IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 18, 2019, accepted January 12, 2020, date of publication January 21, 2020, date of current version March 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2968339

Learning Robust Feature Descriptor for Image Registration With Genetic Programming

YUE WU[1], QINGXIU SU[2], WENPING MA[2], SHAODI LIU[2], AND QIGUANG MIAO[1]

[1]Xi'an Key Laboratory of Big Data and Intelligent Vision, School of Computer Science and Technology, Xidian University, Xi'an 710071, China
[2]Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence, Xidian University, Xi'an 710071, China

Corresponding author: Yue Wu (ywu@xidian.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0807500, and in part by the National Natural Science Foundation of China under Grant 61702392.

ABSTRACT The robustness and accuracy of feature descriptor are two essential factors in the process of image registration. Existing feature descriptors can extract important image features, but it may be difficult to find enough correct correspondences for sophisticated images. And these feature descriptors often require domain expertise and human intervention. The aim of this paper is to utilise Genetic Programming (GP) to automatically evolve feature descriptors which are adaptive to various images including remote sensing images and optical images. In this paper, a novel GP-based method (GPFD) is proposed to extract feature vectors and evolve image descriptors for image registration without supervision. The proposed method designs a set of simple arithmetic operators and first-order statistics to construct feature descriptors in order to reduce noise interference. The performance of the proposed method is evaluated and compared against five methods including SIFT, SURF, RIFT, GLPM and GP. These results demonstrate that the feature descriptors evolved by GPFD are robust to complex geometric transformation, the illumination difference and noise.

INDEX TERMS Image registration, genetic programming, feature descriptor, scale-invariant feature transform (SIFT).

I. INTRODUCTION

Image registration is a process of aligning images of the same scene which are acquired under different conditions, such as different times, various viewpoints or different sensors [1]–[3]. It is a fundamental aspect of many problems in image processing, including image mosaic, image fusion [4], transformation detection [5], three-dimensional terrain reconstruction [6], etc. Moreover, many theories and applications are under the assumed premise that the registration has been done well. Hence, the accuracy and the efficiency of image registration directly affect the follow-up applications.

The existing image registration methods are mainly divided into two categories: intensity-based methods and feature-based methods [7], [8]. Intensity-based methods focus on the image's gray information, and align two images by calculating the similarity between pixel intensities of two images. The common measures of similarity are cross correlation (CC) [9] and mutual information (MI) [10]. These methods consider the global information and can obtain accurate results in some cases. But intensity-based methods suffer from monotonous textures [11] and illumination differences. Feature-based methods look for salient features and use the correlation between those features to determine the optimal parameters of the geometric transformation. In general, these features include point features, line features, as well as region features. Comparing with intensity-based methods, the feature-based methods are robust to complex geometric deformations and large illumination differences [12]. Therefore, feature-based methods are more suitable for image registration.

Developing feature descriptors has attracted many researchers and received increasing attention over the past few decades [13]. The commonly used feature descriptors are Gray-Level Co-occurrence Matrix (GLCM) [14], [15], Histogram of Orientated Gradients (HOG) [16], Local Binary Patterns (LBP) [17], [18], Scale Invariant Feature Transform (SIFT) [19], Speeded-Up Robust Features (SURF) [20]. However, most feature descriptors are designed for specific purpose. For example, GLCM and LBP are mainly used to

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar.

texture images, SIFT and SURF are mainly used to keypoints detection. If there is a new task that needs to be addressed, domain experts will be required in tuning parameters or changing policies to get good results. This will cause a lot of inconvenience to our research.

Genetic Programming (GP) is a promising approach which utilises Evolutionary Computation (EC) principles to automatically evolve a program without human intervention and domain knowledge [13], [21]–[28]. In general, EC imitates Darwin's theory of survival of the fittest by natural selection. It randomly generates the initial population as candidate solution, and chooses the individual according to its fitness. The individuals are used to replication, crossover and mutation operation, so as to produce a new population, and continuously cycle to produce better and better approximate solution.

In recent years, GP has been applied to feature extraction, feature construction, region segmentation, image classification, object detection and image registration [13]. GP can input the gray value, statistical information and neighborhood information of the image to the terminal of the program and optimize it according to different purposes of image processing. And GP has achieved success in image registration, though many existing methods are faced with the challenge of low accuracy and recall rate, especially when the image is complex, such as the presence of various noises, different degrees of rotation, illumination different, distortion deformation, etc., most of the methods can not get good matching performance.

In this paper, a novel method, which can extract feature and evolve image descriptors, is proposed. The proposed method is called genetic programming of feature distance (GPFD). GPFD can automatically construct an image descriptor. The descriptor can detect a set of keypoints (such as angles and edges) and extract informative features from these keypoints. A simple set of arithmetic operators and first-order statistics (such as mid-value and standard deviation) are automatically combined into a set of formula representing image descriptors. What is more important, this method does not require human intervention to design keypoints and features, and it uses image pairs with various transformations to evolve image descriptors. To evaluate the performance of the proposed GPFD method, seven image registration datasets with varying degrees of difficulty will be used. The proposed method is compared to five methods including SIFT [19], SURF [20], RIFT [29], GLPM [30] and GP [13].

The main contributions of this paper are as follows: (1) A method of learning feature descriptors is proposed, which is optimized by GP to achieve excellent matching performance. Robust feature descriptors are obtained by using improved first-order statistics (25*th percentile*, 75*th percentile*, *mid* and *stdev*), the noise problem caused by different sensors in remote sensing image is improved. (2) The proposed method is offline learning, which does not require the addition of new information in the optimization process. And take the feature distance of the training set as the threshold, the initial error point pairs (outliers) are removed, thus reducing the human intervention.

The rest of this paper is organized as follows. Section II describes the related work. In Section III, the proposed method is detailed. Section IV introduces the datasets, the parameter settings, evaluation criterion and computational complexity analysis used in the GPFD process. In Section V, experimental results and analysis are presented. Section VI concludes this paper.

II. RELATED WORK

Several classic feature descriptors have been proposed to extract image features for image registration. Lowe [19] proposed a famous algorithm (SIFT) which can extract distinctive invariant features from images and perform reliable matching between different views of the same scene. To improve the computational efficiency of the algorithm, speed-up robust features algorithm (SURF) was proposed [20]. This algorithm does reduce the run time than SIFT, but it is not stable enough when the image has complex geometric deformation. Ong *et al.* [31] proposed scheme is based on the Scale-Invariant Feature Transform (SIFT) feature extraction and Partial Intensity Invariant Feature Descriptors (PIIFD), and they combined with a new method based on the residual-scaled-weighted Least Trimmed Squares (RSW-LTS) to robustly eliminate incorrect putative matches to achieve better registration results. To remove incorrect matches, random sample consensus (RANSAC) was proposed [32], which selects a sample randomly from the consensus set in each iteration and finds the largest consensus set to calculate the final model parameters. Wu *et al.* [33] developed a new point-matching algorithm called fast sample consensus (FSC), which has higher efficiency and stability than RANSAC. Li *et al.* [29] proposed a radiation-invariant feature matching method (RIFT) that is robust to large nonlinear radiation distortions. Ma *et al.* [30] proposed a novel mismatch removal method for robust feature matching of remote sensing images. The key idea of the approach is to preserve the neighborhood structures of potential true matches between two images. Lee *et al.* [34] presented a low-dimensional step pattern analysis (LoSPA) to conventional feature descriptor-based methods, which tailored to achieve low dimensionality while providing sufficient distinctiveness to effectively align unhealthy multimodal image pairs. This method shows high potential in multimodal retinal image registration applications. Chan *et al.* [35] proposed a multi-layered net-based binary descriptor for texture-less object recognition which is called BIND (Binary Integrated Net Descriptor).It provides precise regional object description through a triple-layered net design to encode edges and internal homogeneous spaces into compact rotation-invariant binary strings.

Some deep learning methods and the strategy of optimization are also used for image registration [36], [37]. For example, Wang *et al.* [36] proposed a deep learning framework for remote sensing image registration which pairs patches from

sensed and reference images, and then learns the mapping directly between these patch-pairs and their matching labels for later registration. And the results show that the deep neural network can effectively improve the registration accuracy.

GP is widely used in many applications because of its hierarchical structural expression [13], [21]–[25]. It is especially effective in solving problems of artificial intelligence, machine learning, control and molecular biology. In the past decades, GP has attracted many researchers to deal with image-related problems, such as keypoints detection, feature extraction, feature selection, classification, object detection, image segmentation, image registration and image processing.

Genetic programming is an evolutionary computation algorithm based on natural selection. It provides a way to find computer programs with the best fitness. The general steps of GP include the following.

