Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Robust feature matching via support-line voting and affine-invariant ratios

Jiayuan Li [a], Qingwu Hu [a,*], Mingyao Ai [a,b], Ruofei Zhong [c]

[a] *School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China*
[b] *State Key Laboratory of Information Engineering, Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China*
[c] *Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing 100048, China*

ARTICLE INFO

Article history:
Received 4 January 2017
Received in revised form 17 July 2017
Accepted 22 August 2017
Available online 5 September 2017

Keywords:
Robust feature matching
Support line voting
AB-SLT
Affine-invariant ratios
Local affine extension
Grid-wise affine model

ABSTRACT

Robust image matching is crucial for many applications of remote sensing and photogrammetry, such as image fusion, image registration, and change detection. In this paper, we propose a robust feature matching method based on support-line voting and affine-invariant ratios. We first use popular feature matching algorithms, such as SIFT, to obtain a set of initial matches. A support-line descriptor based on multiple adaptive binning gradient histograms is subsequently applied in the support-line voting stage to filter outliers. In addition, we use affine-invariant ratios computed by a two-line structure to refine the matching results and estimate the local affine transformation. The local affine model is more robust to distortions caused by elevation differences than the global affine transformation, especially for high-resolution remote sensing images and UAV images. Thus, the proposed method is suitable for both rigid and non-rigid image matching problems. Finally, we extract as many high-precision correspondences as possible based on the local affine extension and build a grid-wise affine model for remote sensing image registration. We compare the proposed method with six state-of-the-art algorithms on several data sets and show that our method significantly outperforms the other methods. The proposed method achieves 94.46% average precision on 15 challenging remote sensing image pairs, while the second-best method, RANSAC, only achieves 70.3%. In addition, the number of detected correct matches of the proposed method is approximately four times the number of initial SIFT matches.

1. Introduction

Image matching, which refers to establishing high-precision correspondences between two or more images with overlapping regions, is a fundamental issue in remote sensing and photogrammetry. Image matching is crucial in many applications, such as image registration (Zitova and Flusser, 2003), bundle adjustment (Triggs et al., 1999), panorama production (Weinmann et al., 2011), and 3D reconstruction (Haala and Kada, 2010). Feature matching is a very important tool for image matching and has been widely studied in the past several decades. Feature-based methods usually share the same framework: first, detect distinct feature points or feature lines; next, describe these features using their local photometric information, such as gradients or intensity values; third, calculate the matching scores between descriptor vectors and use the nearest-neighbor distance ratio (NNDR) (Lowe, 2004) technique to extract potentially good correspondences; finally, filter outliers via RANSAC (Fischler and Bolles, 1981) or graph matching (Conte et al., 2004). In this paper, we focus the on outlier filtering stage and use scale-invariant feature transform (SIFT) (Lowe, 2004) to obtain potentially good correspondences.

The robust feature matching methods can be roughly grouped into two classes:

(1) parametric methods. These methods usually use a parametric model, including rigid, affine, homography, or epipolar transformation, to represent the geometric relationship between an image pair. Matches that are inconsistent with the estimated model are eliminated as outliers. RANSAC is the most popular parametric outlier filtering technique. This method is based on the hypothesize-and-verify strategy. It alternates between two steps until convergence: first, randomly pick a minimum subset of correspondences to compute a specified geometric model; second, verify this

* Corresponding author at: School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China.
E-mail address: huqw@whu.edu.cn (Q. Hu).

http://dx.doi.org/10.1016/j.isprsjprs.2017.08.009
0924-2716/© 2017 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

estimated model by the remaining correspondences. If the number of the matches that support this candidate model is sufficiently large, the transformation can be considered as a reliable solution. RANSAC and its variants (Chum and Matas, 2005; Chum et al., 2003; Raguram et al., 2013; Torr and Zisserman, 2000), however, are not robust to local geometric distortions, such as distortions caused by large elevation differences. They are also not suitable for non-rigid images, such as fisheye images, which have been widely used in close-range photogrammetry. In addition, they also tend to degrade badly if the outlier ratio of the initial matches becomes large (Li and Hu, 2010; Ma et al., 2014).

(2) nonparametric methods. Nonparametric constraints are generally applied in non-rigid image registration and require smooth and slow motion. These constraints usually optimize a cost function. For example, graph matching organizes the features extracted from scene images as graphs and minimizes the structural distortions between graph networks via an energy function (Conte et al., 2004). These methods are suitable for both rigid and non-rigid image matching problems. However, the methods do not apply parametric geometric models as strict constraints; thus, relatively low-precision noisy matches may be difficult to separate from high-precision correct matches. Nonparametric methods may also perform badly if the outlier ratio becomes large.

To address these issues, we present a two-stage robust image matching algorithm based on support line voting and affine-invariant ratios, which considers both photometric and geometric constraints between an image pair. Our method is suitable for both rigid and non-rigid images. It is effective and robust even for cases with large outlier ratios. The support line is used as a photometric constraint in the first stage, which utilizes local region information to distinguish between inliers and outliers. Different from region matching methods, the scale and dominant direction of a support line are given. In addition, the location, shape, and size of the region of a support line are determined by the initial correspondences. This significantly reduces the computational complexity. In the second stage, we use affine-invariant ratios as a geometric constraint. Different from the first stage, this stage can be treated as a parametric strategy. The local affine transformation model can be estimated by the constraint of affine-invariant ratios. The local affine model is robust to local distortions. Thus, it is suitable for both satellite images and close-range photogrammetry images. We use this affine transformation to extract as many high-precision correspondences as possible. Our method also provides a more precise model called the grid-wise affine model for remote-sensing image registration. The primary idea of this paper is illustrated in Fig. 1.

There are three main contributions of our paper. First, we develop a region matching algorithm based on support-line voting. We present a support-line descriptor called adaptive binning support-line transform (AB-SLT). Multiple adaptive gradient histograms are adapted to improve the robustness to distortions. Second, we introduce affine-invariant ratios to refine the matching results. Local affine transformation performs better than global affine transformation, especially for high-resolution remote sensing images and non-rigid images. Third, we propose a grid-wise affine model for image registration.

2. Related work

Image matching and image registration have many applications in computer vision, robotic vision, pattern recognition, and especially in the areas of photogrammetry and remote sensing. Comparative reviews on image matching and registration methods can be found in (Dawn et al., 2010; Zitova and Flusser, 2003). In this section, we briefly review two categories of approaches in remote sensing and computer vision: putative correspondence generation methods and outlier filtering methods.

2.1. Putative correspondence generation

These methods provide initial correspondences, which typically include both inliers and outliers. There are two major categories: area-based methods and feature-based methods.

Area-based methods: Area-based methods only use image intensity information without detecting distinct structures such as corner features. Correlation-based methods usually measure the similarities between a pair of pattern windows (Egels and Kasser, 2003). These methods slide the pattern window in the target image and regard the pattern window with the highest similarity score as a correspondence of the pattern window in the reference image. The typical similarity metrics are sum of absolute differences (SAD), sum of squared differences (SSD), and normalized cross correlation (NCC). These methods are fast and have been applied in many real-time applications of remote sensing such as stereo matching. Their common drawback is that they are sensitive to such changes as rotation changes, scale changes, and viewpoint changes. Mutual information (MI) methods (Chen et al., 2003) often use subsets of or entire images to estimate

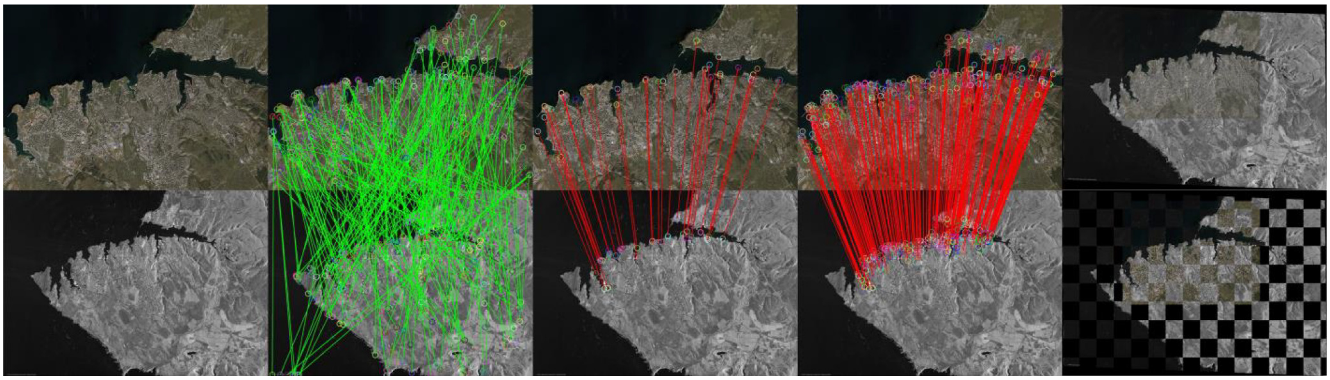

(a) original image pair (b) SIFT matching (c) support line voting (d) refinement and expansion (e) registration

Fig. 1. Schematic diagram of the proposed method. Given an image pair with an overlapping region, we first perform SIFT matching to generate initial matches. We use a support-line voting strategy as a photometric constraint to filter outliers. Then, we adopt local affine-invariant ratios as geometrical constraints to refine and expand the matching results. Finally, the image pair is registered by the established grid-wise affine model.

the transformation model. The MI metric is robust to spectral distortion. Thus, MI-based methods have large competitive superiority for multi-sensor remote sensing and medical image registration. The drawback is that the solutions of MI-based methods are local maximums. Fourier methods (Chen et al., 1994) register an image pair in the frequency domain. To account for the robustness to scale and rotation changes, variants called phase correlation methods are presented in (Foroosh et al., 2002). These approaches are robust to noise while sensitive to spectral distortions.

Feature-based methods: Features, including control points, lines, and salient regions, are more distinct than image intensity values. Generally, feature-based methods have three main stages: feature detection, feature description, and feature matching. SIFT is one of the most popular feature matching methods, which adopts Gaussian scale space and dominant orientation technique for scale, viewpoint, and rotation invariances and uses gradient histogram for illumination invariance. Speeded up robust features (SURF) (Bay et al., 2008) uses a Hessian matrix to detect feature points and introduces an integral image strategy to improve the efficiency. PCA-SIFT (Ke and Sukthankar, 2004) applies the principal component analysis (PCA) technique to reduce the dimension of SIFT descriptor. It reduces the computational cost of SIFT with little loss of robustness. Remote sensing images captured by different sensors, at different time, or from different viewpoints usually suffer from significant local spectral and geometric distortions. Many variants of SIFT and SURF that are more robust to multi-source and multimodal remote sensing images have been developed in the last several decades. Uniform robust SIFT (UR-SIFT) (Sedaghat et al., 2011) studies the distribution of SIFT keypoints and presents a feature selection strategy based on the local entropy distribution. SAR-SIFT (Dellinger et al., 2015) proposes a new gradient definition for SAR images to improve the robustness to speckle noise. Adaptive binning SIFT (AB-SIFT) (Sedaghat and Ebadi, 2015) introduces an adaptive binning histogram strategy to describe feature points. It is robust to local spectral and geometric distortions, which makes it suitable for multisource images. To improve the distinctiveness of SIFT, the AB-SIFT descriptor divides a local circular region into several radial sectors (circular log-polar grids). Log-polar grids have different histogram bins, which assign different weights to pixels according to the distances to the region center. Thus, AB-SIFT is more robust to radial distortions than SIFT. Due to the superiority of the adaptive binning strategy, we develop a support-line descriptor based on multiple adaptive binning histograms.

2.2. Outlier filtering

False correspondence elimination is an important stage of image matching and image registration. As mentioned earlier, outlier filtering methods can be roughly divided into two categories: parametric methods and nonparametric methods.

Parametric methods: RANSAC and its variants, which are based on a hypothesize-and-verify framework, are widely used for false correspondence elimination. MLESAC (Torr and Zisserman, 2000) is a probability-based method, which maximizes the likelihood, while RANSAC maximizes the number of inliers, and it is a robust generalization of RANSAC. PROSAC (Chum and Matas, 2005) improves the first stage of RANSAC. It uses local similarity ordering instead of uniform sampling to draw the minimal subset of correspondences. Locally optimized RANSAC (LO-RANSAC) (Chum et al., 2003) enhances RANSAC through the addition of a local optimization stage, which significantly decreases the number of samples drawn. As a result, PROSAC and LO-RANSAC significantly reduce the computational complexity of RANSAC. USAC (Raguram et al., 2013) is a universal framework for RANSAC-like robust feature matching. It extends the basic RANSAC to incorporate a number of important practical and computational considerations. Hough transform (HT) is another popular tool for robust feature matching. Its core idea is to discretize the geometric parameter space into many bins and accumulate the votes given by each correspondence in the bins (Chin and Suter, 2017). Tolias and Avrithis (Tolias and Avrithis, 2011) propose a variant of HT called Hough pyramid matching (HPM) for multi-object matching. The researchers use a relaxed pyramid matching model to rank the correspondences. Chen et al. (Chen et al., 2013) cast the feature matching problem as a density estimation problem. The method alternates between HT and inverted HT. They use HT to check the correctness of each correspondence and adapt inverted HT to enrich the number of correct correspondences. The major difficulties of HT and its variants lie in discretizing the parameter space. More recently, researchers have proposed some direct methods. Different from the hypothesize-and-verify technique, these methods directly estimate the geometric transformation from initial correspondences contaminated by outliers or noises in one step. Locally linear transforming (LLT) (Ma et al., 2015b) adopts a local geometrical constraint and formulates the outlier filtering problem as a maximum-likelihood estimation of a Bayesian model. Expectation maximization (EM) (Dempster et al., 1977) is then applied to solve this problem. LLT can handle both rigid and non-rigid images. However, it only uses geometric constraints for outlier elimination. Correspondences with relatively low precision may be accepted as inliers. Li et al. (2016) propose an effective, efficient, and robust feature matching method via an l_q-estimator. A new cost function based affine transformation and the l_q-norm is presented. This method is extremely fast compared with RANSAC family. The limitation is that it is not suitable for non-rigid image scenes.

