Remote sensing image matching by integrating affine invariant feature extraction and RANSAC

Liang Cheng *, Manchun Li, Yongxue Liu, Wenting Cai, Yanming Chen, Kang Yang

Department of Geographical Information Science, Nanjing University, Nanjing 210093, China

A new technical framework for remote sensing image matching by integrating affine invariant feature extraction and RANSAC is presented. The novelty of this framework is an automatic optimization strategy for affine invariant feature matching based on RANSAC. An automatic way to determine the distance threshold of RANSAC is proposed, which is a key problem to implement this RANSAC-based automatic optimization. Since affine invariant feature matching technology has been successfully applied to remote sensing image matching, we design an experiment to compare the proposed framework (with optimization) with the standard affine invariant feature matching (without optimization). By using three pairs with different types of imagery, the experimental results indicate that the proposed framework can always get higher correctness of image matching in automatic way, compared to the standard affine invariant feature matching technology.

1. Introduction

Image matching is a hot research topic in image processing, computer vision, photogrammetry and remote sensing, 3D reconstruction [1] and some other related research fields. Facing various remote sensing images (satellite images, aerial images, etc.), there still exist many problems in the existing image matching methods [2]. Fortunately, improvement of image matching has become possible thanks to recent developments in the automatic extraction of local invariant features. Local invariant features, especially affine invariant features, have been shown to be well suited to matching, recognition, and other applications as they are robust to occlusion, background clutter, and content changes [3]. Affine invariant feature extraction technology has been successfully applied to remote sensing image matching [4–6]. This technology can achieve good results of image matching when dealing with images with good texture and small changes of view angle. However, for images with poor texture or large changes of view angle, the power of this technology on image matching significantly decreases.

RANdom SAmple Consensus (RANSAC) is a robust estimation method, which is widely used [7]. The integration of affine invariant feature extraction technology and RANSAC is valuable for the optimization of image matching. In many cases, RANSAC is an effective robust estimator, which can get the correct matches even when more than 50% mismatches exist in the samples. However, there are still some problems in the integration of RANSAC and affine invariant feature extraction. For example, RANSAC algorithm needs to manually determine the optimal distance threshold for different types of remote sensing images, which is a key threshold to obtain good image matching results. The common way is to gradually change the value of distance threshold in manual way until achieving the satisfactory results by visual check. This manual process is arbitrary and can result in inconsistent results by different users.
To address the above problem, this study presents a new approach for remote sensing image matching by integrating affine invariant feature extraction and RANSAC. The novelty of this approach is an automatic optimization strategy for affine invariant feature matching based on RANSAC. To implement this RANSAC-based automatic optimization, an automatic way to determine the distance threshold of RANSAC is proposed, which is a key problem. The basic idea is as following. A robust algorithm, which based on MSER with information Entropy and spatial Dispersion constraints (termed ED-MSER), proposed by us [8] is used for affine invariant feature extraction from remote sensing images. Homography is then used as geometric constraint model of RANSAC, according to the characteristics of remote sensing images. Since the geometric shape of the extracted affine invariant features is polygon, we use homography to overlap these features to calculate the overlap-rate of the corresponding features. Based on this overlap-rate, the correctness of image matching can be calculated. It is used as a quantitative evaluation indicator to automatically change the distance threshold, which leads to the iteration progress of RANSAC for the automatic optimization of image matching.

In this paper, Section 2 introduces a technical framework, which gives the details of the above idea. Fig. 1 shows the main processes of the proposed framework, including two steps: affine invariant feature matching and automatic optimization for matching. Usually, an ideal optimization process of image matching is to remove all the mismatches and keep the right parts. However, it is often very difficult to remove all the mismatches and simultaneously keep all the correct ones. Therefore, two typical optimization objectives are discussed. According to these two objectives, two typical approaches and their corresponding algorithms are designed in Section 3 based on the proposed technical framework. Section 4 gives the experimental results and analysis.

