

Multi-sensor image registration by combining local self-similarity matching and mutual information

Xiaoping LIU¹, Shuli CHEN¹, Li ZHUO (✉)¹, Jun LI¹, Kangning HUANG²

¹ Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-Sen University, Guangzhou 510275, China

² School of Forestry and Environmental Studies, Yale University, New Haven, CT 06511, USA

© Higher Education Press and Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract Automatic multi-sensor image registration is a challenging task in remote sensing. Conventional image registration algorithms may not be applicable when common underlying visual features are not distinct. In this paper, we propose a novel image registration approach that integrates local self-similarity (LSS) and mutual information (MI) for multi-sensor images with rigid/non-rigid radiometric and geometric distortions. LSS is a well-performing descriptor that captures common, local internal layout features for multi-sensor images, whereas MI focuses on global intensity relationships. First, potential control points are identified by using the Harris algorithm and screened based on the self-similarity of their local surrounding internal layouts. Second, a Bayesian probabilistic model for matching the ensemble of the LSS features is introduced. Third, a particle swarm optimization (PSO) algorithm is adopted to optimize the point and region correspondences for maximum self-similarity and MI and, ultimately, a robust mapping function. The proposed approach is compared with several conventional image registration algorithms that are based on the sum of squared differences (SSD), scale-invariant feature transforms (SIFT), and speeded-up robust features (SURF) through the experimental registration of pairs of Landsat TM, SPOT, and RADARSAT SAR images. The results demonstrate that the proposed approach is efficient and accurate.

Keywords automatic registration, multi-sensor images, local self-similarity, mutual information, particle swarm optimization

1 Introduction

In many remote-sensing applications such as multispectral classification, environmental monitoring, image fusion, change detection, and weather forecasting, image registration is an important pre-processing procedure which involves combining relevant information from various types of imageries (Bentoutou et al., 2002; Klein, 2004; Bentoutou and Taleb, 2005a; Bentoutou et al., 2005b; Farah et al., 2008). Image registration is also a challenging task because remotely sensed image data are typically distorted geometrically and radiometrically when recorded by sensors on satellites or aircrafts. The sources of geometric distortion include the rotation of the Earth, the curvature of the Earth's surface, and the uncontrolled variation in the position and attitude of the remote-sensing platform (Richards and Jia, 2006). It becomes more challenging when image registration is applied to multi-sensor images, e.g., multi/hyper-spectral images and synthetic aperture radar (SAR) images because these two types of sensors exhibit large differences in instrumentation effects and imaging principles and reveal different characteristics of the Earth's surface (Mahmudul et al., 2012). Thus, there are increasingly rigid geometric deformations such as small variations in scale, orientation, and shear, and other non-rigid deformations among multi-sensor images. Simultaneously, since the various imaging mechanisms of multi-sensors can affect the measured brightness values of the pixels, thereby distorting their relative brightness and changing the distributions of brightness over an image (Richards and Jia, 2006), such radiometric distortions exacerbate registration difficulties as well. Thus, geometric and radiometric distortions are very common in multi-sensor images. The main purpose of image registration is to remove or suppress geometric distortions between the reference and sensed images to geometrically align the images (Meskine et al., 2010).

In manual registration, the selection of control points (CPs) is usually performed manually, which is inaccurate, time-consuming, and sometimes infeasible due to image complexity (Brook and Ben-Dor, 2011). Therefore, many studies have focused on automatic registration. Most existing automatic image registration techniques can be classified into two categories: intensity-based and feature-based (Zitová and Flusser, 2003). Intensity-based approaches, such as mutual information (MI), estimate transformations (mapping functions) on the basis of the intensity relationship between two whole images (Bentoutou et al., 2002). Feature-based approaches, such as the sum of squared differences (SSD) (Wolberg and Zokai, 2000), scale-invariant feature transform (SIFT) (Lowe, 2004; Yi et al., 2008), and speeded-up robust features (SURF) (Bay et al., 2008; Bouchiha and Besbes, 2013), perform registration by matching features (e.g., points, corners, contours, shapes, and regions) that are extracted from the images (Goshtasby et al., 1986; Li et al., 1995; Belongie et al., 2002; Wong and Clausi, 2007). The most critical issue for image registration, regardless of which approach is used, is the selection of the geometric mapping function (Bentoutou et al., 2002), whereas the accuracy of geometric mapping is strongly affected by the measurement of similarity. However, a perfect similarity measurement for multi-sensor registration is still not possible because radiometric and geometric distortions from multi-sensor images are more complex than those in similar-sensor images, and the reference image and sensed image may not share certain common underlying visual properties (Pratt, 1974; Abdel Sayed et al., 1995; Kim and Fessler, 2004; Shechtman and Irani, 2007; Arévalo and González, 2008; Borzi et al., 2009; Cole-Rhodes and Eastman, 2011).

In 2007, Shechtman and Irani proposed a novel similarity measure that is known as the “local self-similarity (LSS) descriptor” (Shechtman and Irani, 2007). LSS is a local feature descriptor that captures the internal geometric layouts of images based on a log-polar location grid. The LSS descriptor is stable against complex intensity variations, which means that it can capture more meaningful similarities in image patterns. Additionally, it allows local spatial affine distortions and non-affine distortions by using the binned log-polar representation. All of these processes can be achieved without prior learning and thus enable the matching of a wide variety of image and video types (Shechtman and Irani, 2007). With its high efficiency, LSS has been applied in many fields such as image data retrieval (Ken, 2009; Liu et al., 2009; Yang and Hou, 2012), image classification (Zheng, 2011; Zhang et al., 2013), image upscaling (Lee and Kim, 2012), and identification of anatomical landmarks (Ricardo, 2012). Scholars have also extended LSS to free viewpoint action recognition (Jiao, 2012) and multimodal sense stereo correspondence measures (Atousa, 2011). Despite its successful applications in other areas, the LSS method

has received little attention in the multi-sensor registration field. In this paper, we will demonstrate that LSS has great potential in the automatic registration of multi-sensor images.

Although LSS performs very well in capturing internal geometric layouts of local self-similarities with invariance to colour and edge variations, it is more limited by the lower discriminability of local features (Sedaghat and Ebadi, 2015) than by global features that are often observed. This issue may cause problems when locally optimal CPs are not correctly matched between images. Hence, it should be combined with another global similarity measure. Mutual information (MI), which was introduced by Viola (Viola and Wells, 1997) and Collignon (Collignon et al., 1995), could be a good choice because MI does not focus on information about the surface properties of objects (restrictive features) and is robust against variations in illumination. With its strength in obtaining global pattern information, the MI approach has been successfully applied to overcome problems that are associated with multimodal medical image registration (Taleb et al., 2001). Researchers have extended its application to multimodal remote-sensing image registration (Chen et al., 2003a, b; Cole-Rhodes et al., 2003; Arévalo and González, 2008; Suri and Reinartz, 2010; Cole-Rhodes et al., 2012). Therefore, combining LSS and MI can be an effective way to improve multi-sensor image registration.