Nonparametric methods: Nonparametric methods are usually developed for both rigid and non-rigid image matching and have been adopted in many computer vision applications. Torresani et al. (2008) describe a novel optimization technique called dual decomposition for graph matching, in which they define a complex cost function based on the spatial arrangement, texture similarity, and geometric consistency of the keypoints. Cho et al. (2014) introduce a max-pooling strategy for graph matching. In their method, candidate matches are scored by their most promising neighbors, and the scores are then used to update the neighbors. Progressive graph matching (PGM) (Cho and Lee, 2012) is a move-making approach for graph matching, which alternately performs graph probabilistic progression and graph matching steps. In the graph matching step, the reweighted random walk (RRWM) (Cho et al., 2010), integer projected fixed-point (IPFP) (Leordeanu et al., 2009), tensor matching (TM) (Duchenne et al., 2011), or minimum spanning tree induced triangulation (MSTT) (Lian et al., 2012) scheme can be applied. In another work of Cho et al. (2009), they formulate image matching as a clustering problem, and present a novel linkage model and a new dissimilarity metric in the framework of hierarchical agglomerative clustering (ACC). Lian et al. (2016) reduce the robust point matching (RPM) problem to a concave quadratic assignment problem by eliminating the transformation variables and propose a globally optimal branch-and-bound approach based on rectangular subdivision. Jian and Vemuri (2011) use Gaussian mixture models to represent the two initial point sets, and align the two Gaussian mixtures by minimizing a statistical discrepancy measure. Vector field consensus (VFC) (Ma et al., 2014) estimates a consensus of the correct matches based on a vector field. They use the EM algorithm to compute a maximum a posteriori probability of a Bayesian model. The robust L2E estimator (Ma et al., 2015a) is proposed for RPM problems, which assumes that the noise on the inliers obeys a Gaussian distribution with zero mean. As close-range photogrammetry has become increasingly widespread, nonparametric methods have begun to play an important role in photogrammetry and remote sensing applications. Zhao et al. (2013) propose a graph matching method based on bilateral k-nearest neighbors spatial orders around geometric centers. The adjacent relation information and spatial arrangement of feature points are considered. Zhou et al. (2016) propose a probabilistic method for remote sensing feature matching. These researchers use Tikhonov regularizers in kernel Hilbert space to impose nonparametric global geometrical constraints. As mentioned earlier, these methods do not apply parametric geometric models as strict constraints; thus, relatively low-precision noisy matches may be difficult to separate from high-precision correct matches.

In this paper, we propose a novel and robust feature matching method via support-line voting and affine-invariant ratios. Both photometric and geometric constraints are considered in our framework. Thus, the proposed method is robust to spectral and geometric distortions. In addition, the method is suitable for both rigid and non-rigid image matching and image registration problems.

3. Proposed robust feature matching method

This section details the proposed feature matching method for remote sensing. We first define the concept of a support line and develop a support-line descriptor called AB-SLT. Next, we introduce affine-invariant ratios to refine and expand the matching results. In the refinement and expansion stage, we also build a grid-wise affine model for remote sensing image registration.

3.1. Support-line voting

Support line: We first extract the initial correspondence set (P, Q) by applying the SIFT algorithm to an image pair (I_1, I_2). For each feature point $p_i \in P$, its neighbors are found in a circular region with radius r' centered at p_i (see Fig. 2). The support lines are the straight lines that link p_i with its neighbors. As feature point sets P and Q are correspondence, the support lines of q_i (the correspondence of p_i) are also formed once the support lines of p_i are constructed. The support-line pair $(l_{p_ip_j}, l_{q_iq_j})$ is a correspondence only if both (p_i, q_i) and (p_j, q_j) are correct correspondences. In other words, if the support-line pair $(l_{p_ip_j}, l_{q_iq_j})$ is a correspondence, (p_i, q_i) and (p_j, q_j) are more likely to be correct correspondences. Based on this observation, we present a support-line voting strategy for outlier elimination. A support-line descriptor called adaptive binning support-line transform (AB-SLT) is developed. The correspondence (p_i, q_i) gets a vote if a support line is an inlier match.

AB-SLT: SIFT descriptor divides the local feature area into subregions of equal size and describes a feature based on uniform binning histograms. In contrast, AB-SIFT utilizes an improved binning histogram strategy, in which the location grids are incremented while the number of orientation bins is decreased along radial direction. The grids are of different sizes in different radial rings. The main advantage of the adaptive binning histogram technique is its robustness to local geometric distortions. The remote sensing images captured by different sensors, at different time, or from different positions suffer from serious spectral and geometric distortions, especially for wide baseline images with large viewpoint changes. The gradient histogram technique can reduce the effect of spectral distortions. Generally, the local geometric distortion is radially distributed, i.e., the distortions increase with the distance from the centers of the local regions. Thus, the adaptive binning histogram technique is designed to give less emphasis on pixels that are far from the features.

The adaptive binning descriptor is constructed as follows: First, the circular local region of a feature is divided into n non-overlapping rings $R = \{r_1, r_2, \ldots, r_n\}$ along the radial direction. Instead of regular division, an adaptive angular quantization strategy is then applied to these rings. The angular quantization numbers are $M = \{m_1, m_2, \ldots, m_n\}$. Each ring r_i is separated into m_i grids of the same size. Finally, a gradient histogram for each grid is computed. The gradient histogram quantization numbers are $K = \{k_1, k_2, \ldots, k_n\}$. The histogram bin number of grids inside the ring r_i is k_i (see Fig. 3). The histograms of all grids are then concatenated in order to form the final descriptor vector.

We develop a support-line descriptor called AB-SLT based on the adaptive binning histogram technique. From the definition of the support line, we can infer the following: (1) The dominant orientation of a support line is its line direction; (2) The scale of a support line is related to its length. The support-line descriptor has two inherent properties: rotation and scale invariance (see Fig. 4 for more details). Thus, there are no scale-space construction and dominant orientation computing stages in the proposed descriptor. As a result, the computational complexity of AB-SLT is very low compared with traditional gradient-based descriptors. Fig. 3

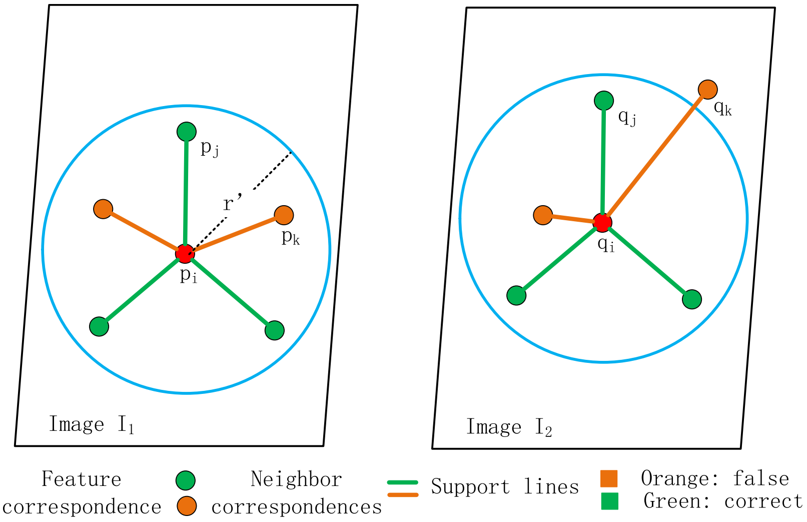


Fig. 2. Illustration of support-line correspondence. The support lines of p_i are the straight lines that link p_i with its neighbors (both orange lines and green lines in the left plot). The green line pairs such as $(l_{p_ip_j}, l_{q_iq_j})$ are correct support-line correspondences; and the orange line pairs such as $(l_{p_ip_k}, l_{q_iq_k})$ are outliers.

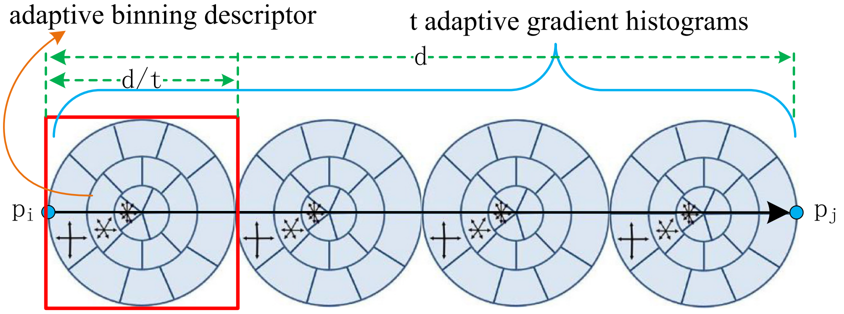


Fig. 3. Illustration of the AB-SLT descriptor. The dominant orientation of adaptive binning descriptor is the support-line direction and the radius of the circular description region is d/t. The adaptive binning descriptor divides the circular region into n rings $R = \{r_1, r_2, \ldots, r_n\}$ along the radial direction, each ring r_i is segmented into m_i grids, and each grid in r_i is described by a histogram with k_i bins. AB-SLT is finally constructed by t adaptive binning descriptors.

illustrates the proposed AB-SLT descriptor. In detail, we first compute the length d of a support line $l_{p_ip_j}$ and divide $l_{p_ip_j}$ into t sublines $SL_{ij} = \{sl_1, sl_2, \ldots, sl_t\}$ of equal length $len = d/t$. Each subline segment sl_i has a circular local support region of radius $len/2$ centered at the midpoint of sl_i for descriptor computation. Using multiple sublines instead of the support line can improve the robustness to local geometric distortions. If we directly use the local region with a radius of $d/2$ centered at the midpoint of $l_{p_ip_j}$, many pixels far from the line will suffer from significant distortions, which will decrease the distinctiveness of the descriptor. Then, for each subline sl_i, an adaptive binning descriptor $D(ls_i)$ with scale $len/2$ and dominant orientation $\vec{l}_{p_ip_j}$ (line direction) is computed. The final AB-SLT descriptor $D(l_{p_ip_j})$ is the concatenation of these multiple subline adaptive binning descriptors,

$$D(l_{p_ip_j}) = \{D(ls_1), D(ls_2), \ldots, D(ls_t)\} \tag{1}$$

Table 1 summarizes the parameter default values of the descriptor, where the radial quantization number, angular grid set, and histogram bin set follow the suggestion of AB-SIFT. The dimension of the proposed descriptor is 1024. The radius of the adaptive binning descriptor is at least 3 pixels because the radial quantization number n is set to 3. Thus, the length of the support line must be longer than $d = t \times 2 \times (len/2) = 8 \times 2 \times 3 = 48$ pixels. Support lines with short lengths are discarded as unreliable.

Different from traditional descriptors, which adopt the nearest-neighbor distance ratio to match descriptors, the proposed method is just to compare the descriptor vectors of a support-line pair. The correspondence relationship of the support lines is already one to one, which avoids the stages with high computational complexity, i.e., neighbor searching and cross matching. We use the Euclidean distance as the similarity metric. A support-line pair $(l_{p_ip_j}, l_{q_iq_j})$ is an inlier if the distance of them is below a certain threshold τ.

The proposed support-line voting strategy works similar to the local region matching methods. The strategy is not sensitive to local geometric distortions because of the carefully designed descriptor. In addition, with the support of local regions, it is more robust to locally repetitive textures. The inliers can be distinguished from outliers by the voting scores, i.e., matches that get more votes are more likely to be inliers. In our experiments, the inliers are the matches whose numbers of votes are larger than η.

3.2. Affine-invariant ratios

Affine-invariant ratios: An affine transformation $T(\cdot)$ is generally represented by,

$$\boldsymbol{y} = T(x) = \boldsymbol{Ax} + \boldsymbol{t} \tag{2}$$

where $\boldsymbol{x}$ are the observations; $\boldsymbol{y}$ are the observations after affine transformation; $\boldsymbol{t}$ is a translation vector; and $\boldsymbol{A}$ is a 2×2 non-singular affine matrix.

