

## Research Article

# A SIFT-Like Feature Detector and Descriptor for Multibeam Sonar Imaging

Wanyuan Zhang,<sup>1,2,3</sup> Tian Zhou,<sup>1,2,3</sup> Chao Xu ,<sup>1,2,3</sup> and Meiqin Liu<sup>4</sup>

<sup>1</sup>Acoustic Science and Technology Laboratory, Harbin Engineering University, Harbin 150001, China

<sup>2</sup>Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), Ministry of Industry and Information Technology, Harbin 150001, China

<sup>3</sup>College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China

<sup>4</sup>Zhejiang University, Hangzhou 310058, China

Correspondence should be addressed to Chao Xu; [xuchao18@hrbeu.edu.cn](mailto:xuchao18@hrbeu.edu.cn)

Received 14 August 2020; Revised 26 April 2021; Accepted 22 June 2021; Published 16 July 2021

Academic Editor: Luca Vollerio

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

Multibeam imaging sonar has become an increasingly important tool in the field of underwater object detection and description. In recent years, the scale-invariant feature transform (SIFT) algorithm has been widely adopted to obtain stable features of objects in sonar images but does not perform well on multibeam sonar images due to its sensitivity to speckle noise. In this paper, we introduce MBS-SIFT, a SIFT-like feature detector and descriptor for multibeam sonar images. This algorithm contains a feature detector followed by a local feature descriptor. A new gradient definition robust to speckle noise is presented to detect extrema in scale space, and then, interest points are filtered and located. It is also used to assign orientation and generate descriptors of interest points. Simulations and experiments demonstrate that the proposed method can capture features of underwater objects more accurately than existing approaches.

## 1. Introduction

In recent years, with the constant exploration of oceans, multi-beam sonar (MBS) is being applied in many scenarios, such as underwater environment mapping [1], underwater target detection [2], and underwater terrain-aided navigation [3]. However, noise interference is strong in water, which makes it more difficult to extract useful information from the images. In order to extract and track features better, it is necessary to match the target features within the sonar image sequences. In the sonar image processing domain, high-resolution sonar imagery is mainly affected by a granular multiplicative noise known as speckle noise, whose presence in MBS images is particularly strong [4–6]. In the feature extraction process, speckle noise pixels and the corner points of targets of interest have similar gradients. This greatly increases the probability of a false match. So, the focus of this work is to develop a new feature detector and descriptor that overcomes the effects of strong speckle noise in MBS imagery.

Recently, Thomas and Zakharov realized object detection by calculating the pixel displacement between continuous sonar images [7], but their method required images of very high quality. For object feature extraction in complex environments such as sonar images, feature-based approaches are more suitable than pixel-based ones for tasks like object detection, change detection, and object classification [8]. Nowadays, a large number of feature-based algorithms have also been adopted in sonar image processing, such as the scale-invariant feature transform (SIFT), speeded-up robust features (SURF), histogram of oriented gradient (HOG), local binary pattern (LBP), and their variants [9–11]. During the collection of sonar data, underwater targets (fish, phytoplankton, and artificial targets) are not stable [12]. Therefore, the requirement of scale and rotation invariance for the application discussed in this study is high. However, the HOG algorithm is quite sensitive to noise and the LBP algorithm is easily influenced by direction information, while SIFT shows better scale and rotation invariance

than SURF. For the purposes of this application, SIFT and its variants are more suitable for feature matching in sonar images.

The SIFT algorithm, presented by Lowe [13], is an interesting option to detect, describe, and match features. Feature detectors and descriptors provide good performance due to their invariance to various transformations. Due to its efficiency in detection, description, and image matching, the SIFT algorithm and its variants are widely used in the field of computer vision and remote optical sensing. Kishu and Rana [14] applied principal component analysis (PCA) to reduce the SIFT descriptor's dimensionality and achieve better results in the field of facial recognition. For image scene classification, Ju et al. [8] worked directly with the two-dimensional matrix and applied generalized PCA (GPCA) to obtain a lower-dimensional matrix. The SIFT algorithm has also been applied in the field of object detection. Tao et al. [15] exploited the clustering information from matched SIFT keypoints and the region information extracted through image segmentation. Urban areas and buildings were detected in satellite images by Sirmacek and Unsalan [16] using SIFT and graph theory tools. Huang et al. [17] presented a novel interest point detection algorithm using a Laplacian-of-Bilateral (LoB) filter to detect extrema in scale space. In order to improve interest point detection, Li et al. [18] proposed a new learning-to-rank framework to eliminate unstable extrema derived from the keypoint detection stage of SIFT.

While the SIFT algorithm has proven its efficiency for various applications in remote optical sensing, the situation is different for sonar images. Sonar image sequences from the same scene appear different because several varying factors affect the imaging technique. These factors include electrical noise, speckle noise, decorrelation due to medium instabilities, and target motion. In particular, the strong speckle noise interference makes image processing very difficult because the gradient by difference operation in the keypoint detection step of the SIFT algorithm is susceptible to this type of noise. Researchers have proposed various methods to improve SIFT's performance in such cases. Depending on the application, some researchers removed or modified some steps of the algorithm to improve its performance [19, 20]. Others suggested denoising or filtering the images during preprocessing to reduce the influence of speckle noise [21, 22]. To suppress false matches, spatial relationships between keypoints have been considered [23]. However, the performance of these newly developed algorithms is still relatively limited, as the number of keypoints detected is not stable and there are insufficient correct matches. In the stage of feature extraction, most of these algorithms are based on Hessian matrices, which rely on the second derivatives. Meanwhile, most of these algorithms do not consider the statistical specificities of speckle noise. These reasons make them easy to adapt to speckle noise. In order to obtain consistent and well distinguishable feature points, it is of primary importance to adapt SIFT-like algorithms to this context.

Considering the gradient definition of existing algorithms [9–11, 19–22], in this study, a new gradient definition

adapted to MBS images is presented. This new gradient definition makes magnitude and orientation robust to speckle noise. It is then used in several steps in the new algorithm for the extraction of local descriptors. An application of this novel algorithm is then presented for target feature matching (including artificial and nonartificial targets) in MBS images.

In this paper, we propose a new algorithm for the extraction of local descriptors, adapted to the statistical specificities of MBS images with noise following a Weibull distribution. This algorithm is largely inspired by SIFT and will be referred to as MBS-SIFT. In Section 2, after analyzing the speckle noise and the background noise distribution of sonar images, a new gradient definition is presented, which yields a magnitude and an orientation robust to speckle noise. We introduce the MBS-SFIT algorithm, including both the detection of keypoints and the computation of local descriptors, in Section 3. Section 4 shows the experimental validation and performance possibilities offered by this new algorithm in MBS images. Finally, we conclude the paper in Section 5.

## 2. A New Gradient Definition for Multibeam Sonar Images

Actually, mostly due to speckle noise, traditional approaches in feature-based detection do not perform well on MBS images. In this section, the speckle noise of MBS images and the statistical distribution of the background noise are first described in detail. We then introduce a new gradient definition to replace the classical gradient by difference for the detection of keypoints and the computation of local descriptors.

*2.1. Background Noise Distribution Estimation of Multibeam Sonar Images.* Speckle noise is one of the primary sources of noise in MBS images. It is mainly caused by interference affecting the returning acoustic wave inside the transducer due to the roughness of the material surface on the wavelength scale. The scattered signal adds coherently, producing patterns of constructive and destructive interference that appear as brighter or darker dots in sonar images. Speckle noise affects the image quality and obscures important information, such as edge and shape and intensity value. Furthermore, background noise distribution estimation is very important for target detection. Compared with the lognormal distribution and  $K$ -distribution, the Weibull distribution provides a trade-off between the accuracy of modeling and computational cost [2]. For these reasons, the Weibull distribution is more suitable to describe the statistical properties of this type of background noise.

The Weibull distribution function is defined as

$$F(x) = 1 - \exp\left(-\frac{x^k}{\lambda}\right), \lambda, k > 0, \quad (1)$$

where  $\lambda$  is the scale parameter and  $k$  is the shape parameter.

