HOMPC: A Local Feature Descriptor Based on the Combination of Magnitude and Phase Congruency Information for Multi-Sensor Remote Sensing Images

Zhitao Fu [1] [iD], Qianqing Qin [1], Bin Luo [1,*], Hong Sun [2] and Chun Wu [1]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; zhitaofu@126.com (Z.F.); qqqin@whu.edu.cn (Q.Q.); cugwhu@126.com (C.W.)

[2] Signal Processing Laboratory, School of Electronic Information, Wuhan University, Wuhan 430072, China; hongsun@whu.edu.cn

[*] Correspondence: luob@whu.edu.cn

check for
updates

Abstract: Local region description of multi-sensor images remains a challenging task in remote sensing image analysis and applications due to the non-linear radiation variations between images. This paper presents a novel descriptor based on the combination of the magnitude and phase congruency information of local regions to capture the common features of images with non-linear radiation changes. We first propose oriented phase congruency maps (PCMs) and oriented magnitude binary maps (MBMs) using the multi-oriented phase congruency and magnitude information of log-Gabor filters. The two feature vectors are then quickly constructed based on the convolved PCMs and MBMs. Finally, a dense descriptor named the histograms of oriented magnitude and phase congruency (HOMPC) is developed by combining the histograms of oriented phase congruency (HPC) and the histograms of oriented magnitude (HOM) to capture the structure and shape properties of local regions. HOMPC was evaluated with three datasets composed of multi-sensor remote sensing images obtained from unmanned ground vehicle, unmanned aerial vehicle, and satellite platforms. The descriptor performance was evaluated by recall, precision, F1-measure, and area under the precision-recall curve. The experimental results showed the advantages of the HOM and HPC combination and confirmed that HOMPC is far superior to the current state-of-the-art local feature descriptors.

Keywords: multi-sensor images; log-Gabor filters; non-linear radiation variations; local feature descriptor; phase congruency and magnitude

1. Introduction

With the rapid development of sensor technology and modern communications, we are now entering a multi-sensor era. Different sensors capture different features, which are useful for a variety of applications, including multi-sensor image registration and fusion [1–3] and pedestrian detection [4–6]. However, the non-linear radiation/intensity variations between multi-sensor images result in the local feature description being a challenging task [7–10].

The traditional approaches based on histograms of oriented gradient descriptors such as scale-invariant feature transform (SIFT) [11] and speeded-up robust features (SURF) [12] perform well on single-sensor images, but generate only a few correct mappings when dealing with multi-sensor images. To address this issue, researchers have proposed many techniques to adapt descriptors based on SIFT/SURF to multi-sensor images. Approaches such as the partial intensity invariant feature

descriptor (PIIFD) [13], R-SIFT [14] orientation-restricted SIFT (OR-SIFT) [15], and multimodal SURF (MM-SURF) [16] use gradient orientation modification to limit the gradient orientation to (0, pi) on the basis of the intensity reversal in certain areas. Saleem et al. [17] proposed NG-SIFT, which employs a normalized gradient to construct the feature vectors, and it was found that NG-SIFT outperformed SIFT on visible and near-infrared images.

Even though these descriptors perform slightly better than the traditional descriptors, the number of mismatches increases due to the orientation reversal, and the total number of matched points is still low. This is because the description ability of these descriptors relies on a linear relationship between images, and they are not appropriate for the significant non-linear intensity differences caused by the radiometric variations between multi-sensor images.

Some descriptors have been designed based on the distribution of edge points, which can be regarded as the common features of multi-sensor images. Aguilera et al. [18] proposed the edge-oriented histogram (EOH) descriptor for multispectral images. Li et al. [19] assigned the main orientation computed with PIIFD to EOH for increased robustness to rotational invariance. Zhao et al. [20] used edge lines for a better matching precision. Shi et al. [21] combined shape context with the DAISY descriptor in a structural descriptor for multispectral remote sensing image registration; however, all the edge points are constrained by contrast and threshold values [22]. Other descriptors have been proposed, based on local self-similarity (LSS) and its extension for multispectral remote sensing images [5–7,23], but the size of the LSS region contributes much to the descriptor's capability. Furthermore, LSS and its extensions usually result in a low number of correctly matched points.

Due to the differences in multi-sensor imaging principles, the intensities among multi-sensor images present non-linear radiation variations, resulting in the above descriptors relying on gradient-based linear intensity variations, i.e., spatial domain information, not performing well for multi-sensor images. In addition to the spatial domain, images can be decomposed into amplitude and phase information by Fourier transform in the frequency domain. The ability of a descriptor can be evaluated by "repeatability" and "distinctiveness", and a trade-off is often made between these measures [19]. To obtain as much useful information as possible from the images is the goal of a descriptor. More information can be obtained by convolving the images using multi-scale and multi-oriented Gabor-based filters, including the Gabor filter and log-Gabor filter.

The Gabor filter responses are invariant to illumination variations [24,25], and the multi-oriented magnitudes transmit useful shape information [26,27]. Since the log-Gabor filters basically consist of a logarithmic transformation of the Gabor domain, researchers have proposed descriptors based on the amplitudes of log-Gabor coefficients. The phase congruency and edge-oriented histogram descriptor (PCEHD) [28] combines spatial information (EOH) and frequency information (the amplitude of the log-Gabor coefficients). The log-Gabor histogram descriptor (LGHD) [29] uses multi-scale and multi-oriented log-Gabor filters instead of the multi-oriented Sobel descriptor, and it divides the region around the point of interest into sub-regions (similar to SIFT with 4 × 4). LGHD has been used to match images with non-linear intensity variations, including visible and thermal infrared images, and has outperformed SIFT, GISIFT [30], EOH [18], and PCEHD. However, LGHD is time-consuming due to its multi-scale computation. Cristiano et al. [31] proposed the multispectral feature descriptor (MFD), which computes the descriptors using fewer log-Gabor filters and, as a result, has a computational efficiency that is better than that of LGHD.

Oppenheim et al. [32] analyzed the function of the phase information, and discovered that phase information is even more important than amplitude information for the preservation of image features. The phase congruency detector is a feature detector based on the local phase information of images. Kovesi et al. [33] proposed a measure of phase congruency that is independent of the overall magnitude of the signal, making it invariant to variations in image illumination and/or contrast. Furthermore, phase congruency is a dimensionless quantity. Therefore, a number of researchers have proposed methods for feature description based on phase congruency, for applications such as template matching for multi-sensor remote sensing images [34,35] and pedestrian detection [36,37].

The distribution of the high-frequency components of log-Gabor filters has been verified to be robust to non-linear intensity variations [29,31], and phase congruency has been successfully used for multispectral image template matching [34–37]. This makes us think that a descriptor combining phase congruency and the distribution of the high-frequency components would be more efficient to capture the common information of multi-sensor images, which is the idea behind the proposed approach. Figure 1 shows a comparison between phase congruency and the distribution of the high-frequency components for a pair of visible and infrared remote sensing images. Vertically, the phase congruency and the distribution of the high-frequency components are similar for the visible and infrared images, even with the non-linear intensity changes existing between the visible and infrared images. Horizontally, the phase congruency is more like the enhanced edges of an image, and the distribution of the high-frequency components is more like the coarse texture of the image.
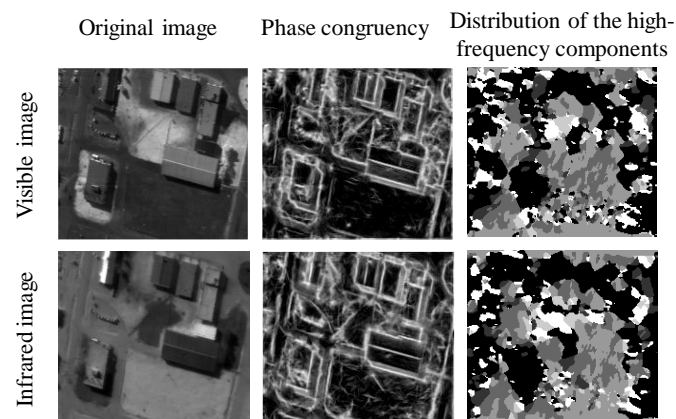

Figure 1. Comparison between phase congruency and the distribution of the high-frequency components.

In this paper, a novel descriptor named the histograms of oriented magnitude and phase congruency (HOMPC) combining the histograms of oriented magnitude (HOM) and the histograms of oriented phase congruency (HPC) is proposed. The HOM and HPC can be efficiently calculated in a dense manner over the local regions of the images through a convolution operation. The HOMPC descriptor reflects the structural and shape properties of local regions, which are relatively independent of the particular intensity distribution pattern across two local regions. To the best of our knowledge, we are the first to combine magnitude and phase congruency information to construct a local feature descriptor to capture the common information of multi-sensor images with non-linear radiation variations.

The main contributions of this paper are as follows:

- We propose oriented phase congruency maps (PCMs) and oriented magnitude binary maps (MBMs) based on log-Gabor filters;
- we have designed a fast method to construct feature vectors through a convolution operation based on the PCMs and MBMs;
- we have developed a novel local feature descriptor based on the magnitude and phase congruency information to capture more common structural and shape properties.

The rest of this paper is organized as follows: Section 2 proposes the HOMPC descriptor based on the HOM and HPC. Section 3 introduces the experimental setup. Section 4 analyzes the parameter sensitivity and the advantages of combining the HOM and HPC and describes the comparison of the proposed HOMPC descriptor with the current state-of-art descriptors. Section 5 presents the conclusions and recommendations for future work.

2. Methodology

The proposed approach is based on the fact that multi-sensor images share similar global appearances for the shape of the objects contained in the scenes, despite having different intensities and textures. The phase congruency and the distribution of the high-frequency components have both been proved suitable to capture the common features of multispectral images [29,31,35]. In this section, a novel local feature descriptor based on the phase and magnitude information is proposed to capture the common feature of multi-sensor images. The PCMs and MBMs are first constructed based on log-Gabor filters, and then the two feature vectors are quickly constructed based on the convolved PCMs and MBMs. Finally, the HOMPC descriptor is developed by combining the HPC and HOM using the structural properties. The details of the proposed method are presented in the following sections.

2.1. The Magnitudes Based on the Log-Gabor Filter

A 2-D log-Gabor filter [38,39] can be obtained using a Gaussian function in the angular direction of the log-Gabor filter. Consequently, the 2-D log-Gabor function is defined as follows:

$$LG_{s,o}(\omega,\theta) = exp\left(\frac{-(\log(\omega/\omega_s))^2}{2(\log(k/\omega_s))^2}\right) exp\left(\frac{-(\theta-\theta_{s,o})}{2\sigma_\theta{}^2}\right), \tag{1}$$

where (ω, θ) represents the polar coordinates of image $I(x, y)$; s and o are the scale and orientation of the 2-D log-Gabor filters; $(\omega_s, \theta_{s,o})$ are the center frequency and center orientation, respectively, for the frequency of the 2-D log-Gabor filters; k/ω_s is kept constant for various values of ω_s; and σ_θ the standard deviation of the Gaussian function in the angular direction.

As a frequency domain filter, the log-Gabor filter can be presented in the spatial domain by inverse Fourier transform. In the spatial domain, the 2-D log-Gabor filter can be represented as:

$$LG_{s,o}(x,y) = LG_{s,o}^{even}(x,y) + i \times LG_{s,o}^{odd}(x,y), \tag{2}$$

where the real part $LG_{s,o}^{even}(x,y)$ and the imaginary part $LG_{s,o}^{odd}(x,y)$ mean the even-symmetric and the odd-symmetric log-Gabor wavelets at scale s and orientation o, respectively. The symbol i is an imaginary unit.

The response vector at scale s and orientation o is obtained by convolution of each quadrature pair with the input signal image $I(x,y)$, and is given by:

$$[E_{s,o}(x,y), O_{s,o}(x,y)] = \left[I(x,y) * LG_{s,o}^{even}(x,y), I(x,y) * LG_{s,o}^{odd}(x,y)\right], \tag{3}$$

where the symbol "$*$" indicates the convolution in the equation.

The amplitude of the response $A_{s,o}(x,y)$ of image $I(x,y)$ at scale s and orientation o are then given by:

$$A_{s,o}(x,y) = \sqrt{E_{s,o}(x,y)^2 + O_{s,o}(x,y)^2}. \tag{4}$$

Hereafter, we define the amplitude of the response as the "magnitude".

