Article

Coarse-to-Fine Image Registration for Multi-Temporal High Resolution Remote Sensing Based on a Low-Rank Constraint

Peijing Zhang^{1,2,3}, Xiaoyan Luo^{4,*}, Yan Ma^{5,6} , Chengyi Wang⁷, Wei Wang⁸ and Xu Qian²¹ School for Informatics and Cyber Security, People's Public Security University of China, Beijing 100038, China; zhangpeijing@ppsuc.edu.cn² School of Mechanical Electronic & Information Engineering, China University of Mining and Technology, Beijing 100083, China; xuqian@cumtb.edu.cn³ College of Geoscience and Surveying Engineering, China University of Mining and Technology, Beijing 100083, China⁴ School of Astronautics, Beihang University, Beijing 102206, China⁵ Beijing Engineering Research Center of Smart Mechanical Innovation Design Service, Beijing Union University, Beijing 100101, China; jqrmayan@buu.edu.cn⁶ College of Robotics, Beijing Union University, Beijing 100101, China⁷ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China; wangcy@radi.ac.cn⁸ Remote Sensing Center of Public Security, People's Public Security University of China, Beijing 100038, China; wwwei@ppsuc.edu.cn

* Correspondence: luoxxy@buaa.edu.cn; Tel.: +13-811-108-902

Abstract: For multi-temporal high resolution remote sensing images, the image registration is important but difficult due to the high resolution and low-stability land-cover. Especially, the changing of land-cover, solar altitude angle, radiation intensity, and terrain fluctuation distortion in the overlapping areas can represent different image characteristics. These time-varying properties cause traditional registration methods with known reference information to fault. Therefore, in this paper we propose a comprehensive coarse-to-fine registration (CCFR) algorithm. First, we design a low-rank constraint-based batch reference extraction (LRC-BRE) method. Under the low-rank constraint, the stable features with highly spatial co-occurrence can be reconstructed via matrix decomposition, and are set as reference images to batch registration. Second, we improve the general feature registration with block feature matching and local linear transformation (BFM-LLT) operators including match outlier filtering (MOF) on regional mutual information and dual-weighted block fitting (DWBF). Finally, based on combining LRC-BRE and BFM-LLT, CCFR is integrated. Experimental results show that the proposed method has a good batch alignment effect, especially in the registration of large difference image pairs. The proposed CCFR achieves a significant performance improvement over many state-of-the-art registration algorithms.

Keywords: registration; multi-temporal remote sensing images; high resolution; low-rank matrix factorization; optical satellite

Citation: Zhang, P.; Luo, X.; Ma, Y.; Wang, C.; Wang, W.; Qian, X.

Coarse-to-Fine Image Registration for Multi-Temporal High Resolution Remote Sensing Based on a Low-Rank Constraint. *Remote Sens.* **2022**, *14*, 573. <https://doi.org/10.3390/rs14030573>

Academic Editor: Chein-I Chang

Received: 16 November 2021

Accepted: 20 January 2022

Published: 25 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Multi-temporal remote sensing image registration is a key preprocessing step for remote sensing image fusion [1,2], change detection [3,4], super-resolution reconstruction [5], and collaborative analysis [6]. In recent years, remote sensing image registration has mostly focused on solving the effects of radiation differences and geometric distortions by constructing robust feature descriptors [7,8], better screening of mismatched features [9,10], integrating multiple methods [11,12], and so on. Although research on multi-temporal image registration in complex terrain areas has been conducted [13,14], there are few studies devoted to accurate registration in complex terrain areas with low-stability land-cover. The low stability of features mentioned in this paper mainly refers to the land-cover changes

caused by seasonal rotation, geological alteration, human activities, and other factors, which lead to huge imaging differences of the same plot in different time series. Common low-stability features include mining, cultivated farmland, and seasonal deciduous woody vegetation under high resolution imaging conditions. Complex terrain mainly refers to undulating mountains and large-scale artificial elevation features (such as dams and buildings) under high resolution conditions. For multi-temporal remote sensing images, radiation differences including the difference of radiation intensity and the change of solar altitude angle are also an important factor leading to differences in ground object imaging; in particular, changes of solar altitude angles lead to different degrees of shadow occlusion and contour diversity of mountainous and elevation ground objects. In addition, differences of shooting inclinations lead to complex geometric deformations of mountainous and elevation ground objects. Mountainous areas with different seasonal vegetation coverage, irregular mining, and beneficiation activities are typical representatives of low-stability and complex terrain, as shown in Figure 1. Due to the mixed influence of the above multiple factors, the remote sensing images have different period features in terms of gray, texture, and contour features, which bring challenges to the accurate temporal registration.

Figure 1. Examples of multi-temporal satellite remote sensing panchromatic images with low-stability and complex terrain: (a) mountainous area covered by seasonal vegetation (700×850 pixels); (b) mine tailings pond (950×1100 pixels).

Currently, remote sensing images are registered based on the geocoding generated by satellite orbit parameters [15]. Affected by the accuracy of satellite positioning and control, these technologies can largely eliminate the global geometric distortion but generate images with ten to dozens of pixel offsets. For mountainous areas with complex terrain, in order to improve the registration accuracy based on geocoding, it is necessary to correct the image point element offset caused by terrain fluctuation with the help of a high-precision digital elevation model (DEM) [16] and a reference image map corrected by ground control points. Limited by the accuracy of the available DEM and the lack of reference images corrected by sufficient ground control points, there are still different degrees of alignment errors in local areas with serious distortion in mountainous areas after ortho-rectification [17,18]. For higher image alignment accuracy, it is necessary to further correct with the help of an automatic image registration method. In addition, the low-stability surface makes it difficult to automatically match the control points between the sensed image and the reference image, in which accurate manual auxiliary matching is also a challenging work.

The performance of an image registration algorithm depends on its image matching step and the transformation function. Common image matching drivers include intensity matching, feature matching, density matching, and other similarity matching algorithms. According to different domains of transformation, spatial transformation estimation methods include global mapping models, local mapping models, and elastic registration [19,20].

The intensity-based method usually uses the gray information of the image to measure the similarity and realize the global spatial transformation in the area. Based on statistical correlation such as normalized cross correlation (NCC) [21,22], mutual information

(MI) [23,24] has a high applicability for remote sensing image registration without prominent detail features, but it is unable to deal with complex area deformation. In order to solve the registration of large-scale geometric deformation and complex terrain, most of the relevant studies focus on the combination of regional intensity similarity and local features. Usually, the feature-based method is used as a coarse registration to calculate parameters, which can be used as the initial condition to perform fine registration using an intensity-based method [11,25]. Some researchers improve the performance of registration by applying gradient features [26] and wavelet-like features [27] to image descriptors.

The feature-based methods separate image feature matching and spatial transformation, which can be used with a variety of local transformation modes to deal with complex deformation. SIFT [28] and its variants (SURF [29], ASIFT [30], and PCA-SIFT [31]) are widely used feature extractors. These extracted features are robust to changes in terms of brightness, rotation, and scaling. The ORB [32] and BRISK [33] algorithms belong to a corner detection method based on FAST [34], which improves the computational efficiency by using binary feature description words. In recent years, for the complex nonlinear radiation intensity difference between multimodal remote sensing images, feature algorithms, such as deny oriented gradient feature (DOGH) [7] and histogram of oriented phase convergence (HOPC) [35], have been proposed and achieved good experimental results in feature matching of optical and SAR image pairs. In order to solve the significant radiation difference between multi-temporal optical remote sensing images, the PSO-SIFT [8] algorithm gives a new gradient definition to each pixel in Gauss scale space to increase the robustness of the gradient direction and size difference in the same region. Aiming at the significant geometric distortion and radiation difference between multi-source optical remote sensing images, Chen et al. [10] proposed an iterative scale invariant feature transformation method to continuously improve the registration accuracy of remote sensing images. For local geometric distortion, Sedaghat and Ebadi [36] used an adaptive histogram quantization strategy to calculate feature position and gradient direction, so that they are robust to local view distortion. However, the above feature matching methods used to improve the robustness of feature matching for significant radiation difference and local geometric distortion achieve good results only in the medium–low resolution multisource remote sensing image registration, but pay less attention to the low stability of ground objects. In view of the impact of clouds, noise, and land-cover change on the accuracy of local registration, Feng et al. [12] proposed a step-by-step registration method combining feature and regional methods. In this method, the first step matching feature results are geometrically corrected by a region-based linear polynomial transformation model, but it is only robust to simple and small radiation differences.

The density-based methods use mostly optical flow estimation. Optical flow [37,38] can achieve per-pixel displacement estimation and is suitable for the accurate alignment of remote sensing images on complex terrain. However, this method has a limited ability to deal with large deformations between registered images, and is sensitive to low-stability surface noise, which causes resampled image texture distortion. Liu et al. [39] employed mesh segmentation to distribute the optical flow field of feature points to deal with complex non-rigid body deformation, so as to improve the robustness of optical flow registration. Xu et al. [40] used the optical flow field method to handle complicated non-rigid changes by matching dense SIFT features. However, when the image texture is inconsistent because of the local land-cover changing, the above method could cause local anomalies in the computed optical flow. In order to realize the high fidelity of the registered image in complex terrain areas, Feng et al. [13] added the detection and correction of abnormal displacement fields after the optical flow estimation.

For high resolution multi-temporal remote sensing images with low-stability land-cover and complex terrain (HMR-LC images), it is difficult to align the images facing severe radiation differences and complex geometric deformation by relying on a single registration method. Multi-level and combined registration methods are needed to solve the problems of this kind of remote sensing image registration. The feature method has

good adaptability to large-scale rotation, displacement and deformation, and this method can be used together with the local transformation model to solve the inconsistency of local deformation in complex terrain. The above research experience in solving complex registration problems shows that the registration performance can be improved by the comprehensive application of feature method and regional intensity information, phase information, and structure information [10,11,25,40]. However, due to the comprehensive influence of land-cover changes, terrain fluctuations, mountain shadows, radiation changes, and other factors, the probability of feature matching errors for HMR-LC is high, which seriously affects the registration results. On the one hand, affected by changes of low land-cover, the texture features of unstable objects in the image pair are significantly different, and even the texture features of partially overlapping areas are completely changed. On the other hand, due to the influence of radiation differences and shooting inclination differences, there are serious geometric deformations and complex radiation differences in the local feature window of elevation objects and complex terrain imaging, which exceeds the invariance of features and the comparability of intensity information, and it is difficult to accurately express the complex geometric relationship of different positions using the surrounding feature information [41]. These greatly reduce the possibility of detecting consistent features in image pairs, resulting in the failure of the application of mainstream feature methods in the studied target image.

Different from the previous research ideas used to solve the influence of surface change, radiation difference, and complex terrain on registration, we no longer focus on improving the registration algorithm of single image pairs, but propose a batch reference extraction into a comprehensive coarse-to-fine registration (CCFR) framework. This framework not only realizes the robust registration of significantly different HMR-LC images, but also solves the problem of a lack of a reliable baseline image for registration.

The main contributions of this paper are as follows:

- (1) Inspired by the excellent image denoising and restoration ability of the low-rank decomposition algorithm, we design a low-rank constraint-based batch reference (LRC-BRE) method to restore the stable features holding highly spatial co-occurrence in the image sequence, and construct a corresponding batch reference, in which each original image has a respective reference.
- (2) To match the original image and the reference image, the regional mutual information is considered to filter the match outliers, named as a match outlier filtering (MOF). Additionally, a dual-weighted block fitting (DWBF) is developed based on the feature inverse distance weight and feature regional similarity weight. The above two operators are integrated to form the block feature matching and local linear transformation (BFM-LLT) registration processing, which has good robustness and alignment accuracy for the coarse registration of multi-temporal remote sensing images with medium and low texture differences and for the registration of original images and stable feature images restored by LRC-BRE.
- (3) A new comprehensive coarse-to-fine registration (CCFR) framework integrated by LRC-BRE and BFM-LLT is proposed for HMR-LC images. By taking the recovered stable feature image as the reference baseline image, the proposed framework transforms the direct registration of large difference HMR-LC image pairs into the indirect registration of small and medium difference image pairs, and realizes the applicability of the mainstream image registration methods.
- (4) On GF-2 and GF-1 satellite remote sensing image datasets with low-stability land-cover and complex terrain, the experimental results show that the comprehensive registration framework CCFR is more effective than the latest registration algorithm in visual quality and quantitative evaluation, owing to the combining of LRC-BRE with good batch reference and BFM-LLT with improved alignment effect.

The rest of the paper is organized as follows: The details of the proposed method are described in Section 2. The experimental results and verification from visual and quantitative perspectives are provided in Section 3. The influencing factors of CCFR

indirect registration and the work of CCFR in different optical satellite images are discussed in Section 4. Finally, our conclusions are summarized in Section 5.