1) Initialising the Population: Generate an initial population consisting of the function set and the terminal set of the problem. The currently commonly used methods for generating random initial population have the full method, the growth method and the half-and-half method.
2) Fitness Evaluation: Each individual of initial population is given a certain fitness according to the ability to solve problems.
3) Selection,Crossover and Mutation: By performing the following operator operations, a new population is generated.
   - Copy the selected individuals into the new population, the copied individuals are randomly selected based on their fitness.
   - By recombining randomly parts of two selected individuals, new individual is generated, and two selected individuals are randomly selected according to fitness.
   - Individuals are selected randomly on the basis of fitness, and selected parts of the individuals are randomly mutated to produce new offspring of a new population.
4) Termination: Selection, crossover, and mutation are performed until termination conditions are met, take the individual with the highest fitness as the optimal solution.

Song *et al.* [21] is one of the earliest people to apply the GP to sophisticated texture analysis. For the task of texture classification, the experimental results show that GP can get high accuracy in texture classification and demonstrate the effectiveness of GP for texture classification. Texture classifiers can evolve directly based on original pixels without the traditional feature extraction. Therefore, a new pattern of texture classification can be established by GP.

Feature selection and feature construction are data preprocessing techniques used to reduce physical memory space and improve the effectiveness of algorithm. GP is introduced to construct feature and select informative features on high-dimensional classification problems [22]. The final features contain a lot of useful information and maintain the the accuracy of classification in most cases. However, it may exist the issue of overfitting when the data has a skewed distribution with many outliers.

Chen *et al.* [23] proposed a new feature selection method, which is named genetic programming with permutation importance (GPPI). GPPI is designed to select features for high-dimensional symbolic regression (SR). The regression results confirm that GPPI is effective in choosing important features along with the presence of relevant features.

Transfer leaning is a type of machine learning approach, it can be used to solve complex tasks. Iqbal *et al.* [24] proposed a novel method based on transfer learning and GP to extract and reuse blocks of information for image classification. And reuse information extracted from similar or different problems is able to improve algorithmic performance on sophisticated image classification problems.

Hindmarsh *et al.* [25] combined Scale-Invariant Feature Transform(SIFT) feature and GP to improve the performance of object recognition. SIFT features can alleviate problems with complex examples involving scale, rotation, and light variations. GP can find the optimal way of describing the keypoints extracted by SIFT algorithm. The experimental results show that the method combining SIFT features with GP achieves good classification performance.

A dynamic GP approach is proposed [13] to evolve rotation-invariant texture image descriptors. Al-Sahaf *et al.* used arithmetic operators and improved first-order statistics to synthesising a set of formulae. The formulae can be utilised to automatically construct a rotation-invariant image descriptor by GP. This method needs only two training instances per class, and reduces the overall complexity of the system in terms of time and physical computer memory. The analysis reveals that the evolved image descriptor can significantly outperform the hand-crafted descriptors. However, this method has some limitations, the descriptor fails to tackle the noise problem.

Motivated by the promising GP method, a novel GPFD method is proposed to extract feature vectors and evolve image descriptors to image registration. The method designs a set of simple arithmetic operators and first-order statistics in order to reduce noise interference. And it is robust and effective which is able to handle complex image transformation.

III. THE PROPOSED METHOD

The proposed method is described in this section. To make this paper comprehensive, and this section provide essential details, the overall algorithm is introduced first, then the function set and the terminal set are illustrated, the fitness function is finally described.

A. OVERALL ALGORITHM

To achieve the goal of our method, first a set of putative matches are constructed. Then, GPFD is used to remove the

false matches contained in the putative set. There are two questions, how to build the set of putative matches and how to generate the test set and the training set. Fortunately, there are several well-designed feature descriptors (e.g., SIFT [19]) can efficiently establish the putative correspondences between two images. Therefore, SIFT operator is briefly introduced to constructing the training set and the test set. Euclidean distance of SIFT feature is calculated, then nearest neighbor distance ratio (NNDR) strategy is used to build correspondences. For each feature point in one image, the method searches the two nearest neighbors in the other image and compares their distances, which is defined as,

$$d_{ratio} = \frac{d_{nearest}}{d_{second\ nearest}} \tag{1}$$

where $d_{nearest}$ is the distance between a certain feature point in one image and its nearest neighbor in the other image, $d_{second\ nearest}$ is the distance between a certain feature point in one image and its second nearest neighbor in the other image. Obviously, when d_{ratio} is equal to 1, all correct correspondences and incorrect correspondences are included, which construct a set of putative matches. And we can obtain the training set. We are able to get the test set by setting a smaller appropriate threshold.

Using the given function set and terminal set, GPFD is able to evolve the individual with the best fitness by selection, crossover and mutation. The individual is called image descriptor. And the terminal set is obtained by the following steps. Take 40 × 40 image blocks around a set of putative matches and calculate the 25*th percentile*, 75*th percentile*, *mid* and *stdev* of each pixel with a 5 × 5 pixels window. 5 × 5 pixels window is used to calculate terminals. There are 25 numbers and sort these numbers, the sixth number is 25*th percentile*, the thirteenth number is *mid*, the twentieth number is 75*th percentile*, and the numbers between 25*th percentile* and 75*th percentile* are used to calculate *stdev*. Notice that 40 × 40 image blocks and 5 × 5 pixels window are all obtained via a lot of experiments. Then the feature vector of each image block is obtained by using the image descriptor of the tree structure in the test set, and the feature distance is calculated. At the same time, the maximum feature distance of the training set is calculated as the threshold. In the test set, the point pairs less than the threshold consist of the correspondence set. The maximum feature distance of different image pair may be different, so using threshold can effectively remove the majority of the incorrect point pairs of different image pair. But few inaccurate correspondences still exist. A classic approach called random sample consensus (RANSAC) [32] will be adopted to remove imprecise point pairs mainly because the algorithm works well when the outliers are smaller than 10%. So we utilize RANSAC to get high registration accuracy. The final results can be obtained through the above steps.

The overall process is shown in Fig. 1-2.

The proposed method is mainly divided into two parts, training and testing. The training process of the overall algorithm is shown in Fig. 1. Firstly, using SIFT [19] to get tentative training set, and selecting manually some point pairs. Then, calculate the first-order statistic of each image block. Finally, getting the optimal image descriptor by the GPFD process, and then obtain the maximum feature distance. The test process is similar to the training process in the previous steps. As shown in Fig. 2, the test process does not need to select point pairs. After calculating the first-order statistic of the image block, using the descriptor to get the feature vector of each image block. And the point pairs whose feature distance is smaller than the maximum feature distance obtained during the training process consist of the final correspondence set.

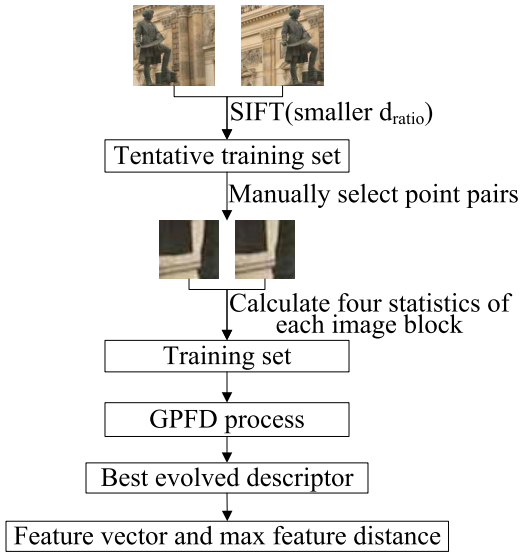


FIGURE 1. The training process of the overall algorithm.

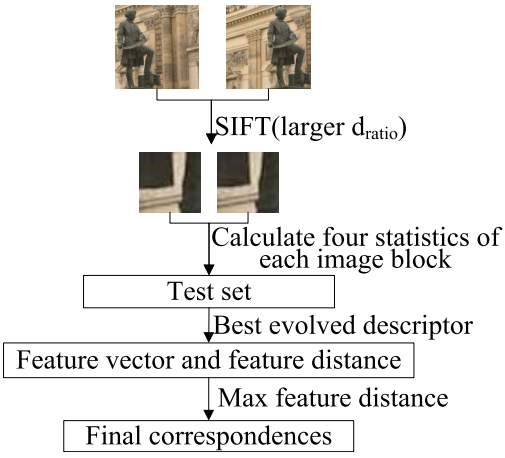


FIGURE 2. The test process of the overall algorithm.

The pseudo-code of the proposed method is shown in Algorithm 1.

B. FUNCTION SET AND TERMINAL SET

In order to keep the individual structure simple, the function set in GPFD consists of five functions. Four of these functions

Algorithm 1 Procedure of GPFD

Input: The reference image and the sensed image.

Output: The model transformation parameter and correspondence set C_{final}.