2.2. The Proposed Oriented Phase Congruency Maps (PCMs)

As a result of its invariance to illumination and contrast variation [33], phase congruency has been used for local feature description [20,40] and template matching [34,35] for non-linear intensity changes. Most of these methods use both the magnitudes of all the scales and the magnitudes of all the orientations to calculate the phase congruency values. Thus, when computing the histograms of local regions, Gaussian convolution and trilinear interpolation methods are needed, which increases the computation time. In order to accelerate the computational efficiency of feature description based

on phase congruency information, we propose the use of PCMs. We calculate the phase congruency of each orientation through the magnitudes of all the scales corresponding to the orientation.

The phase congruency of each orientation is defined as follows:

$$\text{PC}_o(x,y) = \frac{\sum_s W(x,y)\langle A_{s,o}(x,y)\Delta\varnothing_{s,o}(x,y) - T\rangle}{\sum_s A_{s,o}(x,y) + \varepsilon}, \tag{5}$$

where $o = 1, 2, \ldots, N_o$ and $s = 1, 2, \ldots, N_s$ stands for the number of orientations and scales; PC_o is the PCM of orientation o; $\{\text{PC}(x,y)\}_1^{N_o}$ represents all the orientation phase congruency values of point (x,y); T is the estimated noise influence; and ε is a small constant to avoid division by zero. The symbol $\langle\ \rangle$ denotes that the enclosed quantity is equal to itself when its value is positive and is zero otherwise. This means that only energy values that exceed the noise level T are taken into account in the results. $W(x,y)$ is a weighting function that weights the frequency spread. It devalues the phase congruency values at locations where there is a narrow spread of filter response. The weighting function [41] is expressed as:

$$W(x,y) = \frac{1}{1 + e^{\gamma(c - \rho(x,y))}}, \tag{6}$$

where γ is a gain factor controlling the sharpness of the filter; c is the cut-off value below which the phase congruency values are penalized; and $\rho(x,y)$ defines the spread of filter responses. $\Delta\varnothing_{s,o}(x,y)$ is a phase deviation function, whose definition is:

$$\Delta\varnothing_{s,o}(x,y) = (E_{s,o}(x,y)\overline{\varnothing}_{even}(x,y) + O_{s,o}(x,y)\overline{\varnothing}_{odd}(x,y)) - |E_{s,o}(x,y)\overline{\varnothing}_{odd}(x,y) - O_{s,o}(x,y)\overline{\varnothing}_{even}(x,y)| \tag{7}$$

where:

$$(\overline{\varnothing}_{even}(x,y), \overline{\varnothing}_{odd}(x,y)) = \frac{1}{\sqrt{(F(x,y))^2 + (H(x,y))^2}}(F(x,y), H(x,y)),$$

where:

$$F(x,y) = \sum_s E_{s,o}(x,y), \text{ and } H(x,y) = \sum_s O_{s,o}(x,y).$$

To obtain the PCMs, we normalize the phase congruency of all the orientations $\{\text{PC}(x,y)\}_1^{N_o}$ to [36,37] as follows:

$$\text{PCM}_o(x,y) = \Psi(PC_o(x,y)) \times 255, (o = 1, 2, \ldots, N_o), \tag{8}$$

where $\text{PCM}_o(x,y)$ is the *PCM* of orientation o; the operator $\Psi(h)$ normalizes the value of h to [0–1]. Figure 2 shows the six-oriented PCMs based on the log-Gabor filters.
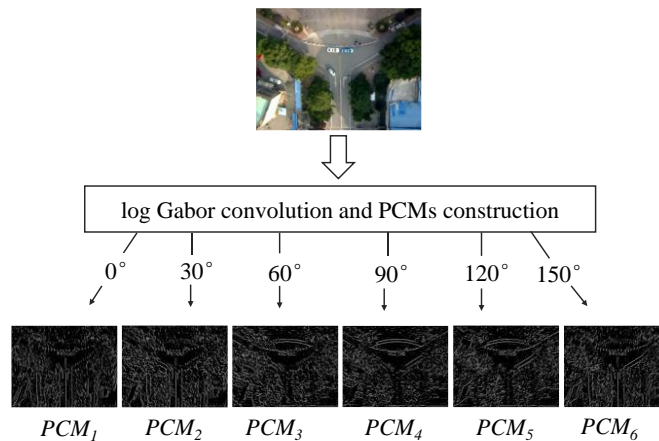

Figure 2. The calculated six-oriented PCMs based on the four-scale and six-oriented log-Gabor filters.

2.3. The Proposed Oriented Magnitude Binary Maps (MBMs)

Aguilera et al. [29] proved that the distribution of the high-frequency components is robust to non-linear intensity variations. The LGHD descriptor calculates the distribution histograms of each sub-region's 4×4 local patch for all scales. Therefore, the computational efficiency of LGHD is very low.

Differing from LGHD, we use the average magnitudes of the different scales to calculate the distribution. In this way, the average magnitudes are more robust to noise. For each orientation *o*, the average magnitudes of all the scales are calculated as:

$$A_o(x, y) = \left(\sum_{s=1}^{N_s} A_{s,o}(x, y) \right) / N_s, \tag{9}$$

where $o = 1, 2, \ldots, N_o$ and $s = 1, 2, \ldots, N_s$ stand for the number of orientations and scales, respectively. A_o is defined as the oriented magnitude map (*MM*) of orientation *o*. We use $\{A(x, y)\}_1^{N_o}$ to represent all the orientation magnitudes of point (x, y). To obtain the *MMs*, we normalize the magnitudes of all the orientations $\{A(x, y)\}_1^{N_o}$ to [0–255] as follows:

$$MM_o(x, y) = \Psi(A_o(x, y)) \times 255, (o = 1, 2, \ldots, N_o), \tag{10}$$

where $MM_o(x, y)$ is the *MM* of orientation *o*; the operator $\Psi(h)$ normalizes the value of *h* to [0–1]. Figure 3a shows the six-oriented *MMs* based on log-Gabor filters.

In order to accelerate the computational efficiency of the feature description based on *MMs*, we propose the *MBMs*, which are consistent with the *PCMs*.

All the orientation magnitudes are compared through each pixel to find the maximum value, and the value of the pixel in the orientation is assigned to 1 if $MM_o(x, y)$ is equal to the maximum of $\{MM(x, y)\}_1^{N_o}$; otherwise, it is assigned to 0. We express the calculation procedure using the following formulation:

$$MBM_o(x, y) = \begin{cases} 1, & if \, MM_o(x, y) == \max(\{MM(x, y)\}_1^{N_o}) \\ 0, & else \end{cases}, \tag{11}$$

where $MBM_o(x, y)$ is the *MBM* of orientation *o*. Figure 3b shows the six-oriented *MBMs* based on *MMs*.
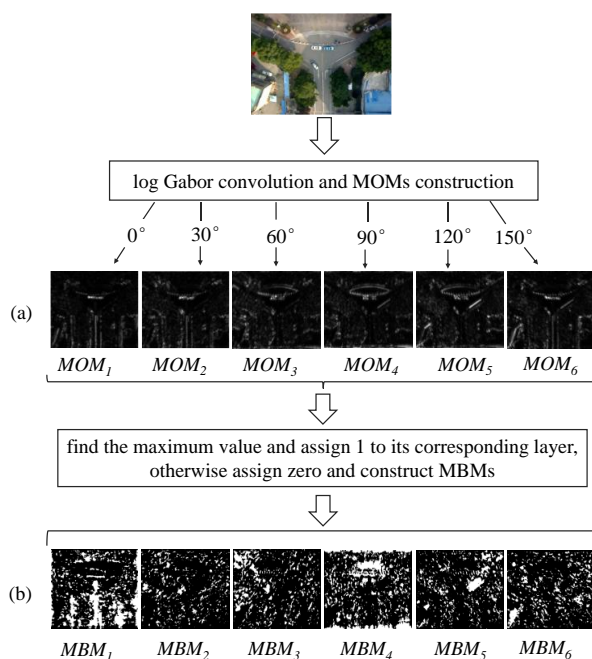

Figure 3. The six-oriented *MMs* (**a**) and the six-oriented *MBMs* (**b**) based on the log-Gabor filters.

2.4. The Proposed HOMPC Descriptor Based on the PCMs and MBMs

In this subsection, the proposed HOMPC descriptor based on the PCMs and MBMs is described. The proposed histograms of oriented phase congruency (HPC) and histograms of oriented magnitude (HOM) are inspired by the HOG method [42] and the methods of Ye et al. [34,35], all of which calculate the descriptors based on a dense grid of local histograms. To accelerate the computational efficiency of the proposed HOMPC descriptor, Gaussian filters and mean filters are introduced to convolve the oriented PCMs and MBMs, instead of the trilinear interpolation method used in the HOG method and the methods of Ye et al. Following this, the HPC and HOM are calculated for a local region. We combine the HPC and HOM to obtain the proposed HOMPC descriptor. Figure 4 presents the main processing chain of the proposed HOMPC descriptor. The detailed steps of this process are as follows:

1. The first step is to apply the four-scale and six-oriented log-Gabor filters to the local region Rx of each point of interest, and then compute the six-oriented PCMs and MBMs (Figures 2 and 3), which provide the common feature information against non-linear intensity changes.

2. The second step is to apply Gaussian filters to the six-oriented PCMs to obtain the convolved PCMs and mean filters to the six-oriented MBMs to obtain the convolved MBMs, based on the same template size. The template size is equal to the patch size of a cell, and four cells combine to form a block. The relationship between blocks and cells is shown in Figure 5.

3. The third step is to divide the local region of each point into *p* overlapping blocks based on the number of interval, and then calculate the feature vector of each block. If the size of local region is m × m, the cell size is $n \times n$, and the number of interval is k, the number of overlapping blocks is the square of the integer part $((m - 2 \times n)/k + 1)$.

   - To calculate the feature vector of a block based on the convolved 6-oriented PCMs. The convoluted values of the six-oriented PCMs in each pixel contribute to the orientation bins of the pixel. We combine the histograms of the four pixels corresponding to the center locations of the four connected cells (Figure 5), which combine to form a block. Figure 6 shows the process of constructing the feature vector for the block based on PCMs. The dimension of the feature vector is 24 bins. We normalize the feature vector by the L2 norm to achieve a better invariance to illumination and shadowing. The character X is used to represent the normalized feature vector based on the PCMs.

   - To calculate the feature vector of the block based on the convolved six-oriented MBMs. The convoluted values of the six-oriented MBMs in each pixel contribute to the orientation bins of the pixel. The process of constructing the feature vector for a block based on the MBMs is similar to the process of constructing the feature vector for the block based on the PCMs. The convolved six-oriented MBMs are used to replace the convolved six-oriented PCMs (Figure 6), and the remaining operation is the same. The dimension of the feature vector is also 24 bins. We also normalize the feature vector by the L2 norm. The character Y is used to represent the normalized feature vector based on the MBMs.

4. We collect all the feature vectors based on the PCMs and MBMs of all the blocks within the dense overlapping grid covering the local feature region into combined feature vectors, named the HPC and HOM, respectively.

