applied sciences

MDPI

Article

An Improved SIFT Underwater Image Stitching Method

Haosu Zhang, Ruohan Zheng, Wenrui Zhang, Jinxin Shao and Jianming Miao *

Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), School of Ocean Engineering and Technology, Sun Yat-sen University, Zhuhai 519000, China; zhanghs7@mail.sysu.edu.cn (H.Z.); zhengrh6@mail2.sysu.edu.cn (R.Z.); zhangwr23@mail2.sysu.edu.cn (W.Z.); shaojx8@mail2.sysu.edu.cn (J.S.)
* Correspondence: miaojm@mail.sysu.edu.cn

Abstract: Underwater image stitching is a technique employed to seamlessly merge images with overlapping regions, creating a coherent underwater panorama. In recent years, extensive research efforts have been devoted to advancing image stitching methodologies for both terrestrial and underwater applications. However, existing image stitching methods, which do not utilize detector information, heavily rely on matching feature pairs and tend to underperform in situations where underwater images contain regions with blurred feature textures. To address this challenge, we present an improved scale-invariant feature transform (SIFT) underwater image stitching method. This method enables the stitching of underwater images with arbitrarily acquired images featuring blurred feature contours and that do not require any detector information. Specifically, we perform a coarse feature extraction between the reference and training images, and then we acquire the target image and perform an accurate feature extraction between the reference and target images. In the final stage, we propose an improved fade-in and fade-out fusion method to obtain a panoramic underwater image. The experimental results show that our proposed method demonstrates enhanced robustness, particularly in scenarios where detecting feature points is challenging, when compared to traditional SIFT methods. Additionally, our method achieves higher matching accuracy and produces higher-quality results in the stitching of underwater images.

Keywords: SIFT; image stitching; image alignment; image preprocessing; point matching

check for updates

Citation: Zhang, H.; Zheng, R.; Zhang, W.; Shao, J.; Miao, J. An Improved SIFT Underwater Image Stitching Method. *Appl. Sci.* **2023**, *13*, 12251. https://doi.org/10.3390/app132212251

Academic Editor: Atsushi Mase

Received: 7 October 2023
Revised: 7 November 2023
Accepted: 8 November 2023
Published: 12 November 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The process of stitching underwater images finds extensive applications in the realms of undersea exploration, marine archaeology, marine biological research, and the visualization of wrecks. Underwater conditions are characterized by the attenuation and backscattering of light, which restricts the field of view that cameras can capture. Consequently, it becomes unfeasible to encompass a larger area within a single frame. Nonetheless, panoramic imaging offers a solution by amalgamating images from diverse viewpoints into a single composite image that spans a wider area [1]. This approach furnishes a more comprehensive reservoir of image data for subsequent research endeavors.

In recent years, significant advancements have been made in the fields of image stitching, mosaic construction, and full three-dimensional (3D) reconstruction. However, it is important to emphasize that these methods are primarily designed for terrestrial scenarios and do not directly address the unique challenges encountered in underwater imaging [2]. Underwater image stitching presents distinct challenges compared to its land-based counterpart. The first major challenge lies in obtaining clear underwater images, given the complex and dynamic underwater environment. Underwater images are susceptible to blurring caused by water flow and the presence of microorganisms in the aquatic medium [3]. The second challenge is associated with the complexity of underwater subjects. Underwater elements such as marine life, rocks, and wrecks often lack distinct texture features, resulting in blurry outlines that are difficult to extract. Additionally, marine organisms like seaweed, oysters, sea snails, earthworms, and sea grass can attach to

shipwrecks, further obscuring their outlines and features. This complicates the accuracy of feature matching. Furthermore, the movement of marine life due to water currents can introduce alignment errors and ghosting artifacts in underwater images. The third challenge pertains to the prevalence of flat areas in underwater images, especially on extensive ocean floors [4]. These regions exhibit similar textures around feature points, making it challenging to extract effective feature points. The fourth challenge is related to the assumptions underlying mosaic construction techniques, which often assume that the images are captured by a purely rotating camera, or that the scene is planar [5]. These assumptions imply the absence of parallax caused by camera movement and a lack of three-dimensional effects. However, underwater images are subject to light attenuation and backscattering, which can result in significant distortions in the mosaic. These distortions undermine the applicability of these assumptions in underwater contexts.

To address the challenges of underwater image stitching, in recent years, many researchers have conducted work on underwater image stitching and video stitching. Underwater mosaicking has predominantly been explored within the realm of vision-based navigation and maintaining a stable position near the seafloor [6]. However, these approaches are typically tailored to specific conditions and are unable to seamlessly stitch together any set of acquired images without external guidance [7]. When it comes to underwater images captured by divers, they often lack the regularity of fixed time or distance intervals, and sensor data are often unavailable. Consequently, to facilitate the stitching of arbitrary underwater images, many researchers have turned to image registration methods like SIFT and SURF. These methods rely solely on image information and do not necessitate the use of detection devices or impose any special requirements specific to underwater images. Nonetheless, it is important to note that these image stitching methods heavily hinge on the information derived from matching pairs of distinctive image features. This reliance places significant demands on both the quantity and accuracy of these feature matches [8]. Insufficient feature matches can pose a considerable challenge, especially in the context of underwater environments. Scale-invariant feature transform (SIFT) and speeded-up robust features (SURF) stand out as potent tools in terms of their ability to handle image transformation and deformation invariance. Lowe [9] proposed a method for extracting unique invariant features from images, which are invaluable for reliable matching across different objects or scenes. These features maintain their invariance to image scaling and rotation, and remarkably, SIFT remains robust even in the presence of variations in image illumination or affine distortion. SIFT's capacity to generate a substantial number of feature vectors, even from a small number of objects, renders it particularly suited for precise matching within large-scale feature databases. Bay et al. [10] introduced the SURF algorithm in response to the slow computational speed of SIFT. However, SURF exhibits lower stability and matching accuracy compared to SIFT, making it less effective in the context of stitching underwater images, particularly in regions with blurred features. In the deep sea, where underwater light rapidly diminishes due to attenuation, natural illumination becomes scarce beyond just a few tens of meters in depth. Consequently, most optical imaging in the deep sea relies on artificial lighting, often resulting in uneven illumination. Therefore, the use of SIFT for image registration in underwater image stitching is highly appropriate. In regions with rich and stable texture features, SIFT excels at extracting a sufficient number of features, enabling the generation of accurate matching pairs. However, underwater stitching frequently encounters challenges in flat areas with limited texture, unclear target contours in captured images, and the presence of underwater microorganisms, all of which can further complicate image quality. Solely relying on SIFT for feature extraction in such cases may result in an insufficient feature set and an inadequate number of accurate matching pairs, ultimately leading to alignment errors and ghosting artifacts in the stitching process. In contrast to existing stitching techniques, our goal is to introduce an underwater image stitching method that is capable of seamlessly merging arbitrarily acquired underwater images. This method remains effective even when the images exhibit blurred feature contours, and it does not depend on detector-related information. In this

paper, we present an enhanced SIFT underwater image stitching method designed to address images with blurred feature contours without relying on detector information.

The overview of our approach is illustrated in Figure 1 and consists of three stages: the underwater image preprocessing stage, the SIFT-based coarse matching stage, and the SIFT-based precise matching stage. In the first stage, we employ the dynamic threshold white balance method and the contrast limited adaptive histogram equalization (CLAHE) algorithm to adjust colors and enhance brightness. In the second stage, we perform coarse feature extraction using the SIFT algorithm on both the reference image (image A) and the training image (image B). The training image represents the image to be transformed into the target image. Then, we match the features in both images using the nearest neighbor ratio algorithm and the random sample consensus (RANSAC) algorithm. Following this matching process, we rotate the training image using the homography matrix and correct the coordinate origin. Finally, any redundant blank areas are removed to obtain the target image (image C). In the third stage, we conduct accurate feature extraction based on the SIFT algorithm for both the reference image (image A) and the target image (image C). This precise feature extraction is followed by matching the features in both images, culminating in an image-paste-type fusion stitching. Ultimately, the panoramic image is created using the enhanced fade-in and fade-out method for seamless fusion. We employ a weighted smoothing algorithm to achieve the stitching of the target image. The weights assigned to the two images intended for merging undergo a gradual transition from 0 to 1 in either the horizontal or vertical direction, thereby achieving a seamless transition in the fused image. In our experiments, we present compelling evidence demonstrating that our proposed method consistently yields excellent results, attesting to its effectiveness in the domain of underwater image stitching.

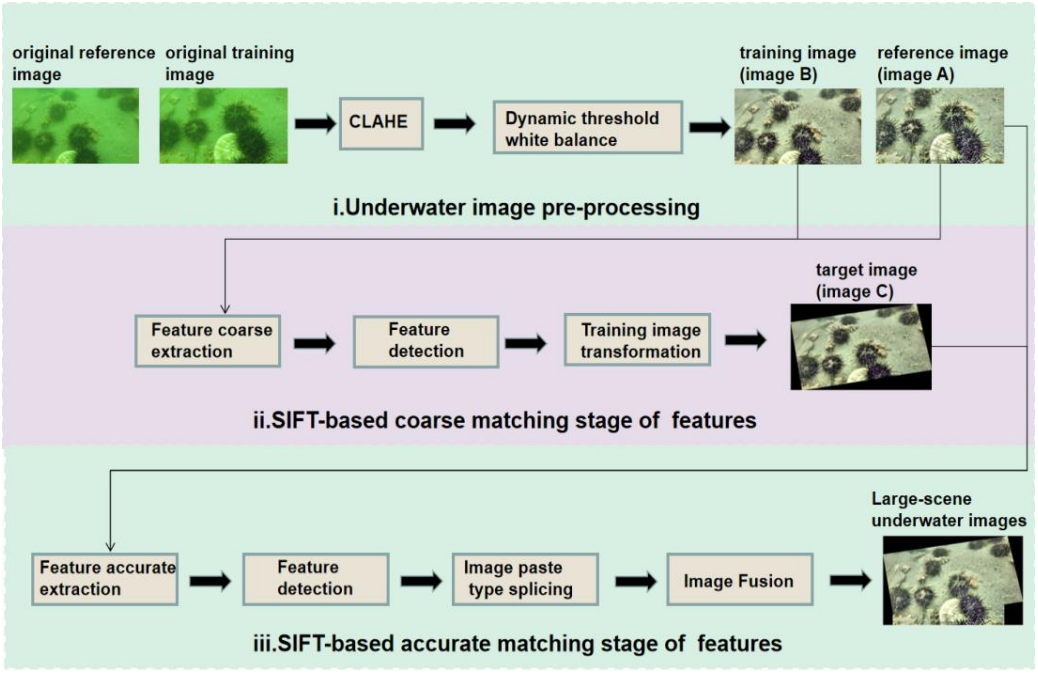


Figure 1. Various stages in the improved SIFT underwater image stitching technique.

