

Article

Feature Point Matching Method Based on Consistent Edge Structures for Infrared and Visible Images

Qi Wang ¹, Xiang Gao ¹, Fan Wang ^{1,*}, Zhihang Ji ^{1,2} and Xiaopeng Hu ¹

¹ School of Computer Science and Technology, Dalian University of Technology, No. 2 Linggong Road, Dalian 116024, China; wangqi_dut@mail.dlut.edu.cn (Q.W.); gaoxiangv1@mail.dlut.edu.cn (X.G.); jizhihang2000@163.com (Z.J.); xphu@dlut.edu.cn (X.H.)

² Information Engineering College, Henan University of Science and Technology, No.263 Kaiyuan Avenue, Luoyang 471023, China

* Correspondence: wangfan@dlut.edu.cn; Tel.: +86-1569-886-8886

Received: 11 February 2020; Accepted: 25 March 2020; Published: 27 March 2020

Abstract: Infrared and visible image match is an important research topic in the field of multi-modality image processing. Due to the difference of image contents like pixel intensities and gradients caused by disparate spectrums, it is a great challenge for infrared and visible image match in terms of the detection repeatability and the matching accuracy. To improve the matching performance, a feature detection and description method based on consistent edge structures of images (DDCE) is proposed in this paper. First, consistent edge structures are detected to obtain similar contents of infrared and visible images. Second, common feature points of infrared and visible images are extracted based on the consistent edge structures. Third, feature descriptions are established according to the edge structure attributes including edge length and edge orientation. Lastly, feature correspondences are calculated according to the distance of feature descriptions. Due to the utilization of consistent edge structures of infrared and visible images, the proposed DDCE method can improve the detection repeatability and the matching accuracy. DDCE is evaluated on two public datasets and are compared with several state-of-the-art methods. Experimental results demonstrate that DDCE can achieve superior performance against other methods for infrared and visible image match.

Keywords: infrared and visible image match; common feature detection; consistent edge structure; edge orientation histogram

1. Introduction

Infrared and visible image match aims to establish the correspondence of feature points between images formed by different spectrums. Visible images can obtain the fine details of scenes and infrared images can obtain coarse structures of scenes even under the condition of limited light like nighttime and fog [1]. Infrared and visible image match can provide complementary information captured by multi-modality images. Abbas et al. [2] utilize infrared thermal images to measure the temperature of neonates. Beauvisage et al. [3] utilize the infrared image to carry out the night-time navigation. These methods use infrared images to accomplish the tasks that are impossible for visible images. The infrared and visible image match has been widely applied in an unmanned aerial vehicle [3,4], remote sensing satellite [5,6], and security monitoring platform [7,8].

The infrared and visible image match is still a difficult problem even though single-modality image matching methods have been extensively studied [9]. Visible images capture reflected light with spectrum 0.4–0.7 μm and infrared images capture thermal radiation with a spectrum of 0.75–15 μm [1]. Due to different imaging mechanisms, there are significant differences in image contents between infrared and visible images. The differences of image contents include nonlinear intensity

variations [10,11] and local detail discrepancies [12]. Single-modality image matching methods can be divided into two types such as hand craft based methods like SIFT (Scale Invariant Feature Transform) [13] and ORB (Oriented FAST and Rotated BRIEF) [14] and deep learning based methods like LIFT (Learned Invariant Feature Transform) [15] and SuperPoint [16]. Hand craft-based methods are designed according to the prior knowledge on single-modality images. Deep learning-based methods are trained from massive annotation data. Due to the difference of infrared and visible images, hand craft-based methods cannot achieve excellent performance for infrared and visible images [17]. Due to the lack of training data, the deep learning-based method can hardly be trained for infrared and visible images.

The extraction and description of common feature points are the basis of the infrared and visible image match. The infrared and visible image match faces two issues as follows:

(1) *Common feature point extraction*. Due to the difference of pixel intensities and local details like textures between infrared and visible images, some feature points do not have corresponding points in another image [10,12]. Matching of these non-common feature points will only induce the wrong matches.

(2) *Common feature point description*. Due to the difference of pixel intensities and local details between infrared and visible images, local descriptions of common feature points may be dissimilar [10,11]. Matching of these dissimilar common feature points will induce the wrong matches.

Extraction of common feature points plays a critical role in the infrared and visible image match [10,17]. Non-common feature points will cause the low repeatability of detected features. In this case, no matter how well the feature description method is designed, a large amount of wrong matches will be brought in. Most of the existing infrared and visible image matching methods utilize the feature point detection methods of single-modality image matching methods [17]. However, these feature point detection methods cannot extract common feature points effectively for infrared and visible images [10,17]. Although phase congruency has been used to extract feature points, phase congruency aims to solve nonlinear intensity variations of multi-modality images [18,19]. Besides the difference of pixel intensities between infrared and visible images, the differences of local details such as textures will also bring in non-common feature points. The extraction of common feature points needs to fully consider the similarity of global structures of infrared and visible images.

Feature description methods for infrared and visible image match can be divided into two categories: gradient-based methods and edge-based methods [20]. After feature points are detected, feature descriptions need to be established for matching. To handle the difference of infrared and visible images, gradient and edge information are used to build feature descriptions.

Gradient-based methods establish feature descriptions by dealing with the gradient orientation reversal and the gradient magnitude discrepancy caused by nonlinear intensity variations of infrared and visible images [20]. To cope with the gradient orientation reversal, the symmetric-SIFT [21] limits the gradient orientation in the interval $[0, \pi)$ during the construction of descriptions. MM-SURF (Multimodal Speeded Up Robust Features) [22] accumulates Haar responses according to their signs. To cope with the gradient magnitude discrepancy, PIIFD (Partial Intensity Invariant Feature Descriptor) [23] and NG-SIFT (Normalized Gradient SIFT) [24] normalizes gradients to the interval $[0, 1]$. PIIFD establishes symmetric feature descriptions further to overcome the gradient orientation reversal. MN-SIFT (Modified Normalized Gradient SIFT) [25] binarizes gradients to $\{0, 1\}$ to tackle the edge strength variation caused by the gradient magnitude discrepancy of multi-modality images. MSCB (Morphological SIFT and Centric Brief) [26] utilizes the morphological gradient method to obtain consistent gradient images of infrared and visible images and establishes descriptions by brief methods [14]. However, gradient-based methods cannot handle feature description variations caused by the discrepancies of image local details because image gradients are essentially dependent on pixel intensities.

Edge-based methods establish feature descriptions by the image edge orientation according to the consistency of edge structures of infrared and visible images [20]. To compute the edge orientation, multi-oriented Sobel spatial filters are introduced in EOH (Edge Oriented Histogram)

[27], PCEHD (Phase Congruency Edge Histogram Descriptor) [28], HoDMs (Histogram of Directional Maps) [29], and HOSM (Histogram of Oriented Structure Maps) [30]. EOH calculates edge orientation responses by multi-oriented Sobel spatial filters and use the orientation corresponding to the maximum response as the edge orientation. The edge orientation histogram is established as feature descriptions. PCEHD extracts edges and feature points of infrared and visible images by using the phase consistency method and calculates the edge orientation histogram for feature points. Based on EOH, HoDMs uses edge strength responses to describe image textures and HOSM enhances image edges by the guided filter. Multi-scale and multi-oriented Log-Gabor filters are introduced in LGHD (Log-Gabor Histogram Descriptor) [31], MFD (Multispectral Feature Descriptor) [32], and RIDLG (Rotation Invariant Feature Descriptor based on Log-Gabor Filters) [33] to compute the edge orientation. These methods establish the edge orientation histogram by the orientation of the maximum Log-Gabor response. On this basis, RIFT (Radiation Invariant Feature Transform) [34] and MSPC (Maximally Stable Phase Congruency) [35] extract image features by phase congruency. To increase the number of feature points in multi-modal images, RIFT adds edge points as feature points further. Log-Gabor filters can obtain richer feature descriptions than Sobel filters. However, Log-Gabor filter suffers from a high computation burden. Because edge-based methods leverage image structures rather than pixel information, these methods are more suitable for infrared and visible image match than gradient-based methods [31].