The background noise can be represented as a set of samples  $X = (x_1, x_2, \dots, x_i, \dots, x_n)$  with a likelihood function defined as

$$L(x, \lambda, k) = \prod_{i=1}^n \left(\frac{k}{\lambda}\right) \left(\frac{x_i}{\lambda}\right)^{k-1} \exp\left(-\frac{x_i^k}{\lambda}\right). \quad (2)$$

After derivation, the maximum-likelihood estimators of  $k$  and  $\lambda$  are

$$\frac{1}{\bar{k}} = \frac{\sum_{i=1}^n x_i^k \ln x_i}{\sum_{i=1}^n x_i^k} - \frac{1}{n} \sum_{i=1}^n \ln x_i, \quad (3a)$$

$$\lambda \bar{k}^k = \frac{1}{n} \sum_{i=1}^n x_i^k. \quad (3b)$$

Figure 1 presents the estimated probability density function of the background noise. Figure 1(a) introduces the empirical histogram of the background noise pixels of the MBS imagery shown in Figure 1(b). Figure 1(b) was obtained using Blueprint's Oculus M750d MBS, whose parameters are described in more detail in Section 4.1. The parameters of the Weibull distribution were estimated from the image in Figure 1(b), whose size is  $1350 \times 2620$  and contains only clutter. As we can see, the data in Figure 1(b) fit the Weibull distribution model well.

**2.2. Gradient Computation for Multibeam Sonar Images.** A large number of edge detectors specifically developed for sonar images are based on the difference of average pixel values or a ratio of average (ROA). However, in speckled images, the ratio edge detector has been found to be much more effective than the classical gradient by difference [24]. Different from the arithmetic mean of the ROA operator, the ratio of exponentially weighted averages (ROEWA) operator is more accurate and stable in edge detection [25]. It computes exponentially weighted local means of the image, as shown in Figure 2.

For example, given a point  $(x_0, y_0)$ , in the two-dimensional ROEWA operator, the means for the vertical gradient direction ( $\theta = \pi/2$ ) are defined as

$$\begin{cases} M_{1,\alpha}\left(\theta = \frac{\pi}{2}\right) = \int_{x=R^-} \int_{y=R^+} I(x_0 + x, y_0 + y) \times f(x, y), \\ M_{2,\alpha}\left(\theta = \frac{\pi}{2}\right) = \int_{x=R^-} \int_{y=R^+} I(x_0 + x, y_0 + y) \times f(x, y), \end{cases} \quad (4)$$

where  $f(x, y) = \exp(-(|x| + |y|)/\alpha)$  is the exponential weight function and  $\alpha$  is the exponential weight parameter.

The ROEWA and its normalization for a direction  $\theta$  are defined as

$$R_{\theta,\alpha} = \frac{M_{1,\alpha}(\theta)}{M_{2,\alpha}(\theta)}, \quad (5a)$$

$$T_{\theta,\alpha} = \max\left(R_{\theta,\alpha}, \frac{1}{R_{\theta,\alpha}}\right). \quad (5b)$$

Therefore, the ratio for the horizontal direction  $T_{0,\alpha}$  and the vertical direction  $T_{\pi/2,\alpha}$  can be calculated using Equation (5b). Moreover, the gradient magnitude for each pixel is defined as

$$\text{Mag}_{\theta,\alpha} = \sqrt{(T_{0,\alpha})^2 + (T_{\pi/2,\alpha})^2}. \quad (6)$$

Compared with the ROA operator, ROEWA is more robust to noise and more accurate in a multiscale edge context. ROEWA provides a good estimate of the gradient magnitude, but it is not a precise method to estimate the gradient orientation. This can be solved by increasing the number of directions, which adds to the computational cost. Inspired by the gradient-based edge detector for optical images [26], the gradient magnitude and orientation are estimated as

$$\begin{cases} \text{Mag}_{\theta,\alpha} = \sqrt{(T_{0,\alpha})^2 + (T_{\pi/2,\alpha})^2}, \\ \text{Ori}_{\theta,\alpha} = \arctan\left(\frac{T_{\pi/2,\alpha}}{T_{0,\alpha}}\right). \end{cases} \quad (7)$$

However, the estimation of orientation shown in (7) is ambiguous.  $T_{0,\alpha}$  and  $T_{\pi/2,\alpha}$  are always positive, and thus,  $\text{Ori}_{\theta,\alpha}$  can only take values between 0 and  $\pi/2$ . This is contradictory to the expectation that  $\text{Ori}_{\theta,\alpha}$  should take values between 0 and  $2\pi$ . Moreover, the gradient computation on a vertical edge with reflectivities  $m_a$  and  $m_b$  ( $m_a < m_b$ ) yields

$$\begin{aligned} T_{0,\alpha} &= \frac{m_b}{m_a}, \\ T_{\pi/2,\alpha} &= 1, \\ \text{Ori}_{\theta,\alpha} &= \arctan\left(\frac{m_a}{m_b}\right). \end{aligned} \quad (8)$$

Therefore, the gradient orientation takes arbitrary values depending on the reflectivities of the areas, when in fact it is equal to zero. In this paper, we use the gradient by ratio (GR), presented by Dellinger et al. [27], to avoid the above shortcomings. We define the horizontal and vertical gradients as

$$\begin{cases} G_{x,\alpha} = \log(R_{0,\alpha}), \\ G_{y,\alpha} = \log(R_{\pi/2,\alpha}). \end{cases} \quad (9)$$

Based on this definition, the gradient magnitude and orientation are as

$$\begin{cases} \text{Mag}_{n,\alpha} = \sqrt{(G_{x,\alpha})^2 + (G_{y,\alpha})^2}, \\ \text{Ori}_{t,\alpha} = \arctan\left(\frac{G_{y,\alpha}}{G_{x,\alpha}}\right), \end{cases} \quad (10)$$

where  $\alpha$  is the exponential weight parameter. The scale space is established by using different values of  $\alpha$  during image filtering.

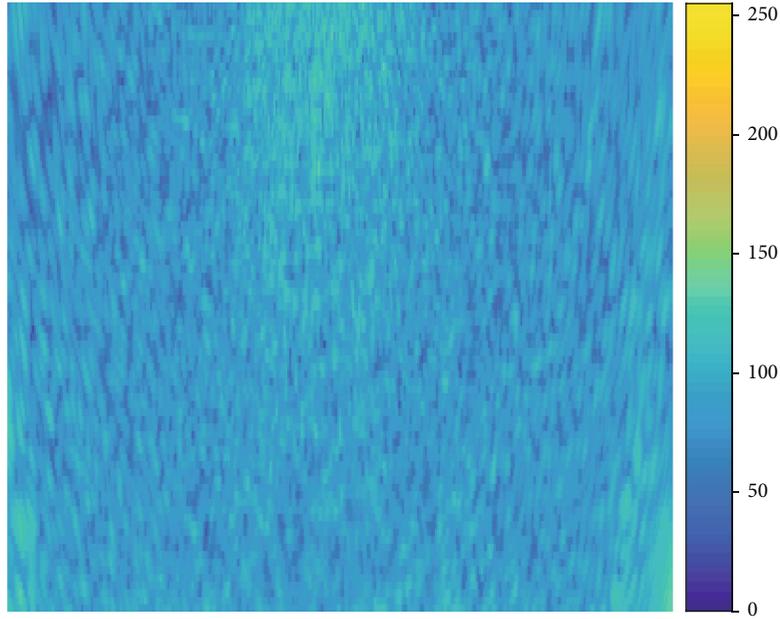
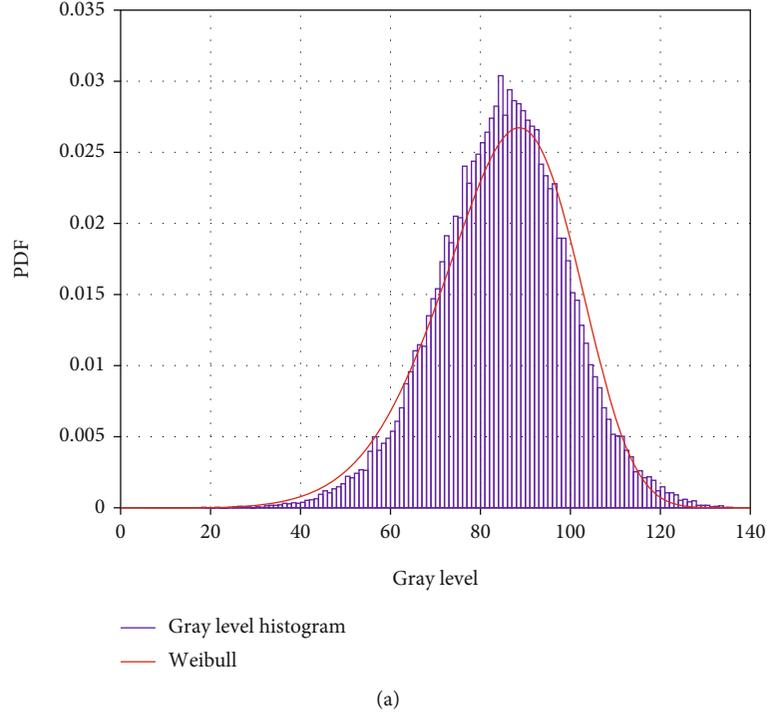


FIGURE 1: Statistical distribution of background noise: (a) histogram and estimated Weibull probability density function; (b) original sonar image.