In this paper, we propose a novel multi-sensor image registration method that combines local self-similarity and mutual information. The integrated method takes both local and global similarity into account, where local similarity is measured by LSS and global information is provided by MI. We also used the Harris corner detector (Harris and Stephens, 1988) to obtain potential CPs and the particle swarm optimization (PSO) (Messerschmidt and Engelbrecht, 2004) algorithm to optimize parameters that are related to the integrated similarity measurement of LSS and MI. A detailed introduction of the proposed registration method will be given in Section 2. In Section 3, we will present evaluation results of the proposed approach on multi-sensor remote-sensing images, including TM, SPOT, and SAR. Comparisons with conventional image registration methods, namely simple PSO with MI, sum of square differences (SSD), SIFT, and SURF, are also included in this section. Conclusions and discussions are presented in Section 4.

2 Methodology

This section describes how we integrated LSS and MI to develop an efficient and accurate multi-sensor image registration method. As shown in Fig. 1, the proposed registration approach consists of three main steps: In the first step, the Harris algorithm (Harris and Stephens, 1988)

is used to detect corner points (potential CPs), and the LSS descriptors are used to screen out less informative CPs. In the second step, matching of CPs on the sensed and reference images is performed by using the Bayesian model (Boiman and Irani, 2007), which has been demonstrated to be effective for matching local self-similarities across images and videos (Shechtman and Irani, 2007). In the third step, parameters of the geometric mapping function are optimized by applying the particle swarm optimization (PSO) algorithm and setting MI maximization as the optimization objective. Then, these parameters are applied to align the sensed and reference images.

2.1 Identification of CPs using LSS

The identification of CPs is a key step in image registration. Most CP extraction methods are based on local invariant features because of their robustness against geometric and illumination differences (Brook and Bendor, 2011). The existing feature-based methods extract a compact set of features (such as edges, lines, points, and shapes) for determining the CPs. They are normally used when the object features are distinct and easy to obtain. However, to be acquired by these methods, features often

must have specific characteristics such as well-defined lines, edges, or shapes. Furthermore, finding corresponding points in reference images for all CPs in a sensed image is difficult because of apparent and dramatic geometric distortions. The distortions increase in complexity when sensed and reference images have been obtained by different types of sensors under different weather, time, and altitude conditions. Matching the potential CPs is impossible if no further information is provided to help define the features of CPs.

The LSS descriptors do not require particular edges, lines, or intensities and effectively capture the local internal spatial layout of self-similarity and tolerate affine and intensity distortions. Matching CPs based on both point and region features is undoubtedly more precise than matching based on individual pixels (Shechtman and Irani, 2007). Therefore, LSS descriptors are introduced to identify CPs in this paper. Prior to that step, the Harris algorithm, which is a classic extraction arithmetic operator that is based on the point features of signals (Harris and Stephens, 1988), was used to select corner points, i.e., potential CPs. Since the number of points that are captured by the Harris algorithm is too large, we use a 10-pixel-radius window and impose a condition for removing less-useful corner points (Fig. 2). The window is used to select

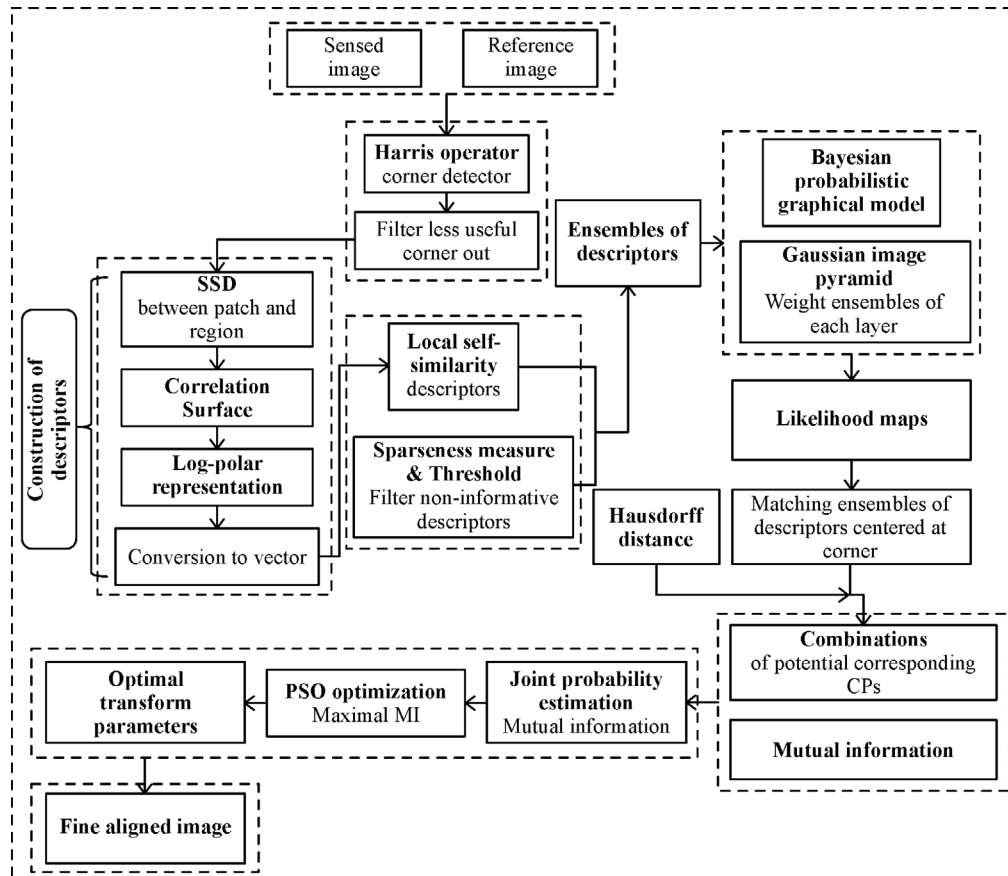


Fig. 1 Flowchart of the proposed method.