However, infrared and visible image matching methods still face the following problems. Gradient-based methods and edge-based methods need to extract common feature points to avoid non-common feature points participating in image matching. Due to the consistency of edge structures of infrared and visible images, edge properties including orientation and length are consistent as well. In the process of establishing feature descriptions by the edge orientation, the edge length should be used to improve the ability of describing the global structure of infrared and visible images further.

In order to overcome these deficiencies of edge-based methods, a feature detection and description method based on consistent edge structures of images (DDCE) is proposed. The main contributions are listed as follows.

- (1) Consistent edges are extracted by selecting long edges to present global structures, which are similar in both infrared and visible images.
- (2) Common feature points are detected according to the constraints of consistent edge structures.
- (3) By using the edge properties including edge length and edge orientation, the edge length weighted edge orientation histogram is computed to build feature descriptions.

The remainder paper is organized as follows. In Section 2, the proposed matching method is described including consistent edge extraction, common feature detection, feature description, and feature matching. In Section 3, experimental results and corresponding analyses are given out. In Section 4, this paper is concluded.

2. Proposed Methods

In this section, DDCE is presented in detail. First, consistent edges of infrared and visible images are extracted to capture the global structures of images. Second, common feature points are detected based on consistent edges. Third, the edge length weighted edge orientation histogram is calculated to build feature descriptions. Lastly, feature correspondences are established according to the obtained feature descriptions.

2.1. Consistent Edge Extraction

Consistent edges are usually the long edges of global structures, which are similar in infrared and visible images [36]. Due to different imaging mechanisms, the consistent edges in infrared and visible images are mainly the long edges that are formed by global structures of images, while the inconsistent edges are mainly the short edges that are formed by local details like textures [12]. Figure 1 illustrates a pair of visible and infrared images and their corresponding edge images. As shown in Figure 1, the long edges of global structures like the building, the crown, and the tent maintain

consistent in infrared and visible images. Since each side of these edges has different reflection and radiation characteristics, these edges can usually be captured in infrared and visible images. The short edges of local details like road textures and leaves are different, as shown in Figure 1. These short edges are a mutual loss or have different position responses in infrared and visible images.

Consistent edges of infrared and visible images are extracted by selecting the long edges of images. First, due to the intense difference between infrared and visible images, a white balance method is utilized to handle infrared and visible images so that they all occupy the maximal possible range [0,255] [37]. Second, histogram equalization is used to enhance the edges of infrared and visible images [38]. Third, the Canny method [39] is used to extract image edges. The key idea of Canny is to use two different thresholds in order to determine which point should belong to an edge: a low threshold $T1$ and a high threshold $T2$. Points with a gradient greater than $T2$ are edge points. Points with a gradient greater than $T1$ and less than $T2$ are edge points if there is a continuous path linking those points to the points with a gradient greater than $T2$. In this paper, the modified Canny method [40] is adopted. The high and low thresholds of the Canny method are set automatically. The detection procedure of the modified Canny method incorporates the same steps as the original Canny method except that $T2$ and $T1$ are set locally. Thresholds ($T2$, $T1$) are determined within a moving window centered on the current pixel. Following the setups used in Reference [40], the window size is chosen to be 20×20 . $T2$ is set to the gradient magnitude that is ranked as the top 30% in the window and $T1$ is set to 40% of $T2$ [40]. The set of the extracted edge is denoted as $E=\{e_1, e_2, \dots, e_M\}$ where M is the number of edges. Lastly, the edge length is calculated by the contour search algorithm [38]. As a result, an edge is denoted as $e_i=\{p_1, p_2, \dots, p_{li}\}$ where p is the edge point and li is the edge length. Consistent edges are generated by selecting edges that are longer than α . The set of consistent edges is denoted as $le = \{e_i \mid |e_i| > \alpha, e_i \in E\}$.

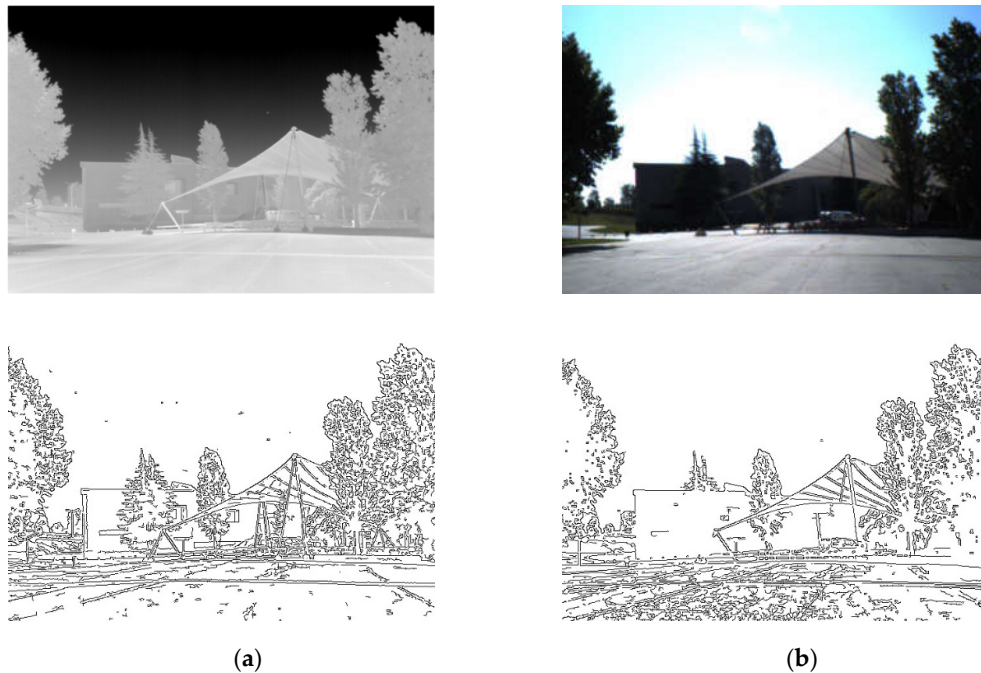


Figure 1. Edge images extracted from infrared and visible images. (a) Infrared image and the corresponding edge image. (b) Visible Image and the corresponding edge image.

2.2. Common Feature Point Detection

Non-common feature points reduce the detection repeatability of feature points. Because local details of infrared and visible images are different from mutual missing and disparate position responses, the feature points extracted on local details are non-common feature points. The performance of feature point matching methods depend on the repeatability of detected feature

points [10,17]. Matching of non-common feature points of images can only bring in the wrong matches, which will bring down the matching performance.