5. Finally, we combine the HPC and HOM to obtain the proposed HOMPC descriptor, which combines the phase congruency and magnitude information for the local feature description.
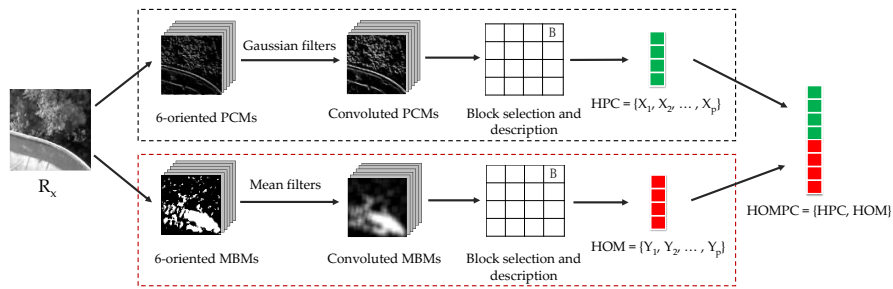
Figure 4. The main processing chain of the HOMPC descriptor. The black dotted line box and red dotted line boxes respectively indicate the construction process of the histograms of oriented phase consistency (HPC) and the histograms of magnitudes (HOM). The HOMPC is made up of them.
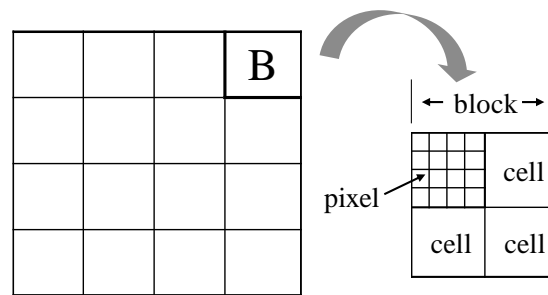

Figure 5. The relationship between blocks and cells. The symbol B represents a block, which contains four cells.
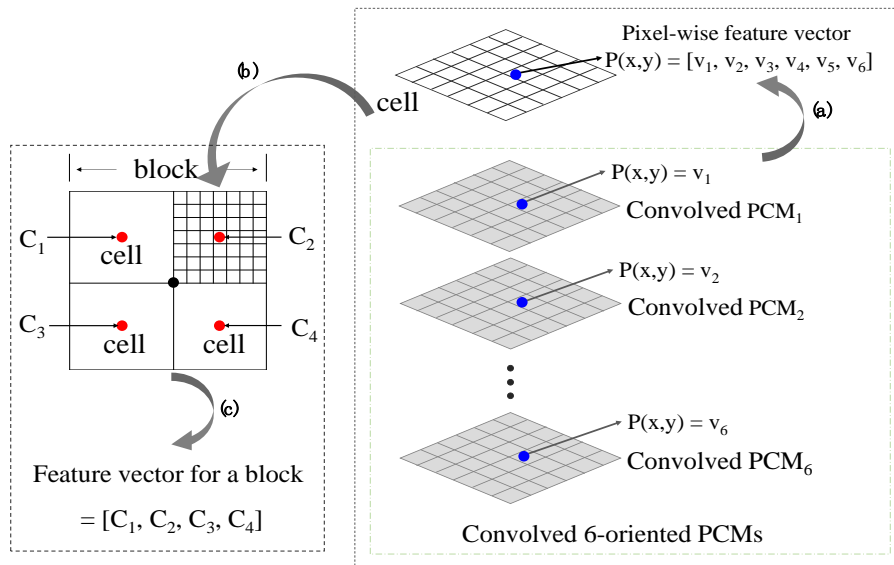

Figure 6. The process of constructing the feature vector for a block. (**a**) Compute the pixel-wise feature vector using the convolved PCMs. P(x, y) means the feature vector of the pixel in location P(x, y), and the vector size is 1×6; (**b**) obtain the central pixel's feature vector of each cell. C_1, C_2, C_3, and C_4 are the corresponding centric pixel's feature vectors of the four cells; (**c**) Combine the four feature vectors to obtain the conjoint feature vector for a block. The feature vector size of the block is 1×24.

3. Experimental Setup

The main contribution of this paper is that we proposed a new descriptor to solve the local region feature description problem of multi-sensor images with non-linear radiation variations. In this

section, the experimental setup is presented to test our proposed HOMPC descriptor. Figure 7 shows the flowchart to evaluate the descriptor performance in our experiments. In order to evaluate the performance of descriptors, some pair of multi-sensor images are chosen and the feature points are detected firstly. Then, the local regions corresponding to the feature points are described using descriptors, and the feature matching is completed with some similarity metric. Then, the correctly matched points and falsely matched points are classified by using the homography and projection error. The homography is calculated by using the manually identified corresponding points of the pair of multi-sensor images. Since two important indicators for evaluating descriptor performance are "repeatability" and "distinctiveness" [19], we finally selected precision and recall to verify the "distinctiveness" and "repeatability" of descriptors, respectively. The F1-measure and area under precision-recall curves (AUCPR) are used to verify the comprehensive performance of precision and recall. A more detailed introduction of the feature detection method is presented in Section 3.1. The details of the feature descriptors are supplied in Section 3.2, and the evaluation criteria and datasets are provided in Section 3.3 and 3.4, respectively. The parameter settings in our experiments are listed in Section 3.5.
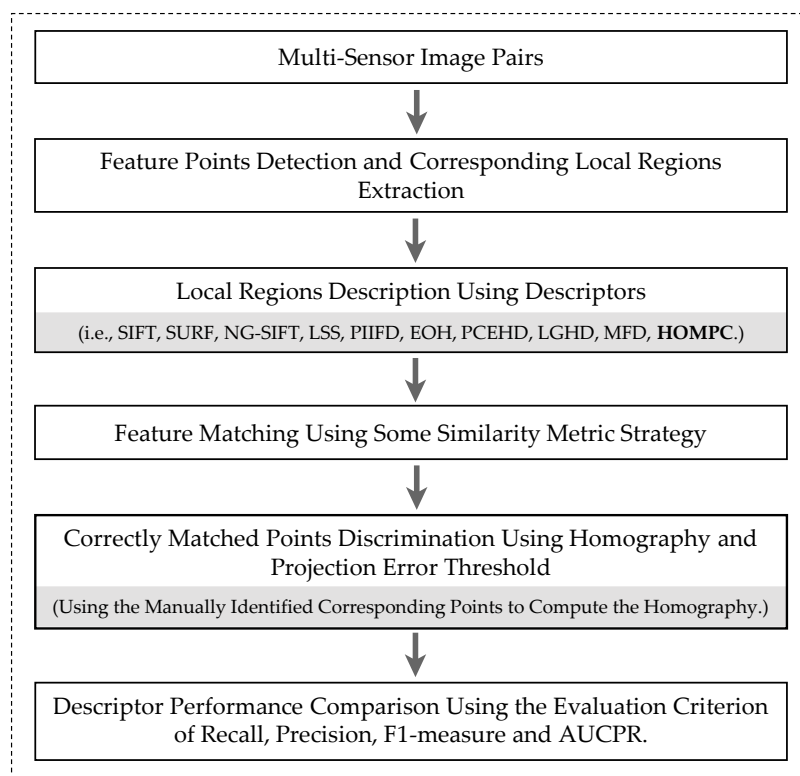

Figure 7. The flowchart to evaluate the descriptor performance in our experiments.

3.1. Feature Detection

The feature detection is the preprocessing for local regions description. Saleem et al. [43] compared different feature detectors and descriptors on four multi-sensor images, and the experimental results showed that the corner feature detectors (e.g., Harris, FAST [44]) could obtain more correct matches than blob detectors (e.g., SIFT, SURF) for multi-sensor images. The aim is to verify the performance of descriptors, and the more the number of the initial corresponding correct matching points, the more beneficial to verify the performance of descriptors.

In our experiment, we selected the phase congruency corner detector [33] to detect feature points. Different from the gradient-based corner detectors (e.g., Harris, FAST), the phase congruency corner detector is based on phase information, which is invariant to contrast and illumination changes.

In addition, the non-maximum suppression method was used to extract local (3×3) significant points. After then, we could obtain more initial corresponding correct matches to verify the descriptor performance. We detected a total of 2500 points for each image of multi-sensor images. However, image texture and scene complexity affect the final detected points.

3.2. Feature Descriptors

In addition to verifying the performance of our descriptors on multi-sensor images, we need to compare our proposed descriptors with classic descriptors through multi-sensor images. We selected some representative descriptors, e.g., SIFT [11], SURF [12], LSS [5], EOH [18], and five state-of-the-art descriptors named NG-SIFT [17], PIIFD [13], PCEHD [28], LGHD [29], and MFD [31] to compare with our proposed HOMPC descriptor. We programmatically implemented NG-SIFT and MFD descriptors and tried to maximize their performance using MATLAB. Whereas the implementation of the remaining algorithms is available online with MATLAB.

3.3. Evaluation Criteria

The ability of a descriptor can be evaluated by "repeatability" and "distinctiveness" [19]. The common evaluation criteria, including recall and precision from [45], F1-measure, and AUCPR from [17], and projection error from [43,46] were chosen to measure the descriptor performance.

Projection error is a Euclidian distance between the reference and projected image feature points. The projected image feature points are transformed by using a known homography H, which is computed by manually selecting the corresponding points of reference and target images. If the projection error of one pair of points is less than some threshold, we regard it as the correctly matched points, otherwise, the falsely matched points.

Descriptor matching is carried out with MATLAB implementation of 'matchFeatures' function, which uses Sum of absolute differences (SAD), sum of squared differences (SSD) or Hamming distance, between descriptor vectors. As introduced in the document of MATLAB, the default SSD metric is suitable for non-binary feature descriptors, so the SSD was selected as the feature similar metric. A reference point of interest is defined as being matched to a test interest point if:

$$D(d^i, d^j) < thresh \times D(d^i, d^k),\tag{12}$$

where $D(.,.)$ is the Euclidean distance; d^i is the feature vector of the reference point of interest; d^j *and* d^k are the two feature vectors of points to be matched and d^k is the second-closest neighbor to d^i. The "*thresh*" is the threshold of the nearest neighbor distance ratio (NNDR). The smaller value of "*thresh*" means tighter matching criterion. The "*thresh*" was set to 0.8–1.0 at an interval of 0.05 in our experiments to test the matching results with different conditions.

Recall is the ratio of the number of correctly matched point pairs of the matching results and the total number of existing correct-match point pairs of the initial match point sets. *Recall* assesses the accuracy of the returning ground-truth image pairs.

Precision is the ratio of the number of correctly matched point pairs of the matching result and the sum of the number of correctly and falsely matched point pairs of the matching results. *Precision* calculates the ability to exclude false positives.

F1-measure captures the fitness of the ground truth and detected points by jointly considering the **Recall** and **Precision**. The F1-measure [31,47] is calculated as follows:

$$\text{F1} - \text{measure} = \frac{2 \times \text{Precison} \times \text{Recall}}{\text{Precision} + \text{Recall}}\tag{13}$$

AUCPR is the area under the precision-recall curve [17], which was also computed for the performance comparison.

To better express the universality of the proposed HOMPC descriptor and the other descriptors, the average precision and recall were calculated from the multi-sensor image pairs of each dataset according to different NNDR thresholds. The F1-measure and AUCPR were calculated from the average precision and recall values.

3.4. Datasets

The descriptor we designed was mainly to solve the problem of common feature description between images with nonlinear radiation differences of multi-sensor images. The image datasets used in our experiments should be acquired from different sensor devices. Three datasets (CVC dataset, UAV dataset, and SAT dataset) were chosen based on the height of the remote sensing platform. Samples from these datasets are shown in Figure 8.

- The CVC dataset includes 44 RGB/LWIR (longwave infrared) [29,48] outdoor image pairs and 100 VS/LWIR (visible and longwave infrared) [18,49] outdoor images of different urban scenarios. These image pairs were captured using the color cameras and infrared camera. For a more detailed description with the two cameras, please refer to [49].
- The UAV dataset includes 27 RGB/LWIR outdoor images specially acquired by ourselves from an unmanned aerial vehicle (UAV) (DJI M100) using the optical and thermal infrared cameras. The image resolution of the thermal infrared camera is 640×480, and the wavelength range is 8–14 μm. The optical camera is industrial grade. Its pixels size is 5.0 μm $\times$ 5.2 μm and image resolution is 1280×720.
- The SAT dataset contains six pairs of remote sensing satellite images. The image pairs cover a variety of low-, medium-, and high-resolution remote sensing satellite images with a ground sample distance (GSD) from 0.5 to 30 m. The supported images came from different satellites. The multi-temporal satellite images were also included. For a more detailed description for the six pairs of images, please refer to [7].
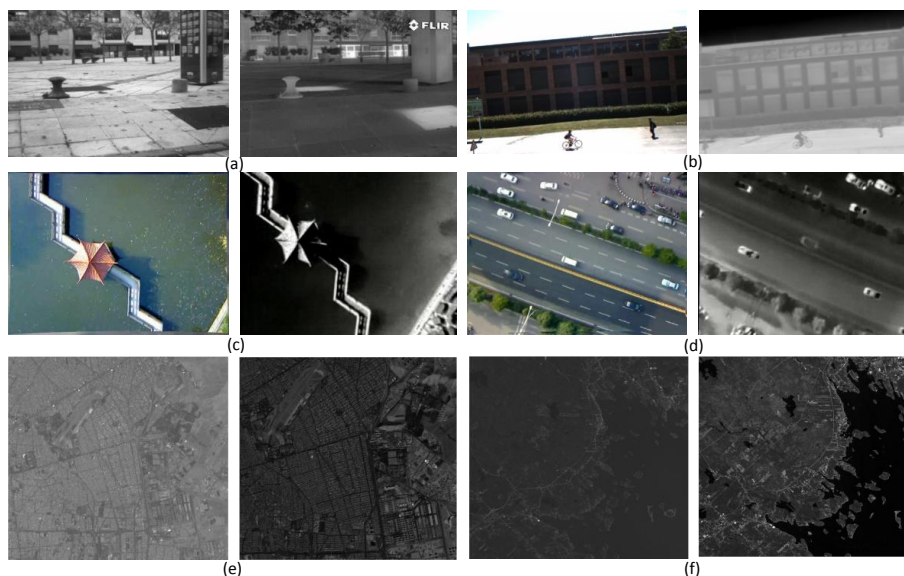

Figure 8. Samples of multi-sensor image pairs in the three datasets. (**a**,**b**) are two image pairs from the CVC dataset using optical camera and infrared camera devices [48,49]; (**c**,**d**) are two image pairs from the UAV dataset using optical and thermal camera devices by ourselves; and (**e**,**f**) are two multi-sensor and multi-temporal image pairs from the SAT dataset using different satellites. Referring to [7], the details of the (left vs. right) images in (**e**) include: satellite type (IRS-1C vs. ASTER), spectral mode (Pan vs. Band3), and acquisition date (1998 vs. 2006); the details of the (left vs. right) images in (**f**) include: satellite type (SPLOT5 vs. LANDSAT ETM+), spectral mode (Band2 vs. Band3), and acquisition date (2006 vs. 1999).