1: The training sets and test sets are constructed by SIFT algorithm.
2: Calculate the statistics around the point (the terminal set of GPFD).
3: Use the terminal to generate population by Rampled-half-and-half.
4: Calculate the fitness value of each descriptor(individual) using eq. (3), then save the best descriptor.
5: By selection, crossover and mutation to get a new population.
6: Go to step 2 if the termination criterion is not met.
7: Best evolved descriptor.
8: Get the feature vector and feature distance of the training sets and the test sets.
9: In the test sets, removing most outliers using the feature distance, and get the correspondence set C_{final}.
10: Calculate the model transformation parameters by least square method (LSM).

are the arithmetic $+$, $-$, $\times$, / operators. The / function will return 0 if the numerator is 0, which is vitally important to avoid the existing of a "division by zero" exception. The fifth of functions is the *code* function, this function makes all negative values are replaced with 0, and all positive and zero values are replaced with 1. Therefore, code nodes cannot appear anywhere except at the root of a program tree.

The terminal set in GPFD consists of the 25*th percentile*, 75*th percentile*, *mid* and *stdev* nodes. Each node performs simple first-order statistics on a set of values. Arrange all numbers from small to large and divide them into four equal parts. The numbers at the three dividing points are the 25*th percentile*, *mid* and 75*th percentile*, respectively. The *stdev* node returns standard deviation between 25*th percentile* and 75*th percentile*. The terrible situation can be avoided as much as possible by selecting 25*th percentile*, 75*th percentile*, *mid* and *stdev*, when some images have complex noises such as remote sensing images.

Example of an evolved GPFD program is shown in Fig.3.

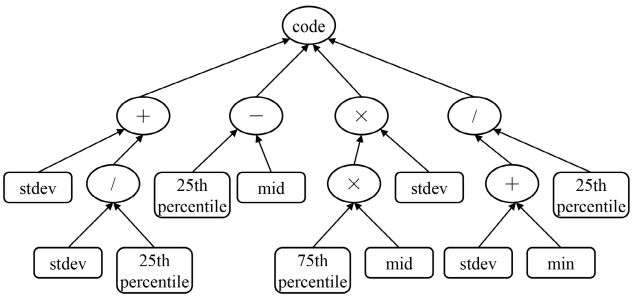


FIGURE 3. Example of an evolved GPFD program.

C. FITNESS FUNCTION

The degree of adaptation of individuals in a population depends on the degree to which they approach the real solution. Therefore, it is necessary to choose a suitable fitness measure that GPFD mechanism can identify image blocks of the same class. Measuring the distance between the feature vectors is the most common approach and is used in the proposed GPFD method. The χ^2 [13] measures the distance between two normalised feature vectors. The two vectors must be normalised and have the same number of elements. The χ^2 is defined as follows.

$$\chi^2 = \frac{1}{2}\sqrt{\frac{1}{m}\sum_{i=1}^{m}\frac{(\vec{x}_i - \vec{y}_i)^2}{(\vec{x}_i + \vec{y}_i)}} \tag{2}$$

where $\vec{x}$ and $\vec{y}$ are vectors, m is the number of elements, and $\vec{x}_i$ is the *ith* element of $\vec{x}$.

The fitness function [13] evaluates the similarity between image blocks and is defined as follows.

$$fitness = \frac{1}{1 + \exp^{-5(W_d - B_d)}} \tag{3}$$

where W_d and B_d are the within-class and between-class respectively. The aim is to seek out a tradeoff between W_d and B_d. That is to say, minimising the distance between two matched image blocks and maximising the distance except two matched image blocks.

IV. EXPERIMENT DESIGN

The experimental design is discussed in this section including the datasets, the parameter settings, evaluation criterion and computational complexity analysis.

A. DATASETS

To evaluate the performance of the feature descriptors by the proposed GPFD method, seven different image registration datasets are used. The part of image pairs are shown in Fig.4.

1) Multi-View Stereo: The dataset includes three 3072 × 2048 images of same places [38]. We choose one scene from a image and create two images for evaluation. The two images are simply rigid transformations.
2) CIAP: The image pair is size 700 × 700 and has already been rectified [30]. The two images are simply rigid transformations. The feature matching tasks for this kind of image pair usually appears in the image Mosaic problem. The images are public available (from the Erdas example data) and were captured over eastern Illinois, IL, USA.
3) VGG: The dataset contains 40 image pairs captured in a fixed position by a planar scene or camera during the acquisition process [39]. Therefore, the image pairs in this dataset always obey homography. The ground truth homographies are provided by the dataset. The images are simple transformations.
4) DTU: The dataset was originally designed for multi-view stereo evaluation, involving many different

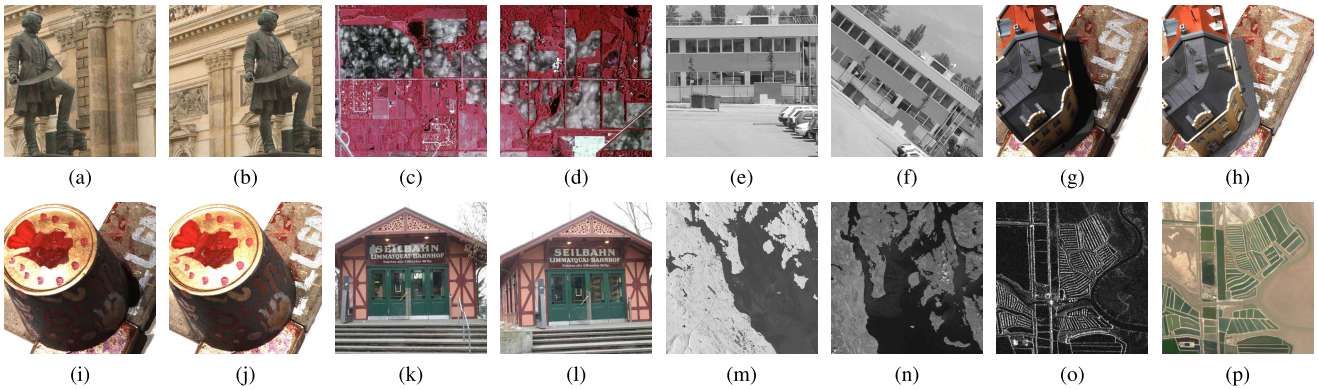


FIGURE 4. Image pairs. (a) (b) Multi-View Stereo. (c) (d) CIAP. (e) (f) VGG. (g) (h) DTU. (i) (j) (k) (l) ZuBuD. (m) (n) SAR. (o) (p) SAR and optical.

scenarios and a wide range of objects [40]. Each scene was shot from 49 or 64 locations, and the position of the ground true-image camera and the internal camera parameters were both highly accurate. We select two scenes from the dataset: 131 images were created to evaluate the image, including a image pair with a large change of viewpoints in the scene.

5) ZuBuD: The ZuBuD image database contains over 1005 images about Zurich city building.[1] Each scene contains five images with complex transformations between different images.
6) SAR: The seventh image pair contains two SAR image of size 800 × 800. The reference image is from the HH mode (L-Band) of a scene taken by the sensor ALOS-PALSAR on June 5,2010, at the region of Campbell River in British Columbia (with an initial spatial resolution of 15m resampled to 30). And the sensed image is from band 5 (1.55-1.75) of a scene taken by the sensor Landsat Enhanced Thematic Mapper Plus (ETM+) on June 26, 1999, at the same region (with a spatial resolution of 30m). The size of this pair of images is 800 × 800 pixels.
7) SAR and optical: The eighth image pair contains a SAR image of size 400 × 400 × 1 and an optical image of size 400 × 400 × 3. SAR image is the C-band of Radarsat-2 image acquired in June 2008, while optical image was acquired from Google Earth in September 2012, including red, green and blue bands. The dataset records the change of land use in ShuGuang village, Dongying city. The spatial resolution is 8 meters.

These datasets contain almost all the different transformations, including translation, rotation, scaling, illumination intensity, and various noise. And it is sufficient for GPFD to evolve image descriptors that are suitable for different types of registration images.

B. PARAMETER SETTINGS

The implementation of all methods including the proposed GPFD method always need suitable parameter. Parameters are vital important to algorithmic performance such as precision, time complexity. GPFD has a random initialization mechanism, so the proposed method is executed 30 times independently, and the average result of 30 times is taken as the final result. A brief summary of the GPFD parameters are shown in Table 1. It is worth mentioning that these parameters are determined by a large number of experiments, so as to better maintain the accuracy and efficiency.

TABLE 1. GP parameter settings.

Parameter	Value	Parameter	Value
Generaions	50	Crossover Rate	0.8
Population Size	100	Mutation Rate	0.2
Minimum Depth	2	Maximum Depth	10
Selection Type	Tournament	Reproduction	Keep the best
Tournament Size	5	Initial Population	Half-and-half

C. EVALUATION CRITERION

The number of correct correspondences and accuracy are important criteria to evaluate the effectiveness of the proposed method. For our experiment results, the evaluation carries out using the following criteria.

