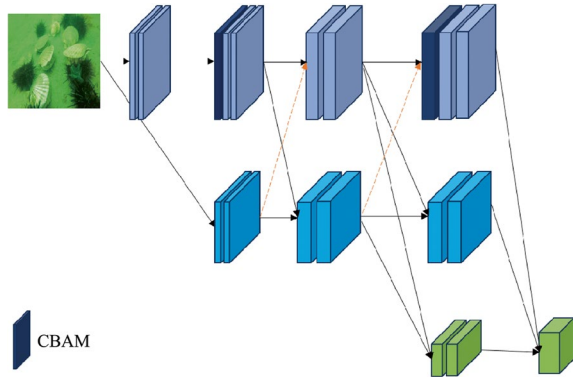


Fig. 5 Overview of the network architecture, underwater feature descriptor network structure based on attenuation invariance with multiscale fusion and attention mechanism

a $\frac{1}{4}H \times \frac{1}{4}W \times 128$ dimensional descriptor is finally obtained. To capture the local information of the image in detail, we adopt a 3×3 convolutional kernel with stride 1 to reduce the size of the receptive field and focus on local information. Additionally, the network introduces the attention mechanism CBAM (Woo et al. 2018) to capture the local information in a more detailed way. As shown in Fig. 6, attention is added to the channel and space to make the network focus more on the target.

The output feature map F of the previous convolution layer is used as the input feature, and the mixed attention output feature F'' is obtained by multiplying the channel attention feature $M_c(F)$ and the spatial attention feature $M_s(F')$.

$$F' = F \otimes M_c(F) \quad (2)$$

$$F'' = F' \otimes M_s(F') \quad (3)$$

3.3 Loss function

For two images I^A and I^B to be matched, where the keypoint $k_{1..N}^a$ of I^A corresponds to the feature descriptor $d_{1..N}^a$, and the keypoint $k_{1..N}^b$ of I^B corresponds to the feature descriptor $d_{1..N}^b$, each keypoint k_i^a in I^A corresponds to each keypoint k_i^b in I^B , and the spatial distance

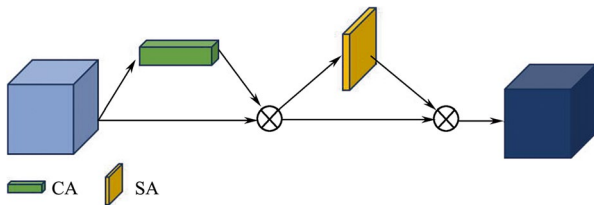


Fig. 6 Overview of CBAM (Convolutional Block Attention Module), the module consists of two sub-modules: channel and spatial

$D(k_i^a, k_i^b)$ of these two key points under L_2 normalization is described as a positive distance pi .

$$D(k_i^a, k_i^b) = \|d_i^a - d_i^b\|_2 \quad (4)$$

$$pi = D(k_i^a, k_i^b) \quad (5)$$

For the negative distance ni between two key points that are not correctly matched.

$$ni = \min(D(k_i^a, k_{n_b(i)}^a), D(k_i^b, k_{n_a(i)}^b)) \quad (6)$$

To maximize the distance between a positive match and the closest nonmatching descriptor, a triplet loss function is used. The triplet loss function is shown in Eq. (7)

$$L_T(I^A, I^B, D) = \frac{1}{N} \sum_{i=1}^N \max(M + (pi)^2 - (ni)^2, 0) \quad (7)$$

For an image I^W where the water body type changes but the illumination and geometry remain unchanged, the change descriptor loss function L_1 between it and the image I can be expressed as follows:

$$L_1 = L_T(I, I^W, \|d - d^W\|_2) \quad (8)$$

For an image $I^{W'}$ transformed in water type, illumination, and geometry, the change descriptor loss function L_2 between it and the image I can be expressed as follows:

$$L_2 = L_T(I, I^{W'}, \|d - d^{W'}\|_2) \quad (9)$$

The final loss function can be expressed as:

$$L = L_1 + L_2 \quad (10)$$

4 Experiments

4.1 Implementation details

To assess the robustness and attenuation invariance of the proposed feature descriptor in degraded underwater images, we use a constructed evaluation dataset. The descriptor is evaluated using qualitative and quantitative analyses. For comparison, we select five advanced feature descriptors: SuperPoint (DeTone et al. 2018), SOS-Net (Tian et al. 2019), GIFT (Liu et al. 2019b), LISRD (Pautrat et al. 2020), and Aslfeat (Luo et al. 2020).

