

# Article A Local Feature Descriptor Based on Oriented Structure Maps with Guided Filtering for Multispectral Remote Sensing Image Matching

Tao Ma $^{1,2}$ , Jie Ma $^{1,2,3,\ast}$  and Kun Yu $^{1,2}$ 

- <sup>1</sup> National Key Laboratory of Science and Technology on Multispectral Information Processing, Huazhong University of Science and Technology, Wuhan 430074, China; matao@hust.edu.cn (T.M.); wh\_ykun@hust.edu.cn (K.Y.)
- <sup>2</sup> School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China
- <sup>3</sup> Science and Technology on Complex System Control and Intelligent Agent Cooperation Laboratory, China Aerospace Science and Industry Corporation third Academy, Beijing 100074, China
- \* Correspondence: majie@mail.hust.edu.cn; Tel.: +86-180-6261-2780

Received: 22 March 2019; Accepted: 18 April 2019; Published: 20 April 2019



Abstract: Multispectral image matching plays a very important role in remote sensing image processing and can be applied for registering the complementary information captured by different sensors. Due to the nonlinear intensity difference in multispectral images, many classic descriptors designed for images of the same spectrum are unable to work well. To cope with this problem, this paper proposes a new local feature descriptor termed histogram of oriented structure maps (HOSM) for multispectral image matching tasks. This proposed method consists of three steps. First, we propose a new method based on local contrast to construct the structure guidance images from the multispectral images by transferring the significant contours from source images to results, respectively. Second, we calculate oriented structure maps with guided image filtering. In details, we first construct edge maps by the progressive Sobel filters to extract the common structure characteristics from the multispectral images, and then we compute the oriented structure maps by performing the guided filtering on the edge maps with the structure guidance images constructed in the first step. Finally, we build the HOSM descriptor by calculating the histogram of oriented structure maps in a local region of each interest point and normalize the feature vector. The proposed HOSM descriptor was evaluated on three commonly used datasets and was compared with several state-of-the-art methods. The experimental results demonstrate that the HOSM descriptor can be robust to the nonlinear intensity difference in multispectral images and outperforms other methods.

**Keywords:** remote sensing image; multispectral image; feature matching; oriented structure maps; guided image filtering

# 1. Introduction

Feature matching is a process that can obtain the correspondence of interest points among two or more images of the same scene with varying degrees of overlap [1]. The correspondence relationships can be used to geometrically align these images [2]. Multispectral images often provide complementary information about the same scene by capturing the different characteristics of different spectral bands. For example, an infrared image captures thermal radiation of the objects in a scene, while a visible image mainly records optical reflection information in a scene. Therefore, feature matching between the multispectral images is at the base of many computer vision and remote sensing applications, such as target recognition [3,4], image registration [5,6], modern military surveillance [7,8],



3D reconstruction [9,10], image fusion [11,12], and medical image processing [13,14]. Due to differences in sensors and spectra between multispectral images, there are significant nonlinear intensity variations, which mean that the grayscales and intensity distributions between multispectral images are nonlinearly correlated. This is because the scene has different reflection characteristics for the electromagnetic spectrum of different bands [15–17]. Therefore, obtaining accurate and robust matching between multispectral images is still a very challenging task. Although there are many different feature matching methods, they all have three steps in common: interest point detection, interest point description, and interest point matching [18–20]. This study focuses on the interest point description, which is the difficulty and core problem in the matching task.

In recent years, many excellent feature descriptors have been proposed for different matching applications. Among these methods, the scale invariant feature transform (SIFT) [21] and speeded-up robust feature (SURF) [22,23] are the most widely used algorithms. Due to the robustness of these two descriptors for scale invariance and rotation invariance, they can achieve good performance in many image matching tasks. However, when these two methods are applied to multispectral image matching works, they often fail to obtain correct correspondences. This is because they are robust to illumination and viewpoint changes but sensitive to nonlinear intensity changes [24–26].

Scholars have designed many feature descriptors to deal with the problem of significant nonlinear intensity differences in multispectral images, in which many methods are improvements to the classic descriptors. Saleem and Sablatnig [24] proposed a modification to SIFT termed normalized gradient SIFT (NG-SIFT), which utilizes normalized gradients to compute the feature vector as the description of interest points to achieve robustness against nonlinear intensity changes in multispectral images. A combined method based on SURF, partial intensity invariant feature descriptor (PIIFD), and robust point matching (RPM), called SURF-PIIFD-RPM [25] are presented to improve the matching performance of multimodal retinal images. This method first employs the SURF detector to extract scale invariant and stable feature points. And then it constructs the local feature descriptors based on PIIFD that can keep robust to intensity changes. Finally, the PRM method is used for feature matching and outliers removing.