2. Methods

Image feature detection and matching performance is the key step to achieve accurate image registration results. The number, accuracy and distribution of matching inner points directly affect the registration accuracy. In this section, we aim to propose a coarse-to-fine registration (CCFR) framework for multi-temporal high resolution remote sensing images, which can not only generate the stable feature reference sequence automatically, but also register the point robustly. A simple flowchart of the CCFR framework is shown in Figure 2. To solve the stable feature reference problem in batch registration processing, we introduce a low-rank constraint-based batch reference (LRC-BRE) extracting method via matrix decomposition under a low-rank constraint. In CCFR, the LRC-BRE is executed based on blocks. We also developed a block feature matching and local linear transformation (BFM-LLT) registration processing via integrating the match outlier filtering (MOF) and the dual-weighted block fitting (DWBF) operators. BFM-LLT is applicable to both batch coarse registration and batch fine registration.

Figure 2. Simple flowchart of the CCFR framework.

2.1. LRC-BRE: A Low-Rank Constraint-Based Batch Reference Extraction

With the rapid development of matrix rank related optimization techniques, the matrix minimum low-rank constraint attempted to replace the image similarity evaluation to realize the batch alignment of natural images [42,43]. Matrix rank correlation optimization technology originated from low-rank matrix factorization (LRMF) [44]. LRMF is an effective method of developing reliable, scalable, and robust algorithms to estimate a low-rank matrix of interest, from potentially noisy, nonlinear, and highly incomplete observations [45]. In view of the above excellent ability of LRMF, we use this model to restore the stable feature images of multi-temporal image sequences to capture a reference image with stable surface features and small radiation difference for multi-temporal remote sensing image registration. Thus, a larger number and higher accuracy of feature pairs are obtained than the direct feature detection and matching of the original image pair.

The texture difference of multi-temporal low-stability images are treated as noise, and based on the relevant hypothesis and inferences of stable feature image restoration, they can be solved from an LRMF problem. The related theorems, hypothesis and inferences are given in the appendix section.

The result of feature matching is directly related to the similarity of texture features between image pairs. The higher the texture similarity of image pairs is, the more matching features and the higher the matching accuracy can be obtained. Let $\text{diff}_{\Omega}(I_i, I_j)$ represent the texture difference degree of the overlapping area Ω of the locally aligned image pair (I_i, I_j) , and index $\text{diff}_{\Omega}(I_i, I_j)$ be the reverse index of the texture similarity of the image pair. $\text{diff}_{\Omega}(I_i, I_j)$ is negatively correlated with the success rate of feature matching. Let T denote spatial transformation, and $g(I_i^{\Omega}) = g(T \circ I_j^{\Omega})$, $I_i^{\Omega} \in I_i$, $I_j^{\Omega} \in I_j$; then there is $\text{diff}_{\Omega}(I_i, I_j) = \text{diff}(I_i^{\Omega}, T \circ I_j^{\Omega})$. When the image pairs (I_i, I_j) are fully aligned, the comprehensive difference of the image pairs is equivalent to the texture difference, that is, $\text{diff}(I_i, I_j) = \text{diff}_{\Omega}(I_i, I_j)$; when not aligned, there is $\text{diff}(I_i, I_j) \geq \text{diff}_{\Omega}(I_i, I_j)$. When there is only a local offset in the image pair, it can be considered that $\text{diff}(I_i, I_j) \approx \text{diff}_{\Omega}(I_i, I_j)$. We assume that there is a threshold β . When $\text{diff}_{\Omega}(I_i, I_j) > \beta$, the commonly used image

feature matching methods fail to obtain enough correct matching features. For high resolution multi-temporal remote sensing images with low stable surface and complex terrain, due to the comprehensive influence of ground feature changes, terrain fluctuations, mountain shadows, radiation changes, and other factors, the texture difference of some images exceeds the threshold β .

According to Inference A3 in the Appendix A, for the image sequence $\{I_i\}_{i=1}^N$ and its restored stable feature image sequence $\{Z_i\}_{i=1}^N$, the probability is $\text{diff}(I_i, I_r) > \text{diff}(I_i, Z_i)$ ($1 \leq i \leq N_{s2} \leq N$). According to Inference A1.1 in the Appendix A, for the image sequence $\{I_i\}_{i=1}^N$ conforming to Hypothesis A2 in the Appendix A and its restored stable feature image sequence $\{Z_i\}_{i=1}^N$, there is $Z_1(x, y) \approx \dots \approx Z_N(x, y)$, $g(Z_1) \approx \dots \approx g(Z_N)$. Then, we only need to take Z_i as the registration reference image and register the image pairs $\{I_i, Z_i\}$ ($1 < i < N$) one by one, which can approximately realize the registration of any image pair $\{I_i, I_j\}$ ($1 < i, j < N$) in the sequence $\{I_i\}_{i=1}^N$. Let $I'_i = T(I_i, Z_i), T(I_i, Z_i) : I_i \rightarrow Z_i$; then there is $I'_1(x, y) \approx \dots \approx I'_N(x, y)$, $g(I'_1) \approx \dots \approx g(I'_N)$. By using the stable feature image recovered by LRMF as the reference image, the registration between the image pairs (I_i, I_j) with large differences $\text{diff}_\Omega(I_i, I_j) > \beta$ is replaced by the registration of the image pairs (I_i, Z_i) and (I_j, Z_j) with medium difference. When $\text{diff}_\Omega(I_i, Z_i) < \beta$ and $\text{diff}_\Omega(I_j, Z_j) < \beta$, the feature-based register method can be effectively applied. Thus, the registration for multi-temporal images with low stability and large difference is realized by indirect registration.

Through the indirect registration of the original image and the corresponding stable feature image in turn, the accurate registration between the low-stability original images with obvious feature differences is realized. For a few serious outlier images I_o in the original image sequence $\{I_i\}_{i=1}^N$, the quality of the restored stable feature image is greatly affected. The stable feature image Z_o^* with high restoration quality and close to the texture features of the original image I_o is selected as the reference image for registration, that is, $I'_o = g(I_o, Z_o^*)$ is used instead of $I'_o = g(I_o, Z_o)$ to improve the robustness of registration to individual outlier image registration.

The flowchart of LRC-BRE and batch registration is shown in Figure 3. First, the alignment degree of the image sequence after coarse registration meets the requirements of Hypothesis A2 in the Appendix A. Moreover, for images with different resolutions, the image sequences need to have the same resolution through the interpolation algorithm. On this basis, the matrix $M = [\text{vec}(I_1) | \dots | \text{vec}(I_N)]$ is constructed by linearly straightening the coarse registration image sequence, and the stable feature image sequence $\{Z_i\}_{i=1}^N$ is restored by Formula (A3) in the Appendix A. Then, from the sequence $\{Z_i\}_{i=1}^N$, the most suitable stable feature image Z_i^* is selected for each image $I_i \in \{I_i\}_{i=1}^N$ as the reference image. Finally, the batch registration of image sequences is realized through the registration of image pairs (I_i, Z_i^*) one by one.

Figure 3. Flowchart of the LRC-BRE (note: the batch registration step is not necessary for batch reference extraction, which is mainly considered for ablation experiments).

According to Inference A2 in the Appendix A, the recovery quality of Z_o is related to the degree of the initial outlier of I_o . The lower the initial outlier degree of I_o , the

more sparse the separated noise matrix e , and the smaller the comprehensive difference $\text{diff}(I_o, Z_o)$ value of the corresponding stable feature image Z_o . When $\text{diff}(I_o, Z_o) > \min_{1 \leq i \leq N, i \neq o} \text{diff}(I_o, Z_i)$, there are stable feature images Z_o^* closer to I_o than Z_o . Due to the low-rank constraint, the texture difference between image pairs (Z_o, Z_i) is small, so the difference between $\text{diff}(I_o, Z_o)$ and $\text{diff}(I_o, Z_i)$ is not obvious. Therefore, we add the constraint of the original image (I_o, I_i) on the similarity of texture features. Finally, in order to ensure that the stable feature image Z_o^* used for substitution has a clear texture, we add the constraint $\min \text{diff} I_i, I_{mean}$ to make the image I_i have a low degree of outlier in the set $\{I_i\}_{i=1}^N$. So far, the most suitable stable feature image Z_o^* can be determined by the following formula:

$$Z_o^* = \text{Best}(I_o, \{I_i, Z_i\}_{i=1}^N) = \arg \min_{\{I_i, Z_i\}_{i=1}^N} \text{diff}(I_o, Z_i) + \gamma_1 * \text{diff}(I_o, I_i) + \gamma_2 * \text{diff} I_i, I_{mean} \quad (1)$$

where, γ_1 and γ_2 is the compromise coefficient, $I_{mean}(x, y) = \sum_{i=1}^N I_i(x, y) / N$. Here, we use the pixel intensity difference 1-norm of the image pair to calculate $\text{diff}(\cdot, \cdot)$. Since the difference of radiation intensity does not change the geometric characteristics of image texture, in order to avoid the excessive interference of radiation intensity on the evaluation index, we uniformly normalize the radiation intensity of the image set in the preprocessing step. By matching the statistics of image intensity probability distribution, the radiation normalization of image sequence is realized. Assuming that the radiation intensity of the image set $\{I_i\}_{i=1}^N$ obeys Gaussian distribution, there are:

$$I'_i(x, y) = \text{Histo}(I_i(x, y)) = \frac{I_i(x, y) - \mu_{I_i}}{\sigma_{I_i}} \times \sigma_{I_{mean}} + \mu_{I_{mean}} \quad (2)$$

where $I'_i(x, y)$ is the pixel gray value after normalization of image radiation intensity, $\mu_{I_{mean}}$ and $\sigma_{I_{mean}}$ are the gray mean and standard deviation of image I_{mean} , respectively, and $I_i(x, y)$, μ_{I_i} , σ_{I_i} are the pixel gray value, gray mean, and standard deviation of image I_i , respectively.

When the image set is not normalized by the unified radiation intensity, we can also calculate $\text{diff}(\cdot, \cdot)$ by constructing the normalized difference 1-norm of the image pair:

$$\text{diff}(I_s, I_r) = \|I'_s - I'_r\|_1 = \frac{1}{\sigma_{I_s} \sigma_{I_r}} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [(I_s(x_i, y_j) - \mu_{I_s}) \sigma_{I_r} - (I_r(x_i, y_j) - \mu_{I_r}) \sigma_{I_s}] \quad (3)$$

$$\text{where, } \mu_{I_s} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I_s(x_i, y_j), \quad \mu_{I_r} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I_r(x_i, y_j), \quad \sigma_{I_s} = \sqrt{\frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I_s(x_i, y_j) - \mu_{I_s}]^2}, \quad \sigma_{I_r} = \sqrt{\frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I_r(x_i, y_j) - \mu_{I_r}]^2}.$$

2.2. BFM-LLT: Block Feature Matching and Local Linear Transformation

Referring to the stable feature image Z_i^* , the texture difference between image pairs (I_i, Z_i^*) is relatively small, but the traditional feature extraction methods still cause matching and location errors due to the interference of complex terrain and noisy background, which affect the final registration accuracy. In order to improve the features-based registration algorithm, we propose the BFM-LLT method. In CCFR, BFM-LLT is also used for coarse registration of image sets, which is suitable for image pairs whose texture difference is less than the threshold β .

The main steps of BFM-LLT are shown in Figure 4. First, BFM-LLT executes the feature matching algorithm in block mode, and uses MOF to filter the matching inner points and optimize the location. Then, the local homography matrix is calculated according to the dual-weighted block fitting (DWBF) method, and each block region executes homography transformation. The sizes of the matching block and fitting block do not need to be

consistent, and it is recommended that the former be larger than the latter. BFM-LLT can better take into account the global alignment and local alignment of complex terrain image registration.

Figure 4. Block feature matching and local linear transformation method (BFM-LLT).

2.2.1. Match Outlier Filtering (MOF)

In the block feature matching phase, to deal with the high resolution, the sensed image and the reference image are divided into $C1 \times C2$ blocks at equal intervals, respectively, and then the features are extracted and matched within the block. Block feature matching can improve the performance in two aspects: (1) through block feature matching, the feature area to be matched is reduced, so as to reduce the impact of low stable surface background noise and high self-similarity of vegetation texture, and improve the number and accuracy of feature-based matching. (2) In terms of computational performance, the block feature matching method reduces the overall computational complexity of the matching algorithm. In addition, the block feature matching method is also convenient for parallel calculation of each block and further improves the execution speed.

In order to improve the robustness of feature matching of complex texture background images, after consistency screening, we use MOF to filter and optimize the position of each inner point pair again, and record their RMI value of each group as the feature weight of the next local fitting. After feature distance matching (FDM) and random sample consensus screening (RANSAC) [46], the matching inner point pair is filtered by the intensity similarity threshold of the local area (taking the location of the feature point as the geometric center of local area). When the number of matching inner points is small, the template matching optimization of matching inner points is carried out based on the local area similarity index to further optimize the location of matching inner points. As shown in Figure 5, this method is suitable for the same resolution image registration with complex texture background.