To better understand the geometric effects, the affine matrix can be decomposed into a rotation term $\boldsymbol{R}(\theta)$ and a deformation term $\boldsymbol{R}(-\phi)\boldsymbol{DR}(\phi)$ (Hartley and Zisserman, 2003),

$$\boldsymbol{A} = \boldsymbol{R}(\theta)\boldsymbol{R}(-\phi)\boldsymbol{DR}(\phi) \tag{3}$$

where $\boldsymbol{D}$ is a 2×2 diagonal matrix formed by scale parameters s_1 and s_2,

$$\boldsymbol{D} = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} \tag{4}$$

There is an important invariant under affine transformations: the ratio of areas. Areas are only scaled by $s_1 \cdot s_2$ because translations and rotations have no influence on areas. $s_1 \cdot s_2$ is equal to $\det(\boldsymbol{A})$ and the ratio of areas is the same after transformation. Based on this property, we can easily derive the affine-invariant ratios. Fig. 5 gives the details.

Given four feature points A_1, B_1, C_1, D_1, we transform them to A_2, B_2, C_2, D_2 by an affine transformation. Let O_1 be the intersection point of line segments A_1C_1 and B_1D_1, and O_2 the intersection point of line segments A_2C_2 and B_2D_2. Based on the invariant, i.e., the ratio of areas, we have,

$$\begin{cases} \dfrac{Area(\Delta S_{A_1B_1O_1})}{Area(\Delta S_{A_1B_1C_1})} = \dfrac{Area(\Delta S_{A_2B_2O_2})}{Area(\Delta S_{A_2B_2C_2})} \\ \dfrac{Area(\Delta S_{A_1B_1O_1})}{Area(\Delta S_{A_1B_1D_1})} = \dfrac{Area(\Delta S_{A_2B_2O_2})}{Area(\Delta S_{A_2B_2D_2})} \end{cases} \tag{5}$$

where ΔS denotes a triangle. Because $\Delta S_{A_1B_1O_1}$ and $\Delta S_{A_1B_1C_1}$ share the same height, we obtain,

$$\begin{cases} \left(\alpha_1 = \dfrac{|A_1O_1|}{|A_1C_1|}\right) = \left(\alpha_1' = \dfrac{|A_2O_2|}{|A_2C_2|}\right) \\ \left(\alpha_2 = \dfrac{|B_1O_1|}{|B_1D_1|}\right) = \left(\alpha_2' = \dfrac{|B_2O_2|}{|B_2D_2|}\right) \end{cases} \tag{6}$$

where $|A_1O_1|$ represents the length of line segment A_1O_1, and (α_1, α_1') and (α_2, α_2') are the affine-invariant ratios.

For rigid remote sensing images, such as satellite and aerial images, the ranges of elevations are very small compared with the flight altitudes of the cameras. Affine transformations are widely used for rigid image registration. Although the distortions of non-rigid images such as close-range oblique UAV images and fisheye images are serious, the relationship between a small local region pair can be still well modeled by a local affine transformation. Generally, for both rigid and non-rigid remote sensing images, the affine-invariant ratios can be satisfied in a small local region.

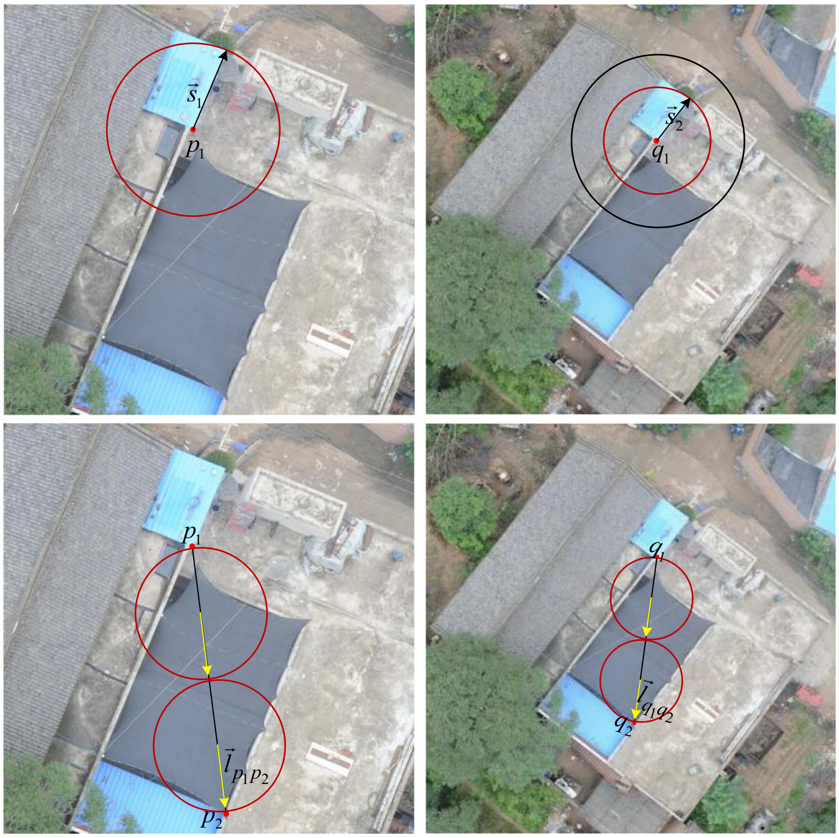


Fig. 4. First row: the scale and dominant orientation of popular histogram-based descriptors such as SIFT. p_1 and q_1 are a pair of matching points; $\vec{s}_1$ and $\vec{s}_2$ are the dominant orientation of p_1 and q_1, respectively; the scales of p_1 and q_1 are the radii of the local patches (red circle regions) centered at p_1 and q_1, respectively. These descriptors must perform the scale-space construction and dominant orientation computing stages to determine the scale and dominant orientation of each feature point. Otherwise, suppose we do not perform scale-space construction and use the same scale for p_1 and q_1. The local patch of q_1 becomes the black circle region. As a result, the description of q_1 becomes unreliable. Second row: the scale and dominant orientation of AB-SLT. $l_{p_1p_2}$ and $l_{q_1q_2}$ are a pair of matching support lines; the line directions $\vec{l}_{p_1p_2}$ and $\vec{l}_{q_1q_2}$ are the dominant orientations of $l_{p_1p_2}$ and $l_{q_1q_2}$, respectively; the scales of $l_{p_1p_2}$ and $l_{q_1q_2}$ are determined by the lengths of $l_{p_1p_2}$ and $l_{q_1q_2}$, respectively. The local patches of $l_{p_1p_2}$ and $l_{q_1q_2}$ are correctly corresponding regions even if there are scale and rotation changes between the image pair. Thus, the support-line descriptor has two inherent properties: rotation and scale invariance.

Table 1
Parameter default values of AB-SLT.

Parameter	Notation	Default value
Radial quantization number	n	3
Angular grid set	M	$M = \{5, 8, 10\}$
Histogram bin set	K	$K = \{8, 6, 4\}$
Number of subline segments	t	8
Descriptor dimension	Dim	$(5 \times 8 + 8 \times 6 + 10 \times 4) \times 8 = 1024$

Refinement and expansion: In the support-line voting stage, most of the outliers can be eliminated. However, it is only based on photometric constraints that the inliers are not quantitatively evaluated. A portion of noisy matches with relatively low precision may be still preserved. To refine the results of support-line voting and provide the registration transformation, a novel method based on affine-invariant ratios is presented. We use the affine-invariant ratios (α_1, α_1') and (α_2, α_2') to establish a basis for local affine transformation estimation in a local region. The affine transformation has six degrees of freedom, i.e., four in the affine matrix and two in the translation vector; it needs three correct non-collinear correspondences to provide a closed-form solution. The basis is formed by a two-line structure, which can provide four inliers for affine transformation estimation. The redundant observations by the basis can improve the robustness to noise compared with closed-form solutions.

The distribution of correspondences obtained by support-line voting may be very uneven. To improve the distribution and place more emphasis on matches with high voting scores, we lay out a grid network with 50 × 50 grids on image I_1 (see Fig. 6). We preserve at most one correspondence with the highest voting score for a grid patch. This step can also largely reduce the computational complexity of the subsequent refinement and expansion stage. We use $(\bar{P}, \bar{Q})$ to denote the preserved correspondence set. For each preserved feature point $\bar{p}_i \in \bar{P}$, we search its neighbors $N(\bar{p}_i)$ in a local circular region with radius r'. We then select four feature points inside $N(\bar{p}_i)$ to construct a two-line basis $A_1C_1 - B_1D_1$ (see Fig. 5). There are three principles for the selection of these points: (1) The affine-invariant ratios should be satisfied, i.e., $|\alpha_1 - \alpha_1'| < \delta$ and $|\alpha_2 - \alpha_2'| < \delta$, where δ is a small value. If these constraints are not satisfied, at least one of the selected points is a false match. (2) The lengths of the two lines A_1C_1 and B_1D_1 should be large. (3) The intersection angle between A_1C_1 and B_1D_1 should be large. The second and third principles ensure that the selected points are far from one another such that the constructed basis represents the local region better. Suppose that the basis is formed by four nearby feature points. The estimated local affine transformation may well model the relationship between the envelope areas of the bases in images I_1 and I_2. However, if the distortion of the image pair is strong, the transformation will be not suitable for the area outside the envelope but inside the local region. In practice, we first link every two feature points inside $N(\bar{p}_i)$ to form line segment set $L(\bar{p}_i)$ and calculate the length of each line segment

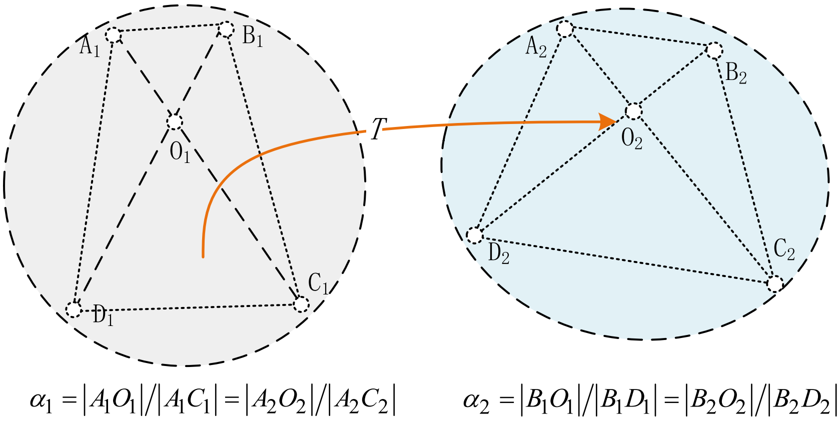

A1
B1
O1
C1
D1
T
A2
B2
O2
C2
D2
α1 = |A1O1|/|A1C1| = |A2O2|/|A2C2|
α2 = |B1O1|/|B1D1| = |B2O2|/|B2D2|

Fig. 5. Affine-invariant ratios. A_2, B_2, C_2, D_2 are the corresponding locations of A_1, B_1, C_1, D_1 after affine transformation T_{local}. O_1 is the intersection of line segments A_1C_1 and B_1D_1; O_2 is the intersection of line segments A_2C_2 and B_2D_2. The ratios $\alpha_1 = |A_1O_1|/|A_1C_1|$ and $\alpha_2 = |B_1O_1|/|B_1D_1|$ are invariant under affine transformation, i.e., $\alpha_1 = |A_1O_1|/|A_1C_1| = |A_2O_2|/|A_2C_2|$ and $\alpha_2 = |B_1O_1|/|B_1D_1| = |B_2O_2|/|B_2D_2|$.

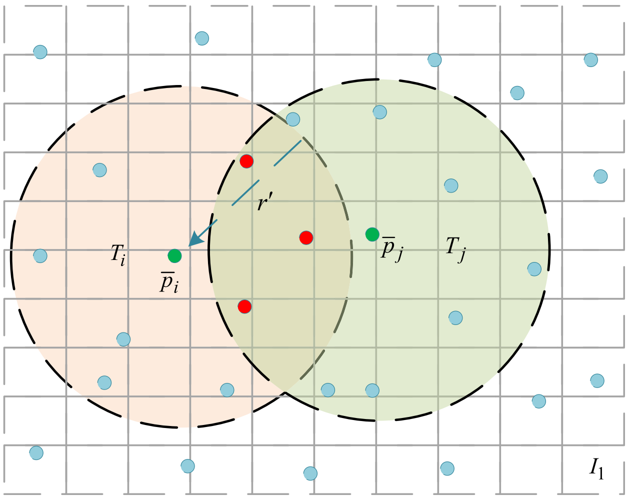


Fig. 6. Correspondence refinement and expansion. We lay out a grid network with 50 × 50 grids on image I_1. We preserve at most one correspondence (blue and red dots in the figure) with the highest voting score for a grid patch. For each preserved feature point $\bar{p}_i$, we search its neighbors $N(\bar{p}_i)$ in a local circular region with radius r'. We then calculate a local affine transformation T_i based on affine-invariant ratios for each local region. Each correspondence will be assigned an identifier, i.e., inlier or outlier, and each grid will be given a local affine transformation.

inside $L(\bar{p}_i)$. Meanwhile, the correspondence of each line segment inside $L(\bar{p}_i)$ is also established because of the one-to-one relationship of $(\bar{P}, \bar{Q})$. Next, we select the ten longest line segment correspondences and compute the affine-invariant ratios for each pair of line segment correspondences. The two-line bases that do not satisfy the first principle are discarded. Finally, we calculate the intersection angle of each two-line basis and select the two-line basis with the largest intersection angle as a reliable one.