Many methods are specially designed to work with multispectral images and can achieve better robustness and efficiency than these progressive SIFT-like or SURF-like algorithms. The paper [26] proposed the edge histogram descriptor (EHD) that has been adopted in the MPEG-7 standard. The EHD employs the spatial distribution of edge points to represent the robust image features, which can keep reliable texture information even when there are significant intensity variations between multispectral images. The paper [27] presented a new feature descriptor called the edge-oriented histogram (EOH), which was used for the matching task between the visible images and the long wave infrared (LWIR) images. The EOH descriptor uses the edge points distribution of four directional edges and one non-directional edge to construct the feature description, which can keep reliable structure information even when there are significant intensity variations between multispectral images. Although the EHD and EOH can describe image contour information, they have difficulty extracting highly similar edges from multispectral images. To solve this deficiency, the paper [28] proposed local descriptor call histograms of directional maps (HoDM), which combined the structure and texture features to descript a keypoint. Different from the EOH descriptor, the Log-Gabor histogram descriptor (LGHD) [29] and multispectral feature descriptor (MFD) [30] use multi-scale and multi-oriented Log-Gabor filters to replace the multi-oriented spatial filters. The LGHD can get richer and more robust feature representation in multispectral images but suffers from high dimensionality and low efficiency. The MFD can significantly reduce the feature dimension while maintaining feature description capabilities.

Figure 1 shows the comparison between the variations of gradient direction in multispectral images and corresponding structure guidance images. It can be seen that the structure guidance images of visible and infrared (IR) image can retain the common structures although the gradient directions are reversed. Many traditional methods usually use the gradient-oriented information to capture the

edges and to build the local feature vectors in multispectral images, which are good at characterizing local shape, but are sensitive to changes of the gradient directions, as shown in Figure 1a1,a2 and Figure 1b1,b2. In order to overcome this deficiency, we propose a new feature descriptor based on oriented Sobel edge maps with guided filtering.



**Figure 1.** Gradient direction variations and structure guidance images between multispectral images (**a**) visible image, and (**b**) infrared image. From top to bottom: (1) source images, (2) gradient directions, and (3) structure guidance images. The green dot indicates the position of the interest point.

The main contribution in this paper consists of three aspects.

(1) We design a new local feature descriptor called histogram of oriented structure maps (HOSM) for multispectral image matching.

(2) We propose a new guidance image extractor, as shown in Figure 1a3,b3, which can extract highly similar structure information from multispectral images.

(3) We provide evaluations on our method in three multi-source datasets and compare it with several state-of-the-art methods, the experiments show that HOSM is more robust to nonlinear intensity variations by keeping more structure features with guided filtering.

The remainder paper is organized as follows. Section 2 introduces the proposed HOSM descriptor. In Section 3, the experimental results and corresponding analyses are exhibited. Finally, we conclude this paper in Section 4.

# 2. Materials and Methods

In this part, we present the proposed HOSM descriptor. This method contains three parts. First, a new local contrast-based operator is proposed to construct the structure guidance image to preserve the significant contour characteristics in source images. Second, the oriented structure maps are computed with guided image filtering to capture the common structure properties in multispectral images. Finally, we calculate the histogram of oriented structure maps to build the feature vector.

#### 2.1. Construct Guidance Image

The significant contour information of an image is the key feature to characterize structures that can usually be expressed by the image gradients. Because of the nonlinear intensity variations in multispectral images, the gradient directions are inconsistent, which leads to the failure to obtain similar structures. To overcome this problem, we propose an isotropic extractor to construct the structure guidance image that can capture the significant contours between the multispectral images. The processing flow is shown in Figure 2 and extractor is expressed as follows.

$$G(x,y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} \frac{|I(x+i,y+j) - I_{mean}|}{\max(I(x+i,y+j), I_{mean})}$$
(1)

where *I* is source image, *G* is the structure guidance image, (x, y) represents the pixel location of filter region center, *k* is the radius of filter window, and *I*<sub>mean</sub> is the mean of intensities of the local region. Then, to get the robust guidance image, we normalize the magnitude of image *G* to the range of 0–255 as follows.



$$G(x, y) = 255 \times \frac{G(x, y) - \min G}{\max G - \min G}$$
<sup>(2)</sup>

**Figure 2.** Processing flow of structure guidance image construction. The red point indicates the center pixel of the local region.

#### 2.2. Compute Oriented Structure Maps

The Sobel operator has good noise-suppression ability and is simple to carry out but can only extract the horizontal and vertical edges. The progressive Sobel filters [26] shown in Figure 3 are able to detect multi-oriented edges in images. They can extract the spatial distribution of one non-directional edge and four oriented edges, including  $0^{\circ}$ ,  $45^{\circ}$ ,  $90^{\circ}$ , and  $135^{\circ}$ . This filter bank follows the MPEG-7 standard [31] and is able to represent the main structure information of the image [27]. Nunes and Padua [30] have pointed out that more orientations have little improvement in description capabilities but can significantly increase time consumption. In order to get a good trade-off between performance and efficiency, we use the five-oriented Sobel filters to construct structure maps. As shown in Figure 4a, let  $f_n(x, y)$ , n = 1, 2, 3, 4, 5 denotes the five Sobel filters of  $0^{\circ}$ ,  $45^{\circ}$ ,  $90^{\circ}$ ,  $135^{\circ}$ , and no orientation. The multi-oriented Sobel edges can be extracted by follows.

$$SE_n(x, y) = |I(x, y) * f_n(x, y)|$$
 (3)

where I(x, y) is pixel of source image, and  $SE_n(x, y)$  represents the Sobel edge of orientation n at the pixel (x, y). "\*" is the convolution operator, and "||" represents the absolute operation.  $SE_n$  is

normalized to the range of 0-255. The largest value at each pixel in different edge images is used to build the edge maps as follows.