The proposed common feature point detection method utilizes the consistent edges of infrared and visible images to detect feature points. Due to the similarity of consistent edge structures, the percentage of common feature points in the detected feature points can be increased. The common feature point detection method contains two steps. First, candidate common feature points and consistent edges of infrared and visible images are detected. Second, consistent edges are leveraged to check whether the candidate feature points are common feature points or not.

Candidate common feature points are detected by the Harris response [41] of images. Saleem et al. [17] prove by experiments that Harris corners have better repeatability than other feature point detection methods for infrared and visible images. Harris Response R of the point p is computed by the equations below.

$$M(p) = \sum_{(x,y) \in N(p)} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (1)$$

$$R(p) = \det(M(p)) - k \cdot (\text{trace}(M(p)))^2 \quad (2)$$

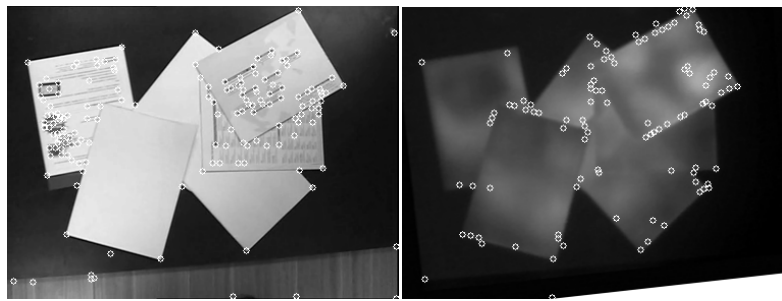
where w is the Gaussian weight function, I_x and I_y are image derivatives, $N(p)$ is the 3×3 neighborhood of point p , and k is a constant in the interval $[0.04, 0.06]$ [39]. Points with response $R > 0$ are marked as candidate feature points [39]. The set of candidate common feature points P' is obtained by local maximum suppression of candidate feature points.

Common feature points of infrared and visible images are obtained by filtering the feature point set P' , according to the constraints of the consistent edges set le . The set of common feature points denoted as P is defined by the equation below.

$$P = \{p_i | p_i \in P' \wedge \exists p_j \in N(p_i) \wedge \exists e_k \in le \wedge p_j \in e_k\} \quad (3)$$

where $N(p)$ is also the 3×3 neighborhood. According to Equation (3), the feature points located on or near the consistent edge are selected as the common feature points. Because the consistent edge reflects the global structure, which is consistent in infrared and visible images, feature points constrained by the consistent edge can be considered common feature points.

An example of common feature point detection is shown in Figure 2. As shown in Figure 2a, the feature points generated by the text on the visible image have no corresponding points on the infrared image. As a result, these feature points are non-common feature points. Matching of these feature points can only get the wrong matches. The feature points extracted by the common feature point detection method are shown in Figure 2b. The number of feature points extracted on the text of visible images is reduced significantly when compared to Figure 2a. The common feature point detection method can improve the matching performance by removing non-common feature points.



(a)

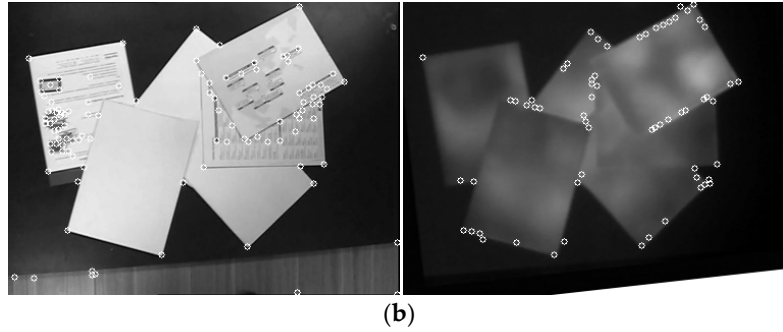


Figure 2. Example of common feature detection. (a) Harris corners. (b) Common feature points.

2.3. Feature Description Establishment

The proposed feature description method leverages the edge length weighted edge orientation histogram to establish feature descriptions based on the global structure of images. Because edges of the global structure are similar in infrared and visible images, feature descriptions need to depict the edge information in the local neighborhood of feature points. The edge length weighted edge orientation histogram can depict the statistics of edges in the local neighborhood of feature points. The steps of establishing feature descriptions based on edge structures include feature point neighborhood partition, edge orientation computation, edge orientation histogram establishment, and histogram normalization.

(1) The local neighborhood of feature points is divided into multiple sub-regions to describe the spatial distribution of edges in the neighborhoods. The size of the neighborhood is 80×80 [27]. Two-layer concentric circles are used to partition the neighborhood as HOG (Histogram of Oriented Gradients) [42]. The outer circle has a radius of r and the inner circle has a radius of $r/2$ where r is 40. Each circle is equally divided by $\pi/9$ radians into 18 sectors. Each of the two adjacent sectors of the same circle are expanded so that their overlap is $\pi/36$. In the case that only edge information is available, which is not as rich as gray and gradients, the neighborhood partition that contains overlapping areas can improve the description ability for feature points [43].

(2) The edge orientation is used to describe the edge information in the neighborhood. During the calculation of the edge orientation, the inconsistent short edges formed by local details are ignored and the consistent long edges formed by global structures are considered. The calculation of the edge orientation is performed by 0° , 45° , 90° , 135° , and non-directional (n.o.) Sobel filters, which are shown in Figure 3 [28]. The edge orientation is calculated for each point p of consistent edges in the neighborhood. The edge orientation of the point p is the orientation of the Sobel filter corresponding to the maximum response value. Let symbols f_i , $i = 1, 2, 3, 4, 5$ denote 0° , 45° , 90° , 135° , and n.o. Sobel filters. The orientation b of the point p can be formulated by the equation below.

$$b(p) = \begin{cases} \operatorname{argmax}_i |I(p) * f_i|, & \exists e_k \in le \wedge p \in e_k \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where $*$ stands for the convolution operation and $||$ stands for the absolute operation.

(3) The edge length weighted edge orientation histogram is used to present the distribution of edges in the neighborhood of feature points. After the edge orientation of each edge pixel in the neighborhood is obtained, the histogram is calculated in each sub-region. Since the long edge has strong discrimination for infrared and visible images, the edge length is utilized in the histogram establishment. In order to enhance feature discrimination, the gradient orientation histogram is weighted by the gradient magnitude in the generation of SIFT and HOG. The length of the edge where the point is located is used as the weight in a similar manner to SIFT and HOG when the edge orientation histogram is calculated. The weight of point p on edge e_i is calculated by the equation below.

$$w(p) = \exp(\min(l_i, l_m)) \delta(l_i \geq \alpha) \quad (5)$$

where l_i is the length of e_i and $\delta(\cdot) = \begin{cases} 1 & \text{true} \\ 0 & \text{false} \end{cases}$. Parameter l_m prevents long edges from suppressing the other edges and dominating the histogram. Parameter l_m is $1/2\max\{H, W\}$ where H and W are image height and width, respectively. Parameter α reduces the effect of inconsistent short edges by only using consistent long edges. The histogram of each sub-region can be formulated by the equation below.

$$h(t) = \sum_p w(p) \delta(b(p) = t) \quad (6)$$

where $t = 1, 2, 3, 4, 5$. The histograms of all sub-regions are connected to obtain a 180-dimensional ($5 \times 18 \times 2$) description vector f .