Similar to LGHD and MFD, it should be noted that the proposed HOMPC descriptor is designed for non-linear radiation variations between multi-sensor images, and we do not consider the geometric changes of rotation and scale variations. Therefore, the multi-sensor remote sensing image pairs should be rectified without significant rotation or scaling transformation. In fact, the image pairs of the CVC dataset were rectified and aligned so that matches should be found in horizontal lines. The image pairs of the UAV dataset included small projective changes, in addition to small rotation transformation. The image pairs of the SAT dataset had been rectified to the same scale size based on GSD values.

3.5. Parameter Settings

The patch size of the proposed HOMPC descriptor was assigned to 80×80, the same as EOH, LGHD, and MFD. The patch sizes of other descriptors were set by default. We use $N_S = 4$ and $N_O = 6$ to express the number of convolution scales and orientations of the log-Gabor filter in the proposed method. The projection error threshold was assigned to 5 for all three datasets. The larger the projection error threshold, the more correctly matched points for the recall and precision calculations. The threshold of NNDR was assigned to the range 0.8–1.0 with an interval of 0.05 in our experiments, where a smaller value means a tighter matching criterion.

4. Experimental Results and Discussion

The parameters of the proposed method are discussed in Section 4.1. The advantages of combining the phase information and magnitude information are evaluated and discussed in Section 4.2. The superiority of the proposed HOMPC descriptor over the current state-of-the-art local feature descriptors is evaluated and discussed in Sections 4.3 and 4.4.

4.1. Parameter Study

The proposed HOMPC method contains four main parameters, namely, N_S, N_O, C_L, and I_L. As mentioned $N_S = 4$ and $N_O = 6$ are used to express the number of convolution scales and orientations of the log-Gabor filter in the proposed method. Parameter C_L is the size of the local cell patch (cell size) used for the cell description. If the cell size is too small, it will contain insufficient information with which it is difficult to reflect the distinctiveness of the feature. In contrast, if the cell size is too large, it is easily affected by the local geometric distortion. Parameter I_L is the number of intervals between blocks. In general, the smaller the number of intervals, the richer the information of the constructed HOM and HPC, and the poorer the robustness and the higher the dimension of the feature vectors. In contrast, if the number of intervals is too large, the HOMPC descriptor will contain less information, which will also affect the distinctiveness of the feature. Therefore, suitable parameters are very important. In this section, we describe the parameter study and sensitivity analysis based on the 44RGB/LWIR dataset. We designed two independent experiments to learn parameters C_L and I_L, where each experiment had only one parameter as a variable, with the other parameters as fixed values. The experimental setup details are summarized in Table 1. For each parameter, we use the average precision and recall, F1-measure, and AUCPR as the evaluation metric. The experimental results are reported in Figures 9 and 10; Tables 2 and 3.

From the experimental results, we can infer that: (1) Larger values of C_L mean that the cell size information is richer, and thus more AUCPR values and F1-measure scores can be obtained; however, due to the effect of local geometric distortion, the AUCPR values and F1-measure scores will decrease as the value of C_L increases. As shown in Table 2 and Figure 9, the HOMPC descriptor achieves the best performance when $C_L = 20$. Therefore, we set $C_L = 20$; (2) From Table 2 and Figure 9, we can see that large values of I_L result in a poor performance, while small values mean richer information of the constructed HOM and HPC feature vectors; however, smaller values of I_L also mean that the distinctiveness of the proposed HOMPC feature vector increases, which will decrease the robustness and increase the dimension of the proposed feature vectors. As shown in Table 3 and

Figure 10, when $I_L = 8$, HOMPC achieves the best performance. Therefore, we set $I_L = 8$. Based on the experimental results and analysis, these parameters were fixed to $C_L = 20$ and $I_L = 8$ in the experiments.

Table 1. Details of the parameter settings.

	Variable	Fixed Parameters
Parameter C_L	$C_L = \{6, 8, 10, 12, 16, 20, 24\}$	$I_L = 8, N_S = 4, N_O = 6$
Parameter I_L	$I_L = \{6, 8, 10, 12, 16, 20\}$	$C_L = 20, N_S = 4, N_O = 6$

Table 2. The AUCPR (%) results of parameter CL.

Metric	$C_L, I_L = 8, N_S = 4, N_O = 6$						
	6	**8**	**10**	**12**	**16**	**20**	**24**
AUCPR	38.43	38.01	40.96	42.99	41.08	**46.40**	45.00

Table 3. The AUCPR (%) results of parameter IL.

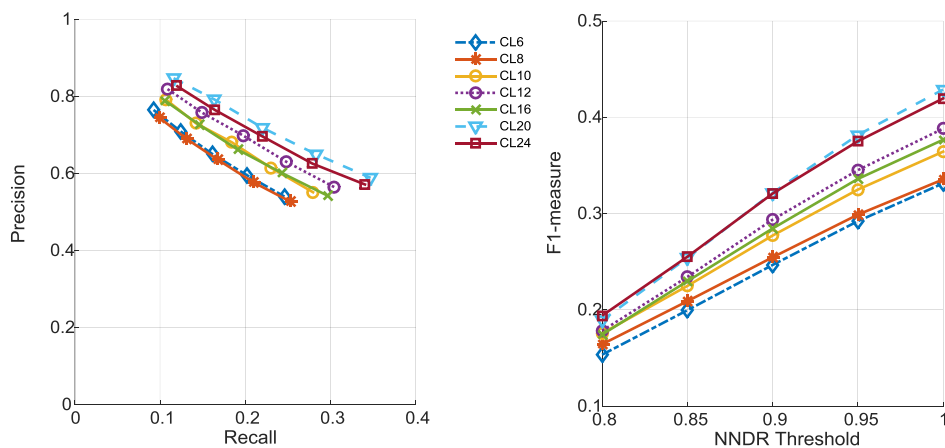| Metric	$I_L, C_L = 20, N_S = 4, N_O = 6$					
	6	**8**	**10**	**12**	**16**	**20**
AUCPR	44.73	**46.40**	41.12	41.01	34.97	29.55

Figure 9. The average PR curve and F1-measure curve results of parameter C_L ($I_L = 8$).
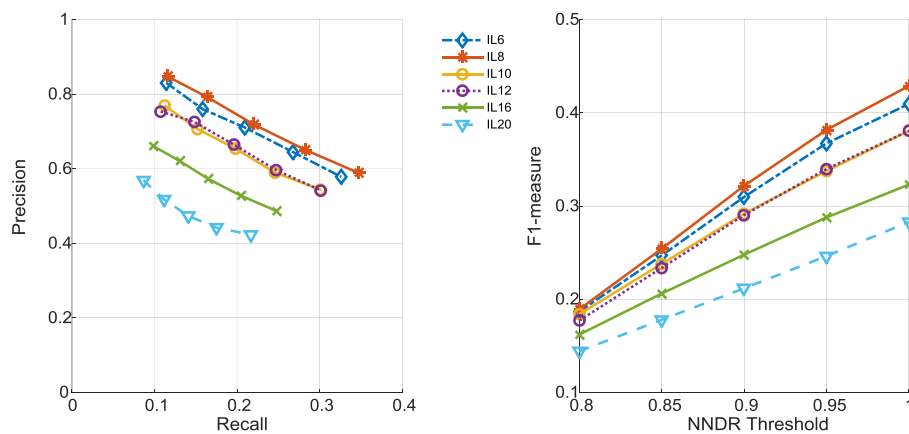

Figure 10. The average PR curve and F1-measure curve results of parameter I_L ($C_L = 20$).

4.2. The Advantages of the Magnitude and Phase Congruency Information Combination

　　To verify the advantages of combining the magnitude and phase congruency information, the HOMPC descriptor was compared with HPC and HOM based on the 44RGB/LWIR dataset. The variable parameters were set as suggested in Section 4.1. The average precision and recall results, as well as the F1-measure and AUCPR results of HOMPC, HOM, and HPC are given in Figures 11 and 12.
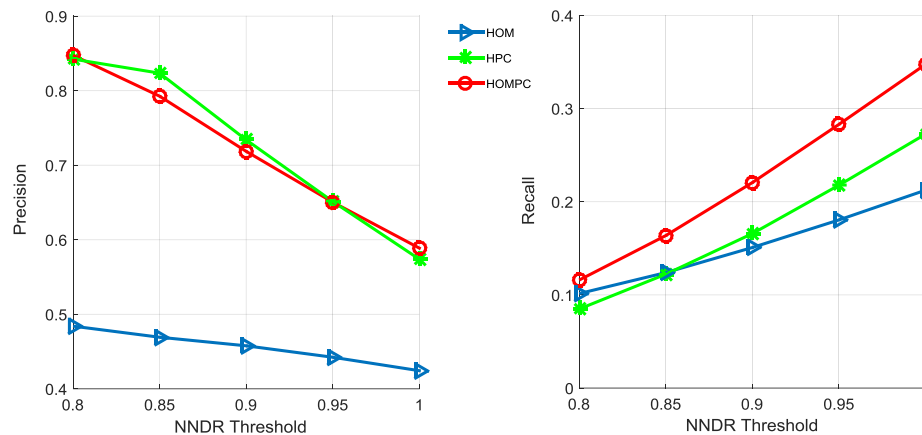

Figure 11. The average precision and recall curve results with different thresholds for HOMPC, HPC, and HOM.

　　Since the larger the NNDR threshold, the looser the matching metric for the point matching, the average curves of the recall values of HOMPC, HOM, and HPC are raised while the average curves of the precision values are decreased as the NNDR threshold increases, as shown in Figure 11. This indicates that the average curve of the recall values of the HOMPC descriptor is much better than the curves of HOM and HPC, and the average curve of the precision values of HOMPC is similar to that of HPC, but is much better than that of HOM. As it is known that the phase information contributes more to the preservation of image features than amplitude information, the precision values of HPC are much better than those of HOM. However, the distribution of the high-frequency component information also contributes to the shape of the objects, and we can see that the recall values of HOM are similar to those of HPC. After combining HOM and HPC, HOMPC keeps the distinctiveness of HPC, but also increases the repeatability by adding HOM.

　　A comprehensive analysis of the average recall and precision is provided in Table 4 and Figure 12. The average PR curves, F1-measure curves, and AUCPR values also indicate that HOMPC outperforms HOM and HPC.

Table 4. AUCPR (%) results for HOM, HPC, and HOMPC.

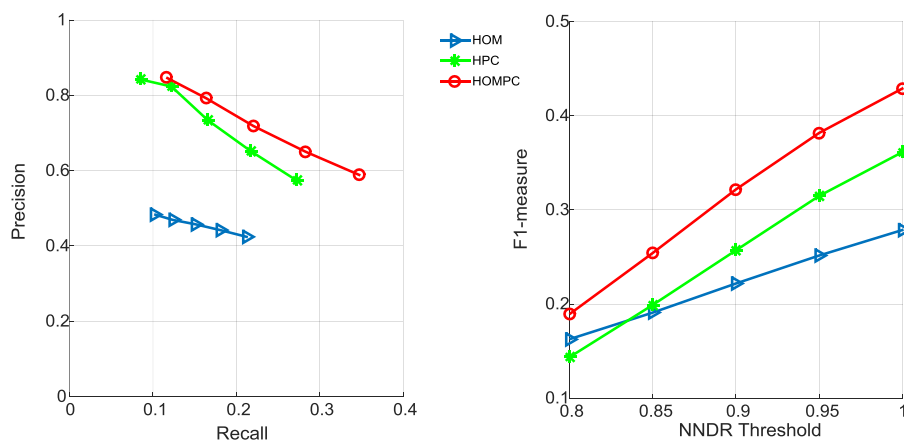| Feature Vector	HOM	HPC	HOMPC
AUCPR	29.27	42.15	**46.40**

Figure 12. The average PR curves and F1-measure curves of HOMPC, HPC, and HOM.

