Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3052

ENHANCED VINS-BASED UNDERWATER LOCALIZATION WITH IMAGE ENHANCEMENT

Fei Liao*, BingLei Bao
Department of Automation, University of Science and Technology of China, Hefei 230026, Anhui, China.
Corresponding Author: Fei Liao, Email: roboleofly@mail.ustc.edu.cn

Abstract: The underwater visual-inertial navigation system (VINS) confronts significant challenges due to the adverse underwater visual environment, including light absorption and scattering, tiny suspended particles, color distortion, and image blurring. To address these issues, this paper introduces a multi-scale fusion-based image enhancement algorithm, integrating it into the front-end of underwater image enhancement techniques. This integration effectively enhances the performance of underwater localization. Experimental results on the Aqualoc underwater dataset demonstrate that the proposed method increases the number of extracted feature points and achieves more stable tracking, thereby reducing localization errors compared to traditional VINS approaches.
Keywords: Underwater image enhancement; Underwater slam

1 INTRODUCTION

Since the beginning of the 21st century, the rapid growth of population has led to a gradual depletion of land resources, and countries have stepped up their exploration and development of the ocean. With the rapid development of computer vision technology, camera sensors have become the main way for autonomous robots and smart wearable devices to perceive the surrounding environment. Rich visual information is extracted and input into the intelligent body to achieve complex tasks such as detection and recognition, positioning and navigation, and planning and decision-making.

Compared with ordinary images, images captured underwater suffer from significantly reduced visibility due to light propagation attenuation [1]. The reduction in underwater visibility is mainly caused by optical phenomena, including absorption and scattering processes. The enhanced scattering caused by suspended particles (such as sediments and plankton) exacerbates the attenuation of light, further reducing the overall visibility [2]. In addition, the selective absorption of specific light wavelengths by water weakens the optical clarity [3]. In view of the above challenges, there is an urgent need to develop effective methods to enhance the image quality of underwater images, thereby improving the effective use of underwater visual data.Image enhancement algorithms are mainly divided into traditional pixel-level processing methods, deep learning-based processing strategies, and image fusion-based methods. Traditional underwater image enhancement methods mainly rely on image processing technology to improve the visual quality of images. These methods usually focus on enhancing contrast, reducing noise, and achieving color balance at the image level. Commonly used methods include histogram equalization, denoising filtering, and color correction.

Histogram equalization (HE) [4] is a common algorithm for enhancing image contrast. This algorithm converts the histogram distribution of an image into an approximately uniform distribution through a cumulative distribution function, thereby expanding the grayscale range of the original image. The purpose of denoising filtering is to reduce the noise information in the image while retaining important details and structural information of the image. Common filtering methods include Gaussian filtering, median filtering, mean filtering, and bilateral filtering. The basic idea is to smooth the image to remove noise.Another pixel-level processing method is to use underwater optical imaging models to deal with the inherent defects of underwater imaging at the optical level. For example, underwater wavelength compensation and image defogging methods, restoration methods based on image blur and light absorption, and underwater image enhancement methods based on Retinex theory.

In order to improve the image quality improvement effect of underwater image enhancement algorithms, researchers have introduced some advanced results in the field of deep learning, such as convolutional neural networks and generative adversarial networks. Deep learning technology can automatically learn and extract image features, so after learning enough sample data, the neural network can restore the features of underwater images. Convolutional neural networks have been widely used in the field of underwater image enhancement in recent years.

Deep learning technology can automatically learn and extract image features, so after learning enough sample data, the neural network can restore the features of underwater images. Convolutional neural networks have been widely used in the field of underwater image enhancement in recent years. Wang et al. [5] proposed an end-to-end framework UIE-net for underwater image enhancement, which can simultaneously perform color correction and dehazing for underwater images. Although this method has achieved good results on benchmark datasets, the performance of the algorithm is greatly affected by the quality of training data when applied in actual scenarios, and more diverse and high-quality datasets are needed to improve the performance of the algorithm. Li et al. [6] proposed an underwater image enhancement network Ucolor, which corrects color cast and enhances contrast of images through a multi-color space embedding method guided by medium transmission. In order to effectively improve the image processing speed, Naik et al. [7] proposed a shallow network structure Sahllow-UWnet, which can significantly improve the processing speed

© By the Author(s) 2025, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

while maintaining performance. UICE2-Net [8] is the first underwater image enhancement algorithm that uses deep learning in both RGB and HSV color spaces. The network consists of three modules: RGB pixel module, HSV global adjustment module and an attention map module.

Generative adversarial networks and diffusion models generate clear output images from underwater raw images based on the principle of image generation. For example, Li et al. [9] proposed a two-stage generative adversarial network WaterNet, which was trained with a dataset of contrasting images in air and water, and finally achieved color correction of monocular underwater images. Islam et al. [11] introduced a conditional generative adversarial network model FUNIE-GAN, which significantly improved the image generation rate and can enhance underwater images in real time. The test results of multiple datasets confirmed the feasibility of the algorithm. Guan et al. [12] used a conditional denoising diffusion probability model DifWater to enhance underwater images and integrated color compensation as a conditional guide. In order to solve the attenuation and scattering problems of underwater light, Wang et al. [13] used an unsupervised generative adversarial network to generate underwater images. The U-net structure in the network was trained on a synthetic underwater image dataset and performed well on real data, but its robustness in processing underwater images of complex scenes was limited.