Figure 5. Template matching optimization based on the similarity of the local region where the matching inner point is located. (a) Local area of reference image (taking a stable feature image as an example, size 90×120 pixels). (b) Local area of sensed image (90×120 pixels).

Considering that regional mutual information (RMI) is an extension of mutual information [47], it combines spatial information in the form of energy function, and uses eight-neighbor information to calculate the joint probability distribution instead of the original gray information. Since RMI is robust to nonlinear intensity differences, and has been successfully applied to multispectral or multi-sensor image registration [23,24], we directly adopt it for the local area similarity measurement index:

$$P_0 = P - \frac{1}{N} \sum_{i=1}^N p_i \quad (4)$$

$$C = \frac{1}{N} P_0 P_0^T \quad (5)$$

$$RMI = H_g(C_A) + H_g(C_B) - H_g(C) \quad (6)$$

where $P = [p_1, p_2, \dots, p_N]$, $N = (m - 2r)(n - 2r)$ is the number of pixels to be counted in a region, the size of the overlapping region is $m \times n$, r is the radius of the local window, and p_i is the pixel's neighborhood vector for the i pixel. The size of P is $d \times N$ with $d = 2 \times r \times r$. $H_g(C)$ is the joint entropy. $H_g(C_A)$ and $H_g(C_B)$ are the marginal entropies, where C_A is the $(d/2) \times (d/2)$ matrix in the upper left of C and C_B is the $(d/2) \times (d/2)$ matrix in the lower left of C .

2.2.2. Dual-Weighted Block Fitting (DWBF)

Due to the low-stability and complex terrain, the accuracy of multi-temporal image feature matching is uncertain, and the uneven distribution of matching features is common. A block homography based on feature inverse distance weight (IDW) and feature region similarity weight (RSW) is proposed, which is called the dual-weighted block fitting (DWBF) method. First, the sensed image and the reference image are divided into $C3 \times C4$ blocks at equal intervals, and each block corresponds to a separate homogeneous space homography transformation matrix $H \in \mathbb{R}^{3 \times 3}$. Then, on the basis of a homography matrix estimation based on inner point pairs, the feature inverse distance weight [48] and feature similarity weight are added for each inner point pair. The feature inverse distance weight of each block is calculated based on the block center point x_* . For any x of the matched inner point set, the farther the distance between x_* and x is, the lower the weight is when estimating the block homography matrix; the feature similarity weight is determined by the RMI of the region where the inner point is located, and the inner point with a high similarity has a higher influence weight.

Let $x = (x, y)^T$ and $x' = (x', y')^T$ be matching points across overlapping images I and I' . A basic method to estimate H from a set of point matches $\{x_i, x'_i\}_{i=1}^K$ across I and I' can be expressed as:

$$\hat{H} = \arg \min_h \sum_{i=1}^K \|a_i H^2\| = \arg \min_h \|AH^2\| \text{ s.t. } \|H\| = 1 \quad (7)$$

where $a_i \in \mathbb{R}^{2 \times 9}$ is computed for the i -th datum $\{x_i, x'_i\}$, and the matrix $A \in \mathbb{R}^{2N \times 9}$ is obtained by stacking vertically a_i for all i . After introducing the IDW and RSW, the block homography matrix H_* is estimated from the problem:

$$H_* = \arg \min_h \sum_{i=1}^K \|w_*^i a_i H\|^2 = \arg \min_h \|W_* A H\|^2 \quad (8)$$

$$w_*^i = \exp\left(-\frac{\|x_* - x_i\|^2}{\eta_i \sigma^2}\right) \quad (9)$$

$$\eta_i = RMI(\text{area}(x_i), \text{area}(x'_i)) \quad (10)$$

where, x_* is the coordinate of the center point of the block, σ is a scale parameter and the weight matrix $W_* = \text{diag}([w_*^1 w_*^1 \dots w_*^N w_*^N])$. The problem in (8) is a weighted SVD problem, and the solution is simply the least significant right singular vector of $W_* A$.

2.3. Algorithm of CCFR

Based on the integration of LRC-BRE and BFM-LLT, CCFR can improve the overall restoration quality of a locally stable feature image sequence, and can select a more suitable

local reference image for the local area of the sensing image rather than the global LRC-BRE application. The specific process of CCFR mainly includes three parts, as shown in Algorithm 1: (1) Batch coarse registration (steps 1–3). First, in order to reduce the complexity of feature matching and minimize the overall cost of spatial transformation of batch coarse registration, the image with the lowest outlier is selected as the common reference image of the whole image sequence G . Then, the BFM-LLT method is used to improve the alignment degree of coarse registration and realize the global alignment of most images. In the BFM step of rough registration, the matching results of each block are recorded as the guidance of whether the registered block participates in LRC-BRE. (2) Block-based LRC-BRE (steps 4 and 5). the LRC-BRE method is applied longitudinally to restore the block stable feature images for each block sequence after coarse alignment, and then they are synthesized into the global stable feature image horizontally. (3) Batch fine registration (step 6). With a globally stable feature image as a reference, the images are registered one by one using the BFM-LLT method.

Algorithm 1: CCFR

Input: images set $\{I_i^0\}_{i=1}^N$

Step 1: $\{I_i^0\}_{i=1}^N$ is aligned by geocoding, cropped and interpolated (optional) to obtain a group of image sequences G with the same size and the same resolution:

$$G = \{I_i\}_{i=1}^N \in \mathbb{R}^{w \times h}$$

Step 2: The lowest outlier image I_φ from image sequence G is calculated by:

$$I_\varphi = \arg \min_{1 \leq i \leq N} \sum_{j=1}^N \text{diff}(I_i, I_{\text{mean}})$$

Step 3: With I_φ as the reference image, based on the BFM-LLT method, the image pairs (I_i, I_φ) $i \in \{1, \dots, N\}$ and $i \neq \varphi$ are coarsely registered in turn to obtain a new image sequence $G' = \{I'_i\}_{i=1}^N$, and the matching results are recorded in the matrix set $\Lambda = \{A_i\}_{i=1}^N \in \text{bool}^{C_1 \times C_2}$:

$$(I'_i, A_i) = \text{BFM-LLT}(I_i, I_\varphi, C_1, C_2)$$

$$\text{where } I_i \in G, I'_i = \begin{bmatrix} \Omega_i^{1,1} & \dots & \Omega_i^{1,C_2} \\ \vdots & \ddots & \vdots \\ \Omega_i^{C_1,1} & \dots & \Omega_i^{C_1,C_2} \end{bmatrix}, A_i = \begin{bmatrix} a_i^{1,1} & \dots & a_i^{1,C_2} \\ \vdots & \ddots & \vdots \\ a_i^{C_1,1} & \dots & a_i^{C_1,C_2} \end{bmatrix}, \Omega_i^{j,k} \in \mathbb{R}^{w_\Omega \times h_\Omega}$$

represents the block region in row j and column k of the image I'_i , and $a_i^{j,k}$ represents the feature matching result of the block region in row j and column k of the image I_i . When the number of matching inner points in the region is insufficient, there is $a_i^{j,k} = \text{true}$.

Step 4: Generate a stable feature image block sequence for each block

sequence: $\{\Omega_i^{j,k}, \Psi_i^{j,k}, \Gamma_i^{j,k}\}_{i=1}^N =$

$$\min_{L_{\Omega^j}, S_{\Omega^j}} \|L_{\Omega^j,k}\|_* + \frac{1}{\sqrt{w_\Omega \times h_\Omega}} \|S_{\Omega^j,k}\|_1 \text{ s.t. } \|M_{\Omega^j,k} - L_{\Omega^j,k} - S_{\Omega^j,k}\|_F \leq \epsilon$$

where $M_{\Omega^j,k} = [\text{vec}(\Omega_1^{j,k}) \dots \text{vec}(\Omega_q^{j,k})] \in \mathbb{R}^{m_\Omega \times q}$, $m_\Omega = w_\Omega \times h_\Omega$, $\Omega_i^{j,k} \in G_{\Omega^j,k}^q$, $G_{\Omega^j,k}^q =$

$\{\Omega_i^{j,k}\}_{i=1}^q$ ($q \leq N$) = $\{\Omega_i^{j,k} | a_i^{j,k} == \text{true}\}_{i=1}^N$, $\Psi_i^{j,k}$ and $\Gamma_i^{j,k}$ respectively represent the block stable feature image and block sparse matrix re-stored corresponding to $\Omega_i^{j,k}$, and ϵ is a penalty factor.

Step 5: For each $I_i \in G$, its globally most suitable stable feature image Z_i^* is synthesized by:

$$Z_i^* = \begin{bmatrix} \Psi_i^{*1,1} & \dots & \Psi_i^{*1,C_2} \\ \vdots & \ddots & \vdots \\ \Psi_i^{*C_1,1} & \dots & \Psi_i^{*C_1,C_2} \end{bmatrix}$$

where $\Psi_i^{*j,k} = \text{Best}(\Omega_i^{j,k}, \{\Omega_l^{j,k}, \Psi_l^{j,k}\}_{l=1}^q)$ is the most appropriate block stable feature image for $\Omega_i^{j,k}$. The Best () is given in Formula (1).

Step 6: Block feature matching and local linear transformation are carried out on the image pair (I'_i, Z_i^*) $i \in \{1, \dots, N\}$ in turn to obtain the accurately registered image sequence:

$$(I''_i, A'_i) = \text{BFM-LLT}(I'_i, Z_i^*, C'_1, C'_2)$$

output: $\{I''_i\}_{i=1}^N$

3. Results and Evaluations

In this section, three representative groups of HMR-LC images in Table 1 were used to evaluate the proposed algorithm. By visual judgment and quantitative evaluation, the superior performance and the broad applicability of the proposed method in batch registration for HMR-LC images was verified.

Table 1. Overview of the experimental data.

Type	Time	No.	Size	Res ¹	Characteristics	Preprocessing	BFM Method
Ecological reserve	2015–2020	18	3600 × 3900 pixels	0.8 m	Mountainous areas with high vegetation coverage; deciduous vegetation is widely distributed.	Geocoding alignment + ortho-rectification (30 m precision DEM)	SURF 6 × 6 block
Mine production	2015–2021	18	2850 × 3300 pixels	0.8 m	Large-scale mining and waste discharge; evergreen vegetation and deciduous vegetation are staggered.	Geocoding alignment + ortho-rectification (30 m precision DEM)	SIFT 3 × 3 block
Mine environmental treatment	2014–2020	22	3900 × 4000 pixels	0.8 m	Small-scale mining activities, greening treatment, deciduous vegetation is widely distributed	Geocoding alignment	SIFT 5 × 5 block

¹ “Res” represents the spatial resolution.

3.1. Experimental Data and Related Algorithms

The high resolution satellite images used in the experiments are from the GF-2 satellite launched by China Aerospace Science and Technology Corporation (CASC). The satellite is a solar synchronous return orbit satellite with a short return cycle. It has been put into use in 2014. So far, it has accumulated a large number of multi-temporal high resolution images in the target area. We selected the panchromatic image of the satellite with high resolution (0.8 m) for correlation registration experiment. The three groups of images were aligned and cropped by using the geocoding information of the satellite images. Among them, the images of the ecological reserve and mine production area were ortho-rectified with 30 m precision DEM. Due to the limitation of positioning and model accuracy, there were still different degrees of alignment errors in the image sequence after course registration.

Three image sequence sets of complex terrain areas with different types of low-stability features were selected in the experiment. Among them, the ecological reserve is an ecological protection area dominated by natural landscape, and the unstable features are mainly the seasonal changes of vegetation such as deciduous trees and shrubs. The mine production area belongs to an iron mine area under production, and the unstable features are mainly concentrated in the mining area and waste discharge area. The mine environmental treatment area had experienced large-scale limestone surface mining before 2010. During the imaging period, there were only sporadic mining activities in this area, and greening treatment was started in some areas. Unstable features are mainly small-scale surface excavation, manual treatment, and seasonal changes of vegetation. The three experimental areas have the characteristics of complex terrain, large imaging time span, and large changes of land-cover in different time sequences.

In the visual evaluation part, we displayed and compared the results of stable feature image recovery, feature matching, and image alignment in three types of regions. In the quantitative evaluation part, the number of feature matching inner points of the global image sequence in each type of region is compared and analyzed. Through normalized correlation coefficient (NCC), mutual information (MI), and root mean square error (RMSE), the registration results of different types of low-stability remote sensing images are compared and analyzed.

The SIFT, SURF, and ORB algorithms are used in feature extraction and matching. After matching based on the sum of squared differences (SSD) of the feature, all three

algorithms use consistency screening to purify matching points. For all the experiments, the RANSAC threshold was 0.1.