It is found that the advantages of combining the magnitude and phase congruency information are significant. More particularly, it is the phase congruency information that makes a greater contribution to the distinctiveness of HOMPC, and the magnitude information adds to the repeatability of HOMPC.

4.3. Descriptor Comparison

We used the three datasets introduced in Section 3.4, i.e., CVC dataset, UAV dataset and SAT dataset to compare the proposed HOMPC descriptor with the current state-of-the-art local feature descriptors of SIFT [11], SURF [12], NG-SIFT [17], LSS [5], PIIFD [13], EOH [18], PCEHD [28], LGHD [29], and MFD [31] in terms of the average precision and recall, as well as F1-measure and AUCPR.

4.3.1. Results Obtained with the CVC Dataset

Figure 13 shows the average precision and average recall curve results obtained with the 144 pairs of visible/longwave infrared images for all the descriptors. As can be seen, on the whole, the average precision curves descend while the average recall curves ascend as the NNDR threshold increases, and the range of the precision values is much larger than that of the recall values. The average precision and recall curves of the proposed HOMPC descriptor show superior results when compared to the other descriptors. Among the other descriptors, the LGHD descriptor performs much better than the remaining descriptors, and PIIFD, EOH, PCEHD, and MFD perform better than LSS, SURF, LSS, and NG-SIFT, all of which present similar results.
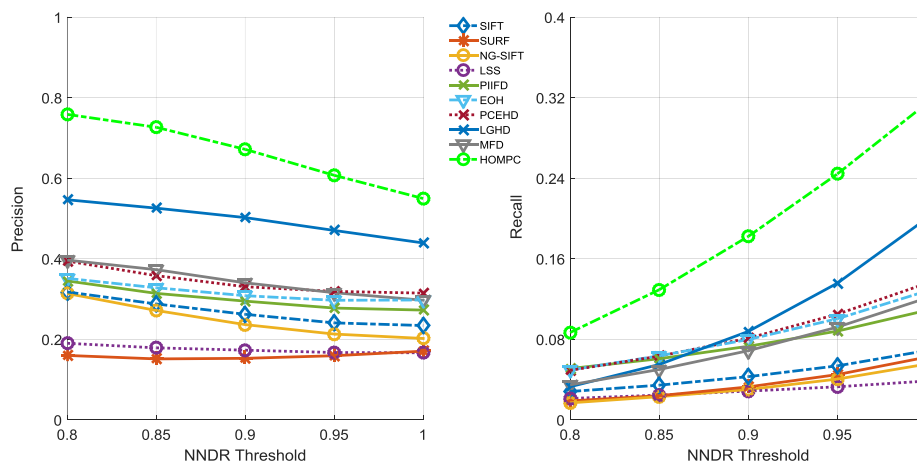
Figure 13. The average precision and recall curve results with different NNDR thresholds for all the descriptors. The HOMPC is our descriptor.

The average PR curves based on the average precision and recall values and the F1-measure curves based on the precision and recall values of the 144 pairs of images for all the descriptors are provided in Figure 14. The AUCPR results based on the average PR curves for all the descriptors are presented in Table 5. We can clearly see that the AUCPR value of HOMPC is much greater than that of the other descriptors. Furthermore, we can see that the shape of the F1-measure curve is much more similar to that of the recall curve than the precision curve. This is because the recall and precision contribute the same weight to the F1-measure score, and the smaller recall values have more impact on the F1-measure scores. Therefore, when the NNDR threshold increases, the F1-measure curve increases at first as the recall curve also increases. Considering the comprehensive evaluation of the descriptor performance, the larger the F1-measure score, the better the performance. It is shown that the F1-measure values are the best for all the descriptors when the NNDR threshold = 1. Figure 15 shows the correctly matched points (green lines) and falsely matched points (red lines) of two typical descriptors (SURF and LGHD) and the proposed HOMPC descriptor of sample image pair (Figure 8a) when the NNDR threshold = 1. The correctly matched points, precision, and recall values are also displayed. Overall, we can see that the proposed HOMPC performs the best of all the descriptors.

Table 5. AUCPR (%) results for all the descriptors.

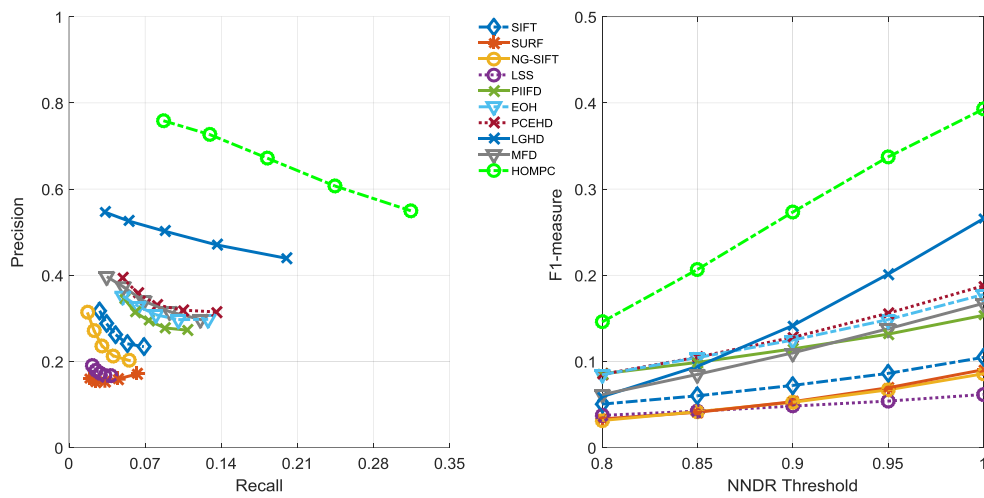| Descriptor	SIFT	SURF	NG-SIFT	LSS	PIIFD	EOH	PCEHD	LGHD	MFD	HOMPC
AUCPR	13.82	9.78	11.56	9.58	17.27	18.77	19.93	28.25	18.40	**41.35**

Figure 14. The average PR curves and F1-measure curve results for all the descriptors.

(**a**) Descriptor = SURF, precision = 0.1729, recall = 0.0639, and correct matches = 37.

(**b**) Descriptor = LGHD, precision = 0.6385, recall = 0.2867, and correct matches = 166.
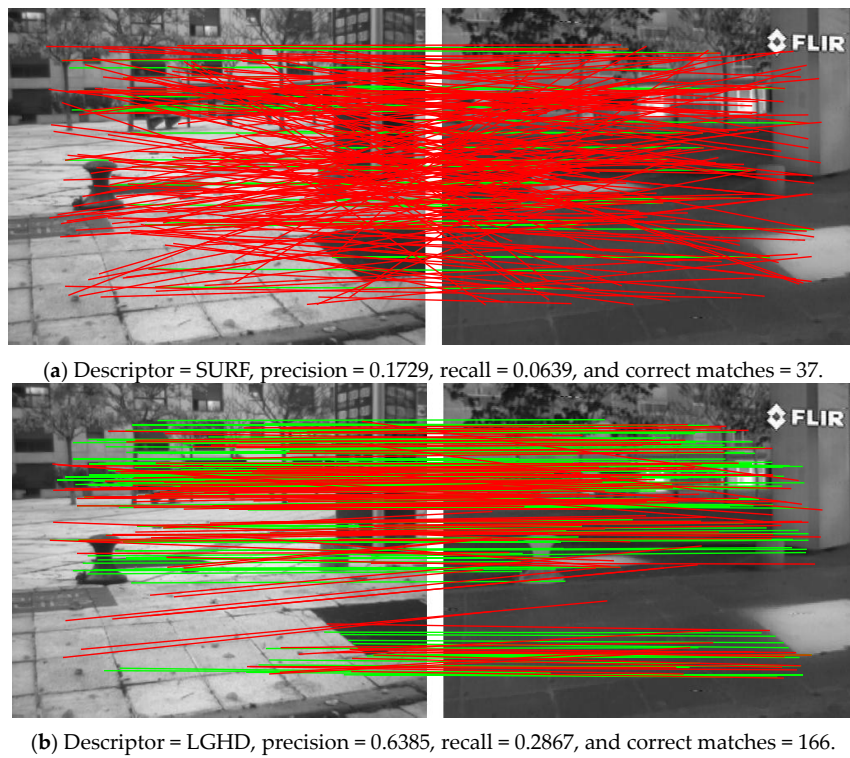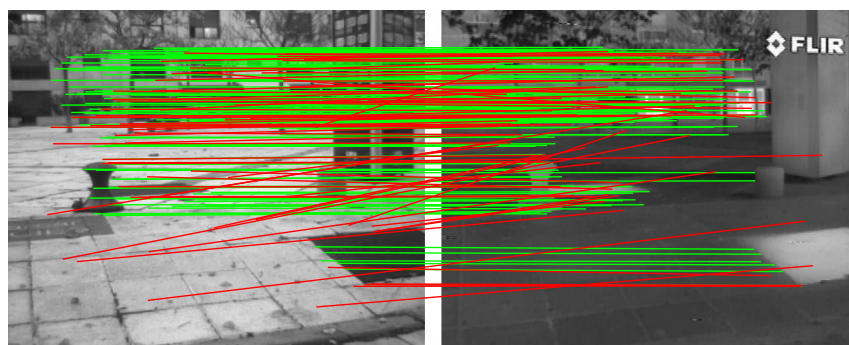
Figure 15. *Cont.*

(**c**) Descriptor = HOMPC, precision = 0.7621, recall = 0.3022, and correct matches = 175.

Figure 15. The correctly matched (green lines) and falsely matched (red lines) results for SURF (**a**), LGHD (**b**), and our HOMPC (**c**) descriptors. The displayed green and red lines are used to visually compare the descriptor performance. The greater number of the green lines, the better. The higher the ratio of the green lines, the better. The quantitative results are also listed below each pair of images.

4.3.2. Results Obtained with the UAV Dataset

Figures 16 and 17, Table 6 show the average precision and recall results, as well as F1-measure and AUCPR results obtained with the 27 pairs of RGB/thermal infrared images. As can be seen in Figure 16, the precision curves of HOMPC are again the best, but the average precision values and recall values are much lower than the average precision and recall results, as well as the F1-measure and AUCPR results for the CVC dataset, especially for the recall values. The AUCPR results of all the descriptors also indicate that the biggest value of HOMPC is just 14.95, which is much lower than most of the results for the CVC dataset. This is because the CVC dataset has been vertically rectified. In contrast, when using the low-resolution thermal infrared camera, the thermal infrared images in the UAV dataset have a lower resolution and there is noise generated in the images. Furthermore, the UAV dataset also has a slight projective transformation caused by the UAV platform. However, we can still clearly see that the proposed HOMPC descriptor is more robust than the other descriptors. Nevertheless, the AUCPR values are very low (below 15%), even if the precision of HOMPC is good when the recall is low. This means that the distances between the feature vectors for local feature regions for true or false matchings are very close but are discriminant enough when there is a need to retrieve only one region. Therefore, the threshold should be large enough to obtain more matching points. Figure 18 shows the correctly matched points (green lines) and false matches (red lines) of two typical descriptors (SURF and LGHD) and the proposed HOMPC descriptor for a pair of images (Figure 8d) when the NNDR threshold = 1. The results of precision and recall, as well as correct matches are also given in Figure 18a–c. It is shown that all three descriptors performed poor, but our proposed HOMPC descriptor also performed best.

Table 6. AUCPR (%) results for all the descriptors.