The primary contribution of this paper is the development of an enhanced SIFT underwater image stitching method strategically designed to minimize the ghosting artifacts commonly associated with stitching. Our approach excels even in areas characterized by blurred texture features, where traditional feature extraction methods struggle. By significantly increasing the number of accurate matching pairs, we enhance the matching accuracy, improve the alignment of feature points between images, reduce stitching errors, and ultimately elevate the quality of underwater panoramic images. In a series of ten un-

derwater image stitching experiments, our proposed method outperformed the traditional SIFT algorithm, increasing the non-reference underwater image quality measure (UIQM) by an average of 6.04%. Furthermore, when compared to references [11–13], our method enhances the UIQM index by 4.73%, 9.01%, and 11.06%, respectively.

The remainder of this paper is structured as follows: We describe the related works in Section 2. In Section 3, we describe our proposed improved SIFT underwater image stitching method. Section 4 describes the experiments. Finally, we present our conclusions in Section 5.

2. Related Work

In this section, we briefly review image stitching methods, both on land and underwater. While research on underwater image stitching has been relatively limited in recent years, notable advancements have been achieved in land-based image stitching methods. In land-based image stitching, a spectrum of approaches has been put forward to enhance local alignment. This includes the computation of multi-content-aware local warping as opposed to a singular global warping strategy. Moreover, some methodologies advocate for the use of seam-driven image stitching strategies aimed at mitigating artifacts resulting from image fusion [14,15].

Lowe et al. [16] proposed the AutoStitch image stitching algorithm, which leverages a global single-response matrix to establish correspondences across all images and generate a panoramic result through multiband fusion. This approach enabled the automated stitching of multiple images. However, a critical constraint of the AutoStitch algorithm is its reliance on the assumption of a consistent scene depth across overlapping image areas. Deviations from this assumption can lead to pronounced ghosting and alignment errors. To mitigate this limitation, researchers have proposed employing more robust matrices or multiple single-response matrices in place of a single global one. Lin et al. [17] proposed the smoothly varying affine (SVA) method, which incorporates smoothly varying affine stitching fields instead of global affine transformations. This technique exhibits superior local deformation and alignment capabilities and can also handle parallax to some extent. Gao et al. [18] proposed the dual-homography warping (DHW) approach, which divides the scene into a background plane and a foreground plane. Two distinct single-response matrices are employed, each aligned to the background and foreground, facilitating seamless stitching of a broader range of realistic scenes.

Zaragoza et al. [11] proposed As-Projective-As-Possible (APAP). Firstly, SIFT features are extracted from two images, and erroneous features are removed using RANSAC. Then, an efficient computational method, moving DLT, is given to estimate the global homography matrix. When two images exhibit variations beyond mere rotation, or when they depict nonplanar scenes, moving DLT employs a location-dependent homography to warp each matching point. In the APAP algorithm, the image is divided into dense grids according to a certain pattern. Higher weights are provided for closer feature points, and lower weights are provided for more distant feature points. Each grid is aligned with a local single-response matrix, which is calculated from the global single-response matrix and the grid weights. Finally, the image stitching is performed using the local homography matrix. The APAP algorithm can align multiple images better via local weight constraint adjustment. Based on this, many researchers have been inspired to propose the use of grid optimization to achieve image stitching. However, since APAP performs affine transformation in non-overlapping regions of the image as well, it can cause severe visual distortion. Chang et al. [12] proposed shape-preserving half-projective (SPHP), which is based on the shape-preserving method of image scaling from the perspective of shape correction. The algorithm performs distorted alignment with single-strain transformations in overlapping regions and maintains the perspective of the image in non-overlapping regions with similarity transformations. At the same time, a smooth transition is formed between the overlapping and non-overlapping regions. The SPHP algorithm can reduce the perspective distortion in non-overlapping regions. The authors also proposed combining

SPHP with APAP to make the proposed warping more robust. Lin et al. [13] proposed adaptive As-Natural-As-Possible (AANAP). This method uses a smooth stitching field over the entire target image while taking into account all local transformation variations. It is less dependent on the choice of parameters and can automatically compute the appropriate global similarity transform. The AANAP algorithm can also mitigate the perspective distortion in non-overlapping regions by linearizing the homography and slowly changing it into a global similarity transform.

Zhang et al. [19] developed an effective method to find a local alignment for optimal image stitching. This local stitching method uses a hybrid alignment model that simultaneously employs homography and content-preserving warping. Jing et al. [20] proposed a parallax-tolerant image stitching method based on robust elastic warping. They constructed an analysis warping function to eliminate parallax errors. Then, the input images are warped over the meshed image plane according to the calculated deformations. However, the effectiveness of these land-based methods in stitching images with blurry underwater conditions remains to be demonstrated.

Due to the complexity of the underwater environment, underwater image stitching techniques pose significant challenges. For more than thirty years, researchers have been studying to create automatic mosaics for underwater applications. For underwater mosaicking, some studies focus on vision-based navigation, while others focus on building large mosaics [21]. Some algorithms can achieve image stitching but require assistance from external detectors or have specific requirements for underwater image acquisition. Fortunately, SIFT-based stitching methods can stitch images using only image information, but they require a high number and accuracy of feature matches [3]. Insufficient feature matching can lead to stitching errors.

One of the earliest computer-assisted underwater image stitching systems was proposed by Haywood [22] in 1986. This is a manual system that uses camera positions to directly obtain the position relationship and coordinates of adjacent frames. The motion parameters between frames can be computed directly from the camera positions, without requiring any registration. However, this system is not automatic or real-time.

Negahdaripour et al. [6] developed a method for estimating camera motion in real time directly from the luminance variations of the image sequence. In [23], direct estimates of motion were successfully applied in underwater mosaic schemes. To achieve video mosaics along unconstrained vehicle paths, Fleischer et al. [24] utilized optimal estimation theory and smoother–follower techniques to identify the errors and reduce the image alignment errors propagating through the image chain. In [25], Fleischer et al. extended the theory of smoother–follower estimation to achieve real-time creation and online improvement of unconstrained underwater video mosaics. However, the image estimates depend on the sensor data and the system geometry. Richard et al. [7] used real-time image registration to achieve real-time creation of underwater video mosaics. Relative offsets between a pair of images can be used to determine when a new image should be acquired and where it belongs in the mosaic. In order to obtain images suitable for creating mosaics, the images are acquired snapping at a fixed time interval, snapping at a fixed position interval, or snapping at a fixed visual interval. Gracias et al. [26] proposed an integrated approach to solve the problem of pose estimation and mosaic construction in underwater mosaicking. This method uses a simple image motion model to compute the image motion and focus on the use of long image sequences with temporal distance overlays, such as those generated by circular trajectories or zigzag scanning patterns. However, this approach is based on the assumption that the dominant motion existing between consecutive frames is translation, and it does not account for lens distortion.

However, these approaches are unable to construct underwater mosaics without detector information or on arbitrarily obtained underwater images. Therefore, researchers have conducted further research on underwater mosaics of arbitrary images by using only the intrinsic information of the images.

Leone et al. [22] introduced a fully automated feature-based method for creating seamless, panoramic 2D mosaics from a sequence of images. Their texture-based approach, combined with statistical filtering, ensures accurate correspondences between adjacent frames and facilitates the removal of outliers. Raut et al. [8] proposed an enhancement to the SIFT algorithm to improve the automatic recognition of underwater images. Their method incorporates a Gabor filter as a pre-filter and utilizes the Hausdorff distance to compute the distance between key points, resulting in the detection of more key points and improved matching. Pizarro et al. [27] presented an approach for large-area underwater mosaicking, addressing challenges related to low overlap, unstructured motion, and varying lighting conditions in underwater images. Their method involves three key steps: determining image homography by estimating and compensating for radial distortion, topology estimation through feature-based pairwise image registration using a multiscale Harris interest point detector and a Zernike moment-based feature descriptor, and global alignment of all images based on the initial alignment obtained from the pairwise estimation. However, this approach heavily relies on effective feature detection and matching phases.

As far as our knowledge extends, there remain significant challenges in detecting a sufficient number of accurate feature pairs when the image feature contours are blurred. Insufficient feature matches can lead to errors in the mosaic.

3. Materials and Methods

To perform the stitching of underwater images with blurred feature contours and without relying on detector information, we present an enhanced stitching method tailored for SIFT underwater images. As illustrated in Figure 1, our approach encompasses three distinct stages, each serving a specific purpose. In this section, we commence by detailing the preprocessing steps applied to underwater images. Subsequently, we provide an overview of the traditional SIFT method, a widely employed technique in image stitching. Finally, we introduce our proprietary stitching algorithm.

3.1. Underwater Image Pre-Processing

When light traverses through water, a portion of it is reflected, while the remainder penetrates the water and is selectively absorbed. Different wavelengths of light experience varying degrees of absorption in water, with red light being significantly attenuated compared to blue–green light. Consequently, underwater images predominantly exhibit blue–green hues and tend to grow darker with increasing depth. Moreover, underwater images are susceptible to blurring and haziness due to the scattering of light by water, which can impair the accuracy of feature point extraction during the stitching process. Therefore, preprocessing of underwater images plays a crucial role, enabling the extraction of a larger number of feature points with enhanced precision.

Underwater image preprocessing can be categorized into two main types: image enhancement methods, and image restoration techniques. While image restoration methods necessitate prior knowledge of the environmental conditions, image enhancement methods offer a more straightforward approach that does not rely on such specific information. As such, this paper employs image enhancement methods. Specifically, dynamic threshold white balance and CLAHE are utilized for the preprocessing of underwater images. The CLAHE method [28] enhances local contrast in images by applying a threshold to limit contrast amplification within localized regions, addressing issues such as noise overamplification that may occur in the adaptive histogram equalization (AHE) method [29]. The dynamic threshold white balance method [30], rooted in the YCbCr color space, offers image enhancement capabilities that enable adjustments to image tones and the restoration of accurate color representation. This ensures faithful rendering of the true colors of the captured subject in the images obtained by the camera.