$$EM_n(x,y) = \begin{cases} SE_n(x,y), & \text{if } SE_n(x,y) = \max\{SE_k(x,y)\}_1^5\\ 0, & \text{others} \end{cases}$$
(4)

where  $EM_n(x, y)$  shown in Figure 4b is edge maps. The edge maps maintain the common structure characteristics of multispectral images but suffer from the deficiencies that are sensitive to noise and can cause aliasing effects. To address these problems, the structure maps are computed by guided image filtering [32–34] on edge maps as follows.

$$SM_n = GF(EM_n, G, r, \varepsilon)$$
 (5)

where *G* is the guidance image,  $SM_n$  represents structure maps as shown in Figure 4c, *r* is the local window radius, and  $\varepsilon$  is the regularization parameter. In the guided filtering operation, the output image is computed by a linear transformation of the guidance image, which can be represented as follows.

$$SM_n(i) = a_m G(i) + b_m, \forall i \in \omega_m \tag{6}$$

where  $\omega_m$  is a local window,  $a_m$  and  $b_m$  are linear coefficients, and *i* represents the pixel location in the local window. To get the linear coefficients, we have to minimize the following cost function.

$$E(a_m, b_m) = \sum_{i \in \omega_m} \left( (a_m G(i) + b_m - EM_n(i))^2 + \varepsilon a_m^2 \right)$$
(7)

 $a_m$  and  $b_m$  can be represented as.

$$a_m = \frac{\frac{1}{|\omega|} \sum_{i \in \omega_m} G(i) E M_n(i) - \mu_m \overline{E M_n}}{\sigma_m^2 + \varepsilon}$$
(8)

$$b_m = \overline{EM_n} - a_m \mu_m \tag{9}$$

where  $\mu_m$  and  $\sigma_m^2$  represent the mean and variance of the guidance image,  $EM_n$  is the mean of input image, and  $|\omega|$  is the number of the pixels in the local window  $\omega_m$ . Pixel *i* can be contained in more than one local windows, in different local windows, the output values are changes. To solve this problem, the mean of all possible coefficients  $a_m$  and  $b_m$  are used as the final coefficients, as follows.

$$SM_n(i) = \overline{a_i}G(i) + \overline{b_i} \tag{10}$$

where  $\overline{a_i}$  and  $\overline{b_i}$  can be computed as follows.

$$\overline{a_i} = \frac{1}{|\omega|} \sum_{m \in \omega_i} a_m \tag{11}$$

$$\overline{b_i} = \frac{1}{|\omega|} \sum_{m \in \omega_i} b_m \tag{12}$$

-1	-2	-1	-2	-1	0	-1	0	1	0	1	2	-1	0	1
0	0	0	-1	0	1	-2	0	2	-1	0	1	0	0	0
1	2	1	0	1	2	-1	0	1	-2	-1	0	1	0	-1
(a)				(b)			(c)			(d)			(e)	

Figure 3. Sobel filters of five orientations at (a)  $0^{\circ}$ , (b)  $45^{\circ}$ , (c)  $90^{\circ}$ , (d)  $135^{\circ}$ , and (e) no orientation.



**Figure 4.** Illustration of oriented structure maps construction. From top to bottom: (a) Sobel edges  $SE_n$ , (b) edge maps  $EM_n$ , and (c) structure maps  $SM_n$ .

## 2.3. Proposed HOSM Descriptor Based on Structure Maps

The proposed HOSM descriptor employs the multi-oriented histogram to build the feature vector based on the oriented structure maps by guided image filtering. Figure 5 illustrates the main flowchart of the proposed HOSM descriptor, where the green point in the region center denotes an interest point. The main processing flow consists of three steps that can be summarized as follows.

- 1. First, a new structure extractor is applied to construct structure guidance images in the local region around the interest point, which can capture the significant contour information and keep robust to nonlinear intensity changes in multispectral images.
- 2. Second, we construct the oriented structure maps by performing guided filtering on multi-oriented Sobel edge maps. The detailed construction process of structure maps is shown in Figure 4. (1) The progressive Sobel filters are applied to construct the multi-oriented edge maps, which can extract the significant structure characteristics of images. (2) We perform guided filtering on edge maps to get the final structure maps. This guided filtering operation can enhance the structures in final structure maps while maintaining the advantages of multiple directions in the edge maps. Therefore, the oriented structure maps are able to remain the common features between multispectral images and reduce aliasing effects with the guided filtering.

3. Finally, we calculate the histogram of oriented structure maps in a local region to build the HOSM descriptor. We first divide the local region around the interest point into 16 sub regions by the  $N \times N$  window. We then compute and normalize the 5-orientation histogram by L2 norm in each sub region. At last, all the histograms of the 16 sub regions are conjoined to form a feature vector.



Figure 5. Main flowchart of the proposed HOSM descriptor. The green point of the region center indicates an interest point.