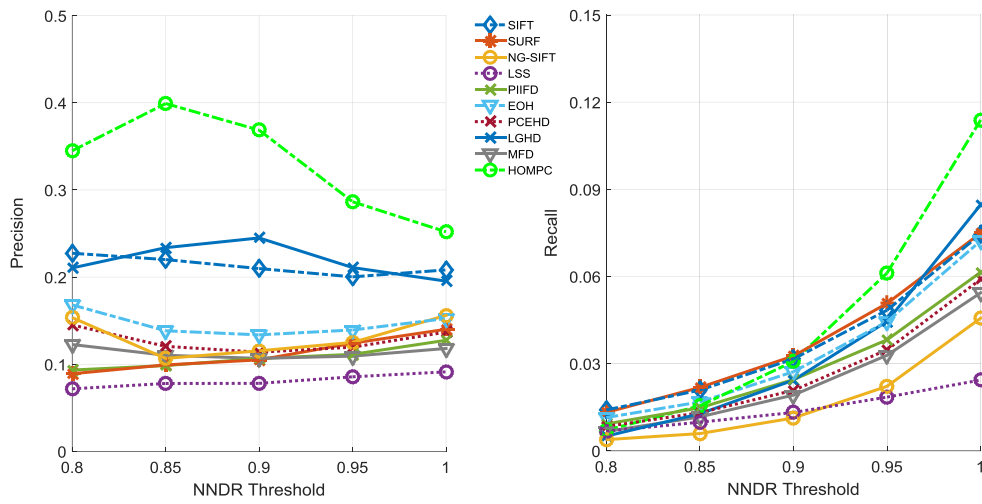| Descriptor	SIFT	SURF	NG-SIFT	LSS	PIIFD	EOH	PCEHD	LGHD	MFD	HOMPC
AUCPR	11.76	7.92	8.22	4.96	7.07	8.60	7.55	10.97	6.49	**14.95**

Figure 16. The average precision and recall curve results with different NNDR thresholds for all the descriptors. The HOMPC is our descriptor.
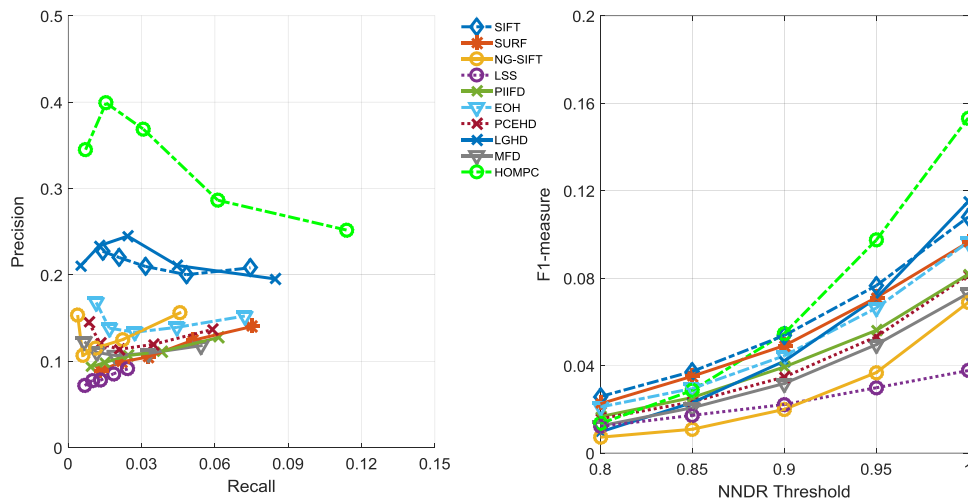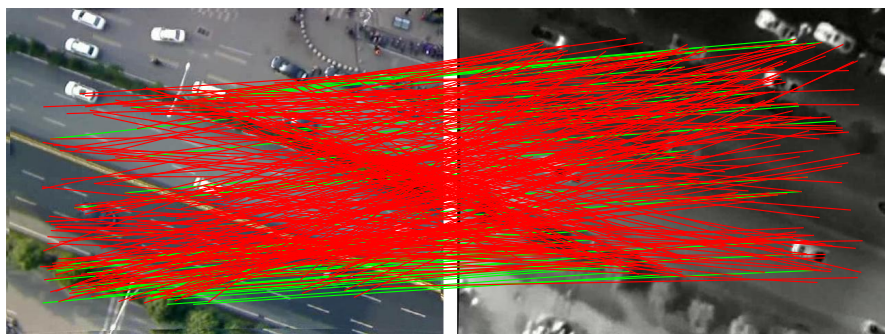

Figure 17. The average PR curves and F1-measure curve results for all the descriptors.

(**a**) Descriptor = SURF, precision = 0.2138, recall = 0.0655, and correct matches = 65.
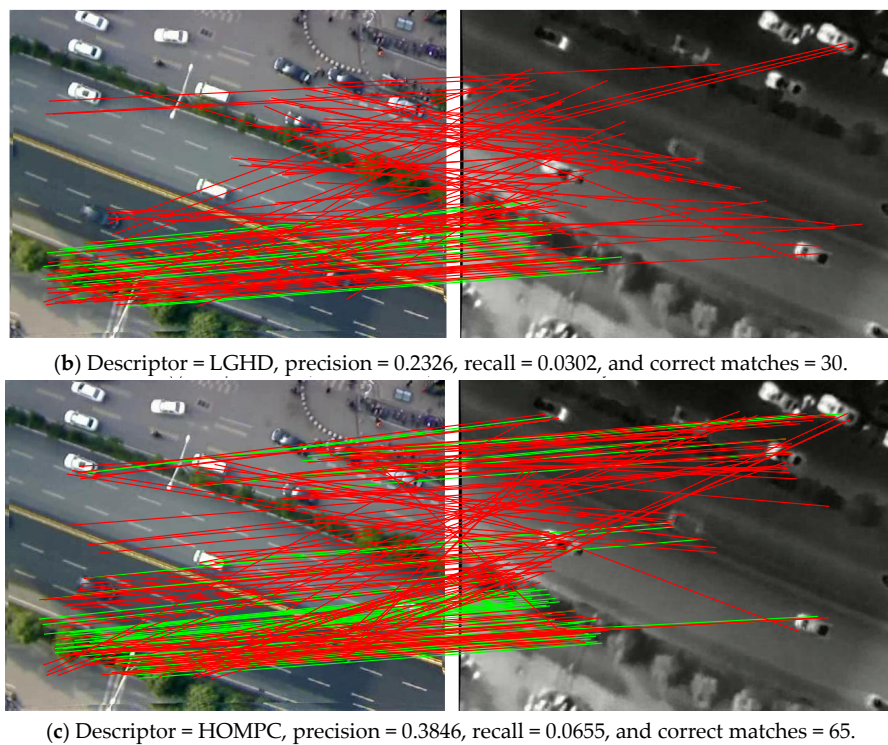
Figure 18. *Cont.*

(**b**) Descriptor = LGHD, precision = 0.2326, recall = 0.0302, and correct matches = 30.

(**c**) Descriptor = HOMPC, precision = 0.3846, recall = 0.0655, and correct matches = 65.

Figure 18. The correctly matched (green lines) and falsely matched (red lines) results for SURF (**a**), LGHD (**b**), and our HOMPC (**c**) descriptors. The displayed green and red lines are used to visually compare the descriptor performance. The greater number of the green lines, the better. The higher the ratio of the green lines, the better. The quantitative results are also listed below each pair of images.

4.3.3. Results Obtained with the SAT Dataset

Figures 19 and 20, Table 7 illustrate the average precision and recall results, as well as F1-measure and AUCPR results obtained with the six pairs of multi-sensor images from the SAT dataset. Compared with the results obtained with the CVC dataset and UAV datasets, it can be seen that the average precision and recall curves for the SAT dataset are much better, as are the F1-measure curves and AUCPR values. This is because the spectral ranges of the SAT dataset are much closer than the visible and thermal infrared images of the CVC dataset and UAV dataset. The greater the spectral range, the greater the difference between two local feature regions and, hence, the less efficient the usual descriptors become. However, the proposed HOMPC descriptor performs much better than the other descriptors, and LGHD, SIFT, NG-SIFT and MFD perform better than the remaining descriptors. The results of SURF, PIIFD, EOH and PCEHD are similar, and they are much better than the results of the LSS descriptor. Figure 21 shows the correctly matched points (green lines) and falsely matched points (red lines) of two typical descriptors (SURF and LGHD) and the proposed HOMPC descriptor of the sample image pair (Figure 8e) when the NNDR threshold = 1. The correctly matched points, precision, and recall values are also listed. It is shown that our proposed HOMPC descriptor is suitable for the feature description of multi-temporal, multi-sensor image pairs.
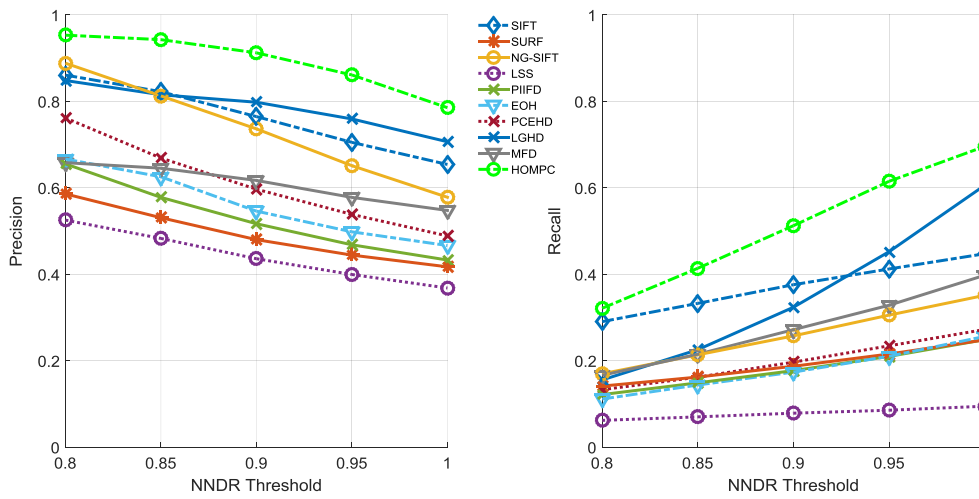
Figure 19. The average precision and recall curve results with different NNDR thresholds for all the descriptors. The HOMPC is our descriptor.
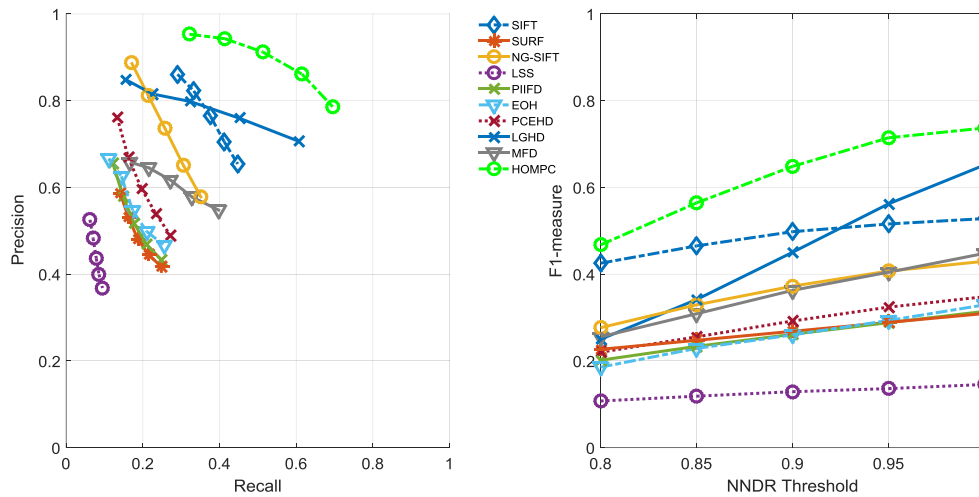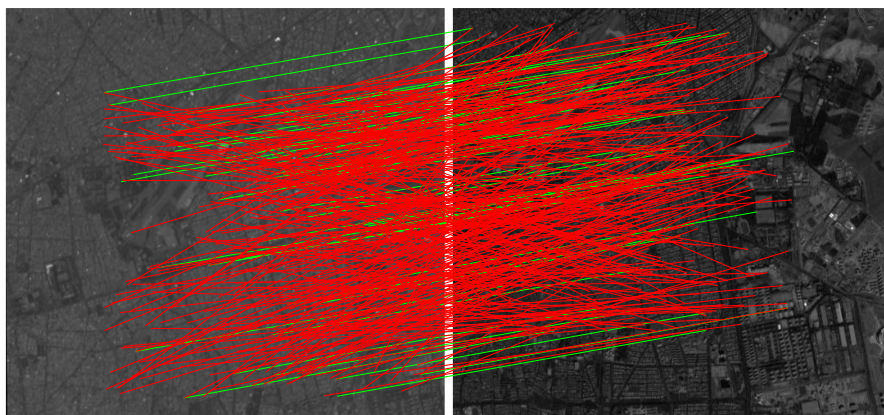

Figure 20. The average PR curves and F1-measure curve results for all the descriptors.