In this paper, we improve the quality of underwater images by optimizing their color and contrast. The dynamic threshold white balance method is used to adjust the color,

while the CLAHE algorithm enhances the contrast and brightness. We set the contrast limit parameter of CLAHE to 2 and use a 4×4 grid size to block the image.

We selected a set of captured images as the preprocessed dataset. The results are shown in Figure 2. As can be seen, there was a significant improvement in image quality. Moreover, we used different methods to preprocess the images and perform image feature point matching. We compared our method with the integrated color model (ICM) [31], underwater light attenuation prior (ULAP) [32], underwater dark channel prior (UDCP) [33], and image blurriness and light absorption (IBLA) [34]. The results of the underwater image quality assessment are summarized in Table 1. The accuracy rate of interest point matching is improved in the preprocessed images, with more features correctly matched. Specifically, our method improves the accuracy rate from 44.76% in the original images to 59.06% in the preprocessed images. Compared with other algorithms, our algorithm has higher accuracy in feature matching. Overall, our proposed method effectively enhances the detection of interest points in underwater images.

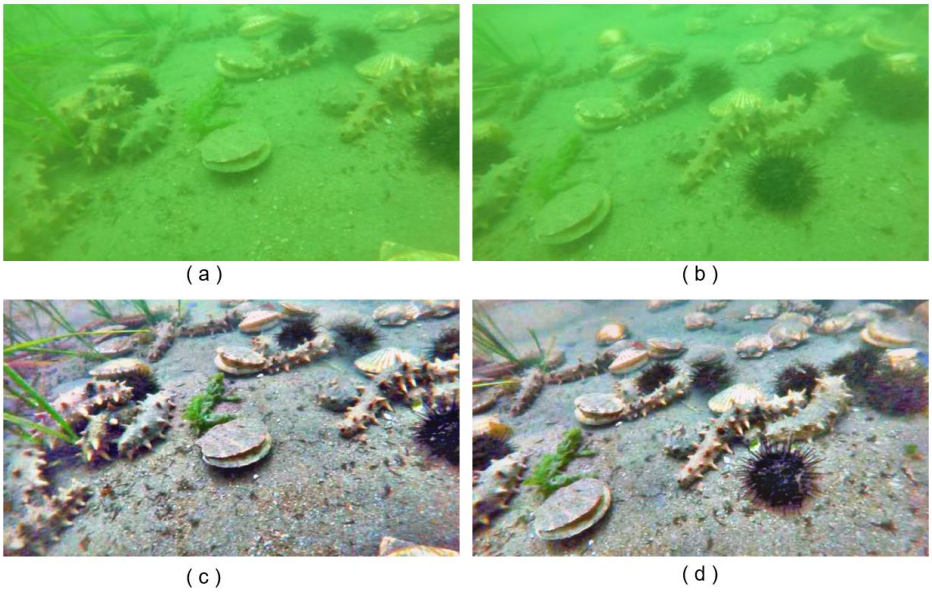


Figure 2. (**a**,**b**) The original underwater images; (**c**,**d**) the preprocessed underwater images generated by our method.

Table 1. Underwater image quality evaluations.

Database	Tentative Matches	Inlier Matches	Accuracy Rate
Original images	210	94	44.76%
ICM	187	103	55.08%
ULAP	191	103	53.93%
UDCP	151	71	47.02%
IBLA	166	89	53.61%
Our method	171	101	59.06%

3.2. Underwater Image Stitching

3.2.1. The Traditional SIFT Method

As described in the literature [9], the SIFT algorithm utilizes a 128-dimensional feature vector to describe feature points in an image with scale, rotation, and translation invariance. The matching process is based on the similarity between these feature point descriptors.

Firstly, SIFT detects preliminary extrema points through the construction of a Gaussian difference pyramid. It then filters low-contrast and edge-response points to obtain robust key points. Subsequently, SIFT assigns an orientation to each key point, calculates the gradient magnitude and orientation, and generates orientation gradient histograms. The

radius of the image area to be calculated is then determined, and the coordinate axis is rotated to align with the main direction of the key point. An 8×8 window surrounding the key point is selected as the feature descriptor, and a feature point descriptor is generated [9]. Finally, SIFT performs feature point matching.

In the feature point matching, the reference image (image A) is denoted by I and the training image (image B) is denoted by I'. The expression is as follows:

$$\mathbf{x}' = \mathbf{H}\mathbf{x}, \tag{1}$$

where $\mathbf{x} = \begin{bmatrix} x & y \end{bmatrix}^T$ and $\mathbf{x}' = \begin{bmatrix} x' & y' \end{bmatrix}^T$ denote the pixel coordinates of the reference image and the pixel coordinates of the training image, respectively, while $\mathbf{H} \in \mathbb{R}^{3\times3}$ represents the homography matrix. The flush coordinates of the target image and the reference image pixel points are expressed as follows:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{2}$$

The homography matrix has a degree of freedom of 8. Therefore, we assume that $h_9 = 1$, and the above equation can be organized as follows:

$$x' = \frac{h_1 x + h_2 y + h_3}{h_7 x + h_8 y + 1} \tag{3}$$

$$y' = \frac{h_4 x + h_5 y + h_6}{h_7 x + h_8 y + 1}. \tag{4}$$

Organize Equations (3) and (4) into the following form:

$$\begin{bmatrix} x_1 y_1 1 & 0 & 0 & 0 & -x_1 x_1' & -y_1 x_1' & -x_1' \\ 0 & 0 & 0 & x_1 y_1 & 1 & -x_1 y_1' & -y_1 y_1' & -y_1' \end{bmatrix} \begin{bmatrix} h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 & 1 \end{bmatrix}^T = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{5}$$

where $\begin{bmatrix} x_1 & y_1 \end{bmatrix}^T$ and $\begin{bmatrix} x_1' & y_1' \end{bmatrix}^T$ indicate the coordinates of the first set of matched pairs. When we obtain the coordinates of the 4 sets of matched pairs, we can solve the above system of equations to obtain the homography matrix.

3.2.2. The Proposed Improved SIFT Method

An improved SIFT method is proposed to address the drawback of only performing feature extraction once in the traditional SIFT algorithm. The reference image (image A) is denoted by I and the target image (image C) is denoted by I''. The expression is as follows:

$$\mathbf{x}'' = \mathbf{H}'\mathbf{H}\mathbf{x} \tag{6}$$

$$\begin{bmatrix} x'' \\ y'' \\ 1 \end{bmatrix} = \begin{bmatrix} h_1' & h_2' & h_3' \\ h_4' & h_5' & h_6' \\ h_7' & h_8' & h_9' \end{bmatrix} \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{7}$$

where $\mathbf{x} = \begin{bmatrix} x & y \end{bmatrix}^T$ and $\mathbf{x}'' = \begin{bmatrix} x'' & y'' \end{bmatrix}^T$ denote the pixel coordinates in the reference image and the pixel coordinates in the target image, respectively, $\mathbf{H} \in \mathbb{R}^{3\times3}$ represents the homography matrix from the training image (image B) to the reference image (image A), and $\mathbf{H}' \in \mathbb{R}^{3\times3}$ represents the homography matrix from the target image (image C) to the training image (image B). Equation (7) shows the projection transformation of the target image (image C) to the reference image (image A). Again, the above equation can be organized as follows:

$$x'' = \frac{(h'_1h_1 + h'_2h_4 + h'_3h_7)x + (h'_1h_2 + h'_2h_5 + h'_3h_8)y + (h'_1h_3 + h'_2h_6 + h'_3)}{(h'_7h_1 + h'_8h_4 + h_7)x + (h'_7h_2 + h'_8h_5 + h_8)y + (h'_7h_3 + h'_8h_6 + 1)} \tag{8}$$

$$y'' = \frac{(h'_4h_1 + h'_5h_4 + h'_6h_7)x + (h'_4h_2 + h'_5h_5 + h'_6h_8)y + (h'_4h_3 + h'_5h_6 + h'_6)}{(h'_7h_1 + h'_8h_4 + h_7)x + (h'_7h_2 + h'_8h_5 + h_8)y + (h'_7h_3 + h'_8h_6 + 1)}. \tag{9}$$

During the stitching phase, we undertake a two-step approach involving coarse feature extraction and precise feature extraction, both based on the SIFT algorithm. Initially, coarse feature extraction is conducted on the reference image (image A) and the training image (image B). Subsequently, we employ the k-nearest neighbor algorithm (KNN) to gauge the similarity of key points between these two images. The KNN method utilizes the Euclidean distance as the measure for assessing feature point similarity. It performs feature point matching by systematically traversing all key points in the two images slated for concatenation. The Euclidean distance between any two corresponding feature points within these two images to be joined (D) can be defined as follows:

$$D(F_Q^i, F_T^j) = \sqrt{\sum_{k=1}^{K} \left(q_k^i - t_k^j\right)^2}, \tag{10}$$

where F_Q^i and F_T^j represent feature points in the reference and training images, respectively, K represents the dimensions of the feature points, $F_Q^i = (q_1^i, q_2^i, \ldots, q_k^i)$, $i \in [1, 2, \ldots, M]$, M is the number of feature points in the reference image, $F_T^j = (t_1^j, t_2^j, \ldots, t_k^j)$, $j \in [1, 2, \ldots, N]$, and N is the number of feature points in the training image. Next, we iterate through all of the feature points in the reference image, identifying those within the training image that possess the closest Euclidean distances t_1 and the second- closest Euclidean distances t_2. After applying this selection process to the feature points, we can establish potential matching pairs. The calculation formula is provided below:

$$result = \frac{D(q, t_1)}{D(q, t_2)} - V_f, \tag{11}$$

where V_f represents a predetermined threshold, typically within the range of 0.4 to 0.8. In this paper, we set V_f to 0.75. For our underwater images, this threshold was thoughtfully selected to maintain a specific quantity of feature matching pairs while ensuring a high level of matching accuracy. If *result* is below 0, it signifies that the distance ratio is below the defined threshold, indicating a successful match of feature pairs. Conversely, if *result* exceeds 0, it implies that the distance ratio surpasses the specified threshold, and subsequently, the removal of the potential matching pair.