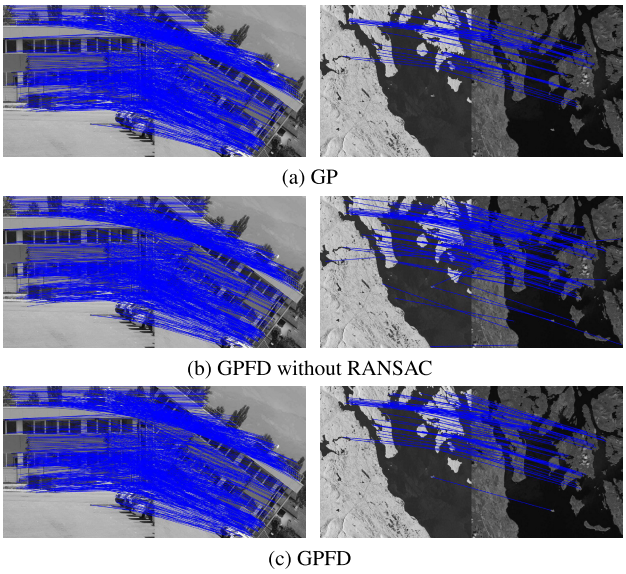
1) Root-mean-square error (RMSE): The accuracy is evaluated by the root-mean-square error (RMSE) criterion. The RMSE can be expressed as follows:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}((X_i - X_i')^2 + (Y_i - Y_i')^2)} \quad (4)$$

where (X_i, Y_i) and (X_i', Y_i') are the coordinates of the ith matching keypoint pair, (X_i', Y_i') is the transformed coordinate, m is the total number of the matching points.

2) Number of correct matches (NCM): The number of correct correspondences is used as the criterion to evaluate the robustness of the proposed GPFD method.

[1] Available at: http://www.vision.ee.ethz.ch/datasets/index.en.html



(a) GP

(b) GPFD without RANSAC

(c) GPFD

FIGURE 5. Match results for two image pairs.

3) Precision (P): Precision is defined by the ratio of correct matches and the sum of correct matches and false matches. It can be expressed as follows:

$$P = \frac{correct\ matches}{correct\ matches + false\ matches} \tag{5}$$

4) Recall rate (R): Recall rate is the number of correctly matched points with respect to the number of corresponding points between two images of the same scene. Correspondences are the total matches that satisfy the transformation matrix within the range of allowable threshold.

$$R = \frac{correct\ matches}{correspondences} \tag{6}$$

D. COMPUTATIONAL COMPLEXITY ANALYSIS

In contrast with the SIFT algorithm, our method mainly increases two step: training descriptor and testing image pair. The computational complexity analysis is described as follows. We use following formula to approximately approach spending time:

$$T_{total} = T_{SIFT} + T_{Train} + T_{Test} \tag{7}$$

T_{SIFT} is the time to calculate the matching points of original images, T_{Train} is the time to train descriptor, and T_{Test} is the time to test image pair. The registration method based on SIFT consists of scale-space extreme detection, keypoints localization, orientation assignment, keypoints descriptor. For an image ($N \times N$ pixels) that has a statistical region with a size of $M \times M$ pixels, its computational complexity is denoted as $O(N^2M^2)$ integrally. For the training process, assuming that depth of the tree is n, that is to say, the length of the descriptor is 2^n, its complexity can be represented as $O(2^n)$. Finally, the complexity of testing image pair is denoted as $O(L^2)$ (L is the number of test set).

TABLE 2. Compare of RMSE, Number of correct matches, Precision and Recall rate of test image pairs (–: without the use of RANSAC).

Methods		GP	GPFD–	**GPFD**
Image Pair 3	RMSE	0.5234	0.5970	**0.5192**
	NCM	807	828	**824**
	Precision	0.9913	0.9915	**0.9976**
	Recall rate	0.8909	0.9143	**0.9153**
Image Pair 7	RMSE	0.7199	0.9573	**0.5246**
	NCM	44	87	**72**
	Precision	0.7728	0.8506	**1**
	Recall rate	0.2982	0.6491	**0.6316**

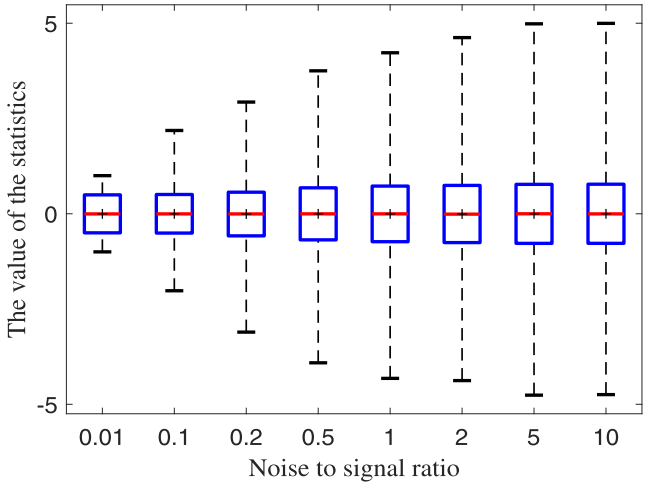


FIGURE 6. Noise robustness test.

V. RESULTS AND DISCUSSIONS

The proposed GPFD method is tested on seven datasets. And the experimental results and analyzes are presented in this section.

The GPFD is compared to five methods including SIFT [19], SURF [20], RIFT [29], GLPM [30] and GP [13]. It is worth mentioning that since the GPFD uses RANSAC to eliminate a small number of mismatches, and to be fair, RANSAC is used in all contrast methods. GPFD has a random initialization mechanism, so the proposed method is executed 30 times independently, and the average result of 30 times is taken as the final result. The root-mean-square error (RMSE), the number of correct matches (NCM), precision and recall rate are used to evaluate the performance of the proposed method.

A. ROBUSTNESS TEST AND DISCUSSIONS

In order to show the effectiveness of the GPFD algorithm more clearly, the results of the two image pairs are shown in Fig. 5.

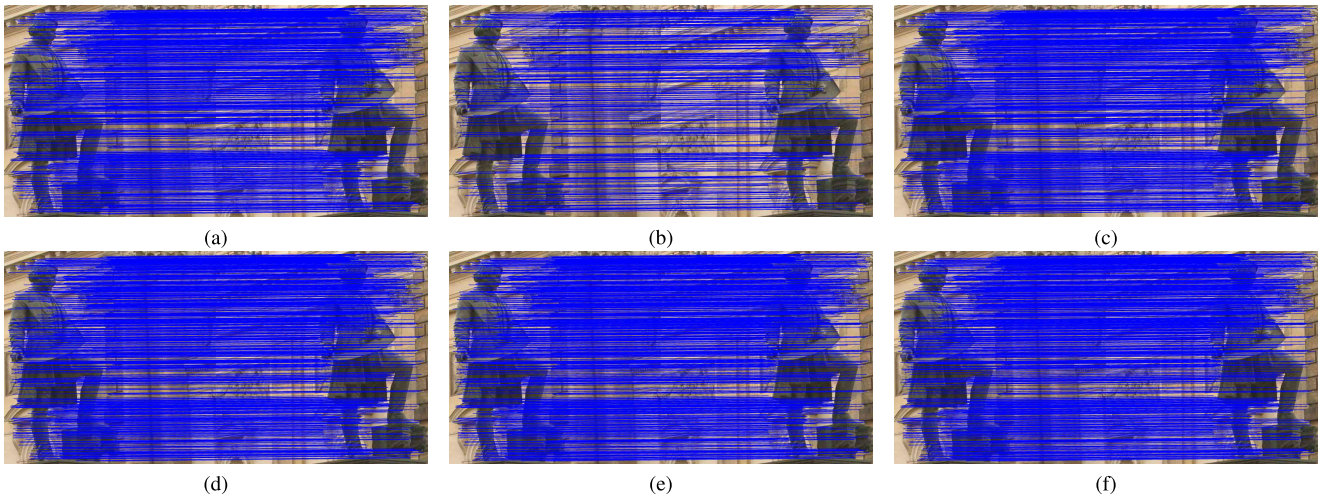


FIGURE 7. Matching results of image pair 1. From top to bottom and left to right: (a) SIFT (0.9963, 0.8687), (b) SURF (0.9962, 0.4230), (c) RIFT (0.9673, 0.8466), (d) GLPM (0.9920, 0.9341), (e) GP (1.0000, 0.9740), (f) GPFD (1.0000, 0.9708). For each algorithm, the first value is the precision, while the second value is recall rate, i.e., (Precision, Recall rate). The two endpoints of each blue line correspond to the positions of feature points in two images, indicating the correct match.