(**a**) Descriptor = SURF, precision = 0.1408, Recall = 0.0875, and correct matches = 49.
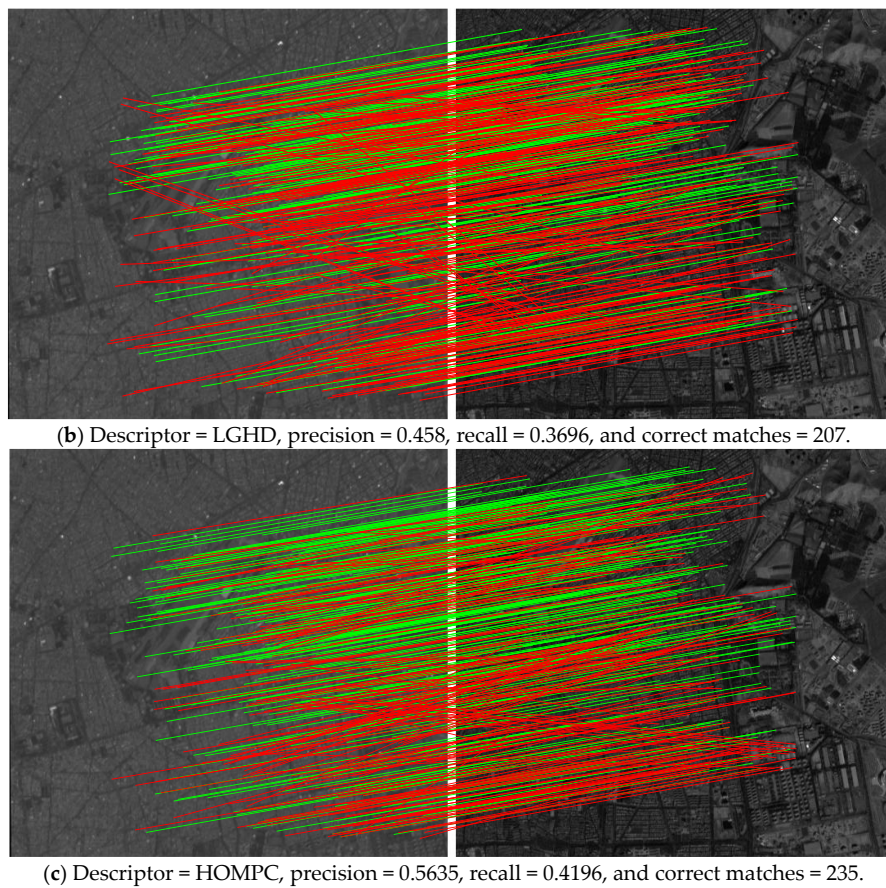
Figure 21. *Cont.*

(**b**) Descriptor = LGHD, precision = 0.458, recall = 0.3696, and correct matches = 207.

(**c**) Descriptor = HOMPC, precision = 0.5635, recall = 0.4196, and correct matches = 235.

Figure 21. The correctly matched (green lines) and falsely matched (red lines) results for SURF (**a**), LGHD (**b**), and our HOMPC (**c**) descriptors. The displayed green and red lines are used to visually compare the descriptor performance. The greater number of the green lines, the better. The higher the ratio of the green lines, the better. The quantitative results are also listed below each pair of images.

Table 7. AUCPR (%) results for all the descriptors.

Descriptor	SIFT	SURF	NG-SIFT	LSS	PIIFD	EOH	PCEHD	LGHD	MFD	HOMPC
AUCPR	57.16	32.05	48.07	22.85	32.93	34.62	37.87	63.33	44.34	**77.01**

4.3.4. Descriptor Computational Efficiency

Based on the experimental results obtained using a PC with an Intel Core i3 CPU @ 2.5 GHz and 8 GB RAM, the average computation times of the descriptors for each feature point of all three datasets are shown in Figure 22. The LGHD descriptor is associated with a high computation time. This is because the LGHD descriptor use the multi-scale and multi-oriented magnitudes of the log-Gabor filters to calculate the feature vectors of all the scales. The MFD and EOH are much faster than LGHD because they use fewer scales and filters. The PIIFD and SIFT use trilinear interpolation in the algorithm and it is, therefore, much slower than SURF, which uses Haar wavelets and integral images for image convolution to obtain a fast feature descriptor. Instead of trilinear interpolation before constructing descriptors, the NG-SIFT uses a uniform feature weighting scheme, and LSS uses the maximum self-similarity values contributes the histograms of LSS. Similarly, the proposed HOMPC uses the values of convolved PCMs and MBMs to directly contribute to the bins of each cell. In this way, it much faster to construct the HOMPC descriptor. The calculations of PCMs and MBMs based on log-Gabor filters made it slightly slower than SURF, NG-SIFT, and LSS. However, the average precision and recall results, as well as the F1-measure and AUCPR results of SURF, NG-SIFT, and LSS for the

datasets performed poorly. The average computation time of the proposed HOMPC descriptor is far lower than that of the other log-Gabor-based descriptors (LGHD and MFD).
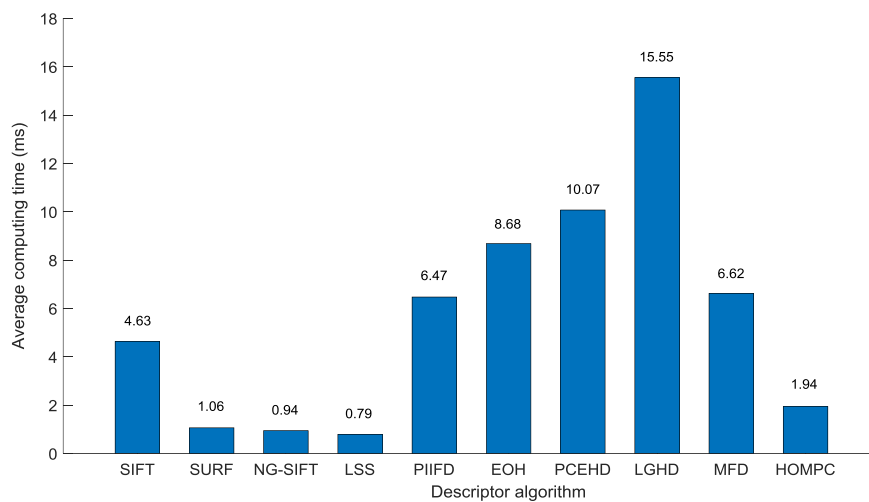

Figure 22. The average computation time per a feature point using different descriptor. The HOMPC is our descriptor.

4.3.5. Influence of Rotation and Scale Variations to our Descriptor

The previous experiments have proved that the proposed HOMPC descriptor is robust to non-linear radiation variations between multi-sensor images, and the proposed HOMPC descriptor outperformed the state-of-the-art descriptors. In this part, the descriptor robustness to rotation and scale variations is evaluated, although the HOMPC descriptor is not designed to be rotationally invariant. We selected one pair of multi-temporal, multi-sensor images (Figure 8e) as the test data.

We firstly tested the descriptor performance with rotation changes. While the left image of Figure 8e remains unchanged, the right image of Figure 8e is rotated from −20 to 20 degrees with an interval of five degrees. The precision, recall, and correct matches of different rotation angles are presented in Table 8. It is shown that the rotation changes that can be tolerated are between −10 and 10 degrees. Subsequently, we tested the descriptor performance with scale variations. The left image of Figure 8e remains unchanged, and the right image of Figure 8e is resized from 0.5 to 1.8. It is revealed that the scale variations that can be tolerated are between 0.8 and 1.2, shown in Table 9.

Table 8. The precision, recall, and correct matches of different rotation angles (degree).

Rotation Angle (°)	−20	−15	−10	−5	0	5	10	15	20
Precision (%)	2.3	6.17	**31.15**	**53.88**	**56.35**	**51.04**	**30.05**	4.5	3.94
Recall (%)	0.36	0.89	**6.79**	**24.82**	**41.96**	**26.25**	**10.36**	0.89	0.89
Correct Matches	2	3	**38**	**139**	**235**	**147**	**58**	5	5

Table 9. The precision, recall, and correct matches of different scales.

Scales	0.5	0.6	0.7	0.8	0.9	1	1.2	1.5	1.8
Precision (%)	7.06	11.04	17.5	**30.2**	**48.97**	**56.35**	**34.5**	17.76	7.38
Recall (%)	2.14	3.21	6.25	**13.21**	**29.64**	**41.96**	**15.89**	4.82	1.61
Correct Matches	12	18	35	**74**	**166**	**235**	**89**	27	9

It can be observed from the results that HOMPC can tolerate rotations between images less than 10 degrees, which could be enough for multi-sensor remote sensing images, because the geometry of remote sensing images can be roughly calibrated by geographical calibration. Additionally,

when dealing with the multi-sensor remote sensing images, the ground sample distance is usually known, and the images can be assigned to the same scale by resampling. Therefore, our HOMPC descriptor can tolerate scales between images from 0.8 to 1.2, which could be enough for multi-sensor remote sensing images.

4.4. Discussion

Comparing the average precision and recall results, as well as F1-measure and AUCPR results of the multi-sensor image pairs among the three datasets, we found that the results for the SAT dataset are better than for the CVC dataset and UAV dataset for all the descriptors, on the whole. This is because that the distance of the spectral range of multi-sensor devices influences the descriptor performance. In detail, the spectral modes of the SAT dataset (i.e., Band2/Band3, Pan/Pan, Pan/Band3, Band2/Band4) and the CVC dataset correspond to visible/longwave infrared, while the UAV dataset corresponds to visible/thermal infrared. Although all of the image pairs are selected from different sensors, the greater the spectral ranges of the sensors, the greater the differences between the two local feature regions are and, hence, the less efficient the local feature descriptors become. The results for the UAV dataset performed worst of all three datasets because the texture difference and the size of overlap regions affect the number of existing correct-match point pairs of the initial match point sets. The great spectral range between the visible and longwave infrared images made their texture difference become large, while the low pixel resolution of the thermal infrared camera and some geometry and perspective transformation between the visible and thermal images made the overlap regions of the reference and target images smaller. The number of existing correct-match point pairs of the initial match point sets of the UAV dataset is very limited, resulting in its low average precision and recall results, as well as the low F1-measure and AUCPR results of all descriptors.

We also found that the gradient-based descriptors, i.e., SIFT, SURF, NG-SIFT, LSS, and PIIFD are more sensitive to spectral differences than sensor differences. Meanwhile, the proposed HOMPC descriptor is more robust when dealing with greater spectral ranges, as well as different sensors. The average descriptor performance of the EOH, PCEHD, MFD, and LGHD descriptors lie between the gradient-based descriptors and the proposed HOMPC descriptor when dealing with visible and longwave infrared images. This is because the gradient-based descriptors rely on a linear relationship between images, and they are not appropriate for the significant non-linear intensity differences caused by radiometric variations.

The reason that the LGHD descriptor performs better than the gradient-based descriptors is that the LGHD descriptor use the four-scale and six-oriented log-Gabor filters (24 filters in total) to capture the multi-scale and multi-oriented magnitude feature information. The LGHD descriptor uses the distribution of the high-frequency components to express the shape and structure information of the objects and is robust to non-linear intensity variations. Since the LGHD descriptor needs to compute the feature vector at four scales, the computational efficiency is poor.

The MFD descriptor attempts to decrease the computation time of LGHD by using fewer log-Gabor filters (10 filters in total), and it shows a descriptor performance that is slightly better than LGHD when the NNDR threshold is 0.80. However, if the NNDR threshold is varied from 0.80 to 1.0 in intervals of 0.05, the descriptor performance of MFD is poorer than that of LGHD. MFD is efficient when the distances of the feature vectors between local regions of multispectral images for true or false matchings are not close, such as visible/near-infrared images. However, when the distances of the feature vectors between local regions are very close to each other, the feature vectors of local regions should be discriminant enough, as with visible/longwave infrared images. The LGHD descriptor using more log-Gabor filters than MFD can capture more feature information and is more discriminative when dealing with feature vectors between local regions that are close to each other. The average precision and recall results, as well as the F1-measure and AUCPR results obtained with the three datasets, also verify the above inference.

When using all the image pairs of the three datasets, the average results of the proposed HOMPC descriptor are much better than the other descriptors. This is because the proposed descriptor is based on the phase congruency and the distribution of the magnitude information and is robust to non-linear radiation variations of multi-sensor images. In addition, the phase congruency information ensures the precision, and the distribution of the magnitude information adds to the correct number of matched points, which is evaluated and discussed in Section 4.2. In fact, the significant performance improvement of HOMPC over the other descriptors demonstrates the effectiveness and advantages of the proposed strategies, including the PCMs and MBMs, the novel measure for calculating the feature vector of each cell based on a convolution operation, and the dense description by overlapping blocks. We have to mention that, in order to make the proposed HOMPC descriptor more efficient, the structure and shape of the common features are captured using the overlapping blocks, in addition to the combination of magnitude and phase congruency information. It is for this reason that our descriptors are more sensitive to rotation transformations.