(4) L2-norm (Euclidean distance) is used to normalize the description vectors. The set of generated feature descriptions is $P\{(p_i, f_i) | i=1, \dots, N\}$ where $f=(h^1, h^2, \dots, h^{180})$.

0°	45°	90°	135°	n.o.
-1	2	-1	-1	-1
-2	2	0	0	0
-1	-1	1	2	1
0	-1	0	-1	0
0	-1	2	-1	0
1	-1	1	-1	-1
2	-1	0	-1	0
1	-1	1	-1	-1

Figure 3. Sobel filters of five orientations at 0°, 45°, 90°, and 135° and no orientation.

2.4. Feature Matching

The bidirectional matching method is utilized to obtain the stable match result [38]. According to the bidirectional matching method, a match is considered valid only if the same pair of feature points in both directions is obtained. Because visible images usually have more content than infrared images, visible images can have more feature points. When searching for similar features of feature points of the visible image in the infrared image, the feature points generated by the extra content of the visible image will only produce the wrong matches. To avoid this situation, the bidirectional matching method searches for similar features of feature points within the infrared image in the visible image. Then similar features of the matched feature points of the visible image are searched in the infrared image.

When determining feature correspondences, the nearest neighbor ratio method is used to select matching points for feature points. The nearest neighbor ratio method is introduced in Reference [13] to improve the matching robustness. When the ratio of the distance of the nearest neighbor over the distance of the second nearest neighbor is less than the specified threshold, the nearest neighbor is regarded as the correct match. The nearest neighbor ratio method is defined by the equation below.

$$\|f_a - f_b\| / \|f_a - f_c\| < r, \quad (7)$$

where f_b and f_c are the nearest and second nearest neighbors of f_a , respectively. Parameter r is set as 0.8 [13]. The distance between two feature descriptions is calculated by the equation below.

$$\|f_a - f_b\| = \sqrt{\sum_i (h_a^i - h_b^i)^2}, \quad (8)$$

where h^i is the histogram bin of the feature description f .

2.5. Algorithm Procedure

The procedure of DDCE is shown in Table 1. Given infrared image I_{ir} and visible image I_{vis} , feature correspondence C of I_{ir} and I_{vis} is output by DDCE. Consistent edge extraction, common feature detection, feature description establishment, and feature matching are described in section

2.1, 2.2, 2.3, and 2.4, respectively. Note that steps (1), (2), and (3) are identical for infrared image I_{ir} and visible image I_{vis} .

Table 1. Algorithm procedure.

Input: infrared image I_{ir} , visible image I_{vis}
Output: feature correspondence C
1. // (1) Consistent Edge Extraction
2. $E\{e_1, e_2, \dots, e_M\} = \text{Canny}(I)$
3. $le = \{e_i \mid e_i > \alpha, e_i \in E\}$
4. // (2) Common Feature Detection
5. $P' = \text{Harris_Corner}(I)$
6. $P = \{p_i \mid p_i \in P' \wedge \exists p_j \in N(p_i) \wedge \exists e_k \in le \wedge p_j \in e_k\}$
7. // (3) Feature Description Establishment
8. $f = \text{Edge_Orientation_Histogram}(P, E)$
9. // (4) Feature Matching
10. $C = \text{Match}(f_{ir}, f_{vis})$

3. Experimental Results

In this section, the matching performance of DDCE is evaluated using visible and infrared images. First, the dataset and evaluation criteria are introduced. Second, images of matching results are given out to illustrate the performance of DDCE on visible and infrared images. Third, matching performance analyses of DDCE are presented quantitatively when compared with state-of-the-art methods. Lastly, running time of DDCE is given out.

3.1. Datasets and Evaluation Criteria

Two public datasets, known as the CVC (Computer Vision Center of Universitat Autònoma de Barcelona) dataset [27] and LWIR (Long Wave Infrared Images) dataset [31], are utilized to validate the performance of DDCE. These two datasets are composed of visible and long wave infrared images. Homography transformations between images are provided by the datasets. Parameter α is utilized in feature detection and description of DDCE. For the CVC dataset and the LWIR dataset, α is set as 20.

Feature detection repeatability, feature matching accuracy, and RANSAC (RANDOM SAMPLE Consensus) [44] estimation result are used to evaluate the matching performance of DDCE. Feature detection repeatability depicts the percentage of repeatable features detected in infrared and visible images. The re-projection error ε of two features with positions p_i and p_j can be expressed by the equation below.

$$\|p_i - Hp_j\| = \varepsilon, \quad (9)$$

where H is the homography transformation. The repeatability of two features is computed by the formula below.

$$C(i, j) = \begin{cases} 1 & \varepsilon < 2 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

where the threshold of the re-projection error ε is 2 pixels [17]. Repeatability is derived from the ratio between the number of repeatable features and the minimum number of features in two images.

$$\text{Repeatability} = \frac{N_a}{\min(\#I_{vis}, \#I_{ir})}, \quad (11)$$

where N_a is the number of all repeatable features.

$$N_a = \sum_i \sum_j C(i, j), \quad (12)$$

and $\#I$ is the number of features detected in image I .

Precision, recall, and the F1 score are used to assess feature matching accuracy. The correspondence of two features can be formulated by the equation below.

$$M(i, j) = \begin{cases} 1 & f_i, f_j \text{ is matched} \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

A pair of points p_i and p_j can be identified as a correct match if they satisfy Equation (7) and Equation (10) simultaneously. Precision is computed by the formula below.

$$Precision = \frac{N_c}{N_c + N_w}, \quad (14)$$

where N_c is the number of correct matches

$$N_c = \sum_i \sum_j C(i, j) \times M(i, j), \quad (15)$$

and N_w is the number of wrong matches.

$$N_w = -N_c + \sum_i \sum_j M(i, j), \quad (16)$$

Recall is computed by the formula below.

$$Recall = \frac{N_c}{N_a}, \quad (17)$$

The F1 score is computed by the equation below.

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall). \quad (18)$$

RANSAC is used to estimate the homography transformation and identifies the correct matches because of the existence of wrong matches. As shown in Reference [28], the matching image formed by identified correct matches is used to display the result of RANSAC estimation.

3.2. Matching Results

3.2.1. Comparison with Single-Modality Image Matching Methods

In this section, a group of matching results on infrared and visible images is given out to show the performance comparison of DDCE against single-modality image matching methods. The single-modality image matching methods used for comparison include SIFT, ORB, LIFT, and SuperPoint. Introductions of these methods are listed in Table 2. This section will present the visual results and performance analyses of DDCE and the matching methods listed in Table 2.

Table 2. Single-modality image matching methods.

Methods	Type
SIFT	Hand Craft Based Method
ORB	Hand Craft Based Method
LIFT	Deep Learning Based Method
SuperPoint	Deep Learning Based Method

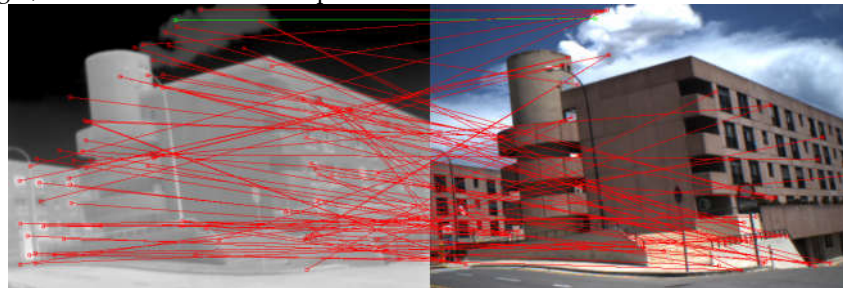
Table 3. The number and percentage of correct matches of single-modality matching methods.