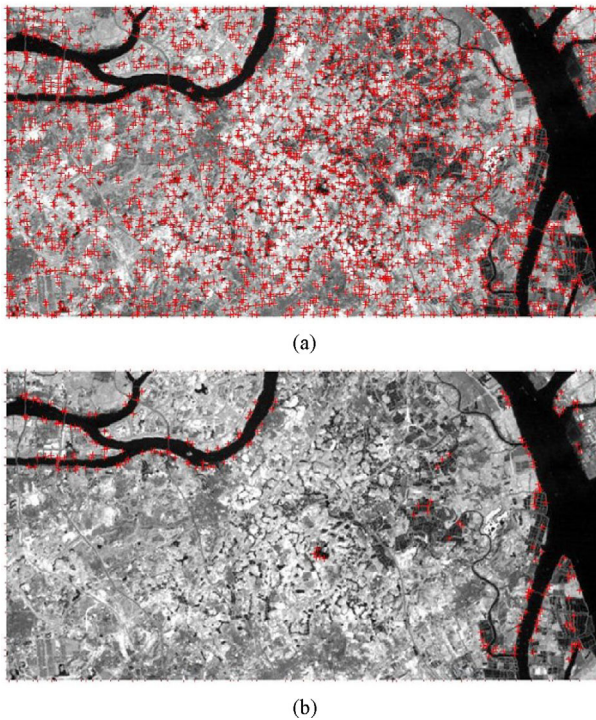


Fig. 2 (a) Corner points that were extracted from a TM image using the Harris algorithm. (b) Remaining corner points after less useful ones were removed.

informative CPs that are as evenly distributed as possible within an image. The condition that is used to filter CPs is that at least half of the window around a corner point must contain dark objects, typically water bodies. Via this filtering, the remaining CPs could have higher potential for generating informative LSS descriptors (Shechtman and Irani, 2007). Then each remaining CP is described by its relative position and invariant local internal layout features, which are defined by an ensemble of LSS descriptors. In this study, such CPs are defined as useful. CPs on the reference and sensed images are matched in pairs only when they share very similar LSS descriptor values and relative positions.

According to Shechtman and Irani's work, the LSS descriptor is expressed as the correlation between a centered patch with its surrounding image region. The calculation is initially achieved by the sum of squared differences (SSD) between a small image patch and its surrounding large image region, which generates a local internal correlation surface that describes the local spatial layouts in detail. To account for radially increasing rigid and non-rigid distortions, the correlation surface is transformed into a log-polar binned form and the maximal value is adopted in every bin (Shechtman and Irani, 2007). In this study, the LSS descriptor refers to a local image region that is centered at p_i (typically of radius 10) and contains the surrounding image patches (typically 3×3), which are transformed to a binned form that contains 80

bins (20 evenly spaced angles and 4 radial intervals of equal length). In addition, the vector of 80 entries is normalized by linearly stretching its values to the range of $[0, 1]$ to make it invariant under intensity distribution differences between corresponding patches and their surrounding image regions. The descriptor is constructed in such a way to capture detailed internal layout information while tolerating affine and non-rigid deformations. This makes it especially suitable for multi-sensor image registration, where the sensed image and reference image have different patterns and intensities. Then, the features of CPs are constructed by searching the LSS descriptors that are within a local image centered at each CP, which is much larger than the region for capturing a LSS descriptor. The CPs will be ready for matching once their LSS descriptors have been constructed.

2.2 Matching CPs using ensembles of LSS descriptors

Matching CPs on the sensed and reference images is performed by using the Bayesian model (Boiman and Irani, 2007), which has been demonstrated to be effective for matching local self-similarities across images and videos (Shechtman and Irani, 2007). To find the pairs that share the most common spatial distribution and relationship of local descriptors, a likelihood map is calculated for each CP in a reference map by using a Gaussian image pyramid weighted Bayesian probabilistic model. The location of the peak value in the likelihood map indicates a potential corresponding CP in the registered image. To ensure the accuracy of the local matching, the Hausdorff distance is also used in this step.

Similarity information of a single descriptor is inadequate for accurate matching of CPs. Therefore, we adopt an ensemble of LSS descriptors to account for the spatial information among descriptors. By matching such an ensemble of LSS descriptors, we not only obtain the local internal spatial layout of a descriptor for CPs but also capture a large range of geometric layouts that are combined by an ensemble of descriptors for CPs.

A is the sensed image, B is the reference image, p_i is the extracted CP, and L_i is the local image that is centered at p_i (typically of radius 20). In each L_i , an ensemble of descriptors is searched within the local image.

A good match is an ensemble of descriptors in the local image L_i of image A that corresponds to a similar ensemble of descriptors in image B . Such pairs of ensembles share similar descriptors in terms of values and relative geometric positions. In this case, a CP that is centered at local image L_i of image A can find its corresponding CP in image B . In addition, small local shifts are permitted to account for small non-rigid deformations during the matching process. To achieve this, an ensemble matching algorithm (Boiman and Irani, 2007), which is mainly based on a Bayesian probabilistic model, is adopted to capture relative geometric relations among the local descriptors

within local images. The procedure of the Bayesian probabilistic model is presented in Fig. 3.

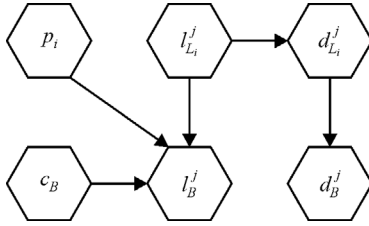


Fig. 3 Bayesian probabilistic model for matching ensembles of descriptors.

Specifically, p_i denotes a centered CP in L_i and c_B denotes a point that may correspond to p_i (Fig. 3). We define $q^1, q^2, q^3, \dots, q^n$ as a region that contains a descriptor that is associated with the following two attributes: 1) the descriptor vector d_j and 2) its location in absolute coordinates l . $d_{L_i}^j$ denotes the j^{th} observed descriptor in L_i , and $l_{L_i}^j$ denotes the location of the j^{th} observed descriptor. Similarly, d_B^j represents the descriptor vector of the j^{th} hidden region in B , and l_B^j represents its location. The likelihood of similarity between this pair of ensembles can be expressed as follows:

$$P(p_i, c_B) = P(p_i, d_{L_i}^1, \dots, l_{L_i}^1, \dots, c_B, d_B^1, l_B^1, \dots). \quad (1)$$

On the basis of the Bayesian probabilistic model, we obtain:

$$\begin{aligned} & P(p_i, d_{L_i}^1, \dots, l_{L_i}^1, \dots, c_B, d_B^1, l_B^1, \dots) \\ &= \alpha \prod_j P(l_B^j | l_{L_i}^j, p_i, c_B) P(d_B^j | d_{L_i}^j) P(d_{L_i}^j | l_{L_i}^j). \end{aligned} \quad (2)$$

Locations with high likelihood values are considered the detected locations of L_i within B . Considering that self-similarity may appear at various scales and in regions of different sizes, we capture LSS descriptors at multiple scales by applying a Gaussian image pyramid. Every ensemble of descriptors is searched at each scale

independently to generate likelihood maps. By normalizing each likelihood map based on the number of descriptors in its scale, the maps can be weighted according to the degree of sparseness (Hoyer, 2004) on a variety of scales. Figure 4 shows examples of likelihood maps for two CPs.

2.3 Optimization of registration based on MI

The LSS approach mainly focuses on local information while the global geometric layout and overall spatial distribution are barely considered. This may lead to inaccurate matching because points with similar surrounding distributions may be incorrectly chosen as corresponding CPs. The test point error (Greenfeld, 2002) and modified Hausdorff distance methodology are often used to solve this problem. However, they both locally compare the geometric relationship of a collection of points to the whole image. To take advantage of both local and global information, we use the maximum MI as the optimization objective in this paper and introduce a particle swarm optimization (PSO) algorithm for searching for the best possible CP matches and parameters of the geometric mapping function.