To further improve the accuracy of feature alignment and stitching, the RANSAC algorithm is employed. Through multiple iterations, it fits the optimal mathematical model to eliminate matching pairs with significant errors. Firstly, a subset of elements is selected from the set of coarsely matched feature points and designated as the inlier group. Within the inlier group, each element is incorporated into the model to compute the model parameters. The calculation formula of the model is shown in Formula (2). Secondly, the remaining points from the feature point set, excluding those in the inlier group, are inserted into the parameter model obtained in the previous steps. The distance from the model is computed for each element point, and points that fall within the allowable error range are marked as inlier elements, while those outside this range are designated as outliers. The error threshold is set to 4. Utilizing all of the element points contained within the inlier group at this stage, we repeat steps 1 and 2 N times, where the expression for the number of iterations, N, is as follows:

$$N = \frac{\lg(1-p)}{\lg(1-\phi^m)}, \tag{12}$$

where p represents the confidence level, typically in the range of 0.95 to 0.99, ϕ represents the proportion of inlier element points in the dataset, and m represents the total number of feature points in the dataset. The best-fitting model, containing the most inlier element points during the iterative process, is chosen as the optimal model. The element points within this inlier group at this stage are considered to be the final feature point matches, and the corresponding homography matrix **H** is calculated.

Once the homography matrix **H** for image A and image B is obtained, multiply this matrix with image B and conduct a coordinate system correction. The correction aligns two of the four vertices of the training image with the X and Y axes, ensuring that all coordinate points are positive. This enhancement significantly accelerates the computational speed of subsequent stitching. Finally, the superfluous blank areas resulting from the rotation are eliminated to yield the target image (image C). The proposed SIFT-based coarse matching stage is shown in Algorithm 1. The SIFT-based coarse matching stage employs the next input parameters—the reference image (image A) $referenceImg$ and the training image (image B) $trainImg$—and employs the next output parameter: the target image (image C) $targetImg$; $kpsReference$ and $kpsTrain$ represent the key points detected in image A and image B, respectively; $referenceDescriptors$ and $trainDescriptors$ represent the feature descriptors for key points detected in image A and image B, respectively; $matches$ represents the matching feature point pairs obtained using the KNN method; **H** represents the estimated perspective transformation matrix; h_q and w_q represent the height and width of image A, respectively; $origin$ represents the boundary corners of the original image; pts represents the coordinates of these corner points; dst represents the positions of these corner points on the target image after perspective transformation.

Algorithm 1: SIFT-based coarse matching stage.

Input: $referenceImg$, $trainImg$
Output: $targetImg$
$kpsReference$, $referenceDescriptors$
$kpsTrain$, $trainDescriptors$
$matches \leftarrow$ matchKeyPointsKNN
$\mathbf{H} \leftarrow$ getHomography
$h_q, w_q, _ \leftarrow$ shape ($referenceImg$)
$origin \leftarrow$ array($[[0,0], [0, h_q - 1], [w_q$-1, $h_q - 1], [w_q$-1, 0]])
$pts \leftarrow$ origin. reshape (-1,1,2)
$dst \leftarrow$ perspectiveTransform (pts, **H**). squeeze ()
if min (dst) < 0 **then**
  if min ($dst[:,0]$) < min ($dst[:,1]$) **then**
    $dst \leftarrow dst -$ min ($dst[:,0]$)
    $dst[:,1] \leftarrow dst[:,1] -$ min ($dst[:,1]$)
  else
    $dst \leftarrow dst -$ min ($dst[:,1]$)
    $dst[:,0] \leftarrow dst[:,0] -$ min ($dst[:,0]$)
  end if
else
  $dst[:,0] \leftarrow dst[:,0] -$ min ($dst[:,0]$)
  $dst[:,1] \leftarrow dst[:,1] -$ min ($dst[:,1]$)
end if
$targetImg \leftarrow$ warpPerspective

In the second phase, precise feature extraction is carried out using the SIFT algorithm on both the target image (image C) and the reference image (image A). The SIFT algorithm is independently applied to extract features within the overlapping regions of these two images. During feature extraction, a bidirectional k-nearest neighbors (KNN) matching algorithm is utilized for feature point matching. This approach is particularly effective in capturing more precise feature points, especially in flat areas of the image. To elucidate the

process, an initial key point is selected from image 1, the two closest feature points in image 2 to this key point in image 1 are identified, and their positions are recorded. If the distance ratio between the nearest feature point and the second-nearest feature point falls below the matching threshold of 0.75, the nearest feature point is retained. A similar procedure is then applied to determine the two closest feature points in image 1 for each feature point in image 2 among the nearest feature points in image 2.

To create a realistic large-scale underwater scene image, this paper introduces an image-pasting fusion stitching method. Firstly, the feature matching pairs are ranked in descending order according to their matching scores. Then, the logarithmic values of the matching feature points derived from the final filter are compared against a predefined threshold. If the values are below this threshold, the top 10% of the feature matching pairs with the highest scores are selected to determine the relative position of the two underwater images, denoted as x_{sub} and y_{sub}. The equations of the relative position correction symbols for the two images are as follows:

$$x_{sub} = \frac{\sum_{i=1}^{n} x_{B(i)} - x_{A(i)}}{n} \tag{13}$$

$$y_{sub} = \frac{\sum_{i=1}^{n} y_{B(i)} - y_{A(i)}}{n}, \tag{14}$$

where x_{sub} is the relative position correction symbol in the horizontal direction, y_{sub} is the relative position correction symbol in the vertical direction, (x_A, y_A) and (x_B, y_B) are the coordinates of matching feature points for the two images, respectively, and n represents the number of feature matching points between two images.

If the number of feature pairs resulting from the final matching process exceeds the predefined threshold, Formulas (13) and (14) are employed to determine the relative positions of the features in all matched pairs. The threshold value is chosen from the range of 50 to 100, depending on the image size. Through experiments conducted on the images presented in this article, it was observed that setting the threshold below 50 leads to a significant increase in matching errors. Conversely, if the threshold exceeds 100, it results in reduced computational efficiency. Therefore, in this study, the threshold was set to 60.

The determination of relative positions is based on the polarity of the relative position correction symbol. When x_{sub} is positive, it indicates that the image to be matched aligns to the left of the reference image, while the reference image is to the right. Conversely, when y_{sub} is positive, it indicates that the image to be matched aligns to the top, while the reference image is below. Once the positions of the two images are established, the relative positions can be utilized for image stitching.

To enhance the quality of the stitched images, we employed an improved fusion algorithm in this study. This approach involves calculating the weighted sum of pixel values within the overlapping regions of two images, taking into account the distance between these regions and the two images. These weights play a critical role in seamlessly blending the overlapping areas, thus reducing the ghosting effects and visible stitch marks in the final image. The calculation formula for the weighted sum is as follows:

$$f = \begin{cases} f_1, & (x,y) \in f_1 \\ c_1 f_1 + c_2 f_2, & (x,y) \in (f_1 \cap f_2) \\ f_2, & (x,y) \in f_2, \end{cases} \tag{15}$$

where f is the pixel value of the fused image, (x, y) is the coordinate of the concatenated image, f_1 and f_2 denote the pixel values of the two original images, respectively, c_1 and c_2 are weighting factors, and we have $c_1 + c_2 = 1, 0 < c_1 < 1, 0 < c_2 < 1$.

After completing the stitching process, the identification of the stitching direction involves the determination of the overlapping area between two images. The fusion algorithm is divided into horizontal and vertical fusion categories, depending on the observed horizontal and vertical stitching characteristics during the image stitching process. The dimensions of the overlapping area, specifically its width and height, are crucial in approximating the stitching direction of the image. When the width of the overlapping part exceeds the height, the two images are categorized as being horizontally stitched, and the horizontal fusion method is applied. Conversely, if the height of the overlapping part is greater than the width, the two images are considered to be vertically stitched, and the vertical fusion method is employed. This algorithm enhances the stitching accuracy by considering the direction of stitching between the images.

In the horizontal fusion method, $f_1(x, y)$ and $f_2(x, y)$ denote the pixel values of the original image on the left and right sides of the overlapping area, respectively, while c_1 and c_2 are calculated as shown below:

$$c_1 = \frac{x_r - x_i}{x_r - x_l} \tag{16}$$

$$c_2 = 1 - c_1 = \frac{x_i - x_l}{x_r - x_l}, \tag{17}$$

where x_i indicates the horizontal coordinate of the current pixel point, x_l indicates the left boundary of the overlapping area, x_r indicates the right boundary of the overlapping area, and $x_r \neq x_l$.

In the vertical fusion method, $f_1(x, y)$ and $f_2(x, y)$ denote the pixel values of the original image at the top and bottom of the overlapping area, respectively; c_1 and c_2 are calculated as shown below:

$$c_1 = \frac{y_d - y_i}{y_d - y_u} \tag{18}$$

$$c_2 = 1 - c_1 = \frac{y_i - y_u}{y_d - y_u}, \tag{19}$$

where y_i indicates the vertical coordinate of the current pixel point, y_u indicates the upper boundary of the overlapping area, y_d indicates the lower boundary of the overlapping area, and $y_d \neq y_u$.

4. Results

In this section, we present a comprehensive evaluation of our method by performing both qualitative and quantitative comparisons with other stitching algorithms. Our algorithm simulations were implemented using Python. We commenced with a qualitative assessment, comparing our improved SIFT-based underwater image stitching technique with several established stitching algorithms. Specifically, we focused on feature matching performance when contrasting our method with the traditional SIFT approach. Additionally, we employed the UIQM metric to assess the performance of our method and of other stitching algorithms in stitching underwater images. To further substantiate the stitching performance, we conducted experimental assessments involving multiple underwater images.

4.1. Qualitative Comparisons

In this paper, we conducted a comparative analysis by selecting several representative image stitching algorithms. Specifically, we categorized the images into two groups (underwater images and land images) based on their content. Subsequently, we compared our method with four other approaches: traditional SIFT, APAP, SPHP, and AANAP. To facilitate alignment comparison across these stitching methods, we employed a straightforward pixel-weighted average fusion technique. To maintain conciseness, we present the results for the Temple [11] and Rock datasets only. Figures 3 and 4 depict the stitching results for

the Rock and Temple images, respectively. The Rock dataset was obtained from the internet, while the Temple dataset was provided by the authors of [18]. As observed in the figures, the results of the other methods exhibited varying degrees of ghosting artifacts within the overlapping regions. In contrast, our method demonstrated superior performance in eliminating artifacts from these overlapping areas. If we focus our attention on the regions enclosed by the red ellipses in the figures, it becomes evident that our results offer a more natural appearance.