# 3. Experiments and Analyses

# 3.1. Datasets and Settings

In order to evaluate the description ability of our proposed HOSM descriptor, we carried out the experiments on three widely used datasets as shown in Figure 6 and compared it with seven state-of-the-art approaches that are SURF, SIFT, NG-SIFT, EOH, LGHD, MFD, and HoDM. Figure 6a is a Potsdam dataset that contains 38 visible (VIS) and near infrared (NIR) aerial image pairs and all image pairs have dimensions of  $6000 \times 6000$ . This dataset is generated from a remote sensing image dataset and can be accessed at [35]. The EPFL dataset was proposed in [36] as shown in Figure 6b. It consists of 477 visible and NIR images and all image pairs have dimensions of  $1024 \times 768$ . The CVC dataset was proposed in [27] as shown in Figure 6c, which contains 100 visible and long wave infrared (LWIR) image pairs and has dimensions  $506 \times 408$  for all images.



**Figure 6.** Samples of the multispectral image pairs from three datasets. (**a**) Potsdam (VIS/NIR) dataset, (**b**) EPFL (VIS/NIR) dataset, and (**c**) CVC (VIS/LWIR) dataset.

All the three datasets satisfy homography, so the ground truth of each image pair can be computed precisely by manually selected checkpoints. The match was regarded as correct correspondence if the residual error computed by the ground truth was less than 3 pixels. The local feature points are extracted by the FAST extractor [37]. The number of correct feature point matches has a great influence on the matching performance of the descriptors.

Euclidean distance is used as a measure of similarity for feature matching. To preserve as many correct matches as possible, the nearest neighbor distance ratio (NNDR) is used as the matching strategy in the experiments. The NNDR can be expressed as follows.

$$D(a_0, b_0) < \eta \bullet D(a_0, b_1)$$
(13)

where  $(\cdot, \cdot)$  indicates the Euclidean distance, *a* and *b* are the feature vectors in two images, and *b*<sub>1</sub> is the second-closest neighbor to *a*<sub>0</sub> compared with the nearest *b*<sub>0</sub>.  $\eta$  is the threshold of the NNDR.

In the experiment implements, the same region size ( $80 \times 80$ ) with EOH was adopted in our descriptor. The guided image filter parameters were set as r = 7, and  $\varepsilon = 0.3$ . The threshold  $\eta$  was set from 0.8 to 1.0 with intervals of 0.05. The parameters of all competing methods were set as the original papers. For all testing images, the same parameters in the proposed method and other competing methods were adopted.

Three metrics, i.e., precision, recall, and F1-score, are used to validate the matching performance, they are defined as follows.

$$precision = \frac{N_c}{N_c + N_f} \tag{14}$$

$$recall = \frac{N_c}{N_c + \overline{N_c}} \tag{15}$$

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall}$$
(16)

where  $N_c$  indicates the number of correct matches,  $N_f$  is the number of false matches, and  $\overline{N_c}$  indicates the number of discard correct matches.

# 3.2. Qualitative Evaluation on Feature Matching

In the qualitative evaluation experiments, we conduct the feature matching tests on the three datasets by our proposed HOSM descriptor. Figures 7 and 8 show some samples of the multispectral image feature matching results by  $\eta$  is set to 0.8 on three datasets, i.e., Potsdam in Figures 7a and 8a, EPFL in Figures 7b and 8b, and CVC in Figures 7c and 8c, for subjective visual analysis.









(c)

**Figure 7.** Sample one of feature matching results by the HOSM descriptor for the multispectral image pairs in three datasets. From top to bottom: matching results on (**a**) Potsdam dataset, (**b**) EPFL dataset, and (**c**) CVC dataset.



**Figure 8.** Sample two of feature matching results by the HOSM descriptor for the multispectral image pairs in three datasets. From top to bottom: matching results on (**a**) Potsdam dataset, (**b**) EPFL dataset, and (**c**) CVC dataset.

As shown in Figure 7, it can be seen that our proposed HOSM descriptor can get a good matching performance by preserving a large number of correct correspondences between the multispectral image pairs on all three datasets. In details, our method can get the maximum number of correct correspondences in EPFL dataset, followed by the results in Potsdam dataset, and the third best performance in CVC dataset. In Figure 7a,b, the matching point pairs are extracted in most areas, while the matching point pairs in Figure 7c are mainly preserved for the vegetation areas. This difference is caused by a variety of factors. First, FAST detector tends to get feature point responses in complex structures in the image. This results in the most responses being obtained in the vegetation areas, with almost no feature point responses in the grayscale smooth region, such as the wall portion, of the image. Second, due to the larger spectral differences in the CVC dataset, the number of initial matching points in Figure 7c is less than that in Figure 7a,b. The last reason is that when the mismatch points are

removed with the NNDR test, some correctly matched points are also removed. Even so, our method still extracts enough correct matching points to get the correct correspondence between the two images.

Figure 8 presents another group of multispectral image matching results on these three datasets. Figure 8a,b are visible and NIR image pairs selected from the Potsdam and EPFL datasets, and Figure 8c is a pair of visible and LWIR images selected from the CVC dataset. In general, the number of matches in Figure 8a,b is greater than that in Figure 8c. It can also be seen from Figure 8c that there are several pairs of obvious mismatches which are marked by red rectangles. There are several factors that cause mismatches. First, the spectral difference between visible light and LWIR is greater than that between visible light and LWIR is greater than that between visible light and NIR, which leads to more challenging for feature matching. Second, the mismatched points have highly similar textures within the local regions, which results in producing highly similar feature vectors. We can also see that these failed matches have similar correspondence relationships to the correct matches, so they are not removed by the NNDR test. However, enough correct matches between the two images are still preserved.