We use Python and Pytorch frameworks via NVIDIA RTX2080Ti on Ubuntu20 to implement the network architecture. The Adam optimization algorithm is utilized for a total of 800 epochs of training with a batch size of 4. The initial learning rate is 0.001, and M in the loss function is 1.

Table 1 Comparison of evaluation metrics for each descriptor method; * and ** data represent optimal and suboptimal results, respectively

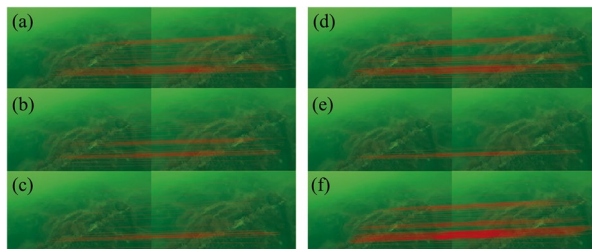
		GIFT	SOSNet	SuperPoint	Aslfeat	LISRD	Ours
Chromatic aberration	Hestimation	0.036	0.041	0.056*	0.001	0.041	0.051**
	Precision	0.115	0.115	0.111	0.093	0.126**	0.146*
	Recall	0.223**	0.200	0.206	0.132	0.205	0.227*
Blurring	Hestimation	0.158	0.164	0.181	0.006	0.230**	0.251*
	Precision	0.216	0.204	0.190	0.104	0.259**	0.266*
	Recall	0.286	0.242	0.228	0.185	0.309**	0.314*
Clear	Hestimation	0.339*	0.338**	0.276	0.154	0.303	0.325
	Precision	0.310	0.288	0.273	0.330	0.334*	0.332**
	Recall	0.388	0.298	0.339	0.337	0.391**	0.393*

4.2 Degraded image descriptor estimation

Qualitative analysis can provide an intuitive assessment of the effectiveness of the descriptor in underwater images that have been severely degraded. For the evaluation dataset, we select images that are heavily impacted by chromatic aberration and blurring. Initially, we use SuperPoint as the key point detector and extract 2000 feature points from the image. Then, we utilize GIFT, SOSNet, SuperPoint, LISRD, Aslfeat, and Ours to generate local feature descriptors and conduct image matching. The results of the qualitative analysis are presented in Table 1.

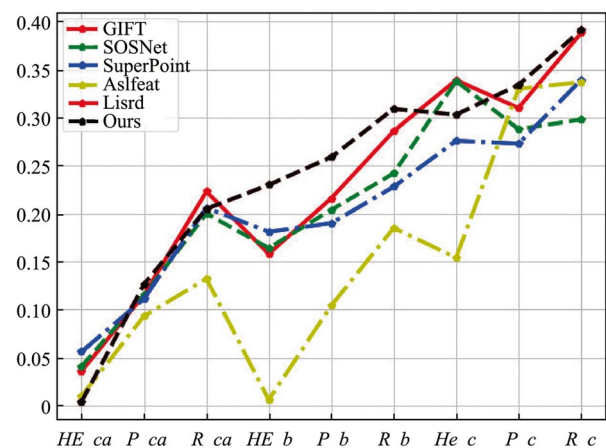
As shown in Fig. 7, this paper introduces descriptors that are highly dense and widely distributed for underwater image matching, outperforming LISRD and GIFT. However, the Aslfeat descriptor demonstrates the poorest matching effect as it only concentrates on areas where the features are more prominent. To provide a more accurate evaluation of the performance of the different descriptors, we employ three indicators: HEstimation, precision, and recall. A higher value of these indicators indicates a closer approximation to the real value and a better evaluation effect.

During the experiment, the SuperPoint keypoint detector is still used to extract 2000 feature points from the evaluation dataset, and then the descriptors

**Fig. 7** Underwater image matching result comparison a–f represent GIFT, SOSNet, SuperPoint, LISRD, Aslfeat, and our descriptors, respectively

generated by various methods are used for evaluation on the dataset. Additionally, the underwater images maintain the same resolution in all methods during the evaluation process. The evaluation results are shown in Table 1.

As shown in Table 1 and Fig. 8, our method outperforms the other five descriptor methods in the overall evaluation. In particular, in degraded underwater images affected by chromatic aberration, our descriptor achieves significantly higher precision and recall than the other methods. Our method demonstrates the best results across all metrics for images with severe blurring, followed by the LISRD and GIFT methods. In the case of clear underwater image matching, the advantage of our method is relatively smaller, and each descriptor algorithm performs well.