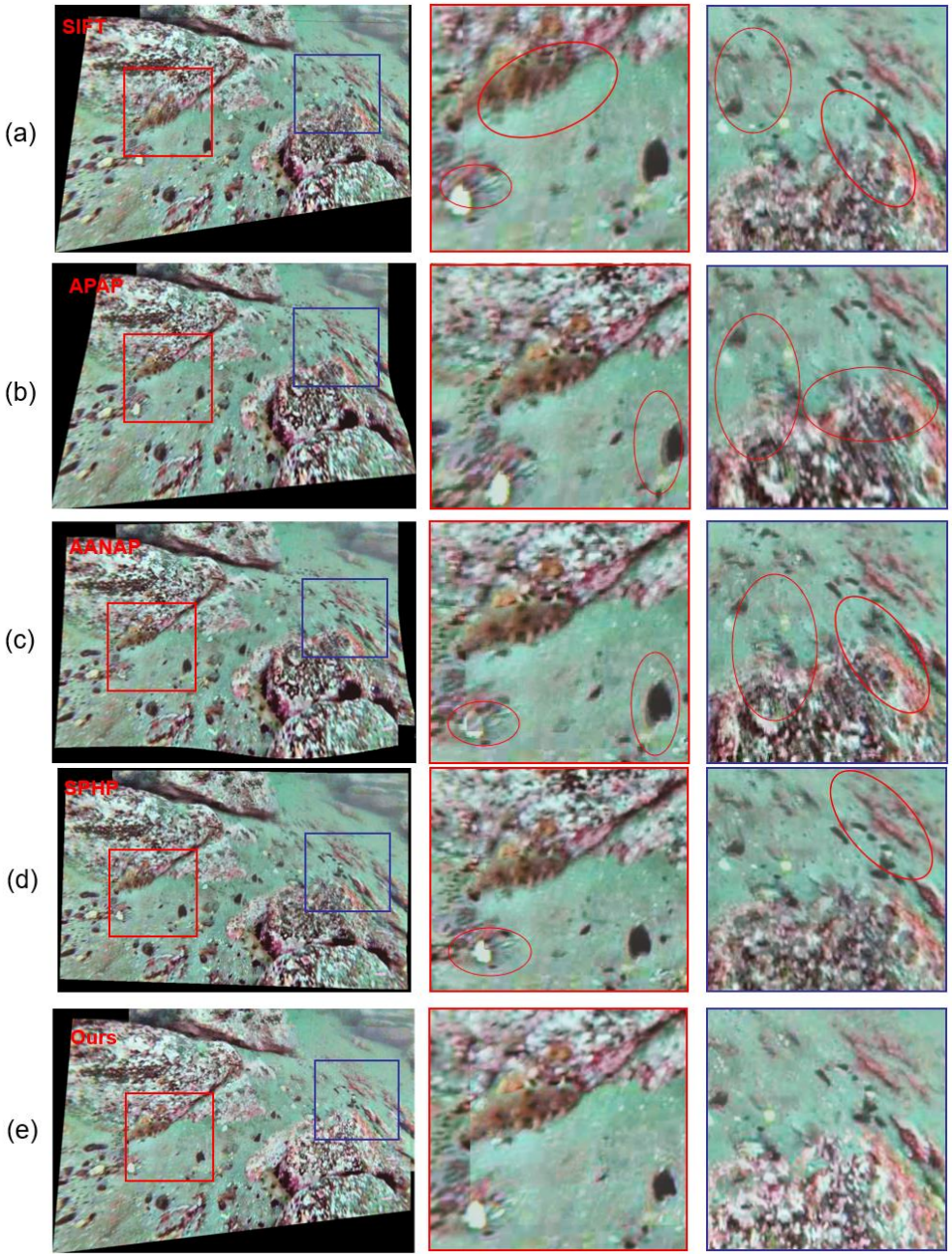


Figure 3. Qualitative comparisons of image stitching on the Rock image pair. Red circles highlight errors. List of acronyms and initialisms: (**a**) SIFT, scale-invariant feature transform; (**b**) APAP, As-Projective-As-Possible warps; (**c**) AANAP, adaptive As-Natural-As-Possible; (**d**) SPHP, shape-preserving half-projective warps; (**e**) our method.

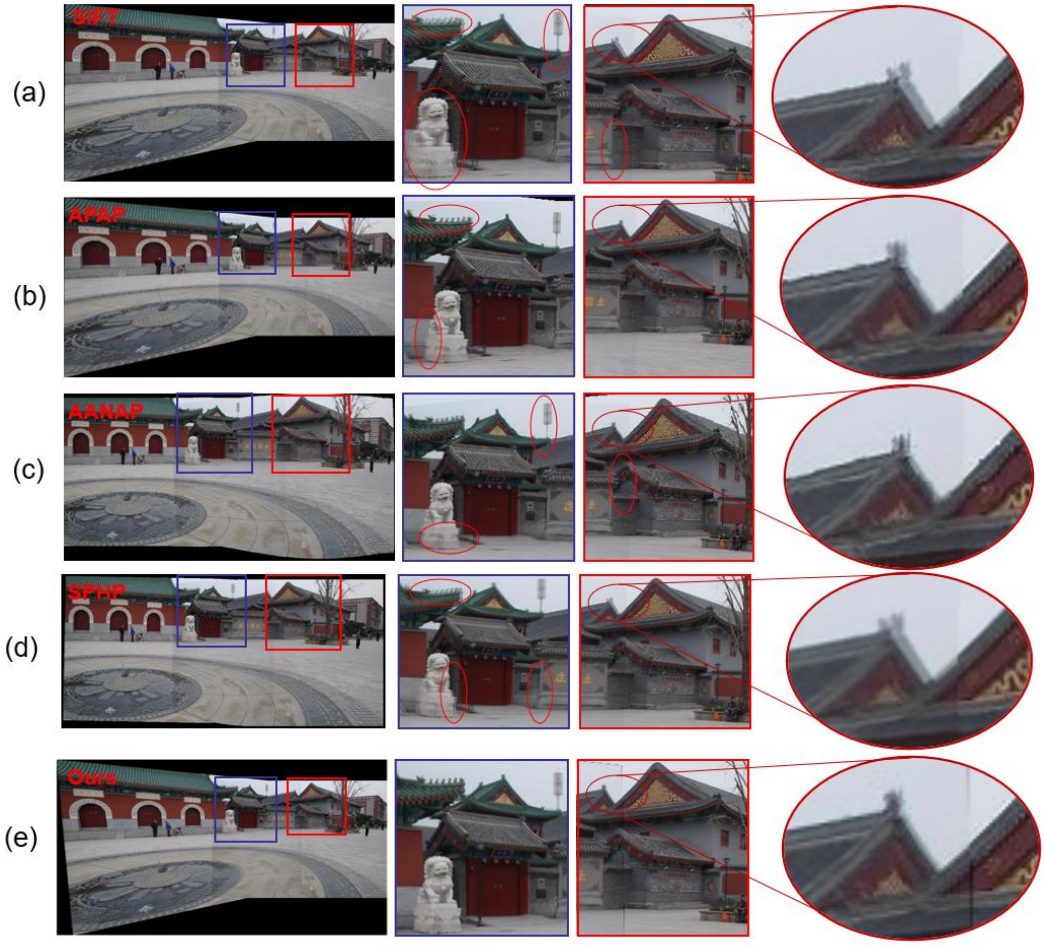


Figure 4. Qualitative comparisons of image stitching on the Temple image pair. Red circles highlight errors. List of acronyms and initialisms: (**a**) SIFT, scale-invariant feature transform; (**b**) APAP, As-Projective-As-Possible warps; (**c**) AANAP, adaptive As-Natural-As-Possible; (**d**) SPHP, shape-preserving half-projective warps; (**e**) our method.

Figure 3 illustrates the results for the Rock dataset, with each row representing the output of a different method. The sequence of results is as follows: traditional SIFT algorithm, APAP, AANAP, SPHP, and our method. The images' intricate textural features, particularly in underwater settings, present a substantial challenge for image stitching. We highlight two aspects for each result: red boxes indicate parallax errors in the overlapping regions where the feature textures are clear, and blue boxes indicate parallax errors in the overlapping regions where the feature textures are blurred. The outcomes of the traditional SIFT method exhibit low alignment accuracy, with ghosting effects appearing at the edges of the rocks, leading to a significant loss of detail information.

The results in the second row indicate that APAP performs well in regions with relatively rich texture features but faces challenges in aligning areas with blurred details. Moving to the third row, we can observe the outcomes of the AANAP method, which exhibit noticeable splicing artifacts and alignment errors at the junction of overlapping and non-overlapping regions. In the fourth row, we present the results of the SPHP method. SPHP demonstrates effective alignment in regions with abundant texture features but encounters difficulties in regions characterized by blurred details.

The results of the Temple dataset are presented in Figure 4, with each row corresponding to a different method. The order of the results is as follows: the traditional SIFT algorithm, APAP, AANAP, SPHP, and our method. Two aspects of each result have been high-

lighted, with blue boxes indicating parallax errors in the overlapping region and red boxes showing alignment errors at the junction of the overlapping and non-overlapping regions.

The traditional SIFT algorithm exhibits poor alignment in the overlapping areas, as evident in both the stone lion and roof areas, which suffer from ghosting issues. The second row displays the results of the APAP algorithm, which effectively aligns the overlapping areas but still exhibits some ghosting problems. In the third row, the results of AANAP reveal a parallax error in the overlapping area. In addition to ghosting, errors in alignment are noticeable, particularly in the houses at the overlapping edges of the two images. The fourth row illustrates the outcomes of the SPHP method. As described in the SPHP paper, this method prioritizes reducing perspective distortion rather than alignment accuracy. Parallax errors are evident in the overlapping area, particularly noticeable in the angles of the power poles, antennas, and houses.