The experimental results demonstrate that our proposed HOSM descriptor can be robust to the intensity nonlinear variations between the multispectral images and obtain good matching performance.

### 3.3. The Advantages of Guided Filtering

In order to verify the advantages of the guided filtering in structure maps construction, we conduct the comparison tests between the HOSM descriptor and the descriptor without guided filtering on EPFL (VIS/NIR) dataset and CVC (VIS/LWIR) dataset. The qualitative and quantitative comparisons are presented in Figure 9 and Table 1.



**Figure 9.** Comparisons of matching results between the HOSM descriptor and the descriptor without guided filtering. The left column is the matching results of the HOSM descriptor, and the right column is the matching results of the descriptor without guided filtering. From top to bottom: matching results on (a) EPFL (VIS/NIR) dataset and (b) CVC (VIS/LWIR) dataset.

Table 1. Precision, recall and F1-score for HOSM and descriptor without guided filtering.

Datua		HOSM		Descriptor Without Guided Filtering				
Pairs	Precision	Recall	F1-Score	Precision	Recall	F1-Score		
Figure 9a	0.785	0.650	0.711	0.542	0.413	0.468		
Figure 9b	0.312	0.243	0.273	0.054	0.045	0.049		

Figure 9 illustrates some samples of matching results. In general, it can be seen that the HOSM descriptor can achieve better matching performance than the descriptor without guided filtering. The

matching results show that the HOSM descriptor has preserved a larger number of correct matching points on the two multispectral image pairs and has fewer false matches. This is because the guidance images have preserved the significant contour information in multispectral images, the guided filtering operation is able to enhance the structures of the structure maps to improve the description ability [38]. In detail, due to the abundant structures in the visible and NIR images, the descriptor without guided filtering can also achieve satisfactory results on the EPFL dataset, as shown in Figure 9a. When performing matching comparisons between visible and LWIR images, the performance of the HOSM descriptor is significantly better than the descriptor without guided filtering, as shown in Figure 9b. This is because the structure information in the LWIR image is weak, the enhancement of the structure by the guided filtering has a great influence on the matching results. Furtherly, the guided filtering operation can reduce aliasing effects to improve the robustness of the descriptor.

Table 1 presents the quantitative metrics of precision, recall, and F1-score. For all metrics, larger values indicate better results. From these objective metrics, we can see that HOSM outperforms the descriptor without guided filtering. Therefore, the experimental results prove that guided filtering can improve matching performance, especially for images with weak structure information.

#### 3.4. Robustness Evaluations to Noise

To evaluate the robustness of the proposed HOSM method to noise, we make qualitative and quantitative tests on Potsdam (VIS/NIR) dataset at different noise levels. Figure 10 and Table 2 present the quantitative comparisons on the average of precision and recall results, and the F1-scores. In Figure 11, some matching results at different noise levels are selected for qualitative comparisons.



**Figure 10.** Quantitative comparisons on the average of precision and recall curves for Potsdam (VIS/NIR) dataset at different noise levels. The Gaussian white noise with a mean of zero is adopted. A large variance  $\sigma$  indicates great noise intensity.

Table 2. Average of F1-score values with different noise levels.

Metrics						
σ	0	0.01	0.05	0.1	0.15	0.2
F1-score	0.512	0.443	0.397	0.365	0.335	0.305



**Figure 11.** Qualitative comparisons of feature matching on Potsdam (VIS/NIR) dataset at different Gaussian white noise levels.  $\sigma$  is the variance of Gaussian function, and large variance  $\sigma$  indicates great noise intensity.

Due to the sensors and imaging principles, noise in IR images usually appears as stationary random white noise, which means that its position and intensity appearing in the image are randomly distributed [39,40]. In detail, the noise in the image usually appears as additive noise, and Gaussian white noise with a mean of 0 is the most common form [41]. For visible images, the intensity of the noise is small due to the short exposure time. Therefore, in the experiments, we only add zero-mean Gaussian white noise with different intensity in the IR images. The variance  $\sigma$  of the Gaussian function can be used to represent the intensity of noise. In the experiment, we set  $\sigma$  to 0, 0.01, 0.05, 0.1, 0.15 and 0.2 respectively. The average of precision and recall curves are illustrated in Figure 10, and the F1-scores are shown in Table 2. The results show that although the values of precision, recall, and F1-score are declining as the noise increases, our method can still get good results. This is because the guided filtering operation can enhance the structures of the structure maps and reduce aliasing effects to improve the robustness of the descriptor.

For subjective evaluations, some qualitative comparison results are shown in Figure 11. These results intuitively show that our HOSM descriptors can perform well in cases of intensive noise. Even when there are some obvious mismatches at  $\sigma = 0.2$ , our method still obtains a lot of correct matching points. The good matching performances show that our method is efficient and robust to noise.