Methods	The Number of Correct Matches	The Percentage of Correct Matches
SIFT	1	0.011
ORB	0	0.000
LIFT	1	0.002
SuperPoint	56	0.144
DDCE	107	0.177

Figure 4 shows the matching results of SIFT, ORB, LIFT, SuperPoint, and DDCE on a pair of infrared and visible images successively. In Figure 4, red lines indicate the wrong matches and green

lines indicate correct matches. To make the matching results easier to see, the number and the percentage of correct matches of each method shown in Figure 4 are listed in Table 3.

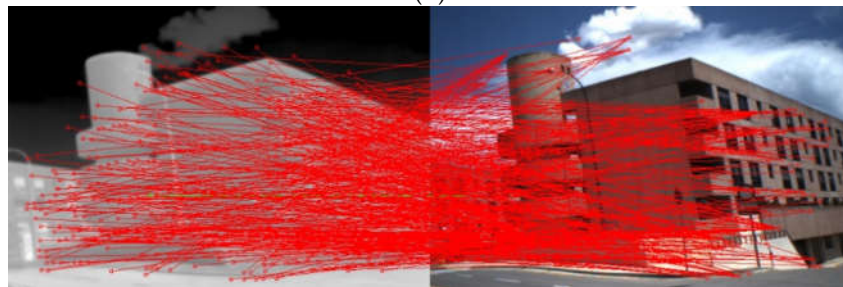
As shown in Table 3, DDCE achieves the largest number and the highest percentage of correct matches. Due to the difference of pixel intensities and gradients between infrared and visible images, SIFT and ORB can only obtain 1 and 0 correct matches, respectively. SIFT and ORB build feature descriptions by gradient orientation and pixel comparison, respectively. Because of the nonlinear intensity variation, SIFT and ORB descriptions for infrared and visible images are dissimilar. LIFT only obtains one correct match. Because the training data of LIFT is obtained from image 3D reconstruction, LIFT aims to tackle the matching of wide baseline visible images. As a result, LIFT cannot handle nonlinear intensity variations between infrared and visible images as well. SuperPoint achieves better matching performance than other single-modality image matching methods. SuperPoint is trained for matching the objects with regular shapes like buildings [16]. Because consistent edge structures reduce the effect of nonlinear intensity variations between infrared and visible images, DDCE achieves the best performance.



(a)



(b)



(c)



(d)

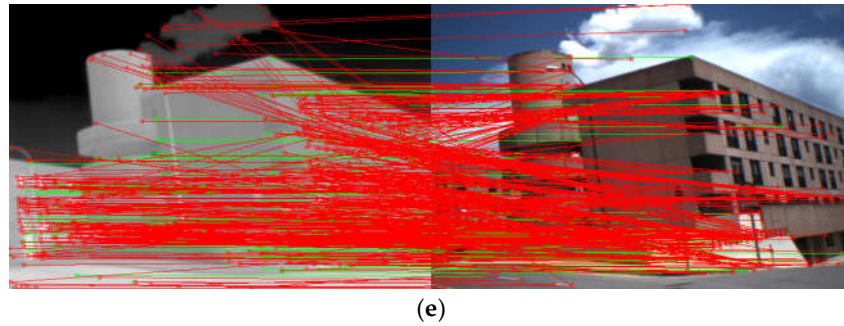


Figure 4. Matching result comparison with single-modality image matching methods. (a) SIFT, (b) ORB, (c) LIFT, (d) SuperPoint, and (e) DDCE. Red lines indicate the wrong matches and green lines indicate correct matches.

3.2.2. Comparison with Multi-Modality Image Matching Methods

In this section, a group of matching results on infrared and visible images is given out to show the performance comparison of DDCE against multi-modality image matching methods. The multi-modality image matching methods used for comparison include PIIFD, MMSURF, PCEHD, LGHD, RIFT, and MSCB. Introductions of these methods are listed in Table 4. This section will present the visual results and performance analyses of DDCE and the matching methods listed in Table 4.

Table 4. Multi-modality image matching methods.

Methods	Type
PIIFD	Gradient-Based Method
MMSURF	Gradient-Based Method
PCEHD	Edge-Based Method
LGHD	Edge-Based Method
RIFT	Edge-Based Method
MSCB	Gradient-Based Method

Figure 5 shows the matching results of PIIFD, MMSURF, PCEHD, LGHD, RIFT, MSCB, and DDCE on a pair of infrared and visible images successively. Similarly, redlines indicate the wrong matches and green lines indicate correct matches. In a similar manner to Section 3.2.1, the number and the percentage of correct matches of each method in Figure 5 are shown in Table 5.

Table 5. The number and percentage of correct matches of multi-modality matching methods.

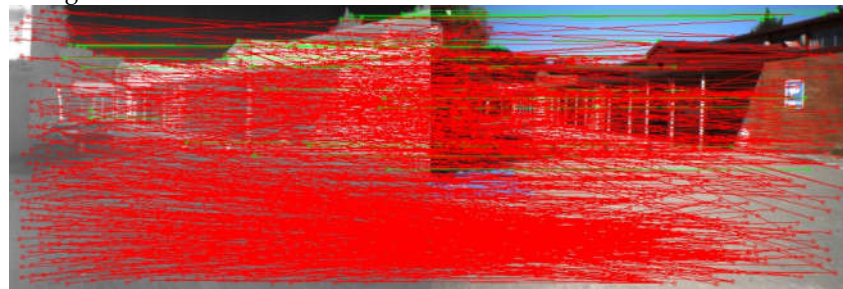
Methods	The Number of Correct Matches	The Percentage of Correct Matches
PIIFD	22	0.027
MMSURF	28	0.085
PCEHD	14	0.173
LGHD	23	0.284
RIFT	255	0.102
MSCB	14	0.052
DDCE	99	0.128

As shown in Table 5, DDCE achieves the second largest number and the third highest percentage of correct matches. Based on the consistency of edge structures, DDCE can achieve superior performance on the number and the percentage of correct matches. According to Table 5 and Figure 5, two conclusions are given out as follows.

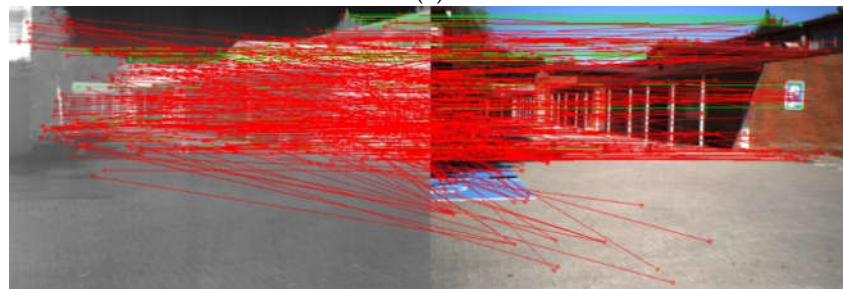
(1) There is a significant difference of gradient information between infrared and visible images. The matching performance of gradient-based methods including PIIFD, MMSURF, and MSCB is worse than edge-based methods. Even though gradient orientation reversal and gradient magnitude

modification are adopted to establish feature descriptions for infrared and visible images, local detail discrepancies can also cause the dissimilarity of feature descriptions. These results indicate that it is difficult to establish similar descriptions through the gradient information.