In the registration display, we compared the proposed method with the following four registration algorithms: The registration method by rank minimization (RRM) [49], piecewise linear mapping (PLM) [50], optical flow modification (OFM) [13], and as-projective-as-possible (APAP) [48]. RRM uses the low-rank structure as an effective constraint for registration tasks. In the experiment, the RRM patch size is set to the same size as the feature matching block. The PLM integrated into ENVI estimates the local affine transformation model for every triangle with three feature points. In PLM, the auto tie point generation matching method is cross-correlation. OFM uses optical flow field estimation and abnormal optical flow modification for remote sensing image registration for complex terrain areas. Due to the weak ability of the optical flow algorithm to deal with large displacement, we first used SIFT to obtain global parameters in the OFM experiment. The APAP algorithm uses the moving direct linear transformation (MDLT) method to estimate the local parameters, by providing higher weights to the closer feature points and lower weights to the farther ones. The SIFT algorithm was used in APAP. The feature method used by CCFR, as well as the parameters for generating block stable feature images, are shown in Table 1.

In order to verify the effectiveness of the LRC-BRE method, we designed ablation experiments in the registration experiments of the ecological reserve and mine production areas. We assume that the image after geocoding alignment and ortho-rectification by 30 m precision DEM can effectively reduce the local distortion of complex terrain and better realize the basic alignment requirements of Hypothesis 2. Based on the above, we directly carried out LRC-BRE and batch registration according to the process shown in Figure 3, which is represented by LRC-BRE-BR subsequently. The feature extraction and fitting method used by LRC-BRE-BR is same to CCFR. Because the initial alignment of the mine environmental treatment area images without ortho-rectification was low, the effect of the ablation experiment was poor.

3.2. Visual Quality

3.2.1. Ecological Reserve

Image restoration result: we used LRC-BRE and CCFR to restore stable feature images. Figure 6 shows the global mean image and local areas of the relevant results. The distortion of ground features in local areas is serious, and the mountains and rocks are raised. There are obvious differences in the contour, shadow and texture of ground features in different seasons. Figure 6a is the original image sequenced by ortho-rectification and geocoding alignment, which shows that there are a few large offset images (marked with a red box), and the offset direction is random. Figure 6b is the stable feature image sequence restored by the LRC-BRE. The results show that the stable feature image is clearly restored, and the unstable feature and temporary noise are removed. Including the image with large offsets in the initial alignment, all of the restored stable feature images are aligned in the spatial position. In addition, the comparison between the red boxes in Figure 6a,b shows that the texture difference between the original image with low alignment and the restored stable feature image is relatively large. Figure 6c is the stable feature image restored by CCFR, which is closer to the original image texture than Figure 6b; the difference is most significant in area V. Due to the failure of block feature matching in the coarse registration stage of CCFR, 6cI (marked with a blue box) did not generate the block stable feature image. Figure 6d is the final result of batch alignment using CCFR, showing that all local images have good spatial alignment.

Figure 6. Global mean image (3600×3900 pixels) and local areas (600×650 pixels) of the remote sensing image sequence and relevant results by LRC-BRE and CCFR in an ecological reserve. The leftmost images of (a–d) are the global mean images of the original image sequence, the stable feature image sequence restored by LRC-BRE, the stable feature image sequence restored by CCFR, and the image sequence registered by CCFR, respectively. The right images of (a–d) are the local area of the above image sequence, whose spatial position corresponds to the yellow box on the left image.

Feature matching result: In the experiment, we selected two cross-seasonal local area images from the original image of the ecological reserve. Figure 7a–c are the results of features extracted and matching of image pairs (Figure 6aI imaged in August 2017 and Figure 6aII imaged in November 2019) through the SURF, SIFT, and ORB algorithms, respectively, showing that the number of matching inner points (yellow line segment connection points) obtained by the three feature matching algorithms was small and the error rate was high. Figure 7d–f are respectively the results of feature matching by the SURF algorithm between Figure 6aI with its corresponding stable feature image restored by LRC-BRE (Figure 6bI), the best alternative stable feature image extracted by LRC-BRE (Figure 6bIV), and the best alternative stable feature image extracted by CCFR (Figure 6cIV).

The results show that the three groups of image pairs achieved correct feature matching, and the number of matching inner points increased in turn. Figure 7g–i respectively show the feature matching by SURF between Figure 6aII and the corresponding stable feature image restored by LRC-BRE (Figure 6bII), the best alternative stable feature image restored by LRC-BRE (Figure 6bIII), and the corresponding stable feature image restored by CCFR (Figure 6cII). The results show that the three groups of image pairs achieved correct feature matching, and the number of matching inner points increased in turn.

Figure 7. Feature extraction and matching results of local area image (600×650 pixels) and restored stable feature image of ecological reserve. The feature algorithms used are (a) SURF, (b) SIFT, (c) ORB, and (d–i) SURF. The roman numeral in the lower right corner is the image number, which corresponds to the image with the same number and background color in Figure 6. The green cross represents the detected key points, the red circle represents the filtered outer points, and the green circle and its associated yellow connecting line are the reserved inner point pairs.

Image registration result: Taking the global image in Figure 8a of Figure 6aI as the reference image and the global image in Figure 8b of Figure 6aII as the sensed image, comparative experiments were carried out using different registration methods. Figure 8c is the overlay of Figure 8a,b, which is fuzzy and shows that there is still much ghosting caused by the offset after geocoding registration and ortho-rectification. Figure 8d is made up of the reference image and the aligned results by CCFR, in which the ghosting is basically eliminated. In order to further evaluate the registration results of different algorithms, the yellow rectangular region of the reference image and the results of the RRM, PLM, APAP, OFM, LRC-BRE-BR, and CCFR algorithms are enlarged, respectively. Figure 8e is an enlarged view of the yellow area of Figure 8c, with a large position offset in both vertical

and horizontal directions. Figure 8f–i show that the results of the relevant registration algorithm have different degrees of offset in both the vertical and horizontal directions. Among them, the RRM method (Figure 8f) has the largest alignment error. Due to the obvious difference between images and textures, the tie point automatically generated by PLM using regional cross-correlation is not ideal, and the alignment result (Figure 8g) has a large offset. As the area is dominated by a natural landscape with few significant features, the correctly matched inner points obtained by SIFT are very few and unevenly distributed, and are mainly concentrated in the upper left area with roads (Figure 8b, red dotted line area). Therefore, the alignment offset of the APAP method (Figure 8h) in area #II is much larger than that in area #I. The OFM algorithm (Figure 8i) corrected and weakened the alignment error of area #II to some extent. After batch registration using the CCFR method (Figure 8k), the alignment of the target image pairs performed well. LRC-BRE-BR (Figure 8j) alignment results take the second place. For example, the stones in the central area (at the junction of gray and green) in Figure 8jII have a small lateral offset to the right with respect to Figure 8kII.

Figure 8. Registration results of image pairs in ecological reserves. (a) Reference image (3600×3900 pixels). (b) Sensed image (3600×3900 pixels). (c) The overlap of the original images (shown in pseudo color). (d) The overlap of the reference image and the result of the proposed algorithm. Magnified images (160×160 pixels) from (e) the original images, (f) RRM, (g) PLM, (h) APAP, (i) OFM, (j) LRC-BRE-BR, and (k) CCFR. In them, the green area of the magnified images is the reference image, and the gray area is the registered image. A red point suggests that the processed image is well-aligned with the reference image.

3.2.2. Mine Production Area

Image restoration result: In the right local areas of the original image sequence shown in Figure 9a, the ground features are mainly a tailings pond (marked by a red circle) with smelting waste. With the discharge of iron ore production waste, the edge area of the tailings pond is constantly changing, and the water content of deposits in tailings pond is also changing. Figure 9a shows that there are a few large offset images (marked with a red box), and the offset direction is random. Figure 9b is the stable feature image sequence

restored by LRC-BRE, which shows that the topographic features of the overlapping area are recovered and the temporary interference noise (such as clouds and shadows in Figure 9bIV) is removed. The textures of some stable feature images in Figure 9b restored from the original large offset images (marked with a red box) are fuzzy. However, the textures of the stable feature images in Figure 9c restored by CCFR from the original large offset images (marked with a red box) are clear, and their spatial positions are aligned. Figure 9d shows that all images after batch registration by CCFR are spatially aligned well.

Figure 9. Global mean image (2850×3300 pixels) and local areas (950×1100 pixels) of the remote sensing image sequence and relevant results by LRC-BRE and CCFR in the mine production area. The leftmost images of (a–d) are the global mean images of the original image sequence, the stable feature image sequence restored by LRC-BRE, the stable feature image sequence restored by CCFR, and the image sequence registered by CCFR, respectively. The right images of (a–d) are the local areas of the above image sequence, whose spatial position corresponds to the yellow box of the left image.

Feature matching result: In the experiment, we selected two local area images with a large time span from the original image sequence of mine production. Figure 10a–c are the results of features extracted and matching of the image pair (Figure 9aI imaged in June 2020 and Figure 9aII imaged in December 2016) through the SURF, SIFT, and ORB algorithms respectively, showing that the number of matching inner points is small and the distribution is very uneven, most of which are concentrated in the tailings dam area at the upper right

(marked by a purple box). Due to the continuous discharge of semi fluid slag waste in the edge area of the tailings pond (marked by a red circle), where the image texture difference is obvious and the number of matching inner points is small. The shadow occlusion area on the right side of the image (marked by a magenta circle) also does not obtain the matching feature inner points. Figure 10d shows that due to the good initial alignment and lower vegetation noise interference in the winter image, the SIFT feature matching inner points of Figure 9aII and its corresponding stable feature image (Figure 9bII) restored by LRC-BRE are dense. Figure 10e–g are the SIFT feature matching of Figure 9aI with its corresponding stable feature image (Figure 9bI) restored by LRC-BRE, the best alternative stable feature image (Figure 9bIII) restored by LRC-BRE, and its corresponding stable feature image (Figure 9cI) restored by CCFR, respectively. The results show that the matching inner points obtained by the three groups of image pairs have a uniform distribution, the number of which increases in turn. Figure 10h,i shows the results of purifying the inner points in Figure 10f,g respectively, by MOF, and the inner points with a large offset in the red circle area are eliminated.

Figure 10. Feature extraction and matching results of local original image (950×1100 pixels) and stable feature image of mine production area. The feature algorithms used are (a) SURE, (b) SIFT, (c) ORB, and (d–i) SIFT. (h) and (i) filtered the inner points again by MOF. The roman numeral in the lower right corner is the image number, which corresponds to the image with the same number and background color as in Figure 9.

Image registration result: Different registration methods were carried out with the global image in Figure 11a of Figure 9aI as the reference image and the global image in Figure 11b of Figure 9aII as the sensed image. Figure 11c is the overlay of Figure 11a,b, which shows that there is still much ghosting caused by certain offsets after geocoding registration and ortho-rectification. Figure 11d shows that the ghosting of the global image pair after registration using the CCFR algorithm is weakened and the texture of the yellow dotted line area is clearer, but the ground surface in the lower left area changes significantly due to surface mining, tailings discharge, land cultivation, and other reasons. Figure 11e is an enlarged view of the yellow area of Figure 11c, with a large position offset in both

vertical and horizontal directions. Figure 11f–i show that the registration results of the RRM, PLM, APAP, and OFM algorithms have different degrees of offset, and the offset degree of region #I is generally larger than that of region #II. Figure 11f shows that the registration result of the RRM method has a large alignment error. Because some artificial features with high feature significance are distributed in the overlapping area, the alignment result of the APAP algorithm (Figure 11h) is relatively good. The alignment effect of the OFM algorithm (Figure 11i) is relatively good, but the registered image on the right side (marked by a red box) of region #II is over-stretched. After batch registration by LRC-BRE-BR (Figure 11j) and the CCFR method (Figure 11k), the alignment of the target image pair is good, and the image texture is not excessively stretched. The offset of the tailings dam section in the red box in area #I of Figure 11j,k belongs to the actual displacement of ground objects, which is related to the continuous heightening of the dam body to the left with the increase of tailings excreta.

Figure 11. Registration results of image pairs in the mine production area. (a) Reference image (2850 × 3300 pixels). (b) Sensed image (2850 × 3300 pixels). (c) The overlap of the original images (shown in pseudo color). (d) The overlap of the reference image and the result of the proposed algorithm. Magnified images (180 × 180 pixels) from (e) the original images, (f) RRM, (g) PLM, (h) APAP, (i) OFM, (j) LRC-BRE-BR, and (k) the CCFR method. In them, the green area of the magnified images is the reference image, and the gray area is the registered image. A red point suggests that the processed image is well aligned with the reference image.