**Fig. 8** Overall evaluation comparison of the descriptor algorithms. HE, P, R represent homography estimation, precision, and recall, respectively. ca, b, c represent chromatic aberration, blurring, and clear, respectively

4.3 Enhanced image descriptor estimation

In literature, there exist approaches to enhance underwater images using common descriptors to improve matching accuracy. To further validate the effectiveness of our proposed method for local feature descriptors in underwater image matching, we compare our method with popular existing UIE methods. We expand the original underwater evaluation dataset by incorporating four additional methods. The performance of the five mainstream descriptors is evaluated using the enhanced dataset and compared with the performance of the method described herein on the original dataset. The evaluation criteria used are HEstimation, precision, and recall. The evaluation results are shown in Table 2 and Fig. 9. Note that higher values indicate better results.

The results of Table 2 and Fig. 9 show that the matching accuracy of the Aslfeat descriptor in the AHE-enhanced color distortion sequence improved by approximately 0.1. Similarly, the homography estimation of the SOSNet and SuperPoint

descriptors in the DCP-enhanced color distortion sequence increased by 0.061 compared with the original underwater dataset. However, it is worth noting that the evaluation results of various descriptor algorithms on the enhanced underwater image are generally lower than those on the original image. This is due to varying degradation degrees in underwater images, which make it challenging for the enhancement algorithms to be universally applicable. In the case of unsuitable underwater images, the image quality does not improve and may even degrade further. However, our proposed underwater image feature descriptor demonstrates strong robustness and attenuation invariance when dealing with degraded underwater images.

4.4 Ablation study

To capture local information in detail and obtain more refined features, we added the CBAM attention mechanism to the network structure. Next, we conducted the ablation experiment of the attention module

Table 2 Comparison of evaluation indicators of each descriptor algorithm after image enhancement; * and ** data represent optimal and suboptimal results, respectively

Descriptor	Dataset	Chromatic aberration			Blurring		
		Hestimation	Precision	Recall	Hestimation	Precision	Recall
GIFT	AHE	0.056**	0.103	0.168	0.169	0.206	0.249
	DCP	0.056	0.107	0.178	0.170	0.202	0.247
	UWCNN	0.051	0.107	0.196	0.170	0.210	0.266
	Deep SESR	0.041	0.092	0.197	0.109	0.151	0.253
	Original	0.036	0.115	0.223**	0.158	0.216	0.286
SOSNet	AHE	0.046	0.097	0.138	0.170	0.188	0.202
	DCP	0.061*	0.104	0.155	0.176	0.189	0.204
	UWCNN	0.051	0.106	0.175	0.151	0.197	0.226
	Deep SESR	0.051	0.090	0.164	0.127	0.129	0.165
	Original	0.041	0.115	0.200	0.164	0.204	0.242
SuperPoint	AHE	0.041	0.095	0.152	0.151	0.182	0.195
	DCP	0.061*	0.102	0.171	0.182	0.183	0.213
	UWCNN	0.041	0.101	0.181	0.157	0.184	0.213
	Deep SESR	0.036	0.086	0.179	0.103	0.133	0.191
	Original	0.056	0.111	0.206	0.181	0.190	0.228
Aslfeat	AHE	0.010	0.190*	0.130	0.121	0.100	0.176
	DCP	0.005	0.090	0.130	0.0	0.104	0.181
	UWCNN	0.010	0.091	0.128	0.0	0.103	0.183
	Deep SESR	0.015	0.076	0.126	0.004	0.064	0.140
	Original	0.010	0.093	0.132	0.006	0.104	0.185
LISRD	AHE	0.046	0.105	0.148	0.200	0.244	0.270
	DCP	0.051	0.116	0.161	0.212	0.228	0.249
	UWCNN	0.051	0.113	0.182	0.224	0.248	0.288
	Deep SESR	0.051	0.102	0.175	0.121	0.150	0.225
	Original	0.041	0.126	0.205	0.230**	0.259**	0.309**
Ours	Original	0.051	0.146**	0.227*	0.251*	0.266*	0.314*

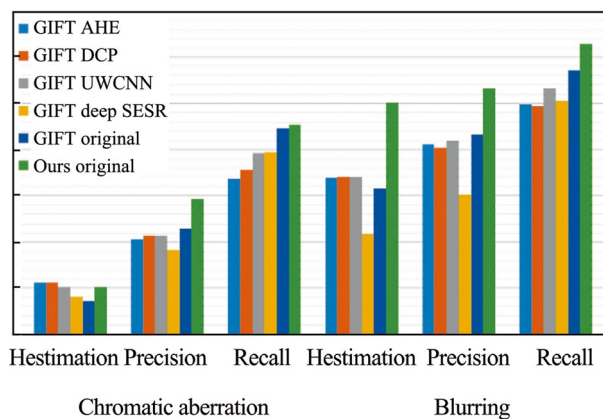


Fig. 9 Comparison histogram of evaluation indicators of each descriptor algorithm after image enhancement

CBAM. The models that do not incorporate attention introduce only the spatial attention module SAM (Laskar and Kannala 2017). Here, we introduced and evaluated the spatial and channel attention module CBAM on the dataset; the evaluation results are presented in Table 3.

