

A Novel Control Point Dispersion Method for Image Registration Robust to Local Distortion

Yuan Yuan, Meng Yu^(✉), Lin Lei, Yang Jian, Chen Jingbo,
Yue Anzhi, and He Dongxu

Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences,
No. 20 North, Datun Road, Chaoyang District, Beijing 100101, China
{yuanyuan, mengyu, linlei, yangjian,
chenjb, yueaz, hedx}@radi.ac.cn

Abstract. Extracting well-distributed and precisely aligned control points (CPs) is extremely important for remote sensing image registration, particularly for high resolution images with large local distortion. Based on a theoretical analysis of estimation perturbation in transformation parameters, a novel CP dispersion approach is proposed to select high quality and uniformly distributed CPs. This approach retains a minimum spanning tree (MST) of the selected CPs during the algorithm and adds a new CP to the tree in each iteration until satisfying the convergence condition. Moreover, to acquire adequate number of CPs for the dispersion process, a coarse-to-fine matching approach is proposed. Experiment results indicate that the proposed method improves the match performance compared to other CP dispersion methods in terms of aligning accuracy.

Keywords: Control point dispersion · Coarse-to-fine matching strategy · Image registration · Local image distortions

1 Introduction

Image registration is a crucial preprocessing step for many subsequent image analysis techniques, such as image mosaicing, change detection, digital elevation model generation and map updating, etc.

The existing automatic image registration methods can be broadly divided into two categories, i.e. area-based methods and feature-based methods [1]. Area-based (or template matching) methods compare similarities between image regions using different measures, such as normalized cross correlation (NCC) [2] and mutual information (MI) [3]. Though area-based methods can attain sub-pixel registration accuracy, they generally can't be applied to the registration task of remote sensing image pairs directly when large rotation or scaling changes exist. Comparing to area-based methods, feature-based methods are more robust to image variations. Feature-based methods extract salient image structures for matching. In particular, point feature is one of the most important groups of feature. Aanæs et al. [4] investigated the performance of some state-of-the-art point feature detectors and found that simple Harris detector

performs very well for small-scale changes while scale invariant feature transform (SIFT) detector is superior to others when considering large-scale changes.

For high-resolution remote sensing image registration, a major difficulty is significant local geometrical distortions [5], which are caused by height displacement and/or platform instability. To overcome these problems, Brown [6] recommended extracting a large number of CPs to applied piecewise mapping between images. For instance, in [7] and [8], Harris detector was used to generate a dense set of CPs after an initial matching stage based on SIFT. In [9], NCC was applied to obtain uniformly spread CPs after the same preliminary stage.

The distribution of CPs also has significant impact on the estimation of transform model [10]. Some approaches have been reported to achieve spatially well-distributed CPs. The strategies employed by those methods can be divided into two categories. The first strategy is to constrain the distribution of interest points in the feature detection step. Brown et al. [11] introduced an adaptive local non-maximal suppression algorithm to select specified number of Harris corner points. Sedaghat et al. [12] proposed the uniform robust (UR) –SIFT, which introduced a grid division step to SIFT and only selected the strongest few keypoints in each grid in both scale and image space. However, though SIFT detector can extract a high number of keypoints in both input and reference images, most of them can't be correctly matched. Reducing the extracted features may result in even fewer matches [13]. The second strategy is to select a subset of well distributed matched CPs. Fonseca et al. [14] suggested clustering the aligned CPs into some clusters and maintaining the strongest one in each cluster. Since the clustering based method was very time-consuming, a faster alternative CP dispersion method was recommended by Fonseca et al. This method divides the input image into subimages and selects the strongest matches within each subimage. An improved approach was proposed by Wang et al. [15] that used least squares iterative fitting method to eliminate false matches in each subimage.

In this letter, a novel CP dispersion method is proposed based on theoretical analysis of the relationship between aligning errors in CPs and parameter estimation. The proposed algorithm utilizes a greedy approach to select high-precision and well-distributed CPs based on a modified minimum spanning tree algorithm. In addition, a coarse-to-fine matching strategy is employed to acquire adequate CPs for the registration of images with local distortions. Coarse registration is implemented to get an initial transform model between input and reference images using SIFT. Then in fine registration step a large number of aligned CPs are obtained with the combination of Harris and NCC. Finally, CP dispersion is carried out to select “good” CPs from them.