3.2.3. Mine Environmental Treatment Area

Image restoration result: The image of the mine environmental treatment area is not ortho-rectified, and the initial alignment degree is low. CCFR was used for stable feature image restoration and batch registration in this area. LRC-BRE was only used to restore the stable feature image. The right side of Figure 12 is the local areas of the original image sequence and the stable feature image sequence recovered by different methods. The upper left corner of this local areas (for example, the red box in Figure 12a) is the quarry still being excavated during the imaging period, and the surface changed significantly over different periods. The right side of Figure 12a is the local areas of the original image sequence aligned

by geocoding, showing that the initial alignment degree is low, and there are generally large offsets in the location of roads and quarries in the area. Its left global mean image texture is fuzzy and contains heavy ghosting. The right side of Figure 12b shows the local areas of the stable feature image recovered by the LRC-BRE method. The result shows that the image texture is blurred and the distortion is high. The right side of Figure 12c shows the local areas of the stable feature image recovered by the CCFR method. The results show that most areas recover the stable feature image, with clear ground objects and consistent spatial position. Only a few regions do not generate stable feature images. Figure 12d shows that all images after batch alignment using CCFR have been spatially aligned well, and the left global mean image texture is clear.

Figure 12. Global mean image (3900×4000 pixels) and local areas (780×800 pixels) of the remote sensing image sequence and relevant results by LRC-BRE and CCFR in the mine environmental treatment area. The leftmost images of (a–d) are the global mean images of the original image sequence, the stable feature image sequence restored by LRC-BRE, the stable feature image sequence restored by CCFR, and the image sequence registered by CCFR, respectively. The right images of (a–d) are the local areas of the above image sequence, whose spatial position corresponds to the yellow box of the left image.

Feature matching result: In the experiment, we selected two local area images with a large time span but similar seasons from the original image sequence of the mine environmental treatment area. Figure 13a–c are the results of features extracted and the matching of image pairs (Figure 12aI imaged in late April 2020 and Figure 12aII imaged in early April 2016) through the SURF, SIFT, and ORB algorithms, respectively, showing that the number of matching inner points is small and the error rate is high. This is the main reason that the quarry area (marked by a red circle) in the image pair has changed significantly,

although the seasonal change of vegetation and the change of solar altitude angle are small. Figure 13d shows that the number of SIFT feature matching inner points of Figure 12aI and the best alternative stable feature image (Figure 12aIV) recovered by CCFR is large and evenly distributed. In Figure 13d, there are still a few internal points with a position offset, which is effectively filtered by MOF in Figure 13e (marked by a red circle). Figure 13f shows that the SIFT feature matching inner points of Figure 12aII and its corresponding stable feature images recovered by CCFR are densely distributed.

Figure 13. Feature extraction and matching results of local original image (780×800 pixels) and stable feature image of mine treatment area. The feature algorithms used are (a) SURF, (b) SIFT, (c) ORB, and (d–f) SIFT. (e) Filtering the inner points again by MOF. The roman numeral in the lower right corner is the image number, which corresponds to the image with the same number and background color in Figure 12.

Image registration result: Taking the global image in Figure 14a of Figure 12aI as the reference image and the global image in Figure 14b of Figure 12aII as the sensed image, comparative experiments were carried out using different registration methods. Figure 14c is the overlay of Figure 8a,b, which shows that there is much ghosting caused by the offset after geocoding registration. Figure 14d shows that the ghosting of the global image pair registered by the CCFR algorithm is significantly weakened, except that the image edge and some mining areas still have color superposition. Figure 14f–i shows respectively that the registration results of RRM, PLM, APAP, and OFM have different degrees of offset in area #I, which is related to the proximity of the mining area with obvious changes. The registration results of RRM (Figure 14f) and PLM (Figure 14g) also have obvious offsets in area #II. Since a small number of artificial objects with high feature significance are distributed near area #II, APAP (Figure 14h) and OFM (Figure 14i) achieve good alignment of trunk roads in area #II. However, the right resampled image (marked by a red box) of OFM (Figure 14i) in area #II has obvious tensile deformation. After batch registration by CCFR, the alignment of the target image pair (Figure 14j) is good, and the image is not excessively stretched.

Figure 14. Registration results of image pairs in the mine treatment area. (a) Reference image (3900 × 4000 pixels). (b) Sensed image (3900 × 4000 pixels). (c) The overlap of the original images (shown in pseudo color). (d) The overlap of the reference image and the result of the proposed algorithm. Magnified images (180 × 180 pixels) from (e) the original images, (f) RRM, (g) PLM, (h) APAP, (i) OFM, and (j) the CCFR method.

From the visualization results of the above three groups of experiments, it can be seen that LRC-BRE and the block-based LRC-BRE (CCFR) are effective for extracting stable feature images as reference images. For the registration between original images and stable feature images extracted, the BFM-LLT method has good results in feature matching and local alignment. The CCFR method can realize the batch alignment of three different types of HMR-LC, and achieve better performance.

3.3. Evaluation of Registration Results

In order to evaluate the accuracy of the image registration, we used the normalized correlation coefficient (NCC), mutual information (MI), and root-mean-square error (RMSE) metrics to conduct the quantitative assessment. Both NCC and MI can reflect the degree of similarity between two images, and are popular evaluation indexes used in the quality assessment of image registration.

The NCC metric is used to provide an overall judgement as to whether the sensed image is well-aligned with the reference image [22,51]:

$$NCC(I_1, I_2) = \frac{\sum_{i=1}^N [I_1(i) - \bar{I}_1] [I_2(i) - \bar{I}_2]}{\sqrt{\sum_{i=1}^N [I_1(i) - \bar{I}_1]^2} \sqrt{\sum_{i=1}^N [I_2(i) - \bar{I}_2]^2}} \quad (11)$$

where $I_1(i)$ and $I_2(i)$ are the intensity values of the i -th pixel in image I_1 and image I_2 , respectively; and $\bar{I}_1$ and $\bar{I}_2$ are the corresponding average intensity values. N is the pixel number. The available range of NCC is $[-1, 1]$, and NCC generally belongs to $(0, 1)$. A larger value of NCC indicates a more accurate registration result.

The MI metric is used as a cost function in similarity measure-based approaches to registration [52]:

$$MI(I_1, I_2) = \sum_{g_1 \in I_1} \sum_{g_2 \in I_2} p_{I_1, I_2}(g_1, g_2) \log \left[\frac{p_{I_1, I_2}(g_1, g_2)}{p_{I_1}(g_1) p_{I_2}(g_2)} \right] \quad (12)$$

The available range of MI is (0, 2). A larger value of MI represents a more precise registration result.

The RMSE metric focuses on evaluating the registration result by calculating the average distance of the corresponding points in the reference and aligned image:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\Delta x_i)^2 + (\Delta y_i)^2} \quad (13)$$

where N is the number of evaluated points, and x_i and y_i are the residual differences of the i -th checkpoint pair in the x and y directions. A smaller RMSE indicates a better result.

The three groups of experiments correspond to Figures 8, 11 and 14 of the visual evaluation experiment, respectively. The quantitative evaluation results are shown in Table 2.

Table 2. Quantitative evaluation of the registration results in the three experiments.

	Ecological Reserve			Mine Production			Mine Environmental Treatment		
	NCC	MI	RMSE/pixels	NCC	MI	RMSE/pixels	NCC	MI	RMSE/pixels
Original	0.28136	0.059138	0.37336	0.14756	0.026849	0.28831	0.28807	0.058247	0.27997
RRM	0.40549	0.10822	0.3533	0.27153	0.054601	0.26371	0.45068	0.13519	0.25229
PLM	0.50274	0.2535	0.31822	0.29541	0.093957	0.27233	0.49631	0.16344	0.24489
APAP	0.63317	0.32688	0.3533	0.34195	0.097451	0.25461	0.53916	0.19488	0.23318
OFM	0.6786	0.34853	0.30167	0.33511	0.09357	0.2557	0.50141	0.17111	0.24879
LRC-BRE-BR	0.68687	0.35448	0.31131	0.36022	0.10722	0.25249	null	null	null
CCFR	0.70928	0.39685	0.28171	0.38415	0.11548	0.24826	0.6413	0.30463	0.22609

According to Table 2, we can draw the following conclusions: (1) Quantitative evaluation of the same type of areas and different registration methods: Due to the global offset in the original image alignment of the three areas, the NCC and MI indicators are low, and the RMSE/pixels indicators are high. The registration result of the RRM algorithm is the worst, and there are generally large offsets. Only in the local area with significant features such as artificial objects, the feature-based APAP method achieves a relatively good alignment result. The registration indicators of the OFM algorithm are better than RRA and APAP. The proposed CCFR methods achieved the best quantitative evaluation. In the ecological reserve and mine production area by geocoding registration and ortho-rectification, the quantitative evaluation of LRC-BRE-BR is second only to CCFR. (2) Quantitative evaluation of the same registration method and different type areas: Since there are only seasonal vegetation changes in land-cover, the NCC and MI indicators of multiple registration methods in the ecological reserve are overall higher than those in the other two areas. Due to the large surface changes caused by mining and waste discharge, the NCC and MI indicators of multiple registration methods in the mine production area are lower than in the other two areas. The overall similarity of RMSE/pixels indicators to different regions is slightly inconsistent with NCC and MI. For example, the RMSE/pixels indicators of ecological reserves are the highest. However, the change trend of RMSE/pixels in the same area and the same registration method is consistent with NCC and MI. That is, the larger the NCC, the larger the MI and the smaller the RMSE/pixels.

In general, the CCFR method is effective for HMR-LC images' registration. Especially when the land-cover changes and influence of radiation differences are significant, the robustness of CCFR is superior to other registration methods of comparison.

4. Discussion

The previous section validates the effect of the proposed method in image registration. This section discusses the advantages and influencing factors of indirect registration with stable feature images restored by LRC-BRE as the reference, and introduces the work of CCFR in different satellite images.

4.1. Quantitative Comparison of Feature Matching Results

In order to objectively verify the influence of the stable feature image as a reference image on feature matching, Figure 15 shows the results of direct feature matching and indirect feature matching with the reference of a stable feature image as the reference for the multi-temporal global image sequence in the three experimental areas in Section 3.2. For the image pairs with a small texture difference from the specified image I_φ in the sequence, the effect of direct feature matching is better; however, for the HMR-LC images sequence, the image pairs with small texture differences are relatively rare. Details are as follows:

Figure 15. Global multi-temporal image sequence feature matching quantization results in the three experiments. (a) Ecological reserve. (b) Mine production area. (c) Mine environmental treatment area. The abscissa is the imaging date of the sensing image, and the ordinate is the number of feature matching inner points. The blue polyline, gray polyline, and yellow polyline are, respectively, the number of feature matching inner points with the lowest outlier image I_φ , the stable feature image Z_i^{*L} extracted by LRC-BRE, and the stable feature image Z_i^{*C} extracted by CCFR.

For the ecological reserve image sequence, the statistics of matched inner points with BFM(SURF) are shown in Figure 15a. The results show that the number of inner points

obtained by image pair (I_i, I_φ) direct feature matching fluctuates greatly. The number of inner points between image pairs (I_i, I_φ) imaged in the same or similar seasons is the largest among the three types of broken lines, but that between image pairs (I_i, I_φ) imaged across seasons is lower than those of (I_i, Z_i^{*L}) and (I_i, Z_i^{*C}) , or even close to 0. This is related to the large seasonal variation of deciduous woody vegetation in the area. The fluctuation of the number of inner points of the image pair (I_i, Z_i^{*L}) and (I_i, Z_i^{*C}) sequence is small, and the situations of broken line valleys is better than (I_i, I_φ) .

For of mine production area image sequence, the statistics of feature matching of inner points with BFM(SIFT) are shown in Figure 15b. The results show that only a few (I_i, I_φ) with similar seasons have more inner points than those of (I_i, Z_i^{*L}) and (I_i, Z_i^{*C}) in the same period. In most cases, the number of inner points between (I_i, Z_i^{*C}) is the largest, and the number of inner points between (I_i, Z_i^{*L}) is the second. In addition to seasonal fluctuation of the number of inner points, due to the surface changes caused by the continuous large-scale mine production in the area during the imaging period, with increasing of the time interval from I_φ (red dot), the number of inner points of (I_i, I_φ) shows an overall downward trend. However, the number of inner points in (I_i, Z_i^{*L}) and (I_i, Z_i^{*C}) is relatively stable.

For the mine environmental treatment area image sequence, the statistics of feature matching of inner points with BFM(SIFT) is shown in Figure 15c. The results show that the number of inner points between (I_i, I_φ) with a large seasonal span is smaller than that between (I_i, Z_i^{*C}) . In addition, the number of inner points of (I_i, I_φ) and (I_i, Z_i^{*C}) decreases as the time interval from I (red dot) increases. This is related to the land-cover changes caused by the strengthening of mine ecological restoration in the region since 2018.