4.2. Quantitative Comparisons

4.2.1. Comparison with the Traditional SIFT Algorithm

Our method demonstrates remarkable robustness when compared to the traditional SIFT algorithm, especially in scenarios where the textures surrounding the features exhibit high similarity. Figure 5 presents a visual comparison of the matching outcomes between our method and the conventional SIFT algorithm. The colored lines represent feature matching between two images, while the red boxes represent better stitching images obtained by our algorithm compared to traditional SIFT algorithms. Image A refers to the reference image and image B refers to the training image. As evident in the figure, the traditional SIFT algorithm struggled to obtain a satisfactory number of feature matches in the experimental image. Out of the 133 detected matching pairs, only 73 were accurate, resulting in an accuracy rate of 54.89%. This inaccuracy leads to improper alignment and significant ghosting artifacts in the traditional SIFT algorithm's results. Similar to the traditional SIFT algorithm, the outcomes of other stitching methods are also suboptimal. In contrast, our algorithm utilizes the projected transformed image for precise feature extraction and matching. This approach resulted in the identification of 85 correct matching pairs out of 139 detected, corresponding to an accuracy rate of 61.15%. Consequently, our algorithm excels in detecting a greater number of feature pairs and achieving higher matching accuracy, even in cases where feature detection is challenging in underwater images. The stitching results further underscore the superior alignment performance of our algorithm.

Figure 6 illustrates the comparative matching results between our method and the conventional SIFT algorithm using a different set of images. The colored lines represent feature matching between two images, while the red boxes represent better stitching images obtained by our algorithm compared to traditional SIFT algorithms. Image A refers to the reference image and image B refers to the training image. As depicted in Figure 6, the traditional SIFT algorithm successfully identified 277 correct matches out of 430 pairs, achieving an accuracy rate of 64.42%. In contrast, our algorithm identified 304 correct matches out of 434 pairs, yielding an accuracy rate of 70.05%. The results highlighted within the red box emphasize the limitations of the traditional SIFT algorithm, which excels in matching regions with distinct texture features but falters in regions with less prominent textures. Conversely, our algorithm exhibits the capability to accurately detect feature matches in regions with both prominent feature points and areas where the image contours are less distinct, leading to an enhanced quantity and accuracy of matches. This experiment underscores the heightened robustness and improved matching performance of our approach, stemming from our precise feature extraction and matching methodology.

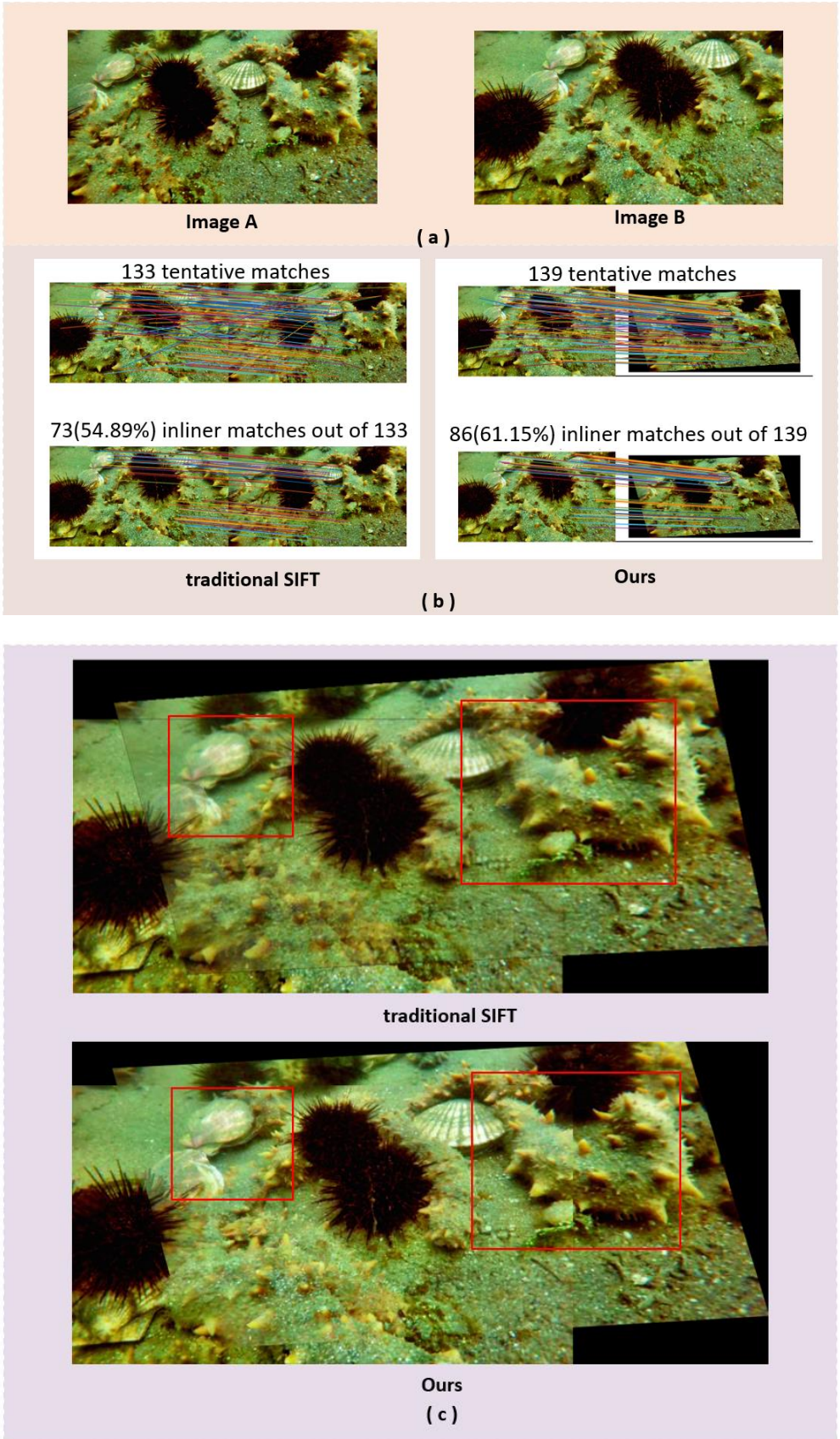


Figure 5. An image feature matching case: (**a**) Image inputs to be stitched. (**b**) Our feature matching and the traditional SIFT algorithm. (**c**) Stitching results of the traditional SIFT algorithm and ours.

4.2.2. Comparisons with Image Stitching Algorithms

UIQM is grounded in the framework proposed by Karen Panetta in 2015 [35], which takes into consideration three vital metrics for assessing underwater image quality: colorfulness measure (UICM), sharpness index (UISM), and contrast measure (UIConM). UIQM is computed as a linear combination of these three metrics. The higher the UIQM value, the more favorable the color balance, sharpness, and contrast of the image. UIQM serves as a comprehensive standard for evaluating the quality of underwater images, with a specific focus on their degradation characteristics. Widely recognized and authoritative in the realm of underwater image quality assessment, UIQM played a pivotal role in our study. Consequently, we conducted ten experiments on an image dataset and employed the UIQM metric to compare the performance of our proposed stitching method with that of several other stitching algorithms.

The formula for calculating the UIQM is shown below:

$$UIQM = c_1 \times UICM + c_2 \times UISM + c_3 \times UIConM. \quad (20)$$

Based on the results of [35], we took the scale factors as $c_1 = 0.0282$, $c_2 = 0.2953$, and $c_3 = 3.5753$. The results are shown in Table 2. Generally speaking, the value range of UIQM was between 0 and 3. For images with good imaging quality and clarity, the value of UIQM was generally above 1.5. As shown in Table 2, except for image 2, the UIQM value was relatively low due to the blurriness and poor quality of the image itself. Our method showed significant improvements in UIQM metrics compared to the traditional SIFT algorithm, with all datasets having higher scores and an average improvement of 6.04%. This demonstrates that our algorithm enhances the quality and sharpness of underwater panoramic images. Additionally, our method outperformed other algorithms such as APAP, AANAP, and SPHP, with improvements in UIQM of 4.73%, 9.01%, and 11.06%, respectively.