2. Technical framework

2.1. Affine invariant feature matching

Mikolajczyk et al. [9] presented the state-of-the-art on affine covariant region detectors and have compared their performance, including (1) Harris–Affine detector [3,10]; (2) Hessian–Affine detector [3]; (3) the maximally stable extremal region (MSER) [11]; (4) edge-based region (EBR) [12]; (5) intensity extremal-based region (IBR) (IBR detector) [13]; and (6) salient regions detector [14]. They also identified that, in many cases, the MSER algorithm obtains the best results compared to the

![Flow chart of the proposed method.](image-url)
other approaches. Mikolajczyk and Schmid [15] also presented a comparative study of several popular local descriptors and identified scale-invariant feature transform (SIFT) [16] algorithm as being the most resistant to common image deformations.

Based on the integration of the MSER (the best detector) and SIFT (the best descriptor), a new algorithm (ED-MSER) for robust affine invariant feature extraction was presented by us [8]. In this algorithm, the key section is a feature-related filtering strategy. It is a hierarchical filtering strategy for affine invariant feature detection based on information entropy and spatial dispersion constraints. The key idea of the filtering strategy is to evaluate the information entropy and spatial dispersion of all features detected by MSER, remove the features with low information entropy and bad distribution, and just select the features with high information entropy and good distribution, thus benefiting the subsequent image matching and many other applications.

The ED-MSER algorithm can be summarized in the following steps:

(a) Use standard MSER to detect local affine invariant region.
(b) Evaluate and select local regions detected by MSER. Considering the shapes of regions extracted by MSER are irregular polygons, a convenient method to calculate information entropy of irregular polygon is used.
   1. Take the local region as an ROI (Region of Interest) and create its corresponding mask;
   2. Multiply the original image with the mask to get a new image, named as ROI image;
   3. Set the probability of gray value 0 in the ROI image to 0;
   4. Calculate entropy of the result image using formula (1), which is entropy of local region.

\[
H = - \sum_i P_i \log_2 P_i
\]

where \( P_i \) is the probability of \( i \).

To quantify the spatial distribution of features, the concept of spatial dispersion quality is introduced. We calculate spatial dispersion quality of features using formula (2). Smaller \( \text{Disp} \) the smaller spatial dispersion (more concentrated distribution); larger \( \text{Disp} \) the larger spatial dispersion (more discrete distribution).

\[
\text{Disp} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x}_{wmc})^2 + \sum_{i=1}^{n} (y_i - \bar{y}_{wmc})^2}{n}}
\]

\[
(\bar{x}_{wmc}, \bar{y}_{wmc}) = \left( \frac{\sum_{i=1}^{n} W_i x_i}{\sum_{i=1}^{n} W_i}, \frac{\sum_{i=1}^{n} W_i y_i}{\sum_{i=1}^{n} W_i} \right)
\]

where \( w_i \) is weight of point \( i \). Entropy of local region is taken as weight.

If a local region is not qualified according to a criterion of information content and spatial dispersion quality, it will be removed.

(c) Use SIFT as ED-MSER descriptor. From the above steps, the feature vectors can be created and used in image matching algorithms.

In this study, the ED-MSER algorithm is used for affine invariant feature extraction. Features with evenly spatial distribution and rich information content can be selected by this algorithm for the following matching and optimization processes. After feature selection, Euclidean distance is used as the similarity measure for feature matching. When matching two feature groups, calculate the distance between one feature in image A and all features in its corresponding image A'. A pair of matched features can be obtained by comparing the shortest distance and second-shortest distance. If the ratio of shortest distance and the second-shortest distance is less than 0.6, the two features are considered to be matched.

2.2 Automatic optimization for feature matching

2.2.1 Correctness of matching

Homography, which refers to the invertible homogeneous transformations between two planes, plays a very important role in multi-view geometry. It has been widely used in camera calibration, three-dimensional reconstruction, image mosaic and related research fields. In theory, the homography can only be applied to transferring features between two planes. But for satellite images or aerial images, the change of terrain or height of surface feature is quite small compared to the height of the sensor. Therefore, homography can be used in this study as geometric constraint model in RANSAC algorithm. Affine invariant features are extracted by the ED-MSER algorithm. Since the geometric shape of the extracted affine invariant features is polygon, we use homography to overlap these features to calculate the overlap-rate of the corresponding features. Based on this overlap-rate, the correctness of feature matching can be calculated. Two steps are included in the correctness calculation.