4.2. Correlation Factors of Restored Stable Feature Image Quality

According to Theorem A1 in the Appendix A, the restored stable feature image is composed of the first R principal component vectors $\{u_i\}_{i=1}^R$ generated by the low-rank decomposition from the initial image sequence. Moreover, the difference of the coefficient $\{\sigma_i\}_{i=1}^R$ corresponding to the principal component vector sequence is exponential. Therefore, the restored quality of the stable feature image sequence is closely related to the principal component vectors sorted in the front. In addition, according to the Inference A2 in the Appendix A, when the texture difference noise remains unchanged, the restoration quality of the stable feature image sequence is positively correlated with the spatial consistency of the initial image sequence.

To facilitate the discussion on the correlation factors of the restored stable feature image quality, taking the local image sequences of ecological reserves in Section 3.2.1 as an example, the texture states of images D_i corresponding to the principal component vector u_i extracted by low-rank decomposition from different initial image sequences are compared. The image D_i is generated by the following:

$$D_i = \text{vec}^{-1}(\text{abs}(u_i)) \quad (14)$$

where $\text{abs}(\cdot)$ means taking an absolute value for each element of the vector, and vec^{-1} indicates the reverse operation of stacking. Figure 16a is the first four principal component images generated by low-rank decomposition from nine local images of the first row in Figure 6a, in which the proportion of large offset images (marked by a red box) is 2/9. Figure 16b was generated by low-rank decomposition from all 18 local images in Figure 6a, in which the proportion of large offset images (marked by a red box) is also 2/9. Figure 16c was generated from 16 local images (excluding severely outlying block images after coarse registration by CCFR) in Figure 6a.

Among the three groups of results in Figure 16, the texture features of principal component images (especially the first two) (Figure 16c) generated by the CCFR method are the clearest through rough registration based on blocks and eliminating the severely outlying block images used by low-rank decomposition. When the proportion of large offset images in the initial image set is the same, the principal component images (Figure 16b)

with more images participating in low-rank decomposition have a clearer texture and less ghosting than principal component images (Figure 16a), with less images participating. This shows that under the same initial alignment, the more initial images are involved in batch alignment, the texture spatial consistency of extracted stable feature image is better.

Figure 16. Rows (a–c) of the images correspond to the first four principal component vectors, which were extracted by low-rank decomposition from 9 local images of the first row in Figure 6a, all 18 local images in Figures 6a and 17 local images (excluding severely outlier block images after coarse registration by CCFR) in Figure 6a, respectively. Each block size is 600×650 pixels.

4.3. Registration with Different Optical Satellite Images

For the registration of remote sensing image sets with different resolutions, CCFR can still work after that the low-resolution remote sensing images are interpolated and resampled in accordance with high resolution images. To check how the proposed algorithm works in different optical satellite images, we collected images (Figure 17a) from the GF-1 and GF-2 optical satellites, which have different resolutions.

Figure 17. The mine areas of the remote sensing images from the GF-2 and GF-1 optical satellites. (a) The original image set. (b) A local magnification (32×40 pixels) of the GF-1 image; (c) The result of sampling on (b); (d) A local magnification (80×100 pixels) of the GF-2 image.

The dataset and pretreatment are described in Table 3. The imaging area belongs to the mine area, with mainly underground mining. The surface changes of multi-temporal images are mainly reflected by the continuous discharge and leakage of mine waste, as well as radiation differences. These images were preliminarily aligned based on geographic information, but there are still different degrees of offset. Because the images have different resolutions, we up-sampled the GF-1 satellite images according to the resolution of GF-2 satellite images.

Table 3. Overview of the experimental data of GF-1 and GF-2 satellites.

Type	Time	Sensor	No.	Size	Res ¹	Characteristics	Preprocessing
Mine production	2014–2021	GF-1	8	960×1200 pixels	2 m	Large-scale underground mining, continuous discharge and leakage of mine waste.	Geocoding alignment, up sampling
		GF-2	16	2400×3000 pixels	0.8 m		Geocoding alignment

¹ “Res” represents the spatial resolution.

For comparison, the results registered for GF-1 (Figure 18b) referring to GF-2 (Figure 18a) are shown in Figure 18 by different methods. The imaging time of Figure 18a is winter, and the partial area of the tailings pond, road, and ore storage yard are covered with snow, which is significantly different from the summer image of Figure 18b. Figure 18c is the overlay of Figure 18a,b, which shows that there is still much ghosting caused by offset. Figure 18d is made up of the reference image and the aligned results by CCFR, in which the ghosting is eliminated. Figure 18e is an enlarged view of the yellow area of Figure 18c, with a large position offset in both the vertical and horizontal directions. Figure 18f–i show the results of RRM, PLM, APAP, and OFM, respectively. There is a large offset in areas #I and

#II of Figure 18f, and the registered image is overstretched (marked with a red box). Due to the influence of complex terrain and low-stability land-cover, the feature points extracted by PLM and APAP have small location errors. In Figure 18g,h, the offset is reduced, but not completely eliminated. OFM (Figure 18i) can eliminate the offset, but the registered image has obvious over-stretching (marked with a red box). CCFR (Figure 18j) achieved a precisely aligned result, and the registered image was not excessively stretched.

Figure 18. Registration results of image pairs in the mine production area from the GF-1 and GF-2 satellites. (a) Reference image from GF-2 (2400 × 3000 pixels). (b) Sensed image from GF-1 (original size 960 × 1200 pixels). (c) The overlap of the original images (shown in pseudo color). (d) The overlap of the reference image and the result of the proposed algorithm. Magnified images (200 × 200 pixels) from (e) the original images, (f) RRM, (g) PLM, (h) APAP, (i) OFM, and (j) the CCFR method. In them, the green area of the magnified images is the reference image, and the gray area is the registered image. A red point suggests that the processed image is well-aligned with the reference image.

Table 4 lists the quantitative evaluations of the above registration results in terms of NCC, MI, and RMSE. The experimental results show that CCFR achieved good alignment in both visual and quantitative aspects.

Table 4. Quantitative evaluation of the registration results in the experiment with the GF-1 and GF-2 satellites.

	Ecological Reserve		
	NCC	MI	RMSE/pixels
Original	0.28861	0.095284	0.39303
RRM	0.30568	0.09869	0.39306
PLM	0.47723	0.29056	0.36538
APAP	0.50373	0.44417	0.39686
OFM	0.50852	0.41378	0.37514
CCFR	0.51342	0.44967	0.3638

Therefore, when the difference in image resolution is less than four-fold, the same pixel image set can be obtained after interpolation, and the CCFR method can be effectively used.

5. Conclusions

To decrease the influence of low-stability land-cover and large radiation differences on the conventional image registration methods, we constructed a new comprehensive registration framework for robust registration of HMR-LC combined with the stable feature image restoration strategy. By the LRC-BRE method proposed, a set of stable feature images can be restored from the multi-temporal image sequences, and the most suitable stable feature image can be selected as the reference image for each sensed image. The proposed BFM-LLT method is effective for coarse registration of HMR-LC images with medium and low texture differences and registration with stable feature images as a reference. In the BFM step, the MOF operator can reduce the impact of the high background noise and matching inner point position offset caused by distorted terrain on the feature matching accuracy. In the LLT step, the DWBF operator can solve the influence of uneven feature distribution and the similarity difference of the inner point area on the accuracy of complex terrain fitting. By integrating BFM-LLT and block-based LRC-BRE, the CCFR method can improve the quality of the extracted stable feature image as a reference, and has good robustness for the registration of HMR-LC images with large formats and significant differences. For the three types of realistic HMR-LC image sequences used in the experiments, the results qualitatively and quantitatively show that the proposed algorithm can achieve a reliable registration accuracy. For registration with different optical satellites, the CCFR method can be effectively used after the data pre-processing. For example, low-resolution remote sensing images should be interpolated and resampled in accordance with high resolution images in the set.

In future work, the CCFR framework can set $C1 = 1$ and $C2 = 1$ as the initial parameters of feature matching blocks to align the image sequences that have not undergone geocoding registration, and can introduce an iterative strategy to continuously improve the alignment result by increasing the number of blocks and executing CCFR many times.

Author Contributions: Conceptualization, P.Z. and X.L.; methodology, P.Z.; validation, P.Z. and X.L.; formal analysis, P.Z., X.L. and C.W.; investigation, P.Z. and W.W.; resources, P.Z. and W.W.; data curation, P.Z. and W.W.; writing—original draft preparation, P.Z.; writing—review and editing, P.Z., X.L. and C.W.; funding acquisition, P.Z. and Y.M.; supervision, X.Q. and C.W.; project administration, P.Z. and Y.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the R&D Program of the Beijing Municipal Education Commission (under grant number: KM202111417007).

Data Availability Statement: Not applicable.

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

Appendix A

Hypothesis A1. It is assumed that $\{I_i\}_{i=1}^N \in \mathbb{R}^{w \times h}$ is a set of fully aligned low-stability multi-temporal remote sensing image sequences of the same region. The texture difference noise of image I_i is recorded as e_i , which is mainly composed of the superposition of noise e_i^c caused by the change of land-cover and noise e_i^r caused by radiation difference, $e_i = e_i^c \oplus e_i^r$. $\{E_i^0\}_{i=1}^N \in \mathbb{R}^{w \times h}$ represents the set of matrices that complement 0 for the vacant area of surface texture noise e_i according to the size of I_i . The elements of matrix E_i^0 contain positive and negative values. $\{Z_i^0\}_{i=1}^N \in \mathbb{R}^{m_1 \times m_2}$ is the set of images removing texture noise e_i that is expected to be recovered from the image sequence $\{I_i\}_{i=1}^N$. Because the noise caused by the unstable surface is eliminated, we call image Z_i^0 a stable feature image. There are $I_i = Z_i^0 + E_i^0$, $E_i^0 = E_i^c \oplus E_i^r$, $g(I_1) = \dots = g(I_N)$, $g(Z_1^0) = \dots = g(Z_N^0)$, $Z_1^0(x, y) = \dots = Z_N^0(x, y)$, where $i \in \{1, \dots, N\}$, “+” represents the sum of matrices and “ $\oplus$ ”

represents the noise superposition, $g(\mathbf{Z})$ represents the two-dimensional spatial position of image $\mathbf{Z}$, and $\mathbf{Z}(x, y)$ represents the pixel gray value of image $\mathbf{Z}$ at two-dimensional spatial coordinates (x, y) .

If $\text{vec}: \mathbb{R}^{w \times h} \rightarrow \mathbb{R}^m$ denotes the operator that selects an m -pixel region of interest (typically $m \gg N$) and stacks it as a vector, then there are $\mathbf{M} = [\text{vec}(\mathbf{I}_1), \dots, \text{vec}(\mathbf{I}_N)] \in \mathbb{R}^{m \times N}$, $\mathbf{L}^0 = [\text{vec}(\mathbf{Z}_1^0), \dots, \text{vec}(\mathbf{Z}_N^0)] \in \mathbb{R}^{m \times N}$ and $\mathbf{S}^0 = [\text{vec}(\mathbf{E}_1^0), \dots, \text{vec}(\mathbf{E}_N^0)] \in \mathbb{R}^{m \times N}$. Because of $\mathbf{I}_i = \mathbf{Z}_i^0 + \mathbf{E}_i^0$, there is $\mathbf{M} = \mathbf{L}^0 + \mathbf{S}^0$. In the case of Hypothesis A1, since $\{\mathbf{Z}_i^0\}_{i=1}^N$ is the image set expected to be restored after removing noise in the fully aligned same region, the $\mathbf{L}^0$ constructed by it is a low-rank matrix with a rank close to 1. For large-scale global images, $\mathbf{e}^c$ is a local image noise; after radiation normalization, $\mathbf{e}^r$ is mainly reflected in the local imaging noise caused by the change of solar altitude angle. Therefore, $\mathbf{S}^0$ is a sparse matrix, and $\mathbf{M}$ is an approximate low-rank matrix. Restoring the low-rank structure of $\mathbf{M}$ is a typical low-rank matrix factorization (LRMF) problem, which can be described as:

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{S}\|_2 \quad \text{s.t.} \quad \text{rank}(\mathbf{L}) \leq r, \mathbf{M} = \mathbf{L} + \mathbf{S} \quad (\text{A1})$$

where $\|\cdot\|_2$ denotes the 2-norm and $1 \leq r \ll \min(m, N)$. Then, through each column of matrices $\mathbf{L}$ and $\mathbf{S}$, the stable feature image sequence estimation $\{\mathbf{Z}_i\}_{i=1}^N$ and noise matrix sequence estimation $\{\mathbf{E}_i\}_{i=1}^N$ are recovered.