Table 3 shows that the matching accuracy in the three underwater image sequences increases if only the spatial attention module is introduced. The channel attention module (CBAM) is introduced herein on the basis of spatial attention. In this case, the matching accuracy is greatly improved compared with the network without the attention module. Further, there are slight improvements compared to the network model that only adds the spatial attention module.

5 Conclusions

Herein, we constructed an underwater dataset of over 30000 images (namely the MWT dataset) and a UIFD evaluation dataset to compensate for the lack of evaluation datasets for underwater visual feature matching. Further,

we propose an underwater feature descriptor network based on attenuation invariance. Qualitative and quantitative comparisons of the proposed descriptors on the evaluation dataset confirmed that the proposed method outperforms advanced descriptor algorithms.

Authors' contributions

Jinghao Sun and Hong Zhang carried out the experiment. Yuan Rao and Jinghao Sun wrote the manuscript with support from Hui Yu. Hui Yu conceived of the study and participated in its design and coordination. All authors read and approved the final manuscript.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests. Hui Yu is one of the Editorial Board Members, but he was not involved in the journal's review of, or decision related to, this manuscript.

Received: 18 August 2023 Revised: 13 September 2023 Accepted: 14 September 2023

Published online: 09 November 2023

References

- Albawi S, Mohammed TA, Al-Zawi S (2017) Understanding of a convolutional neural network. 2017 International Conference on Engineering and Technology (ICET), Akdeniz Univ, Antalya, Turkey, 21–21 August 2017, pp 1–6
- Bay H, Tuytelaars T, Van Gool L (2006) SURF: speeded up robust features. Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, 7–13 May 2006, pp 404–417
- Calonder M, Lepetit V, Strecha C, Fua P (2010) BRIEF: binary robust independent elementary features. Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, 5–11 September 2010, pp 778–792
- Dai Z, Huang XH, Chen WN, He L, Zhang H (2019) A comparison of CNN-based and hand-crafted keypoint descriptors. 2019 International Conference on Robotics and Automation (ICRA), Montreal, Canada, 20–24 May 2019, pp 2399–2404
- DeTone D, Malisiewicz T, Rabinovich A (2018) SuperPoint: self-supervised interest point detection and description. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, 18–22 June 2018, pp 337–349
- Ding XY, Wang YF, Yan Y, Liang Z, Mi ZT, Fu XP (2019) Jointly adversarial network to wavelenght compensation and dehazing of underwater images. Preprint at [arXiv:1907.05595](https://arxiv.org/abs/1907.05595)
- Fabbri C, Islam MJ, Sattar J (2018) Enhancing underwater imagery using generative adversarial networks. 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018, pp 7159–7165
- He KM, Sun J, Tang XO (2010) Single image haze removal using dark channel prior. IEEE Trans Pattern Anal Mach Intell 33(12):2341–2353
- Islam MJ, Luo PG, Sattar J (2020) Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. Preprint at [arXiv:2002.01155](https://arxiv.org/abs/2002.01155)
- Jerlov NG (1977) Classification of sea water in terms of quanta irradiance. ICES J Mar Sci 37(3):281–287
- Laskar Z, Kannala J (2017) Context aware query image representation for particular object retrieval. 20th Scandinavian Conference on Image Analysis, Tromsø, Norway, 12–14 June 2017, pp 88–99

Table 3 Attention module ablation experiment evaluation results

		W/o CBAM	SAM	CBAM
Chromatic aberration	Hestimation	0.040	0.048	0.051
	Precision	0.137	0.142	0.146
	Recall	0.176	0.213	0.227
Blurring	Hestimation	0.181	0.236	0.251
	Precision	0.235	0.249	0.266
	Recall	0.283	0.307	0.314
Clear	Hestimation	0.317	0.320	0.325
	Precision	0.289	0.312	0.332
	Recall	0.361	0.384	0.393