In addition to the advantages of average precision and recall, as well as the F1-measure and AUCPR, the computational efficiency of the proposed HOMPC descriptor is far better than that of LGHD, which can be considered the second-best descriptor when comparing the results of the multi-sensor image pairs of the three datasets. Details of the computational efficiency are provided in Section 4.3.4. Additionally, the influences of rotation and scale variations to the proposed HOMPC descriptor is discussed in Section 4.3.5 in detail.

Summarizing the quantitative experimental results described in Sections 4.3 and 4.4, we can draw the following summaries:

- The proposed HOMPC descriptor is designed for the description problem of multi-sensor images with non-linear radiation variations.
- In the experiments undertaken in this study, HOMPC achieved very good average precision and recall on the three datasets.
- The descriptor performance of HOMPC is far superior to that of the classical local feature descriptors to describe the local regions of multi-sensor images.
- The time consumption of HOMPC is far lower than that of the other log-Gabor-based descriptors (LGHD and MFD).
- The proposed HOMPC descriptor can tolerate small amounts of rotation and scale variations.

5. Conclusions

In this paper, we proposed a novel descriptor (named HOMPC) based on the combination of magnitude and phase congruency information to capture the common information of multi-sensor images with non-linear radiation variations. We first introduce the concept of magnitudes of log-Gabor filters, and we then propose the PCMs and MBMs, preparing for the convolution operation. To accelerate the computational efficiency of each feature vector, we apply Gaussian filters and mean filters to the PCMs and MBMs, respectively. To capture the structure and shape properties of local regions, we describe the local regions using the HPC and HOM based on a dense grid of local histograms. Finally, we combine the HOM and HPC to obtain the proposed HOMPC descriptor, which can capture more common features of multi-sensor images from the combination of magnitude and phase congruency information.

To make a fair comparison between the local feature descriptors, we used the same feature detection method and the same similarity metric to uniformly test the descriptor performance in the experiments. In the experiments, we first studied the parameters and tested the advantages of integrating the HOM and HPC. The HOMPC descriptor was then evaluated using three datasets (CVC dataset, UAV dataset, and SAT dataset) and compared to the state-of-the-art local feature descriptors of SIFT, SURF, NG-SIFT, LSS, PIIFD, EOH, PCEHD, LGHD, and MFD. The experimental results confirmed that HOMPC outperforms the other local feature descriptors. Moreover, by designing a fast

method of constructing the feature vectors for each block, HOMPC has a much lower run time than LGHD, which achieved the second-highest F1-measure and AUCPR values in the experiments. Finally, the rotation and scale variations to the proposed HOMPC descriptor are evaluated and the results show that our HOMPC descriptor tolerates small amounts of rotation and scale variations, although the purpose is to address the non-linear radiation variations between multi-sensor images.

The HOMPC descriptor can be applied to change detection, target recognition, image analysis, image registration, and fusion of multi-sensor images. In our future work, we will test our HOMPC descriptor on more multi-sensor images with non-linear radiation variations, such as optical and SAR images, and optical and LiDAR images.

Author Contributions: Z.F., Q.Q., and B.L. conceived the approach and designed the experiments; Z.F. performed the experiments and analyzed the data; C.W. contributed the data analysis; Z.F. wrote the paper; and H.S. reviewed the paper.

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

References

1. Ye, Y.; Shan, J. A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences. *ISPRS J. Photogramm. Remote Sens.* **2014**, *90*, 83–95. [CrossRef]
2. Khaleghi, B.; Khamis, A.; Karray, F.O. Multisensor data fusion: A review of the state-of-the-art. *Inf. Fusion* **2013**, *14*, 28–44. [CrossRef]
3. Ma, J.; Chen, C.; Li, C. Infrared and visible image fusion via gradient transfer and total variation minimization. *Inf. Fusion* **2016**, *31*, 100–109. [CrossRef]
4. Deselaers, T.; Ferrari, V. Global and efficient self-similarity for object classification and detection. In Proceedings of the IEEE Conference on CVPR, San Francisco, CA, USA, 13–18 June 2010; pp. 1633–1640.
5. Shechtman, E.; Irani, M. Matching Local Self-Similarities across Images and Videos. In Proceedings of the IEEE Conference on CVPR, Minneapolis, MN, USA, 18–23 June 2007; pp. 1–8.
6. Torabi, A.; Bilodeau, G.A. Local self-similarity-based registration of human ROIs in pairs of stereo thermal-visible videos. *Pattern Recognit.* **2013**, *46*, 578–589. [CrossRef]
7. Sedaghat, A.; Ebadi, H. Distinctive Order Based Self-Similarity descriptor for multi-sensor remote sensing image matching. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 62–71. [CrossRef]
8. Mouats, T.; Aouf, N.; Sappa, A.D. Multispectral Stereo Odometry. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 1210–1224. [CrossRef]
9. Li, Y.; Jin, H.; Wu, J. Establishing Keypoint Matches on Multimodal Images with Bootstrap Strategy and Global Information. *IEEE Trans. Image Process.* **2017**. [CrossRef] [PubMed]
10. Aguilera, C.A.; Aguilera, F.J.; Sappa, A.D. Learning cross-spectral similarity measures with deep convolutional neural networks. In Proceedings of the IEEE Conference on CVPR Workshops, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 267–275.
11. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
12. Bay, H.; Tuytelaars, T.; Van Gool, L. Speeded-Up Robust Features. *Comput. Vis. Image Underst.* **2008**, *110*, 404–417. [CrossRef]
13. Chen, J.; Tian, J.; Lee, N. A partial intensity invariant feature descriptor for multimodal retinal image registration. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1707–1718. [CrossRef] [PubMed]
14. Li, Q.; Wang, G.; Liu, J. Robust Scale-Invariant Feature Matching for Remote Sensing Image Registration. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 287–291. [CrossRef]

15. Vural, M.F.; Yardimci, Y.; Temizel, A. Registration of multispectral satellite images with orientation-restricted SIFT. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Cape Town, South Africa, 12–17 July 2009; pp. 243–246.

16. Zhao, D.; Yang, Y.; Ji, Z. Rapid multimodality registration based on MM-SURF. *Neurocomputing* **2014**, *131*, 87–97. [CrossRef]

17. Saleem, S.; Sablatnig, R. A robust sift descriptor for multispectral images. *IEEE Signal Process. Lett.* **2014**, *21*, 400–403. [CrossRef]

18. Aguilera, C.; Barrera, F.; Lumbreras, F. Multispectral Image Feature Points. *Sensors* **2012**, *12*, 12661–12672. [CrossRef]

19. Li, Y.; Shi, X.; Wei, L. Assigning Main Orientation to an EOH Descriptor on Multispectral Images. *Sensors* **2014**, *15*, 15595–15610. [CrossRef] [PubMed]

20. Zhao, C.; Zhao, H.; Lv, J. Multimodal image matching based on Multimodality Robust Line Segment Descriptor. *Neurocomputing* **2015**, *177*, 290–303. [CrossRef]

21. Shi, Q.; Ma, G.; Zhang, F. Robust Image Registration Using Structure Features. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 2045–2049. [CrossRef]

22. Mortensen, E.N.; Deng, H.; Shapiro, L. A SIFT descriptor with global context. In Proceedings of the IEEE Conference on CVPR, San Diego, CA, USA, 21–23 September 2005; Volume 1, pp. 184–190.

23. Kim, S.; Ryu, S.; Ham, B. Local self-similarity frequency descriptor for multispectral feature matching. In Proceedings of the IEEE International Conference on Image Processing, ICIP2014, Paris, France, 27–30 October 2014; pp. 5746–5750.

24. Osadchy, M.; Jacobs, D.; Lindenbaum, M. Surface dependent representations for illumination insensitive image comparison. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 98–111. [CrossRef] [PubMed]

25. Zambanini, S.; Kampel, M. A Local Image Descriptor Robust to Illumination Changes. In Proceedings of the Scandinavian Conference, SCIA 2013, Espoo, Finland, 17–20 June 2013; pp. 11–21.

26. Lades, M.; Vorbruggen, J.; Buhmann, J. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.* **1993**, *42*, 300–311. [CrossRef]

27. Wiskott, L.; Fellous, J.; Kuiger, N. Face Recognition by Elastic Bunch Graph Matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *19*, 775–779. [CrossRef]

28. Mouats, T.; Aouf, N. Multimodal stereo correspondence based on phase congruency and edge histogram descriptor. In Proceedings of the International Conference on Information Fusion, FUSION 2013, Istanbul, Turkey, 9–12 July 2013; pp. 1981–1987.

29. Aguilera, C.A.; Sappa, A.D.; Toledo, R. LGHD: A feature descriptor for matching across non-linear intensity variations. In Proceedings of the IEEE International Conference on Image Processing, ICIP 2015, Quebec City, QC, Canada, 27–30 September 2015; pp. 178–181.

30. Firmenichy, D.; Brown, M. Multispectral interest points for RGB-NIR image registration. In Proceedings of the IEEE International Conference on Image Processing, ICIP2011, Brussels, Belgium, 11–14 September 2011.

31. 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]

32. Oppenheim, A.V.; Lim, J.S. The importance of phase in signals. *Proc. IEEE* **1981**, *69*, 529–541. [CrossRef]

33. Kovesi, P. Phase Congruency Detects Corners and Edges. In Proceedings of the International Conference on Digital Image Computing: Techniques and Applications, DICTA 2003, Macquarie University, Sydney, Australia, 10–12 December 2003; pp. 309–318.

34. Ye, Y.; Shan, J.; Bruzzone, L. Robust Registration of Multimodal Remote Sensing Images Based on Structural Similarity. *IEEE Trans. Geosci. Remote Sens.* **2017**, 1–18. [CrossRef]

35. Ye, Y.; Li, S. HOPC: A Novel Similarity Metric Based on Geometric Structural Properties for Multi-Modal Remote Sensing Image Matching. In Proceedings of the ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Prague, Czech Republic, 12–19 July 2016.

36. Ragb, H.K.; Asari, V.K. Histogram of oriented phase (HOP): A new descriptor based on phase congruency. *SPIE Commer. Sci. Sens. Imaging* **2016**. [CrossRef]

37. Ragb, H.K.; Asari, V.K. Chromatic Domain Phase Features with Gradient and Texture for Efficient Human Detection. *Electron. Imaging* **2017**, *2017*, 74–79. [CrossRef]

38. Fischer, S.; Perrinet, L.; Redondo, R. Self-Invertible 2D Log-Gabor Wavelets. *Int. J. Comput. Vis.* **2007**, *75*, 231–246. [CrossRef]

39. Arróspide, J.; Salgado, L. Log-Gabor Filters for Image-Based Vehicle Verification. *IEEE Trans. Image Process.* **2013**, *22*, 2286. [CrossRef] [PubMed]

40. Fan, D.; Ye, Y.; Pan, L. A remote sensing adapted image registration method based on SIFT and phase congruency. In Proceedings of the International Conference on Image Analysis and Signal Processing, Wuhan, China, 21–23 October 2011; pp. 326–331.

41. Kovesi, P. *Invariant Measures of Image Features from Phase Information*; University of Western Australia: Perth, Australia, 1996.

42. Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of the IEEE Conference on CVPR, San Diego, CA, USA, 20–26 June 2005; pp. 886–893.

43. Saleem, S.; Bais, A.; Sablatnig, R.; Ahmad, A.; Naseer, N. Feature points for multisensor images. *Comput. Electr. Eng.* **2017**, *62*, 511–523. [CrossRef]

44. Rosten, E.; Drummond, T. Machine learning for high-speed corner detection. In Proceedings of the 9th European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; Volume 1, pp. 430–443.

45. Mikolajczyk, K.; Schmid, C. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1615. [CrossRef] [PubMed]

46. Heinly, J.; Dunn, E.; Frahm, J.-M. *Comparative Evaluation of Binary Features*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 759–773.

47. Zhao, W.L.; Ngo, C.W. Scale-rotation invariant pattern entropy for keypoint-based near-duplicate detection. *IEEE Trans. Image Process.* **2009**, *18*, 412–423. [CrossRef] [PubMed]

48. 44RGB/LWIR Dataset. Available online: https://owncloud.cvc.uab.es/owncloud/index.php/s/1Wx715yUh6kDAO7 (accessed on 1 August 2018).

49. 100VS/LWIR Datasets. Available online: http://adas.cvc.uab.es/projects/simeve/index539a.html?q=node/2 (accessed on 1 August 2018).