In Section 2.1, we saw that the speckle noise of MBS images is randomly distributed. Here, we denote the pixel value and speckled noise of a point  $(x, y)$  as  $I(x, y)$  and  $N(x, y)$ , respectively. We then proceed to calculate the gradient magnitude as

$$\begin{aligned}
 G_\alpha &= \log (R_{\theta, \alpha}) = \log \left( \frac{\bar{I}_1(x, y) \cdot \bar{N}_1(x, y)}{\bar{I}_2(x, y) \cdot \bar{N}_2(x, y)} \right) \\
 &= \log (\bar{I}_1(x, y)) + \log (\bar{N}_1(x, y)) \\
 &\quad - \log (\bar{I}_2(x, y)) - \log (\bar{N}_2(x, y)),
 \end{aligned} \tag{11}$$

where  $\bar{I}_1(x, y)$ ,  $\bar{N}_1(x, y)$  and  $\bar{I}_2(x, y)$ ,  $\bar{N}_2(x, y)$  are the exponentially weighted values of the horizontal and vertical directions of the image, respectively. In the case of speckle noise, the direction distinction is unnecessary, as the noise intensities are assumed to be approximately equal in the horizontal and vertical directions. Thus, Equation (11) can be approximated as

$$G_\alpha \approx \log (\bar{I}_1(x, y)) - \log (\bar{I}_2(x, y)) = \log \left( \frac{\bar{I}_1(x, y)}{\bar{I}_2(x, y)} \right). \tag{12}$$

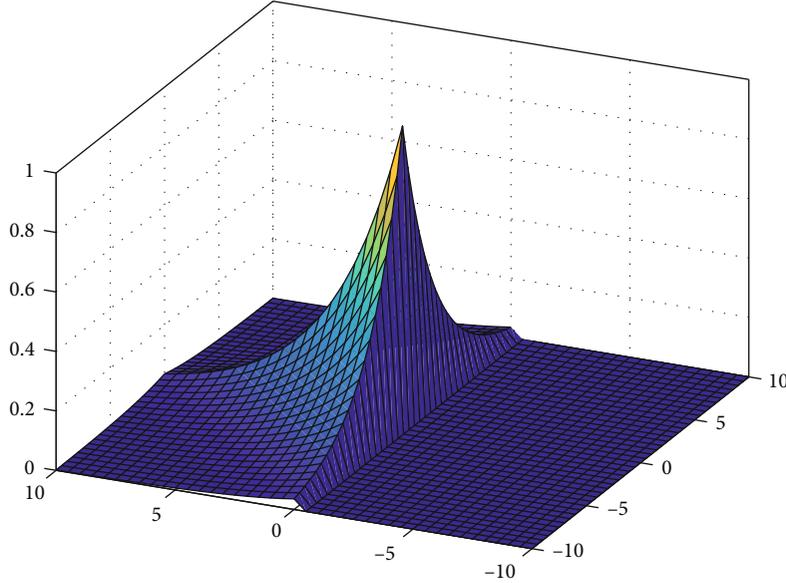


FIGURE 2: Exponential function for computation of weighted means.

From Equation (12), we see that the gradient information of the image is independent of the region's value on both sides of the pixel and only related to the weighted ratio of the intensities in the two directions.

By using the logarithm, the aforementioned problem of determining the gradient orientation on a vertical edge is avoided, since the computation yields

$$\begin{aligned} G_{x,\alpha} &= \log(m_a) - \log(m_b), \\ G_{y,\alpha} &= 0, \\ \text{Ori}_{t,\alpha} &= 0. \end{aligned} \quad (13)$$

In order to obtain both negative and positive gradient values, no normalization is performed between the ratio and its inverse. With this approach, all possibilities of orientation values are taken into account.

Figure 3 presents the gradient magnitude computed by difference and ratio on a rectangle corrupt using speckle noise. Compared with the region without noise, the values produced using differences (Figure 3(b)) are higher and inhomogeneous within the rectangle. Unlike the gradient by difference, the GR method does not produce higher values on the rectangle corrupt by noise, while the values are homogeneous (Figure 3(c)). This shows that gradient by difference is susceptible to noise, while GR is only affected by the neighborhood weighted ratio of pixels and the gradient magnitude is hardly affected by noise.

### 3. SIFT-Like Algorithm Adapted to Multibeam Sonar Images

In this section, we first introduce the original SIFT algorithm. It includes three steps, keypoint detection, orientation assignment, and descriptor representation. In order to adapt the SIFT algorithm to MBS images, we present the MBS-SIFT

algorithm, the corresponding steps of which are keypoint detection using the MBS-Harris detector, orientation assignment, and descriptor extraction via the GR method.

**3.1. Original SIFT Algorithm.** The SIFT algorithm has been introduced to characterize local gradient information. The SIFT descriptor is a sparse feature representation that combines two operators: a feature detector and a feature descriptor. Detection involves selecting points of interest, and descriptors are then calculated to describe these features. The SIFT algorithm mainly consists of keypoint detection, orientation assignment, and descriptor representation.

**3.1.1. Keypoint Detection.** The first stage of keypoint detection is to select and identify position  $(x, y)$ , scale  $\sigma$ , and orientation  $\theta$  that can be repeatedly assigned under different conditions.

The difference-of-Gaussian (DoG) scale space, as a close approximation of the scale-normalized Laplacian of Gaussian (LoG), is constructed with scales  $\sigma_l = \sigma_0 \cdot r^l$  and  $l \in \{0, 1, \dots, l_{\max} - 1\}$ . Each sample point is compared to its eight neighbors in the current image and nine neighbors in the scale above and below. Local extrema are then selected to obtain keypoints defined through their location and scale. Candidates with low contrast or located on edges are filtered using a threshold on the multiscale Harris corner detector, which is based on the Harris matrix. The Harris matrix and corner detector are, respectively, defined as

$$\begin{aligned} C(x, y, \sigma) &= \sigma^2 \cdot G_{\sqrt{2}\sigma} * \begin{bmatrix} \left(\frac{\partial I_\sigma}{\partial x}\right)^2 & \left(\frac{\partial I_\sigma}{\partial x}\right) \cdot \left(\frac{\partial I_\sigma}{\partial y}\right) \\ \left(\frac{\partial I_\sigma}{\partial x}\right) \cdot \left(\frac{\partial I_\sigma}{\partial y}\right) & \left(\frac{\partial I_\sigma}{\partial y}\right)^2 \end{bmatrix}, \\ R(x, y, \sigma) &= \det(C(x, y, \sigma)) - t \cdot \text{tr}(C(x, y, \sigma))^2, \end{aligned} \quad (14)$$