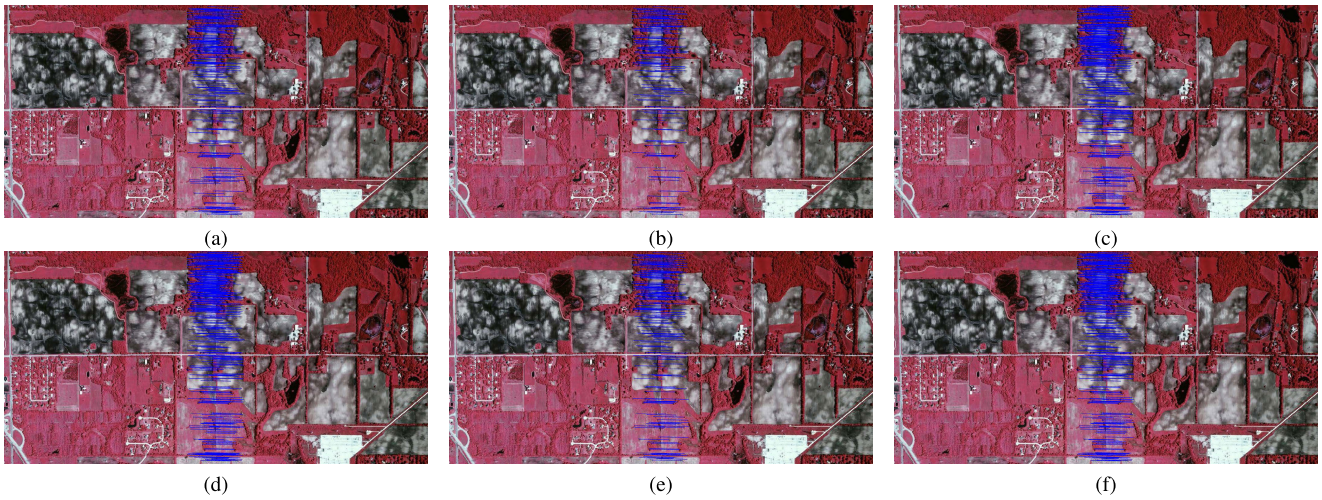


FIGURE 8. Matching results of image pair 2. From top to bottom and left to right: (a) SIFT (1.0000, 0.8379), (b) SURF (1.0000, 0.5379), (c) RIFT (0.9782, 1.0000), (d) GLPM (1.0000, 1.0000), (e) GP (1.0000, 0.9586), (f) GPFD (1.0000, 1.0000). For each algorithm, the first value is the precision, while the second value is recall rate, i.e., (Precision, Recall rate). The two endpoints of each blue line correspond to the positions of feature points in two images, indicating the correct match.

As shown in the Table 2 and Fig. 4, the GPFD obviously obtains more matches and more efficient results than GP. To be more convincing, we employ the GPFD method without RANSAC as a comparison and confirm that the image descriptor evolved by GPFD is robust to various noise. GP performs well for simple optical image, but it will be significantly affected in the presence of noise. There are lots of speckle noise between image pair 7. Due to the speckle noise, some correct matches cannot be found and some pseudo matches are regarded as correct matches. This directly affects the matching performance of GP. While GPFD always achieves good registration performance. Because the difference between GP and GPFD is terminal set, the terminals of GP are the *max*, *min*, *mean* and *stdev*. And the terminals of GP are easily influenced by noise. A simple experiment is used to verify this statement. Randomly generate 1000 numbers between -1 and 1, add the gaussian noise with the noise to signal ratio of 0.01, 0.1, 0.2, 0.5, 1, 2, 5, 10, and compare the stability between the statistics. The noise robustness test results are shown in Fig. 5.

In Fig. 6, the red line is the *mid*, the black + is the *mean*, the top and bottom of box are the 25*th percentile* and 75*th percentile*, respectively, the top and bottom of figure are the *max* and *min*, respectively. It is evident that the *max* and *min* fluctuate greatly after adding the noise of normal distribution. While the 25*th percentile* and 75*th percentile* are almost impervious to noise. The experiment explains that 25*th percentile*, 75*th percentile*, *mid* are very representative

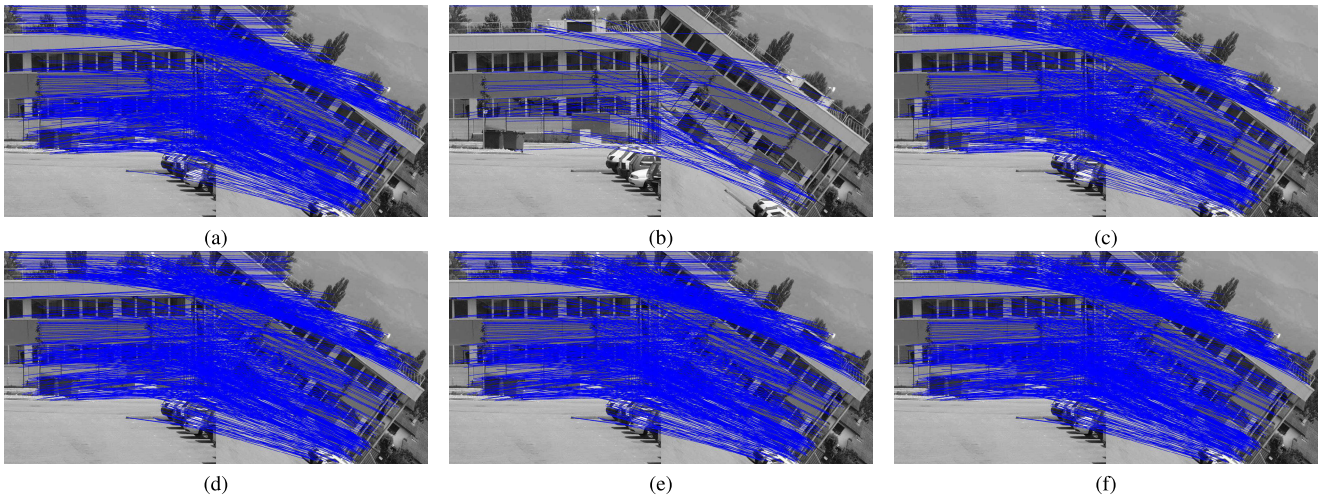


FIGURE 9. Matching results of image pair 3. From top to bottom and left to right: (a) SIFT (0.9827, 0.8363), (b) SURF (0.9223, 0.1058), (c) RIFT (0.9276, 0.6704), (d) GLPM (0.9897, 0.8574), (e) GP (0.9913, 0.8909), (f) GPFD (0.9976, 0.9153). For each algorithm, the first value is the precision, while the second value is recall rate, i.e., (Precision, Recall rate). The two endpoints of each blue line correspond to the positions of feature points in two images, indicating the correct match.

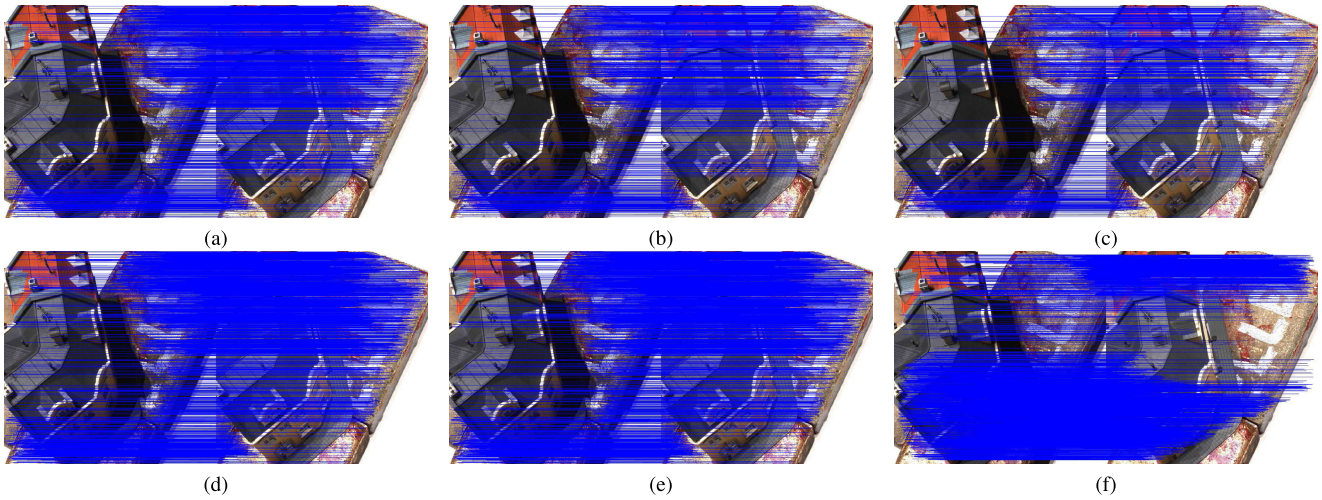


FIGURE 10. Matching results of image pair 4. From top to bottom and left to right: (a) SIFT (0.8817, 0.4247), (b) SURF (0.9900, 0.3165), (c) RIFT (0.9872, 0.2360), (d) GLPM (0.9927, 0.7193), (e) GP (0.9336, 0.7054), (f) GPFD (0.9996, 1.0000). For each algorithm, the first value is the precision, while the second value is recall rate, i.e., (Precision, Recall rate). The two endpoints of each blue line correspond to the positions of feature points in two images, indicating the correct match.

and robust first-order statistics, and can be used to evolve robust image descriptors.