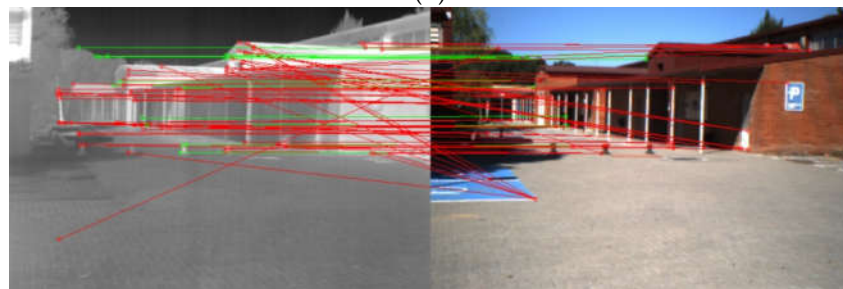
(2) Structure information maintains a certain degree of similarity in infrared and visible images. By using image structure information, edge-based methods including PCEHD, LGHD, and DDCE achieve better matching performance than other methods. Although LGHD and PCEHD achieve higher percentage of correct matches, LGHD and PCEHD obtain fewer correct matches than other methods. RIFT obtains the largest number of correct matches. RIFT extracts dense features of images by detecting corners and edge points. However, the percentage of correct matches of RIFT is lower than the other edge-based methods.



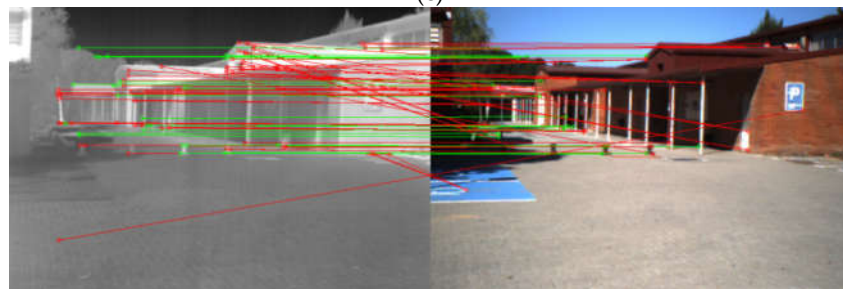
(a)



(b)



(c)



(d)

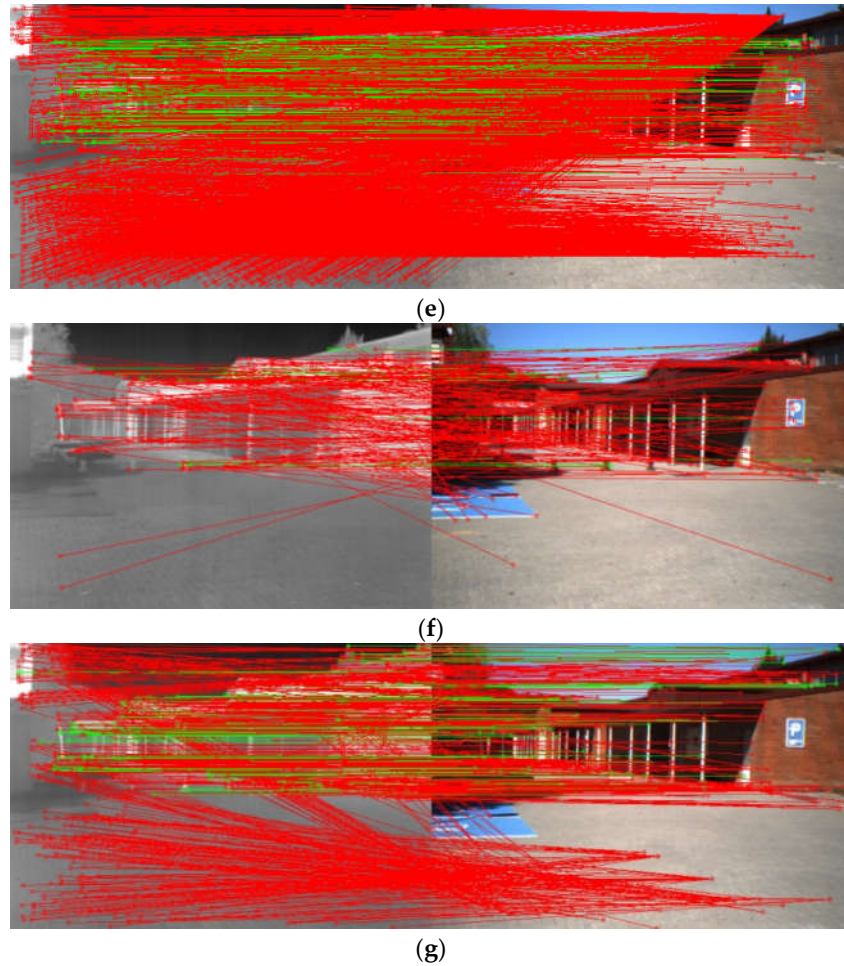


Figure 5. Matching result comparison with multi-modality image matching methods. (a) PIIFD, (b) MMSURF, (c) PCEHD, (d) LGHD, (e) RIFT, (f) MSCB, and (g) DDCE. Red lines indicate wrong matches and green lines indicate the correct matches.

3.3. Matching Performance Analysis

3.3.1. Feature Detection Performance Analysis

Repeatability is used to evaluate the feature detection performance of DDCE and the feature detection methods listed in Table 2. Because multi-modality image matching methods focus on feature descriptions, only single-modality image matching methods listed in Table 2 are adopted for the performance evaluation of feature detection.

The quantitative comparison of the feature detection repeatability of each method is shown in Table 6. DDCE achieves the second highest feature detection repeatability among these methods. Because DDCE extracts common feature points based on consistent edge structures, DDCE achieves the superior feature detection repeatability. SIFT leverages DOG (Difference of Gaussian) to detect image blobs. ORB and DDCE leverage FAST and Harris, respectively, to detect image corners. The repeatability of SIFT is lower than that of ORB and DDCE for long-wave infrared and visible images. This result is consistent with the conclusions presented in the reference [17]. Repeatability of LIFT is lower than the other methods. LIFT essentially extracts image patches that cannot locate feature points accurately. It can be found that some LIFT features are located in smooth regions like chimney in Figure 4. SuperPoint is trained by endpoints of simple artificial geometric shapes during the initialization phase and is extended to real images afterward. SuperPoint can detect objects with regular geometric shapes such as buildings and roads. SuperPoint can achieve excellent detection

repeatability for CVC and the LWIR dataset in which the infrared and visible images contain buildings and roads.

Table 6. Comparison of feature detection.

Method	SIFT	ORB	LIFT	SuperPoint	DDCE
Repeatability	0.204	0.311	0.119	0.582	0.515

3.3.2. Feature Match Performance Analysis

Precision, recall, and the F1 score are used to evaluate the feature matching performance of DDCE and the feature match methods listed in Table 2 and Table 4. Single-modality and multi-modality image matching methods are adopted for the performance evaluation of the feature matching method.