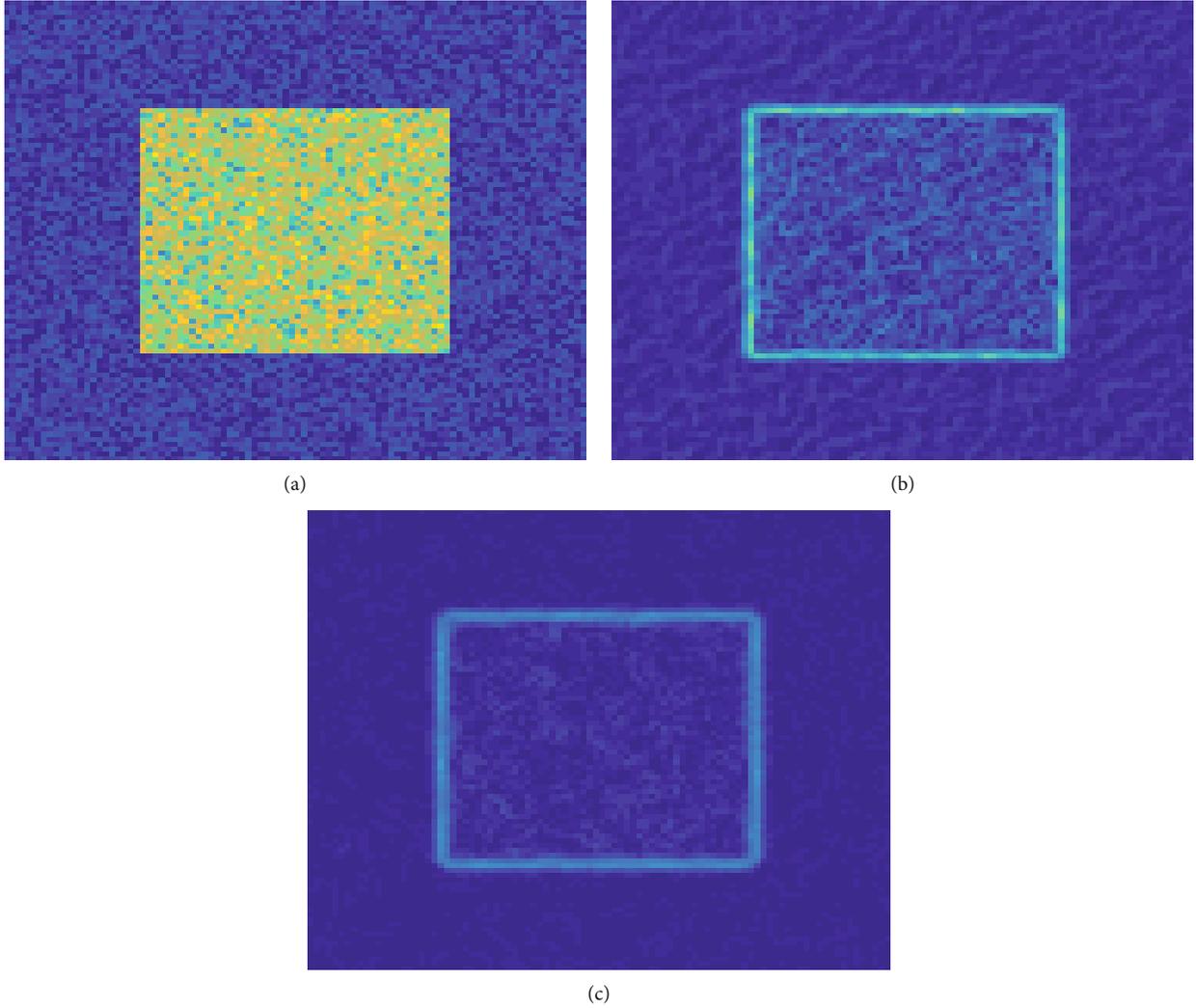


FIGURE 3: A rectangle with speckle noise and its gradient magnitude for two gradient computation methods: (a) original image; (b) gradient by difference; (c) gradient by ratio.

where  $G_{\sqrt{2}\cdot\sigma}$  is a Gaussian kernel with standard deviation  $\sqrt{2}\cdot\sigma$ ,  $I_{\sigma}$  is the convolution of the original image with a Gaussian kernel with standard deviation  $\sigma$ ,  $t$  is an arbitrary parameter, and  $*$  is the convolution operator in  $x$  and  $y$ .

In this study, the keypoints are detected as local extrema in the LoG scale space. To eliminate points lying on edges or having low contrast, the multiscale Harris criterion is applied [28]. We refer to this approach as the LoG-Harris detector.

**3.1.2. Orientation Assignment.** By assigning a consistent orientation to each keypoint based on local image properties, the keypoint descriptor can be expressed relative to this orientation. In this manner, we achieve invariance to image rotation. After substantial experimental comparisons, Lowe presented a method for the calculation of a local histogram of gradient orientations, weighted by the gradient magnitudes and a Gaussian-weighted circular window [29]. The orientation histogram is formed on a scale-dependent region around the keypoint. The highest peak in the histogram is determined, and then, any other local peak above 80% of

the maximum is also selected as an orientation. Therefore, there will be multiple keypoints detected at the same position and scale but with different orientations. In [13], the orientation histogram was divided into 8 bins covering the  $360^{\circ}$  range of orientations.

**3.1.3. Descriptor Representation.** Figure 4 shows a keypoint descriptor calculated to describe the local geometry in each keypoint  $P(x, y, \sigma, \theta)$ . First, a square neighborhood is defined around the keypoint with the scale of the keypoint to achieve translation and scale invariance. The gradient orientations are then rotated relative to the keypoint orientation to achieve orientation invariance, illustrated using small arrows at each sample location in Figure 4(a). The gradient magnitudes are weighted using a Gaussian function to avoid sudden changes and to give less emphasis to gradients that are far from the center of the descriptor. This is indicated using a circular window in Figure 4(a).

This normalized neighborhood is then divided into  $4 \times 4$  sample regions, as shown in Figure 4(b). Each region consists of eight directions for each orientation histogram,

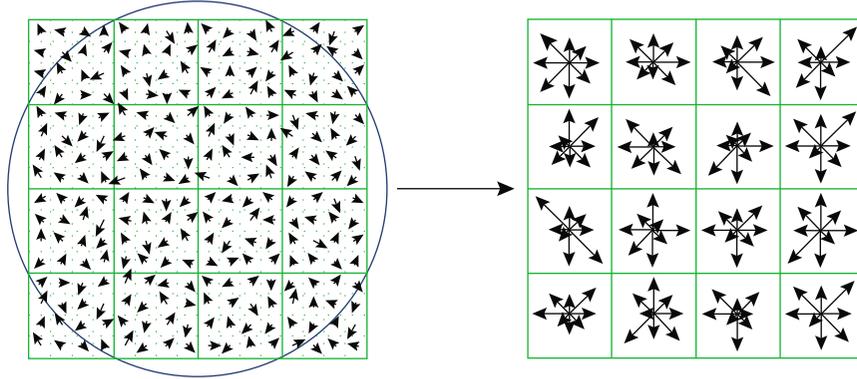


FIGURE 4: The SIFT descriptor is created by calculating the gradient magnitude and orientation at each point in the neighborhood of a keypoint. It is weighted by a Gaussian window, indicated by the overlaid circle shown in (a). This normalized neighborhood is divided into  $4 \times 4$  sample regions with 8 orientation bins in each, as shown in (b). Therefore, a 128 element feature vector is obtained as the SIFT descriptor of the keypoint.

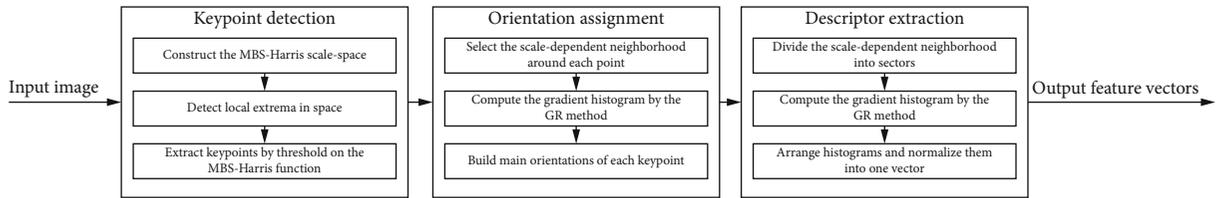


FIGURE 5: Outline of the MBS-SIFT algorithm, which consists mainly of three parts: keypoint detection, orientation assignment, and descriptor extraction.

with the length of each arrow corresponding to the magnitude of that histogram entry. For each keypoint, the SIFT descriptor is obtained by concatenating and normalizing these histograms.

The original SIFT algorithm was designed to detect structures that are specific to optical photography with low noise conditions. In the LoG-Harris detector, the multiscale Harris criterion allows the suppression of low-contrast and edge points by applying a threshold  $r_H$  on  $R(x, y, \sigma)$ . However, the speckle noise in MBS images leads to a stronger gradient magnitude on homogeneous areas with high reflectivity than on areas with low reflectivity. The SIFT algorithm does not perform well on MBS images, since the calculation of gradient and orientation relies on a classical gradient by difference.

**3.2. Proposed MBS-SIFT Algorithm.** To adapt the SIFT algorithm to MBS images, it is necessary to consider their statistical properties. The new gradient calculation method for MBS images presented in Section 2.2 is applied in several steps of the proposed algorithm. Both the magnitude and the orientation obtained are robust to speckle noise.

This new algorithm, named MBS-SIFT, operates in a similar way as the original SIFT algorithm, i.e., with a feature detector followed by a feature descriptor. In this section, a new keypoint detector is introduced, as well as a new orientation assignment and a MBS-adapted feature descriptor. The outline of this algorithm is presented in Figure 5. The main contributions of this paper are the keypoint detection based on the MBS-Harris detector and the histogram of gradient orientation obtained using the GR method.

**3.2.1. Keypoint Detection Based on MBS-Harris Detector.** LoG matrices rely on the second derivative, so they are not convenient to adapt to speckle noise. In contrast, the multiscale Harris function is based on the first derivative. Based on the new gradient computation adapted to MBS images mentioned in Section 2, we present a new algorithm to detect the keypoints.