(a) Overlap the corresponding affine invariant features using homography and calculate the overlap-rate of the feature pairs by formula (3).
Overlap = \frac{R_{a\cap R_{bH}}}{R_{a} \cup R_{bH}}

where \(a,b\) is a image pair, \(H\) is homography, \(R_{a}\) is the area of feature on image \(a\), \(R_{bH}\) is the area of corresponding feature on image \(b\) which is converted from image \(a\) by homography, \(R_{a\cap R_{bH}}\) is the intersection region of corresponding features, \(R_{a} \cup R_{bH}\) is the union region of corresponding features.

(b) Ground truth is needed for the calculation of matching correctness. We assume the matching is correct if the overlap-rate of a feature pair is 50% or more. The matching correctness is computed by formula (4) as the ratio between the number of correct matches and the number of the total matches.

\[
\text{Ratio}_{CM} = \frac{NB_{CM}}{NB_{TM}}
\]

where \(NB_{CM}\) is the number of correct matches, \(NB_{TM}\) is the total number of matches.

2.2.2. Automatic optimization for matching

According to the matching results derived from affine invariant feature matching method in Section 2.1, the RANSAC-based automatic optimization of matching is as following:

(a) Randomly select four feature pairs from the results of matching.
(b) Compute homography \(H_{tmp}\) from the selected pairs.
(c) Compute inliers where Sum of Squared Difference (SSD) < \(\text{thresh}\) (assume an initial value to this distance threshold according to a specific objective, details discussed in next section).
(d) Keep largest set of inliers and re-compute homography \(H\) based on all of the inliers.
(e) Calculate the matching correctness using homography \(H\) based on the method in Section 2.2.1.
(f) If the matching correctness meets the optimization objective, end the process; if the matching correctness does not meet the optimization objective, adjust the distance threshold \(\text{thresh}\) automatically (difference strategies given in next section based on the specific optimization objectives), and start a new RANSAC estimation from step a.

To achieve the automatic process of the matching optimization, the distance threshold is automatically adjusted based on the matching correctness, until reaching the expected objective of the optimization. In the process of RANSAC iteration, a judgment should be made to determine whether optimization is feasible, if it is infeasible, directly terminate the optimization. Given in this section is a general process framework of automatically optimization for matching, different algorithms are derived in next section based on this framework, according to specific optimization objectives.

3. Two typical optimization approaches and algorithms

In this section, two typical optimization approaches (Approach A and B) with different optimization objectives are discussed. To implement these two approaches, the corresponding algorithms (Algorithm A and B) are developed based on the proposed technical framework. Usually, it is often very difficult to remove all the mismatches and simultaneously keep all the correct ones. On the other hand, it is possible to reach the objective of removing all the mismatches, but some correct matches may also be removed. The main reason includes two aspects: (1) The complexity of remote sensing image matching. For images with large geometric distortion, complex texture (forest, dense buildings, etc.), and small overlapping rate, automatic image matching process is very difficult and there may exist a number of mismatches after matching. (2) The limitation of homography capability. It should be pointed out that the change of terrain or height of surface feature on remote sensing images will more or less influence the effect of homography. When the elevation between surface features is large, the power of the homography will decrease, and further affect the accuracy of matching optimization in some extent.

The aim of Approach A is to remove all the mismatches, no matter whether some correct matches are also removed. To implement this Approach A, the corresponding Algorithm A is introduced:

(a) For a pair of images, extract affine invariant features by using ED-MSER algorithm, respectively; and then conduct affine invariant feature matching.
(b) Randomly select four feature pairs from the results of matching (check whether these features are colinear, if so, redo the selection).
(c) Compute homography \(H_{tmp}\) from the selected pairs.
(d) For each pair, a feature \(p_i = (x_i, y_i, 1)^T\) and its corresponding feature \(p'_i = (x'_i, y'_i, 1)^T\), calculate distance \(d_i = d(p_i, H_{tmp} p'_i) + d(p'_i, H_{tmp}^{-1} p_i)\); and compute SSD of the selected pairs.
(e) Compute inliers where SSD < \(\text{thresh}\) (Based on the objectives of Approach A, assume the initial distance threshold using a minimum value (1E-6)).
(f) Keep largest set of inliers and re-compute least-squares homography \(H\) based on all of the inliers.
IKONOS images covering a forest region, with spatial resolution of 1 m and size of 600 x 700 pixels, taken in Silkeborg, Denmark. Fig. 4 shows a pair of aerial images covering a region with mixed texture (trees and building), with a very high spatial resolution of 5 cm, a small overlapping rate and size of 3200 x 2000 pixels.