The quantitative comparison of the feature matching accuracy of each method is shown in Table 7. It can be found that multi-modality image matching methods achieve better performance than single-modality image matching methods. Due to the difference of pixel intensities and gradients between infrared and visible images, it is difficult for single-modality image matching methods to match infrared and visible images. To match infrared and visible images, hand craft methods need to be modified according to characteristics of infrared and visible images. Deep learning methods need to be trained by massive training data.

Among multi-modality image matching methods, edge-based methods achieve better performance than gradient-based methods. DDCE leverages consistent edge structures to establish feature descriptions and achieves the second highest F1 score among these methods. Although the similarity of image structures decreases as the spectral difference increases, image structures still maintain better consistency than image pixel intensities and gradients.

Precision and recall of multi-modality image matching methods on infrared and visible images are lower than those of single-modality image matching methods on visible images. Due to the large difference between infrared and visible images, the similarity of descriptions of common feature points is low. Even though common feature points can be detected, the number of common feature points that can be correctly matched is still small. As a result, the match recall of feature points is low. Due to the existence of non-common feature points and the low similarity of common feature points, the match precision of feature points is low as well.

Table 7. Comparison of feature match.

Type	Method	Precision	Recall	F1
single-modality image matching methods	SIFT	0.028	0.024	0.026
	ORB	0.001	0.003	0.002
	LIFT	0.002	0.019	0.004
	SuperPoint	0.022	0.031	0.027
	PIIFD	0.054	0.024	0.033
multi-modality image matching methods	MMSURF	0.082	0.205	0.117
	PCEHD	0.143	0.333	0.200
	LGHD	0.164	0.327	0.218
	RIFT	0.107	0.192	0.137
	MSCB	0.142	0.217	0.172
	DDCE	0.156	0.301	0.205

3.3.3. RANSAC Estimation Performance Analysis

RANSAC estimation is used to evaluate the matching performance of LGHD and DDCE. The performance of RANSAC estimation depends on both the percentage and the number of correct matches [44]. A group of experimental results of LGHD and DDCE is illustrated in Figure 6. It can be found that LGHD achieves a small number of correct matches, as presented in Table 5. As shown in

Figure 6(a), the experimental result of LGHD contains wrong matches indicated by non-horizontal lines. In addition, the building that is the main scene of infrared and visible images is missing. As shown in Figure 6(b), however, DDCE successfully identifies the building as indicated by horizontal lines. Although LGHD achieves a slightly higher percentage of correct matches than DDCE, DDCE obtains a significantly larger number of correct matches than LGHD, which leads to a better RANSAC estimation performance by DDCE.

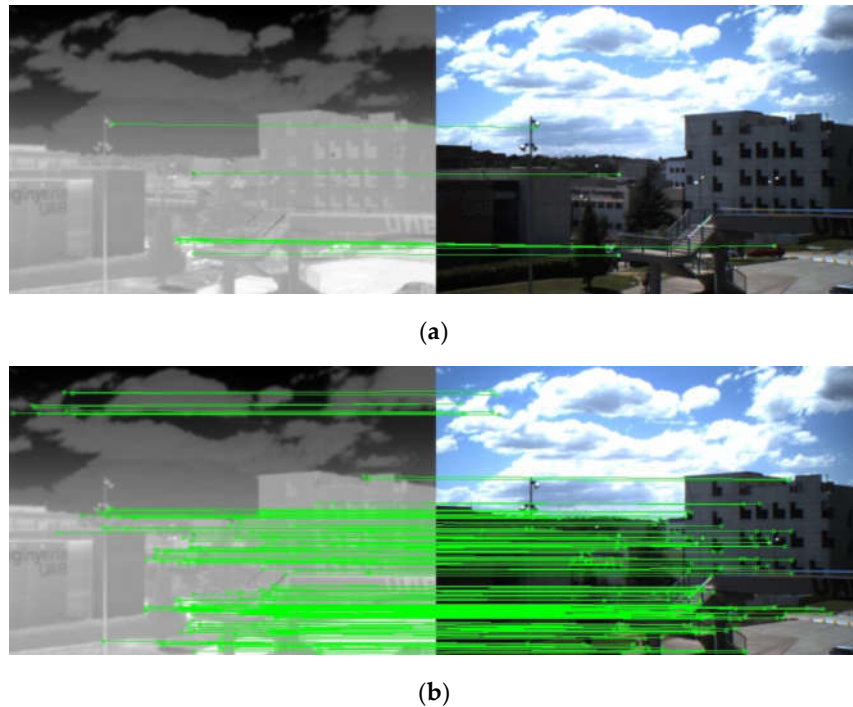


Figure 6. Feature match after RANSAC. (a) Matching Results of LGHD. (b) Matching Results of DDCE.

3.4. Running Time Performance

Table 8 presents the Average Running Time (ART) of the multi-modality image matching methods listed in Table 4. The unit of ART is second. All experiments are performed on a desktop with 64Byte Windows 7 OS. The hardware of the desktop comprises 3.30 GHz Intel i5 4 core CPU and 8G memory. The software of the desktop comprises vs2010 and opencv2.4. DDCE is implemented by C++ and all methods used for comparison are provided by the authors.

As shown in Table 8, DDCE achieves the second best performance on running time. MMSURF is inherited from SURF and achieves the best performance on running time. The methods including PCEHD, LGHD, and RIFT are time-consuming because phase congruency need to be computed.

Table 8. Average running time.

Method	PIIFD	MMSURF	PCEHD	LGHD	RIFT	MSCB	DDCE
ART	6.5	2.4	7.2	8.5	13.5	3.7	2.9

4. Conclusions

In this paper, a feature detection and description method based on consistent edge structures known as DDCE was proposed for infrared and visible image match. First, consistent edge structures were detected to address nonlinear intensity variations and local detail discrepancies of infrared and visible images. Second, common feature points of infrared and visible images were extracted based on the consistent edge structures to improve the repeatability of feature points. Lastly, feature

descriptions were established according to edge attributes including length and direction to enhance the description ability. In order to validate the performance DDCE, two public datasets CVC and LWIR were employed for matching test and several state-of-the-art methods were used for comparison. Experimental results showed that DDCE could achieve superior matching performance compared with PIIFD, MMSURF, MSCB, PCEHD, and RIFT. Although LGHD achieved the highest percentage of correct matches, DDCE could obtain better RANSAC estimation performance than LGHD.

In the future, more infrared and visible images including different kinds of targets and scenes will be acquired under a variety of meteorological conditions. Specific matching strategies for different targets and scenes will be designed to improve the matching reliability of DDCE. DDCE will be modified to be an invariant to rotation and scale. DDCE will be attempted to apply on practical platforms like unmanned aerial vehicles and remote sensing satellites.

Author Contributions: Writing—original draft, Q.W. Writing—review & editing, X.H. and F.W. Software, X.G. and Z.J. All authors have read and agreed to the published version of the manuscript.

Funding: The National Major Special Funding Project, grant number 2018YFA0704605, and 13th Five-Year Major Special Funding Project, grant number 2017ZX05064, funded this research.