Combining the multiscale Harris matrix, the function defined in (14), and the GR calculation, we propose the multiscale MBS-Harris matrix and the multiscale MBS-Harris function, respectively, as

$$C_{MH}(x, y, \sigma) = G_{\sqrt{2}\sigma} * \begin{bmatrix} (G_{x,\alpha})^2 & (G_{x,\alpha}) \cdot (G_{y,\alpha}) \\ (G_{x,\alpha}) \cdot (G_{y,\alpha}) & (G_{y,\alpha})^2 \end{bmatrix},$$

$$R_{MH}(x, y, \sigma) = \det(C_{MH}(x, y, \sigma)) - d \cdot \text{tr}(C_{MH}(x, y, \sigma))^2, \quad (15)$$

where the derivatives  $G_{x,\alpha}$  and  $G_{y,\alpha}$  are calculated in Equation (9) and  $d$  is an arbitrary parameter.

In the MBS-Harris detector, instead of the LoG scale space, we use a multiscale representation of the original image, which is achieved by the multiscale MBS-Harris function at different scales  $\beta_m = \beta_0 \cdot c^m$ , where  $m \in \{0, 1, \dots, m_{\max} - 1\}$ . Local extrema are then detected and compared with the 8 adjacent points of the same scale and the 18 points corresponding to the scales adjacent above and below. Here, a bilinear interpolation is used to refine the subpixel location of the keypoints, while we suppress low-contrast and edge

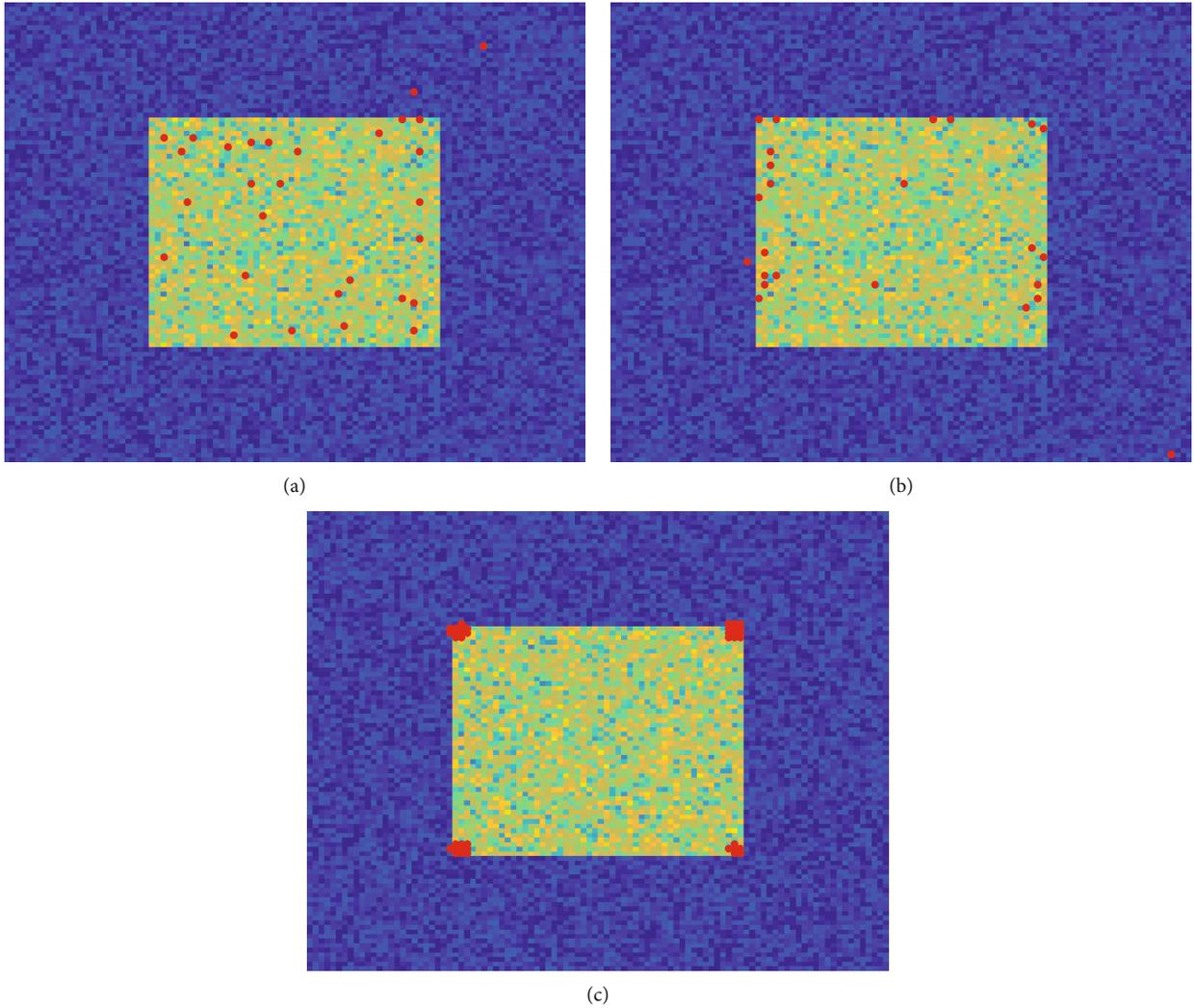


FIGURE 6: Detection of keypoints on a rectangle corrupt by speckle noise with the LoG-Harris detector applied on the amplitude image and the logarithm image, respectively, compared with the MBS-Harris detector applied on the logarithm image. (a) LoG-Harris detector applied on amplitude image (29 keypoints). (b) LoG-Harris detector applied on the logarithm image (24 keypoints). (c) MBS-Harris detector applied on the logarithm image (39 keypoints).

points by applying a threshold  $r_{MH}$  on the multiscale MBS-Harris function  $R_{MH}(x, y, \sigma)$ . Finally, the keypoints are characterized using their location and scale. The MBS-Harris detector improves robustness to speckle noise due to its use of the first derivative instead of the second derivatives adopted in the LoG-Harris detector. In Figure 6(c), it is easy to verify the efficiency of the MBS-Harris detector in the image with a rectangle corrupt using speckle noise; keypoints are only detected on the corners without spurious detections.

In Figure 6, the LoG-Harris detection is applied on the amplitude and the logarithm images, respectively. With this approach, keypoints are indeed detected near corners in the amplitude image (see Figure 6(a)), but they are poorly located with fewer spurious detections on homogeneous areas and on edges. Although LoG-Harris applied on the logarithm image performs well in dealing with additive noise, it is not robust enough to speckle noise, and there are still some spurious detections (Figure 6(b)). By adjusting the param-

eters on the multiscale Harris criterion, the number of false detections can be reduced, but the number of correct ones will be also decreased.

*3.2.2. Orientation Assignment and Descriptor Extraction via the GR Method.* In the original SIFT algorithm, histograms of gradient orientation are calculated in the neighborhood of each keypoint and weighted using the gradient magnitude and a Gaussian window. As a result, both orientation assignment and descriptor extraction are insensitive to speckle noise. In this paper, we calculate these histograms using the GR method, as introduced in Section 2.2. First, we build a histogram of gradient orientations in the neighborhood of each point and segment it to obtain reference directions. A set of 9 circular histograms of gradient orientations with respect to the reference direction is then built, as shown in Figure 7. In log-polar coordinates, these histograms correspond to 9 disjoint regions (a central region, 4 regions on

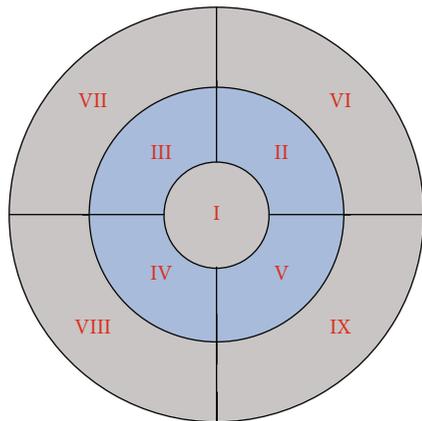


FIGURE 7: Scheme of the circular descriptor. The log-polar grid shows nine location bins: a central region, 4 regions on the first ring, and 4 more regions on the second ring.

the first ring, and 4 more regions on the second ring) of the neighborhood of each interest point. The complete  $360^\circ$  range of orientations is divided into 12 bins. We select the main modes of the local orientation histogram using the contrario mode selection method mentioned in [30], and each keypoint is allowed to have at most two orientations.