B. MATCHING PERFORMANCE TEST AND DISCUSSIONS

We compare the proposed method with the SIFT [19], SURF [20], RIFT [29], GLPM [30] and GP [13] algorithms, and to be fair, RANSAC is used in all contrast methods. SIFT is a most representative method. SIFT and SIFT-based methods like SURF have been widely used in image registration. RIFT is suitable for a variety of multi-modal images owing to its rotation invariance. GLPM is a novel mismatch removal method for robust feature matching. And GP is a widely used evolutionary algorithm. Experimental results show that the proposed GPFD method can get better performance than compared methods for various image registration task.

The matching results for eight image pairs are shown in Fig. 7-14. And the root-mean-square error (RMSE), the number of correct matches (NCM), precision, recall rate and running time are listed in the Table 3. Owing to the adoption of stable first-order statistics, our method can get better result, but the running time is large.

It can be seen that the proposed method achieves significantly better or similar performance on almost all evaluation criteria than other methods. Feature descriptor shows good results on test samples. This means that the evolved descriptors are suitable for many types of image pairs. And detailed analysis is described in the following.

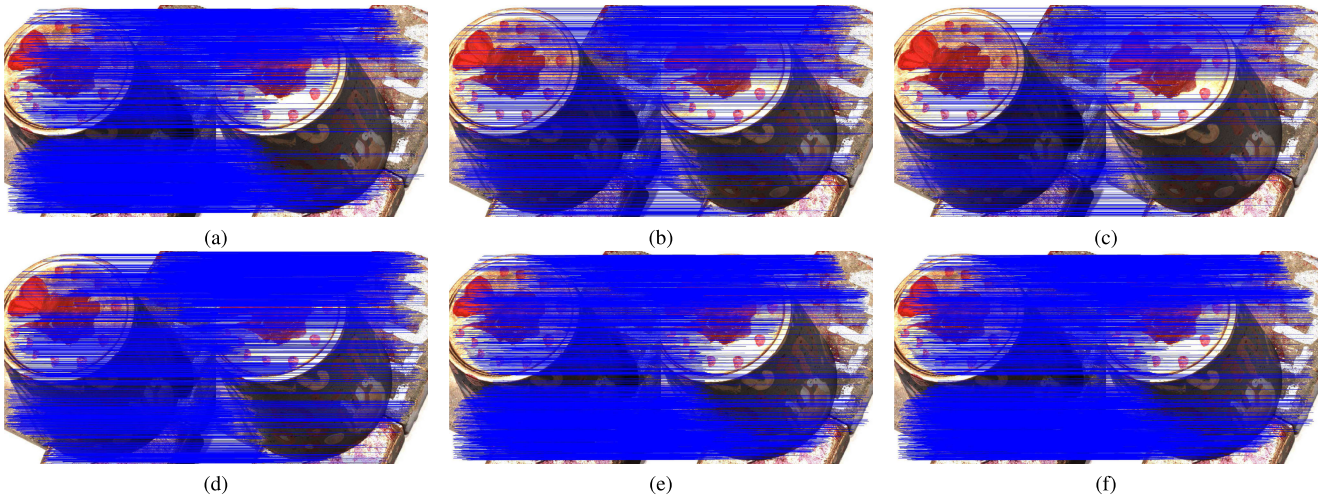


FIGURE 11. Matching results of image pair 5. From top to bottom and left to right: (a) SIFT (0.9574, 0.7121), (b) SURF (0.9833, 0.3236), (c) RIFT (0.9623, 0.1928), (d) GLPM (0.9902, 0.6475), (e) GP (0.9912, 0.8290), (f) GPFD (0.9954, 0.8740). For each algorithm, the first value is the precision, while the second value is recall rate, i.e., (Precision, Recall rate). The two endpoints of each blue line correspond to the positions of feature points in two images, indicating the correct match.

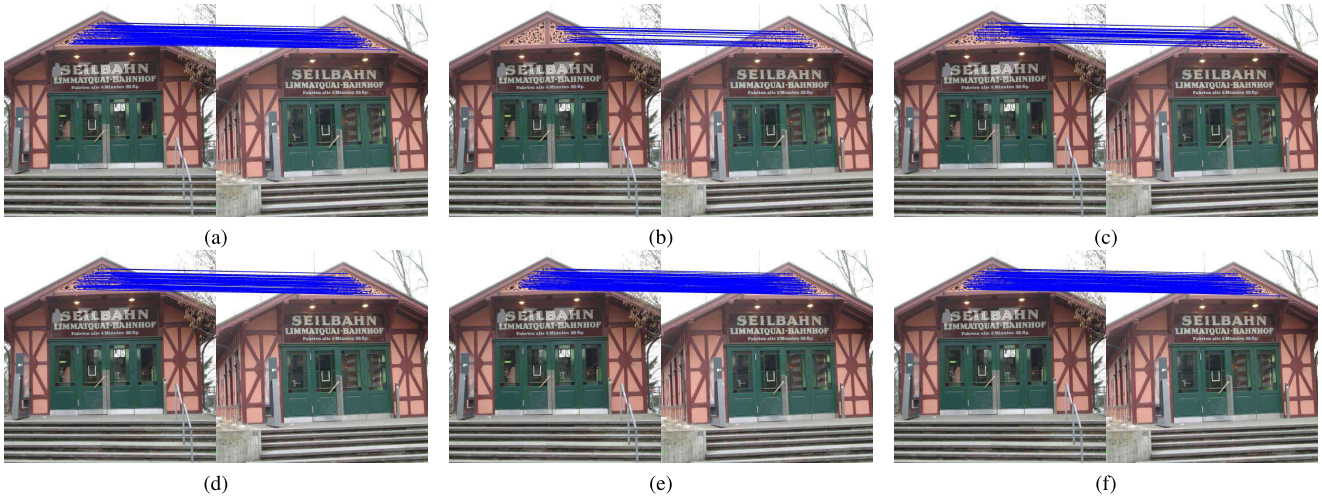


FIGURE 12. Matching results of image pair 6. From top to bottom and left to right: (a) SIFT (0.8831, 0.5151), (b) SURF (1.0000, 0.1137), (c) RIFT (0.8462, 0.2500), (d) GLPM (0.9664, 0.4978), (e) GP (0.9412, 0.6060), (f) GPFD (0.9785, 0.6894). For each algorithm, the first value is the precision, while the second value is recall rate, i.e., (Precision, Recall rate). The two endpoints of each blue line correspond to the positions of feature points in two images, indicating the correct match.

The image pair 1 has simple translation, so the matching of feature points is relatively easy, but our method can obtain more correct correspondences. And the Table 3 and Fig. 7 show that all the methods can align image pair 1. But robust terminal set is adopted in our method, and more correct correspondences are obtained. The proposed method gets superior RMSE. The number of correct matches, precision and recall rate are worse than GP but better than others. It is because the image pair 1 has less noise and our method have not significant advantage. However, it also fully demonstrates the effectiveness of genetic programming in image registration.

The image pair 2 suffers from rigid transformation and noise. So there are many wrong matches or similar pixels around the keypoints. And one to more matches are found because two keypoints are close enough in position or they have similar descriptors. This causes poor matching performance. By adopting robust first-order statistic, all the evaluation criterion of the proposed method are better than other algorithm in addition to the running time. And the experiment results instruction that our method is robust to noise.

The image pair 3 has simple transformations such as translation and rotation. As shown in the Table 3 and Fig. 9, the precision of all methods are close to or equal to 1. GPFD obtains the highest RMSE, number of correct matches, precision and recall rate. In addition, the terminal set used in the proposed method is not affected by rotation, so the image

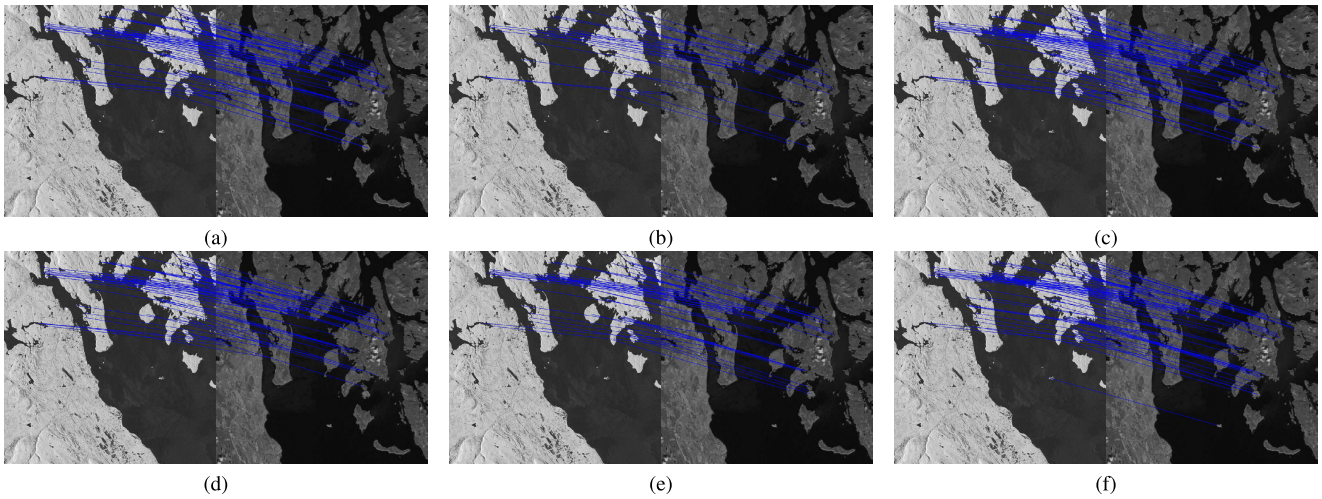


FIGURE 13. Matching results of image pair 7. From top to bottom and left to right: (a) SIFT (0.9531, 0.5351), (b) SURF (1.0000, 0.2368), (c) RIFT (0.9412, 0.5862), (d) GLPM (0.9836, 0.5263), (e) GP (0.7728, 0.2982), (f) GPFD (1.0000, 0.6316). For each algorithm, the first value is the precision, while the second value is recall rate, i.e., (Precision, Recall rate). The two endpoints of each blue line correspond to the positions of feature points in two images, indicating the correct match.