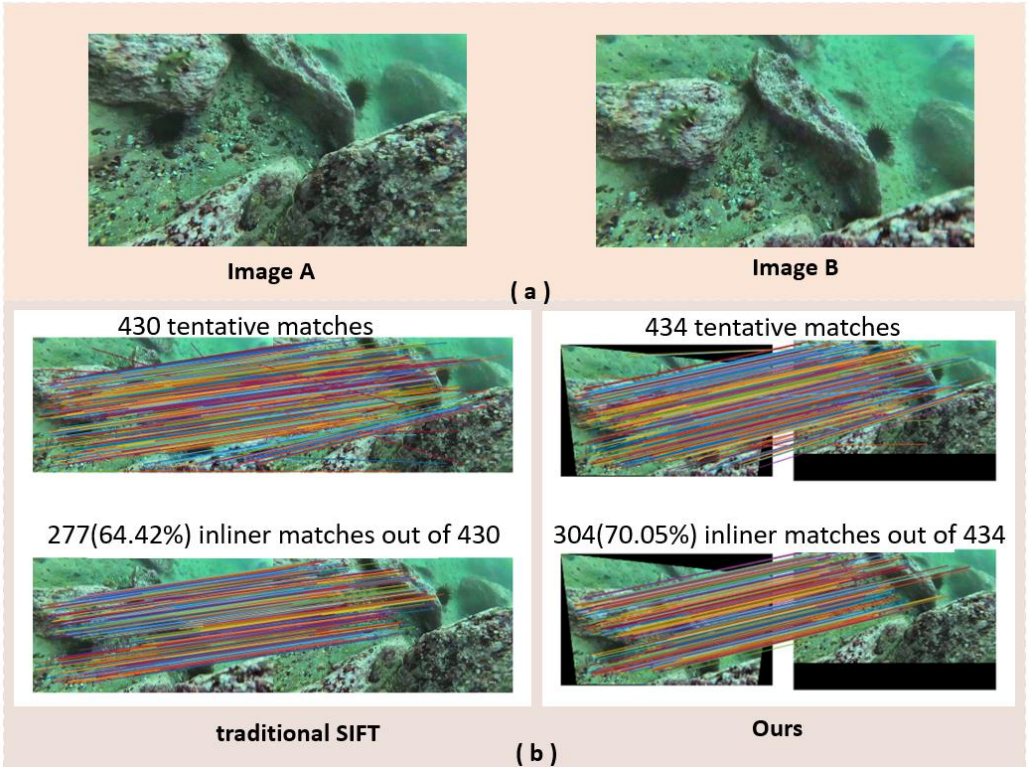


Figure 6. *Cont.*

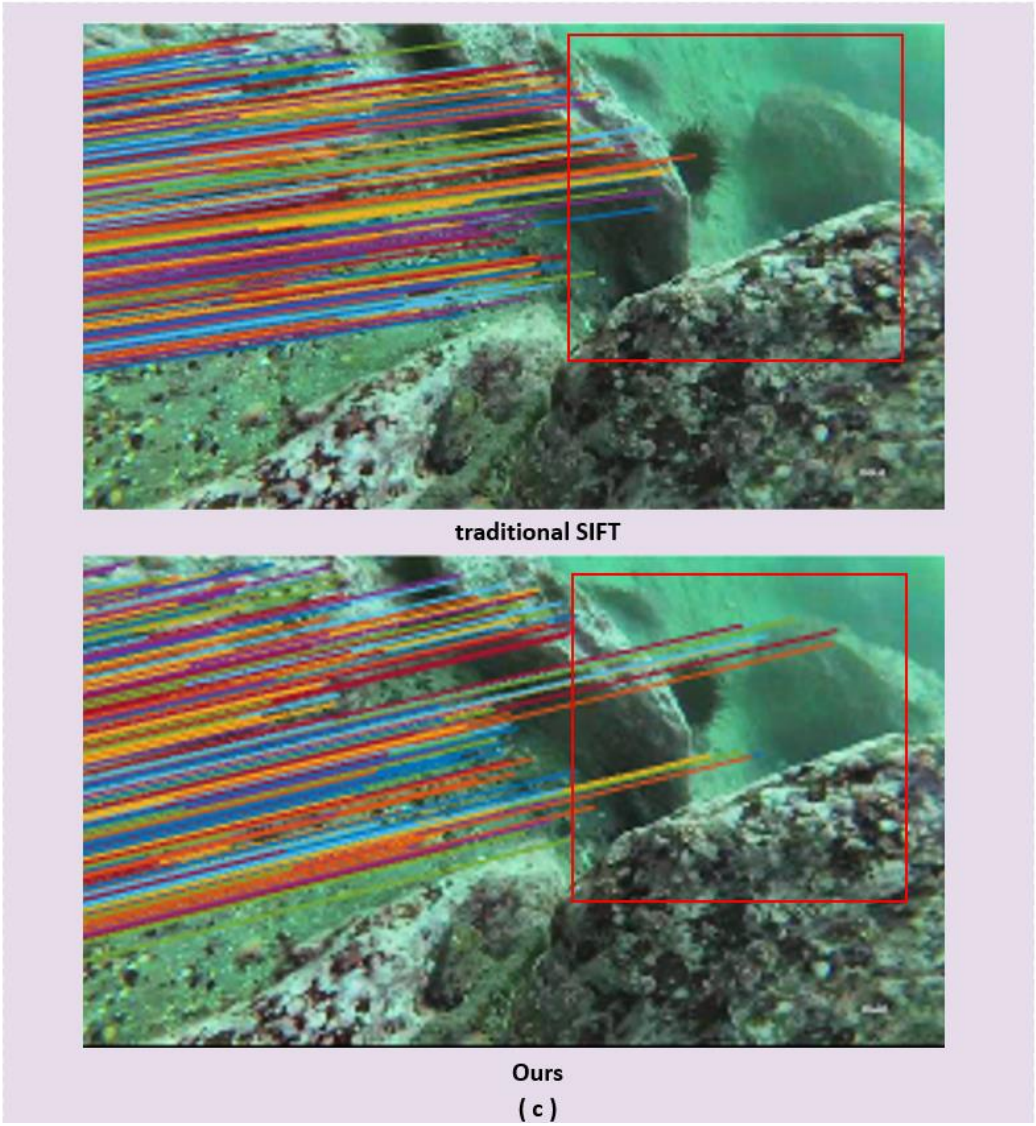


Figure 6. An image feature matching case: (**a**) Image inputs to be stitched. (**b**) The feature matching of image B. (**c**) Our feature matching and the traditional SIFT algorithm.

Table 2. UIQM index of different stitching methods.

Image	SIFT	APAP	SPHP	AANAP	Ours
1	1.2994	1.3745	1.2997	1.3953	1.4524
2	0.5841	0.6366	0.6616	0.6695	0.6342
3	1.8733	1.9963	1.9578	1.7377	1.9047
4	1.6443	1.7072	1.6050	1.7192	1.7440
5	1.4720	1.5522	1.4220	1.5256	1.6588
6	1.7069	1.7134	1.6298	1.6833	1.7364
7	1.6783	1.6375	1.5809	1.6719	1.6966
8	2.0174	2.0437	1.8865	1.9139	2.0564
9	1.7277	1.7150	1.7910	1.5551	1.7855
10	1.5662	1.5659	1.4754	1.6446	1.7473

4.3. Stitching of Multiple Images

To demonstrate the effectiveness of our method in stitching multiple images, we present a panorama example consisting of multiple input images. In order to achieve better visual effects, our proposed fade-in and fade-out fusion algorithm was employed

to handle the stitching gaps during the process. Figure 7 showcases a large underwater shipwreck scene.

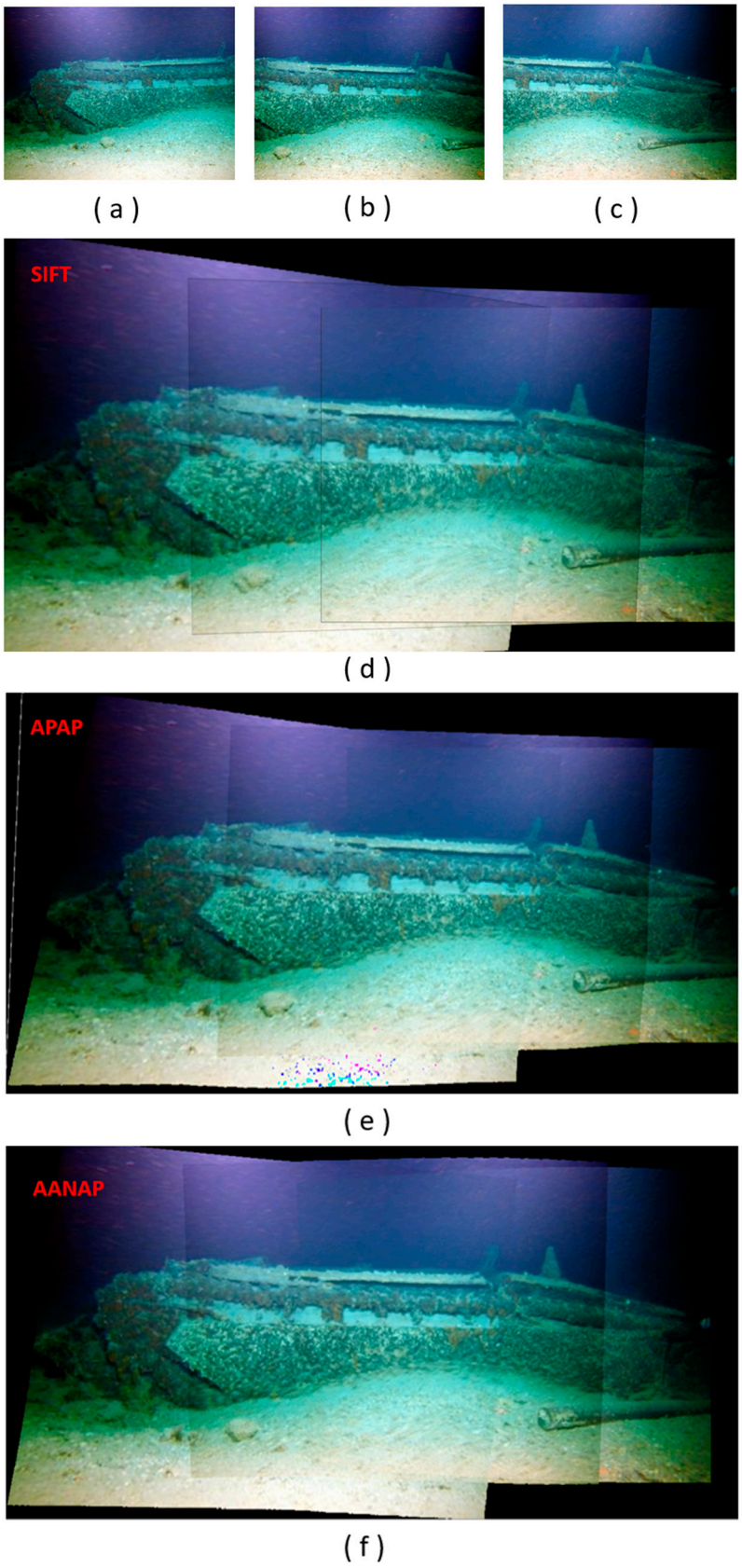


Figure 7. *Cont.*

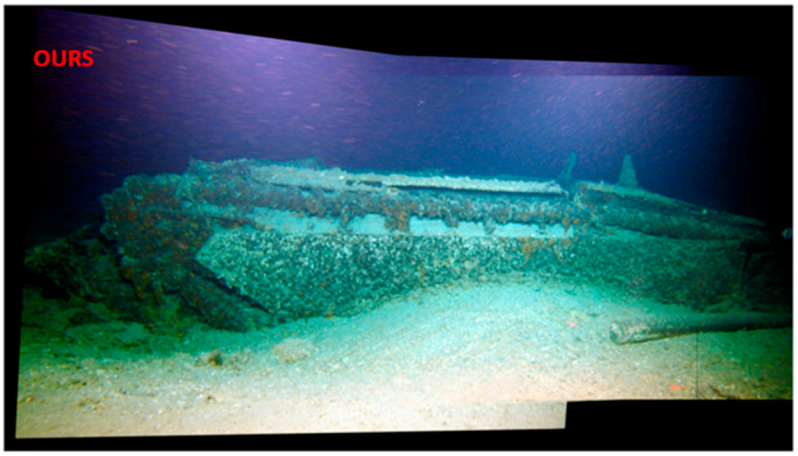


Figure 7. Stitching results of multiple images in perspective projection on the Ship image pair: (**a**–**c**) Image inputs to be stitched. (**d**) SIFT. (**e**) APAP, As-Projective-As-Possible warps. (**f**) AANAP, adaptive As-Natural-As-Possible;.(**g**) Our method.

As shown in Figure 7, when compared with SIFT, APAP, and AANAP, our approach accomplishes the seamless stitching of the underwater shipwreck images while preserving the integrity of the image content and presenting a natural panorama. Parallax errors and perspective distortions are imperceptible in the resulting image. Even in the presence of a weak feature texture and blurred surface contours of the shipwreck caused by numerous attachments, our method effectively tackles the demanding challenge of feature extraction and matching. This showcases the resilience and comprehensiveness of our approach in managing the fusion of multiple images.