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

References

1. Morris, N.J.W.; Avidan, S.; Matusik, W. Statistics of Infrared Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–7.
2. Abbas, A.K.; Leonhardt, S. Intelligent neonatal monitoring based on a virtual thermal sensor. *BMC Med. Imaging* **2014**, *14*, 9.
3. Beauvisage, A.; Aouf, N.; Courtois, H. Multi-Spectral Visual Odometry for Unmanned Air Vehicles. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Budapest, Hungary, 9–12 October 2016; pp. 1994–1999.
4. Khattak, S.; Papachristos, C.; Alexis, K. Visual-Thermal Landmarks and Inertial Fusion for Navigation in Degraded Visual Environments. In Proceedings of the IEEE Aerospace Conference, Big Sky, MT, USA, 2–8 March 2019; pp. 345–357.
5. Wang, P.; Qu, Z.; Wang, P. A Coarse-to-Fine Matching Algorithm for FLIR and Optical Satellite Image Registration. *IEEE Geosci. Remote Sens. Lett.* **2012**, *99*, 599–603.
6. Vural, M.F.; Yardimci, Y.; Temelzel, 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. III-243.
7. Tang, C.; Tian, G.Y.; Chen, X.; Wu, J.B.; Li, K.J.; Meng, H. Infrared and Visible Images Registration with Adaptable Local-global Feature Integration for Rail Inspection. *Infrared Phys. Technol.* **2017**, *87*, 31–39.
8. Zhao, B.; Xu, T.; Chen, Y.; Li, T.; Sun, X. Automatic and Robust Infrared-Visible Image Sequence Registration via Spatio-Temporal Association. *Sensors* **2019**, *19*, 997–1014.
9. Schonberger, J.L.; Hardmeier, H.; Sattler, T. Comparative Evaluation of Hand-crafted and Learned Local Features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6959–6968.
10. Kelman, A.; Sofka, M.; Stewart, C.V. Keypoint Descriptors for Matching across Multiple Image Modalities and Non-linear Intensity Variations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 3257–3264.
11. 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.
12. Ye, Y.; Shan, J.; Bruzzone, L.; Shen, L. Robust Registration of Multimodal Remote Sensing Images based on Structural Similarity. *IEEE Trans. Geosci. Remote Sensing* **2017**, *55*, 2941–2958.
13. Lowe, D.G. Distinctive Image Features from Scale-invariant Key-points. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.
14. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An Efficient Alternative to SIFT or SURF. In Proceedings of the IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2564–2571.

15. Yi, K.M.; Trulls, E.; Lepetit, V.; Fua, P. Lift: Learned Invariant Feature transform. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer, Cham, Switzerland, 2016; pp. 467–483.
16. DeTone, D.; Malisiewicz, T.; Rabinovich, A. Superpoint: Self-supervised Interest Point Detection and Description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018.
17. Saleem, S.; Bais, A.; Sablatnig, R.; Ahmad, A.; Naseer, N. Feature Points for Multi-sensor Images. *Comput. Electr. Eng.* **2017**, *62*, 511–523.
18. Zhang, L.; Li, B.; Tian, L.F.; Zhu, W. LPPCO: A Novel Multimodal Medical Image Registration Using New Feature Descriptor Based on the Local Phase and Phase Congruency of Different Orientations. *IEEE Access* **2018**, *6*, 71976–91987.
19. Fu, Z.; Qin, Q.; Luo, B.; Sun, H.; Wu, C. HOMPC: A Local Feature Descriptor Based on the Combination of Magnitude and Phase Congruency Information for Multi-Sensor Remote Sensing Images. *Remote Sens.* **2018**, *10*, 1234.
20. Liu, X.; Ai, Y.; Tian, B.; Cao, D. Robust and Fast Registration of Infrared and Visible Images for Electro-Optical Pod. *IEEE Trans. Ind. Electron.* **2018**, *66*, 1335–1344.
21. Chen, J.; Tian, J. Real-time Multi-Modal Rigid Registration based on a Novel Symmetric-SIFT Descriptor. *Prog. Nat. Sci.* **2009**, *19*, 643–651.
22. Zhao, D.; Yang, Y.; Ji, Z.; Hu, X. Rapid Multimodality Registration based on MM-SURF. *Neurocomputing* **2014**, *131*, 87–97.
23. Chen, J.; Tian, J.; Lee, N.; Zheng, J.; Smith, R.T.; Laine, A.F. A Partial Intensity Invariant Feature Descriptor for Multimodal Retinal Image Registration. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1707–1718.
24. Saleem, S.; Sablatnig, R. A Robust Sift Descriptor for Multispectral Images. *IEEE Signal Process. Lett.* **2014**, *21*, 400–403.
25. Saleem, S.; Bais, A.; Sablatnig, R. Towards Feature Points based Image Matching between Satellite Imagery and Aerial Photographs of Agriculture land. *Comput. Electron. Agric.* **2016**, *126*, 12–20.
26. Zeng, Q.; Adu, J.; Liu, J.; Yang, J.; Xu, Y.; Gong, M. Real-time Adaptive Visible and Infrared Image Registration based on Morphological Gradient and C-SIFT. *J. Real Time Image Process.* **2019**, doi:10.1007/s11554-019-00858-x.
27. Aguilera, C.; Barrera, F.; Lumbreras, F.; Sappa, A.D.; Toledo, R. Multispectral Image Feature Points. *Sensors* **2012**, *12*, 12661–12672.
28. Mouats, T.; Aouf, N. Multimodal Stereo Correspondence based on Phase Congruency and Edge Histogram Descriptor. In Proceedings of the IEEE Proceedings of the 16th International Conference on Information Fusion, Istanbul, Turkey, 9–12 July 2013; pp. 1981–1987.
29. 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.* **2018**, *16*, 100–104.
30. Ma, T.; Ma, J.; Yu, K. A Local Feature Descriptor Based on Oriented Structure Maps with Guided Filtering for Multispectral Remote Sensing Image Matching. *Remote Sens.* **2019**, *11*, 951.
31. 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, Quebec City, QC, Canada, 27–30 September 2015; pp. 178–181.
32. 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.
33. Chen, H.; Xue, N.; Zhang, Y.; Lu, Q.; Xia, G. Robust Visible-Infrared Image Matching by Exploiting Dominant Edge Orientations. *Pattern Recognit. Lett.* **2019**, *127*, 3–10.
34. Li, J.; Hu, Q.; Ai, M. RIFT: Multi-modal Image Matching Based on Radiation-Invariant Feature Transform. *arXiv* **2018**, arXiv:1804.09493.
35. 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.
36. Li, Y.; Jin, H.; Wu, J.; Liu, J. Establishing Keypoint Matches on Multimodal Images with Bootstrap Strategy and Global Information. *IEEE Trans. Image Process.* **2017**, *26*, 3064–3076.
37. Nicolas, L.; Jose, L.L.; Jean, M.M.; Ana, B.P.; Catalina, S. Simplest Color Balance. *Image Process. Line* **2011**, *1*, 297–315.

38. Laganière, R. *OpenCV 3 Computer Vision Application Programming Cookbook*; Packt Publishing Ltd.: Birmingham, UK, 2017; pp. 119–256.
39. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *8*, 679–698.
40. Simonson, K.M.; Drescher, S.M.; Tanner, F.R. A statistics-based approach to binary image registration with uncertainty analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 112–125.
41. Harris, C.G.; Stephens, M. A combined corner and edge detector. In Proceedings of the 4th Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988; pp. 189–192.
42. Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005; pp. 886–893.
43. Zhao, C.; Zhao, H.; Lv, J.; Sun, S.; Li, B. Multimodal Image Matching based on Multimodality Robust Line Segment Descriptor. *Neurocomputing* **2016**, *177*, 290–303.
44. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* **1981**, *24*, 381–395.



© 2020 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/>).