The remainder of this paper is organized as follows. In Sect. 2, the proposed CP dispersion method is presented. Then the coarse-to-fine matching approach is described in Sect. 3. Section 4 presents the experiments and discussion, whereas conclusions are drawn in Sect. 5.

2 Control Point Dispersion

To figure out how the aligning errors contained in the CPs may affect the final registration result, some previous theoretical analysis is introduced at first.

Suppose the transformation between two images is only rotation with an angle of θ , let \mathbf{x}_1 and \mathbf{x}_2 be the accurate coordinates of two CPs in the input image and suppose \mathbf{X}_1 and \mathbf{X}_2 are their correspondences in the reference image. Then

$$\cos(\theta) = \frac{(\mathbf{x}_2 - \mathbf{x}_1) \cdot (\mathbf{X}_2 - \mathbf{X}_1)}{\|\mathbf{x}_2 - \mathbf{x}_1\| \|\mathbf{X}_2 - \mathbf{X}_1\|} \quad (1)$$

where $\|\cdot\|$ donates vector norm and \cdot donates inner product. Assuming that the CP pairs $(\mathbf{x}_1, \mathbf{X}_1)$ and $(\mathbf{x}_2, \mathbf{X}_2)$ are detected with errors $(d\mathbf{x}, d\mathbf{X})$, Fonseca and Kenney [14] have proven that perturbation in $\cos(\theta)$ varies to first order with the errors divided by the interpoint distances (shown in (2)):

$$d(\cos(\theta)) \leq 2 \left(\frac{\|d\mathbf{x}\|}{\|\mathbf{x}_2 - \mathbf{x}_1\|} + \frac{\|d\mathbf{X}\|}{\|\mathbf{X}_2 - \mathbf{X}_1\|} \right) \quad (2)$$

Moreover, they claimed that similar result produced when calculating the scaling parameter between images.

According to the previous discussion, it can be seen that the distribution of CPs has significant impact on the estimation of transformation parameters. In the case of two CPs, the estimation error of transformation parameters is approximately linearly correlated with the aligning errors of CPs divided by the distance between them.

Consequently, assuming that adequate matched CPs have been obtained, the aim of CP dispersion is to select a subset of them with the highest accuracy. Meanwhile, the distance between any two CPs is larger than a certain size. If a constant distance threshold is adopted to exclude CPs, many high-precision ones may be discarded, let alone the risk of too sparse CPs. Taking into account of the relationship described in Eq. 2, we suggest using the aligning error of each CP as weight to adjust a constant interpoint distance threshold.

Supposing an initial transform model \mathbf{H} is estimated using all the CPs (in this study, quadratic polynomial model is adopted for its simplicity), the aligning error of the i th CP is computed by (3) and the adaptive distance threshold is defined in (4)

$$e_i = \|\mathbf{H}\mathbf{x}_i - \mathbf{X}_i\| \quad (3)$$

$$T_{d_i} = e_i T \quad (4)$$

where T_{d_i} represents the minimum Euclidean distance from the i th CP to all the selected CPs. T is a constant, which denotes the ‘base’ interpoint distance. It is determined empirically according to the expected number of CPs: the larger T is, the less CPs are retained. If e_i is zero, T_{d_i} is also equal to zero and thus the i th CP is ought to be selected. In the meantime, for a CP located distant from others, even if its aligning error is a little larger, it could be selected as well.

A novel CP dispersion method based on the proposed adaptive distance threshold is proposed. Inspired by the minimum spanning tree (MST) problem in graph theory, a simple and fast greedy approach revised from Prim algorithm is designed. During the algorithm, we maintain an output tree which stands for a group of CPs already selected.

At the beginning, the tree is empty. At each iteration, we choose a new CP and determine whether to add it to the output tree. The proposed approach is described as follows.