In addition to the above methods, strategies based on image fusion have also shown good performance in image enhancement tasks. Ancuti et al. [14] first introduced a fusion strategy to improve the quality of underwater images. In their proposed algorithm, color-corrected and contrast-enhanced images are generated from the original blurred underwater images, and these processed images are used as inputs in the fusion stage to synthesize enhanced images through multi-scale pyramid fusion theory. Subsequently, they improved the performance of the algorithm by optimizing the white balance correction method and introducing the defogging operation, making it more stable in extreme environments such as turbid seawater [15]. Gao et al. [16] also proposed a new underwater image enhancement algorithm based on multi-scale fusion theory, and improved the weight map in the fusion process by imitating the human visual system. Song et al. proposed a strategy that combines multi-scale fusion with global stretching of the model, and adopted an updated saliency weight coefficient method to fuse contrast and spatial cues to improve fusion quality. Kang et al. [17] combined multi-path input, multi-feature fusion, and attention mechanism to propose a high underwater image enhancement framework SPDF, which significantly improved image quality on the dataset.

Considering the cost, weight, and convenience of information acquisition, cameras based on optical imaging are one of the important sensors for spatial position prediction. Intelligent robots, unmanned vehicles, augmented reality devices, etc. on land are generally equipped with one or more cameras to predict their own movement. The optimization-based strategy is widely used in visual SALM. The optimization algorithm can make full use of historical poses and landmarks, and show better performance in large-scale and long-time series tasks. Qin et al. [18] proposed a visual-inertial real-time positioning algorithm VINS-Mono, which only requires a very low-cost monocular camera and an inertial measurement unit IMU to achieve state estimation of the six-degree-of-freedom pose of the body. The algorithm implements a visual-inertial tightly coupled nonlinear optimizer, which calculates a more accurate spatial pose change by minimizing marginal information, inertial measurement residuals, and visual reprojection errors. On this basis, Qin et al. [19] integrated multiple sensors such as GPS signals and depth cameras to realize a multi-sensor pose state estimation algorithm VINS-Fusion based on optimization strategies, which further promoted and applied this strategy.

Although the above algorithms show good performance on land, they have not been tested and verified much in underwater environments. In order to improve underwater visual positioning, some scholars have tried to improve the quality of images. For example, Xin et al. [20] proposed an end-to-end network for SLAM preprocessing in underwater low-light environments. By enhancing low-light images and self-supervised learning to improve feature point matching, the performance of VSLAM based on feature point extraction was effectively improved.

The harsh underwater visual conditions bring certain challenges to the visual positioning algorithm. Although there are related studies on underwater image enhancement, there are not many systems that add image enhancement to underwater visual positioning. Based on this, this paper introduces a multi-scale image enhancement algorithm into the traditional visual inertial positioning algorithm to improve the underwater positioning accuracy. The specific contributions are as follows:

1. The proposed underwater image enhancement preprocessing algorithm consists of two straightforward stages. Initially, a white balance color correction is applied to each sub-image based on the gray world assumption, followed by a restrained contrast enhancement, yielding two optimized sub-images. In the subsequent stage, these sub-images are integrated using multi-scale fusion techniques to produce a final image that is of higher quality and enhanced for clarity.
2. To address the problems of poor underwater image quality, small number of feature extractions and unstable tracking, a multi-scale fusion algorithm is used to enhance image features in the image preprocessing stage, which significantly increases the number of feature point extractions and improves the stability of system positioning.
3. In order to compare the performance of the improved algorithm, this paper runs the VINS algorithm on the real underwater dataset Aqualoc. In the dynamic low-light underwater visual environment, the dataset test results show that the proposed algorithm.

2 METHOD

2.1 Fusion Based Underwater Image Enhancement Algorithm

2.1.1 Color correction

Light in water is absorbed and reflected by the medium and gradually attenuates. Different wavelengths of light attenuate in water to different degrees. As the absorption of visible light by water increases with the increase of wavelength, the blue with the shortest wavelength propagates the longest distance in water, while the red with the longest wavelength propagates the shortest distance in water. Therefore, most underwater images we see appear blue-green [21]. In the color correction stage, we introduced the grayscale world theory, which assumes that in a natural environment, regardless of the color type, the average brightness of the overall image tends to be neutral gray, that is, the mean values of the red, green and blue channels of the image should be approximately equal. If the three channels of the original image satisfy the grayscale world hypothesis, then the following equations are satisfied:

$$\frac{1}{H\times W}\sum_{i=0}^{H-1}\sum_{j=0}^{W-1}R(u,v)=\frac{1}{H\times W}\sum_{i=0}^{H-1}\sum_{j=0}^{W-1}G(u,v)=\frac{1}{H\times W}\sum_{i=0}^{H-1}\sum_{j=0}^{W-1}B(u,v) \tag{1}$$

where i and j represent the values of the horizontal and vertical pixel position index values of the image, respectively, and H and W represent the height and width of the image. When an underwater image I_0 is acquired, its color correction can be performed using the above assumptions. First, the target grayscale value needs to be calculated. In order to obtain the target grayscale value, the average values of the three channels are calculated respectively.

$$\begin{cases}\mu_R=\dfrac{1}{H\times W}\displaystyle\sum_{i=0}^{H-1}\sum_{j=0}^{W-1}I_0^R(u,v)\\ \mu_G=\dfrac{1}{H\times W}\displaystyle\sum_{i=0}^{H-1}\sum_{j=0}^{W-1}I_0^G(u,v)\\ \mu_B=\dfrac{1}{H\times W}\displaystyle\sum_{i=0}^{H-1}\sum_{j=0}^{W-1}I_0^B(u,v)\end{cases} \tag{2}$$

When an underwater image I_0 is acquired, its color correction can be performed using the above assumptions. First, the target grayscale value needs to be calculated. In order to obtain the target grayscale value, the average values of the three channels are calculated respectively.
and then the average value of the three channels is taken:

$$\lambda=\frac{\mu_R+\mu_G+\mu_B}{3} \tag{3}$$

each element is normalized according to the mean to achieve color correction

$$I_0^c(i,j)=\left(\frac{I_0^c(i,j)}{\mu_c}\right)\cdot\lambda(c\in\{R,G,B\}) \tag{4}$$

2.1.2 Contrast increasement

Light in water medium is affected by absorption, scattering and refraction, so the attenuation rate is much faster than that in land environment. In the case of insufficient light, the contrast of the image is low. At the same time, water contains a large number of suspended particles, forming water mist, which makes the object more blurred. In order to solve this problem, the strategy of grayscale equalization is introduced. Artificial light sources are often used for fill light in underwater images, so there will be obvious bright and dark areas. In order to prevent detail loss and highlight overflow caused by excessive enhancement, limited contrast histogram equalization is used. The main idea is to introduce contrast limitation parameters on the basis of the principle histogram equalization, so as to control the output dynamic range, so as to ensure that the detail characteristics of the bright part can be retained while the brightness of the dark part is improved.