MI is defined as the information that is contained in two random variables A and B about each other. MI can be expressed as follows:

$$I(A, B) = H(A) + H(B) - H(A, B), \quad (3)$$

where $I(A, B)$ is the MI of A and B , and $H(A)$ and $H(B)$ are the entropies of A and B , respectively, while $H(A, B)$ is their joint entropy.

When the MI method is used in image auto-registration, the sensed image and the reference image can be considered as variables A and B , respectively. The two images should be registered when $I(A, B)$ reaches its maximum value. The entropies of the two images and their joint entropy can be calculated by the following equations:

$$H(A) = - \sum_a p_A(a) \log_{AP}(a), \quad (4)$$

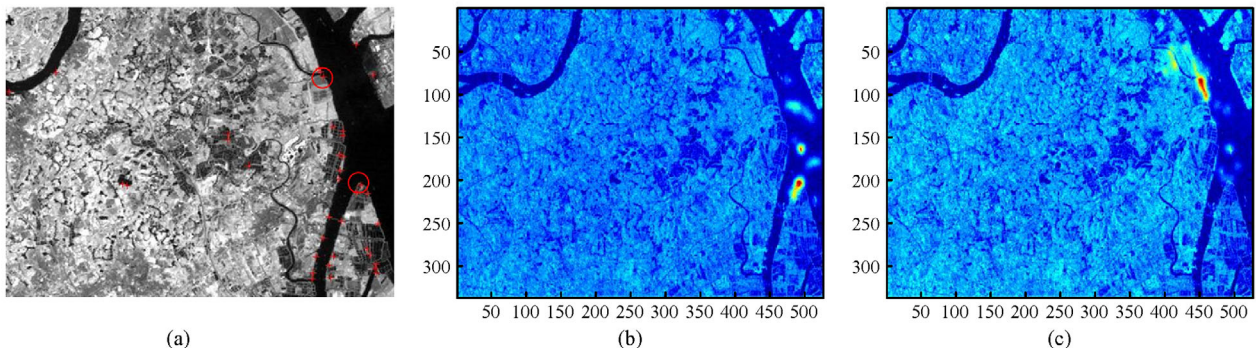


Fig. 4 Likelihood maps (b, c) for matching two control points (circled in red) in reference image (a).

$$H(B) = -\sum_b p_B(b) \log_{BP}(b), \quad (5)$$

$$H(A,B) = -\sum_{a,b} p_{A,B}(a,b) \log_{A,BP}(a,b), \quad (6)$$

where $p_A(a)$ and $p_B(b)$ are marginal probability functions and $p_{A,B}(a,b)$ is the joint probability function. $p_{A,B}(a,b)$ can be obtained via the following equation:

$$p_{A,B}(a,b) = \frac{h(a,b)}{\sum_{a,b} h(a,b)}, \quad (7)$$

where h is the joint histogram, which is a two-dimensional matrix that indicates the numbers of intensity pairs in the reference and sensed images (Chen et al., 2003b). $p_A(a)$ and $p_B(b)$ can be computed in a similar manner.

$$h(a,b) = \begin{pmatrix} h(0,0) & h(0,1)\dots & h(0,N-1) \\ h(1,0) & h(1,1)\dots & h(1,N-1) \\ \dots & \dots & \dots \\ h(M-1,0) & h(M-1,0)\dots & h(M-1,N-1) \end{pmatrix}. \quad (8)$$

In the joint histogram, M and N are the ranges of the intensity values of the two images; $h(a,b)$ is the number of pixel pairs with intensity value a in A and intensity value b in B . The primary characteristic of the joint histogram is its increasing dispersion level with the mis-registration of the two images (Liang et al., 2014). The value of MI reaches its maximum when the sensed and reference images are accurately registered.

PSO is a population-based evolutionary computation technique (Kennedy and Eberhart, 1995, 2001) with strong search capabilities. In comparison with genetic algorithms, which exploit the competitive characteristics of biological evolution (e.g., survival of the fittest), PSO exploits cooperative and social aspects, such as the flocking of birds and the swarming of insects (Wachowiak et al., 2004). Starting from a diffused status, populations (particles) tend to move in the search space. All PSO particles in an N -dimensional space are searching for their own best fit. The investigation of the theoretical properties of PSO is an active research area (Clerc and Kennedy, 2002).

In PSO, velocity iteration and particle location can be expressed as follows:

$$v(t+1) = \omega v(t) + c_1 r_1 (pbest - x(t)) + c_2 r_2 (gbest - x(t)), \quad (9)$$

$$x(t+1) = x(t) + v(t+1), \quad (10)$$

where $v(t+1)$ is the velocity of each particle in the next iteration, $v(t)$ is its current velocity, $pbest$ is the personal

best particle, and $gbest$ is the global best particle. Furthermore, $x(t+1)$ represents the new location of the particle; ω represents the inertial weight, which is the effect of the current velocity on the next iteration; and c_1 and c_2 are study factors that represent the information exchange between each particle in the whole population; they are usually given the value of 2 in calculations. r_1 and r_2 are “acceleration coefficients”, which are random numbers that are uniformly distributed in the range of $[0,1]$ and are used to increase the randomness of particle movement.

In this paper, we set combinations of possible CPs that are found in image A as particles and update each particle’s location by randomly choosing other matched points in image A . The optimization objective is to identify the optimal combinations of CPs and to maximize the MI of the two images. Thus, we limit the space and constrain particles by finding feasible solutions in a specific space; the results of this approach indicate that better solutions should be searched on the basis of previous iterations. This approach can also improve the calculation speed and accuracy of the algorithm and solve the integration problem of the polynomial mapping function with MI. By finding the maximum MI using combinations of matched pairs of points as particles in PSO, we obtain the following benefits:

1) Existing registration methods that are based on MI use PSO only to focus on affine deformations, such as translation, rotation, and scaling. However, non-affine distortions are excluded; this situation is unsuitable for the registration of multi-sensor images. By adopting the cluster pairs of points as PSO particles, the resulting polynomial mapping function can correct non-rigid deformations. Then, the polynomial mapping function is used to align sensed images with reference images.

2) By integrating LSS and MI, the global and local information are well-balanced. We guarantee not only the accuracy of local point matching but also the best ensemble of point matching geometrically and intensively from a global perspective.