- **Step 1.** First, compute the aligning errors $\{e_i\}$, $i = 1, 2, \dots, N$ and adaptive distance thresholds $\{T_{d_i}\}$ of all the CPs, where N stands for the total number of CPs. Then sort the CPs in ascending order according to their errors. Let j be the index of the first unselected CP, which is set to 1.
- **Step 2.** Add the j th CP to the output tree. Calculate and record the distance between each of the other CP and the j th CP. Set $j = j + 1$.
- **Step 3.** Move to the j th CP. If its distance to the output tree is less than T_{d_j} , it is regarded as ‘disconnected’ with the output tree. Then set $j = j + 1$ and go on to step 4 directly. Otherwise, this CP is viewed as ‘connected’ and is added to the output tree. Then calculate the distances from the remaining CPs to the newly added one. For each remaining CP, if its distance to the j th CP is less than its recorded distance, update the distance to be equal to the smaller one. Otherwise, keep it unchanged. Let $j = j + 1$.
- **Step 4.** Repeat Step 3 until all the CPs have been traversed.

3 Registration with Control Point Dispersion

We employ a coarse-to-fine matching strategy to match images with large local distortions. The proposed approach can be summarized as three stages, i.e. pre-registration, dense CP generation and CP dispersion, described in more detail below.

• Stage 1. Pre-Registration

In this study, the same pre-registration process based on SIFT algorithm is carried out [7]. First, SIFT feature detection is conducted in both input and reference images to extract candidate keypoints. Then a 128-dimension vector is assigned to each keypoint based on its neighbor subregions. After SIFT descriptors are generated, the nearest neighbor distance ratio matching strategy is used to find potential correspondences. Finally, RANdom Sample Consensus (RANSAC) is adopted to remove outliers. Then the founded CPs are used to estimate an initial homography matrix. The detail of SIFT algorithm is discussed in [16].

• Stage 2. Dense CP Generation

The CPs acquired in the first stage is often insufficient for the requirement of image registration. To obtain more CPs, Harris corner detection is conducted in the input image. For each corner point, a template matching procedure is used to search for its correspondence in the reference image. We adopt NCC as the similarity measure. Assuming a template window is defined in the input image, center of the window is the Harris corner point. Each template window is mapped to a corresponding window in the reference image. NCC similarity measure is defined as

$$\rho(x, y) = \frac{\sum_x \sum_y (I(x, y) - \bar{I})(I'(x', y') - \bar{I}')}{\sqrt{\sum_x \sum_y (I(x, y) - \bar{I})^2} \sqrt{\sum_{x'} \sum_{y'} (I'(x', y') - \bar{I}')^2}} \quad (5)$$

where $I(x, y)$ and $I'(x', y')$ are gray values in the template window and the corresponding window respectively. \bar{I} and \bar{I}' are the mean gray values in the two windows. The corresponding window is moving within a searching radius in the reference image. When the maximum value of NCC is achieved and larger than T_n , the center of the corresponding window is treated as a correspondence. In this study, the radiuses of the template window and the search window are set to 15 and 21 respectively. T_n is set to 0.8.

To achieve sub-pixel precision in NCC, we adopted the correlation interpolation method, which uses bi-cubic convolution to refine the locations of the CPs. A two-dimensional cubic convolution of the correlation coefficients is applied to the neighbor 4×4 pixels of each correspondence. The peak is set as its refined location. The advantage of this method is that it can achieve desirable result with a relatively lower computational amount comparing to the other sub-pixel approaches [2].

• Stage 3. Control Point Dispersion

Once enough CPs have been obtained, their aligning errors are re-examined to eliminate false matches using least squares iteration fitting. Then the proposed CP dispersion is employed to select well-distributed ones. Finally, the transform model is re-estimated. Some previous studies suggested using non-rigid transformation for the registration of images with large local distortions [7, 8], which is beyond the scope of this paper.

4 Experiments and Discussion

In the experiments, we applied the coarse-to-fine matching algorithm to high-resolution optical satellite images with simulated distortions and real local distortions. The performance of the proposed CP dispersion method is also compared with the correspondence error checking (CEC) method suggested in [2] and the subdivision method [8]. CEC is a traditional and commonly used method for CP exclusion which iteratively discards CPs with the largest aligning errors and re-estimates the transform model until either all the CPs' RMSEs are less than a threshold (0.5 pixel in this experiment) or the number of CPs is less than three. Subdivision method is the foundation of a group of existing CP dispersion methods. The main difference between subdivision method and the proposed method is that the latter uses adaptive distance threshold. For subdivision method, the expected number of subimages was set to 256, and for the proposed method the base distance T was set to 20 pixels in our experiment.