Calculate the histogram $H[k]$ of the input image I_0, which represents the probability of the gray value k appearing, and then calculate the cumulative distribution function $F(k)$.

In order to limit the excessive enhancement of contrast, a truncation threshold of contrast limitation is introduced to truncate the part exceeding the threshold to prevent the final image from being overexposed. The cumulative distribution function after processing is:

$$F'(k)=\begin{cases}F(k), & F(k)\le T,\\ T, & F(k)>T.\end{cases} \tag{5}$$

map the corrected cumulative distribution function to the target pixel value range:

$$y(x)=min\big(max\big(M\cdot F'(x),0\big),M\big) \tag{6}$$

where M is the maximum value of the target grayscale range (usually $M=255$). $F'(x)$ is the truncated CDF value (normalized to $[0,1]$).

2.1.3 Image fusion

Complex underwater scenes pose challenges to the effective acquisition of visual information. To address the problems of color distortion, water mist, and low contrast in images, we proposed a white balance algorithm based on the grayscale vision hypothesis and a limited contrast map histogram equalization algorithm. In order to retain the enhanced effective feature information, we introduced a multi-scale image fusion algorithm.

Image Pyramid is a theory that uses a multi-scale hierarchical structure to represent images. The size and clarity of an image are representations of its scale. The bottom of the image pyramid is the original high-resolution image, and each layer upward is the result of downsampling the image. Therefore, the resolution and size of the image are gradually

reduced, and the visual features are gradually refined. The theory was first proposed by Burt et al. [22], who first introduced the image pyramid to achieve efficient decoding and encoding of images.

Assume that an image is described by a matrix with R rows and C columns. Each pixel represents the light intensity I at the corresponding position. The value of light I ranges from 0 to $K-1$. The original image is defined as the 0th layer G_0 of the Gaussian pyramid.

The first layer G_1 of the pyramid is a reduced low-pass filtered version of G_0. Here, the image G_0 is convolved with the Gaussian kernel for smoothing:

$$G_1^{'}(u,v) = \sum_{i=-k}^{k} \sum_{i=-k}^{k} w(i,j) I(u+i, v+j) \tag{7}$$

Where $w(i,j)$ is the weight of the Gaussian kernel.

Then the resulting image $G_1^{'}$ is scaled down, usually by removing all even-numbered rows and columns, to obtain

$$G_1(u,v) = G_1^{'}(2u, 2v) \tag{8}$$

Repeat the above process until the preset Nth layer image G_N is obtained. The Gaussian pyramid performs low-pass filtering on the image in terms of frequency, removes high-frequency information, and retains the overall structure and low-frequency information of the image. On this basis, the Laplacian pyramid can be calculated.

Each layer of the Laplacian pyramid is obtained by subtracting two adjacent layers of the Gaussian pyramid. The G_1 layer image of the Gaussian pyramid is upsampled and smoothed to obtain

$$\widehat{G_0}(u,v) = 4 \sum_{i=-2}^{2} \sum_{j=-2}^{2} w(i,j)\, G_1\left(\frac{u+i}{2}, \frac{v+i}{2}\right) \tag{9}$$

The 0th level L_0 of the Laplacian pyramid is

$$L_0 = G_0 - G_1 \tag{10}$$

Going up layer by layer, we can get L_1, ... , L_{N-1}. The Laplacian pyramid obtains the high-frequency part of the image, that is, the edge details of the original image at different scales (Figure 1).

$$G_k = L_k + \text{upsample}\,(G_{k+1}) \tag{11}$$

The decomposition of the image pyramid is reversible. Through the top-level Gaussian image G_N and the Laplacian pyramid, the following formula is used to calculate layer by layer:

$$G_k = L_k + \text{upsample}\,(G_{k+1}) \tag{12}$$

Finally can be restored to the original original image:

$$G_0 = L_0 + \text{upsample}\,(L_1 + \text{upsample}\,(L2 + \cdots)) \tag{13}$$

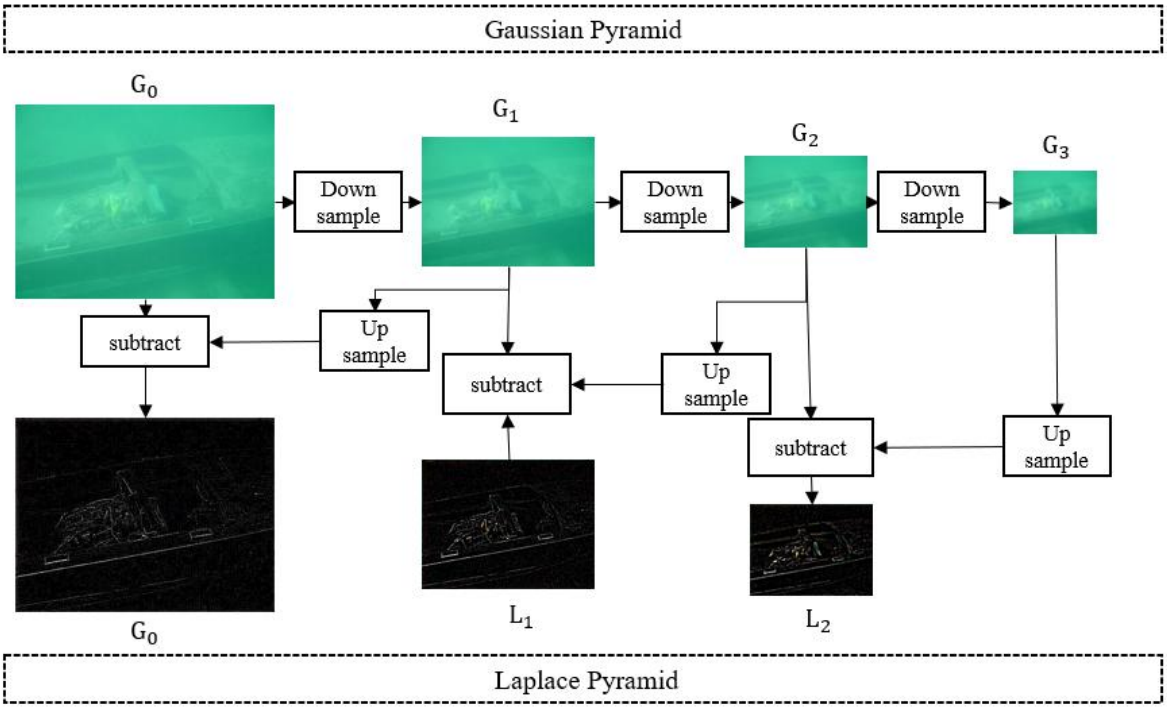

Figure 1 Image Gaussian Pyramid and Laplacian Pyramid Decomposition

The fusion process of the two sub-images requires the introduction of appropriate weight factors. The weight factor here is a matrix of the same size as the image, and the value of each element represents the proportion of the sub-image in the final fused image. In the fusion process, in order to retain high-quality features, the weight is calculated from the sub -image's salient features W_S, local contrast W_{LC}, global contrast W_L, and exposure W_E [15]. Calculate the above weights for the two subgraphs and get the total weight factor of each graph by adding them up

$$W^k = W_L + W_{LC} + W_S + W_E \tag{14}$$

In order to avoid the impact of different orders of magnitude in the weight calculation process, the final weight factor is normalized

$$\overline{W}^k = \frac{W^k}{\sum_{k=1}^{K} W^k} \tag{15}$$

Each input sub-image is decomposed layer by layer into a multi-scale Gaussian pyramid G_1, ... , G_{N-1} by convolution with a Gaussian kernel and downsampling, and then the Laplacian pyramid L_1, ... , L_N .is obtained by subtracting the low-pass filtered image. In the entire pyramid data structure, G_l and L_l each represent the corresponding l-th layer. The final image is calculated by mixing the input image and weights in a multi-layer pyramid manner

$$R_l(x) = \sum_{k=1}^{K} G_l\{\overline{W}_k(x)\} L_l\{I_k(x)\} \tag{16}$$

Where l represents the number of layers of the pyramid, and k is the index value of the input sub-image. Here, the normalized weight factor is decomposed into the Gaussian pyramid $G_l\{\overline{W}_k(x)\}$, and the input sub-image is decomposed into the Laplacian pyramid $L_l\{I_k(x)\}$.

The advantage of using the multi-scale image fusion method is that it can retain the dominant features in the input image during the fusion process. As shown in Figure 2, the upper left corner of the chart shows the input image, and the lower right corner shows the output image. The input image is a typical underwater photo, which is characterized by defects such as color deviation, fog, blur, and insufficient lighting. Through the multi-scale image fusion method, these defects can be effectively improved while retaining or enhancing the important features in the image, thereby obtaining a higher quality output image.

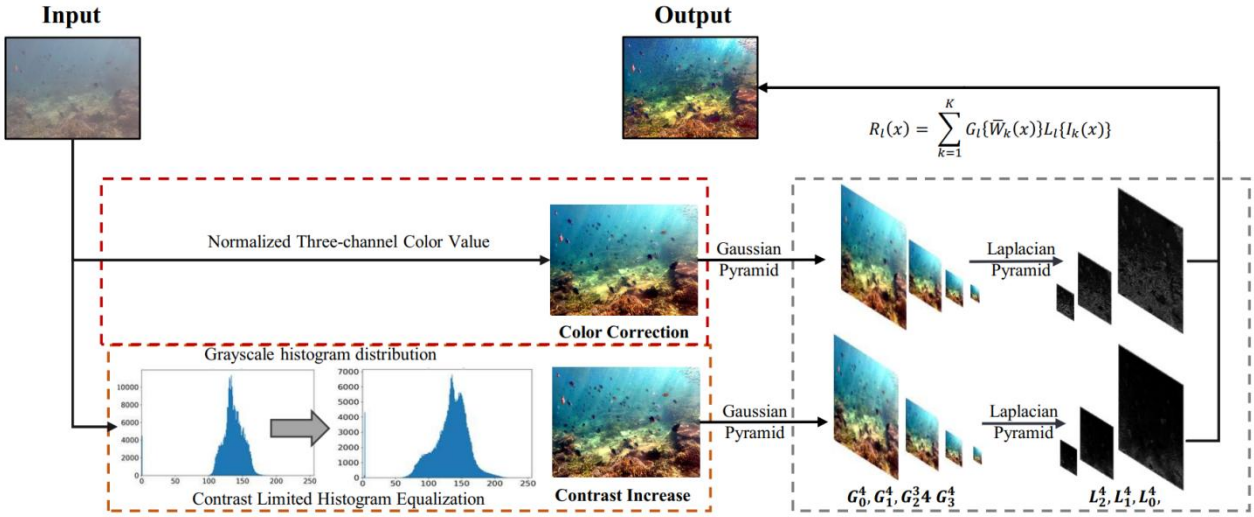

Figure 2 Underwater Fusion Based Image Enhancement Algorithm

2.2 Improve Underwater Visual Front by Image Enhancement

The method of feature point extraction in the VINS algorithm is the Shi-Tomasi algorithm. Its basic idea is to detect feature points from points in the image where the brightness changes dramatically, such as edge intersections, and return the positions of points with large gradient changes in the two-dimensional image as the coordinates of the feature points. However, this algorithm has problems such as low feature density and lack of rich feature description information.