## 3.5. Quantitative Evaluation on Metrics

Figure 12 illustrates the average values of precision-recall pairs, and Table 3 shows the average values of the F1-score for the eight methods on all the three datasets. For all metrics, including precision, recall, and F1-score, larger numbers indicate better performance. It can be seen that the HOSM descriptor outperforms the other methods in the metrics on all the datasets. In general, the

matching metrics of all the methods get the best values on EPFL dataset, the second-best values on Potsdam dataset, and the last results on CVC dataset. Compared to NIR images in the Potsdam and EPFL datasets, the LWIR images in the CVC dataset have lower resolution, while having greater nonlinear intensity differences with corresponding visible images. Therefore, it is most challenging for the CVC dataset on all descriptors to extract the common structure features from multispectral images.



**Figure 12.** Quantitative comparisons on precision-recall averages with NNDR thresholds from 0.8 to 1.0 in three datasets. (**a**) Potsdam dataset, (**b**) EPFL dataset, and (**c**) CVC dataset.

Detecto	n	Methods									
Datasets	4	SIFT	SURF	NG-SIFT	EOH	LGHD	MFD	HoDM	HOSM		
	0.8	0.258	0.171	0.283	0.275	0.396	0.417	0.452	0.471		
	0.85	0.265	0.179	0.295	0.292	0.412	0.422	0.454	0.480		
Potsdam	0.9	0.257	0.179	0.294	0.309	0.408	0.424	0.466	0.484		
	0.95	0.261	0.175	0.284	0.315	0.412	0.418	0.462	0.479		
	1.0	0.244	0.174	0.274	0.323	0.419	0.414	0.443	0.482		
	0.8	0.671	0.611	0.668	0.471	0.705	0.727	0.736	0.761		
	0.85	0.687	0.623	0.686	0.508	0.712	0.731	0.741	0.767		
EPFL	0.9	0.691	0.637	0.699	0.537	0.723	0.734	0.745	0.773		
	0.95	0.698	0.646	0.714	0.557	0.731	0.742	0.751	0.778		
	1.0	0.709	0.658	0.717	0.563	0.735	0.746	0.762	0.780		
	0.8	0.06	0.032	0.082	0.101	0.119	0.129	0.147	0.151		
	0.85	0.078	0.055	0.082	0.104	0.126	0.135	0.152	0.166		
CVC	0.9	0.075	0.058	0.082	0.107	0.126	0.134	0.156	0.174		
	0.95	0.077	0.067	0.082	0.107	0.126	0.137	0.159	0.184		
	1.0	0.08	0.074	0.080	0.112	0.132	0.144	0.149	0.191		

Table 3. Average of F1-score values for all eight methods on all three datasets.

As shown in Figure 12a–c, the HOSM descriptor achieves the best performance in all precision-recall curves, followed by HoDM. Due to the oriented structure maps preserving the common features and structure guidance images overcoming nonlinear variations in the gradient direction between the multispectral images, the HOSM descriptor can be robust to get high precision and recall values in these datasets. The HoDM can get good matching performance due to the combination of structure and texture information. The MFD and LGHD have similar performance because both descriptors are progressive methods based on the EOH, in which the multi-scale and multi-oriented Log-Gabor filters were used to replace multi-directional spatial filters. Because the Log-Gabor filters are better at retaining the oriented edge characteristics of multispectral images, the MFD and LGHD methods can achieve better matching performance than the original EOH. It can be seen from the Figure 12a,c that the EOH is superior to SIFT, SURF and NG-SIFT, but the results are opposite in Figure 12b. This is because in the EPFL dataset, the spectra of the image pairs are close, SIFT, SURF, and NG-SIFT can obtain stable descriptors to get better matching results than EOH. However, EOH is more robust to image pairs with significant nonlinear intensity variations, as shown in Figure 12a,c.

From the average of F1-score values in Table 3, where the best scores are marked in bold, we can more accurately compare the matching performance of all methods. Table 3 indicates that the HOSM method can obtain the highest average scores in all cases, followed by HoDM. LGHD can get good performance similar to MFD on all datasets, while EOH can achieve third best scores on Potsdam and CVC datasets. For EPFL dataset, these three SIFT-like methods are able to obtain better performance than EOH. The high F1-score values demonstrate that the HOSM descriptor not only achieves good matching performance, but also achieve a good balance between precision and recall.

### 3.6. Quantitative Comparisons on Running Time

A desktop with 4 GB memory and 2.5 GHz Intel Core i3 CPU is used to carry out the experiments and all these methods are implemented by the MATLAB codes. Figure 13 shows the average computation time of the methods in each feature point on all the three datasets, in details, Figure 13a is the average running time of Potsdam dataset, Figure 13b is the average running time of EPFL dataset, and Figure 13c is the average running time of CVC dataset. It can be seen that the HOSM is faster than other methods and achieves the best time consumption. This is because we have previously established the structure maps with linear guided filtering operation. The values of structure maps are directly used to build the HOSM descriptor, without other preprocessing. From all experimental results, it can be seen that our HOSM descriptor is robust to the nonlinear intensity variations in multispectral images and is superior to other methods.









**Figure 13.** Quantitative comparisons on average computation time for all the descriptors. (**a**) Potsdam dataset, (**b**) EPFL dataset, and (**c**) CVC dataset.