Once the basis correspondence is established, a local affine transformation $T_{local}(\cdot)$ can be estimated by linear least squares. The local affine transformation has three main functions: (1) It provides a quantitative evaluation of the correspondences inside the local region. Each correspondence is classified as an inlier or outlier according to its residual error v after transformation, i.e., $inliers = \{v | v < \varepsilon\}$. (2) It extracts as many high-precision matches as possible. In this paper, we use SIFT to obtain the initial correspondence set. However, there are still many inliers that are not matched with SIFT. We perform an expansion stage for the remaining SIFT keypoints (feature points unmatched with SIFT) to extract as many high-precision matches as possible. In detail, for each keypoint kp_i inside the local region, we use $T_{local}(\cdot)$ to predict the ideal position kq_i' of its correspondence kq_i. We then find the keypoint set KQ in a small circular region with radius ε centered at kq_i'. The similarity score between keypoint $kq_j \in KQ$ and feature kp_i is computed as follows,

$$Dist(kp_i, kq_j) = e^{-\varepsilon/d_{ij}} \cdot \| D_{SIFT}(kp_i) - D_{SIFT}(kq_j) \| \tag{7}$$

where d_{ij} is the Euclidean distance between keypoint kq_j and the ideal position kq_i', $D_{SIFT}(kp_i)$ represents the SIFT descriptor of feature kp_i, and $\| \cdot \|$ is the norm operator. In this similarity metric, both geometric and photometric information is considered. The first term $e^{-\varepsilon/d_{ij}}$ is a geometric constraint whose role is to give more emphasis on the candidate matches that are close to the ideal position kq_i'. The second term $\| D_{SIFT}(kp_i) - D_{SIFT}(kq_j) \|$ is a photometric difference. If the lowest distance is below τ, the feature with the lowest distance is accepted as the true correspondence of kp_i. (3) It assigns a transformation for each grid inside the local region (see Fig. 6).

For efficiency, we do not perform local affine-invariant basis establishment for neighbors $N(\bar{p}_i)$ that have been assigned identifiers (inliers or outliers). The radius r' is set to be much larger than the grid size. Thus, we can search for sufficiently many correspondences for affine-invariant ratios computation. With large radius r', each correspondence $(\bar{p}_i, \bar{q}_i) \in (\bar{P}, \bar{Q})$ will be assigned more than one identifier. In fact, the identifiers of a true correspondence should always be inliers. In other words, if a correspondence is classified as an inlier in one local region and an outlier in another one, the correspondence will be treated as an unreliable correspondence, which should be further investigated.

3.3. Grid-wise affine registration

Typically, given a reliable feature correspondence set $(\bar{P} = \{\bar{p}_i\}_1^N,\ \bar{Q} = \{\bar{q}_i\}_1^N)$ of an image pair $(I_1,\ I_2)$, the image pair can be registered by the global affine model,

$$\tilde{\boldsymbol{q}}_i = \boldsymbol{H}\tilde{\boldsymbol{p}}_i \tag{8}$$

where $\tilde{\boldsymbol{q}}_i$ and $\tilde{\boldsymbol{p}}_i$ are the homogeneous coordinates of $\bar{p}_i$ and $\bar{q}_i$, respectively, and $\boldsymbol{H} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{t} \\ \boldsymbol{0}_{1\times 2} & 1 \end{bmatrix}$ is a 3 × 3 affine transformation matrix. The registration problem can be efficiently solved by least squares (LS). However, if the image scene is not a planar scene, using a global affine warp inevitably yields misalignment. Motivated by Zaragoza et al. (2013), we first use a weighted global affine transformation to alleviate the problem. We assign each grid G_{ij} inside I_1 a location-dependent affine transformation $\boldsymbol{H}_{ij}$,

$$\tilde{\boldsymbol{y}}_{*_{ij}} = \boldsymbol{H}_{ij}\tilde{\boldsymbol{x}}_{*_{ij}} \tag{9}$$

where $\tilde{\boldsymbol{x}}_{*_{ij}}$ is the homogeneous coordinates of the center pixel of G_{ij} and $\tilde{\boldsymbol{y}}_{*_{ij}}$ is the homogeneous coordinates of the pixel corresponding to $\tilde{\boldsymbol{x}}_{*_{ij}}$ in image I_2. $\boldsymbol{H}_{ij}$ can be estimated from the weighted least squares (WLS) problem,

$$\boldsymbol{H}_{ij} = \arg\min_{\boldsymbol{H}} \sum_{i=1}^{N} \|w_*^i(\boldsymbol{H}\tilde{\boldsymbol{p}}_i - \tilde{\boldsymbol{q}}_i)\|^2 \tag{10}$$

where $\{w_*^i\}_1^N$ is a weight set that places more emphasis on correspondences that are closer to $\tilde{\boldsymbol{x}}_{*_{ij}}$, and the weights are calculated as,

$$w_*^i = \exp(-\|\tilde{\boldsymbol{x}}_{*_{ij}} - \tilde{\boldsymbol{p}}_i\|^2/\sigma^2) \tag{11}$$

where σ is a scale parameter, which is set to 8.5 as in the literature. Intuitively, the projective warp $\boldsymbol{H}_{ij}$ better respects the local structure in G_{ij} than global $\boldsymbol{H}$ because Eq. (11) assigns larger weights to correspondences closer to $\tilde{\boldsymbol{x}}_{*_{ij}}$. However, $\boldsymbol{H}_{ij}$ is solved by all correspondences, and it is still not very accurate when local geometric distortion cannot be ignored. Fortunately, we have assigned a local affine transformation to grids inside some local regions in the last section. Thus, the grids inside some local regions are assigned the same local affine transformations as the local regions, while the grids that are not overlapped by any local regions of the constructed bases are assigned a weighed global affine transformation by Eq. (10). The image pair is then registered based on the developed grid-wise affine model, which is more robust to local geometric distortions.

4. Experimental results and discussions

In this study, we evaluate the performance of the proposed method on both rigid and non-rigid remote-sensing image datasets. We compare our approach with six other state-of-the-art robust feature matching methods, i.e., VFC, LLT, RANSAC, USAC, PGM + RRWM, and ACC. The parameters are set according to their literature's suggestion and fixed throughout all experiments (Table 2). The implementations of these algorithms are obtained from the authors' websites (Table 2). The dataset and the demo software of the proposed method are publicly available.[1]

4.1. Data set

Three categories of remote sensing image pairs, including both rigid and non-rigid datasets, are used to evaluate the proposed method. The first two categories are used for rigid image matching evaluation and the last one is used for non-rigid evaluation.

Data set 1: This dataset consists of 15 image pairs formed by different types of multi-sensor and multi-temporal satellite or aerial images. The details of these image pairs are summarized in Table 3. In this dataset, the ground sample distance (GSD) ranges from 0.5 m to 30 m, including high-, medium-, and low-resolution remote sensing images. These image pairs suffer from serious geometric distortions, photometric distortions, and extremely small overlapping regions. For example, multi-temporal image pairs suffer from high temporal changes; the spectral information is significantly different between multi-sensor and multi-band image pairs; and the overlapping region of image pair 14 is less than 5% of the image size.

Table 2
Parameter settings for compared methods.

Methods	Parameter setting	Source code
VFC	$\beta = 0.1$; $\lambda = 3$; $\tau = 0.75$; $\gamma = 0.9$; $a = 10$.	https://sites.google.com/site/jiayima2013/
LLT	$K = 15$; $\lambda = 1000$; $\tau = 0.5$; $\gamma = 0.9$; $\beta = 0.1$; $M = 15$; $a = 10$.	https://sites.google.com/site/jiayima2013/
RANSAC	$m = 3$; $t = 3$; $\eta_0 = 0.99$; $\lambda = 3$; $K = 50000$	http://www.peterkovesi.com/matlabfns/index.html#robust
USAC	$m = 3$; $t = 3$; $\eta_0 = 0.99$; $\lambda = 3$; $K = 50000$; $\delta = 0.01$; $\varepsilon = 0.2$.	http://www.cs.unc.edu/~rraguram/usac/
PGM	$\alpha = 50$; $k_1 = 25$; $k_2 = 5$.	http://cv.snu.ac.kr/research/~ProgGM/
RRWM	$\alpha = 0.2$; $\beta = 30$.	http://cv.snu.ac.kr/research/~RRWM/
ACC	$K_{AP} = 10$; $r_{AP} = 0.05$; $\delta_D = 25$; $\tau_a = 1\%$; $\tau_m = 1$.	http://cv.snu.ac.kr/research/~acc/

Note: the parameter symbols of each method listed in the table are the same as in the corresponding paper, and the meaning of each parameter can be found in the corresponding paper.

Data set 2: This dataset is simulated by the RGB channels and the infrared channel of a Worldview 2 image that was taken over Guangzhou, China. Specifically, we use the RGB channels as the reference image and perform affine or rotational transformation on the infrared channel to produce the target images. The first six image pairs suffer affine transformations, including rotation, translation, and non-isotropic scaling, and image pairs 7 and 8 only suffer 15 °and 75° rotations, respectively. Fig. 7 shows the reference and target images of data set 2.

Data set 3: There are four image pairs in this dataset. The first image pair is cropped from two 360° × 180° panoramic photos captured by an unmanned aerial vehicle in Wuhan, China, 2016 (see Fig. 13(a)). The second pair is formed by two fisheye images captured by an unmanned aerial vehicle in Baoding, China, 2013. These two image pairs are used for non-rigid image matching evaluation. The third pair consists of a GoPro Hero 2 wide-angle image and a Canon EOS 5D Mark II image taken over Nova Scotia, Canada (see Fig. 15(a)). The last pair is formed by two Nikon D800 images with a GSD of approximately 0.1 m, which were taken over Henan, China by an unmanned aerial vehicle (see Fig. 16(a)). These two image pairs are used for non-rigid image registration evaluation.

We establish a ground-truth transformation for each image pair in data set 1 and data set 2. For each image pair inside data set 1, five evenly distributed correspondences with sub-pixel accuracy are manually selected and an accurate affine transformation is then estimated by linear least squares. Data set 2 is a simulated dataset for which the ground truth can be easily computed with the known geometric transformation. Then, the established ground-truth transformation is used to determine the inliers. Matches with residual errors larger than a certain threshold $\lambda = 3$ are regarded as outliers. Five metrics are used for evaluation on data set 1 and data set 2: precision, number of inliers, root-mean-square deviation (RMSE), mean error, and maximum error. Precision is the percentage of correct matches out of all detected matches. The number of inliers is the product of the precision and the total number of detected matches. RMSE is calculated by the residuals of all detected matches, mean error is the average of the absolute residuals of all detected matches, and maximum error is the maximum residual among all detected matches. RMSE, mean error, and maximum error measure the location accuracy of correspondences. For the deformable dataset, we randomly select 100 detected correspondences of each method on each image pair. Then, we manually find the corresponding matching points for the 100 keypoints in the first image of each image pair. These manually

[1] https://sites.google.com/site/jiayuanli2016whu/home

Table 3
Input image pairs of Data set 1.