The algorithm performs well on land, but the contrast of underwater images is low due to light scattering and absorption, which leads to smaller gradient changes in the image, weaker corner points, unstable feature detection or insufficient key points detected. In addition, underwater suspended matter can cause visual interference, and Tomasi feature detection is less stable against noise and motion blur, which can lead to detection errors. Based on the above factors, the underwater image enhancement proposed in this paper is adopted. As shown in Figure 3, the enhanced image has been significantly improved in contrast and feature details.

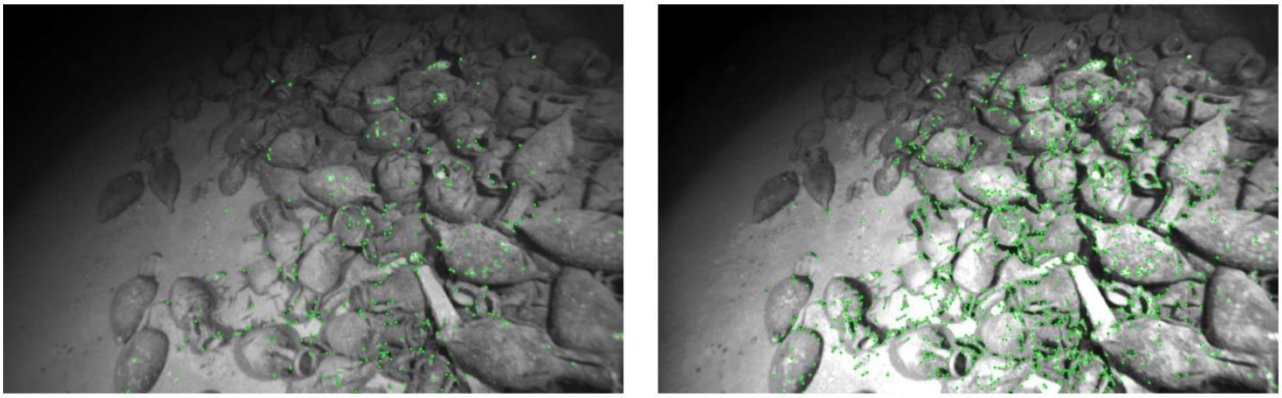

Figure 3 Shi-Tomasi Corner Detection of Images before and after Enhancement

In the original image on the left, the maximum number of feature points detected by Shi-Tomasi is 276, while in the enhanced image on the right, the maximum number of feature points returned by Shi-Tomasi is increased to 1215, which is a 340% increase in the maximum number of feature detections. Due to the enhancement of contrast and the prominence of edge characteristics, the number and quality of corner points detected have been significantly improved. In addition, the temporal consistency of corner points obtained in the enhanced image is improved, which is conducive to subsequent visual tracking and position solving.

In the process of matching visual feature points, although the tracking accuracy is significantly improved by the image enhancement algorithm, there are still a lot of matching errors caused by illumination distortion, weak texture, non-rigid deformation, suspended particles or object occlusion in long-term dynamic changing scenes. The discrete points generated by the above situations will destroy the consistency of geometric constraints, resulting in the failure of subsequent camera pose estimation tasks. In order to solve this problem, after visual feature matching, the random sampling consensus algorithm (RANSAC) is used to filter out the internal points that meet the geometric constraints from the set of candidate matching points containing noise, which is beneficial to the subsequent spatial solution.

The matching feature points in adjacent frames are the projections of the same spatial feature in different images, thus satisfying the epipolar constraint as shown in Figure 4. The matching pixel points in the previous frame image I_A, and the subsequent frame image I_B are denoted as p_1 and p_2. According to the pinhole camera model, the following relationship can be obtained:

$$s_1 p_1 = KP, s_2 p_2 = K(RP + t) \tag{17}$$

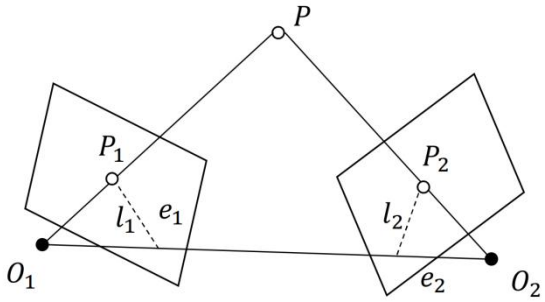

Figure 4 Epipolar Constraints of Feature Points in Adjacent Frames

where K is the camera intrinsic matrix, and R, t represent the rotation matrix and translation vector between the two coordinate systems. Through geometric constraints, we obtain:

$$E = t \times R, F = K^{-T} E K^{-1}, x_2^T E x_1 = p_2^T F p_1 = 0 \tag{18}$$

where F is called the fundamental matrix (Fundamental Matrix), E is called the essential matrix, and by solving the essential matrix, the spatial motion R, t of the camera can be estimated.

$$[x_i \quad y_i \quad 1] F \begin{bmatrix} x_j \\ y_j \\ 1 \end{bmatrix} = 0 \tag{19}$$