Instead of using a square neighborhood and  $4 \times 4$  square sectors with 8 orientation bins as in the original SIFT descriptor, we rely on a circular neighborhood and 9 disjoint regions with 12 orientation bins. Again, the GR method is used to compute the orientations. This descriptor is obtained in a very similar way to the circular-SIFT descriptor, i.e., by connecting the orientation histogram corresponding to the log-polar coordinates. The only difference is the use of GR instead of the gradient by difference. We refer to this resulting descriptor as the ratio descriptor.

#### 4. Experimental Validation of the MBS-SIFT Algorithm

In this section, we first introduce the MBS and the parameter settings of algorithms presented in this paper. After processing the MBS images, the LoG-Harris detector and MBS-Harris detector are compared and analyzed. Finally, we combine the LoG-Harris detector and MBS-Harris detector with SIFT and the ratio descriptors, respectively, and evaluate their performance using ROC curves.

**4.1. Data Acquisition and Parameter Settings.** MBS images were acquired using Blueprint's Oculus M750d MBS, which is intended for near-field target identification. Acoustic image sequences for fish detection were selected to assess the method proposed in this paper. The sonar covers a region over  $70^\circ$  (horizontal)  $\times$   $12^\circ$  (vertical) and allows for the collection of 512 beams with an operating frequency of 1.2 MHz. Its angular resolution is  $0.6^\circ$ , and the maximum update rate is 40 Hz.

For the LoG-Harris detector, the scale-space representation started from a detection scale of 0.5 and the scale factor between two levels of resolution was set to  $r = 2^{1/3}$ . We used

17 scale levels. For the multiscale Harris criterion, the parameter  $t$  was empirically set to 0.04 and the threshold  $r_H$  was set to 4000 for 8-bit sonar images but was adapted for each MBS image to obtain the same number of keypoints.

For the MBS-Harris detector, the parameters chosen were as follows: the first scale was set to  $\beta_0 = 2$ ; the ratio between the two scales was  $c = 2^{1/3}$ ; the number of scales was  $m_{\max} = 8$ ; the arbitrary parameter of the MBS-Harris criterion was set to  $d = 0.04$ . The remaining scales were then obtained by setting  $\beta_m = \beta_0 \cdot c^m$ . After an experimental study of interest points in sonar images, the threshold  $r_{MH}$  applied on the multiscale MBS-Harris function was set to 0.05.

**4.2. Results and Analysis of Keypoint Detection.** We used 54 MBS image pairs to verify the proposed algorithm. Single frame and sequences of MBS images were used to compare the different keypoint detectors. In Figure 8, we see sonar images of a single fish target and a metal chain. As expected, using the LoG-Harris detector (Figures 8(a) and 8(c)), the keypoints detected are mainly in high-reflectivity areas, but many false detections occur in the background corrupt by speckle noise. However, the keypoints detected using the MBS-Harris detector (Figures 8(b) and 8(d)) are mostly located on bright points and corners of the fish target and the metal chain. The number of false detections is considerably lower than that of the LoG-Harris detector.

Figure 9 presents the keypoint detection of two fish targets in a sonar image sequence using the LoG-Harris and MBS-Harris detectors. In Figures 9(a)–9(d), many false detections are obtained using the LoG-Harris detector due to the influence of speckle noise. In Figures 9(e)–9(h), we see that the MBS-Harris detector performs better and the keypoints are detected on the corners and the bright points of the two fish targets.

**4.3. Matching Performance Evaluation.** Keypoints of two sonar images are matched according to their descriptors. One approach is to select for each descriptor in the first image the nearest neighbor in the second image based on the minimum Euclidean distance. However, many interest points from the first image will not have any correct match in the second image because they were the result of the background's speckle noise influence. To match the two images robustly, a more effective measure is obtained by comparing the distance of the closest neighbor to that of the second-closest neighbor. This matching criterion is named the nearest neighbor distance ratio method [13]. First, for each descriptor in the first image, we select the nearest and second-nearest neighbor in the second image based on the Euclidean distance. To filter false matches, the distances to the closest and second-closest neighbors are then compared. A threshold  $th$  is applied on the ratio of those respective distances. This measure performs well because correct matches need to have the closest neighbor significantly closer than the closest incorrect match to achieve reliable matching.

ROC curves were computed for different combinations of keypoint detectors and descriptors by varying  $th$ . The keypoints of two images were matched according to their

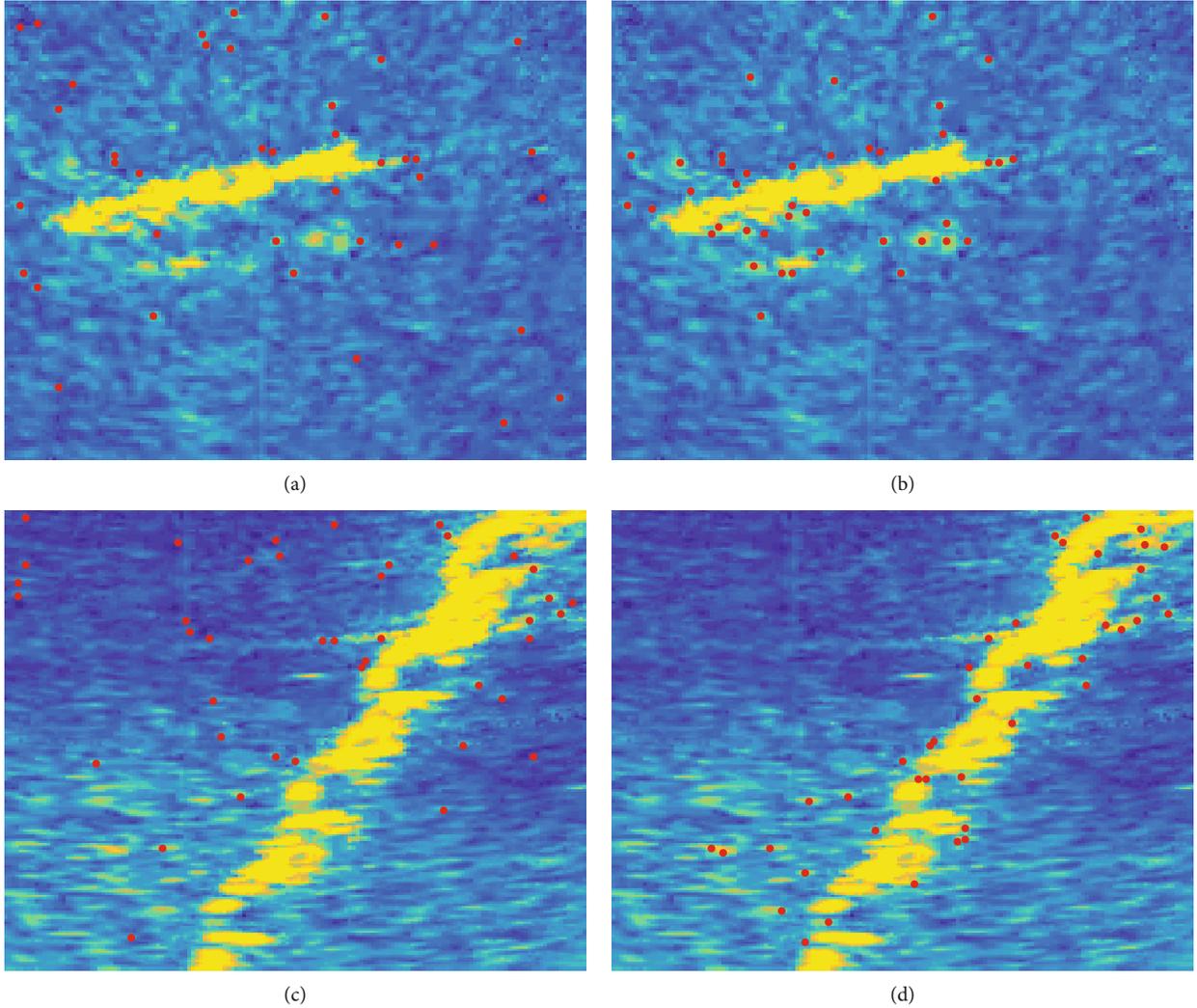


FIGURE 8: Detection of keypoints on the single frame MBS image using the LoG-Harris and the MBS-Harris detectors.  $r_H$  and  $r_{MH}$  were adjusted to obtain the same number of keypoints with both detectors: (a) single fish target using LoG-Harris; (b) single fish target using MBS-Harris; (c) metal chain using LoG-Harris; (d) metal chain using MBS-Harris.

respective descriptors. The percentage of correctly matched keypoints  $P_{CM}$  and the false match rate  $P_{FM}$  were defined as

$$P_{CM} = \frac{N_{CM}}{N_T}, \quad (16a)$$

$$P_{FM} = \frac{N_{FM}}{N - N_T}, \quad (16b)$$

where  $N$  is the total number of possible matches,  $N_T$  is the number of true correct matches, and  $N_{CM}$  and  $N_{FM}$  are the numbers of correct and false matches, respectively, at a certain value of the threshold.