Due to the influence of multiple factors, the low-stability complex terrain multi-temporal image sequence coarsely aligned by geocoding or image registration usually has different degrees of alignment errors. When the image alignment error conforms to the local random distribution and global contingency, we call this situation basic alignment. Differently from RASL [42] and RRM [49], we regard the imaging difference caused by the offset between the sensed image $\mathbf{I}$ and the stable feature image $\mathbf{Z}$ expected to be restored as noise. We can express this situation as follows:

Hypothesis A2. It is assumed that $\{\mathbf{I}_i\}_{i=1}^N \in \mathbb{R}^{w \times h}$ is a set of low-stability multi-temporal remote sensing image sequences near the same region with basic alignment. Taking the spatial position of the expected restored stable feature image $\mathbf{Z}_i^0$ as a reference, the pseudo texture noise caused by the alignment error is recorded as $\mathbf{e}_i^d$. $\mathbf{e}_i^d$ includes and is limited to the following two cases: (1) There is offset noise caused by alignment errors in the local areas randomly distributed in the space of multi-temporal images. For complex terrain remote sensing images, local offset noise caused by terrain distortion is more common. (2) A few images $\mathbf{I}_i$ ($1 \leq i \leq N_d \ll N$) have offset noise caused by global (or large range) alignment errors. If the texture difference noise of the image sequence conforms to Hypothesis A1, the image $\mathbf{I}_i$ comprehensive noise $\mathbf{e}_i = \mathbf{e}_i^c \oplus \mathbf{e}_i^r \oplus \mathbf{e}_i^d$.

In the case of Hypothesis A2, since $\{\mathbf{Z}_i^0\}_{i=1}^N$ is the image set expected to be restored after removing noise and alignment errors, the $\mathbf{L}^0$ constructed by it is still a low-rank matrix in the basically aligned same region. According to the two limiting conditions of $\mathbf{e}_i^d$ in Hypothesis A2, the matrix $\mathbf{S}^0 = [\text{vec}(\mathbf{E}_1^0), \dots, \text{vec}(\mathbf{E}_N^0)] \in \mathbb{R}^{m \times N}$ still can be considered as sparse. When $\mathbf{S}$ is a sparse large noise matrix, Problem (1) is transformed into a bi-objective optimization problem:

$$\min_{\mathbf{L}, \mathbf{S}} (\text{rank}(\mathbf{L}), \|\mathbf{S}\|_0) \quad \text{s.t.} \quad \mathbf{M} = \mathbf{L} + \mathbf{S} \quad (\text{A2})$$

where $\|\cdot\|_0$ is 0-norm. In Formula (2), $\text{rank}(\mathbf{L})$ and $\|\mathbf{S}\|_0$ are nonlinear and nonconvex, so it is difficult to optimize. Using the robust principal component analysis (RPCA) [44] method, question (A2) is converted into a tractable convex optimization program:

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \quad \text{s.t.} \quad \mathbf{M} = \mathbf{L} + \mathbf{S} \quad (\text{A3})$$

where $\|\cdot\|_*$ is nuclear-norm, $\|\cdot\|_1$ is 1-norm and $\lambda (>0)$ is a compromise factor. By solving the optimization problem (A3), we can recover the low-rank matrix estimate L and the coefficient matrix estimate S . Then the stable characteristic image sequence $\{Z_i\}_{i=1}^N$ and noise sequence matrix $\{E_i\}_{i=1}^N$ are recovered.

Next, combined with the algorithm implementation of low-rank decomposition, we analyze the composition and characteristics of the restored low-rank matrix and the stable feature image. The accelerated proximal gradient (APG) [53], augmented Lagrangian method (ALM) [54], and inexact ALM [55] algorithms can be used to solve the optimization problem (A2). The algorithm generally adopts a singular value threshold (SVT) to iteratively optimize the kernel norm of a low-rank matrix, and uses a soft threshold operator to optimize the 0-norm of a sparse matrix. Through alternating iterative calculation, it gradually approximates to the low-rank matrix L and sparse matrix S . The SVT operator is calculated by partial SVD. The final recovered low-rank matrix can be expressed as:

$$L^* = \sum_{i=1}^N (\sigma_i - \delta)_+ \mathbf{u}_i \mathbf{v}_i^T = \sum_{i=1}^R \sigma_i \mathbf{u}_i \mathbf{v}_i^T \quad (\text{A4})$$

where δ is the threshold of the penalty item, $(x)_+ = \max(x, 0)$. $\mathbf{u}_i$, $\mathbf{v}_i$ and σ_i are the left singular vectors, the right singular vector, and the singular values, respectively. N is the column number of matrix L , and R is the number of $(\sigma_i - \delta)_+ > 0$. According to the composition and low rank of L , it is concluded that:

Theorem A1. For an approximate low-rank data matrix M , through the low-rank decomposition based on the column vector, any column j of the recovered low-rank matrix L can be expressed as $L_{(j)} = \sum_{i=1}^R \sigma_i v_i^j \mathbf{u}_i$, where $\{\mathbf{u}_i\}_{i=1}^R \in \mathbb{R}^m$ and $\{\mathbf{v}_i\}_{i=1}^R \in \mathbb{R}^N$ are unit vectors, v_i^j represents the j -th element of unit vector $\mathbf{v}_i$, and $R = \text{rank}(L)$. Because the matrix L is low-rank, the columns of L are highly correlated and have an approximately linear correlation.

According to the sparsity requirements of Formula (A3) on the recovery matrix S , it is concluded that:

Theorem A2. For an approximate low-rank data matrix M , through the low-rank decomposition based on the column vector, the sparse matrix S recovered must have a partial region S^Ω whose value is equal to or close to 0. According to $M = L + S$, it can be considered that the region Ω of matrix M and L is equal, that is, $M^\Omega = L^\Omega$. At the same time, in order to achieve sparse optimization, the iterative process makes a sparse approximation to the region with the maximum column co-occurrence probability of M .

In the low-rank decomposition process, each iterative update is the overall optimization of L and S estimates. This process does not take the difference of spatial location into account. It is only a principal component purify (PCP) solution based on the matrix value, that is, the intensity information of the image. Therefore, according to Theorems A1 and A2 and Hypothesis A2, we draw Inferences A1.1, A1.2, and A2.

Inference A1.1. For a group of multi-temporal remote sensing image sequences $\{I_i\}_{i=1}^N$ satisfying Hypothesis A2, a group of stable feature image sequences $\{Z_i\}_{i=1}^N$ and a group of corresponding sparse noise matrix sequences $\{E_i\}_{i=1}^N$ can be recovered by low-rank decomposition of the matrix M composed of the images straightened in columns. There are $I_i = Z_i + E_i$, $E_i = E_i^c \oplus E_i^r \oplus E_i^d$, where $\{Z_i\}_{i=1}^N \in \mathbb{R}^{w \times h}$ is highly correlated and has similar texture features at the same spatial location. For any coordinate (x, y) in the global region, there is $Z_1(x, y) \approx \dots \approx Z_N(x, y)$. In addition, there is $g(Z_1) \approx \dots \approx g(Z_N)$ in the spatial position.

Inference A1.2. For the stable feature image sequence $\{Z_i\}_{i=1}^N$ recovered from a set of multi-temporal remote sensing image sequences $\{I_i\}_{i=1}^N$ conforming to Hypothesis A2, there must be some

stable feature images $\mathbf{Z}_i (1 \leq i \leq N_{S1} \leq N)$, in which several local regions $\Omega_j (j \geq 1)$ are completely aligned. For any coordinate (x, y) in the region Ω_j , there is $\mathbf{Z}_1^{\Omega_j}(x, y) = \dots = \mathbf{Z}_N^{\Omega_j}(x, y)$. In addition, there is $g(\mathbf{Z}_1^{\Omega_j}) = \dots = g(\mathbf{Z}_N^{\Omega_j})$ in the spatial position.

According to Theorem A1, each column of the low-rank matrix $\mathbf{L}$ is a different linear combination of the same unit eigenvector group $\{\mathbf{u}_i\}_{i=1}^R$. Therefore, the degree of the outlier of the initial image sequence directly affects the generation of the principal component vector $\{\mathbf{u}_i\}_{i=1}^R$ in low-rank decomposition, the overall restoration quality of the stable feature image sequence $\{\mathbf{Z}_i\}_{i=1}^N$, and the restoration of the local offset region of the individual image. Outliers include texture differences and spatial offsets. We define the overall alignment degree of an image sequence as spatial consistency, and draw a conclusion:

Inference A2. When the texture difference noise remains unchanged, the restoration quality of the stable feature image sequence $\{\mathbf{Z}_i\}_{i=1}^N$ is positively correlated with the spatial consistency of the initial image sequence $\{\mathbf{I}_i\}_{i=1}^N$. The higher the spatial consistency of the initial image sequence $\{\mathbf{I}_i\}_{i=1}^N$, the smaller the overall distortion of the stable feature image restored by low-rank decomposition, and the more sparse the separated noise matrix. For any image $\mathbf{I}_i \in \{\mathbf{I}_i\}_{i=1}^N$, the restoration quality of the stable characteristic image $\mathbf{Z}_i$ and the outlier degree of the initial image $\mathbf{I}_i$ is inversely correlated. The lower the initial outlier of image $\mathbf{I}_i$ in the image sequence, the smaller the distortion of the restored stable feature image $\mathbf{Z}_i$, the clearer the texture, and the closer $\mathbf{Z}_i$ is to the expected stable feature image $\mathbf{Z}_i^0$. At the same time, the more sparse the separated noise matrix $\mathbf{E}_i$ is, the smaller the value of $\|\mathbf{E}_i\|_1 = \|\mathbf{I}_i - \mathbf{Z}_i\|_1$ is.

According to the principle of low-rank sparse decomposition, there are $\sum_{i=1}^N \|\mathbf{I}_i - \mathbf{Z}_i\|_1 = \sum_{i=1}^N \|\mathbf{E}_i\|_1 = \|\mathbf{S}\|_1$. According to Theorem A2, determined by the sparsity of $\mathbf{S}$, the matrix sequence $\{\mathbf{E}_i\}_{i=1}^N$ has overall sparsity. Therefore, the image pair $(\mathbf{I}_i, \mathbf{Z}_i)$ has small comprehensive differences. Let $\text{diff}(\cdot, \cdot)$ represent the comprehensive difference of image pairs, and the following inference can be drawn:

Inference A3. For a set of multi-temporal remote sensing image sequences $\{\mathbf{I}_i\}_{i=1}^N$ conforming to Hypothesis A2 and its restored stable feature image sequence $\{\mathbf{Z}_i\}_{i=1}^N$, when the noise is widespread in the original image sequence, for any selected image $\mathbf{I}_r \in \{\mathbf{I}_i\}_{i=1}^N$, there is $\sum_{i=1}^N \text{diff}(\mathbf{I}_i, \mathbf{I}_r) > \sum_{i=1}^N \text{diff}(\mathbf{I}_i, \mathbf{Z}_i)$; when N takes a larger value, there must be a set of image sequences $\{\mathbf{I}_i\}_{i=1}^{N_{S2}} \subseteq \{\mathbf{I}_i\}_{i=1}^N$, and then there is $\text{diff}(\mathbf{I}_i, \mathbf{I}_r) > \text{diff}(\mathbf{I}_i, \mathbf{Z}_i) (1 \leq i \leq N_{S2} \leq N)$.

The derivation process of Inference A3 is as follows:

Here, we use the pixel intensity difference 1-norm $\|\mathbf{I}_i - \mathbf{I}_j\|_1$ of the image pair to approximately calculate $\text{diff}(\mathbf{I}_i, \mathbf{I}_j)$. It is assumed that the image set $\{\mathbf{I}_i\}_{i=1}^N$ conforms to Hypothesis 2, the image subset $\varphi = \{\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_t\} \subseteq \{\mathbf{I}_i\}_{i=1}^N$ has comprehensive noises $\{e_1, e_2, \dots, e_t\}$, and low-rank decomposition can eliminate these noises, that is, $\mathbf{I}_i = \mathbf{Z}_i + e_i$. For any image $\mathbf{I}_r \in \{\mathbf{I}_i\}_{i=1}^N$, ignoring the coincidence cancellation in the noise area, there are:

$$\sum_{i=1}^N \text{diff}(\mathbf{I}_i, \mathbf{Z}_i) = \|\mathbf{S}\|_1 = \sum_{i=1}^t \|e_i\|_1 \quad (\text{A5})$$

$$\sum_{i=1}^N \text{diff}(\mathbf{I}_i, \mathbf{I}_r) = \begin{cases} \sum_{i=1}^t \|e_i\|_1 = \sum_{i=1}^N \text{diff}(\mathbf{I}_i, \mathbf{Z}_i) & \mathbf{I}_r \notin \varphi \\ \sum_{i=1, i \neq r}^t \|e_i\|_1 + (N-1) * \|e_r\|_1 = \sum_{i=1}^t \|e_i\|_1 + (N-2) * \|e_r\|_1 & \mathbf{I}_r \in \varphi \end{cases} \quad (\text{A6})$$

When comprehensive noise is widespread, then $\mathbf{I}_r \in \{\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_t\} (t \mapsto N)$. There is the equation:

$$\sum_{i=1}^N \text{diff}(\mathbf{I}_i, \mathbf{I}_r) = \sum_{i=1}^N \text{diff}(\mathbf{I}_i, \mathbf{Z}_i) + (N-2) * \|e_r\|_1 \quad (\text{A7})$$

according to Equation (A7), when $N > 2$, there is $\sum_{i=1}^N \text{diff}(I_i, I_r) > \sum_{i=1}^N \text{diff}(I_i, Z_i)$. When N is large, there is $\sum_{i=1}^N \text{diff}(I_i, I_r) \gg \sum_{i=1}^N \text{diff}(I_i, Z_i)$.