In order to find the point pairs that satisfy the geometric constraints in the paired point set, a random sampling consistency check is performed on the point set. Firstly, 8 pairs are randomly selected from the input point pair set, and the geometric constraint formula is expanded into a homogeneous coordinate form.

$$Af = 0 \tag{20}$$

So get the following overdetermined linear equation

Where $A \in R^{n\times 9}$, $f \in R\text{^}9$ is the expansion vector of F. The above equation is solved by the singular value decomposition of A^TA to obtain the least squares solution, so that $\sum_{i=1}^{n} ||Af||^2$ is minimized, and the candidate basic matrix F' is calculated. All point pairs in the pairing point set are verified using geometric constraints:

$$x_i^T \widehat{F} x_j < \varepsilon \tag{21}$$

Where ε is the threshold of the reprojection error. If it is less than this threshold, it is an inlier, and if it is greater than this threshold, it is an outlier. Repeat the above steps and select the model with the largest number of inliers as the final solution. In underwater scenes, a large number of matching points may be mismatched points caused by bubbles, suspended particles or uneven lighting. By introducing the RANSAC algorithm to filter discrete values, only high-confidence matching points that meet geometric constraints are retained.

In order to improve the visual extraction and tracking performance under harsh underwater conditions, this paper preprocesses the original image obtained underwater with a fusion-based image enhancement algorithm. The comparison of the corner point detection results of the original image and the enhanced image shows that the algorithm can greatly improve the quantity and quality of feature point detection. Considering the underwater dynamic environment, a random sampling consistency algorithm is used to perform geometric constraint filtering on the set of matching points to reduce the false pairing caused by suspended particles and lighting effects.

3 EXPERIMENTS & ANALYSIS

This paper proposes a method to improve the VINS visual front-end based on image enhancement to improve its underwater visual feature tracking effect. This section will analyze the algorithm performance in terms of the number of visual feature points extracted and positioning accuracy.

To more fully demonstrate the algorithm's adaptability improvement effect in underwater scenes, we conducted comparative experiments on the Aqualoc dataset. The Aqualoc dataset was collected underwater by a remotely operated vehicle (ROV) equipped with a monocular camera and inertial sensors [23]. The true value of its movement trajectory was reconstructed and estimated in three dimensions using Colmap, providing a benchmark for algorithm evaluation.

3.1 Feature Point Extraction Performance

The Aqualoc dataset consists of sensory data collected in various underwater scenarios using a monochromatic camera, a low-cost MEMS-IMU, and an embedded computer. The archaeological site sequences were recorded in the Mediterranean Sea, off the coast of Corsica. To validate the effectiveness of the aforementioned image enhancement algorithm as a visual front-end for feature extraction and tracking, the third and seventh trajectories from the archaeological site were selected. Shi-Tomasi corner detection was performed on the images from these trajectories to compute the maximum number of high-quality feature points detectable in the images before and after enhancement. The test results are shown in Figure 5 below, where from top to bottom the 3rd and 7th sequences of the archaeological site.

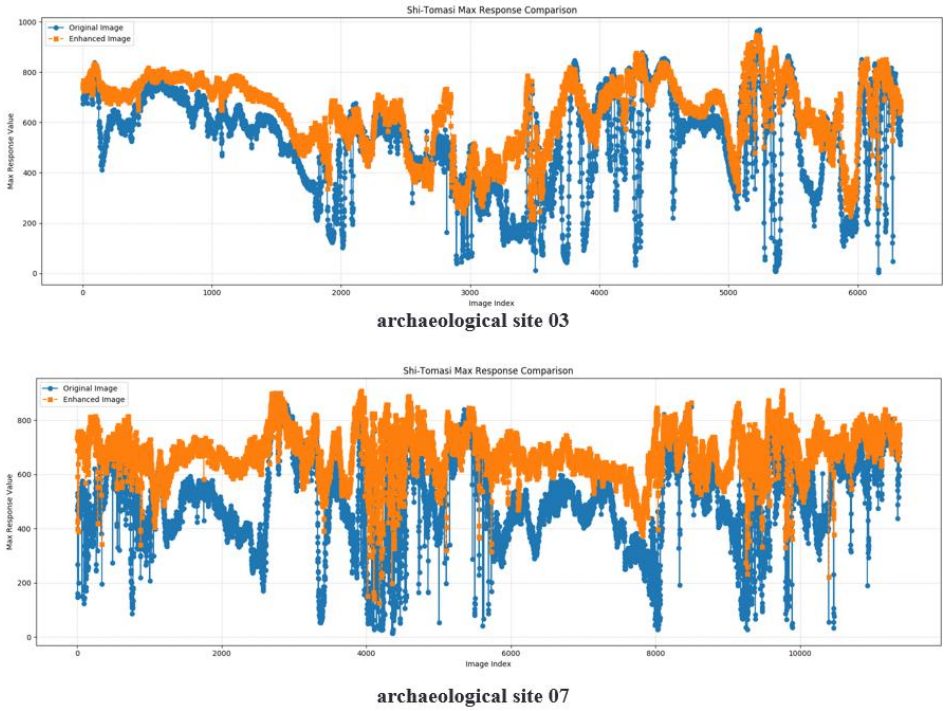

Figure 5 Comparison of Shi-Tomasi Feature Detection Max Num on Archeological Site 03 and 07

The horizontal axis in the figure is the frame index value of the image, and the vertical axis is the maximum number of feature points that can be extracted by the shi-tomasi feature detector with the same parameters. The blue line represents the original image, and the orange line represents the enhanced image. The third scene of the archaeological site in the above figure is located at a depth of about 270m on the real seabed. There is a wreckage of an antique shipwreck here. Most of the ground in this scene is flat, and there are many small rocks, so the visual texture is mainly repeated. In this scene, the turbidity is low, and the rolling of tiny sand grains further hinders the line of sight, making visual feature extraction more difficult. In general, the orange line is higher than the blue line in most cases, especially when the image index is about 4000 and 5300, the improvement effect is very significant, indicating that it is easier to extract high-quality feature points from the enhanced image.