5. Discussion

The effectiveness of image stitching heavily depends on the image quality and the number of matched features. The proposed method maintains alignment accuracy and shows robustness in all of the abovementioned challenging examples. It performs better in aligning relatively texture-rich regions and can achieve effective image stitching in regions with blurred details.

The traditional SIFT algorithm requires that the targets to be stitched are in or approximately in the same plane. When there are occlusion and angle differences, ghosting of the stitched image occurs. APAP uses a local homography matrix for the calculation, which improves the performance of the stitching compared with the traditional SIFT algorithm approach. However, there are still some alignment errors. Moreover, the local model of APAP is vulnerable to the number and distribution of matching features. AANAP reduces the perspective error based on APAP, but it brings certain alignment error problems. AANAP linearizes the homology in the non-overlapping regions and uses global similarity transformation with automatic estimation in the overlapping regions. Smooth interpolation is used for extrapolation between the overlapping and non-overlapping regions. In this process, it is able to achieve a slow change between the overlapping and non-overlapping regions. However, the compensation effect of smooth interpolation is limited when there is already a large alignment error between the overlapping and non-overlapping regions. The current parameter selection of the SPHP algorithm does not take into account the image content (e.g., line features), which can lead to the algorithm failing to reduce the distortion of the line.

6. Conclusions

In this paper, we introduce an enhanced SIFT method designed for the stitching of underwater images characterized by smeared feature contours. Our method stands out

because it does not rely on detector information for feature extraction. Even in regions with blurred texture features, where traditional feature extraction methods falter, our approach significantly augments the number of accurately matched pairs, thereby elevating the precision of feature matching. Our method incorporates two critical phases for the stitching process, each rooted in the SIFT algorithm: coarse feature extraction and precise feature extraction. To enhance the image contrast and quality, we integrate a dynamic threshold white balance algorithm and CLAHE. Additionally, we have refined the fade-in and fade-out technique for generating panoramic images. A series of experiments were conducted to validate the effectiveness of our algorithm. Our results underscore the superiority of our method over the conventional SIFT algorithm. It effectively mitigated ghosting issues in stitched images and enhanced the UIQM measure by an average of 6.04% across ten underwater image stitching experiments. When compared to alternative algorithms such as APAP, SPHP, and AANAP, our method outperformed them, boosting the UIQM index by 4.73%, 9.01%, and 11.06%, respectively. Furthermore, we demonstrated the efficacy of our method in seamlessly stitching multiple images, preserving the integrity of the image content in expansive scenes without significant ghosting artifacts. Our future research endeavors will concentrate on compensating for parallax in cases of substantial movements and addressing the influence of moving objects within the scene on stitching quality.

Author Contributions: Conceptualization, J.M. and W.Z.; methodology, J.M. and W.Z.; software (Pycharm 2020.1.2), R.Z.; validation, R.Z., W.Z. and H.Z.; formal analysis, H.Z.; investigation, H.Z.; resources, H.Z.; data curation, H.Z.; writing—original draft preparation, R.Z.; writing—review and editing, J.S.; visualization, R.Z.; supervision, J.M.; project administration, J.M.; funding acquisition, J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (42227901 and 52371358), the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (311020011), the Key-Area Research and Development Program of Guang-dong Province (2020B1111010004), and the Special Project for Marine Economy Development of Guangdong Province (GDNRC [2022] 31).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jatmiko, D.A.; Prini, S.U. Study and Performance Evaluation Binary Robust Invariant Scalable Keypoints (BRISK) for Underwater Image Stitching. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *879*, 012111. [CrossRef]
2. Faugeras, O.; Luong, Q. *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications*; MIT Press: Cambridge, MA, USA, 2001.
3. Kim, D.; Song, J. Underwater Image Preprocessing and Compression for Efficient Underwater Searches and Ultrasonic Communications. *Int. J. Precis. Eng. Manuf.* **2007**, *8*, 38–45.
4. Pandey, A.; Pati, U.C. Image mosaicing: A deeper insight. *Image Vis. Comput.* **2019**, *89*, 236–257. [CrossRef]
5. Szeliski, R. Image mosaicing for tele-reality applications. In Proceedings of the 1994 IEEE Workshop on Applications of Computer Vision, Sarasota, FL, USA, 5–7 December 1994; IEEE: New York, NY, USA, 1994.
6. Negahdaripour, S.; Lanjing, J. Direct recovery of motion and range from images of scenes with time-varying illumination. In Proceedings of the International Symposium on Computer Vision-ISCV, Coral Gables, FL, USA, 21–23 November 1995; IEEE: New York, NY, USA, 1995.
7. Marks, R.L.; Rock, S.M.; Lee, M.J. Real-time video mosaicking of the ocean floor. *IEEE J. Ocean. Eng.* **1995**, *20*, 229–241. [CrossRef]
8. Raut, S.; Pati, U.C. Underwater image registrtion with improved SIFT algorithm. In Proceedings of the 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 19–20 May 2017; IEEE: New York, NY, USA, 2017.
9. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
10. Bay, H.; Ess, A.; Tuytelaas, T.; Van Goal, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [CrossRef]

11. Zaragoza, J.; Cesar, H. As-projective-as-possible image stitching with moving DLT. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013.
12. Chang, C.; Sato, Y.; Chuang, Y. Shape-preserving half-projective warps for image stitching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.
13. Lin, C.C.; Pankanti, S.U.; Natesan Ramamurthy, K.; Aravkin, A.Y. Adaptive as-natural-as-possible image stitching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
14. Gao, J.; Li, Y.; Chin, T.J.; Brown, M.S. Seam-driven image stitching. In *Eurographics (Short Papers)*; The Eurographics Association: Eindhoven, The Netherlands, 2013.
15. Lin, K.; Jiang, N.; Cheong, L.F.; Do, M.; Lu, J. Seagull: Seam-guided local alignment for parallax-tolerant image stitching. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016.
16. Brown, M.; Lowe, D.G. Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vis.* **2007**, *74*, 59–73. [CrossRef]
17. Lin, W.; Liu, S.; Matsushita, Y.; Ng, T.T.; Cheong, L.F. Smoothly varying affine stitching. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; IEEE: New York, NY, USA, 2011.
18. Gao, J.; Kim, S.J.; Brown, M.S. Constructing image panoramas using dual-homography warping. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; IEEE: New York, NY, USA, 2011.
19. Zhang, F.; Liu, F. Parallax-tolerant image stitching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.
20. Li, J.; Wang, Z.; Lai, S.; Zhai, Y.; Zhang, M. Parallax-tolerant image stitching based on robust elastic warping. *IEEE Trans. Multimed.* **2017**, *20*, 1672–1687. [CrossRef]
21. Leone, A.; Distante, C.; Mastrolia, A.; Indiveri, G. A fully automated approach for underwater mosaicking. In Proceedings of the OCEANS 2006, Boston, MA, USA, 18–21 September 2006; IEEE: New York, NY, USA, 2006.
22. Haywood, R. Acquisition of a micro scale photographic survey using an autonomous submersible. In *Oceans*; IEEE: New York, NY, USA, 1986; pp. 1423–1426.
23. Xu, X.; Negahdaripour, S. Vision-based motion sensing for underwater navigation and mosaicing of ocean floor images. In Proceedings of the Oceans'97. MTS/IEEE Conference Proceedings, Halifax, NS, Canada, 6–9 October 1997; IEEE: New York, NY, USA, 1997.
24. Fleischer, S.D.; Wang, H.H.; Rock, S.M.; Lee, M.J. Video mosaicking along arbitrary vehicle paths. In Proceedings of the Symposium on Autonomous Underwater Vehicle Technology, Monterey, CA, USA, 2–6 June 1996; IEEE: New York, NY, USA, 1996.
25. Fleischer, S.D.; Marks, R.L.; Rock, S.M.; Lee, M.J. Improved real-time videomosaicking of the ocean floor. In Proceedings of the 'Challenges of Our Changing Global Environment'. Conference Proceedings. OCEANS'95 MTS/IEEE, San Diego, CA, USA, 16 June 1995; IEEE: New York, NY, USA, 1995.
26. Gracias, N.; Santos-Victor, J. Underwater mosaicing and trajectory reconstruction using global alignment. In *MTS/IEEE Oceans 2001. An Ocean Odyssey. Conference Proceedings (IEEE Cat. No. 01CH37295), Honolulu, HI, USA 5–8 November 2001*; IEEE: New York, NY, USA, 2001.
27. Pizarro, O.; Singh, H. Toward large-area mosaicing for underwater scientific applications. *IEEE J. Ocean. Eng.* **2003**, *28*, 651–672. [CrossRef]
28. Zuiderveld, K. Contrast limited adaptive histogram equalization. *Graph. Gems* **1994**, 474–485. [CrossRef]
29. Stark, J.A. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Trans. Image Process.* **2000**, *9*, 889–896. [CrossRef] [PubMed]
30. Weng, C.; Chen, H.; Fuh, C. A novel automatic white balance method for digital still cameras. In Proceedings of the 2005 IEEE International Symposium on Circuits and Systems (ISCAS), Kobe, Japan, 23–26 May 2005; IEEE: New York, NY, USA, 2005.
31. Kashif, I.; Salam, R.A.; Osman, A.; Talib, A.Z. Underwater Image Enhancement Using an Integrated Colour Model. *IAENG Int. J. Comput. Sci.* **2007**, *34*, 239–244.
32. Song, W.; Wang, Y.; Huang, D.; Tjondronegoro, D. A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration. In Proceedings of the Pacific Rim Conference on Multimedia, Hefei, China, 21–22 September 2018; Springer: Berlin/Heidelberg, Germany, 2018.
33. Paulo Drews, J.R.; Nascimento, E.; Moraes, F.; Botelho, S.; Campos, M. Transmission Estimation in Underwater Single Images. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Sydney, Australia, 2–8 December 2013.
34. Peng, Y.T.; Cosman, P.C. Underwater Image Restoration Based on Image Blurriness and Light Absorption. *IEEE Trans. Image Process.* **2017**, *26*, 1579–1594. [CrossRef] [PubMed]
35. Panetta, K.; Gao, C.; Agaian, S. Human-visual-system-inspired underwater image quality measures. *IEEE J. Ocean. Eng.* **2015**, *41*, 541–551. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.