No.	Image pair	Spectral mode	Image size	GSD (m)	Acquisition date	Location	Description
1	World View 2	Pan	405 × 350	0.5	2011	USA-	Multi-temporal
	World View 2	Pan	405 × 350	0.5	2014	California	
2	TM	Band 5	512 × 512	30	1992	Brazil-	Multi-temporal
	TM	Band 5	512 × 512	30	1994	Amazon	
3	JERS-1	Radar	256 × 256	18	1995	Brazil-	Multi-temporal
	JERS-1	Radar	256 × 256	18	1996	Amazon	
4	TM	Band 5	512 × 512	30	1990	USA-	Multi-temporal
	TM	Band 5	512 × 512	30	1994	Iowa	
5	SPOT 2	Band 3	256 × 256	20	1995	Brazil-	Multi-temporal
	TM	Band 4	256 × 256	30	1994	Brasilia	Multi-sensors
6	Pleiades-1A	Pan-sharpened	2000 × 1400	0.5	2014	Ukraine-	Multi-sensors
	TerraSAR-X	Radar	2000 × 1400	1	2014	Sevastopol	
7	Pleiades-1A	Pan-sharpened	800 × 800	0.5	2014	Ukraine-	Multi-sensors
	TerraSAR-X	Radar	800 × 800	1	2014	Sevastopol	
8	SPOT 5	Pan-sharpened	800 × 800	2.5	2002	China-	Multi-temporal
	SPOT 6	Pan-sharpened	800 × 800	1.5	2012	Beijing	Multi-sensors
9	SPOT 5	Pan-sharpened	800 × 800	2.5	2002	China-	Multi-temporal
	SPOT 6	Pan-sharpened	800 × 800	1.5	2012	Beijing	Multi-sensors
10	SPOT 5	Pan-sharpened	800 × 800	2.5	2003	France-	Multi-temporal
	SPOT 7	Pan-sharpened	800 × 800	1.5	2014	Paris	Multi-sensors
11	SPOT 5	Pan	1000 × 1000	2.5	2008	China-	Multi-temporal
	SPOT 5	Pan	1000 × 1000	2.5	2012	Shanghai	
12	World View 2	Pan-sharpened	1200 × 1200	0.5	2012	China-	Multi-temporal
	ZY-3	Pan	1200 × 1200	2.1	2013	Hong Kong	Multi-sensors
13	TM	Band 1	1450 × 1480	30	2000	Unknown	Multi-bands
	TM	Band 4	1450 × 1480	30	2000		
14	Aerial	Color-infrared	1400 × 1375	0.5	2011	USA-	Small overlaps
	Aerial	Color-infrared	1400 × 1375	0.5		Illinois	
15	Radarsat-2	Radar	800 × 800	3	2013	China-	Multi-sensors
	Airborne SAR	Radar	800 × 800	3	2013	Jiangsu	

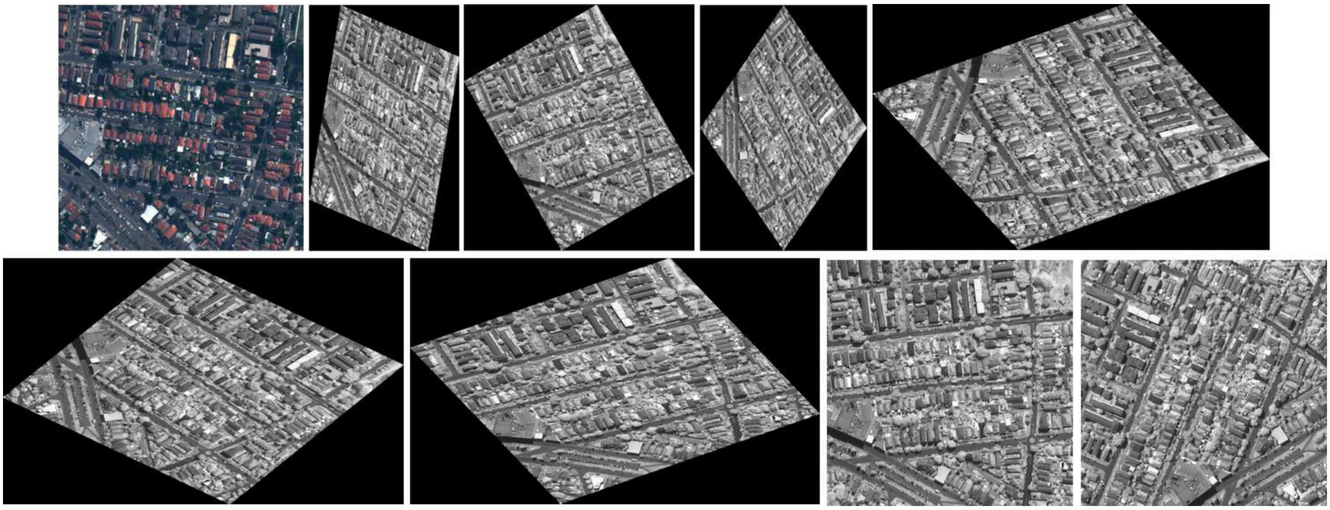


Fig. 7. Data set 2. The first RGB image is the reference image and the last eight infrared images are target images. They form six image pairs.

established correspondences have subpixel accuracy and are regarded as ground-truth inliers. We compute the precision, RMSE, mean error, and maximum error of each method based on the selected 100 correspondences. The number of inliers of each method is then calculated by multiplying precision and the total number of detected matches. In all experiments, the initial matches are obtained by the SIFT algorithm with an NNDR ratio of 0.85, implemented in OPENCV. We fix $t = 8$, $\tau = 0.35$, $r' = \max(w, h)/10$, $\delta = 0.04$, $\eta = 3$, and $\varepsilon = 3$ for the following experiments.

4.2. Rigid feature matching

Qualitative comparison: We first evaluate the proposed method on several typical image pairs in data set 1, including image pairs 1, 9, and 14. Image pair 1 suffers from high land-use changes; image pair 9 is formed by images that were recorded by different sensors with a ten-year interval; and the overlapping region of image pair 14 is extremely small, i.e., less than 5% of the image size. The matching problem on these image pairs, thus, becomes quite challenging because of the illumination, viewpoint, scale (GSD), and temporal differences. The initial inlier ratios of these three image pairs are 18.85%, 2.06%, and 7.47%, respectively. The results are shown in Figs. 8–10.

From the figures, VFC, LLT, and RANSAC achieve good results on image pairs (such as image pairs 1) with relatively high inlier ratios, but poor results on image pairs (such as image pair 9) with low inlier ratios. In addition, the correspondences detected by these methods still contain many low-precision matches (noise). This is expected, since VFC is a non-parametric method that does not use an accurate geometric model for noise removal, and both LLT and RANSAC solve the registration transformation by closed-form solutions that are sensitive to noise. USAC achieves even better precision accuracy than our method on image pairs 1. However, it is very sensitive to the inlier ratios of initial matches, as indicated by its complete failure on image pairs 9 and 14. PGM + RRWM performs poorly; it obtains results that are even worse than the initial matches on some image pairs, such as image pairs 1. It is not suitable for high-precision matching problems. ACC is not sensitive to inlier ratios. It achieves similar precision accuracy to our method on image pairs 9 and 14. However, it obtains low precision on image pair 1. In contrast, the proposed method achieves very impressive performances on all three challenging image pairs. Only several matches of the 100 displayed correspondences for each image pair are low-precision noise. Our method performs support-line voting to filter most of the outliers. This strategy is a photometric constraint and is not sensitive to low inlier ratio. In addition, the affine-invariant ratios stage refines the matches and evaluates each match by its local affine residual. Thus, the proposed method is more robust to low inlier ratio and noise than the compared methods.

Quantitative comparison: Fig. 11 shows the quantitative comparisons on data set 1, where Fig. 11(a–e) plot the precision accuracy, number of inliers, RMSE, mean error, and maximum error, respectively. RMSE and mean error larger than 10 pixels are shown as 10 pixels in Fig. 11(c) and (d), and maximum error larger than 20 pixels is shown as 20 pixels in Fig. 11(e). This dataset is very challenging because some initial inlier ratios are very low (8 of 15 image pairs have initial inlier ratios that are lower than 10%). The average initial inlier ratio of this dataset is 16.94% and the average initial number of correct matches is 50.

From Fig. 11(a), we can see that RANSAC and USAC achieve quite satisfying precisions on image pairs with high initial inlier ratios, e.g., image pairs 1–5. However, their performances decrease significantly as the initial outlier ratio increases. For example, RANSAC fails on image pairs 6, 7, and 9, and USAC fails on image pairs 6–12, and 14. The precision of LLT is less satisfied, and it may even fail on image pairs with relatively high inlier ratios, such as image pair 3. VFC obtains similar performance to LLT. Both are sensitive to low-precision noisy correspondences; thus, their precisions under 3 pixels are not very high. In most cases, PGM + RRWM performs poorly and its results are only superior to the initial matches. We find that if we regard matches with residuals of less than 15 pixels as ground-truth inliers, i.e., $\lambda = 15$, the average precision of PGM + RRWM improves from 38.35% to 72.28%. Thus, in the results of PGM + RRWM, there are many low-precision noisy matches. ACC performs much better than PGM + RRWM. This method seems to be insensitive to inlier ratios. The method achieves good results on some image pairs with low inlier ratios (such as image pairs 7–11), but poor results on some with high inlier ratios (such as image pairs 1–5). In contrast, our method achieves the best precision accuracy on most of the image pairs, especially for image pairs with low initial inlier ratios, such as image pairs 6, 7, 9, and 12. The average precision and number of inliers of each method are reported in Table 4. As shown, the proposed method obtains 24.16% higher precision accuracy compared with RANSAC, which ranks second place among all methods. Benefitting from the expansion stage, our method finds as many high-precision correspondences as possible, which can be easily observed from Fig. 11(b). For example, the initial number of inliers of image pair 10 is 20, while the proposed method extracts 220 high-precision matches. The number of inliers of the proposed method is about four times the number of initial matches. Fig. 11

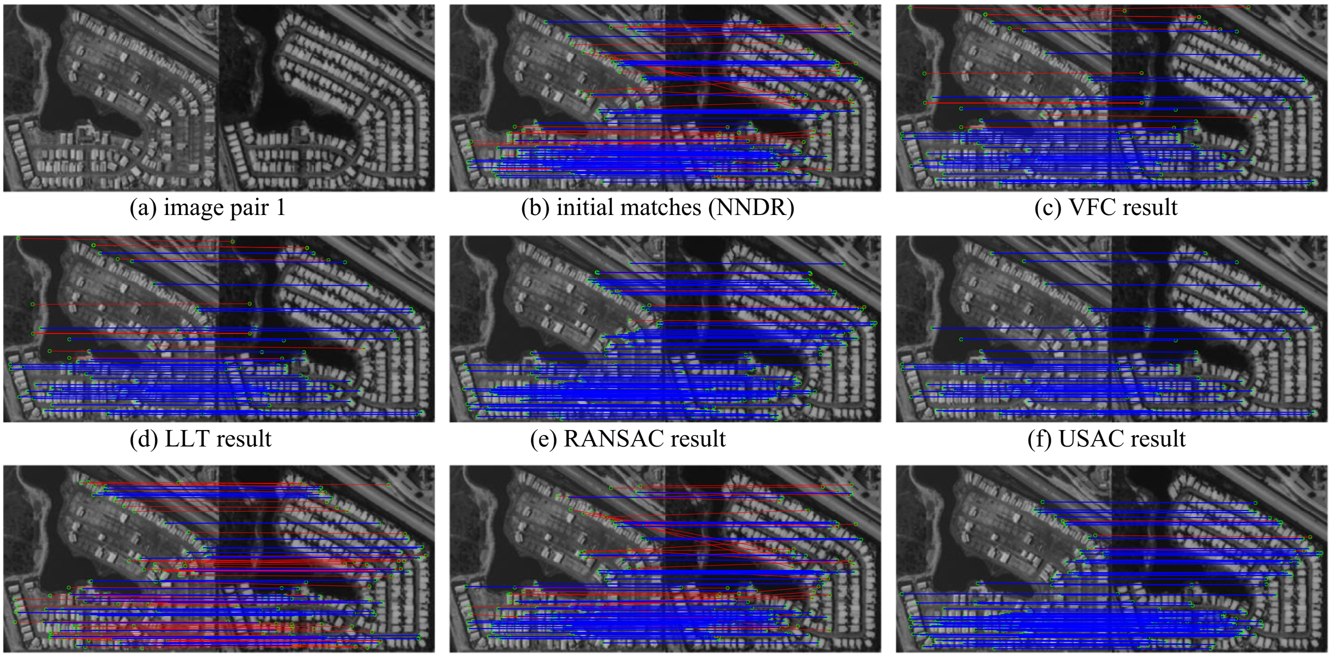
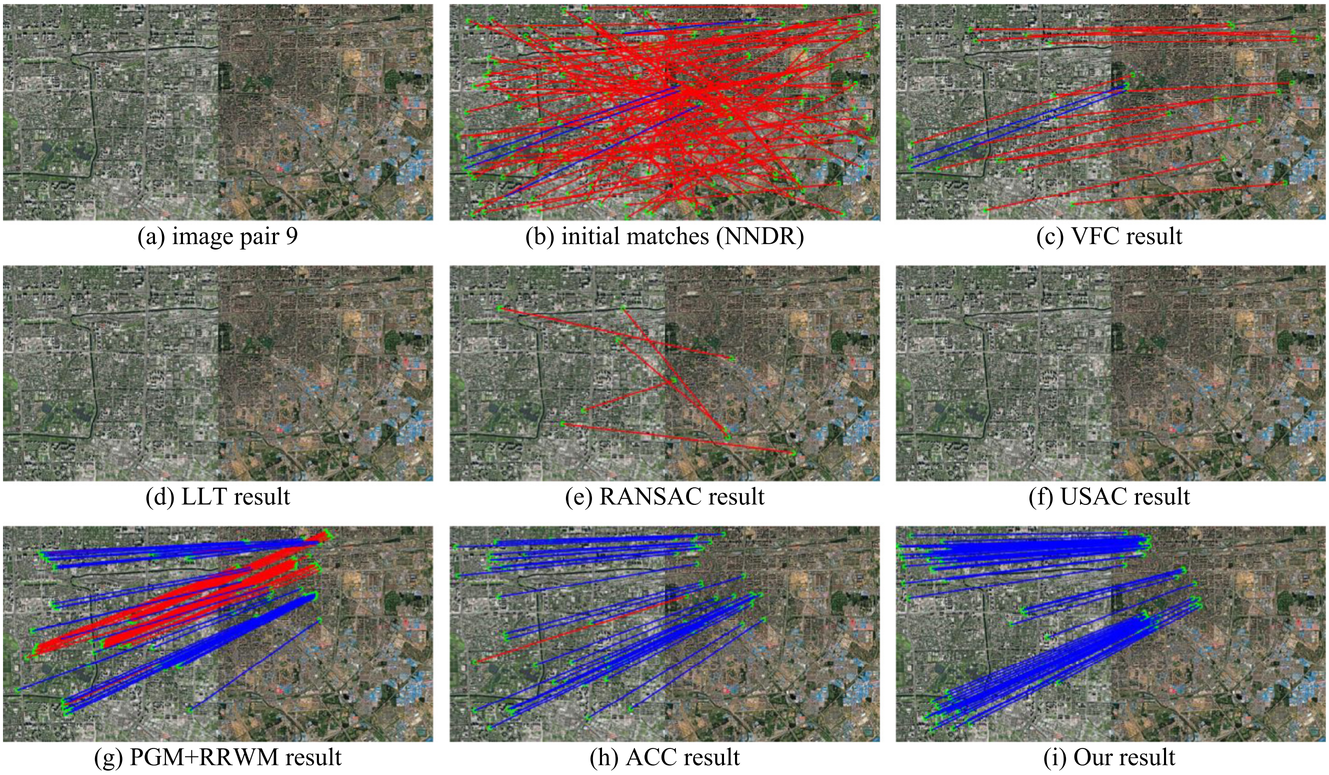
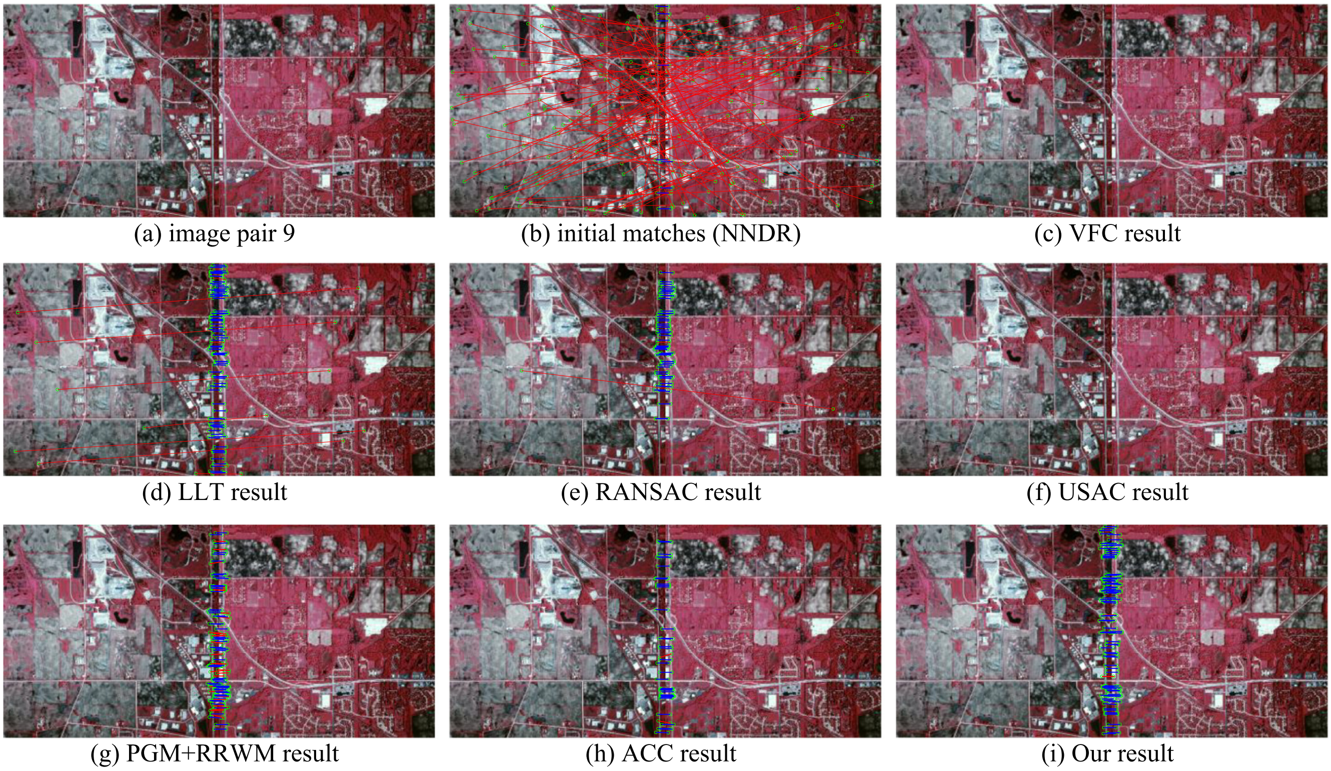