### 4. Conclusions

In this paper, we propose a new local feature descriptor termed HOSM for multispectral remote sensing image matching task. First, we propose a new local contrast based operator to construct the structure guidance images for keeping the significant contour features in multispectral images. Then, construct the oriented structure maps based on multi-oriented Sobel edges with guided filtering. Guided image filtering can achieve better computational efficiency than traditional time-consuming trilinear interpolation methods. Finally, the histograms of oriented structure maps are computed to build the feature vectors. In order to verify the HOSM descriptor, three widely used datasets named Potsdam, EPFL, and CVC are employed to conduct feature matching tests, and seven state of the art methods are used for comparison. The experimental results show that our method is outstanding to address the nonlinear intensity changes in multispectral images and can achieve better matching performance.

In the future, we will further study how to enhance the description ability of our method, because the HOSM descriptor is robust to nonlinear intensity changes but sensitive to rotation and scale invariance. We can assign a main orientation to the HOSM to cope with the rotation variance and use a multiscale keypoint detector to keep scale invariance.

Author Contributions: Writing—original draft, T.M.; Writing—review & editing, J.M; Software, K.Y.

Funding: This research received no external funding.

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

## References

- Zhou, W.; Wang, C.; Xiao, B.; Zhang, Z. SLD: A Novel Robust Descriptor for Image Matching. *IEEE Signal Process. Lett.* 2014, 21, 339–342. [CrossRef]
- Ma, J.; Ma, Y.; Zhao, J.; Tian, J. Image Feature Matching via Progressive Vector Field Consensus. *IEEE Signal Process. Lett.* 2015, 22, 767–771. [CrossRef]
- 3. Zhu, B.; Xin, Y. Effective and robust infrared small target detection with the fusion of poly directional first order derivative images under facet model. *Infrared Phys. Technol.* **2015**, *69*, 136–144. [CrossRef]
- 4. Chen, Y.; Ogata, T.; Ueyama, T.; Takada, T.; Ota, J. Automated Field-of-View, Illumination, and Recognition Algorithm Design of a Vision System for Pick-and-Place Considering Colour Information in Illumination and Images. *Sensors* **2018**, *18*, 1656. [CrossRef] [PubMed]
- Shi, Q.; Ma, G.; Zhang, F.; Chen, W.; Qin, Q.; Duo, H. Robust Image Registration Using Structure Features. IEEE Geosci. Remote Sens. Lett. 2014, 11, 2045–2049.
- 6. Guislain, M.; Digne, J.; Chaine, R.; Monnier, G. Fine scale image registration in large-scale urban LIDAR point sets. *Comput. Vis. Image Underst.* **2017**, *157*, 90–102. [CrossRef]
- Song, J.; Liu, L.; Huang, W.; Li, Y.; Chen, X.; Zhang, Z. Target detection via HSV color model and edge gradient information in infrared and visible image sequences under complicated background. *Opt. Quantum Electron.* 2018, *50*, 171–175. [CrossRef]
- 8. Li, Y.; Tao, C.; Tan, Y.; Shang, K.; Tian, J. Unsupervised Multilayer Feature Learning for Satellite Image Scene Classification. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 157–161. [CrossRef]
- 9. Jung, J.; Sohn, G.; Bang, K.; Wichmann, A.; Armenakis, C.; Kada, M. Matching Aerial Images to 3D Building Models Using Context-Based Geometric Hashing. *Sensors* **2016**, *16*, 932. [CrossRef]
- Zhao, C.; Sun, L.; Purkait, P.; Duckett, T.; Stolkin, R. Dense RGB-D Semantic Mapping with Pixel-Voxel Neural Network. *Sensors* 2018, 18, 3099. [CrossRef] [PubMed]
- 11. Ma, T.; Ma, J.; Fang, B.; Hu, F.; Quan, S.; Du, H. Multi-scale decomposition based fusion of infrared and visible image via total variation and saliency analysis. *Infrared Phys. Technol.* **2018**, *92*, 154–162. [CrossRef]
- 12. Feng, Q.; Hao, Q.; Chen, Y.; Yi, Y.; Wei, Y.; Dai, J. Hybrid Histogram Descriptor: A Fusion Feature Representation for Image Retrieval. *Sensors* **2018**, *18*, 1943. [CrossRef] [PubMed]
- Nguyen, D.; Baek, N.; Pham, T.; Park, K. Presentation Attack Detection for Iris Recognition System Using NIR Camera Sensor. *Sensors* 2018, 18, 1315. [CrossRef]