Due to $\text{diff}(I_r, I_r) = 0$, the mean difference between the image sequence $\{I_i\}_{i=1}^N$ and image I_r is $\frac{\sum_{i=1}^N \text{diff}(I_i, I_r)}{N-1}$. The mean difference of the image pair $\{I_i, Z_i\}_{i=1}^N$ is $\frac{\sum_{i=1}^N \text{diff}(I_i, Z_i)}{N}$. When N is large, which means $N - 1 \approx N$, there is $\frac{\sum_{i=1}^N \text{diff}(I_i, I_r)}{N-1} \approx \frac{\sum_{i=1}^N \text{diff}(I_i, I_r)}{N} > \frac{\sum_{i=1}^N \text{diff}(I_i, Z_i)}{N}$. Thus, we can draw Inference A3.

References

- Shen, H.; Meng, X.; Zhang, L. An Integrated Framework for the Spatio-Temporal-Spectral Fusion of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7135–7148. [CrossRef]
- Zhou, Y.; Rangarajan, A.; Gader, P.D. An Integrated Approach to Registration and Fusion of Hyperspectral and Multispectral Images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3020–3033. [CrossRef]
- Tang, Y.; Wang, L.-J.; Ma, G.-C.; Jia, H.-J.; Jin, X. Emergency monitoring of high-level landslide disasters in Jinsha River using domestic remote sensing satellites. *J. Remote Sens.* **2019**, *23*, 252–261.
- Song, F.; Yang, Z.; Gao, X.; Dan, T.; Yang, Y.; Zhao, W.; Yu, R. Multi-Scale Feature Based Land Cover Change Detection in Mountainous Terrain Using Multi-Temporal and Multi-Sensor Remote Sensing Images. *IEEE Access* **2018**, *6*, 77494–77508. [CrossRef]
- Bordone Molini, A.; Valsesia, D.; Fracastoro, G.; Magli, E. DeepSUM: Deep Neural Network for Super-Resolution of Unregistered Multitemporal Images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3644–3656. [CrossRef]
- Guan, X.-B.; Shen, H.-F.; Gan, W.-X.; Zhang, L.-P. Estimation and spatiotemporal analysis of winter NPP in Wuhan based on Landsat TM/ETM+ Images. *Remote Sens. Technol. Appl.* **2015**, *30*, 884–890.
- Ye, Y. Fast and Robust Registration of Multimodal Remote Sensing Images via Dense Orientated Gradient Feature. *ISPRS—Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *XLII-2/W7*, 1009–1015. [CrossRef]
- Ma, W.; Wen, Z.; Wu, Y.; Jiao, L.; Gong, M.; Zheng, Y.; Liu, L. Remote Sensing Image Registration with Modified SIFT and Enhanced Feature Matching. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 3–7. [CrossRef]
- Chang, H.-H.; Wu, G.-L.; Chiang, M.-H. Remote Sensing Image Registration Based on Modified SIFT and Feature Slope Grouping. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1363–1367. [CrossRef]
- Chen, S.; Zhong, S.; Xue, B.; Li, X.; Zhao, L.; Chang, C.-I. Iterative Scale-Invariant Feature Transform for Remote Sensing Image Registration. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 3244–3265. [CrossRef]
- Chen, S.; Li, X.; Zhao, L.; Yang, H. Medium-Low Resolution Multisource Remote Sensing Image Registration Based on SIFT and Robust Regional Mutual Information. *Int. J. Remote Sens.* **2018**, *39*, 3215–3242. [CrossRef]
- Feng, R.; Du, Q.; Li, X.; Shen, H. Robust Registration for Remote Sensing Images by Combining and Localizing Feature- and Area-Based Methods. *ISPRS Journal of Photogrammetry and Remote Sensing* **2019**, *151*, 15–26. [CrossRef]
- Feng, R.; Du, Q.; Luo, H.; Shen, H.; Li, X.; Liu, B. A Registration Algorithm Based on Optical Flow Modification for Multi-temporal Remote Sensing Images Covering the Complex-terrain Region. *J. Remote Sens.* **2021**, *25*, 630. [CrossRef]
- Feng, R.; Du, Q.; Shen, H.; Li, X. Region-by-Region Registration Combining Feature-Based and Optical Flow Methods for Remote Sensing Images. *Remote Sens.* **2021**, *13*, 1475. [CrossRef]
- Schubert, A.; Small, D.; Jehle, M.; Meier, E. COSMO-SkyMed, TerraSAR-X, and RADARSAT-2 Geolocation Accuracy after Compensation for Earth-System Effects. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012.
- Aguilar, M.A.; del Mar Saldana, M.; Aguilar, F.J. Assessing Geometric Accuracy of the Orthorectification Process from Geoeye-1 and Worldview-2 Panchromatic Images. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *21*, 427–435. [CrossRef]
- Chen, L.-C.; Li, S.-H.; Chen, J.-J.; Rau, J.-Y. Method of Ortho-Rectification for High-Resolution Remote Sensing Image, October 9, 2008. Available online: <https://www.patentsencyclopedia.com/app/20080247669> (accessed on 25 October 2021).
- Hasan, R.H. Evaluation of the Accuracy of Digital Elevation Model Produced from Different Open Source Data. *J. Eng.* **2019**, *25*, 100–112. [CrossRef]
- Nag, S. Image Registration Techniques: A Survey. *engrXiv* **2017**. [CrossRef]
- Zitová, B.; Flusser, J. Image Registration Methods: A Survey. *Image Vis. Comput.* **2003**, *21*, 977–1000. [CrossRef]
- Martinez, A.; Garcia-Consuegra, J.; Abad, F. A Correlation-Symbolic Approach to Automatic Remotely Sensed Image Rectification. In Proceedings of the IEEE 1999 International Geoscience and Remote Sensing Symposium, IGARSS'99 (Cat. No.99CH36293), Hamburg, Germany, 28 June–2 July 1999.
- Hel-Or, Y.; Hel-Or, H.; David, E. Fast Template Matching in Non-Linear Tone-Mapped Images. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011.
- Chen, H.-M.; Varshney, P.K.; Arora, M.K. Performance of Mutual Information Similarity Measure for Registration of Multitemporal Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 2445–2454. [CrossRef]
- Kern, J.P.; Pattichis, M.S. Robust Multispectral Image Registration Using Mutual-Information Models. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1494–1505. [CrossRef]

25. Zhao, L.-Y.; Lü, B.-Y.; Li, X.-R.; Chen, S.-H. Multi-source remote sensing image registration based on scale-invariant feature transform and optimization of regional mutual information. *Acta Physica Sinica*. **2015**, *64*, 124204. [\[CrossRef\]](#)
26. Ravanbakhsh, M.; Fraser, C.S. A Comparative Study of DEM Registration Approaches. *J. Spat. Sci.* **2013**, *58*, 79–89. [\[CrossRef\]](#)
27. Murphy, J.M.; Le Moigne, J.; Harding, D.J. Automatic Image Registration of Multimodal Remotely Sensed Data with Global Shearlet Features. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1685–1704. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [\[CrossRef\]](#)
29. Bay, H.; Tuytelaars, T.; Van Gool, L. SURF: Speeded Up Robust Features. In *Computer Vision—ECCV 2006*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 404–417.
30. Morel, J.-M.; Yu, G. ASIFT: A New Framework for Fully Affine Invariant Image Comparison. *SIAM J. Imaging Sci.* **2009**, *2*, 438–469. [\[CrossRef\]](#)
31. Ke, Y.; Sukthankar, R. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004, Washington, DC, USA, 27 June–2 July 2004.
32. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An Efficient Alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011.
33. Leutenegger, S.; Chli, M.; Siegwart, R.Y. BRISK: Binary Robust Invariant Scalable Keypoints. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011.
34. Rosten, E.; Porter, R.; Drummond, T. Faster and Better: A Machine Learning Approach to Corner Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 105–119. [\[CrossRef\]](#)
35. Ye, Y.; Shan, J.; Bruzzone, L.; Shen, L. Robust Registration of Multimodal Remote Sensing Images Based on Structural Similarity. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2941–2958. [\[CrossRef\]](#)
36. Sedaghat, A.; Ebadi, H. Remote Sensing Image Matching Based on Adaptive Binning SIFT Descriptor. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5283–5293. [\[CrossRef\]](#)
37. Brox, T.; Bruhn, A.; Papenberger, N.; Weickert, J. High Accuracy Optical Flow Estimation Based on a Theory for Warping. In *Lecture Notes in Computer Science*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2004; pp. 25–36.
38. Ren, Z.; Li, J.; Liu, S.; Zeng, B. Meshflow Video Denoising. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017.
39. Liu, L.; Chen, F.; Liu, J.-B. Optical Flow and Feature Constrains Algorithm for Remote Sensing Image Registration. *Comput. Eng. Des.* **2014**, *35*, 3127–3131.
40. Xu, F.; Yu, H.; Wang, J.; Yang, W. Accurate Registration of Multitemporal UAV Images Based on Detection of Major Changes. In Proceedings of the 2018 21st International Conference on Information Fusion (FUSION), Cambridge, UK, 10–13 July 2018.
41. Brigot, G.; Colin-Koeniguer, E.; Plyer, A.; Janez, F. Adaptation and Evaluation of an Optical Flow Method Applied to Coregistration of Forest Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2923–2939. [\[CrossRef\]](#)
42. Peng, Y.; Ganesh, A.; Wright, J.; Xu, W.; Ma, Y. RASL: Robust Alignment by Sparse and Low-Rank Decomposition for Linearly Correlated Images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2233–2246. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Wu, Y.; Shen, B.; Ling, H. Online Robust Image Alignment via Iterative Convex Optimization. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012.
44. Candes, E.; Li, X.; Ma, Y.; Wright, J. Robust Principal Component Analysis?: Recovering Low-Rank Matrices from Sparse Errors. In Proceedings of the 2010 IEEE Sensor Array and Multichannel Signal Processing Workshop, Jerusalem, Israel, 4–7 October 2010.
45. Chi, Y.; Lu, Y.M.; Chen, Y. Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview. *IEEE Trans. Signal Processing* **2019**, *67*, 5239–5269. [\[CrossRef\]](#)
46. Torr, P.H.S.; Murray, D.W. The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix. *Int. J. Comput. Vis.* **1997**, *24*, 271–300. [\[CrossRef\]](#)
47. Russakoff, D.B.; Tomasi, C.; Rohlfing, T.; Maurer, C.R., Jr. Image Similarity Using Mutual Information of Regions. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 596–607.
48. Zaragoza, J.; Tat-Jun, C.; Tran, Q.-H.; Brown, M.S.; Suter, D. As-Projective-as-Possible Image Stitching with Moving DLT. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1285–1298. [\[PubMed\]](#)
49. Hu, T.; Zhang, H.; Shen, H.; Zhang, L. Robust Registration by Rank Minimization for Multiangle Hyper/Multispectral Remotely Sensed Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2443–2457. [\[CrossRef\]](#)
50. Goshtasby, A. Piecewise Linear Mapping Functions for Image Registration. *Pattern Recognit.* **1986**, *19*, 459–466. [\[CrossRef\]](#)
51. Han, Y.; Bovolo, F.; Bruzzone, L. An Approach to Fine Coregistration between Very High Resolution Multispectral Images Based on Registration Noise Distribution. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6650–6662. [\[CrossRef\]](#)
52. Gharbia, R.; Ahmed, S.A.; Hassanien, A.E. Remote Sensing Image Registration Based on Particle Swarm Optimization and Mutual Information. In *Advances in Intelligent Systems and Computing*; Springer: New Delhi, India, 2015; pp. 399–408.
53. Lin, Z.; Ganesh, A.; Wright, J.; Wu, L.; Chen, M.; Ma, Y. Fast Convex Optimization Algorithms for Exact Recovery of a Corrupted Low-Rank Matrix. 2009. Available online: <https://www.ideals.illinois.edu/handle/2142/74352> (accessed on 20 January 2021).
54. Lin, Z.; Chen, M.; Ma, Y. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices. *arXiv* **2010**, arXiv:1009.5055.
55. Yuan, X.-M.; Yang, J.-F. Sparse and Low Rank Matrix Decomposition via Alternating Direction Method. *Preprint* **2009**, 12.