Here, we compare two feature detectors, the LoG-Harris detector and the proposed MBS-Harris detector, and two feature descriptors, the proposed ratio descriptor and the SIFT descriptor. The ROC curves of the four considered situations were calculated for the 54 MBS image pairs and are shown in Figure 10. We see from the figure that the best performance was achieved by the combination of the MBS-Harris detector

and the ratio descriptor. Nearly 60% of correct matches were obtained at a false rate of 1%. The corresponding rate of the MBS-Harris detector/SIFT descriptor combination was 48%. For the other two configurations, this rate was less than 40%. In conclusion, the MBS-Harris detector was more robust to speckle noise in sonar images than the LoG-Harris detector, and the ratio descriptor achieves better results than the SIFT descriptor.

**4.4. Computational Efficiency.** To evaluate the computational efficiency of the proposed MBS-Harris detector and ratio descriptor, the 54 MBS image pairs collected using Blueprint's Oculus M750d MBS were used to test and compare our method with the other three methods (LoG-Harris detector+SIFT descriptor, LoG-Harris detector+ratio descriptor, and MBS-Harris detector+SIFT descriptor). The MBS images' size was  $129 \times 166$ . The hardware and software resources (i.e., Intel® Core™ i5-7200U CPU @ 2.50 GHz, 12 GB RAM, HDD 500 GB, MATLAB 2016a, and Windows 10 64-bit OS) were the same for all the above algorithms.

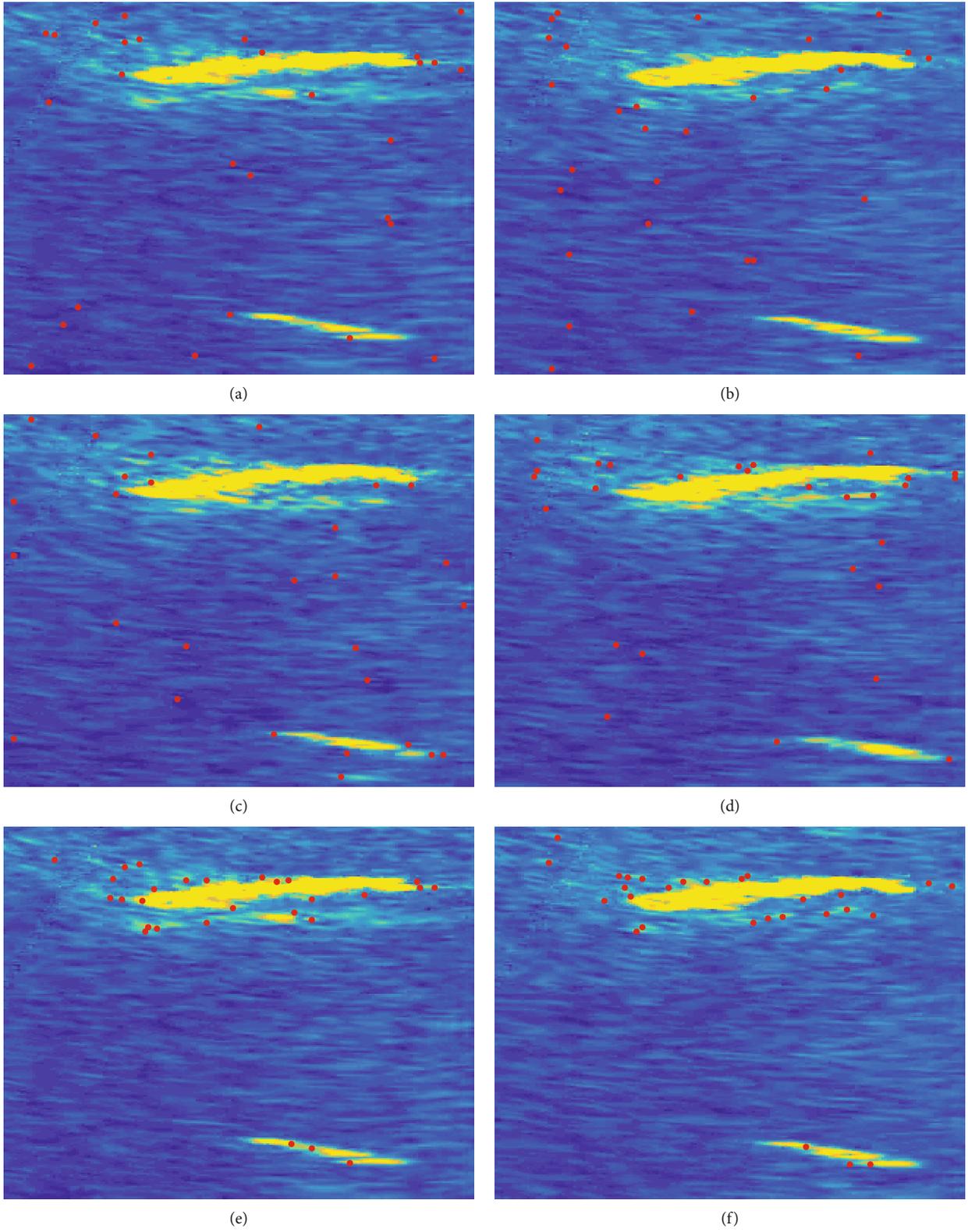


FIGURE 9: Continued.

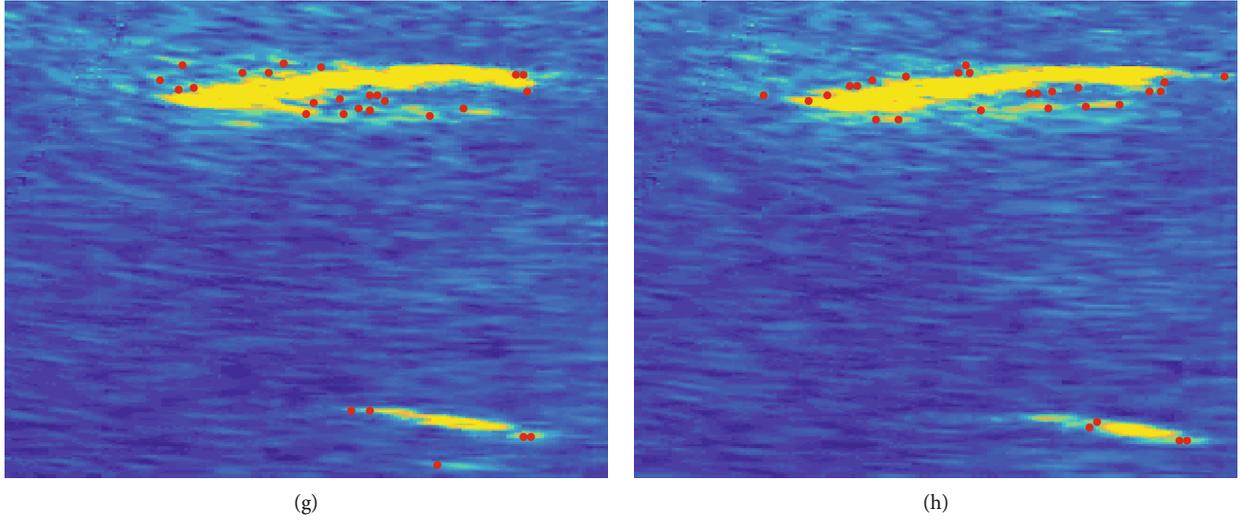


FIGURE 9: Detection of keypoints on the MBS image sequences using the LoG-Harris and MBS-Harris detectors.  $r_H$  and  $r_{MH}$  were adjusted to obtain the same number of keypoints with all keypoint detectors: (a–d) two fish targets using LoG-Harris; (e–h) two fish targets using MBS-Harris.