(a) image pair 1
(b) initial matches (NNDR)
(c) VFC result
(d) LLT result
(e) RANSAC result
(f) USAC result
(g) PGM+RRWM result
(h) ACC result
(i) Our result

Fig. 8. Results on image pair 1 of data set 1. Green dots represent feature points, red lines represent false matches, and blue lines represent correct matches. For better visualization, no more than 100 randomly selected matches are presented. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



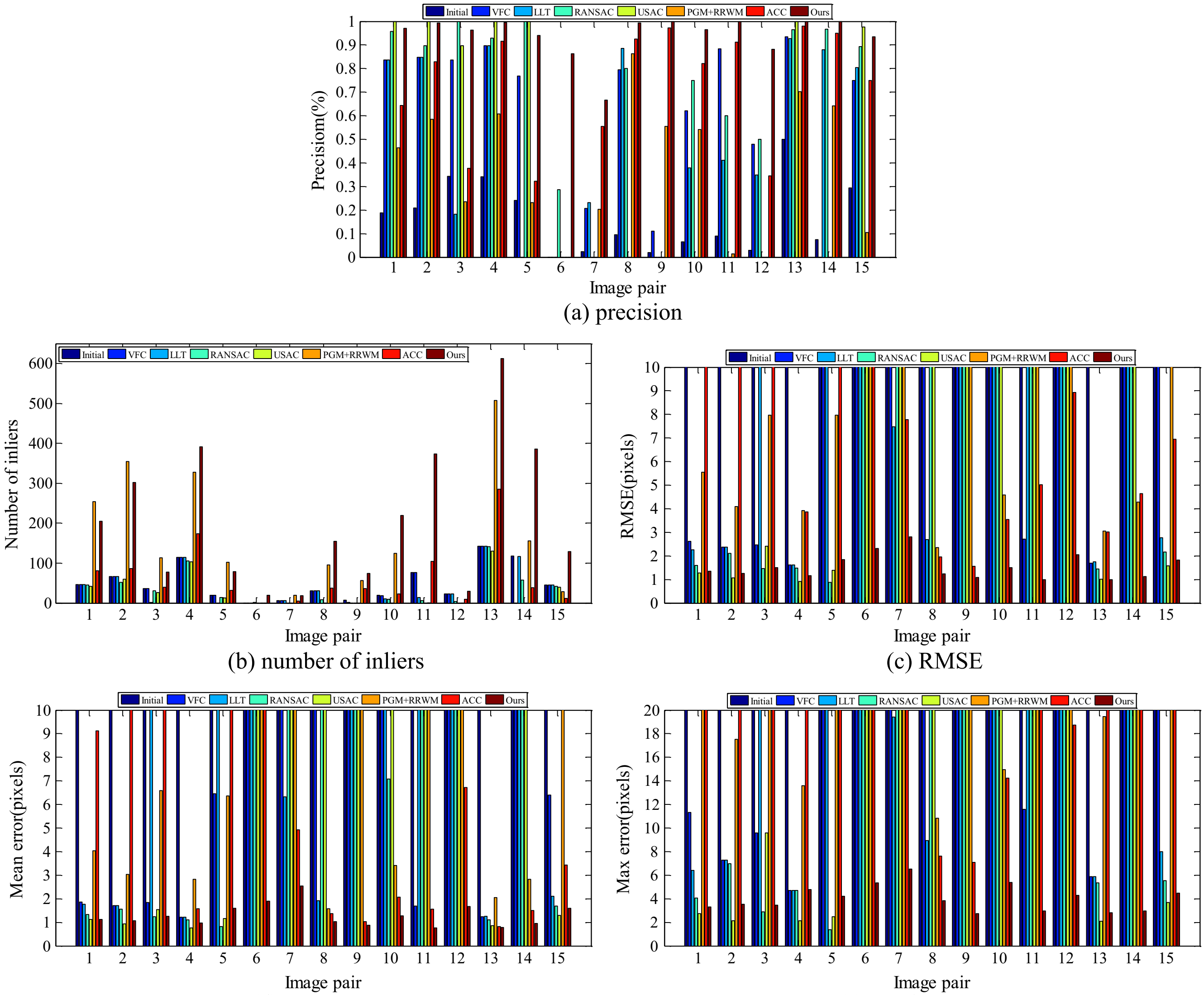
(a) image pair 9
(b) initial matches (NNDR)
(c) VFC result
(d) LLT result
(e) RANSAC result
(f) USAC result
(g) PGM+RRWM result
(h) ACC result
(i) Our result

Fig. 9. Results on image pair 9 of data set 1. Green dots represent feature points, red lines represent false matches, and blue lines represent correct matches. For better visualization, no more than 100 randomly selected matches are presented.



(a) image pair 9
(b) initial matches (NNDR)
(c) VFC result
(d) LLT result
(e) RANSAC result
(f) USAC result
(g) PGM+RRWM result
(h) ACC result
(i) Our result

Fig. 10. Results on image pair 14 of data set 1. Green dots represent feature points, red lines represent false matches, and blue lines represent correct matches. For better visualization, no more than 100 randomly selected matches are presented. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



(a) precision

(b) number of inliers

(c) RMSE

(d) mean error

(e) maximum error

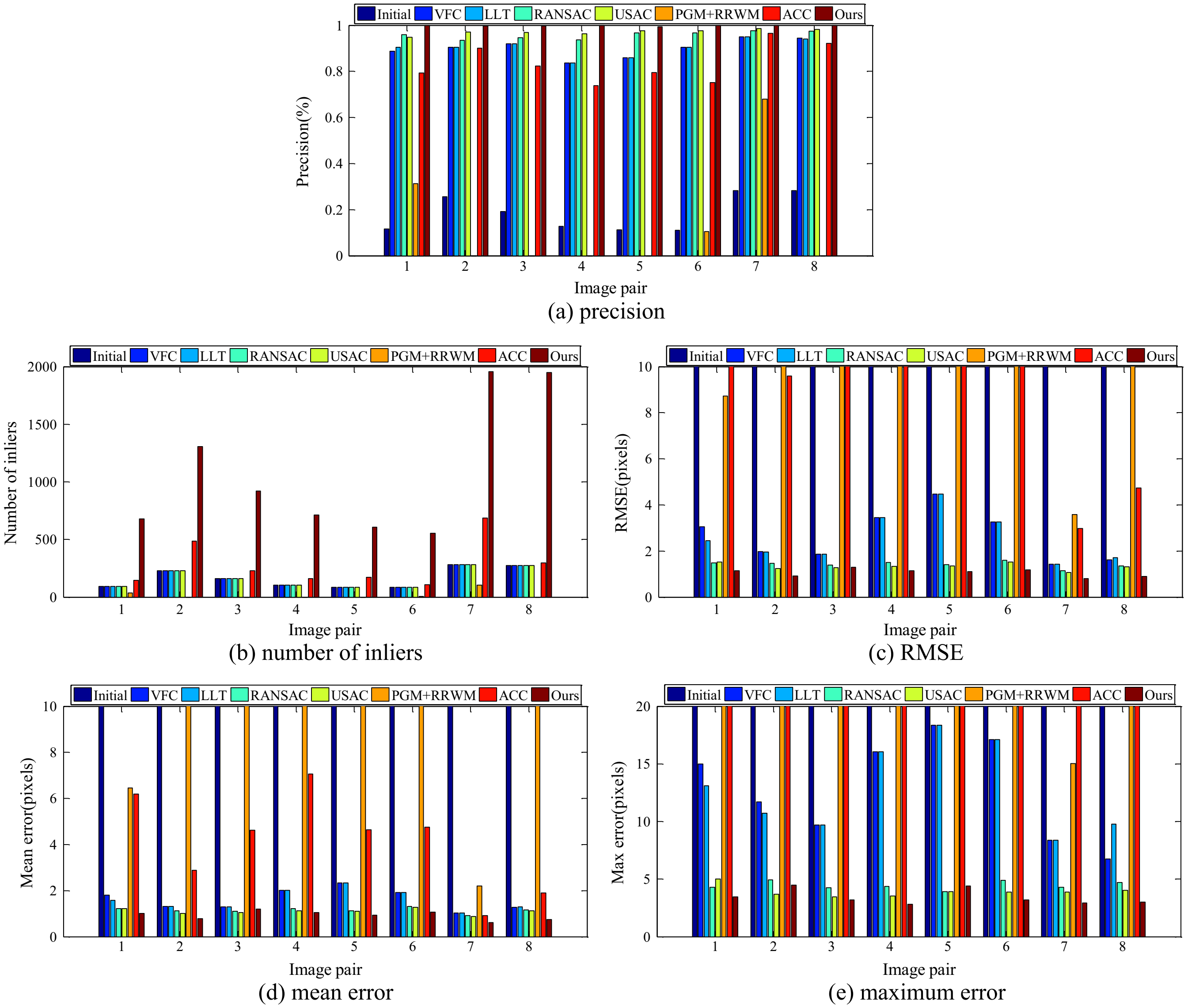
Fig. 11. Quantitative comparisons on data set 1.