4. Experiment and analysis

4.1. Dataset

In order to test the validity and applicability of the proposed algorithms, different types of remote sensing images are selected and applied to the two optimization approaches respectively. The experiment data includes one pair of satellite images and two pairs of aerial images, which containing different view angles, different texture, different overlapping rate and different spatial resolution. The satellite images are downloaded from ISPRS official website (www.isprs.org). They are IKONOS images covering a forest region, with spatial resolution of 1 m and size of 600 x 700 pixels, shown in Fig. 2. Fig. 3 shows a pair of aerial images covering a region with dense buildings, with a big change of view angle (46°) and size of 650 x 600 pixels, taken in Silkeborg, Denmark. Fig. 4 shows a pair of aerial images covering a region with mixed texture (trees and building), with a very high spatial resolution of 5 cm, a small overlapping rate and size of 3200 x 2000 pixels.

4.2. Experimental results and analysis

ED-MSER algorithm is used to extract the affine invariant features from the three pairs of remote sensing images. Based on these extracted features, feature matching is conducted. Fig. 2 shows the results of image matching by the proposed algorithms on the pair of satellite imagery (IKONOS imagery with forest texture). (1) Without the automatic optimization, the results from affine invariant feature matching technology are shown in Fig. 2a. Through manually visual inspection, more results from affine invariant feature matching technology are shown in Fig. 2a. Through manually visual inspection, more
than 10 mismatches are found. (2) Following by Approach A, the matching correctness is set to 100% and Algorithm A is applied to these images. One hundred and nineteen pairs of matches are automatically obtained and the distance threshold is automatically set to 0.005. No mismatches are found, and all of the corresponding features are correctly matched, shown in Fig. 2b. The whole process takes 270s, including feature matching and matching optimization (this algorithm is implemented by MATLAB). The optimization process is automatically iterated at seven times. Fig. 5a illustrates that match numbers change with the changes of the distance threshold during this iteration processes. (3) Following by Approach B, the matching correctness is set to be more than 40%. Shown in Fig. 2c, 179 pairs of matches are automatically obtained and the distance threshold is automatically set to 16. Five mismatches are found. The whole process takes about 320s and the optimization
process is automatically iterated at five times. Fig. 5b illustrates that mismatch numbers change with the changes of the distance threshold during this iteration processes.

Fig. 3 shows the results of image matching by the proposed algorithms on the pair of aerial imagery (with dense building texture and large change of view angle). (1) Without the automatic optimization, the results only by affine invariant feature matching technology are shown in Fig. 3a. Through manually visual inspection, more than 10 mismatches are found. (2) Following by Approach A, the matching correctness is set to 100% and Algorithm A is applied to these images. One hundred and fourteen pairs of matches are automatically obtained and the distance threshold is automatically set to 0.01. No mismatches are found, and all of the corresponding features are correctly matched, shown in Fig. 3b. The whole process takes about 360s and the optimization process is automatically iterated at eight times. Fig. 5c illustrates that match numbers change with the changes of the distance threshold during this iteration processes. (3) Following by Approach B, the matching correctness is set to be more than 40%. Shown in Fig. 3c, 135 pairs of matches are automatically obtained and the distance threshold is automatically set to 16. Four mismatches are found. The whole process takes about 300s and the optimization process is automatically iterated at five times. Fig. 5d illustrates that mismatch numbers change with the changes of the distance threshold during this iteration processes.