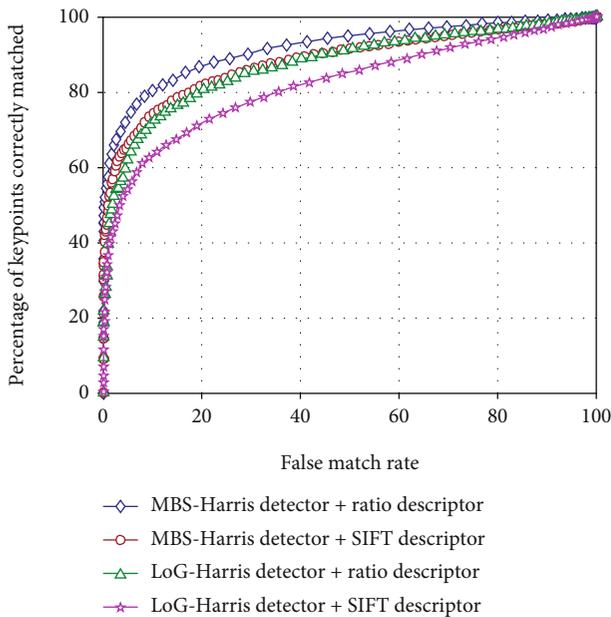


FIGURE 10: ROC curves, calculated on 54 MBS image pairs, for evaluating the performance of the SIFT and ratio descriptors and the LoG-Harris and MBS-Harris detectors.

TABLE 1: Running time (in seconds) for feature detection and description.

LoG-Harris detector+SIFT descriptor	LoG-Harris detector+ratio descriptor	MBS-Harris detector+SIFT descriptor	MBS-Harris detector+ratio descriptor
0.4881	0.4219	0.3613	0.3084

The average running times of the proposed approach and the other three methods are presented in Table 1. The proposed MBS-Harris detector and ratio descriptor method

took significantly less time to complete the required task and are therefore more computationally efficient.

## 5. Conclusion

In this paper, we present the MBS-SIFT algorithm, which combines a keypoint detector with a feature descriptor adapted to MBS images. A new gradient calculation approach specifically tailored to MBS images, the GR method, is applied in several steps of the proposed algorithm to make the proposed algorithm robust to speckle noise. Based on the multiscale Harris detector, this new keypoint detector offers consistent keypoints of various targets in sonar images. The robust gradient magnitudes and orientations obtained using GR allow the implementation of an efficient feature descriptor for MBS images.

From the ROC curves, we observe that the MBS-SIFT algorithm achieves better results than the traditional SIFT algorithm. In single frame sonar images, the keypoints are mainly detected on the corners and bright points of the targets. In the continuous multiframe sonar images, the features of multiple objects are detected consistently. Thus, a better matching effect is obtained. In future work, we will apply the algorithm to target tracking and the measurement of target velocity in sonar images. In addition, sonar image matching under different conditions will be considered.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. U1709203, Grant No. 52001097) and the National Science and Technology Major Project of China (Grant No. 2016ZX05057005).

## References

- [1] K. Haris, B. Chakraborty, B. Ingole, A. Menezes, and R. Srivastava, "Seabed habitat mapping employing single and multi-beam backscatter data: a case study from the western continental shelf of India," *Continental Shelf Research*, vol. 48, pp. 40–49, 2012.
- [2] J. Gao, H. Li, B. Chen, T. Zhou, C. Xu, and W. Du, "Fast two-dimensional subset censored CFAR method for multiple objects detection from acoustic image," *IET Radar, Sonar & Navigation*, vol. 11, no. 3, pp. 505–512, 2017.
- [3] D. Peng, T. Zhou, C. Xu, W. Zhang, and J. Shen, "Marginalized point mass filter with estimating tidal depth bias for underwater terrain-aided navigation," *Journal of Sensors*, vol. 2019, 15 pages, 2019.
- [4] F. Chaillan, C. Fraschini, and P. Courmontagne, "Speckle noise reduction in SAS imagery," *Signal Processing*, vol. 87, no. 4, pp. 762–781, 2007.
- [5] T. Celik and T. Tjahjadi, "A novel method for sidescan sonar image segmentation," *IEEE Journal of Oceanic Engineering*, vol. 36, no. 2, pp. 186–194, 2011.
- [6] J. Grabek and B. Cyganek, "Speckle noise filtering in side-scan sonar images based on the Tucker tensor decomposition," *Sensors*, vol. 19, no. 13, 2019.
- [7] H. B. Thomas and Y. V. Zakharov, "Attitude-trajectory estimation for forward looking multi-beam sonar based on acoustic image registration," *IEEE Journal of Oceanic Engineering*, vol. 44, no. 3, pp. 753–766, 2019.
- [8] L. Ju, K. Xie, H. Zheng, B. Zhang, and W. Yang, "GPCA-SIFT: a new local feature descriptor for scene image classification," *Communications in Computer and Information Science*, vol. 663, no. 4, pp. 286–295, 2016.
- [9] N. Abbas, T. Saba, S. Khan, Z. Mehmood, and R. Tabasum, "Reduced reference image quality assessment technique based on DWT and path integral local binary patterns," *Arabian Journal for Science and Engineering*, vol. 45, no. 4, pp. 3387–3401, 2020.
- [10] R. Bibi, Z. Mehmood, R. M. Yousaf, T. Saba, M. Sardaraz, and A. Rehman, "Query-by-visual-search: multimodal framework for content-based image retrieval," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, pp. 5629–5648, 2020.
- [11] R. Bibi, Z. Mehmood, R. M. Yousaf et al., "BoVW model based on adaptive local and global visual words modeling and log-based relevance feedback for semantic retrieval of the images," *EURASIP Journal on Image and Video Processing*, vol. 2020, no. 27, pp. 1–30, 2020.
- [12] Z. Tang, G. Ma, J. Lu, Z. Wang, F. Bin, and Y. Wang, "Sonar image mosaic based on a new feature matching method," *IET Image Processing*, vol. 14, no. 10, pp. 2149–2155, 2020.
- [13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [14] D. R. Kisku and S. Rana, "Multithread face recognition in cloud," *Journal of Sensors*, vol. 2016, 21 pages, 2016.
- [15] C. Tao, Y. Tan, H. Cai, and J. Tian, "Airport detection from large IKONOS images using clustered SIFT keypoints and region information," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 1, pp. 128–132, 2011.
- [16] B. Sirmacek and C. Unsalan, "Urban-area and building detection using SIFT keypoints and graph theory," *IEEE Transactions on Geoscience & Remote Sensing*, vol. 47, no. 4, pp. 1156–1167, 2009.
- [17] M. Huang, Z. Mu, H. Zeng, and H. Huang, "A novel approach for interest point detection via Laplacian-of-bilateral filter," *Journal of Sensors*, vol. 3, 9 pages, 2015.
- [18] B. Li, R. Xiao, Z. Li, R. Cai, B.-L. Lu, and L. Zhang, "Rank-SIFT: learning to rank repeatable local interest points," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 1737–1744, Providence, RI, USA, June 2011.
- [19] J. Liu and X. Yu, "Research on SAR image matching technology based on SIFT," in *21st International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 37no. Part B1, pp. 403–408, Beijing, China, 2008.
- [20] V. Wang and M. P. Hayes, "Image registration of simulated synthetic aperture sonar images using SIFT," in *Proceedings of the 29th International Conference on Image and Vision Computing*, pp. 31–36, New Zealand, 2014.
- [21] P. Schwind, S. Suri, P. Reinartz, and A. Siebert, "Applicability of the SIFT operator to geometric SAR image registration," *International Journal of Remote Sensing*, vol. 31, no. 8, pp. 1959–1980, 2010.
- [22] G. Tesfaye, B. Marchand, J. D. Tucker, D. D. Sternlicht, T. M. Marston, and M. R. Azimi-Sadjadi, "Automated change detection for synthetic aperture sonar," in *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XIX*, vol. 9072, International Society for Optics and Photonics, 2014.
- [23] B. Fan, C. Huo, C. Pan, and Q. Kong, "Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 657–661, 2013.
- [24] A. C. Bovik, "On detecting edges in speckle imagery," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 10, pp. 1618–1627, 1988.
- [25] R. Fjortoft, A. Lopes, P. Marthon, and E. Cubero-Castan, "An optimal multiedge detector for SAR image segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 793–802, 1998.
- [26] S. Suri, P. Schwind, J. Uhl, and P. Reinartz, "Modifications in the SIFT operator for effective SAR image matching," *International Journal of Image and Data Fusion*, vol. 1, no. 3, pp. 243–256, 2010.
- [27] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: a SIFT-like algorithm for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 453–466, 2015.
- [28] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 1, pp. 525–531, Vancouver, BC, Canada, 2001.

- [29] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, pp. 1150–1157, Kerkyra, Greece, 1999.
- [30] A. Desolneux, L. Moisan, and J.-M. Morel, *From Gestalt Theory to Image Analysis: A Probabilistic Approach*, vol. 34, Springer-Verlag, New York, NY, USA, 2008.