Table 4
Average precision and number of inliers on data set 1.

Metric	NNDR	VFC	LLT	RANSAC	USAC	PGM + RRWM	ACC	OURS
Precision/%	16.76	59.79	50.88	70.3	45.81	38.35	68.66	94.46
Number of inliers	50	42	41	34	28	143	64.27	205

(c–e) indicate similar results to Fig. 11(a). RANSAC and USAC achieve small RMSEs, mean errors, and maximum errors, while the performances of VFC and LLT are less satisfactory on image pairs 1–5. As mentioned earlier, VFC is a non-parametric method that does not utilize residual errors to distinguish high-precision correspondences from relatively low-precision correspondences. LLT is solved in closed form and the estimated transformation may be skewed by noise. The performances of all four methods drop rapidly as the initial inlier rate decreases, especially that of USAC. ACC performs better than these four methods on image pairs 7–15, but worse on image pairs 1–5. In contrast, our method is very robust and accurate. The average RMSE, mean error, and maximum error of our method are 1.54 pixels, 1.29 pixels, and 4.05 pixels, respectively. From the maximum error metric, we know that the results of the proposed method contain almost no gross errors. In other words, the outliers of our method are correspondences with relatively low precision.

Fig. 12 shows the quantitative comparisons on data set 2, including precision accuracy, number of inliers, RMSE, mean error, and maximum error comparisons. This simulated dataset is not very challenging because it only suffers from affine transformation and intensity differences. There is no local geometrical distortion in this dataset. Thus, the image pairs in data set 2 can be perfectly aligned using the correct affine transformation. The average initial inlier ratio of this dataset is 18.51%, and the average initial number of correct matches is 164. Every method achieves remarkable performance on this dataset, except PGM + RRWM, since PGM + RRWM is very sensitive to non-isotropic scaling inside affine



(a) precision

(b) number of inliers

(c) RMSE

(d) mean error

(e) maximum error

Fig. 12. quantitative comparisons on data set 2.

Table 5
Average precision and number of inliers on data set 2.

Metric	NNDR	VFC	LLT	RANSAC	USAC	PGM + RRWM	ACC	OURS
Precision/%	18.51	90.08	90.26	95.74	97.16	13.73	83.62	99.81
Number of inliers	164	163	163	164	163	18	286	1086

transformations (image pairs 1–6 of data set 2) and large rotations (image pair 7 of data set 2). The precisions of RANSAC and USAC are slightly lower than those of our method. VFC and LLT cannot effectively distinguish noisy matches (relatively low-precision matches) from inliers. The maximum errors of VFC and LLT are generally larger than 8 pixels. ACC performs considerably better on image pairs 7–8 than on image pairs 1–6. Thus, ACC may be sensitive to non-isotropic scaling. The average precision and number of inliers of each method are reported in Table 5. Our method obtains 2.65% higher precision accuracy compared with the second-best method, USAC. The average number of inliers of our method is 1086, which is approximately 6.5 times the number of initial matches. The average RMSEs of USAC and our method are 1.33 pixels and 1.06 pixels, respectively; their average mean errors are 1.1 pixels and 0.93 pixels, respectively; and their average maximum errors are 3.93 pixels and 3.44 pixels, respectively. The proposed method achieves the best RMSE accuracy because most of the relatively low-precision correspondences are discarded by the affine-invariant ratios constraint.

4.3. Non-rigid feature matching

We also evaluate the proposed method on the first two non-rigid image pairs of data set 3. Both image pairs suffer from serious local geometric distortions. The initial inlier rates are 44% and 45% for image pair 1 and image pair 2, respectively, and the initial numbers of inliers are 799 and 404, respectively. The qualitative result of image pair 1 are reported in Fig. 13.

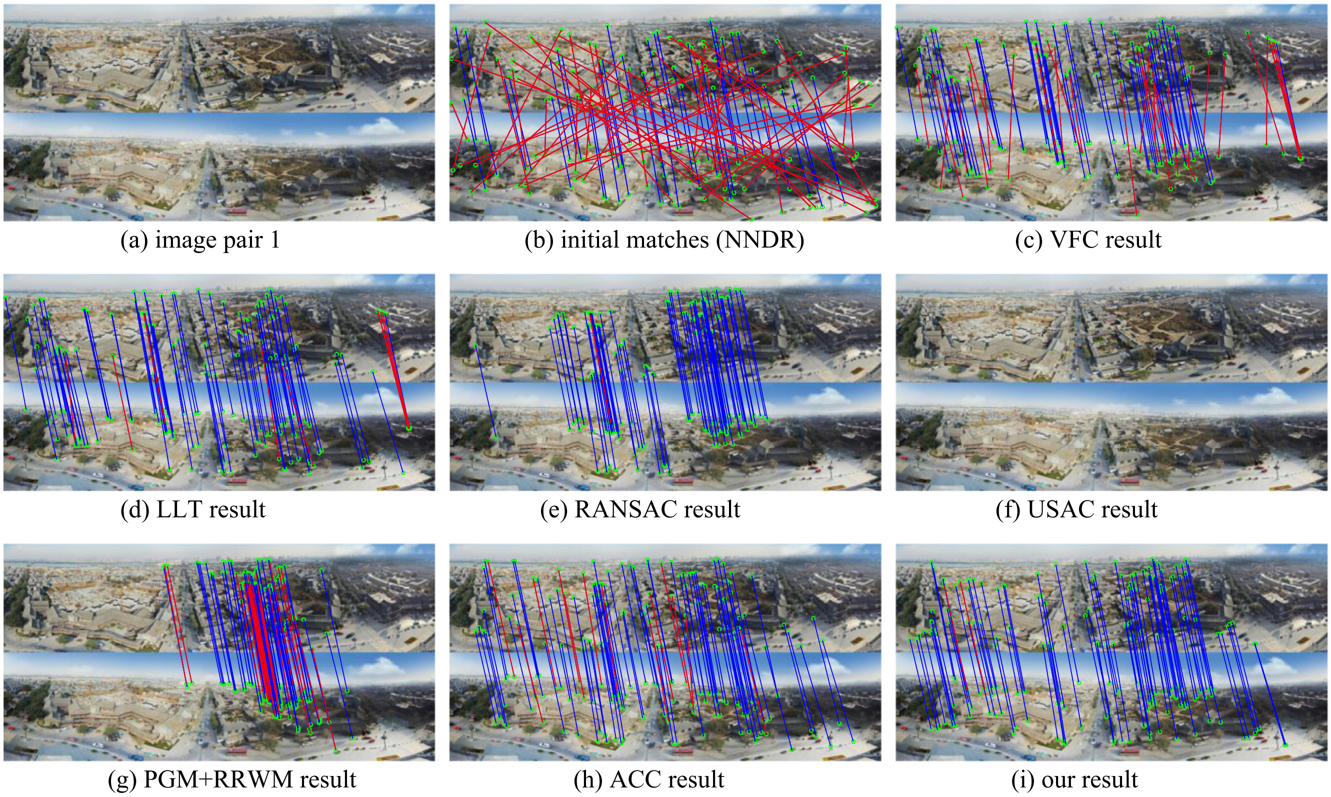


Fig. 13. Results on image pair 1 of data set 3. Green dots represent feature points, red lines represent false matches, and blue lines represent correct matches. For better visualization, no more than 100 randomly selected matches are presented.

Table 6
Precision and number of inliers results on data set 3.

Image pair	Metric	NNDR	VFC	LLT	RANSAC	USAC	PGM + RRWM	ACC	OURS
1	Precision/%	44	70	91	**98**	0	79	87	**98**
	Number of inliers	799	664	788	338	0	254	463	**1672**
2	Precision/%	45	68	87	93	0	57	76	**100**
	Number of inliers	404	361	299	208	0	113	192	**469**

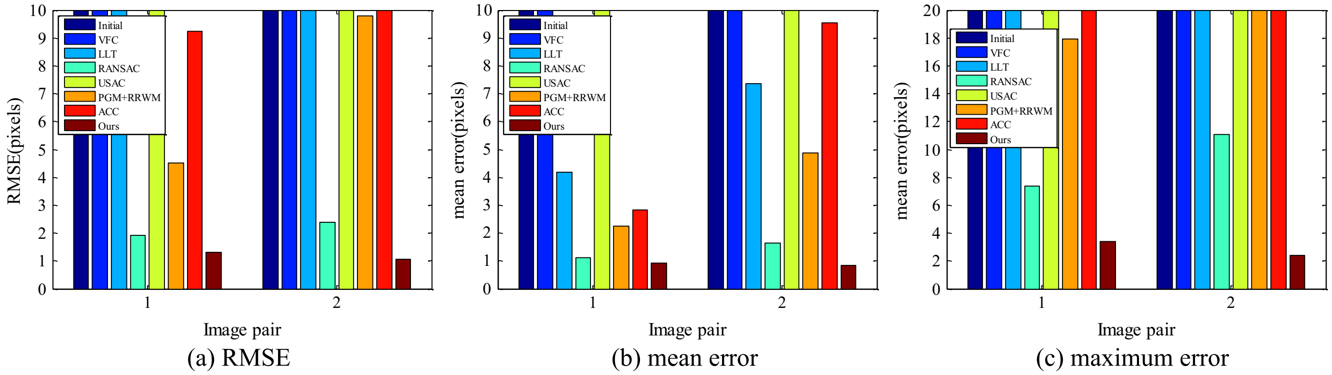


Fig. 14. RMSE, mean error, and maximum error results on data set 3.

According to the results, USAC fails to detect any matches on image pair 1; both PGM + RRWM and VFC preserve many outliers; and LLT and ACC are better than VFC. However, the results of these methods are less satisfying than those of RANSAC and our method. Both RANSAC and our method achieve very impressive accuracy performances. However, RANSAC only detects matches that satisfy its estimated model. Many correct matches are identified as outliers by RANSAC due to the local geometric distortions. As a result, the preserved matches are not evenly distributed in the overlapping region as our method. The precision and number of inliers of each method are reported in Table 6. On image pair 2, our method obtains 7% higher precision accuracy compared with RANSAC. RANSAC only detects approximately 50% of the correct correspondences in the initial matches, while the proposed method

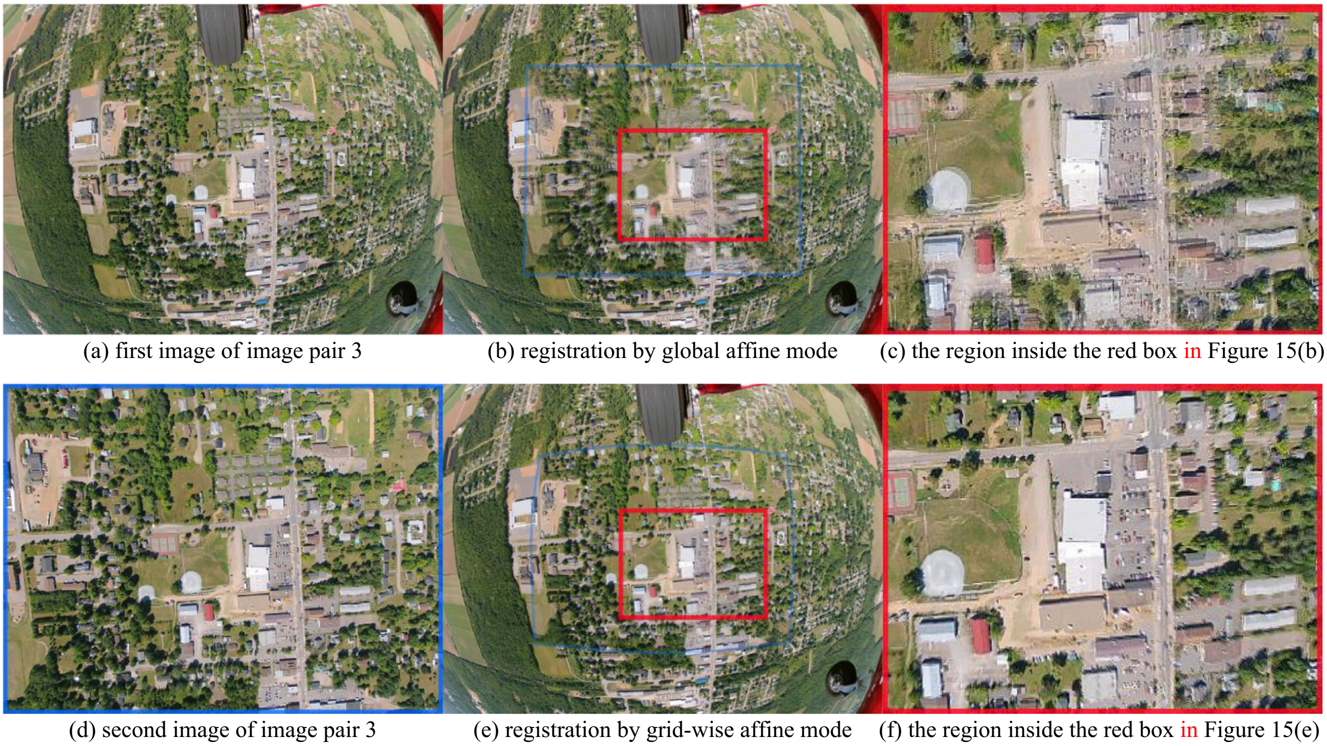

(a) first image of image pair 3
(b) registration by global affine mode
(c) the region inside the red box in Figure 15(b)
(d) second image of image pair 3
(e) registration by grid-wise affine mode
(f) the region inside the red box in Figure 15(e)

Fig. 15. Image registration comparison on image pair 3 of data set 3. The misalignment artifacts can be significantly reduced by using the grid-wise affine model instead of the global affine model.