3 Experiments and results

The proposed scheme is tested by registering various types of remote-sensing images from multi-sensors, including optical data from Landsat TM and microwave imagery of RADARSAT SAR (Fig. 5). These images cover the same study area in Guangzhou, which is the largest city in south China. The TM imagery and SAR imagery have uncorrelated illumination and different data collection times. The optical sensors operate in the visible and infrared regions of the electromagnetic spectrum to provide information, whereas the SAR sensor generates a directed beam of pulses that illuminate terrain to produce high-resolution back-scattering of radar-frequency energy. On the SAR image, large bright areas exist where some objects

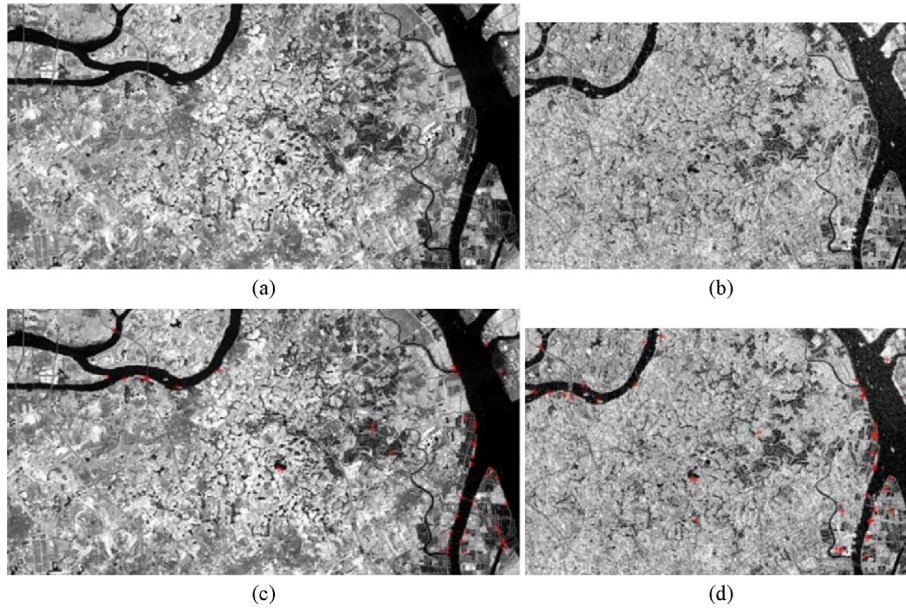


Fig. 5 Multi-sensor images that were used to evaluate the LSS-MI registration method. (a) The TM (band 4398 \times 763) image. (b) The SAR (HV) image (335 \times 526). CPs that were extracted from remote-sensing images. (c) CPs that were extracted from the TM image. (d) CPs that were extracted from the SAR image.

are oriented in divergent positions to the satellite track direction. Thus, strong corner reflections are generated; however, this type of reflection does not exist in the optical image. Reflectance from these areas is relatively lower in the optical image than the SAR image.

During the preprocessing step, all the images are resampled to the same resolution of the finer image so that registration can be performed and the proposed method can be assessed.

Section 3 consists of three parts: In the first part, the procedure of applying the proposed method is presented using Landsat TM and SAR images. The second part shows the performances of the proposed method in registering various types of imagery, such as Landsat TM, SPOT, and RADARSAT SAR. The third part compares the proposed method with three different feature-based methods, i.e., SSD, SIFT, and SURF.

3.1 Procedure of the proposed registration method

A Landsat TM image (398 \times 763) and a close-range RADARSAT SAR image (335 \times 526) were used to demonstrate the steps of applying the LSS-MI. Acquired by different platforms, these two images have different spectral responses, spatial resolutions, and observation times. The TM image, which has a resolution of 30 m, was manually pre-registered and resampled to the same resolution as the SAR image, which is 12.5 m.

First, a coarse result was obtained by using an ensemble of LSS descriptors. Then, PSO optimization was performed. The parameters of PSO were set as follows: The particle population in the iteration was initially restricted to

30 individuals with 30 dimensions. Iterations were set to stop if the *gbest* value remains unchanged for 20 runs or the iteration count reaches a maximum value of 200. The hybrid PSO uses combinations of possible CPs as particles. To test the performance of the PSO parameters, we compared our algorithm with an original PSO for MI by using the TM and SAR images that are shown in Fig. 5. Figure 6 shows the evolution of the best solution during iterations. The MI index indicates the fitness value of the best particle that was obtained during the runs of PSO. A higher maximum fitness provides better accuracy in estimating the optimal value, which was identified by searching with the newly proposed PSO with MI. Then, the novel algorithm was applied to the pair of TM and SAR images.

As shown in Fig. 6, our algorithm outperformed the original algorithms. When the simple PSO algorithm is used, the rate of the MI of the sensed and reference images

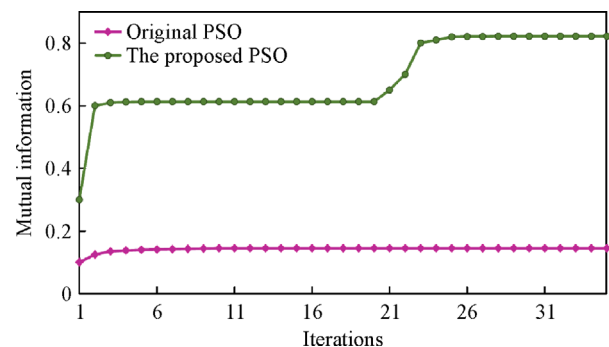


Fig. 6 Evolution of best solution during the run of PSOs.

increases slowly. The maximum MI value is also low (approximately 0.14), compared with that of our modified PSO algorithm (approximately 0.8). The comparison results demonstrate that the modified PSO algorithm for MI has a high chance of finding the best set of transform parameters.

3.2 Effectiveness of the LSS-MI method

In this section, remote-sensing images of various spatial resolutions from various sensors such as Landsat TM, SPOT, and RADARSAT SAR images were used to demonstrate the effectiveness of the proposed LSS-MI method. First, we used a single-band (band 4) TM image and a single-band (band 4) SPOT image. The spatial resolutions of the TM image and the SPOT image are 30 m and 10 m, respectively. Although their spectral responses are similar, they were measured at different times. As stated in Section 2, these two images were resampled to the same resolution before performing the LSS-MI registration. As shown in Fig. 7(a) and 7(b), the textures and intensities of these two images are different. However, most pairs of control points are still accurately matched. The second pair of images includes a band 4 TM image and a panoramic band SPOT image which has a spatial resolution of 2.5 m. As shown in Figs. 7(c) and 7(d), the

differences between the two images are even more apparent than with the first pair, which makes it very difficult to select control points manually. However, the proposed LSS-MI method performs very well under these conditions and matches most pairs of control points accurately. The third pair of multi-sensor images includes the TM image and the SAR image that were considered in Section 2. The registration results (Fig. 8) show that the LSS-MI method successfully matched the two spectral response images.