To assess the performance quantitatively, the RMSE of the test points (TPs) is used to evaluate the registration accuracy. The TPs were uniformly generated or manually selected spread out the whole image. In addition, the number of CPs is also considered as an important measure.

4.1 Images with Simulated Distortion

Two SPOT-5 panchromatic images (geometrically and radiometrically preprocessed at Level 2B in WGS 84) taken on October 1, 2002 and October 16, 2007 were used as test dataset, which have been precisely aligned. Two pairs of test images of size 512×512 pixels were clipped from both images [see Fig. 1(a), (b)]. In order to simulate local image distortion, we transformed one image in a pair by the following sinusoidal function:

$$\begin{aligned} x' &= x - 2 \sin\left(\frac{y}{32}\right) \\ y' &= y + 2 \sin\left(\frac{x}{32}\right) \end{aligned} \quad (6)$$

where (x, y) and (x', y') are pixel locations in the original image and the transformed image respectively. 16×16 TPs evenly spaced in the reference image were generated. Their correspondences in the transformed image were computed by (8).

After the dense CP generation stage, 282 pairs of CPs were found in dataset 1 and 198 pairs were found in dataset 2. The final experiment results are shown in Table 1, from which we can see that the proposed method clearly outperformed its competitors with the lowest RMSEs and moderate number of CPs. Note that in dataset 2, the RMSE of CEC was extremely large. It was because that all the CPs located in the homogenous farmland area were pruned to minimize the least square error, which resulted in serious CP concentration.

4.2 Remote Sensing Images with Local Distortion

To further demonstrate the feasibility of the proposed method in real applications, two representative image pairs with significant terrain relief were chosen. One was located in the mountainous region and the other was in densely built city area. In dataset 1, two SPOT-5 panchromatic images were captured on December 1, 2006 and March 20, 2007, with sizes of 554×640 and 710×905 respectively [see Fig. 1(c)]. In dataset 2, two ZY1-02C panchromatic images taken on August 18, 2013 and October 24, 2012, with sizes of 684×703 and 787×865 respectively were chosen [see Fig. 1(d)]. 20 TPs were selected manually for each dataset.

After the dense CP generation stage, 367 CPs were found in dataset 1 and 653 were found in dataset 2. The performance results were listed in Table 2, from which we can see that the registration accuracy of the proposed method is higher than those of the other methods. In addition, it also provided sufficient number of CPs for rigid or non-rigid transformation estimation. Figure 2 demonstrated the CP location in both cases. We can see that the CPs selected by CEC were rather spatial concentrated in Fig. 2(b), since it pruned CPs by looking only at their aligning errors. Though it eliminated most of the inferior matches, some correct CPs located in the area with local distortions were also removed. On the contrary, subdivision method over pruned some CPs with high precision by only retaining the strongest one in each subimage, as shown in Fig. 2(c). It should be noticed that its registration accuracy in case 2 is even lower