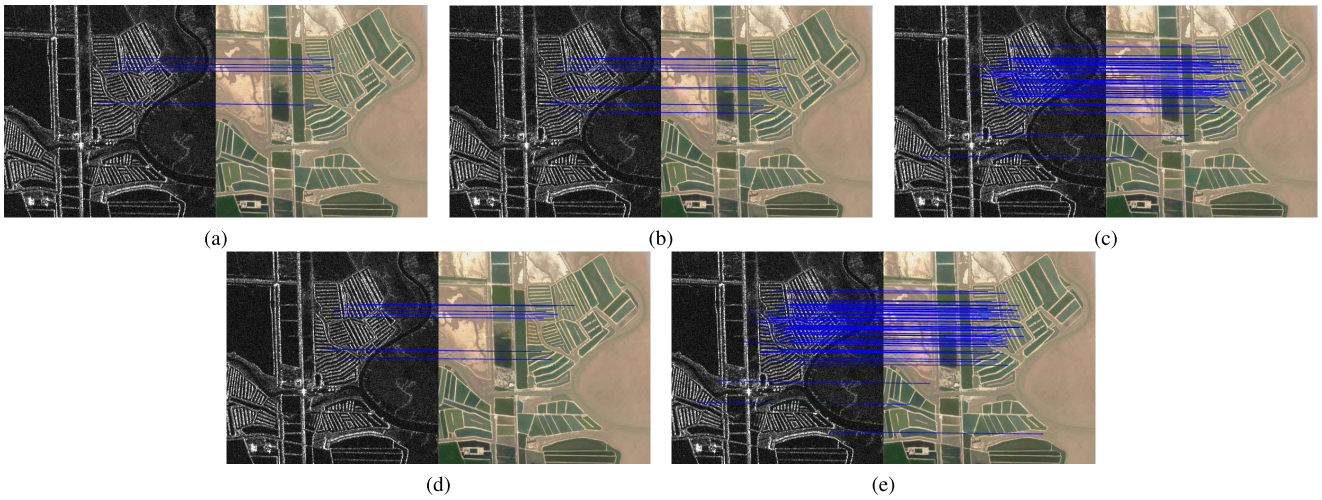


FIGURE 14. Matching results of image pair 8. From top to bottom and left to right: (a) SIFT (0.6667, 0.0556), (b) SURF (0.7273, 0.1111), (c) RIFT (0.9667, 0.8056), (d) GP (0.7500, 0.0833), (e) GPFD (0.9859, 0.9700). For each algorithm, the first value is the precision, while the second value is recall rate, i.e., (Precision, Recall rate). The two endpoints of each blue line correspond to the positions of feature points in two images, indicating the correct match.

descriptor evolved by GPFD is rotation-invariant. In a word, GPFD can work very well and get good performance in some simple image pairs.

The image pair 4 and 5 have the transformation of illumination intensity. This will make image data information missing or vague and the mapping relation between these two images complex, which lead to different registration process. The results in the Table 3 indicate that GPFD is able to obtain the best RMSE, number of correct matches, precision and recall rate. And number of correct matches in our method is almost four times as many as RIFT. The recall rate of GPFD is obviously higher than other methods. And from the experimental results of two image pairs, GPFD is robust to illumination change.

As shown in the Table 3 and Fig. 12, although SIFT, SURF, RIFT, GLPM and GP can get match results for image pair 6, the number of correct matches is small, and the recall rate is poor. The image pair 6 is obtained from different viewpoints and has complex non-rigid transformation. On the other hand, the image pair have subtle illumination change and image changes caused by glass reflection. These make the matching of feature points more difficult, and some real keypoints can not be found and some pseudo keypoints are regarded as real keypoints, so the number of correct matches are concentrated in the image with small changes. However, the proposed method still achieves better results than SIFT, SURF, RIFT, GLPM and GP, GPFD also is affected by complex image transformations, and the recall rate only reaches 0.6894.

TABLE 3. Compare of Root-mean-square error (RMSE), Number of correct matches (NCM), Precision, Recall rate and Time (sec) of test image pairs (*: fails to get correct registration result).

Methods / Datasets / Criteria		Pair1	Pair2	Pair3	Pair4	Pair5	Pair6	Pair7	Pair8
RMSE	SIFT	0.3937	0.1676	0.5305	0.7635	0.5326	0.5358	0.5482	0.5883
	SURF	0.3736	0.1557	0.6328	0.4567	0.4997	0.4915	0.5872	0.5950
	RIFT	0.5679	0.1569	0.5356	0.5798	0.5998	0.5213	0.7065	0.6320
	GLPM	0.4726	0.1758	0.5451	0.4444	0.5478	0.6544	0.5691	*
	GP	0.3572	0.1645	0.5234	0.5382	0.4853	0.4832	0.7199	0.7005
	GPFD	**0.3569**	**0.1539**	**0.5192**	**0.4422**	**0.4594**	**0.4742**	**0.5246**	**0.5512**
NCM	SIFT	538	243	751	1362	2439	77	64	6
	SURF	262	156	103	904	1079	15	27	11
	RIFT	540	412	649	676	654	39	71	60
	GLPM	581	433	778	2049	2144	68	61	*
	GP	601	220	807	2137	2742	85	44	8
	GPFD	**599**	**445**	**824**	**2829**	**2879**	**93**	**72**	**71**
Precision	SIFT	0.9963	1	0.9827	0.8817	0.9574	0.8831	0.9531	0.6667
	SURF	0.9962	1	0.9223	0.9900	0.9833	1	1	0.7273
	RIFT	0.9673	0.9782	0.9276	0.9872	0.9623	0.8462	0.9412	0.9667
	GLPM	0.9920	1	0.9897	0.9927	0.9902	0.9664	0.9836	*
	GP	1	1	0.9913	0.9336	0.9912	0.9412	0.7728	0.7500
	GPFD	**1**	**1**	**0.9976**	**0.9996**	**0.9954**	**0.9785**	**1**	**0.9859**
Recall rate	SIFT	0.8687	0.8379	0.8363	0.4247	0.7121	0.5151	0.5351	0.0556
	SURF	0.4230	0.5379	0.1058	0.3165	0.3236	0.1137	0.2368	0.1111
	RIFT	0.8466	1	0.6704	0.2360	0.1928	0.2500	0.5862	0.8056
	GLPM	0.9341	1	0.8574	0.7193	0.6475	0.4978	0.5263	*
	GP	0.9740	0.9586	0.8909	0.7054	0.8290	0.6060	0.2982	0.0833
	GPFD	**0.9708**	**1**	**0.9153**	**1**	**0.8740**	**0.6894**	**0.6316**	**0.9722**
Time	SIFT	10.199	12.765	13.867	48.325	45.924	6.544	16.826	6.617
	SURF	13.678	9.984	10.118	26.216	25.478	5.693	8.861	3.678
	RIFT	9.955	10.158	13.114	19.188	18.319	10.669	13.839	5.683
	GLPM	10.155	8.458	15.503	23.065	17.583	4.256	11.918	4.254
	GP	14.367	13.545	15.274	60.923	56.738	7.036	19.239	7.468
	GPFD	**14.259**	**13.017**	**14.965**	**58.361**	**55.294**	**7.812**	**18.153**	**7.592**

And later we will improve this problem by extracting more image information.

As we all know, remote sensing images have sophisticated noise. The image pair 7 is two SAR image and is acquired by different sensor with significant intensity difference. The result in Table 3 shows that SIFT, SURF, RIFT, GLPM and GP fail to get good result especially SURF algorithm. Because the descriptors of the common methods are seriously influenced by significant difference of the image intensity. Therefore, the descriptors of compared methods may not extract valuable features. And the distance of corresponding feature descriptors will not measure the real relation of two matching points. So these methods can not get enough reliable correspondences to compute

the transformation matrix parameters. The proposed method obtains superior performance owing to the robustness of our method.