3.3 Performance of the LSS-MI method

Typical feature-based image registration methods such as SSD (Wolberg and Zokai, 2000), SIFT (Yi et al., 2008), and SURF (Bouchiha and Besbes, 2013) are compared with our method. The matching performances of the SSD-based method, SURF-based method, and our proposed method are presented in Fig. 9. The SIFT-based method was omitted from the figure because it failed to match the corner points. Two statistical indicators, the number of obtained matches in the outputs and the RMSE of the registration results, are listed in Table 1. Given that the images we used have different sizes (398×763 and 335×526), we modified the RMSE equation as follows:

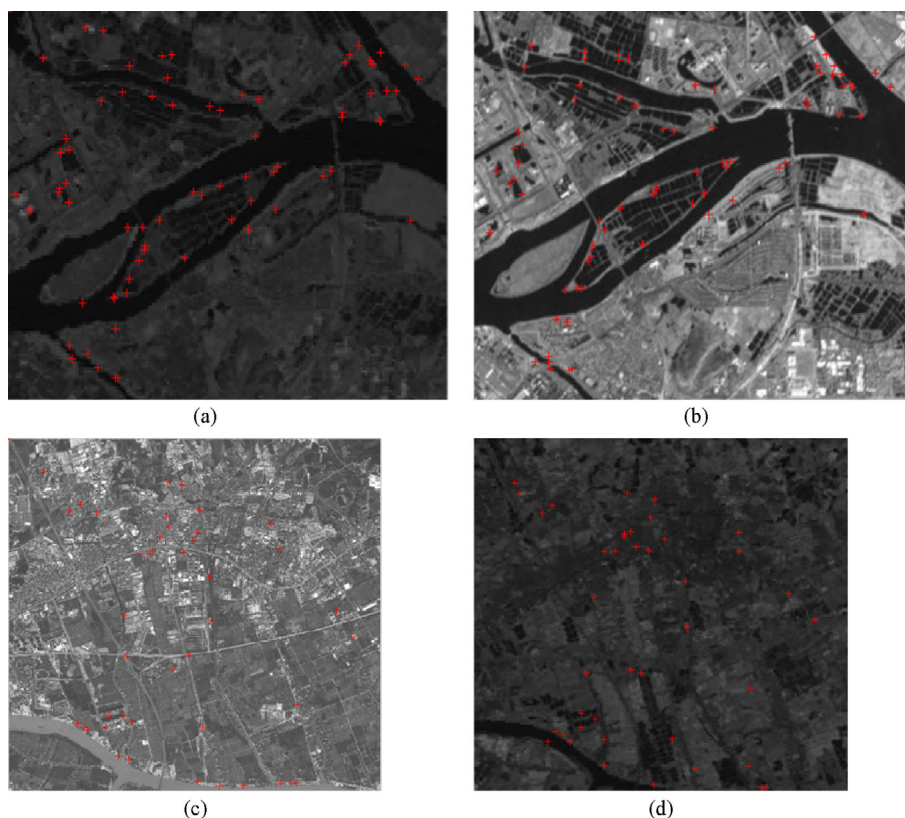


Fig. 7 (a) Landsat 5 band 4 image (30 m). (b) SPOT band 4 image (10 m). (c) Landsat 5 band 4 image (30 m). (d) SPOT Panchromatic band image (2.5 m).

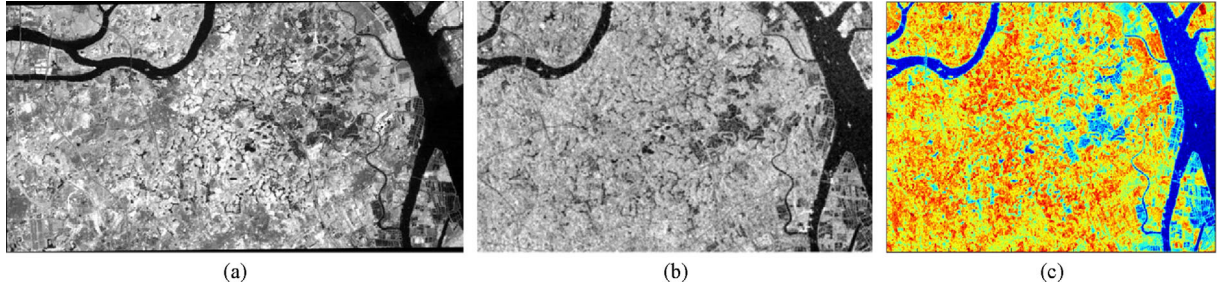


Fig. 8 (a) The aligned image. (b) The reference image. (c) Merged image of the aligned image and the reference image.

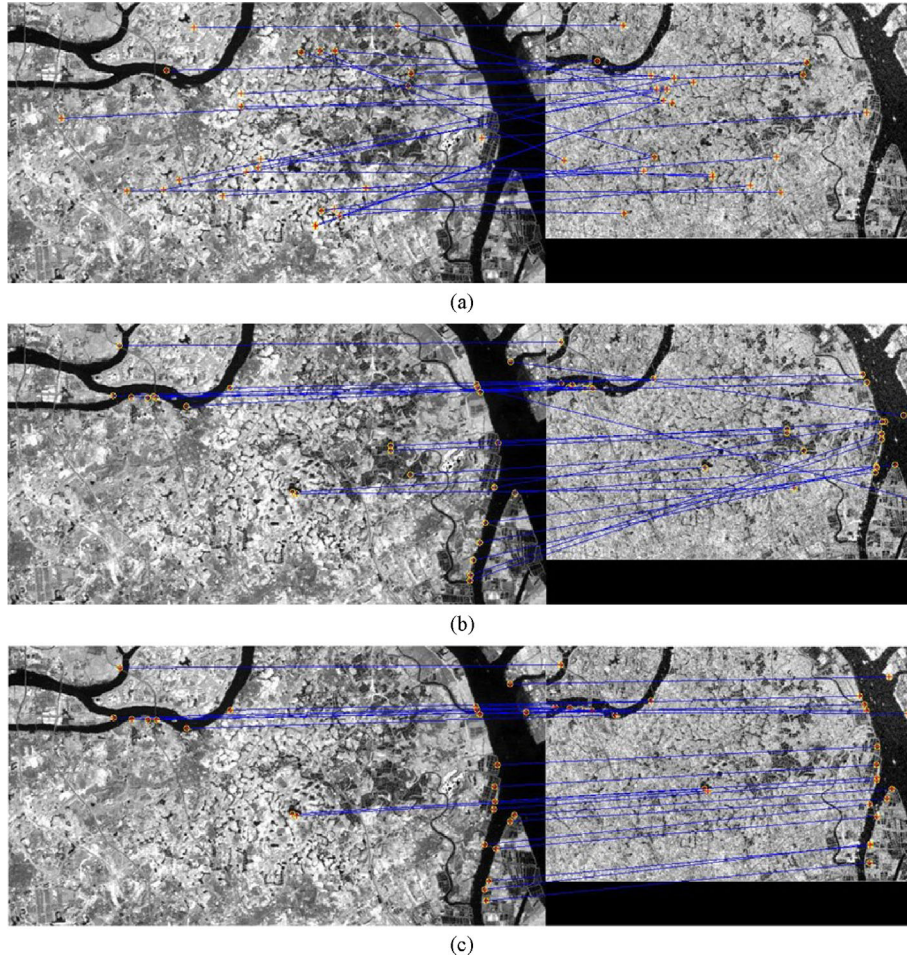


Fig. 9 Comparison between the SURF-based, SSD-based methods and our proposed method. (a) Matching performance based on SURF. (b) Matching performance based on SSD. (c) Matching performance based on integrated LSS and MI.

$$\text{RMSE} = \sqrt{\sum \left(\left((x_i^A - x_i^B)^2 + (y_i^A - y_i^B)^2 \right) / n \right)}, \quad (11)$$

where x_i^A and x_i^B denote the relative x -coordinates of the i^{th} CP in images A and B , respectively; y_i^A and y_i^B denote the relative y -coordinates of the i^{th} CP in images A and B , respectively; and n is the number of captured pairs of CPs.