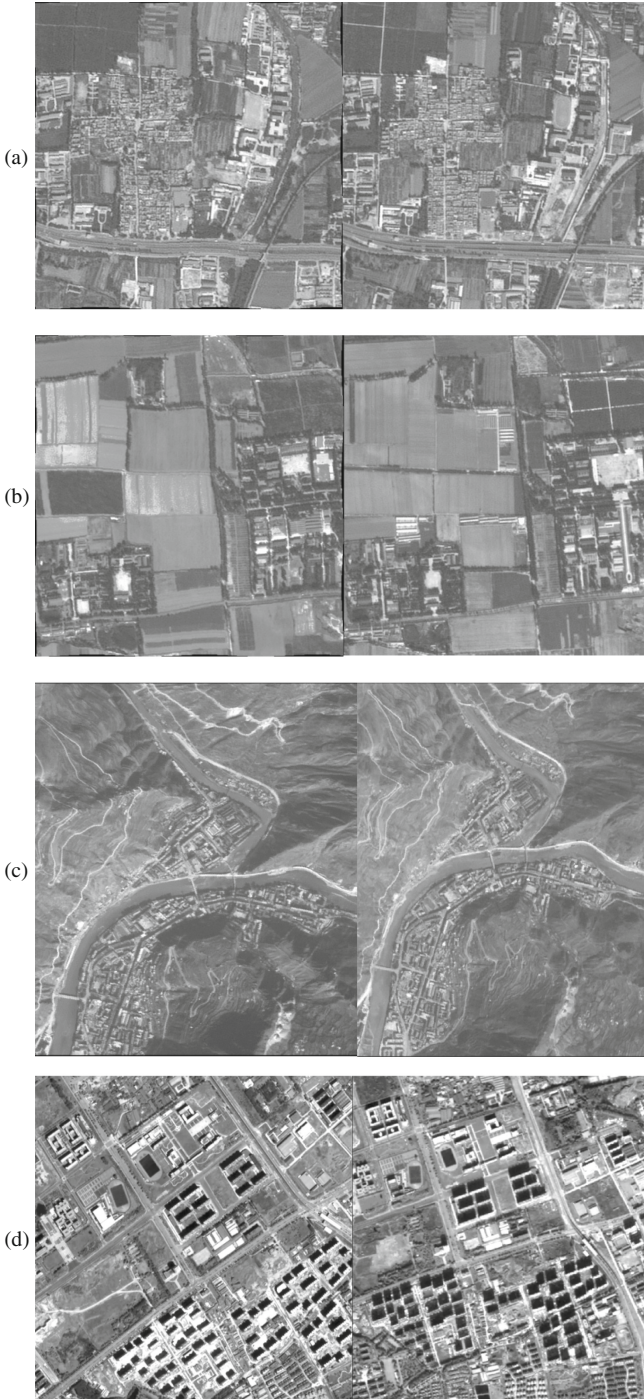


Fig. 1. Test image pairs. (a), (b) SPOT-5 images taken on 2002 (left) and 2007 (right). (c) SPOT-5 images taken on 2006 (left) and 2007 (right). (d) ZY1-02C images taken on 2013 (left) and 2012 (right).

than CEC, which may due to its excessive pursuit of “even distribution” at the cost of allowing some low-precision CPs. In contrast, the proposed method tended to select more high-precision CPs in the residential area with complex texture while also retained some CPs located in the mountainous regions and cultivated lands at the same time.

Table 1. Performance comparison on images with simulated local distortion

		The proposed method	CEC	Subdivision
Dataset 1	Number	42	26	129
	RMSE	2.548	3.426	3.082
Dataset 2	Number	38	18	90
	RMSE	2.604	54.490	3.278

Table 2. Performance comparison on images with real distortion

		The proposed method	CEC	Subdivision
Dataset 1	Number	156	140	158
	RMSE	2.818	3.032	3.012
Dataset 2	Number	108	55	222
	RMSE	1.355	1.481	2.317

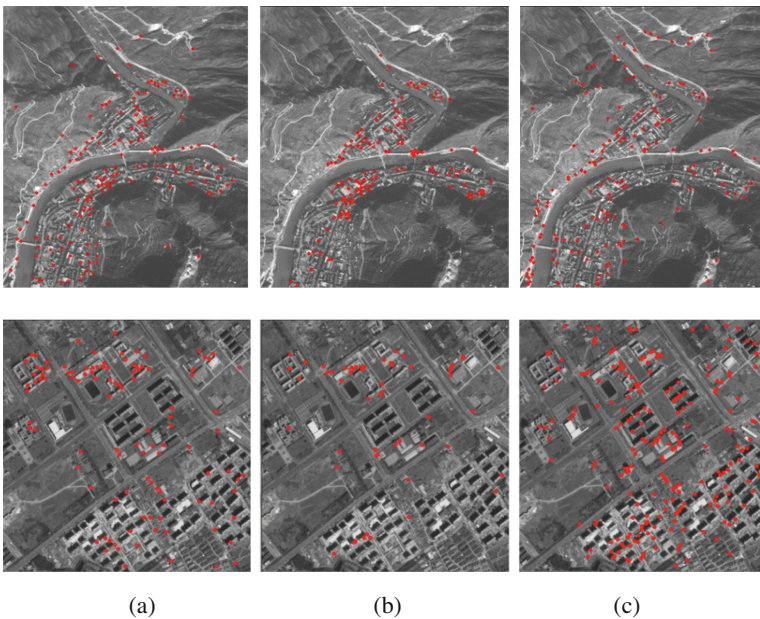


Fig. 2. CPs selected by different approaches in case 1 (above) and case 2 (below). (a) The proposed method. (b) CEC method. (c) Subdivision method.