The image pair 8 is a SAR image and a optical image from Google Earth with different imaging principle. GPFD gets the most optimal RMSE, number of correct matches, precision and recall rate. Precision by the proposed method is close to 1 and recall rate is evidently better than contrast methods. However, GP fails to achieve good performance because it is not robust to noise. And the presence of noise affects the matching result of image descriptor, and makes incorrect correspondences may be obtained. GPFD employs terminal set with rich statistical information, so image descriptors evolved by our method are more robust to noise and reliable.

VI. CONSLUSION

This paper proposes a novel GPFD method which can automatically evolve feature descriptors. Unlike other feature descriptors, the proposed method does not need human intervention to tune parameters. In the GPFD method, a set of simple arithmetic operators and improved first-order statistics are designed as the function set and the terminal set to evolve feature descriptors for complex image registration. Our algorithm includes three parts: 1) using SIFT to construct the training set and the test set; 2) utilizing GPFD to generate feature descriptors to remove a mass of false correspondences; and 3) removal of imprecise points to improve accurate by RANSAC algorithm. Experiments on different image datasets confirm that GPFD performs well for image with highly distortion and noise. Compared with other representative methods, GPFD is able to receive better performance including RMSE, the number of correct matches, precision and recall rate. In conclusion, the proposed GPFD is a robust and accurate registration method.

In this paper, GPFD has presented promising results in difficult image pairs. In the future, we will focus on studying a new approach to construct the training set and the test set to reduce the running time.

REFERENCES

[1] B. Zitov and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[2] W. Cao, F. Lyu, Z. He, G. Cao, and Z. He, "Multimodal medical image registration based on feature spheres in geometric algebra," *IEEE Access*, vol. 6, pp. 21164–21172, 2018.

[3] Y. Dong, T. Long, W. Jiao, G. He, and Z. Zhang, "A novel image registration method based on phase correlation using low-rank matrix factorization with mixture of Gaussian," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 446–460, Jan. 2018.

[4] B. Lin, X. Tao, Y. Duan, and J. Lu, "Hyperspectral and multispectral image fusion based on low rank constrained Gaussian mixture model," *IEEE Access*, vol. 6, pp. 16901–16910, 2018.

[5] L. Ke, Y. Lin, Z. Zeng, L. Zhang, and L. Meng, "Adaptive change detection with significance test," *IEEE Access*, vol. 6, pp. 27442–27450, 2018.

[6] B. Tian, S. Shi, Y. Liu, S. Xu, and Z. Chen, "Image registration of interferometric inverse synthetic aperture radar imaging system based on joint respective window sampling and modified motion compensation," *J. Appl. Remote Sens.*, vol. 9, no. 1, Jan. 2015, Art. no. 095097.

[7] M. Gong, S. Zhao, L. Jiao, D. Tian, and S. Wang, "A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4328–4338, Jul. 2014.

[8] J. Ma, J. Zhao, Y. Ma, and J. Tian, "Non-rigid visible and infrared face registration via regularized Gaussian fields criterion," *Pattern Recognit.*, vol. 48, no. 3, pp. 772–784, Mar. 2015.

[9] G. Dahman, J. Flordelis, and F. Tufvesson, "Cross-correlation of large-scale parameters in multi-link systems: Analysis using the box-Cox transformation," *IEEE Access*, vol. 6, pp. 13555–13564, 2018.

[10] S. Suri and P. Reinartz, "Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 2, pp. 939–949, Feb. 2010.

[11] A. Sedaghat, M. Mokhtarzade, and H. Ebadi, "Uniform robust scale-invariant feature matching for optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4516–4527, Nov. 2011.

[12] J. Jiang and X. Shi, "A robust point-matching algorithm based on integrated spatial structure constraint for remote sensing image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1716–1720, Nov. 2016.

[13] H. Al-Sahaf, M. Zhang, A. Al-Sahaf, and M. Johnston, "Keypoints detection and feature extraction: A dynamic genetic programming approach for evolving rotation-invariant texture image descriptors," *IEEE Trans. Evol. Comput.*, vol. 21, no. 6, pp. 825–844, Dec. 2017.

[14] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vols. SMC–3, no. 6, pp. 610–621, Nov. 1973.

[15] R. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, May 1979.

[16] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jul. 2005, pp. 886–893.

[17] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," in *Proc. 12th Int. Conf. Pattern Recognit.*, vol. 1, Dec. 2002, pp. 582–585.

[18] B. Xue and M. Zhang, "Evolutionary feature manipulation in data mining/big data," *ACM Sigevolution*, vol. 10, no. 1, pp. 4–11, May 2017.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[20] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vision-ECCV*, 2006, pp. 404–417.

[21] A. Song, T. Loveard, and V. Ciesielski, "Towards genetic programming for texture classification," in *Proc. Austral. Joint Conf. Artif. Intell.*, 2001, pp. 461–472.

[22] B. Tran, B. Xue, and M. Zhang, "Genetic programming for feature constructin and selection in classfication on high-dimensional data," *Memetic Comput.*, vol. 8, no. 1, pp. 3–15, 2016.

[23] Q. Chen, M. Zhang, and B. Xue, "Feature selection to improve generalization of genetic programming for high-dimensional symbolic regression," *IEEE Trans. Evol. Comput.*, vol. 21, no. 5, pp. 792–806, Oct. 2017.

[24] M. Iqbal, B. Xue, H. Al-Sahaf, and M. Zhang, "Cross-domain reuse of extracted knowledge in genetic programming for image classification," *IEEE Trans. Evol. Comput.*, vol. 21, no. 4, pp. 569–587, Aug. 2017.

[25] S. Hindmarsh, P. Andreae, and M. Zhang, "Genetic programming for improving image descriptors generated using the scale-invariant feature transform," in *Proc. 27th Conf. Image Vis. Comput. New Zealand-IVCNZ*, 2012, pp. 85–90.

[26] Y. Wu, W. Ma, Q. Miao, and S. Wang, "Multimodal continuous ant colony optimization for multisensor remote sensing image registration with local search," *Swarm Evol. Comput.*, vol. 47, pp. 89–95, Jun. 2019.

[27] Z. Wang, Y.-S. Ong, J. Sun, A. Gupta, and Q. Zhang, "A generator for multiobjective test problems with difficult-to-approximate Pareto front boundaries," *IEEE Trans. Evol. Comput.*, vol. 23, no. 4, pp. 556–571, Aug. 2019.

[28] Z. Wang, Y.-S. Ong, and H. Ishibuchi, "On scalable multiobjective test problems with hardly dominated boundaries," *IEEE Trans. Evol. Comput.*, vol. 23, no. 2, pp. 217–231, Apr. 2019.

[29] J. Li, Q. Hu, and M. Ai, "RIFT: Multi-modal image matching based on radiation-invariant feature transform," 2018, *arXiv:1804.09493*. [Online]. Available: https://arxiv.org/abs/1804.09493

[30] J. Ma, J. Jiang, H. Zhou, J. Zhao, and X. Guo, "Guided locality preserving feature matching for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4435–4447, Aug. 2018.

[31] E. P. Ong, J. A. Lee, J. Cheng, G. Xu, and B. H. Lee, "A robust outlier elimination approach for multimodal retina image registration," in *Proc. MICCAI*, vol. 2, 2015, pp. 329–337.

[32] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[33] Y. Wu, W. Ma, M. Gong, L. Su, and L. Jiao, "A novel point-matching algorithm based on fast sample consensus for image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 43–47, Jan. 2015.

[34] J. A. Lee, J. Cheng, B. H. Lee, E. P. Ong, G. Xu, D. W. K. Wong, J. Liu, A. Laude, and T. H. Lim, "A low-dimensional step pattern analysis algorithm with application to multimodal retinal image registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1046–1053.

[35] J. Chan, J. A. Lee, and Q. Kemao, "BIND: Binary integrated net descriptors for texture-less object recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2068–2076,

[36] S. Wang, D. Quan, X. Liang, M. Ning, Y. Guo, and L. Jiao, "A deep learning framework for remote sensing image registration," *ISPRS J. Photogram. Remote Sens.*, vol. 145, pp. 148–164, Nov. 2018.

[37] Y. Wu, Q. Miao, W. Ma, M. Gong, and S. Wang, "PSOSAC: Particle swarm optimization sample consensus algorithm for remote sensing image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 242–246, Feb. 2018.

[38] R. Fransens, C. Strecha, and L. Van Gool, "A mean field EM-algorithm for Coherent Occlusion handling in MAP-estimation prob," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jul. 2006, pp. 300–307.

[39] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, nos. 1–2, pp. 43–72, Nov. 2005.

[40] H. Aanæs,, R. R. Jensen, G. Vogiatzis, E. Tola, and A. B. Dahl, "Large-scale data for multiple-view stereopsis," *Int. J. Comput. Vis.*, vol. 120, no. 2, pp. 153–168, Nov. 2016.