- 14. So, R.W.K.; Chung, A.C.S. A novel learning-based dissimilarity metric for rigid and non-rigid medical image registration by using Bhattacharyya Distances. *Pattern Recognit.* **2017**, *62*, 161–174. [CrossRef]
- 15. Ye, Y.; Shan, J.; Bruzzone, L.; Shen, L. Robust registration of multimodal remote sensing images based on structural similarity. *Ieee Trans. Geosci. Remote Sens.* **2017**, *55*, 2941–2958. [CrossRef]
- 16. Liu, Y.; Mo, F.; Tao, P. Matching Multi-Source Optical Satellite Imagery Exploiting a Multi-Stage Approach. *Remote Sens.* **2017**, *9*, 1249. [CrossRef]
- 17. He, C.; Fang, P.; Xiong, D.; Wang, W.; Liao, M. A Point Pattern Chamfer Registration of Optical and SAR Images Based on Mesh Grids. *Remote Sens.* **2018**, *10*, 1837. [CrossRef]
- Ma, J.; Zhao, J.; Tian, J.; Bai, X.; Tu, Z. Regularized vector field learning with sparse approximation for mismatch removal. *Pattern Recognit.* 2013, 46, 3519–3532. [CrossRef]
- 19. Yu, M.; Deng, K.; Yang, H.; Qin, C. Improved WαSH Feature Matching Based on 2D-DWT for Stereo Remote Sensing Images. *Sensors* **2018**, *18*, 3494. [CrossRef]
- 20. Wang, G.; Wang, X.; Fan, B.; Pan, C. Feature Extraction by Rotation-Invariant Matrix Representation for Object Detection in Aerial Image. *Ieee Geosci. Remote Sens. Lett.* **2017**, *14*, 851–855. [CrossRef]
- 21. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
- 22. Bay, H.; Tuytelaars, T.; Van Gool, L. SURF: Speeded Up Robust Features. In Proceedings of the European Conference on Computer Vision (ECCV), Graz, Austria, 7–13 May 2006; pp. 404–417.
- Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* 2008, 110, 346–359. [CrossRef]
- 24. Saleem, S.; Sablatnig, R. A Robust SIFT Descriptor for Multispectral Images. *IEEE Signal Process. Lett.* **2014**, 21, 400–403. [CrossRef]
- 25. Wang, G.; Wang, Z.; Chen, Y.; Zhao, W. Robust point matching method for multimodal retinal image registration. *Biomed. Signal Process. Control.* **2015**, *19*, 68–76. [CrossRef]
- 26. Manjunath, B.S.; Ohm, J.; Vasudevan, V.V.; Yamada, A. Color and texture descriptors. *IEEE Trans. Circuits Syst. Video Technol.* **2001**, *11*, 703–715. [CrossRef]
- 27. Aguilera, C.; Barrera, F.; Lumbreras, F.; Sappa, A.D.; Toledo, R. Multispectral Image Feature Points. *Sensors* 2012, 12, 12661–12672. [CrossRef]
- Fu, Z.; Qin, Q.; Luo, B.; Wu, C.; Sun, H. A Local Feature Descriptor Based on Combination of Structure and Texture Information for Multispectral Image Matching. *IEEE Geosci. Remote Sens. Lett.* 2019, 16, 100–104. [CrossRef]
- 29. Aguilera, C.A.; Sappa, A.D.; Toledo, R. LGHD: A feature descriptor for matching across non-linear intensity variations. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 178–181.
- 30. Nunes, C.F.G.; Pádua, F.L.C. A Local Feature Descriptor Based on Log-Gabor Filters for Keypoint Matching in Multispectral Images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1850–1854. [CrossRef]
- 31. Sikora, T. The MPEG-7 visual standard for content description—An overview. *IEEE Transations Circuits Syst. Video Technol.* **2001**, *11*, 696–702. [CrossRef]
- 32. He, K.; Sun, J.; Tang, X. Guided Image Filtering. In Proceedings of the European Conference on Computer Vision (ECCV), Heraklion, Crete, Greece, 5–11 September 2010; pp. 1–14.
- He, K.; Sun, J.; Tang, X. Guided Image Filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* 2013, 35, 1397–1409. [CrossRef]
- 34. Zhao, H.; Jiang, L.; Jin, X.; Hui, D.; Li, X. Constant time texture filtering. *Vis. Comput.* **2018**, *34*, 83–92. [CrossRef]
- 35. Potsdam Dataset of Remote Sensing Images, Distributed by the International Society for Photogrammetry and Remote Sensing. Available online: http://www2.isprs.org/commissions/comm3/wg4/2dsem-label-potsdam. html (accessed on 8 September 2018).
- Brown, M.; Süsstrunk, S. Multi-spectral SIFT for scene category recognition. In Proceedings of the 2011 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Fort Collins, CO, USA, 20–25 June 2011; pp. 177–184.
- 37. Rosten, E.; Porter, R.; Drummond, T. Faster and Better: A Machine Learning Approach to Corner Detection. *Ieee Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 105–119. [CrossRef] [PubMed]

- 38. Choi, J.; Park, H.; Seo, D. Pansharpening Using Guided Filtering to Improve the Spatial Clarity of VHR Satellite Imagery. *Remote Sens.* **2019**, *11*, 633. [CrossRef]
- 39. Kim, Y.S.; Lee, J.H.; Ra, J.B. Multi-sensor image registration based on intensity and edge orientation information. *Pattern Recognit.* **2008**, *41*, 3356–3365. [CrossRef]
- 40. Liu, X.; Ai, Y.; Zhang, J.; Wang, Z. A Novel Affine and Contrast Invariant Descriptor for Infrared and Visible Image Registration. *Remote Sens.* **2018**, *10*, 658. [CrossRef]
- 41. Keller, Y.; Averbuch, A. Multisensor image registration via implicit similarity. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 794–801. [CrossRef] [PubMed]



© 2019 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 (http://creativecommons.org/licenses/by/4.0/).