5 Conclusion

An efficient approach for CP dispersion is proposed in the paper for remote sensing image registration. This method balances the impact of distances between CPs and their aligning errors by using adaptive distance thresholds to select high-precision and spatially well-distributed CPs. In addition, a coarse-to-fine registration approach is introduced to achieve registration result robust to local image distortion. Experimental results for images with simulated distortion and real local distortions show that proposed method significantly improves the matching accuracy, which contribute to the higher accuracy of subsequent image stitching and analysis.

Acknowledgements. This work is partially supported in part by the National Natural Science Foundation of China (Grant No. 41401474), and the Young Scholars of Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences (Y3SJ7100CX), and the High Resolution Earth Observation Major Project (No. 04-Y30B01-9001-12/15).

References

1. Zitova, B., Flusser, J.: Image registration methods: a survey. *Image Vis. Comput.* **21**, 977–1000 (2003)
2. Debella-Gilo, M., Kääb, A.: Sub-pixel precision image matching for measuring surface displacements on mass movements using normalized cross-correlation. *Remote Sens. Environ.* **115**(1), 130–142 (2011)
3. Gong, M., Zhao, S., Jiao, L., Tian, D., Wang, S.: A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information. *IEEE Trans. Geosci. Remot. Sen.* **52**(7), 4328–4338 (2014)
4. Aanæs, H., Dahl, A.L., Pedersen, K.S.: Interesting interest points. *Int. J. Comput. Vis.* **97**(1), 18–35 (2012)
5. Han, Y., Choi, J., Byun, Y., Kim, Y.: Parameter optimization for the extraction of matching points between high-resolution multisensor Images in Urban Areas. *IEEE Trans. Geosci. Remot. Sen.* **52**(9), 5612–5621 (2014)
6. Brown, L.G.: A survey of image registration techniques. *ACM Comput. Surv.* **24**(4), 325–376 (1992)
7. Yu, L., Zhang, D., Holden, E.J.: A fast and fully automatic registration approach based on point features for multi-source remote-sensing images. *Comput. Geosci.* **34**(7), 838–848 (2008)
8. Sedaghat, A., Hamid, E., Mokhtarzade, M.: Image matching of satellite data based on quadrilateral control networks. *Photogramm. Rec.* **27**(140), 423–442 (2012)
9. Ma, J., Chan, J.C., Canters, F.: Fully automatic subpixel image registration of multiangle CHRIS/Proba data. *IEEE Trans. Geosci. Remote Sen.* **48**(7), 2829–2839 (2010)
10. Gonçalves, H., Gonçalves, J.A., Corte-Real, L.: Measures for an objective evaluation of the geometric correction process quality. *IEEE Geosci. Remote Sens. Lett.* **6**(2), 292–296 (2009)
11. Brown, M., Szeliski, R., Winder, S.: Multi-image matching using multi-scale oriented patches. In: *Proceedings of CVPR*, pp. 510–517 (2005)

12. Sedaghat, A., Mokhtarzade, M., Ebadi, H.: Uniform robust scale-invariant feature matching for optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **49**(11), 4516–4527 (2011)
13. Yu, X., Lyu, Z., Hu, D., Xu, J.: Scale-invariant feature transform based on the frequency spectrum and the grid for remote sensing image registration. *Gisci. Remote Sens.* **50**(5), 543–561 (2013)
14. Fonseca, L.M.G., Kenney, C.S.: Control point assessment for image registration. In: *Proceedings of VII Brazilian Symposium Computer Graphic and Image Processing*, pp. 125–132 (1999)
15. Wang, L., Niu, Z., Wu, C., Xie, R.: A robust multisource image automatic registration system based on the SIFT descriptor. *Int. J. Remote Sens.* **33**(12), 3850–3869 (2012)
16. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)