The SSD- and SURF-based methods were able to obtain matches (Table 1 and Fig. 9). However, the RMSE of the

registration results that were acquired by using these two algorithms were large and, hence, unacceptable. The LSS-based method outperformed the SSD- and SURF-based methods in terms of both the number of obtained matched pairs and RMSE. However, its RMSE value was still larger than that of our proposed method. The SIFT-based method delivers no matches in outputs because the features in the images that correspond to different sensor types share different properties (e.g., intensity, shape, line, and

Table 1 Performances of SSD, SIFT, SURF, LSS and our proposed algorithms

| Similarity measure | Number of obtained matches in the output | RMSE (modified) |
|--------------------|--|-----------------|
| SSD | 15 | 0.2112 |
| SIFT | 0 | - |
| SURF | 10 | 0.3193 |
| LSS | 17 | 0.2082 |
| LSS-MI | 25 | 0.0929 |

gradient). In conclusion, our proposed method produces the most matched pairs and the lowest RMSE, which ensures the best registration results for the test set of multi-sensor images. The aligned image and the reference image that are acquired by using our proposed method are shown in Fig. 8.

4 Discussion and conclusions

In this paper, we present a new method for automated multi-sensor image registration which exploits similarity from both local and global viewpoints. The LSS method is used to capture the local internal layout, thereby allowing both affine and non-affine deformations. Corner point matching based on an ensemble of descriptors is used to capture the geometric layout on various scales (local region to local image). From a global viewpoint, we optimize MI by using a modified PSO algorithm which improves the MI rate significantly and, hence, ensures the accuracy and robustness of the mapping function.

We evaluated the proposed approach by applying it to various types of remote-sensing imagery in Guangzhou, China. The results show that the LSS-MI method can accurately and effectively register multi-sensor images with various resolutions and imaging principles. A comparison with other existing image registration methods, namely, SSD, SIFT, SURF, and LSS, shows that our proposed method is the most robust and efficient in dealing with greyscale and geometric discrepancies between corresponding pixels and regions, and affine and non-affine deformations.

The proposed method still has a few limitations: First, the quality of initially selected potential CPs by the Harris detector may influence the performance of the proposed method. Therefore, it is necessary to screen out many points that are not useful, which may otherwise consume substantial computational resources. Second, when the LSS descriptors' discriminability is relatively low, it may influence how reliably different features are distinguished and matched. Although we introduce the Bayesian model and the Gaussian image pyramid to help solve this problem, the computational cost inevitably rises. In

addition, several empirical parameters should be set in advance before applying the proposed method. Typically, the radius of the window for filtering CPs was set as 10 in the CP identification step. The local self-similarity descriptor was defined with radius of 10 based on a unit of patches of size 3×3 . In the CP matching step, the local image radius was generally set to 20. During the MI optimization, the particle population in the iteration was initially restricted to 30 individuals with 30 dimensions. Iterations were set to stop if the maximum fitness remained unchanged for 20 runs or the iteration count reached a maximum value of 200. Nevertheless, these empirical parameters seemed to perform well with the multi-sensor images that were considered in our experiments.

Future research may focus on topics such as finding a more robust way of selecting initial corner points and better balancing between calculation accuracy and efficiency. For further applications, it will also be necessary to examine the performance of the LSS-MI method in more challenging multi-sensor image registration tasks, such as registering a visible band image with a LIDAR image or registering a thermal band image with a visible band image, where similarity of the geometry or intensity distributions is extremely low.

Acknowledgements The work was supported by the National Natural Science Foundation of China (Grant No. 41371499) and the Natural Science Foundation of Guangdong Province (No. 2015A030313505).

References

- Abdel Sayed S, Ionescu D, Goodenough D (1995). Matching and registration method for remote sensing images. In: Proceedings of Geoscience and Remote Sensing Symposium, 2, 1029–1031
- Arévalo V, González J (2008). Improving piecewise linear registration of high-resolution satellite images through mesh optimization. *IEEE Trans Geosci Remote Sens*, 46(11): 3792–3803
- Atousa T (2011). Local self-similarity as a dense stereo correspondence measure for thermal-visible video registration. In: Proceedings of the 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, Washington, DC, USA
- Bay H, Ess A, Tuytelaars T, Van Gool L (2008). Speeded-up robust features (SURF). *Comput Vis Image Underst*, 110(3): 346–359
- Belongie S, Malik J, Puzicha J (2002). Shape matching and object recognition using shape contexts. *IEEE Trans Pattern Anal Mach Intell*, 24(4): 509–522
- Bentoutou Y, Taleb N (2005a). A 3-D space-time motion detection for an invariant approach image registration approach in digital subtraction angiography. *Comput Vis Image Underst*, 97(1): 30–50
- Bentoutou Y, Taleb N (2005b). Automatic extraction of control points for digital subtraction angiography image enhancement. *IEEE Trans Nucl Sci*, 52(1): 238–246
- Bentoutou Y, Taleb N, Chikr El Mezouar M, Taleb M, Jetto L (2002). An invariant approach for image registration in digital subtraction angiography. *Pattern Recognit*, 35(12): 2853–2865