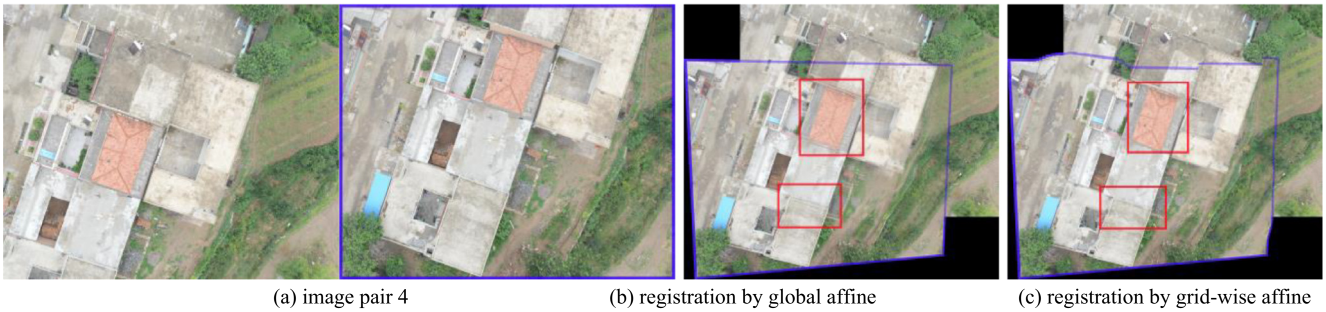

(a) image pair 4
(b) registration by global affine
(c) registration by grid-wise affine

Fig. 16. Image registration comparison on image pair 4 of data set 3. The misalignment artifacts can be significantly reduced by using the grid-wise affine model instead of the global affine model.

extracts many good matches in addition to the initial matches. The average number of inliers identified by our method is 1071, which is approximately 1.8 times the number of inliers in the initial matches. The RMSE, mean error, and maximum error of each method are shown in Fig. 14. Again, RANSAC and our method achieve better performances than the other methods.

4.4. Grid-wise affine registration

In our method, we also calculate a grid-wise affine model for image registration. We compare this model with the global affine model to demonstrate its superiority on a non-rigid image pair and a UAV image pair, i.e., image pair 3 and image pair 4 of data set 3. We transform the second image into the first one according to the transformation models for each image pair. No post-processing stage is applied to eliminate the ghosting phenomenon. Thus, the registered image with less severe ghosting phenomenon is better. The results are shown in Figs. 15 and 16.

Image pair 3 consists of a wide-angle UAV image and a pinhole UAV image. The wide-angle image suffers from serious radial geometric distortions, i.e., the distortions increase with the distance from the image center. The geometric relationship of this image pair cannot be well modeled by a global affine transformation because of the local geometric distortions. As shown in Fig. 15 (b) and (c), the ghosting phenomenon is serious inside the red box region. For example, the edges of buildings and trees are doubled and the image is seriously blurred. In contrast, the registration result (Fig. 15(e) and Fig. 15(f)) obtained with the grid-wise affine model is much better. The edges of buildings and trees are clear.

Image pair 4 consists of two low-altitude UAV images, where the effect caused by elevation differences should not be ignored. The building roofs and the grounds in image pair 4 should satisfy different affine models. Thus, the misalignment artifacts are serious in Fig. 16(c), which registers the buildings and grounds using the same transformation model. For instance, the edges inside the red box of Fig. 16(c) are doubled and the grass is blurred. In contrast, the grid-wise affine model can significantly reduce the misalignment artifacts because it assigns different best local affine models with building roofs and grounds.

5. Conclusions

In this paper, we proposed a novel robust feature matching method based on support-line voting and affine-invariant ratios for remote sensing image matching and registration. The support-line voting technique can eliminate most of the outliers and is not sensitive to low initial inlier ratios since it utilizes a robust photometric constraint. In this stage, we introduce adaptive binning histograms for support-line descriptor construction that are more robust to local geometric distortions. We observe that affine-invariant ratios are generally satisfied by remote sensing images, including both rigid and non-rigid images, and we use them as geometric constraints to refine the matching results. Each correspondence in this stage is classified as an inlier or outlier multiple times, which further improve the robustness of the proposed method. We also use the estimated local affine transformations to extract as many high-precision correspondences as possible and establish a grid-wise affine model for image registration. We show that the proposed method significantly outperforms the six compared state-of-the-art methods, i.e., VFC, LLT, RANSAC, USAC, PGM + RRWM, and ACC, on several rigid and non-rigid data sets.

Acknowledgments

The authors would like to express their gratitude to the editors and the reviewers for their constructive and helpful comments for substantial improvement of this paper. This work was supported by the National Natural Science Foundation of China (NSFC) (41271452 and 41701528), the Fundamental Research Funds for the Central Universities (2042017KF0235), and the Key Technologies R&D Program of China (2015BAK03B04).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.isprsjprs.2017.08.009.

References

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-up robust features (SURF). Comput. Vis. Image Underst. 110, 346–359.

Chen, H.-M., Varshney, P.K., Arora, M.K., 2003. Performance of mutual information similarity measure for registration of multitemporal remote sensing images. IEEE Trans. Geosci. Remote Sens. 41, 2445–2454.

Chen, H.-Y., Lin, Y.-Y., Chen, B.-Y., 2013. Robust feature matching with alternate hough and inverted hough transforms. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2762–2769.

Chen, Q.-S., Defrise, M., Deconinck, F., 1994. Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition. IEEE Trans. Pattern Anal. Mach. Intell. 16, 1156–1168.

Chin, T.-J., Suter, D., 2017. The maximum consensus problem: recent algorithmic advances. Synth. Lect. Comput. Vis. 7, 1–194.

Cho, M., Lee, J., Lee, K., 2010. Reweighted random walks for graph matching. Comput. Vision–ECCV 2010, 492–505.

Cho, M., Lee, J., Lee, K.M., 2009. Feature correspondence and deformable object matching via agglomerative correspondence clustering. In: IEEE 12th International Conference on Computer vision, 2009. IEEE, pp. 1280–1287.

Cho, M., Lee, K.M., 2012. Progressive graph matching: Making a move of graphs via probabilistic voting. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012. IEEE, pp. 398–405.

Cho, M., Sun, J., Duchenne, O., Ponce, J., 2014. Finding matches in a haystack: a max-pooling strategy for graph matching in the presence of outliers. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2083–2090.

Chum, O., Matas, J., 2005. Matching with PROSAC-progressive sample consensus. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, pp. 220–226.

Chum, O., Matas, J., Kittler, J., 2003. Locally optimized RANSAC. In: Joint Pattern Recognition Symposium. Springer, pp. 236–243.

Conte, D., Foggia, P., Sansone, C., Vento, M., 2004. Thirty years of graph matching in pattern recognition. Int. J. Pattern Recognit Artif Intell. 18, 265–298.

Dawn, S., Saxena, V., Sharma, B., 2010. Remote sensing image registration techniques: a survey. In: International Conference on Image and Signal Processing. Springer, pp. 103–112.

Dellinger, F., Delon, J., Gousseau, Y., Michel, J., Tupin, F., 2015. Sar-sift: a sift-like algorithm for sar images. IEEE Trans. Geosci. Remote Sens. 53, 453–466.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc. Ser. B (Methodol.), 1–38.

Duchenne, O., Bach, F., Kweon, I.-S., Ponce, J., 2011. A tensor-based algorithm for high-order graph matching. IEEE Trans. Pattern Anal. Mach. Intell. 33, 2383–2395.

Egels, Y., Kasser, M., 2003. Digital Photogrammetry. CRC Press.

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24, 381–395.

Foroosh, H., Zerubia, J.B., Berthod, M., 2002. Extension of phase correlation to subpixel registration. IEEE Trans. Image Process. 11, 188–200.

Haala, N., Kada, M., 2010. An update on automatic 3D building reconstruction. ISPRS J. Photogram. Remote Sens. 65, 570–580.

Hartley, R., Zisserman, A., 2003. Multiple View Geometry in Computer Vision. Cambridge University Press.

Jian, B., Vemuri, B.C., 2011. Robust point set registration using gaussian mixture models. IEEE Trans. Pattern Anal. Mach. Intell. 33, 1633–1645.

Ke, Y., Sukthankar, R., 2004. PCA-SIFT: a more distinctive representation for local image descriptors. Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, vol. 502. IEEE, pp. II-506–II-513.

Leordeanu, M., Hebert, M., Sukthankar, R., 2009. An integer projected fixed point method for graph matching and map inference. Adv. Neural. Inf. Process. Syst., 1114–1122

Li, J., Hu, Q., Ai, M., 2016. Robust feature matching for remote sensing image registration based on $ L_ q $-estimator. IEEE Geosci. Remote Sens. Lett. 13, 1989–1993.

Li, X., Hu, Z., 2010. Rejecting mismatches by correspondence function. Int. J. Comput. Vision 89, 1–17.

Lian, W., Zhang, L., Yang, M.-H., 2016. An efficient globally optimal algorithm for asymmetric point matching. IEEE Trans. Pattern Anal. Mach. Intell.

Lian, W., Zhang, L., Zhang, D., 2012. Rotation-invariant nonrigid point set matching in cluttered scenes. IEEE Trans. Image Process. 21, 2786–2797.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60, 91–110.

Ma, J., Qiu, W., Zhao, J., Ma, Y., Yuille, A.L., Tu, Z., 2015a. Robust L2E estimation of transformation for non-rigid registration. IEEE Trans. Sig. Proc. 63, 1115–1129.

Ma, J., Zhao, J., Tian, J., Yuille, A.L., Tu, Z., 2014. Robust point matching via vector field consensus. IEEE Trans. Image Process. 23, 1706–1721.

Ma, J., Zhou, H., Zhao, J., Gao, Y., Jiang, J., Tian, J., 2015b. Robust feature matching for remote sensing image registration via locally linear transforming. IEEE Trans. Geosci. Remote Sens. 53, 6469–6481.

Raguram, R., Chum, O., Pollefeys, M., Matas, J., Frahm, J.-M., 2013. USAC: a universal framework for random sample consensus. IEEE Trans. Pattern Anal. Mach. Intell. 35, 2022–2038.

Sedaghat, A., Ebadi, H., 2015. Remote sensing image matching based on adaptive binning SIFT descriptor. IEEE Trans. Geosci. Remote Sens. 53, 5283–5293.

Sedaghat, A., Mokhtarzade, M., Ebadi, H., 2011. Uniform robust scale-invariant feature matching for optical remote sensing images. IEEE Trans. Geosci. Remote Sens 49, 4516–4527.

Tolias, G., Avrithis, Y., 2011. Speeded-up, relaxed spatial matching. In: IEEE International Conference on Computer Vision (ICCV), 2011. IEEE, pp. 1653–1660.

Torr, P.H., Zisserman, A., 2000. MLESAC: a new robust estimator with application to estimating image geometry. Comput. Vis. Image Underst. 78, 138–156.

Torresani, L., Kolmogorov, V., Rother, C., 2008. Feature correspondence via graph matching: models and global optimization. In: European Conference on Computer Vision. Springer, pp. 596–609.

Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W., 1999. Bundle adjustment—a modern synthesis. In: International Workshop on Vision Algorithms. Springer, pp. 298–372.

Weinmann, M., Weinmann, M., Hinz, S., Jutzi, B., 2011. Fast and automatic image-based registration of TLS data. ISPRS J. Photogram. Rem. Sens. 66, S62–S70.

Zaragoza, J., Chin, T.-J., Brown, M.S., Suter, D., 2013. As-projective-as-possible image stitching with moving DLT. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2339–2346.

Zhao, M., An, B., Wu, Y., Lin, C., 2013. Bi-SOGC: a graph matching approach based on bilateral KNN spatial orders around geometric centers for remote sensing image registration. IEEE Geosci. Remote Sens. Lett. 10, 1429–1433.

Zhou, H., Ma, J., Yang, C., Sun, S., Liu, R., Zhao, J., 2016. Nonrigid feature matching for remote sensing images via probabilistic inference with global and local regularizations. IEEE Geosci. Remote Sens. Lett. 13, 374–378.

Zitova, B., Flusser, J., 2003. Image registration methods: a survey. Image Vis. Comput. 21, 977–1000.