Fig. 3. The results of image matching by the proposed method on the pair of aerial imagery (dense building texture, large change of view angle). (a) The matching results without optimization (number of total matches: 139; number of mismatches: more than 10). (b) The matching results by Algorithm 1 (number of total matches: 114; number of mismatches: 0; distance threshold: 0.01). (c) The matching results by Algorithm 2 (number of total matches: 135; number of mismatches: 4; distance threshold: 16).
Fig. 4 shows the results of image matching by the proposed algorithms on the pair of aerial imagery (with small overlapping rate and very high spatial resolution). (1) Without the automatic optimization, the results only by affine invariant feature matching technology are shown in Fig. 4a. Through manually visual inspection, more than 10 mismatches are found. (2) Following by Approach A, the matching correctness is set to 100% and Algorithm A is applied to these images. Twenty nine pairs of matches are automatically obtained and the distance threshold is automatically set to 0.001. No mismatches are found, and all of the corresponding features are correctly matched, shown in Fig. 3b. The whole process takes about 280s and the optimization process is automatically iterated at eight times. Fig. 5e illustrates that match numbers change with the changes of the distance threshold during this iteration processes. (3) Following by Approach B, the matching correctness is set to be more than 40%. Shown in Fig. 3c, 47 pairs of matches are automatically obtained and the distance threshold is automatically set to 0.5. No mismatches are found. The whole process takes about 390s and the optimization process is automatically iterated at ten times. Fig. 5f illustrates that mismatch numbers change with the changes of the distance threshold during this iteration processes.

From the experimental results, the proposed algorithms can produce much better matching results comparing to the standard affine invariant feature matching (without optimization). Although Algorithm A removes some correct matching features, all the mismatches are completely removed. Algorithm B preserves more correct matching features compared to Algorithm A, in spite of some mismatches are also kept. The experiments using remote sensing images with different texture, different view angle changes, and different spatial resolution demonstrated that the technical framework presented in Section 2 is effective because the two algorithms derived from this framework can reach the optimization objective of two approaches.

5. Conclusion

A new technical framework on remote sensing image matching by integrating affine invariant feature extraction and RANSAC is presented. The novelty of this framework is a strategy on automatic optimization for affine invariant feature matching based on RANSAC. An automatic way to determine the distance threshold of RANSAC is proposed, which is a
key problem to implement this RANSAC-based automatic optimization. After affine invariant feature extraction using EDMSER, homography is selected as geometric constraint model of RANSAC and then used to overlap these features to calculate the overlap-rate of the corresponding features. Based on this overlap-rate, the correctness of image matching can be computed, which is used as a quantitative evaluation indicator to automatically change the distance threshold to implement the iteration progress of RANSAC for the automatic optimization of image matching. Furthermore, two typical optimization approaches with two specific optimization objectives are introduced. Based on the proposed framework, two corresponding algorithms are developed to implement these two optimization approaches. The comparative experiment by using the proposed framework (with optimization) with the standard affine invariant feature matching (without optimization) indicates that the proposed framework can always get much higher matching score compared to the standard affine invariant feature matching method.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (Grant No. 41001238). Sincere thanks are given for the comments and contributions of anonymous reviewers and members of the Editorial team.
References


Liang Cheng received the Ph.D. degree in Photogrammetry and Remote Sensing, from State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China, in 2008. Currently, he is an associate professor with the Department of Geographical Information Science, Nanjing University, China. His areas of interest are remote sensing and photogrammetry, image understanding, pattern recognition, LiDAR data processing, and 3D reconstruction.

Manchun Li received the Ph.D. degree in Cartography and Geographic Information System, from School of Geographic and Oceanographic Sciences, Nanjing University, China, in 1997. Currently, he is a professor with the Department of Geographical Information Science, Nanjing University, China. His areas of interest are cartography, geographic information system, and image processing.

Yongxue Liu received the Ph.D. degree in Cartography and Geographic Information System, from School of Geographic and Oceanographic Sciences, Nanjing University, China, in 2004. Currently, he is an associate professor with the Department of Geographical Information Science, Nanjing University, China. His interest areas cover geographic information system and image processing.

Yanning Chen is a Ph.D. candidate in the Department of Geographical Information Science, Nanjing University, China. His areas of interest are LiDAR data processing, 3D reconstruction and Geographic Information Science.

Wenting Cai is a Master student in the Department of Geographical Information Science, Nanjing University, China. Her areas of interest are machine learning and remote sensing.

Kang Yang received the B.S. degree in Cartography and Geographic Information System, from School of Geographic and Oceanographic Sciences, Nanjing University, China, in 2008. Currently, he is a Ph.D. candidate with the Department of Geographical Information Science, Nanjing University, China. His interest areas cover remote sensing and image processing.