The seventh scene of the archaeological site in the figure below is 380m below the seabed. There is a two-ear bottle mountain and the top of the mountain is several meters higher than the surrounding seabed level, so there is a certain degree of ups and downs. There are low texture characteristics on the sand around the two-ear bottle mountain. Due to the presence of two-ear bottles, there are more marine wildlife, so the environment is very dynamic. It can also be seen in the figure that the orange line is higher than the blue line in most cases, especially when the real image index is 2000, 4000, 6000 to 8000, and there are a large number of blue lines where the maximum number of feature points is less than 200. This situation will be extremely unfavorable for subsequent pose solution. Most of the orange lines of the enhanced image are above 200.

In summary, through the test of the image sequence taken underwater on the real seabed, the results show that after the original image is enhanced based on fusion, the maximum number of features that can be extracted by the same shi-tomasi detector has been improved, which effectively proves the positive role of the image preprocessing algorithm in feature extraction.

3.2 Underwater Visual Positioning Performance

Integrating the fusion-based image enhancement algorithm into the VINS system can effectively improve the stability of the positioning system. The original algorithm and the improved algorithm were tested in multiple scenarios from the underwater dataset of the archaeological site. The operational status is shown in Figure 6. As can be seen from the figure, the improved image algorithm extracts more feature points and obtains richer point clouds, which is beneficial for subsequent pose estimation.

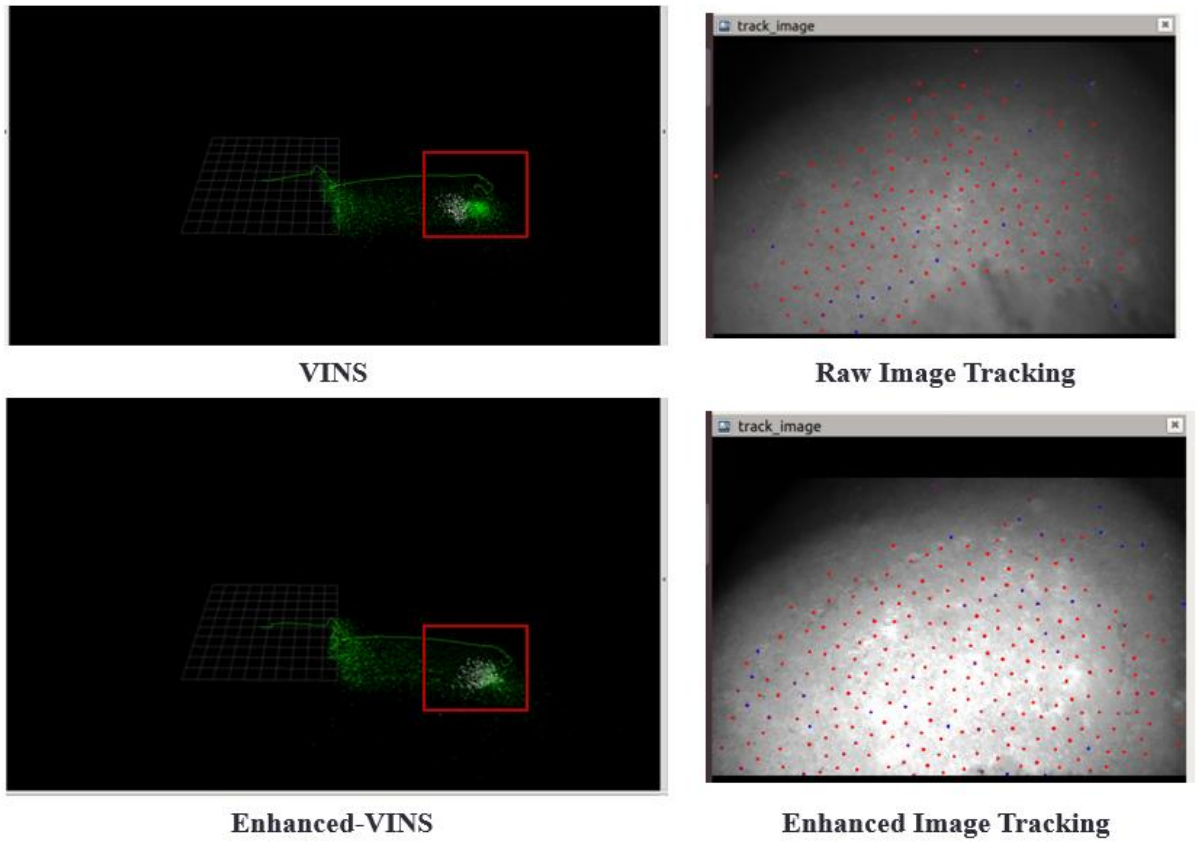

Figure 6 Comparison of VINS and Enahnced-VINS on archaeological site sequence 05

In order to further compare the improvement effect of the image enhancement algorithm on the VINS algorithm, we will perform the improved algorithm and the VINS algorithm on multiple test sequences on the archaeological scene to evaluate their performance differences under different conditions. The test sequences include A2, A5, A6, A7, A8 and

A10. For each sequence, we calculated the minimum (Min), median (Medium) and maximum (Max) absolute trajectory error to comprehensively evaluate the stability and reliability of the two algorithms. The experimental results are shown in Table 1.