- Boiman O, Irani M (2007). Detecting irregularities in images and in video. *Int J Comput Vis*, 74(1): 17–31
- Borzi A, Bisceglie M D, Galdi C, Giangregorio G (2009). Robust registration of satellite images with local distortions. In: *Proceedings of 2009 IEEE International Geoscience and Remote Sensing Symposium*, 3: III-251–III-254
- Bouchiha R, Besbes K (2013). Automatic remote-sensing image registration using SURF. *International Journal of Computer Theory and Engineering*, 5(1): 88–92
- Brook A, Ben-Dor E (2011). Automatic registration of airborne and spaceborne image topology map matching with SURF processor algorithm. *Remote Sens*, 3(1): 65–82
- Chen H M, Arora M K, Varshney P K (2003a). Mutual information-based image registration for remote sensing data. *Int J Remote Sens*, 24(18): 3701–3706
- Chen H M, Varshney P K, Arora M K (2003b). Performance of mutual information similarity measure for registration of multitemporal remote sensing images. *IEEE Trans Geosci Remote Sens*, 41(11): 2445–2454
- Clerc M, Kennedy J (2002). The particle swarm—Explosion, stability, and convergence in a multidimensional complex space. *IEEE Trans Evol Comput*, 6(1): 58–73
- Cole-Rhodes A A, Eastman R D (2011). Gradient descent approaches to image registration. In: Moigne J L, Netanyahu N S, Eastman R D, eds. *Image Registration for Remote Sensing*. Cambridge: Cambridge University, 265–276
- Cole-Rhodes A, Johnson K L, Moigne J L, Zavorin I (2003). Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. *IEEE Transactions on Image*, 12(12): 1495–1511
- Cole-Rhodes A, Johnson K, Le Moigne J (2012). Multiresolution registration of remote sensing images using stochastic gradient. In: Szu H H, Buss J R, eds. *Wavelet and Independent Component Analysis Applications IX*. SPIE Proceedings Vol. 4738
- Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P, Marchal G (1995). Automated multimodality image registration based on information theory. *Inf Process Med Imaging*, 3: 263–274
- Farah I R, Boulila W, Etabaâ K S, Solaiman B, Ahmed M B (2008). Interpretation of multisensor remote sensing images: multiapproach fusion of uncertain information. *IEEE Trans Geosci Remote Sens*, 46(12): 4142–4152
- Goshtasby A, Stockman G C, Page C V (1986). A region-based approach to digital image registration with subpixel accuracy. *IEEE Trans Geosci Remote Sens*, GE-24(3): 390–399
- Greenfeld J S (2002). Matching GPS Observation to Location on a Digital Map. In: *Proceedings of the 81st Annual Meeting of the Transportation Research Board*, (3): 13
- Harris C, Stephens M (1988). A combined corner and edge detector. In: *Proceedings of Fourth Alvey Vision Conference*, 147–151
- Hoyer P O (2004). Non-negative matrix factorization with sparseness constraints. *J Mach Learn Res*, 5: 1457–1469
- Jiao W (2012). Free Viewpoint Action Recognition based on Self-similarities. In: *Proceedings of the 11th International Conference on Signal Processing (ICSP)*, 2, 1131–1134
- Ken C (2009). Efficient Retrieval of Deformable Shape Classes using Local Self-Similarities. In: *Proceedings of 2009 IEEE 12th International Conference on Computer Vision Workshops*, 264–271
- Kennedy J, Eberhart R C (1995). Particle swarm optimization. In: *Proceedings of the IEEE International Conference on Neural Networks*, 4, 1942–1948
- Kennedy J, Eberhart R C (2001). *Swarm Intelligence*. San Francisco: Morgan Kaufmann Publisher
- Kim J, Fessler J A (2004). Intensity-based image registration using robust correlation coefficients. *IEEE Trans Med Imaging*, 23(11): 1430–1444
- Klein L A (2004). *Sensor and Data Fusion: A Tool for Information Assessment and Decision Making*. Bellingham: SPIE Press, 8–10
- Lee H K, Kim T C (2012). Local self-similarity based backprojection for image upscaling. In: *Proceedings of 2012 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1215–1218
- Li H, Manjunath B S, Mitra S K (1995). A contour-based approach to multisensor image registration. *IEEE Trans Image Process*, 4(3): 320–334
- Liang J, Liu X, Huang K, Li X, Wang D, Wang X (2014). Automatic registration of multisensor images using an integrated spatial and mutual information (SMI) metric. *IEEE Trans Geosci Remote Sens*, 52(1): 603–615
- Liu S, Du X Y, Zhang J H (2009). Structure extracting and matching based on similarity-pictorial structure model for microscopic images. In: *Proceedings of International Conference on Artificial Intelligence*, 3: 181–185
- Lowe D G (2004). Distinctive image features from scale-invariant key points. *Int J Comput Vis*, 60(2): 91–110
- Meskine F, Mezouar M C E, Taleb N (2010). A rigid image registration based on the non subsampled contourlet transform and genetic algorithms. *Sensors (Basel)*, 10(9): 8553–8571
- Messerschmidt L, Engelbrecht A P (2004). Learning to play games using a PSO-based competitive learning approach. *IEEE Trans Evol Comput*, 8(3): 280–288
- Pratt W K (1974). Correlation techniques of image registration. *IEEE Trans Aerosp Electron Syst*, AES-10(3): 353–358
- Ricardo G (2012). Landmark localisation in brain MR images using feature point descriptors based on 3D local self-similarities. In: *Proceedings of the 9th IEEE International Symposium on Biomedical Imaging*, 1535–1538
- Richards J A, Jia X (2006). *Remote Sensing Digital Image Analysis* (4th ed). Berlin: Springer-Verlag, 56–58
- Sedaghat A, Ebadi H (2015). Distinctive order based self-similarity descriptor for multi-sensor remote sensing image matching. *ISPRS J Photogramm Remote Sens*, 108: 62–71
- Shechtman E, Irani M (2007). Matching local self-similarities across images and videos. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1–8
- Suri S, Reinartz P (2010). Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas. *IEEE Trans Geosci Remote Sens*, 48(2): 939–949
- Taleb N, Bentoutou Y, Deforges O, Taleb A (2001). A 3-D space-time motion evaluation for image registration in digital subtraction angiography. *Comput Med Imaging Graph*, 25(3): 223–233
- Viola P, Wells W M III (1997). Alignment by maximization of mutual information. *Int J Comput Vis*, 24(2): 137–154
- Wachowiak M P, Smolikova R, Zheng Y, Zurada J M, Elmaghraby A S

- (2004). An approach to multimodal biomedical image registration utilizing particle swarm optimization. *IEEE Trans Evol Comput*, 8 (3): 289–301
- Wolberg G, Zokai S (2000). Robust image registration using log-polar transform. In: *Proceedings of IEEE International Conference on Image Processing*, 1: 493–496
- Wong A, Clausi D A (2007). ARRSI: automatic registration of remote sensing images. *IEEE Trans Geosci Remote Sens*, 45(5): 1483–1493
- Yang H, Hou X (2012). Local self-similarity based texture classification. In: *Proceedings of the 5th International Congress on Image and Signal Processing (CISP)*, 795–799
- Yi Z, Chen Z, Yang X (2008). Multi-spectral remote image registration based on SIFT. *Electron Lett*, 44(2): 107–108
- Zhang H G, Bai X, Zheng H X, Zhao H J, Zhou J, Cheng J, Lu H (2013). Hierarchical remote sensing image analysis via graph laplacian energy. *IEEE Geosci Remote Sens Lett*, 10(2): 396–400
- Zheng H (2011). A novel approach for satellite image classification using local self-similarity. In: *Proceedings of Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, 2888–2891
- Zitová B, Flusser J (2003). Image registration methods: a survey. *Image Vis Comput*, 21(11): 977–1000