- Leng CC, Zhang H, Li B, Cai GR, Pei Z, He L (2018) Local feature descriptor for image matching: a survey. *IEEE Access* 7:6424–6434
- Li CY, Anwar S, Porikli F (2020a) Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognit* 98(107):107038
- Li J, Skinner KA, Eustice RM, Johnson-Roberson M (2017) Watergan: unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robot Autom Lett* 3(1):387–394
- Li TY, Rong S, Cao XT, Liu YB, Chen L, He B (2020b) Underwater image enhancement framework and its application on an autonomous underwater vehicle platform. *Opt Eng* 59(8):083102
- Lin Y, Liu B (2014) Underwater image bidirectional matching for localization based on sift. *J Mar Sci Appl* 13(2):225–229
- Lindeberg T (2012) Scale invariant feature transform. *Scholarpedia* 7(5):10491
- Liu P, Wang GY, Qi H, Zhang CF, Zheng HY, Yu ZB (2019a) Underwater image enhancement with a deep residual framework. *IEEE Access* 7:94614–94629
- Liu Y, Shen ZH, Lin ZX, Peng SD, Bao HJ, Zhou XW (2019b) Gift: learning transformation-invariant dense visual descriptors via group CNNs. 33rd Conference on Neural Information Processing Systems (NeurIPS), Vancouver, Canada, 08–14 December 2019, pp 6992–7003
- Luo ZX, Zhou L, Bai XY, Chen HK, Zhang JH, Yao Y et al (2020) ASLFeat: learning local features of accurate shape and localization. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 13–19 June 2020, pp 6588–6597
- McCartney EJ (1976) *Optics of the atmosphere: scattering by molecules and particles*. Wiley, New York
- Mishchuk A, Mishkin D, Radenovic F, Matas J (2017) Working hard to know your neighbor's margins: local descriptor learning loss. 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, 4–9 December 2017, pp 4827–4838
- Oliver K, Hou WL, Wang S (2010) Image feature detection and matching in underwater conditions. *Conference on Ocean Sensing and Monitoring II*, Orlando, FL, 5–6 April 2010, pp 185–196
- Pautrat R, Larsson V, Oswald MR, Pollefeys M (2020) Online invariance selection for local feature descriptors. *Computer Vision–ECCV 2020: 16th European Conference*, Glasgow, UK, 23–28 August 2020, pp 707–724
- Peng LT, Zhu CL, Bian LH (2023) U-shape transformer for underwater image enhancement. *IEEE Trans Image Proc* 32:3066–3079
- Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer TH et al (1987) Adaptive histogram equalization and its variations. *Comput Vis Graph Image Proc* 39(3):355–368
- Qi Q, Zhang YC, Tian F, Wu QMJ, Li KQ, Luan X et al (2021) Underwater image co-enhancement with correlation feature matching and joint learning. *IEEE Trans Circ Syst Video Technol* 32(3):1133–1147
- Revaud J, Weinzaepfel P, De Souza C, Pion N, Csurka G, Cabon Y et al (2019) R2d2: repeatable and reliable detector and descriptor. Preprint at [arXiv:1906.06195](https://arxiv.org/abs/1906.06195)
- Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015, pp 234–241
- Rosten E, Drummond T (2006) Machine learning for high-speed corner detection. *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision*, Graz, Austria, 7–13 May 2006, pp 430–443
- Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: an efficient alternative to SIFT or SURF. 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011, pp 2564–2571
- Rui Y, Huang TS, Chang SF (1999) Image retrieval: past, present, and future. *J Vis Commun Image Represent* 10(1):1–23
- Silberman N, Hoiem D, Kohli P, Fergus R (2012) Indoor segmentation and support inference from rgb-d images. *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision*, Florence, Italy, 7–13 October 2012, pp 746–760
- Solonenko MG, Mobley CD (2015) Inherent optical properties of Jerlov water types. *Appl Opt* 54(17):5392–5401
- Sun K, Xiao B, Liu D, Wang JD (2019) Deep high-resolution representation learning for human pose estimation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019, pp 5686–5696
- Tian Y, Fan B, Wu F (2017) L2-net: deep learning of discriminative patch descriptor in euclidean space. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017, pp 6128–6136
- Tian Y, Yu X, Fan B, Wu F, Heijnen H, Balntas V (2019) SOSNet: second order similarity regularization for local descriptor learning. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019, pp 11008–11017
- Tola E, Lepetit V, Fua P (2009) DAISY: an efficient dense descriptor applied to wide-baseline stereo. *IEEE Trans Pattern Anal Mach Intell* 32(5):815–830
- Uplavikar PM, Wu Z, Wang Z (2019) All-in-one underwater image enhancement using domain-adversarial learning. Preprint at [arXiv:1905.13342](https://arxiv.org/abs/1905.13342)
- Woo S, Park J, Lee JY, Kweon IS (2018) CBAM: convolutional block attention module. *Computer Vision–ECCV 2018: 15th European Conference on Computer Vision*, Munich, Germany, 8–14 September 2018, pp 3–19
- Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. 16th IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017, pp 2242–2251

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.