Table 1 Absolute Trajectory Error Comparison between VINS and EN-VINS

Sequence	VINS			EN-VINS		
	Min	Medium	Max	Min	Medium	Max
A2	0.110000	1.037429	2.198876	0.181079	0.871403	2.199951
A5	0.172183	0.542419	2.322855	0.029175	0.320848	2.109739
A6	0.169899	0.291294	4.542645	0.008474	0.140198	3.271086
A7	0.067507	0.920890	5.552000	0.110673	1.161439	5.233549
A8	/	/	/	0.062034	0.513904	1.538490
A10	/	/	/	0.130057	0.729954	3.357582

According to the chart analysis, the performance of the EN-VINS algorithm on multiple sequences is better than that of VINS. Specifically, in the four sequences A2, A5, A6 and A7, the minimum, median and maximum ATE values of EN-VINS either surpassed VINS in all aspects or most indicators were ahead of VINS. Especially in the A5 and A6 sequences, EN-VINS not only performed outstandingly in terms of stability (reflected by the smaller maximum value), but also showed significant advantages in terms of accuracy (reflected by the lower median and minimum values). In addition, for the two sequences A8 and A10, VINS system failed to obtain the location information, while EN-VINS maintained the integrity of the data in these two sequences, further demonstrating its higher reliability.

Overall, EN-VINS showed higher comprehensive performance than VINS throughout the test process, providing more reliable results both under normal and extreme conditions. It is particularly noteworthy that EN-VINS can still provide stable and accurate positioning results in the face of situations that may cause VINS positioning failure, which shows that EN-VINS has stronger robustness and adaptability.

4 CONCLUSION

The harsh underwater visual environment is a major challenge for visual positioning algorithms. This paper introduces an image multi-scale fusion strategy to enhance the underwater image preprocessing, which significantly improves the image contrast and detail features. Combining this image enhancement algorithm with the traditional visual inertial positioning algorithm VINS greatly increases the number of feature point extraction, improves the quality of feature point extraction, and achieves more stable visual tracking.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Xia H, Liao F, Bao B, et al. Perspective on Wearable Systems for Human Underwater Perceptual Enhancement, IEEE Transactions on Cybernetics, 2024.
[2] Raveendran S, Patil M D, Birajdar G K. Underwater image enhancement: a comprehensive review, recent trends, challenges and applications, Artificial Intelligence Review, 2021, 54: 5413-5467.
[3] Jian M, Liu X, Luo H, et al. Underwater image processing and analysis: A review, Signal Processing: Image Communication, 2021, 91: 116088.
[4] Pizer S M, Amburn E P, Austin J D, et al. Adaptive histogram equalization and its variations, Computer vision, graphics, and image processing, 1987, 39(3): 355-368.
[5] Wang Y, Zhang J, Cao Y, et al. A deep CNN method for underwater image enhancement, 2017 IEEE international conference on image processing (ICIP). IEEE, 2017: 1382-1386.
[6] Li C, Anwar S, Hou J, et al. Underwater image enhancement via medium transmission-guided multi-color space embedding, IEEE Transactions on Image Processing, 2021, 30: 4985-5000.
[7] Naik A, Swarnakar A, Mittal K. Shallow-uwnet: Compressed model for underwater image enhancement (student abstract), Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(18): 15853-15854.
[8] Wang Y, Guo J, Gao H, et al. UIEC^ 2-Net: CNN-based underwater image enhancement using two color space. Signal Processing: Image Communication, 2021, 96: 116250.
[9] Li J, Skinner K A, Eustice R M, et al. WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. IEEE Robotics and Automation letters, 2017, 3(1): 387-394.

[10] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks//Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.
[11] Islam M J, Xia Y, Sattar J. Fast underwater image enhancement for improved visual perception, IEEE Robotics and Automation Letters, 2020, 5(2): 3227-3234.
[12] Guan M, Xu H, Jiang G, et al. DiffWater: Underwater image enhancement based on conditional denoising diffusion probabilistic model, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, 17: 2319-2335.
[13] Wang N, Zhou Y, Han F, et al. UWGAN: Underwater GAN for real-world underwater color restoration and dehazing, arXiv preprint arXiv:1912.10269, 2019.
[14] Ancuti C, Ancuti C O, Haber T, et al. Enhancing underwater images and videos by fusion, 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012: 81-88.
[15] Ancuti C O, Ancuti C, De Vleeschouwer C, et al. Color balance and fusion for underwater image enhancement, IEEE Transactions on image processing, 2017, 27(1): 379-393.
[16] Guo P, Zeng D, Tian Y, et al. Multi-scale enhancement fusion for underwater sea cucumber images based on human visual system modelling, Computers and Electronics in Agriculture, 2020, 175: 105608.
[17] Kang Y, Jiang Q, Li C, et al. A perception-aware decomposition and fusion framework for underwater image enhancement, IEEE Transactions on Circuits and Systems for Video Technology, 2022, 33(3): 988-1002.
[18] Qin T, Li P, Shen S. Vins-mono: A robust and versatile monocular visual-inertial state estimator, IEEE transactions on robotics, 2018, 34(4): 1004-1020.
[19] Qin T, Cao S, Pan J, et al. A general optimization-based framework for global pose estimation with multiple sensors, arXiv preprint arXiv:1901.03642, 2019.
[20] Xin Z, Wang Z, Yu Z, et al. ULL-SLAM: underwater low-light enhancement for the front-end of visual SLAM, Frontiers in Marine Science, 2023, 10: 1133881.
[21] Singh N, Bhat A. A systematic review of the methodologies for the processing and enhancement of the underwater images, Multimedia Tools and Applications, 2023, 82(25): 38371-38396.
[22] Burt P J, Adelson E H. The Laplacian pyramid as a compact image code, Readings in computer vision. Morgan Kaufmann, 1987: 671-679.
[23] Ferrera M, Creuze V, Moras J, et al. AQUALOC: An underwater dataset for visual–inertial–pressure localization, The International Journal of Robotics Research, 2019, 38(14): 1549-1559.